Active DNA demethylation in the mononuclear phagocyte system

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.) der Fakultät für Biologie und Vorklinische Medizin der Universität Regensburg

vorgelegt von

Sandra Schmidhofer

aus

Neuburg an der Donau

im Jahr

2014

Das Promotionsgesuch wurde eingereicht am: 14.11.2014

Die Arbeit wurde angeleitet von: Prof. Dr. Michael Rehli

Prüfungsausschuss:

Vorsitzender:	Prof. Dr. Herbert Tschochner
Erstgutachter:	Prof. Dr. Michael Rehli
Zweitgutachter:	Prof. Dr. Gernot Längst
Drittprüfer:	Prof. Dr. Richard Warth
Ersatzprüfer:	PD Dr. Attila Németh

Table of contents

List of figures

List of tables

List of abbreviations

°C	degree Celsius
µg	microgram
µl	microliter
µM	micromolar
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5hmU	5-hydroxymethyluracil
5mC	5-methylcytosine
AB	antibody
AID	activation-induced cytidine deaminase
AM	active modification
APC	antigen-presenting cell
APOBEC	apolipoprotein B mRNA editing enzyme
APS	Ammoniumperoxidisulfat
AR	active removal
BER	base excision repair
BM	bone marrow
bp	base pair
BS	bisulfite sequencing
BSA	bovine serum albumin
C	cytosine
cDC	classical dendritic cell
cDNA	complementary DNA
CGI	CpG island
ChIP	Chromatin Immunoprecipitation
ChIP-seq	Chromatin Immunoprecipitation-sequencing
chr	chromosome
CMC	chromatin modifying complexes
CMML	chronic myelomonocytic leukemia
CXXC	Cysteine-X-X-Cysteine zinc-finger domain
Da	Dalton
DC	dendritic cell
ddH_2O	double-distilled water
DME	Demeter
DMR	differentially methylated region
DNA	desoxyribonucleic acid
DNMT1	DNA methyltransferase 1
dNTP	deoxyribonucleotide triphosphate
ds	double strand
E. coli	Escherichia coli
eRNA	enhancer RNA
ESC	embryonic stem cell

EtBr	Ethidium bromide
EtOH	Ethanol
$EtONH_2$	*O*-ethylhydroxylamine
ETS	E26 transformation-specific
eTSS	enhancer transcription start site
F	forward
FACS	flourescence-activated cell sorting
fCAB-seq	chemically assisted bisulfite sequencing of 5fC
FCS	fetal calf serum
FITC	flourescein isothyiocyanate
g	gram
GC	guanine/cytosine
gDNA	genomic DNA
GFE	GO term fold enrichment
GGSD	glycosylation of 5hmC and glycosyl-sensitive digest of DNA
GM-CSF	granulocyte macrophage colony stimulating factor
h	hour
H	histone
HAT	histone acetyltransferase
HDAC	histone deacetylase
hESC	human embryonic stem cell
hMeDIP	hydroxymethylated DNA Immunoprecipitation
IDAX	inhibition of the Dvl and axin complex
iDC	immature dendritic cell
IFN	type I interferon
IgG	Immunoglobulin G
IL	interleukin
IP	Immunoprecipitation
kb	kilobase
LB	lysogeny broth (Luria-Bertani)
lncRNA	long non-coding RNA
LPS	lipopolysaccharide
M	molar
MAC	macrophage
MBD	methyl-binding domain
MBD3/NuRD	methyl-CpG binding domain protein 3/ nucleosome remodeling deacetylase
me	methylation
MeCP2	methyl CpG-binding protein 2
MeDIP	methylated DNA Immunoprecipitation
mESC	murine embryonic stem cell
mg	milligram
min	minute
mio	million
MLL5	mixed lineage leukemia 5
MO	monocyte
moDC	monocyte-derived DCs

MO-DC	monocyte-derived dendritic cell
MPS	mononuclear phagocyte system
mRNA	messenger RNA
MSRE	DNA modification-sensitive restriction assay
ng	nanogram
NGS	next generation sequencing
NuRD	nucleosome remodeling deacetylase
O/N	over night
O-GlcNAc	O-linked β-D-N-acetylglucosamine
OGT	*O*-linked β-$_D$-*N*-acetylglucosamine transferase
p300/CBP	E1A binding protein p300/CREB-binding protein
PB-MNC	peripheral blood mononuclear cells
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PD	passive DNA demethylation mechanism
pDC	plasmatoid dendritic cell
PEST	peptide sequence rich in proline, glutamic acid, serine, threonine
PMA	phorbol 12-myristate 13-acetate
PU.1	PU box binding protein 1
qPCR	quantitative real-time PCR
qRT-PCR	quantitative reverse transcriptase real-time PCR
R	reverse
RNA	ribonucleic acid
RNAse	ribonuclease
ROS1	repressor of silencing 1
rpm	rounds per minute
RT	room temperature
SAM	S-Adenosyl methionine
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulfate
sec	second
Set1	set domain containing 1a
Set1/COMPASS	complex associated with Set1 complex
SIN3A	SIN3 transcription regulator family member A
siRNA	small interfering RNA
siRNA TF	transfection of siRNAs
SMUG1	single-strand-selective monofunctional uracil-DNA glycosylase 1
ss	single strand
Ta	annealing temperature
TDG	thymine DNA glycosylase
TEMED	N,N,N',N'-Tetramethylethylenediamine
TET	ten-eleven translocation
TF	transcription factor
Tip-DC	TNFα and iNOS expressing DC
Tm	melting temperature
Treg	regulatory T cell

TSS	transcription start site
U	unit
UCSC	University of California, Santa Cruz
UDPG	Uridine Diphosphoglucose
UHRF1	ubiquitin-like plant homeo-domain and RING finger domain 1
V	volt
vol	volume
W	watt

1 Summary

DNA demethylation processes have been studied for many years and entered the focus of extensive research with the discovery of active demethylation mechanisms (He *et al.*, 2011; Ito *et al.*, 2011; Iyer *et al.*, 2009; Kriaucionis and Heintz, 2009; Tahiliani *et al.*, 2009). These processes contribute to the regulation of cell type-specific gene expression patterns and the dynamics of other epigenetic mechanisms (Wu and Zhang, 2014). The investigation of the different types of mechanisms and their role in different cell types or developmental stages is an important challenge to understand the complex regulatory processes in mammals. The data presented in this work allowed further insights into the active demethylation processes and contributed to the understanding of regulatory mechanisms in different hematopoietic cell types. Using an *in vitro* model system, representing the human mononuclear phagocyte system, we were able to characterize the active DNA demethylation mechanism in the absence of passive demethylation events. The data revealed that the targeted, locus-specific active DNA demethylation process is initiated by the modification of 5mC to 5hmC. Further experiments based on the knockdown of candidate enzymes identified TET2 as the initiator of the active DNA demethylation process and as being responsible for the conversion of 5mC to 5hmC. Investigation of further possible players like TDG, MBD4, OGT, and HELLS gave first insights into a possible contribution to the process and so far the data indicated that none of the enzymes is involved in the first conversion step. Functional investigation of the demethylated regions in reporter gene assays linked the local binding of TFs like PU.1 and synchronous demethylation events to the activation of potential enhancer elements. The data demonstrated that their activation depended on the methylation level and that demethylation led to enhancer activation in a cell type-specific manner. Moreover the results indicated that the activation of cell type-specific enhancer elements requires a corresponding set of TF to open the regions, which may include the removal of 5mC in this process.

The validation and adaption of a 5hmC-enrichment method to next generation sequencing allowed us to investigate the active demethylation processes on a genome-wide level. Using the Hydroxymethyl Collector^{TM} kit we assessed the global dynamics of DNA demethylation and its association with the key hematopoietic transcription factor PU.1 in differentiating monocytes. The global screen illustrated

dynamic patterns of 5hmC and confirmed its role as an intermediate of active demethylation events accompanying the transition into another cell type. Local binding of PU.1 at demethylated sites further supported the theory of a correlation between demethylation events and the recruitment of PU.1. However, active DNA demethylation events were not altogether dependent on PU.1 binding, since several regions accumulated 5hmC in the absence of this TF, indicating the involvement of other factors and thus a site-specific recruitment of PU.1. The data further hinted at a possible regulatory role of 5hmC as an epigenetic mark, actively recruiting or passively impeding other factors. Further gene ontology analyses confirmed the immunological background of the cells and presented genes involved in the immune response and inflammation to be associated with active demethylation processes and the local appearance of PU.1. Corresponding expression changes suggested an involvement of PU.1 in the regulation of transcriptional changes during monocyte differentiation. However, regions with increasing or stable 5hmC levels displayed transcriptional changes independent of demethylation or PU.1 and supported the involvement of other factors in their regulation as well as possible regulatory functions of 5hmC.
A global screen of PU.1 distribution in differentiating monocytes illustrated dynamic PU.1 binding patterns upon the transition into another cell type and confirmed the association with demethylation events at subsets of PU.1 target regions. Comparing the 5hmC and PU.1 dynamics during monocyte differentiation we presented first evidence for a distinct chronology of PU.1 and demethylation events. In a subset of PU.1 target regions demethylation was present in monocytes but recruited PU.1 primarily on the transition into a new cell type. It is still unclear, if PU.1 generally profits from the opening of demethylated regions or if it administrates various functions at different target regions. The localization of the PU.1 patterns to active, cell type-specific regulatory elements revealed distinct distribution dynamics during cell differentiation. PU.1 is mainly targeted to promoter and promoter-distal regulatory regions that are activated in a cell type-specific manner. The active nature of the regions supported an involvement of PU.1 in hematopoietic cell differentiation. PU.1 binding was associated with marginally dynamic, demethylated states and indicated a role for PU.1 in the maintenance of an active transcriptional state or its independent recruitment to demethylated regulatory regions.

In summary the data presented in this work contributed to the understanding of the active DNA demethylation mechanism and revealed dynamic global association of demethylation events and PU.1 binding accompanying cell fate decisions in hematopoietic cells.

2 Introduction to epigenetics

In multicellular organisms embryonic development starts from one zygote and results in a complete organism with various lineages of distinct cell types. Although almost all of those cell types (with few exceptions) carry the same DNA sequence they adopt different functions, phenotypes and have diverse potential to mature into other cell types in response to a changing environment (Kohli and Zhang, 2013; Probst *et al.*, 2009). This diversity is created by changes to the epigenome, resulting in tissue-specific gene expression patterns, which are established in the early embryo, maintained during cell divisions and adjusted upon developmental or environmental changes. The epigenome is defined as heritable changes in genome activity that are independent of changes in the DNA sequence (Probst *et al.*, 2009; Russo, 1996) and represents a complex network of molecular mechanisms involved in genome regulation. Those epigenetic mechanisms mostly affect and reorganize the chromatin structure. Chromatin is a complex of DNA, histones and non-histone proteins which organizes DNA and regulates its accessibility (Margueron and Reinberg, 2010). Post-translational modifications of histone tails, like methylation or acetylation, function as major epigenetic marks. They facilitate a dynamic regulation of the chromatin compaction and help to recruit TFs and other regulatory proteins (Laugesen and Helin, 2014; Luger *et al.*, 1997). Chromatin regulators are mostly organized in large protein complexes and act as “readers”, “writers” or “erasers” of the histone marks. While the latter two add or remove modifications from the histone tails, “readers” are able to recognize and interpret the histone marks and thus recruit other factors to the chromatin (Badeaux and Shi, 2013). Additionally, variants of the core histones (H2A, H2B, H3, and H4) contribute to flexible regulation of gene expression. They take over histone functions, but tend to be less stable and thus easier to remove or reposition via chromatin remodeling (Jin and Felsenfeld, 2007; Weber *et al.*, 2014). Additionally histone variants keep genomic regions accessible and enable the binding of activating or repressive complexes and thus regulate transcription (Hu *et al.*, 2013a; Weber *et al.*, 2014). Long non-coding RNAs (lncRNA) can function as epigenetic modulators, too. Among many features they are able to target activating and repressive chromatin modifying complexes (CMC) to the genome (Yang *et al.*, 2014; Zhao *et al.*, 2010) and regulate the DNA-binding activity of interacting TFs (Vance and Ponting, 2014). They

play an important role for organization and modification of chromatin structure and regulate gene expression programs (Vance and Ponting, 2014). The DNA itself carries another epigenetic mark, the 5-methylcytosine (5mC). This mark plays a central role in epigenetic genome regulation and is described more detailed in the following chapter.

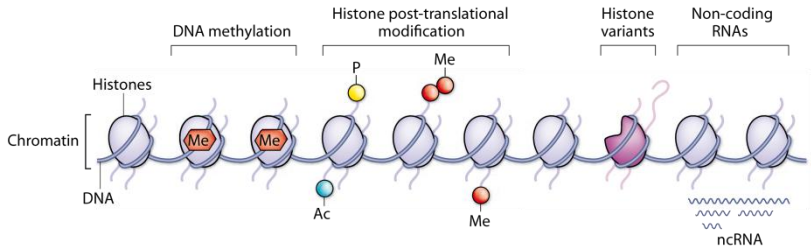

Figure 2.1: Epigenetic mechanisms of genome regulation
Schematic overview of major mechanisms regulating the epigenome (Modified from (Dulac, 2010))

2.1 DNA methylation

Methylation of the fifth position of cytosine is a key epigenetic modification and highly conserved in most plants, animals and fungi (Law and Jacobsen, 2010). In mammals DNA methylation occurs mainly at palindromic CG dinucleotide sequences (CpG) and is catalyzed by the *de novo* DNA methyltransferases DNMT3A and DNMT3B (Bird, 2002). Other non-CpG DNA methylations like CpC, CpT and CpA have been reported in plants where they occur frequently (Law and Jacobsen, 2010). In mammals those events are rather rare and, so far, restricted to oocytes, embryonic stem cells (ESC) and brain cells (Lister *et al.*, 2013; Lister *et al.*, 2009; Shirane *et al.*, 2013; Xie *et al.*, 2012). Although the overall abundance of CpGs is low in mammalian genomes (~28 million in human), CpG methylation is distributed all over the genome and covers 60-80% of all CpGs (Smith and Meissner, 2013). However, a small fraction of CpGs (<10%) is clustered together in CpG islands (CGI), which remain largely unmethylated (Wu and Zhang, 2014). 5mC is a heritable epigenetic mark and maintained through the cell cycle by the maintenance DNA methyltransferase 1 (DNMT1). The ubiquitin-like plant homeo-domain and RING finger domain 1 (UHRF1) recruits DNMT1 to hemi-methylated DNA where it transfers a methyl group to the unmethylated cytosine (Bostick *et al.*, 2007; Sharif *et al.*, 2007). The stable inheritance of 5mC plays an important role in genomic imprinting, where the methylation pattern of the parental alleles are passed on to the daughter cells, and additionally in silencing of transposable elements, in order to maintain genome stability (Bird, 2002; Law and Jacobsen, 2010). DNA methylation is primarily associated with the suppression of gene expression and

exhibits its repressing effects by blocking the binding of transcription activators or by the recruitment of transcriptional repressors (Tate and Bird, 1993). Thus although some transcription factors favor methylated binding sites, many others, that activate gene expression upon binding to their binding motifs, fail to bind when their consensus binding sequences are methylated (Campanero *et al.*, 2000; Rozenberg *et al.*, 2008; Sunahori *et al.*, 2009). Other proteins recognize and bind the 5mC mark. Among those readers are several members of the methyl-CpG-binding domain (MBD) family of proteins, which aid in transcription repression (Hendrich and Bird, 1998). They selectively bind 5mC and recruit repressive complexes with chromatin modifying components like histone deacetylases (HDAC) and histone methyltransferases (Robertson and Wolffe, 2000; Sarraf and Stancheva, 2004; Spruijt and Vermeulen, 2014)

A correct regulation of DNA methylation is an important feature for the prevention of malignant diseases. Many types of cancer show aberrant DNA methylation patterns and acquire a genome-wide hypomethylated state. The global loss of 5mC affects oncogenes, which are originally methylated and thus silenced, but become activated during cancer progression. Other genes, so called tumor suppressor genes, are generally located at non-methylated CGI and associated with tumor repression. Those regions are targets for aberrant DNA methylation, which occurs frequently in and contributes to cancer progression (Baylin and Jones, 2011; Jones and Baylin, 2002, 2007). These mechanisms of 5mC-mediated transcriptional repression are crucial for the cell and add to a complex repertoire of repressive mechanisms to ensure correct development from zygote to the adult organism and to prevent tumorigenesis.

2.2 DNA demethylation

DNA methylation regulates gene expression in different ways leading to a silenced and repressed environment. To overcome and reverse repressive regulation and to open and activate silenced genes, DNA methylation can be removed in a process called DNA demethylation. In the past few years several distinct pathways have been discovered which can be global or restricted to specific loci and play important roles in different stages of cell differentiation. One distinctive feature of demethylation mechanisms is the dependence on cell proliferation and thus replication of DNA. In the passive DNA demethylation mechanism (PD), DNMT1 and UHRF1 are absent or impeded by modified 5mC and fail to methylate the nascent strand, resulting in the loss

of DNA methylation after a few rounds of replication. This mechanism plays an important role in zygote development where PD mainly, but not exclusively, contributes to demethylation processes in the paternal, and also in the maternal genome (Guo *et al.*, 2014; Shen *et al.*, 2014).

In order to respond quickly to environmental and intrinsic signals, the cell uses active DNA demethylation mechanisms, which involve enzymatic activity and are in most cases independent of DNA replication (Kohli and Zhang, 2013). Active DNA demethylation mechanisms have been discovered in both plants and animals. The direct excision of 5mC by DNA glycosylases is mainly found in flowering plants, where DME (Demeter) and ROS1 (repressor of silencing 1) in combination with the BER (base excision repair) excise and remove 5mC (Zhu, 2009). In mammals, no orthologues of DME and ROS1 exist, but several studies support the idea that modified derivatives of 5mC are removed via BER in animals (Wu and Zhang, 2010). In zebrafish AID/APOBEC (activation-induced cytidine deaminase/ apolipoprotein B mRNA editing enzyme) deaminates 5mC which can be excised by methyl-CpG-binding domain protein 4 (MBD4), a member of the BER family of proteins (Rai *et al.*, 2008). A similar process was also found in mammals, where deamination of 5mC leads to T:G mismatches, which are excised by thymine DNA glycosylase (TDG) or MBD4 and replaced by BER (Bhutani *et al.*, 2010; Popp *et al.*, 2010; Rai *et al.*, 2008). Alternatively in mouse brain, AID/APOBEC deaminates 5hmC (5-hydroxymethylcytosine) to 5hmU (5-hydroxymethyluracil), which can be excised by the glycosylases TDG and single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1) (Cortellino *et al.*, 2011; Guo *et al.*, 2011). At the moment the significance of these processes in mammals is discussed controversially, since the affinity of AID/APOBEC to 5hmC and 5mC *in vitro* is much lower than to C (cytosine) and the knockout of those enzymes did not show any developmental defects in mice (Muramatsu *et al.*, 2000; Nabel *et al.*, 2012). Other pathways of active demethylation, which include the oxidation of 5mC to 5hmC or further to 5fC (5-formylcytosine) and 5caC (5-carboxylcytosine), have been discovered. One example is the decarboxylation of 5caC to C, which has been demonstrated for mESC lysates, but did not reveal a possible candidate enzyme yet (Schiesser *et al.*, 2012). Other reports suggest a dehydroxymethylation function of DNMT3A and DNMT3B under oxidizing conditions (Chen *et al.*, 2012; Liutkeviciute *et al.*, 2009). Those mechanisms have been identified *in vitro* but their biological relevance has yet to be determined. The best studied example of active demethylation mechanisms in mammals includes the active modification of 5mC followed by active removal of the modified cytosine in a BER mechanism (AM-AR) (Kohli and Zhang, 2013). In the first steps TET (ten-eleven translocation) proteins convert 5mC to 5hmC

and further to 5fC and 5caC (He *et al.*, 2011; Ito *et al.*, 2011; Tahiliani *et al.*, 2009). Those modifications can then be subsequently excised by TDG via BER (He *et al.*, 2011; Maiti and Drohat, 2011). It has been proposed that in some cases the intermediates can also be diluted in a replication-dependent manner, combining active modification and passive dilution (AM-PD) (Kohli and Zhang, 2013). The AM-PD pathway contributes to the early loss of 5mC in both parental genomes after fertilization of the zygote (Guo *et al.*, 2014) whereas the AM-AR mechanism is abundant in many cell types (reviewed by Wu and Zhang, 2014).

Both pathways require the initial conversion step, which is catalyzed by the TET proteins. Therefore the following section focusses on the important role of TET proteins in cellular processes.

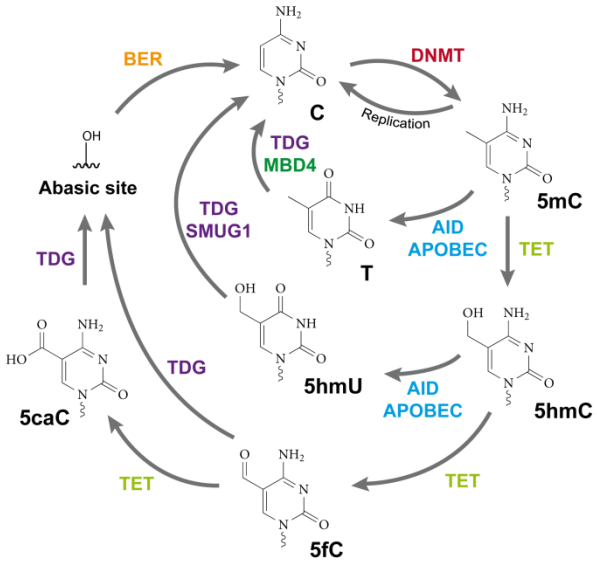

Figure 2.2: DNA demethylation pathways
Schematic overview of possible pathways for DNA demethylation (modified from (Branco *et al.*, 2012))

TET proteins

All three mammalian members of the family of ten-eleven translocation proteins TET1, TET2, and TET3 are Fe^{2+}- and 2-oxoglutarate-dependent dioxygenases which directly convert 5mC to 5hmC and further to 5fC and 5caC in the active demethylation process. All TET proteins share similar structural features, for example a c-terminal catalytic domain, containing a double-stranded β-helix fold and an adjacent cysteine-rich domain. Both domains can form a compact catalytic fold, which converts 5mC to 5hmC (Hu *et al.*, 2013b). A conserved Cysteine-X-X-Cysteine (CXXC) zinc-finger domain is located at the n-terminus of TET1 and TET3. The CXXC domain facilitates DNA binding at unmodified, CpG-rich sequences and is a common feature of DNA binding

proteins like DNMT1. In TET proteins the CXXC regulates the stability of the proteins, but different functional properties have been reported. TET1 CXXC preferentially binds CpG sequences regardless of the modification status of the cytosine or, controversially, no DNA at all (Frauer *et al.*, 2011; Xu *et al.*, 2011; Zhang *et al.*, 2010) and mutations of the domain don't alter the expression or activity of TET1 (Ko *et al.*, 2013). The CXXC domain of TET3 targets unmodified cytosines in both CpG and non-CpG context (Xu *et al.*, 2012) and regulates protein levels and tethering to DNA (Ko *et al.*, 2013). In the case of TET2, the CXXC domain has been separated in a chromosomal inversion event and now encodes for the IDAX (inhibition of the Dvl and axin complex) protein (Iyer *et al.*, 2011; Iyer *et al.*, 2009), which regulates TET2 turnover. ES cells depleted of TET2 failed to downregulate its protein levels and overexpression of IDAX in a myeloid cell line resulted in elevated TET2 and 5hmC levels (Ko *et al.*, 2013). The CXXC domain of IDAX targets unmethylated CpGs at CGI and promoters, and plays an important role in the recruitment of TET2 to its target sites, which is significantly enriched at CpG-rich promoters (Chen *et al.*, 2013; Deplus *et al.*, 2013; Ko *et al.*, 2013)). A recent study reported a direct DNA-binding potential of TET2 via its cysteine-rich domain. The interaction is sequence specific and shows substrate preference for 5mC in a CpG context. Interestingly the binding is independent of the methyl-group and does not distinguish methylated or unmethylated CpGs, indicating an involvement of other TET-interacting proteins in the correct targeting of TET2 to effector regions (Hu *et al.*, 2013b). Taken together TET2 binding to DNA is dependent on the presence or absence of specific targeting proteins (Pastor *et al.*, 2013), and further investigation of possible candidate enzymes is needed to clarify the issue.

The interaction of TET proteins with their target regions results in regulation of transcriptional activity, which is mostly associated with local demethylation events (see below) and may involve the recruitment of chromatin modifying complexes. Recent studies identified the *O*-linked β-$_D$-*N*-acetylglucosamine (*O*-GlcNAc) transferase OGT to be one of the direct binding partners of TET proteins (Balasubramani and Rao, 2013; Chen *et al.*, 2013; Deplus *et al.*, 2013; Vella *et al.*, 2013). OGT catalyzes the addition of *O*-GlcNAc sugars to a variety of proteins *in vivo*, including chromatin-modifying enzymes (Kreppel *et al.*, 1997; Vosseller *et al.*, 2002). It has been shown that the *O*-GlcNAcylation of H2B, which is associated with transcriptional activation, can be found at TET2 binding sites (Chen *et al.*, 2013). Other interactions of OGT and TET proteins involve the recruitment of the Set1/COMPASS (complex associated with Set1) complex, which results in H3K4me3 deposition and thus activates transcription (Deplus *et al.*, 2013). Those studies suggest an important role of TET proteins in transcriptional regulation, but whether it is of active or repressive nature depends on

the interaction partners. TET1 for example is highly abundant in murine ESC (mESC) and can be found at active, CpG-rich promoters and at repressed bivalent promoters in the same cell type. On the one hand, TET1 facilitates oxidation of 5mC to maintain an active state at promoters or distal-regulatory elements (Ficz *et al.*, 2011; Shen *et al.*, 2013; Wu *et al.*, 2011b). At poised genes, on the other hand, which are marked by bivalent chromatin modifications (H3K4me3 and H3K27me3) TET1 recruits the PRC2 complex, associated with transcriptional repression (Wu and Zhang, 2011). Interaction of TET1 with other repressing complexes was detected for the MBD3/NuRD (methyl-CpG binding domain protein 3/ nucleosome remodeling deacetylase) (Yildirim *et al.*, 2011) and SIN3A (SIN3 transcription regulator family member A) complex (Williams *et al.*, 2011), but further insights into the selection of active or repressive interaction partners are needed to determine their importance for gene regulation (Wu and Zhang, 2014).

Several studies in the past years have pointed out a link between functional disruption of TET proteins and human malignancies. While TET1 and TET3 mutations are rare in hematological malignancies, over 700 TET2 mutations have been identified to be associated with leukemia (Abdel-Wahab *et al.*, 2009; Euba *et al.*, 2012; Liu *et al.*, 2014; Weissmann *et al.*, 2012). They mostly cause a loss-of-function of the protein and result in decreased 5hmC levels, for example in the bone marrow, as well as aberrant DNA methylation (Ko *et al.*, 2010; Konstandin *et al.*, 2011). Taken together these findings indicate an important role for TET2 as a tumor suppressor in the hematopoietic lineage (Delhommeau *et al.*, 2009). This hypothesis is supported by several murine TET2 knock-out models. Mice lacking TET2 are viable, fertile and develop a mostly normal phenotype, but acquire hematopoietic malignancies with age. Their disease state is associated with an increased number of myeloid and lymphoid progenitor cells that failed to differentiate. This phenotype resembles the human CMML (chronic myelomonocytic leukemia) and confirms the crucial role of TET2 in hematopoietic lineage commitment and differentiation (Li *et al.*, 2011; Moran-Crusio *et al.*, 2011; Quivoron *et al.*, 2011).

TDG

The thymine-DNA glycosylase (TDG) is a member of the base excision repair system and has previously been linked to DNA demethylation processes in mammals. It recognizes and excises pyrimidine base T:G or U:G mismatches in a base-flipping mechanism and initiates the BER mechanism to restore the proper base (Cortazar *et al.*, 2007; Neddermann and Jiricny, 1994). Two possible TDG-mediated pathways are initiated by the deamination of 5mC or 5hmC via AID/APOBEC. TDG is responsible for

subsequent removal of the resulting 5hmU and T, followed by base excision repair to restore C (Guo *et al.*, 2011; Popp *et al.*, 2010). It is also involved in the AM-AR demethylation process, where TDG is responsible for the removal of 5fC and 5caC modifications (He *et al.*, 2011; Maiti and Drohat, 2011). In contrast to C, 5mC and 5hmC the N-glycosidic bond of 5fC and 5caC is much weaker (Bennett *et al.*, 2006) and allows an efficient removal of those modifications at low energy cost (Williams *et al.*, 2012). Knockdown and catalytic mutation experiments revealed a crucial role for TDG in embryonic development. $TDG^{-/-}$ mice are lethal with distinct developmental defects in the embryo and show aberrant DNA methylation and decreased expression of developmental transcription factors (Cortazar *et al.*, 2011; Cortellino *et al.*, 2011). The association with activating histone modifiers like MLL1 (mixed-lineage leukemia 1) methyltransferase and p300/CBP (E1A binding protein p300/CREB-binding protein) at promoters was impaired upon TDG depletion, indicating an important role of TDG in the recruitment of transcriptional activators and chromatin maintenance (Cortazar *et al.*, 2011; Tini *et al.*, 2002). Subsequent studies in TDG-depleted mESC detected about 10-fold elevated levels of 5fC and 5caC (He *et al.*, 2011; Shen *et al.*, 2013; Song *et al.*, 2013), which were enriched at inactive, poised promoters and regions of lineage-specific TF binding. Accumulation of 5fC and 5caC was also found at promoter-distal regulatory elements associated with H3K4me1. The poised state of these enhancers suggested a role for 5fC and 5caC in enhancer priming (see chapter 2.2.1 and (Shen *et al.*, 2013; Song *et al.*, 2013). These findings indicate a role for the TET/TDG-mediated active demethylation process at regulatory elements in the early development. Further investigation of demethylation processes at those elements is crucial for the complete understanding of cell type-specific regulation of gene expression.

2.2.1 Genomic distribution of oxidized 5-methylcytosines

To interpret the function of active demethylation, a number of studies have mapped oxidized modifications of 5mC throughout the genomes of various cell types (Wu and Zhang, 2014) and detected the presence of intermediates of the TET/TDG-mediated DNA demethylation process at specific genomic elements.

Promoters and gene bodies

Promoters are often associated with CpG islands and lack DNA methylation. DNMTs are blocked by the presence of active H3K4me3 and H3K27ac histone marks and the histone variant H2A.Z, which keep CGI promoters transcriptionally active (Ooi *et al.*, 2007; Wu and Zhang, 2014; Zilberman *et al.*, 2008).

In line with the lack of 5mC at mESC TSS, CpG-rich promoters are also devoid of 5hmC, 5fC and 5caC, the modification products of 5mC. In contrast, promoters with low or intermediate CpG-density are enriched for 5hmC in mouse and human ESC (Pastor *et al.*, 2011; Shen *et al.*, 2013; Song *et al.*, 2013; Szulwach *et al.*, 2011; Yu *et al.*, 2012). Among those, promoters of lineage-specific genes (mESC) carry a bivalent chromatin signature in (H3K4me3, H2K27me3). The accumulation of 5hmC, 5fC and 5caC at those sites, indicates maintaining functions of TET/TDG-mediated demethylation at poised promoters. Similar enrichment of 5hmC, 5fC and 5caC at gene bodies further supports the significance of active demethylation events in mESC transcription regulation (Shen *et al.*, 2013; Song *et al.*, 2013).

Enhancer

Of all genomic locations 5hmC is most abundant at cis-regulatory elements, called enhancers (Stadler *et al.*, 2011; Stroud *et al.*, 2011; Szulwach *et al.*, 2011; Yu *et al.*, 2012). These cis-acting, regulatory elements are ~200-500 bp in length, and marked by active chromatin marks H3K4me1 and H3K27ac in the active state (Shen *et al.*, 2012; Spitz and Furlong, 2012). Enhancers establish long-range interactions with promoters and recruit activating TF to regulate gene expression. RNA polymerase II occupancy is another feature of active enhancers and leads to the transcription of enhancer RNAs (eRNA) (Natoli and Andrau, 2012). Several studies suggest that those eRNAs are important for the formation and stabilization of the promoter-enhancer-interaction, but controversial data exists (Hah *et al.*, 2013; Lai *et al.*, 2013; Li *et al.*, 2013; Melo *et al.*, 2013).

Active enhancers are occupied by a variety of common and cell type-specific TFs at proximal clusters of binding sites, which is associated with nucleosomal depletion, loss of DNA methylation and increased DNase hypersensitivity (He *et al.*, 2010; Mito *et al.*, 2007; Schmidl *et al.*, 2009; Song *et al.*, 2011). In line with this observation 5hmC is absent from the binding site of TFs but enriched in the surrounding area, indicating a constant turnover of 5mC to keep those regions open (Lister *et al.*, 2013; Yu *et al.*, 2012). TFs themselves recruit co-activator proteins which lack the ability to bind DNA and act as chromatin remodelers, histone modifiers or as mediator of long-range

crosstalk with the basal transcriptional machinery at the promoter (Borggrefe and Yue, 2011; Calo and Wysocka, 2013; D'Alessio *et al.*, 2009; Roeder, 2005; Weake and Workman, 2010). TF occupancy is increased at active and poised enhancers (Creyghton *et al.*, 2010; Heinz *et al.*, 2010). In ESC poised state enhancers are transcriptionally inactive, characterized by low nucleosome density and surrounded by bivalent histone marks H3K27me3 and H3K4me1 (Creyghton *et al.*, 2010; Rada-Iglesias *et al.*, 2011). The poised state is similar to activated enhancer states and indicates a priming or pre-marking at those regions for rapid activation upon future cell-fate decisions (Buecker and Wysocka, 2012). Association of poised enhancers with key developmental TF, like OGT4 in human ESC (hESC) supports this notion and suggests that they are important drivers of differentiation (Loh and Lim, 2010; Thomson *et al.*, 2011).

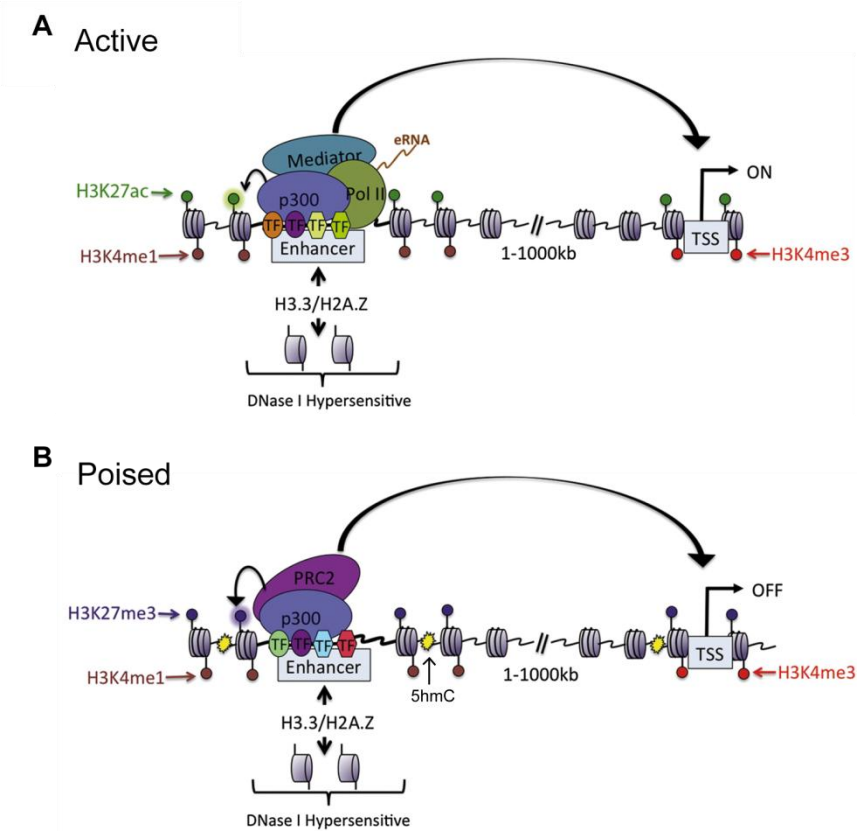

Figure 2.3: Hallmarks of active and poised enhancers
Schematic overview of characteristic factors and chromatin at active and poised enhancers. (Modified from (Calo and Wysocka, 2013))

Accumulation of active demethylation intermediates, 5hmC and 5fC at poised and 5caC preferentially at active enhancers further supported the involvement of TET/TDG-mediated active demethylation processes at distal regulatory elements in mESC (see previous chapter and (Shen *et al.*, 2013; Song *et al.*, 2013). A recent study further

confirmed this hypothesis, showing that hypermethylation following a triple TET knockout in mESC mainly occurred at active and poised enhancer elements (Lu *et al.*, 2014). The authors demonstrated a direct link of TET/TDG-mediated demethylation at distal regulatory elements with the regulation of cell type-specific transcription to establish and mediate cell identity.

2.3 The transcription factor PU.1

Among the key transcription factors regulating gene expression patterns, PU.1 (PU box binding protein 1) is one of the most studied examples and plays a major role in hematopoiesis. It is responsible for the regulation and conduction of this multi-step process and controls lineage decisions and cell-fate of a variety of specialized blood cell types arising from one hematopoietic stem cell precursor (Arinobu *et al.*, 2007; Back *et al.*, 2005; Friedman, 2007; Singh *et al.*, 1999).

PU.1 is a member of the E26 transformation-specific (ETS) transcription factor family. Its functional structure consists of several distinct domains like an ETS domain next to a helix-turn-helix domain, which both facilitate binding to specific DNA sequence motifs (Kodandapani *et al.*, 1996; Pio *et al.*, 1996; Wei *et al.*, 2010). PU.1 further contains an N-terminal acidic and a glutamine-rich domain for transcription activation and a PEST (peptide sequence rich in proline, glutamic acid, serine and threonine) domain which is involved in protein-protein-interactions. Due to its unique structural properties, PU.1 belongs to a specialized group of transcription factors with pioneering characteristics (Zaret and Carroll, 2011). It has the particular ability to access and bind DNA in DNase I-inaccessible heterochromatin regions. In cooperation with chromatin remodeling and -modifying interaction partners it facilitates repositioning of nucleosomes and opens chromatin structure to enable occupancy of other TF (Ghisletti *et al.*, 2010; Pham *et al.*, 2013; Schonheit *et al.*, 2013; Serandour *et al.*, 2011).

Together with other TF PU.1 is able to alter cell type-specific expression patterns (Ghisletti *et al.*, 2010; Heinz *et al.*, 2010; Pham *et al.*, 2013). And different combinations of lineage-specific TFs at given cell states, determine the PU.1-dependent regulation of the differentiation process in the hematopoietic lineage. For example, in concert with the transcription factor C/EBPα PU.1, guides differentiation of the myeloid lineage via alteration of the expression pattern towards myeloid-specific genes (Hohaus *et al.*, 1995). The interactions can also be antagonistic, as was shown for the erythroid lineage-specific TF GATA-1. The specific interplay of these factors

determines the cell-fate of precursor cells and drives their commitment towards the myeloid or erythroid-megakaryocytic lineage (Zhang *et al.*, 1999; Zhang *et al.*, 2000).
The ability of PU.1 to recruit and assembly chromatin or DNA modifying complexes indicates another important role in the regulation of gene expression. In cooperation with the histone acetyltransferases CBP and p300, PU.1 promotes active transcription (Bai *et al.*, 2005; Yamamoto *et al.*, 1999). Other reports have found repressive interaction partners of PU.1. In combination with the methyl CpG-binding protein 2 (MeCP2) or with the DNA methyltransferases DNMT3A and 3B, PU.1 target sites are transcriptionally repressed (de la Rica *et al.*, 2013; Imoto *et al.*, 2010; Suzuki *et al.*, 2006). Studies in hematopoietic cell systems, which express PU.1 endogenously, have displayed controversial data regarding the involvement of PU.1 in DNA methylation changes during differentiation. Our group identified a general association of PU.1 binding with local demethylation events in differentiating monocytes and macrophages (Pham *et al.*, 2013). This finding is in line with work of de la Rica et al. who proposed that an interaction of PU.1 with TET2 regulates transcriptional activation in monocyte-derived osteoclasts (de la Rica *et al.*, 2013). Interestingly they also linked PU.1 to DNA methylation in cooperation with DNMT3B, which is opposite to our finding that PU.1 binding is inversely correlated to 5mC appearance (Pham *et al.*, 2013). Further investigation is needed to address this controversy.

2.4 The mononuclear phagocyte system

2.4.1 The mononuclear phagocyte system

In the past years several cell systems in mammals have been discovered, isolated and recapitulated for *in vitro* studies. One intensively studied example is the mononuclear phagocyte system (MPS). It was initially defined by van Furth and Cohn in 1968, comprising a network of non-granulocytic, myeloid cells with extraordinary phenotypic plasticity (van Furth and Cohn, 1968), and it belongs to the hematopoietic lineage tree (Abbas *et al.*, 2007). Peripheral blood-monocytes, the non-dividing progenitor cells of the MPS, give rise to a variety of cell types *in vivo*. Upon distinct stimuli they are able to differentiate into tissue macrophages, myeloid dendritic cells and osteoclasts (Geissmann *et al.*, 2010; Hume, 2008). In this process PU.1 plays an important role. It regulates gene expression in myeloid cells, like macrophages or monocytes (De Kleer *et al.*, 2014; Ghisletti *et al.*, 2010) and is essential for DC development (Carotta *et al.*, 2010a). As a fundamental component of the adaptive and innate immune response,

the MPS is in charge of important stimulatory and regulatory functions as well as functions in tissue remodeling and homeostasis (Yona and Jung, 2010). It is involved in chronic inflammatory diseases like rheumatoid arthritis, inflammatory bowel disease, neurodegenerative disorders, and arteriosclerosis (Valledor *et al.*, 2010). Although the MPS acts as a dynamically interacting network of cells with similar functional features, it comprises individual lineages with distinct roles as well as different origins and maintenance strategies (Jenkins and Hume, 2014).

The unique differentiation steps of the MPS can be recapitulated *in vitro* using a system published by Sallusto and Lanzavecchia (Sallusto and Lanzavecchia, 1994). Upon stimulation with granulocyte-macrophage colony stimulating factor (GM-CSF) and IL-4 (interleukin 4), isolated peripheral blood-monocytes can be differentiated into immature dendritic cells (iDC) and further matured by adding toll-like receptor ligands, TNFα or LPS (De Smedt *et al.*, 1996). However, in the presence of human AB-serum, isolated monocytes differentiate into macrophages after seven days of cell culture (Andreesen *et al.*, 1983) (see Figure 3.1).

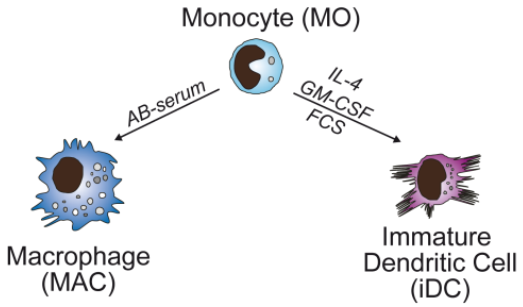

Figure 2.4: Cell model of the MPS *in vitro*
Schematic overview of differentiating monocytes *in vitro*. (Modified from (Klug *et al.*, 2013))

One of the unique features of this *in vitro* system is the absence of proliferation during the differentiation of monocytes into iDC or MAK (Cheung and Hamilton, 1992; Clanchy *et al.*, 2006; Finnin *et al.*, 1999; Pickl *et al.*, 1996; van Furth *et al.*, 1979). Several proliferation assays confirmed, that the transition of one cell type into another occurs independent of any replication events (Klug *et al.*, 2010). This cellular setting therefore represents an ideal model system to study molecular mechanisms involving enzymatic activity in the absence of DNA replication.

2.4.2 Monocytes

Monocytes belong to the group of white blood cells or leukocytes and originate from hematopoietic precursor cells in the bone marrow (BM) (Abbas *et al.*, 2007). They are defined as non-proliferative, circulating cells in the peripheral blood system (Hume, 2006; van Furth *et al.*, 1979) and represent 10% of peripheral leukocytes in human (4% in mice). Monocytes can be classified into three subpopulations by their distinctive expression of the lipopolysaccharide (LPS)-coreceptor CD14 and the Fcγ III receptor CD16 (Passlick *et al.*, 1989; Ziegler-Heitbrock *et al.*, 2010). Around 90% of blood monocytes are referred to as "classical" monocytes due to their high expression of CD14 and lack of CD16 ($CD14^{++}CD16^{-}$). The other two subgroups are referred to as "non-classical" ($CD14^{dim}CD16^{+}$) and "intermediate" monocytes ($CD14^{+}CD16^{+}$). Each subgroup is characterized by differential gene-expression patterns, transcriptional regulation and specific functions (Schmidl *et al.*, 2014; Zawada *et al.*, 2011).

In the MPS, monocytes constitute the most mobile and dynamic population (Yona and Jung, 2010). Via the bloodstream these myeloid precursor cells have access to all tissues in the body and supply local sites of immune response with effector cells of the MPS (Yona and Jung, 2010). To reach the inflamed areas they use chemokine and adhesion receptors in order to cross tissue barriers and produce cytokines to mediate local inflammatory reactions. In the current view monocytes in the blood stream represent a reservoir to maintain small subpopulations of the macrophage (MAK) and dendritic cell (DC) compartments (Geissmann *et al.*, 2003; Randolph *et al.*, 1998; Serbina and Pamer, 2006). In general both cell types originate from other precursor cells from the bone marrow or fetal tissues (Ginhoux *et al.*, 2010; Yona *et al.*, 2013). However, in certain settings and under inflammatory conditions, monocytes are able to differentiate into MAK and DCs (Ingersoll *et al.*, 2011; Serbina *et al.*, 2008; Shi and Pamer, 2011). In mice, for example, dendritic cells in the mucosal epithelium, called Langerhans' cells originate from bone marrow precursors, but can be derived from $LY6C^{hi}$ monocytes in an inflammatory milieu (Iijima *et al.*, 2007). Each new study adds more details to the classification of immune cells, but makes it difficult to clearly define the borders of their definition. A subset splenic DC, called Tip-DCs (TNFα and iNOS expressing DCs) are derived from MO (Serbina *et al.*, 2003) in mice. But according to recent characterizations of their phenotype, they represent activated effector monocytes rather than dendritic cells (Mildner *et al.*, 2013).

2.4.3 Macrophages

Macrophages are stationary, phagocytotic cells residing in a variety of lymphoid and non-lymphoid tissues, like microglia in the brain (Mildner *et al.*, 2007), alveolar MAK (Sawyer *et al.*, 1982; Tarling *et al.*, 1987), splenic white pulp and metallophilic MAK (Wijffels *et al.*, 1994), and Kupffer cells in the liver (Crofton *et al.*, 1978). Tissue macrophages are important effector cells of the immune system. Their broad spectrum of receptors is important for pathogen-specific phagocytosis and they are able to regulate and maintain the inflammatory response via inflammatory cytokines. In addition, macrophages have various functions in tissue homeostasis. They produce growth factors for tissue remodeling and repair after injury and are responsible for the elimination of apoptotic cells (Gilroy *et al.*, 2004; Gordon, 2002, 2007; Lawrence *et al.*, 2002). Macrophages are derived from different origins. Although monocytes can differentiate into MAK *in vitro* they are not the main source of MAK *in vivo*. Most tissue macrophages, like microglia (Ginhoux *et al.*, 2010; Schulz *et al.*, 2012), alveolar MAK (Guilliams *et al.*, 2013) or Kupffer cells (Schulz *et al.*, 2012), originate from fetal precursors and are replenished by local proliferation. Monocyte-derived macrophages are mainly inflammatory MAK, which stem from peripheral blood monocytes infiltrating the inflamed tissue (Ingersoll *et al.*, 2011; Shi and Pamer, 2011). Interestingly several studies detected additional MO-derived MAK sub-populations in a variety of other tissues, like liver, kidney, and peritoneal cavity (Ginhoux *et al.*, 2009; Goldszmid *et al.*, 2012). And intestinal macrophages, in the lamina propria, are completely replenished from monocytes originating from BM or fetal precursors (Varol *et al.*, 2009). These findings suggest that, independent of infection, several subsets of tissue MAK are permanently replenished from monocytes in the steady state (Jenkins and Hume, 2014). The majority of these observations have been obtained from murine systems and still await confirmation in the human setting. Although both organisms share similarities in their hematopoietic organization (Tavian and Peault, 2005) humans have different life spans than rodents and the MAK homeostasis in aging humans is still unclear.

2.4.4 Dendritic cells

The dendritic cell compartment comprises a variety of specialized immune cells subsets which were first discovered in 1973 as potent stimulators of naïve T cells (Steinman and Cohn, 1973). Dendritic cells are widely distributed in lymphoid and non-lymphoid organs and characterized by different functions and phenotypes (Carotta *et al.*, 2010b). Per definition they functionally belong to the family of antigen-presenting cells (APC) and have important effector functions on microbes and pathogens in the immune response (Pulendran, 2004). But recent advances in DC classification suggest a different definition of DC according to their $Flt3^+$ positive precursor cells. Although those DC arise from a different progenitor in the bone marrow, compared to MO and MAK, they are still considered to compose a large fraction of the MPS (Jenkins and Hume, 2014).

Dendritic cells are mainly divided into two, specialized subgroups; “classical” (cDC) and “plasmatoid” (pDC) DCs. Classical dendritic cells include a variety of non-lymphoid and lymphoid tissue DCs which recognize invading pathogens and are specialized for the maintenance of self-tolerance (Banchereau and Steinman, 1998; Merad and Manz, 2009; Steinman *et al.*, 2003). In contrast, pDC are characterized by their ability to produce high quantities of type I interferons (IFNs) in response to viral infections (Nakano *et al.*, 2001; Siegal *et al.*, 1999). The emerging amount of studies of DC subsets revealed further subgroups and overlapping classifications due to their functions, phenotypes and locations. Thus a third group of “non-classical” or “monocyte-derived DCs” (moDCs) was identified (Mildner *et al.*, 2013; Segura and Amigorena, 2013). Those cells are phenotypically similar to cDC but differ in their gene expression patterns (Xu *et al.*, 2007). Those studies further identified PU.1 as a key driver of DC development. It alters higher-order chromatin structure to regulate the expression of Irf8, an important factor for DC development (Schonheit *et al.*, 2013). And it controls Flt3 expression in DC precursor cells in mice, to maintain cDC and pDC populations (Carotta *et al.*, 2010a).

Taken together, these studies show that PU.1 is a crucial TF for all members of the MPS and questions regarding its regulatory function or mechanistic details can be addressed in the *in vitro* model of this cell system.

3 Results

3.1 The active DNA demethylation mechanism in the human mononuclear phagocyte system

The main aim of this work was to characterize active demethylation in the mononuclear phagocyte system and to identify the factors involved in the process. Furthermore we wanted to gain insights into global patterns and its role in regulation of cell fate and lineage determination.

The data presented in the following chapters (3.1.1. and 3.1.2.) have been published in Genome Biology (Klug *et al.*, 2013). To present a complete picture of the story, work from both co-authors is included in this work. When this is the case, the figures are cited and can be found online in (Klug *et al.*, 2013). The figures taken from the publication are all slightly adapted in size, format, etc. And at some points additional data is shown to give a complete and updated overview of the story. In any case the individual co-author's contribution is clearly stated.

3.1.1 Identification of DMR in the mononuclear phagocyte system

The cell system: The human mononuclear phagocyte system

The mononuclear phagocyte system (MPS) represents a distinct part of the hematopoietic cells. Its members are non-granulocytic, myeloid cells derived from monocytes, which play a crucial role in the innate and adaptive immune response (Hume, 2006; Steinman and Inaba, 1999; van Furth and Cohn, 1968). The phenotypic diversity of CD14+ positive blood monocytes contributes to a variety of myeloid cell types, like macrophages, dendritic cells and osteoclasts, depending on the surrounding environment. This system can be recapitulated *in vitro* as depicted in Figure 3.1. Peripheral blood monocytes (MO), isolated from blood, differentiate into macrophages (MAK) in the presence of human serum (Andreesen *et al.*, 1983) and immature

dendritic cells (iDC) can be derived by adding IL-4 and granulocyte-macrophage colony-stimulating factor (GM-CSF) (Sallusto and Lanzavecchia, 1994) to the culture medium.

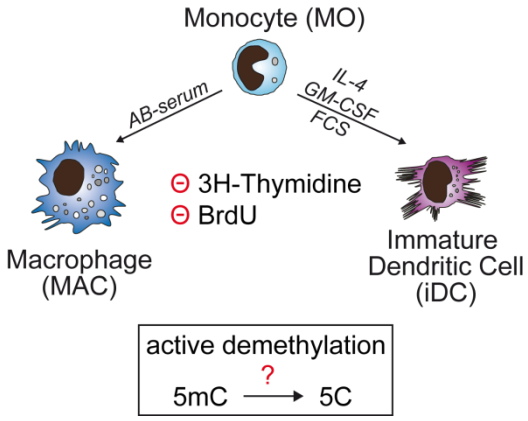

Figure 3.1: Monocyte differentiation *in vitro*
Schematic overview of the *in vitro* monocyte differentiation model. Monocytes don't proliferate in this setup (as demonstrated by the lack of nucleotide incorporation). Therefore DNA demethylation in this system requires an active process.

Earlier studies used BrdU and ^{3}H-Thymidine incorporation assays to show that monocytes do not proliferate during differentiation. This finding excludes the possibility of passive demethylation and confirms that the MPS is a suitable cell system to address questions on active DNA demethylation (Klug *et al.*, 2010) (overview in Figure 3.1). Recently our lab identified a number of differentially methylated regions (DMR) (Figure 3.2) in differentiating monocytes using MCIp and microarray techniques (Klug *et al.*, 2010). The changes in methylation were highly reproducible and not restricted to promoter regions. Further ChIP experiments characterized the surrounding histone environment as active. H3K4me3, an active histone mark, was found at demethylated TSS and the active enhancer marks H3K4me1 and H3K4me2 were both found at promoter-distal sites of demethylation. Taken together, these findings indicated recruitment of a demethylation machinery to the DMRs facilitating active DNA demethylation and participating in histone-modification processes. In turn these events lead to gene activation and transcriptional priming, (Klug *et al.*, 2010) and are the basis for the work presented in this thesis (Figure 3.1 and Figure 3.2 are published in (Klug *et al.*, 2013)).

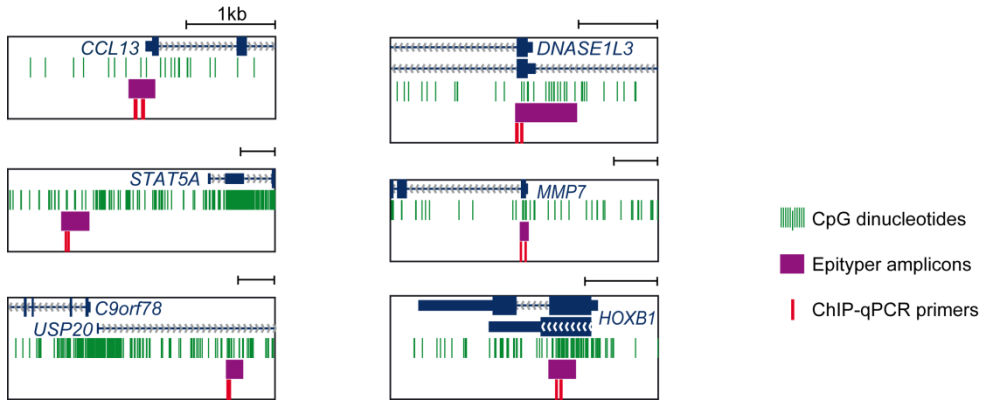

Figure 3.2: UCSC Genome Browser tracks of DMRs
Positions of regions (purple) measured by MALDI-TOF analysis of bisulfite converted DNA (MassARRAY, shown in Figure 3.3 B) and of primers (red) used for hMeDIP qPCR (see Figure 3.3) are shown relative to positions of CpG dinucleotides (green) and neighboring genes (blue). The tracks were generated using the UCSC Genome Browser.

3.1.2 Identification of key players in the DNA demethylation process

3.1.2.1 Time course of 5mC and 5hmC changes in DC

The conversion of 5mC to 5hmC has recently been proposed to be a possible initiator of active DNA demethylation processes. This step is catalyzed by a family of Ten-Eleven Translocation proteins (TET1-3) (Tahiliani *et al.*, 2009). To test for a possible involvement of TET proteins in the demethylation process we followed the methylation changes at previously defined sites (Figure 3.2) of active DNA demethylation in time courses of differentiating monocytes as displayed in Figure 3.3A. Bisulfite treated DNA was measured with the MassARRAY system by using mass spectrometry and six exemplary DMRs are shown as heatmaps displaying the differentiation time course of every CpG dinucleotide present in the amplicons. As expected all regions show DNA demethylation at several CpG residues. In two regions, *CCL13* and *USP20*, methylation is removed at early time points whereas *DNase1L3* and *Stat5* are demethylated at later stages of differentiation (“late” DMR). Two control regions are included in the figure which show no methylation (*MMP7*) at all or are constitutively methylated (*HOXB1*) throughout the seven day time course.

