Dissertation

Deconvolution of Hematopoietic Commitment Decisions by
Genome-wide Analysis of Progressive DNA Methylation Changes

Sina Stäble

2019

Dissertation

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by

Master of Science, Diplom-Ingenieurin (FH) Sina Stäble
born in Gießen, Germany

Oral examination: December 2nd 2019

Deconvolution of Hematopoietic Commitment Decisions by Genome-wide Analysis of Progressive DNA Methylation Changes

Referees:
PD Dr. Odilia Popanda
Dr. Michael Milsom

Das Geheimnis aller Erfinder ist, nichts für unmöglich anzusehen.

Justus Freiherr von Liebig (1803 - 1873)

This work is dedicated to my parents, Carmen und Uwe Zimmermann.

Summary

Hematopoietic stem cells (HSCs) are responsible for the life-long production of all mature blood cells. The classical hierarchical model of hematopoiesis has been revised based on single-cell transcriptome analyses to suggest a continuous rather than a step-wise differentiation process. While differentiation trajectories can be inferred from single-cell transcriptional snapshots, it remains a challenge to identify definitive points of lineage commitment. The molecular characterization of such commitment decisions is required to accurately model the hematopoietic system.

The analysis of 5-methylcytosine may facilitate the identification of such molecular commitment marks due to the progressive nature of how DNA methylation is programmed during differentiation. In this doctoral thesis, a genome-wide DNA methylome map of murine hematopoiesis was generated using tagmentation-based whole genome bisulfite sequencing, encompassing 26 hematopoietic cell populations. Across all populations, 147,232 differentially methylated regions (DMRs) were identified and grouped into coordinately regulated regions by hierarchical clustering. These dynamically regulated regions showed progressive and unidirectional DNA methylation programming during normal hematopoietic differentiation. The DNA methyltion programs can be interpreted as pan-hematopoietic, lineage- and cell type-specific, indicating a molecular mechanism of cell fate restriction. The lineage specificity of the DNA methylation programs was confirmed by the enrichment of hematopoietic transcription factor binding motifs and lineage-specific enhancer programs. Strikingly, lineage-specific DMRs were already identified within the multipotent hematopoietic stem and progenitor compartment, supporting the concept of early lineage restriction. Furthermore, a model of murine hematopoiesis could be inferred in form of a diffusion map that is purely based on DNA methylation dynamics during hematopoietic differentiation. To gain further insights into how DNA methylation dynamics relate to the regulation of gene expression, a single-cell transcriptome map of the entire murine hematopoietic system was generated. The integration of DNA methylation dynamics with single-cell gene expression patterns provided evidence for an anti-correlation between DNA methylation and cell-type specific gene expression patterns. However, loss of DNA methylation was not invariably associated with an increase in gene expression. This suggests that DNA methylation has more a permissive rather than an instructive role in regulating lineage-specific transcriptional programs.

Finally, the DNA methylome map was used as a resource to investigate epigenetic patterns that underlie the myeloid and megakaryocytic lineage bias observed in aged mice. By comparing the DNA methylomes of young and aged HSCs, 3,275 DMRs were identi-

fied, which were predominantly associated with a loss of DNA methylation in genes involved in integrin signaling, platelet activation, and platelet aggregation. Notably, 46% of these DMRs overlapped with DNA methylation patterns identified in normal hematopoietic differentiation and significantly affected DNA methylation programs showing low DNA methylation levels in megakaryocyte-primed cell populations. Together, these findings suggest that HSC aging alters the DNA methylome *in vivo*, in a manner that is associated with a increased differentiation towards the megakaryocytic lineage.

In summary, the DNA methylome map of murine hematopoiesis generated in this thesis provides novel insights into the epigenetic regulation of hematopoietic differentiation and complements recent findings from single-cell transcriptome studies. Furthermore, it represents a rich resource to investigate DNA methylation patterns in hematopoiesis across a broad range of conditions.

Zusammenfassung

Hämatopoetische Stammzellen (HSZ) sind für die lebenslange Produktion aller reifen Blutzellen verantwortlich. Das klassische hierarchische Modell der Hämatopoese wurde basierend auf Einzelzell-Transkriptomanalysen weiterentwickelt, so dass nun ein kontinuierlicher Differenzierungsprozess postuliert wird. Obwohl Differenzierungstrajektorien aus Einzelzell-Transkriptomen abgeleitet werden können, bleibt es eine Herausforderung, den definitiven Differenzierungspunkt zu bestimmen. Die molekulare Charakterisierung solcher Differenzierungspunkte ist jedoch essenziell, um das hämatopoetische System im Detail zu verstehen.

Die Analyse von 5-Methylcytosin könnte aufgrund der kontinuierlichen und robusten DNA-Methylierungsveränderungen während des Differenzierungsprozesses die Identifizierung solcher molekularen Differenzierungspunkte ermöglichen. Im Rahmen dieser Doktorarbeit wurde ein genomweiter DNA-Methylierungsdatensatz bestehend aus 26 murinen hämatopoetischen Zellpopulationen generiert. Im gesamten hämatopoetischen System wurden 147.232 differentiell methylierte Regionen (DMRs) identifiziert, welche durch hierarchische Clusteranalyse in koordiniert regulierte Regionen gruppiert wurden. Diese dynamisch regulierten Regionen zeigten progressive und unidirektionale DNA-Methylierungsveränderungen während der normalen hämatopoetischen Differenzierung und wurden als pan-hämatopoetische, linien- und zelltypspezifische DNA-Methylierungsprogramme interpretiert. Die biologische Relevanz dieser Programme wurde durch die Anreicherung von hämatopoetischen Transkriptionsfaktor-Bindungsmotiven und linienspezifischen Enhancer-Programmen bestätigt. Hervorzuheben ist, dass bereits im multipotenten hämatopoetischen Stamm- und Vorläuferzellkompartiment linienspezifische DMRs identifiziert wurden, was auf eine frühe hämatopoetische Differenzierung hindeutet. Des Weiteren konnte mithilfe des sogenannten Diffusions-Algorithmus ein Modell der murinen Hämatopoese abgeleitet werden, welches lediglich auf DNA-Methylierungsdynamiken während der Blutbildung basiert. Um den Zusammenhang zwischen DNA-Methylierung und Genexpression zu untersuchen, wurde ein Einzelzell-Transkriptomdatensatz des gesamten hämatopoetischen Systems erstellt. Die Integration von DNA-Methylierungs- und Transkriptionsdynamiken weißt auf eine allgemeine Anti-Korrelation dieser molekularen Mechanismen hin. Jedoch war ein DNA-Methylierungsverlust nicht zwingend mit einer Zunahme der Genexpression verbunden, was vermuten lässt, dass ein Verlust der DNA-Methylierung die Zelle lediglich auf potenzielle Transkriptionsprogramme vorbereitet, welche im Verlauf der Differenzierung an-, aber auch wieder abgeschaltet werden können.

Im abschließenden Teil dieser Doktorarbeit wurden die generierten DNA-Methylome als

Ressource verwendet, um epigenetische Muster zu untersuchen, die der verstärkten myeloiden und megakaryozytären Differenzierung in der Blutbildung gealterter Mäuse zugrunde liegen. Beim Vergleich des Methyloms von jungen und gealterten HSZ wurden 3.275 DMRs identifiziert. Diese DMRs waren überwiegend durch Methylierungsverlust charakterisiert und betrafen Gene, die an Integrin-Signalwegen, Blutplättchenaktivierung und Blutplättchenaggregation beteiligt sind. Bemerkenswerterweise überlappten 46% dieser DMRs mit den DNA-Methylierungsveränderungen der normalen hämatopoetischen Differenzierung, welche vorwiegend eine niedrige DNA-Methylierung in megakaryozytären Zellpopulationen aufweisen. Insgesamt deuten diese Ergebnisse darauf hin, dass der Alterungsprozess von HSZ mit DNA-Methylierungsveränderungen einhergeht, die eine megakaryozytäre Differenzierung begünstigen.

Zusammenfassend liefert der vorliegende DNA-Methylomdatensatz der murinen Hämatopoese neue Erkenntnisse über die epigenetische Regulation der hämatopoetischen Differenzierung und ergänzt die kürzlich gewonnenen Erkenntnisse aus Einzelzell-Transkriptomstudien. Darüber stellt dieser Datensatz eine umfassende Ressource dar, um Veränderungen der DNA-Methylierungsmuster in der normalen oder malignen Hämatopoese zu untersuchen.

Contents

1 Introduction

The hematopoietic system is maintained by a pool of multipotent hematopoietic stem cells (HSCs) that are responsible for the life-long production of all mature blood cells. The process of hematopoietic differentiation is accompanied by tightly regulated epigenetic changes that affect gene expression patterns and in turn, the lineage specification of differentiating cells. Lineage commitment precedes differentiation and is defined as the process in which a cell makes a decision for a certain cell fate.

1.1 Epigenetics

The term 'epigenetics' was introduced by the embryologist Conrad Hal Waddington in 1942 to describe causal interactions between the environment and the genome leading to the establishment of a phenotype [Waddington, 1942]. Moreover, Waddington published the 'epigenetic landscape' as a metaphor of epigenetic mechanisms that define the cell fate of undifferentiated cells. In this metaphor, cells are represented as marbles rolling down a landscape of bifurcation valleys separated by ridges (Figure 1.1 a). These valleys illustrate alternative cell fates that are shaped by 'guy-ropes' (gene and transcription factor interactions) underlying the epigenetic landscape [Waddington, 1957] (Figure 1.1 b).

Nowadays, 'epigenetics' is broadly defined as *"the study of changes in gene function that are meiotically and/or mitotically heritable and that do not entail changes in the DNA sequence"* [Wu and Morris, 2001]. External stimuli as well as normal developmental and differentiation processes can cause epigenetic changes that affect

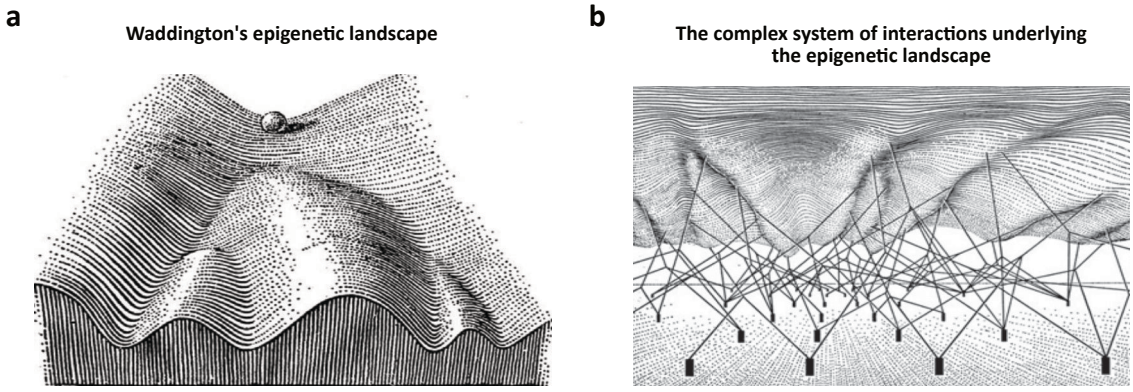

Figure 1.1: Waddington's epigenetic landscape. **a**, Waddington's epigenetic landscape wherein a cell represented as a marble can roll down a landscape of different cell fates illustrated by valleys. **b**, 'Guy-ropes' (complex interactions of gene and transcription factor networks) underlying the epigenetic landscape shape the valleys and ridges in the mountain. [Waddington, 1957]

gene expression patterns. The interaction of multiple epigenetic mechanisms defines the epigenome of a cell [Stricker et al., 2017].

1.1.1 Epigenetic layers

The epigenome of a cell is composed of different epigenetic layers, including DNA modifications, post-translational histone modifications, histone variants, nucleosome occupancy, RNA modifications, non-coding RNAs, and the three-dimensional chromatin conformation (Figure 1.2). These epigenetic mechanisms interact with transcription factors to regulate gene expression [Stricker et al., 2017].

At the single nucleotide level, covalent modifications of cytosines and adenines have been identified as epigenetic mechanisms that interfere with binding of transcription factors and affect gene expression. Methylation at carbon 5 (C5) of cytosine (5-methylcytosine, 5-mC) is by far the most investigated DNA modification and, in this thesis, is referred to as DNA methylation [Stricker et al., 2017].

At the nucleosome level, histone modifications are covalent post-translational modifications of histone tails, which include methylation, acetylation, and phosphorylation. Histone modifications as well as histone variants regulate the binding of DNA around the nucleosome core, thereby influencing the chromatin structure and the accessibility of the transcription machinery to the DNA. Histone modification can be profiled using chromatin immunoprecipitation sequencing (ChIP-seq) [Stricker et al., 2017].

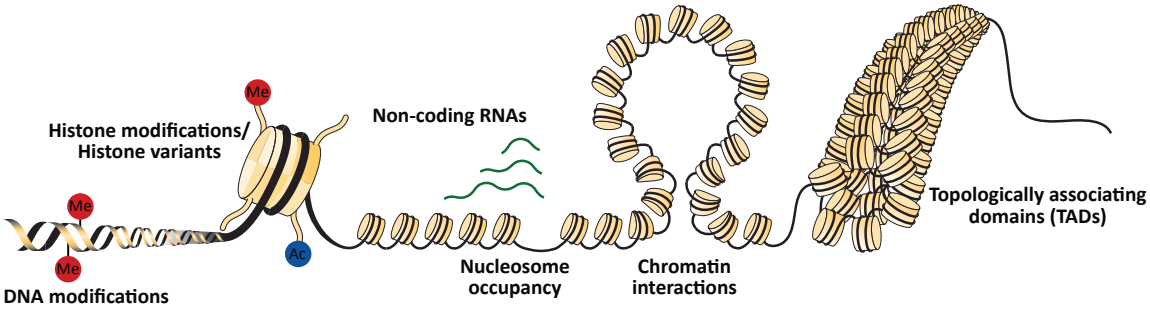

Figure 1.2: Epigenetic layers. The epigenome of a cell is composed of different epigenetic layers, including DNA modifications, post-translational histone modifications, histone variants, nucleosome occupancy, RNA modifications, non-coding RNAs, and the three-dimensional chromatin conformation (modified from Stricker *et al.* [Stricker et al., 2017].

At the chromatin level, nucleosome occupancy impacts on chromatin accessibility to *cis*-regulatory regions [Klemm et al., 2019]. Open chromatin can be profiled by assay for transposase-accessible chromatin sequencing (ATAC-seq) [Buenrostro et al., 2015]. Furthermore, the spatial chromatin organization within the nucleus is critical for cell type-specific interactions between enhancer and promoter regions that control gene expression [Zheng and Xie, 2019]. Genome-wide chromosome conformation capture assays (Hi-C) identified megabase-sized topological associating domains (TADs) of self-interacting chromatin [Lieberman-Aiden et al., 2009]. While TAD boundaries are conserved between cell types, intra-TAD interactions are highly dynamic during differentiation and cell type-specific [Zheng and Xie, 2019].

At the transcript level, RNA modifications and non-coding RNAs (ncRNAs) regulate gene expression by interference at the transcriptional or post-transcriptional level [Holoch and Moazed, 2015].

1.1.2 *Cis*-regulatory regions

Cis-regulatory regions, such as promoters and enhancers, play a pivotal role in the regulation of cell type-specific gene expression during development and cellular differentiation. Promoters are located directly upstream of the transcription start site (TSS) of genes, while enhancers can be located in variable distances up- or downstream of the TSS. *Cis*-regulatory elements contain binding sites for cell type-specific transcription factors. Chromatin loops bring enhancers and promoters into close proximity in order to enhance the activity of the transcription machinery. These

interactions are highly dynamic and occur mainly within TADs [Levo and Segal, 2014, Allis and Jenuwein, 2016].

Cis-regulatory sequences are identified by characteristic molecular features in the DNA. Whole-genome bisulfite sequencing, ChIP-seq, and ATAC-seq have enabled genome-wide mapping of *cis*-regulatory features across a variety of cell types and organisms [Stricker et al., 2017, Davis et al., 2018, Zerbino et al., 2015]. *Cis*-regulatory regions are usually characterized by low DNA methylation levels and are devoid of nucleosomes. Adjacent nucleosomes are marked by characteristic histone modifications. For example, histone H3 lysine 4 trimethylation (H3K4me3) and lysine 27 acetylation (H3K27ac) are found at active promoter regions, while histone H3 lysine 4 monomethylation (H3K4me1) and H3K27ac are associated with active enhancer regions [Zhou et al., 2011, Allis and Jenuwein, 2016].

1.1.3 DNA methylation

Covalent modifications of DNA bases were first discovered in 1948 [Hotchkiss, 1948]. In 1975, Holliday and Riggs proposed that DNA methylation is an epigenetic mechanism that plays an important role in the regulation of gene expression (Figure 1.3 a) [Riggs, 1975, Holliday and Pugh, 1975]. Indeed, early studies correlated DNA methylation with gene silencing. Accordingly, DNA methylation has been identified to be involved in X-chromosome inactivation, virus silencing, genetic imprinting, and the regulation of gene expression [Li et al., 1993, Mohandas et al., 1981, McGhee and Ginder, 1979, Ben-Hattar and Jiricny, 1988]. Although DNA methylation is broadly accepted as a repressive epigenetic mark, the correlation between DNA methylation and gene expression is certainly more complex than initially proposed [Stricker et al., 2017].

DNA methylation mainly occurs in the context of cytosine-phosphate-guanine (CpG) dinucleotides (Figure 1.3 b) [Doskocil and Sorm, 1962, Stricker et al., 2017]. More than 80% of all 28 million CpG sites in the human genome (21 million GpG sites in the murine genome) are methylated in somatic tissues [Smith and Meissner, 2013]. CpG-dense regions are found in so-called CpG-islands (CGIs) that frequently overlap with transcription start sites (TSS) of actively transcribed genes [Smith and Meissner, 2013, Luo et al., 2018]. CGIs commonly show DNA methylation values of less than 10%, while active distal regulatory regions, such as enhancers, reveal DNA methylation levels ranging from 10 to 50% [Stadler et al., 2011, Luo et al., 2018].

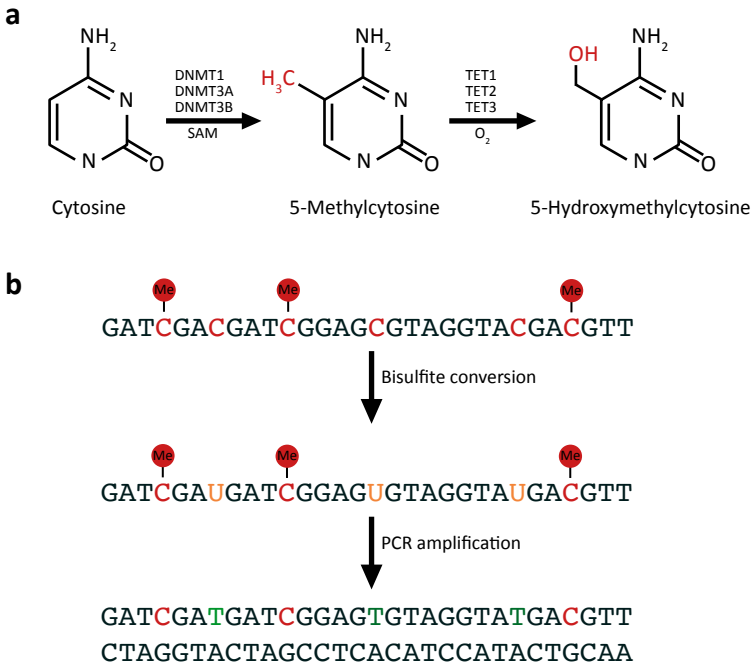

Figure 1.3: DNA methylation. **a**, DNMT1, DNMT3A, and DNMT3B methylate carbon 5 (C5) of cytosine using S-adenosyl methionine (SAM) as a methyl donor. TET enzymes oxidize the methyl group of 5-methylcytosine to 5-hydroxymethylcytosine as part of the active demethylation process. **b**, Bisulfite treatment of genomic DNA deaminates unmethylated cytosines to uracils, while methylated cytosines are protected. During PCR amplification, uracils are changed to thymidines.

CpGs in gene bodies of transcribed genes are usually methylated. Under physiological conditions, only 15 to 21% of all CpG sites show plasticity in DNA methylation among different somatic tissues [Ziller et al., 2013, Stadler et al., 2011, Luo et al., 2018]. Accordingly, the DNA methylome of different cell types can be described by locus-specific DNA methylation patterns [Bock et al., 2012, Meissner et al., 2008]. In contrast to somatic tissues, genome-wide erasure and re-establishment of DNA methylation levels are observed during early embryonic development [Greenberg and Bourc'his, 2019, Monk et al., 1987, Sanford et al., 1987].

DNA methylation is catalyzed by conserved DNA methyltransferases (DNMTs) that transfer methyl groups to cytosines. DNMT3A and DNMT3B establish *de novo* DNA methylation, while DNMT1 maintains methylation upon DNA replication using hemi-methylated CpG palindromes as a template (Figure 1.3 a) [Okano et al., 1999, Li et al., 1992, Smith and Meissner, 2013]. Active DNA demethylation is mediated by ten-eleven translocation (TET) methylcytosine dioxygenases, which

catalyze the sequential oxidation of 5-mC to 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC), and 5-carboxylcytosine (5-caC), which is followed by thymine DNA glycosylase-induced base excision repair [He et al., 2011, Ito et al., 2011, Wu and Zhang, 2017]. Passive DNA demethylation occurs progressively during cell division caused by imperfect DNA methylation maintenance of DNMT1 (Figure 1.3 a) [Wu and Zhang, 2017]. Loss of Dnmt or Tet enzymes leads to developmental abnormalities and embryonic lethality in mice [Li et al., 1992, Okano et al., 1999, Dawlaty et al., 2014]. Notably, both enzymes are recurrently mutated in the context of cancer, causing deregulation of DNA methylomes [Ley et al., 2010, Genovese et al., 2014, Delhommeau et al., 2009], a common feature of malignant transformation [Feinberg and Vogelstein, 1983].

Given that DNA methylation is a stable molecular mark that can be transferred to daughter cells, this epigenetic modification provides the molecular basis for an epigenetic memory. As a result, DNA methylation can be exploited to dissect differentiation trajectories [Bock et al., 2012, Farlik et al., 2016] and to identify the cell-of-origin in the context of malignant transformation [Kulis et al., 2015, Bormann et al., 2018, Oakes et al., 2014, Krivtsov et al., 2013].

1.1.4 DNA methylation profiling technologies

DNA methylation plays an important role in many physiological processes and pathological conditions. Hence, profiling of DNA methylation patterns is essential to understand how these biological processes are regulated at the molecular level. Over the past decades, a plethora of methodologies has been developed to investigate DNA methylation, including DNA methylation assays based on methylation-sensitive restriction enzymes (MRE) (e.g. MRE-seq), on 5-methylcytosine-specific antibodies (e.g. methylation DNA immunoprecipitation (MeDIP-seq)), or on bisulfite conversion (e.g. whole genome bisulfite sequencing (WGBS)) [Barros-Silva et al., 2018, Frommer et al., 1992]. The advent of next generation sequencing (NGS) enabled genome-wide DNA methylation profiling at single CpG resolution. Nowadays, WGBS is considered as the gold standard for DNA methylation profiling. Bisulfite treatment of genomic DNA deaminates unmethylated cytosines to uracils, while methylated cytosines are protected (Figure 1.3 b). Reduced-representation bisulfite sequencing (RRBS) offers a cost-effective alternative with single-CpG resolution that integrates MRE digestion followed by size-selection, bisulfite conversion, and NGS to investi-

gate CpG-rich regions, such as promoters and CpG-islands [Gu et al., 2011]. In addition, the microarray-based Infinium MethylationEPIC Bead Chip is a widely used bisulfite-based method that analyses 850,000 CpG sites of the human genome [Moran et al., 2016].

Given that bisulfite conversion is associated with a high degree of DNA degradation, classical WGBS requires large amounts of DNA (>1 µg) [Barros-Silva et al., 2018, Lister et al., 2009, Lipka et al., 2014]. Recent advances in WGBS have reduced the amount of starting material required to nanograms, which has enabled genome-wide DNA methylation profiling of rare cell populations, such as HSCs [Wang et al., 2013, Miura et al., 2012]. For instance, tagmentation-based whole genome bisulfite sequencing (TWGBS) requires only 5 to 30 ng of genomic DNA. TWGBS uses a Tn5 transposase to fragment genomic DNA and ligate sequencing adapters in a single step that is followed by bisulfite conversion and PCR-amplification of sequencing libraries [Wang et al., 2013, Lipka et al., 2014]. In this doctoral thesis, TWGBS was used to generate genome-wide and high coverage DNA methylation profiles starting from 10,000 FACS-sorted hematopoietic cells per replicate.

Recent technological advances have made it possible to profile genome-wide DNA methylation patterns at the single-cell level, covering up to 50% of all CpG dinucleotides [Smallwood et al., 2014, Farlik et al., 2015, Clark et al., 2017, Mulqueen et al., 2018]. For example, single-cell bisulfite sequencing (scBS-seq) established by Clark *et al.* uses post-bisulfite adaptor tagging (PBAT), a method that bisulfite converts and fragments the genomic DNA in a single step before sequencing adaptor ligation [Clark et al., 2017]. Building on single-cell DNA methylation profiling, single-cell multi-omics approaches have been developed to generate transcriptomic and epigenetic profiles from a single cell [Clark et al., 2018]. Despite the sparsity of single-cell DNA methylation profiles, studies have revealed epigenetic heterogeneity within cell populations [Macaulay et al., 2017, Angermueller et al., 2016].

Nanopore sequencing, a third generation sequencing approach, records changes to an electrical current as a single nucleotide sequence is passed through a protein nanopore. The resulting signal can be translated into a DNA sequence and can even be used to distinguish modified bases, including methylated cytosines. Although this technology is still in the early stages of development and not yet feasible on low input samples or on single cells, nanopore sequencing is a promising technology, which might allow direct analysis of DNA modifications without the currently needed bisulfite conversion step [Laszlo et al., 2013, Schatz, 2017].

1.1.5 Analysis of genome-wide DNA methylation data

Beta-values are a measure to describe the DNA methylation levels of single CpG sites or of genomic regions. Beta-values, ranging from 0 to 1, are defined as the ratio of methylated CpGs to the sum of methylated and unmethylated CpGs [Du et al., 2010].

Various bioinformatic and statistical approaches have been developed to analyze large-scale and genome-wide DNA methylation data. The analysis consists of two major steps: first, alignment of bisulfite-converted sequencing reads and DNA methylation calling. Second, the detection of differentially methylated CpGs, also referred to as differentially methylated loci (DML), between two different biological conditions. These DMLs are then merged to define differentially methylated regions (DMRs), which are genomic regions of multiple adjacent CpG sites with different methylation levels across multiple samples. DMRs often overlap with *cis*-regulatory regions and therefore influence the regulation of gene expression. The size of DMRs is typically in the range of a few hundred to a few thousand bases [Bock, 2012].

1.2 The hematopoietic system

The mammalian blood system consists of multiple mature hematopoietic cell types, including erythrocytes, platelets, monocytes, granulocytes, mast cells, dendritic cells, T and Blymphocytes, and natural killer (NK) cells [Seita and Weissman, 2010]. These mature blood cells fulfill essential cellular functions in the mammalian body. For instance, erythrocytes are required for the transport of oxygen and nutrients, while platelets play an important role in wound healing. A variety of immune cells, such as granulocytes, monocytes, dendritic cells, and lymphocytes, are responsible for naïve and adaptive immunity. Given that most mature blood cells are short-lived (hours to a few weeks), the hematopoietic system needs to be continually replenished in a process referred to as hematopoiesis. Approximately, $3x10^5$ erythrocytes and $3x10^4$ leukocytes (white blood cells) are generated per second in an adult human under homeostatic conditions [Takizawa et al., 2012].

1.2.1 Hematopoietic stem cells

Hematopoietic stem cells (HSCs) reside at the apex of the hematopoietic system and are responsible for the production of all blood cells. [Seita and Weissman, 2010]. HSCs are a rare cell population, which represent approximately 0.001% of the bone marrow of an adult mouse [Harrison et al., 1993]. By the classical definition, HSCs have two unique features: first, HSCs are multipotent cells that possess the capability of producing all hematopoietic cell types. Second, in a process called self-renewal HSCs are able to divide and generate novel multipotent HSCs without apparent differentiation. The self-renewal capacity of HSCs is essential for the life-long maintenance of the hematopoietic system [Seita and Weissman, 2010, Haas et al., 2018]. In the mouse, HSCs are immunophenotypically defined as Lin^- $Sca1^+$ $cKit^+$ (LSK) $CD150^+$ $CD48^-$ bone marrow cells [Wilson et al., 2008, Cabezas-Wallscheid et al., 2014].

Under homeostatic conditions, murine HSCs reside in a quiescent state in the bone marrow and infrequently contribute to blood production [Wilson et al., 2008]. *In situ* lineage tracing studies have revealed that steady-state hematopoiesis is predominantly maintained by multipotent progenitor cells (MPPs), which possess restricted self-renewal capacity compared to HSCs [Busch et al., 2015, Sun et al., 2014]. In response to infection, inflammation, or blood loss, dormant HSCs are forced into cycle and actively contribute to the recovery of the hematopoietic system [Essers et al., 2009, Takizawa et al., 2011, Baldridge et al., 2010, Cheshier et al., 2007]. However, exit from dormancy is associated with the production of reactive oxygen species (ROS) that induce DNA damage, consequently resulting in impaired stem cell function [Walter et al., 2015].

For many years, the HSC pool was considered to be a homogenous cell population of multipotent cells with balanced differentiation output. However, comprehensive single-cell transplantation assays and the advent of single-cell transcriptomics have revealed considerable heterogeneity within the HSC pool with respect to their self-renewal capacity and lineage potential [Muller-Sieburg et al., 2004, Dykstra et al., 2007, Sanjuan-Pla et al., 2013, Guo et al., 2013, Yamamoto et al., 2013, Velten et al., 2017]. Subsequent studies have illustrated the prospective isolation of lymphoid-, myeloid-, and megakaryocytic-biased HSCs [Beerman et al., 2010, Challen et al., 2010, Morita et al., 2010, Sanjuan-Pla et al., 2013, Gekas and Graf, 2013]. Nowadays, it is accepted that the majority of HSCs have uni- or bilineage differentiation

potential. Notably, the existence of lineage-biased HSCs has been confirmed under unperturbed conditions, showing that megakaryopoiesis is the predominant fate of HSCs [Rodriguez-Fraticelli et al., 2018]. The lineage bias of individual HSCs can be propagated in serial transplantation assays, suggesting that the observed HSC heterogeneity is not a stochastic phenomenon [Dykstra et al., 2007, Muller-Sieburg et al., 2002, Yu et al., 2017, Challen et al., 2010, Naik et al., 2013], but rather determined by cell-intrinsic epigenetic mechanisms [Yu et al., 2017].

1.2.2 The classical model of the hematopoietic hierarchy

The classical model of hematopoiesis postulates a hierarchical and tree-like bifurcation model with HSCs located at the apex. HSCs give rise to progenitor cell populations in a step-wise differentiation process that is accompanied by a gradual loss of both, multipotency and self-renewal capacity (Figure 1.4) [Seita and Weissman, 2010, Akashi et al., 2000, Kondo et al., 1997].

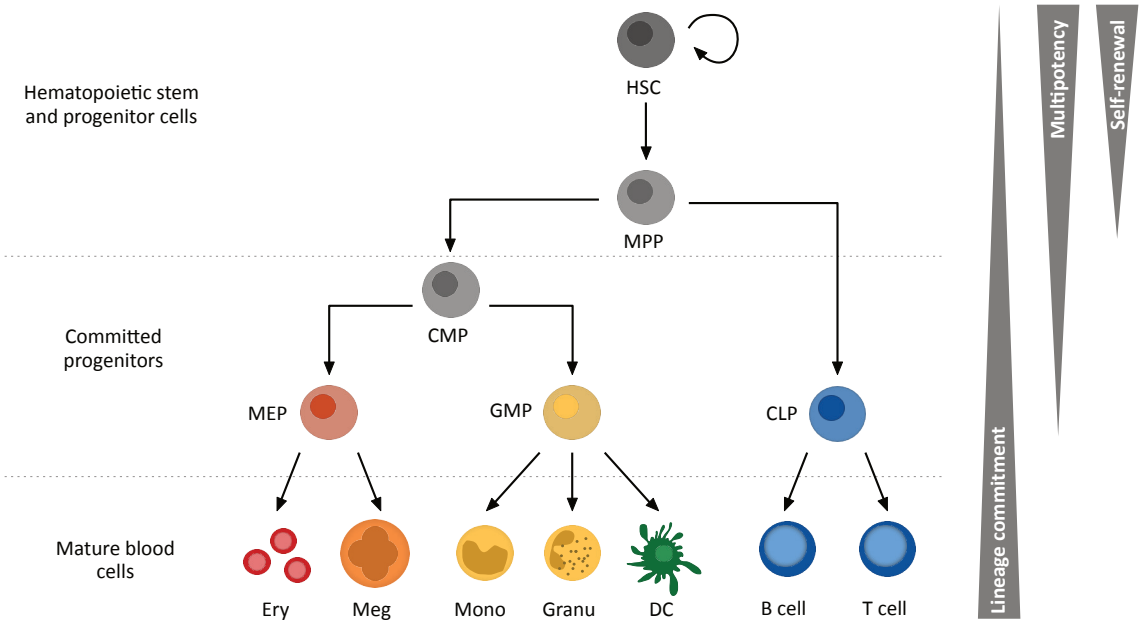

Figure 1.4: The classical hierarchical model of the hematopoietic system. In the classical model of hematopoiesis, HSCs reside at the apex of the hierarchy and give rise to homogenous progenitor cell populations in a step-wise differentiation process (Figure modified from Laurenti *et al.* [Laurenti and Gottgens, 2018]).

Immunophenotypic definitions of hematopoietic cell populations allow the prospective purification of hematopoietic cell types using fluorescence activated cell sorting (FACS). According to the classical view of hematopoiesis, these cell surface marker-defined hematopoietic cell populations are considered to be discrete and homogenous cell populations. Notably, the classical model of hematopoiesis was developed based on these surface marker-defined cell populations and the subsequent functional characterization of their lineage potential using *ex vivo* and *in vivo* assays [Laurenti and Gottgens, 2018].

Within the HSC pool, two distinct HSC subpopulations can be distinguished, which differ in their repopulating capacity. Long-term HSCs (in this thesis referred to as HSCs: LSK $CD150^+$ $CD48^-$ $CD34^-$) reconstitute and sustain the hematopoietic system throughout life, while short-term HSCs (in this thesis referred to as MPP1: LSK $CD150^+$ $CD48^-$ $CD34^+$) maintain the hematopoietic system for only a few weeks post transplantation [Morrison and Weissman, 1994, Morrison et al., 1997, Yang et al., 2005]. Short-term HSCs (or MPP1 cells) further restrict their self-renewal capacity as they differentiate to produce multipotent progenitor cells (MPPs), which show limited to no engraftment in transplantation experiments [Adolfsson et al., 2001].

According to the classical hierarchical model, MPPs differentiate into oligopotent common myeloid progenitors (CMPs) or common lymphoid progenitors (CLP), which is the first lineage commitment step that separates the myelo-erythroid and the lymphoid branches. CMPs can give rise to more restricted oligopotent and bipotent progenitor cells, namely megakaryocyte erythroid progenitors (MEPs) and granulocyte macrophage progenitors (GMPs). These oligopotent progenitor cells further differentiate into unipotent populations that ultimately give rise to mature blood cells, which execute the essential biological functions within the hematopoietic system (Figure 1.4) [Akashi et al., 2000, Kondo et al., 1997, Na Nakorn et al., 2002, Haas et al., 2018, Laurenti and Gottgens, 2018].

1.2.3 Findings that challenged the classical model of hematopoiesis

The implementation of additional surface marker combinations identified hematopoietic subpopulations and suggested several modifications of the classical hierarchical tree of hematopoiesis [Adolfsson et al., 2005, Sanjuan-Pla et al., 2013, Yamamoto

et al., 2013, Pietras et al., 2015]. For instance, the MPP population can be subdivided into MPP2, MPP3, and MPP4 cells [Wilson et al., 2008], which have been shown to exhibit distinct lineage potentials. MPP2 cells show a predominantly megakaryocytic and erythroid differentiation bias, while MPP3 and MPP4 cells display a myeloid and lymphoid differentiation bias, respectively (Figure 1.5) [Pietras et al., 2015, Rodriguez-Fraticelli et al., 2018]. Furthermore, the identification of lymphoid-primed multipotent progenitors (LMPP) revealed that the myeloid and lymphoid branches remain associated further down in the hierarchy [Adolfsson et al., 2005]. In addition, HSCs with megakaryocyte-restricted lineage potential have been identified. These primed HSCs are able to produce megakaryocyte-restricted progenitors in a direct shortcut from HSCs without transitioning through oligopotent progenitor populations [Haas et al., 2015, Notta et al., 2016, Yamamoto et al., 2013, Sanjuan-Pla et al., 2013, Rodriguez-Fraticelli et al., 2018].

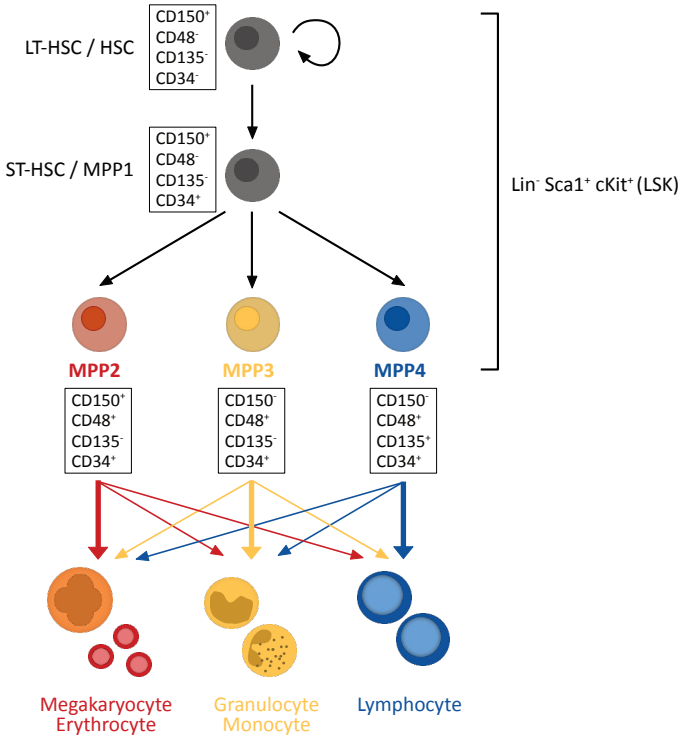

Figure 1.5: Lineage-biased subpopulations of multipotent progenitor cells. Multipotent progenitors (MPPs) can be subdivided into functionally and molecularly distinct subpopulations, namely MPP2, MPP3, and MPP4. MPP2 are erythroid- and megakaryocyte-primed cells. MPP3 are myeloid-primed cells and MPP4 are lymphoid-primed cells (modified from Pietras *et al.* [Pietras et al., 2015]).

1.2.4 The revised continuum model of hematopoiesis

Recent studies based on single-cell transplantation assays, *in situ* lineage tracing in unperturbed hematopoiesis, and single-cell transcriptome profiling have challenged fundamental aspects of the tree-like model and suggested a revised continuum model of hematopoietic commitment [Velten et al., 2017, Paul et al., 2015, Macaulay et al., 2016, Nestorowa et al., 2016, Notta et al., 2016].

In contrast to the homogenous HSC population suggested by the classical model, the revised model takes into account recent findings that the immunophenotypic HSC compartment is functionally heterogeneous and in fact harbors lineage-biased HSCs, indicating early lineage segregation [Dykstra et al., 2007, Morita et al., 2010, Yamamoto et al., 2013, Muller-Sieburg et al., 2004, Velten et al., 2017, Rodriguez-Fraticelli et al., 2018]. Further evidence for early commitment decisions is evident in the oligopotent progenitor populations, including CMPs, GMPs, and MEPs. Based on various methods, these progenitor cell populations have been shown to be functionally and molecularly heterogeneous with unilineage transcriptional profiles [Paul et al., 2015, Notta et al., 2016, Karamitros et al., 2018, Perie et al., 2015, Buenrostro et al., 2018, Rodriguez-Fraticelli et al., 2018]. Importantly, *in situ* lineage tracing confirmed that oligopotent progenitor cell populations have predominantly unilineage output during steady-state hematopoiesis [Rodriguez-Fraticelli et al., 2018]. Collectively, these studies showed that surface marker-defined cell populations are heterogeneous and consequently proposed that lineage commitment occurs early at the level of immunophenotypic HSCs and MPPs.

Transcriptional snapshots of thousands of individual hematopoietic cells enabled the *in silico* reconstruction of differentiation trajectories that were purely based on the similarity of single-cell transcriptomes, and independent of known surface markers. These single-cell RNA sequencing (scRNA-seq) studies revealed that HSCs continuously acquire lineage-committed transcriptional states without passing through discrete progenitor cell populations [Velten et al., 2017, Macaulay et al., 2016, Nestorowa et al., 2016, Karamitros et al., 2018, Giladi et al., 2018] and indicated the presence of early lineage priming in the hematopoietic stem and progenitor cell (HSPC) compartment [Velten et al., 2017, Nestorowa et al., 2016]. Interestingly, the finding of continuous hematopoietic differentiation was consistent across different species including humans, mice, and zebrafish [Velten et al., 2017, Macaulay et al., 2016, Nestorowa et al., 2016, Karamitros et al., 2018, Giladi et al., 2018].

In summary, the revised model of hematopoiesis illustrates a differentiation continuum with lineage-restriction occurring as early as in the immunophenotypic HSPC compartment (Figure 1.6) [Laurenti and Gottgens, 2018, Haas et al., 2018].