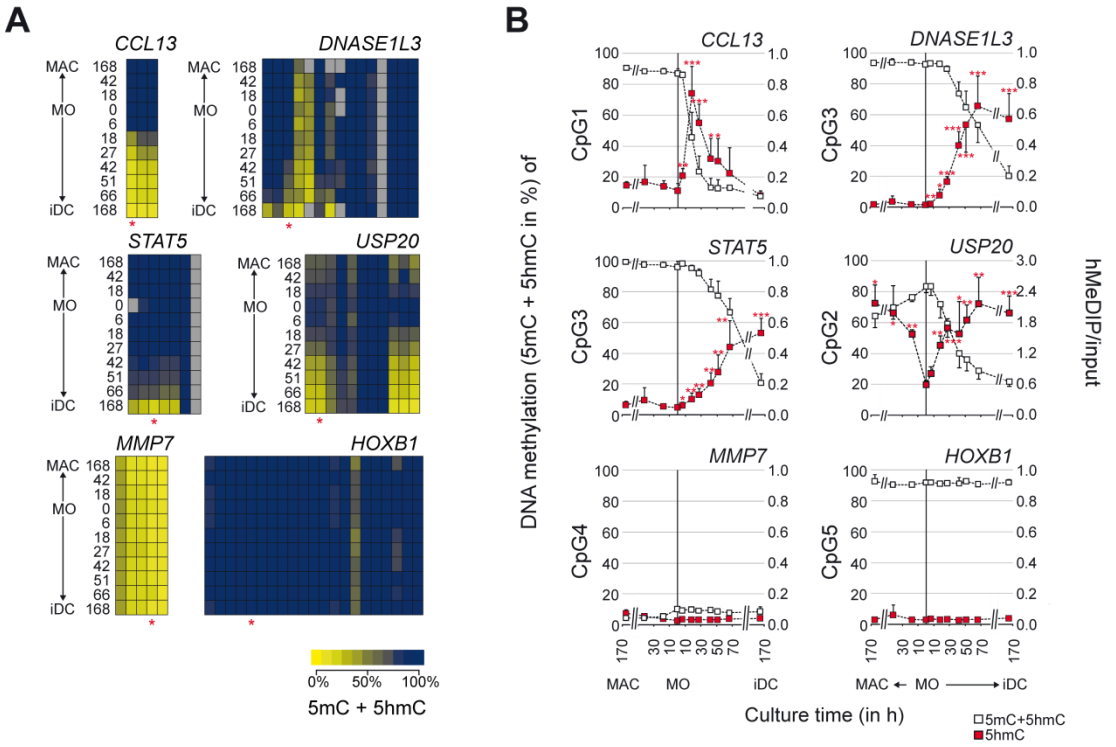

Figure 3.3: 5mC deposition precedes active DNA demethylation in human monocytes
(**A**) MassARRAY analysis of bisulfite-converted DNA at four loci that show active DNA demethylation during monocyte to DC differentiation, as well as four two control regions (values are mean of n≥4). Data are presented as heatmaps. The methylation content (including both 5mC and 5hmC) is indicated by coloring (yellow: no methylation, dark blue: 100% methylation) with each box representing a single CpG dinucleotide and each row representing the succession of CpGs measured. Grey boxes indicate CpGs that were not detected by MALDI-TOF MS. Red asterisks mark the CpGs that are shown in (B). Methylation ratios of single CpG units for individual donors are also available online (Klug *et al.*, 2013). (**B**) Dynamics of DNA methylation (5mC+5hmC) and 5-hydroxymethylation (5hmC) during monocytic differentiation. DNA methylation levels of single CpGs as measured by MassARRAY (open squares) are compared with 5hmC enrichment (measured by hMeDIP, red squares) at the same loci shown in (A) (*n*≥4, values are mean + or – SD). Exact genomic positions of analyzed CpG residues are given in Table S3, Additional File 3 (available online, Klug *et al.*, 2013).

In Figure 3.3B we compared the demethylation of single CpG dinucleotides (measured by mass spectrometry, indicated by red asterisks in Figure 3.3A) to the appearance of 5hmc at those regions. 5hmC was detected using 5hmC methyl-DNA immune precipitation (hMeDIP) and plotted against methylation levels. The progression of DNA demethylation during the differentiation time course coincides with the simultaneous appearance of 5hmC which is inversely proportional to methylation levels. In one DMR, the *CCL13* promoter region, not only 5mC but also 5hmC disappears completely towards the end of the time course, indicating completion of the demethylation process. DNA demethylation is not restricted to DCs only but can also be observed in MAK as displayed at the *USP20* region. This finding will be further investigated in later chapters (see section 3.2)

The MassARRAY is a stable and reproducible system to study DNA methylation at single CpGs. Unfortunately bisulfite treatment doesn't differentiate between 5mC and

5hmC; both DNA modifications are protected from bisulfite conversion and appear in the methylated fraction. To verify the loss of DNA methylation at the example regions we measured 5mC using MeDIP and compared it to 5hmC measured with hMeDIP (Figure 3.4) using the same samples and amplicons (Figure 3.2). Although the methylcytidine antibody displayed a lower affinity compared to the 5hmC antibody progressive demethylation was detected at all loci shown except the control regions. (The figures (Figure 3.3, Figure 3.4) are available online in (Klug *et al.*, 2013). The experiments were performed by M. Klug.)

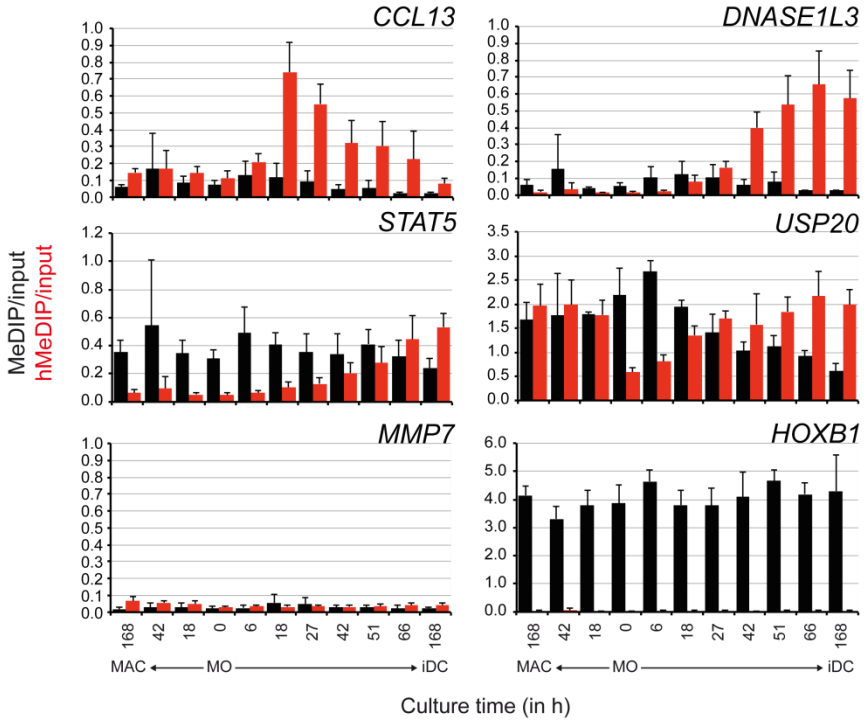

Figure 3.4: Comparison of hMeDIP and MeDIP results
The same samples were analyzed for 5mC and 5hmC enrichment by MeDIP and hMeDIP, respectively. Enrichment is shown relative to the signal observed with the genomic input material. Sensitivities of both antibodies were strikingly different with the 5hmC antibody being much more efficient in precipitating DNA. Especially at low CpG content regions, signals were difficult to quantify. (*n*=5, values are means ± SD).

In general antibody based pull-down techniques (like hMeDIP and MeDIP) are likely to show bias towards CpG dinucleotide density (Pastor *et al.*, 2011) in the enriched regions. Since the methylcytidine antibody did perform with lower affinity at regions with low CpG content, a chemical approach was used to independently confirm the data. Specific glycosylation of 5hmC followed by glycosyl-sensitive digest with *Msp*I and *Hpa*II is shown in Figure 3.5 for five example regions. The results showed similar patterns of demethylation and synchronous appearance of 5hmC as observed before. The *MMP7* region lacks *Msp*I restriction sites and could therefore not be analyzed with

this technique. (Figure 3.4 and Figure 3.5 are published in (Klug *et al.*, 2013). Experiments have been performed by S. Schmidhofer and M. Klug.).

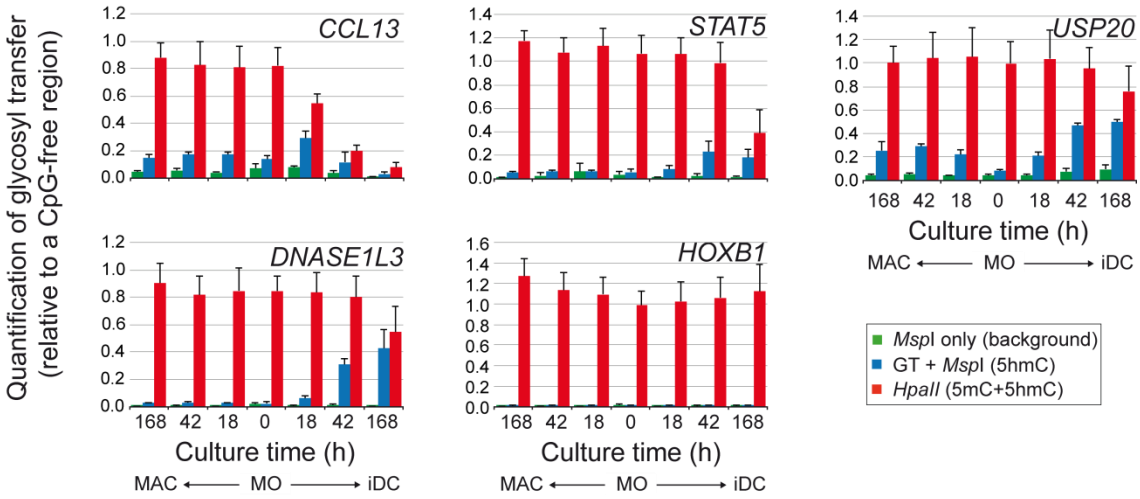

Figure 3.5: Alternative 5hmC measurements
Quantification of 5hmC using glycosyl transfer to 5hmC, followed by glycosylation-sensitive restriction (GGSD) with *Msp*I and subsequent amplification of protected regions. DNAs were also digested with *Hpa*II, which is sensitive to both 5hmC and 5mC. QPCR results are shown relative to the amplification of a CpG-free region. Values represent mean ± SD ($n \geq 3$) for control digests (*Msp*I only), digests after glycosyl transfer (GT+ *Msp*I) or digests with *Hpa*II (average for DNAs digested before and after glycosyl transfer). Exact genomic positions of analyzed CpG residues are given in Table S2, Additional File 2 of (Klug *et al.*, 2013). The *MMP7* region lacks *Msp*I recognition sequences and was not analyzed.

3.1.2.2 Knockdown experiments of possible factors involved in active DNA demethylation (Part I)

Taken together the above results strongly indicated the local conversion of 5mC to 5hmC to be the first step of the demethylation process which suggests a possible involvement of TET hydroxylases. Other studies have linked members of the base excision repair such as MBD4 and TDG to the active DNA demethylation process. The glycosylases are capable of binding and removing deaminated 5mC as well as 5hmC or its derivative oxidation products 5caC and 5fC (He *et al.*, 2011; Maiti and Drohat, 2011; Metivier *et al.*, 2008; Otani *et al.*, 2013). The first part of this chapter focuses on the characterization of TET proteins, MBD4 and TDG as prospect candidates responsible for the active DNA demethylation mechanism.

To investigate the involvement of each candidate enzyme in the active demethylation process, we established a transfection protocol for freshly isolated primary blood monocytes. Therefore the cells were transiently transfected with siRNA complementary

to candidate mRNAs or a control siRNA and differentiated into dendritic cells under iDC culture conditions (for the complete protocol see section 5.2.1.3.3).

3.1.2.2.1 mRNA expression levels of candidate enzymes

Before analyzing a causal link between DNA demethylation and possible candidates involved in the process the overall expression of TET proteins as well as DNA glycosylases was measured using qRT-PCR. In Figure 3.6 the levels of mRNA in a differentiation time course of monocytes to dendritic cells is shown relative to the expression of a housekeeping gene (*HPRT1*). In panel 3A we could observe an abundant expression of *TET2* as well as *TET3*, however the latter was expressed at much lower levels. Whereas MO and DC mainly express *TET2*, *TET1* was not detectable in those cell types, neither on mRNA (Results not shown) nor on protein level (see section 3.1.2.2.3). Since *TET3* was expressed at low levels and not detectable with commercial antibodies in Western blotting (see 3.1.2.2.3) we focused on TET2 as a candidate enzyme for further experiments. (Figure 3.6 is modified from (Klug *et al.*, 2013). Experiments were performed by S. Schmidhofer and M. Klug.)

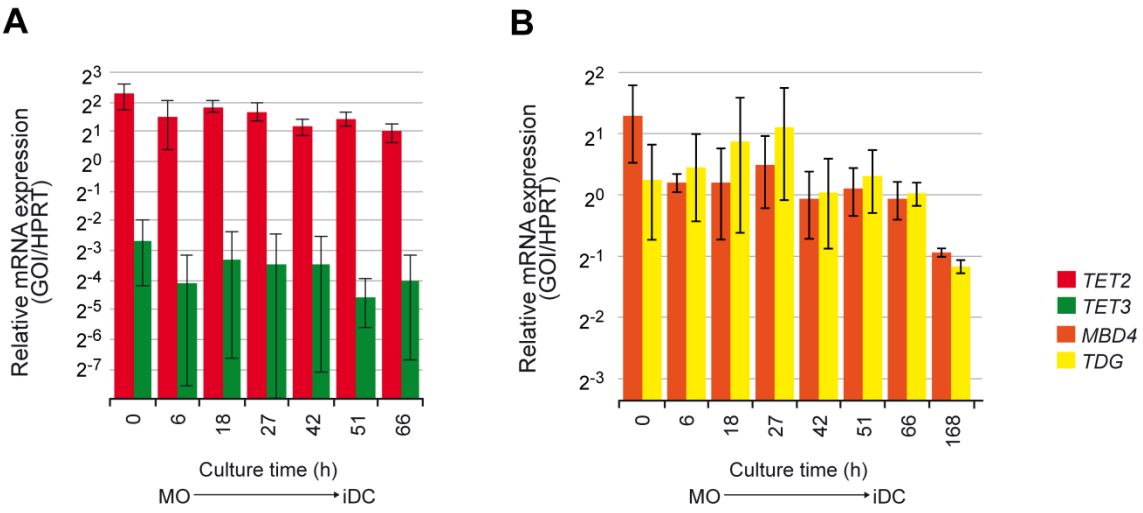

Figure 3.6: Candidate enzymes are expressed in human monocytes
The expression profile of *TET2* and *TET3* (**A**) or *MBD4* and *TDG* (**B**) during monocyte differentiation into dendritic cells is shown. Quantitative RT-PCR results are shown relative to *HPRT1* expression and represent mean values ± SD (*n*=2-6). No mRNA expression was detected for *TET1* in monocytes or monocyte-derived cells.

MBD4 and *TDG* expression was measured in time courses of differentiating monocytes as well. Both glycosylases were abundantly expressed at every time point of differentiation at similar amounts. The results are displayed in Figure 3.6B. These

findings confirmed both glycosylases as potential contributors to the active DNA demethylation process. Thus both enzymes were included in further experiments.
(S. Schmidhofer performed the experiments.)

3.1.2.2.2 Cell conditions after siRNA transfection

Monocytes and dendritic cells are members of the immune system and react sensitively to foreign nucleic acids and transfection procedures. We therefore choose to study early time points to avoid severe transfection-induced effects on survival and differentiation. To monitor cell conditions after knockdown the cells were stained and analyzed by flow cytometry. As shown in the DAPI staining in Figure 3.7 the large majority (>98%) of cultured cells were viable after 42 h of transfection. Two thirds of untreated cells displayed induced levels of the surface marker CD1a which is specific for dendritic cells, indicating ongoing DC differentiation. In contrast, all transfected cells showed reduced numbers of cells bearing CD1a, while most cells presented CD14 (a marker for MO) only, representing a delay in differentiation.

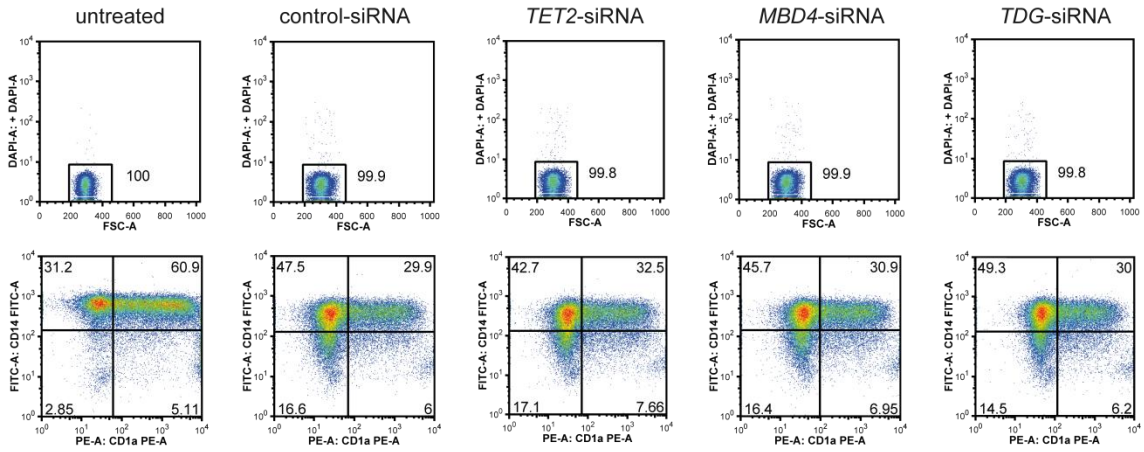

Figure 3.7: Survival of siRNA transfected monocytes
Untreated or siRNA-transfected cells after 42h were stained with DAPI (Sigma), anti- CD14-Fitc (My4a, Coulter Clone, a marker for monocytes) and anti-CD1a-PE (T6-RD1; Coulter Clone; a marker for iDC) and analyzed by flow cytometry using the LSRII from BD (Heidelberg, Germany). The large majority of cells (>98%) was viable (as demonstrated by the exclusion of DAPI in most cells, upper panels) and entered the DC differentiation pathway (as indicated by the induction of CD1a, lower panels), although numbers of differentiating cells were consistently reduced in all transfected samples.

On mRNA level the differentiation marker CD1a as well as the expression of the *CCL13* region was also reduced in transfected cells compared to untreated cells (qPCR levels shown in Figure 3.8). These data showed that survival was largely unaffected by transfection, while differentiation was impaired by the procedure.

(Figure 3.7 and Figure 3.8 are modified from (Klug *et al.*, 2013). Expression assays were performed by S. Schmidhofer and M. Klug. Cells for flow cytometry were prepared and stained by S. Schmidhofer. The flow cytometry measurements were conducted by Kristina Doser (AG Edinger/Hoffmann).)

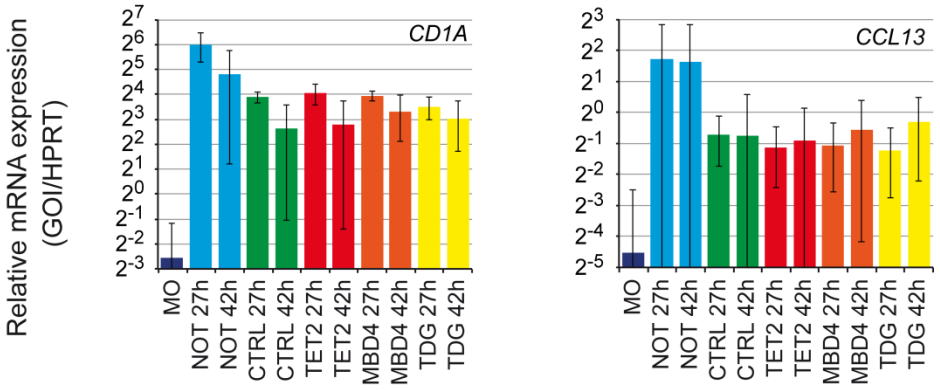

Figure 3.8: Marker gene expression of siRNA transfected monocytes
Expression of *CD1A* and *CCL13* marker genes in monocytes either non-transfected (NOT) or transfected with TET2-, MBD4-, TDG-siRNA or control siRNA were measured after 27h and 42h of differentiation culture. qRT-PCR results were normalized to *HPRT1* expression. Values represent mean ± SD (n≥4).

3.1.2.2.3 siRNA knockdown efficiency

Figure 3.9 illustrates knockdown efficiencies on mRNA (normalized to HPRT1 expression, top panel) and protein level (lower panel).

The TET2 siRNA (Figure 3.9A) treated cells showed significantly reduced levels of TET2 mRNA at 27 h and at 42 h after transfection as well as reduced amounts of protein at both time points. While the knockdown of MBD4 (Figure 3.9B) was even more distinct after 27 h and 42 h compared to TET2 siRNA, TDG mRNA depletion shown in Figure 3.9C was significant after 42 h. On protein level both glycosylases displayed decreased levels of protein.

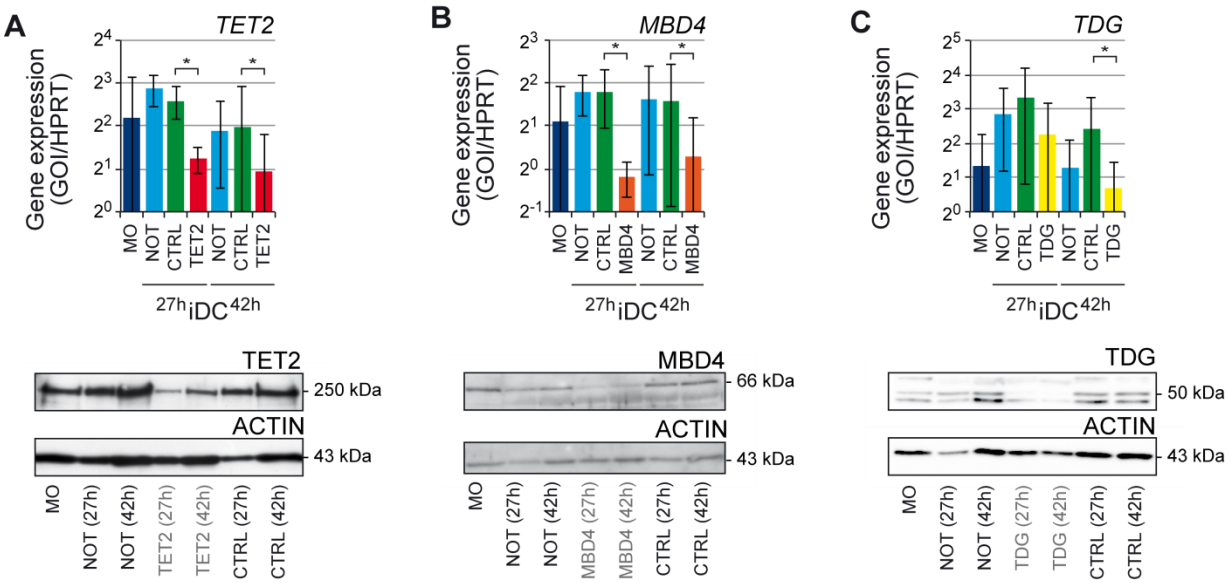

Figure 3.9: Knockdown efficiency on mRNA and protein level
mRNA (upper panels) and protein expression (lower panels) of (A) *TET2*, (B) *MBD4*, or (C) *TDG* in monocytes left untreated or transfected with the corresponding TET2-, MBD4-, TDG-, or control siRNA after 27 h and 42 h of differentiation culture. qRT-PCR results were normalized to *HPRT1* expression (n≥4, values are mean ± SD, *P<0.05 Student's T-test, paired, two-sided). Protein levels of TET2, MBD4, or TDG were analyzed using Western blotting (results are representative of *n*=3 independent experiments).

(Figure 3.9 is published in (Klug *et al.*, 2013); qPCR was performed by S. Schmidhofer and M. Klug; Western blotting was done by S. Schmidhofer.)

3.1.2.2.4 Effect of candidate knockdown on methylation kinetics

Since knockdown efficiency of all candidates reached a sufficient level after 27 h and 42 h of transfection we analyzed the effect of siRNA transfection on the demethylation process at both time points. Bisulfite treated DNA isolated from the transfected cells was measured by mass spectrometry and is displayed in Figure 3.10A as heatmaps showing methylation levels for each CpG in the DMRs. Asterisks indicate selected CpGs illustrated as bar charts in Figure 3.10B.

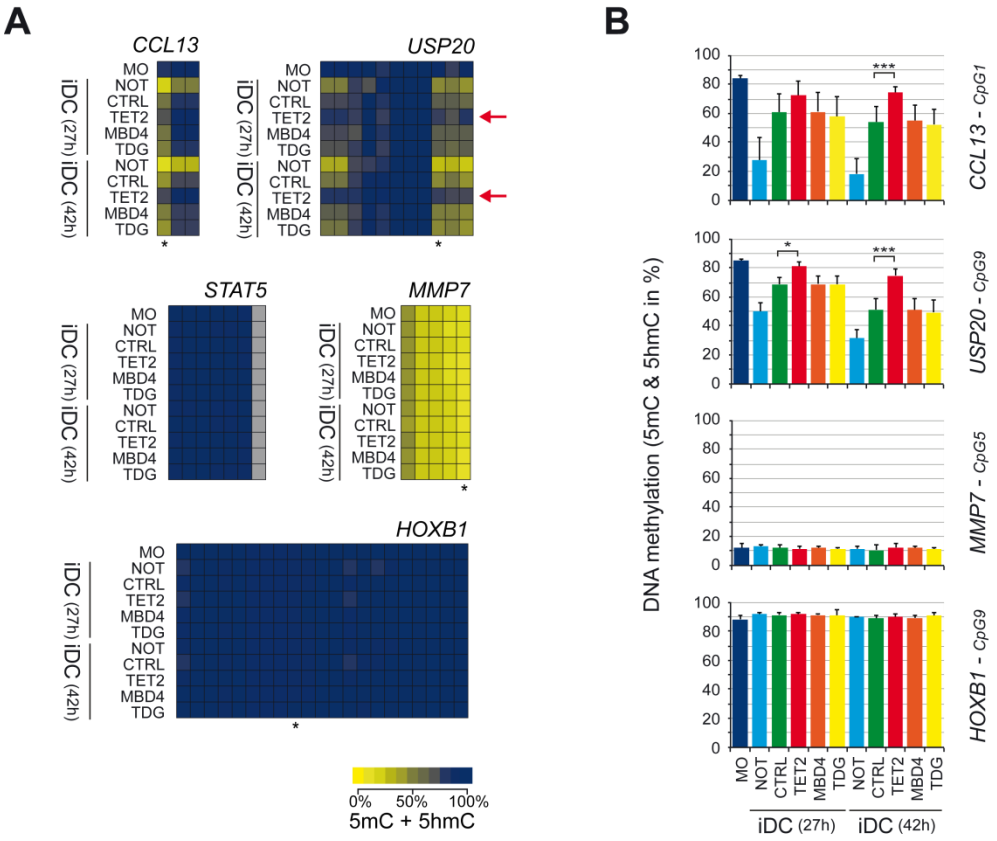

Figure 3.10: TET2 is required for active demethylation in human monocytes
(**A**) MassARRAY analysis of bisulfite-converted DNA at five loci that show active DNA demethylation during monocyte to DC differentiation, as well as for two control regions (values are mean of *n*≥4). Data are presented as heatmaps. The methylation content (including both 5mC and 5hmC) is indicated by coloring (yellow: no methylation, dark blue 100% methylation) with each box representing a single CpG dinucleotide and each row representing the succession of CpGs measured. Grey boxes indicate CpGs that were not detected by MassARRAY. Asterisks mark the CpGs that are shown in (B). Red arrows mark TET2-siRNA treated samples that show a specific decrease in demethylation. (*DNase1L3* methylation didn't change during the first 42h, but spectra were of low quality and are not shown.). Methylation ratios of single CpG units for individual donors are provided in Table S4 in Additional File 4 (available online, Klug *et al.*, 2013). (**B**) Bar charts for MassARRAY results of the indicated CpG residues of actively demethylated (*CCL13*, *USP20*) or control loci (*MMP7*, *HOXB1*). Values are mean ± SD, (*n*≥4; * P<0.05, *** P<0.001 Student's T-test, paired, two-sided).

Both regions characterized by early loss of methylation (*CCL13* and *USP20*) showed significantly higher levels of methylation in TET2 depleted cells after 27 h and/or 42 h compared to cells treated with control siRNA. This indicates a delay of demethylation after TET2 depletion in both "early" regions. In contrast the "late" regions *Stat5* and *DNase1L3* displayed no loss of methylation at 27 h and 42 h time points (due to low spectra quality the results for *DNase1L3* are not shown.). These findings are in line with results shown above. Those two regions get demethylated at late stages of the monocyte to iDC differentiation and thus don't show changes in methylation at these early time points. As expected, in the control regions *MMP7* and *HOXB1* no changes in methylation were observed.

Cells with reduced levels of MBD4 and TDG displayed similar levels of methylation compared to control siRNA and no significant changes in demethylation rates could be detected (Figure 3.10B). In all siRNA transfected cells a slight delay in demethylation compared to untreated cells could be observed. This seems to be a method-based effect and is in line with the deferred differentiation detected in 3.1.2.2.2. Nonetheless, the demethylation process was clearly demonstrated as well as the effects of TET2 knockdown which still allows addressing questions on the demethylation process using this method.

These experiments show that TET2 is required for the active demethylation process in monocyte to DC differentiation, whereas no involvement of MBD4 and TDG in the first step of 5mC to 5hmC conversion could be detected with this method.

(Figure 3.10 is available online in (Klug *et al.*, 2013). MassARRAY experiments for TET2 were performed by S. Schmidhofer and M. Klug. MBD4 and TDG were analyzed by S. Schmidhofer.)

3.1.2.2.5 5caC or 5fC accumulation at DMRs

Bisulfite treatment of DNA does not only fail to discriminate between 5mC and 5hmC but does also not distinguish other 5mC modifications (5fC and 5caC) from unmodified cytosines. To detect a possible accumulation of 5fC or 5caC at DMRs in cells expressing low levels of MBD4 or TDG this method is not suitable. In a restriction digest based approach we used *MspI* to measure restriction efficiencies after knockdown at single restriction sites covering one demethylated CpG dinucleotide each. The restriction enzyme is inhibited by the presence of 5fC and 5caC (Ito *et al.*, 2011) and detection of PCR products after digestion may indicate evidence for 5fC or 5caC at the DMRs. In Figure 3.11B the cutting efficiency of *Msp*I after depletion of MBD4 and TDG as well as TET2 at indicated regions is displayed. All DNAs were digested at similar levels at all regions and thus did not show significant differences in restriction efficiency of *Msp*I.

To test if 5fC and 5caC generally appear at DMRs we digested DNAs from dendritic cells at day 7, a late time point of the demethylation process. The results are illustrated in Figure 3.11A. All regions were cut with similar restriction efficiencies, even at the *CCL13* locus where 5mC as well as 5hmC are almost completely removed at that time point and the demethylation process is likely to have reached completeness.

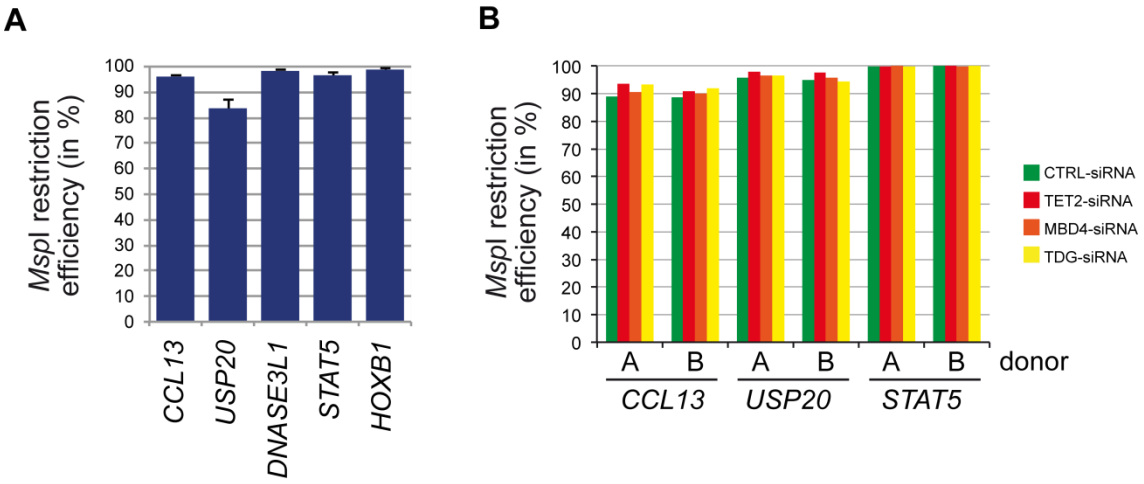

Figure 3.11: *Msp*I restriction efficiency at demethylated CpG residues
DNA from the indicated untreated (iDC at 7 days of culture, in A) or siRNA transfected cells (27h time point, in B) was subjected to MSRE restriction digests as outlined in chapter 5.2.2.7.3.1. qPCR was applied to all samples to quantify the amplification of fragments amplicons across single *MspI* sites. Amplification values for the indicated target regions for *MspI*-digested and control-digested samples were normalized against values obtained for a region in the *GAPDH* locus lacking an *MspI* restriction site. Cutting efficiency is calculated by subtracting the ratio (in percent) of normalized values for *MspI*-digested and control-digested samples from 100%.

Restriction efficiency was similar in all DNAs tested, indicating no accumulation of 5fC and 5caC at the CpGs in the restriction sequence. It has to be noted that this method only covers one CpG per region and may therefore not represent the modification state of all CpGs in the regions. Furthermore restriction efficiencies may also depend on methylation of the outer cytosine of the restriction site (5mCCGG). In this case restriction efficiency was not inhibited at all, but nevertheless this method may only give a first hint. Other methods have to be used to get further insights into modification states after knockdown of candidate enzymes.
(Figure 3.11 is available online (Klug *et al.*, 2013), S. Schmidhofer performed the experiments.)

3.1.2.2.6 Effect of candidate knockdown on 5hmC kinetics

Upon TET2 depletion 5mC demethylation rates were clearly delayed. This finding indicated an important contribution of TET2 to the demethylation process. To test whether TET2 is responsible for the conversion of 5mC to 5hmC we checked if 5hmC kinetics were affected as well. Based on pull-down using the hMeDIP method, 5hmC levels were measured by qPCR of DNA from transfected cells after 27 h and 42 h. Figure 3.12 illustrates 5hmC levels at the indicated demethylated regions. All DMRs showed either a tendency or significant reduction of 5hmC in cells with reduced TET2 expression compared to control siRNA transfected cells. The control regions *MMP7*

and *HOXB1* showed minor levels of 5hmC which presumably represents background noise since 5mC levels in Figure 3.10 showed no demethylation at the analyzed time points. As expected, due to the lack of methylation differences, 5hmC levels of cells transfected with either MBD4 or TDG siRNA did not show significant differences in comparison with control siRNA treated cells.

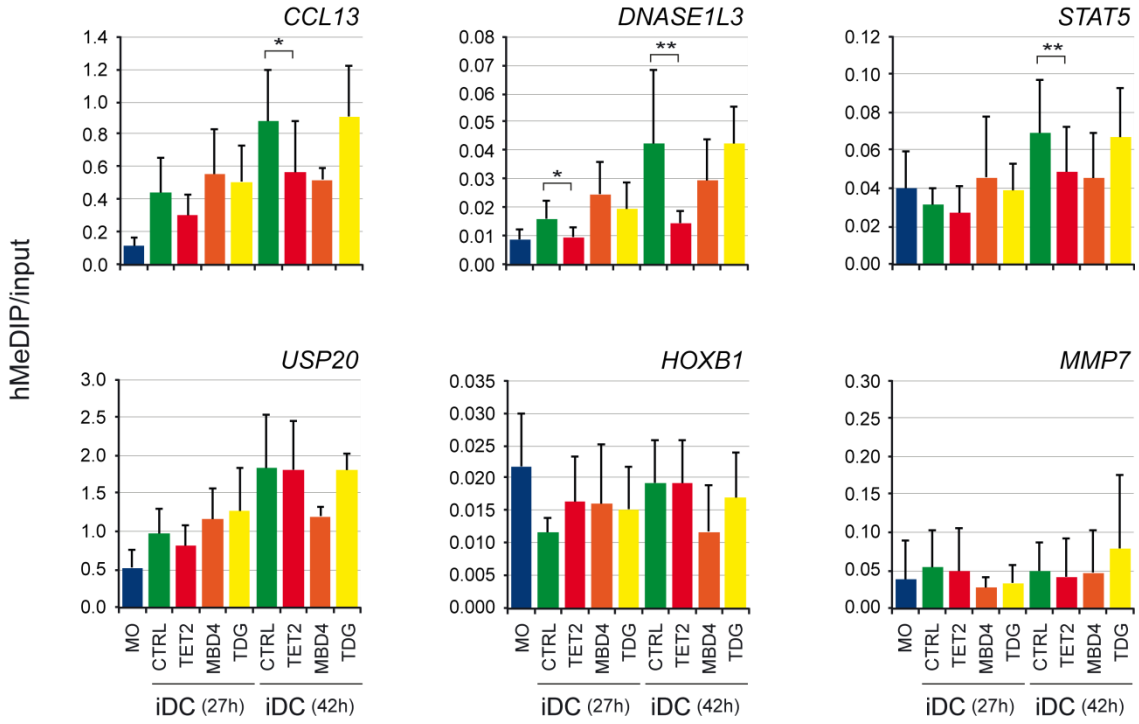

Figure 3.12: 5hmC levels after siRNA-mediated knockdown of TET2, MBD4 and TDG.
DNA from the indicated siRNA transfections was subjected to hMeDIP and the enrichment of demethylation targets and control regions was measured using qPCR (*n*=4, values are means ± SD; * P<0.05, ** P<0.01 Student's T-test, paired, two-sided).

To verify the results 5hmC was also quantified with glycosyl transfer followed by glycosylation-sensitive restriction (GGSD) with *Msp*I and *Hpa*II. Data is illustrated in Figure 3.13 for four DMRs and one control region (*MMP7* lacks restriction site.). Significant reduction of 5hmC levels in TET2 siRNA treated cells was detected at the *DNase1L3* and *USP20* locus, whereas *CCL13* and *Stat5* displayed non-significant tendencies only. 5hmC levels in MBD4 and TDG siRNA transfected cells were again not significantly altered.

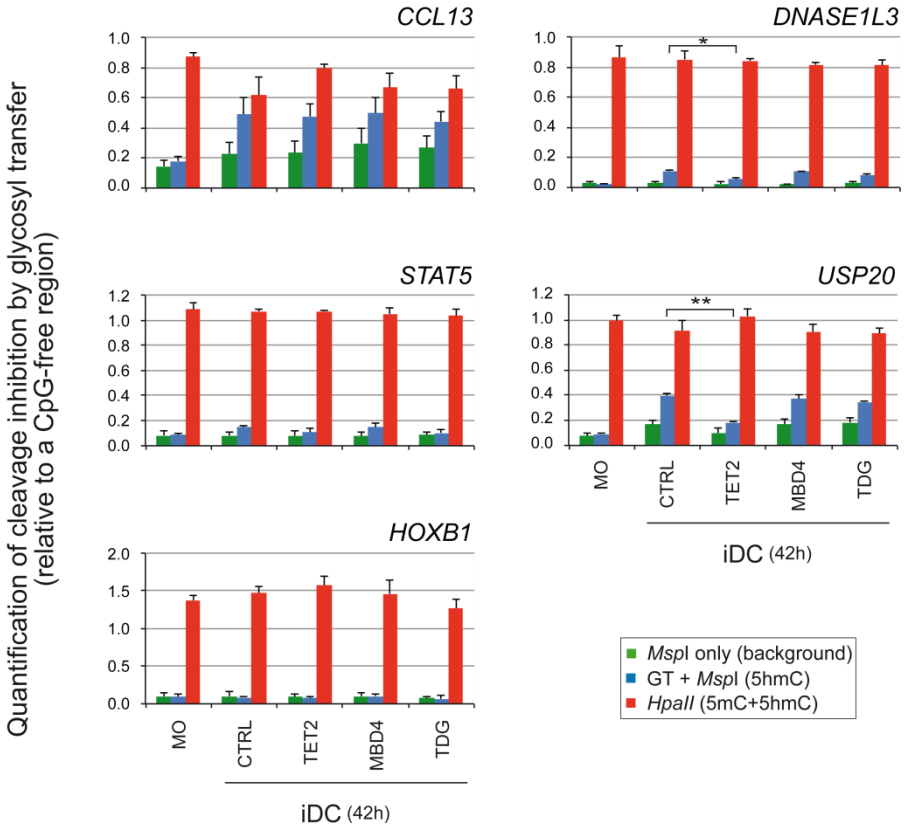

Figure 3.13: Alternative 5hmC measurements after siRNA-mediated knockdown of TET2, MBD4 and TDG

Quantification of 5hmC using glycosyl transfer to 5hmC followed by glycosylation-sensitive restriction (GGSD) with *Msp*I and subsequent amplification of protected regions. DNAs were also digested with *HpaII*, which is sensitive to both 5hmC and 5mC. qPCR results are shown relative to the amplification of a CpG-free region. Values represent mean ± SD (*n*=3) for control digests (*Msp*I only), digests after glycosyl transfer (GT+ *Msp*I) or digests with *Hpa*II (average for DNAs digested before and after glycosyl transfer), which reflects the presence of both 5mC and 5hmC. Exact genomic positions of analyzed CpG residues are given in Table S2, Additional File 2 (available online, Klug *et al.*, 2013). The *MMP7* region lacks *Msp*I/*Hpa*II recognition sequences and was not analyzed.

These results clearly indicated TET2 to be responsible for the conversion of 5mC to 5hmC whereas MBD4 and TDG are not important for this conversion step.

(Figure 3.12 and Figure 3.13 are available online (Klug *et al.*, 2013). hMeDIP experiments were performed by S. Schmidhofer and M. Klug. S. Schmidhofer conducted GGSD experiments.)

3.1.2.3 Knockdown experiments of possible factors involved in active DNA demethylation (Part II)

So far the first step in the active DNA demethylation mechanism in differentiating monocytes could be unraveled and TET2 was found to be responsible for the conversion of 5mC to 5hmC. Since the succeeding steps of the process still remained

unclear we further focused on the characterization of two other possible candidate enzymes involved in the demethylation process in the second part of this chapter.

Two recent publications presented a number of DNA readers that are able to recognize and physically bind to specific DNA modifications, such as 5mC, 5hmC, 5fC and 5caC (Iurlaro *et al.*, 2013; Spruijt *et al.*, 2013). The authors showed that the recognition of the 5mC oxidation derivatives was target- and cell type-specific. A list of specific readers for 5hmC included several DNA glycosylases (Neil1, Neil3) as well as helicases (HELLS, Harp, CHD4) and we hypothesized that a direct recognition of 5hmC by specific readers may be initiating the final excision steps of demethylation (Klug *et al.*, 2013). Helicases are involved in DNA repair (Brosh, 2013) and DNA repair was shown to play an important role in active DNA demethylation mechanisms (Wu and Zhang, 2014). Since HELLS was independently identified as a 5hmC reader by both studies, it represented a suitable candidate enzyme.
As was outlined in the introduction, TET proteins are able to regulate transcription activity by the recruitment of interacting proteins, like chromatin-modifying enzymes (Chen *et al.*, 2013; Deplus *et al.*, 2013; Williams *et al.*, 2011; Yildirim *et al.*, 2011). One possible candidate is the OGT enzyme, which has previously been identified to be recruited by all three TET proteins (Chen *et al.*, 2013; Deplus *et al.*, 2013; Vella *et al.*, 2013). It *O*-GlcNAcylates TET1 post-translational and is a crucial factor for TET1 protein stability (Shi *et al.*, 2013). In a complex with TET proteins it is involved in transcriptional activation (Chen *et al.*, 2013; Deplus *et al.*, 2013).

In order to study a potential role in the active DNA demethylation process, we used a siRNA-based approach to deplete OGT and Hells, respectively, in differentiating monocytes as described above (3.1.2.2).

3.1.2.3.1 mRNA expression levels of OGT and HELLS in differentiating monocytes

To get a first overview of the availability of OGT and HELLS we performed mRNA expression analysis in differentiation time courses using qRT-PCR to quantify mRNA levels. The results are illustrated relative to the expression of a housekeeping gene (*HPRT1*) in Figure 3.14. Both *OGT* and *HELLS* are abundantly expressed at every time point to similar amounts, while overall *OGT* is expressed at a higher level compared to *HELLS*. Although HELLS levels were considerably lower than OGT levels, both

enzymes had reasonable potential to be involved in the DNA demethylation process and were therefore further investigated.

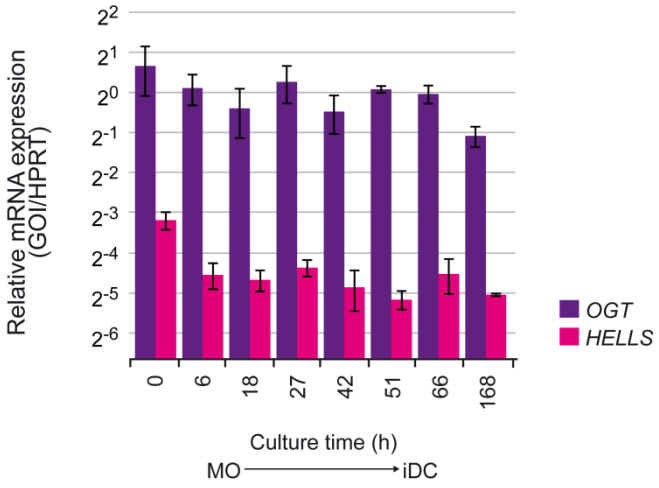

Figure 3.14: *OGT* and *HELLS* are expressed in monocytes
Quantitative expression levels of *OGT* and *HELLS* in differentiating monocytes and dendritic cells are shown. RT-PCR levels are shown relative to *HPRT1* expression. Values are mean ± SD (*n*=2-4).

3.1.2.3.2 Cell conditions and knockdown efficiency after siRNA transfection

After 42 h transfection, cell condition was surveyed. The cells were stained with DAPI, CD1a and CD14 antibodies and analyzed by flow cytometry. The results are shown in Figure 3.15 and are similar to those shown earlier in Figure 3.7 and Figure 3.8. The cells were mostly viable (>97%) and untreated cells were in advanced stages of DC differentiation after 42 h cell culture. All cell populations treated with siRNA, displayed comparable numbers of cells in a monocytic state (upper left box in panels) and cells entering the DC differentiation pathway (upper right box in panels). Compared to the untreated cells (left panels) they showed a delay in differentiation kinetics.

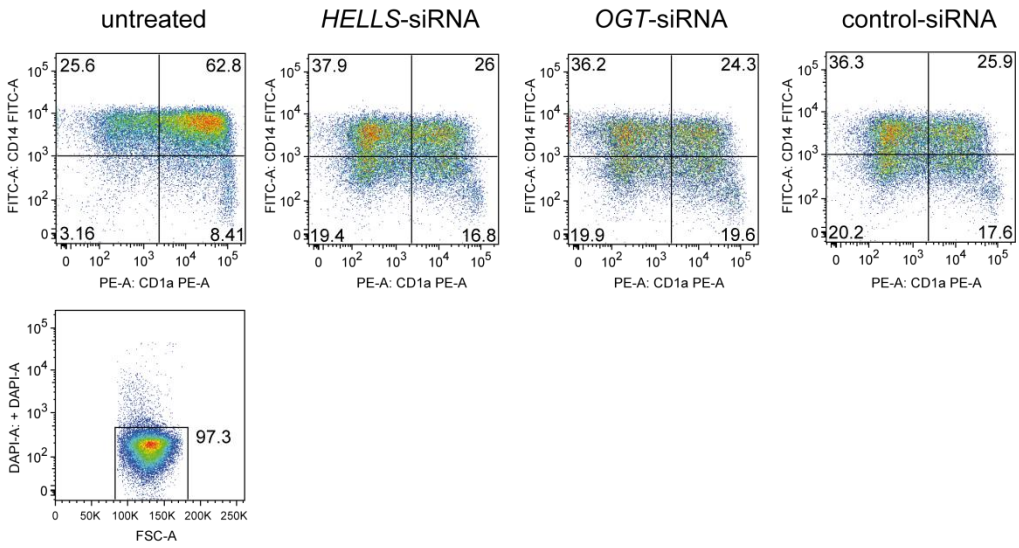

Figure 3.15: Survival of siRNA transfected monocytes
Untreated or siRNA-transfected cells after 42h were stained with DAPI, anti- CD14-Fitc and anti-CD1a-PE and analyzed by flow cytometry using the LSRII from BD (Heidelberg, Germany). The large majority of cells (>97%) was viable (lower panels) and entered the DC differentiation pathway (upper panels).

On mRNA level the expression of the DC marker CD1A and of a control region *CCL13* was also found to be consistently reduced in transfected versus untreated cells (illustrated in Figure 3.16). This is in accordance with the flow cytometry data.

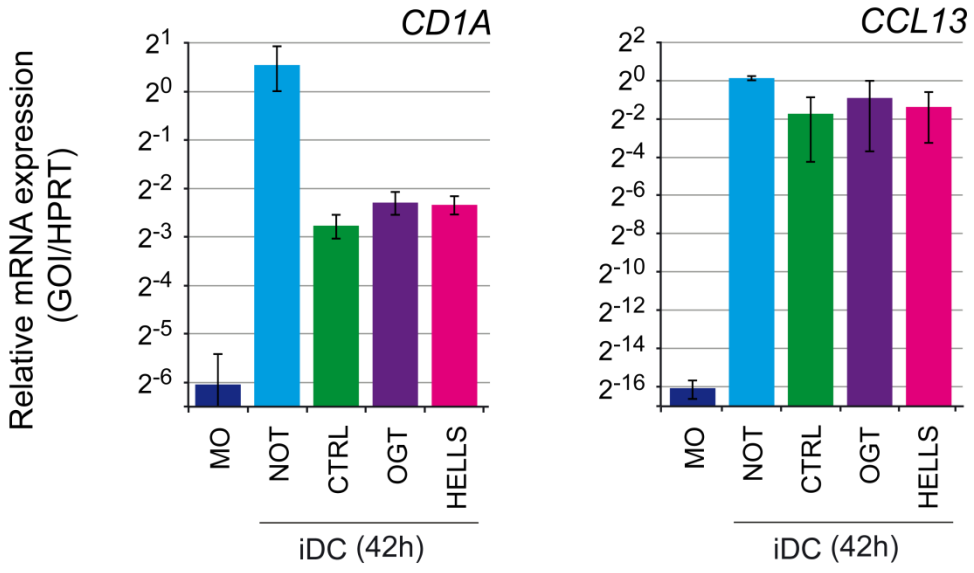

Figure 3.16: Marker gene expression of siRNA transfected monocytes
Expression of *CD1a* and *CCL13* marker genes in differentiating monocytes after 42 h post-transfection. Cells were either not transfected (NOT) or transfected with OGT-, HELLS-, or control siRNA and qRT-PCR results were normalized to *HPRT1* expression. Values represent mean ± SD (*n*=2).

The data confirmed that in contrast to untreated cells the transfection has an impact on differentiation, but does not alter the overall tendency towards iDC maturation and also cell survival was mostly unaffected.

Figure 3.17 displays mRNA levels of the candidate enzymes after siRNA-mediated depletion in monocytes and differentiating cells (either transfected or untreated) relative to the expression of a housekeeping gene (*HPRT1*). OGT siRNA transfection shows tendencies of 43% to 65% of knockdown, whereas HELLS levels are significantly depleted after siRNA transfection compared to HELLS levels in cells transfected with a control siRNA (63% to 71% knockdown).

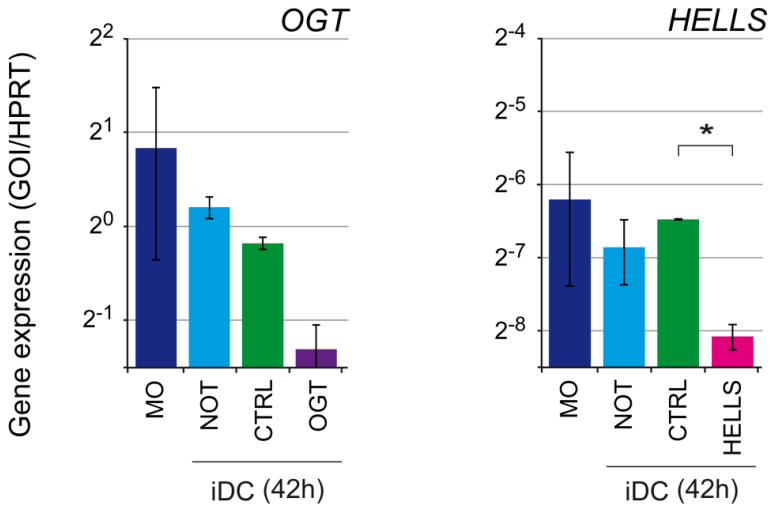

Figure 3.17: Knockdown efficiency on mRNA level
mRNA expression of candidate genes in monocytes either non-transfected (NOT) or transfected with OGT-, HELLS- or control siRNA was measured after 42 h of differentiation culture. qRT-PCR results were normalized to *HPRT1* expression (*n*=2, values are mean ± SD, * P<0.05 Student's T-test, paired, two-sided).

Taken together all siRNA transfection experiments achieved good knockdown efficiencies, while showing only minor effects on the differentiation potential of these cells.

3.1.2.3.3 Effect of candidate knockdown on 5mC and 5hmC kinetics

To test if OGT or HELLS are involved in the active demethylation process we analyzed the effects of siRNA transfection on methylation levels. Therefore bisulfite converted DNA from cells harvested 27 h or 42 h after transfection was measured by MALDI-TOF (MassARRAY EpiTYPER system). Methylation levels are shown for selected CpGs as bar charts.

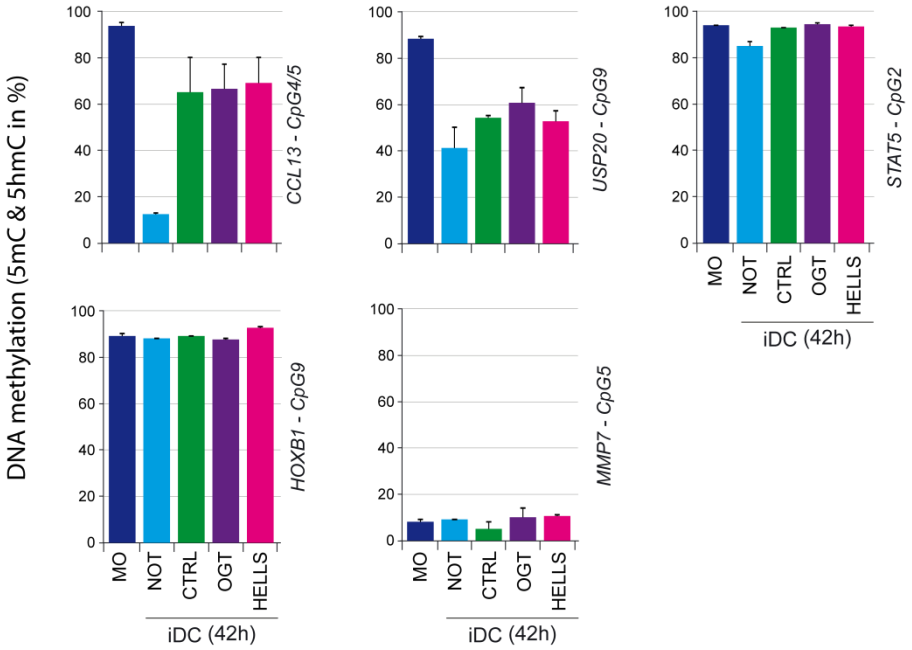

Figure 3.18: 5mC levels after siRNA-mediated knockdown of OGT and HELLS
MassARRAY analysis of bisulfite-converted DNA at five loci including two control regions (*MMP7*, *HOXB1*) is shown. Selected CpGs are presented as bar charts. Values are mean ± SD (*n*=2).

In both OGT and HELLS-depleted cells the methylation levels are similar compared to control siRNA treated cells. All siRNA treated cells show delayed demethylation at the indicated time points, which is a result of the transfection protocol, as was observed earlier (see chapter 3.1.2.2.4).

The data presented indicates that decreased levels of neither OGT nor HELLS influence the methylation pattern at the analyzed DMRs. As mentioned earlier, the MassARRAY method does not distinguish 5mC from 5hmC. To rule out 5mC detection problems, we measured 5hmC levels using hMeDIP at 27 h and 42 h time points and quantified 5hmC in the DMRs by qPCR. Figure 3.19 summarizes the results. OGT as well as HELLS depletion showed no 5hmC changes in any analyzed region compared to control siRNA treated cells. This is in line with the results for 5mC levels in those regions. The results indicate that neither OGT nor HELLS are involved in the 5mC to 5hmC conversion step. However, the question if OGT and HELLS play a role in the removal of further demethylation intermediates (5fC, 5caC) cannot be addressed with this method. The MassARRAY system does not detect those modifications, which appear unrecognized in the C fraction. To test the involvement of OGT or HELLS in the excision of the modifications, for example a restriction-based efficiency assay with a 5fC- and 5caC-sensitive enzyme can be performed.

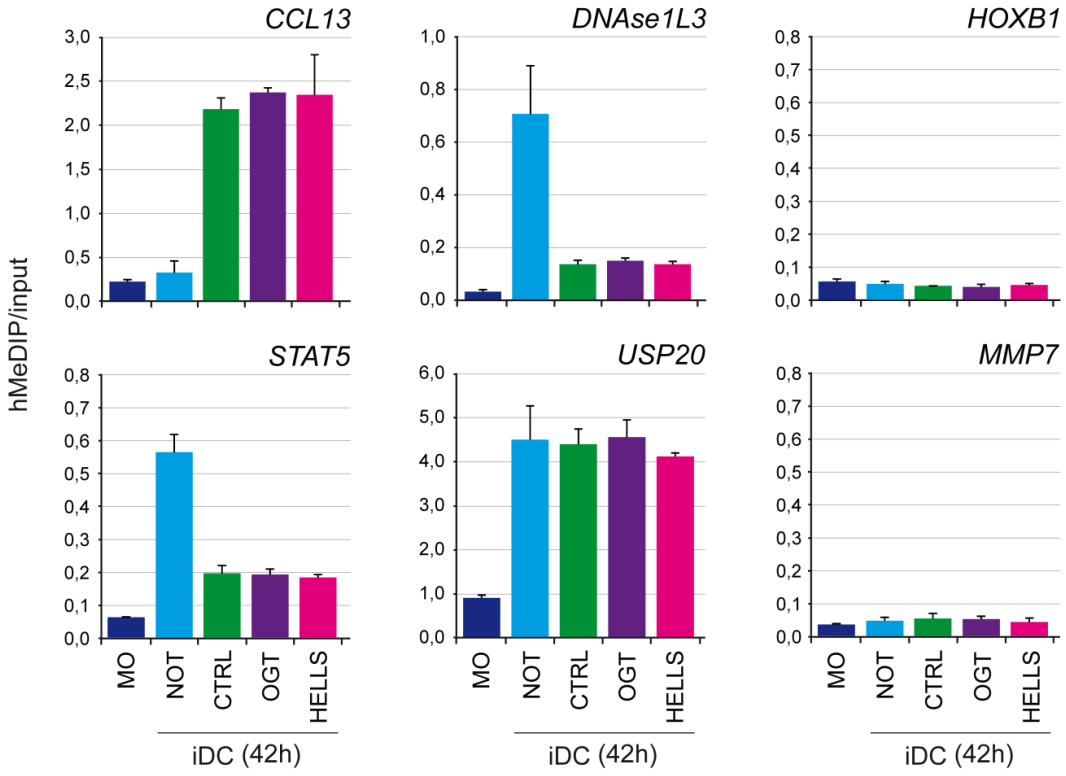

Figure 3.19: 5hmC levels after siRNA-mediated knockdown of OGT and HELLS
DNAs from transfected cells 42 h post transfection were measured with hMeDIP and the enrichment of selected regions and control regions was quantified using qPCR (*n*=2, values are mean ± SD)

With the methods used, we were not able to show an involvement of OGT or HELLS in the active DNA demethylation process. Nevertheless other approaches may be better suited to identify or rule out a role of those enzymes in this process.

3.1.2.4 Detection of 5fC and 5caC in differentiating monocytes

In the process of active DNA demethylation, TET proteins do not only convert 5mC to 5hmC, but in addition are also able to further oxidize those modifications to 5fC and 5caC (He *et al.*, 2011; Ito *et al.*, 2011; Pfaffeneder *et al.*, 2011). Following this, the resulting modifications are excised by TDG and finally replaced by cytosine via the BER (Cortazar *et al.*, 2011; Cortellino *et al.*, 2011; He *et al.*, 2011; Maiti and Drohat, 2011; Zhang *et al.*, 2012b).

The low abundance of 5fC and 5caC (Ito *et al.*, 2011; Pfaffeneder *et al.*, 2011) complicates the investigation of those intermediates in the active demethylation process. Common detection methods for cytosine modifications like the MassARRAY system are based on bisulfite conversion, and are not able to distinguish 5mC from 5hmC or 5fC and 5caC from C. To overcome those issues, a variety of new methods have been published in the last three years (Lu *et al.*, 2013; Raiber *et al.*, 2012; Song

et al., 2013). One method, published by Song *et al*. in 2013, provided a solution for both issues at the same time (Song *et al.*, 2013). Utilizing a specific, chemical treatment the fCAB-seq method detects 5fC at single-base resolution. In untreated cells the 5fC, 5caC and unmodified cytosines are converted by bisulfite treatment and appear indistinguishable in the C fraction. In cells treated with the chemical *O*-ethylhydroxylamine ($EtONH_2$) 5fC is selectively modified and thus protected from bisulfite conversion. In the readout it shifts to the 5mC fraction and can be distinguished by comparing untreated to treated cells. To measure 5fC in our cell system we adjusted this bisulfite-based method to our MassARRAY EpiTYPER system (fCAB-epi, see chapter 5.2.2.7.3.2). To verify the reproducibility of the adjusted method we designed control oligos, each containing only two unmodified CpG sites or two CpGs carrying specific modifications (5fC, 5caC, 5mC, 5hmC).

Figure 3.20 shows 5mC levels of oligos untreated or treated with $EtONH_2$ and measured by the MassARRAY EpiTYPER system (fCAB-epi). As expected the oligos containing 5mC and 5hmC modifications were detected as fully methylated, and the CpGs of the unmodified oligo showed no methylation at all. The untreated and therefore unprotected 5fC-oligo displayed ~25% of methylation, which indicated incomplete bisulfite conversion of the 5fC modification. Although the protocol was followed exactly as recommended by the authors (Song *et al.*, 2013) the bisulfite conversion did not proceed to completeness. Nevertheless, after treatment with $EtONH_2$ the 5fC-oligos were protected from bisulfite conversion and appeared in the 5mC fraction.

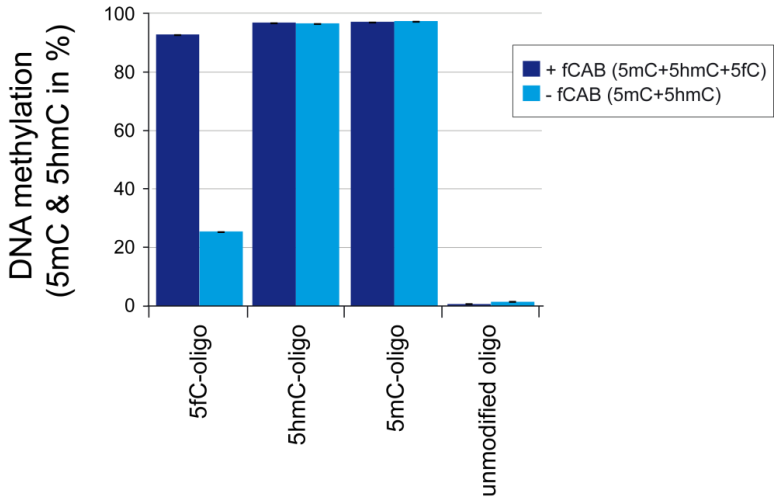

Figure 3.20: 5mC levels of modified and unmodified control oligos
MassARRAY measurement (fCAB-epi) of bisulfite-converted synthetic DNA oligos containing two CpG sites with 5fC, 5hmC, 5mC modifications or unmodified cytosines (unmodified oligo). The bar chart illustrates averaged DNA methylation (5mC+5hmC) of two CpGs in each oligo with (dark blue bars) or without (light blue bars) fCAB treatment. Values are ± SD (*n*=2).

In the experiments presented in Figure 3.21A we measured a region showing early, active DNA demethylation in a time course of differentiating monocytes. The DNAs were treated with the fCAB method or left untreated before bisulfite conversion and MassARRAY analyses. The results show a small increase of the methylation levels in the fCAB-treated samples accounting for 5mC, 5hmC and in contrast to the untreated samples also 5fC. To check if the finding was reproducible we analyzed DNA of differentiating monocytes 42 h after TDG depletion. TDG is responsible for excision of 5fC at regions harboring this intermediate during demethylation, and a depletion of this enzyme results in the accumulation of 5fC (Shen *et al.*, 2013; Song *et al.*, 2013). Therefore, it is easier to detect 5fC in samples deprived of this glycosylase. Figure 3.21B illustrates 5mC and 5hmC levels of two actively demethylated regions measured with or without chemical 5fC protection (+/- fCAB treatment). Both regions showed inconsistent methylation levels which is probably due to technical variation. The level of variation was similar to expected changes resulting from the scarce modification 5fC, indicating that the method has to be further adjusted to be able to detect the expected low levels of 5fC. First of all, since 5fC is only present at low abundances the bisulfite conversion efficiency has to be optimized to be able to detect any changes at a high resolution. Furthermore the data presented resulted from pioneering experiments in DNA from only one donor. Thus it is necessary to repeat those experiments to confirm the results. (Due to exhaustion of TDG-depleted DNA samples we didn't have enough material to repeat the experiments.)

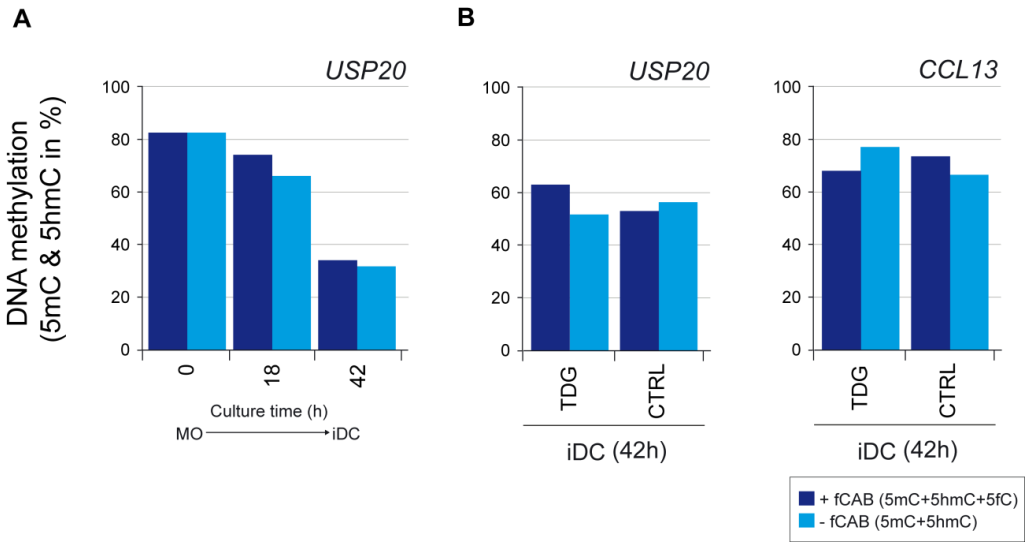

Figure 3.21: 5fC detection in differentiating monocytes
MassARRAY analysis (fCAB-epi) of bisulfite-converted DNA from differentiating monocytes is shown in a time course (**A**) or after 42 TDG siRNA transfection (**B**). 5mC and 5hmC levels of DNA with (dark blue) or without (light blue) fCAB treatment are depicted as bar charts (*n*=1).

3.2 Functional characterization of active DNA demethylation at DMRs

Using ChIP-seq data of several hematopoietic cell types our group recently identified several promoter-distal DMRs in the myeloid lineage which are associated with cell type specific transcription factor binding, e.g. PU.1. Corresponding DNA methylation profiles in the DMRs showed a direct correlation between demethylation processes and local binding of major hematopoietic transcription factors. Both events co-occurred at the same stages of macrophage differentiation (Pham *et al.*, 2013). The synchronous appearance of the enhancer mark H3K4me1 at those regions indicated a possible enhancer function and led to the hypothesis that methylation levels regulate the usage of those elements. In this context the master transcription factor PU.1 seems to be necessary for demethylation of those cell type specific elements and thus facilitates the accession for further proteins (Pham *et al.*, 2013).
In order to characterize the DMRs functionally and to test a possible link between function, methylation, and cell type specific TFs further experiments were performed and are presented in the following chapters.

3.2.1 Reporter Assay reveals enhancer function of DMRs

To test the ability of those promoter-distal regions to enhance the activity of a promoter we performed luciferase reporter assays. This method is a common tool to study the regulatory capacity of DNA elements to activate promoter activity in a heterologous context. In brief, the regions of interest are arranged on a plasmid upstream of a basal heterologous promoter EF1α and a luciferase gene. This plasmid is transfected into the cells of interest and the expression of the luciferase gene is analyzed. Our group developed a reporter vector that lacks CpG dinucleotides (Klug and Rehli, 2006), which we used to design reporter plasmids containing the DMRs of interest. Since previous results hint at a methylation dependent activity we constructed and tested both *in vitro* *Sss*I-methylated and unmethylated reporter plasmids. Two control-constructs,

containing the basal EF1α promoter (pCpGL-EF1) or both the EF1α promoter and CMV enhancer (pCpGL-cmv) were used as negative and positive control.

3.2.1.1 Reporter plasmid activity in THP-1 and HeLa cells

For the transfection experiments we chose the human myeloid cell line THP-1 to test the possible enhancer activity of the DMRs in a heterologous MO and MAK context.
Since monocytes are fragile and prone to in-vitro stimulation due to their sensitive immune cell nature they are difficult to transfect and suffer from severe impairment by the transfection. On the other hand the THP-1 cell line does not only share a human, myeloid background but is also easy to transfect at sufficient quantities. Furthermore, stimulation of THP-1 with PMA (phorbol 12-myristate 13-acetate) drives the cells into a monocyte- and macrophage-like phenotype where the cells become adherent and stop proliferation (Abrink *et al.*, 1994; Tsuchiya *et al.*, 1982). Thus questions regarding enhancer activity in MO and MAK can be addressed in this cell system.

THP-1 cells were transfected and luciferase activities were measured as described above. Normalized levels of luciferase measurements from three independent experiments are depicted as green bars in Figure 3.22. Of all tested constructs only one (ADAP1) showed significantly enhanced activation of the basal promoter compared to the promoter alone (significant elevation of luciferase levels compared to the basal vector pCpGL-EF1 is not shown in this figure). The lack of significance in other constructs may be caused by considerable luciferase level variation between experiments. Since the mean over all experiments is used to calculate the statistics, visible differences in single experiments are likely to be masked by the deviation. We therefore also considered constructs which showed significant enhancer activation compared to the basal vector in 2 out of 3 single experiments. In this case two additional regions (PIM3, CACNA1C) fulfilled this criterion.
On average (shown in the left panel) the enhancer activity significantly decreased in two regions (ADAP1, PIM3), when plasmids were methylated before transfection. One methylated construct (*ADAP1*) still showed significant luciferase activity compared to the basal vector pCpGL-EF1, indicating that methylation influences the activity level of the construct but does not suffice to completely abolish reporter activity. The *PIM3* construct lost the ability to enhance the basal EF1α promoter when methylated, indicating that its enhancer activity is dependent on the CpG methylation status. Methylation of the CACNA1C construct did not alter its potential to enhance the activity

of the basal promoter, indicating either a methylation-independent activity or the disposal of the mark by enzymes present in the THP1-cells.

In the right panel the averaged activity of the methylated constructs compared to the corresponding, unmethylated constructs is shown. The illustration confirms the results shown in the left panel. It displays a significantly reduced enhancer activity for both, ADAP1 and PIM3 (but not CACNA1C) in the presence of methylation relative to the unmethylated construct.

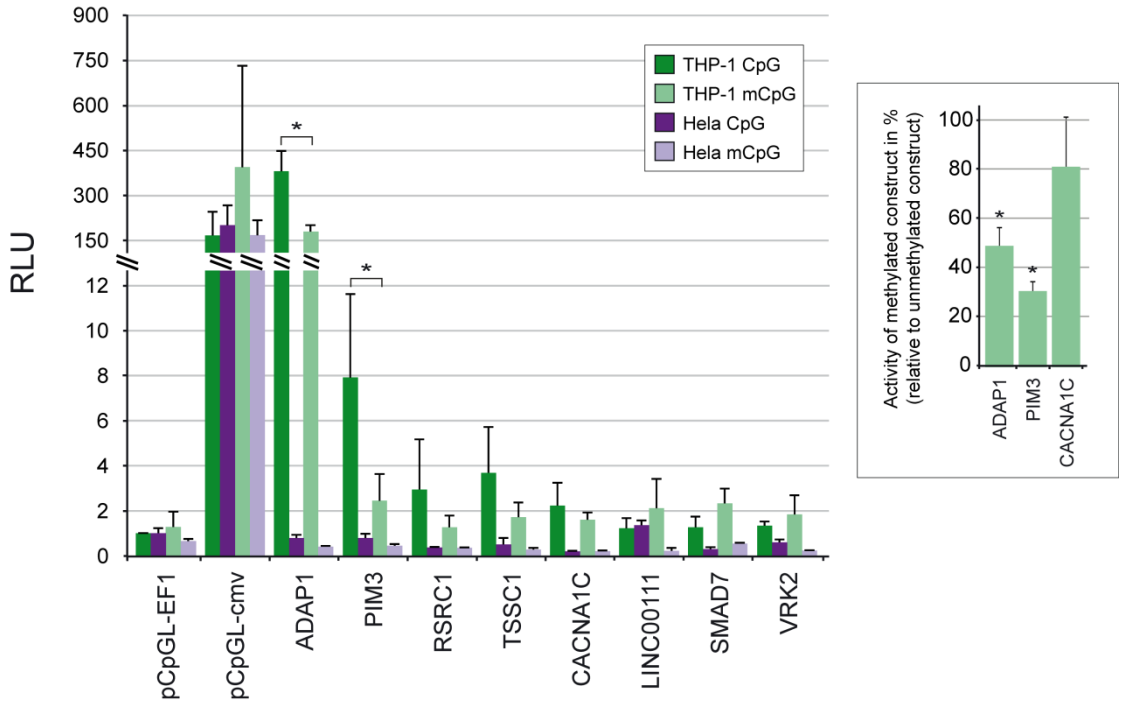

Figure 3.22: Enhancer activity in DMRs
Eight DMRs were cloned upstream of a basic EF1-promoter into the CpG-free luciferase vector pCpGL. The indicated plasmids were in *vitro Sss*I-methylated (mCpG, light colors) or unmethylated (CpG, dark colors) and transiently transfected into THP-1 (green bars, *n*=3, mean values are ± SD, biological replicates) or HeLa cells (purple bars, *n*=1, mean ± SD of 3 technical replicates). Luciferase activity was normalized against the activity of a cotransfected *Renilla* construct and values ± SD are shown (left panel) relative to the unmethylated pCpGL-EF1 construct. Asterisks indicate significant difference between methylated and unmethylated plasmids ($P < 0.05$, paired Student's t-test). The right panel shows averaged, relative activity of methylated compared to the unmethylated constructs (100%) in percent. Only plasmids which displayed significant activity compared to the basal vector in at least 2 out of 3 single experiments are shown. Values are mean ± SD, (*n*=3, $P<0.05$ Student's paired T-test). Significance between constructs and basal vector pCpGL-EF1 is not shown in this figure.

The results above identified three regions with potential enhancer function in reporter assays, in the THP-1 cell line.

Based on first analyses of the DMRs (Pham *et al.*, 2013) our group proposed that demethylation at promoter-distal elements as well as enhancer usage may be cell type specific. In addition the published results suggested a cell type-specific role of TFs in the correct utilization of relevant genomic elements.

To check if a transcription factor set resembling a monocyte- and macrophage phenotype is able to activate a different set of our constructs we differentiated THP-1 cells with PMA as described in section (5.2.1.2.1.1) and repeated the reporter assays.

The results are displayed in Figure 3.23 and show by trend elevated basal promoter activation, compared to the EF1α promoter alone, for four regions (*RSRC1*, *TSSC1*, *CACNA1C*, *LINC00111*). Except for the positive control vector, none of these observations were significant. Due to variable luciferase levels between the experiments we calculated the statistics within the single experiments. All four regions were able to significantly activate the basal promoter, compared to the promoter alone, in every single experiment. One additional region (PIM3), which showed enhancer activity in 2 out of 3 experiments, was also included. Methylation of plasmids prior to transfection had no significant impact on enhancing properties in the experiment average when compared to corresponding unmethylated constructs. Single experiments, though, displayed visible reduction of activity in methylated compared to unmethylated constructs. This finding is further supported by an illustration of methylated construct activities shown relative to the activity of unmethylated regions in the right panel of Figure 3.23. All five regions depicted in the chart show significantly reduced reporter activity when methylated prior to transfection.
This way of illustrating the results relative to their reference construct is independent of technical variation between the experiments. It confirmed that the averaged values calculated for the experiments are actually biased by method deviation. It further supported the finding that in both experiments different regions showed their enhancer potential depending on the cell context. In the case of *CACNA1C* the different results in regard to the methylation sensitivity of the region hint at a cell type-specific removal of the methylation, which is also dependent on cell type-specific conditions.

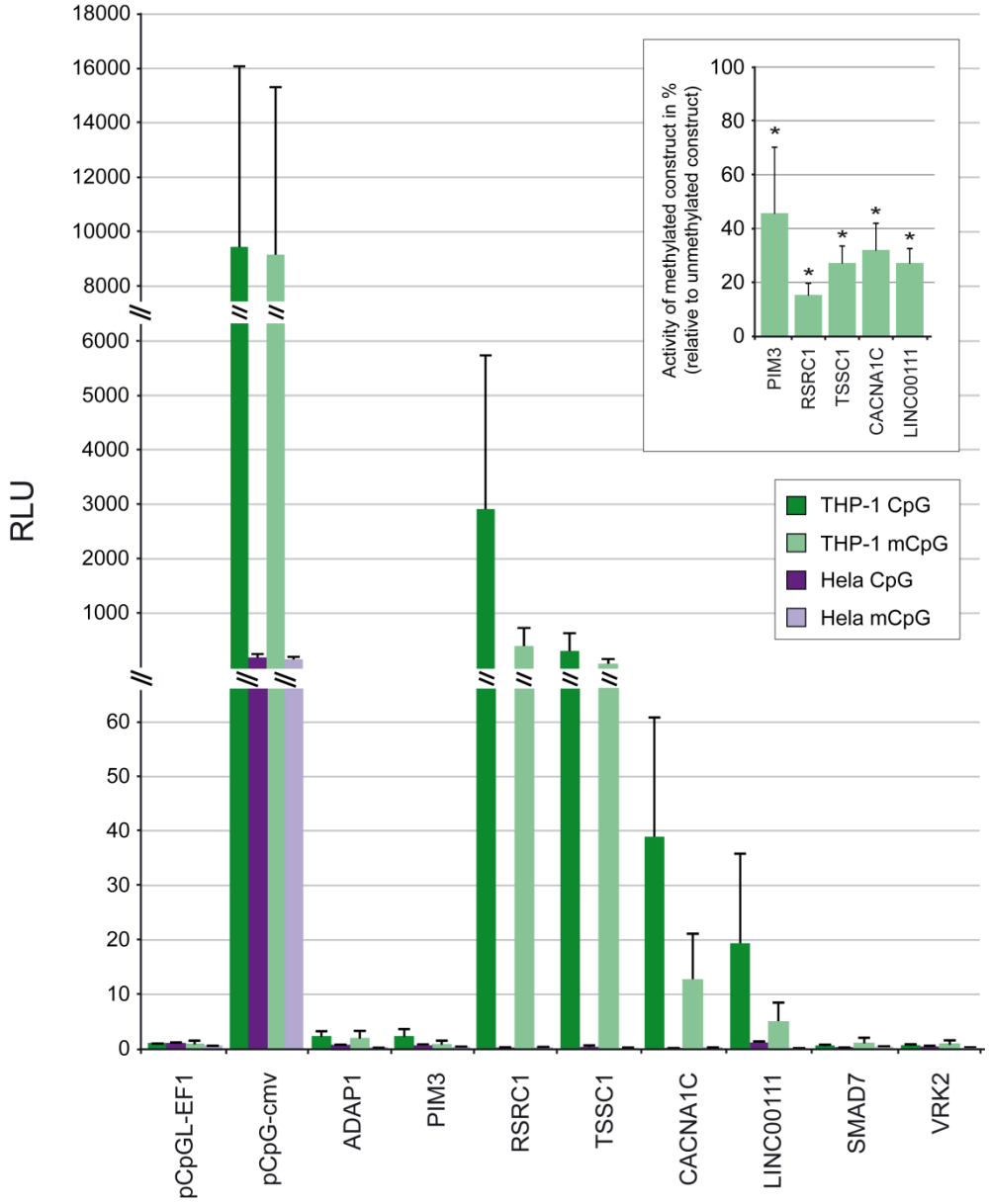

Figure 3.23: Enhancer activity in DMRs in the presence of PMA
Eight DMRs were cloned upstream of a basic EF1α-promoter into the CpG-free luciferase vector pCpGL. The indicated plasmids were in *vitro Sssl*-methylated (mCpG, light colors) or unmethylated (CpG, dark colors) and transiently transfected into THP-1 (green bars, *n*=3) that were stimulated with PMA 12h after transfection or HeLa cells (purple bars, *n*=1). Luciferase activity was normalized against the activity of a cotransfected *Renilla* construct and mean values ± SD of 3 biological replicates (THP-1) and 3 technical replicates (HeLa) are shown relative to the unmethylated pCpGL-EF1 construct. Significance between constructs and basal vector pCpGL-EF1 is not shown.

Taken together these results suggest a cell type specific opening and activation of regulatory promoter-distal elements. This led to the question if TFs present in the cells play a role in activation of those regions. Since most regions featured PU.1 binding closely located to the DMRs (Pham *et al.*, 2013) we repeated the experiments in a non-hematopoietic human cell line (HeLa, derived from cervical cancer cells) expressing no endogenous PU.1 (Kwon *et al.*, 2005).

The results are presented in Figure 3.22 (and in Figure 3.23) and show no significant activation of the basal promoter EF1α by any of the constructs except for the positive control pCpGL-cmv. This finding is in line with the hypothesis that only a cell type-specific repertoire of TF is able to activate distal regulatory elements (Heinz *et al.*, 2010; Pham *et al.*, 2012) and further discussed in chapter (4.1).

3.2.2 Active demethylation in differentiating MAK

To study a general causal connection between active demethylation and cell type specificity in promoter-distal regions we first measured methylation levels of additional time points in differentiating monocytes at previously defined sites of DNA demethylation. The cells were cultured under appropriate conditions to obtain iDC or MAC, respectively. DNAs were bisulfite converted and quantified using the MassARRAY system (Sequenom). Results of four representative regions are shown as histograms (blue squares) (in Figure 3.24) displaying methylation level time courses of one representative CpG each. The figure clearly shows a subsequent loss of DNA methylation in all indicated CpG dinucleotides during the differentiation into MAK which is in line with previous findings (Pham *et al.*, 2013).
The demethylation of those regions was not restricted to macrophages but also visible in differentiating iDC. Two DMRs, ACP2, and RSRC1, displayed an equal demethylation pace whereas three out of six regions showed delayed demethylation compared to MAK.

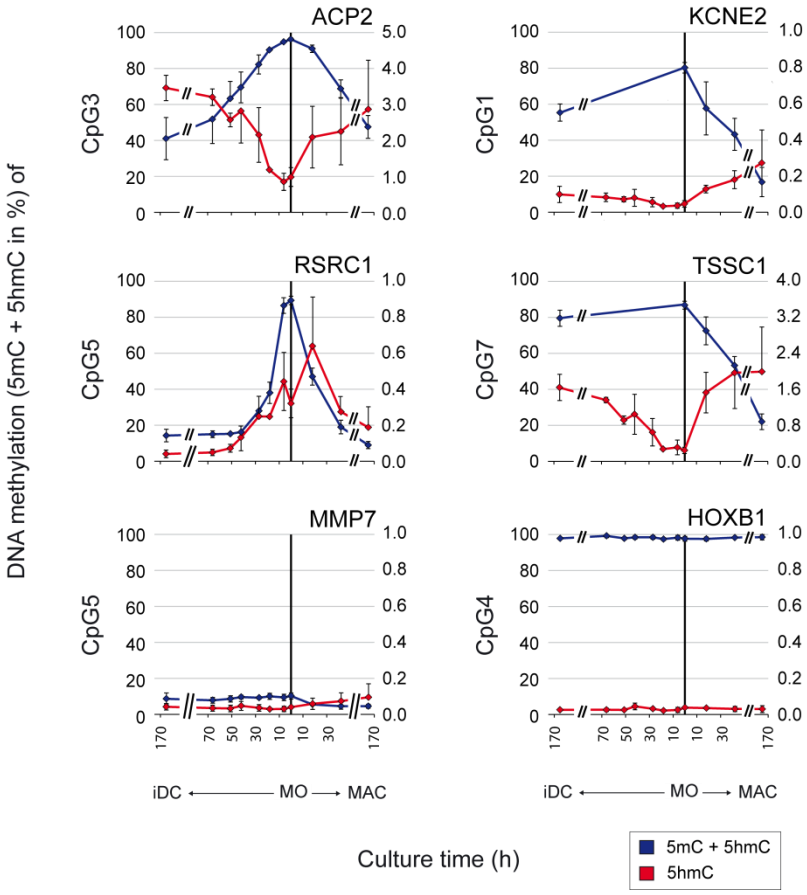

Figure 3.24: 5mC deposition precedes active DNA demethylation in differentiating monocytes
Dynamics of DNA methylation (5mC+5hmC) and 5-hydroxymethylation (5hmC) during monocyte differentiation. DNA methylation levels of single CpGs as measured by MassARRAY (blue squares, *n*≥4) are compared with 5hmC enrichment (measured by hMeDIP, red squares, *n*=2) at four DMRs (values are mean + or – SD).

The progressive loss of DNA methylation led to further investigation of the demethylation process. In order to verify an active process in those regions we studied the appearance of 5hmC at the indicated loci using hMeDIP. Four DMRs were quantified using qPCR and compared to methylation levels of single CpG dinucleotides measured by mass spectrometry. The results are displayed as red squares in the same Figure 3.24.

In differentiating MAK the experiments clearly show an increase in 5hmC levels in all tested regions, which is in line with the corresponding MassARRAY result. Whenever 5mC levels decrease 5hmC is present at the same locus. The appearance and removal of 5hmC was simultaneous to 5mC depletion in all DMRs.

The dendritic cells displayed similar patterns of loss of 5mC and simultaneous gain of 5-hydroxymethylation in five out of 6 regions. With three of them showing decelerated

demethylation kinetics compared to the same loci in macrophages although 5hmC levels increased likewise.

In order to validate the 5hmC detection and to rule out density bias of the antibody-based method we performed a chemical labeling-based method (GGSD) in one of the regions (only one PCR primer pair covered a *Msp*I site necessary for restriction digest, region *TSSC1*). Adequate amounts of 5hmC were detected by qPCR quantification (results not shown). This finding reduced the probability of antibody-based bias of the method (at least in this region). Furthermore it hinted at a constant rate of 5mC to 5hmC conversion, with slow excision kinetics at those regions, which can't be detected using the MassARRAY method.

In one of the tested DMRs we did not detect 5hmC levels relative to the background signal despite reduced 5mC levels (data not shown). Further investigation will be necessary to determine if high 5hmC turnover or other demethylation mechanisms are involved.

3.3 Global 5hmC and PU.1 distribution in differentiating monocytes

With the data presented so far we were able to confirm a functional link between cell-type specific TF binding and local demethylation events. These findings support the idea that PU.1 and other TFs are directly involved in the demethylation process (de la Rica *et al.*, 2013; Pham *et al.*, 2013; Stadler *et al.*, 2011). PU.1 is a key regulator of the hematopoietic compartment and responsible for the development of lymphoid and myeloid lineages (Friedman, 2007). In cooperation with other interaction partners and TFs it modifies surrounding chromatin structure and regulates transcription activity (for a review see Burda *et al.*, 2010; Kihara-Negishi *et al.*, 2001)

To investigate active demethylation and its association with master TFs on a global scale we performed whole genome sequencing of the first intermediate of active demethylation, 5hmC, and the master transcription factor PU.1.

3.3.1 Comparison of two whole genome 5hmC-enrichment methods

Global analysis of 5hmC can be performed using a variety of different methods available as published methods or as complete kits for purchase. The techniques range from pull-down methods, which are antibody- or chemical labeling-based to single-base resolution approaches.

For our purpose we chose to test and compare an antibody-based method hMeDIP and the Hydroxymethyl Collector™ kit, a chemical labeling-based method available from ActiveMotif. The protocols were adjusted to the next generation library preparation protocol established in our laboratory (5.2.2.7.2.3, 5.2.2.7.2.4). Sequencing was performed (as described in 5.2.2.9, 5.2.2.10) on several time points of differentiating monocytes and analyzed as described in chapter 5.2.5.

3.3.1.1 Validation of two different 5hmC-enrichment methods

To compare pull-down efficiencies of the two methods, two time points of differentiating monocytes of one donor were subjected to 5hmC-specific pull-down. The reads obtained from the library preparations were individually aligned to the hg19 reference genome. For each time point (MO and iDC) peaks were defined and all peaks obtained from both methods at one time point were combined. To estimate the pull-down efficiency, the individual read (tag) coverage at the defined peaks was estimated and compared in the scatterplots depicted in Figure 3.25. The low Pearson correlation coefficients (r) of r=-0.26 (MO) and r=-0.07 (iDC) suggest no correlation between the 5hmC tag count sets enriched by each IP method.

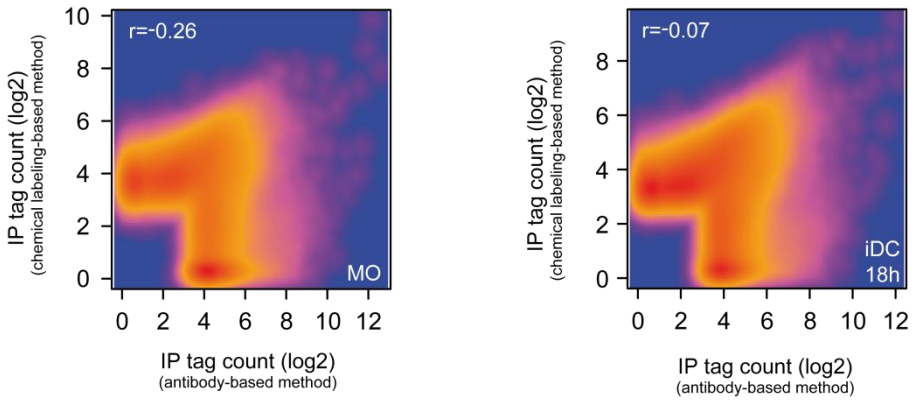

Figure 3.25: Global distribution of 5hmC tags by two different IP methods
Scatterplot representation of 5hmC tag counts from two IP methods, hMeDIP-seq (antibody-based) and Hydroxymethyl Collector™-seq (chemical labeling-based).Tag counts of peaks are compared between different time points of differentiating monocytes (MO: left panel, iDC: right panel). The colors represent the distribution of log2-transformed tag counts of all peaks found at the indicated time points (blue, orange and red represent low, medium and high values). The Pearson correlation coefficients (r) are based on log2-transformed values.

To validate the diverging 5hmC peak sets enriched by both methods we performed 5hmC-sensitive restriction digests (GGSD, 5.2.2.7.2.1) as an alternative 5hmC detection method.

We chose several peak regions detected uniquely by each of the methods and analyzed 5hmC levels in DNA samples from the same donor and differentiation time course used for sequencing. Figure 3.26 shows the result of six representative regions measured with the alternative 5hmC detection method. In the left panel unique regions enriched by the Hydroxymethyl Collector™ kit are presented. All four tested regions showed rising levels of 5hmC during the time course depicted in two exemplary regions in this figure.

5hmC in hMeDIP-seq specific peak regions (displayed in the center panels) could only be reproduced in one out of three tested regions (two of them shown in this figure). The two regions on the right hand side represent a positive control region (*USP20*) and one region, which is methylated during DC differentiation and serves as a negative control (*HOXB1*).

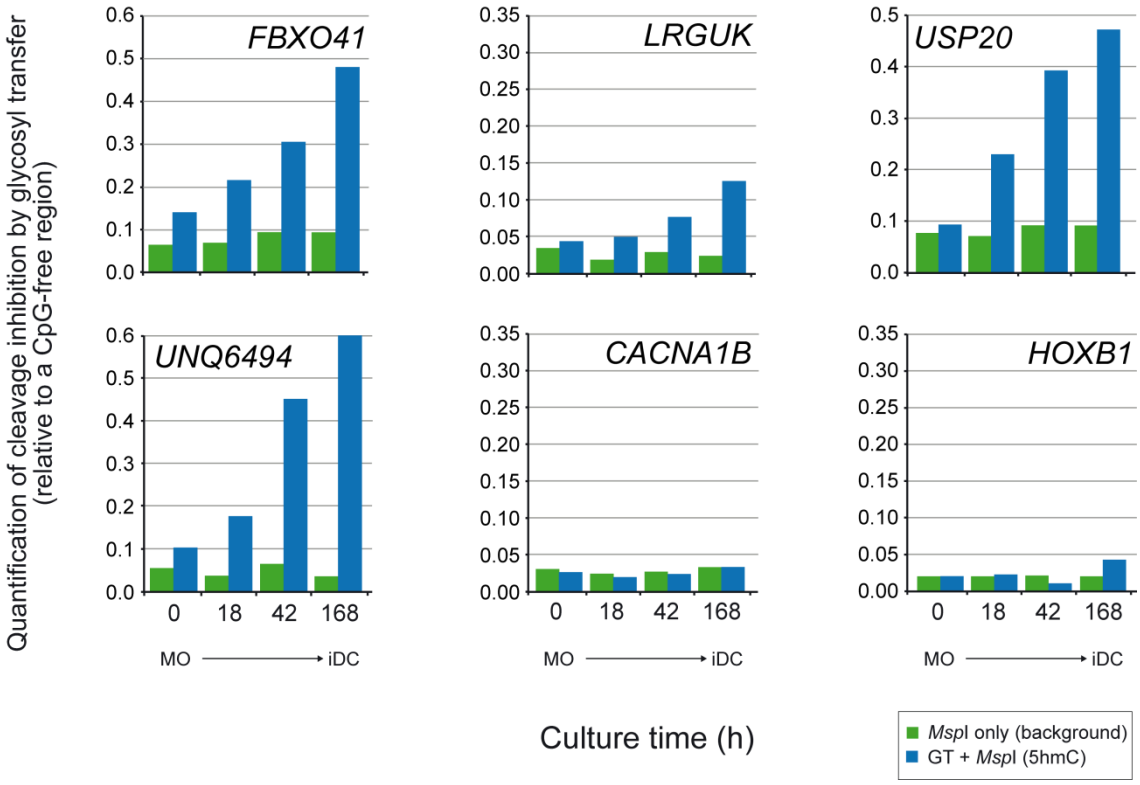

Figure 3.26: Validation of 5hmC measurements in differentiating DCs
Quantification of 5hmC using glycosyl transfer to 5hmC followed by glycosylation-sensitive restriction (GGSD) with *Msp*I and subsequent amplification of protected regions in differentiating monocytes. qPCR results are shown for control digests (background, green bars) or digests after glycosyl transfer (5hmC, blue bars) relative to the amplification of a CpG-free region (*n*=1).

In summary, only results produced by the chemical labeling-based method were reproducible with another method.

We further analyzed the reproducibility between technical and biological replicates obtained with the Hydroxymethyl Collector™ kit. The comparison of the replicates is presented in Figure 3.27. In (A) a technical replicate of the same time point is shown and both replicates are compared to biological replicates in (B). The Pearson coefficients indicate a good correlation between biological and technical replicates and confirm a stable reproducibility of the method. The higher r value (r=0.80) for the comparison of donor 1 (technical replicate 2) and donor 2 may be explained by technical variation. Samples were produced in one batch, suggesting that replicates of this method are more reproducible when generated in the same experiment. It has to be noted, that the differences between the two batches may also be resulting from protocol alterations. One component (the Biotin Conjugate Solution) was used at unknown concentrations in the procedure (in the second batch) and thus the incubation time had to be prolonged. Although subsequent qPCR analyses showed equal enrichment efficiencies for a control region compared to the first batch, interference with the results cannot be ruled out. Further replicates are necessary to clarify the

issue. Biological replicates are compared in (Figure 3.27C) and corresponding Pearson coefficients (DC18h, r=0.73 and DC 168h, r=0.77) confirm good correlation between the samples and stable reproducibility of the method.

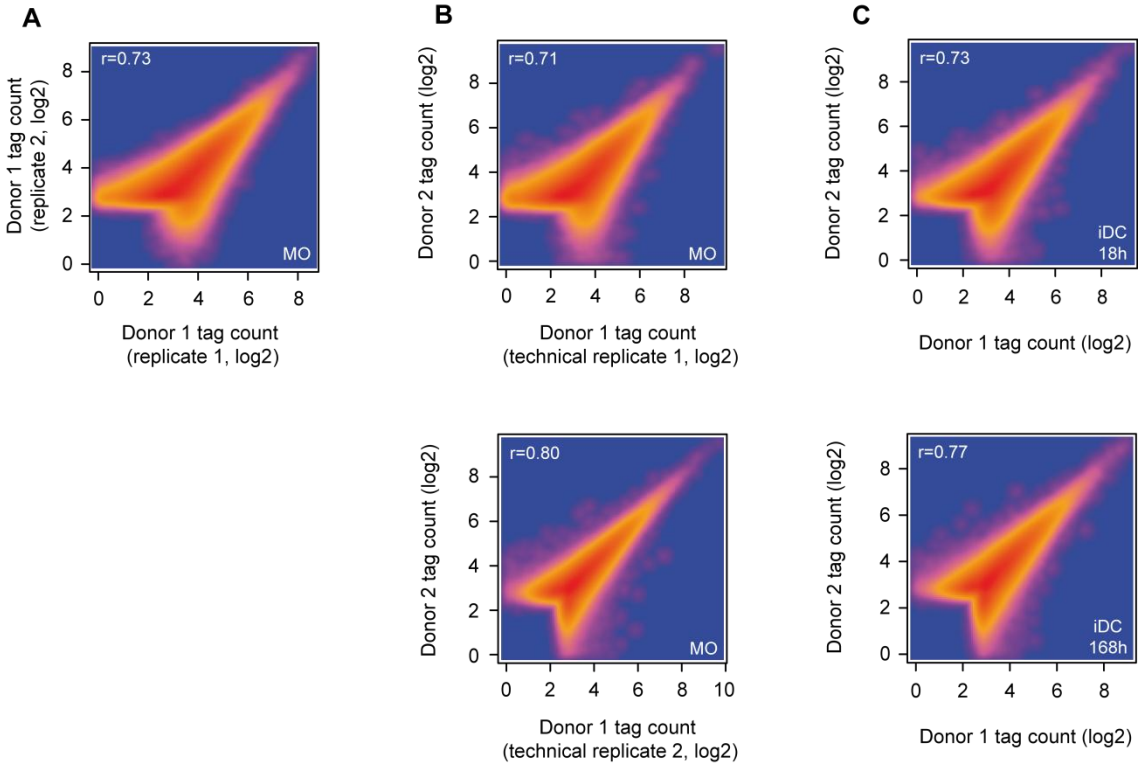

Figure 3.27: Global distribution of 5hmC tags in replicates
Scatterplot comparison of 5hmC tag counts covered by Hydroxymethyl Collector™-seq (chemical labeling-based method) of two independent experiments (donor1, 2).Tag counts of peaks are compared between technical replicates (**A**), biological and technical replicates (**B**) and biological replicates (**C**) of several time points of differentiating monocytes. The colors represent the distribution of log2-transformed tag counts of all peaks in the indicated time points (blue, orange and red represent low, medium and high values). The Pearson correlation coefficients (r) are based on log2-transformed values.

In summary the glycosyl-sensitive restriction digest experiments validated the Hydroxymethyl Collector™ method, which is characterized by high levels of reproducibility. Thus the Hydroxymethyl Collector™ method was used for the subsequent whole genome 5hmC analyses.

3.3.2 Global genome analysis of 5hmC

3.3.2.1 Genome-wide 5hmC dynamics in DC differentiation

To get a global picture of 5hmC dynamics, a complete peak file of time point-specific 5hmC peaks from two biological replicates (donor 1 and 2) was generated (5.2.5.2).

The significantly enriched differential peaks (11909 peaks) were clustered using the k-means method and grouped into four subsets. All peaks were evenly distributed among the clusters with around 2200-3300 peaks in each subset.

For each time point the averaged tag coverage was either visualized as heatmaps of genomic distance distributions (sorted by decreasing tag counts) or as histograms for the separate clusters and is shown in Figure 3.28. In addition to 5hmC peaks, corresponding PU.1 ChIP-seq tag coverage is presented in the same figures.

Figure 3.28A displays heatmaps of both 5hmC and PU.1 for three time points of DC differentiation separated into the four clusters estimated by k-means clustering. In the left part of the figure two cluster sets (#2=2597 peaks and #4=2231 peaks) show a high 5hmC enrichment in monocytes which decreases during DC differentiation (in cluster #4 to a higher degree than cluster #2). Corresponding PU.1 levels (right part of figure) display scattered patterns of binding in the surrounding 1 kb area which become more distinct upon transition to DC concentrating at the 5hmC peak center. Irregular distributions of 5hmC peaks complicate the mapping of PU.1 in their vicinity, which may in part explain the finding. Later analyses of 5hmC in the surrounding of well-defined PU.1 peaks are much more informative (see chapter 3.3.3.1). At least in cluster #4, PU.1 is present in the areas of 5hmC enrichment, indicating a connection between the two features. In contrast to clusters #2 and #4, cluster #1 (3692 peaks) and #3 (3389 peaks) both gain 5hmC upon differentiation into dendritic cells. PU.1 tag counts are almost absent in monocytes and gain only low amounts with no distinct distribution pattern over time. The same data is visualized for each cluster as histograms in Figure 3.28B showing tag coverage of 5hmC and of PU.1 ChIP-seq. In concordance with the results from Figure 3.28A the progressive depletion of 5hmC upon transition into DCs, in cluster #2 and #4, is also visible in the corresponding histograms. In cluster #4, PU.1 tags seem to accumulate in the 5hmC peak center region at the 18h time point (DC 18h) and decrease with further differentiation. The overall PU.1 levels in cluster #2 are also elevated at DC18h but show no peak in the center regions. In contrast, monocytes leave a small dip in the center region with low levels of PU.1 at the shores of 5hmC peaks. For cluster #1 and #3, 5hmC dynamics are also visible in the left panels of the histograms but again no distinct PU.1 binding patterns could be observed.

These results revealed different 5hmC dynamics in pairs of two clusters (#2/ #4 and #1/ #3) and represent events of active demethylation in cluster #2/#4, at least. In order to confirm if 5hmC levels are associated with demethylation events in the peak cluster sets, 5hmC peaks of each cluster were annotated with CpG methylation data, obtained from bisulfite-sequencing of MO and DC at day 7 (see chapter 5.2.5.9.2 for details).

Since bisulfite treatment does not distinguish between 5mC and 5hmC, the indicated methylation levels account for a combination of 5mC and 5hmC levels.

The four methylation histograms presented in Figure 3.28C show similar 5mC patterns in cluster set #2/ #4 and #1/ #3. Cluster #2 and #4 both display a major reduction of 5mC upon DC differentiation most prominent at the center of 5hmC peaks. In combination with the observed local loss of 5hmC (Figure 3.28A) the decrease in 5mC levels most likely represents DNA demethylation events at 5hmC peaks in those clusters. In contrast, cluster #1 and #3 only depict a moderate overall reduction of methylation during monocyte differentiation without local demethylation in the peak center.

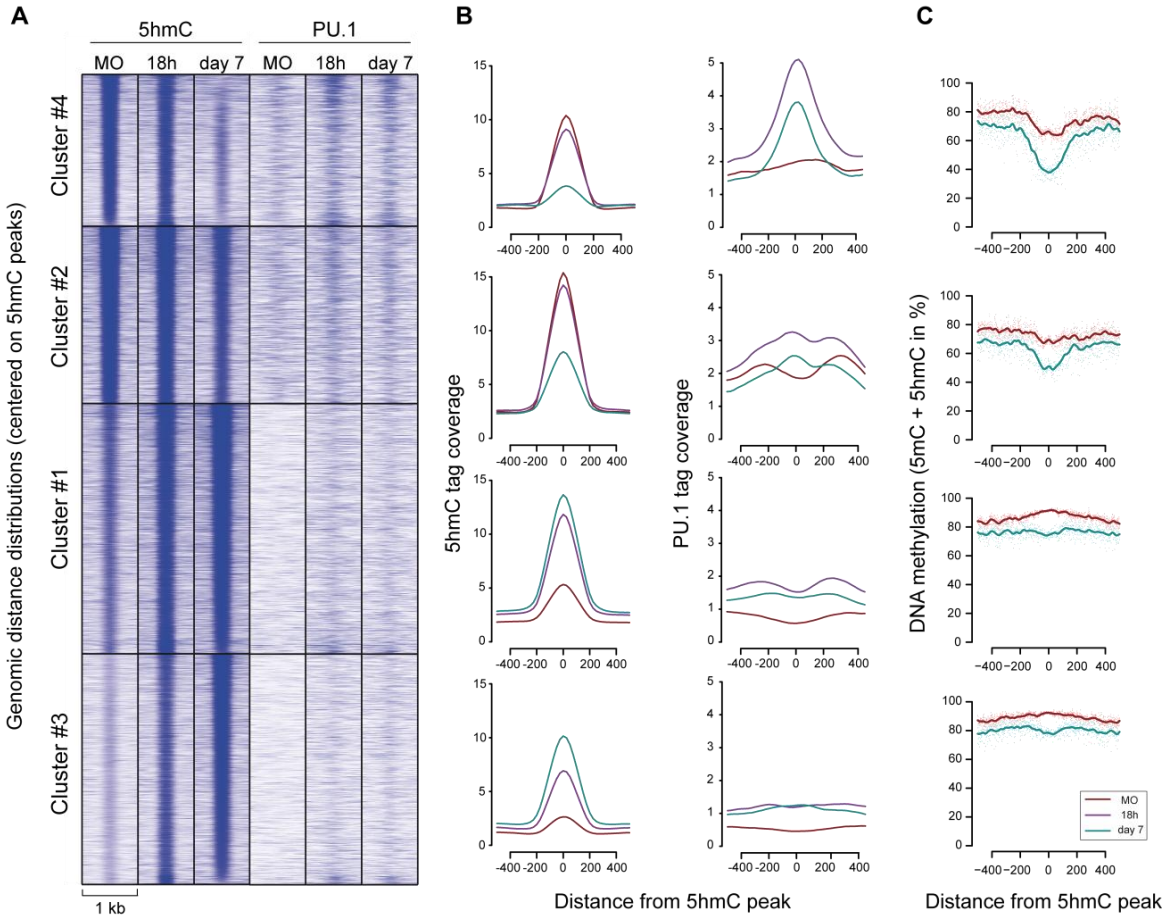

Figure 3.28: 5hmC peak dynamics and corresponding PU.1 and 5mC levels
5hmC peaks were clustered according to their Hydroxymethyl Collector^{TM}-seq profiles. Tag counts of two biological replicates were combined and annotated to associated 5hmC peaks. The genomic distance distribution of corresponding 5hmC tag counts of two biological replicates (combined, left panel) or PU.1 ChIP-seq tags (right panel) of several time points of differentiating monocytes are shown in heatmaps (**A**) or histograms (**B**) (regions 1 kb) relative to the peak center. (**C**) Histograms of methylation levels as assessed by bisulfite sequencing of differentiating monocytes (MO *n*=4, DC *n*=6). 5mC and 5hmC levels of all valid CpGs in 1kb regions of the indicated cluster sets in are shown in % of methylation relative to 5hmC peak centers.

Since bisulfite data, presented in Figure 3.28C, were generated in different laboratories, the observed methylation changes may be resulting from technical variation. To check if the demethylation events were specific for the 5hmC peak set we compared it to 5mC distribution in a set of 7899 random peaks (for details see chapter 5.2.5.3) illustrated in Figure 3.29. In A and B the genomic distance distribution of 5hmC and PU.1, respectively, is shown as heatmaps or histograms. As expected neither 5hmC nor PU.1 were enriched on a global level in differentiating monocytes. This finding supports the previous findings of 5hmC specific PU.1 accumulation in the 5hmC peak set. In Figure 3.29C methylation levels of MO and DC (at day 7) in the proximity of random peaks are shown in a histogram. The methylation dynamics in this figure are similar to the changes observed in the 5hmC specific peak set. This finding indicates that the overall loss of methylation during monocyte differentiation observed in the four 5hmC peak subsets is not a specific feature of 5hmC peaks but a global phenomenon based on methodical diversity. Nevertheless, the prominent loss of 5mC in the center of 5hmC peaks, observed in cluster #2/#4, can only be in part explained by a global loss of 5mC altogether and does account for demethylation events present at the peaks in those subsets.

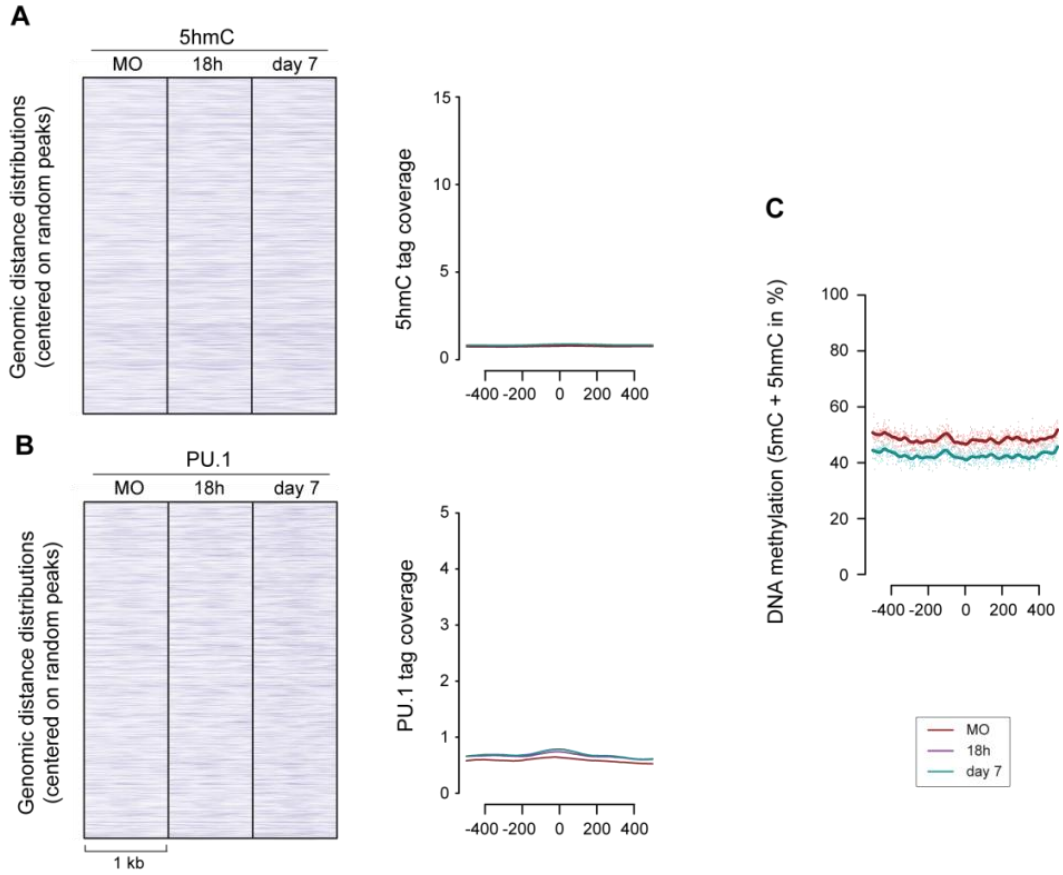

Figure 3.29: Global 5hmC, PU.1 and 5mC distributions
A set of random peaks was annotated with corresponding PU.1 ChIP-seq (**A**) or 5hmC tag counts (Hydroxymethyl Collector^{TM}-seq) (**B**) and is shown as heatmaps of genomic distance distributions or histograms in a 1 kb genomic region relative to the peak centers. (**C**) Histogram of methylation levels obtained by bisulfite sequencing at random peaks in differentiating monocytes (MO *n*=4, DC *n*=6). 5mC and 5hmC levels of all valid CpGs across a 1 kb genomic interval of the indicated random peak set (A, B) are shown in % of methylation relative to peak centers.

Taken together the results show different 5hmC- as well as corresponding PU.1- and 5mC dynamics in the four clusters. In the next chapter the functions of the cluster sets were analyzed and compared.

3.3.2.2 Gene ontology analysis of 5hmC peak cluster sets

To further characterize the biological functions of regions present in the different 5hmC clusters (Figure 3.28), the peak sets were subjected to gene ontology analyses using the Homer GO analysis tool (described in chapter 5.2.5.4).

Figure 3.30 lists significantly enriched gene ontology (GO) terms related to biological processes. The top 38 terms enriched in at least one of the four cluster sets (Figure 3.30A) are shown as heatmaps of log2 transformed GO term fold enrichment (GFE) with the corresponding generic terms and GO accession numbers.

The major fraction of enriched GO terms is related to immunological processes, which is in line with known roles of monocytes and dendritic cells. As members of the immune system they are enriched for GO terms corresponding to immune response and inflammation. DCs are involved in T cell activation, whereas endocytosis and wound healing are important functions of monocytes (Martin, 1997). A minor fraction of enriched GO terms is involved in cell signaling and hemostasis. Most terms listed in (A) were enriched in Cluster #2 and #4 only. Since the top GO terms did not contain any terms uniquely enriched in set #1/#3, only a selection of uniquely enriched terms in cluster #2/#4 is presented in Figure 3.30B. Unique terms of set #2/#4 are exclusively enriched for terms related to immune response and for one term involved in hemostasis. The observed local demethylation in the sets indicates selective opening and activation of important, cell type-relevant genes via active demethylation.

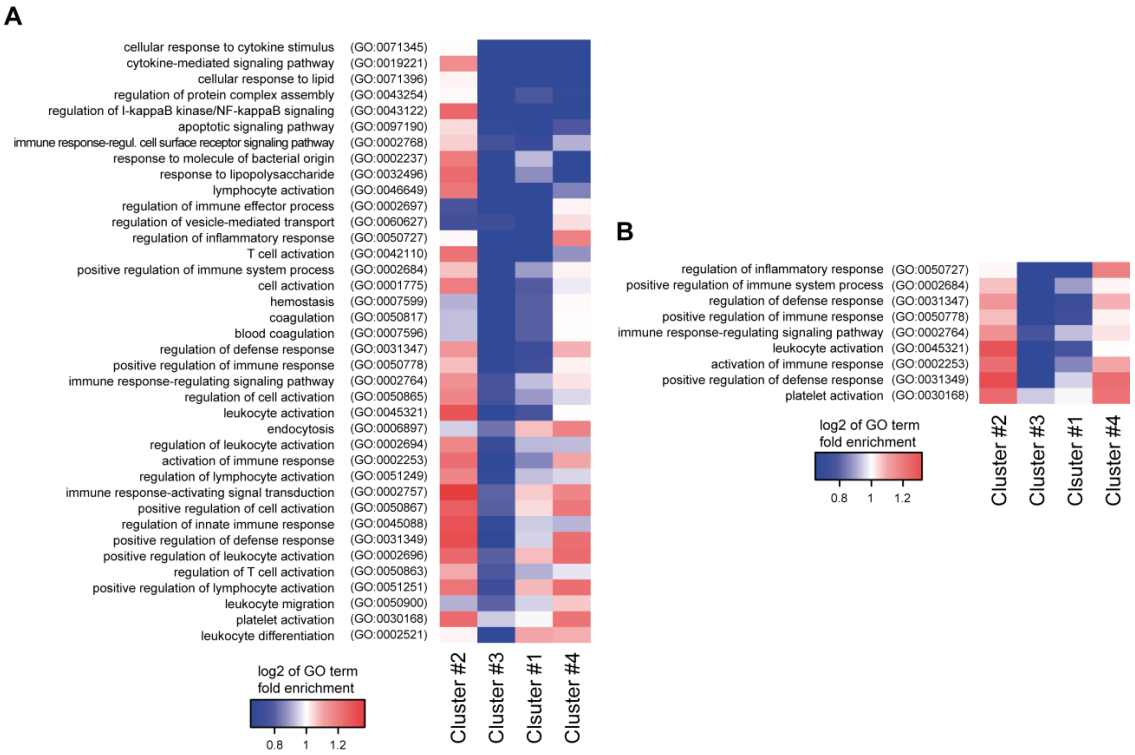

Figure 3.30: 5hmC peak cluster-based GO and pathway analyses
Heatmaps (log2 scale) of fold enrichment values for GO terms related to biological processes. Fold enrichment for GO terms was calculated for cluster-specific gene sets using the Homer tool. And log2 transformed values for each term were analyzed using the hypergeometric test and corrected for multiple hypothesis testing using the Benjamini-Hochberg method. Term enrichment values are indicated by color code intensity (blue=depletion, red=enrichment). Generic terms and GO accession number are given for the top 38 enriched terms of all cluster-specific sets (**A**) and a selection of GO terms enriched uniquely in cluster set #2 and #4 is shown in (**B**).