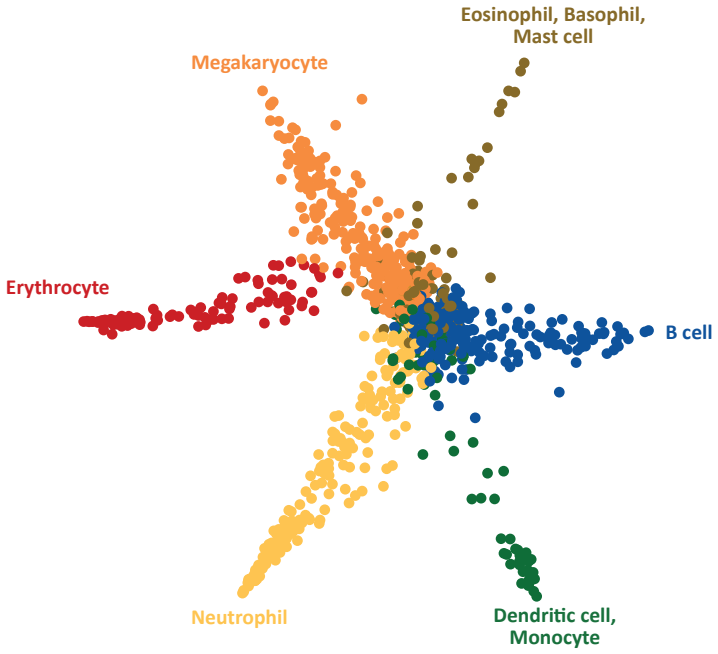

Figure 1.6: The continuum model of hematopoiesis. The continuum model of hematopoiesis indicates a continuous differentiation process with early lineage priming and the presence of unipotent cells within the HSPC compartment. Several representations of the continuum model have been suggested (Laurenti2018). This illustration has been modified from Velten *et al.* They developed STEMNET, a dimensionality reduction algorithm, to visualize the continuous differentiation landscape of human hematopoietic stem and progenitor cells [Velten et al., 2017].

1.3 Epigenetics in hematopoiesis

Hematopoiesis is tightly regulated at both the epigenetic and transcriptional levels. Epigenetic integrity is essential for normal hematopoietic differentiation and defects in the epigenetic machinery have been implicated in hematopoietic abnormalities.

The first experimental proof that DNA methylation plays a fundamental role in the regulation of hematopoietic differentiation came from Dnmt1-hypomorphic mice, which show perturbed multilineage differentiation with impaired lymphoid output [Broske et al., 2009]. Along these lines, conditional inactivation of epigenetic regulators, such as *Dnmt3a*, *Dnmt3b*, and *Tet2*, in murine HSCs resulted in enhanced self-renewal capacity and in impaired differentiation potential [Challen

et al., 2011, Challen et al., 2014, Jeong et al., 2018, Moran-Crusio et al., 2011]. More importantly, in humans, recurrent mutations in epigenetic modifiers, such as *DNMT3A*, *TET2*, and *ASXL1*, have been identified to play a role in the development of age-related clonal hematopoiesis (ARCH) and hematological malignancies, including acute myeloid leukemia (AML) and myelodysplastic syndromes (MDS) [Genovese et al., 2014, Jaiswal et al., 2014, Xie et al., 2014, Ley et al., 2010, Walter et al., 2011, Delhommeau et al., 2009, Steensma et al., 2015].

1.3.1 Epigenome dynamics during hematopoietic differentiation

Early array-based studies demonstrated a high plasticity of the DNA methylome during hematopoiesis and identified characteristic DNA methylation patterns of terminally differentiated blood cell types. The exploration of differential DNA methylation revealed an enrichment of lineage-specific transcription factor binding sites in hypomethylated regions and linked DNA methylation dynamics to gene expression changes. As a result, several genes and pathways have been identified that regulate lymphoid versus myeloid differentiation [Ji et al., 2010, Hodges et al., 2011].

The development of RRBS enabled a more comprehensive DNA methylation map of murine hematopoietic cell populations. This demonstrated that lineage- and cell type-specific DNA methylation patterns were maintained in terminally differentiated cells, suggesting that DNA methylation changes function as epigenetic switches that determine the cellular identity by preventing aberrant transcription factor binding [Bock et al., 2012].

The first genome-wide DNA methylation map of the murine HSPC compartment was generated using TWGBS. This study demonstrated that DNA methylation changes already occur during early hematopoietic commitment. The majority of DMRs were found at distal *cis*-regulatory regions that affect the expression of stemness- and lineage-specific genes, substantiating the role of DNA methylation in lineage specification. Importantly, the majority of DNA methylation changes were unidirectional, either showing a progressive loss or gain of DNA methylation during early hematopoietic differentiation [Cabezas-Wallscheid et al., 2014, Lipka et al., 2014].

Recently, genome-wide DNA methylation profiles of human hematopoietic cells showed not only high DNA methylome plasticity between hematopoietic cells types,

but also revealed variations in DNA methylation patterns dependent on the tissue source. Furthermore, these data have been used to bioinformatically infer a methylome-based model of human hematopoietic differentiation [Farlik et al., 2016], however, none of the currently available human DNA methylation maps covers all hematopoietic lineages.

In line with DNA methylation dynamics, genome-wide mapping of histone modifications using ChIP-seq identified lineage-specific enhancer programs that were enriched for lineage-specific transcription factor binding motifs. The enhancer landscape significantly changed during hematopoietic differentiation and was accompanied by the *de novo* establishment of lineage-specific enhancers [Lara-Astiaso et al., 2014]. Furthermore, genome-wide chromatin accessibility studies using bulk and single-cell ATAC-seq identified cell type-specific open chromatin regions enriched for binding motifs of key hematopoietic transcription factors. In addition, these studies detected epigenetic heterogeneity within surface marker-defined hematopoietic progenitor cell populations [Corces et al., 2016, Buenrostro et al., 2018]. Finally, analysis of promoter-associated genomic architecture of human hematopoietic cells using Hi-C revealed cell type-specific promoter interactomes that reflect known hematopoietic cellular relationships [Javierre et al., 2016].

Collectively, epigenetic changes during hematopoiesis revealed that the epigenome reflects and probably instructs hematopoietic lineage commitment. Various studies imply that epigenetic patterns are more precise in defining the cell identity and the differentiation stage as compared to gene expression profiles [Corces et al., 2016, Bock et al., 2012, Cabezas-Wallscheid et al., 2014, Lipka et al., 2014]. This also indicates that the epigenetic landscape facilitates and coordinates transcriptional responses. Future work integrating recently developed single-cell technologies for epigenomics, transcriptomics, proteomics, and lineage tracing will provide molecular insights into how hematopoietic differentiation is regulated at the single-cell level [Shema et al., 2019, Macaulay et al., 2017].

1.4 Aged hematopoiesis

In the landmark publication 'Hallmarks of Aging', the process of aging was broadly defined as *"the time-dependent functional decline that affects most living organisms"* [Lopez-Otin et al., 2013]. Further to this, aging is characterized by the progressive functional degeneration of many tissues and increased risk of age-related diseases, such as cancer, diabetes, cardiovascular disorders, and neurodegenerative abnormalities. The hallmarks of aging include *"genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication."* [Lopez-Otin et al., 2013]. Furthermore, aging is associated with chronic, low-grade inflammation of several tissues and organs, a phenomenon referred to as 'inflammaging', which contributes to age-related pathologies [Kovtonyuk et al., 2016]. Within regenerating tissues, such as the hematopoietic system, the progressive age-related tissue decline is likely driven by the functional exhaustion of adult stem cells, which is accompanied by abnormalities in the stem cell epigenome [Rossi et al., 2008, Beerman and Rossi, 2014].

1.4.1 DNA methylation changes associated with aging

The epigenetic landscape of a cell gradually shifts during aging [Michalak et al., 2019]. Initially, global DNA hypomethylation was suggested as a general hallmark of aging in various tissues [Wilson and Jones, 1983]. However, recent studies have shown that aging is associated with both DNA methylation loss in heterochromatin regions and locus-specific gain of DNA methylation in promoter regions [Unnikrishnan et al., 2018, Beerman et al., 2013, Rakyan et al., 2010, Teschendorff et al., 2010]. The so-called epigenetic drift likely establishes an age-related transcriptional response, causing the onset of aging phenotypes [Hannum et al., 2013, Beerman et al., 2013].

Based on aging-associated epigenetic drift, a number of DNA methylation clocks have been described in order to estimate the chronological age of an organism. These single-tissue or multi-tissue age estimators are based on DNA methylation values of a few specific CpG sites, ranging from 3 to 353 CpGs in murine and human methylation clocks [Bocklandt et al., 2011, Garagnani et al., 2012, Hannum et al., 2013, Horvath, 2013, Weidner et al., 2014, Petkovich et al., 2017, Stubbs et al., 2017, Wang et al., 2017].

1.4.2 Hematopoietic stem cell aging

The hematopoietic system experiences functional and molecular changes during aging [Akunuru and Geiger, 2016, Choudry and Frontini, 2016]. In humans, aged hematopoiesis is clinically characterized by peripheral blood cytopenias, decreased bone marrow cellularity, immune dysfunction, increased incidence of thrombosis, and clonal hematopoiesis with an increased risk for myeloid malignancies [Choudry and Frontini, 2016]. Age-related clonal hematopoiesis (ARCH) is defined as the expansion of one or more HSC clones that sustain the production of mature blood cells, preferentially of myeloid cells, in elderly people. ARCH is further characterized by the acquisition of somatic mutations in HSCs. These mutations frequently affect known oncogenic drivers in AML and MDS, including epigenetic modifiers such as *DNMT3A*, *TET2*, and *ASXL1* [Genovese et al., 2014, Jaiswal et al., 2014, Xie et al., 2014, Ley et al., 2010, Walter et al., 2011, Delhommeau et al., 2009, Steensma et al., 2015]. It is possible that many of these aging-related phenotypes result from functionally impaired cells within the aged HSC pool [Rossi et al., 2008].

Aging of HSCs has been extensively investigated in laboratory mice, identifying profound changes in the HSC self-renewal capacity and differentiation potential [Dykstra et al., 2011, Sudo et al., 2000, Rossi et al., 2005, Akunuru and Geiger, 2016]. The self-renewal capacity of aged HSCs is decreased compared to their young counterparts as shown by reduced reconstitution in serial transplantation studies. The self-renewal capacity of HSCs appears to be inversely correlated with the replication history of HSCs [Dykstra et al., 2011, Sudo et al., 2000, Bernitz et al., 2016]. Furthermore, immunophenotypic HSCs significantly expand over time [Sudo et al., 2000], which is accompanied by alterations in the composition of the HSC pool. Accordingly, HSC aging is associated with an increase in myeloid- and megakaryocytic-biased HSCs causing enhanced myeloid and megakaryocytic lineage output at the expense of lymphoid cells (Figure 1.7) [Beerman et al., 2010, Challen et al., 2010, Dykstra et al., 2011, Gekas and Graf, 2013, Grover et al., 2016, Rundberg Nilsson et al., 2016]. In line with this, a recent study tracking label retention of HSCs *in vivo* demonstrated that phenotypic HSCs that have proliferated and lost label show a myeloid- and megakaryocytic-restricted lineage potential [Bernitz et al., 2016]. Similar to mice, human immunophenotypic HSCs expand with age and become functionally compromised showing a myeloid bias in xenotransplantation studies [Pang et al., 2011, Rundberg Nilsson et al., 2016].

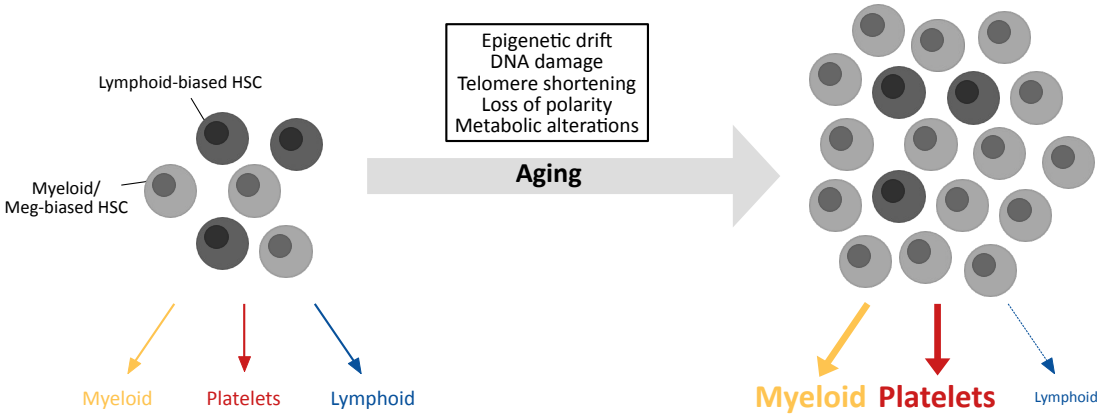

Figure 1.7: Hematopoietic stem cell aging. Aging of the HSC pool is associated with decreased self-renewal capacity and an expansion of myeloid- and megakaryocyte-biased HSCs leading to a skewed myeloid and megakaryocytic lineage outcome in aged hematopoiesis. Molecular mechanisms that drive the aging phenotype include changes in the epigenetic landscape ('epigenetic drift'), accumulation of DNA damage, telomere shortening, and loss of the polar distributions of cytoskeletal proteins and epigenetic markers.

Several molecular mechanisms have been identified to be involved in the functional decline of HSCs, including accumulation of DNA damage, telomere shortening, decreased cellular polarity, and metabolic alterations (Figure 1.7) [Akunuru and Geiger, 2016, Florian et al., 2012, Beerman et al., 2014]. In addition, HSC aging is associated with a remarkable epigenetic drift, including changes in the DNA methylome and altered histone modifications [Choudry and Frontini, 2016, Beerman and Rossi, 2014]. Of note, reprogramming of HSCs via induced pluripotent stem cells rejuvenated aged murine HSCs, suggesting that the impaired function of aged HSCs is primarily driven by the epigenetic drift [Wahlestedt et al., 2017]. Indeed, studies investigating the DNA methylome of young and aged HSCs identified locus-specific alterations, which affected genomic regions associated with hematopoietic lineage potential. Specifically, hypermethylated DMRs overlapped with genes involved in lymphoid and erythroid lineage specification [Beerman et al., 2013, Taiwo et al., 2013, Sun et al., 2014]. Recently, changes in the epigenetic landscape of aged human HSCs have been detected, which mainly affect developmental and cancer pathways [Adelman et al., 2019]. Despite the evidence demonstrating an epigenetic drift in aged HSCs, the molecular mechanisms by which these epigenetic alterations cause the functional decline and lineage bias of aged HSCs remains poorly understood.

Overall, DNA methylation dynamics play a fundamental role during hematopoietic differentiation, however, it is not fully understood how DNA methylation changes are globally programmed throughout the entire hematopoietic system. In this doctoral thesis, whole-genome bisulfite sequencing was used to generate a DNA methylome map of murine hematopoiesis, which provided novel insights into DNA methylation programming during hematopoietic differentiation. The understanding of how normal hematopoiesis is regulated at the epigenetic level is the basis for the investigation of altered DNA methylation patterns in aberrant hematopoiesis.

2

Aims of the Thesis

The classical hierarchical model of hematopoiesis has been modified based on recent single-cell transcriptome analyses, which indicate a continuous rather than a stepwise differentiation process. While differentiation trajectories can be inferred from single-cell transcriptomic snapshots of differentiating cells [Velten et al., 2017, Giladi et al., 2018, Nestorowa et al., 2016, Dahlin et al., 2018, Macaulay et al., 2016], it remains a challenge to identify definitive points of commitment along these trajectories. However, the identification of such molecular commitment decisions is essential to accurately model the hematopoietic system.

5-Methylcytosine is a stable epigenetic mark that upon cell division is inherited by the daughter cells in the form of an epigenetic memory, but at the same time, is dynamically regulated during the process of cellular differentiation [Kim and Costello, 2017]. Whole genome DNA methylation data jointly published by the laboratories of my supervisors Daniel Lipka and Michael Milsom revealed progressive epigenetic changes during early hematopoietic commitment, showing either a continuous loss or gain of methylation from hematopoietic stem cells (HSCs) to the most differentiated multipotent progenitor populations (MPPs) [Cabezas-Wallscheid et al., 2014, Lipka et al., 2014]. These early methylation changes were conserved in mature blood organs like the bone marrow, spleen, or thymus, but non-hematopoietic tissues and immature murine embryonic stem cells lacked this epigenetic pattern [Lipka et al., 2014].

Based on the observed DNA methylation dynamics during early hematopoiesis [Cabezas-Wallscheid et al., 2014, Lipka et al., 2014], we hypothesize that DNA methylation dynamics are progressive and unidirectional during the process of hemato-

poietic differentiation. The progressive nature of DNA methylation during hematopoietic differentiation, in combination with an epigenetic memory, seem to be unique features of this epigenetic mark that might be beneficial for the identification of definitive molecular commitment points and the deconvolution of hematopoietic differentiation trajectories. The aim of this doctoral thesis was to generate the first genome-wide DNA methylation map of the murine hematopoietic system at single-CpG resolution using tagmentation-based whole genome bisulfite sequencing (TWGBS).

The primary objectives of this doctoral thesis were:

1. Establish sorting schemes for hematopoietic cell populations
2. Sort hematopoietic cell populations at high purity and generate TWGBS libraries
3. Identification of cell type- and lineage-specific DNA methylation programs
4. Validation of the progressivity and unidirectionality concept of DNA methylation dynamics during hematopoietic differentiation
5. Model the hematopoietic system based on DNA methylation dynamics
6. Correlate identified DNA methylation programs with gene expression profiles, binding motifs of hematopoietic transcription factors, and published hematopoietic enhancer programs
7. Investigate DNA methylation reprogramming events during HSC aging in the context of the murine DNA methylation map of normal hematopoiesis

Collectively, the resulting DNA methylation map of murine hematopoiesis will enhance our knowledge about how DNA methylation is programmed during hematopoietic lineage commitment and will function as a rich resource to study aberrant DNA methylation patterns in perturbed hematopoiesis.

3
Results

3.1 Generation of a genome-wide DNA methylation map of murine hematopoiesis

Hematopoietic differentiation is tightly regulated by epigenetic changes that reflect hematopoietic lineage specification. Several studies of immunophenotypically defined hematopoietic cell populations demonstrated high plasticity of the DNA methylome during hematopoietic differentiation and identified DNA methylation patterns that are characteristic for hematopoietic lineages and cell types [Farlik et al., 2016, Lipka et al., 2014, Cabezas-Wallscheid et al., 2014, Bock et al., 2012, Hodges et al., 2011, Ji et al., 2010]. Moreover, tagmentation-based whole genome bisulfite sequencing (TWGBS) enabled the first genome-wide DNA methylation map of early hematopoiesis. This study revealed early DNA methylation dynamics at the hematopoietic stem and progenitor cell (HSPC) level and suggested that hematopoietic differentiation is accompanied by progressive and unidirectional DNA methylation dynamics [Lipka et al., 2014, Cabezas-Wallscheid et al., 2014].

As a continuation of this project, we applied TWGBS to generate a genome-wide DNA methylation map of the entire hematopoietic system. This map encompasses 26 surface marker-defined hematopoietic cell populations comprising all blood lineages, including HSPC populations (subpopulations of the Lin^- $Sca1^+$ $cKit^+$ (LSK) compartment), committed progenitors, and terminally differentiated cell types (Figure 3.1 a). The hematopoietic cell populations were isolated from the bone marrow or spleen of 8-12 weeks-old C57BL/6J mice by fluorescence-activated cell sorting

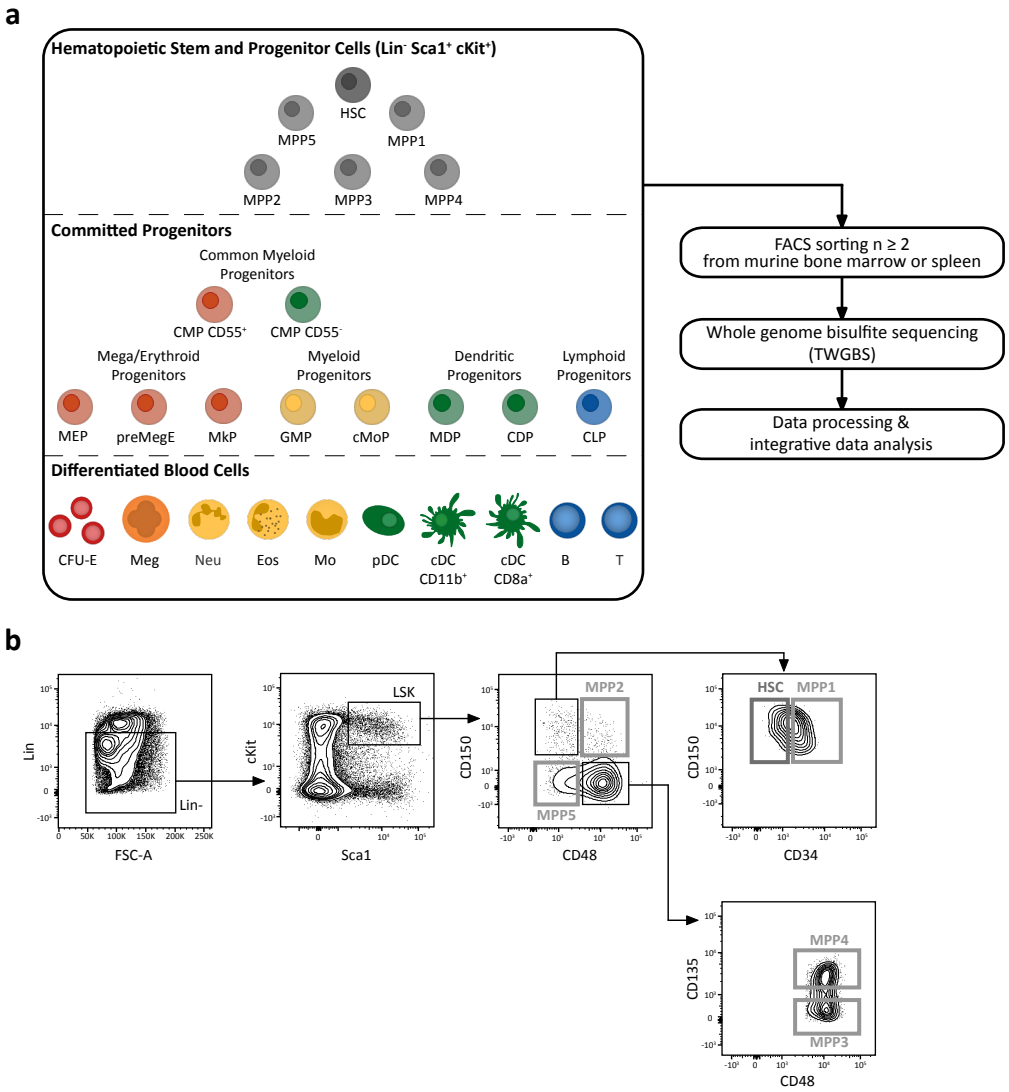

Figure 3.1: Generation of a genome-wide DNA methylation map of the murine hematopoietic system. **a**, Hematopoietic cell populations were isolated by FACS from the bone marrow or spleen of 8-12 weeks-old C57BL/6J mice, including hematopoietic stem and progenitor cell populations, committed progenitors, and differentiated blood cells. In total, 26 cell populations were sorted from pools of mice in at least biological duplicates (in collaboration with Dr. Ruzhica Bogeska and Dr. Melinda Czeh) and subjected to TWGBS, followed by data processing and integrative data analysis. HSC, hematopoietic stem cell; MPP, multipotent progenitor population; MEP, megakaryocyte erythroid progenitor; preMegE, pre-megakaryocyte-erythroid progenitor; MkP, megakaryocyte progenitor; GMP, granulocyte macrophage progenitor; cMoP, common monocyte progenitor; MDP, monocyte-dendritic cell progenitor; CDP, common dendritic cell progenitor; CLP, common lymphoid progenitor; CFU-E, colony forming unit-erythroid; Meg, megakaryocyte; Neu, neutrophil; Eos, eosinophil; Mo, monocyte; pDC, plasmacytoid dendritic cell; cDC, conventional dendritic cell; B, B cell; T, T cell. **b**, Representative sorting scheme for the isolation of primary HSPC populations (Lin$^-$ Sca1$^+$ cKit$^+$): HSC (CD150$^+$ CD48$^-$ CD135$^-$ CD34$^-$), MPP1 (CD150$^+$ CD48$^-$ CD135$^-$ CD34$^+$), MPP5 (CD150$^-$ CD48$^-$ CD135$^-$ CD34$^+$), MPP2 (CD150$^+$ CD48$^+$ CD135$^-$ CD34$^+$), MPP3 (CD150$^-$ CD48$^+$ CD135$^-$ CD34$^+$), MPP4 (CD150$^-$ CD48$^+$ CD135$^+$ CD34$^+$). Sorting schemes for committed progenitors and differentiated blood cells are shown in Figure A.1 and A.2, respectively.

(FACS) using antibody panels targeting cell type-specific surface marker proteins (Figure 3.1 b, Figure A.1 and A.2). As part of this doctoral thesis, several sorting schemes for hematopoietic cell populations were established.

For each hematopoietic cell population, at least two biological replicates were isolated from a pool of mice (in collaboration with Dr. Ruzhica Bogeska and Dr. Melinda Czeh) and subjected to library preparation for TWGBS, followed by data processing and integrative data analysis (Figure 3.1 a). The purity of FACS-sorted cell populations was verified by a re-sort of purified cells. Sorted populations with a purity less than 80% were excluded subsequent TWGBS.

Of note, the CMP compartment was further separated into two functionally and transcriptionally distinct subpopulations based on the expression of the surface marker CD55, with CMP $CD55^-$ cells predominantly producing myeloid cells and CMP $CD55^+$ cells mainly generating erythroid and megakaryocytic cells (Figure 3.1 a) [Guo et al., 2013].

Overall, TWGBS provided robust DNA methylation data with a merged genome-wide CpG coverage ranging from 13x to 78x per hematopoietic cell population and a bisulfite conversion rate >99.9%, as determined by CH methylation levels of Chromosome 1 (Figure 3.2 a, Table A.1). Subopulations of the hematopoietic stem and progenitor cell compartment revealed the highest global DNA methylation levels of 81% to 82%, which gradually decreased during differentiation in all blood lineages. Consistent with previous studies [Shearstone et al., 2011], cell types of the erythroid and megakaryocytic lineages showed the most prominent genome-wide loss of DNA methylation (>10%) as compared to HSCs (Figure 3.2 b, Table A.1).

Importantly, principal component analysis (PCA) based on DNA methylation levels measured in multicellular 'Ensembl Regulatory Features' (promoters, promoter flanking regions, enhancers, CTCF binding sites, transcription factor binding sites, and open chromatin regions) demonstrated that DNA methylation profiles cluster predominantly by cell type, indicating high reproducibility between biological replicates. The PCA also reflected known hematopoietic cellular relationships, indicating that methylation dynamics in regulatory regions might be critical for cell type specification. Specifically, principal component (PC) 1 (26.1%) separated hematopoietic populations based on their cellular differentiation status (immature vs. mature cell types), while PC2 (13.0%) separated the different blood lineages from each other (Figure 3.2 c).

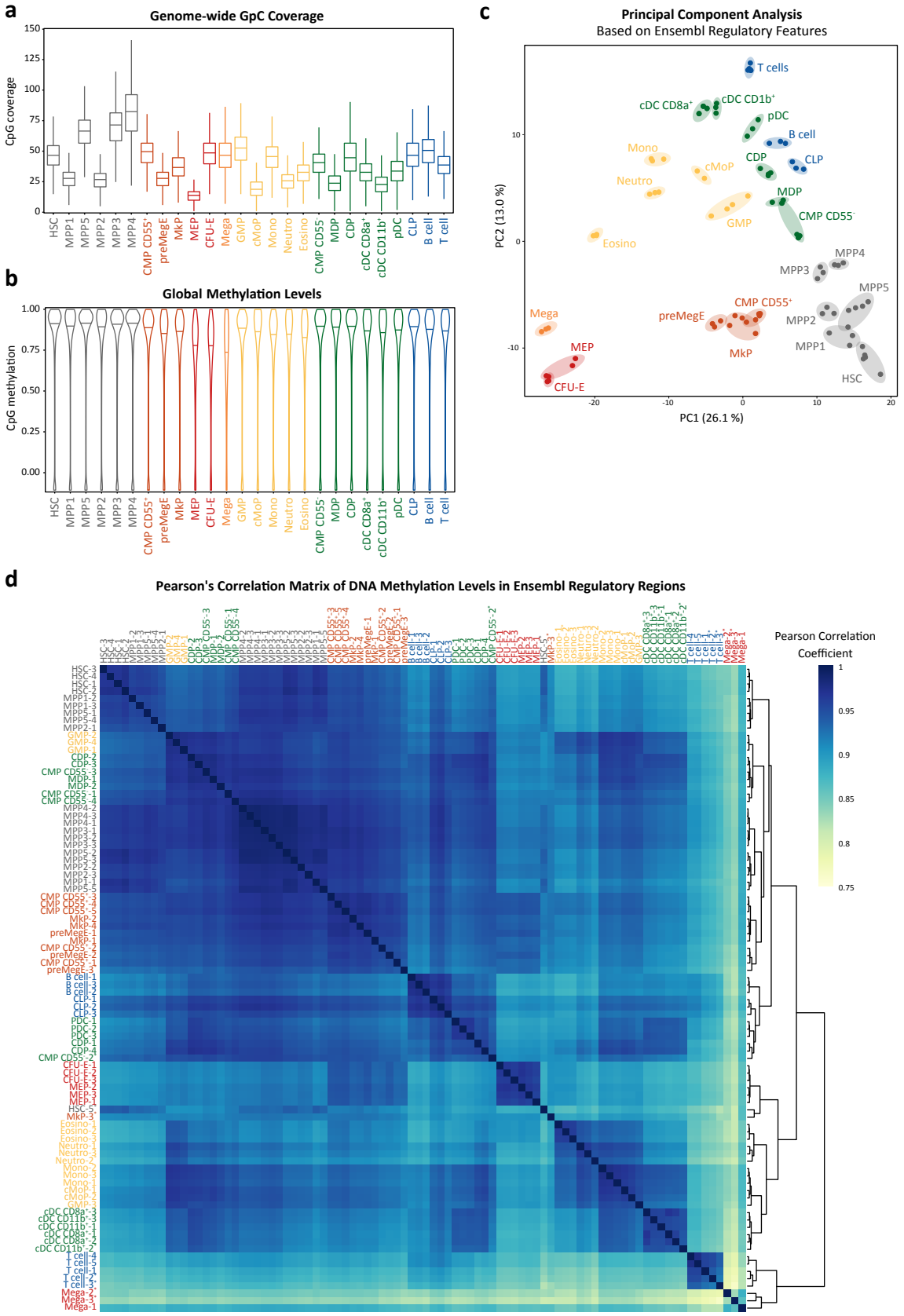

Figure 3.2: Quality control of hematopoietic DNA methylation profiles.

Figure 3.2: Quality control of hematopoietic DNA methylation profiles. TWGBS was performed to generate DNA methylation profiles of 26 hematopoietic cell types (n $\geq$ 2). **a**, Boxplot showing the median genome-wide CpG coverage of 1x10^6 randomly sampled CpG sites per cell population. **b**, Violin plot showing the distribution of CpG methylation levels across 1x10^6 randomly sampled CpG sites per cell population. **c**, PCA based on DNA methylation levels in 'Ensembl Regulatory Features' (promoters, promoter flanking regions, enhancers, CTCF binding sites, transcription factor binding sites, open chromatin regions). Each dot represents a biological replicate. **d**, Pearson's correlation matrix of biological replicates based on DNA methylation levels in 'Ensembl Regulatory Features'. Rows and columns of the Pearson's correlation matrix were arranged by Ward's hierarchical agglomerative clustering. Replicates marked with asterisks indicate low CpG coverage (i.e. <5).

The reproducibility between biological replicates was further assessed by calculating the Pearson's correlation coefficient r between all replicates based on DNA methylation values in 'Ensembl Regulatory Features'. The Pearson's r was >0.95 for the majority of the analyzed cell types, with the exception of a few biological replicates with low CpG coverage (i.e. <5). Clustering of the Pearson's correlation matrix confirmed the high concordance of DNA methylation measurements between biological replicates (Figure 3.2 d).

Taken together, we generated high quality DNA methylation profiles of 26 murine hematopoietic cells types, which can be used to investigate methylation dynamics during hematopoietic differentiation.

3.2 Hierarchical clustering of differentially methylated regions identifies cell type- and lineage-specific DNA methylation programs

In order to detect genomic regions that are dynamically regulated during the process of hematopoietic differentiation, differentially methylated regions (DMRs) were identified by pairwise comparison of HSCs versus all downstream hematopoietic cell populations (DMR calling was performed by Stephen Kraemer). As expected, the number of DMRs correlated with the differentiation status of a hematopoietic cell population. Early commitment steps within the HSPC compartment were associated with relatively low numbers of DMRs (32 to 6,496 DMRs). In contrast, committed progenitors and terminally differentiated cell types showed a gradual increase of DMRs, ranging from 7,000 DMRs in MkPs to 46,510 DMRs in the CFU-E popula-

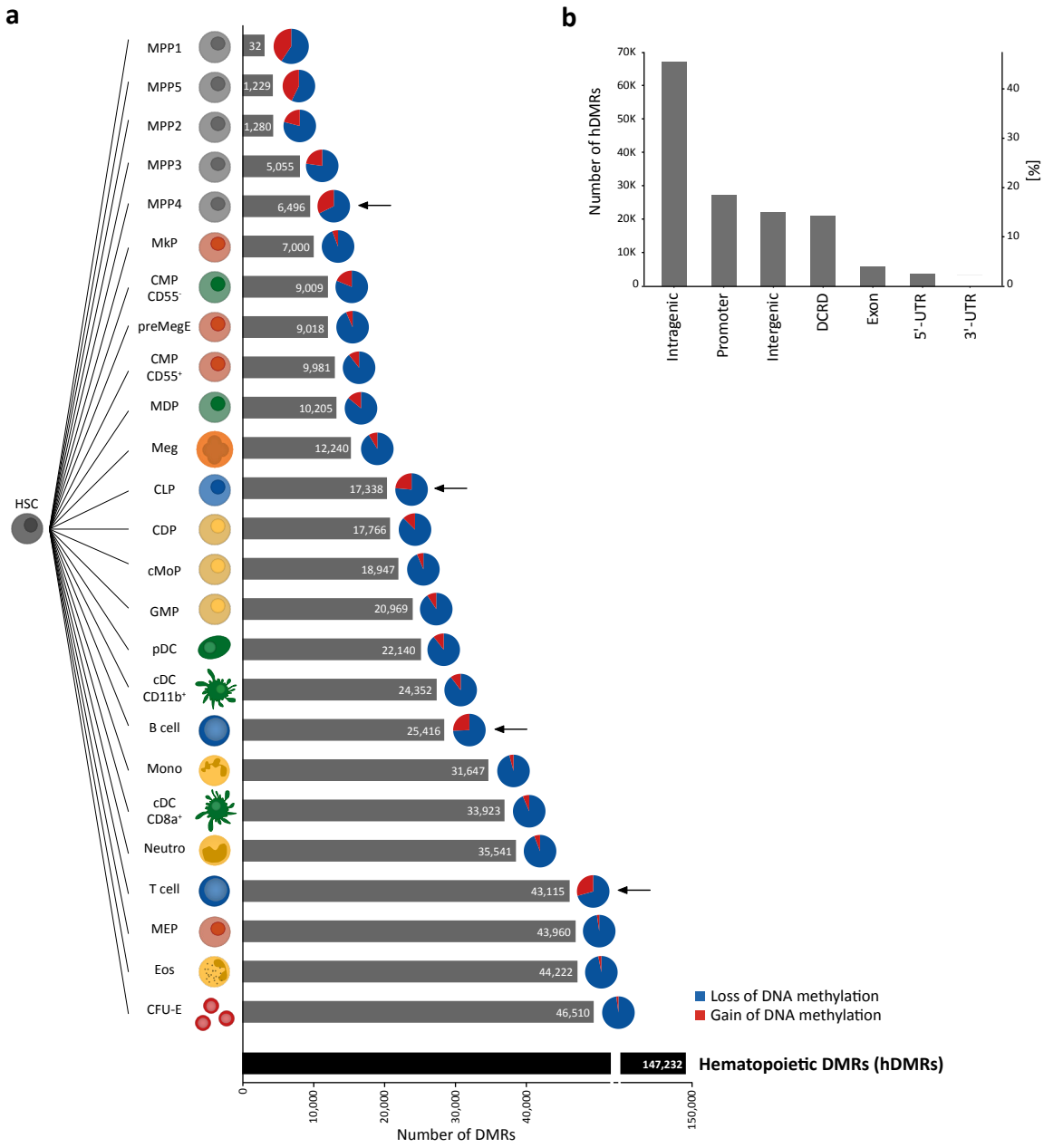

Figure 3.3: DNA methylation changes during hematopoietic differentiation are predominantly associated with loss of DNA methylation. a, DMRs were identified in pairwise comparisons between HSCs and each of the hematopoietic cell populations (DMR calling was done by Stephen Kraemer). Bar plot showing the number of DMRs for the individual pairwise comparisons. The pie charts illustrate the gain (red) or loss (blue) of DNA methylation. Arrows highlight the higher number of gain of methylation DMRs in lymphoid cell populations compared to other blood lineages. By merging DMRs resulting from the individual pairwise comparisons, 147,232 unique DMRs ('hematopoietic DMRs'; hDMRs) were identified throughout the hematopoietic system. **b**, Bar plot showing the genomic regions annotation of the hDMRs. DCRD, distant *cis*-regulatory domain; UTR, untranslated region.

tion (Figure 3.3 a). The majority of DNA methylation changes during hematopoietic differentiation are characterized by a loss of DNA methylation. In line with previous studies [Broske et al., 2009, Bock et al., 2012, Farlik et al., 2016], cell populations of the lymphoid lineage, including CLPs, B and T cells, revealed more DMRs associated with a gain of DNA methylation compared to other blood lineages. Strikingly, the lymphoid-specific gain of DNA methylation seems to be already detectable in the lymphoid-primed MPP4 population (Figure 3.3 a).

By merging DMRs resulting from the individual pairwise comparisons, we identified 147,232 unique DMRs throughout the hematopoietic system, which we designated as so-called 'hematopoietic DMRs' (hDMRs) (Figure 3.3 a). The majority of hDMRs are located in intragenic regions (45.6%), followed by promoter regions (18.6%), intergenic regions (15.0%), distal *cis*-regulatory domains (DCRD, 50,000 - 5,000 upstream of TSS) (14.2%), exons (4.0%), and finally 5'-UTRs (2.5%). No hDMRs were found in the 3'-UTRs (Figure 3.3 b).

Unsupervised hierarchical clustering of the 147,232 hDMRs using z-score transformed methylation beta-values identified 29 clusters of hDMRs that revealed cell type- or lineage-specific methylation programs (in collaboration with Stephen Kraemer). DNA methylation clusters with similar methylation dynamics were fused to eight cell type- or lineage-specific methylation cluster sets, namely erythroid, HSPC, T lymphoid, megakaryocytic, B lymphoid, dendritic, pan-hematopoietic, and myeloid cluster sets (Figure 3.4 a). In general, the methylation programs were defined based on the cell type- or lineage-specific loss of DNA methylation. For instance, the B lymphoid cluster 13 shows a specific DNA methylation loss in B cells. Notably, the B lymphoid-specific demethylation is already present in the CLP population (Figure 3.4 a). This loss of DNA methylation is further highlighted by beta-value deltas calculated in relation to HSCs (Figure 3.4 b). Similarly, the erythroid, T lymphoid, megakaryocytic, dendritic, and myeloid cluster sets comprise hDMRs with lineage-specific loss of DNA methylation in the respective blood cell lineages. In addition, the HSPC-specific cluster set is characterized by hDMRs with low methylation levels in the HSPC populations, while the pan-hematopoietic clusters are comprised of hDMRs, which exhibit high DNA methylation levels in HSCs and low DNA methylation levels in the majority of the more differentiated hematopoietic cell populations (Figure 3.4 a, b).