3.3.2.3 mRNA expression profiles in 5hmC peak cluster sets

Several groups have shown a direct link between active DNA demethylation and gene expression levels in several cell types (Mellen *et al.*, 2012; Szulwach *et al.*, 2011; Wu *et al.*, 2011a). In differentiating monocytes the loss of DNA methylation at DMRs coincides with activated chromatin and leads to gene activation in approximately 50% of the cases, including the *CCL13* promoter region (Klug *et al.*, 2010). The remaining 50% of the tested regions did not show altered mRNA expression, thus no global correlation between demethylation and transcription levels could be detected.

To test whether the observed individual 5hmC cluster demethylation dynamics are linked to expression levels, the peak sets (cluster #1-#4) were annotated (see section 5.2.5.5) with mRNA expression data obtained by whole genome microarray analyses of monocyte differentiation time courses (Klug *et al.*, 2010). For each time point-specific 5hmC peak, expression levels of the closest gene were annotated. The results are displayed as boxplots in Figure 3.31. In line with previous observations similar patterns were observed for cluster #1 and #3 as well as for #2 and #4. Cluster #1 and #3 showed a significant increase of gene expression at cluster-specific peaks upon differentiation of MO into DC. Cluster #2 and #4 display a significant reduction in transcription levels upon differentiation initiation. In summary the observed cluster-specific expression patterns indicate an expression-dependent clustering.

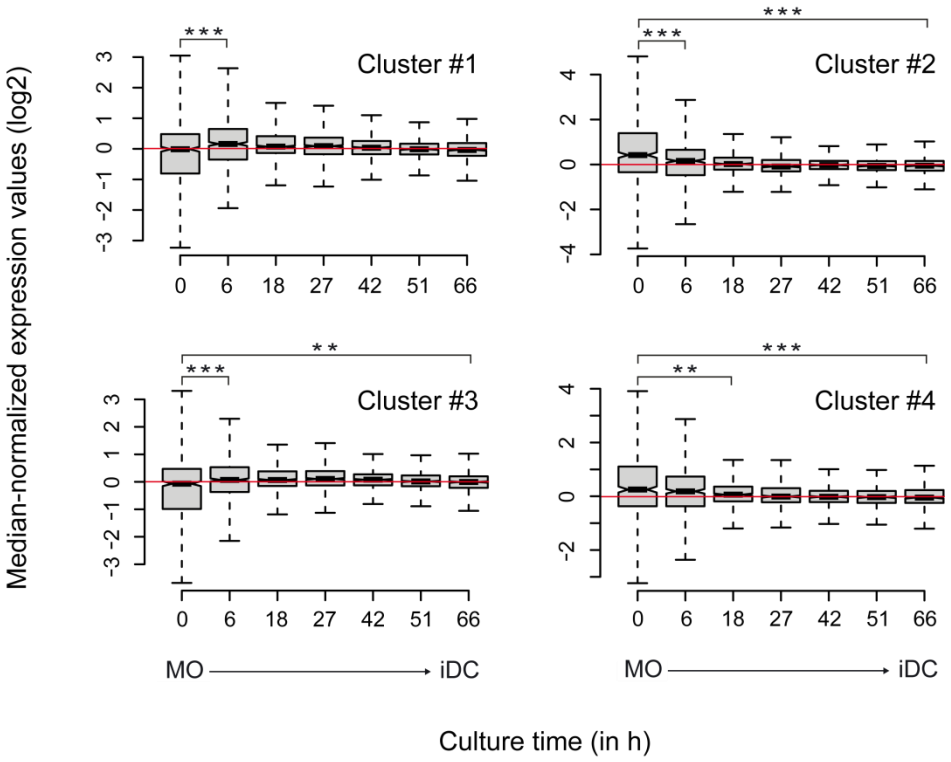

Figure 3.31: Gene expression level dynamics during monocyte to dendritic cell differentiation
Box plots showing the distribution of median-normalized mRNA expression levels for genes closest to differentiation time point-specific 5hmC peak regions. Box frames display the interquartile ranges (25%-75%) with an intersection representing the median; whiskers extend to the data extremes; outliers are not shown. The base line is plotted in red. Asterisks indicate significant difference of mRNA expression levels for MO compared to DC time points (**P < 0.01 and ***P < 0.001, Mann-Whitney-U test). Only significant differences upon transition of MO to DC and compared to differentiated iDC are shown.

3.3.3 Global analysis of PU.1 distribution

3.3.3.1 Genome-wide PU.1 distribution analyses of differentiating monocytes

To get further insights into global PU.1 binding dynamics a set of all differential, time point-specific PU.1 ChIP-seq peaks (16208 peaks) was further analyzed (5.2.5.6). Performing k-means clustering on the PU.1 peak set resulted in 6 clusters of different sizes. Illustrated as heatmaps of genomic distance distributions in Figure 3.32A two clusters, #4 and #6 (#4=3227 peaks, #6=3237 peaks), show a steady loss of PU.1 peaks in the progress of differentiation. In contrast, a constant gain of PU.1 binding during monocyte differentiation is visible in peaks of the remaining clusters. Regions at the MO time point in cluster #5 (1773 peaks) almost completely lack PU.1 binding at these sites but display a similar amount of PU.1 enrichment on day 7 of the time

course. The differences in pace and quantity of PU.1 binding dynamics is even more obvious in histograms of the same data (Figure 3.32B, left panels). Regions in cluster #3 and #2 (#3=3124 peaks, #2=2534 peaks) accumulate PU.1 binding much faster than regions in cluster #1 (2313 peaks). Whereas cluster #3 and #1 enrich more PU.1 until DC day 7, PU.1 binding at regions in cluster #2 decreases from after 18h of differentiation. These findings indicate different PU.1 binding dynamics at regions of the different cluster sets.

Among others, our group observed PU.1 binding events at differentiation time point-specific DMRs and suggested that local binding of sequence specific TFs is involved in active DNA demethylation processes in differentiating macrophages (Pham *et al.*, 2013). To check if PU.1 involvement was associated with active DNA demethylation events in differentiating DC, we compared the localization of 5hmC around PU.1 binding sites. Therefore we annotated the individual PU.1 cluster sets with 5hmC tag counts obtained from whole genome sequencing (see chapter 5.2.5.6). The results are illustrated in Figure 3.32A (heatmaps) and Figure 3.32B (histograms). Clusters characterized by progressive loss of PU.1 binding (#4, #6) showed a constant depletion of 5hmC at the center of PU.1 binding sites. In contrast, in clusters with increasing PU.1 binding during monocyte differentiation, 5hmC accumulates in the direct proximity of PU.1 binding sites and shifts from the center to the shores of PU.1 binding in DCs at day 7 of differentiation. In cluster #5, 5hmC accumulation precedes the appearance of PU.1 at the regions in this set. This finding hints at a distinct chronology of PU.1 and 5hmC accumulation in, at least, one subset of the peaks.

PU.1 binding events, in all clusters of PU.1 accumulation during the time course, seem to be correlated to 5hmC appearance supporting the theory of a connection between PU.1 binding and active DNA demethylation events. Based on this assumption, cluster #4 and #6 may represent regions at later stages of the demethylation process, where PU.1 is not needed anymore and removed. To test this hypothesis CpG methylation in the area surrounding PU.1 binding sites was analyzed. Again, individual PU.1 binding clusters were annotated with 5mC tag counts obtained from whole genome bisulfite sequencing (data processing described in 5.2.5.9.2) and plotted in Figure 3.32C. In all clusters an overall reduction of CpG methylation was observed. This is in line with earlier findings of a global reduced level of methylation levels due to method discrepancies. However, the depletion of 5mC in the direct vicinity of the PU.1 peak center of differentiating monocytes cannot be entirely explained by technical differences. In cluster #4 and #6, 5mC levels in the peak center were already depleted in monocytes and the levels did not change during differentiation. This finding indicates that the demethylation process in the PU.1 peak centers is completed. The remaining

clusters show a 5mC reduction in the proximity of PU.1 binding in monocytes. 5mC further decreases in the center regions with progression of demethylation. In regions of cluster #5, which are devoid of PU.1 in monocytes, 5mC levels are reduced but to a lesser extent than in clusters marked by PU.1 binding in monocytes. Taken together PU.1 dynamics seem to be linked to 5hmC in differentiating monocytes.

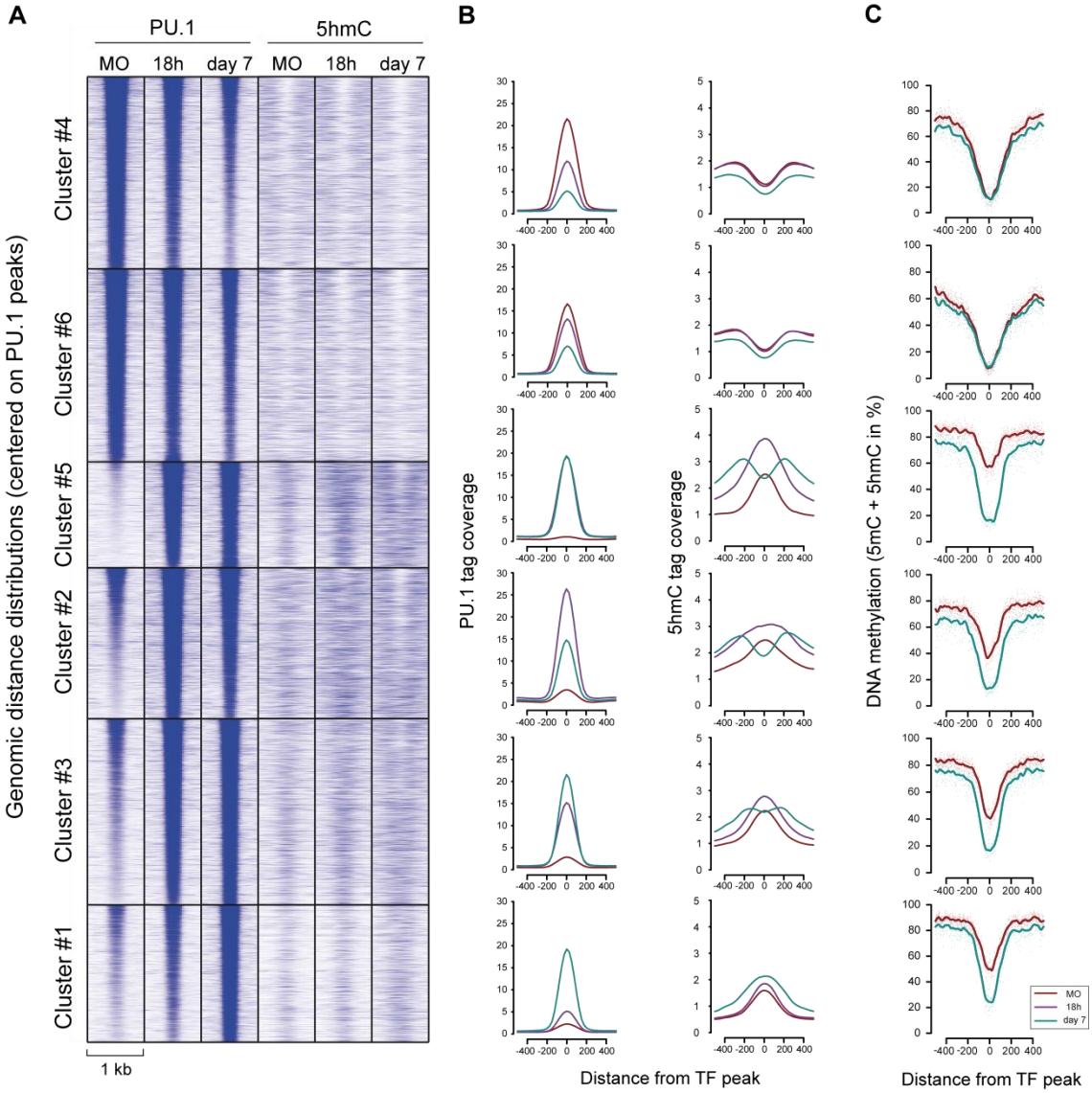

Figure 3.32: PU.1 binding dynamics and corresponding 5hmC and 5mC levels
PU.1 peaks were clustered according to their dynamic binding profile in differentiating monocytes. Genomic tag count distance distribution of PU.1 ChIP-seq (left panel) and Hydroxymethyl Collector™-seq tags (right panel) of several differentiation time points are shown in heatmaps (**A**) or histograms (**B**) across a 1 kb genomic interval relative to the transcription factor peak center. (**C**) Histograms of methylation levels assessed by bisulfite sequencing at PU.1 binding sites of differentiating monocytes (MO *n*=4, DC *n*=6). 5mC and 5hmC levels of all valid CpGs in 1kb regions of the indicated PU.1 cluster sets in (A) is shown in % of methylation relative to peak centers.

3.3.3.2 PU.1 binding dynamics at promoter elements

So far PU.1 binding has been investigated in several human and murine tissues (Ghisletti *et al.*, 2010; Heinz *et al.*, 2010; Pham *et al.*, 2012; Ridinger-Saison *et al.*, 2012; Wilson *et al.*, 2010; Wontakal *et al.*, 2011; Zhang *et al.*, 2012a) and was found to be associated with promoters and enhancers where it is regulating gene expression (Kihara-Negishi *et al.*, 2001). To assess a possible link of PU.1 binding and transcription levels in our cell system we sorted an existing set or promoter regions produced by the FANTOM consortium (Consortium *et al.*, 2014) for actively transcribed promoters and compared it to local PU.1 ChIP-seq and Hydroxymethyl Collector^{TM}-seq tag counts from our own analyses (details in chapter 5.2.5.7 and 5.2.5.9.3).

3.3.3.2.1 PU.1 binding dynamics in the context of gene transcription

The genomic PU.1 tag distance distribution for two sets of actively transcribed promoters (MO- and DC-specific) is shown as heatmaps in Figure 3.33A (left panels) and as histograms in Figure 3.33B (left panels). In monocytes, the number of promoter regions (5882) is more than twice as high as the number of DC promoter regions (2666). The majority of these promoters are associated with PU.1 accumulation. This finding confirms a link between PU.1 binding and active promoter usage. However a small set of MO-specific (and also DC-specific) promoters lack PU.1. CpG density coverage (illustrated in the right panel of Figure 3.33A) indicates that those peaks mainly represent CpG-rich promoter regions or CpG islands (CGI), which are generally devoid of this TF. In both promoter sets, PU.1 enrichment is slightly shifted to the left of the center, indicating preferential binding upstream of the TSS. At monocyte-specific promoters, PU.1 occupancy is progressively receding during differentiation. In contrast, DC-specific promoters display a rapid increase of PU.1 binding upon MO to DC transition. These findings indicate specific binding dynamics of PU.1 at cell type-specific promoters.

In the MO-specific promoter set, 5hmC occurrence is rare (Figure 3.33A and B, right panel), and completely absent at preferential binding sites of PU.1. In contrast, 5hmC appearance is not directly related to PU.1 binding in the DC-specific promoter set. At the beginning of the time course, the modification is randomly distributed in the analyzed region surrounding the TSS but seems to clear the center region of PU.1 binding and shift to the shores until day 7 of differentiation. This may be due to the fact that the core PU.1 consensus motif lacks CpG dinucleotides and therefore 5hmC

(Pham *et al.*, 2013). Promoters devoid of PU.1 binding support this idea, since they lack any regions of complete 5hmC depletion in both promoter sets.

In Figure 3.33C both sets of cell-specifically transcribed promoters were annotated with CpG methylation data, obtained from bisulfite sequencing of MO and DC at day 7 (see chapter 5.2.5.9.3 for details). In line with the 5hmC distribution, 5mC levels in both sets are overall depleted in the TSS region during differentiation. This observation confirms that while proximal promoter regions may acquire 5mC and 5hmC, the centers at the TSS are generally devoid of those marks (Szulwach *et al.*, 2011; Weber *et al.*, 2007; Wu and Zhang, 2011). Taken together these results show distinct PU.1 dynamics at cell type-specific promoters, which are not correlated with demethylation events.

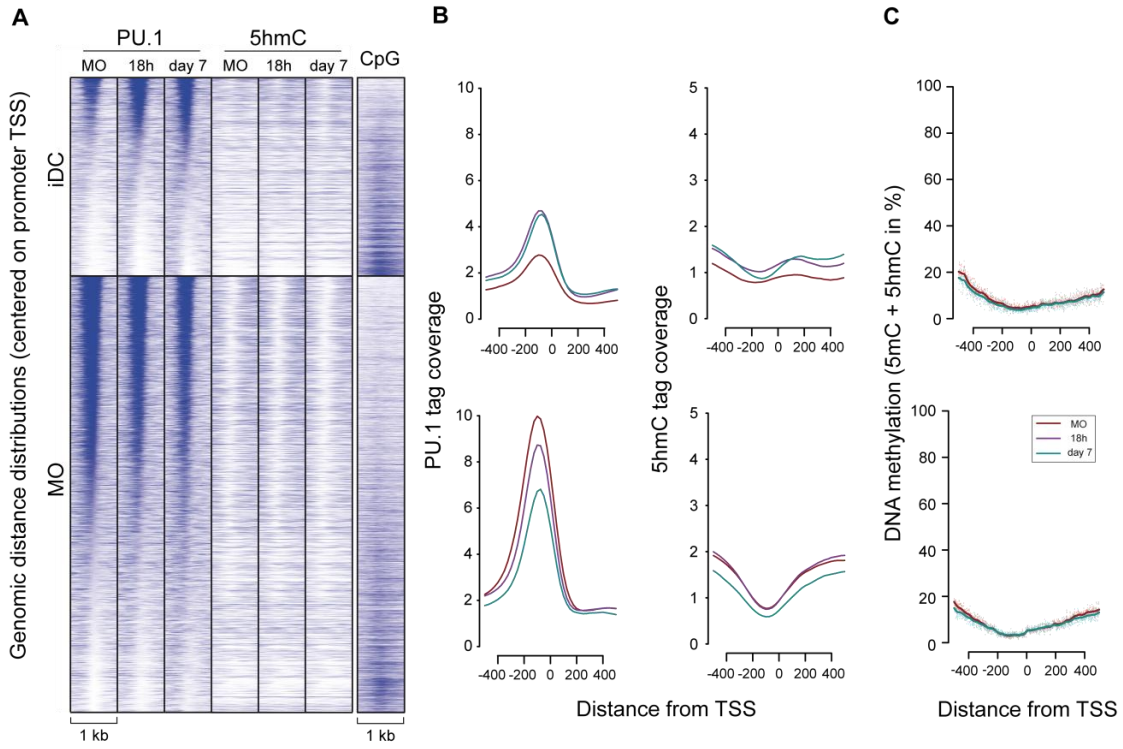

Figure 3.33: Correlation of PU.1 accumulation with promoter expression levels
A CAGE-based promoter set was divided into time point-specific expression sets (MO, iDC, *n*=3) and annotated with associated tag counts of PU.1 ChIP-seq, Hydroxymethyl Collector^{TM}-seq and global CpG-coverage. (**A**) CpG density and coverage in 1 kb regions around PU.1 peak centers is illustrated as heatmaps in the right panel and the genomic distance distribution (PU.1, 5hmC) of three time points of differentiating monocytes (1-kb regions centered on the TSS) are shown in the left panels (A) or histograms (**B**). (**C**) Methylation levels obtained by bisulfite sequencing of differentiating monocytes (MO *n*=4, DC *n*=6) are shown as histograms. 5mC and 5hmC levels of all valid CpGs in 1kb regions of the time point-specific sets is shown in % of methylation in relative distance to the TSS.

3.3.3.3 PU.1 binding dynamics at promoter-distal transcribed elements

Enhancers are promoter-distal regulatory elements that drive cell type specific gene expression. They are associated to a variety of cell type-specific TFs which aid in the

regulation of transcription (Heintzman *et al.*, 2009; Ong and Corces, 2011). Pioneering TF like PU.1 are initiating the access to and open new enhancer elements (Heinz *et al.*, 2010; Serandour *et al.*, 2011). In fact, cell type-specific PU.1 binding sites are much more abundant at promoter-distal sites than at promoters (Pham *et al.*, 2013).

To assess the PU.1 binding dynamics at enhancers in monocyte differentiation, a set of bidirectionally transcribed enhancers (Andersson *et al.*, 2014) was sorted for cell type-specific enhancer elements in MO and DC and compared to PU.1 tag counts (details in chapter 5.2.5.8 and 5.2.5.9.3). Heatmaps (Figure 3.34A) and histograms (Figure 3.34B) illustrate genomic PU.1 tag density distribution in the two sets of MO (3018 enhancer regions) and DC (278 enhancer regions) specific enhancers. MO-specific enhancers show decreasing PU.1 binding intensity during the differentiation time course. In contrast, active DC enhancers enrich PU.1 during monocyte differentiation. This observation is similar to promoter PU.1 dynamics.

In summary, PU.1 targeting is mainly found at promoter and promoter-distal regions that are activated in a cell type-specific manner. Accordingly, upon transition into a new cell type PU.1 is removed from cell type-specific regulatory elements of the former cell type. These findings support the notion of general acquisition of the key regulators, like PU.1 to cell type-specific sites (Heintzman *et al.*, 2009; Ong and Corces, 2011; Pham *et al.*, 2012). Compared to promoters, the fraction of enhancer regions bound by PU.1 is elevated in the MO-specific set. This is in line with other data sets that show cell type-specific PU.1 binding mainly at promoter-distal regulatory sites (Pham *et al.*, 2013). PU.1 binding is mainly located in the center of the enhancer transcription start sites (eTSS). This finding reflects the nature of the enhancer set, which was centered between the two transcription start sites of the bidirectional eRNA transcripts.

The small numbers of DC-specific enhancers (10-fold lower compared to number of MO enhancers), may be explained by methodical problems in enhancer detection using CAGE; overall low tag counts in the DC samples impede proper detection of those elements.

Compared to promoters, 5hmC levels in the vicinity of actively transcribed enhancers (shown in Figure 3.34A and B, right panels) are increased and display a similar, but more distinct pattern. At PU.1-acquiring DC-specific enhancers, 5hmC is randomly distributed in monocytes and moves to the shores of PU.1 binding sites until day 7 of differentiation. However, in the MO-specific enhancer set 5hmC is absent from PU.1 peak sites at all time points and its levels decrease until day 7 of monocyte differentiation.

Analysis of 5mC (Figure 3.34B) in general, confirms the distribution of 5hmC. 5mC is also constantly absent at the core PU.1 binding sites in the MO- and DC-specific

enhancer set, and shows a slight reduction at the shores during differentiation. This is in line with earlier findings and may be explained by technical variation (3.3.2.1).

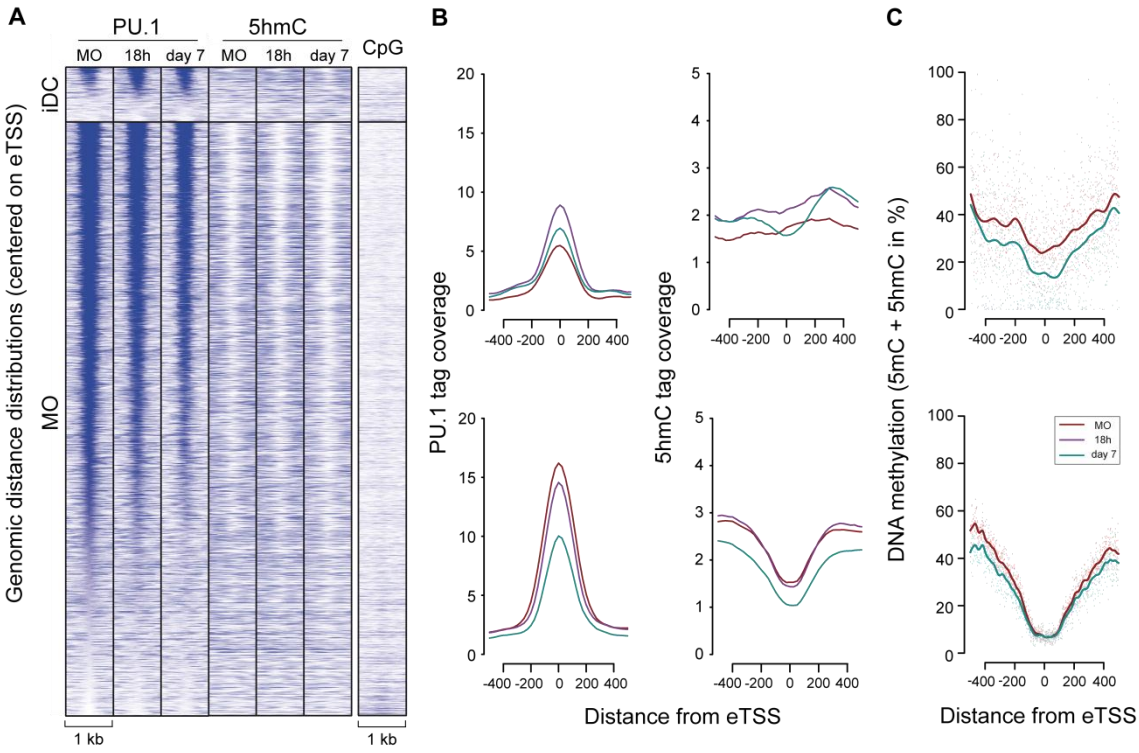

Figure 3.34: PU.1 dynamics at MO and DC specific enhancer regions
A CAGE-based enhancer set was divided into time point-specific expression sets (MO, iDC, $n \geq 2$) and annotated with associated tag counts of PU.1 ChIP-seq (left panel), Hydroxymethyl Collector™-seq (middle panel) and genomic CpG coverage (right panel). Genomic 5hmC and PU.1 tag distribution at three time points of differentiating monocytes as well as CpG density is shown in heatmaps across a 1 kb genomic interval relative to the eTSS or the center of the enhancer. (**C**) Histograms of methylation levels as assessed by bisulfite sequencing of differentiating monocytes (MO n=4, DC n=6). 5mC and 5hmC levels of all valid CpGs in 1 kb regions of the time point-specific sets is shown in % of methylation in relative distance to the eTSS.

4 Discussion

4.1 The active DNA demethylation mechanism and its functional relevance

At beginning of this thesis the oxidative enzymatic modification of 5mC had just been discovered as a part of active demethylation processes in mammals. Until now, the active DNA demethylation process has been the topic of ongoing research. In the previous years, numerous publications emerged and contributed to the understanding of the process (He *et al.*, 2011; Ito *et al.*, 2011; Iyer *et al.*, 2009; Kriaucionis and Heintz, 2009; Tahiliani *et al.*, 2009). To date the relative importance of the various pathways is still under debate and it is not clear which components are associated with the demethylation machinery or how these mechanisms contribute to gene regulation in different cell types (Kohli and Zhang, 2013; Wu and Zhang, 2014). In this work we further characterized the active DNA demethylation processes in the hematopoietic compartment and gained insights into the functional relevance and spatio-temporal patterns of the demethylation events. At the beginning, most studies regarding DNA demethylation focused on proliferating or pharmaceutically arrested cells (Kangaspeska *et al.*, 2008; Metivier *et al.*, 2008). Both approaches are not ideal suited for the investigation of this process. First, drugs are globally interfering with the cells and may influence the molecular processes. Second, demethylation processes observed in proliferating cell systems do not exclude a possible contribution of passive demethylation events and are therefore not suited to study active demethylation. The data presented in this work are obtained from a non-proliferating *in vitro* cell system. This *in vitro* setting provides many advantages compared to other model systems. It utilizes primary monocytes directly obtained from human blood donors and thus simulates the original *in vivo* setting of monocytes. The cells are easily isolated from whole blood, providing abundant studying material from the organism of interest. Most important, the cells differentiate in the absence of DNA replication (Klug *et al.*, 2010), thereby excluding passive demethylation processes to interfere with the experiments. Taken together, this model system is ideally suited to address questions regarding active DNA demethylation processes. Using this system we were able to gain further

insights into the active demethylation mechanism in differentiating monocytes and to identify TET2 as a key player in this process.

Our group had previously detected active demethylation at several DMRs in differentiating monocytes. DMRs were identified as differentially methylated regions in monocyte-derived DC compared to MAK and characterized by loss of methylation during differentiation (Klug *et al.*, 2010). These demethylation events occurred preferentially but not exclusively at promoter-distal regions and featured the synchronous appearance of activating histone marks. The findings indicated the recruitment of an active DNA demethylation machinery to the DMRs in the context of histone-modifying processes (Klug *et al.*, 2010). The main aim of this thesis was to further characterize this active DNA demethylation process and to identify the enzymes involved. Although many demethylation pathways had been described to that date, the active demethylation pathway, initiated by the oxidation of 5mC by the TET proteins, was predominantly observed in mammalian cell systems (Bruniquel and Schwartz, 2003; Ma *et al.*, 2009; Mayer *et al.*, 2000; Morgan *et al.*, 2005; Oswald *et al.*, 2000). It was further noted that the oxidation of 5mC is more likely to occur than direct removal of 5mC, since the oxidation of 5mC has much lower energy costs (Kohli and Zhang, 2013). To test if the loss of 5mC observed at DMRs in our cell system was associated with the conversion into 5hmC we followed (in chapter 3.1.2.1) 5mC and 5hmC levels and observed a loss of 5mC accompanied by a synchronous appearance of 5hmC with different kinetics at those loci. Using alternative methods to clearly distinguish both marks, and to exclude bias towards CpG-density or off-target effects due to unspecific cross-reaction of the AB (Matarese *et al.*, 2011; Pastor *et al.*, 2011), we could confirm the results. As outlined in the introduction, the conversion to 5hmC is a major step in the demethylation mechanism in many other human cell types (reviewed in Wu and Zhang, 2014). This modification as well as further oxidation steps of 5hmC into 5fC or 5caC are catalyzed by TET proteins (He *et al.*, 2011; Ito *et al.*, 2011; Tahiliani *et al.*, 2009), indicating an involvement of these enzymes in the process. A restriction-based approach (3.1.2.2.5) confirmed the complete removal of possible further oxidation products like 5fC or 5caC at the demethylated regions. Since this method only covered one CpG it is possible that this result does not reflect the modification state of all CpGs in this region. Thus, other methods to detect 5fC and 5caC, like the adapted version of fCAB-seq, will be useful to clarify this issue in the future. In summary, these findings confirmed that the conversion of 5mC to 5hmC is the first step in the demethylation process in the utilized cell system and indicated a possible involvement of TET hydroxylases.

After the discovery of their potential to initiate active demethylation processes, TET proteins have been the focus of many studies. They were identified as key players in developmental as well as cell type-specific processes and their expression in different tissues leads to overlapping and distinct functions at different stages of development (reviewed in Branco *et al.*, 2012; Pastor *et al.*, 2013). As a major regulator of differentiation, TET2 is ubiquitously expressed in the hematopoietic lineage (Ko *et al.*, 2011; Li *et al.*, 2011; Moran-Crusio *et al.*, 2011; Quivoron *et al.*, 2011). This finding is in line with the expression data obtained in our setting. In chapter 3.1.2.2.1 and 3.1.2.2.3 we could show that of all TET proteins only TET2 was abundantly expressed and showed stable protein levels at all time points of the differentiation time course. Thus TET2 represented a potential candidate enzyme for the demethylation events observed in differentiating monocytes. The depletion of TET2 (3.1.2.2.4 and 3.1.2.2.6) confirmed a lagged loss of 5mC as well as reduced 5hmC levels indicating a delay in the active demethylation process. Since TET2 is the major TET protein expressed in the hematopoietic lineage, it was not surprising to identify an involvement in the demethylation process. As illustrated in the introduction, TET2 has been linked to the demethylation process in other cell types before and the loss of this protein resulted in malignant transformation of the hematopoietic compartment, indicating important functions as a tumor repressor (Huang and Rao, 2014; Pastor *et al.*, 2013; Wu and Zhang, 2014). Indeed loss of 5hmC and aberrant 5mC patterns are hallmarks of malignant hematopoietic cells (Ko *et al.*, 2010; Konstandin *et al.*, 2011). Taken together we identified TET2 as the initiator of the active DNA demethylation process in differentiating monocytes and as being responsible for the conversion of 5mC to 5hmC. In the next steps, we investigated an involvement of further factors in the active DNA demethylation mechanism, in particular in the remaining steps following the conversion of 5mC to 5hmC.

Members of the BER have gained a lot of attention in the last years as important players in the last steps of active DNA demethylation processes (Kohli and Zhang, 2013). The glycosylase TDG for example was shown to be involved in the active DNA demethylation process and responsible for the excision of 5mC oxidation derivatives 5fC and 5caC (He *et al.*, 2011; Zhang *et al.*, 2012b). Other studies linked TDG to embryonic development (Cortazar *et al.*, 2011; Cortellino *et al.*, 2011), where TDG is crucial for the assembly and function of activating histone complexes, thus regulating the expression of developmental genes (Cortazar *et al.*, 2011). Combining the ability to contribute to the demethylation pathway and its potential to recruit chromatin modifying complexes, TDG presented a possible candidate for the demethylation mechanism in our cell system. In mammals, the glycosylase MBD4 was so far known to be involved

in a demethylation mechanism initiated by the deamination of 5mC (Bhutani *et al.*, 2010; Popp *et al.*, 2010; Rai *et al.*, 2010) and a possible role in other active processes was hypothesized (Chen and Riggs, 2011). Although both glycosylases were ubiquitously expressed in differentiating monocytes (3.1.2.2.1), siRNA experiments (3.1.2.2.4 and 3.1.2.2.6) indicated that neither TDG nor MBD4 are responsible for the conversion of 5mC into 5hmC, which is in line with previous studies (Zhang *et al.*, 2012b). It is still unclear, if these enzymes are involved in the later steps of the DNA demethylation pathway. We tested if 5hmC was converted to 5fC or 5caC, further downstream oxidation products of 5mC. This process is catalyzed by TET proteins (He *et al.*, 2011; Ito *et al.*, 2011) and in cells where TDG is involved in the demethylation mechanism, TDG depletion was shown to result in accumulation of those modifications (He *et al.*, 2011; Shen *et al.*, 2013; Song *et al.*, 2013). The generally impeded base and DNA double-strand break repair in blood monocytes (Bauer *et al.*, 2011) may have been responsible for a delayed conversion to 5fC and 5caC and thus for a delay in the demethylation process, as was observed. But we could not observe any impairment in restriction efficiency (3.1.2.2.5) and thus no enrichment of 5fC or 5caC after TDG or MBD4 depletion at the DMRs, which suggested the complete removal of 5fC and 5caC at least at one CpG. A different approach using the fCAB method (3.1.2.4) resulted in inconclusive results, due to technical variation. Further adjustments of the method or more sensitive detection methods for 5fC and 5caC, will be necessary to gain further insights into these processes. Although a direct involvement in the demethylation process, via the excision of 5fC and 5caC, was so far only shown for TDG (He *et al.*, 2011; Maiti and Drohat, 2011; Zhang *et al.*, 2012b), a possible role for MBD4 cannot be ruled out at this stage. A recent study detected the interaction of TDG as well as MBD4 with all three TET proteins, further supporting a possible involvement of both MBD4 and TDG in the demethylation process (Muller *et al.*, 2014). However, the authors show that TDG, but not MBD4, is able to activate oxidation-dependent reporter gene expression and suggest different regulatory roles for both enzymes. In support of this theory both glycosylases were shown to play a role in different mechanisms of transcriptional regulation. TDG regulates the maintenance of active and bivalent chromatin via the recruitment of histone modifying complexes (Cortazar *et al.*, 2011) and MBD4 is involved in transcriptional repression in cooperation with Sin3A and HDAC1 (Kondo *et al.*, 2005). Since we were previously able to link demethylation processes in our cell system to the simultaneous appearance of active histone modifications (Klug *et al.*, 2010), it will be important to investigate histone modifications surrounding the DMRs in cells, depleted of TDG or MBD4. This may give further insights into the recruitment of the demethylation machinery, its association with

histone-modifying processes and a possible function for TDG and MBD4 in this process.

The 5mC mark is a known epigenetic regulator and recruits 5mC binding proteins to DNA (Hendrich and Bird, 1998; Meehan *et al.*, 1989). Given the vast abundance of 5mC oxidation derivatives in the genome it has been hypothesized that those may as well act as epigenetic regulators (Delatte *et al.*, 2014). Two recent publications have supported that theory (Iurlaro *et al.*, 2013; Spruijt *et al.*, 2013) by presenting a number of DNA readers capable of targeted recognition and physical binding to specific DNA modifications, such as 5mC, 5hmC, 5fC and/or 5caC. These results suggested potential candidates involved in the active demethylation process, such as the helicase HELLS, which was shown to bind directly to 5hmC in both studies (Iurlaro *et al.*, 2013; Spruijt *et al.*, 2013). Furthermore helicases are involved in DNA repair (Brosh, 2013), which plays an important role in the active DNA demethylation mechanism in several cell types (for a review see Wu and Zhang, 2014). Another candidate enzyme, O-linked β-D-N-acetylglucosamine transferase (OGT), was recently identified as an interaction partner of TET proteins (Chen *et al.*, 2013; Deplus *et al.*, 2013; Vella *et al.*, 2013). It could be shown, that these proteins are both forming a complex, which is involved in transcriptional activation (Chen *et al.*, 2013; Deplus *et al.*, 2013). Using a siRNA-based method we could show that neither of the candidate enzymes is involved in the conversion of 5mC to 5hmC (3.1.2.3.3). Nevertheless, a possible involvement of OGT or HELLS in the demethylation process cannot be ruled out altogether. Helicases play a role in DNA repair (Brosh, 2013) and HELLS may be involved in the excision of further 5mC derivatives. To test this possibility a modification-sensitive restriction-digest (MSRE, as described in chapter 5.2.2.7.3.1) or the fCAB method (Song *et al.*, 2013) could be performed in cells depleted of HELLS or OGT. OGT is an important factor of TET1 protein-stabilization and OGT-inhibition was shown to result in a 70% depletion of TET1 (Shi *et al.*, 2013). Since we did not see any effect of OGT-depletion on the 5mC to 5hmC conversion rate, the protein-stabilizing effect may be limited to TET1 and not be important for TET2. This is in line with previously published data (Chen *et al.*, 2013). The authors reported that OGT does neither affect the TET2-mediated 5hmC synthesis nor convert TET2-derived 5hmC into other derivatives *in vitro*. Of note, Shi *et al*. reported that the stabilization of TET1 was due to the *O*-GlcNAcylation of TET1 by OGT (Shi *et al.*, 2013). *O*-GlcNAcylation of TET2 and TET3 has been controversially discussed. While several groups detected an *O*-GlcNAcylation of TET2 or TET3 by OGT (Ito *et al.*, 2014; Vella *et al.*, 2013; Zhang *et al.*, 2014), Deplus et al. reported the contrary (Deplus *et al.*, 2013). It is thus still unclear if the *O*-GlcNAcylation of TET2 and TET3 occurs in general and if it is a prerequisite for the

stability of all TET proteins. Alternatively, TET2 levels may also be sustained by other unknown factors. In addition to the catalytic activity of TET2 in the active demethylation process, the recruitment of chromatin modifiers, such as OGT may present a possible link between active DNA demethylation and the simultaneous accumulation of activating histone marks, like H3K4 methylation, we observed earlier (Klug *et al.*, 2010). This hypothesis is supported by several studies reporting that OGT controls the activity of interacting histone methyltransferases, like CARM1 (co-activator associated arginine methyltransferase) or MLL5 (mixed lineage leukemia 5) (Cheung *et al.*, 2008; Deplus *et al.*, 2013; Fujiki *et al.*, 2009; Sakabe and Hart, 2010). One of these studies detected a recruitment of OGT by TET2 and TET3. OGT promoted the stabilization of an associated Set1/COMPASS complex and targeted the SETD1A methyltransferase within the complex to chromatin. The depletion of TET or OGT resulted in a loss of H3K4me3 (Deplus *et al.*, 2013). Taken together, OGT may be involved in later steps of the active demethylation process in our cell system and/or it may be responsible for the maintenance of the surrounding chromatin in a “hydroxymethylation independent” model of gene regulation (Delatte *et al.*, 2014). It will be interesting to investigate histone mark profiles at the DMRs after depletion of OGT and TET2 to address this question.

In summary, we were able to characterize the active DNA demethylation mechanism in differentiating monocytes in the first part of the thesis. We discovered that the targeted, locus-specific active DNA demethylation process is initiated by the conversion of 5mC to 5hmC and identified TET2 is required for this step. We further presented first insights into a possible involvement of TDG, MBD4, HELLS and OGT in the demethylation pathway. Further investigation of this issue will be important to define their importance for the active demethylation process. For the results of MBD4, TDG, OGT, and HELLS it might be possible, that minor effects on the demethylation process were obscured due to the heterogeneous nature of the transfected samples. As previously discussed, the transfection procedure affected the differentiation progress and the majority of cells did not yet enter the differentiation pathway at the early time points measured. To overcome this problem, a cell sort using flow cytometry may enrich for cells entering the differentiation pathway and thus provide bias-free results in subsequent experiments.

In monocyte-derived macrophages, derived from the same *in vitro* model of the MPS, our group has also detected DMRs (Pham *et al.*, 2013). Since the differentiation of MO into MAK takes place in the absence of cell proliferation we hypothesized that those are active demethylation events (Pham *et al.*, 2013). In chapter 3.2.2 we could confirm

this notion. The loss of 5mC was correlated with the synchronous appearance of 5hmC. These results were in line with earlier findings in DMRs in monocyte-derived dendritic cells (see chapter 3.1.2.1), but the functional relevance of the demethylation events was still unclear. Active DNA demethylation is, in cooperation with other processes, involved in the transcriptional regulation of adjacent genes and makes the DNA accessible for methylation-sensitive transcriptional regulators (Pastor *et al.*, 2013; Wu and Zhang, 2014). These changes in methylation are not always immediately associated with gene activation (Klug *et al.*, 2010). DNA demethylation has previously been linked to promoter-distal enhancer regions, where it contributes to the establishment of enhancers which are either active or pre-marked for later activation (Buecker and Wysocka, 2012; Pham *et al.*, 2012; Schmidl *et al.*, 2009). Since the loss of methylation in the macrophage DMRs was paralleled by an enrichment of H3K4me1, which marks enhancer regions (Heintzman *et al.*, 2007), this led to the assumption that those regions may be activated for enhancer functions. The question was addressed functionally in reporter gene assays. For human dendritic cells there is no suitable cell line available to be subjected to reporter gene assays. So we chose a human cell line that can be transformed into a related, macrophage-like cell type (Abrink *et al.*, 1994; Tsuchiya *et al.*, 1982). Enhancer reporter assays testing DMRs upstream of a heterologous promoter in THP-1 cells (section 3.2.1) revealed enhancer activity for several regions, identifying those demethylated sites as potential enhancers. Strikingly, the enhancing properties were dependent on the methylation level of those regions. In almost all regions *Sss*I-mediated methylation of the plasmid DNA impaired or completely abolished the enhancing potential, supporting the hypothesis that demethylation is necessary for the activation of distal-regulatory enhancer regions. This hypothesis is consistent with the previously reported link between demethylation events and enhancer activity, both in a related and un-related cell type (Klug *et al.*, 2010; Schmidl *et al.*, 2009). Klug et al. showed enhancer activity for several DC-specific, demethylated regions, which was lost after *in vitro Sss*I-methylation of the plasmid DNAs (Klug *et al.*, 2010). The demethylation events at the DMRs in MAK were correlated with the local binding of key TF, like PU.1, C/EBPβ and EGR2, co-occurring at the same stages of MAK differentiation that displayed demethylation (Pham *et al.*, 2013). It was therefore likely that the local binding of cell type-specific TF initiates the demethylation process (Pham *et al.*, 2013). In support of this theory, most of the regions displayed enhancer potential, but their activation depended on the cell context. THP-1 cells developed into a monocyte-macrophage like cell type in the presence of PMA and enabled (compared to THP-1 cells cultured without PMA) a different set of regions to enhance a basal promoter, indicating cell type-specific enhancer usage

(chapter 3.2.1). In line with these findings, one region showed methylation-sensitivity in one cell context but not in another, indicating a cell type-specific removal of the 5mC mark. This is consistent with reporter gene assays in regulatory T cells (Treg). Treg-specific enhancer elements were activated in a related cell type, suggesting the presence of required activating factors in both cell types. The authors further suggested an important role of DNA methylation in the restriction of lineage-specific promoter-distal regulatory elements in other cell types than Tregs (Schmidl *et al.*, 2009). Our results in a non-myeloid cell system lacking the key TF PU.1 (Kwon *et al.*, 2005) (3.2.1) confirmed this theory. The activation of cell type-specific enhancer elements requires a corresponding set of TF to open the regions, and may include the removal of 5mC in this process. In support or this finding it has been shown, that C/EBPβ and EGR2 are upregulated during MO to MAK differentiation and that the binding of each of those TFs is associated with MAK-specific gene regulation (Pham *et al.*, 2012; Pham *et al.*, 2007). The demethylation events in the DMRs were not restricted to MAK, but were also seen in differentiating dendritic cells, albeit with different demethylation kinetics (3.2.2). It seems likely that related cell types are able to activate a common set of regulatory elements (Schmidl *et al.*, 2009), but with different kinetics, depending on the present set of transcriptional activators.

Taken together we were able to confirm active DNA demethylation in monocyte-derived macrophages and link these events to enhancer function. We demonstrated that the enhancer regions were methylation-sensitive and that demethylation led to enhancer activation in a cell type-specific manner. It has to be noted that those findings were obtained in an artificial cell system only resembling the actual TF content of MAK and lacking the surrounding chromatin structure. Nevertheless, the experiments gave first hints at cell type-specificity and enhancer usage and supported the results of other colleagues (Pham *et al.*, 2013).

4.2 Global dynamics of active demethylation and the master transcription factor PU.1

The first part of this work further characterized the demethylation mechanism in differentiating monocytes and identified TET2 as a key player in this process. Functional analysis of selected DMRs, associated with key TF, linked the local binding of these factors and the simultaneous demethylation events to functional purposes. These findings support the theory suggested by us (Pham *et al.*, 2013) and others

(reviewed in Ong and Corces, 2011) that master TFs like PU.1 are directly involved in the demethylation process. Since active demethylation events occur on a global level and are not restricted to enhancer elements (Klug *et al.*, 2010), we further investigated the previous results on a global scale. The following chapters give a genome-wide overview on the dynamics of 5mC oxidation and the association with the key TF PU.1 in differentiating monocytes.

In the last years a variety of methods have been developed to map 5hmC distribution across the genome. They can be divided into pull-down methods including antibody-or chemical labeling-based methods and single-base resolution approaches, such as bisulfite-sequencing. The single-base resolution strategies are able to cover scattered CpGs that may escape the detection of affinity-based methods. Base-resolution methods measure the average modification of each single cytosine and are dependent of the sequencing depth and coverage of 5hmC. Improved protocols like the oxBS-seq and TAB-seq method (Booth *et al.*, 2012; Yu *et al.*, 2012) overcome the former problem of discrimination between 5mC and 5hmC. The main disadvantage of base-resolution methods is the sequencing costs which increase with the sequencing depth necessary to detect rare cytosine modifications. Affinity-based approaches on the other hand are cost-effective and cover the majority of cytosines which are mostly clustered in the genome (Yu *et al.*, 2012). Those methods are limited by their resolution, which is dependent on the DNA fragment size (reviewed in Wu and Zhang, 2014). To measure 5hmC levels genome-wide, we tested and validated two affinity-based methods. First, the hMeDIP based on a 5hmC antibody and second, the Hydroxymethyl Collector^{TM} kit, which includes selective, chemical labeling of 5hmC. The comparison of sequencing read counts from both methods, presented in chapter 3.3.1.1 revealed highly diverging results. Although they did not show any correlation, we were able to detect the DC-derived example DMRs in both sets. The methods did however show different spatial accuracies, which may be due to the low resolution of the method (data not shown). Another criterion we took into account was that antibody-based methods, compared to chemical labeling, may be prone to non-specific enrichment of simple repeats (Matarese *et al.*, 2011). However, both sets derived from our methods equally enriched for simple repeats, and LINE1 repeats (data not shown). Since those features were inconclusive, we validated a number of regions that were uniquely detected by each method with an alternative 5hmC detection method (3.3.1.1). In conclusion only the results from the Hydroxymethyl Collector^{TM} kit were reproducible with the other method. It has to be noted that the alternative method is also based on glycosylation of 5hmC and may therefore not present an entirely independent approach. Nevertheless, also within biological and technical replicates the kit showed good correlations and high

reproducibility, even after slight alterations to the protocol. Interestingly other groups compared the same methods (with a slightly altered hMeDIP protocol, as well as a different polyclonal AB from the same company) in murine neuronal progenitor cells, mature neurons, brain and liver cells and obtained similar enrichment between the two methods (Hahn *et al.*, 2013; Thomson *et al.*, 2013). In contrast to our approach both studies used the same polyclonal antibody from Active Motif, indicating a consensus of the data obtained using this antibody with results from the Hydroxymethyl Collector™ kit. In order to validate selected regions enriched by both tested methods Thomson *et al*. used glycosyl transfer to 5hmC followed by glycosylation-sensitive restriction digest. This approach is similar to the validation protocol we used (see chapter 5.2.2.7.2.1). They were able to reproduce the results from both methods (hMeDIP, Hydroxymethyl Collector™), supporting the earlier notion. The authors further detected significant bias of the hMeDIP method towards simple repeats as well as LINE1 elements (Thomson *et al.*, 2013), whereas Hahn *et al*. did not detect any bias regarding CA-repeats (Hahn *et al.*, 2013). Our results obtained with the AB-based method were consistent with the results of the second group (Hahn *et al.*, 2013). The discrepancies may be explained by the different antibodies used in the studies and variations in the protocol. In any case, the results for the Hydroxymethyl Collector™ kit were concordant and supported our data. Taken together, the Hydroxymethyl Collector™ kit presented itself as an accurate and stable method and was therefore chosen for further 5hmC analyses.

Genome-wide mapping of 5hmC has previously been performed in a number of cell types to gain insights into the functional relevance of demethylation processes. In combination with global maps of other 5mC oxidation derivatives and distribution profiles of TET proteins those studies contributed to a better understanding of the purposes of active DNA demethylation events in various tissues and developmental stages (Pastor *et al.*, 2013; Wu and Zhang, 2014). This work presents a genome-wide profile of 5hmC dynamics and its association with the key TF for hematopoiesis, PU.1, in a hematopoietic cell context. In chapter 3.3.2.1 we focused on the question, if active demethylation events in monocyte to DC differentiation can be clustered into functional entities and if there is a connection between 5mC oxidation and PU.1 binding. The k-means clustering indeed classified the 5hmC peak set into four sub-clusters with different 5hmC dynamics during the differentiation time course. The subsets exhibited similar dynamics in pairs of two clusters, either gaining or loosing 5hmC over time. The loss of 5hmC was accompanied by a simultaneous loss of 5mC, indicating ongoing active DNA demethylation at those regions. The quality-control of 5mC data presented in the same chapter identified the global reduction of 5mC to be introduced by

biological or technical variation due to the different methods and donors used for whole genome 5mC analysis (see chapter 3.3.2.1). Nevertheless the distinct local loss of 5mC and 5hmC at 5hmC peak centers, respectively, was still clearly detected and not explicable by method or donor variation. On the basis of this control data analysis we considered the dynamics extending the global variations as specific demethylation events. Further investigation of methylation levels in selected regions by the MassARRAY system may be used to validate the active demethylation events in those regions and thus to confirm the assumption. The recruitment of PU.1 to sites of 5hmC enrichment and 5mC loss supports the idea of a correlation between demethylation and PU.1, which has been hypothesized in previous chapters and is supported by several publications. Pham and colleagues also detected loss of methylation at sites of PU.1 accumulation in macrophages, a related cell type. DNA methylation and PU.1 binding were found to be mutually exclusive and even the majority of non-bound PU.1 motifs did not correlate with 5mC (Pham *et al.*, 2013). Similar observations were made for another transcription factor, CTCF, in methylome studies of murine neuronal and ESC. The binding of this TF resulted in local loss of methylation at the binding site and indicated an involvement of CTCF in the initiation of demethylation (Stadler *et al.*, 2011). For the regulation of demethylation via transcription factors two opposing modes of action are discussed at the moment. First, TF like PU.1 may induce local demethylation processes to open the DNA for other 5mC-sensitive TF as was previously shown by Stadler *et al.* (Stadler *et al.*, 2011). Or second, the demethylation process removes 5mC to provide access for PU.1 and other factors (Pham *et al.*, 2013; Stadler *et al.*, 2011). The results presented in chapter 3.3.2.1 speak in favor of the latter theory. One subset of PU.1 target regions recruit PU.1 to sites with abundant 5hmC accumulation, indicating the involvement of other factors in the initiation of the demethylation process, at least for regions in this cluster. This is further supported by observations in chapter 3.3.3.1 and will be discussed later. Regions gaining 5hmC during monocyte differentiation did not show any loss of 5mC, which may indicate that the active demethylation proceeds at a slower pace, similar to the observations at single DMRs with “slow” kinetics (see chapter 3.1.2.1). It is also possible that the demethylation process pauses at the 5hmC step. Previous studies suggested this possibility for the male and to a lesser extent for the maternal pronucleus of preimplantation embryos. In the early zygote 5hmC is in general not further oxidized and may serve as a stable epigenetic mark before it is passively diluted in a replication-dependent manner (Inoue and Zhang, 2011; Iqbal *et al.*, 2011; Shen *et al.*, 2014; Wang *et al.*, 2014). Similar results have been observed in several tumor tissues, like brain, lung, breast, uterus and pancreatic tissue (Jin *et al.*, 2011). In murine neurogenesis

5hmC levels increase during differentiation, and consistent with our results are associated with stable 5mC levels (Hahn *et al.*, 2013). In our cell system the 5hmC mark may therefore not only serve as an intermediate of the demethylation process. Its presence at many sites that are not subjected to demethylation suggests additional functions, e.g. as an epigenetic mark with regulatory potential, which has been discussed previously (Calo and Wysocka, 2013; Szulwach and Jin, 2014). Those studies suggest that 5hmC may recruit specific "readers" to the DNA, which recognize 5hmC and may contribute to transcription regulation. In support of this theory two recent studies identified several specific binding proteins for 5hmC, including glycosylases and DNA repair proteins (Iurlaro *et al.*, 2013; Spruijt *et al.*, 2013). Spruijt *et al*. detected different sets of 5hmC binders in mESC, neural progenitor cells or adult mouse brain tissue, suggesting different functions of 5hmC depending on the cell context (Spruijt *et al.*, 2013). The second study confirmed many 5hmC readers in an independent approach using mESC extracts and detected sequence-specific binding preference of several reader proteins. These findings added an additional characteristic to the distinct recognition of 5hmC, further supporting its functional diversity (Iurlaro *et al.*, 2013). 5hmC may also represent a negative signal, blocking 5mC-specific readers. Most 5mC-binding proteins do not recognize 5hmC and may therefore be removed from the DNA when 5mC is converted into 5hmC (Jin *et al.*, 2010; Valinluck *et al.*, 2004). UHRF1 is involved in methylation-maintenance in cooperation with DNMT1 and is known to bind both 5hmC and 5mC albeit with much lower affinity to 5hmC (Hashimoto *et al.*, 2012; Valinluck *et al.*, 2004). The accumulation of 5hmC may therefore lead to impaired methylation maintenance during replication, as several other studies have shown (Shearstone *et al.*, 2011; Valinluck and Sowers, 2007). Shearstone *et al*. reported a replication-dependent loss of 5hmC in erythropoiesis, in the presence of functional DNMT1 and UHRF1 (Shearstone *et al.*, 2011). In addition Valinluck *et al.* observed the prohibition of DNMT1-mediated methylation in the presence of 5hmC, pointing at a role of this mark in the maintenance of an unmethylated state (Valinluck and Sowers, 2007). There are several other DNA binding proteins targeting both 5hmC and 5mC *in vitro*, like MeCP2, MBD1, MBD2, and MBD4 (Hashimoto *et al.*, 2014; Jin *et al.*, 2010; Mellen *et al.*, 2012; Otani *et al.*, 2013). *In vivo* experiments confirmed 5mC but not 5hmC as a target modification for those readers, so far (Baubec *et al.*, 2013). Further studies will be necessary to investigate the functional relevance of 5hmC readers and to what extent they contribute to 5hmC-specific transcriptional regulation. It will be interesting to test a possible role for 5hmC as a stable mark in differentiating monocytes and to identify possible 5hmC readers in the corresponding 5hmC-accumulating regions.

Gene ontology analyses of the four 5hmC subsets confirmed the immunological background of monocytes and dendritic cells, displaying mainly GO terms related to immunological processes (3.3.2.2). Along with common functions of MO and DC as positive regulators of the immune system (Abbas *et al.*, 2007), GO terms related to distinct functions of either MO or DC were also enriched. The majority of GO terms covered genes involved in immune response and inflammation, which are common characteristics of monocytes (Ginhoux and Jung, 2014). Those terms were mainly enriched in clusters correlated with active demethylation and PU.1 recruitment. Interestingly, previous studies reported an important role of PU.1 in the regulation of genes involved in inflammation (Rehli, 2000; Smith *et al.*, 1998) and PU.1 binding sites are enriched in regulatory regions of inflammatory pathways in MAK (Ghisletti *et al.*, 2010; Heinz *et al.*, 2010; Turkistany and DeKoter, 2011). Therefore these sets may represent PU.1 target genes that regulate MO function. Lymphocyte activation, a common feature of DCs, was predominantly enriched in one of the two clusters. Given the important function of PU.1 as a key driver of DC development (reviewed in Carotta *et al.*, 2010b), this cluster may further comprise a set of PU.1 target genes that are activated upon differentiation into DC. In both sets, active demethylation correlated with an overall repression of gene expression (chapter 3.3.2.3) in the corresponding regions. This finding is interesting, since active demethylation is mostly but not exclusively associated with transcriptional activation (Wu and Zhang, 2014). However, our observation is supported by earlier analyses of the same expression data set, which associated active demethylation with gene activation as well as with gene repression (Klug *et al.*, 2010). The accumulation of PU.1 at these sites may indicate an involvement in the down-regulation of those genes. In support of that theory, studies in other myeloid cells show, that PU.1 is not only involved in the regulation of gene expression (Carotta *et al.*, 2010b; Ghisletti *et al.*, 2010), but is further able to recruit activating as well as repressing interaction partners (Bai *et al.*, 2005; Imoto *et al.*, 2010; Suzuki *et al.*, 2006). In addition, recent observations in monocyte-derived osteoclasts linked PU.1 to repressive as well as actively transcribed regions. However PU.1 binding was associated with hypermethylation in the context of transcriptional repression (de la Rica *et al.*, 2013). This feature differs from the observations in MO differentiating into DC, presented in this work, as well as from studies in monocyte-derived MAK (Pham *et al.*, 2013), where PU.1 binding is predominantly associated with hypomethylation. This discrepancy may indicate different regulatory roles of PU.1 in related, yet different cell types. It will be interesting to assess the surrounding chromatin structure, identify further TF bound at the sites of demethylation and

possible binding partners of PU.1 in the different cell types in the future. This may provide further insight into the regulatory functions of PU.1 and its association and cooperation with cell type-specific factors in different cell types. In macrophages this has previously been done and revealed distinct TF motif signatures corresponding to differentiation-associated epigenetic changes (Pham *et al.*, 2012; Pham *et al.*, 2013). Nevertheless, our data indicate that PU.1 may be involved in the regulation of transcriptional changes during differentiation of MO into DC. However, it has to be noted that the expression analyses using Homer, associate 5hmC enrichment with expression changes of the nearest gene. The next gene can be located at a certain distance from the demethylated CpGs, which may introduce a bias for false positive genes flanking large intergenic regions (Lowe *et al.*, 2007; Taher and Ovcharenko, 2009). Additionally we can only assume that local demethylation affects the closest gene. TF binding for example usually spans from proximal to several kb distal binding events and consideration of the proximal peaks, only, may neglect effects of distal binding events (McLean *et al.*, 2010). If the nearest gene is affected by the demethylation it is possible that it is not or not yet associated with transcriptional changes. These regions may be pre-marked for later activation or exhibit other regulatory functions (Buecker and Wysocka, 2012). The other two sets of regions were mostly depleted of the major GO terms and did not enrich any of the terms uniquely (3.3.2.2). This finding may result from the nearest gene approach used by Homer gene ontology analyses as was discussed above and thus represent only a fraction of the affected genes. Additionally, for the analysis mainly major GO terms were considered. Since PU.1 did also not accumulate to those regions it is tempting to speculate that those sites are a heterogeneous set of regions maintaining important, but less prominent genes, which did not appear in the top GO terms, and are regulated independent of PU.1. When taking the gene expression analyses (3.3.2.3) into account, these sets were associated with an overall increase in gene expression. As was discussed in detail earlier in this chapter, the demethylation process in both sets presumably paused at the 5hmC mark. In combination with the transcriptional activation these results support the idea that 5hmC attracts specific readers, which are able to regulate transcription (Delatte *et al.*, 2014; Iurlaro *et al.*, 2013; Spruijt *et al.*, 2013; Yildirim *et al.*, 2011). In support of this finding Yildirim *et al.* identified MBD3 as a potential reader of 5hmC *in vitro* and observed a repressive gene regulatory function in association with the MBD3/NuRD complex (Yildirim *et al.*, 2011). This notion could not be reproduced by other groups, who did not find evidence for 5hmC binding *in vitro* (Hashimoto *et al.*, 2012; Spruijt *et al.*, 2013). Further investigation of 5hmC function in differentiating monocytes will clarify, if there is one (or more) possible reader of 5hmC

directly recruited to activate the expression at those loci, or if an accumulation of 5hmC simply repels 5mC-binding proteins in order to maintain an unmethylated, de-repressed state (Shearstone *et al.*, 2011; Valinluck *et al.*, 2004). In this case other methylation-sensitive DNA binding proteins may benefit from a conversion of 5mC to 5hmC and be recruited to the unmethylated regions to activate transcription.
Taken together, the global screen of 5hmC revealed functionally distinct clusters of regions, which are partly associated with the key transcription factor PU.1 and differential gene expression profiles. We were able to further characterize the previously detected bimodal regulation of gene expression (Klug *et al.*, 2010) and correlate it to distinct 5hmC dynamics, TF binding, as well as biological functions.

To deepen the understanding of PU.1 recruitment to its target genes and the association with active demethylation events, we analyzed PU.1 binding patterns on a genome-wide level. The results revealed different binding dynamics of PU.1 at several sets of regions during monocyte differentiation and led to the assumption that each cluster of regions represents different target sites with varying functions of PU.1. On the basis of PU.1 dynamics the six clusters could be divided into two groups (3.3.3.1). One group of PU.1 target sites showed increased PU.1 binding during the differentiation time course and simultaneously gained 5hmC at PU.1 sites. With rising PU.1 levels the 5hmC pattern became more distinct and localized next to the PU.1 binding site, which is in line with observations from other TFs (Lister *et al.*, 2013; Yu *et al.*, 2012). The accumulation of both marks was accompanied by the distinct loss of 5mC (and 5hmC) confirming the direct correlation between PU.1 and demethylation events as was observed in previous chapters. This is in line with earlier findings in MAK where our group detected only small overlap between PU.1 and 5mC and assumed a possible recruitment of the demethylation machinery by PU.1, leading to local demethylation (Pham *et al.*, 2013). In support of this notion de la Rica *et al.* recently reported a direct interaction between PU.1 and TET2 at demethylated genes resulting in gene expression changes (de la Rica *et al.*, 2013). Interestingly observations in one cluster of PU.1 target regions argued against this theory, and confirmed earlier observations in a subset of 5hmC-enriched regions (see section 3.3.2.1). Regions in this PU.1 cluster clearly showed demethylation events at the monocyte stage, but recruited PU.1 primarily upon the transition into dendritic cells. This finding indicated, that at those regions other factors than PU.1 are responsible for the initiation of the demethylation process and that PU.1 was possibly either recruited by those factors or benefited from the opening of those regions. Targeted depletion of PU.1 or TET2 at selected loci will be helpful to validate this finding. A recent study from

our group showed that most potential binding sites of PU.1 in hematopoietic cells are not occupied, which is in part due to their location in inaccessible heterochromatin (Pham *et al.*, 2013). Further generation of global chromatin maps in combination with motif analyses may help to understand, if those sites are located in heterochromatic regions and therefore not accessible for PU.1. TET2 binding to DNA was shown to be dependent on PU.1 levels in differentiating osteoclasts (de la Rica *et al.*, 2013), indicating that PU.1 is prerequisite for TET2 recruitment to demethylated regions in this cell type. The aforementioned distinct chronology of PU.1 and 5hmC in a subset of PU.1 target region in dendritic cells may not necessarily display a discrepancy but rather one possible function of PU.1. It may still be possible that PU.1 and 5hmC exhibit different spatio-temporal patterns in different cell types or even at different target regions of the same cell type, depending on the cooperation of local factors. Cooperation was shown for PU.1 binding patterns in other cell types and may determine PU.1 function at bound loci (Heinz *et al.*, 2010; Pham *et al.*, 2012; Pham *et al.*, 2013). In the remaining subsets of this group it is still not clear, which mark precedes which and PU.1 may initiate the demethylation process in association with TET2 recruitment at those sites. Characterization of selected regions in the sub-clusters may give further insight into the chromatin context and the involvement of other factors. Further analyses of PU.1 and 5hmC dynamics in monocyte progenitor cells will help to solve the question of the “chicken and the egg” at regions in these clusters. These experiments may help to confirm a dual role for PU.1 depending on the set of target sites or a general recruitment of PU.1 to demethylated sites in differentiating monocytes.

The other group of regions lost PU.1 peaks present in monocytes during the differentiation into DC. The rare 5hmC abundance as well as the absence of 5mC (and 5hmC) in the vicinity of PU.1 binding sites supported the idea that those regions represent later stages of the demethylation process where the demethylation is completed and PU.1 is dispensable and removed from those sites. This theory implies a functional role of PU.1 in the demethylation process as was hypothesized by Pham and colleagues who suggested an involvement of sequence-specifically recruited TFs like PU.1 in the initiation of this process (Pham *et al.*, 2013). In differentiating monocytes there is so far no evidence for this theory, but it cannot be excluded either. The close association to demethylation events observed in differentiating monocytes and the data from other studies support an involvement of PU.1 in the demethylation process at least in subsets of PU.1 target genes. To date it is still not clear if PU.1 is capable of the initiation of the demethylation process by the recruitment of the demethylation machinery, or if it is required for the stabilization of other factors and

complexes in this cell system. Since cooperation of TFs is an important feature for binding site selection (Pham *et al.*, 2013; Tallack *et al.*, 2010) it may also be possible that PU.1 binding recruits other relevant TFs or selectively blocks their binding to the site of demethylation. In support of this theory several studies reported a dynamic change of TF composition at cell type specific elements during development (Lin *et al.*, 2010; Pham *et al.*, 2012; Pilon *et al.*, 2011). The loss of PU.1 at this group of regions may thus indicate the establishment of a different TF signature at the demethylated sites upon transition into a new cell type, which does not include PU.1. Further analyses of motif occurrence and occupancy by other TF may resolve these questions.

During monocyte differentiation we observed distinct PU.1 patterns upon the transition into a new cell type. The gain or loss of PU.1 was associated with distinct 5hmC dynamics resulting in or maintaining a demethylated state at the PU.1 target regions.
In general, key regulators are recruited to cell type-specific sites, such as enhancers where they regulate cell type-specific gene expression (Heintzman *et al.*, 2009; Ong and Corces, 2011; Pham *et al.*, 2012). In differentiating monocytes we observed similar binding patterns for the TF PU.1 at regulatory elements. We divided two sets of actively transcribed promoters and enhancers into groups of MO- or DC-specific elements and assayed surrounding PU.1 and 5hmC distribution. The results showed that whereas DC-specific promoter regions acquired PU.1, MO-specific promoters exhibited receding levels of PU.1 binding during monocyte differentiation (3.3.3.2.1). This has previously been observed in macrophages, too, where PU.1 was lost at promoter regions during MO to MAK differentiation and PU.1 displayed distinct cell type-specific binding patterns in both cell types (Pham *et al.*, 2012). Consistent with the PU.1 patterns at promoters, cell type-specific, active enhancer regions acquired PU.1 at DC-specific sites and lost the TF at MO-specific regions upon transition into a new cell type. Taken together, PU.1 is mainly targeted to promoter and promoter-distal regulatory regions that are activated in a cell type-specific manner. The active nature of the regions in both enhancer and promoter sets may indicate a functional relevance of the corresponding PU.1 dynamics. Sérandour and colleagues reported dynamic 5hmC patterns at cell type-specific enhancer regions in mice, which is associated with active histone marks, chromatin opening and an intermediate CpG density (Serandour *et al.*, 2012). The authors suggested an important role of demethylation in the regulation of enhancer elements. Consistent with this report, cell type-specifically activated enhancer regions in differentiating monocytes were associated with low to intermediate CpG-content and distinct 5hmC patterns (3.3.3.3). In contrast the 5hmC distribution was less dynamic than observed at other PU.1 target regions and indicated that PU.1

recruitment to cell type-specific enhancer or promoter elements did not significantly alter the demethylated state of those regions. Yet it may be involved in the maintenance of an active transcriptional state. In support of this finding the loss of PU.1 binding at macrophage-specific enhancer regions was associated with reduction of the active enhancer histone signature at those loci in earlier studies of macrophages. This finding indicated that PU.1 may be responsible for the maintenance of active enhancer histone marks (Heinz *et al.*, 2010; Pham *et al.*, 2012) and supported the hypothesis of PU.1 participating in the transcriptional regulation of the transition into a new cell type as was discussed earlier in this chapter. During monocyte differentiation PU.1 is removed from regions that characterize the former cell type. The loss of PU.1 may indicate the decommissioning of the regions which was shown to proceed with the loss of activating histone marks at enhancer region in differentiating mESC (Whyte *et al.*, 2012). This is consistent with the theory that PU.1 may be involved in transcriptional regulation via histone marks, but further investigation will be necessary to validate this hypothesis. Mapping of histone marks in the vicinity of the cell type-specific enhancer and promoter regions may help to get insights into the processes at regulatory regions. Given the vast abundance of PU.1 in myeloid cells like MO and MAK (Carotta *et al.*, 2010b; Pham *et al.*, 2012) its specific binding to target sites may be dependent on the cooperative binding of other TF (Pham *et al.*, 2013). It is still not clear if PU.1 actively interferes in transcriptional regulation of the regulatory target genes. To address the question a conditional knockout may be helpful to assess the gene expression levels in the absence (or presence) of PU.1 at example regions. It has to be noted that the depletion of a key hematopoietic TF in hematopoietic cells is difficult to interpret. Secondary effects on cell homeostasis and development may interfere with global mechanisms, but a careful interpretation of the approach may still help to resolve this question.

In summary the global distribution of PU.1 showed distinct binding profiles at different PU.1 target regions. These dynamics were again directly correlated with demethylation events and represented target gene specific PU.1 patterns accompanying the transition into a new cell type. Our results present evidence for a distinct chronological succession of TF binding and demethylation events at a subset of effector regions. Dynamic PU.1 binding was further linked to the active usage of functionally relevant elements, which supported an involvement of PU.1 in hematopoietic cell differentiation.

The data presented in this work contributed to the understanding of the active DNA demethylation mechanism and revealed distinct patterns of transcription regulation upon cell fate decisions in hematopoietic cells. We could show first evidence of PU.1

as a beneficiary of demethylation events at certain subsets of PU.1 target regions and further confirmed a direct correlation between this master transcription factor and active DNA demethylation.

5 Material and Methods

5.1 Material & Equipment

5.1.1 Technical devices

8-Channel PipettorImpact2 Equalizer 384	Thermo Fisher Scientific, Hudson, USA
Autoclave	Technomara, Fernwald, Germany
Bioanalyzer 2100	Agilent Technologies, Böblingen, Germany
Biofuge fresco	Heraeus, Osterode, Germany
Caliper LapChip XT	Perkin Elmer, Waltham, USA
Casy Cell Counter	Innovatis/Roche, Basel, Schweiz
Centrifuges	Heraeus, Hanau; Eppendorf, Hamburg, Germany
Chemi Doc XRS+ System	Bio-Rad, Hercules, USA
Covaris S2	Covaris, Woburn, USA
Electrophoresis equipment	Biometra, Göttingen, Germany
Electrophoresis equipment	Bio-Rad, Hercules, USA
Fast-Blot machine	Agfa, Köln
Heat sealer	Eppendorf, Hamburg, Germany
Heatblock	Stuart Scientific, Staffordshire, UK
Hemocytometer	Marienfeld, Lauda-Königshofen, Germany
Illumina HiSeq1000	Illumina, San Diego, USA
Incubators	Heraeus, Hanau, Germany
Intelli-Mixer RM-2L	Elmi-Tech, Riga, Latvia
J6M-E centrifuge	Beckmann, Munich, Germany
Laminar air flow cabinet	Heraeus, Hanau, Germany
LSRII	BD, Heidelberg, Germany
Luminometer (Sirius)	Berthold Detect. Systems, Pforzheim, Germany
Magnetic Particle Concentrator	Life Technologies, Carlsbad, USA
MassARRAY Compact System	Sequenom, San Diego, USA
MassARRAY MATRIX Liquid Handler	Sequenom, San Diego, USA
MassARRAY Phusio chip module	Sequenom, San Diego, USA
Megafuge 3,0 R	Heraeus, Osterode, Germany
Microscopes	Zeiss, Jena, Germany
Microwave Privileg 8020	Privileg
Minigel gelelectrophoresis equipment	Biometra, Göttingen
Multifuge 3S-R	Heraeus, Osterode, Germany
Multipipettor Multipette plus	Eppendorf, Hamburg, Germany
NanoDrop 1000	PeqLab, Erlangen, Germany
PCR Thermocycler PTC-200	MJ-Research/Biometra, Oldendorf, Germany

pH-Meter	Knick, Berlin, Germany
Picofuge	Heraeus, Osterode, Germany
Pipetboy	Integra Biosciences, Fernwald, Germany
Pipettes	Gilson, Milddleton, USA
Pipettes	Eppendorf, Hamburg, Germany
Power supplies	Biometra, Göttingen, Germany
Qubit 2.0 Flourometer	Life Technologies, Carlsbad, USA
Realplex Mastercycler epGradient S	Eppendorf, Hamburg, Germany
Rocking plattform HS250	IKA Labortechnik, Staufen, Germany
Rotilabo®-mini-centrifuge	Carl Roth, Karlsruhe, Germany
Sigma 2 – Sartorius	Sartorius, Göttingen, Germany
Sonifier 250	Branson, Danbury, USA
Sonorex Ultrasonic Bath	Bandelin, Berlin, Germany
Sorvall RC 6 plus	Thermo Fisher Scientific, Hudson, USA
Thermomixer	Eppendorf, Hamburg, Germany
Thermomixer compact	Eppendorf, Hamburg, Germany
Typhoon 9200	Molecular Dynamics, Krefeld, Germany
Typhoon™	Amersham Biosciences, Germany
Ultracentrifuge Optima L-70	Beckman, Munich, Germany
UVT-20 M Transilluminator	Herolab, Wiesloch, Germany
Veriti 384 well Thermal Cycler	Life Technologies, Carlsbad, USA
Vortexer	Scientific Industries Ink., Bohemia, USA
Water purification system	Millipore, Eschborn, Germany
Waterbath	Julabo, Seelstadt, Germany

5.1.2 Consumables

384-well PCR plate	Thermo Fisher Scientific, Hudson, USA
8-channel pipettor tips Impact 384	Thermo Fisher Scientific, Hudson, USA
Adhesive PCR sealing film	Thermo Fisher Scientific, Hudson, USA
Amersham Hyperfilm™ ECL	GE Healthcare, Chalfont St Giles, UK
AMPure XP Magnetic Beads	Beckman Coulter Genomics, Krefeld, Germany
Cell culture dishes	Nunc/Thermo Fisher Scientific, Hudson, USA
Cell culture dishes	Greiner Bio-one, Frickenhausen, Germany
Cell culture flasks and pipettes	Costar, Cambridge, USA
CLEAN resin	Sequenom, San Diego, USA
Cryo tubes	Corning, Corning, USA
Disposable scalpel	Feather, Osaka, Japan
DNA LoBind Tubes	Eppendorf, Hamburg, Germany
Einweg-Pasteurpipette	Carl Roth, Karlsruhe, Germany
ep Dualfilter T.I.P.S.®	Eppendorf, Hamburg, Germany
GenePulser Electroporation Cuvette (1mm)	Bio-Rad, Hercules, USA
Heat sealing film	Eppendorf, Hamburg, Germany

LabChip XT DNA 300 assay kit	PerkinElmer, Waltham, USA
MATRIX Liquid Handler D.A.R.Ts tips	Thermo Fisher Scientific, Hudson, USA
Micro test tubes (0.5 ml, 1.5 ml, 2 ml)	Eppendorf, Hamburg, Germany
Micro test tubes DNALoBind (0.5, 1.5, 2 ml)	Eppendorf, Hamburg, Germany
Multiwell cell culture plates and tubes	Falcon, Heidelberg, Germany
Nitrocellulose Membrane (0.45 µm)	Life Technologies, Carlsbad, USA
Non-Adhesive sealing film (LightCycler)	Roche, Basel, Switzerland
nProteinA Sepharose 4 FastFlow	GE Healthcare, Munich, Germany
Nylon Transfer membrane	MSI, Westboro, USA
PCR plate Twin.tec 96 well	Eppendorf, Hamburg, Germany
Pipettes	Costar, Cambridge, USA
Plug molds	Bio-Rad, Hercules, USA
ProteinA Dynabeads	Life Technologies, Carlsbad, USA
rProteinA Sepharose 4 FastFlow	GE Healthcare, Munich, Germany
Sepharose CL-4 Beads	Sigma-Aldrich, St. Louis, USA
SpectroCHIP bead array	Sequenom, San Diego, USA
Sterile combitips for Eppendorf multipette	Eppendorf, Hamburg, Germany
Sterile micropore filters	Millipore, Darmstadt, Germany
Sterile plastic pipettes	Costar, Cambridge, USA
Syringes and needles	Becton Dickinson, Heidelberg, Germany
Teflon foils	Heraeus, Hanau, Germany
Tubes (5ml, 15 ml, 50 ml, 220 ml)	Falcon, Heidelberg, Germany
Whatman® Chromatography Paper	Sigma-Aldrich, St. Louis, USA

5.1.3 Chemicals

All chemicals were purchased from Sigma-Aldrich (St. Louis, USA) or Merk (Darmstadt, Germany) if not otherwise stated.

5.1.4 Molecular tools

5.1.4.1 Enzymes and reagents

Agarase	NEB, Frankfurt, Germany
Amersham™ ECL™ Prime Western Blotting Detection Reagent	GE Healthcare, Chalfont St Giles, UK
Ammoniumperoxidisulfat	Merck, Darmstadt, Germany
Beetle-Juice KIT	PJK, Kleinblittersdorf, Germany
Bio-Safe™ Coomassie	Bio-Rad, Hercules, USA
Biozym Plaque GP Agarose	Biozym, Oldendorf, Germany

BSA	Sigma-Aldrich, St. Louis, USA
Cell Culture Medium Supplement	Life Technologies, Carlsbad, USA
dATP Solution	NEB, Frankfurt, Germany
Deoxynucleotide (dNTP) Solution Set	NEB, Frankfurt, Germany
DNA Ladder 1 kb plus	Life Technologies, Carlsbad, USA
DNA ladder 50 bp	NEB, Frankfurt, Germany
DNA ladder Generuler 50 bp	Thermo Fisher Scientific, Waltham, USA
DNA light loading dye (6x)	NEB, Frankfurt, Germany
Dual-Luciferase® Reporter Assay System	Promega, Madison, USA
Enzymatics enzymes (Library prep)	NEB, Frankfurt, Germany
Ethidium bromide	Sigma-Aldrich, St. Louis, USA
Fermentas DNA loading dye (6x)	Thermo Fisher Scientific, Waltham, USA
Fetal Bovine Serum	Gibco/Life Technologies, Carlsbad, USA
	GE Healthcare, Chalfont St Giles, UK
Glycogen	Ambion/Life Technologies, Carlsbad, USA
Lipofectamin 2000	Invitrogen/Life Technologies, Carlsbad, USA
M-MLV Reverse Transcriptase kit	Promega, Madison, USA
N,N,N',N'-Tetramethylethylenediamine	Sigma-Aldrich, St. Louis, USA
NEXTflex™ DNA Barcodes	HiSS Diagnostics, Freiburg, Germany
Nuclease-free water	Gibco/Life Technologies, Carlsbad, USA
Oligo Only Kit	Illumina, San Diego, USA
peqGOLD Protein-Marker VII	Peqlab, Erlangen, Germany
Phusion Hot Start High-Fidelity DNA Polymerase	NEB, Frankfurt, Germany
Phusion Taq Polymerase	Thermo Fisher Scientific, Waltham, USA
Phorbol 12-myristate 13-acetate	Sigma-Aldrich, St. Louis, USA
Ponceau S-Solution	AppliChem, Darmstadt, Germany
Precision Plus Protein™ Kaleidoscope™	Bio-Rad, Hercules, USA
Proteinase K	Roche, Basel, Switzerland
QuantiFast SYBR green	Qiagen, Hilden, Germany
Random Decamers	Life Technologies, Carlsbad, USA
Restriction endonucleases	NEB, Frankfurt, Germany
	Roche, Basel, Switzerland
RNase	Roche, Basel, Switzerland
RNase A	Sequenom, San Diego, USA
Rotiphorese® Gel 30 (37.5:1)	Carl Roth, Karlsruhe, Germany
SAM (S-adenosyl-methionine)	NEB, Frankfurt, Germany
Shrimp Alkaline Phosphatase (SAP)	Sequenom, San Diego, USA
*Sss*I CpG methyltransferase	NEB, Frankfurt, Germany
Taq DNA Polymerase	Roche, Basel, Switzerland
T-Cleavage MassCleave Reagent kit	Sequenom, San Diego, USA
Transcriptor High fidelity cDNA Synthesis kit	Roche, Basel, Switzerland
Trypan blue solution	Merck, Darmstadt, Germany
Tween® 20	Sigma-Aldrich, St. Louis, USA

5.1.4.2 Antibiotics

Antibiotic	Stock concentration	Dilution	Final concentration	Company
Ampicillin	100 mg/ml	1:1000	100 µg/ml	Ratiopharm, Ulm, Germany
Zeocin	100 mg/ml	1:4000	25 µg/ml	Invitrogen, Karlsruhe, Germany

5.1.4.3 Antibodies

Antibody for FACS	Product	Company
CD14-Fitc	My4a	Coulter Clone, Brea, USA
CD1a-PE	T6-RD1	Coulter Clone, Brea, USA
DAPI	4′,6-Diamidino-2-phenylindole dihydrochloride	Sigma, St. Louis, USA

Antibody for ChIP	Company
hmC	ActiveMotif, Carlsbad, USA
PU.1	SantaCruz, Dallas, USA
H3K4me1	Abcam, Cambridge, UK
IgG	Merck, Darmstadt, Germany

Antibody for Western blotting	Source/Company
TET2	Dr. Olivier Bernard, Institut Gustave Roussy, Villejuif, France
b-Actin	Sigma-Aldrich, St. Louis, USA
TDG	Prof. Dr. Primo Schär, University of Basel, Switzerland
MBD4	Diagenode, Seraing, Belgium
anti-rabbit IgG-HRP	Dako, Glostrup, Denmark

5.1.4.4 Molecular biology kits

Agilent DNA 1000	Agilent, St. Clara, USA
Agilent High Sensitivity DNA kit	Agilent, St. Clara, USA
Agilent RNA 6000 Nano kit	Agilent, St. Clara, USA
Amaxa® Human Monocyte Nucleofector® Kit	Lonza, Basel, Switzerland
Blood & Cell Culture DNA Kit	Qiagen, Hilden, Germany
Clean and concentrator(TM) kit	Zymo Research, Irvine, USA
DNeasy Blood & Tissue Kit	Qiagen, Hilden, Germany
EndoFree Plasmid Midi Kit	Qiagen, Hilden, Germany
EpiTect Bisulfite Kit	Qiagen, Hilden, Germany
EZ DNA methylation kit	Zymo Research, Irvine, USA
Genomic DNA Clean & Concentrator	Zymo Research, Irvine, USA
Hydroxymethyl Collector™	ActiveMotif, Carlsbad, USA
MinElute Gel Extraction Kit	Qiagen, Hilden, Germany
MinElute PCR Purification Kit	Qiagen, Hilden, Germany
MycoAlert™ Mycoplasma Detection Kit	Lonza, Basel, Switzerland

NucleoSpin® Plasmid	Macherey-Nagel, Düren, Germany
QIAamp DNA Micro/Mini Kit	Qiagen, Hilden, Germany
QIAGEN Plasmid Plus Kit	Qiagen, Hilden, Germany
QIAquick Nucleotide Removal Kit	Qiagen, Hilden, Germany
QIAquick PCR Purification Kit	Qiagen, Hilden, Germany
QuantiFast SYBR® Green PCR Kit	Qiagen, Hilden, Germany
Qubit® dsDNA HS Assay Kit	Life technologies, Carlsbad, USA
Qubit® ssDNA Assay Kit	Life technologies, Carlsbad, USA
Quest 5-hmC Detection Kit™	Zymo Research, Irvine, USA
REPLI-g Mini/Midi Kit	Qiagen, Hilden, Germany
Rneasy Mini Kit	Qiagen, Hilden, Germany

5.1.5 Oligonucleotides

Oligonucleotides for qRT-PCR and qPCR were synthesized by Eurofins MWG (Ebersberg, Germany). Oligonucleotides for the MassARRAY system were purchased from Sigma-Aldrich (St. Louis, USA).

5.1.5.1 Oligonucleotide primers

5.1.5.1.1 qPCR primers

Gene		**Primer sequence** (F, forward; R, reverse)	**Assay**
ACP2	*F*	5' CAGTAGAGAAGCAGAGGCGTCAG 3'	a
	R	5' AATTGGCTCGTTCTCCCACC 3'	
C9ORF78/USP20	*F*	5' AGTACGCGTGTTTGGATTTGGG 3’	a,b,c
	R	5' TGCTCTGAACACTCAGCTCCAC 3’	
CACNA1B	*F*	5' AAGTGATGAGAAATACCAGTGCGG 3'	b
	R	5' CAAGCCACAATGAGCCCTTCC 3'	
CCL13	*F*	5' CTAGAAAAGTCTTTGGTGCCCAG 3'	a,e,g
	R	5' CTTGGAACTCTCAGAGGACCTTG 3'	
CCL13	*F*	5' CAAGTGAGATTTAAGAGGAGGGAGGG 3'	d
	R	5' AGTGTCTGGTACATATTCGGTGCT 3'	
CCL13	*F*	5' ACTTGGTCAACGCCCTGCT 3'	b,c
	R	5' TCTGCTCCTCTGGCTGTTCC 3'	
DNase1L3	*F*	5' GGGCTCACCTTCACAATGACATCC 3'	a,g
	R	5' TCTGCTCCTTCAACGTCAGGTC 3'	
DNase1L3	*F*	5' CGCCATCCTCCACAAAGTCAC 3'	b
	R	5' GACCACTTCCGAGTTCATCCAC 3'	

Empty	*F*	5´ GAAACCCTCACCCAGGAGATACAC 3´	b,g
	R	5´ TGCAGTGGGACTTTATTCCATAGAAGAG 3´	
FBX041	*F*	5' TTCCAAATGAAAGTTGGCAGGAC 3'	b
	R	5' CAAAGGATCAGGTCTCAGAACCA 3'	
GAPDH	*F*	5´ AGGCTGGATGGAATGAAAGGCAC 3´	c
	R	5´ CTCCCACAAAGGCACTCCTG 3´	
HCGP39	*F*	5' GATCTGATCTGCAGCCCACC 3'	d
	R	5' TGCACCATCAGAACCAGAACC 3'	
HCGP39	*F*	5' CTTGTCGGATTGTAGGTCATAAG 3'	d
	R	5' GACAAGAAGGGTGTGGTCTAATA 3'	
HOXB1	*F*	5' CCTCGGAGAGGAGATCAGCA 3'	a,b
	R	5' CCATATCCTCCGCAGCATCC 3'	
KCNE2	*F*	5' CATTCAGAAGGTCACAGAGGGCA 3'	a
	R	5' CCACAAACTCACTCTCAGGCGA 3'	
LRGUK	*F*	5' TGACTTGGTGCAGAAACTTGCT 3'	b
	R	5 'TGGTATCTGTTGCTAAGTCTTGGG 3'	
MMP7	*F*	5' GCTGTGACATACCTGAGCCTG 3'	a
	R	5' CTATGCGACTCACCGTGCTG 3'	
RSRC1	*F*	5' CTGCCATGTGAGAAAGACCTG 3'	a
	R	5' CAGAAGGAGGTGGAATGAGGA 3'	
Spi1	*F*	5´ CCAGTCACCACAGGAAGCATG 3´	d
	R	5´ CGTTTCTCTGGGCCGCTGTG 3´	
STAT5	*F*	5' GCACAGACTCTGCATCCTCTTCTC 3'	a,b,c,g
	R	5' AGTCCTGCTTCCTCTGCCCA 3'	
TSSC1	*F*	5' TCCCACGGAGTTGACTGATCCA 3'	a
	R	5' GTGCCGGTGGTGATCAATGTG 3'	
TSSC1	*F*	5' GACTTCCCTGGTCCTTTCCC 3'	b
	R	5' CCATGAGTGGGAGGTGTGAG 3'	
UNQ6496	*F*	5' AAACAGAGGCAGCTGACTCTCAGG 3'	b
	R	5' TGCGCCCTTCACAGATGGAG 3'	
VRK2	*F*	5' TTTCCGTTCCTCTGTGTTTATGCT 3'	a
	R	5' ACCTCTTCTGACACCCATCCTC 3'	

[a] hMeDIP-qPCR

[b] GGSD

[c] MSRE

[d] ChIP-qPCR

[e] hmC-enrichment qPCR

[f] Mass spectrometry analysis of bisulfite-converted DNA

[g] Amplicons described in Ref. (Klug *et al.*, 2010)

5.1.5.1.2 Sequencing primers

Name	**Primer sequence**
pCpG-LUC_Seq_AS	5' CACAGACATCTCAAAGTATTCAGC 3'
pCGS	5' TAAATCTCTTTGTTCAGCTCTCTG 3'

5.1.5.1.3 Primer for molecular cloning

Name/Gene		**Primer sequence** (F, forward; R, reverse)
CACNA1C_PstI	*F*	tgtatctgcagGAAGCTGAGAACTGAATCAATGAATGCC
CACNA1C_SpeI	*R*	tgataactagtATTACACAGACAAATTCCACTTGGTCCC
RSRC1_NsiI	*F*	tgtatatgcatGGAGATTTCCCGTCTGTGCCTG
RSRC1_XbaI	*R*	tgatatctagaGGAAGCTCAGCATTAAGAAGCAAGGG
TSSC1_PstI	*F*	tgtatctgcagATAAATGTCAAGGCAAAGAGGGAAAGGG
TSSC1_XbaI	*R*	tgatatctagaTGAGTGGGAGGTGTGAGCTCGT
ADAP1_NsiI	*F*	tgtatatgcatAGCATCTCCCTAGTCCAGAGCC
ADAP1_XbaI	*R*	tgatatctagaCTGAGCCTGGATGTTGAAGACCC
SMAD7_NsiI	*F*	tgtatatgcatCTAACACCCAGCTCACCCAGGA
SMAD7_XbaI	*R*	tgatatctagaCTAGAAGATGAGGCAGGAGGCAGAG
LINC00111_NsiI	*F*	tgtatatgcatGACTTGAAGCCGGGACACTTAAGGA
LINC00111_XbaI	*R*	tgatatctagaGCCCTGCGGATTTCAGAGAAGAC
PIM3_PstI	*F*	tgtatctgcagCCAGACTGAGTCCTTCTCAGCTCC
PIM3_XbaI	*R*	tgatatctagaAGGCCCTTCCTGCTTGACCA
VRK2_PstI	*F*	tgtatctgcagAATTGCCATTCTCTGGGAACTATGAGTG
VRK2_XbaI	*R*	tgatatctagaGCAGCAATGTGAAGGAGAAACTTTGAG

5.1.5.1.4 qRT-PCR primers

Gene		**Primer sequence** (F, forward; R, reverse)	**Assay**
CCL13	*F*	5' GAAGATCTCCTTGCAGAGGCT 3'	qRT-PCR [g]
	R	5' GGGTCAGCACAGATCTCCTTG 3'	
CD1a	*F*	5' TATTCCGTATACGCACCATTCGGTC 3'	qRT-PCR
	R	5' CAGCCTCCTGTCACCTGTATCTC 3'	
HELLS	*F*	5' TTCAGCAGGCTTGATGGGTC 3'	qRT-PCR
	R	5' AGATTAATGCCCAGGCCACCA 3'	
HPRT	*F*	5' AAGTTTGTTGTAGGATATGCCC 3'	qRT-PCR [g]
	R	5' GAACATTGATAATTTTACTGGCG 3'	
MBD4	*F*	5' TGGCTCTGAAATGGACAACAACTG 3'	qRT-PCR
	R	5' TTCTATCTGTGTTCGTGGGATGGT 3'	
OGT	*F*	5' TCCGAGTGAAGGTGATGGCAG 3'	qRT-PCR
	R	5' TTCGAGCGCCCTTAGTATAGCC 3'	
TDG	*F*	5' AAATCCATGCAGCAGTGAACCT 3'	qRT-PCR
	R	5' CACTGCCCATTAGGAATGCCA 3'	

TET1	*F*	5' GCTCTCATGGGTGTCCAATTGCT 3'	qRT-PCR
	R	5' ATGAGCACCACCATCACAGCAG 3'	
TET2	*F*	5' AAGGCTGAGGGACGAGAACGA 3'	qRT-PCR
	R	5' TGAGCCCATCTCCTGCTTCCA 3'	
TET3	*F*	5' CCTGCCGATGACAAGCTGGA 3'	qRT-PCR
	R	5' GAGTTCCCGGATAGAGGCGA 3'	

5.1.5.1.5 MassARRAY primers

Amplicon		**Primer sequence**
Epi00103_STAT5A.1	10F	aggaagagagAGTTGTTTGGTTTTGTGTGTTTTTT
	T7R	cagtaatacgactcactatagggagaaggctAAAAAAATCCTACTTCCTCTACCCA
Epi00104_STAT5A.2	10F	aggaagagagAAAGTGATTTTTTTGAAGAGTGGTG
	T7R	cagtaatacgactcactatagggagaaggctTCCAAAAAAACAAATCAAAACCTAA
Epi00108_STAT5A.1	10F	aggaagagagTTTATAGGGAGGTATTAGGGTTTGG
	T7R	cagtaatacgactcactatagggagaaggctTCCCTTCTTCAAAAAAAATTCCTAT
Epi00109_CCL13.1	10F	aggaagagagTTTGTGGTTTGAATAGTTAGAAGGA
	T7R	cagtaatacgactcactatagggagaaggctCAACAAACACAAAAACACTACAAAAA
Epi00110_CCL13.2	10F	aggaagagagTTTATGGTTTTTTATGGTGAATGGT
	T7R	cagtaatacgactcactatagggagaaggctAAAATAACTTACCTAACTAAACAAATCCC
Epi00111_P2RY6.1	10F	aggaagagagTTTGGTTATGTTTGGAGTTTGTAGA
	T7R	cagtaatacgactcactatagggagaaggctAAAAAAATACCCTTACCAACCATTT
Epi00112_P2RY6.2	10F	aggaagagagGAGTGTAAATGGTTGGTAAGGGTAT
	T7R	cagtaatacgactcactatagggagaaggctAAATCCCAAATATCTTCAAAAAACC
Epi00123_DNASE1L3.1	10F	aggaagagagTTTTTTAGGAAAGGGGTTTATTTTT
	T7R	cagtaatacgactcactatagggagaaggctAAAATCCAACACTCCAAACACTACT
Epi00124_DNASE1L3.2	10F	aggaagagagGGAGGAGAAGTAGTAGTGGGGTTAG
	T7R	cagtaatacgactcactatagggagaaggctCACCCCAAATACCCTCTAAAATAAA
Epi00125_DNASE1L3.3	10F	aggaagagagTGGTTTATTTTAGAGGGTATTTGGG
	T7R	cagtaatacgactcactatagggagaaggctTCTCTAACAACACACTCCTAATATTTATAC
Epi00147_C9ORF78.2	10F	aggaagagagAGAGGTTTTTGTGAGGAAGTTTTTT
	T7R	cagtaatacgactcactatagggagaaggctACTACCCACACACTTCTATATCTCCTC
Epi00148_C9ORF78.3	10F	aggaagagagGGAATTTTGTTATTTTTTAGGGTGG
	T7R	cagtaatacgactcactatagggagaaggctAAAACCACCATCCTCTAACTCTC
Epi00162_MMP7.1	10F	aggaagagagGGAATTTTAAGTAAGTGGGTTGTGA
	T7R	cagtaatacgactcactatagggagaaggctACAATCACTAACAAAAAACACCAAA
Epi00184_CLEC10A.1	10F	aggaagagagGAAGATAAGGTTGGAAATGGGTTAT
	T7R	cagtaatacgactcactatagggagaaggctACCTCTAATCCTTACAACACAACCA
Epi00185_CLEC10A.2	10F	aggaagagagTATTATTATTTGTGGGAGGTTTGGA
	T7R	cagtaatacgactcactatagggagaaggctAACAAAACTAACCTCAAACCCAACT
Epi00193_HOXB1_01	10F	aggaagagagTTTGAGTTTTATTTGTTTTGGGTGG
	T7R	cagtaatacgactcactatagggagaaggctCCTAAAAATACCCTTCCCCAACTC
Epi00194_HOXB1_02	10F	aggaagagagTTATTTTTTGGGTTAATTAGAAGGAGA
	T7R	cagtaatacgactcactatagggagaaggctAAATTTCTCTTAACCTTCATCCAATC

Epi00195_HOXB1_03	10F	aggaagagagTATTTTTTGGGTTAATTAGAAGGAGA
	T7R	cagtaatacgactcactatagggagaaggctAAATTTCTCTTAACCTTCATCCAATC
SQ00011_fCAB-epi_ctrl	10F	aggaagagagGGGAGGGTTTATTATTTGGTTTTAG
	T7R	cagtaatacgactcactatagggagaaggctCACCAATCACAAAAAAACATCTTAC

5.1.5.2 Plasmids

Internal number	Gene/Plasmid	Name	Backbone
437	*Renilla* luciferase	phRL-TK Vector	pRL (Promega)
865	pCpGL-cmv	pCpG-CMV	pCpG-mcs (Invivogen)
995	pCpGL-EF1	pCpGL-EF1	pCpG-mcs (Invivogen)
1146	CACNA1C	pCpG-CACNA1C-RA	pCpG-mcs (Invivogen)
1147	LINC00111	pCpG-NCRNA00111-RA	pCpG-mcs (Invivogen)
1148	PIM3	pCpG-PIM3-RA	pCpG-mcs (Invivogen)
1149	RSRC1	pCpG-RSRC1-RA	pCpG-mcs (Invivogen)
1150	SMAD7	pCpG-SMAD7-RA	pCpG-mcs (Invivogen)
1151	VRK2	pCpG-VRK2-RA	pCpG-mcs (Invivogen)
1152	TSSC1	pCpG-TSSC1-RA	pCpG-mcs (Invivogen)
1153	ADAP1	pCpG-ADAP1-RA	pCpG-mcs (Invivogen)

5.1.5.3 siRNAs

All siRNAs (siGENOMEsiRNA SMARTpool) were purchased from Thermo Scientific Dharmacon (Lafayette, USA).

5.1.6 Cell lines and E. coli strains

Name	Origin	Source/Company
THP-1	human	DSMZ (ACC 16)
Daudi	human	DSMZ (ACC 78)
HeLa	human	DSMZ (ACC 57)
PIR1 (F- Δlac169 rpoS(Am) robA1 creC510	E.coli strain	Invitrogen, Karlsruhe, Germany

5.2 Methods

5.2.1 Cell culture

5.2.1.1 General cell culture conditions

In general all cells were cultivated according to “Current protocols in molecular biology” (Ausubel, 1987) if not otherwise stated. The cells were handled in a sterile lamina flow hood with sterile consumables only.
Cells lines were cultivated in the indicated medium (see Table 5.1) and primary cells were cultivated in RPMI (Gibco). All media were supplemented with 10% FCS, L-glutamine (2mM), sodium pyruvate (1mM), antibiotics (50 U/ml penicillin, 50 µg/ml streptomycin), 2 ml vitamins, non-essential amino acids and 50 µM ß-mercaptoethanol, if not otherwise stated. All supplements were purchased from Gibco and Biochrome, respectively. FCS (fetal calve serum) was heat inactivated before use for 30 min at 56°C and each new batch of FCS or medium was tested before use. All cells were cultured at general cell culture conditions in a standard incubator (37°C, 5% of CO_2, 95 % of relative humidity). All suspension cells were harvested by gentle rocking of the flask, transferring the suspension into a falcon tube and removal of the supernatant after centrifugation at 300xg for 8 min at 4°C. In all washing steps the cells were resuspended in PBS and centrifuged again.

5.2.1.1.1 Cell number and viability

To asses cell number and viability cell solutions were stained with Trypan blue solution and counted in a Neubauer hemocytometer according to the manufacturer’s recommendation. Since cell membranes of living cells are still intact Trypan blue is only absorbed in dead cells, which turn blue and can be easily detected.

5.2.1.1.2 Mycoplasma assay

All cells primary or cell lines were routinely checked for mycoplasma contamination with the MycoAlert™ Mycoplasma Detection kit (Lonza).
Cell culture supernatants were measured according to manufacturer’s recommendations.

5.2.1.2 Cell lines

All cell lines were obtained from DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen) and cultivated according to Table 5.1.

Table 5.1: Culture conditions

Name	Culture medium	Adherence	DSMZ accession number
THP-1	RPMI 1640	no	ACC 16
Daudi	RPMI 1640	no	ACC 78
HeLa	DMEM	yes	ACC 57

Suspension cell lines were split 1:3 to 1:6 into fresh medium every 2-3 days. Adherent cells were washed with PBS and detached from the surface using 0.05% trypsin/0.2% EDTA/PBS (3 ml per 75 cm^2 per culture vessel area) incubated 5-10 min at 37°C with slight rocking of the plate. The reaction was stopped using 5-15 ml of medium including 10% FCS. After one washing step the cells were seeded in the appropriate medium with all supplements and FCS.

5.2.1.2.1 Design, assembly, and transfection of reporter constructs

To verify regulatory functions of the DMRs reporter constructs were designed for luciferase assays.

The predicted enhancer sequences were extracted from the UCSC Genome Browser and PCR primers covering the enhancer and control regions were designed. The primers included distinct *Pst*I, *Nsi*I, *Spe*I or *Xba*I restriction sites (listed in 5.1.5.1) which were introduced into the sequence during PCR. Enhancer fragments ranged from 291 to 586 bp covering the site of transcription factor binding as well as the CpGs showing methylation dynamics.

The design of the pCpGL-cmv vector used in this study is described elsewhere (Klug, 2010, urn:nbn:de:bvb:355-epub-134293). The CpG-free vector contains a basal EF1α promoter and a CMV enhancer which was removed using *Pst*I and *Spe*I and replaced with the DMRs.

Enhancer and control regions were amplified from genomic DNA, ligated into the vector and cloned into the E. coli strain *PIR 1* matching the plasmids origin of replication. The amplified Plasmids were isolated from the host cells, purified and controlled with restriction digest followed by agarose gel electrophoresis. Glycerol stocks were made from positive clones with verified sequences only.

Human cell lines (THP-1, HeLa) were co-transfected with either *in vitro Sss*I-methylated or unmethylated reporter plasmids including an internal *Renilla* control plasmid (Rehli, 2000). DEAE dextran or Lipofectamine® reagent were used according to manufacturer's recommendations (see following chapters). The cells were cultivated for 48 h with or without 10^{-8} of PMA 12h after transfection until harvesting at 48h post transfection. Individual activity of firefly and *Renilla* luciferase was measured in duplicates and triplicates in three independent experiments and firefly luciferase was normalized to *Renilla* luciferase activity. A control construct containing the basal EF1α promoter only was included in the transfections to further correct enhancer activity for basal luciferase activity originating from the promoter element. Vice versa the original pCpGL-cmv vector containing both EF1α promoter and CMV enhancer was used as a positive control.

5.2.1.2.1.1 Transfection with DEAE dextran

This method was used to transfect THP-1 cells with reporter plasmid constructs. In principle dextran (polycation) forms complexes with the DNA (anionic charge) which are then transported into the cells via endocytosis.
THP-1 cells can be differentiated into a mature monocyte- and macrophage-like phenotype by adding phorbol myristate acetate. Upon stimulation the cells become adherent and stop proliferation (Abrink *et al.*, 1994; Tsuchiya *et al.*, 1982). Luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega) or the Beetle-Juice KIT (PJK) according to manufacturer's recommendations.

Throughout the first part of the protocol the cells were handled at sterile conditions. One day before transfection confluent cells were split and diluted to a concentration of 5×10^5 cells/ml. After incubation for 24 h cells were pooled, divided into 6 ml samples (3×10^6 cells), washed with 6 ml of STBS buffer and centrifuged 12 min at 1200 rpm (14 ml round bottom tubes, polypropylene). Meanwhile the *Renilla* control plasmid was diluted in STBS buffer (20 ng/70 µl), divided into 70 µl samples and mixed with 200 ng of plasmid DNA (either methylated or non-methylated). DEAE stock solution was prepared dissolving DEAE dextran in STBS buffer to a final concentration of 800 µg/ml. Cell pellets were now transfected with a fresh mix of 70 µl DEAE stock solution and 70 µl of plasmid-STBS-mix and vortexed. After 20 min of incubation at 37°C the transfection was terminated adding 10 ml of STBS buffer and centrifuged 12 min at

1200 rpm. The pellets were resuspended in 6 ml of RPMI (10% FCS), transferred to a cell culture dish (Ø 60mm) and incubated 48 h at 37°C.
If necessary, 24 h after incubation the cells were supplemented with a final concentration of 10^{-8} µg/µl PMA and incubated for another 24 h.

48 h after transfection without PMA the cells were harvested using a Pasteur's pipet into 14 ml round bottom tubes (poly-styrene). Remaining liquid in the dishes was recovered by rinsing the dishes with 5 ml of PBS and added to the cell suspension. After centrifugation (1200 rpm, 10 min, 4°C) the wash step was repeated with another 5 ml of PBS, the supernatant was discarded and cells were resuspended in 1 ml of PBS. The suspension was then transferred to a smaller tube (1.5 ml, Eppendorf), centrifuged (1200 rpm, 10 min, 4°C) and mixed thoroughly with 300 µl of 1 x Passive lysis buffer.
If PMA was added to the cells, the harvesting process was slightly different. The adherent cells were washed with 5 ml PBS, the supernatant was discarded and the cells were detached from the surface by adding 300 µl of 1 x Passive lysis buffer and slight scraping with a cell scraper. The solution was recovered and centrifuged.
After precipitation (13000 rpm, 1 min) of remaining debris, luciferase and firefly activity of the cell lysates were measured on a luminometer, respectively.

Required buffers and solutions:

STBS Buffer (10x)	25 ml	(25 mM)	Tris (1 M), pH 7.4
	8 g	(137 mM)	NaCl
	6.372 g	(5 mM)	KCl
	0.160 g	(0.6 mM)	$Na_2HPO4/7H_2O$
	0.102 g	(0.7 mM)	$CaCl_2$
	Add ddH_2O to 1 l, autoclave.		
DEAE Dextran	10 mg/ml STBS		

5.2.1.2.1.2 Transfection with Lipofectamine® Reagent

Lipofectamine® Reagent (Invitrogen) is a lipid transfection reagent which contains polycationic and neutral lipids in a 3:1 ratio. In principle it forms complexes with the negatively charged DNA which are incorporated into the cells via endocytosis. This method was used to transfect the HeLa cell line for luciferase reporter assays. The Lipofectamine® 2000 assay was performed according to manufacturer's instructions.

Since the reagent is sensitive to antibiotics, cells were transfected in Opti-MEM® I medium (Gibco) lacking antibiotics.
One day before transfection HeLa cells were seeded at 0.4 x 10^6 cells in 2 ml of DMEM (including 10% FCS, PAA) per well into 6-well cell culture plates.
Shortly before transfection the cells were washed two times with 2 ml of PBS while 100 µl Opti-MEM® were mixed with 20 ng and 1 µg plasmid DNA per reaction in 14 µl round bottom tubes (poly-styrene, Falcon). A mixture of 100 µl Opti-MEM® and 7 µl of Lipofectamine® reagent was added to the DNA solution, mixed and incubated 30 min at room temperature. While 800 µl of Opti-MEM® were added to the reagent-mix, the last PBS wash buffer was removed from the cell pellets. The cells were now resuspended and incubated with the transfection mixture at 37°C for 5 h. The reaction was stopped with 1 ml of DMEM (incl. 10% FCS) and further incubated 43 h. 24 h after transfection the medium was changed and fresh 2 ml of DMEM were added per well.

48 h after transfection the cells were harvested. First the medium was removed and the cells were washed with 4 ml of PBS. After discarding of the wash buffer, the cells were scraped from the surface (with a cell scraper) and at the same time lysed with 300 µl of 1 x Passive lysis buffer. The solution was transferred to a 1.5 µl tube, centrifuged at 13000 x rpm for 1 min before measuring transgene expression. Luciferase and firefly activity of the cell lysates were measured on a luminometer using the Beetle-Juice KIT (PJK) according to manufacturer's instructions.

5.2.1.3 Primary cells

5.2.1.3.1 Isolation of PB-MNCs

The collection of blood cells was performed in compliance with the Helsinki Declaration and all donors signed an informed consent.
Peripheral blood mononuclear cells (PB-MNC) were isolated from healthy donors by leukapheresis and separated by density gradient centrifugation over Ficoll/Hypaque (Johnson *et al.*, 1977). Monocytes were isolated from the MNC fraction using counter current centrifugal elutriation as described in Sanderson *et al.* (Sanderson *et al.*, 1977). Usually monocyte yields ranged between 10-20% of total MNCs depending on donor variation.

5.2.1.3.2 Monocyte cell culture

Blood monocytes isolated from healthy donor blood were cultivated at 1 x 10^6/m in RPMI supplemented as described in 5.2.1.1. The cells were further stimulated adding 20 U/ml recombinant human IL-4 (Promokine, Heidelberg, Germany) and 280 U/ml GM-CSF (Berlex, Seattle, USA) in order to differentiate the cells into immature dendritic cells (iDC) and with 2% human AB serum to obtain macrophages. The cells were cultured for 7 days at according conditions and harvested after 6 h, 18 h, 27 h, 42 h, 66 h and after 7 days of stimulation or immediately after elutriation. At 66 h after stimulation the cells were supplemented with fresh differentiation promoting factors.
Macrophages were cultivated in Teflon foil and harvested by gentle rocking of the bag after cooling it down to 4°C (described in Andreesen *et al.*, 1983).

5.2.1.3.3 Small interfering RNA (siRNA) transfection

Principle

Depletion of the protein of interest can be facilitated using a variety of different methods. RNA interference is a very effective yet easy way to degrade mRNAs before their translation. Small interfering RNAs (siRNAs) are a pool of small single stranded RNAs targeting different epitopes on the target mRNA. Once injected into the cell, they form stable double stranded RNA molecules with the target mRNAs and thus block further processing of the mRNAs.

In order to transmit siRNA in to primary cells the Amaxa device and the Human Monocyte Nucleofector® Kit (Lonza) were used according to manufacturer's protocol with modifications.

Protocol

Freshly isolated blood monocytes were washed once and pelleted at 6 x 106 cells per transfection. After removing the supernatant completely 600 nM of siRNA (Thermo Scientific Dharmacon, Lafayette, USA) were added to the cells, resuspended carefully in 100 µl of Human Monocyte Nucleofector® solution (82 µl of Nucleofector® solution, 18 µl supplement) and transferred to a cuvette. The cuvette was immediately placed into the Nucleofector® device. After transfection using the Y-01 program the cells were resuspended in 500 µl pre-warmed RPMI medium (supplemented with 10% FCS, antibiotic-free). The suspension was further transferred into a cell flask containing more medium (10% FCS, 20 U/ml IL-4, 280 U/ml GM-CSF). The cells were cultured at a

density of 1 x10^6 cells/ml for 27 h or 42 h. RPMI medium without antibiotics was used in the whole procedure to avoid increased cell death. The cells were harvested as described in 5.2.1.1 and subjected to DNA, RNA and protein isolation as well as flow cytometry measurements.

5.2.1.4 E. coli culture and methods

5.2.1.4.1 Culture

Glycerol stocks of the E. coli strain PIR1 (F- Δlac169 rpoS(Am) robA1 creC510 (Invitrogen, Karlsruhe, Germany) were streaked on a LB agar-plate and grown overnight at 37°C to obtain isolated colonies. Single colonies were picked and one colony per 30 ml of LB medium containing the appropriate antibiotics was cultured in a shaking device (200 rpm) over night at 37°C.

Antibiotic	Stock concentration	Dilution	Final concentration	Company
Ampicillin	100 mg/ml	1:1000	100 µg/ml	Ratiopharm, Ulm, Germany
Zeocin	100 mg/ml	1:4000	25 µg/ml	Invitrogen, Karlsruhe, Germany

Required medium:

LB medium	10 g	BactoTrypton (Difco)
	5 g	BactoYeastExtract (Difco)
	10 g	NaCl
	Add ddH_2O to 1 l, autoclave and adjust pH to 7.5	

5.2.1.4.2 Freezing

In order to preserve cells, glycerol stocks were prepared. The cells were cultivated to confluency and 500 µl glycerol (80% in LB medium) were added to 500 µl of cell suspension. The vials were mixed thoroughly and stored at -80°C.

5.2.1.5 Molecular cloning

5.2.1.5.1 Construction of reporter plasmids

Insert sequences for cloning were generated from genomic DNA using PCR and precipitated according to chapter 5.2.1.5.2. The used primers introduced two matching restriction sites (*PstI* or *NsiI* and *SpeI* or *Xba*I) in order to generate sticky ends after restriction digest (see section 5.2.2.4). The first two and the last two restriction

enzymes produced identical overhangs non-complementary to the other, which prevented self-ligation of the digested vector.

The CpG-free vector pCpGL-cmv was isolated from cryo cultures and digested overnight using *Pst*I and *Spe*I restriction enzymes to cut out the CMV enhancer and to match the inserts' sticky ends. After size validation of inserts and vector on an agarose gel, both were purified from the gel (see chapter 5.2.2.3) and mixed in a ligation reaction with a 3:1 molar excess of insert compared to vector. Each insert was ligated into the vector over night at 16°C as well as one vector control without insert DNA (see section 5.2.1.5.4.). As described in chapter 5.2.1.5.5, 5 µl of the ligation mix were transformed into competent E. coli *PIR1* cells (50 µl, Invitrogen). To validate accurate insertion, selected clones were picked, cultured and plasmids were isolated according to section 5.2.1.4.1 and 5.2.1.5.6. After control digest on an agarose gel the constructs were sequenced to check sequence integrity (Geneart, Regensburg, Germany) and preserved for long term storage (see chapter 5.2.1.4.2).

5.2.1.5.2 Polyethylene glycol (PEG) precipitation

PCR products for cloning were cleaned and precipitated with polyethylene glycol. DNA and PEG were mixed at equal amounts and incubated for 15 min at room temperature. After centrifugation for 15 min at 15000 rpm (RT) the supernatant was carefully discarded and 100 µl ethanol added without disturbing the pellet. Another centrifugation step followed (15 min, 15000 rpm, RT). The pellet was left to dry at RT for 5 min and dissolved in 20-25 µl H_2O.

Required solution:

PEG-mix	26.2 g	(26.2%)	PEG 8000
	20 ml	(0.67 M)	NaOAc (3M), pH 5.2
	660 µl	(0.67 mM)	$MgCl_2$ (1M)
	Add ddH_2O to 100 ml		

5.2.1.5.3 Dephosphorylation of DNA

To prevent self-ligation of the vector after digestion the plasmids were treated with 1 U alkaline phosphatase (calf intestine, Roche) per 1 pmol of vector for 1 h at 37°C. The enzyme excises phosphate groups and thus allows vector ligation with the phosphate containing insert only.