The number of hDMRs per cluster varied from 936 hDMRs in the pan-hematopoietic

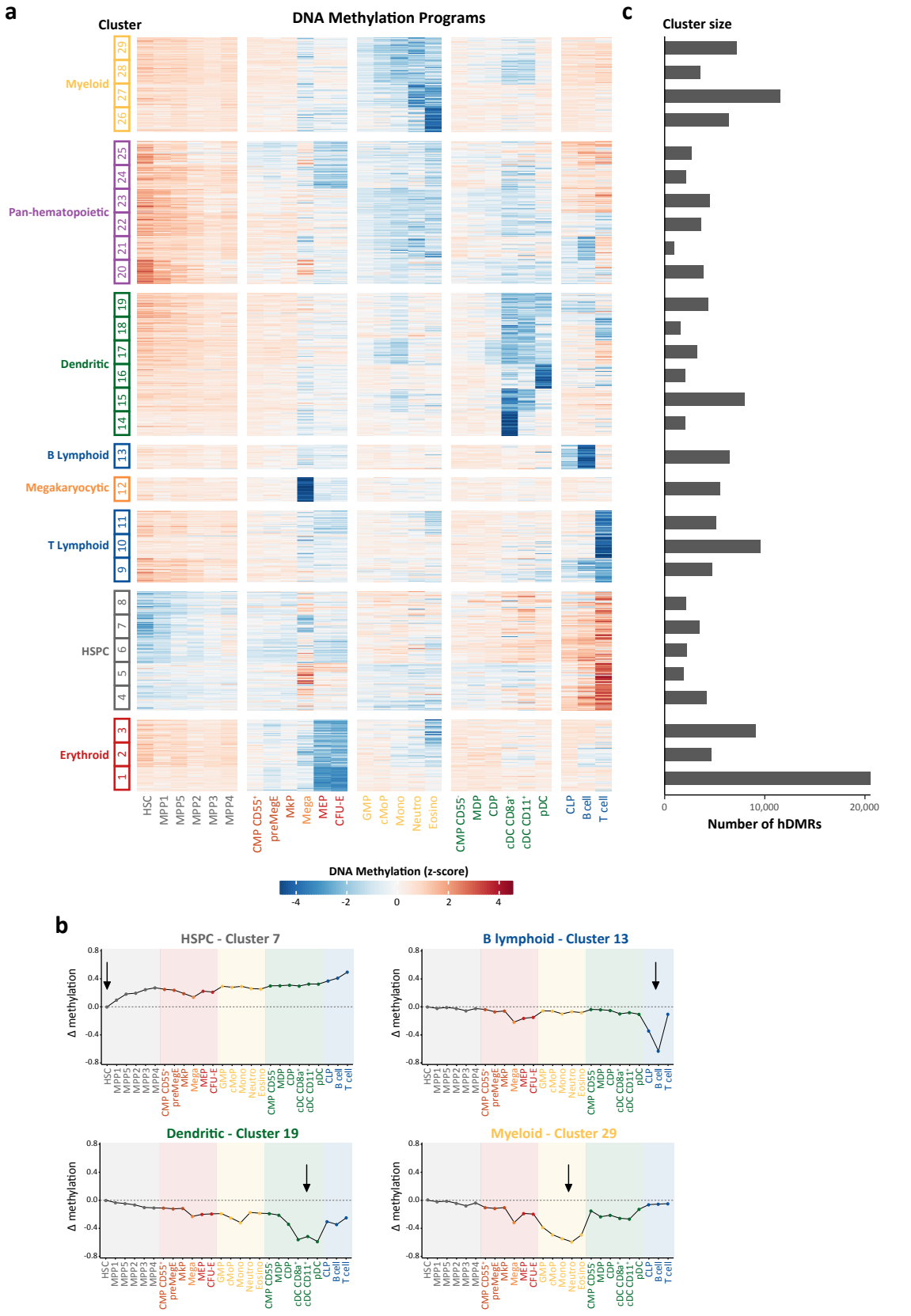

Figure 3.4: Clustering of hematopoietic DMRs identifies cell type- and lineage-specific DNA methylation programs.

Figure 3.4: Clustering of hematopoietic DMRs identifies cell type- and lineage-specific DNA methylation programs. **a**, Heatmap showing hDMRs clustered with Ward's hierarchical agglomerative clustering method using z-score transformed beta-values (in collaboration with Stephen Kraemer). In total, 29 distinct cell type- and lineage-specific DNA methylation programs have been identified, which can be further combined to eight cluster sets. Annotation of clusters and cluster sets are shown on the left. Annotation of the analyzed cell populations is shown at the bottom. **b**, Line plots depicting the beta-value deltas calculated in relation to HSCs for the following DNA methylation programs: HSPC cluster 7, B lymphoid cluster 13, dendritic cluster 19, and myeloid cluster 29. **c**, Bar plot showing the number of hDMRs per DNA methylation cluster.

cluster 21 to 20,552 hDMRs in the erythroid cluster 1 (Figure 3.4 c). Erythropoiesis was associated with the broadest DNA methylation programming, as demonstrated by the high number of hDMRs in the erythroid-specific methylation clusters 1-3 (34,257 hDMRs). In addition, erythroid cell populations (MEP and CFU-E) exhibited a rather unspecific DNA methylation loss in nearly all other DNA methylation clusters (Figure 3.4 a-c).

Given the identification of cell type- and lineage-specific DNA methylation programs, we first assessed whether genomic regions surrounding TSS of well-characterized hematopoietic genes are dynamically regulated in the analyzed hematopoietic cell populations. Indeed, a number of hDMRs overlapped with TSS of key hematopoietic regulators (Lara-Astiaso2014, Giladi2018). DNA methylation changes at such hDMRs were defined by the gradual loss of DNA methylation in the respective lineage, either showing a continuous broadening or deepening of the DNA methylation valley (Figure 3.5 a-d). The enzyme myeloperoxidase (Mpo) is synthesized during myeloid differentiation. Accordingly, the TSS of *Mpo* overlapped with a hDMR of the myeloid cluster 29, which showed decreasing DNA methylation levels and increased in width from HSCs over GMPs towards monocytes and neutrophils. In contrast, cell populations from other lineages revealed only a minor or no DNA methylation loss (Figure 3.5 a). Similarly, the TSS of *Klf1*, a transcription factor important for erythropoiesis, overlapped with a hDMR of the erythroid cluster 1. The demethylation of this erythroid-specific hDMR was initiated in the erythroid precursor populations (CMP $CD55^{+}$, preMegE) and peaked in the more mature erythroid cell populations (MEP, CFU-E) (Figure 3.5 b). As another example, a hDMR of the dendritic cluster 17, which is in close proximity to the TSS of the transcription factor *Irf8*, showed a specific DNA methylation loss in monocytes and dendritic cells, consistent with its function in the corresponding differentiation trajectories (Figure 3.5 c). Finally,

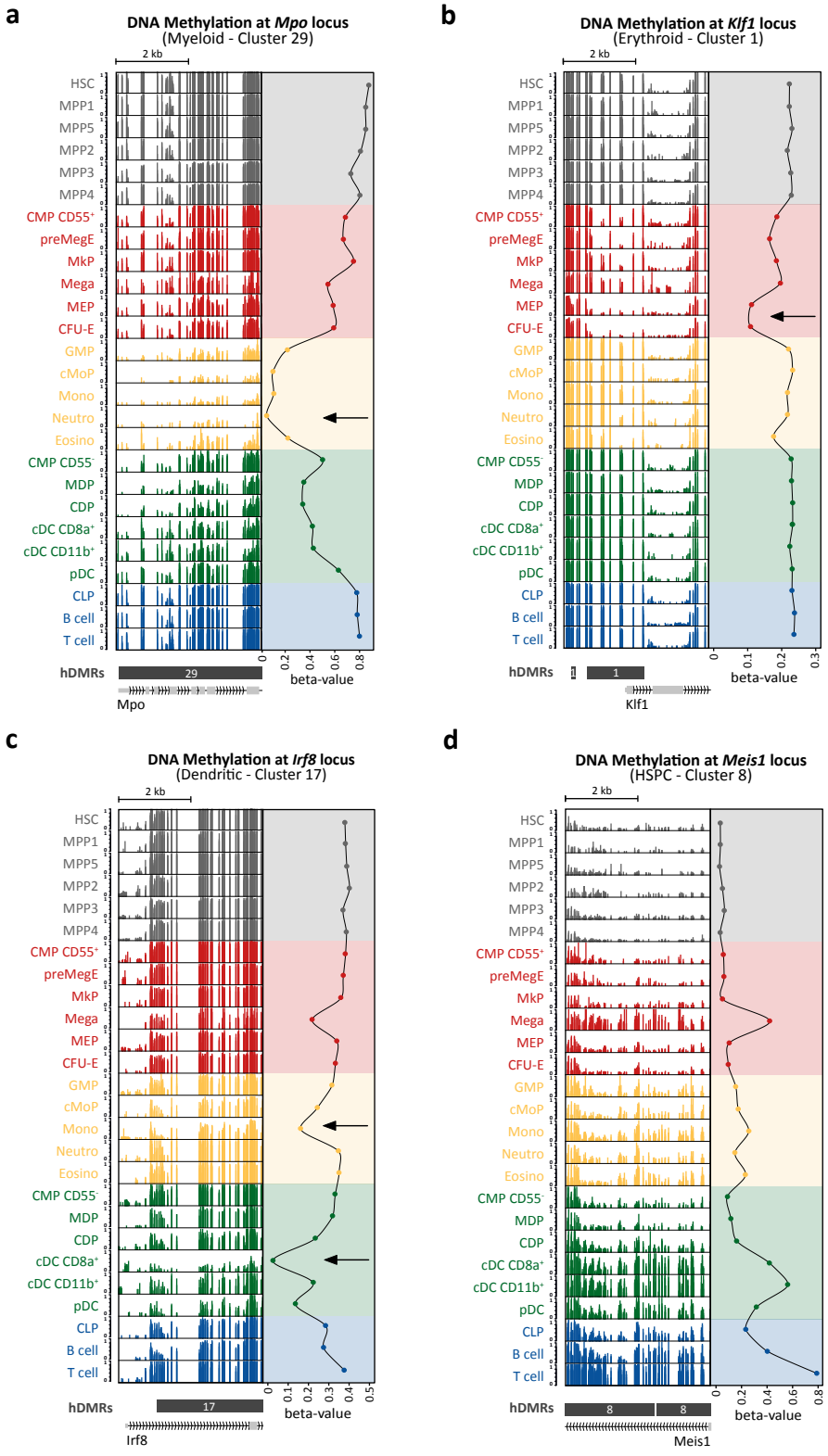

Figure 3.5: Dynamic DNA methylation changes at transcription start sites of known hematopoietic regulators overlap with hDMRs of cell type- and lineage-specific DNA methylation programs.

Figure 3.5: Dynamic DNA methylation changes at transcription start sites of known hematopoietic regulators overlap with hDMRs of cell type- and lineage-specific DNA methylation programs. Genome browser track snapshots depicting DNA methylation levels at TSS of known hematopoietic regulators in 26 hematopoietic cell populations: *Mpo* **(a)**, *Klf1* **(b)**, *Irf8* **(c)**, and *Meis1* **(d)**. DNA methylation levels of individual CpGs are shown on the left hand side of each panel. Aggregated beta-values per genomic window (5 kb) are depicted in the adjacent line plots. hDMRs and genes are annotated at the bottom.

two hDMRs of the HSPC cluster 8, which are located at the TSS of *Meis1*, a key regulator of stem cell maintenance, showed a gradual gain of DNA methylation from the HSPCs over the progenitor populations to the terminally differentiated cell populations of all hematopoietic lineages (Figure 3.5 d).

To further assess the biological relevance of the identified DNA methylation programs, the 'Genomic Regions Enrichment of Annotations Tool' (GREAT) was used to annotate hDMRs to nearby genes and to predict biological functions of the individual DNA methylation programs [McLean et al., 2010]. Using GREAT, we found that hDMRs are preferentially located near genes associated with lineage-specific hematopoietic cell function. As a result, Gene Ontology (GO) Biological Process terms that correlate with hematopoietic lineage specification and function were enriched in the corresponding DNA methylation clusters. Accordingly, the myeloid DNA methylation cluster 29 was enriched for GO terms associated with immune response of myeloid cells against pathogens, while the T lymphoid cluster 9 was enriched for GO terms related to leukocyte activation and differentiation (Figure 3.6 a, b). Moreover, the HSPC-specific DNA methylation cluster 6 correlated with GO terms of immune system development, homeostasis of cell number, and negative regulation of metabolic processes (Figure 3.6 c). In summary, with the exception of the erythroid cluster 2 and the megakaryocytic cluster 12, which did not show any enrichment, the identified DNA methylation clusters were enriched for GO terms characteristic of the respective lineage specification (Figure A.3).

Collectively, hematopoietic differentiation is characterized by DNA methylation changes that reflect the cellular differentiation status. These DNA methylation changes can be clustered into cell type- and lineage-specific DNA methylation programs that seem to be essential for lineage specification and cellular function.

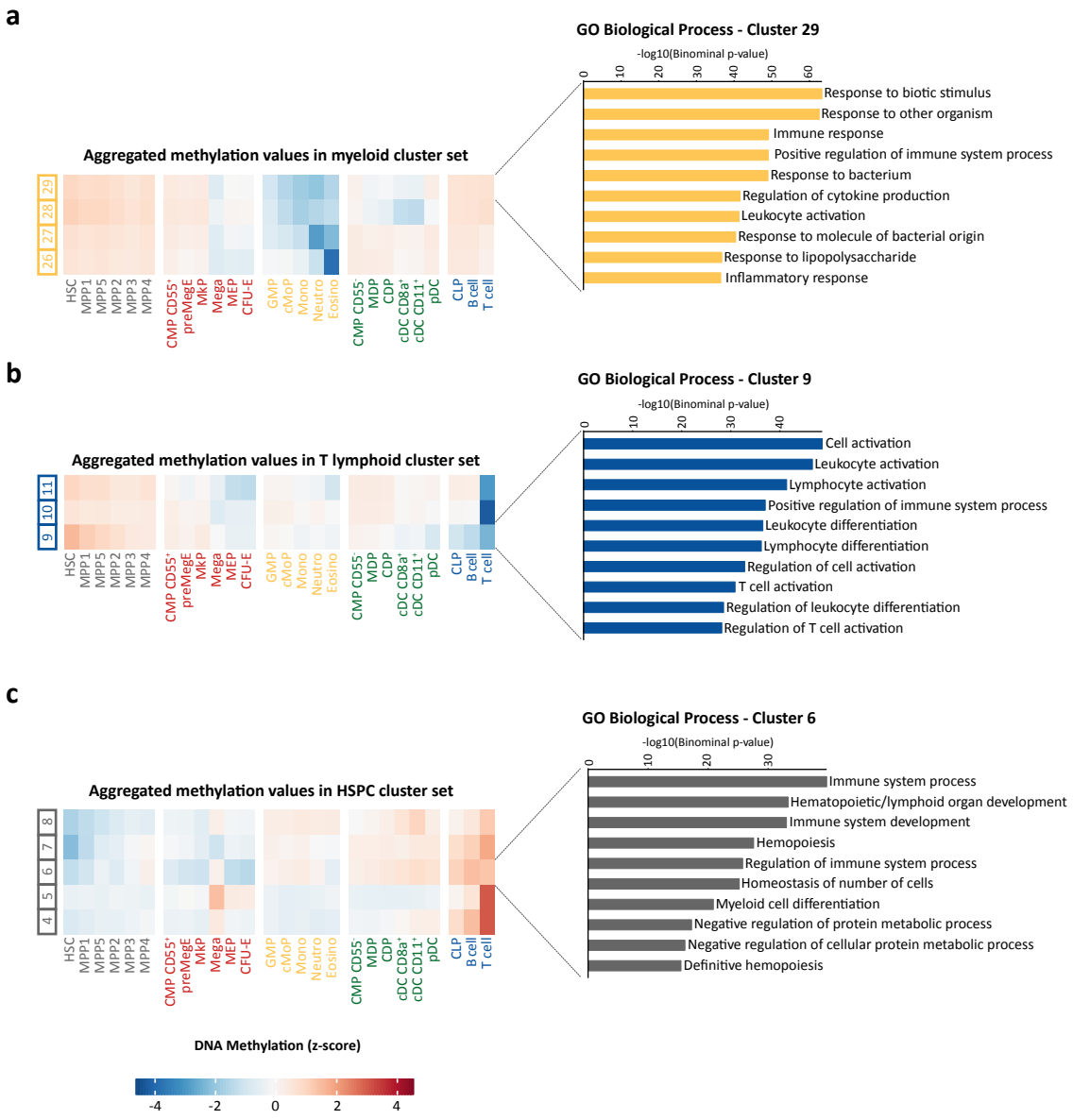

Figure 3.6: Cell type- and lineage-specific DNA methylation programs are enriched for gene ontology terms associated with corresponding hematological processes. GO enrichment analysis of DNA methylation programs was performed using the GREAT web-tool. The heatmaps show the aggregated DNA methylation levels per cluster. Bar plots show the enrichment for the ten most significantly enriched 'GO Biological Process' terms. Displayed are -$\log_{10}$(binominal p-values) for GO terms with a binominal fold enrichment >2. GO term enrichment analysis for the myeloid cluster 29 **(a)**, the T lymphoid cluster 9 **(b)**, and the HSPC cluster 6 **(c)**.

3.3 Hematopoietic differentiation trajectories can be inferred based on progressive and unidirectional DNA methylation changes

Recent technological advancements have enabled the interrogation of transcriptomes and epigenomes at single-cell resolution. Single-cell transcriptomic snapshots and single-cell chromatin accessibility profiles have been used to infer differentiation trajectories in the hematopoietic system [Giladi et al., 2018, Velten et al., 2017, Nestorowa et al., 2016, Macaulay et al., 2016, Cabezas-Wallscheid et al., 2017]. For this purpose, several algorithms have been developed to order individual cells along their differentiation trajectories. In particular, the diffusion algorithm, is well suited to infer differentiation trajectories in systems with multiple end points. Diffusion maps are widely used and can be seen as a non-linear dimensionality reduction method [Angerer et al., 2016, Haghverdi et al., 2015].

With the aim of using DNA methylation dynamics to infer differentiation trajectories and to model the murine hematopoietic system, we applied the diffusion algorithm implemented in the R package 'destiny' to our bulk DNA methylation map of hematopoiesis [Angerer et al., 2016, Haghverdi et al., 2015]. In line with the observed massive DNA methylation reprogramming during erythropoiesis, the first diffusion component (DC1) is mainly described by erythroid differentiation (Figure 3.7 a). However, plotting the second diffusion component (DC2) versus the third diffusion component (DC3) resulted in a diffusion map that separated all different hematopoietic cell types and blood lineages (lineage annotation visualized by color) (Figure 3.7 b). The resulting diffusion map indicates known hematopoietic differentiation trajectories with a branching of the megakaryocytic/erythroid, lymphoid, dendritic, and myeloid lineages. The observation that the megakaryocytes cluster apart from the megakaryocytic progenitors suggests that the sorted megakaryocyte population has differentiated directly from the HSC compartment [Notta et al., 2016, Rodriguez-Fraticelli et al., 2018, Haas et al., 2015, Yamamoto et al., 2013] (Figure 3.7 b, Figure A.4).

Furthermore, unsupervised Louvain clustering, a graph-based clustering method, grouped the hematopoietic cell populations into six clusters (Figure 3.7 c):

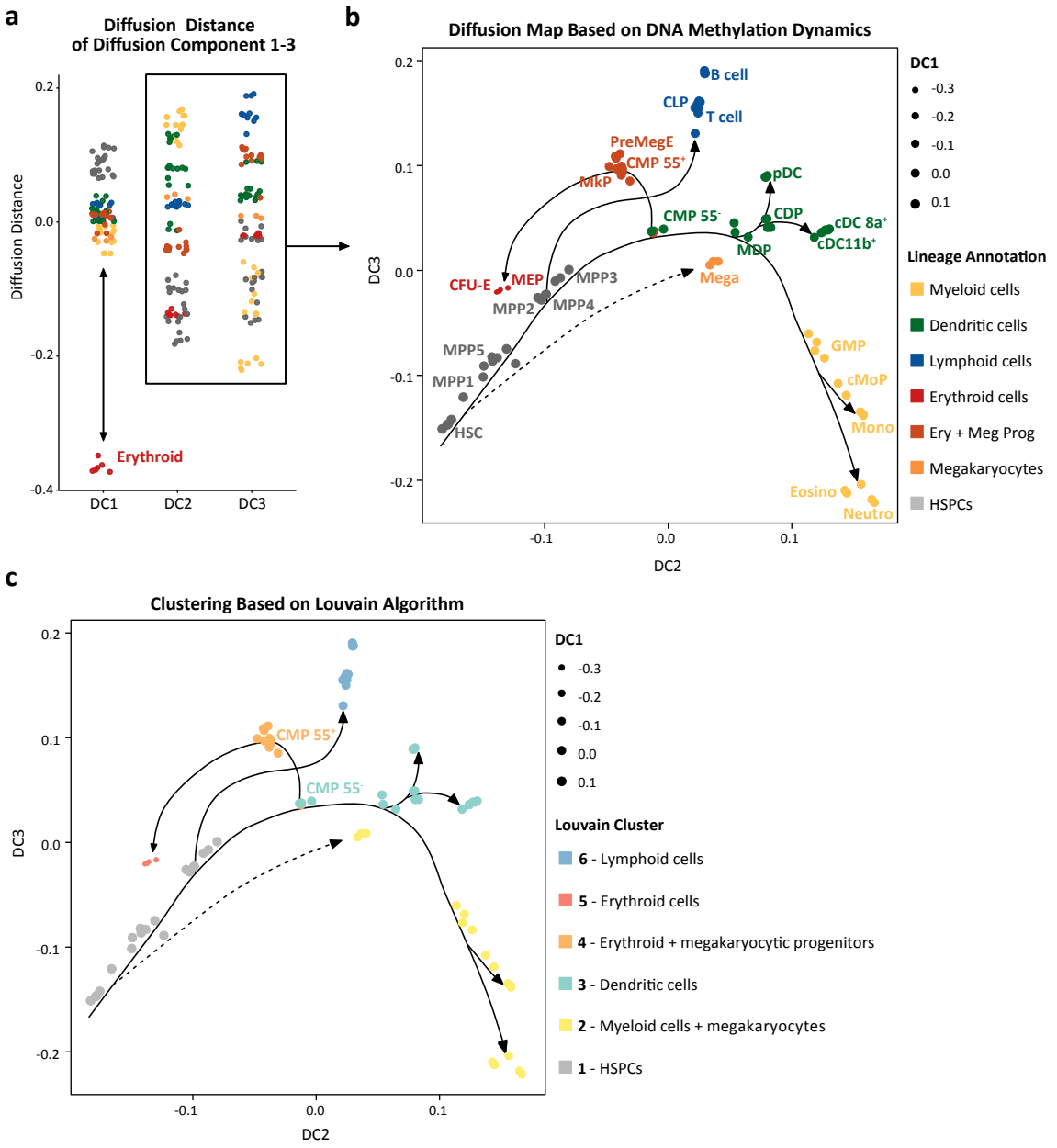

Figure 3.7: Diffusion map based on DNA methylation remodeling events during hematopoietic differentiation reflects known cellular relationships. The diffusion algorithm implemented in the R package 'destiny' was used to infer a model of the hematopoietic system in the form of a diffusion map, which is purely based on DNA methylation dynamics during hematopoietic differentiation. **a**, Jitter plot showing the diffusion distances for diffusion components (DCs) 1-3. Each dot represents a replicate. **b**, Shown is the diffusion map of DC2 versus DC3. The dot sizes represent DC1. The lines indicate presumed differentiation trajectories. **c**, Louvain clustering identifies related cell types. Depicted are the color-coded Louvain clusters as identified by the Louvain algorithm ('destiny' R package). The Louvain clusters are projected onto the diffusion map (DC2 versus DC3).

1. HSPC populations
2. Myeloid cells and megakaryocytes
3. Dendritic cells
4. Erythroid and megakaryocytic progenitors
5. Erythroid cells
6. Lymphoid cells

Intriguingly, the previously defined subpopulations of CMPs showed two clearly distinct DNA methylation patterns. Accordingly, the $CD55^{+}$ CMPs were assigned to the Louvain cluster of erythroid and megakaryocytic progenitors, while $CD55^{-}$ CMPs were assigned to the Louvain cluster of dendritic cells (Figure 3.7 c). These findings confirmed the existence of two distinct CMP subpopulations at the epigenetic level and demonstrated that even closely related cell populations harbor unique pigenetic patterns.

To substantiate our hypothesis that hematopoietic differentiation is accompanied by progressive and unidirectional DNA methylation changes, we projected the aggregated DNA methylation values of the individual DNA methylation programs onto the diffusion map. In line with our hypothesis, we could demonstrate that the DNA methylation programs continuously lose or gain DNA methylation during hematopoietic differentiation. Strikingly, the DNA methylation changes were invariably unidirectional from HSCs to the most differentiated cell populations, suggesting that DNA methylation fates remain stable under homeostatic conditions (Figure 3.8, Figure A.5 and A.6). These unidirectional DNA methylation dynamics can be demonstrated by the myeloid cluster 29, which reveals a progressive and unidirectional loss of DNA methylation along the myeloid differentiation path from HSPCs over myeloid progenitors (CMP $CD55^{-}$, MDP, GMP, cMoP) towards differentiated myeloid cells (monocytes, eosinophils, neutrophils) (Figure 3.8 a, b). Cell populations from other blood lineages also lose DNA methylation in cluster 29, however, in these populations the DNA methylation patterns remain stable after having passed commitment points into other lineages (Figure 3.8 a, b). For example, terminally differentiated dendritic cells show a moderate loss of DNA methylation in hDMRs of cluster 29. The DNA methylation levels within these DMRs remain stable from MDPs onwards, while they become further demethylated in terminally differentiated myeloid cells (Figure 3.8 a, b). These findings recapitulate the idea of an epigenetic landscape that was first

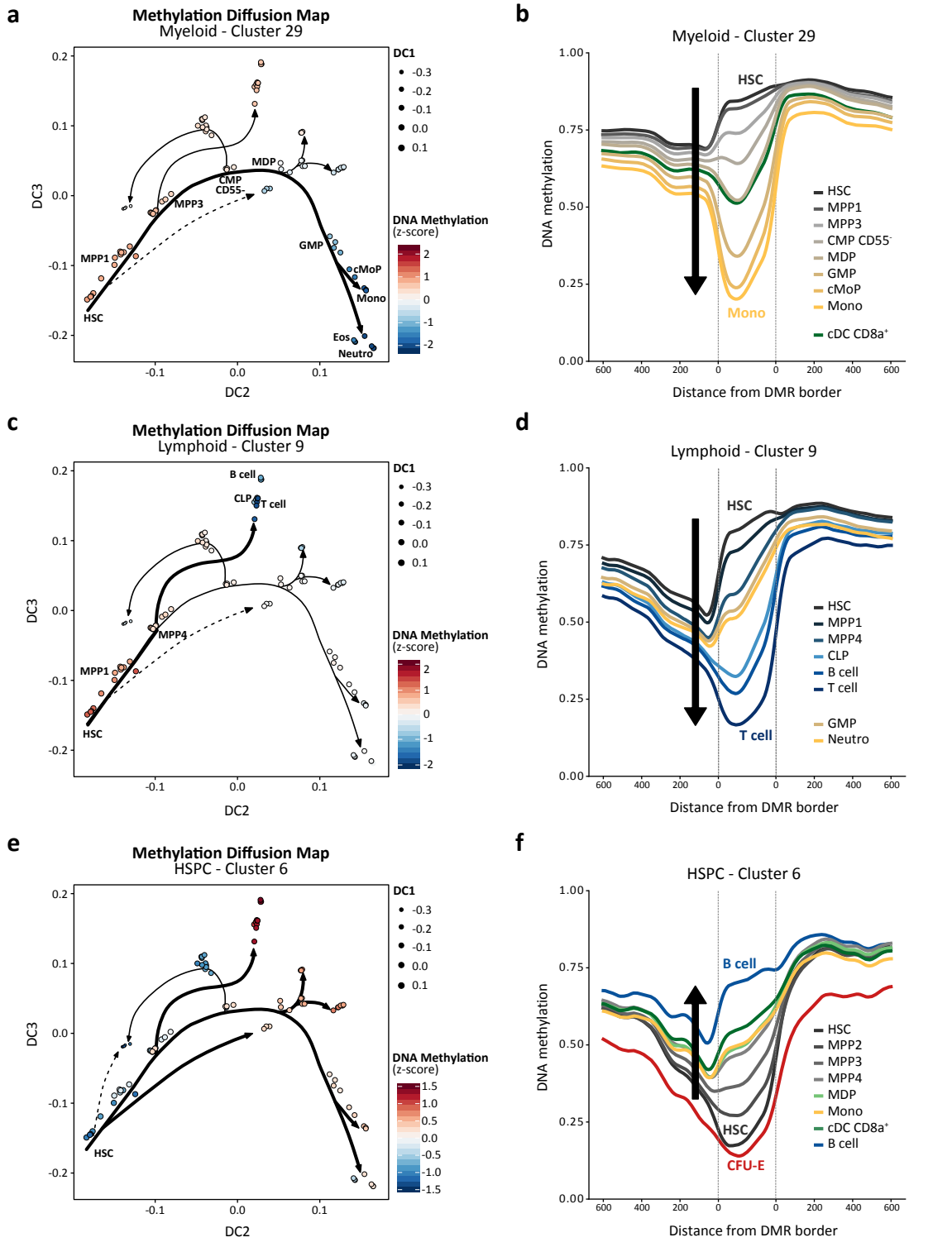

Figure 3.8: DNA methylation changes are progressive and unidirectional during hematopoietic differentiation. DNA methylation levels (z-score transformed beta-values) were aggregated per methylation cluster and projected onto the diffusion map. hDMR-centric line plots show average DNA methylation levels (beta-values) in hDMRs +/- 600 bp. The line plots were generated by Stephen Kraemer. Together, these plots show DNA methylation pseudo-dynamics for the myeloid cluster 29 **(a, b)**, the lymphoid cluster 9 **(c, d)**, and the HSPC cluster 6 **(e, f)**.

proposed by Conrad Hal Waddington [Waddington, 1957].

As another example, hDMRs of the lymphoid methylation cluster 9 show a progressive DNA demethylation during differentiation from HSCs to lymphoid cell populations, while populations of other lineages reveal only a minor DNA methylation loss at these genomic loci (Figure 3.8 c, d).

Lastly, hDMRs of the HSPC methylation cluster 6 show progressive and unidirectional DNA methylation gain in all lineages, with the notable exception of erythroid cells (Figure 3.8 e, f). This can either be explained by the global loss of methylation during erythropoiesis or one might speculate that a subset of erythrocytes originate directly from the HSCs via a shortcut as observed for megakaryocytes [Notta et al., 2016, Rodriguez-Fraticelli et al., 2018, Haas et al., 2015, Yamamoto et al., 2013].

In conclusion, we could confirm that progressive and unidirectional DNA methylation changes occur during the process of hematopoietic differentiation, reminiscent of Waddington's epigenetic landscape [Waddington, 1957]. DNA methylation pseudodynamics allowed us to infer a model of hematopoiesis in form of a diffusion map, which is solely based on DNA methylation changes.

3.4 Integration of DNA methylation profiles with single-cell gene expression snapshots

DNA methylation changes have been shown to impact on the regulation of gene expression. However, the correlation between DNA methylation and gene expression is rather weak at the global level [Bock et al., 2012, Cabezas-Wallscheid et al., 2014, Lipka et al., 2014]. To further examine how the DNA methylation programming impacts on the regulation of gene expression during murine hematopoiesis, we generated a comprehensive multi-layer single-cell gene expression map of the murine hematopoietic system (in collaboration with Dr. Mark Hartman, Jens Langstein, and Katharina Bauer). To enrich rare cell populations, such as the HSPCs, we isolated three cell layers, namely Lin^- $Sca1^+$ $cKit^+$ (LSK) cells, Lin^- $Sca1^-$, $cKit^+$ (LK) cells, and $CD45^+$ total bone marrow cells. Single-cell gene expression profiles were generated using a droplet-based platform (Figure 3.9 a). Overall, 8,957 single cells were captured, including 1,086 cells sorted from the LSK compartment, 3,497 cells sorted from the LK compartment, and 4,374 total bone marrow cells.

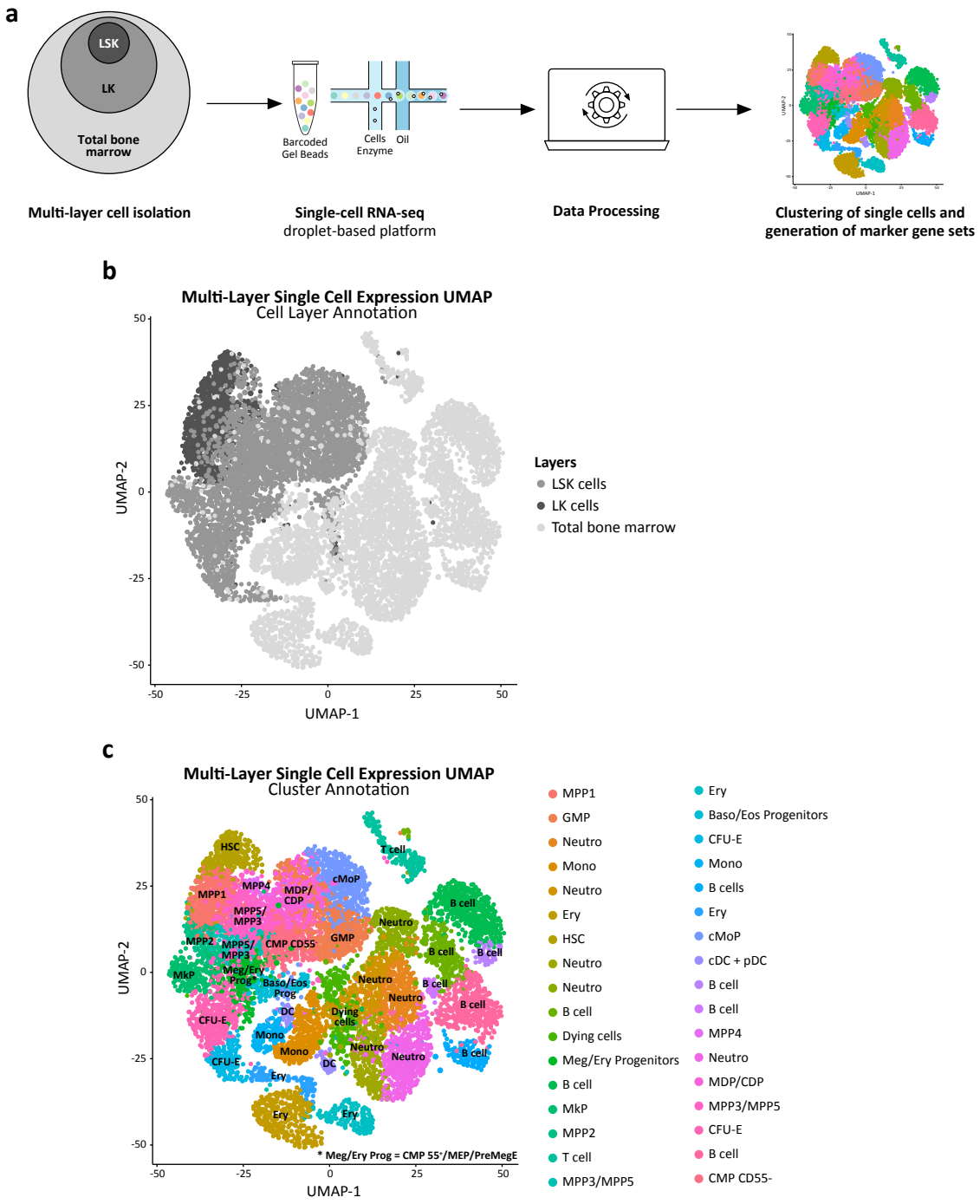

Figure 3.9: Generation of a single-cell multi-layer gene expression map of the murine hematopoietic system.

Figure 3.9: Generation of a single-cell multi-layer gene expression map of the murine hematopoietic system. a, Experimental workflow for the generation of a single-cell gene expression map of murine hematopoiesis. Three cell layers were sorted from the murine bone marrow, namely the LSK (Lin$^-$ Sca1$^+$ cKit$^+$) compartment, LK (Lin$^-$ Sca1$^-$ cKit$^+$) compartment, and total bone marrow (CD45$^+$). The isolated cell layers were subjected to scRNA-seq using a droplet-based platform (10X Genomics; in collaboration with Dr. Mark Hartmann, Jens Langstein, and Katharina Bauer). Data pre-processing, identification of single-cell clusters, and non-linear dimension reduction (UMAP) was performed using the R package 'Seurat' (data pre-processing was performed by Abdelrahman Mahmoud). The clusters were annotated based on the expression of known hematopoietic marker genes as well as the expression of surface marker genes, which were used to isolate hematopoietic cell populations for DNA methylation profiling (cluster annotation was done by Dr. Simon Hass, Dr. Mark Hartmann, and Jens Langstein). **c+b**, UMAP of single-cell gene expression profiles with cell layer **(b)** and cluster annotation **(c)**.

The R package 'Seurat' was used to identify 34 clusters of single-cell expression snap-hots consisting of 39 to 575 single cells and to visualize the single cells in a Uniform Manifold Approximation and Projection (UMAP) (data pre-processing was performed by Abdelrahman Mahmoud) (Figure 3.9 b, c). As expected, the annotation of the three bone marrow layers revealed a hierarchical order of LSK, LK, and total bone marrow cells (Figure 3.9 b).

The single-cell transcriptome clusters were manually annotated based on the expression of known hematopoietic marker genes. If possible, clusters were annotated by means of classical surface markers as defined earlier in the TWGBS experiments (annotation of single-cell expression clusters was performed by Dr. Simon Haas, Dr. Mark Hartman, and Jens Langstein). Nearly all sorted cell populations could be matched to one or more single-cell expression clusters, with the exception of CLPs, which were probably too few in number to build a discrete single-cell expression cluster (Figure 3.9 c). Consistent with previous publications, the resulting single-cell expression UMAP suggests a continuous differentiation process during hematopoiesis [Macaulay et al., 2016, Nestorowa et al., 2016, Velten et al., 2017].

In order to assess the correlation between DNA methylation changes and the regulation of gene expression, we used the gene expression profiles of the 34 single-cell gene expression clusters as marker gene sets for enrichment analyses based on a two-sided Fisher's exact test. Using this approach, we found a strong enrichment of cell type-specific marker gene sets in corresponding cell type- and lineage-specific DNA methylation programs (Figure 3.10 a, b). Briefly, the scRNA-seq marker gene sets of myeloid progenitors and terminally differentiated cells were significantly overrepresented in the myeloid DNA methylation programs 26 to 29. Similarly, the scRNA-seq

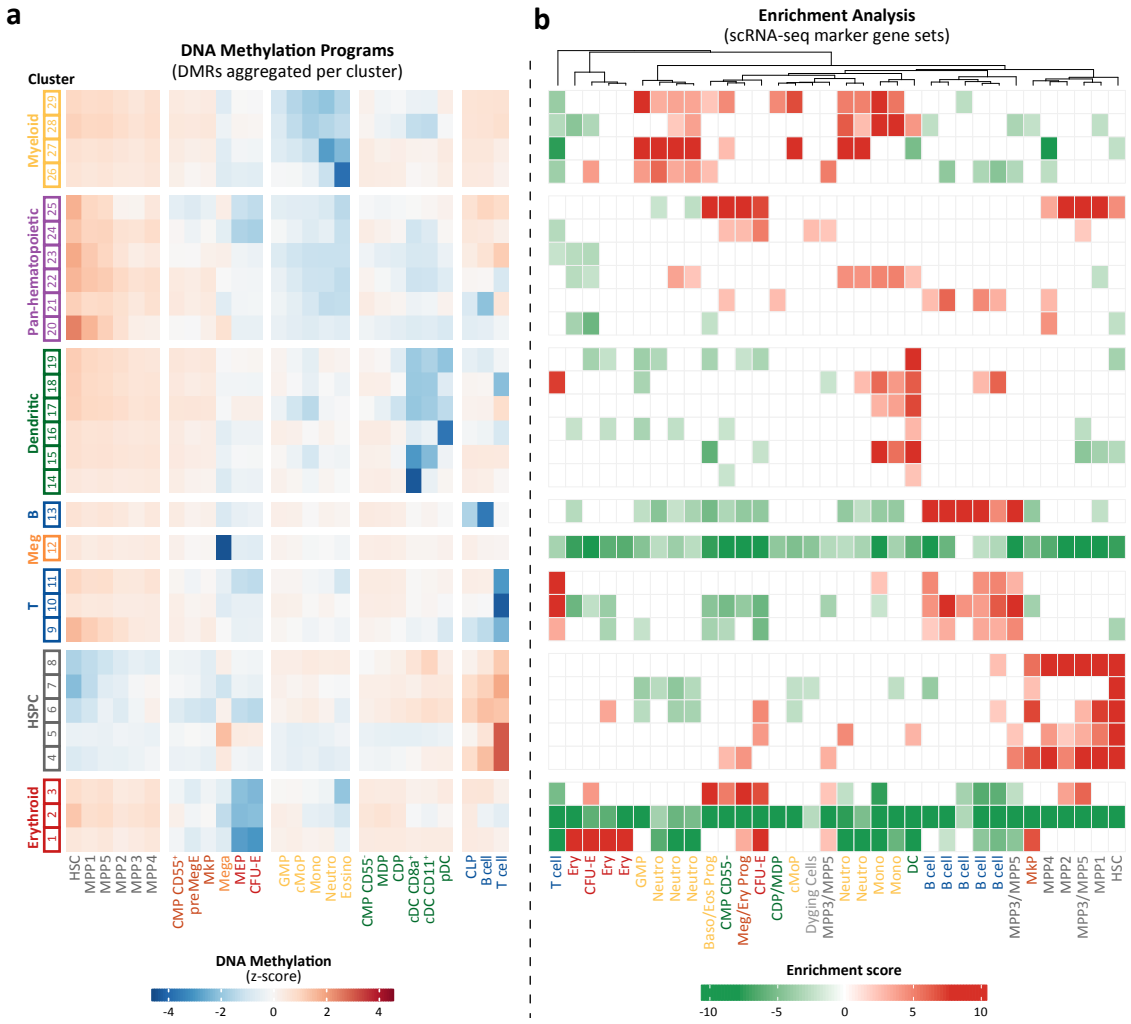

Figure 3.10: Cell type-specific expression programs are overrepresented in corresponding DNA methylation programs. Enrichment analysis of scRNA-seq marker gene sets was performed using a two-sided Fisher's exact test (in collaboration with Stephen Kraemer). **a**, Heatmap showing aggregated scaled DNA methylation levels per DNA methylation cluster. **b**, Heatmap showing the enrichment score of scRNA-seq gene sets per DNA methylation program. The enrichment score was calculated by multiplying the product of the $\log_{10}$(p-value) by the sign of the log odds ratio. Red indicates a significant enrichment ($p < 0.01$) of marker gene sets, green indicates significant depletion ($p < 0.01$). Clustering of columns was performed with hierarchical clustering using Ward's method.

marker gene set of T cells was highly enriched in the T lymphoid DNA methylation program 9, while non-lymphoid scRNA-seq marker gene sets were underrepresented in this methylation cluster. As expected, the pan-hematopoietic DNA methylation cluster showed a significant enrichment for scRNA-seq marker gene sets of various blood cell types, while the HSPC-specific DNA methylation cluster revealed an enrichment of HSPC marker gene sets. Notably, the HSPC-specific DNA methylation clusters were further enriched for scRNA-seq marker genes of MkPs (Figure 3.10 a, b), consistent with recent findings suggesting that HSCs are the major source of megakaryocytes in steady-state hematopoiesis [Rodriguez-Fraticelli et al., 2018]. As previously observed in the GREAT enrichment analysis of GO terms (Figure 3.6), the erythroid DNA methylation cluster 2 and the megakaryocytic DNA methylation cluster 12, both lacked overrepresentation of any scRNA-seq marker gene set. This emphasizes again the global loss of methylation during erythropoiesis. The lack of enrichment for the megakaryocytic cluster 12 can be explained by the size of megakaryocytes, which are probably too big to be captured in a droplet. Thus, we have generated gene expression data of bulk sorted megakaryocytes that we will be used to verify the megakaryocytic methylation cluster 12 (data not analyzed yet).

Of note, although this enrichment analysis was done based on hDMRs that overlap with promoter regions, comparable enrichment patterns were obtained when we used all hDMRs from each DNA methylation program (data not shown).

Given the remarkable overrepresentation of cell type-specific marker gene sets in corresponding DNA methylation programs, we next investigated the correlation between DNA methylation pseudo-dynamics and gene expression. For this purpose, we projected the aggregated promoter DNA methylation levels of the individual DNA methylation programs onto the single-cell gene expression UMAP. Gene expression values per cell were calculated as the mean of the normalized gene expression score of all genes that show dynamic promoter methylation in a specific DNA methylation cluster (Figure 3.11).

Consistent with our previous findings, all DNA methylation programs showed progressive and unidirectional DNA methylation pseudo-dynamics. Loss of DNA methylation at promoter regions was mostly associated with an increase of gene expression levels, while a gain of DNA methylation was commonly associated with a decrease in gene expression levels (Figure 3.11). This anti-correlation is clearly illustrated by DNA methylation and gene expression pseudo-dynamics in cluster 6,

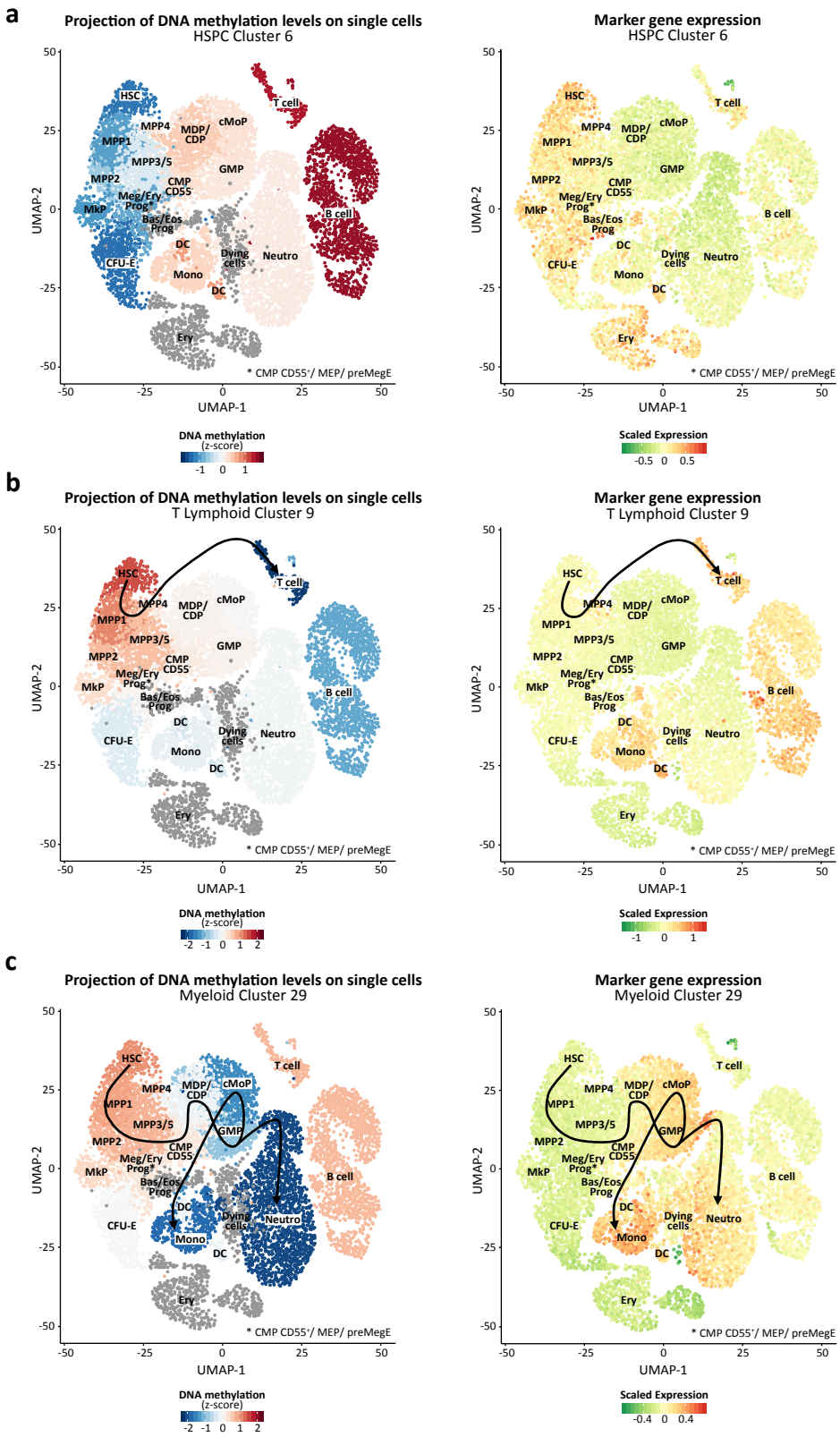

Figure 3.11: Projection of DNA methylation levels onto the single-cell RNA-seq UMAP reveals a complex relationship between DNA methylation programming and gene expression patterns.

Figure 3.11: Projection of DNA methylation levels onto the single-cell RNA-seq UMAP reveals a complex relationship between DNA methylation programming and gene expression patterns. Aggregated z-score transformed DNA methylation values per DNA methylation cluster were projected onto the single-cell gene expression UMAP. Gene expression values were calculated as the mean of the scaled and normalized expression of all marker genes that show variable DNA methylation at promoter regions in the individual DNA methylation clusters. Shown are the UMAPs with the projection of promoter DNA methylation values and the UMAPs with the averaged gene expression values for the HSPC cluster 6 **(a)**, T lymphoid cluster 9 **(b)**, and myeloid cluster 29 **(c)**.

showing low DNA methylation levels in HSPCs, erythroid and megakaryocytic cells and conversely high gene expression levels in the same cell populations (Figure 3.11 a). Furthermore, cluster 9 reveals a progressive loss of DNA methylation during hematopoietic differentiation, with the lowest DNA methylation levels observed in B cells and T cells. This lymphoid-specific loss of DNA methylation is accompanied by an upregulation of genes, whose expression is already detectable in lymphoid-primed MPP4 cells (Figure 3.11 b). Interestingly, the early loss of DNA methylation at the transition step from HSC to MPP1 might suggest that DNA methylation changes precede changes in gene expression (Figure 3.11 b). However, a statement about the temporal sequence of DNA methylation dynamics and gene expression changes requires further evaluation using data on DNA methylation and gene expression coming from the same single cell.

Importantly, DNA methylation changes did not always correlate with gene expression changes. Besides the clear lymphoid-specific loss of methylation, hDMRs of cluster 9 became also slightly demethylated in erythroid cells, dendritic cells, monocytes, and neutrophils; however, only dendritic cells and monocytes showed a clear upregulation of the corresponding genes (Figure 3.11 b). These observations support the view that DNA methylation patterns do not directly translate into transcriptional patterns. Low or high levels of DNA methylation at gene regulatory elements reflect a certain chromatin structure that either facilitates transcriptional activity or not. Additional factors need to be recruited to lowly methylated gene regulatory regions in order to activate gene expression programs.

While DNA methylation changes are invariably unidirectional, gene expression patterns reveal more complex dynamics. As already demonstrated, the myeloid cluster 29 shows a unidirectional loss of DNA methylation during the differentiation process towards monocytes and neutrophils (Figure 3.8 a, b; Figure 3.11 c). This loss of DNA methylation is associated with a gradual increase in gene expression

during monocytic differentiation. However, the neutrophil differentiation trajectory shows only a temporal upregulation of this expression program at the transition from GMPs to neutrophils, while mature neutrophils again exhibit low expression levels (Figure 3.11 c).

In order to validate the observed DNA methylation and gene expression dynamics at the level of individual genes, we further analyzed the correlation of promoter DNA methylation and gene expression of selected hematopoietic genes. Similar to the DNA methylation programs, hDMRs in promoter regions of individual genes showed progressive DNA methylation changes that mostly correlated with the transcript level (Figure 3.12).

Meis1 codes for a transcription factor required for erythropoiesis, megakaryopoiesis, and stem cell expansion. In line with this, the hDMR overlapping with the promoter region of *Meis1* revealed a progressive gain of DNA methylation in the myeloid, dendritic, and lymphoid lineage, associated with silencing of *Meis1* in these cell populations (Figure 3.12 a).

Similarly, the promoter of *Cd19* showed a strong demethylation in B cells, which was accompanied by an upregulation of *Cd19*, a gene essential for B cell specification (Figure 3.12 b).

We also identified a number of genes that lack the tight anti-correlation between DNA methylation and gene expression. For instance, DNA methylation at the promoter of *Mpo*, a gene important for myeloid differentiation, was associated with a unidirectional demethylation in the myeloid lineage, with the lowest methylation levels in monocytes and neutrophils. However, this loss of DNA methylation was only accompanied by a temporal increase in *Mpo* expression in myeloid progenitor cell populations (Figure 3.12 c), indicating that other epigenetic layers are involved in the initiation of gene expression.

Altogether, the integration of DNA methylation and gene expression dynamics during hematopoietic differentiation uncovered two intriguing results: first, while gene expression can alternatingly change during hematopoietic differentiation, we could show that DNA methylation programming is invariably unidirectional, indicating that DNA methylation changes are irreversible under physiological conditions. Second, the loss of DNA methylation is not inevitably accompanied by an increase of gene expression. This might also explain the weak anti-correlation between DNA methy-

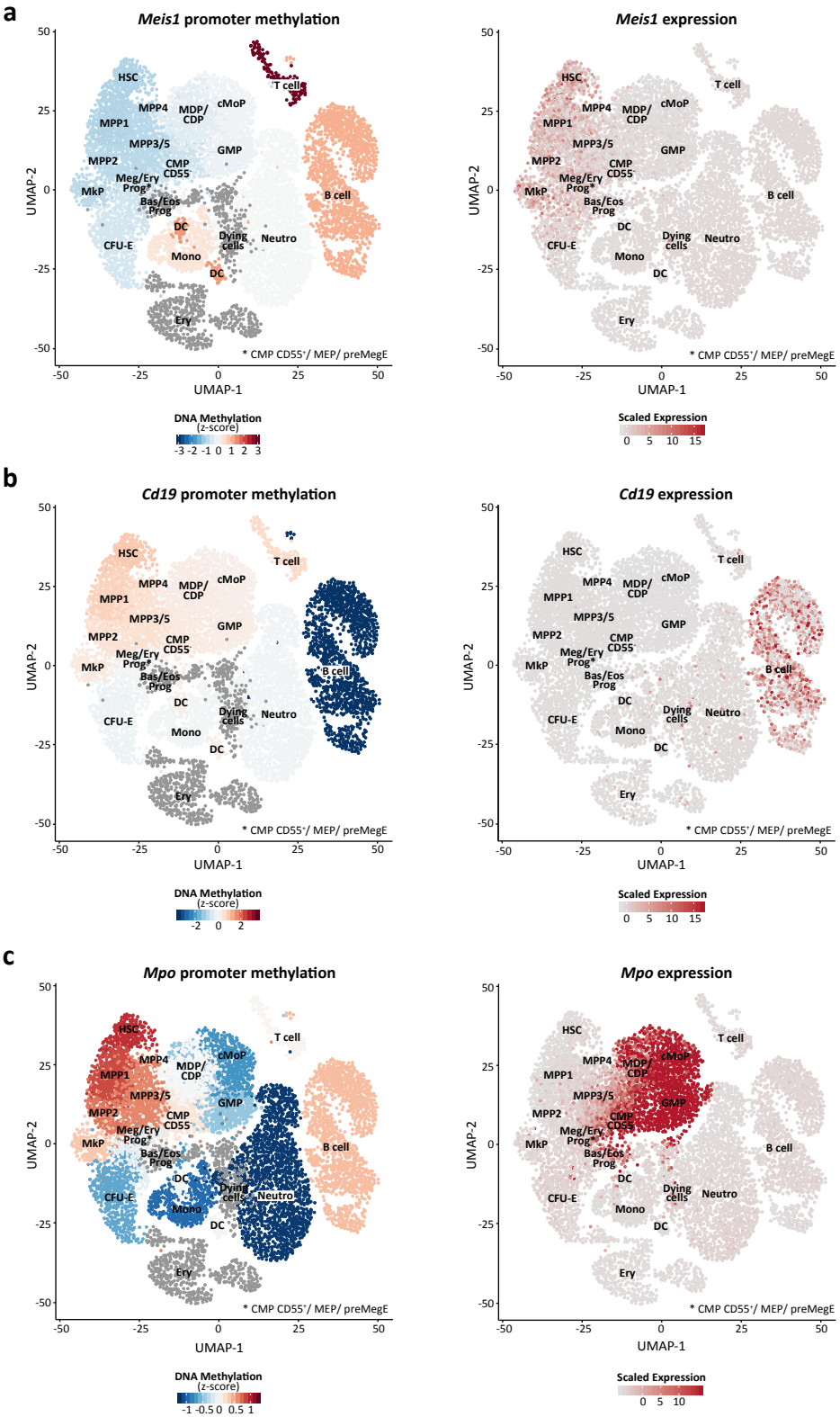

Figure 3.12: Projection of promoter DNA methylation of *Meis*1, *Cd19*, and *Mpo* onto the single-cell RNA-seq UMAP.

Figure 3.12: Projection of promoter DNA methylation of *Meis1*, *Cd19*, and *Mpo* onto the single-cell RNA-seq UMAP. Z-score transformed DNA methylation values of hDMRs overlapping with promoters of selected genes were projected onto the single-cell gene expression UMAP. Shown are the UMAPs with the projection of promoter DNA methylation levels on the left and the UMAPs with the projection of gene expression values on the right for *Meis1* **(a)**, *Cd19* **(b)**, and *Mpo* **(c)**.

lation and gene expression observed in previous studies [Bock et al., 2012, Cabezas-Wallscheid et al., 2014, Lipka et al., 2014, Hodges et al., 2011]. Furthermore, this finding suggests that DNA methylation changes are rather permissive than instructive for the regulation of gene expression and lineage specification.

3.5 Lineage-specific DNA methylation programs overlap with *cis*-regulatory regions

Previous studies have demonstrated that DNA methylation changes associated with early hematopoietic differentiation occur predominantly at *cis*-regulatory elements, including distal transcription factor binding sites or enhancer regions [Cabezas-Wallscheid et al., 2014, Lipka et al., 2014].

To further characterize the biological relevance of the individual DNA methylation programs, we searched for overrepresented transcription factor binding motifs in the individual DNA methylation clusters. We performed an enrichment analysis of 239 transcription factor binding motifs (in collaboration with Stephen Krämer), which were downloaded from the HOMER webpage. Overall, we found 167 transcription factor binding motifs that were significantly ($p < 1x10^{-20}$) enriched in at least one of the 29 DNA methylation programs, including binding motifs of known hematopoietic transcription factors (Figure 3.13) [Orkin and Zon, 2008, Lara-Astiaso et al., 2014].

Accordingly, binding motifs of transcription factors essential for erythroid and megakaryocytic lineage specification, such as Meis1, Kfl4, and GATA-like proteins, were strongly enriched in the erythroid cluster set, while binding motifs for transcription factors essential for of other lineages (e.g. Cebp, Irf4, Ebf) were depleted. The strong enrichment of binding motifs of multiple non-hematopoietic transcription factors in the erythroid cluster 2 can again be explained by a broad unspecific loss of DNA methylation during terminal erythropoiesis. Binding sites for key transcription factors of T cell differentiation (E2A, Tcf12, Tcf3, Runx1) were enriched in the T lymphoid cluster set, whereas binding sites for myeloid transcription factor (Cepb,

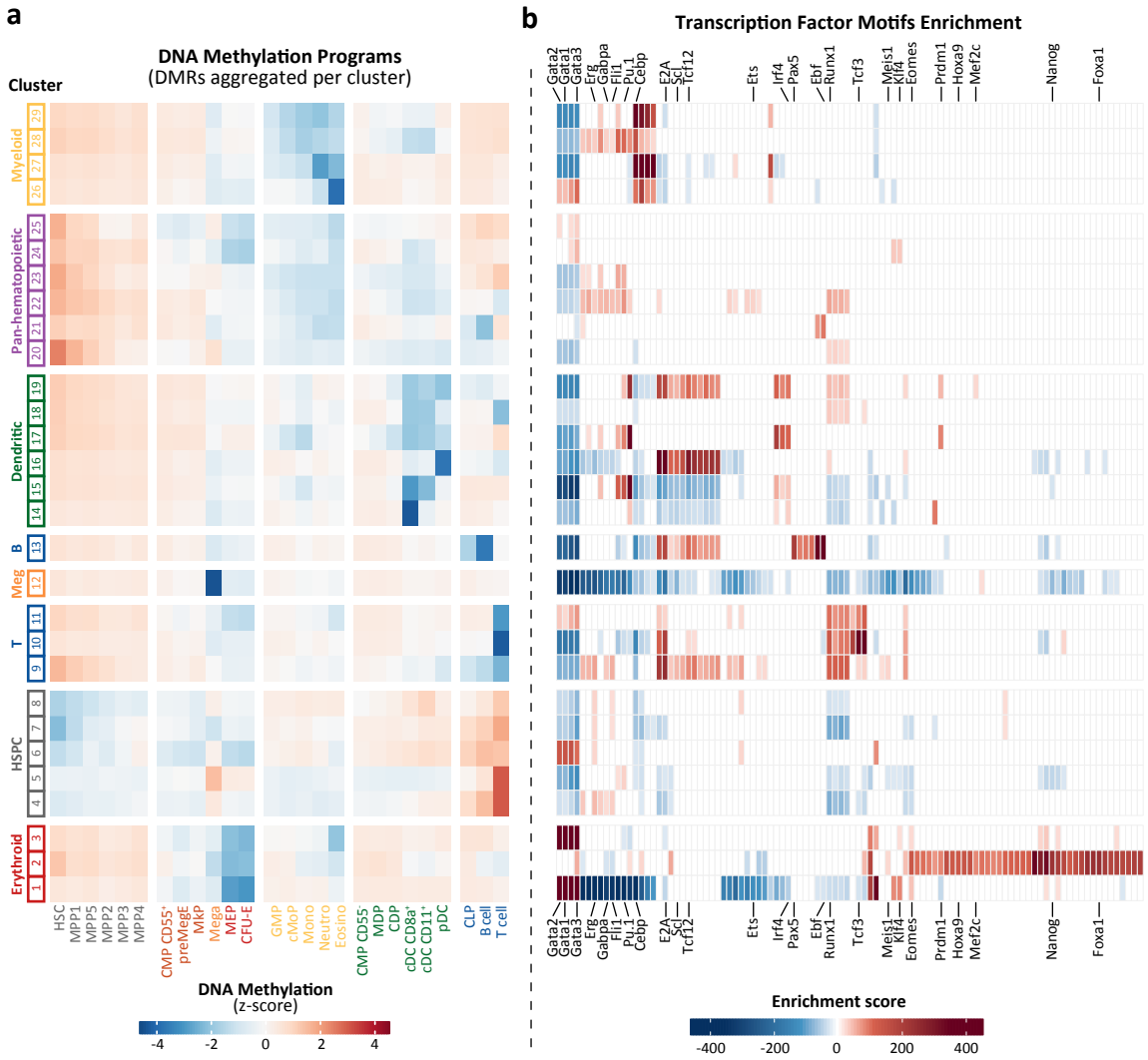

Figure 3.13: Binding motifs of key hematopoietic transcription factors are differentially enriched across DNA methylation programs. Enrichment analysis of HOMER transcription factor binding motifs was performed using a Fisher's exact test (in collaboration with Stephen Kraemer). **a**, Heatmap (from Figure 3.10 a) showing the scaled DNA methylation levels aggregated per DNA methylation cluster. **b**, Heatmap showing the enrichment scores of the 100 most significantly enriched transcription factor binding motifs. The enrichment score was calculated by multiplying the $\log_{10}$(p-value) by the sign of the log odds ratio. Red color indicates significant enrichment (p $<1x10^{-20}$) of a binding motif, whereas blue color indicates a significant depletion (p $<1x10^{-20}$). Clustering of columns was performed with hierarchical clustering using Ward's method and Euclidean distance.

Pu.1) were depleted, indicating closing of myeloid transcription factor binding sites in T cells. Furthermore, binding motifs of transcription factors required for B cell maturation (E2A, Pax5, Ebf) were significantly overrepresented in the B lymphoid cluster, while binding motifs other transcription factors were again depleted. Along these lines, binding sites for transcription factors involved in dendritic cell specification (Pu.1, Irf4) were overrepresented in the dendritic cluster set. Strikingly, the dendritic clusters 16 and 19, which show a specific loss of DNA methylation in pDCs, were further enriched for binding motifs of lymphoid transcription factors (Tcf12, E2A, Runx1). This observation corresponds with current research indicating that pDCs may arise from an alternative differentiation path via a lymphoid progenitor population [Rodrigues et al., 2018]. Finally, the myeloid cluster set revealed an enrichment of transcription factor binding sites of key myeloid transcription factors, like Pu.1 and Cepb-like transcription factors. In agreement with previous enrichment analyses, the megakaryocytic methylation cluster 12 lacked any motif enrichment. Thus, it will be important to validate the integrity of the DNA methylome of megakaryocytes using bulk RNA-seq data. Of interest, the HSPC cluster set showed only a weak enrichment for binding motifs of transcription factors that are important for stem cell maintenance (Erg, Meis1, Scl); however, binding motifs of transcriptions factors involved in erythroid and megakaryocytic lineage priming (GATA-like transcription factors, Fli1) were found to be enriched in HSPCs. This finding further substantiates data demonstrating that the majority of phenotypic HSCs are megakaryocyte-primed HSCs (Figure 3.13, Figure A.7) [Rodriguez-Fraticelli et al., 2018, Haas et al., 2015, Yamamoto et al., 2013].

To study the impact of DNA methylation changes on the actual binding of transcription factors, we downloaded chromatin immunoprecipitation sequencing (ChIP-seq) data from the CODEX database. Enrichment analysis of ChIP-seq peaks was performed using data from 215 experiments investigating the binding of 58 transcription factors in primary hematopoietic cell populations (Figure A.8). Although the ChIP-seq data was generated in diverse hematopoietic cell populations, the DNA methylation programs showed lineage-specific enrichments of hematopoietic transcription factor binding. For instance, the myeloid cluster set revealed an overrepresentation of ChIP-seq peaks of the myeloid transcription factors PU.1, Cebpa, and Cebpb, while B lymphoid transcription factors, such as Pax5, Ebf2, and Ebf3, were enriched in the B lymphoid cluster. Additionally, transcription factors involved in hematopoietic stem cell maintenance, such as Tal1, Lmo2, Erg, and Gata2, were

highly enriched in the HSPC cluster set (Figure A.8).

Next, we investigated the correlation of the expression of hematopoietic transcription factors and the DNA methylation status of hDMRs overlapping with transcription factor binding sites. As expected, we found key hematopoietic transcription factors differentially expressed in a lineage- and cell type-specific manner. However, DNA methylation of transcription factor binding motifs did not always correlate with the expression of the corresponding transcription factor, indicating again a permissive role of DNA methylation in lineage specification.

For example, *Gata1* expression and low DNA methylation at hDMRs overlapping with Gata1 binding motifs coincide in erythrocyte- and megakaryocyte-primed cells to instruct erythropoiesis and megakaryopoiesis (Figure 3.14 a). In contrast, hDMRs overlapping with binding motifs of the key myeloid transcription factor Pu.1 are demethylated in the myeloid and erythroid lineages, however, Pu.1 is only significantly expressed in myeloid cells (Figure 3.14 b). Similarly, DMRs overlapping with Pax5 binding motifs lose DNA methylation in all blood lineages during hematopoietic differentiation, but only B cells express Pax5 (Figure 3.14 c). Genes encoding transcription factors that are important for stem cell maintenance such as *Hoxa9* are expressed in the HSPC compartment and get continuously downregulated during differentiation. However, HSCs neither show a loss of DNA methylation at binding motifs of transcription factors regulating stem cell maintenance, such as Hoxa9, Tal1, or Erg, nor at binding motifs of lineage-specifying transcription factors, such as Gata1, Pu.1, or Pax5 (Figure 3.14 d), suggesting that the low expression of lineage-specifying transcription factors is critical for stem cell maintenance.

In summary, we found that the DNA methylation status of transcription factor binding motifs is important for hematopoietic lineage specification. Our analysis indicates that the co-occurrence of transcription factor expression and DNA methylation loss at corresponding binding sites is essential for lineage specification.

Previous studies have demonstrated that *de novo* establishment of hematopoietic enhancers is characteristic for hematopoietic differentiation. Lara-Astiaso *et al.* performed a comprehensive analysis of chromatin marks to identify 48,415 murine hematopoietic enhancers throughout the hematopoietic system, which were grouped into nine clusters by k-means clustering. Based on the activity of the enhancer regions in the different cell types investigated, they designated the nine clusters as follows: I. Common, II. Myeloid + Progenitors, III. Lymphoid + Progenitors, IV.

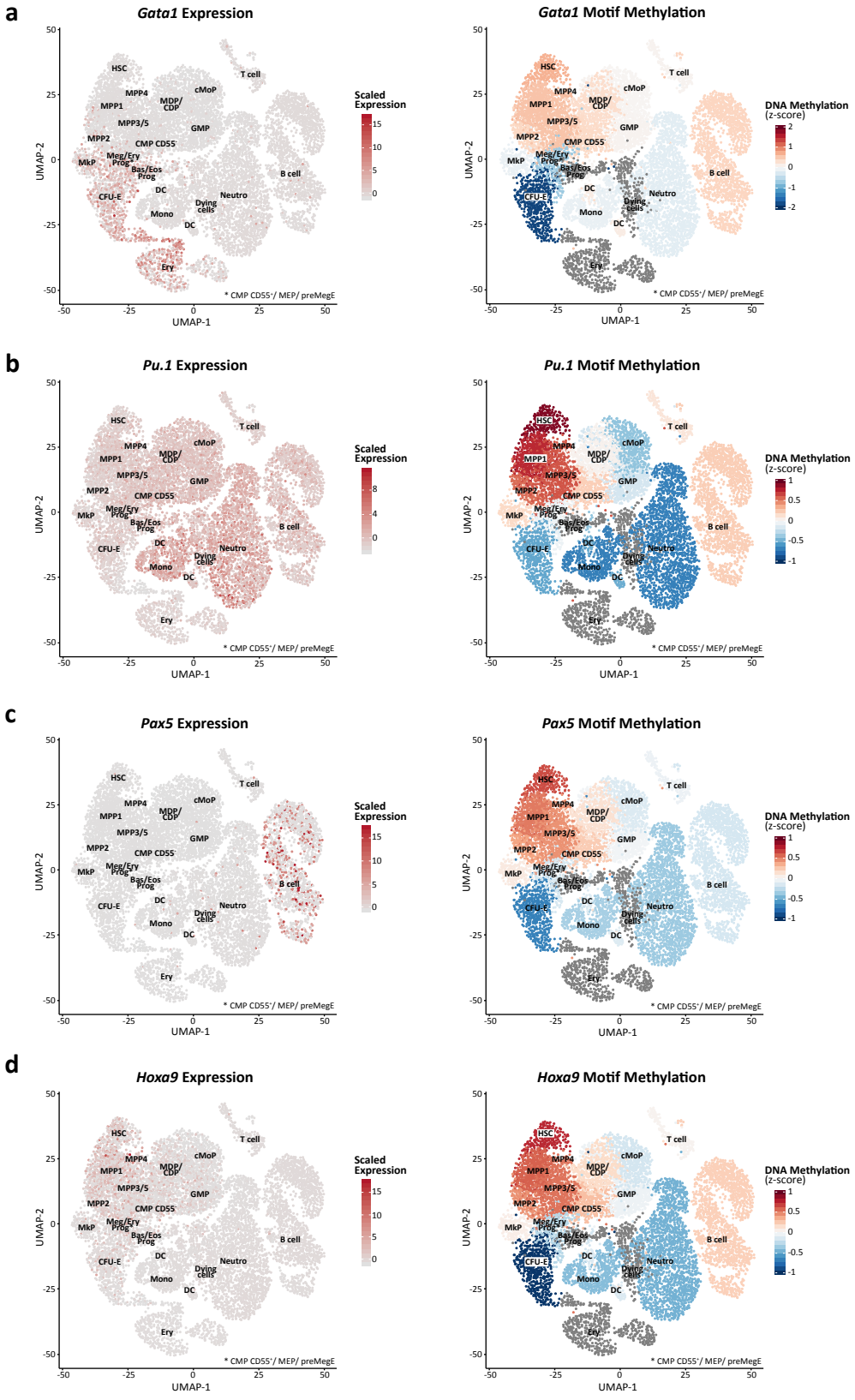

Figure 3.14: Correlation of transcription factor expression and DNA methylation of corresponding transcription factor binding motifs.

Figure 3.14: Correlation of transcription factor expression and DNA methylation of corresponding transcription factor binding motifs. Depicted are UMAPs showing the expression levels of selected transcription factors on the left and UMAPs with projected z-score transformed DNA methylation values of hDMRs overlapping with corresponding transcription factor binding motifs on the right. Expression values and DNA methylation of binding motifs are depicted for *Gata1* **(a)**, *Pu.1* **(b)**, *Pax5* **(c)**, and *Hoxa9* **(d)**.

Erythroid + Progenitors, V. Progenitors, VI. Myeloid, VII. B cells, VIII. T/NK cells, IX. Erythroid [Lara-Astiaso et al., 2014].

We recapitulated the k-means clustering and used the identified enhancer programs to perform enrichment analysis (Figure A.9). Overall, we found a rather small overlap (8.65%) of hDMRs and enhancer regions compared to the overlap with promoter regions (18.58%) (Figure 3.15 a). This minor overlap can be explained by the low quality of the enhancer detection in HSPCs and committed progenitor populations using ChIP-seq and by a smaller panel of hematopoietic cell populations investigated in this study [Lara-Astiaso et al., 2014].

In order to investigate whether lineage-specific enhancer programs are overrepresented in the DNA methylation programs, we performed an enrichment analysis based on the Fisher's exact test (in collaboration with Stephen Kraemer) (Figure 3.15).

Overall, we found varying enrichment of all enhancers in the individual DNA methylation programs. While the T lymphoid cluster 9 showed a significant overlap with enhancers, the dendritic cluster 14 was depleted for enhancer regions. The small overlap with enhancer regions might explain why methylation cluster 14 lacks any enrichment of lineage-specific enhancers.

Lineage-specific enhancer programs were enriched in corresponding DNA methylation programs. The B cell enhancer cluster (VII.), for instance, was highly enriched in the B lymphoid DNA methylation cluster. Moreover, the myeloid enhancer programs (II. Myeloid + Progenitors and VI. Myeloid) were significantly overrepresented in the myeloid DNA methylation cluster set (Figure 3.15 b, c).

Altogether, we could show that lineage-specific hDMRs overlap with enhancer regions of lineage-specific enhancer programs. The data further indicates that DNA methylation dynamics are involved in the *de novo* establishment of hematopoietic enhancer regions during hematopoietic differentiation.

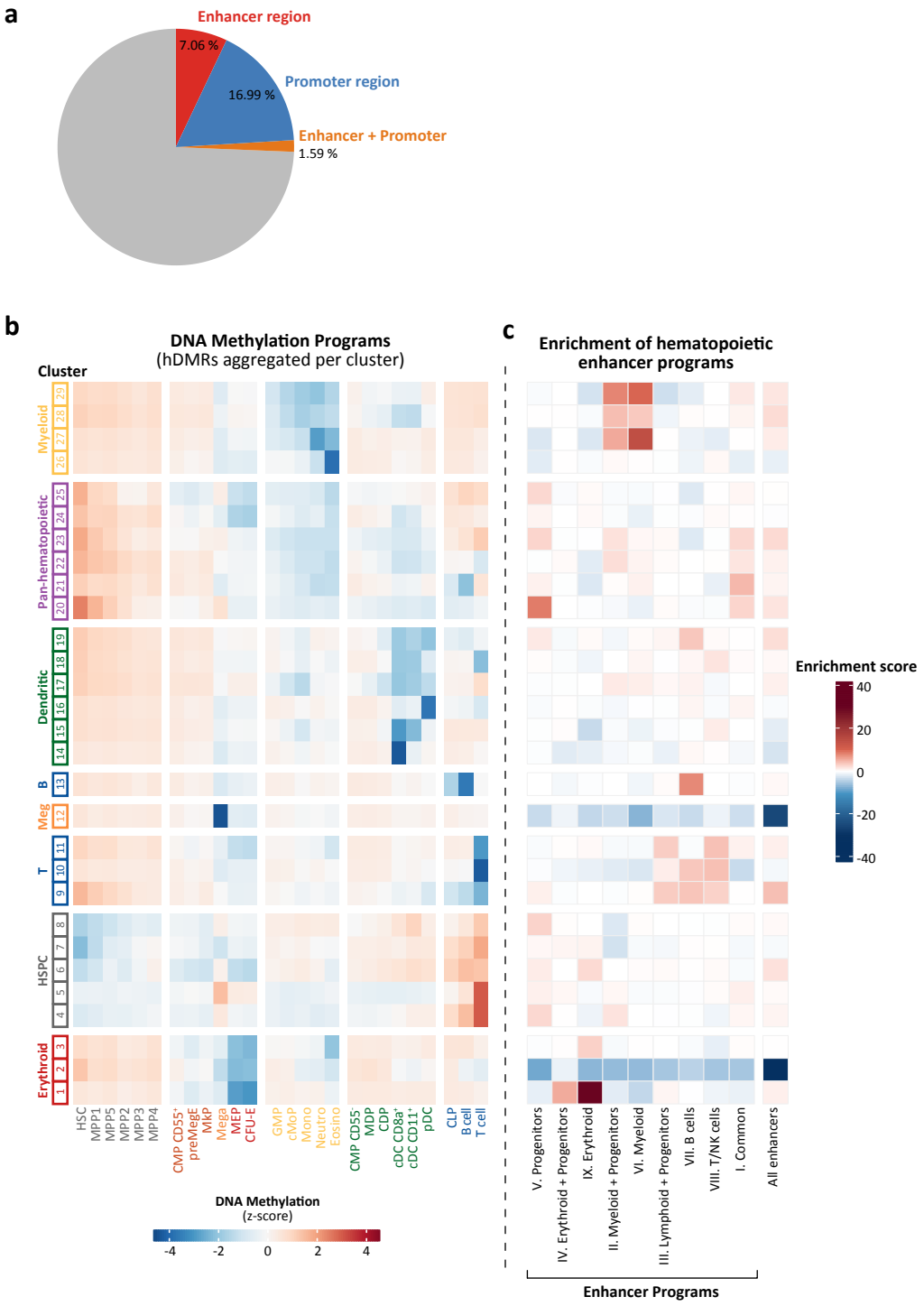

Figure 3.15: Lineage-specific enhancer programs overlap with lineage-specific methylation programs. **a**, Pie chart showing the percentage of hDMRs overlapping with hematopoietic enhancer regions (red, ChIP-seq data from Lara-Astiaso *et al.*) and with annotated promoter regions (blue). Grey indicates overlap of hDMRs with other genomic regions (Figure 3.3 b). **b**, Heatmap (from Figure 3.10 a) showing the scaled DNA methylation levels aggregated per DNA methylation cluster. **c**, Heatmap showing the enrichment scores of hematopoietic enhancer programs. Enrichment analysis was performed using the Fisher's exact test (in collaboration with Stephen Kraemer). Red indicates enrichment and blue indicates depletion of enhancer programs.

3.6 Early DNA methylation changes allow the segregation of terminal cell fates

With the intention to explore early hematopoietic lineage priming at the level of DNA methylation, we further dissected the HSPC compartment analyzed by Cabezas-Wallscheid *et al.* (HSC, MPP1, MPP2, MPP3/4) into six cell populations [Cabezas-Wallscheid et al., 2014]: HSC, MPP1, MPP5, MPP2, MPP3, and MPP4 (Figure 3.1.b).

In accordance with the previous studies by Cabezas-Wallscheid *et al.* and Lipka *et al.* [Cabezas-Wallscheid et al., 2014, Lipka et al., 2014], we observed early DNA methylation dynamics within the HSPC compartment in several DNA methylation clusters. For instance, the megakaryocyte- and erythrocyte-primed MPP2 population revealed lower DNA methylation levels in hDMRs of cluster 6 as compared to either the myeloid-primed MPP3 or the lymphoid-primed MPP4 populations. This loss in DNA methylation became more evident in megakaryocyte and erythrocyte progenitor cell populations (Figure 3.16 a). Similarly, lymphoid-primed MPP4 cells showed higher DNA methylation levels in cluster 25 as compared to the MPP2 and MPP3 populations. The high DNA methylation values in cluster 25 are likely to reflect the lymphoid priming of MPP4 cells, as all other hematopoietic cell types, except for those of the lymphocytic lineage, show lower DNA methylation levels in these regions (Figure 3.16 b). Incidentally, cluster 25 is probably responsible for the repression of myeloid genes in lymphoid cell populations.

Given the differential DNA methylation patterns in the HSPC compartment, we assessed whether DNA methylation signatures of differentiated cell types were already present in populations of the HSPC compartment. To this end, we performed a non-negative matrix factorization (NMF), a matrix decomposition method to reduce the complexity of a matrix. The factorization rank (n = 6) was determined empirically. The factorization rank represents six methylation signatures, which we designated as: 'Stem cells', 'Mega/Ery', 'Myeloid', 'Dendritic', 'B lymphoid', and 'T lymphoid'. The resulting exposure matrix confirmed that lineage-specific DNA methylation signatures are differentially exposed across the individual MPP subpopulations. For instance, the megakaryocytic and erythroid populations are highly exposed to the Mega/Ery signature. Strikingly, within the HSPC compartment this lineage signature shows the strongest exposure in the megakaryocyte- and

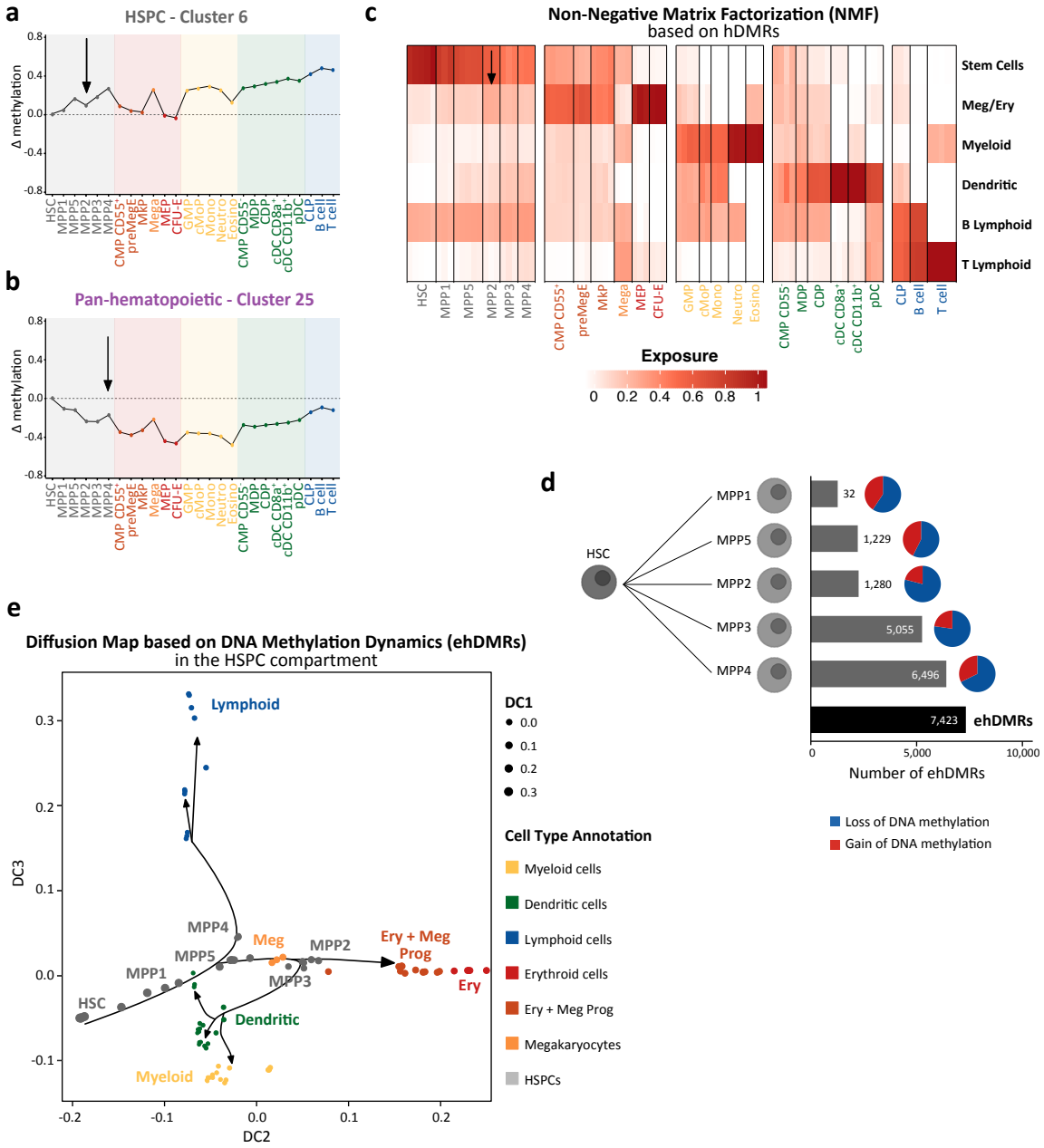

Figure 3.16: Lineage-specific DNA methylation changes are detectable within the primitive hematopoietic stem and progenitor cell compartment. a, **b**, Line plots depicting the absolute change in DNA methylation levels as compared to HSCs (delta methylation) for the HSPC cluster 6 **(a)** and for the pan-hematopoietic cluster 25 **(b)**. Arrows indicate early lineage priming in MPP2 and MPP4, respectively. **c**, NMF reveals lineage-specific signatures with differential exposures in the primitive HSPC compartment. NMF was performed with the 'Bratwurst' Bioconductor package with a factorization rank of 6. **d**, DMRs were identified by pairwise comparison of HSCs and the different MPP populations 1-5 (DMR calling was done by Stephen Kraemer). Bar plot illustrating the number of DMRs for the different pairwise comparisons. The pie charts illustrate the gain (red) or loss (blue) of DNA methylation. In total, 7,423 unique DMRs ('early hematopoietic DMRs'; ehDMRs) were identified. **e**, Diffusion map based on DNA methylation pseudo-dynamics in ehDMRs. Depicted are DC2 versus DC3. The dot size represents DC1. Lines illustrate presumed differentiation trajectories.

erythrocyte-primed MPP2 population (Figure 3.16 c), indicating early commitment decisions in the HSPC compartment.