5.2.1.5.4 Ligation of vector and insert

After preparation of corresponding sticky ends on insert sequences and vector the two DNA fragments were ligated. Therefore a mixture of 1 U of T4 DNA ligase (NEB, Frankfurt, Germany) in a final concentration of 1 x T4 DNA ligase reaction buffer and in a final volume of 10-20 µl, including insert and vector DNA, was prepared. The insert DNA was added in a 3:1 molar excess to 100-200 ng of plasmid DNA and the mixture was incubated over night at 16°C. To monitor insert ligation compared to vector self-ligation a vector control containing plasmid DNA only was included in the procedure.

5.2.1.5.5 Transformation of competent bacteria

To amplify plasmid constructs chemically competent E.coli PIR1 cells were transformed according to the following protocol.
50 µl of competent bacteria were thawed on ice and carefully mixed with 5 µl of ligated vector or vector control and incubated 30 min on ice. The cells were heat-shocked for 30 seconds at 42°C in a water bath and immediately stored on ice for 2-5 min. To select transformed cells only, the cells were incubated for 90 min at 37°C (shaking) in 250 µl SOC medium and plated on an LB agar-plate supplemented with the antibiotic zeocin. Only plasmid-containing clones acquired the resistance and were able to grow. Colonies were grown over night at 37°C and DNA was isolated (see chapter 5.2.1.5.6) and stored at - 20°C until further usage.

Required medium:

SOC medium	20 g	(2%)	BactoTrypton (Difco)
	5 g	(0.5%)	BactoYeastExtract (Difco)
	0.6 g	(10 mM)	NaCl
	0.2 g	(3 mM)	KCl
	Add ddH_2O to 1l, autoclave and add:		
	10 ml	(10 mM)	$MgCl_2$ (1 M, sterile)
	10 ml	(10 mM)	$MgSO_4$ (1 M, sterile)
	10 ml	(20 mM)	Glucose (2 M, sterile)

5.2.1.5.6 Isolation of Plasmids

Plasmid isolation from 2 ml liquid bacterial culture was performed using the NucleoSpin® Plasmid Quick Pure kit from Macherey & Nagel. To isolate larger amounts of plasmid DNA for cloning or transfection experiments, 30 ml of cultured

bacteria were purified using the EndoFree Plasmid Midi kit and Plasmid Plus kit (Qiagen).

5.2.2 Working with DNA

All standard molecular procedures were performed according to “Current protocols in molecular biology” (Ausubel, 1987) if not otherwise stated. The same applies for all purchased kits which have been used according to the manufacturer’s manual.

5.2.2.1 Isolation of genomic DNA

Cultured cells were trypsinated if necessary and/or washed twice. After removal of the supernatant the pellets were stored at -20°C or subjected to DNA isolation using either the Qiagen Blood and Tissue kit or the Qiagen Blood and Cell culture kit.

5.2.2.2 DNA quantity and quality

5.2.2.2.1 Agarose gel electrophoresis

To routinely check the quality of purified genomic DNA agarose gel electrophoresis was the method of choice. Electrophoresis separates nucleic acids in porous agarose gels according to their size and shape.
Depending on the size of the DNA 1% to 2% of agarose dissolved in 1 x TAE was melted in a microwave and mixed with 2,5 µl ethidium bromide (10 mg/ml) /50 ml gel after cooling down to ~50°C. The gel was poured into the gel chamber to polymerize. Meanwhile the DNA samples were mixed 4:1 with DNA loading dye (1:5). The solid gel was covered with 1 x TAE running buffer, loaded with the DNA samples and run for 30 min to 1.5 h at 80 to 120 V.

Required buffers:

TAE (50x)	252.3 g	(2 M)	Tris/HCL, pH 7.8
	20.5 g	(250 mM)	NaOAc/HOAc, pH 7.8
	18.5 g	(50 mM)	EDTA
	Add ddH_2O to 1l		

DNA loading dye (5x)	500 µl	(50 mM)	Tris/HCL, pH 7.8
	500 µl	(1%)	SDS (20%)
	1 ml	(50 mM)	EDTA (0.5 M), pH 8.0
	4 ml	(40%)	Glycerol
	10 mg	(1%)	Bromphenol blue
	Add ddH_2O to 10 ml		

5.2.2.2.2 NanoDrop spectrophotometer

DNA concentrations were measured using the NanoDrop spectrophotometer device according to the manufacturer's instructions.

5.2.2.2.3 Qubit®

The Qubit® system was used to detect very low DNA concentrations with the Qubit® dsDNA HS Assay Kit or single stranded DNA using the Qubit® ssDNA Assay kit. Samples were prepared and measured according to the manufacturer's manual.

5.2.2.2.4 2100 Bioanalyzer

For fragmented DNA the 2100 Bioanalyzer (Agilent Technologies) was an alternative method to check quality, quantity and fragment length at the same time. DNA samples were prepared and measured with the Agilent DNA 1000 or Agilent High Sensitivity DNA kit. The device was also used to detect RNA quality and quantity using the Agilent RNA 6000 Nano kit.

5.2.2.3 Purifying DNA with Gel extraction

To obtain homogenous, size-selected PCR fragments and plasmid DNA the whole sample was separated on an agarose gel. The band of interest was excised from the gel with a scalpel and DNA was purified and recovered with the QIAquick Gel Extraction kit (Qiagen) according to the manufacturer's instructions.

5.2.2.4 Enzymatic restriction digest

To check or prepare DNA in several methods enzymatic restriction digests were performed. All enzymes were purchased from New England Biolabs or Roche. If not

otherwise stated the enzyme was mixed in excess (see table Table 5.2) in a final concentration of 1 x corresponding buffer containing BSA, if required, and DNA. The solution volume was adjusted to the volume of enzyme in a 9:1 ratio and incubated at 37°C for 1-2 hours or overnight.

Table 5.2: Parameters for enzymatic restriction digest of DNA

Method	Amount DNA	Amount enzyme	Volume	Incubation time
Inserts for cloning	1-2.5 µg	15 U each	35-45 µl	1 h
Vector for cloning	10 µg	20 U each	40 µl	O/N
Check ligation	8 µl	10 U each	20 µl	1 h 20 min
GGSD	400-500 ng	30 U/60 U	50 µl	O/N
hMeDIP after TF	4 µg	30 U	100 µl	O/N
MSRE	100 ng	20 U each	12 µl	O/N

5.2.2.5 Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) is a simple method to perform in vitro synthesis and amplification of DNA. In brief 100 ng DNA was amplified using the Phusion Hot Start High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, USA) and the appropriate primers in the following mix. The annealing time was adjusted to the lowest primer T_M (annealing temperature = $T_{M,lowest}$+3°C).

10 µl 5 x Phusion HD Buffer
1 µl dNTPs (10 mM)
1 µl Primer S (10 pmol/µl)
1 µl Primer AS (10 pmol/µl)
0.5 µl Phusion Hot Start Taq Polymerase (2 U/µl)
X µl Template DNA
➔ Add to 50 µl with nuclease-free H_2O

Table 5.3: Reaction parameter for general PCR

PCR step		Cycling parameter
Initial melting		98°C, 30 s
35 cycles	Melting	98°C, 10 s
	Annealing	69°C, 20 s
	Extension	72°C, 20 s
Final extension		72°C, 7 min
Cooling		12°C, forever

5.2.2.6 *In-vitro* methylation and removal of methylation

DNA was in-vitro methylated using 4 U *Sss*I CpG methyltransferase (NEB) per 1 µg DNA supplemented with 320 µM of S-adenosyl methionine (SAM). The reaction was incubated 5 h at 37°C, replenished with enzyme and SAM after 2 hours and terminated by heating it to 65°C for 10 min. To clean-up the DNA the QIAamp DNA Kit (Qiagen) was used according to the manual.

In order to erase all methyl marks present on genomic DNA, whole genome amplification was performed using the Repli-g Mini or Midi (Qiagen).

To obtain DNA with a distinct level of methylation; for example 50%, both 100% and 0% methylated DNAs were mixed in a ratio of 1:1. This DNA was used as control DNA for methylation analysis on the MassARRAY system.

5.2.2.7 Detection of DNA modifications

5.2.2.7.1 Detection of 5-methylcytosine (5mC)

5.2.2.7.1.1 Quantitative DNA methylation analysis using the MassARRAY system

The MassARRAY system is a bisulfite-based method to detect and quantify DNA methylation levels at single-base resolution. Most methods to measure DNA methylation require an amplification step to ensure sufficient resolution. Since the methyl mark is lost within two PCR steps, the methylation patterns have to be transferred into gene code information. This is accomplished by sodium bisulfite treatment. This chemically deaminates all regular cytosines and converts them into uracil but it doesn't affect methylated cytosines which stay unchanged. This treatment results in methylation dependent sequence variation from C to T in the PCR products. The change in the underlying sequence leads to a G/A variation on the reverse strand and a mass difference of 16 Da per CpG-residue in the cleavage products which can be detected by the MassARRAY system (MALDI-TOF mass spectrometry). A complete overview of the method can be found in Ehrich *et al*., 2005 or in the EpiTYPER User Guide (Ehrich *et al.*, 2005, www.sequenom.com).

Protocol

The experiments were performed according to the EpiTYPER application guide and protocols (Sequenom). DNA bisulfite treatment was carried out using either the EZ

DNA methylation kit (Zymo) or the EpiTect Plus kit (Qiagen). Cycling times were adjusted to following conditions if not otherwise stated:

1) 95°C for 30 min
2) 50°C for 15 min
3) Go back to step 1, 19 times
4) 4°C for ever

PCR, SAP treatment, *in vitro* transcription, RNA cleavage and MALDI-TOF mass spectrometry was done as recommended by Sequenom. The EpiDesigner tool (Sequenom, Hamburg, Germany) was used to design PCR primers including a T7 polymerase promoter tag added to the 5' end of the reverse primer and a 10-mer overhang added to the forward primer to balance melting temperature differences.
Raw data was processed using the EpiTYPER software (v1.2, Sequenom).

5.2.2.7.2 Detection of hydroxymethylcytosine (5hmC)

5.2.2.7.2.1 Glycosylation and glycosyl-sensitive digest of DNA (GGSD)

This method was adapted from the Quest 5-hmC Detection Kit™ (Zymo). The kit permits sequence specific detection of 5hmC using a 5hmC glycosyltransferase enzyme which specifically tags and therefore protects residual hydroxymethylcytosine by glycosylation. In a subsequent digestion step glycosylated 5hmCs are protected from a glycosyl-sensitive restriction enzyme (*Msp*I), whereas all other non-protected cytosines are cleaved. We optionally utilized a second enzyme (*Hpa*II), which is sensitive to 5mC, 5hmC, and glycosyl-5hmC to obtain absolute levels of 5mC and 5hmC. The locus specific cleavage or protection of DNA can be quantified by qPCR.

Protocol

Instructions of the manufacturer's manual were followed with slight alterations. 400-500 ng gDNA was glycosylated for 3 h at 37°C in the following reaction setup:

400-500 ng DNA
5 µl 10 x Glycosyltransferase Reaction Buffer
5 µl 10 x Uridine Diphosphoglucose (UDPG, 1 mM)
2 µl Glycosyltransferase (2 U/µl)
➔ Fill up with ddH_2O to a final volume of 50 µl

In the next step the DNA was purified using the Zymo Clean and Concentrator^{TM}-5 kit and eluted in 22 µl EB (70°C) after incubation of the elution buffer on the membrane of the column for 2 min. The DNA was now digested with 30 U of *Msp*I or 30 u *Hpa*II (NEB) overnight at 37°C in the corresponding buffer. After incubation another clean up step followed as described above with an elution volume of 20 µl and was eluted from the column twice. The samples were now randomly measured in the NanoDrop device and diluted in elution buffer to a final concentration of ~1-2 ng/µl. In the following qPCR step the digestion and protection of the DNA was measured, respectively.

5.2.2.7.2.2 MeDIP and hMeDIP (Hydroxy-/methylated DNA Immunoprecipitation)

The MeDIP (hMeDIP) method is an antibody based immunoprecipitation to enrich for (hydroxyl-) methylated DNA fragments. It was originally adapted from a protocol of Mohn *et al*. (Mohn *et al.*, 2009) which was designed to pull-down methylated DNA. The enriched DNA fragments can be further quantified using qPCR.

In principle 4 µg of gDNA was digested with *Mse*I in a final volume of 100 µl for 1,5 h at 37°C, then purified using the Qiaquick PCR Purification kit (Qiagen) and eluted twice in 50 µl pre-heated (55°C) EB after incubation for two minutes.
In the next step 1-2 µg DNA was diluted in 300 µl TE buffer (pH 8,0) and denatured for 10 min at 95°C followed by immediate cooling on ice for another 10 min. 10% of input DNA was set aside at 4°C. 30 µl of 10 x IP buffer were now added to the denatured DNA together with 1-2 µg of 5-methylcytidine or 5-hydroxy-methylcytidine antibody (Diagenode and ActiveMotif, respectively) and incubated 2. In the meantime 30 µl ProteinA Dynabeads (Invitrogen) per sample were washed with 800 µl of 0,1% BSA/PBS for 5 min at room temperature to block unspecific binding of the antibody. The beads were separated from the wash buffer on a magnetic rack, the supernatant was discarded and the wash step was repeated once. The blocked beads were resuspended in 30 µl of 1 x IP buffer, added to the antibody-DNA mixture and incubated 2 h at 4°C with overhead shaking.
After 2 to maximum 3 h of rotation the beads were separated on a magnetic rack, the supernatant was discarded and the beads were washed three times with 700 µl 1 x IP-buffer in a shaking Thermomixer (800 rpm, Eppendorf) for 10 min at RT. After discarding the supernatant the beads were resuspended in 250 µl Proteinase K digestion buffer, which was also added to the input sample, together with 3.5 µl of Proteinase K (20 mg/ml, Roche) and incubated 3 h at 50°C in a shaking Thermomixer

(Eppendorf). After incubation the supernatant was separated from the beads on a magnetic rack and transferred to a new Eppendorf tube.
Finally the enriched DNA and input were purified using the Qiaquick PCR Purification Kit (Qiagen), eluted twice in 30 µl EB (55°C, 2 min incubation) and further quantified with qPCR.

Required buffers and solutions:

Na-Phosphate buffer (1 M), pH 7.0	39 ml 2 M NaH_2PO_4 (276 g/L) 61 ml 2 M Na_2HPO_4 (284 g/L) Add ddH_2O to 100 ml
IP buffer (10x)	100 mM Na-Phosphate buffer, pH 7.0 1.4 M NaCl 0.5% Triton X-100
TE buffer (1x)	10 mM Tris-HCl pH 8.0 1 mM EDTA pH 8.0
IP buffer (1x)	1 volume 10×IP buffer 9 volumes TE buffer
BSA/PBS solution	9 ml PBS 1 ml BSA (10 mg/ml)
Proteinase K digestion buffer	50 mM Tris/HCl, pH 8.0 10 mM EDTA 0.5% SDS

5.2.2.7.2.3 hMeDIP-seq

In order to measure 5hmC levels genome-wide it was part of this thesis to adapt the hMeDIP protocol for hMeDIP-sequencing (hMeDIP-seq). The protocol was altered with slight modifications and performed as already described in chapter 5.2.2.7.2.2.
First, the DNA was sonicated using the Covaris device instead of digestion to obtain a homologous fragmentation range of approximately 200 bp. Therefore gDNA was diluted in TE buffer and fragmented according to the amount using the following settings.

Covaris settings	4-12 µg DNA	2-3 µg DNA
Duty cycle	10%	10%
Intensity	5	5
Cycles per burst	200	200
Duration	180 s	120 s
Water level	12	12
Tube	DNA micro	DNA micro
Volume	130 µl	50 µl

Fragment length was confirmed on an agarose gel and 2 µg of sonicated DNA further subjected to immunoprecipitation. The purification step was also modified to improve DNA yield. The DNA was incubated in binding buffer for 15 min. After centrifugation the flowthrough was loaded on the column again and centrifuged one more time. The wash buffer was incubated for 2 min on the column and centrifuged one minute from every side. Residual wash buffer was aspirated from the inner ring and the membrane was dried one minute at RT. Then 37 µl of pre-heated (55°C) EB were pipetted on the center of the membrane, incubated for 2 min and eluted twice from the column. Of the remaining 33 µl of eluate 3 µl were quantified using qPCR to check enrichment performance and 30 µl were further processed.

During the procedure the DNA was denaturated by heat resulting in single stranded DNA fragments. Before subjecting the DNA to library preparation, the second strand had to be generated in order to perform effective adapter ligation further downstream of the procedure. This was done using the NEBNext® Second Strand Synthesis Module following the manufacturer's instructions with slight alterations. Random primer were taken from the Retroscript® kit (50 µM, Life Technologies) and 450 ng added to the following mixture in a final volume of 80 µl.

Mastermix:

30 µl 5hmC-enriched DNA
2.9 µl Random Decamers (50 µM, Ambion)
8 µl 10 x Second Strand Synthesis Reaction Buffer
4 µl Second Strand Synthesis Enzyme Mix
35.1 µl H_2O (nuclease-free)

The solution was mixed thoroughly and incubated 2.5 h at 16°C in a thermal cycler. After incubation the DNA was purified as described above and eluted in 44 µl of EB which accounts for 4 µl loss of DNA and the remaining 40 µl were further processed using the library preparation protocol described in chapter 5.2.2.9.

5.2.2.7.2.4 Hydroxymethyl Collector™-seq

This method was used to measure 5-hydroxymethylcytosine on a genome-wide level and is based on the Hydroxymethyl Collector™ kit (ActiveMotif). The kit chemically modifies 5hmC residues and thus enriches for hydroxymethylated DNA fragments. Hydroxymethylated cytosines are specifically glycosylated and further tagged with biotin at the glycosyl group. Streptavidin-conjugated magnetic beads are used to pull-down captured DNA which can be further quantified by qPCR. Samples enriched for 5hmC are then subjected to library preparation as described in section 5.2.2.9 and sequenced according to chapter 5.2.2.10.

Protocol

First the DNAs were sonicated to an approx. fragment length of 200 bp using the Covaris S2 with the following settings:

Covaris settings	
Duty cycle	10%
Intensity	5
Cycles per burst	200
Duration	180 s
Water level	12
Tube	DNA micro
DNA	12 µg
Volume	130 µl

After checking fragment size on an agarose gel (2%) and quantity using the NanoDrop, 2 µg DNA of each source were further processed using the Hydroxymethyl Collector™ kit. 5hmC enrichment was performed according to manufacturer's manual (version B2) including the negative control but with slight alterations to the protocol. The biotin solution was incubated for 1 h or 1.5 h. To improve the yield elution buffer was heated to 50°C and incubated 2 min on the column before the elution centrifugation. End-to-end rotation was done on an Intelli-Mixer (program F5). In the last step DNA was eluted in 45 µl elution buffer, quantified by qPCR and subjected to library preparation.

5.2.2.7.3 Detection of formylcytosine (5fC), carboxylcytosine (5caC)

5.2.2.7.3.1 DNA modification-sensitive restriction efficiency (MSRE)

This protocol was used to detect 5fC and 5caC levels at single CpGs by quantification of the cutting-efficiency of a modification-sensitive enzyme. 100 ng of iDC DNA was digested overnight with 20 U *Msp*I and 20 U *Hha*I or *Hha*I alone (control) at 37°C. *Msp*I is sensitive to 5fC and 5caC (Ito *et al.*, 2011) whereas *Hha*I digests both modifications. After digestion, the samples were diluted 1:10 with ddH_2O and digested fragments were quantified using qPCR. The digested and control-digested sample amplification values were normalized against the GAPDH region, which lacks a *Msp*I restriction site. The cutting efficiency was calculated by subtracting the normalized values for *Msp*I-digested and *Hha*I-digested (control) samples from 100%.

5.2.2.7.3.2 fCAB-epi

This chemical modification assisted BS-seq protocol is adapted from the original fCAB-seq method developed by Song and colleagues (Song *et al.*, 2013). In principle, 5fC is modified by *O*-ethylhydroxylamine ($EtONH_2$) in a chemical reaction to protect it from bisulfite-conversion. The subtraction of standard BS-converted sample values from fCAB-seq samples allows detection of 5fC at single-base resolution.

Design of control oligos

To test the adapted protocol (fCAB-epi) four DNA oligos (61 bp) of non-human origin, containing only two CpG sites, were designed. The sequence was adapted from the pGEM®-3Z vector (Promega). Each oligo contains one of the following modifications at both CpGs: 5mC, 5hmC, 5fC, 5caC, or unmodified cytosines. The oligos were ordered and synthesized at Eurofins MWG (Ebersberg, Germany).

Table 5.4: Control oligos for fCAB-epi

Name	Sequence (5'->3')
ctrl_5fC	GGGAGGGCTTACCATCTGGCCCCAGGZGGATGCZGGGTAAGATGCTTTTCTGTGACTGGTG
ctrl_5hmC	GGGAGGGCTTACCATCTGGCCCCAGGZGGATGCZGGGTAAGATGCTTTTCTGTGACTGGTG
ctrl_5mC	GGGAGGGCTTACCATCTGGCCCCAGGZGGATGCZGGGTAAGATGCTTTTCTGTGACTGGTG
ctrl_C	GGGAGGGCTTACCATCTGGCCCCAGGCGGATGCCGGGTAAGATGCTTTTCTGTGACTGGTG

Protocol

Instructions of the published protocol were followed with slight alterations (Song *et al.*, 2013). 1 µg of DNA was treated for 2 h at 37°C in the following reaction setup to protect 5fC or in a control setup without $EtONH_2$:

1 µg DNA
2 µl MES buffer (1 M)
2 µl *O*-ethylhydroxylamine ($EtONH_2$, 0.1 M)
➔ Fill up with ddH_2O to a final volume of 20 µl

In the next step the DNA was purified using the QIAquick Nucleotide Removal Kit (Qiagen) and eluted in 30 µl EB (37°C). DNA concentrations were measured using the Nanodrop device and 200 ng of sample and control-sample DNA were subjected to bisulfite-conversion using the EpiTect Bisulfite Kit according to the manufacturer's manual (Qiagen). The samples were further processed in the MassARRAY EPITYPER protocol and measured using mass spectrometry as was described in chapter 5.2.2.7.1.1.

5.2.2.8 ChIP (Chromatin Immunoprecipitation)

The ChIP (Chromatin immunoprecipitation) is a method to study protein interactions with DNA. The majority of regulatory proteins interact directly with their DNA target regions. With this method they are crosslinked to their specific genomic location, using formaldehyde. This chemical substance directly targets amino groups on amino acids and nucleotides forming a Schiff's base. The resulting DNA fragments can be detected and quantified after immuno-based pull-down. Depending on the aim of the experiment ChIP samples can be quantified using qPCR for site specific analysis of the regions of interest or by whole genome sequencing to study the protein of interest on a genome-wide level.

Principle

In the first step formaldehyde was used to cross-link DNA-protein-interactions according to a protocol published earlier (Metivier *et al.*, 2003). The cells were incubated in media with all supplements and 1% formaldehyde for 10 min at room temperature. By adding 0,125 M glycine the reaction was stopped and the cells were further washed in 1 x PBS including 1 mM PMSF and 10 mM NaBut twice. The pellet was stored immediately at -80°C or further processed.

In the next step the pellets were resuspended in L1A lysis buffer (250 µl/10 x 10^6 cells) and lysed in L1B lysis buffer (250 µl/10 x 10^6 cells) for 10 min on ice. After centrifugation the pellets were resuspended in Nuclear Lysis Buffer L2 (400 µl/10 x 10^6 cells) and sonicated using the Covaris device with the following settings.

Covaris settings	
Duty cycle	20%
Intensity	8
Cycles per burst	200
Duration	1800 s
Water level	15
Tube	Chromatin
Cells	20 x 10^6
Volume	800 µl

After centrifugation to clear the lysate, fragment size was checked on an agarose gel and 5% of IP volume was separately stored on 4°C and used as input DNA. 2 x 10^6 cells were pre-cleared with 50 µl of Sepharose CL-4B beads (blocked in DB supplemented with 0,5% BSA and 20 µg/ml glycogen, 2 h, RT) per sample in DB (including protease inhibitors) for 2 h at 4°C and after centrifugation incubated with 2,5 µg appropriate antibody over night at 4°C. Chromatin samples were precipitated using 50 µl of blocked ProteinA Sepharose beads (in DB including protease inhibitors, supplemented with 0,5% BSA and 20 µg/ml glycogen, overnight, RT) per IP for 3 h at 4°C. In the next step the bead-complexes were washed according to the following conditions. After incubation of wash buffer (5 min, rotating, room temperature) the samples were centrifuged for 5 min at 500 g and 4°C and the wash buffer was discarded. This step was repeated with all wash buffers listed in the following table.

Buffer	**Wash steps**	**Volume**
WB I	2 x	400 µl
WB II	2 x	400 µl
WB III	2 x	400 µl
TE, pH 8.0	3 x	400 µl

The antibody-DNA-complexes were now eluted from the beads by incubation with freshly prepared 110 µl EB twice (20 and 10 min, RT, shaking at 850 rpm in a Thermomixer). After centrifugation (500 g, 3 min, 4°C) 2 x 100 µl supernatant were recovered and together with the input DNA (filled up to 200 µl with EB) incubated with 0,5 µg/µl of Proteinase K over night at 65°C. After reverse cross-linking the samples

were digested with 0.33 µg/µl RNAse for 2 h at 37°C and then purified using the Qiaquick PCR purification kit with extended incubation of PB (30 min) and elution in 52 µl pre-warmed EB (55°C) after 2 min incubation at RT. 10 µl of the eluate were diluted 1:3 in H2O and quantified using qPCR and the remaining 40 µl were subjected to library preparation.

Required buffers and solutions:

Glycine	9 .85g	(2.625 M)	Glycine
	Add ddH_2O to 50 ml		
Cell Buffer Mix	1 ml	(10 mM)	HEPES / KOH (1 M), pH 7.9
	4.25 ml	(85 mM)	KCL (2 M)
	200 µl	(1 mM)	EDTA (0.5 M), pH 8.0
	Add ddH_2O to 97 ml *		
Nuclear Lysis Buffer (L2)	5 ml	(50 mM)	Tris/HCl (1 M), pH 7.4
	5 ml	(1%)	SDS (20%)
	1,43 ml	(0.5%)	Empigen BB (35%)
	2 ml	(10 mM)	EDTA (0.5 M), pH 8.0
	Add ddH_2O to 97 ml *		
Dilution Buffer (DB)	2 ml	(20 mM)	Tris/HCl (1 M), pH 7.4
	2 ml	(100 mM)	NaCl (5 M)
	400 µl	(2 mM)	EDTA (0.5 M), pH 8.0
	5 ml	(0.5%)	Triton X-100 (10%)
	Add ddH_2O to 97 ml *		
Wash Buffer I (WB I)	2 ml	(20 mM)	Tris/HCl (1 M), pH 7.4
	3 ml	(150 mM)	NaCl (5 M)
	500 µl	(0.1%)	SDS (20%)
	10 ml	(1%)	Triton X-100 (10%)
	400 µl	(2 mM)	EDTA (0.5 M), pH 8.0
	Add ddH_2O to 100 ml		
Wash Buffer II (WB II)	2 ml	(20 mM)	Tris/HCl (1 M), pH 7.4
	10 ml	(500 mM)	NaCl (5 M)
	10 ml	(1%)	Triton X-100 (10%)
	400 µl	(2 mM)	EDTA (0.5 M, pH 8.0)
	Add ddH_2O to 100 ml		

Wash Buffer III (WB III)	1 ml	(10 mM)	Tris/HCl (1 M), pH 7.4
	10 ml	(250 mM)	LiCl (2.5 M)
	10 ml	(1%)	NP-40 (10%)
	10 ml	(1%)	Deoxycholate (10%)
	200 µl	(1 mM)	EDTA (0.5 M), pH 8.0
	Add ddH_2O to 100 ml		
Elution Buffer (EB)	500 µl	(0.1 M)	$NaHCO_3$ (1M)
	250 µl	(1%)	SDS (20%)
	Add ddH_2O to 5 ml		
	*add immediately before use (per ml of buffer):		
	10 µl	(1 mM)	PMSF (100 mM)
	20 µl		50x Roche Inhibitor Cocktail
	10 µl	(10 mM)	NaBut (100 mM)

5.2.2.9 Library preparation

Principle

In order to perform next generation sequencing, DNA libraries of each DNA source have to be generated. Therefore the DNAs are fragmented to ~200 bp to ensure a good coverage and mapping, respectively. To limit DNA input and therefore sample preparation costs the DNA is enriched for the mark of interest by antibody pull-down or other affinity purification methods before library preparation. The first step of the library protocol performs a repair of 3' and 5' ends to form blunt ended, phosphorylated DNA molecules which serve as anchors for a poly-A tails. Next one adenosine is added to the forward and reverse strand. This enables adapter ligation at both ends in order to ensure specific bar coding of all DNA fragments of one source. In the next step DNAs are purified from adapter dimers and non-ligated fragments by size selection; this can be done using agarose gel electrophoresis followed by gel extraction or using automated nucleic acid fractionation devices. After PCR amplification and quality check the DNAs are ready for sequencing.

Protocol

Library preparation was performed according to the following two library protocols. During this work the size selection of the original protocol was adapted to the Caliper LabChIP XT to enhance the yield of DNA. Both protocols are listed on the following pages. For the whole procedure DNA low bind consumables were used. All enzymes were purchased from Enzymatics (via NEB, Frankfurt, Germany) unless otherwise

stated. All components were thawed on ice and all samples were thoroughly mixed on a vortexer (speed 6) and spun down before incubation. Incubation steps were carried out in a thermal cycler with heated lid if block was heated > 20°C.

First part

In the first step 40 µl of DNA was polished to generate blunt ended fragments. For each donor one fragmented, non-enriched input sample from the identical DNA source was diluted (10 ng into 40 µl) and included in the library preparation procedure. All samples were mixed with 10 µl of the following mastermix and incubated 30 min at 20°C.

Table 5.5: Mastermix for DNA end-repair

Component	*Volume*	*Final concentration*
H2O	3,34 µl	n.a.
10x T4 ligase buffer	5,0 µl	1x
dNTP mix (10 mM)	1,0 µl	2 mM
T4 DNA polymerase	0,30 µl	n.a.
Klenow fragment	0,06 µl	n.a.
T4 Polynucleotide kinase	0,30 µl	n.a.
Total volume	10 µl	

Next the DNAs were purified using the MinElute PCR Purification kit from Qiagen according to manufacturer's instructions with slight alterations to improve DNA yield. All centrifugation steps were carried out at 13000 x rpm except for the binding on column and the elution step (11000 x rpm). Samples were washed with PE twice and the membrane was dried 2 min at full speed. Remaining wash buffer on the purple rim inside the column was removed and air dried one minute. The DNA was eluted twice in 10.2 µl and 10.3 µl of pre-warmed elution buffer (50°C) and incubated 5 min each at RT before elution.

In the next step A-tailing was performed. Sample volume was measured, 2.6 µl of mastermix added to 18 µl of sample, mixed and incubated 30 min at 37°C.

Table 5.6: 3'-dA addition by Exo-Klenow

Component	*Volume*	*Final concentration*
10 x Blue buffer	2,0 µl	1x
dATP (10mM)	0,40 µl	195 µM
Klenow (3'-5' exo-)	0,20 µl	n.a.
Total volume	2,6 µl	

Next the Sample volume was calculated and mixed with (sample volume + 2 µl) of 2 x Rapid ligation buffer. NEXTflex DNA Barcodes (HiSS Diagnostics, Freiburg, Germany) were diluted 1:2 in 1 x T4 DNA ligase buffer, mixed well and 1 µl of adapter mix (1:2) was added to the corresponding DNA sample. Finally 1 µl of T4 Rapid DNA ligase was added to the mixture and ligated 10 min at 30°C.
Before the next step, the sample volume was measured and adjusted to 50 µl.

In this clean up step the samples were pre-size-selected according to a ratio of sample volume to bead volume of 1:1.1 (Agencourt AMPure XP, Beckman Coulter).
The tubes were vortexed two times, flicked and vortexed again before incubation for 5 min at RT followed by precipitation of the beads to the vessel wall on a magnetic rack for 2 min at RT. The supernatant was discarded and the beads were carefully washed two times with 70% ethanol for 30 seconds. After removal of the ethanol the tubes were air dried at 37°C for 2-5 min to evaporate remaining liquid. The dry pellet was resuspended in 50 µl ddH_2O, incubated 5 min at RT and spun down for one second at 13,000 x rpm. To separate the beads from the liquid the tubes were placed on a magnetic holder for 2 min at RT and the supernatant was transferred to a new tube.

Next the previous clean up step was repeated with the same sample to beads ratio and eluted in 37.5 µl for ChIP samples and 25 µl for 5hmC-enriched samples.

From this step the two library preparation protocols differ.

Second part (using agarose gels for size selection)
The samples were now amplified by performing 18 cycles of PCR according to the following mastermix.

Table 5.7: PCR mix (18 cycles) for ChIP-seq

Component	*Volume*	*Final concentration*
Phusion HF buffer (5x)	10 µl	1x
BiooPrimer Mix (100 µM)	2.0 µl	0.5 µM
dNTP mix (10 mM each)	1.5 µl	300 µM
Phusion polymerase	0.5 µl	n.a.
Total Volume	50 µl	

Table 5.8: Parameters for PCR cycling (18 cycles)

	98°C	30 sec
18x	98°C	10 sec
	65°C	30 sec
	72°C	30 sec
	72°C	5 min
	4°C	forever

After PCR the samples were purified according to the cleanup step described above using AMPure XP beads at a sample to beads ratio of 1:1.1 and the DNA was eluted in 50 µl for a second clean up.

The previous step was repeated with a sample to beads ratio of 1:1.1 and the DNA was eluted in 12 µl for size selection on an agarose gel.
2 µl of 5 x DNA loading dye was added to the samples before loading them on a 2% low melting agarose gel (50 ml) together with 1 µg of DNA ladder (NEB) and running it at 60 V for 100 min in pre-cooled running buffer (1 x TAE).
The gel was then placed on an UV screen and 200-350 bp fragments were excised with a scalpel.

DNA was purified from the gel pieces using the QIAquick Gel Extraction kit with slight alterations to the protocol.
First the weight of the gel pieces was determined and 6 volumes of QG per volume of gel were added. In a 5 min incubation step the slices were dissolved at RT in a rotator (speed F1, 6 rpm). After adding 2 volumes of isopropanol per volume of gel the tubes were mixed thoroughly and applied on a MinElute column, spun for 30 seconds at 11000 x rpm and the flowthrough was transferred back onto the column for another centrifugation step. The remaining liquid was discarded and the column was washed with 500 µl of Buffer QG and spun at maximum speed. Two more wash steps were performed with 750 µl of buffer QG and centrifugation at max speed for 45 seconds. After removal of the flowthrough the empty column was centrifuged at maximum speed for 2 min and transferred into a fresh 1.5 ml Eppendorf tube. Remaining liquid was aspirated from the purple ring inside the column and air dried for 1 min at RT. To elute the DNA, 13 µl of EB buffer were heated to 50°C, applied to the membrane and incubated for 5 min at RT. The column was centrifuged for 1 min at 11000 rpm.
DNA quality and fragment size were tested on the Bioanalyzer device (DNA 1000 assay).

This protocol was applied to the PU.1 ChIP-seq samples 7, 8, and 9 (5.2.5.6).

Second part (using the Caliper LabChIP XT for size selection)

The samples were now amplified by performing 4 cycles of PCR according to the following mastermix and cycling protocol.

Table 5.9: PCR mix (12/4 cycles) for 5hmC-seq

Component	*Volume*	*Final concentration*
Phusion HF buffer (5x)	10 µl	1x
Betaine (5 M)	13 µl	1.3 M
BiooPrimer1 (100 µM)	0.25 µl	0.5 µM
BiooPrimer2 (100 µM)	0.25 µl	0.5 µM
dNTP mix (10 mM each)	1 µl	200 µM
Phusion polymerase	0.5 µl	n.a.
Total Volume	50 µl	

Table 5.10: Parameters for PCR cycling (4/12 cycles)

	98°C	2 min
4x / 12x	98°C	30 sec
	65°C	30 sec
	72°C	45 sec
	72°C	5 min
	4°C	forever

After PCR the samples were purified according to the cleanup step described above using AMPure XP beads at a sample to beads ratio of 1:1.8 and the DNA was eluted in 10 µl for Caliper. Measure recovered sample volume and adjust it to a final volume of 10 µl with H_2O.

DNA samples were prepared for size selection on the Caliper LabChip XT according to the manufacturer's manual and separated using the LabChip XT DNA 300 assay kit including one sizing ladder on each chip. Size selection ranges are depicted in Table 5.11.

Table 5.11: Size range for Caliper size selection

Enrichment method	Sample source	Range (in bp)	Range (in %)	Applied to Lib. Nr.
ChIP	Sample	234-316	275 bp +/- 15%	#7-9, 55-63
	Input sample	213-287	250 bp +/- 15%	
5hmC enrichment	Sample	234-316	275 bp +/- 15%	#44-50
	Input sample	213-287	250 bp +/- 15%	
	Sample	238-322	280 bp +/- 15%	#38-43
	Input sample	238-322	280 bp +/- 15%	

The DNA was recovered from the chip in 20 µl elution buffer and filled up to 37.5 µl for ChIP samples and to 25 µl for 5hmC-enriched samples with ddH_2O.
In a second PCR step the DNA fragments were further amplified in 12 PCR cycles according to Table 5.9 and Table 5.10.

After amplification the samples were purified with a 1:1.1 sample to beads ratio and eluted in 50 µl H_2O as described before.
The previous clean up step was repeated with a sample to beads ratio of 1:1.1 and the DNA was recovered in 11 µl H_2O if DNA fragment size and amount was measured using the Bioanalyzer DNA 1000 assay or in 16 µl for the DNA Hyper Sensitivity assay.
In the last step the Bioanalyzer was performed.

This protocol was applied to the PU.1 ChIP-seq, hMeDIP-seq, and Hydroxymethyl Collector™ samples #38-50, 55-63 (5.2.5.2, 5.2.5.6).

Samples from both protocols were now ready for whole genome sequencing. Sequencing was done by the KFB center (Biopark, Regensburg) on an Illumina HiSeq 1000 device (see next chapter).

5.2.2.10 Next generation sequencing (NGS)

Next generation sequencing was carried out at the service facility “KFB - Center of excellence for Fluorescent Bioanalytics” (Regensburg, Germany; www.kfb-regensburg.de). The sample libraries were quantified (KAPA SYBR FAST ABI Prism Library Quantification Kit, Illumina) and cluster generation was performed on the cBot (TruSeq SR Cluster Kit v3, Illumina). Next, the samples were sequenced on an Illumina HiSeq 1000 device. The resulting data provided as .fastq-files and further processed as outlined in chapter (5.2.5).

5.2.3 Working with RNA

5.2.3.1 RNA preparation and isolation

Whole RNA lysates from human blood cells were obtained using the following protocol. Cultured cells were washed and pelleted. Cell pellets were lysed in 350 µl RLT buffer/5 x 10^6 cells containing ß-mercaptoethanol and sheared 10 times using a 1 ml syringe. The lysates were kept on ice at all times and immediately frozen to -80°C for storage.
Total RNA was extracted from the thawed samples using the Qiagen RNeasy Midi or Mini kit according to manufacturer's recommendation. Quality and concentration of the RNA was then verified using agarose gel electrophoresis and the NanoDrop device or the Bioanalyzer.

5.2.3.2 Formaldehyde agarose gel

To check RNA quality, samples were analyzed on a formaldehyde gel. Adding formaldehyde to the system prevents intramolecular baseparing and thus secondary structures which would interfere with the proper separation of the RNA in the gel. 1% of agarose was dissolved in 1 x MOPS buffer by heating in a microwave, cooled down to ~60°C and mixed with 6.5% of formaldehyde. Meanwhile RNA sample were heated at 37°C for 30 min and four volumes of RNA loading dye (1:4) were added. After heating to 65°C the samples were cooled down on ice and ready for loading. Now the polymerized gel was put inside an agarose gel chamber, filled with 1 x MOPS buffer, and loaded with the prepared samples. The gel was run 1 hour at 120 V and analyzed on an UV screen. The gel was handled under a fume hood at all times.

5.2.3.3 qRT-PCR

In the quantitative reverse transcriptase real-time PCR (qRT-PCR) a reverse transcriptase is used to transcribe the RNA to cDNA which is further amplified. For this purpose two kits were used depending on the amount of starting material of RNA.

M-MLV Reverse Transcriptase kit (Promega):

1 µg of RNA

1µl Random decamer primer

1 µl dNTP's (10pmol/ml)

add H_2O

➔ Incubate for 5 min at 65°C

4 µl of 5 x M-LLV Buffer

➔ Mix, centrifuge and incubate 2 min at 42°C

1 µl M-LLV Reverse transcriptase

➔ Incubate 50 min at 42°C and 15 min at 70°C

Transcriptor High fidelity cDNA Synthesis kit (Roche):

100 ng to 1 µg of RNA

2 µl Random hexamer primer

add H_2O

➔ Incubate for 10 min at 65°C, put on ice

4 µl 5 x High fidelity reaction buffer

0.5 µl Protector RNase inhibitor (40U/µl)

2 µl dNTP mix (10 mM each)

1 µl DTT

1.1 µl High Fidelity Reverse transcriptase

➔ Mix, centrifuge and incubate 30 min at 50°C

➔ Inactivate 5 min at 85°C

After a final dilution step (1:5) the samples are now ready for RT-qPCR or storage at -20°C.

5.2.4 Working with Proteins

5.2.4.1 Sample preparation and isolation

Whole protein lysates from human blood cells were prepared according to the following protocol. Isolated cells were washed, counted and all PBS removed from pellets. Fresh cell pellets were then lysed in 200 µl 2x SDS lysis buffer à 3 x 10^6 cells, sheared with a 1 ml syringe to homogenize the sample and boiled for 10 min at 95°C in a shaking Thermomixer (Eppendorf) at 850 rpm. After centrifugation (1min, 13000 rpm, 4°C) the supernatant was transferred to a new tube and stored at -80°C.

Required buffers and solutions:

2 x SDS lysis buffer	10 ml		Glycerin
	6.25 ml	(125 mM)	Stacking gel buffer (1 M Tris, pH 6.8)
	2 g	(4%)	SDS
	5 ml	(10%)	2-Mercaptoethanol (14.4 M)
	10 mg	(0.02%)	Bromphenolblue
	Add ddH_2O to 50 ml.		

5.2.4.2 SDS Page, Western blotting and Immunostaining

Protein samples were separated in a discontinuous gel system consisting of two layers of gel with different salt and polyacrylamide concentrations. The 8% or 10% gels were prepared freshly (Table 5.12) and overlaid with isopropanol alcohol until polymerization.

Table 5.12: SDS-PAGE stock solutions

Gel stock solutions	**Separating gel**		**Stacking gel**
Final AA concentration	8%	10%	5%
Stacking gel buffer	-	-	25 ml
Separating gel buffer	25 ml	25 ml	-
SDS (10%)	1 ml	1 ml	1 ml
Acrylamide (30%)	26,4 ml	33 ml	16,7 ml
Aqua bidest	Adjust to 100 ml		

Table 5.13: SDS-PAGE gel mix solutions

	Separating gel	**Stacking gel**
Separating gel stock solution	6 ml	-
Stacking gel stock solution	-	2,5 ml
TEMED	6 µl	2,5 µl
APS (10%)	30 µl	20 µl

To denature secondary protein structures samples were boiled (95°C, 10 min, shaking 850 rpm), loaded on the gel (1,5 x 10^5 to 5 x 10^5 cells per lane) together with a protein marker (Bio-Rad or Peqlab) and separated for 30-60 min at 100 V and another 1,5-2 h at 120 V in 1 x Laemmli buffer.

According to their size proteins were further transferred to a nitrocellulose membrane using either a semi-dry system (Biometra Fastblot semi-dry blotter) or a wet system (Biorad Mini Transblot Cell).

The semi-dry system is a discontinuous Western blotting system consisting of three different buffers. While the membrane was equilibrated in anode buffer B three Whatman paper soaked with buffer A were stacked into the chamber. Starting on the anode, followed by three Whatman paper soaked in buffer B and then the pre-soaked membrane was added on top of the stack. Next, the gel was placed on the membrane and covered by three Whatman paper soaked in cathode buffer C. With the cathode on top, the chamber was closed and blotted at 11 V for one hour at RT.

In the wet blot system there is only one buffer which surrounds the gel and membrane stack completely. At first three pre-soaked Whatman paper, membrane and gel were pre-soaked in the Wettransfer buffer. Next three Whatman paper were placed on the chamber device followed by the membrane and the gel. On top of the stack followed another three Whatman paper before the chamber was closed and arranged in the right direction in the tank, which was filled up with tank buffer. The proteins were blotted onto the membrane at 100 V for 4 hours at 4°C.
After one hour of blotting the membranes were recovered, marked at standard band sizes with a permanent pen and blocked one hour in 5% dry milk in TBS-T (TBS-(TDG)-T was used for the TDG antibody) at room temperature. The membranes were washed three times in TBS-T before Immunostaining was carried out over night at 4°C (antibody concentrations listed in table (Table 5.14). Second antibody (coupled with horseradish peroxidase (HRP), anti-rabbit, 1:5000) incubation was performed one hour at room temperature. After chemical activation of the antibody (Amersham™ ECL™ kit) fluorescence signals were detected exposing the blot to autoradiography film (Amersham Hyperfilm™ ECL) and developing it in a Fast-Blot machine (Agfa) or by using a fluorescence scanner (Chemi Doc XRS+ System, Bio-Rad).

Table 5.14: Antibodies for Western blotting

Antibody	*Membrane*	*Dilution*	*Species*	*Source/ Company*
TET2	Nitrocellulose	1:2000	rabbit	Dr. Olivier Bernard, Institut Gustave Roussy, Villejuif, France
MBD4	Nitrocellulose	1:2000	rabbit	Diagenode
TDG	Nitrocellulose	1:10000	rabbit	Prof. Dr. Primo Schär, University of Basel, Switzerland

Required buffers and solutions:

Stacking gel buffer	60.6 g	Tris/HCL	
	Add ddH_2O to 500 ml		
Separating gel buffer	90.93 g	Tris/HCL	
	Add ddH_2O to 500 ml		
APS	1 g	Ammonium persulfate	
	Add ddH_2O to 10 ml and store in aliquots at -20°C.		
5 x Laemmli buffer	216 g	(0.95 M)	Glycine
	15 g	(40 mM)	Tris
	15 g	(0.5%)	SDS
	Add ddH_2O to 3 l		
Buffer A	36.3 g	Tris	
	200 ml	Methanol	
	Add ddH_2O to 1 l		
Buffer B	3.03 g	Tris	
	200 ml	Methanol	
	Add ddH_2O to 1 l and adjust pH to 10.4		
Buffer C	5.2 g	ε-Amino-n-caproic acid	
	200 ml	Methanol	
	Add ddH_2O to 1 l		
10 x Tank buffer	140 g	Glycine	
	30 g	Tris	
	5 g	SDS	
	Add ddH_2O to 1 l		
Wettransfer buffer	200 ml	Methanol	
	80 ml	Tank buffer (10x)	
	Add ddH_2O to 1 l		
10 x TBS	45.8 g	(20 mM)	Tris
	175.5 g	(150 mM)	NaCl
	Add ddH_2O to 2 l and adjust pH to 7.4		

1 x TBS (TDG)	100 mM	Tris-HCL pH 8.0	
	150 mM	NaCl	
	0.1%	Tween 20	
	Add ddH_2O to 1 l		
1 x TBS-T	1 l		TBS (1x)
	1 ml	(0.1%)	Tween 20
Blocking solution	5 g	(5%)	dry milk
	Add 100 ml TBS-T		

5.2.4.3 Staining of SDS gels

5.2.4.3.1 Ponceau S staining

After Western blotting the membranes were stained 5 to 10 min with Ponceau S (AppliChem) to check protein transfer. To remove the staining solution membranes were incubated in 5% dry milk in TBS-T on a shaker and further subjected to immunodetection.

5.2.4.3.2 Coomassie staining

In order to check complete protein transfer after Western blotting, gels were incubated in Bio-Safe™ Coomassie solution (Bio-Rad) on a shaker for 5-10 min. Washing the gel in H_2O for 15 min removed the staining.

5.2.4.4 FACS

FACS (Flourescence-activated cell sorting) is a flow cytometry based method to detect intra- and extracellular proteins on/in cells and to sort them accordingly. Therefore cells are washed and labeled with different antibodies at the same time. The LSRII (BD) detects and separates the fluorescence and light scattering signals at single cell resolution and sorts the cells into separate tubes for further processing.
In order to test cell condition after transfection 0,5 x 10^6 to 1 x 10^6 cells were washed in cold FACS buffer and incubated 20 min at 4°C with 5 µl of CD14-Fitc antibody (My4a, Coulter Clone) and 2,5 µl of CD1a-PE antibody (T6-RD1, Coulter Clone) per 500,000 cells (no light exposure). After another washing step with FACS buffer cells were resuspended in 250 µl of fixation buffer. Before subjecting the cells to FACS

measurement on the LSRII (BD) 0.2 µg DAPI (Sigma)/mio cells was added to the suspension.

5.2.5 Next generation sequencing analysis

Analysis of next generation sequencing data sets was performed on an Apple Mac Pro (OS X, v10.9.5, processor 2x2.93 GHz, 6-Core Intel Xeon) as described in the following chapters. The raw sequencing data files were provided in the .fastq file format from the sequencing facility (KFB, Regensburg, Germany) and further processed according to the following scripts. Tools and software used for analysis are given in chapter 5.2.6 and additional data sets as well as in-house pearl scripts are further described in chapter 5.2.5.9.

5.2.5.1 Comparison of two whole genome 5hmC IP methods

For global analyses of 5hmC we tested and compared two 5hmC-enrichment methods, hMeDIP and the Hydroxymethyl Collector™ kit. Data generation is described in chapter 5.2.2.7.2.3 and 5.2.2.7.2.4.
Data processing was done for all hMeDIP-seq and Hydroxymethyl Collector™-seq samples using the following command line and tools as outlined:

#All files:
hMeDIP_MO_Lib38.fastq.gz
hMeDIP_DC18h_Lib39.fastq.gz
hmCcol_MO_Lib40.fastq.gz
hmCcol_DC18h_Lib41.fastq.gz
hmCcol_INPUT_MO_Lib42.fastq.gz
hmCcol_INPUT_DC18h_Lib43.fastq.gz
hmCcol_MO_Lib44.fastq.gz
hmCcol_DCd7_Lib45_GTAGAG.fastq.gz
hmCcol_MO_Lib46.fastq.gz
hmCcol_DC18h_Lib47.fastq.gz
hmCcol_DCd7_Lib48.fastq.gz
hmCcol_INPUT_MO_Lib49.fastq.gz
hmCcol_INPUT_DC18h_Lib50.fastq.gz

###Mapping to hg19 and creating individual tag directories with Bowtie2:

#INPUT:
$myMap-bowtie2.pl -x hg19 -p 23 -name <target folder for bowtie2> <INPUT path to raw data>
$makeTagDirectory <target folder for tag directory> <target folder for bowtie2> -genome hg19 -checkGC ;
#Samples:
$myMap-bowtie2.pl -x hg19 -p 23 -name <target folder for bowtie2> <experiment path to raw data>
$makeTagDirectory <target folder for tag directory> <target folder for bowtie2> -genome hg19 -checkGC ;
Generate combined tag directory files:
$makeTagDirectory <path to combined output tag directory file> -d <path to tag directory 1> <path to tag directory 2>

###Create UCSC file:
$makeUCSCfile <path to tag directory> -o auto

Figure 3.25 und Figure 3.27:

####Generation of tag count distribution scatterplots for tag counts of hMeDIP and Hydroxymethyl Collector™ kit, to compare the two 5hmC-enrichment methods and method replicates:
#Find peaks using the INPUT as control tag directory (Homer tools with options -region -s200 -mD200 -L0):
$findPeaks <path to tag directory> -i <path to INPUT tag directory> -region -size 200 -minDist 200 -L 0 -o <path/filename for output>;

##Remove unmappable regions using Homer tools and BEDtools:
#Conversion to .bed file:
$pos2bed.pl <path to peak file> > <path to output .bed file>

#Remove unmappable regions:
#Set of unmappable regions was produced earlier in the Rehli laboratory.
$intersectBed -v -a <path to .bed file> -b <path to set of unmappable regions > <path to output .bed> ;

###Generate merged peak file with unique and common peaks, without duplicate peaks:
#Merging peak file with reference file to obtain all common and unique peaks without duplicates:
#Find common peaks with minimum overlap of 50%:
$IntersectBed -a <path to file1> -b <path to file2> -wa -wb -f 0.5 <path to output .bed file>

#Find unique peaks in peak file with maximum overlap of 50%:
$IntersectBed -a <path to file1> -b <path to file2> -v -f 0.5 <path to output .bed file>

#Exclude duplicates in common peak files and find common peaks missing in the reference file:
intersectBed -a <path to file1> -b <path to file2> -v -f > <path to output .bed file>

#Merge common peak files of two corresponding files:
$cat <path to file1> <path to file2> > <path to output .bed file>

#Sorting complete common peak file in Excel:

#Sort for minimum start position and maximum stop position, concatenate chromosome-start-stop and #remove duplicates.

#Merge unique peak files and complete common peak file to one file:
$cat <path to file1> <path to file2> > <path to output .bed file>

#Convert .bed file back to peak file .txt:
$bed2pos.pl <bed file> > <path to output peak file .txt>

#Annotate peak files using Homer (with options -log -size given):
$annotatePeaks.pl <path to peak file> <genome hg19> -log -size given > <path to output file>

#Save positions and tag counts for R:
#Generate scatterplots in R:
$setwd("path to peak file")
$data <- read.table("peak file name .txt", header=T, sep="\t")
$Lab.palette <- colorRampPalette(c("blue", "orange", "red"), space = "Lab")
$attach(data)
$per <- cor(D1_38_hMeDIP, D1_40_hmCcol)
$pdf(file="file name .pdf", height=4, width=4)
$par(mar=c(5,5, 0.5, 0.5),cex.lab=1.4, cex.axis=1.1, pty="s", las=1)
$smoothScatter(name x-axis,name y-axis,nrpoints = 0,colramp = Lab.palette,ylab="D1_40_hmCcol", xlab="D1_38_hMeDIP_log2", cex=0.5)
$mtext(per,family="Helvetica", col="black", side=1,line=2,cex=1.3,padj=0.2)
$dev.off()

5.2.5.2 Genome-wide 5hmC analyses

To get a global picture of 5hmC, PU.1 and 5mC dynamics around 5hmC peaks in differentiating monocytes, global analysis of 5hmC was performed using the Hydroxymethyl Collector™ kit. Data generation is described in chapter 5.2.2.7.2.4. PU.1 ChIP-seq and bisulfite-sequencing is described in chapters 5.2.2.8, 5.2.2.9, 5.2.2.10, and 5.2.5.9.2.
Data processing was done for all Hydroxymethyl Collector™-seq samples using the following command line and tools as outlined:

#All files:
hmCcol_MO_Lib40.fastq.gz
hmCcol_DC18h_Lib41.fastq.gz
hmCcol_INPUT_MO_Lib42.fastq.gz
hmCcol_INPUT_DC18h_Lib43.fastq.gz
hmCcol_MO_Lib44.fastq.gz
hmCcol_DCd7_Lib45_GTAGAG.fastq.gz

hmCcol_MO_Lib46.fastq.gz
hmCcol_DC18h_Lib47.fastq.gz
hmCcol_DCd7_Lib48.fastq.gz
hmCcol_INPUT_MO_Lib49.fastq.gz
hmCcol_INPUT_DC18h_Lib50.fastq.gz

####Mapping to hg19 and creating individual tag directories with Bowtie2:
#INPUT:
$myMap-bowtie2.pl -x hg19 -p 23 -name <target folder for bowtie2> <INPUT path to raw data>
$makeTagDirectory <target folder for tag directory> <target folder for bowtie2> -genome hg19 -checkGC ;
#Samples:
$myMap-bowtie2.pl -x hg19 -p 23 -name <target folder for bowtie2> <experiment path to raw data>
$makeTagDirectory <target folder for tag directory> <target folder for bowtie2> -genome hg19 -checkGC ;
Generate combined tag directory files:
$makeTagDirectory <path to combined output tag directory file> -d <path to tag directory 1> <path to tag directory 2>

###Create UCSC file:
$makeUCSCfile <path to tag directory> -o auto

###To assess 5hmC, PU.1, and 5mC dynamics around 5hmC peaks a complete peak file of unique 5hmC peaks of all time points from both donors was generated:
Find differential peaks using another time point (TP) as control tag (option -style factor -L 0 -F 2 -s200) using Homer tools:
$findPeaks <path to tag directory> -i <path to control tag directory> -style factor -F 2 -L 0 -size 200 -o <path/filename for output>;

##Remove unmappable regions using Homer tools and BEDtools:
#Conversion to .bed file:
$pos2bed.pl <path to peak file> > <path to output .bed file>

#Remove unmappable regions with BEDtools:
#Set of unmappable regions was produced earlier in the Rehli laboratory.
$intersectBed -v -a <path to .bed file> -b <path to set of unmappable regions > <path to output .bed>

#Convert .bed file back to peak file .txt:
$bed2pos.pl <bed file> > <path to output peak file .txt>

###Merging all TP into common peak file for cluster analysis with Homer tools:
#Use option -d = maximum distance between peak centers to merge = average peak size/2 = 100
$mergePeaks -d 100 <peak file1> <peak file2> [peak file3] ... > <path to newPeakFile.txt>

#Annotating peaks with all tag counts of Hydroxymethyl Collector^{TM}-seq tag counts including INPUTs:
$annotatePeaks.pl <path to merged peak file> hg19 -size given -d <path to tag directories> > <path to output file>

##Sorting annotated peak file in Excel:
#Sort set for peak enrichment > 4-fold enriched over INPUT.
#Sort for replicate time points enriched > 10 tags.

#Annotating positions with all Hydroxymethyl Collector™-seq tag counts including INPUTs (with option -noadj) for EdgeR:
$annotatePeaks.pl <path to peak file> hg19 -size 200 -noadj -d <path to tag directories> > <path to output file>
#Save IDs and tag counts for R.