In order to investigate the concept of early epigenetic priming further, we focused our analysis on only those hDMRs, which were identified within the HSPC compartment (n = 7,423 DMRs). We designated these DMRs as 'early hematopoietic DMRs' (ehDMRs). The majority of these ehDMRs were associated with a loss of DNA methylation compared to HSCs (Figure 3.16 d). Remarkably, the DNA methylation dynamics detected in ehDMRs were sufficient to separate terminal cell fate decision in a diffusion map and a classical phylogenetic tree (Figure 3.16 e; Figure A.10). Notably, the diffusion map revealed differentiation biases towards different hematopoietic lineages across the different MPP populations: MPP2 shows a bias towards megakaryocytic and erythroid progenitors, MPP3 is biased towards myeloid and dendritic populations and, lastly, MPP4 shows a lineage bias towards the lymphoid populations (Figure 3.16 e).

Collectively, lineage-specific DMRs were identified within the primitive HSPC cell compartment, supporting the concept of early lineage restriction and providing a potential molecular basis for lineage priming.

3.7 The DNA methylome of aged HSCs is primed towards the megakaryocytic lineage

Our DNA methylation map of murine hematopoiesis provides a rich resource to study DNA methylation programming across a broad range of conditions, such as leukemic transformation, inflammatory challenge, or aging. As a proof-of-concept, we applied our DNA methylation map of normal hematopoiesis to the exploration of the phenomenon of myeloid- and platelet-biased differentiation that has been described in aged HSCs [Dykstra et al., 2011, Sudo et al., 2000, Grover et al., 2016].

We analyzed the hematologic parameters of ten aged mice (>2 years old), five of which were used as donors for DNA methylation profiling of aged HSCs. Aged mice with clear macroscopic abnormalities, such as splenomegaly (>200 mg), strong leukopenia, or infections, were considered as diseased mice and were neither included in the phenotypic analysis nor were they used for DNA methylation profiling of aged HSCs. The hematological parameters of aged HSCs showed high variability, indicat-

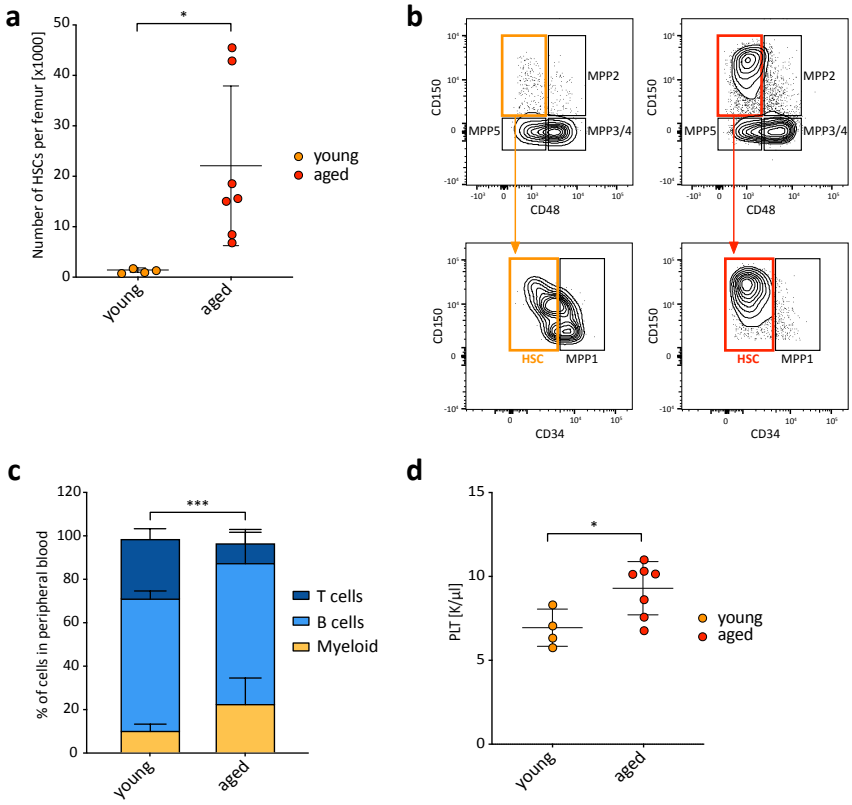

Figure 3.17: Hematopoietic aging is characterized by an increased HSC pool and a differentiation bias towards the myeloid and megakaryocytic lineage. **a**, Absolute number of HSCs per femur in young and aged mice (mean ± standard deviation [SD]). Individual mice are represented as dots. **b**, Representative FACS plots of the HSPC compartment of young and aged mice. **c**, Peripheral blood cell analysis of young and aged mice showing the percentage of myeloid ($Gr1^+$, $Mac1^+$), B cells ($B220^+$), and T cells ($CD4^+$, $CD8a^+$). Mean ± SD are shown for young (n = 4) and aged (n = 7) mice. **d**, Platelet counts in peripheral blood of young and aged mice (mean ± SD). Individual mice are represented as dots. *** = p >0.005, ** = p >0.01, * = p >0.05, unpaired *t*-test with Welch's correction.

ing that aging in general is associated with heterogeneous phenotypes (Table A.2). Consistent with previous studies, our aging cohort showed a 5- to 32-fold expansion of immunophenotypically defined HSCs as compared to young mice (Figure 3.17 a, b). Furthermore, aged mice exhibited a myeloid differentiation bias, which was accompanied by an increase of myeloid cells and a decrease of lymphoid cells in the peripheral blood (Figure 3.17 c). The myeloid differentiation bias was already evident in the hematopoietic progenitor populations, as shown by an expansion of both the myeloid-primed MPP3 population and GMPs (data not shown). In addition to the myeloid bias, the majority of aged mice revealed an increased number of platelets in the peripheral blood, indicating a bias towards megakaryocytic differentiation with

increasing age (Figure 3.17 d). In summary, the analyzed aging cohort recapitulates known phenotypes associated with aging of the murine hematopoietic system.

In order to investigate the remodeling of the DNA methylome in aged HSCs, TWGBS of HSCs isolated from both young and aged mice was performed. Given the fact that phenotypically defined HSCs increase with age, we were able to generate DNA methylomes of aged HSCs isolated from individual two-year-old mice (Figure 3.18 a). The global CpG coverage of aged HSCs was in the range of the young HSC pools (young HSCs: 8.5x - 14.7x; aged HSCs: 4.8x - 19.8x; Table A.2). Furthermore, the mean methylation levels were nearly identical in young and aged HSCs (young: 82.2%; aged: 82.7%). In contrast to previous studies, neither a significant increase of global methylation levels with aging [Sun et al., 2014, Beerman et al., 2013] nor a decrease [Taiwo et al., 2013] of global methylation levels was observed in our samples (Figure 3.18 a).

In total, 3,275 aging-related DMRs (aDMRs) were identified by the pairwise comparison of young and aged HSCs, of which 2,708 lose and 567 gain DNA methylation during the process of aging (Figure 3.18 c). Using the GREAT annotation and enrichment tool, we found hypomethylated aDMRs to be enriched for GO Biological Process terms associated with leukocyte activation, apoptotic signaling, immune response, and myeloid differentiation (Figure 3.18 d). Furthermore, the hypomethylated DMRs were enriched for GO Molecular Function terms associated with small GTPase activity and integrin binding (Figure 3.18 e), both of which are important functions for HSC adhesion and migration as well as for the activation of platelets [Wagers et al., 2002, Cancelas and Williams, 2009, Huveneers and Danen, 2009]. Indeed, aging-specific DNA methylation loss affected genes involved in megakaryopoiesis, such as *Vwf*, *Itgb3*, and *Selp*, and small GTPases, such as *RohJ* (Figure 3.18 f). Interestingly, promoter regions of classical myeloid genes showed similar DNA methylation levels in young and aged mice, including *Spi1*, *Cepba*, *Mpo*, and *Csf1r* (data not shown). Loss of DNA methylation was further observed at the promoter region of the hematopoietic stem cell marker *Slamf1* (Figure 3.18 f), consistent with previous studies showing an upregulation of Slamf1 (CD150) surface marker expression in aged HSCs [Beerman et al., 2010]. Hypermethylated aDMRs were enriched for GO terms associated with developmental processes, such as gastrulation and embryonic organ morphogenesis (data not shown).

Altogether, this data suggests that HSC aging results in remodeling of the DNA

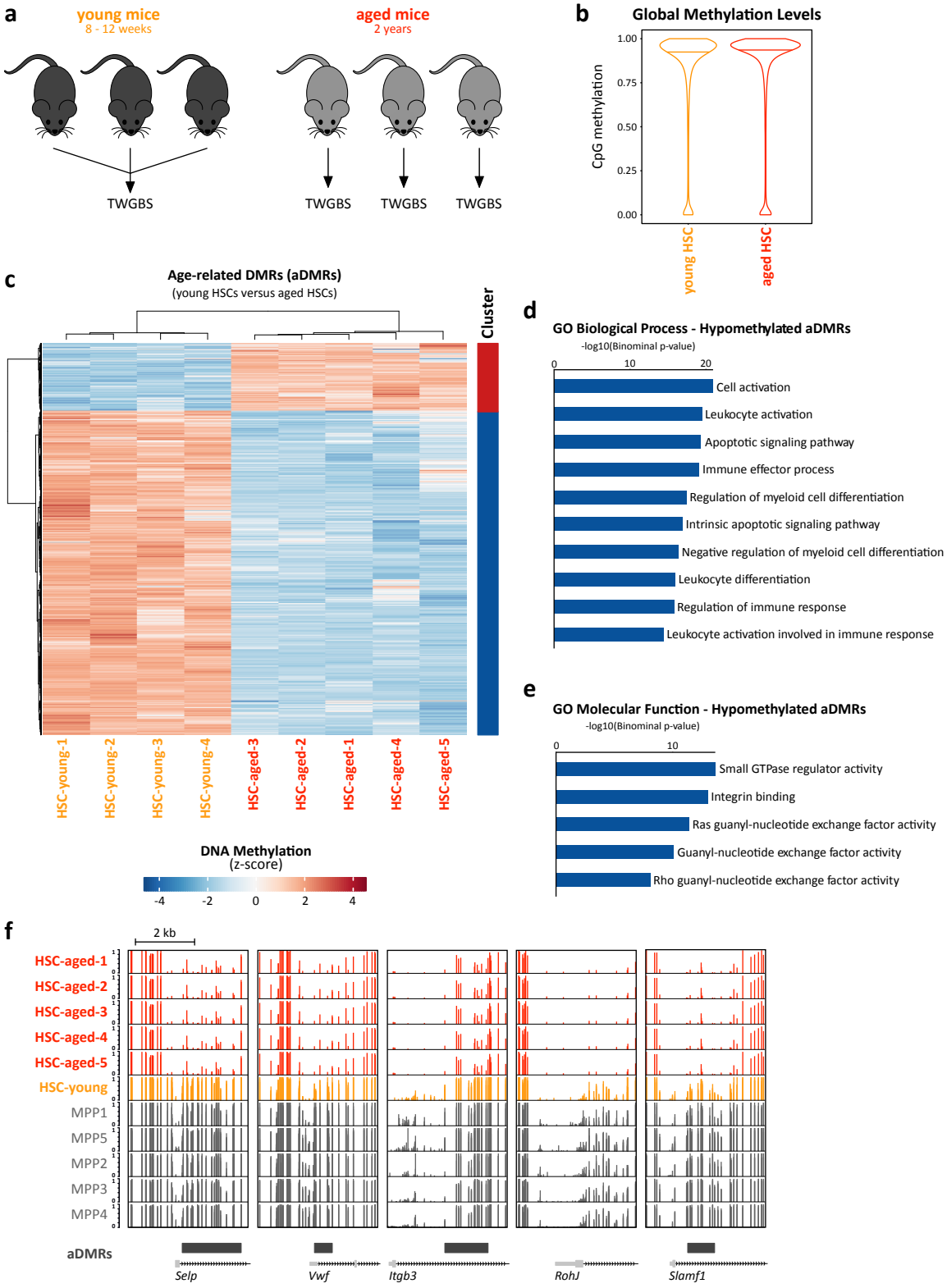

Figure 3.18: Epigenome remodeling of aged HSCs involves DNA methylation loss at megakaryocytic marker genes.

Figure 3.18: Epigenome remodeling of aged HSCs involves DNA methylation loss at megakaryocytic marker genes. TWGBS was performed to generate DNA methylation profiles of young and aged HSCs. **a**, Experimental workflow demonstrating the sample pooling strategy for DNA methylation profiling of young (8-12 weeks) and aged (>2 years) HSCs. **b**, Violin plot show the median CpG methylation levels of $1x10^6$ randomly sampled CpG sites. **c**, DMRs were identified by comparing young and aged HSCs. In total, 3,275 aging-related DMRs (aDMRs) were identified, of which 2,708 were associated with a loss and 567 with a gain of DNA methylation. Heatmap showing aDMRs clustered with Ward's hierarchical agglomerative clustering method using z-score transformed beta-values. Annotation of hypo- (blue) and hyper- (red) methylated aDMRs are shown on the right. **d**, **e**, GO enrichment analysis of hypomethylated aDMRs was performed using GREAT. Bar plot showing the enrichment for the ten most significantly enriched GO Biological Process terms **(d)** and the five most significantly enriched GO Molecular Function terms **(e)**. Displayed are -$\log_{10}$(binominal p-values) for GO terms with a binominal fold enrichment >2. **f**, Genome browser track snapshots depicting DNA methylation levels at TSS of *Selp*, *Vwf*, *Itgb3*, *RohJ*, and *Slamf1*. For aged HSCs, each replicate is shown in a separate track, while merged replicates are depicted for young HSC and MPP populations.

methylome, which is characterized by a molecular differentiation bias towards the megakaryocytic lineage.

To further assess the molecular priming of aged HSCs towards the megakaryocytic lineage, we investigated the DNA methylation profiles of aged HSCs in the context of our DNA methylation map of murine hematopoiesis. Notably, 46% of aDMRs (1,502 aDMRs) overlapped with the DNA methylation programs identified in normal hematopoietic differentiation (Figure 3.19 a). The overlapping aDMRs were found to be enriched for MSigDB pathways associated with integrin signaling, platelet activation, and platelet aggregation (Figure 3.19 b). Along these lines, HOMER motif enrichment analysis revealed an overrepresentation of Gata-like binding motifs, consistent with Gata1 acting as a key transcription factor for megakaryocyte differentiation (Figure 3.19 c). PCA based on DNA methylation levels in hDMRs demonstrated that aged HSCs are closely related to young HSCs. However, it is worth to mention that aged HSCs show a trend towards the megakaryocytic and erythroid progenitor populations on PC2 (Figure 3.19 d).

To identify the lineage-specific DNA methylation programs that are predominantly affected by aging-associated DNA methylome remodeling, we calculated the overlap of hypomethylated aDMRs with the 29 DNA methylation differentiation programs. We focused the analysis on hypomethylated aDMRs, since aging was mainly associated with loss of DNA methylation. Methylation programs 6, 7, and 25 overlapped significantly with hypomethylated aDMRs (5.5%, 4.6%, and 2.9%) as compared to other clusters (mean overlap of all clusters: 1.1%) (Figure 3.20 a-c). Strikingly,

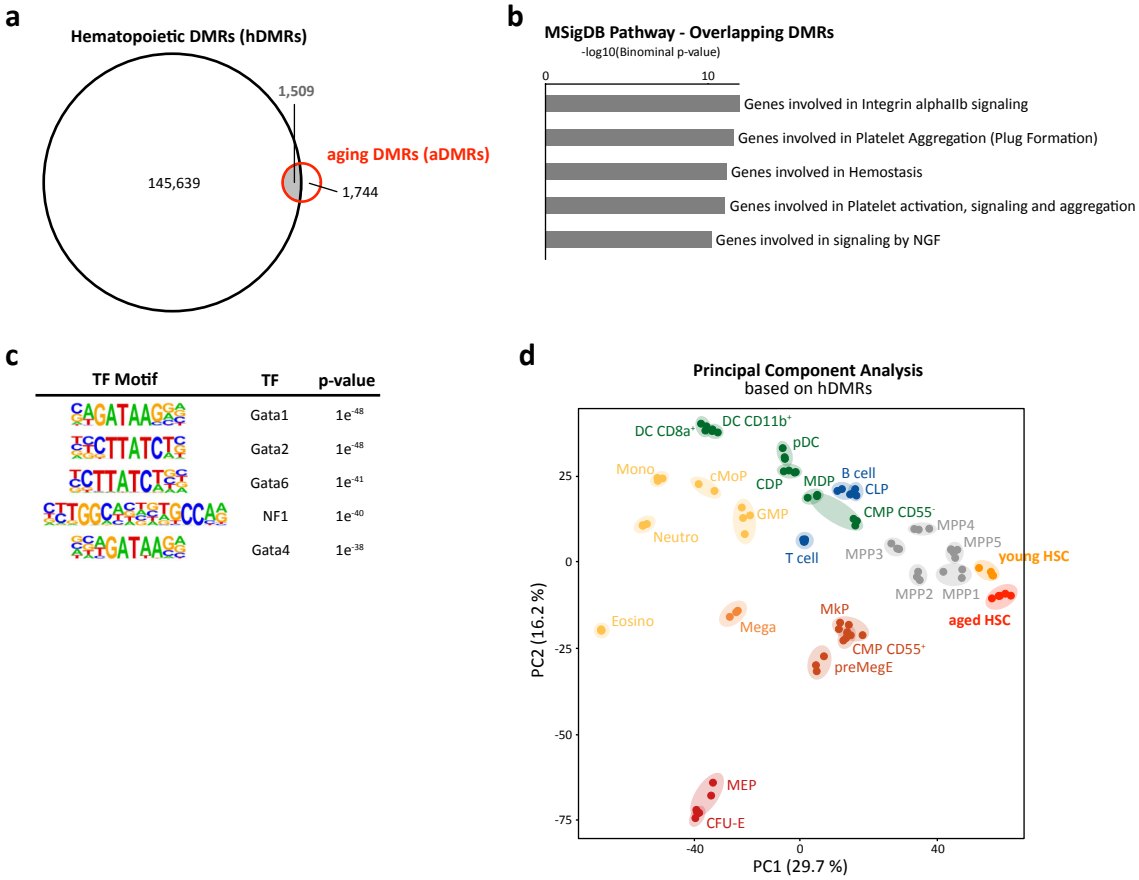

Figure 3.19: Age-related DNA methylation changes overlap with hematopoiesis DMRs and recapitulate features of megakaryocyte differentiation. **a**, Venn diagram illustrating the overlap of hDMRs and aDMRs. **b**, MSigDB pathway enrichment analysis of overlapping DMRs was performed using the GREAT annotation and enrichment tool. Bar plot showing the enrichments for the five most significantly enriched GO Molecular Function terms. Displayed are -$\log_{10}$(binominal p-values) for GO terms with a binominal fold enrichment >2. **c**, Transcription factor motif enrichment analysis of overlapping DMRs was performed using HOMER. Shown are the transcription factor motifs and the corresponding p-values. **d**, PCA based on DNA methylation levels in hDMRs. Each dot represents a biological replicate.

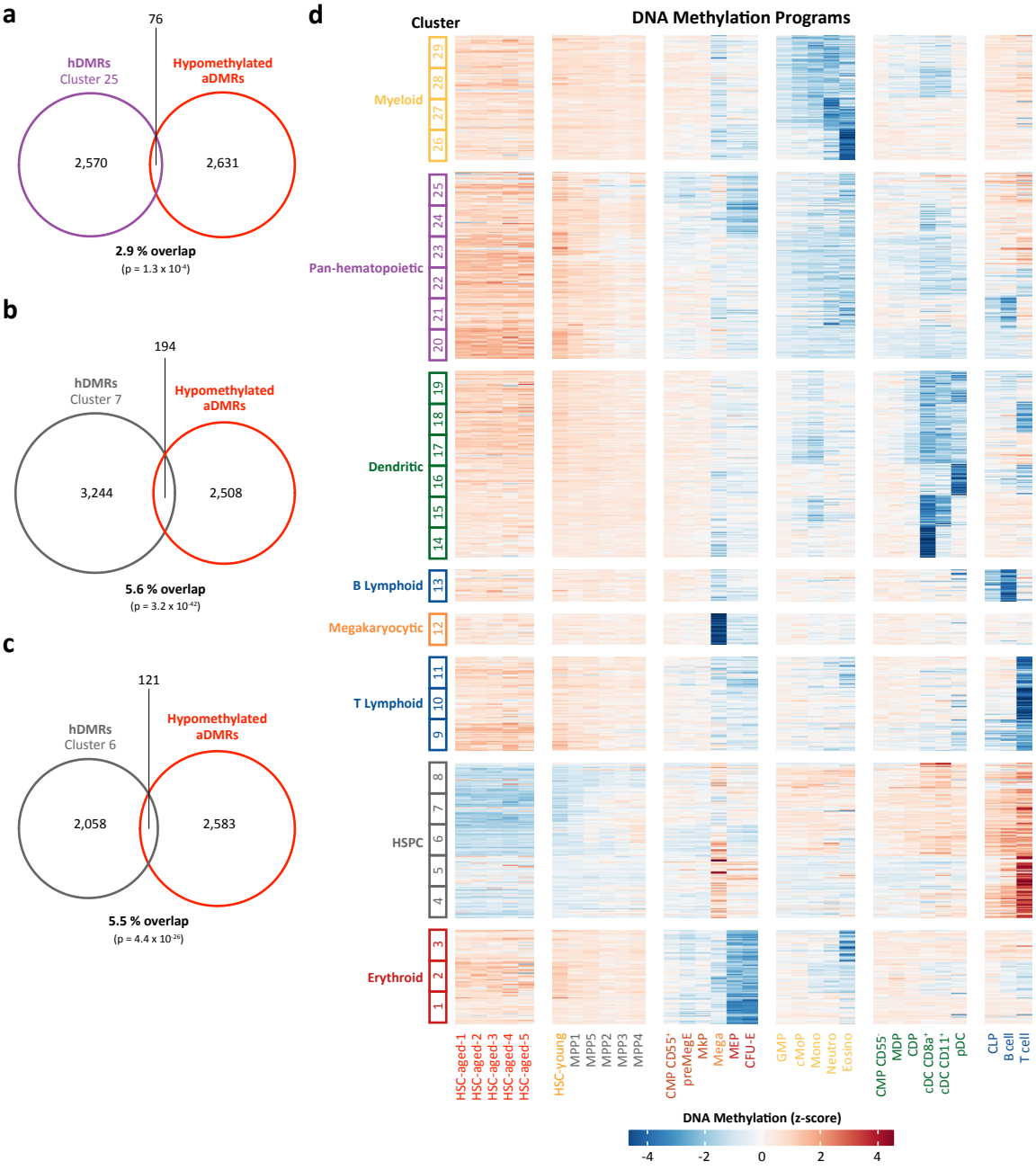

Figure 3.20: Age-related loss of DNA methylation overlaps with hematopoietic DMRs that are characteristic for megakaryocytic differentiation. a-c, Venn diagram illustrating the overlap of aDMRs and hDMRs of DNA methylation cluster 25 **(a)**, cluster 7 **(b)**, and cluster 6 **(c)**. p-values were calculated using a one-sided Fisher's exact test. **d**, Heatmap showing the z-score transformed beta-values in the twenty-nine DNA methylation programs. Annotation of clusters and cluster sets are shown on the left. Annotation of the analyzed cell populations is shown at the bottom.

these DNA methylation programs are characterized by low DNA methylation levels in HSCs, megakaryocytic progenitors, or myeloid populations (Figure 3.20 d).

Collectively, HSC aging causes locus-specific DNA methylation changes that are associated with megakaryocyte-lineage priming. Our DNA methylation data support previous publications suggesting an expansion of megakaryocyte-primed HSCs with age [Grover et al., 2016, Yamamoto et al., 2018, Rundberg Nilsson et al., 2016, Gekas and Graf, 2013].

3.8 The DNA methylome mediates cellular memory of chronic inflammatory stress

During their entire lifespan, organisms are repeatedly exposed to inflammatory stress that drives HSCs out of dormancy and into active cell cycle in order to compensate the loss of mature blood cells. However, experimental mice are housed under restricted pathogen-free conditions, which keep the majority of murine HSCs in a long-term quiescent state. The lack of inflammatory stress might be a simple reason why murine laboratory models fail to recapitulate key features of human hematopoietic aging, such as blood cytopenias and bone marrow hypoplasia. Bogeska *et al.* demonstrated that chronic treatment with the pro-inflammatory agonist pI:pC in early- to mid-life causes a decline of HSC potency and recapitulates clinically relevant features of aged human hematopoiesis in experimental mice (unpublished data from the Milsom group). We hypothesize that inflammatory treatment induces HSC cycling and accelerates epigenetic remodeling, which drives the development of age-related phenotypes.

With the intention to test this hypothesis and to identify stress-induced changes in the HSC methylome, we treated laboratory mice with three rounds of pI:pC- or PBS-injections. Four weeks after the last pI:pC-injection, HSCs were sorted and subjected to TWGBS (Figure 3.21 a). Pairwise DMR calling of pI:pC- versus PBS-treated HSCs identified 468 inflammation-associated DMRs (iDMRs), of which 202 lose and 266 gain DNA methylation upon inflammatory exposure (Figure 3.21 b). GREAT analysis demonstrated that hypomethylated iDMRs were enriched for genes associated with immune response processes and interferon signaling, confirming that pI:pC-treatment indeed reprograms the DNA methylome of HSCs in order to induce

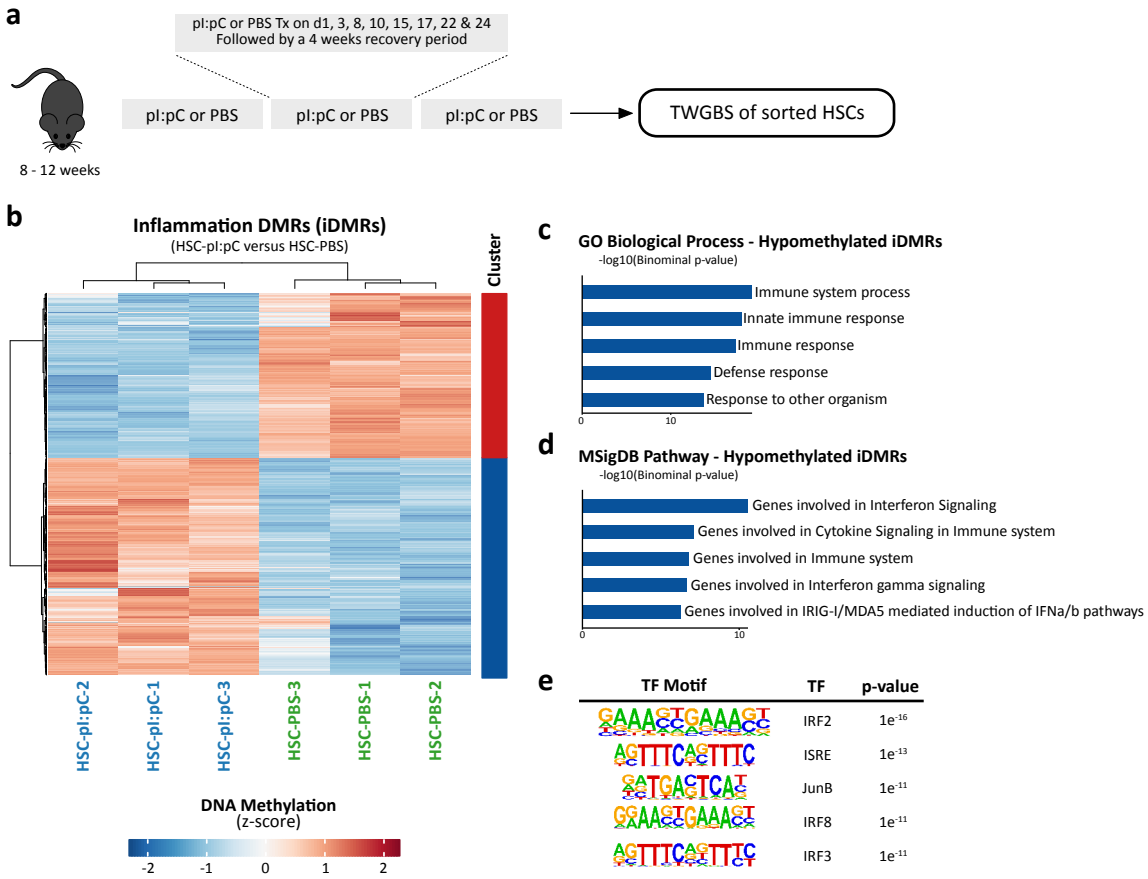

Figure 3.21: Chronic inflammatory stress remodels the DNA methylome of HSCs. a, Schematic of pI:pC-treatment to investigate remodeling of the DNA methylome as a consequence of inflammatory stress. **b**, DMRs were called by pairwise comparison of pI:pC- and PBS-treated HSCs. In total, 468 inflammation-related DMRs (iDMRs) were identified, of which 202 were associated with a loss and 266 with a gain of DNA methylation. Heatmap showing iDMRs clustered with Ward's hierarchical agglomerative clustering method using z-score transformed beta-values. Annotation of hypo- (blue) and hyper- (red) methylated iDMRs is shown on the right. **c**, **d**, GREAT analysis of hypomethylated iDMRs. Bar plot showing the enrichment for the five most significantly enriched GO Biological Process terms **(c)** and MSigDB pathways **(d)**. Displayed are $-\log_{10}$(binominal p-values) for GO terms with a binominal fold enrichment >2. **e**, Transcription factor motif enrichment analysis of hypomethylated iDMRs was performed using HOMER. Shown are the transcription factor motifs and the p-values.

an anti-inflammatory response (Figure 3.21 c, d). Furthermore, motif enrichment analysis revealed an overrepresentation of transcription factor binding motifs, which are involved in regulatory networks of inflammation, such as interferon response factors (IRFs) (Figure 3.21 e). In contrast, hypermethylated iDMRs were enriched for genes associated with developmental processes, such as blood vessel development (data not shown).

In order to assess whether the exposure to inflammatory stress accelerates epigenetic aging of the hematopoietic system in laboratory mice, we next analyzed the DNA methylome of pI:pC-treated HSCs in the context of aDMRs. PCA based on aDMRs suggests that chronic inflammatory stress does not accelerate hematopoietic aging. PC1 (72.5%) reflected hematopoietic aging and ordered the analyzed HSC samples according to the age of their donor mice. Importantly, PC1 did not separate PBS-treated HSCs from pI:pC-treated HSCs. PC2 (9.1%) separated pI:pC-treated HSCs from all other HSC samples, indicating that DNA methylation changes caused by inflammatory stress are distinct from those occurring during chronological hematopoietic aging (Figure 3.22 a). Hierarchical clustering of aDMRs identified four DNA methylation clusters that reveal continuous DNA methylation changes from young (8-12 weeks) to mid-aged (8 months) and finally to aged HSCs (2 years) (Figure 3.22 b). Remarkably, DNA methylation reprogramming in cluster 1 and 2 seemed to be more advanced towards the DNA methylome of aged HSCs in pI:pC-treated HSCs compared to PBS-treated HSCs (Figure 3.22 b). This observation shows that a subset of aDMRs is affected by inflammatory stress, which might be responsible for an accelerated aging phenotype in mice exposed to chronic inflammatory stress. Indeed, enrichment analysis demonstrated that aDMRs of cluster 2 are enriched for biological processes involved in inflammatory signaling and immune response (Figure 3.22 d, c), while aDMRs with a gain of DNA methylation (cluster1) were associated with developmental processes. aDMRs that show only a significant loss of DNA methylation in aged HSCs (cluster 3 and 4) were again associated with leukocyte activation, platelet activation, and small GTPases (data not shown).

Finally, we analyzed the inflammation-induced DNA methylation changes in the context of normal hematopoietic differentiation. Overall, 61% of iDMRs overlapped with hDMRs. Hypomethylated iDMRs significantly overlapped with hDMRs of the pan-hematopoietic clusters 20 ($p = 7.2x10^{-2}$), 23 ($p = 5.1x10^{-3}$), and 25 ($p = 4.5x10^{-13}$)

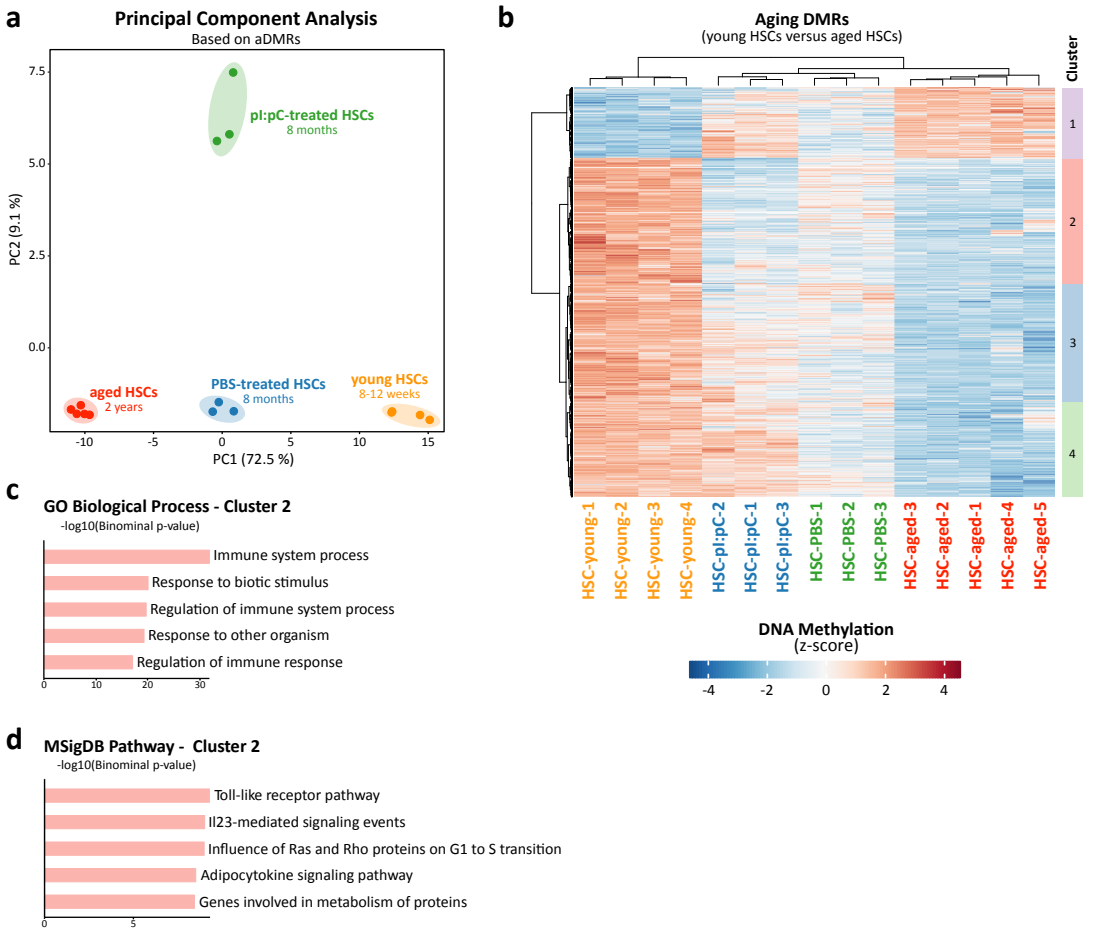

Figure 3.22: A subset of aging-related remodeling events is accelerated upon chronic inflammatory challenge in the HSC epigenome. **a**, PCA based on DNA methylation levels aDMRs. Each dot represents a biological replicate. **b**, Heatmap showing aDMRs clustered with Ward's hierarchical agglomerative clustering method using z-score transformed beta-values. Annotation of DNA methylation clusters is shown on the left. **c**, **d**, GREAT analysis of aDMRs of DNA methylation cluster 2. Bar plot showing the enrichment for the five most significantly enriched GO Biological Process terms **(c)** and MSigDB pathways **(d)**.

(data not shown), which are associated with a loss of DNA methylation in myeloid cell populations, indicating the activation of innate immunity by inflammatory challenge. In contrast, hypermethylated iDMRs significantly overlapped with hDMRs of the HSPC clusters 6 ($p = 1.3x10^{-5}$), 7 ($p = 3.1\ x10^{-308}$), and 8 ($p = 1.0x10^{-20}$) (data not shown).

We further analyzed the inflammation-induced DNA methylation changes in the context of early hematopoiesis. Hierarchical clustering of DNA methylation levels in ehDMRs identified 5 methylation clusters, of which cluster 1 and 3 revealed dif-

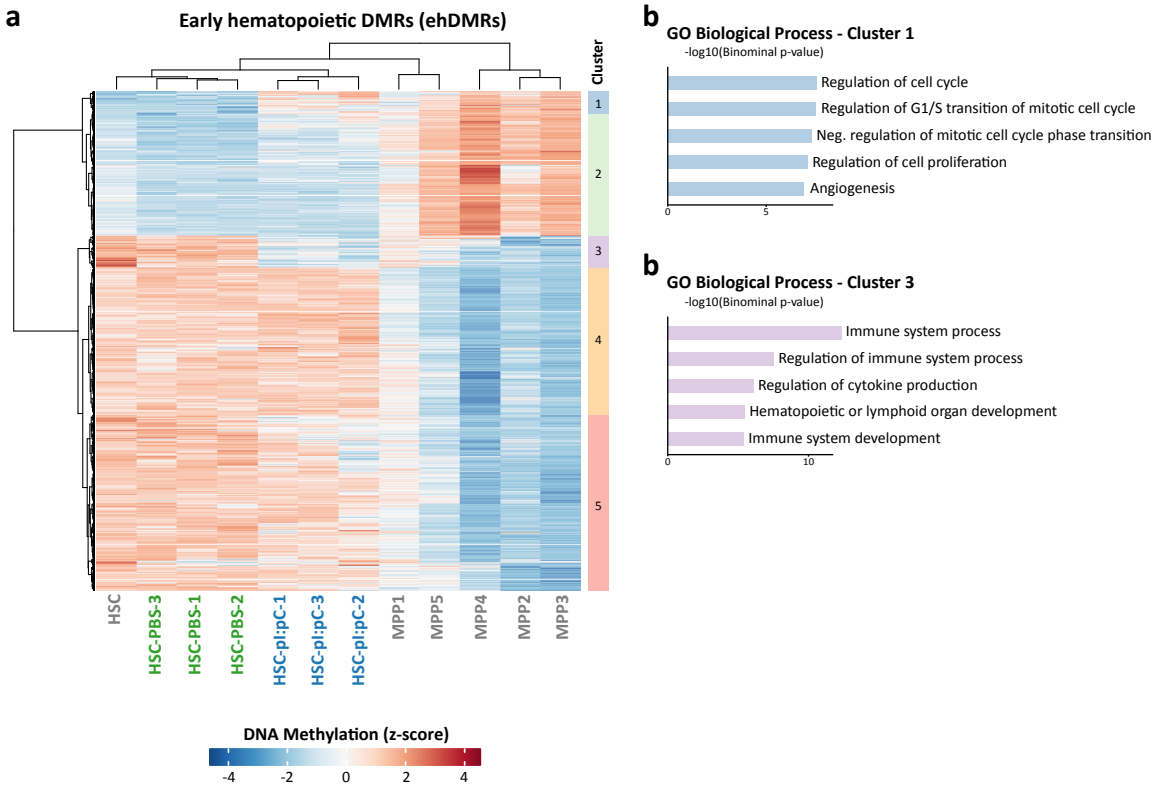

Figure 3.23: pI:pC-treated HSCs acquire DNA methylation patterns of multipotent progenitors. a, Heatmap showing 'early hematopoietic DMRs' (ehDMRs) clustered with Ward's hierarchical agglomerative clustering method using z-score transformed beta-values. Annotation of methylation clusters is shown on the left. **b**, **d**, GREAT enrichment analysis of ehDMR clusters 1 and 3. Bar plot showing the enrichment for the five most significantly enriched GO Biological Process terms for methylation cluster 1 **(b)** and methylation cluster 3 **(c)**. Displayed are -$\log_{10}$(binominal p-values) for GO terms with a binominal fold enrichment >2.

ferential DNA methylation patterns between pI:pC- and PBS-treated HSCs (Figure 3.23 a). Notably, the methylome of pI:pC-treated HSCs resembles the methylomes of MPP populations in cluster 1 and 3. EhDMRs that gain methylation (cluster 1) in pI:pC-treated HSCs compared to PBS-treated HSCs were associated with the regulation of cell cycle, consistent with the exit out of stem cell dormancy (Figure 3.23 b), while ehDMRs that lose DNA methylation (cluster 3) were enriched for GO terms associated with immune system processes (Figure 3.23 c).