###Normalize data, calculate statistics, and estimate clusters in R:
#Setting directory and create a new object in R:
$setwd("path to file")
$library(edgeR)
$data <- read.delim("file name", row.names="ID")
$group <- factor(c(rep("group 1",3), rep("group 2",2), rep("group 3",2)))
$d <- DGEList(counts=data,group=group)
$colnames(d) <- c("column 1","column 2", "column 3", "column 4", "column 5","column 6", "column 7")

###Normalization of data:
$d <- calcNormFactors(d, method=c("upperquartile"))
###Calculating degree of inter-sample variability:
$d <- estimateCommonDisp(d, verbose=TRUE)
#Use normalization with smallest BCV:
upperquartille = Disp = 0.05511 , BCV = 0.2348
#Calculating single values and distribution:
$d <- estimateTagwiseDisp(d, trend="none")
#Transformation to log2:
$y <- cpm(d, prior.count=2, log=TRUE)
#Printing data:
$write.table(as.matrix(y),file="log2-transformed data file name",sep="\t", col.names=NA)

###Calculating statistical analysis between MO and DC time points:
$de.com <- exactTest(d, pair=c("TP 1","TP 2"))
$results <- topTags(de.com,n = Inf)
$write.table(as.matrix(results$table),file="output file name",sep="\t", col.names=NA)

##Sorting log2-transformed file in Excel:
#First the TPs were median-normalized
#and sorted for FDR.
#Sort statistical analysis files for FDR <0.05 and
#delete all peaks in median file that don't have FDR <0.05.

###Performing k-means clustering in R:
##Estimate cluster-count with k-means:
$library(gplots)
$library(RColorBrewer)

```
$s <- read.table("significant data file name", header=TRUE, sep="	", row.names="ID")
$data <- as.matrix(s)
$wss <- (nrow(data)-1)*sum(apply(data,2,var))
$for (i in 2:15) wss[i] <- sum(kmeans(data,
		centers=i)$withinss)
$plot(1:15, wss, type="b", xlab="Number of Clusters",
 ylab="Within groups sum of squares")
$fit <- kmeans(data, estimated cluster-number)
$aggregate(data,by=list(fit$cluster),FUN=mean)

#Appending cluster assignment:
$mydata <- data.frame(data, fit$cluster)
$write.table(as.matrix(mydata),file="cluster file name",sep="\t", col.names=NA)

#The 5hmC cluster file was used for the following analyses.
```

Figure 3.28 A:

####Illustrating the dynamics of 5hmC and PU.1 in distance distribution heatmaps:
#Sort cluster file for cluster-number in Excel.
##Sorting clusters for maximum enrichment:
#Annotating sorted cluster file with Hydroxymethyl Collector™-seq tag and PU.1 ChIP-seq tag counts to retrieve tag information using Homer tools:
$annotatePeaks.pl <path to peak file> hg19 -size given -d <path to tag directories> > <path to output file>
#Sorting for maximum tag count in averaged replicate time points in Excel, and save position data.

###Generating distance distribution heatmaps of 5hmC and PU.1 around 5hmC peaks using Homer tools and R:
#Annotate position file with combined 5hmC and PU.1 tag counts of replicate TPs:
$annotatePeaks.pl <path to peak file> hg19 -size 1000 -hist 25 -ghist -d <path to combined tag directories> > <path to output file>

```
#Use generated files for R:
#For PU.1 heatmaps:
$setwd("path to ghist file")
$data <- read.delim("ghist file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,3,length=100), seq(3,6,length=100))
$png(filename="file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, density.info="none", trace="none", lhei = c(0.05,0.95), lwid = $c(0.05,0.95))
$dev.off()
```

#For 5hmc heatmaps:

```
$setwd("path to ghist file")
$data <- read.delim("ghist file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,5,length=100), seq(5,10,length=100))
$png(filename="file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, density.info="none", trace="none", lhei = c(0.05,0.95), lwid = $c(0.05,0.95))
$dev.off()
```

Figure 3.28 B:

####Illustrating the dynamics of 5hmC and PU.1 in histograms:

#Sort file for cluster-number in Excel.

###Sorting clusters for maximum enrichment:

#Annotating sorted cluster file with Hydroxymethyl Collector^{TM}-seq and PU.1 ChIP-seq tag counts to retrieve tag information using Homer tools:

$annotatePeaks.pl <path to peak file> hg19 -size given -d <path to tag directories> > <path to output file>

#Sorting for maximum tag count in averaged replicate time points in Excel, and save position data.

###Generating histograms of 5hmC and PU.1 tag counts around 5hmC peaks using Homer tools and R:

#Split sorted cluster position file into separate clusters.

#Annotate split cluster files with combined 5hmC and PU.1 tag counts of replicate TPs:

$annotatePeaks.pl <path to peak file> hg19 -size 1000 -hist 25 -d <path to combined tag directories> > <path to output file>

#Split annotated files into 5hmC and PU.1, reduce to coverages,

#and generate histograms of 5hmC and PU.1 distribution around 5hmC peaks in R:

```
$setwd("path to file")
$data <- read.table("hmC or PU.1 file name", header=T, sep="\t")
$attach(data)
$plotcolors <- c("color 1", "color 2", "color 3")
$pdf(file="file name .pdf", height=4, width=8)
$par(mar=c(3,3,1,3))
$plot(Distance,TP 1,type="l",col=plotcolors[3],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16), xlim=c(-500,500))
$par(mar=c(3,3,1,3), new=TRUE)
$plot(Distance,TP 2,type="l",col=plotcolors[2],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16), xlim=c(-500,500))
$par(mar=c(3,3,1,3), new=TRUE)
$plot(Distance,TP 3,type="l",col=plotcolors[1],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16), xlim=c(-500,500))
$axis(1,padj=-1.0,family="Helvetica",cex.axis=1,at=c(-500,-400,-300,-200,-100,0,100,200,300,400,500))
$axis(2,padj=0.8,family="Helvetica",cex.axis=1, col="black", col.axis="black")
$mtext("name y-axis",family="Helvetica", col="black", side=2,line=2,cex=1.3,padj=0.2)
```

```
$mtext("name x-axis",family="Helvetica",ps=12,side=1,line=2,cex=1.3,padj=-0.6)
$dev.off()
```

Figure 3.28 C:

```
###Plotting heatmaps of 5mC distribution around 5hmC peaks:
#Use cluster file generated in this chapter, reduce it to positions, and split into separate clusters.
#Use processed files from chapter 5.2.5.9.2.

#Reformat data and plot graph using R:
#Command averages methylation level over all valid CpGs at each position:
$myMethylHist_MODC.pl <path to split cluster peak file> hg19 <path to rescaled MO file> <path to rescaled DC file> -scaled -o <path to output file>

$x <- read.table("path to reformatted file", header=T, sep="    ")
$attach(x)
$sS1 = smooth.spline(Distance,BG1, spar=0.7)
$sS2 = smooth.spline(Distance,BG2, spar=0.7)
$pdf(file="path and file name .pdf", width=3, height=3)
$par(mar=c(1.4,1.6,0.2,0.2))
$plot(Distance,BG1,type="p",col=adjustcolor("brown1",alpha=0.3), xaxt="n",yaxt="n",xlab="",
ylab="",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))
$par(mar=c(1.4,1.6,0.2,0.2), new=TRUE)
$plot(Distance,BG2,type="p",col=adjustcolor("cyan3",alpha=0.3), xaxt="n",yaxt="n",xlab="",
ylab="",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))
$axis(1,padj=-1.2,family="Helvetica",cex.axis=0.8)
$axis(2,padj=0.8,family="Helvetica",cex.axis=0.8)
$lines(sS1, col="color 1", lwd=2)
$lines(sS2, col="color 2", lwd=2)
$dev.off()
```

5.2.5.3 Quality control of 5hmC, PU.1 and 5mC data

To check if the sequencing data produced in different laboratories was comparable and thus if the observed dynamics were specific or due to technical variation, we compared the sequencing data with a set of random peaks. The random peak set was produced by chance using Homer tools (findMotifsGenome.pl).

Figure 3.29 A:

```
###Generating distance distribution heatmaps of 5hmC and PU.1 around random peaks using Homer tools and R:
#Use random peak file generated with Homer and
#annotate it with combined 5hmC and PU.1 tag counts of replicate TP:
```

$annotatePeaks.pl <path to random peak file> hg19 -size 1000 -hist 25 -ghist -d <path to combined tag directories> > <path to output file>

#Use ghist files for R:
#For PU.1 heatmaps:

```
$setwd("path to ghist file")
$data <- read.delim("ghist file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,3,length=100), seq(3,6,length=100))
$png(filename="file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, density.info="none", trace="none", lhei = c(0.05,0.95), lwid = $c(0.05,0.95))
$dev.off()
```

#For 5hmC heatmaps:

```
$setwd("path to ghist file")
$data <- read.delim("ghist file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,5,length=100), seq(5,10,length=100))
$png(filename="file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, density.info="none", trace="none", lhei = c(0.05,0.95), lwid = $c(0.05,0.95))
$dev.off()
```

Figure 3.29 B:

###Generating histograms of 5hmC and PU.1 tag counts around random peaks using Homer tools and R:
#Annotate random peak file with combined 5hmC and PU.1 tag counts of replicate TPs:
$annotatePeaks.pl <path to random peak file> hg19 -size 1000 -hist 25 -d <path to combined tag directories> > <path to output file>

#Split annotated files into 5hmC and PU.1, reduce to coverages,
#and generate histograms for 5hmC and PU.1 distribution around random peaks in R:

```
$setwd("path to 5hmC or PU.1 file")
$data <- read.table("file name", header=T, sep="\t")
$attach(data)
$plotcolors <- c("color 1", "color 2", "color 3")
$pdf(file="file name .pdf", height=4, width=8)
$par(mar=c(3,3,1,3))
$plot(Distance,TP 1,type="l",col=plotcolors[3],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16), xlim=c(-500,500))
$par(mar=c(3,3,1,3), new=TRUE)
```

```
$plot(Distance,TP 2,type="l",col=plotcolors[2],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16),
xlim=c(-500,500))
$par(mar=c(3,3,1,3), new=TRUE)
$plot(Distance,TP 3,type="l",col=plotcolors[1],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16),
xlim=c(-500,500))
$axis(1,padj=-1.0,family="Helvetica",cex.axis=1,at=c(-500,-400,-300,-200,-100,0,100,200,300,400,500))
$axis(2,padj=0.8,family="Helvetica",cex.axis=1, col="black", col.axis="black")
$mtext("name y-axis",family="Helvetica", col="black", side=2,line=2,cex=1.3,padj=0.2)
$mtext("name x-axis",family="Helvetica",ps=12,side=1,line=2,cex=1.3,padj=-0.6)
$dev.off()
```

Figure 3.29 C:

###Generating histograms of 5mC distribution around random peaks in R:
#Use files generated in this chapter and from chapter 5.2.5.9.2.
#Reformat data and plot graph using R:
#Command averages methylation level over all valid CpGs at each position:
$myMethylHist_MODC.pl <path to random peak file> hg19 <path to rescaled MO file> <path to rescaled DC file> -scaled -o <path to output file>

```
$x <- read.table("path to previous reformatted file", header=T, sep="     ")
$attach(x)
$sS1 = smooth.spline(Distance,BG1, spar=0.7)
$sS2 = smooth.spline(Distance,BG2, spar=0.7)
$pdf(file="path and file name .pdf", width=3, height=3)
$par(mar=c(1.4,1.6,0.2,0.2))
$plot(Distance,BG1,type="p",col=adjustcolor("brown1", alpha=0.3), xaxt="n",yaxt="n",xlab="",
ylab="",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))
$par(mar=c(1.4,1.6,0.2,0.2), new=TRUE)
$plot(Distance,BG2,type="p",col=adjustcolor("cyan3", alpha=0.3), xaxt="n",yaxt="n",xlab="",
ylab="",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))
$axis(1,padj=-1.2,family="Helvetica",cex.axis=0.8)
$axis(2,padj=0.8,family="Helvetica",cex.axis=0.8)
$lines(sS1, col="brown4", lwd=2)
$lines(sS2, col="darkcyan", lwd=2)
$dev.off()
```

5.2.5.4 Gene ontology (GO) analyses of 5hmC peaks

Gene ontology (GO) analyses were performed in the 5hmC cluster sets generated in chapter 5.2.5.2.

####Gene ontology analyses in cluster sets of 5hmC peaks using R:
#Use 5hmC cluster file generated in 5.2.5.2.
#Sort file for cluster-number in Excel

###Sorting clusters for maximum enrichment:
#Annotating sorted cluster file with all 5hmC tag counts using Homer tools:
$annotatePeaks.pl <path to peak file> hg19 -size given -d <path to tag directories> > <path to output file>

#Sort for maximum enrichment in averaged replicate time points in Excel, save position data,
#and split into separate cluster files.
#Annotating sorted cluster file to generate gene ontology information using Homer tools:
$annotatePeaks.pl <path to split cluster peak file> hg19 -go <path to gene ontology output file> > <path to annotated output file>

##Calculating GeneFoldEnrichment of GO terms and transform data into log2:
#Calculate statistics using the Hypergeomvert test in Excel:
$hypgeomvert(target genes in term; total target genes; genes in term; total genes)
#genes in term = all genes with annotation
#target genes in term = genes in set with annotation
#total genes = all genes
#total target genes = all genes in set

#Calculate FDR (= (p-value*N)/rank) and
#sort for peaks with a significant FDR.
#Sort for GO terms with > 2-fold enriched GFE.

##Clustering GO terms and generate dendrogram in R:
$setwd("path to sign. GO term file")
$library(edgeR)
$library(gplots)
$library(RColorBrewer)
$data <- read.delim("file name of sign. GO term file", row.names="ID")
#Transform into log2:
$y <- log2(data)
$write.table(as.matrix(y),file="log2 file name",sep="\t", col.names=NA)
#Add random jitter to correct for zeros.

$z <- read.table("corrected log2 file", header=TRUE, sep="	", row.names="ID")
$data <- as.matrix(z)
$hc <- hclust(dist(t(data), method = "manhattan"), method="ward")
$hr <- hclust(dist(data, method = "manhattan"), method="ward")
$mycol <- colorRampPalette(c("blue","white","red"))(299)
$col_breaks = c(seq(-2,-0.33333,length=100), seq(-0.33333,0.33333,length=100), seq(0.33333,2,length=100))
$pdf(file="file name .pdf", height=8, width=8)
$heatmap.2(data, Rowv=as.dendrogram(hr), Colv=as.dendrogram(hc), col = mycol, breaks = col_breaks, na.rm=TRUE, scale="none", margins=c(10,10), cexRow=0.5, cexCol=1.0, key=TRUE, density.info="none", trace="none")
$dev.off()
#Extract dendrogram info and GFE
$write.table(t(data[rev(hr$labels[hr$order]), hc$labels[hc$order]]), file = "dendrogram file name", sep = "", col.names=NA)

##Sort GO terms in Excel:
#Sort for major GO terms consisting of min. 200 genes.
#Sort major GO terms for terms enriched exclusively in set #2/#4 in Excel,
#and print heatmaps in R:

```
$setwd("path to folder")
$library(gplots)
$library(RColorBrewer)
$z <- read.delim("GOterm file name", row.names="ID")
$data <- as.matrix(z)
$mycol <- colorRampPalette(c("blue","white","red"))(299)
$col_breaks = c(seq(0.65,0.9,length=100), seq(0.9,1.1,length=100), seq(1.1,1.5,length=100))
$pdf(file="file name .pdf", height=8, width=8)
$heatmap.2(data, Rowv=NA, Colv=NA, col = mycol, breaks = col_breaks, dendrogram = "none", na.rm=TRUE, scale="none", margins=c(10,10), cexRow=0.5, cexCol=1.0, key=TRUE, density.info="none", trace="none")
$dev.off()
```

5.2.5.5 Global gene expression during MO differentiation

To investigate a possible correlation of demethylation events with transcriptional changes, the 5hmC cluster peak set (generated in chapter 5.2.5.2) was compared with publically available gene expression data obtained from differentiating MO (5 donors) and iDC (3 donors) (see chapter 5.2.5.9.1).

####Gene expression analyses in cluster sets of 5hmC peaks using R:
#Use 5hmC cluster file generated in 5.2.5.2.
#Sort file for cluster-number in Excel.
###Sorting clusters for maximum enrichment:
#Annotating sorted cluster file with all Hydroxymethyl Collector™-seq tag counts using Homer tools:
$annotatePeaks.pl <path to peak file> hg19 -size given -d <path to tag directories> > <path to output file>

#Sort for maximum enrichment in averaged replicate time points using Excel, save position data,
#and split into separate cluster files.

###Adjust microarray expression data file of MO and DC [GEO:GSE19236] (see chapter 5.2.5.9.1):
#Extract and average normalized TP of donors and
#retrieve genebank accession numbers from Agilent homepage
(http://www.chem.agilent.com/cag/bsp/gene_lists.asp).

#Annotate split cluster files with expression data file using Homer:
$annotation.pl <path to split cluster file> hg19 -gene <expression data file> > <path to output file>
#Sort for annotated peaks and save positions for R.

##Calculate statistics in R:

```
#Use Man-Whitney-U test to calculate significant changes between MO and DC time points:
$setwd("path to annotated file")
$x <- read.table("annotated file name", header=T, sep="\t")
$attach(x)
$wilcox.test(TP 1,TP 2,data=x)

#Generate boxplots without outliers to improve visualization of expression changes:
$setwd("path to expression data file")
$x <- read.table("expression data file", header=T, sep="\t")
$logx <- log2(x)
$pdf(file="boxplot file name .pdf", height=4, width=5)
$boxplot(logx,what=c(1,1,1,0),axes = FALSE, col = "lightgray", border = "black",overallline = "median", notch=TRUE,
outline=FALSE, range=2)
$axis(1,padj=-0.8,family="Helvetica",cex.axis=1,at=1:7,lab=c("MO","6h","18h","27h","42h","51h","66h"))
$axis(2,padj=0.8,family="Helvetica",cex.axis=1)
$mtext("name y-axis",family="Helvetica",side=2,line=2,cex=1.2,padj=0.4)
$mtext("name x-axis",family="Helvetica",ps=12,side=1,line=2,cex=1.2,padj=-0.6)
$dev.off()
```

5.2.5.6 Genome-wide PU.1 analyses

To get further insights into global PU.1, 5hmC and 5mC dynamics around PU.1 peaks a set of all differential, time point-specific PU.1 ChIP-seq peaks was analyzed. The ChIP-seq procedure and generation of the data is described in chapter 5.2.2.8, 5.2.2.9, and 5.2.2.10. For data generation of 5hmC and 5mC tags see chapter 5.2.2.7.2.4, 5.2.2.9, 5.2.2.10, and 5.2.5.9.2.

Data processing was done for all ChIP-seq samples using the following command line and tools as outlined:

```
#All files:
PU1ChiP_MN_Lib55.fastq.gz
PU1ChIP_DC18h_Lib56.fastq.gz
PU1ChIP_INPUT_DC18h_Lib57.fastq.gz
PU1ChIP_DC18h_Lib58.fastq.gz
PU1ChIP_DC42h_Lib59.fastq.gz
PU1ChIP_DCd7_Lib60.fastq.gz
PU1ChIP_MN_Lib61.fastq.gz
PU1ChIP_DC27h_Lib62.fastq.gz
PU1ChIP_DC51h_Lib63.fastq.gz
PU1ChIP_DC66h_S7.fastq.gz
PU1ChIP_DCd7_S8.fastq.gz
PU1ChIP_INPUT_DCd7_S9.fastq.gz
```

###Mapping of .fastq files to hg19 and creating individual tag directories with Bowtie2:
#INPUT:
$myMap-bowtie2.pl -x hg19 -p 23 -name <target folder for bowtie2> <INPUT path to raw data>
$makeTagDirectory <target folder for tag directory> <target folder for bowtie2> -genome hg19 -checkGC ;
#Samples:
$myMap-bowtie2.pl -x hg19 -p 23 -name <target folder for bowtie2> <Experiment path to raw data>
$makeTagDirectory <target folder for tag directory> <target folder for bowtie2> -genome hg19 -checkGC ;
Generate combined tag directory files:
$makeTagDirectory <path to combined output tag directory file> -d <path to tag directory 1> <path to tag directory 2>

###Create UCSC file:
$makeUCSCfile <path to tag directory> -o auto

#####Define transcription factor peaks in ChiP-seq data:
#Find differential peaks using another time point (TP) as control tag directory (Homer tools with option -style factor = for TF ChiP-seq analysis, centers peaks on maximum tag overlap):
$findPeaks <path to tag directory> -i <path to control tag directory> -style factor -o <path to output directory>

###Compiling a complete set of all differential peaks for cluster generation:
#Merging all differential peak sets from all donors with Homer tools:
#Use option -d = maximum distance between peak centers to merge = average peak size/2 = 104
$mergePeaks -d 104 <peak file1> <peak file2> [peak file3] ... > <path to output peak file>

#Annotating merged peak file with all PU.1 ChIP-seq tag counts inclusive INPUT:
#Use option -size = average peak size = 207
$annotatePeaks.pl <path to peak file> hg19 -size 207 -d <path to PU.1/INPUT tag directories> > <path to output file>
#Sort annotated peak file in Excel:
#Sort for replicates that are enriched > 4-fold over the INPUT and
#for replicate-pairs enriched > 10 tags.
#Annotate position file with option -noadj (= skips tag count normalization based on total tags sequenced) for EdgeR:
$annotatePeaks.pl <path to peak file> hg19 -size 207 -noadj -d <path to PU.1 tag directories> > <path to output file>

###Normalize data, calculate statistics, and estimate clusters in R:
#Setting directory and create a new object in R:
$setwd("path to file")
$library(edgeR)
$data <- read.delim("file name", row.names="ID")
$group <- factor(c(rep("group 1",3), rep("group 2",2), rep("group 3",2)))
$d <- DGEList(counts=data,group=group)
$colnames(d) <- c("column 1","column 2", "column 3", "column 4", "column 5","column 6", "column 7")

###Normalization of data:

```
$d <- calcNormFactors(d, method=c("upperquartile"))
```

###Calculating degree of inter-sample variability:

```
$d <- estimateCommonDisp(d, verbose=TRUE)
```

#Using normalization with smallest BCV:

upperquartille = Disp = 0.11732 , BCV = 0.3425

#Calculating single values and distribution:

```
$d <- estimateTagwiseDisp(d, trend="none")
```

#Transformation to log2:

```
$y <- cpm(d, prior.count=2, log=TRUE)
```

###Calculating statistical analysis between MO and DC time points:

```
$de.com <- exactTest(d, pair=c("TP 1","TP 2"))
$results <- topTags(de.com,n = Inf)
$write.table(as.matrix(results$table),file="output file name",sep="\t", col.names=NA)
```

##Sorting log2-transformed file in Excel:

#Calculate medians of all TPs.

#Sort statistical analysis files for FDR <0.05 and

#delete all peaks in median file that don't have FDR <0.05.

###Performing k-means clustering in R:

##Estimate cluster-count with k-means:

```
$library(gplots)
$library(RColorBrewer)
$s <- read.table("significant data file name", header=TRUE, sep="	", row.names="ID")
$data <- as.matrix(s)
$wss <- (nrow(data)-1)*sum(apply(data,2,var))
$for (i in 2:15) wss[i] <- sum(kmeans(data, centers=i)
$withinss)
$plot(1:15, wss, type="b", xlab="Number of Clusters", ylab="Within groups sum of squares")
$fit <- kmeans(data, estimated cluster-number)
$aggregate(data,by=list(fit$cluster),FUN=mean)
```

#Appending cluster assignment:

```
$mydata <- data.frame(data, fit$cluster)
$write.table(as.matrix(mydata),file="cluster file name",sep="\t", col.names=NA)
```

#Sort cluster file for cluster-number in Excel

#Annotating sorted cluster file with PU.1 ChIP-seq and Hydroxymethyl Collector^{TM}-seq tag counts to retrieve tag information using Homer tools:

$annotatePeaks.pl <path to peak file> hg19 -size 207 -d <path to PU.1 tag directories> > <path to output file>

#The annotated cluster file was used for the following analyses.

Figure 3.32 A:

###Illustrating the dynamics of PU.1 and 5hmC in distance distribution heatmaps:
##Sorting clusters in cluster file for maximum enrichment:
#Sort averaged replicates for maximum tag count and reduce to positions.

#Generating distance distribution heatmaps of PU.1 and 5hmC around PU.1 peaks using Homer tools and R:
#Annotate position file with combined PU.1 and 5hmC tag counts of replicate TP:
$annotatePeaks.pl <path to peak file> hg19 -size 1000 -hist 25 -ghist -d <path to combined tag directories> > <path to output file>

#Use generated files for R:
#For PU.1 heatmaps:

```
$setwd("path to ghist file")
$data <- read.delim("ghist file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,3,length=100), seq(3,6,length=100))
$png(filename="file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, density.info="none", trace="none", lhei = c(0.05,0.95), lwid = $c(0.05,0.95))
$dev.off()
```

#For 5hmc heatmaps:

```
$setwd("path to ghist file")
$data <- read.delim("ghist file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,5,length=100), seq(5,10,length=100))
$png(filename="file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, density.info="none", trace="none", lhei = c(0.05,0.95), lwid = $c(0.05,0.95))
$dev.off()
```

Figure 3.32 B:

###Illustrating the dynamics of PU.1 and 5hmC in histograms:
###Sorting clusters in cluster file for maximum enrichment:
#Sort averaged replicates for maximum tag count and reduce to positions.

###Generating histograms of 5hmC and PU.1 tag counts using Homer tools and R:
#Split cluster position file into 6 separate clusters.
#Annotate split cluster files with combined PU.1 and 5hmC tag counts of replicate TP:

$annotatePeaks.pl <path to peak file> hg19 -size 1000 -hist 25 -d <path to combined tag directories> > <path to output file>

#Split annotated files into 5hmC and PU.1, reduce to coverages
#and generate histograms of 5hmC and PU.1 distribution around 5hmC peaks in R:

```
$setwd("path to file")
$data <- read.table("hmC or PU.1 file name", header=T, sep="\t")
$attach(data)
$plotcolors <- c("color 1", "color 2", "color 3")
$pdf(file="file name .pdf", height=4, width=8)
$par(mar=c(3,3,1,3))
$plot(Distance,TP 1,type="l",col=plotcolors[3],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16), xlim=c(-500,500))
$par(mar=c(3,3,1,3), new=TRUE)
$plot(Distance,TP 2,type="l",col=plotcolors[2],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16), xlim=c(-500,500))
$par(mar=c(3,3,1,3), new=TRUE)
$plot(Distance,TP 3,type="l",col=plotcolors[1],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16), xlim=c(-500,500))
$axis(1,padj=-1.0,family="Helvetica",cex.axis=1,at=c(-500,-400,-300,-200,-100,0,100,200,300,400,500))
$axis(2,padj=0.8,family="Helvetica",cex.axis=1, col="black", col.axis="black")
$mtext("name y-axis",family="Helvetica", col="black", side=2,line=2,cex=1.3,padj=0.2)
$mtext("name x-axis",family="Helvetica",ps=12,side=1,line=2,cex=1.3,padj=-0.6)
$dev.off()
```

Figure 3.32 C:

###Plotting heatmaps of 5mC distribution around PU.1 peaks:
#Use annotated cluster file generated in this chapter, reduce it to positions and split into separate clusters.
#Use processed files from chapter 5.2.5.9.2.

##Reformat bisulfite-data and plot graph in R:
#Command averages methylation level over all valid CpGs at each position:
$myMethylHist_MODC.pl <path to split cluster peak file> hg19 <path to rescaled MO file> <path to rescaled DC file> -scaled -o <path to output file>

```
$x <- read.table("path to reformatted file", header=T, sep="    ")
$attach(x)
$sS1 = smooth.spline(Distance,BG1, spar=0.7)
$sS2 = smooth.spline(Distance,BG2, spar=0.7)
$pdf(file="path and file name .pdf", width=3, height=3)
$par(mar=c(1.4,1.6,0.2,0.2))
$plot(Distance,BG1,type="p",col=adjustcolor("brown1",alpha=0.3), xaxt="n",yaxt="n",xlab="",
ylab="",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))
$par(mar=c(1.4,1.6,0.2,0.2), new=TRUE)
$plot(Distance,BG2,type="p",col=adjustcolor("cyan3",alpha=0.3), xaxt="n",yaxt="n",xlab="",
ylab="",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))
$axis(1,padj=-1.2,family="Helvetica",cex.axis=0.8)
```

$axis(2,padj=0.8,family="Helvetica",cex.axis=0.8)
$lines(sS1, col="color 1", lwd=2)
$lines(sS2, col="color 2", lwd=2)
$dev.off()

5.2.5.7 Promoter set analysis

To assess the binding dynamics of PU.1, 5hmC and 5mC around promoters we used an openly available promoter set published by the FANTOM consortium (Consortium *et al.*, 2014) (see chapter 5.2.5.9.3). As described in (Consortium *et al.*, 2014) the promoter set was generated based on transcription, which indicated active usage.

#To obtain cell type-specific sets of actively used promoters, we annotated the promoter set with CAGE enhancer expression data from MO (3 donors) and DC (3 donors) (Consortium *et al.*, 2014).
$annotatePeaks.pl <path to promoter file> hg19 -size given -len 0 -strand + -d <path to CAGE tag directories>-noadj -noann -nogene > <path to output file>

#Sort promoter file for at least one TP per row with a minimum tag count of 1 in Excel.
#Calculate statistics (FDR, logFoldChange, logCPM, p-value) in R:
$setwd("path to file")
$library(edgeR)
$data <- read.delim("annotated file name", row.names="ID")
$group <- factor(c(rep("TP 1",3), rep("TP 2",4)))
$lib.size <- c(1000000,1000000,1000000,1000000,1000000,1000000,1000000)
$d <- DGEList(counts=data,group=group,lib.size=as.numeric(lib.size))
$colnames(d) <- c("column name 1", "column name 2","column name 3","column name 4","column name 5","column name 6","column name 7")
$d <- estimateCommonDisp(d, verbose=TRUE)
#Disp = 0.05591 , BCV = 0.2365
$d <- estimateTagwiseDisp(d, trend="none")
$de.com <- exactTest(d, pair=c("MO","mDC"))
$results <- topTags(de.com,n = Inf)
$write.table(as.matrix(results$table),file="EdgeR.results.txt",sep="\t", col.names=NA)

##Sorting annotated file in Excel:
#Sort for FDR<0.05 and >2fold change in log-Fold-Change and
#sort for cell type-specific expression and divide set into cell type-specific MO or DC sets.
#Sort for max tag count and reduce to positions.
#The generated cell type-specific promoter sets are used for the following analyses.

Figure 3.33 A:

###Illustrating the dynamics of PU.1 and 5hmC in distance distribution heatmaps:

#Annotate the promoter sets with local PU.1 ChIP-seq and Hydroxymethyl Collector™-seq tag counts using Homer tools:
$annotatePeaks.pl <path to promoter file> hg19 -size 207 -d <path to tag directories> > <path to output file>
#Sort for maximum PU.1 tag count in averaged replicate time points (in Excel) and save position data.

###Generating distance distribution heatmaps of PU.1 and 5hmC around promoter TSS using Homer tools and R:
#Annotate cell type-specific promoter sets with tag counts of PU.1 ChIP-seq and Hydroxymethyl Collector™-seq samples using Homer tools:
$annotatePeaks.pl <path to promoter file> hg19 -ghist -hist 25 -s1000 -d <path to tag directories> > <path to output file>

#Use generated files for R:
#For PU.1 heatmaps:
$setwd("path to ghist file")
$data <- read.delim("ghist file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,3,length=100), seq(3,6,length=100))
$png(filename="file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, density.info="none", trace="none", lhei = c(0.05,0.95), lwid = $c(0.05,0.95))
$dev.off()

#For 5hmc heatmaps:
$setwd("path to ghist file")
$data <- read.delim("ghist file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,5,length=100), seq(5,10,length=100))
$png(filename="file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, density.info="none", trace="none", lhei = c(0.05,0.95), lwid = $c(0.05,0.95))
$dev.off()

Figure 3.33 A (right panel):

###Investigation of CpG density coverage:
#Annotate cell type-specific promoter sets with tag counts of PU.1 ChIP-seq and Hydroxymethyl Collector™-seq samples to retrieve tag count using Homer tools:
$annotatePeaks.pl <path to promoter file> hg19 -size 207 -d <path to tag directories> > <path to output file>
#Sort for maximum PU.1 tag count in averaged replicate time points (in Excel) and save position data.

###Generating distance distribution heatmaps of CpG-content around cell type-specific promoters using Homer tools and R:
#Use a reference set of all valid CpGs of the genome, containing CpG-locations extended to 100 bp each, in order to create a tag directory file.
#Separate the promoter file into cell type-specific sets of positions and
#annotate the sets with the reference CpG tag directory using Homer tools:
$annotatePeaks.pl <path to cell type-specific promoter file> hg19 -size 1000 -hist 25 -ghist -d <path to CpG tag directory> > <path to output file>

#Use generated files for R:
$setwd("path to annotated file")
$data <- read.delim("annotated file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,3,length=100), seq(3,6,length=100))
$png(filename="distance distribution heatmap file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, $density.info="none", trace="none", lhei = c(0.05,0.95), lwid = c(0.05,0.95))
$dev.off()

Figure 3.33 B:

###Illustrating the dynamics of PU.1 and 5hmC in histograms:
#Annotate cell type-specific promoter sets with tag counts of PU.1 ChIP-seq and Hydroxymethyl Collector™-seq samples to retrieve tag count using Homer tools:
$annotatePeaks.pl <path to promoter file> hg19 -size 207 -d <path to tag directories> > <path to output file>
#Sort for maximum PU.1 tag count in averaged replicate time points (in Excel) and save position data.

##Generating histograms of PU.1and 5hmC tag distribution around cell type-specific promoters using Homer tools and R:
#Separate into cell type-specific sets of positions and
#annotate cell type-specific promoter sets with tag counts of PU.1 ChIP-seq and Hydroxymethyl Collector™-seq samples using Homer tools:
$annotatePeaks.pl <path to promoter file> hg19 -size 1000 -hist 25 -d <path to tag directories> > <path to output file>

#Split annotated file into PU.1 or 5hmC coverages for R.
##Generate histograms in R:
$setwd("path to file")
$data <- read.table("hmC or PU.1 file name", header=T, sep="\t")
$attach(data)
$plotcolors <- c("color 1", "color 2", "color 3")
$pdf(file="file name .pdf", height=4, width=8)
$par(mar=c(3,3,1,3))

```
$plot(Distance,TP 1,type="l",col=plotcolors[3],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16),
xlim=c(-500,500))
$par(mar=c(3,3,1,3), new=TRUE)
$plot(Distance,TP 2,type="l",col=plotcolors[2],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16),
xlim=c(-500,500))
$par(mar=c(3,3,1,3), new=TRUE)
$plot(Distance,TP 3,type="l",col=plotcolors[1],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16),
xlim=c(-500,500))
$axis(1,padj=-1.0,family="Helvetica",cex.axis=1,at=c(-500,-400,-300,-200,-100,0,100,200,300,400,500))
$axis(2,padj=0.8,family="Helvetica",cex.axis=1, col="black", col.axis="black")
$mtext("name y-axis",family="Helvetica", col="black", side=2,line=2,cex=1.3,padj=0.2)
$mtext("name x-axis",family="Helvetica",ps=12,side=1,line=2,cex=1.3,padj=-0.6)
$dev.off()
```

Figure 3.33 C:

###Plotting heatmaps of 5mC distribution in the vicinity of cell type-specific promoter regions:
#Use files generated in this chapter and from chapter 5.2.5.9.2.

#Reformat data and plot graph using R:
#Command averages methylation level over all valid CpGs at each position:
$myMethylHist_MODC.pl <path to cell type-specific promoter file> hg19 <path to rescaled MO file> <path to rescaled DC file> -scaled -o <path to output file>

```
$x <- read.table("path to reformatted file", header=T, sep="    ")
$attach(x)
$sS1 = smooth.spline(Distance,BG1, spar=0.7)
$sS2 = smooth.spline(Distance,BG2, spar=0.7)
$pdf(file="path and file name .pdf", width=3, height=3)
$par(mar=c(1.4,1.6,0.2,0.2))
$plot(Distance,BG1,type="p",col=adjustcolor("brown1",alpha=0.3), xaxt="n",yaxt="n",xlab="",
ylab="",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))
$par(mar=c(1.4,1.6,0.2,0.2), new=TRUE)
$plot(Distance,BG2,type="p",col=adjustcolor("cyan3",alpha=0.3), xaxt="n",yaxt="n",xlab="",
ylab="",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))
$axis(1,padj=-1.2,family="Helvetica",cex.axis=0.8)
$axis(2,padj=0.8,family="Helvetica",cex.axis=0.8)
$lines(sS1, col="color 1", lwd=2)
$lines(sS2, col="color 2", lwd=2)
$dev.off()
```

5.2.5.8 Enhancer set analysis

To investigate the PU.1 binding dynamics around active, cell type-specific enhancers during monocyte differentiation a publically available set of enhancer elements produced by the FANTOM consortium (Andersson *et al.*, 2014) was utilized (see

chapter 5.2.5.9.3). As described in (Andersson *et al.*, 2014) the FANTOM5 CAGE expression atlas (available from Consortium *et al.*, 2014) was used to identify transcribed and thus active enhancer elements in several cell types. Enhancers were identified due to their bidirectional transcription and centered between the bidirectional transcripts.

#To obtain cell type-specific sets of actively used enhancers, we annotated the enhancer set with CAGE enhancer expression data from MO (3 donors) and DC (2 donors, one was left out due to insufficient tag coverage) (available from Consortium *et al.*, 2014).
$annotatePeaks.pl <path to enhancer file> hg19 -size given -len 0 -strand + -d <path to CAGE tag directories>-noadj -noann -nogene > <path to output file>

#Sort annotated file for TPs with a minimum tag count of 1 in Excel.
#Calculation of statistics in R was too stringent. Instead t.test (unpaired, two-sided, <0.05) was used (Excel) to calculate statistics:
#Sorting annotated file in Excel for p<0.05 and separate into two cell type-specific sets.
#Sort for maximum tag count in MO and DC sets and reduce to positions.
#The cell type-specific enhancer files were used for the following analyses.

Figure 3.34 A:

###Illustrating the dynamics of PU.1 and 5hmC in distance distribution heatmaps:
#Annotate cell type-specific enhancer sets with tag counts of PU.1 ChIP-seq and Hydroxymethyl Collector^{TM}-seq samples to retrieve tag count using Homer tools:
$annotatePeaks.pl <path to enhancer file> hg19 -size 207 -d <path to tag directories> > <path to output file>
#Sort for maximum PU.1 tag count in averaged replicate time points (in Excel) and save position data.
#Generating distance distribution heatmaps of PU.1 and 5hmC around MO- and DC-specific enhancers using Homer tools and R:
#Annotate cell type-specific enhancer sets with tag counts of PU.1 ChIP-seq and Hydroxymethyl Collector^{TM}-seq samples using Homer tools:
$annotatePeaks.pl <path to cell type-specific enhancer file> hg19 -ghist -hist 25 -s1000 -d <path to tag directories> > <path to output file>

#Use generated files for R:
#For PU.1 heatmaps:
$library(gplots)
$library(RColorBrewer)
$setwd("path to ghist file")
$data <- read.delim("ghist file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,3,length=100), seq(3,6,length=100))

$png(filename="file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, density.info="none", trace="none", lhei = c(0.05,0.95), lwid = $c(0.05,0.95))
$dev.off()

#For 5hmc heatmaps:
$library(gplots)
$library(RColorBrewer)
$setwd("path to ghist file")
$data <- read.delim("ghist file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,5,length=100), seq(5,10,length=100))
$png(filename="file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, density.info="none", trace="none", lhei = c(0.05,0.95), lwid = $c(0.05,0.95))
$dev.off()

Figure 3.34 A (right panel):

###Investigation of CpG density coverage:
#Annotate cell type-specific enhancer sets with tag counts of PU.1 ChIP-seq and Hydroxymethyl Collector™-seq samples to retrieve tag count using Homer tools:
$annotatePeaks.pl <path to cell type-specific enhancer file> hg19 -size 207 -d <path to tag directories> > <path to output file>
#Sort for maximum PU.1 tag count in averaged replicate time points (in Excel) and save position data.

###Generating distance distribution heatmaps of CpG-content around cell type-specific enhancers using Homer tools and R:
#Use a reference set of all valid CpGs of the genome, containing CpG-locations extended to 100 bp each, in order to create a tag directory file.
#Annotate cell type-specific enhancer sets with the reference CpG tag directory using Homer tools:
$annotatePeaks.pl <path to cell type-specific enhancer file> hg19 -size 1000 -hist 25 -ghist -d <path to CpG tag directory> > <path to output file>

#Use generated files for R:
$setwd("path to annotated file")
$data <- read.delim("annotated file name", row.names="Gene")
$d <- data.matrix(data)
$mycol <- colorRampPalette(c("white","blue"))(199)
$col_breaks = c(seq(0,3,length=100), seq(3,6,length=100))
$png(filename="distance distribution heatmap file name .png", height=24000, width=18000)
$heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks=col_breaks, dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=FALSE, $density.info="none", trace="none", lhei = c(0.05,0.95), lwid = c(0.05,0.95))
$dev.off()

Figure 3.34 B:

###Illustrating the dynamics of PU.1 and 5hmC in histograms:
#Annotate cell type-specific enhancer sets with tag counts of PU.1 ChIP-seq and Hydroxymethyl Collector™-seq samples to retrieve tag count using Homer tools:
$annotatePeaks.pl <path to cell type-specific enhancer file> hg19 -size 207 -d <path to tag directories> > <path to output file>
#Sort for maximum PU.1 tag count in averaged replicate time points (in Excel) and save position data.

##Generating histograms of PU.1 and 5hmC tag distribution around cell type-specific enhancers using Homer tools and R:
#Annotate cell type-specific enhancer sets with tag counts of PU.1 ChIP-seq and Hydroxymethyl Collector™-seq samples using Homer tools:
$annotatePeaks.pl <path to enhancer file> hg19 -size 1000 -hist 25 -d <path to tag directories> > <path to output file>

#Split annotated file into PU.1 or 5hmC coverages for R.
##Generate histograms in R:

```
$setwd("path to file")
$data <- read.table("hmC or PU.1 file name", header=T, sep="\t")
$attach(data)
$plotcolors <- c("color 1", "color 2", "color 3")
$pdf(file="file name .pdf", height=4, width=8)
$par(mar=c(3,3,1,3))
$plot(Distance,TP 1,type="l",col=plotcolors[3],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16), xlim=c(-500,500))
$par(mar=c(3,3,1,3), new=TRUE)
$plot(Distance,TP 2,type="l",col=plotcolors[2],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16), xlim=c(-500,500))
$par(mar=c(3,3,1,3), new=TRUE)
$plot(Distance,TP 3,type="l",col=plotcolors[1],xaxt="n",yaxt="n",xlab="",axes=FALSE, ylab="",lwd=1, lty=1,ylim=c(0,16), xlim=c(-500,500))
$axis(1,padj=-1.0,family="Helvetica",cex.axis=1,at=c(-500,-400,-300,-200,-100,0,100,200,300,400,500))
$axis(2,padj=0.8,family="Helvetica",cex.axis=1, col="black", col.axis="black")
$mtext("name y-axis",family="Helvetica", col="black", side=2,line=2,cex=1.3,padj=0.2)
$mtext("name x-axis",family="Helvetica",ps=12,side=1,line=2,cex=1.3,padj=-0.6)
$dev.off()
```

Figure 3.34 C:

###Plotting heatmaps of 5mC distribution in the vicinity of cell type-specific enhancer regions:
#Use files generated in this chapter and from chapter 5.2.5.9.2.
#Reformat data and plot graph using R:
#Command averages methylation level over all valid CpGs at each position:
$myMethylHist_MODC.pl <path to divided cell type-specific enhancer file> hg19 <path to rescaled MO file> <path to rescaled DC file> -scaled -o <path to output file>

```
$x <- read.table("path to reformatted file", header=T, sep="    ")
$attach(x)
$sS1 = smooth.spline(Distance,BG1, spar=0.7)
$sS2 = smooth.spline(Distance,BG2, spar=0.7)
$pdf(file="path and file name .pdf", width=3, height=3)
$par(mar=c(1.4,1.6,0.2,0.2))
$plot(Distance,BG1,type="p",col=adjustcolor("brown1",alpha=0.3), xaxt="n",yaxt="n",xlab="",
ylab="",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))
$par(mar=c(1.4,1.6,0.2,0.2), new=TRUE)
$plot(Distance,BG2,type="p",col=adjustcolor("cyan3",alpha=0.3), xaxt="n",yaxt="n",xlab="",
ylab="",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))
$axis(1,padj=-1.2,family="Helvetica",cex.axis=0.8)
$axis(2,padj=0.8,family="Helvetica",cex.axis=0.8)
$lines(sS1, col="color 1", lwd=2)
$lines(sS2, col="color 2", lwd=2)
$dev.off()
```

5.2.5.9 Additional data sets (NGS, Microarray)

In addition to NGS data sets generated in this work, other data sets from public sources or from scientific cooperation partners were used to compare with our data. When this is the case, the data source is stated in the command line instructions or referenced to the corresponding publication.

5.2.5.9.1 Whole genome expression data of MO and DC

The expression data set of differentiating MO and DC was previously produced in our laboratory (Klug *et al.*, 2010) and is publicly available from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) repository (accession number [GEO:GSE19236].
In brief, mRNAs of time courses from differentiating monocytes were quantified on Whole Human Genome Expressionarrays (4×44 K Agilent). The normalized (according to Klug 2010) expression data of 7 time points of differentiating monocytes from 3-5 donors were averaged and used in our analyses.

5.2.5.9.2 Bisulfite-sequencing data of MO and DC

MethylC-seq data of DC (Barreiro laboratory, Canada)

Unpublished bisulfite-sequencing data obtained from dendritic cells of 6 donors were kindly provided from Dr. Luis Barreiro (CHU Sainte-Justine, Montreal, Canada). In

brief, peripheral blood monocytes were isolated from healthy blood donors as described in Barreiro *et al.* (Barreiro *et al.*, 2012). The cells were cultivated in RPMI 1640 (Invitrogen) in the presence of 10% FCS (Dutscher), L-glutamine (Invitrogen), GM-CSF (20 ng/ml, Immunotools), and IL-4 (20 ng/ml, Immunotools) to differentiate them into dendritic cells *in vitro*. After approximately 7 days the cells were harvested, and the MethylC-seq protocol (Lister *et al.*, 2009) was performed. DNAs were subjected to library preparation (TruSeq RNA Sample Preparation kit v2, Illumina) and bisulfite-conversion (MethylCode™ Bisulfite Conversion Kit, Invitrogen). The sequencing was performed on an Illumina HiSeq 2000 according to the manufacturer's manual. The sequencing reads were aligned to the human reference genome (GRCh37/hg19) and lambda phage genome (Bismark tools) using a bisulfite-converted reference genome. The data were quality-checked and methylation levels of each CpG were estimated using the number of C (methylated) and T (unmethylated) reads (C/C+T) (Lister *et al.*, 2009).

Bisulfite-seq data for $CD14^{+}CD16^{-}$ classical monocyte samples (BLUEPRINT Epigenome Project)

Within the framework of the BLUEPRINT Epigenome Project Whole Genome Bisulfite Sequencing was applied to several peripheral, $CD14^{+}CD16^{-}$ classical monocyte samples as described on the homepage (http://www.blueprint-epigenome.eu/). BS-seq data sets from 4 donors were downloaded from the homepage for data analysis (http://ftp.ebi.ac.uk/pub/databases/blueprint).

Bisulfite data (MO and DC) from both sources (see above) was processed according to the following command line using the indicated tools:

###Adjust bisulfite-sequencing data sets of MO and DC before comparison with our data using Homer tools:

#Conversion of BigWig to bedGraph format:
$bigWigToBedGraph <path to BigWig file> <path to output bedGraph file>

#Merging MO data files:
$unionBedGraphs -i <path to MO file 1> < path to MO file 2>... -filler NA > <path to output .bedGraph file>
$myCombine3BedGraph.pl <path to previous output .bedGraph file> > <path to output .bedGraph file>

#Rescale data for uniformity:
$myScaleBedGraph.pl <path to merged MO or DC file> -S 1 > <path to output file>

The resulting files (MO and DC) were used for NGS analyses in chapter 5.2.5.

5.2.5.9.3 Sets of active enhancers and promoters and CAGE expression data of MO and DC

For the analyses of PU.1 dynamics in MO and DC we utilized a publically available promoter set, enhancer set and CAGE expression data generated by the FANTOM consortium. The data sets are published in the following publications (Andersson *et al.*, 2014; Consortium *et al.*, 2014) and can be downloaded from the homepage (http://fantom.gsc.riken.jp/data/).
We downloaded a promoter set, an enhancer set, and CAGE expression data of monocytes (3 donors) and immature dendritic cells (3 donors) and processed them according to the instructions in chapters 5.2.5.7 and 5.2.5.8.

5.2.5.9.4 Additional pearl scripts (AG Rehli)

In chapter 5.2.5 several in-house pearl scripts were used to analyze the indicated data sets. These pearl scripts were designed by Prof. Dr. Michael Rehli and are listed on the following pages.

myMap-Bowtie2.pl

```
#!/usr/bin/perl -w
sub printCMD() {
	print STDERR "\n\tUsage: map-bowtie2.pl [options] <FASTQ file1> [FASTQ file2]...\n";
	print STDERR "\n\t\t\tpaired end: <FASTQ end1>,<FASTQ end2> [FASTQ file2 end1],[end2]...\n";
	print STDERR "\n\tRequired Options:\n";
	print STDERR "\t\t-index <path-to-bt2 index> (path to bowtie2 index to use for mapping)\n";
	print STDERR "\t\t-x <path...> (also works same as -index)\n";
	print STDERR "\t\t\ti.e. -index /bioinf/bowtie2/bowtie2-2.0.0-beta6/indexes/hg19\n";
	print STDERR "\n\tAlignment Type:\n";
	print STDERR "\t\t-bowtie2 (map with bowtie2,default)\n";
	print STDERR "\t\t-tophat2 (map with tophat2)\n";
	print STDERR "\t\t\t-path <path-to-program-file> (executable file to run if not in path/diff name)\n";
	print STDERR "\n\tCPU options:\n";
	print STDERR "\t\t-cpu <#> (Number of instances to run at once, default:1)\n";
	print STDERR "\t\t-p <#> (Number of cpus per instance, default: 1)\n";
	print STDERR "\t\t-un (will output unaligned reads)\n";
	print STDERR "\t\tRecommended examples (for 8 cores):\n";
	print STDERR "\t\t\tFor bowtie2: -cpu 1 -p 8 (align each file quickly 1 at a time)\n";
```

```
		print STDERR "\t\t\tFor tophat2: -cpu 8 -p 1 (because tophat has several non-parallel steps...)\n";
		print STDERR "\n\tBowtie options:\n";
		print STDERR "\t\t--local (local alignment, default: global/end-to-end)\n";
		print STDERR "\t\t-bam (convert output file to sorted bam file, default: sam)\n";
		print STDERR "\n\tTophat options:\n";
		print STDERR "\t\t--library-type <type> (library type for tophat, default: fr-firststrand\n";
		print STDERR "\t\t\t\tOther optoins: fr-unstranded, fr-secondstrand)\n";
		print STDERR "\t\t-G <GTF file> (Use as guide for splice mapping)\n";
		#print STDERR "\t\t-mis <#> (Max number of mismatches, default: 2)\n";
		print STDERR "\n\tGeneral Options to pass along to alignment program:\n";
		print STDERR "\t\t-f (Input is FASTA files, default expects FASTQ)\n";
		print STDERR "\t\t-pass \"...\" (need to include quotes)\n";
		#print STDERR "\t\t-remap (remap unaligned reads with bowtie afterwards, returning random position)\n";
		print STDERR "\t\t-name (path/outputfile name w/o extension)\n";
		print STDERR "\n";
		exit;
}
if (@ARGV < 2) {
		printCMD();
}
my $bt2Index = "";
my $maxCPUs = 1;
my $pCPUs = 1;
my $remapFlag = 0;
my @files = ();
my $libraryType = "fr-firststrand";
my $maxMultihits = 20;
my $guideGTF = "";
my $program = 'bowtie2';
my $exe = '';
my $local = '';
my $pass = "";
my $unFlag = 0;
my $n = 3;
my $maxMisMatches = " -n $n --genome-read-mismatches $n --read-mismatches $n ";
my $bamFlag = 0;
my $outputFile = '';
my $outputFileName = '';
my $bamFile = '';
my $sortedBamFile = '';

for (my $i=0;$i<@ARGV;$i++) {
		#print STDERR "$ARGV[$i] $i\n";
		if ($ARGV[$i] eq '-index' || $ARGV[$i] eq '-x') {
```

```
                $bt2Index = $ARGV[++$i];
        } elsif ($ARGV[$i] eq '-bam') {
                $bamFlag= 1;
        } elsif ($ARGV[$i] eq '-mis') {
                my $n = $ARGV[++$i];
                $maxMisMatches = " -n $n --genome-read-mismatches $n --read-mismatches $n ";
        } elsif ($ARGV[$i] eq '-bowtie2') {
                $program = 'bowtie2';
        } elsif ($ARGV[$i] eq '-tophat2') {
                $program = 'tophat2';
        } elsif ($ARGV[$i] eq '-f') {
                $pass .= " -f ";
        } elsif ($ARGV[$i] eq '--local') {
                $local = '--local';
        } elsif ($ARGV[$i] eq '-path') {
                $exe = $ARGV[++$i];
        } elsif ($ARGV[$i] eq '-cpu') {
                $maxCPUs = $ARGV[++$i];
        } elsif ($ARGV[$i] eq '-name') {
                $outputFileName = $ARGV[++$i];
        } elsif ($ARGV[$i] eq '-un') {
                $unFlag = 1;
        } elsif ($ARGV[$i] eq '-p') {
                $pCPUs = $ARGV[++$i];
        } elsif ($ARGV[$i] eq '-remap') {
                $remapFlag = 1;
        } elsif ($ARGV[$i] eq '--library-type') {
                $libraryType = $ARGV[++$i];
        } elsif ($ARGV[$i] eq '-pass') {
                $pass = $ARGV[++$i];
        } elsif ($ARGV[$i] eq '-G') {
                $guideGTF = " -G $ARGV[++$i]";
        } elsif ($ARGV[$i] =~ /^\-/) {
                print STDERR "\n!!! \"$ARGV[$i]\" not recognized...\n";
                printCMD();
        } else {
                push(@files, $ARGV[$i]);
        }
}
#$program .= " -f ";
if ($exe eq '') {
        $exe = $program;
        print STDERR "\tWill run $exe...\n";
}
if ($bt2Index eq '') {
```

```
		print STDERR "!!! Need to specify bowtie2 index and/or CPUs !!!\n";
		printCMD();
}
my $genomeName = $bt2Index;
$genomeName =~ s/^.+\///;
print STDERR "\tbt2Index = $bt2Index ($genomeName)\n";
print STDERR "\tNumber of instances at once: $maxCPUs\n";
print STDERR "\tNumber of cpus per instance: $pCPUs\n";
if ($program eq 'bowtie2') {
		print STDERR "\n\tbowtie2 will be used to align the following files:\n";
} elsif ($program eq 'tophat2') {
		print STDERR "\tlibrary type: $libraryType\n";
		print STDERR "\n\tTophat2 will be used to align the following files:\n";
		$maxMisMatches = '';
}
foreach(@files) {
		print STDERR "\t\t$_\n";
}
print STDERR "\t\n";
my @pids = ();
my $cpus = 0;
for (my $j=0;$j<@files;$j++) {
		my $file = $files[$j];
		my $pid = fork();
		$cpus++;
		if ($pid == 0) {
			#child process
			my $peflag = 0;
			my %delete = ();
			my $file1 = "";
			my $file2 = "";
			if ($file =~ /\,/) {
				my @f = split /\,/, $file;
				$file1 = $f[0];
				$file2 = $f[1];
				print STDERR "\t\tPaired end files: $file1 and $file2\n";
				$peflag = 1;
				if ($file1 =~ /\.gz$/) {
					my $newfile = $file1;
					$newfile  =~ s/\.gz$//;
					`gunzip -c "$file1" > "$newfile"`;
					$delete{$newfile} = 1;
					$file1 = $newfile;
				}
				if ($file2 =~ /\.gz$/) {
```

```
                        my $newfile = $file2;
                        $newfile  =~ s/\.gz$//;
                        `gunzip -c "$file2" > "$newfile"`;
                        $delete{$newfile} = 1;
                        $file2 = $newfile;
                }
                if ($program eq 'tophat2') {
                        $file = "\"$file1\" \"$file2\"";
                } elsif ($program eq 'bowtie2') {
                        $file = "-1 \"$file1\" -2 \"$file2\"";
                }
        } else {
                if ($file =~ /\.gz$/) {
                        my $newfile = $file;
                        $newfile  =~ s/\.gz$//;
                        `gunzip -c "$file" > $newfile`;
                        $delete{$newfile} = 1;
                        $file = $newfile;
                }
                $file1 = $file;
                $file = "\"$file\"";
        }
        if ($program eq 'tophat2') {
                my $outputDir = "$file1.$genomeName.tophat2";
                my $outputFile = "$file1.$genomeName.tophat2.bam";
                my $outputJunc = "$file1.$genomeName.tophat2.junc";
                my $logFile = "$file1.$genomeName.tophat2.log";
                `$exe --library-type $libraryType -p $pCPUs -g $maxMultihits $guideGTF -o
$outputDir $pass $maxMisMatches "$bt2Index" $file 2> $logFile`;
                `mv "$outputDir/accepted_hits.bam" "$outputFile"`;
        `parseTophatJunctions.pl "$outputDir/junctions.bed" > "$outputJunc"`;
                `rm -r "$outputDir"`;
        if ($unFlag) {
                my $samFile = "$file1.$genomeName.tophat2.sam";
                `samtools view -h "$outputFile" > "$samFile" 2>> $logFile`;
                my $unalignedFile = "$file1.$genomeName.tophat2.unaligned.fq";
                my $unalignedFile2 = "$file2.$genomeName.tophat2.unaligned.fq";
                `getUnalignedReadsSam.pl $samFile $file1 > $unalignedFile 2>> $logFile`;
                        if ($peflag) {
                `getUnalignedReadsSam.pl $samFile $file2 > $unalignedFile2 2>> $logFile`;
                        }
                        `rm "$samFile"`;
                }
        } else {
                my $logFile = "$file1.$genomeName.bowtie2.log";
```

```
                        my $unFile = " --no-unal";
                        if ($outputFileName eq '') {
                                $outputFile = "$file1.$genomeName.bowtie2.sam";
                        } else {
                                $outputFile = "$outputFileName.sam";
                        }
                        if ($unFlag) {
                                $unFile = "--un $file1.$genomeName.bowtie2.unaligned.fq";
                        }
                `$exe $local $unFile -p $pCPUs $pass -x "$bt2Index" $file > "$outputFile" 2> "$logFile"`;
                        if ($bamFlag) {
                                if ($outputFileName eq '') {
                                        $bamFile = "$file1.$genomeName.bowtie2.tmp.bam";
                                        $sortedBamFile = "$file1.$genomeName.bowtie2";
                                } else {
                                        $bamFile = "$outputFileName.tmp.bam";
                                        $sortedBamFile = "$outputFileName.bowtie2";
                                }
                `samtools view -S -b -@ $pCPUs "$outputFile" > "$bamFile" 2>> "$logFile"`;
                `samtools sort -m 5000000000 -@ $pCPUs "$bamFile" "$sortedBamFile" 2>> $logFile`;
                                `rm "$outputFile" "$bamFile"`;
                        }
                }
                foreach(keys %delete) {
                        `rm "$_"`;
                }
                exit(0);
        }
        push(@pids, $pid);
        if ($cpus >= $maxCPUs) {
                my $id = wait();
                $cpus--;
        }
}
my $id = 0;
while ($id >= 0) {
        $id = wait();
        if ($id == -1) {
        } else {
        }
}
print STDERR "\n";
```

myMethylHist_MODC.pl

```
#my $directory = "path to directory";

if (@ARGV < 1) {
	print STDERR "<input file> <genome> <MO DNA methylation bedGraph file> <DC DNA methylation bedGraph file> <options>  -o <outputfile>\n";
	print STDERR "\toptions:\n";
	print STDERR "\t\t-size <#> histogram size (default 1000)\n";
	print STDERR "\t\t-scaled <#> bedgraph is already transformed\n";
	exit;
}
my $size = 1002;
my $scaled = 0;
for (my $i=1;$i<@ARGV;$i++) {
	if ($ARGV[$i] eq '-size') {
		$size = $ARGV[++$i] + 2;
		}
		elsif ($ARGV[$i] eq '-scaled') {
		$scaled = 1;
		}
		elsif ($ARGV[$i] eq '-o') {
		$outputfile = $ARGV[++$i];
		}
}
my $posFile = $ARGV[0];
my $genome = $ARGV[1];
my $bedGraph = $ARGV[2];
my $bedGraph2 = $ARGV[3];
my $uppersize = $size / 2;
my $lowersize = $size / -2;
my $shift = 1;
my $sum = 0;
my $count = 0;
$rand = rand();
my $tmpFile = "$directory/" . $rand . ".bedGraph.tmp";
my $tmpFile1 = "$directory/" . $rand . ".hist.tmp";
my $tmpFile2 = "$directory/" . $rand . ".TP.tmp";
my $tmpFile3 = "$directory/" . $rand . ".out.tmp";
my $tmpScript = $rand . ".R";
my $tmpOut1 = "$directory/" . $rand . ".out1.tmp";
my $tmpOut2 = "$directory/" . $rand . ".out2.tmp";

	print STDERR "\tprocessing first sample\n";
	if ($scaled == 1) {
	print STDERR "\t...using scaled bedGraph file1\n";
```

```
		$tmpFile = $bedGraph
		} else {
		print STDERR "\t...converting bedGraph file\n";
		`myScaleBedGraph.pl "$bedGraph" -S 1 >"$tmpFile"`;
		}
		print STDERR "\t...creating ghist file\n";
`annotatePeaks.pl "$posFile" "$genome" -size "$size" -hist 1 -ghist -bedGraph "$tmpFile" >"$tmpFile1"`;

		print STDERR "\t...transposing ghist file\n";
		`myTP.pl "$tmpFile1" > "$tmpFile2"`;

#open OUT, ">$outputfile";
open OUT, ">$tmpOut1";
open IN, $tmpFile2 ;

while (<IN>) {
		chomp;
		s/\r//g;
		my @line= split /\t/;
		if ($line[0] eq 'Gene') {
		print STDERR "\t...printing table\n";
		print OUT "Distance\tBG1\n";
		} elsif ($line[0] eq $uppersize || $line[0] eq $lowersize) {
		} else {
		print OUT "$line[0]";
		$columns = @line ;
for (my $i=1;$i<$columns;$i++) {
		if ($line[$i] == 0) {
		}
		else {
		$sum = $sum + $line[$i];
		$count = $count +1;
		}
}
		if ($count == 0) {
		print OUT "\t\n";
		}
		else {
		$mean = $sum / $count - 1;
		print OUT "\t$mean\n";
		}
}
$sum = 0;
$count = 0;
}
```

```
close IN;
close OUT;
		print STDERR "\tprocessing second sample\n";
		if ($scaled == 1) {
		print STDERR "\t...using scaled bedGraph file2\n";
		$tmpFile = $bedGraph2
		} else {
		print STDERR "\t...converting bedGraph file\n";
		`myScaleBedGraph.pl "$bedGraph2" -S 1 >"$tmpFile"`;
		}
		print STDERR "\t...creating ghist file\n";
`annotatePeaks.pl "$posFile" "$genome" -size "$size" -hist 1 -ghist -bedGraph "$tmpFile" >"$tmpFile1"`;
		print STDERR "\t...transposing ghist file\n";
		`myTP.pl "$tmpFile1" > "$tmpFile2"`;

open OUT2, ">$tmpOut2";
open IN2, $tmpFile2 ;
while (<IN2>) {
		chomp;
		s/\r//g;
		my @line= split /\t/;
		if ($line[0] eq 'Gene') {
		print STDERR "\t...printing table\n";
		print OUT2 "Distance\tBG2\n";
		} elsif ($line[0] eq $uppersize || $line[0] eq $lowersize) {
		} else {
		print OUT2 "$line[0]";
		$columns = @line ;

for (my $i=1;$i<$columns;$i++) {
		if ($line[$i] == 0) {
		}
		else {
		$sum = $sum + $line[$i];
		$count = $count +1;
		}
}
		if ($count == 0) {
		print OUT2 "\t\n";
		}
		else {
		$mean = $sum / $count - 1;
		print OUT2 "\t$mean\n";
		}
}
```

```
$sum = 0;
$count = 0;
}
close IN2;
close OUT2;
	`paste \"$tmpOut1\" \"$tmpOut2\" | cut -f1,2,4 > $outputfile`;

# `rm -f \"$tmpFile1\" \"$tmpFile2\"`;
	if ($scaled == 0) {
	`rm -f \"$tmpFile\"`;
	}
	print STDERR "\t...printing pdf\n";

open OUT, ">$tmpScript";
print OUT "x <- read.table(\"$outputfile\", header=T, sep=\"\t\")\n";
print OUT "attach(x)\n";
print OUT "sS1 = smooth.spline(Distance,BG1, spar=0.5)\n";
print OUT "sS2 = smooth.spline(Distance,BG2, spar=0.5)\n";
print OUT "pdf(file=\"$outputfile.pdf\", width=3, height=3)\n";
print OUT "par(mar=c(1.4,1.6,0.2,0.2))\n";
print OUT "plot(Distance,BG1,type=\"p\",col=adjustcolor(\"brown1\",
alpha=0.3),xaxt=\"n\",yaxt=\"n\",xlab=\"\", ylab=\"\",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))\n";
print OUT "par(mar=c(1.4,1.6,0.2,0.2), new=TRUE)\n";
print OUT "plot(Distance,BG2,type=\"p\",col=adjustcolor(\"cyan3\",
alpha=0.3),xaxt=\"n\",yaxt=\"n\",xlab=\"\", ylab=\"\",pch=46,axes=FALSE,lwd=3,ylim=c(0,100))\n";
print OUT "axis(1,padj=-1.2,family=\"Helvetica\",cex.axis=0.8)\n";
print OUT "axis(2,padj=0.8,family=\"Helvetica\",cex.axis=0.8)\n";
print OUT "lines(sS1, col=\"color 1\", lwd=2)\n";
print OUT "lines(sS2, col=\"color 2\", lwd=2)\n";
print OUT "dev.off()\n";
close OUT;
`R --no-save < "$tmpScript"`;
#`rm -f \"$tmpScript\" \"$tmpFile\" \"$tmpFile2\" \"$tmpOut1\" \"$tmpOut2\"`;
exit;
```