In summary, DNA remodeling events were detectable in the DNA methylome of HSCs serially exposed to inflammatory stress even four weeks after the last pI:pC-injection, indicating that extrinsic inflammatory stimuli leave an epigenetic memory

in the DNA methylome of HSCs. Moreover, inflammatory challenge accelerated a subset of aging-related DNA methylation reprogramming events. This finding suggests that aging of the DNA methylome is a function of multiple factors, including inflammatory stimuli over a lifetime. This might explain why HSCs lose functionality as a consequence of inflammation-induced gene expression programs that progressively and irreversibly overwrite the native DNA methylome present in quiescent HSCs. Inflammatory challenge further affected DNA methylation patterns involved in normal hematopoiesis, which reflect the exit from dormancy and the activation of the innate immunity.

4 Discussion

4.1 Mapping global DNA methylation dynamics during hematopoietic differentiation

Hematopoietic differentiation is a dynamic process, which involves widespread but tightly regulated remodeling of the epigenome [Farlik et al., 2016, Corces et al., 2016, Buenrostro et al., 2018, Bock et al., 2012, Cabezas-Wallscheid et al., 2014, Lipka et al., 2014, Lara-Astiaso et al., 2014, Ji et al., 2010, Hodges et al., 2011, Javierre et al., 2016]. Dysregulation of the epigenetic machinery causes abnormal hematopoiesis as exemplified by Dnmt1-hypomorphic HSCs, which show an impaired lymphoid differentiation potential [Broske et al., 2009]. Despite extensive investigations of transcriptional [Velten et al., 2017, Giladi et al., 2018] and epigenetic dynamics during hematopoietic differentiation [Lara-Astiaso et al., 2014, Corces et al., 2016, Buenrostro et al., 2018, Farlik et al., 2016, Bock et al., 2012, Javierre et al., 2016], it is still unclear how lineage commitment is regulated at the molecular level.

The DNA methylation map of murine hematopoiesis, which was established as part of this doctoral thesis, complements molecular maps of murine and human hematopoiesis, which are based on gene expression dynamics [Giladi et al., 2018, Velten et al., 2017, de Graaf et al., 2016], chromatin accessibility profiles [Corces et al., 2016, Buenrostro et al., 2018], histone modifications [Lara-Astiaso et al., 2014], and promoter-enhancer interactions [Javierre et al., 2016]. Future research integrating different molecular layers will enhance our knowledge of lineage commitment and might change the current view of the hematopoietic system.

4.1.1 The DNA methylation map of murine hematopoiesis in the context of published data sets

With the intention of increasing our knowledge of DNA methylation dynamics and its role in lineage commitment, we leveraged tagmentation-based whole genome bisulfite sequencing (TWGBS) to establish a genome-wide DNA methylation map of the entire murine hematopoietic system, encompassing DNA methylomes of 26 immunophenotypically-defined hematopoietic cell populations. To our knowledge, this is the first complete DNA methylome map covering cell populations from all hematopoietic lineages (myeloid, dendritic, erythroid, megakaryocytic, and lymphoid). Previous DNA methylation studies of murine and human hematopoiesis covered only fractions of the DNA methylome map and were mainly focused on the identification of DNA methylation patterns separating myeloid from lymphoid cell populations [Farlik et al., 2016, Bock et al., 2012, Ji et al., 2010]. Notably, we further dissected the DNA methylomes of hematopoietic stem and progenitor cells [Cabezas-Wallscheid et al., 2014, Lipka et al., 2014] in order to investigate the epigenetic patterns underlying early lineage decisions [Velten et al., 2017, Notta et al., 2016, Pietras et al., 2015, Perie et al., 2015, Paul et al., 2015, Rodriguez-Fraticelli et al., 2018].

Whole genome DNA methylation profiling is still a cost-intensive technique and commercial kits for low-input genome-wide DNA methylation analysis are not available. Furthermore, the interpretation of genome-wide DNA methylation data needs advanced and sophisticated bioinformatical skills. This explains why so far only a limited number of genome-wide DNA methylome datasets of hematopoietic cell populations exists.

The previously published DNA methylation map of murine hematopoiesis comprising 13 hematopoietic cell populations was generated using RRBS [Bock et al., 2012], a DNA methylation profiling method that covers only 8% to 10% of all CpGs with a bias towards CpG-rich regions [Lipka et al., 2014]. More importantly, by comparing TWGBS and RRBS data, Cabezas-Wallscheidt *et al.* could demonstrate that 85% of DMRs are exclusively identified by TWGBS [Cabezas-Wallscheid et al., 2014], emphasizing the additional information provided by TWGBS-based DNA methylome profiling. Farlik *et al.* applied µWBGS to generate a genome-wide DNA methylation map of human hematopoiesis comprising 17 cell populations. However, the coverage of this DNA methylation map is much lower compared to our TWGBS-based DNA

methylomes [Farlik et al., 2016].

Overall, we generated highly reproducible biological replicates that were merged to achieve a genome-wide coverage ranging from 12x to 81x per hematopoietic cell population. This is, to our knowledge, the highest coverage of hematopoietic methylomes published so far, enabling the investigation of DNA methylation patterns at the single-molecule level, so-called epialleles [Landan et al., 2012]. Epialleles provide a measure of epigenetic heterogeneity that can be used to identify previously undetected subpopulations and probably to map hematopoietic commitment points.

In accordance with previous DNA methylation studies of the hematopoietic system [Farlik et al., 2016, Bock et al., 2012], DNA methylomes of hematopoietic cell populations accurately reflect their differentiation stage and lineage choice. By pairwise comparison of HSCs with all downstream hematopoietic cell populations, we identified 147,232 unique 'hematopoietic DMRs' (hDMRs). Overall, we found that murine hematopoietic differentiation is predominantly associated with a loss of DNA methylation, as previously observed in human hematopoiesis [Farlik et al., 2016], suggesting that stem cell maintenance is mainly regulated by the suppression of genes involved in hematopoietic differentiation.

4.1.2 Massive loss of DNA methylation during erythropoiesis

Concurrent with data published by Shearstone *et al.* [Shearstone et al., 2011], we observed a massive DNA methylation loss during erythroid differentiation. Interestingly, enrichment analysis demonstrated that the majority of the demethylation events are located near genes that are not involved in erythroid maturation. Along these lines, Bartholdy *et al.* found that DNA demethylation in erythroid cells occurred primarily in intergenic regions and inactive gene bodies forming large partially methylated domains [Bartholdy et al., 2018]. The function of global demethylation during erythropoiesis is still unclear. Based on current data, one may speculate that the global loss of DNA methylation facilitates the rapid proliferation during erythroid differentiation, which is not sustainable with concurrent maintenance of DNA methylation. Moreover, it is conceivable that regulated DNA methylation patterns are secondary owing to the enucleation in terminally differentiated erythrocytes.

4.1.3 Locus-specific DNA hypermethylation is characteristic for lymphoid differentiation

Populations of the lymphoid branch revealed a higher number of hDMRs associated with a gain of DNA methylation as compared to all other blood lineages, a phenomenon that has already been observed by other groups [Farlik et al., 2016, Bock et al., 2012, Ji et al., 2010, Tejedor et al., 2018, Rodriguez-Fraticelli et al., 2018]. Dnmt1-hypomorphic mice have been shown to exhibit reduced DNA methylation maintenance, causing myeloid-biased hematopoiesis with impaired lymphoid output [Broske et al., 2009]. Indeed, HSPCs isolated from Dnmt1-hypomorphic mice showed a global loss of DNA methylation compared to wild-type mice (unpublished data by the Lipka group, generated by Sina Stäble). Hence, the loss of lymphoid-specific hypermethylation events in HSPCs might explain why Dnmt1-hypomorphic mice have an impaired lymphoid output [Broske et al., 2009]. Our data support the hypothesis that locus-specific DNA methylation is essential for lymphoid differentiation and protects HSCs from differentiating into the myeloid or erythroid lineages.

4.1.4 DNA methylation dynamics complement single-cell RNA sequencing data

Single-cell RNA-seq (scRNA-seq) studies have revised the classical hierarchical model of the hematopoietic system, suggesting that hematopoiesis is a continuous rather than a step-wise differentiation process. Furthermore, these studies support the concept of early lineage restriction and indicate that the HSPC compartment is composed of multiple unipotent progenitor cells as opposed to the classical view, which assumed multipotency of HSPCs [Velten et al., 2017, Giladi et al., 2018, Nestorowa et al., 2016, Macaulay et al., 2016]. Despite the single-cell resolution of gene expression data, bulk DNA methylation profiling provides additional molecular information that complements single-cell transcriptome data, as described below.

First, expression data lack any direct information about *cis*-regulatory elements, such as enhancer regions. Accordingly, genome-wide profiling of chromatin accessibility demonstrated that distal regulatory elements more precisely define the identity of hematopoietic cell populations as opposed to gene expression profiles [Corces et al., 2016]. Indeed, many hDMRs are located outside of promoter regions and overlap with *cis*-regulatory features that provide additional information about the cell iden-

tity [Cabezas-Wallscheid et al., 2014, Lara-Astiaso et al., 2014].

Second, gene expression profiles represent a snapshot of commitment decisions that have already been taken and lack any information about when exactly a cell became committed to a particular cell state. In view of its progressive and unidirectional nature, DNA methylation features a molecular barcode ('epigenetic memory') that can be used to map definitive points of hematopoietic commitment. Given that DNA methylation at CpG sites is a binary process (methylated or unmethylated) [Stricker et al., 2017], such epigentic commitment points could still be compatible with stepwise hematopoietic differentiation. Vice versa, one might even speculate that DNA methylation patterns can be used to forecast cell type-specific transcriptional programs and thus the ultimate lineage fate of a differentiating cell.

4.2 DNA methylation programs reflect hematopoietic lineage decisions

In order to identify lineage-specific DNA methylation programs, we performed hierarchical clustering of the 147,232 hDMRs. Using z-score transformed beta-values, we detected twenty-nine DNA methylation clusters, each comprised of hDMRs with similar DNA methylation dynamics throughout the hematopoietic system.

The DNA methylation status of known hematopoietic marker genes, such as *Klf1*, *Mpo*, *Irf8*, and *Cd19*, verified the biological relevance of the identified DNA methylation programs, which were further validated by enrichment analyses of GO terms, transcription factor binding motifs, scRNA-seq marker genes, and hematopoietic enhancer programs. Using these tools, the biological relevance of all DNA methylation clusters, except for the erythroid cluster 2 and the megakaryocyte cluster 12, could be confirmed.

The lack of enrichment for the erythroid cluster 2 can be explained by the global and rather unspecific loss of DNA methylation during erythropoiesis. It is conceivable that a global DNA methylation loss also occurs during megakaryopoiesis. However, the lack of enrichment for the megakaryocyte cluster 12 might question the integrity of the megakaryocyte DNA methylome. If we assume that the sorted cells were indeed megakaryocytes, their DNA methylation profiles are fundamentally different from the profiles of other megakaryocyte progenitors ($CD55^+$ CMP, preMegE, MkP), a phenomenon that has been previously observed by gene expression profiling [Yang

et al., 2017]. Considering this observation, one might speculate that the isolated megakaryocyte population originates directly from HSCs via a shortcut bypassing megakaryocyte progenitors [Notta et al., 2016, Yamamoto et al., 2013, Rodriguez-Fraticelli et al., 2018, Haas et al., 2015]. This possibility could be tested by *in situ* lineage tracing and subsequent analysis of all megakaryocytic cell populations [Rodriguez-Fraticelli et al., 2018]. Therefore, analysis of hDMRs in cluster 12 might uncover novel regulators essential for shortcut megakaryopoiesis.

Future research using epigenome-editing techniques based on deactivated Cas9 (dCas9) fused to an epigenetic effector protein such as Dnmt3a will be valuable to explore the causal function of the identified hDMRs during lineage commitment [Stepper et al., 2017, Huang et al., 2017, Amabile et al., 2016].

The following sections focus on DNA methylation dynamics within the identified methylation programs that substantiate recent findings based on functional and transcriptional data.

4.2.1 DNA methylation patterns indicate the existence of megakaryocyte-primed HSCs at the apex of the hematopoietic system

Multipotency is considered to be a key feature of HSCs. However, nowadays it is accepted that the HSC pool is a heterogenous cell population with differential lineage potentials for the myeloid, lymphoid, and megakaryocytic lineages [Sanjuan-Pla et al., 2013, Yamamoto et al., 2013, Dykstra et al., 2007, Rodriguez-Fraticelli et al., 2018, Velten et al., 2017]. This cellular heterogeneity might be reflected by distinct DNA methylation profiles. Indeed, *in vivo* clonal tracking approaches have shown that the HSC pool consists of functionally distinct HSC clones that are determined by clone-specific DNA methylation profiles [Yu et al., 2017].

In our DNA methylation data set, the HSPC clusters showed, as expected, a significant enrichment for HSPC marker genes. Strikingly, we also found a strong enrichment of MkP marker genes and an overrepresentation of binding motifs for erythroid and megakaryocytic transcription factors. These data suggest that the HSC pool is either contaminated with megakaryocyte progenitors or that the majority of HSCs possess an intrinsic megakaryocytic lineage priming. The second

possibility substantiates by several functional and transcriptional studies. Using a Vwf-eGFP reporter mouse model, Sanjuan-Pla *et al.* identified a Vwf^+ cell population within the HSC pool that shows platelet-specific gene expression patterns and exhibits megakaryocytic-biased lineage output in transplantation assays [Sanjuan-Pla et al., 2013]. Single-cell transplantations and single-cell gene expression analysis of phenotypic HSCs confirmed the existence of a platelet-biased HSC population at the apex of the hematopoietic system [Yamamoto et al., 2013, Guo et al., 2013, Carrelha et al., 2018]. Recently, Rodriguez-Fraticelli *et al.* used transposon tagging to show that megakaryopoiesis is the primary fate of the majority of HSCs under unperturbed conditions [Rodriguez-Fraticelli et al., 2018], a finding that is further substantiated by our DNA methylation data.

4.2.2 The CMP population is a heterogeneous cell population with distinct DNA methylation profiles

Single-cell gene expression profiling and transplantation assays resolved high levels of heterogeneity within surface marker-defined hematopoietic progenitor populations. The CMP population is by far the most heterogeneous population in terms of gene expression and lineage output, questioning the existence of a population with erythroid, megakaryocytic, myeloid, and dendritic lineage potential [Guo et al., 2013, Paul et al., 2015, Buenrostro et al., 2018, Drissen et al., 2019, Perie et al., 2015]. Guo *et al.* identified two functionally distinct CMP subpopulations based on the expression of the surface marker CD55. While $CD55^+$ CMPs mainly produce erythroid and megakaryocytic cells *in vitro*, $CD55^-$ CMPs generate predominantly myeloid cells [Guo et al., 2013]. To illustrate this heterogeneity at the epigenetic level, we generated DNA methylation profiles of $CD55^+$ and $CD55^-$ CMPs. Indeed, we found that these CMP subpopulations feature two different DNA methylomes. Accordingly, $CD55^+$ CMPs clustered with megakaryocytic and erythroid progenitor populations in the DNA methylation diffusion map, while $CD55^-$ CMPs clustered with myeloid and dendritic progenitors. Altogether, our data confirmed the existence of at least two functionally and molecular distinct subpopulations in the CMP compartment. Notably, epiallele analysis might uncover further subpopulations, which potentially reveal unipotent lineage commitment [Paul et al., 2015, Velten et al., 2017].

4.2.3 DNA methylation profiles question the bipotent lineage output of MEPs

MEPs have been defined as bipotent progenitors with both megakaryocytic and erythrocytic lineage potential [Akashi et al., 2000]. Analysis of hematopoietic DNA methylation profiles revealed highly similar patterns for the MEP and the CFU-E subpopulations. Both of these cell populations share DNA methylomes that are significantly distinct from the DNA methylomes of megakaryocyte progenitors. In line with single-cell transcriptome data by Paul *et al.* [Paul et al., 2015], our DNA methylation profiles challenge the bipotency of MEPs and rather suggest a unipotent MEP population with exclusive erythroid lineage potential. The discrepancy between molecular profiles and functional data by Akashi *et al.* can be explained by the fact that *in vitro* differentiation and transplantation experiments do not entirely reflect the lineage fate under physiological conditions. In addition, this might be explained by a minor subpopulation of cells that retained megakaryocyte potential [Akashi et al., 2000].

4.2.4 DNA methylation patterns trace back to the origin of mature blood cells

The epigenetic memory of DNA methylation can be exploited to identify progenitor populations of mature cell types and to infer differentiation routes based on DNA methylation pseudo-dynamics.

For instance, the developmental origin of pDCs is still uncertain. Data published by several groups indicate that pDCs do not only originate from common dendritic progenitors (CDPs) but also from lymphoid progenitors via an alternative differentiation route [Rodrigues et al., 2018, Sathe et al., 2013, Shigematsu et al., 2004]. In agreement with this data, the pDC-specific methylation clusters (cluster 16 and 19) are enriched for binding motifs of lymphoid transcription factors, such as Tcf12, E2a, and Runx1. Furthermore, pDCs cluster between dendritic and lymphoid cell types in the DNA methylation diffusion map. It is conceivable that pDCs originating from different progenitor cell populations represent subpopulations within the sorted pDCs [O'Keeffe et al., 2002, Reizis, 2019]. Single-cell transcriptome or epigenome profiling of murine dendritic cells might further facilitate the mapping of native dendritic differentiation trajectories.

As another example, according to the current models of hematopoiesis, GMPs give rise to all myeloid cell types, including monocytes, neutrophils, eosinophils, basophils, and mast cells. Recently, Drissen *et al.* identified two distinct myeloid differentiation trajectories by single-cell transcriptome profiling that separate before the segregation from other hematopoietic lineages and can be characterized by the differential expression of *Gata1*. Accordingly, progenitor cells (preGMs, pre-granulocyte-macrophage progenitor) expressing *Gata1* generated mast cells, eosinophils, megakaryocytes, and erythrocytes, while progenitor cells (preGMs) lacking *Gata1* generated monocytes, neutrophils, and lymphocytes [Drissen et al., 2016]. In line with the data published by Drissen *et al.*, we observed that eosinophils as well as megakaryocytic and erythroid progenitors share similar DNA methylation patterns (cluster 3, 6, 11, 24, 25 and 26), which were enriched for binding motifs of Gata-like transcription factors, supporting a common differentiation branch of eosinophils, megakaryoctes, and erythrocytes. However, our data provide no evidence that monocytes, neutrophils, and lymphoid cells transition through a common progenitor population. This discrepancy might be explained by the fact that the Gata1^{-} preGM population analyzed by Drissen *et al.* represents a mixture of Flt3^{+} and Flt3^{-} cells [Drissen et al., 2016].

4.3 DNA methylation dynamics enable modeling of the hematopoietic system

During early hematopoietic differentiation, DNA methylation changes are progressive, either showing a continuous loss or gain of DNA methylation from the HSCs to the most differentiated MPP population [Cabezas-Wallscheid et al., 2014, Lipka et al., 2014]. Strikingly, these early DNA methylation changes are conserved in mature blood organs and are specific to hematopoietic cells [Lipka et al., 2014].

4.3.1 Hematopoietic differentiation is defined by progressive and unidirectional DNA methylation changes

Based on the observed DNA methylation dynamics during early hematopoiesis, we hypothesized that hematopoietic differentiation as a whole is associated with progressive and unidirectional DNA methylation changes that are irreversible under physiological conditions. Indeed, we found that DNA methylation dynamics in the

identified DNA methylation programs are invariably unidirectional, either showing a progressive loss or gain of DNA methylation from the HSCs to the most differentiated cell populations in all hematopoietic lineages. These unidirectional DNA methylation changes can be used to order cell types along their hematopoietic differentiation trajectory based on so-called DNA methylation pseudo-dynamics. The methylation pseudo-dynamics were sufficient to infer a model of murine hematopoiesis [Farlik et al., 2016], highlighting the capability of DNA methylation dynamics to deconvolute cellular relationships within regenerating tissues such as the hematopoietic system.

The ultimate proof that DNA methylation changes are irreversible is limited by the lack of *in vivo* real-time DNA methylation capturing methods. Future development of *in vivo* DNA methylation reporters might help to resolve this question [Stelzer et al., 2015].

4.3.2 DNA methylation dynamics during murine hematopoiesis are reminiscent of Waddington's epigenetic landscape

As a result of recent scRNA-seq studies on hematopoiesis, several models describing hematopoietic differentiation at the molecular level have been introduced, including modified models of Waddington's epigenetic landscape [Velten et al., 2017, Laurenti and Gottgens, 2018, Haas et al., 2018, Waddington, 1957]. In this thesis, the progressive and unidirectional DNA methylation changes during hematopoietic differentiation were used to infer a model of hematopoiesis in form of a diffusion map (Figure 3.7 b, Figure 4.1 a). These DNA methylation pseudo-dynamics, as exemplified by a myeloid DNA methylation program, can be illustrated on the basis of the classical epigenetic landscape (Figure 4.1 a, b):

Differentiating cells roll down a mountain and commit towards different cell fates, which are visualized by ridges in the mountain. During this process, differentiating cells progressively acquire DNA methylation commitment marks of lineage- and cell type-specific DNA methylation programs (e.g. myeloid DNA methylation program), as illustrated by arrows in Figure 4.1. The DNA methylation program of the final cell fate becomes more distinct (e.g. myeloid DNA methylation program), while acquired DNA methylation signatures of other lineage-specific methylation programs remain stable after having passed commitment points to other lineages. The mountain slope represents the unidirectional gain or loss of methylation during hematopoiesis.

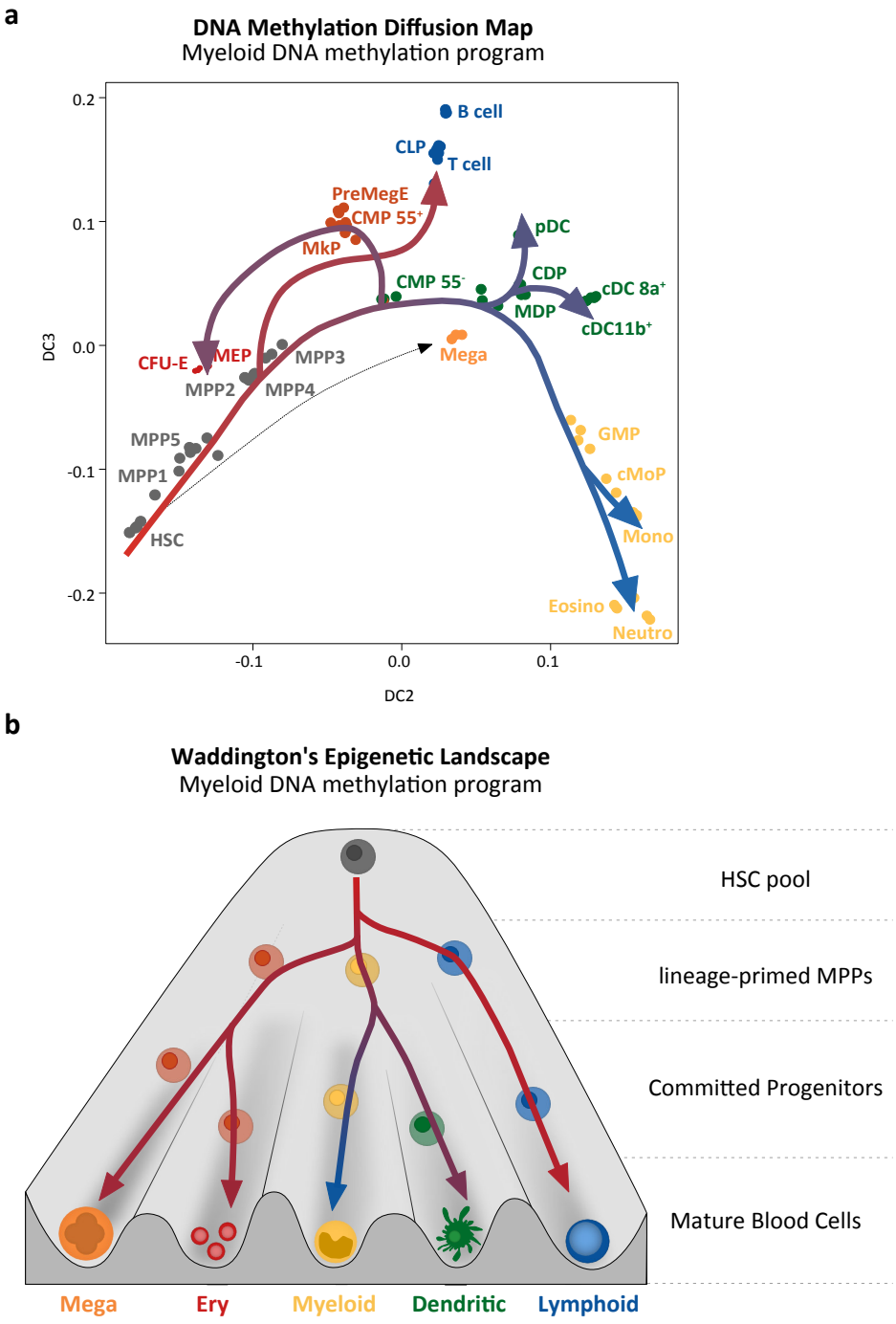

Figure 4.1: The DNA methylation dynamics during murine hematopoiesis can be visualized in Waddington's epigenetic landscape. a, DNA methylation diffusion map with projected progressive and unidirectional DNA methylation changes of a myeloid DNA methylation program. **b**, Waddington's epigenetic landscape with projected progressive and unidirectional DNA methylation changes of a myeloid DNA methylation program. Arrows illustrate DNA methylation dynamics. Red indicates high methylation values, blue indicates low methylation values.

4.4 Integration of epigenome and transcriptome indicates a permissive role for DNA methylation during lineage commitment

The relationship between the epigenome and the transcriptome is relevant to our understanding of the molecular mechanisms underlying cell fate decisions. Initial data on the correlation between DNA methylation and gene expression suggested that promoter DNA methylation inversely correlates with gene expression [McGhee and Ginder, 1979, Ben-Hattar and Jiricny, 1988, Jones, 2012]. However, recent studies revealed that the correlation between promoter methylation and gene expression is rather weak [Farlik et al., 2016, Bock et al., 2012, Cabezas-Wallscheid et al., 2014, Stricker et al., 2017], indicating a more complex relationship between the epigenome and the transcriptome.

To further investigate how DNA methylation changes affect the regulation of gene expression, we integrated our DNA methylation map with a multi-layer single-cell gene expression map of murine hematopoiesis. Remarkably, we found a strong enrichment of cell type-specific marker gene sets in corresponding DNA methylation programs. This finding supports a global anti-correlation between promoter DNA methylation and gene expression. Indeed, projection of DNA methylation levels onto the single-cell gene expression UMAP revealed an inverse correlation between promoter DNA methylation and gene expression for the majority of genes. However, this analysis also indicated that promoter DNA methylation and gene expression does not always correlate within the hematopoietic system. While DNA methylation programs were exclusively programmed in unidirectional manner, a number of gene expression programs were only temporarily activated and lacked the negative correlation with DNA methylation. This finding provides an explanation for the weak correlation between DNA methylation and gene expression programs observed in previous studies [Farlik et al., 2016, Bock et al., 2012, Cabezas-Wallscheid et al., 2014]. Furthermore, loss of DNA methylation was not always associated with an increase in gene expression, suggesting that DNA methylation changes are rather permissive as opposed to being instructive for the activation of gene expression. This observation highlights the interplay of multiple factors and layers required to induce gene expression, including DNA methylation of *cis*-regulatory elements, transcription factor ex-

pression levels, chromatin accessibility, and histone modifications. Along these lines, analysis of transcription factor expression levels and DNA methylation of transcription factor binding motifs indicated that the co-occurrence of transcription factor expression and demethylation of binding motifs is essential for lineage specification. As a final note, data from individual cells covering multiple molecular layers will be necessary to fully understand the temporal sequence of DNA methylation, chromatin accessibility, and gene expression dynamics and to elucidate their complex interactions [Clark et al., 2018, Angermueller et al., 2016].

4.5 Lineage-specific DNA methylation changes in the HSPC compartment reveal early commitment decisions

The multipotent progenitor compartment (MPPs) is composed of distinct subpopulations that are different with respect to their molecular profile and their lineage potential [Oguro et al., 2013, Pietras et al., 2015, Cabezas-Wallscheid et al., 2014, Lipka et al., 2014]. Transplantation assays demonstrated that MPP2 cells predominantly produce erythroid and megakaryocytic cells, while MPP3 and MPP4 cells mainly produce myeloid and lymphoid cells, respectively [Pietras et al., 2015]. The Camargo lab confirmed the functional behavior of the different MPP populations in unperturbed hematopoiesis using transposon lineage tracing [Rodriguez-Fraticelli et al., 2018].

As part of this thesis, the HSPC compartment investigated by Cabezas-Wallscheid *et al.* was further dissected into six subpopulations (HSC, MPP1, MPP2, MPP3 MPP4, and MPP5) with the aim to investigate the existence of early commitment decisions at the epigenetic level.

Differential DNA methylation dynamics of MPP2, MPP3, and MPP4 were observed in the cell type- and lineage-specific DNA methylation programs. Based on this finding, a focus was set on a more detailed analysis of the DNA methylation changes that were exclusively identified within the HSPC compartment. Strikingly, these early DNA methylation changes were sufficient to segregate terminal cell fate decisions in a diffusion map as well as in a phylogenetic tree, indicating that lineage commitment decisions are already taken in the primitive HSPC compartment. In line with functional data [Pietras et al., 2015, Rodriguez-Fraticelli et al., 2018], the

visualization of early DNA methylation dynamics in the diffusion map revealed a lineage bias of MPP2 towards erythropoiesis and megakaryopoiesis, of MPP3 towards myelopoiesis, and of MPP4 towards lymphopoiesis. However, due to the lack of single-cell methylomes, it remains unclear whether progenitor populations downstream of HSCs indeed consist of unilineage-primed progenitor cells [Velten et al., 2017, Nestorowa et al., 2016] or whether at least some cells retain bi- or multilineage potential [Karamitros et al., 2018].

Previous studies demonstrated that the MPP1 and MPP5 populations possess multipotent lineage potential, but had reduced engraftment efficiency compared to HSCs [Oguro et al., 2013, Yamamoto et al., 2013]. This finding is consistent with our DNA methylation data ordering MPP1 and MPP5 downstream of HSCs in the DNA methylation diffusion map. In addition, both populations lack DNA methylation signatures of mature cell types, suggesting that both populations possess multipotent differentiation capacity. However, similar to MPPs, single-cell DNA methylation profiling might uncover heterogeneity and unipotent lineage priming in HSC, MPP1, and MPP5 populations.

4.6 The DNA methylation map represents a rich resource to study DNA methylation reprogramming in abnormal hematopoiesis

The comprehensive DNA methylation map of murine hematopoiesis provides a rich resource to study aberrant DNA methylation patterns across a broad range of conditions. We applied the DNA methylation map to investigate the phenomenon of myeloid- and platelet-biased differentiation that has been observed in aged HSCs [Dykstra et al., 2011, Sudo et al., 2000, Grover et al., 2016, Morrison et al., 1996].

4.6.1 DNA methylomes of aged HSCs indicate an expansion of megakaryocyte-primed HSCs with age

Aging is associated with functional and molecular alterations of HSCs [Dykstra et al., 2011, Sudo et al., 2000, Grover et al., 2016, Morrison et al., 1996]. Despite the stability of the DNA methylome, the HSC epigenome progressively shifts over a lifetime [Sun et al., 2014, Beerman et al., 2013, Adelman et al., 2019, Florian et al.,

2012]. Accordingly, DNA methylation changes have been used as an epigenetic clock to measure the chronological age of an organism [Stubbs et al., 2017, Horvath, 2013].

The epigenetic drift during aging certainly plays a fundamental role in the age-related dysfunction of HSCs. Thus, we have generated genome-wide DNA methylomes of young and aged HSCs. Our DNA methylation data showed no significant differences in global DNA methylation levels between young and aged HSCs, a finding that is inconsistent with published data either showing a gain [Beerman et al., 2013, Sun et al., 2014] or loss [Taiwo et al., 2013] of global DNA methylation levels with age. This discrepancy is most likely due to the use of different DNA methylation profiling technologies and sorting strategies of HSCs.

Overall, 3,275 aging-related DMRs (aDMRs) were identified in aged HSCs. Of these aging-related remodeling events, more than 80% were associated with a loss of DNA methylation and affected genes that are involved in megakaryopoiesis, such as *Vwf*, *Itgb3* (*CD61*), *Selp*, and small GTPases [Bianchi et al., 2016]. In line with this, Florian *et al.* demonstrated that HSC aging is characterized by an increased activity of the Rho GTPase Cdc42, which is accompanied by the loss of its polar distribution [Florian et al., 2012]. Besides their function in megakaryopoiesis and platelet activation, Rho GTPases play a fundamental role in HSC adhesion and migration [Williams et al., 2008]. This is controversial, as aged HSCs show a reduced repopulation potential in transplantation experiments. However, this discrepancy can be explained by the overall impaired fitness of aged HSCs.

Remarkably, 46% of aging-related DNA methylation changes overlapped with hDMRs identified in normal hematopoiesis. The overlapping aDMRs were associated with genes involved in integrin signaling and platelet activation. Moreover, they significantly affected DNA methylation programs that are characterized by low DNA methylation levels in megakaryocyte-primed progenitor populations. In other words, HSC aging results in remodeling of the DNA methylome *in vivo*, in a manner that is associated with a molecular differentiation bias towards megakaryopoiesis. This is in accordance with previous studies using single-cell transcriptomics and functional single-cell assays, demonstrating an expansion of platelet-biased HSCs with age [Grover et al., 2016, Yamamoto et al., 2018, Rundberg Nilsson et al., 2016]. Along these lines, Gekas *et al.* revealed that Itga2b (CD41), a platelet integrin that interacts with Itgb3 (CD61), is expressed on a subtype of HSCs that increases with age [Gekas and Graf, 2013].

Despite the myeloid-lineage bias characteristic of aged HSCs, aging-related DNA

methylation changes did not affect key myeloid genes or programs. One might speculate that extrinsic factors, such as cytokines and chemokines, produced by the aged bone marrow niche are responsible for the myeloid bias of aged HSCs [Ho et al., 2019, Ergen et al., 2012], while the megakaryocytic-bias is intrinsically programmed.

4.6.2 Chronic inflammatory stress accelerates aging of the HSC methylome

Experimental mice are housed under restricted pathogen-free conditions, which keep murine HSCs predominantly in a dormant state. In this context, murine HSCs contribute only minimally to the blood production under homeostatic conditions [Busch et al., 2015, Sun et al., 2014], but can be forced out of dormancy by inflammatory challenge [Essers et al., 2009, Baldridge et al., 2010, Takizawa et al., 2011, Walter et al., 2015].

The lack of HSC division in experimental mice limits the accumulation of DNA damage [Walter et al., 2015] and likely remodeling of the epigenome. This in turn prevents the onset of age-related phenotypes observed in human hematopoietic aging, such as blood cytopenias, bone marrow hypoplasia, and the development of hematological malignancies. Along these lines, Adelman *et al.* recently demonstrated that the epigenome of aged human HSCs is associated with reprogramming of cancer pathways that are also affected in acute myeloid leukemia (AML) [Adelman et al., 2019].

Using a pulsed H2B-GFP label to track the proliferative history of HSCs *in vivo*, the Moore group indeed showed that slowly cycling aged HSCs, which retain the H2B-GFP label (label retaining cell; LRC), are functionally similar to young HSCs. In contrast, non-LRC HSCs from aged mice were functionally compromised and showed a megakaryocyte-restricted potential *in vitro* [Bernitz et al., 2016, Qiu et al., 2014]. In addition, Beerman *et al.* have shown that increased proliferation induced by serial 5-Fluorouracil (5-FU) treatment causes functional decline in HSCs and DNA methylation reprogramming that recapitulates physiological HSC aging [Beerman et al., 2013]. Unpublished data from the Milsom group (Bogeska *et al.*) revealed that chronic inflammatory challenge of laboratory mice with pI:pC leads to a decline in HSC potency and recapitulates clinical features of aged human hematopoiesis. Altogether, these data suggest that HSC aging can be accelerated by external stimuli that force HSCs into cycle and, as a consequence of increased proliferation, can

subsequently reprogram the HSC epigenome.

Remodeling of the HSC methylome caused by inflammatory challenge was investigated in mice that have been serially treated with pI:pC. Despite a four-week recovery period after the last pI:pC-injection, the HSC methylomes exhibited remnants of an inflammatory signature. This finding indicated that gene expression programs activated by repeated inflammatory stimuli leave molecular scars in form of aberrant DNA methylation patterns.

Furthermore, inflammatory challenge accelerated aging-related DNA methylation reprogramming, which might be responsible for the premature aging phenotype observed in pI:pC-treated mice (Bogeska *et al.*, unpublished data). In line with these findings, studies from other groups have shown that the epigenetic clock is affected by biological interventions such as diet [Stubbs et al., 2017] or infection [Horvath and Levine, 2015, Gross et al., 2016]. For example, HIV^+ patients were shown to have an accelerated epigenetic clock compared to healthy individuals [Gross et al., 2016]. As a final note, it is possible that the inflammation-associated DNA methylome reflects an HSC subpopulation that has expanded in response to the inflammatory challenge. Indeed, Mann *et al.* identified an inflammation responsive subpopulation of HSCs marked by CD61 expression, which expands with age and upon inflammatory challenge [Mann et al., 2018].

Lastly, we could show that inflammatory challenge affected DNA methylation patterns involved in normal hematopoiesis, reflecting the exit from dormancy and the activation of myeloid methylation programs in order to induce innate immunity.

Altogether, our data suggests that aging of the HSC methylome is, among other factors, a function of inflammatory stimuli per life cycle and indicates that inflammatory challenges leave an epigenetic memory of activated transcriptional programs, which irreversibly overwrite the pristine DNA methylome of HSCs resulting in their functional decline.

4.7 Conclusion & Perspective

In this doctoral thesis, a genome-wide DNA methylation map of the murine hematopoietic system was generated, which provided broad insights into how DNA methylation is programmed during hematopoietic commitment. This DNA methylation map will serve as a resource to investigate altered DNA patterns that emerge from all kinds

of extrinsic and extrinsic influences such as aging, inflammatory stress, or malignant transformation. The current technological developments of single-cell omics will enable the generation of single-cell DNA methylomes, which will be essential to accurately model the hematopoietic system and to investigate the existence of multipotent cells in the HSPC compartment.

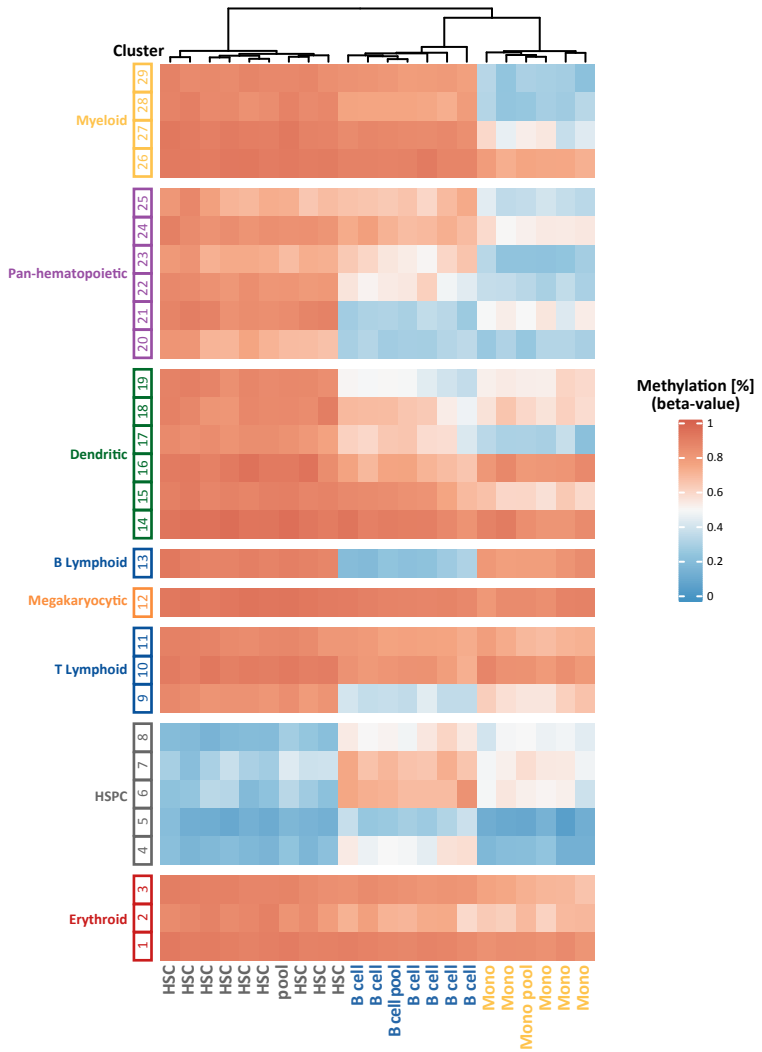

Figure 4.2: Aggregated DNA methylation patterns in lineage-specific DNA methylation programs separates single-cell DNA methylomes according to cell type. Single-cell DNA methylome libraries were generated from sorted HSCs, B cells, and monocytes. The CpGs was merged per DNA methylation program. Columns were arranged by hierarchical clustering based on Euclidean distance.