---

myCombine3BedGraph.pl

```
if (@ARGV < 1) {
	print STDERR "<input union.bedGraph file> <options> <outputfile>\n";
	exit;
}
open IN, $ARGV[0] or die "Could not open file $ARGV[0]\n";
my $count = 0;
my $value = 0;
```

```
my $adjustedcount = 0;
while (<IN>) {
	chomp;
	s/\r//g;
	my @line= split /\t/;

	for (my $i=3;$i<@line;$i++) {
		if ($line[$i] eq 'NA') {
		} else {
		$count++;
		$value = $value + $line[$i];
		}
	}
	$adjustedcount = int( $value *1000 / $count) / 10 ;
	print $line[0];
	print "\t$line[1]\t$line[2]\t$adjustedcount\n";
	$count = 0;
	$value = 0;
}
close IN;
exit;
```

---

myScaleBedGraph.pl

```
if (@ARGV < 1) {
	print STDERR "<input bedGraph file> <options> <outputfile>\n";
	print STDERR "\toptions:\n";
	print STDERR "\t\t-S <#> add factor\n";
	exit;
}
open IN, $ARGV[0] or die "Could not open file $ARGV[0]\n";
my $scale = 1;
for (my $i=1;$i<@ARGV;$i++) {
	if ($ARGV[$i] eq '-S') {
		$scale = $ARGV[++$i];
	}
}
while (<IN>) {
	chomp;
	s/\r//g;
	my @line= split /\t/;
	$adjustedcount = $line[3]+$scale;
	print $line[0];
```

```
        print "\t$line[1]\t$line[2]\t$adjustedcount\n";
}
close IN;
exit;
```

5.2.6 Software tools and databases

Software/Database	Version	Source/Company
Adobe Illustrator	CS3	Adobe Systems, Mountain View, USA
BEDtools	v2.20.1 or later	http://bedtools.readthedocs.org/en/latest/
BioEdit	v6.4.5	http://www.mbio.ncsu.edu/BioEdit/bioedit.html
BioGPS	-	http://www.ncbi.nlm.nih.gov/geo/
Bowtie2	v2.2.1	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
ControlMate	v1.4.0	Life Technologies, Carlsbad, USA
EpiDesigner	-	http://www.epidesigner.com/
Epityper	v1.2	Sequenom, Hamburg, Germany
GeneRunner	v3.0.5	http://www.generunner.com
Homer	v4.5 or later	http://homer.salk.edu/homer/
ImageLab	v4.0.1	Bio-Rad, Munich, Germany
MethPrimer	-	http://www.urogene.org/methprimer
Microsoft Excel	2010	Microsoft, Redmond, USA
PerlPrimer	v1.1.19	http://perlprimer.sourceforge.net/
PubMed	-	www.ncbi.nlm.nih.gov/entrez
R	3.0.2	http://www.r-project.org/
USCS Genome Browser	-	www.genome.ucsc.edu

6 Publications

Neumeier, M., Krautbauer, S., **Schmidhofer, S.**, Hader, Y., Eisinger, K., Eggenhofer, E., Froehner, S.C., Adams, M.E., Mages, W., and Buechler, C. (2013). Adiponectin receptor 1 C-terminus interacts with PDZ-domain proteins such as syntrophins. Experimental and molecular pathology *95*, 180-186.

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

Pham, T.H., Minderjahn, J., Schmidl, C., Hoffmeister, H., **Schmidhofer, S.**, Chen, W., Langst, G., Benner, C., and Rehli, M. (2013). Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. Nucleic acids research *41*, 6391-6402.

Bauer, S., Wanninger, J., **Schmidhofer, S.**, Weigert, J., Neumeier, M., Dorn, C., Hellerbrand, C., Zimara, N., Schaffler, A., Aslanidis, C., et al. (2011). Sterol regulatory element-binding protein 2 (SREBP2) activation after excess triglyceride storage induces chemerin in hypertrophic adipocytes. Endocrinology *152*, 26-35.

Wiest, R., Weigert, J., Wanninger, J., Neumeier, M., Bauer, S., **Schmidhofer, S.**, Farkas, S., Scherer, M.N., Schaffler, A., Scholmerich, J., et al. (2011). Impaired hepatic removal of interleukin-6 in patients with liver cirrhosis. Cytokine *53*, 178-183.

*Equal contributors

7 References

Abbas, A.K., Lichtman, A.H., and Pillai, S. (2007). Cellular and molecular immunology, 7th ed., Elsevier/Saunders.

Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R.*, et al.* (2009). Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood *114*, 144-147.

Abrink, M., Gobl, A.E., Huang, R., Nilsson, K., and Hellman, L. (1994). Human cell lines U-937, THP-1 and Mono Mac 6 represent relatively immature cells of the monocyte-macrophage cell lineage. Leukemia *8*, 1579-1584.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T.*, et al.* (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-461.

Andreesen, R., Picht, J., and Lohr, G.W. (1983). Primary cultures of human blood-born macrophages grown on hydrophobic teflon membranes. Journal of immunological methods *56*, 295-304.

Arinobu, Y., Mizuno, S., Chong, Y., Shigematsu, H., Iino, T., Iwasaki, H., Graf, T., Mayfield, R., Chan, S., Kastner, P.*, et al.* (2007). Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. Cell stem cell *1*, 416-427.

Ausubel, F.M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J.A., Struhl, K. (1987). Current Protocols in Molecular Biology. John Wiley & Sons.

Back, J., Allman, D., Chan, S., and Kastner, P. (2005). Visualizing PU.1 activity during hematopoiesis. Experimental hematology *33*, 395-402.

Badeaux, A.I., and Shi, Y. (2013). Emerging roles for chromatin as a signal integration and storage platform. Nature reviews. Molecular cell biology *14*, 211-224.

Bai, Y., Srinivasan, L., Perkins, L., and Atchison, M.L. (2005). Protein Acetylation Regulates Both PU.1 Transactivation and Ig 3' Enhancer Activity. The Journal of Immunology *175*, 5160-5169.

Balasubramani, A., and Rao, A. (2013). O-GlcNAcylation and 5-methylcytosine oxidation: an unexpected association between OGT and TETs. Molecular cell *49*, 618-619.

Banchereau, J., and Steinman, R.M. (1998). Dendritic cells and the control of immunity. Nature *392*, 245-252.

Barreiro, L.B., Tailleux, L., Pai, A.A., Gicquel, B., Marioni, J.C., and 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.

Baubec, T., Ivanek, R., Lienert, F., and Schubeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell *153*, 480-492.

Bauer, M., Goldstein, M., Christmann, M., Becker, H., Heylmann, D., and Kaina, B. (2011). Human monocytes are severely impaired in base and DNA double-strand break repair that renders them vulnerable to oxidative stress. Proceedings of the National Academy of Sciences of the United States of America *108*, 21105-21110.

Baylin, S.B., and Jones, P.A. (2011). A decade of exploring the cancer epigenome - biological and translational implications. Nature reviews. Cancer *11*, 726-734.

Bennett, M.T., Rodgers, M.T., Hebert, A.S., Ruslander, L.E., Eisele, L., and Drohat, A.C. (2006). Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. Journal of the American Chemical Society *128*, 12510-12519.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042-1047.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & development *16*, 6-21.

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science *336*, 934-937.

Borggrefe, T., and Yue, X. (2011). Interactions between subunits of the Mediator complex with gene-specific transcription factors. Seminars in cell & developmental biology *22*, 759-768.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

Branco, M.R., Ficz, G., and Reik, W. (2012). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nature reviews. Genetics *13*, 7-13.

Brosh, R.M., Jr. (2013). DNA helicases involved in DNA repair and their roles in cancer. Nature reviews. Cancer *13*, 542-558.

Bruniquel, D., and Schwartz, R.H. (2003). Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nature immunology *4*, 235-240.

Buecker, C., and Wysocka, J. (2012). Enhancers as information integration hubs in development: lessons from genomics. Trends in genetics : TIG *28*, 276-284.

Burda, P., Laslo, P., and Stopka, T. (2010). The role of PU.1 and GATA-1 transcription factors during normal and leukemogenic hematopoiesis. Leukemia *24*, 1249-1257.

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

Campanero, M.R., Armstrong, M.I., and Flemington, E.K. (2000). CpG methylation as a mechanism for the regulation of E2F activity. Proceedings of the National Academy of Sciences of the United States of America *97*, 6481-6486.

Carotta, S., Dakic, A., D'Amico, A., Pang, S.H., Greig, K.T., Nutt, S.L., and Wu, L. (2010a). The transcription factor PU.1 controls dendritic cell development and Flt3 cytokine receptor expression in a dose-dependent manner. Immunity *32*, 628-641.

Carotta, S., Wu, L., and Nutt, S.L. (2010b). Surprising new roles for PU.1 in the adaptive immune response. Immunological reviews *238*, 63-75.

Chen, C.C., Wang, K.Y., and Shen, C.K. (2012). The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. The Journal of biological chemistry *287*, 33116-33121.

Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013). TET2 promotes histone O-GlcNAcylation during gene transcription. Nature *493*, 561-564.

Chen, Z.X., and Riggs, A.D. (2011). DNA methylation and demethylation in mammals. The Journal of biological chemistry *286*, 18347-18353.

Cheung, D.L., and Hamilton, J.A. (1992). Regulation of human monocyte DNA synthesis by colony-stimulating factors, cytokines, and cyclic adenosine monophosphate. Blood *79*, 1972-1981.

Cheung, W.D., Sakabe, K., Housley, M.P., Dias, W.B., and Hart, G.W. (2008). O-linked beta-N-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins. The Journal of biological chemistry *283*, 33935-33941.

Clanchy, F.I., Holloway, A.C., Lari, R., Cameron, P.U., and Hamilton, J.A. (2006). Detection and properties of the human proliferative monocyte subpopulation. Journal of leukocyte biology *79*, 757-766.

Consortium, F., the, R.P., Clst, Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Lassmann, T., Itoh, M.*, et al.* (2014). A promoter-level mammalian expression atlas. Nature *507*, 462-470.

Cortazar, D., Kunz, C., Saito, Y., Steinacher, R., and Schar, P. (2007). The enigmatic thymine DNA glycosylase. DNA repair *6*, 489-504.

Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F.*, et al.* (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419-423.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D.*, et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67-79.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A.*, et al.* (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.

Crofton, R.W., Diesselhoff-den Dulk, M.M., and van Furth, R. (1978). The origin, kinetics, and characteristics of the Kupffer cells in the normal steady state. The Journal of experimental medicine *148*, 1-17.

D'Alessio, J.A., Wright, K.J., and Tjian, R. (2009). Shifting players and paradigms in cell-specific transcription. Molecular cell *36*, 924-931.

De Kleer, I., Willems, F., Lambrecht, B., and Goriely, S. (2014). Ontogeny of myeloid cells. Frontiers in immunology *5*, 423.

de la Rica, L., Rodriguez-Ubreva, J., Garcia, M., Islam, A.B., Urquiza, J.M., Hernando, H., Christensen, J., Helin, K., Gomez-Vaquero, C., and 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 Smedt, T., Pajak, B., Muraille, E., Lespagnard, L., Heinen, E., De Baetselier, P., Urbain, J., Leo, O., and Moser, M. (1996). Regulation of dendritic cell numbers and maturation by lipopolysaccharide in vivo. The Journal of experimental medicine *184*, 1413-1424.

Delatte, B., Deplus, R., and Fuks, F. (2014). Playing TETris with DNA modifications. The EMBO journal *33*, 1198-1211.

Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Masse, A., Kosmider, O., Le Couedic, J.P., Robert, F., Alberdi, A.*, et al.* (2009). Mutation in TET2 in myeloid cancers. The New England journal of medicine *360*, 2289-2301.

Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Mendez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E.*, et al.* (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. The EMBO journal *32*, 645-655.

Dulac, C. (2010). Brain function and chromatin plasticity. Nature *465*, 728-735.

Ehrich, M., Nelson, M.R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G., Cantor, C.R., Field, J.K., and van den Boom, D. (2005). Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America *102*, 15785-15790.

Euba, B., Vizmanos, J.L., Garcia-Granero, M., Aranaz, P., Hurtado, C., Migueliz, I., Novo, F.J., and Garcia-Delgado, M. (2012). A meta-analysis of TET2 mutations shows a distinct distribution pattern in de novo acute myeloid leukemia and chronic myelomonocytic leukemia. Leukemia & lymphoma *53*, 1230-1233.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398-402.

Finnin, M., Hamilton, J.A., and Moss, S.T. (1999). Direct comparison of the effects of CSF-1 (M-CSF) and GM-CSF on human monocyte DNA synthesis and CSF receptor expression. Journal of interferon & cytokine research : the official journal of the International Society for Interferon and Cytokine Research *19*, 417-423.

Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenoder, S., Wang, M., Qin, W., Soding, J., Spada, F.*, et al.* (2011). Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. PloS one *6*, e16627.

Friedman, A.D. (2007). Transcriptional control of granulocyte and monocyte development. Oncogene *26*, 6816-6828.

Fujiki, R., Chikanishi, T., Hashiba, W., Ito, H., Takada, I., Roeder, R.G., Kitagawa, H., and Kato, S. (2009). GlcNAcylation of a histone methyltransferase in retinoic-acid-induced granulopoiesis. Nature *459*, 455-459.

Geissmann, F., Jung, S., and Littman, D.R. (2003). Blood monocytes consist of two principal subsets with distinct migratory properties. Immunity *19*, 71-82.

Geissmann, F., Manz, M.G., Jung, S., Sieweke, M.H., Merad, M., and Ley, K. (2010). Development of monocytes, macrophages, and dendritic cells. Science *327*, 656-661.

Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.L.*, et al.* (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity *32*, 317-328.

Gilroy, D.W., Lawrence, T., Perretti, M., and Rossi, A.G. (2004). Inflammatory resolution: new opportunities for drug discovery. Nature reviews. Drug discovery *3*, 401-416.

Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., Mehler, M.F., Conway, S.J., Ng, L.G., Stanley, E.R.*, et al.* (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. Science *330*, 841-845.

Ginhoux, F., and Jung, S. (2014). Monocytes and macrophages: developmental pathways and tissue homeostasis. Nature reviews. Immunology *14*, 392-404.

Ginhoux, F., Liu, K., Helft, J., Bogunovic, M., Greter, M., Hashimoto, D., Price, J., Yin, N., Bromberg, J., Lira, S.A.*, et al.* (2009). The origin and development of nonlymphoid tissue CD103+ DCs. The Journal of experimental medicine *206*, 3115-3130.

Goldszmid, R.S., Caspar, P., Rivollier, A., White, S., Dzutsev, A., Hieny, S., Kelsall, B., Trinchieri, G., and Sher, A. (2012). NK cell-derived interferon-gamma orchestrates cellular dynamics and the differentiation of monocytes into dendritic cells at the site of infection. Immunity *36*, 1047-1059.

Gordon, S. (2002). Pattern recognition receptors: doubling up for the innate immune response. Cell *111*, 927-930.

Gordon, S. (2007). The macrophage: past, present and future. European journal of immunology *37 Suppl 1*, S9-17.

Guilliams, M., De Kleer, I., Henri, S., Post, S., Vanhoutte, L., De Prijck, S., Deswarte, K., Malissen, B., Hammad, H., and Lambrecht, B.N. (2013). Alveolar macrophages develop from fetal monocytes that differentiate into long-lived cells in the first week of life via GM-CSF. The Journal of experimental medicine *210*, 1977-1992.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.P., Hu, B.*, et al.* (2014). Active and passive demethylation of male and female pronuclear DNA in the Mammalian zygote. Cell stem cell *15*, 447-458.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423-434.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome research *23*, 1210-1223.

Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.G., Jiang, Y., Pfeifer, G.P.*, et al.* (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. Cell reports *3*, 291-300.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic acids research *40*, 4841-4849.

Hashimoto, H., Pais, J.E., Zhang, X., Saleh, L., Fu, Z.Q., Dai, N., Correa, I.R., Jr., Zheng, Y., and Cheng, X. (2014). Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. Nature *506*, 391-395.

He, H.H., Meyer, C.A., Shin, H., Bailey, S.T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M.*, et al.* (2010). Nucleosome dynamics define transcriptional enhancers. Nature genetics *42*, 343-347.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303-1307.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W.*, et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108-112.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (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, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (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., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Molecular and cellular biology *18*, 6538-6547.

Hohaus, S., Petrovick, M.S., Voso, M.T., Sun, Z., Zhang, D.E., and Tenen, D.G. (1995). PU.1 (Spi-1) and C/EBP alpha regulate expression of the granulocyte-macrophage colony-stimulating factor receptor alpha gene. Molecular and cellular biology *15*, 5830-5845.

Hu, G., Cui, K., Northrup, D., Liu, C., Wang, C., Tang, Q., Ge, K., Levens, D., Crane-Robinson, C., and Zhao, K. (2013a). H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. Cell stem cell *12*, 180-192.

Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y.G., Zhu, J., Wang, P., and Xu, Y. (2013b). Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. Cell *155*, 1545-1555.

Huang, Y., and Rao, A. (2014). Connections between TET proteins and aberrant DNA modification in cancer. Trends in genetics : TIG *30*, 464-474.

Hume, D.A. (2006). The mononuclear phagocyte system. Curr Opin Immunol *18*, 49-53.

Hume, D.A. (2008). Macrophages as APC and the dendritic cell myth. Journal of immunology *181*, 5829-5835.

Iijima, N., Linehan, M.M., Saeland, S., and Iwasaki, A. (2007). Vaginal epithelial dendritic cells renew from bone marrow precursors. Proceedings of the National Academy of Sciences of the United States of America *104*, 19061-19066.

Imoto, A., Okada, M., Okazaki, T., Kitasato, H., Harigae, H., and Takahashi, S. (2010). Metallothionein-1 isoforms and vimentin are direct PU.1 downstream target genes in leukemia cells. The Journal of biological chemistry *285*, 10300-10309.

Ingersoll, M.A., Platt, A.M., Potteaux, S., and Randolph, G.J. (2011). Monocyte trafficking in acute and chronic inflammation. Trends in immunology *32*, 470-477.

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

Iqbal, K., Jin, S.G., Pfeifer, G.P., and Szabo, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proceedings of the National Academy of Sciences of the United States of America *108*, 3642-3647.

Ito, R., Katsura, S., Shimada, H., Tsuchiya, H., Hada, M., Okumura, T., Sugawara, A., and Yokoyama, A. (2014). TET3-OGT interaction increases the stability and the presence of OGT in chromatin. Genes to cells : devoted to molecular & cellular mechanisms *19*, 52-65.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and 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, E.A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S., and Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. Genome biology *14*, R119.

Iyer, L.M., Abhiman, S., and Aravind, L. (2011). Natural history of eukaryotic DNA methylation systems. Progress in molecular biology and translational science *101*, 25-104.

Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. Cell Cycle *8*, 1698-1710.

Jenkins, S.J., and Hume, D.A. (2014). Homeostasis in the mononuclear phagocyte system. Trends in immunology *35*, 358-367.

Jin, C., and Felsenfeld, G. (2007). Nucleosome stability mediated by histone variants H3.3 and H2A.Z. Genes & development *21*, 1519-1529.

Jin, S.G., Jiang, Y., Qiu, R., Rauch, T.A., Wang, Y., Schackert, G., Krex, D., Lu, Q., and Pfeifer, G.P. (2011). 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. Cancer research *71*, 7360-7365.

Jin, S.G., Kadam, S., and Pfeifer, G.P. (2010). Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic acids research *38*, e125.

Johnson, W.D., Jr., Mei, B., and Cohn, Z.A. (1977). The separation, long-term cultivation, and maturation of the human monocyte. The Journal of experimental medicine *146*, 1613-1626.

Jones, P.A., and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. Nature reviews. Genetics *3*, 415-428.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. Cell *128*, 683-692.

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F., and Reid, G. (2008). Transient cyclical methylation of promoter DNA. Nature *452*, 112-115.

Kihara-Negishi, F., Yamamoto, H., Suzuki, M., Yamada, T., Sakurai, T., Tamura, T., and Oikawa, T. (2001). In vivo complex formation of PU.1 with HDAC1 associated with PU.1-mediated transcriptional repression. Oncogene *20*, 6039-6047.

Klug, M. (2010). Dynamics of DNA methylation in differentiating hematopoietic cells.

Klug, M., Heinz, S., Gebhard, C., Schwarzfischer, L., Krause, S.W., Andreesen, R., and 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., and Rehli, M. (2006). Functional analysis of promoter CpG methylation using a CpG-free luciferase reporter vector. Epigenetics : official journal of the DNA Methylation Society *1*, 127-130.

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

Ko, M., An, J., Bandukwala, H.S., Chavez, L., Aijo, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lahdesmaki, H.*, et al.* (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature *497*, 122-126.

Ko, M., Bandukwala, H.S., An, J., Lamperti, E.D., Thompson, E.C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S.B., and Rao, A. (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proceedings of the National Academy of Sciences of the United States of America *108*, 14566-14571.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R.*, et al.* (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839-843.

Kodandapani, R., Pio, F., Ni, C.Z., Piccialli, G., Klemsz, M., McKercher, S., Maki, R.A., and Ely, K.R. (1996). A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. Nature *380*, 456-460.

Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. Nature *502*, 472-479.

Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. Molecular and cellular biology *25*, 4388-4396.

Konstandin, N., Bultmann, S., Szwagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., Herold, T., Mulaw, M., Kakadia, P.M., Schneider, S.*, et al.* (2011). Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. Leukemia *25*, 1649-1652.

Kreppel, L.K., Blomberg, M.A., and Hart, G.W. (1997). Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. The Journal of biological chemistry *272*, 9308-9315.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929-930.

Kwon, O.H., Lee, C.K., Lee, Y.I., Paik, S.G., and Lee, H.J. (2005). The hematopoietic transcription factor PU.1 regulates RANK gene expression in myeloid progenitors. Biochemical and biophysical research communications *335*, 437-446.

Lai, F., Orom, U.A., Cesaroni, M., Beringer, M., Taatjes, D.J., Blobel, G.A., and Shiekhattar, R. (2013). Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature *494*, 497-501.

Laugesen, A., and Helin, K. (2014). Chromatin repressive complexes in stem cells, development, and cancer. Cell stem cell *14*, 735-751.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nature reviews. Genetics *11*, 204-220.

Lawrence, T., Willoughby, D.A., and Gilroy, D.W. (2002). Anti-inflammatory lipid mediators and insights into the resolution of inflammation. Nature reviews. Immunology *2*, 787-795.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X.*, et al.* (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516-520.

Li, Z., Cai, X., Cai, C.L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood *118*, 4509-4518.

Lin, Y.C., Jhunjhunwala, S., Benner, C., Heinz, S., Welinder, E., Mansson, R., Sigvardsson, M., Hagman, J., Espinoza, C.A., Dutkowski, J.*, et al.* (2010). A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. Nature immunology *11*, 635-643.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D.*, et al.* (2013). Global epigenomic reconfiguration during mammalian brain development. Science *341*, 1237905.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315-322.

Liu, W.J., Tan, X.H., Luo, X.P., Guo, B.P., Wei, Z.J., Ke, Q., He, S., and Cen, H. (2014). Prognostic significance of Tet methylcytosine dioxygenase 2 (TET2) gene mutations in adult patients with acute myeloid leukemia: a meta-analysis. Leukemia & lymphoma.

Liutkeviciute, Z., Lukinavicius, G., Masevicius, V., Daujotyte, D., and Klimasauskas, S. (2009). Cytosine-5-methyltransferases add aldehydes to DNA. Nature chemical biology *5*, 400-402.

Loh, K.M., and Lim, B. (2010). Recreating pluripotency? Cell stem cell *7*, 137-139.

Lowe, C.B., Bejerano, G., and Haussler, D. (2007). Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proceedings of the National Academy of Sciences of the United States of America *104*, 8005-8010.

Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014). Role of Tet proteins in enhancer activity and telomere elongation. Genes & development *28*, 2103-2119.

Lu, X., Song, C.X., Szulwach, K., Wang, Z., Weidenbacher, P., Jin, P., and He, C. (2013). Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. Journal of the American Chemical Society *135*, 9315-9317.

Luger, K., Rechsteiner, T.J., Flaus, A.J., Waye, M.M., and Richmond, T.J. (1997). Characterization of nucleosome core particles containing histone proteins made in bacteria. Journal of molecular biology *272*, 301-311.

Ma, D.K., Jang, M.H., Guo, J.U., Kitabatake, Y., Chang, M.L., Pow-Anpongkul, N., Flavell, R.A., Lu, B., Ming, G.L., and Song, H. (2009). Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. Science *323*, 1074-1077.

Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. The Journal of biological chemistry *286*, 35334-35338.

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nature reviews. Genetics *11*, 285-296.

Martin, P. (1997). Wound healing--aiming for perfect skin regeneration. Science *276*, 75-81.

Matarese, F., Carrillo-de Santa Pau, E., and Stunnenberg, H.G. (2011). 5-Hydroxymethylcytosine: a new kid on the epigenetic block? Molecular systems biology *7*, 562.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501-502.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology *28*, 495-501.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*, 499-507.

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

Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Oude Vrielink, J.A., Elkon, R., Melo, S.A., Leveille, N., Kalluri, R.*, et al.* (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. Molecular cell *49*, 524-535.

Merad, M., and Manz, M.G. (2009). Dendritic cell homeostasis. Blood *113*, 3418-3427.

Metivier, R., Gallais, R., Tiffoche, C., Le Peron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G.*, et al.* (2008). Cyclical DNA methylation of a transcriptionally active promoter. Nature *452*, 45-50.

Metivier, R., Penot, G., Hubner, M.R., Reid, G., Brand, H., Kos, M., and Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. Cell *115*, 751-763.

Mildner, A., Schmidt, H., Nitsche, M., Merkler, D., Hanisch, U.K., Mack, M., Heikenwalder, M., Bruck, W., Priller, J., and Prinz, M. (2007). Microglia in the adult brain arise from Ly-6ChiCCR2+ monocytes only under defined host conditions. Nature neuroscience *10*, 1544-1553.

Mildner, A., Yona, S., and Jung, S. (2013). A close encounter of the third kind: monocyte-derived cells. Advances in immunology *120*, 69-103.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. Science *315*, 1408-1411.

Mohn, F., Weber, M., Schubeler, D., and Roloff, T.C. (2009). Methylated DNA immunoprecipitation (MeDIP). Methods in molecular biology *507*, 55-64.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X.*, et al.* (2011). Tet2 loss leads to

increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer cell *20*, 11-24.

Morgan, H.D., Santos, F., Green, K., Dean, W., and Reik, W. (2005). Epigenetic reprogramming in mammals. Human molecular genetics *14 Spec No 1*, R47-58.

Muller, U., Bauer, C., Siegl, M., Rottach, A., and Leonhardt, H. (2014). TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. Nucleic acids research *42*, 8592-8604.

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell *102*, 553-563.

Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nature chemical biology *8*, 751-758.

Nakano, H., Yanagita, M., and Gunn, M.D. (2001). CD11c(+)B220(+)Gr-1(+) cells in mouse lymph nodes and spleen display characteristics of plasmacytoid dendritic cells. The Journal of experimental medicine *194*, 1171-1178.

Natoli, G., and Andrau, J.C. (2012). Noncoding transcription at enhancers: general principles and functional models. Annual review of genetics *46*, 1-19.

Neddermann, P., and Jiricny, J. (1994). Efficient removal of uracil from G.U mispairs by the mismatch-specific thymine DNA glycosylase from HeLa cells. Proceedings of the National Academy of Sciences of the United States of America *91*, 1642-1646.

Ong, C.T., and Corces, V.G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. Nature reviews. Genetics *12*, 283-293.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D.*, et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Current biology : CB *10*, 475-478.

Otani, J., Arita, K., Kato, T., Kinoshita, M., Kimura, H., Suetake, I., Tajima, S., Ariyoshi, M., and Shirakawa, M. (2013). Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. The Journal of biological chemistry *288*, 6351-6362.

Passlick, B., Flieger, D., and Ziegler-Heitbrock, H.W. (1989). Identification and characterization of a novel monocyte subpopulation in human peripheral blood. Blood *74*, 2527-2534.

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

Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P.*, et al.* (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Nature *473*, 394-397.

Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deiml, C.A., Hagemeier, C., and Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. Angewandte Chemie *50*, 7008-7012.

Pham, T.H., Benner, C., Lichtinger, M., Schwarzfischer, L., Hu, Y., Andreesen, R., Chen, W., and Rehli, M. (2012). Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states. Blood *119*, e161-171.

Pham, T.H., Langmann, S., Schwarzfischer, L., El Chartouni, C., Lichtinger, M., Klug, M., Krause, S.W., and Rehli, M. (2007). CCAAT enhancer-binding protein beta regulates constitutive gene expression during late stages of monocyte to macrophage differentiation. The Journal of biological chemistry *282*, 21924-21933.

Pham, T.H., Minderjahn, J., Schmidl, C., Hoffmeister, H., Schmidhofer, S., Chen, W., Langst, G., Benner, C., and Rehli, M. (2013). Mechanisms of in vivo binding site selection of the hematopoietic master transcription factor PU.1. Nucleic acids research *41*, 6391-6402.

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

Pilon, A.M., Ajay, S.S., Kumar, S.A., Steiner, L.A., Cherukuri, P.F., Wincovitch, S., Anderson, S.M., Mullikin, J.C., Gallagher, P.G., Hardison, R.C.*, et al.* (2011). Genome-wide ChIP-Seq reveals a dramatic shift in the binding of the transcription factor erythroid Kruppel-like factor during erythrocyte differentiation. Blood *118*, e139-148.

Pio, F., Kodandapani, R., Ni, C.Z., Shepard, W., Klemsz, M., McKercher, S.R., Maki, R.A., and Ely, K.R. (1996). New insights on DNA recognition by ets proteins from the crystal structure of the PU.1 ETS domain-DNA complex. The Journal of biological chemistry *271*, 23329-23337.

Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature *463*, 1101-1105.

Probst, A.V., Dunleavy, E., and Almouzni, G. (2009). Epigenetic inheritance during the cell cycle. Nature reviews. Molecular cell biology *10*, 192-206.

Pulendran, B. (2004). Modulating vaccine responses with dendritic cells and Toll-like receptors. Immunological reviews *199*, 227-250.

Quivoron, C., Couronne, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.H.*, et al.* (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer cell *20*, 25-38.

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

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201-1212.

Rai, K., Sarkar, S., Broadbent, T.J., Voas, M., Grossmann, K.F., Nadauld, L.D., Dehghanizadeh, S., Hagos, F.T., Li, Y., Toth, R.K.*, et al.* (2010). DNA demethylase activity maintains intestinal cells in an undifferentiated state following loss of APC. Cell *142*, 930-942.

Raiber, E.A., Beraldi, D., Ficz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. Genome biology *13*, R69.

Randolph, G.J., Beaulieu, S., Lebecque, S., Steinman, R.M., and Muller, W.A. (1998). Differentiation of monocytes into dendritic cells in a model of transendothelial trafficking. Science *282*, 480-483.

Rehli, M. (2000). PU.1 and Interferon Consensus Sequence-binding Protein Regulate the Myeloid Expression of the Human Toll-like Receptor 4 Gene. Journal of Biological Chemistry *275*, 9773-9781.

Ridinger-Saison, M., Boeva, V., Rimmele, P., Kulakovskiy, I., Gallais, I., Levavasseur, B., Paccard, C., Legoix-Ne, P., Morle, F., Nicolas, A.*, et al.* (2012). Spi-1/PU.1 activates transcription through clustered DNA occupancy in erythroleukemia. Nucleic acids research *40*, 8927-8941.

Robertson, K.D., and Wolffe, A.P. (2000). DNA methylation in health and disease. Nature reviews. Genetics *1*, 11-19.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS letters *579*, 909-915.

Rozenberg, J.M., Shlyakhtenko, A., Glass, K., Rishi, V., Myakishev, M.V., FitzGerald, P.C., and Vinson, C. (2008). All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. BMC genomics *9*, 67.

Russo, V.E.A., Martienssen, R. A., Riggs, A.D. (1996). Epigenetic mechanisms of gene regulation. Cold Spring Harbor Laboratory Press, Plainview, N.Y.

Sakabe, K., and Hart, G.W. (2010). O-GlcNAc transferase regulates mitotic chromatin dynamics. The Journal of biological chemistry *285*, 34460-34468.

Sallusto, F., and Lanzavecchia, A. (1994). Efficient presentation of soluble antigen by cultured human dendritic cells is maintained by granulocyte/macrophage colony-stimulating factor plus interleukin 4 and downregulated by tumor necrosis factor alpha. The Journal of experimental medicine *179*, 1109-1118.

Sanderson, R.J., Shepperdson, R.T., Vatter, A.E., and Talmage, D.W. (1977). Isolation and enumeration of peripheral blood monocytes. Journal of immunology *118*, 1409-1414.

Sarraf, S.A., and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Molecular cell *15*, 595-605.

Sawyer, R.T., Strausbauch, P.H., and Volkman, A. (1982). Resident macrophage proliferation in mice depleted of blood monocytes by strontium-89. Laboratory investigation; a journal of technical methods and pathology *46*, 165-170.

Schiesser, S., Hackner, B., Pfaffeneder, T., Muller, M., Hagemeier, C., Truss, M., and Carell, T. (2012). Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. Angewandte Chemie *51*, 6516-6520.

Schmidl, C., Klug, M., Boeld, T.J., Andreesen, R., Hoffmann, P., Edinger, M., and Rehli, M. (2009). Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. Genome research *19*, 1165-1174.

Schmidl, C., Renner, K., Peter, K., Eder, R., Lassmann, T., Balwierz, P.J., Itoh, M., Nagao-Sato, S., Kawaji, H., Carninci, P.*, et al.* (2014). Transcription and enhancer profiling in human monocyte subsets. Blood *123*, e90-99.

Schonheit, J., Kuhl, C., Gebhardt, M.L., Klett, F.F., Riemke, P., Scheller, M., Huang, G., Naumann, R., Leutz, A., Stocking, C.*, et al.* (2013). PU.1 level-directed chromatin structure remodeling at the Irf8 gene drives dendritic cell commitment. Cell reports *3*, 1617-1628.

Schulz, C., Gomez Perdiguero, E., Chorro, L., Szabo-Rogers, H., Cagnard, N., Kierdorf, K., Prinz, M., Wu, B., Jacobsen, S.E., Pollard, J.W.*, et al.* (2012). A lineage of myeloid cells independent of Myb and hematopoietic stem cells. Science *336*, 86-90.

Segura, E., and Amigorena, S. (2013). Inflammatory dendritic cells in mice and humans. Trends in immunology *34*, 440-445.

Serandour, A.A., Avner, S., Oger, F., Bizot, M., Percevault, F., Lucchetti-Miganeh, C., Palierne, G., Gheeraert, C., Barloy-Hubler, F., Peron, C.L.*, et al.* (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. Nucleic acids research *40*, 8255-8265.

Serandour, A.A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Metivier, R.*, et al.* (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. Genome research *21*, 555-565.

Serbina, N.V., Jia, T., Hohl, T.M., and Pamer, E.G. (2008). Monocyte-mediated defense against microbial pathogens. Annual review of immunology *26*, 421-452.

Serbina, N.V., and Pamer, E.G. (2006). Monocyte emigration from bone marrow during bacterial infection requires signals mediated by chemokine receptor CCR2. Nature immunology *7*, 311-317.

Serbina, N.V., Salazar-Mather, T.P., Biron, C.A., Kuziel, W.A., and Pamer, E.G. (2003). TNF/iNOS-producing dendritic cells mediate innate immune defense against bacterial infection. Immunity *19*, 59-70.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.*, et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.

Shearstone, J.R., Pop, R., Bock, C., Boyle, P., Meissner, A., and Socolovsky, M. (2011). Global DNA demethylation during mouse erythropoiesis in vivo. Science *334*, 799-802.

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., and Zhang, Y. (2014). Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. Cell stem cell *15*, 459-470.

Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell *153*, 692-706.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V.*, et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Shi, C., and Pamer, E.G. (2011). Monocyte recruitment during infection and inflammation. Nature reviews. Immunology *11*, 762-774.

Shi, F.T., Kim, H., Lu, W., He, Q., Liu, D., Goodell, M.A., Wan, M., and Songyang, Z. (2013). Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. The Journal of biological chemistry *288*, 20776-20784.

Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., and Sasaki, H. (2013). Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. PLoS genetics *9*, e1003439.

Siegal, F.P., Kadowaki, N., Shodell, M., Fitzgerald-Bocarsly, P.A., Shah, K., Ho, S., Antonenko, S., and Liu, Y.J. (1999). The nature of the principal type 1 interferon-producing cells in human blood. Science *284*, 1835-1837.

Singh, H., DeKoter, R.P., and Walsh, J.C. (1999). PU.1, a shared transcriptional regulator of lymphoid and myeloid cell fates. Cold Spring Harbor symposia on quantitative biology *64*, 13-20.

Smith, M.F., Jr., Carl, V.S., Lodie, T., and Fenton, M.J. (1998). Secretory interleukin-1 receptor antagonist gene expression requires both a PU.1 and a novel composite NF-kappaB/PU.1/ GA-binding protein binding site. The Journal of biological chemistry *273*, 24272-24279.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nature reviews. Genetics *14*, 204-220.

Song, C.X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H.*, et al.* (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell *153*, 678-691.

Song, L., Zhang, Z., Grasfeder, L.L., Boyle, A.P., Giresi, P.G., Lee, B.K., Sheffield, N.C., Graf, S., Huss, M., Keefe, D.*, et al.* (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome research *21*, 1757-1767.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nature reviews. Genetics *13*, 613-626.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F.*, et al.* (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell *152*, 1146-1159.

Spruijt, C.G., and Vermeulen, M. (2014). DNA methylation: old dog, new tricks? Nature structural & molecular biology *21*, 949-954.

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

Steinman, R.M., and Cohn, Z.A. (1973). Identification of a novel cell type in peripheral lymphoid organs of mice. I. Morphology, quantitation, tissue distribution. The Journal of experimental medicine *137*, 1142-1162.

Steinman, R.M., Hawiger, D., and Nussenzweig, M.C. (2003). Tolerogenic dendritic cells. Annual review of immunology *21*, 685-711.

Steinman, R.M., and Inaba, K. (1999). Myeloid dendritic cells. Journal of leukocyte biology *66*, 205-208.

Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., and Jacobsen, S.E. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome biology *12*, R54.

Sunahori, K., Juang, Y.T., and Tsokos, G.C. (2009). Methylation Status of CpG Islands Flanking a cAMP Response Element Motif on the Protein Phosphatase 2Ac Promoter Determines CREB Binding and Activity. The Journal of Immunology *182*, 1500-1508.

Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T., Hara, E., Tenen, D.G., Hozumi, N., and Oikawa, T. (2006). Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. Oncogene *25*, 2477-2488.

Szulwach, K.E., and Jin, P. (2014). Integrating DNA methylation dynamics into a framework for understanding epigenetic codes. BioEssays : news and reviews in molecular, cellular and developmental biology *36*, 107-117.

Szulwach, K.E., Li, X., Li, Y., Song, C.X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K.*, et al.* (2011). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. PLoS genetics *7*, e1002154.

Taher, L., and Ovcharenko, I. (2009). Variable locus length in the human genome leads to ascertainment bias in functional inference for non-coding elements. Bioinformatics (Oxford, England) *25*, 578-584.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Tallack, M.R., Whitington, T., Yuen, W.S., Wainwright, E.N., Keys, J.R., Gardiner, B.B., Nourbakhsh, E., Cloonan, N., Grimmond, S.M., Bailey, T.L.*, et al.* (2010). A global role for KLF1 in erythropoiesis revealed by ChIP-seq in primary erythroid cells. Genome research *20*, 1052-1063.

Tarling, J.D., Lin, H.S., and Hsu, S. (1987). Self-renewal of pulmonary alveolar macrophages: evidence from radiation chimera studies. Journal of leukocyte biology *42*, 443-446.

Tate, P.H., and Bird, A.P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. Current opinion in genetics & development *3*, 226-231.

Tavian, M., and Peault, B. (2005). Embryonic development of the human hematopoietic system. The International journal of developmental biology *49*, 243-250.

Thomson, J.P., Hunter, J.M., Nestor, C.E., Dunican, D.S., Terranova, R., Moggs, J.G., and Meehan, R.R. (2013). Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. Nucleic acids research *41*, e206.

Thomson, M., Liu, S.J., Zou, L.N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell *145*, 875-889.

Tini, M., Benecke, A., Um, S.J., Torchia, J., Evans, R.M., and Chambon, P. (2002). Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. Molecular cell *9*, 265-277.

Tsuchiya, S., Kobayashi, Y., Goto, Y., Okumura, H., Nakae, S., Konno, T., and Tada, K. (1982). Induction of maturation in cultured human monocytic leukemia cells by a phorbol diester. Cancer research *42*, 1530-1536.

Turkistany, S.A., and DeKoter, R.P. (2011). The transcription factor PU.1 is a critical regulator of cellular communication in the immune system. Archivum immunologiae et therapiae experimentalis *59*, 431-440.

Valinluck, V., and Sowers, L.C. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer research *67*, 946-950.

Valinluck, V., Tsai, H.H., Rogstad, D.K., Burdzy, A., Bird, A., and Sowers, L.C. (2004). Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic acids research *32*, 4100-4108.

Valledor, A.F., Comalada, M., Santamaria-Babi, L.F., Lloberas, J., and Celada, A. (2010). Macrophage proinflammatory activation and deactivation: a question of balance. Advances in immunology *108*, 1-20.

van Furth, R., and Cohn, Z.A. (1968). The origin and kinetics of mononuclear phagocytes. The Journal of experimental medicine *128*, 415-435.

van Furth, R., Raeburn, J.A., and van Zwet, T.L. (1979). Characteristics of human mononuclear phagocytes. Blood *54*, 485-500.

Vance, K.W., and Ponting, C.P. (2014). Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends in genetics : TIG *30*, 348-355.

Varol, C., Vallon-Eberhard, A., Elinav, E., Aychek, T., Shapira, Y., Luche, H., Fehling, H.J., Hardt, W.D., Shakhar, G., and Jung, S. (2009). Intestinal lamina propria dendritic cell subsets have different origin and functions. Immunity *31*, 502-512.

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K.*, et al.* (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Molecular cell *49*, 645-656.

Vosseller, K., Sakabe, K., Wells, L., and Hart, G.W. (2002). Diverse regulation of protein function by O-GlcNAc: a nuclear and cytoplasmic carbohydrate post-translational modification. Current Opinion in Chemical Biology *6*, 851-857.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W.*, et al.* (2014). Programming and inheritance of parental DNA methylomes in mammals. Cell *157*, 979-991.

Weake, V.M., and Workman, J.L. (2010). Inducible gene expression: diverse regulatory mechanisms. Nature reviews. Genetics *11*, 426-437.

Weber, C.M., Ramachandran, S., and Henikoff, S. (2014). Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Molecular cell *53*, 819-830.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature genetics *39*, 457-466.

Wei, G.H., Badis, G., Berger, M.F., Kivioja, T., Palin, K., Enge, M., Bonke, M., Jolma, A., Varjosalo, M., Gehrke, A.R.*, et al.* (2010). Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. The EMBO journal *29*, 2147-2160.

Weissmann, S., Alpermann, T., Grossmann, V., Kowarsch, A., Nadarajah, N., Eder, C., Dicker, F., Fasan, A., Haferlach, C., Haferlach, T.*, et al.* (2012). Landscape of TET2 mutations in acute myeloid leukemia. Leukemia *26*, 934-942.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221-225.

Wijffels, J.F., de Rover, Z., Beelen, R.H., Kraal, G., and van Rooijen, N. (1994). Macrophage subpopulations in the mouse spleen renewed by local proliferation. Immunobiology *191*, 52-64.

Williams, K., Christensen, J., and Helin, K. (2012). DNA methylation: TET proteins-guardians of CpG islands? EMBO reports *13*, 28-35.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343-348.

Wilson, N.K., Foster, S.D., Wang, X., Knezevic, K., Schutte, J., Kaimakis, P., Chilarska, P.M., Kinston, S., Ouwehand, W.H., Dzierzak, E.*, et al.* (2010). Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. Cell stem cell *7*, 532-544.

Wontakal, S.N., Guo, X., Will, B., Shi, M., Raha, D., Mahajan, M.C., Weissman, S., Snyder, M., Steidl, U., Zheng, D.*, et al.* (2011). A large gene network in immature erythroid cells is controlled by the myeloid and B cell transcriptional regulator PU.1. PLoS genetics *7*, e1001392.

Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011a). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes & development *25*, 679-684.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011b). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature *473*, 389-393.

Wu, H., and Zhang, Y. (2011). Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. Cell Cycle *10*, 2428-2436.

Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. Cell *156*, 45-68.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. Nature reviews. Molecular cell biology *11*, 607-620.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell *148*, 816-831.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S.*, et al.* (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Molecular cell *42*, 451-464.

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T.*, et al.* (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for Xenopus eye and neural development. Cell *151*, 1200-1213.

Xu, Y., Zhan, Y., Lew, A.M., Naik, S.H., and Kershaw, M.H. (2007). Differential development of murine dendritic cells by GM-CSF versus Flt3 ligand has implications for inflammation and trafficking. Journal of immunology *179*, 7577-7584.

Yamamoto, H., Kihara-Negishi, F., Yamada, T., Hashimoto, Y., and Oikawa, T. (1999). Physical and functional interactions between the transcription factor PU.1 and the coactivator CBP. Oncogene *18*, 1495-1501.

Yang, Y.W., Flynn, R.A., Chen, Y., Qu, K., Wan, B., Wang, K.C., Lei, M., and Chang, H.Y. (2014). Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. eLife *3*, e02046.

Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell *147*, 1498-1510.

Yona, S., and Jung, S. (2010). Monocytes: subsets, origins, fates and functions. Current opinion in hematology *17*, 53-59.

Yona, S., Kim, K.W., Wolf, Y., Mildner, A., Varol, D., Breker, M., Strauss-Ayali, D., Viukov, S., Guilliams, M., Misharin, A.*, et al.* (2013). Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. Immunity *38*, 79-91.

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B.*, et al.* (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell *149*, 1368-1380.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors: establishing competence for gene expression. Genes & development *25*, 2227-2241.

Zawada, A.M., Rogacev, K.S., Rotter, B., Winter, P., Marell, R.R., Fliser, D., and Heine, G.H. (2011). SuperSAGE evidence for CD14++CD16+ monocytes as a third monocyte subset. Blood *118*, e50-61.

Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y.G. (2010). TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. Cell research *20*, 1390-1393.

Zhang, J.A., Mortazavi, A., Williams, B.A., Wold, B.J., and Rothenberg, E.V. (2012a). Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell *149*, 467-482.

Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.-L., Luo, C., Jiang, H., and He, C. (2012b). Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nature chemical biology *8*, 328-330.

Zhang, P., Behre, G., Pan, J., Iwama, A., Wara-Aswapati, N., Radomska, H.S., Auron, P.E., Tenen, D.G., and Sun, Z. (1999). Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proceedings of the National Academy of Sciences of the United States of America *96*, 8705-8710.

Zhang, P., Zhang, X., Iwama, A., Yu, C., Smith, K.A., Mueller, B.U., Narravula, S., Torbett, B.E., Orkin, S.H., and Tenen, D.G. (2000). PU.1 inhibits GATA-1 function and erythroid differentiation by blocking GATA-1 DNA binding. Blood *96*, 2641-2648.

Zhang, Q., Liu, X., Gao, W., Li, P., Hou, J., Li, J., and Wong, J. (2014). Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked beta-N-acetylglucosamine transferase (OGT). The Journal of biological chemistry *289*, 5986-5996.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Molecular cell *40*, 939-953.

Zhu, J.K. (2009). Active DNA demethylation mediated by DNA glycosylases. Annual review of genetics *43*, 143-166.

Ziegler-Heitbrock, L., Ancuta, P., Crowe, S., Dalod, M., Grau, V., Hart, D.N., Leenen, P.J., Liu, Y.J., MacPherson, G., Randolph, G.J.*, et al.* (2010). Nomenclature of monocytes and dendritic cells in blood. Blood *116*, e74-80.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature *456*, 125-129.

8 Acknowledgment

Zum Schluss möchte ich mich bei einigen Menschen sehr herzlich bedanken:

- bei meinem Betreuer Prof. Dr. Michael Rehli für seine Geduld, sein allzeit offenes Ohr für Fragen auch in stressigen Zeiten, für seine Erreichbarkeit in der wohlverdienten Urlaubszeit und die Möglichkeit in seiner Arbeitsgruppe diese Doktorarbeit anzufertigen.

- bei unserem Abteilungsleiter Prof. Dr. Reinhard Andreesen und auch bei seinem Nachfolger Prof. Dr. Wolfgang Herr, die beide mit Ihrer ruhigen Gelassenheit entspannte Betriebsausflüge, Weihnachtsfeiern und Retreats ermöglicht haben.

- bei meinen Mentoren, Prof. Dr. Dirk Schübeler für hilfreiche Anregungen und bei Prof. Dr. Gernot Längst für gute Ratschläge, für die Möglichkeit ein Praktikum in Chile zu betreuen und in der RIGeL Graduiertenschule mitzuwirken.

- bei Julia und Almut für viele schöne Momente, die den Laboralltag ungemein versüßt haben, für eure Unterstützung in allen Labor- und Lebenslagen und trotz des Arbeitspensums für eine unglaublich schöne Zeit.

- bei der gesamten AG Rehli: Thu Hang P., Chris S., Claudia G., Maja K., Lucia S., Ute A., Ireen R., Dagmar G., Johanna R., Julia M., Sandra P., Saskia P., Daniel H., Claudia K., vielen Dank für die gute Zusammenarbeit, für kurzweilige Mittagessen und Kaffee-Pausen und ein tolles Arbeitsklima!

- bei unseren Nachbarn, der AG Kreutz für die „Beherbergung“ im Carreras-Anbau, für kleine Pläusche zwischendrin und eine stets volle Schokoladen-Schublade. Bei Kristina und Jacky und dem Rest der AG Edinger/Hoffmann, beim JCC-Team und der AG Holler, bei Susanne und Kinga und allen anderen, die zum Gelingen dieser Arbeit in irgendeiner Form beigetragen haben.

Ein großer Dank gilt meiner Familie, meinen Freunden und besonders meinem Freund, die mich mit viel Geduld und Verständnis unterstützt und mir im Studium und im Leben zur Seite gestanden haben.