The current challenge of analyzing single-cell DNA methylomes lies in the sparsity of CpG methylation data, covering only 5% to 10% of all CpGs [Clark et al., 2017]. In a pilot experiment, we have generated single-cell DNA methylomes of HSCs, monocytes, and B cells according to the protocol published by Clark *et al.* (in collaboration with Dr. Mark Hartmann and Maximilian Schönung) [Clark et al., 2017]. Hierarchical clustering of the single-cell DNA methylomes based on CpG sites covered in all single cells failed to cluster the DNA methylomes by cell type (data not shown). On the contrary, merging of the CpGs per DNA methylation cluster separated the single-cell DNA methylomes according to their cell type and recapitulated the DNA methylation patterns of bulk-sorted HSCs, monocytes, and B cells (Figure 6.1). This is in line with recently published single-cell DNA methylome data showing that HSCs and LSKs only clustered apart using unsupervised hierarchical clustering based on CpG sites within DMRs identified in early hematopoiesis by Cabezas-Wallscheid *et al.* [Hui et al., 2018, Cabezas-Wallscheid et al., 2014]. Hence, the identified hDMRs will serve as a reference of genomic regions to analyze sparse single-cell DNA methylation profiles of hematopoietic cells.

In follow-up experiments for this project, single-cell DNA methylome and transcriptome data will be generated from the same cells of index-sorted HSPCs [Clark et al., 2018]. The resulting single-cell DNA methylome data will allow us to resolve the question whether epigenetic commitment marks of different hematopoietic lineages can co-exist in individual hematopoietic stem and progenitor cells. This information is essential to fully understand hematopoietic differentiation. Furthermore, this data will allow us to investigate the temporal sequence of DNA methylation and gene expression dynamics in a single cell.

5

Materials & Methods

5.1 Animals

All mice were housed under specified pathogen-free (SPF) conditions in individually ventilated cages (IVC) in the animal facilities of the German Cancer Research Center (DKFZ) or the University of Ulm. The experimental procedures were approved by the German Regierungspräsidium Karlsruhe für Tierschutz und Arzneimittelüberwachung. C57BL/6J mice were purchased from Harlan Laboratories. Aged C57BL/6J were provided by the laboratory of Hartmut Geiger (University of Ulm).

5.1.1 Chronic treatment with pI:pC

To induce cycling of HSCs *in vivo*, 8- to 12-week-old C57BL/6J mice were intraperitoneally (i.p.) injected with 5 mg kg^{-1} pI:pC (InvivoGen). Injections of PBS (Sigma-Aldrich) served as negative controls.

5.2 Analysis of murine blood and bone marrow

Blood was taken from the murine facial vein and collected in K3 EDTA coated tubes (Sarstedt). Bone marrow cells were isolated by flushing dissected femora, tibiae, and hips using a 23 gauge needle (Medoject) and 1 mL syringe (Hemke Sass Wolf) and collecting the cells in PBS (Sigma-Aldrich) supplemented with 2% FCS (Gibco).

5.2.1 Blood counts and bone marrow cellularity

Peripheral blood cell counts and bone marrow cellularity per femur were determined using a HemaVet 950 FS veterinary hematology system (Drew Scientific).

5.2.2 Flow cytometric analysis of peripheral blood and bone marrow

Cells were stained with respective antibody panels in PBS supplemented with 2% FCS (PBS/FSC) for at least 20 min at 4°C. All FACS antibodies were titrated and applied in appropriate dilutions. For flow cytometric analysis, samples were diluted in PBS/FCS, filtered, and analyzed with a LSRII or Fortessa flow cytometer (BD Biosciences). FACS data was evaluated using FlowJo10.

Periperhal blood staining For the analysis of blood composition, 30 µL of peripheral blood was stained with 70 µL of antibody mix in PBS/FSC as detailed in Table 5.1. After 1 h of incubation, red blood cells were lysed with 1 mL of Ammonium-Chloride-Potassium (ACK)-lysis buffer (Lonza) for 10 min. Cells were washed once with PBS/FCS, resuspended in 200 µL of PBS/FCS containing 5 µg/mL of 7-amino actinomycin D (7-AAD, Thermo Fisher Scientific), and finally FACS analyzed. 7-AAD staining was used for the exclusion of dead cells.

Table 5.1: Antibody panel for flow cytometric analysis of peripheral blood.

Antigen	Clone	Fluorophore	Company
CD11b	M1/70	APC	Invitrogen
Gr1	RB6-8C5	APC	eBioscience
B220	RA3-6B2	APC	eBioscience
CD4	GK1.5	PE-Cy7	Invitrogen
CD8a	53-6.7	PE-Cy7	eBioscience
B220	RA3-6B2	PE-Cy7	eBioscience

Bone marrow stainings Bone marrow cells (maximal 2x10^7 cells in 200 µl PBS/FCS) were stained with antibody panels as detailed in Table 5.2 and 5.3. After 30 min of incubation, red blood cells were lysed for 10 min with 1 mL of ACK-lysis buffer. Cells were washed once with PBS/FCS, resuspended in 200 µL of PBS/FCS, and finally FACS analyzed.

Table 5.2: Antibody panel for flow cytometric analysis of HSPC populations in the bone marrow.

Antigen	Clone	Fluorophore	Company
CD4	GK1.5	PE-Cy7	Invitrogen
CD8a	53-6.7	PE-Cy7	eBioscience
CD11b	M1/70	PE-Cy7	eBioscience
B220	RA3-6B2	PE-Cy7	eBioscience
Gr1	RB6-8C5	PE-Cy7	eBioscience
Ter119	TER-119	PE-Cy7	eBioscience
cKit (CD117)	2B8	APC	Invitrogen
Sca1	D7	APC-Cy7	BD Bioscience
CD150 (SLAM)	TC15-12F12.2	PE-Cy5	BioLegend
CD48	HM48-1	Pacific Blue	BioLegend
CD34	RAM34	FITC	eBioscience
CD135	A2F10	PE	eBioscience

Table 5.3: Antibody panel for flow cytometric analysis of committed progenitor populations in the bone marrow.

Antigen	Clone	Fluorophore	Company
CD4	GK1.5	PE-Cy7	Invitrogen
CD8a	53-6.7	PE-Cy7	eBioscience
CD11b	M1/70	PE-Cy7	eBioscience
B220	RA3-6B2	PE-Cy7	eBioscience
Gr1	RB6-8C5	PE-Cy7	eBioscience
Ter119	TER-119	PE-Cy7	eBioscience
cKit (CD117)	2B8	APC	Invitrogen
Sca1	D7	APC-Cy7	BD Bioscience
CD16/32	93	eFluor 450	eBioscience
CD34	RAM34	FITC	eBioscience
CD65	RIKO-3	PE	BioLegend

5.3 Isolation of surface marker-defined hematopoietic cell populations

5.3.1 Isolation of murine bone marrow cells

Bone marrow cells were collected from dissected femora, tibiae, hips, and spines of sacrificed mice by crushing bones in Iscove´s modified Dulbecco´s medium (IMDM, Gibco) and filtering the cell suspension through a 40 µm cell strainer (Greiner Bio-One).

5.3.2 Enrichment of lineage negative bone marrow cells

Bone marrow cells were collected as described in Section 5.3.1. Next, low-density mononuclear cells (LDMNCs) were enriched by density gradient centrifugation using Histopaque 1083 (Sigma-Aldrich). Briefly, 6 mL cell suspension with a maximal cell density of $3x10^7$ cells mL^{-1} were gently layered onto 6 mL of Histopaque 1083 in a 15

mL tube and centrifuged with 400 g for 20 min at 20°C and brake off. LDMNCs were harvested by carefully aspirating the upper layer, interphase, and Histopaque phase. To increase the number of LDMNCs, the remaining cell pellet was resuspended in 6 mL IMDM and subjected to another round of density gradient centrifugation.

LDMNCs were collected in 3 mL of PBS/FCS and then incubated with a cocktail of biotin-conjugated rat anti-mouse lineage antibodies (Table 5.4) at a concentration of 112.6 µL of lineage cocktail per $1x10^8$ cells for 30 min at 4°C. The labeled LDMNCs were washed once with PBS/FCS and incubated with anti-rat IgG-coated magnetic Dynabeads from the Dynabeads Untouched Mouse T Cells Kit (Invitrogen) at a concentration of 1 mL beads per 1 x 10^8 cells in PBS/FCS for 45 min at 4°C. Finally, lineage negative bone marrow cells were enriched by magnetic depletion of lineage positive cells using the DynaI MPC-6 magnet (Invitrogen). The resulting supernatant containing the purified lineage negative bone marrow cells was used for further purification of hematopoietic stem and progenitor cells by flow cytometric cell sorting.

Table 5.4: Cocktail of biotin-conjugated rat anti-mouse antibodies.

Antigen	Clone	Concentration	Company
CD5	53-7.3	4.2 µg mL^{-1}	BD Bioscience
CD45R/B220	RA3-6B2	2.8 µg mL^{-1}	BD Bioscience
CD11b/Mac1α	M1/70	2.6 µg mL^{-1}	BD Bioscience
CD8a	53-6.7	4.2 µg mL^{-1}	BD Bioscience
Ly-6G and Ly-6C/Gr1	RB6-8C5	2.4 µg mL^{-1}	BD Bioscience
Ter119	TER-119	2.6 µg mL^{-1}	BD Bioscience

5.3.3 Isolation of murine splenic cells

Splenic cells were isolated by pressing dissected spleens from sacrificed mice through a 40 µm cell strainer using a 2 mL syringe plunger and collecting the cells in PBS/FCS.

5.3.4 Fluorescence activated cell sorting of hematopoietic cell populations

Cells were stained with respective antibody panels in PBS/FSC for at least 20 min at 4°C. Before cell sorting, samples were diluted in PBS/FCS and filtered through a 40 µm cell strainer. All hematopoietic cell populations were sorted into 1.5 µL tubes containing 300 µL of PBS/FCS using an Aria I, Aria II, or Fusion cell sorter (BD Biosciences) and a 100 µm nozzle. An aliquot of sorted cells was taken to validate

the purity of the sorted cell population in a re-sort.

Gating strategies are shown in Figure 3.1 b, A.1, and A.2. For TWGBS, sorted cells were pelleted and snap-frozen in liquid nitrogen. Cell pellets were stored at -80°C until further processing.

Cell sorting of hematopoietic cell populations from total bone marrow The following hematopoietic cell populations were sorted from total bone marrow cells (5.3.1): monocytes, neutrophils, eosinophils, B cells, and T cells. For this purpose, total bone marrow cells at a concentration $1x10^7$ cells mL^{-1} were incubated with different panels of antibodies diluted in PBS/FCS for 20 min at 4°C as detailed in Table 5.5. After washing, the labeled total bone marrow cells were resuspended in PBS/FCS and finally subjected to fluorescence activated cell sorting.

Cell sorting of hematopoietic cell population from lineage negative cells The following hematopoietic cell populations were sorted from lineage negative bone marrow cells (5.3.2): HSC, MPP1, MPP5, MPP2, MPP3, MPP4, GMP, MEP, CMP $CD55^+$, CMP $CD55^-$, CLP, preMegE, MkP, CFU-E, MDP^*, CDP^*, and $cMoP^*$ (* enrichment of lineage negative cells without density gradient centrifugation step). For this purpose, $1x10^7$ cells mL^{-1} lineage negative cells were incubated with different panels of antibodies diluted in PBS/FCS for 20 min at 4°C as detailed in Table 5.5. After washing, the labeled lineage negative cells were resuspended in PBS/FCS and finally subjected to fluorescence activated cell sorting.

Cell sorting of hematopoietic cell populations from spleen The following hematopoietic cell populations were sorted from splenic cell suspensions (Section 5.3.3): pDC, cDC $CD8a^+$, and cDC $CD11b^+$ (cells were isolated by Dr. Melinda Czeh). For this purpose, splenocytes at a concentration $1x10^7$ cells mL^{-1} were incubated with different panels of antibodies diluted in PBS/FCS for 20 min at 4°C as detailed in Table 5.5. After washing, the labeled cells were resuspended in PBS/FCS and finally subjected to fluorescence activated cell sorting. 7-AAD was used for the exclusion of dead cells.

Cell sorting of murine megakaryocytes from total bone marrow With the aim to isolate diploid megakaryocytes, total bone marrow cells were fixed, permeabilized and stained with an intercalating DNA dye. In detail, murine bone marrow cells

of one mouse were diluted in 10 mL PBS/FCS and labeled with 100 µL of self-made rat anti-mouse lineage cocktail (provided by the lab of Marieke Essers) for 1 h at 4°C. After washing with PBS/FCS, the cells were incubated with 1.5 mL of anti-rat IgG-coated magnetic Dynabeads for 30 min at 4°C. Lineage negative bone marrow cells were enriched as described before by magnetic depletion of lineage positive cells using the DynaI MPC-6 magnet (Invitrogen). For cell surface staining, $1x10^7$ cells mL^{-1} lineage negative cells were incubated with 200 µL of antibody mix diluted in PBS/FCS for 45 min at 4°C as detailed in Table 5.5. For fixation, the labeled cells were once washed with PBS/FCS and then fixed with BD Cytofix/Cytoperm (BD Bioscience) for 15 min on ice. The fixed cells were washed twice with PermWash buffer (BD Bioscience) and incubated with 200 µL of PermWash buffer containing 25 µg/mL of Hoechst 33342 (Invitrogen) for 10 min on ice. After washing, the cells were resuspended in PBS/FCS and subjected to fluorescence activated cell sorting.

Table 5.5: Antibodies panels used for the isolation of 26 hematopoietic populations.

Sorted Populations	Antigen	Clone	Conjugation	Company
HSC, MPP1, MPP5, MPP2, MPP3, MPP4	CD5	53-7.3	Biotin	BD Bioscience
	B220	RA3-6B2	Biotin	BD Bioscience
	CD11b	M1/70	Biotin	BD Bioscience
	CD8a	53-6.7	Biotin	BD Bioscience
	Gr1	RB6-8C5	Biotin	BD Bioscience
	Ter119	TER-119	Biotin	BD Bioscience
	Streptavidin	-	PE-Cy7	BioLegend
	cKit (CD117)	2B8	APC	Invitrogen
	Sca1	D7	APC-Cy7	BD Bioscience
	CD150 (SLAM)	TC15-12F12.2	PE-Cy5	BioLegend
	CD48	HM48-1	Pacific Blue	BioLegend
	CD34	RAM34	FITC	eBioscience
	CD135 (Flt3)	A2F10	PE	eBioscience
GMP, MEP, CMP $CD55^-$, CMP $CD55^+$	CD5	53-7.3	Biotin	BD Bioscience
	B220	RA3-6B2	Biotin	BD Bioscience
	CD11b	M1/70	Biotin	BD Bioscience
	CD8a	53-6.7	Biotin	BD Bioscience
	Gr1	RB6-8C5	Biotin	BD Bioscience
	Ter119	TER-119	Biotin	BD Bioscience
	Streptavidin	-	PE-Cy7	BioLegend
	cKit (CD117)	2B8	APC	Invitrogen
	CD16/32	93	eFluor 450	eBioscience
	CD34	RAM34	FITC	eBioscience
	CD155	RIKO-3	PE	BioLegend
CLP	CD5	53-7.3	Biotin	BD Bioscience
	B220	RA3-6B2	Biotin	BD Bioscience
	CD11b	M1/70	Biotin	BD Bioscience
	CD8a	53-6.7	Biotin	BD Bioscience
	Gr1	RB6-8C5	Biotin	BD Bioscience
	Ter119	TER-119	Biotin	BD Bioscience
	Streptavidin	-	PE-Cy7	BioLegend
	cKit (CD117)	2B8	APC	Invitrogen
	Sca1	D7	APC-Cy7	BD Bioscience
	CD127	SB/199	PE	BD Bioscience

Sorted Populations	Antigen	Clone	Conjugation	Company
preMegE, MkP, CFU-E	CD5	53-7.3	Biotin	BD Bioscience
	B220	RA3-6B2	Biotin	BD Bioscience
	CD11b	M1/70	Biotin	BD Bioscience
	CD8a	53-6.7	Biotin	BD Bioscience
	Gr1	RB6-8C5	Biotin	BD Bioscience
	Streptavidin	-	PerCP-Cy5.5	BD Bioscience
	cKit (CD117)	2B8	APC	Invitrogen
	Sca1	D7	APC-Cy7	BD Bioscience
	CD150 (SLAM)	TC15-12F12.2	PE-Cy5	BioLegend
	CD41	TMWReg30	FITC	BD Bioscience
	CD16/32	93	PE	eBioscience
	CD105	MJ7/18	Pacific Blue	Invitrogen
	Ter119	TER-119	PE-Cy7	eBioscience
MDP, CDP	CD3	145-2C11	PerCP-Cy5.5	BioLegend
	CD4	GK1.5	PerCP-Cy5.5	BioLegend
	B220	RA3-6B2	PerCP-Cy5.5	BioLegend
	CD11b	M1/70	PerCP-Cy5.5	BioLegend
	CD11c	N418	PerCP-Cy5.5	BioLegend
	CD19	6D5	PerCP-Cy5.5	BioLegend
	CD8a	53-6.7	PerCP-Cy5.5	BioLegend
	Gr1	RB6-8C5	PerCP-Cy5.5	BioLegend
	Ter119	TER-119	PerCP-Cy5.5	BioLegend
	NK1.1	PK136	PerCP-Cy5.5	BioLegend
	cKit (CD117)	2B8	Pacific Blue	BioLegend
	CD115	AFS98	PE	BioLegend
	CD135 (Flt3)	A2F10	APC	BioLegend
cMoP	CD3	145-2C11	PerCP-Cy5.5	BioLegend
	CD4	GK1.5	PerCP-Cy5.5	BioLegend
	B220	RA3-6B2	PerCP-Cy5.5	BioLegend
	CD11c	N418	PerCP-Cy5.5	BioLegend
	CD19	6D5	PerCP-Cy5.5	BioLegend
	CD8a	53-6.7	PerCP-Cy5.5	BioLegend
	Ly6G	1A8	PerCP-Cy5.5	BioLegend
	Ter119	TER-119	PerCP-Cy5.5	BioLegend
	NK1.1	PK136	PerCP-Cy5.5	BioLegend
	cKit (CD117)	2B8	APC	Invitrogen
	CD115	AFS98	PE	BioLegend
	CD135 (Flt3)	A2F10	Biotin	eBioscience
	Streptavidin	-	PE-Cy7	eBioscience
	Ly6C	HK1.4	APC -Cy7	BioLegend
	CD11b	V450	Pacific Blue	BD Bioscience
Monocytes, B cells, T cells	Ly6C	HK1.4	APC-Cy7	BioLegend
	CD11b	M1/70	FITC	eBioscience
	B220	RA3-6B2	APC	eBioscience
	CD4	GK1.5	PE	eBioscience
	CD48a	53-6.7	PE	eBioscience
Neutrophils, Eosinophils	CD5	53-7.3	Biotin	BD Bioscience
	B220	RA3-6B2	Biotin	BD Bioscience
	CD8a	53-6.7	Biotin	BD Bioscience
	Ter119	TER-119	Biotin	BD Bioscience
	Streptavidin	-	PE-Cy7	BioLegend
	CD45	30-F11	PE-Cy5	eBioscience
	CD11b	M1/70	APC-Cy7	eBioscience
	SiglecF	E50-2440	PE	BD Bioscience
	Ly6G	1A8-Ly6g	FITC	eBioscience
	FcεRI	MAR-1	Pacific Blue	BioLegend
	cKit (CD117)	2B8	APC	Invitrogen
pDC	CD11c	N418	APC	eBioscience
	PDCA	eBio827	PE	eBioscience
	7-AAD	-	-	Thermo Fisher

Sorted Populations	Antigen	Clone	Conjugation	Company
cDC CD8a$^+$, cDC CD11b$^+$	CD11c	N418	APC	eBioscience
	MHCII	M5/114.15.2	PE	eBioscience
	CD11b	M1/70	PE-Cy7	BioLegend
	CD8a	53-6.7	V450	BD Bioscience
	7-AAD	-	-	Thermo Fisher
Megakaryocyte (2N)	CD3	-	-	self-made antibody
	CD4	-	-	self-made antibody
	B220	-	-	self-made antibody
	CD11b	-	-	self-made antibody
	CD8a	-	-	self-made antibody
	Gr1	-	-	self-made antibody
	Ter119	-	-	self-made antibody
	CD4	GK1.5	PE-Cy7	Invitrogen
	B220	RA3-6B2	PE-Cy7	eBioscience
	CD11b	M1/70	PE-Cy7	BioLegend
	CD8a	53-6.7	PE-Cy7	eBioscience
	Gr1	RB6-8C5	PE-Cy7	eBioscience
	Ter119	TER-119	PE-Cy7	eBioscience
	CD42d	1C2	APC	eBioscience
	CD41	eBioMWReg30	PE	eBioscience
	CD9	eBioKMC8	FITC	eBioscience
	Hoechst33342	-	-	Invitrogen

5.4 Whole genome DNA methylation profiling of hematopoietic cell populations

5.4.1 DNA extraction from snap-frozen cell pellets

DNA was isolated from snap-frozen cell pellets (10.000 - 50.000 sorted cells) using the QIAamp DNA Micro Kit (Qiagen) according to the manufacturer's instructions. The purified DNA was eluted twice with 15 µL of UltraPure nuclease-free water (Thermo Fisher Scientific). The concentration of isolated DNA was measured with a Qubit 2.0 Fluorometer (Thermo Fisher Scientific) using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Of the isolated DNA, 10 to 30 ng were used for subsequent TWGBS.

5.4.2 Generation of tagmentation-based whole genome bisulfite sequencing libraries

Tagmentation-based whole genome bisulfite sequencing (TWGBS) was performed as previously published by Wang *et al.* [Wang et al., 2013]. Briefly, pre-adapters (Tn5mC-Apt1 and Tn5mC1.1-A1block) were annealed to generate a 10 µM load adaptor mixture. The transposome was assembled by transferring 11 µL of load

adaptor to 10 µL of Ez-Tn5 transposase (Epicentre). For tagmentation of the genomic DNA, 2 µL of assembled transposome were added to a mixture of 10 µL of 2x tagmentation buffer (20 mM Tris(hydroxymethyl)aminomethane; 10 mM MgCl2; 20% (vol/vol) dimethylformamide; pH 7.6) and 8 µL of DNA (10 to 30 ng) in a 0.2 mL tube and incubated for 8 min at 55°C. The tagmentation reaction was stopped by adding 15 µL of 5 M Guanidinium thiocyanate (Sigma-Aldrich). The resulting tagmented DNA was purified with AmPure XP beads (Beckman Coulter) using a SX-8G IP-Star robot (Diagenode). Oligonucleotide replacement and gap repair was performed by adding 2 µL of dNTP mix (10 mM, Fermentas), 2 µL of 10x Ampligase buffer (Biozym), 2 µL of replacement oligo, 1 µL of T4 DNA polymerase (New England BioLabs), and 2.5 µL of Ampligase (Biozym) to 11 µL of eluate and incubating the mix as indicated in the published protocol. After AmPure XP Bead purification, the tagmented DNA was bisulfte-converted using the EZ DNA methylation kit (Zymo) according to the manufacturer's instructions. For the generation and amplification of TWGBS sequencing libraries, 11 µL of bisulfite-converted DNA were mixed with 12.5 µL of 2x Kapa 2G Robust HotStart ReadyMix (Kapa Biosystems), 0.75 µL of primer Tn5mCP1 (10 µM), 0.75 µL of barcoded Tn5mC reverse primer, and 0.25 µL of 100x SYBR Green (Life Technologies) and amplified in a real-time PCR (95°C, 3 min; 12 to 15 cycles of 95°C, 20 s; 62°C, 15 s; 72°C, 40 s) using Lightcycler 480 with a 96-well block (Roche). The PCR reaction was stopped when the amplification curve reached the plateau phase. For each sample two to four differently barcoded sequencing libraries were generated in distinct real-time PCR reactions in order to reduce the number of PCR duplicates. After a final AmPure XP Bead purification, the library concentration was quantified with a Qubit 2.0 Fluorometer using the Qubit dsDNA HS Assay Kit and the fragment size distribution (expected mean fragment size: 300 to 500 bp) was determined with a Bioanalyzer 2100 (Agilent) using the Quant-iT dsDNA HS assay kit (Agilent). Finally, the resulting libraries were pooled in an equimolar ratio and subjected to 125 bp paired-end sequencing on a HiSeq2000 platform (Illumina). Sequencing was performed by the Genomics and Proteomics Core Facility at the DKFZ.

Table 5.6: Oligonucleotides for the generation of TWGBS libraries.

Oligonucleotide	Sequence (5' to 3')	Modification
Tn5mC-Apt1	TcGTcGGcAGcGTcAGATGTGTATAAGAGAcAG	c: 5C-methylated
Tn5mC1.1-A1block	pCTGTCTCTTATACAddC	p: phosphate ddC: dideoxycytidylate
Replacement oligo	pcTGTcTcTTATAcAcATcTccGAGccCAcGAGAcinvT	p: phosphate c: 5C-methylated invT: inverted deoxythymidylate
Tn5mCP1	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTC	
Tn5mCBar-X	CAAGCAGAAGACGGCATACGAGAT-barcode(XXXXXXXXX)-GTCTCGTGGGCTCGG	

5.4.3 TWGBS read alignment

Read alignment was performed by the Omics IT and Data Management Core Facility (ODCF) at the DKFZ using an updated version of the pipeline published by Wang *et al.* [Wang et al., 2013], which was implemented as a Roddy Workflow[1] in the automated 'One Touch Pipeline' (OTP) [Reisinger et al., 2017]. Briefly, adaptor sequences of raw reads were trimmed using 'Trimmomatic' [Bolger et al., 2014]. The sequencing reads were then *in silico* bisulfite-converted; with a C-to-T conversion in the first read and a G-to-A conversion in the second read of each read pair. The software package 'BWA-MEM' [Li, 2013] was used with default parameters to align the converted reads to the *in silico* bisulfite-converted reference mm10 genome. Reads aligned to the anti-sense strand of the reference genome were removed and the remaining aligned reads were converted back to the original sequence. PCR duplicates were identified and removed using 'Picard MarkDuplicates'[2]. Finally, the sequencing reads from all libraries per sample were merged. The alignment quality was validated by computing the mapping rate using 'SAMtools' [Li et al., 2009], insert size distributions, and genome coverage statistics.

5.4.4 Methylation calling and quality control

Methylation calling and M-bias trimming was performed by Stephen Kraemer from the Bioinformatics and Omics Data Analytics group at the DKFZ with the methylation caller 'bistro' (version 0.2.0)[3]. Automatic M-bias detection was done using the 'binomp' algorithm to remove the gap repair nucleotides introduced by the tag-

[1]https://github.com/DKFZ-ODCF/AlignmentAndQCWorkflows
[2]https://github.com/broadinstitute/picard
[3]https://github.com/stephenkraemer/bistro

mentation reaction and additional nucleotides affected by M-bias. Sequencing reads with a mapping rate ≥25 and nucleotides with a Phred-scaled quality score ≥25 were considered for further analysis. In general, cytosines converted to thymine were considered as unmethylated, while non-converted cytosines were considered as methylated.

Bisulfite conversion rates were estimated using the CH methylation levels on Chromosome 1. Data reproducibility of biological replicates was estimated by PCA using the R package 'FactoMineR' [Lê, S. and Josse, J. and Husson, F., 2008] and by calculating the Pearson's correlation coefficient r based on methylation levels in murine 'Ensembl regulatory features' downloaded from the Ensembl database[4].

5.4.5 Detection of differentially methylated regions

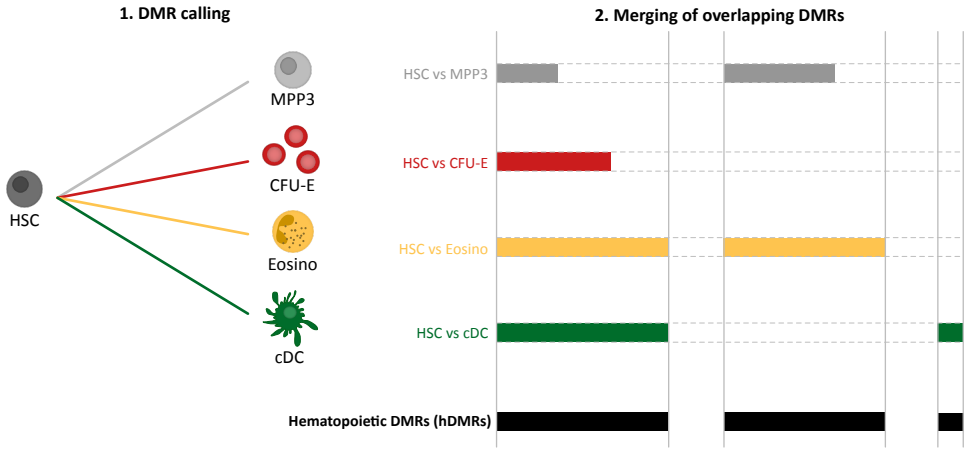

Figure 5.1: Workflow for the identification of hematopoietic DMRs. First, DMRs were called by pairwise comparisons of HSCs versus all downstream hematopoietic populations (illustrated here by only four populations for simplicity). Subsequently, overlapping DMRs from the individual pairwise comparisons were merged to the so-called 'hematopoietic' DMRs (hDMRs).

Differentially methylated regions (DMRs) were called by the pairwise comparison using the R package 'DSS' [Park and Wu, 2016]. First, the statistical test using the function 'DMLtest' without smoothing was run for two samples with at least two biological replicates. Second, DMRs were identified with the function 'callDMR'. DMRs were defined as at least 3 consecutive CpGs with a minimal DNA methylation difference of 10% and a p-value of 0.01 (hematopoietic DMRs) or 0.05 (aging DMRs and inflammation DMRs).

[4]http://www.ensembl.org/info/genome/funcgen/regulatory_features.html

The final set of 'hematopoietic DMRs' (hDMRs) was generated by merging overlapping DMRs from the individual pairwise comparisons (HSC vs MPP1, HSC vs MPP2, etc.) as exemplified by 4 pairwise comparisons in Figure 5.1.

5.4.6 Definitions of genomic features

The software tool 'gtfanno'[5] was used to annotate DMRs to RefSeq genes and to define genomic regions. Coordinates of genomic features (gene, transcript, exon, CDS, UTR, start codon, stop codon) of RefSeq genes were downloaded from the GENCODE database[6]. Genomic features were defined as follows: promoter: -5,000 bp upstream and 1,000 bp downstream of RefSeq TSS; 5'-UTR: UTR as defined by GENCODE; 3'-UTR: UTR as defined by GENCODE, not overlapping with start codon; exon: as defined by GENCODE; DCRD (Distant Cis Regulatroy Domain): -50,000 bp to -5,000 bp upstream of TSS; intergenic: all genomic regions that do not overlap with the other genomic features.

5.5 Single-cell methylation profiling of hematopoietic cells

5.5.1 Single-cell sorting of hematopoietic cells

Bone marrow cell suspensions were prepared for cell sorting as described in 5.3. Cells were index-sorted into 0.2 mL tubes containing 3 µL of RLT Plus Buffer (Qiagen) using an Aria I cell sorter (BD Bioscience) and a 100 µm nozzle. After cell sorting, the sorted single cells were immediately frozen on dry ice and stored at -80°C until further processing. The fluorescence intensities per channel and cell were recorded for retrospective analysis of surface marker expression.

5.5.2 Generation of single-cell bisulfite sequencing libraries

Post-bisulfite adapter tagging (PBAT) libraries for single-cell bisulfite sequencing (scBS-seq) were generated as previously described by Clark *et al.* (in collaboration with Dr. Mark Hartmann and Maximilian Schönung) [Clark et al., 2017]. Briefly, 7

[5]https://github.com/stephenkraemer/gtfanno
[6]https://www.gencodegenes.org/mouse/release_M19.html

µL of water were added to 3 µL of the cell lysate from 5.5.1. Bisulfite conversion and purification was carried out using the EZ Methylation Direct Kit (Zymo) according to manufacturer's instructions but with half of the indicated volumes. The bisulfite-converted DNA was eluted directly into 40 µL of preamplification mixture (4 µL 10x Blue buffer (Enzymatics), 1.6 µL dNTP mix (10 mM, Roche), 1.6 µL Preamp primer (10 µM, IDT), 32.8 µL water) and incubated at 65°C for 3 min. For the following preamplification, 1 µL of Klenow exo- (50 U μL^{-1}, Enzymatics) was added to the mixture, which was slowly heated from 4°C to 37°C, incubated at 37°C for 30 min, and then heated to 95°C for 45 sec. The preamplification step was repeated three times with the addition of 2.5 µL of preamplification mixture (0.25 µL 10x Blue buffer, 0.1 µL dNTP mix, 1 µL Preamp primer (10 µM), 0.5 µL Klenow exo- (50 U μL^{-1}), 0.65 µL water). For exonuclease I treatment, the reaction mixture was incubated with 2 µL of Exonuclease I (NEB) and 48 µL of water at 37°C for 1 h. The DNA was purified with AMPpure XP beads (Beckman Coulter) using 0.8x bead to sample ratio according to the manufacturer's guideline and immediately eluted into 49 µL of adaptor 2 tagging mix (5 µL 10x Blue buffer, 2 µL dNTP mix, 2 µL Adapter 2 Oligo (10 µM, IDT), 40 µL water). The mixture was heated to 95°C for 45 sec and cooled on ice. For second strand synthesis, 1 µL of Klenow exo- (50 U μL^{-1}) was added and the mixture was slowly heated from 4°C to 37°C and then incubated at 37°C for 90 min. The double-tagged products were purified with AMPpure XP beads using 0.8x bead to sample ratio and immediately eluted into 50 µL of PCR reaction mix (10 µL 5X KAPA HiFi Fidelity Buffer (Anachem), 1 µL dNTP mix, 1 µL PE1.0 primer (10 µM, IDT), 1 µL iPCRTag (10 µM), 1 µL KAPA HotStart polymerase (1 U μL^{-1}, Anachem), 36 µL water). The scRNA-seq libraries were amplified in a PCR reaction (95°C, 2 min; 12 to 16 cycles of 94°C, 80 s; 65°C, 30 s; 72°C, 30 s; and finally 72°C, 3 min) of 12 to 16 cycles. After a final AmPure XP Bead purification with a 0.8x bead to sample ratio, the scBS-seq library concentration was quantified with a Qubit 2.0 Fluorometer using the Qubit dsDNA HS Assay Kit and the fragment size distribution was determined with a Bioanalyzer 2100 using the Quant-iT dsDNA HS assay kit. Finally, 19 scBS-seq libraries, four mini-bulk libraries (10 to 100 cells), and one negative control were multiplexed in equimolar ratios and subjected to 75 bp paired-end sequencing on a HiSeq2000 platform. Sequencing was done by the Genomics and Proteomics Core Facility at the DKFZ.

Table 5.7: Oligonucleotides for the generation of scBS-seq libraries. All oligonucleotides were purchased form IDT.

Oligonucleotide	Sequence (5' to 3')	Modification
Preamp primer	CTACACGACGCTCTTCCGATCTNNNNNN	
Adapter 2 Oligo	TGCTGAACCGCTCTTCCGATCTNNNNNN	
PE1.O	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACG ACGCTCTTCCGATC*T	*: phosphorothioate
iPCRTag	CAAGCAGAAGACGGCATACGAGATA-barcode(XXXXXXXX)- GAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T	*: phosphorothioate

5.5.3 scBS-seq read alignment and methylation calling

Read alignment and methylation calling was performed by Anand Mayakonda from the division of Cancer Epigenomics at the DKFZ using a modified workflow published by Clark *et al.* [Clark et al., 2017]. Briefly, adaptors and low-quality sequences were trimmed with 'fastp' [Chen et al., 2018]. Subsequently, read alignment, PCR duplicate exclusion, and methylation calling was done with the software program 'Bismark' (v0.22.1) [Krueger and Andrews, 2011]. The sequencing reads were aligned to the mm10 reference genome in the single-end non-directional mode. After duplicate exclusion, methylation calling was performed with default parameters.

5.6 Single-cell gene expression profiling of hematopoietic cell layers

5.6.1 Sorting of hematopoietic cell layers for single-cell RNA sequencing

Three murine hematopoietic cell layers from one donor mouse, including Lin$^-$ Sca1$^+$ cKit$^+$ (LSK) cells, Lin$^-$ Sca1$^-$ cKit$^+$ (LK) cells, and total bone marrow cells (CD45$^+$), were isolated for single-cell RNA sequencing (scRNA-seq). LSK cells and LK cells were sorted from a suspension of lineage negative cells and bone marrow cells from a suspension of total bone marrow cells (Section 5.3.1, 5.3.2, 5.3.4). Overall, 3,366 LSK cells, 10,044 LK cells, and 10,004 total bone marrow cells were sorted into PBS/FCS and subjected to scRNA-seq library preparation.

5.6.2 Generation of single-cell RNA sequencing libraries

Single-cell RNA-seq libraries were prepared in the scOpenLab at the DKFZ by Dr. Mark Hartmann and Katharina Bauer using the Chromium Single-Cell 3' Reagent Kit v2 (10X Genomics) and following the manufacturer's instructions. Briefly, sorted LSK, LK, and total bone marrow cells were loaded on a Chromium Single-Cell A Chip and processed using the Chromium Single-Cell Controller (10X Genomics) to generate single-cell Gel Bead-In-Emulsions (GEMs) and full-length barcoded cDNA. Subsequently, the GEMs were broken and the released barcoded cDNA was amplified by PCR. Finally, scRNA-seq libraries were generated in a PCR reaction and subjected to paired-end 100 cycles sequencing on a HiSeq4000 platform (Illumina). Sequencing was performed by the Genomics and Proteomics Core Facility at the DKFZ.

5.6.3 Analysis of single-cell transcriptomes

Chromium scRNA-seq output was processed by Abdelrahman Mahmoud from the division of Applied Bioinformatics at the DKFZ. Briefly, the 10X Genomics 'Cell Ranger' pipeline version 3.1 (10X Genomics) was used with default parameters for demultiplexing of raw reads, alignment to the mm10 reference genome, and generation of the unique molecular identifiers (UMI) count matrix. The R package 'Seurat' version 3 was used for further pre-processing of the scRNA-seq data [Butler et al., 2018, Stuart et al., 2019]. Quality control (QC) filtering of single cells was performed with default parameters. The data was normalized and scaled using the 'sctransform' wrapper implemented in Seurat. Clustering of single cells was performed based on Louvain clustering using Seurat's 'FindClusters' function. Differentially expressed genes were called and used to manually annotate the scRNA-seq clusters (annotation of single-cell clusters was performed by Dr. Simon Haas, Dr. Mark Hartman, and Jens Langstein). Finally, to visualize scRNA-seq data set, the UMAP was plotted using Seurat's 'RunUMAP' function.

5.7 Bioinformatical methods for downstream analysis

5.7.1 Hierarchical clustering of DMRs

For hierarchical clustering of DMRs, the beta-values of each DMR were z-score transformed. The z-score, also known as standard score, scales the values of a column around the mean. If a data point is equal to the mean, the z-score will be scaled to 0. The z-score is defined by the following equation (x, data point; µ, mean; σ, standard deviation):

$$z = \frac{x - \mu}{\sigma} \tag{5.1}$$

Unsupervised hierarchical clustering of z-score transformed DNA methylation values was performed based on Euclidean distance and Ward's linkage. Heatmaps were plotted with the R/Bioconductor package 'ComplexHeatmap' [Gu et al., 2016]. To determine an appropriate number of DNA methylation clusters, the resulting dendrogram was cut using the function 'cutreeHybrid' implemented in the R package 'dynamicTreeCut' [Langfelder et al., 2016]. The 'deepSplit' and 'minGap' parameters were adjusted to modify the size and number of clusters.

5.7.2 Genome browser tracks

Genome browser tracks were generated using the R/Bioconductor package 'Gviz' [Hahne and Ivanek, 2016]. Murine RefSeq genes were extracted form the 'TxDb. Mmusculus.UCSC.mm10.knownGene TxDB' R object based on the UCSC build mm10.

5.7.3 Genomic region set enrichment analysis using GREAT

The 'Genomic Regions Enrichment of Annotations Tool' (GREAT, version 3.0.0) with default parameters was used to annotate DMRs to nearby genes and perform Gene Ontology and MSigDB pathway enrichment analysis based on a binomial test over genomic regions (e.g. DMRs) [McLean et al., 2010]. The mm10 mouse genome

was used as background. Enriched gene sets with a binominal fold enrichment >2.0 were ranked according to their -$\log_{10}$(binominal p-value).

5.7.4 Transcription factor motif enrichment analysis using HOMER

The function 'findMotifsGenome.pl' implemented in the 'Hypergeometric Optimization of Motif EnRichment' tool (HOMER) was used to find known transcription factor binding motifs in DMRs and perform motif enrichment analysis based on a hypergeometric test over genomic regions (e.g. DMRs) [Heinz et al., 2010]. The mm10 mouse genome was used as background. Enriched transcription factor binding motifs were ranked according to their hypergeometric p-value.

5.7.5 Overlap enrichment analysis of genomic features based on the Fisher's exact test

Overlap enrichment analyses of genomic region sets (single-cell marker genes, transcription factor motifs, ChIP-seq peaks, enhancer regions) and methylation programs (hDMRs) were performed with 'regionset_profiler'[7]. This tool is based on a two-sided Fisher's exact test calculating the p-value and log odds ratio of the overlap. The whole set of hDMRs was used as the background region set. The enrichment score was calculated by multiplying the product of the $\log_{10}$(p-value) by the sign of the log odds ratio.
Genomic regions set were curated as follows:

1. The single-cell marker gene sets were extracted from the single-cell clusters identified with the R package 'Seurat' (see 5.5.2).
2. Transcription factor binding motifs were downloaded from the HOMER webpage[8]
3. Genomic regions of ChIP-seq peaks were downloaded from the 'Codex' database and filtered for ChIP-seq data generated in untreated primary hematopoietic cell populations[9] [Sanchez-Castillo et al., 2015].

[7]https://github.com/stephenkraemer/regionset_profiler
[8]http://homer.ucsd.edu/homer/data/motifs/
[9]http://big.databio.org/regiondb/LOLACore_180412.tgz

4. Hematopoietic enhancer regions have been defined as previously published by Lara-Astiaso *et al.*[10] [Lara-Astiaso et al., 2014].

The overlap of genomic regions (transcription factor motifs, ChIP-seq peaks, enhancer regions) and hDMRs was computed with 'BEDtools' [Quinlan and Hall, 2010].

5.7.6 Diffusion map representation

Diffusion map representation, a spectral non-linear reduction method, was generated using the R package 'destiny' [Angerer et al., 2016]. Louvain clusters were identified with the 'Louvain' algorithm implemented in 'destiny'.

5.7.7 Non-negative factorization

Non-negative matrix factorization (NMF), a matrix decomposition model, was run with a factorization rank of six using the R/Bioconductor package 'Bratwurst'[11].

5.7.8 Phylogenetic reconstruction

Methylation-based phylogenetic trees were generated calculating 'Manhattan' distance matrices based on mean DNA methylation levels within early hematopoietic or hematopoietic DMRs. Phylogenies were inferred by the minimal evolution method [Desper and Gascuel, 2002] using the 'fastme.bal' function and plotted using 'plot.phylo' function, both implemented in the R package 'ape' [Paradis and Schliep, 2019].

5.7.9 Overlap between two sets of DMRs

The overlap of two sets of DMRs was calculated with the 'findOverlapsOfPeak' function implemented in the R/Bioconductor package 'ChIPpeakAnno' [Zhu et al., 2010, Zhu, 2013]. DMRs with a minimum overlap of one nucleotide were considered to be overlapping. Venn diagrams were plotted with the 'makeVennDiagram'. p-values were calculated with a one-sided Fisher's exact test with all hDMRs as genomic background.

[10] http://compbio.cs.huji.ac.il/blood-chromatin/Data.html
[11] https://github.com/wurst-theke/bratwurst

5.7.10 Data wrangling, visualization, and statistical analyses

Data wrangling, data visualization, and statistical analyses were performed using R. Wrangling of data was mainly conducted with functions of the R package 'dplyr' [Wickham et al., 2019]. Data was visualized using the R packages 'ggplot2' or 'ComplexHeatmap' [Wickham, 2016, Gu et al., 2016]. 3D plots were generated using the R package 'rgl' [Adler and Murdoch, 2019]. Statistical tests were calculated using functions of the R package 'stats'. The type of statistical test, sample size (n), and p-values are stated in the respective figure legend. Statistical significance is indicated by *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.005$. Error bars indicate standard deviation (SD). All figures were designed using Affinity Designer 1.5.5. This thesis was written in LaTeX.

A
Appendix

A.1 Supplementary Tables

Table A.1: Quality parameters of DNA methylomes of hematopoietic cell populations.

Cell Type	n	Genomic Coverage	Mean Genomic Coverage	Mean Methylation Level [%]	Bisulfite Conversion Rate [%]
HSC	1	9.4	44.3	82.0	99.38
	2	14.7			
	3	8.5			
	4	9.7			
	5	1.9*			
MPP1	1	14.1	25.9	82.1	99.35
	2	6.1			
	3	5.8			
MPP5	1	6.7	62.8	82.0	99.50
	2	22.2			
	3	17.5			
	4	6.0			
	5	10.4			
MPP2	1	6.3	25.2	81.0	99.13
	2	9.6			
	3	9.3			
MPP3	1	24.4	67.0	81.3	99.61
	2	22.9			
	3	19.8			
MPP4	1	28.2	77.9	81.8	99.68
	2	24.8			
	3	25.0			
CMP $CD55^+$	1	5.1	46.2	79.6	99.38
	2	6.1			
	3	10.7			
	4	13.2			
	5	11.1			
MEP	1	11.4	46.1	71.3	99.51
	2	19.7			
	3	14.9			
preMegE	1	13.7	26.0	0.41	77.5
	2	7.8			
	3	4.4*			

Cell Type	n	Genomic Coverage	Mean Genomic Coverage	Mean Methylation Level [%]	Bisulfite Conversion Rate [%]
MkP	1	9.6	35.2	78.2	99.59
	2	10.9			
	3	1.2*			
	4	13.5			
Megakaryocyte	1	7.7	12.6	70.8	98.87
	2	3.1*			
	3	1.8*			
CFU-E	1	15.5	44.7	71.0	99.60
	2	16.4			
	3	12.8			
GMP	1	13.7	50.3	79.3	99.30
	2	16.1			
	3	5.9			
	4	14.6			
cMoP	1	10.8	18.7	78.7	99.25
	2	7.9			
Monocytes	1	11.6	43.2	76.3	99.62
	2	16.7			
	3	14.9			
Neutrophils	1	11.3	24.0	77.0	99.58
	2	4.4*			
	3	8.2			
Eosinophils	1	10.7	30.7	75.0	99.30
	2	11.3			
	3	8.7			
CMP $CD55^{-}$	1	9.5	38.3	80.7	99.35
	2	3.5*			
	3	12.7			
	4	12.5			
MDP	1	14.3	23.0	80.8	99.15
	2	8.7			
CDP	1	7.2	44.2	80.4	99.21
	2	15.5			
	3	13.1			
	4	8.4			
cDC $CD8a^{+}$	1	10.5	31.3	78.4	99.57
	2	7.1			
	3	13.6			
cDC $CD811b^{+}$	1	9.5	21.8	78.2	99.56
	2	4.3*			
	3	7.9			
pDC	1	11.6	32.5	79.0	99.59
	2	9.8			
	3	11.1			
CLP	1	13.1	45.0	80.4	99.36
	2	23.1			
	3	8.8			
B cells	1	14.4	48.1	79.2	99.36
	2	12.6			
	3	21.1			
T cells	1	5.0	36.9	78.7	99.33
	2	3.7*			
	3	1.7*			
	4	13.0			
	5	13.5			

* Replicates marked with asterisks indicate low coverage samples (i.e. < 5)

Table A.2: Overview of young and aged mice analyzed. Splenomegaly was defined by a spleen weight >200 mg. Leukopenia was defined by WBC <2.0 K µl^{-}. WBC, white blood counts [K µl^{-}]; NE, neutrophils [K µl^{-}]; LY, lymphocytes [K µl^{-}]; MO, monocytes [K µl^{-}]; EO, eosinophils [K µl^{-}]; BA, basophils [K µl^{-}]; RBC, red blood counts [M µl^{-}]; PLT, platelets [K µl^{-}]; WGBS, whole genome bisulfite sequencing

Mouse ID	age	No of HSC	WBC	NE	LY	MO	EO	BA	RBC	PLT	Spleen in mg	Macroscopic Abormality	WGBS	Genomic Coverage	Methylation Level [%]
Young-1	8-12	1,960	10.84	1.32	9.24	0.27	0.01	0.00	9.38	714	82	-	no	-	-
Young-2	8-12	1,552	8.48	1.07	7.19	0.21	0.00	0.00	9.85	840	77	-	no	-	-
Young-3	8-12	1,174	8.76	1.09	7.08	0.57	0.01	0.00	9.46	641	58	-	no	-	-
Young-3	8-12	989	8.18	1.68	5.73	0.76	0.00	0.00	10.67	583	75	-	no	-	-
Young-pool-1	8-12	-	-	-	-	-	-	-	-	-	-	-	yes (=HSC-young-1)	9.4	82.1
Young-pool-2	8-12	-	-	-	-	-	-	-	-	-	-	-	yes (=HSC-young-2)	14.7	82.3
Young-pool-3	8-12	-	-	-	-	-	-	-	-	-	-	-	yes (=HSC-young-3)	8.5	82.2
Young-pool-4	8-12	-	-	-	-	-	-	-	-	-	-	-	yes (=HSC-young-4)	9.7	82.3
Aged-1	111	18,788	6.76	2.45	3.87	0.41	0.01	0.01	10.1	871	127	-	yes (=HSC-aged-1)	12.2	82.8
Aged-2	111	15,302	10.3	2.96	6.48	0.73	0.1	0.03	7.23	685	214	splenomegaly	no	-	-
Aged-3	111	8,708	5.74	2.40	3.08	0.22	0.03	0.00	9.45	1025	96	-	yes (=HSC-aged-2)	19.8	83.0
Aged-4	114	8,996	1.94	1.28	0.51	0.12	0.01	0.01	7.39	724	1,090	splenomegaly, leukopenia	no		
Aged-5	119	15,844	6.30	2.39	3.72	0.17	0.02	0.01	9.14	1,023	163	-	yes (=HSC-aged-3)	16.6	83.2
Aged-6	119	7,077	6.08	1.57	4.05	0.4	0.05	0.01	8.16	766	406	splenomegaly	no	-	-
Aged-7	130	4,976	2.00	1.15	0.62	0.14	0.06	0.03	3.80	246	103	intestinal infection, leukopenia	no	-	-
Aged-8	130	45,744	20.54	6.87	11.83	1.35	0.43	0.06	9.42	1,042	99	-	yes (=HSC-aged-4)	5.4	82.5
Aged-9	130	43,140	9.12	5.10	3.55	0.40	0.07	0.00	9.50	1,110	110	-	yes (=HSC-aged-5)	4.8	82.0
Aged-10	130	729	1.92	0.53	1.08	0.29	0.01	0.00	5.19	49	307	splenomegaly, leukopenia	no	-	-

A.2 Supplementary Figures

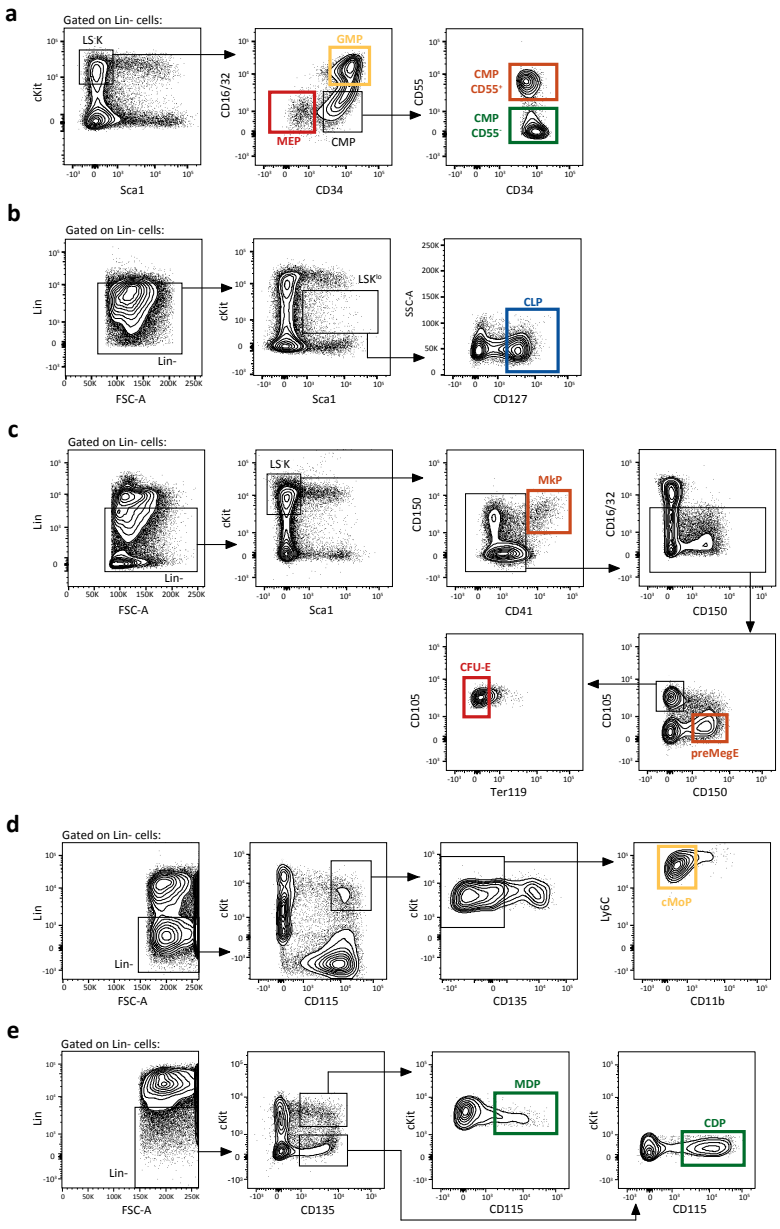

Figure A.1: Representative sorting schemes used to isolate primary progenitor populations. Committed progenitor populations were collected from the bone marrow of 8-12 weeks-old CD57BL/6J mice in at least two biological replicates. Shown are sorting schemes for: GMP, MEP, CMP CD55^{-}, CMP CD55^{+} **(a)**; CLP **(b)**; MkP, preMegE, CFU-E **(c)**; cMoP **(d)**; MDP, CDP **(e)**.

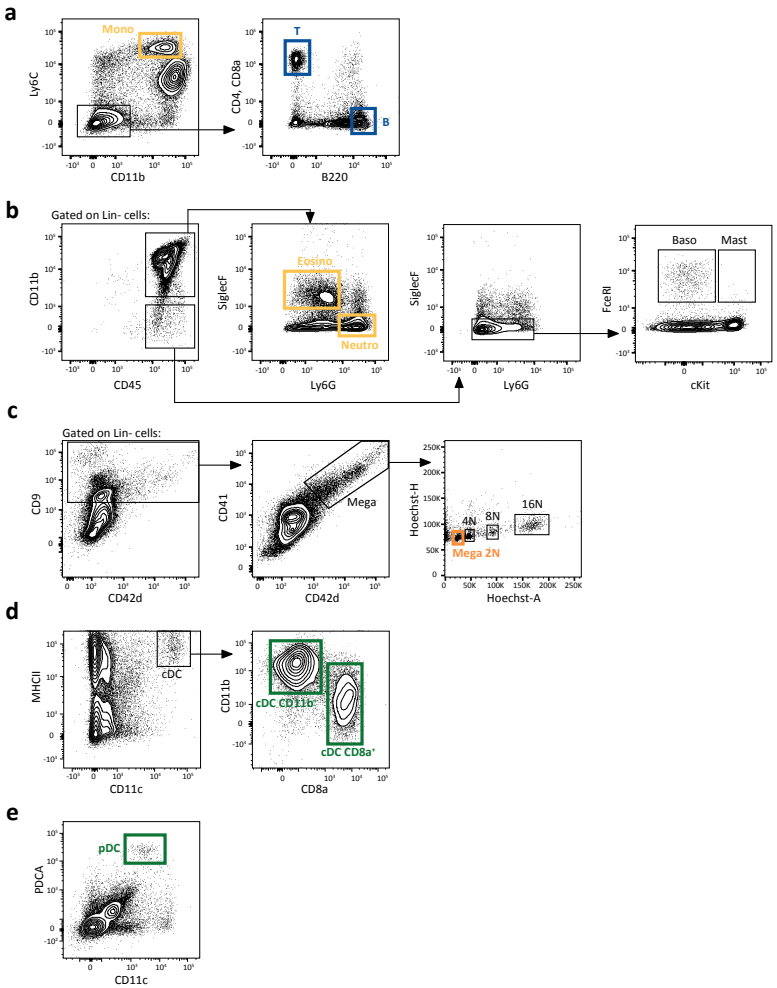

Figure A.2: Representative sorting schemes used to isolate primary differentiated blood cell types. Blood cell types were collected from the bone marrow or spleen of 8-12 weeks-old CD57BL/6J mice in at least two biological replicates. Shown are sorting schemes for: monocytes, B cells, T cells **(a)**; eosinophils, neutrophils **(b)**; megakaryocytes **(b)**; cDCs $CD8a^+$, cDCs $CD11b^+$ (collected from spleen) **(d)**; pDCs (collected from spleen) **(e)**.

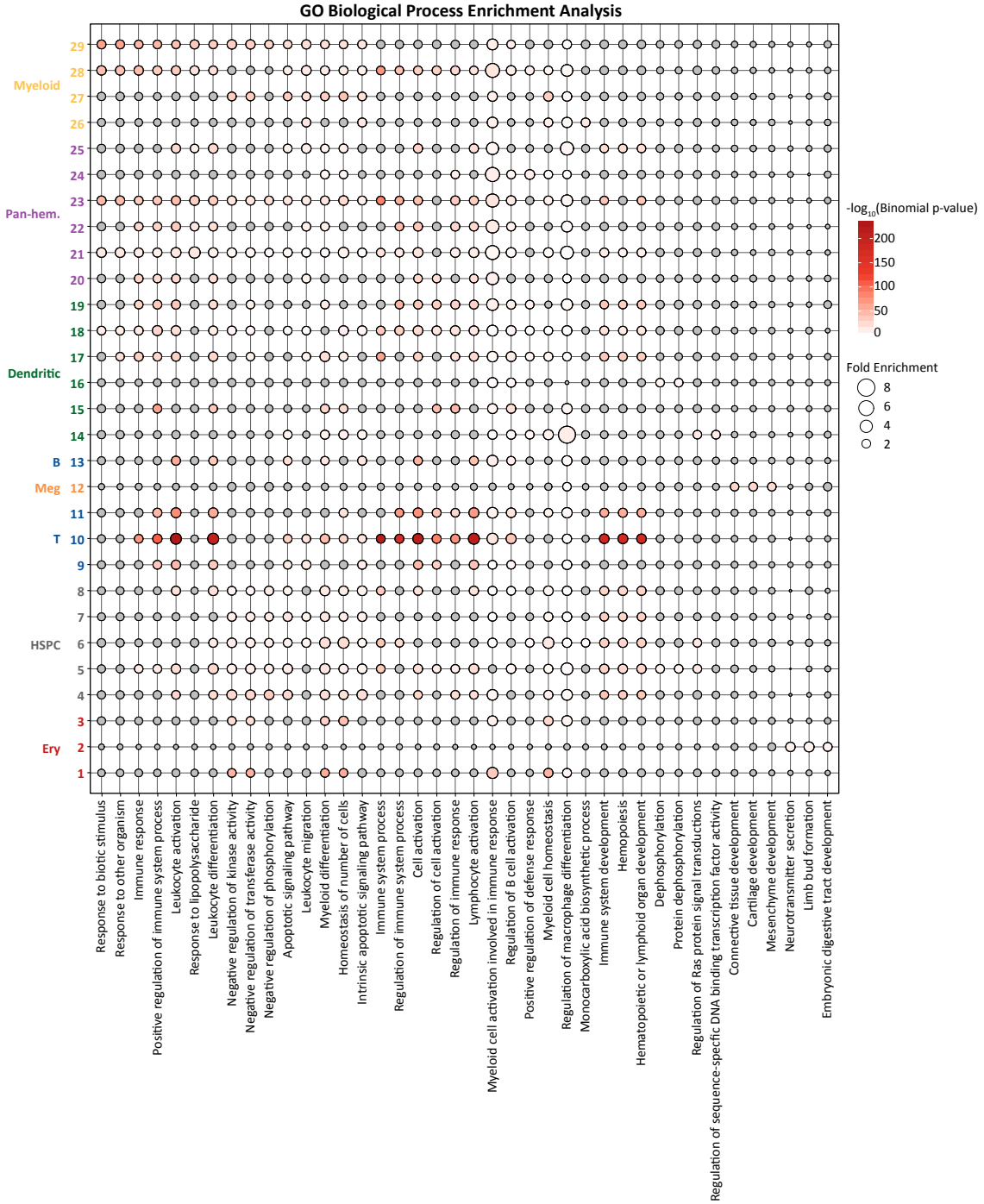

Figure A.3: Cell type- and lineage-specific DNA methylation programs are enriched for gene ontology terms associated with corresponding hematological processes. Bubble chart showing the enrichment for the 3 most significantly enriched GO Biological Process terms for all twenty-nine DNA methylation programs. Color represents the $-\log_{10}$(binomial p-value) and size represents the binominal fold enrichment. GO terms with a binominal fold enrichment < 2 are colored in grey.

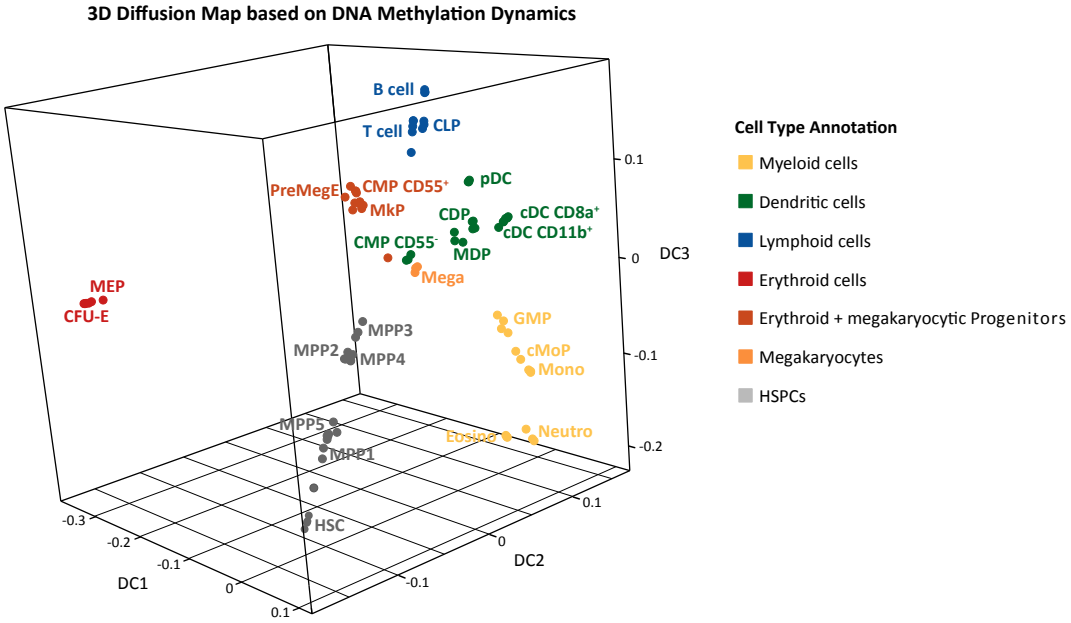

Figure A.4: 3D diffusion map based on DNA methylation remodeling events during hematopoietic differentiation reflects known hematopoietic cellular relationships. The diffusion algorithm implemented in the R package 'destiny' was used to infer a model of the hematopoietic system in form of a diffusion map, which is purely based on DNA methylation dynamics. Shown is a snapshot of the 3D methylation diffusion map of diffusion component 1-3.

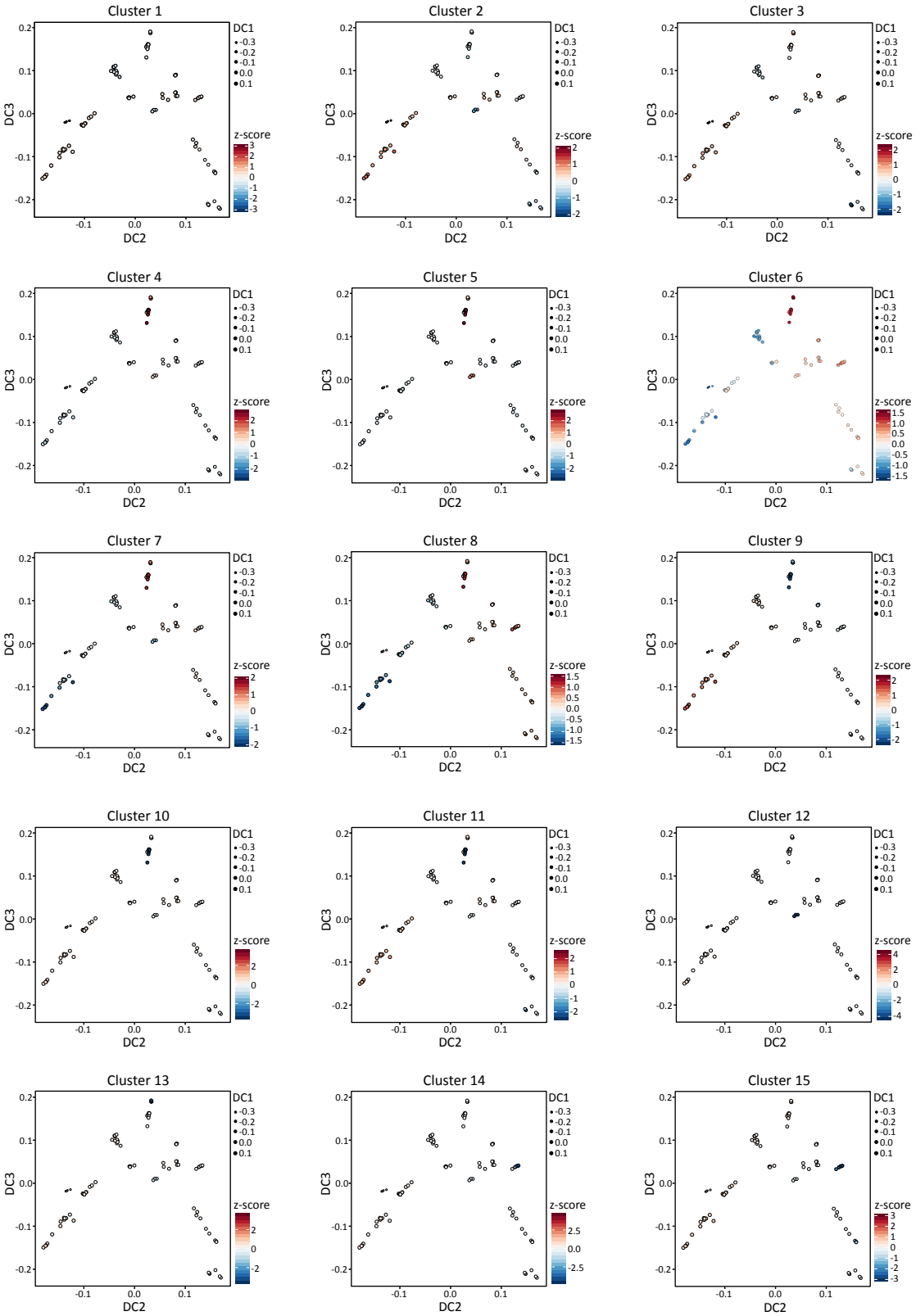

Figure A.5: DNA methylation changes are progressive and unidirectional during hematopoietic differentiation. DNA methylation levels (z-score transformed beta-values) were aggregated per methylation cluster and projected onto the diffusion map. Shown are DNA methylation programs 1 to 15.

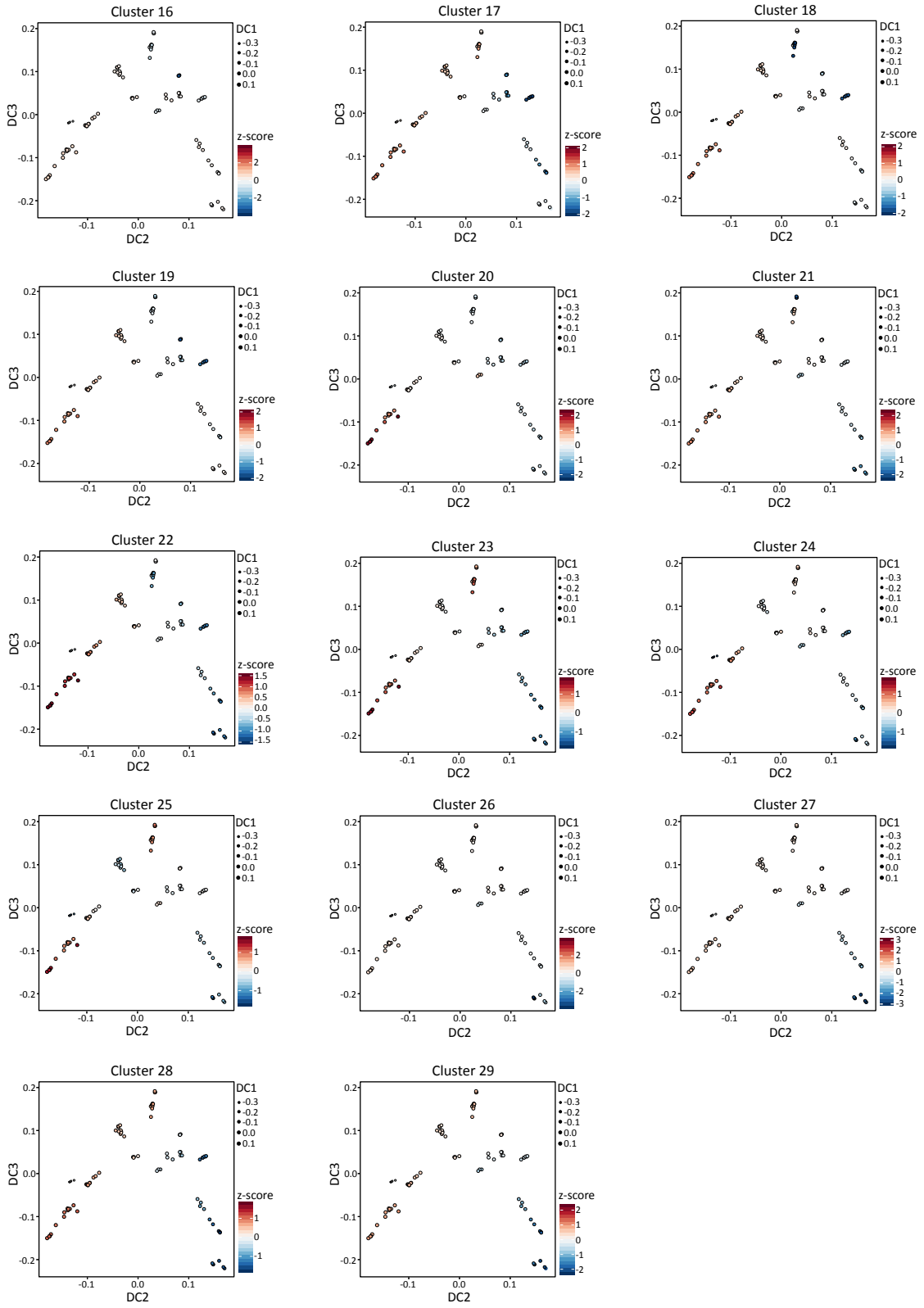

Figure A.6: DNA methylation changes are progressive and unidirectional during hematopoietic differentiation. DNA methylation levels (z-score transformed beta-values) were aggregated per methylation cluster and projected onto the diffusion map. Shown are DNA methylation programs 16 to 29.

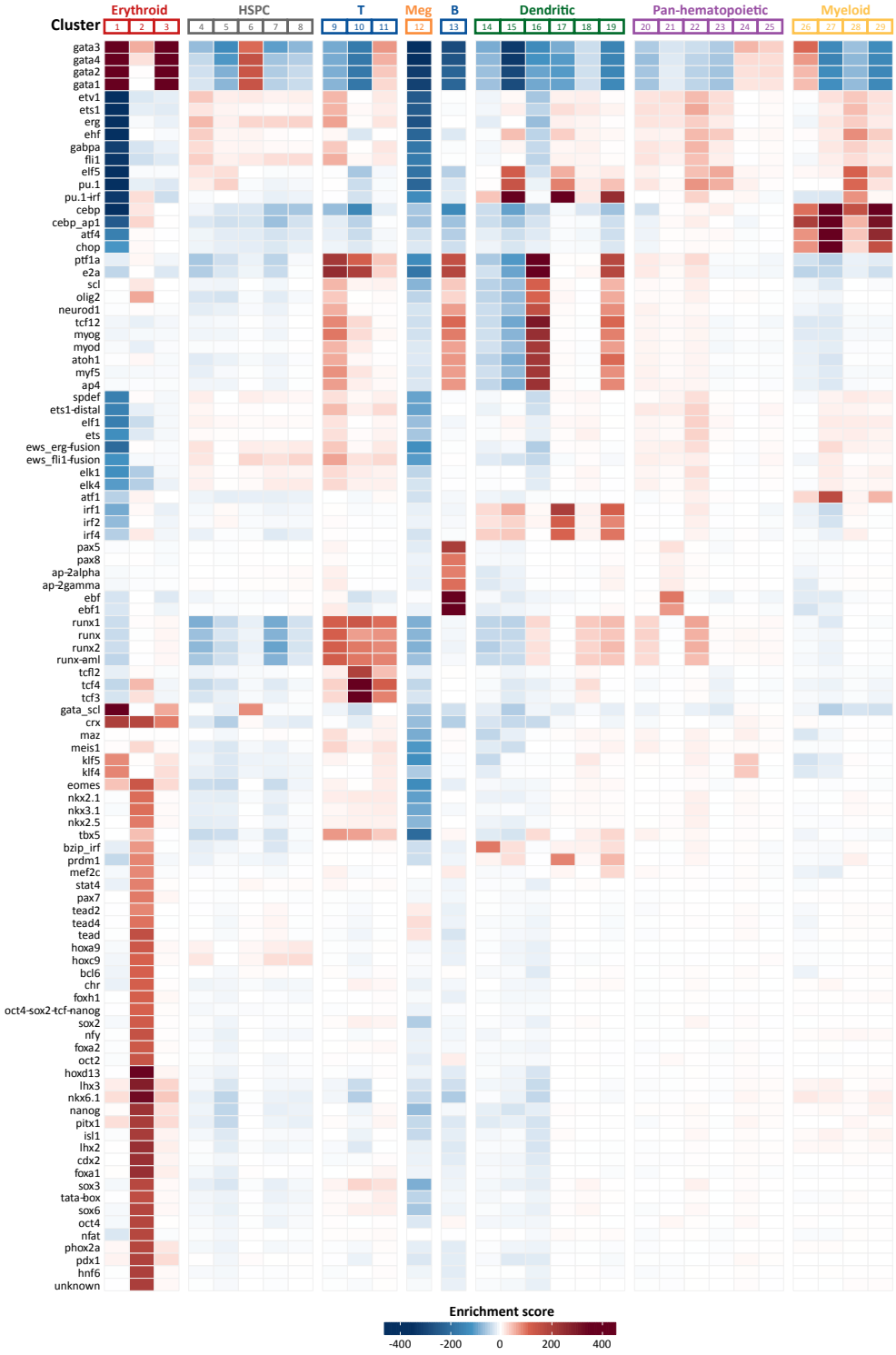

Figure A.7: Binding motifs of key hematopoietic transcription factors are differentially enriched across DNA methylation programs. Enrichment analysis of HOMER transcription factor binding motifs (same as Figure 3.13). Heatmap showing the enrichment scores of the 100 most significantly enriched transcription factor binding motifs.

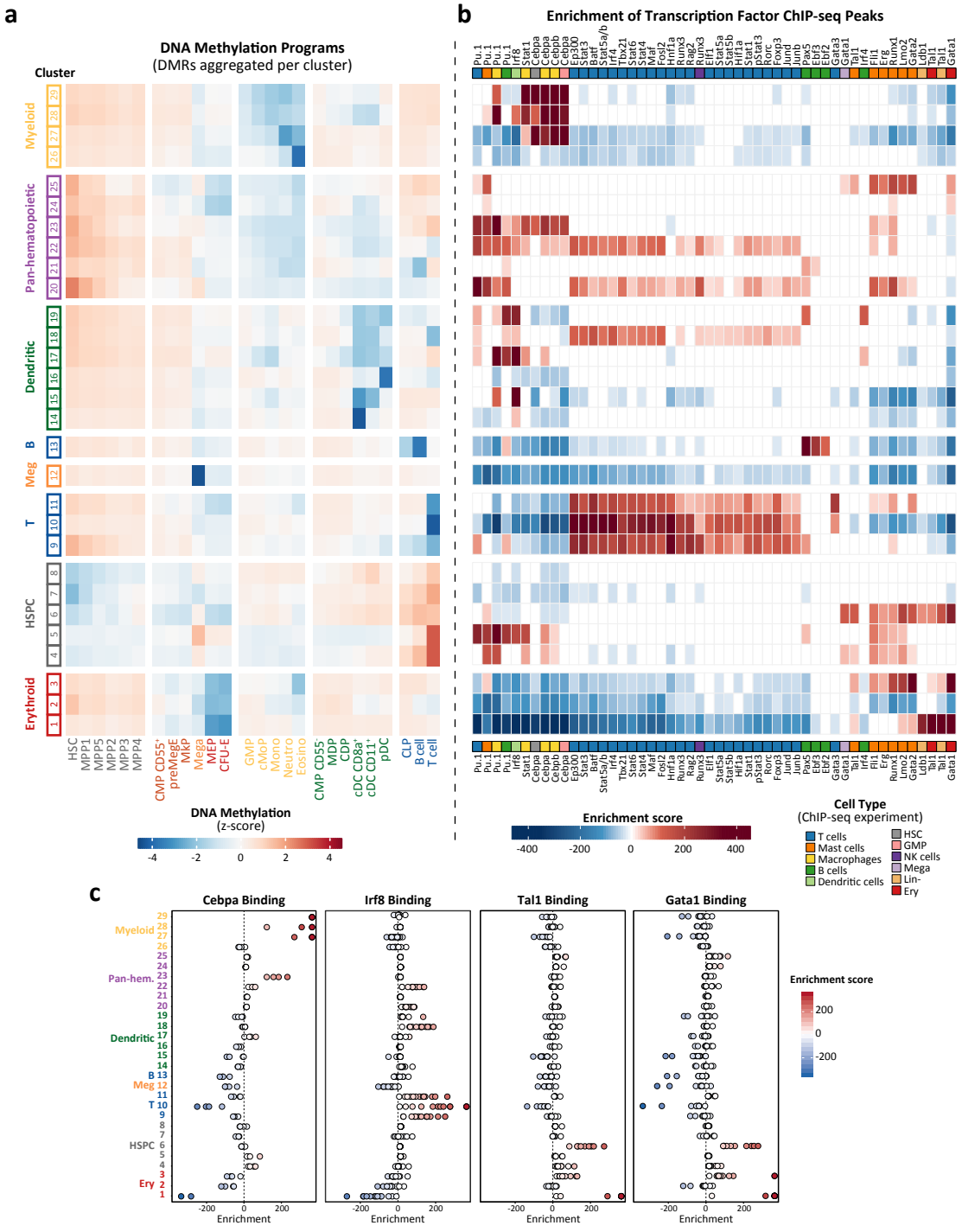

Figure A.8: ChIP-seq peaks of key hematopoietic transcription factors are enriched in DNA methylation programs. ChIP-seq data generated in primary hematopoietic cells was downloaded from the CODEX database. Enrichment analysis of ChIP-seq peaks of different transcription factor binding sites was performed using the Fisher's exact test (in collaboration with Stephen Kraemer). **a**, Heatmap showing the aggregated methylation levels per methylation cluster. **b**, Heatmap showing the enrichment scores of the 50 most significant transcription factor binding sites. Red indicates significant enrichment ($p < 1x10^{-20}$) of binding motif, blue indicates depletion ($p < 1x10^{-20}$). Clustering of columns was performed with hierarchical clustering using Ward's method. **c**, Bar plots showing the enrichments for all downloaded ChIP-seq experiments of Cepba, Irf8, Tal1, and Gata1.

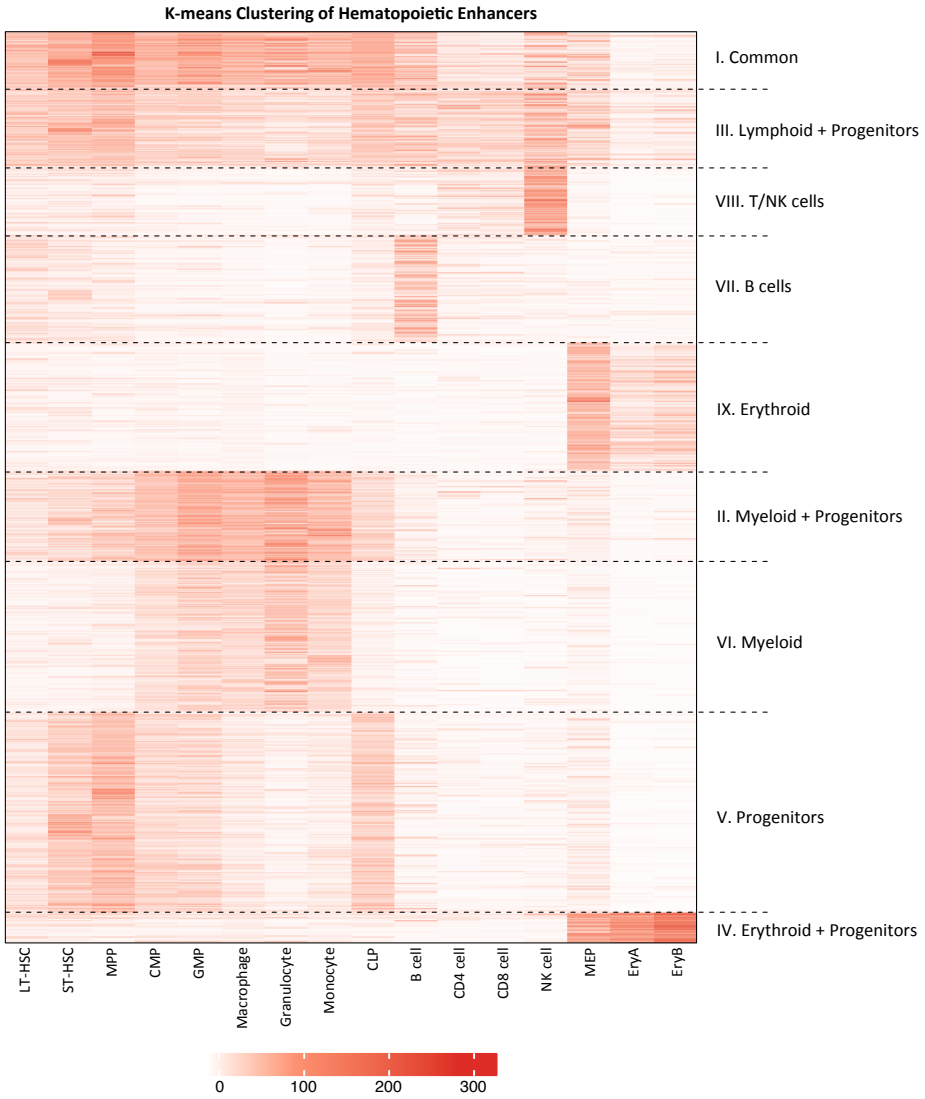

Figure A.9: Recapitulation of K-means clustering of hematopoietic enhancer regions identified by Lara-Astiaso *et al.* Heatmap showing 48,415 hematopoietic enhancer regions clustered by K-means clustering (K = 9) based on H3K4me1 average read counts. Cluster annotation is depicted on the right.

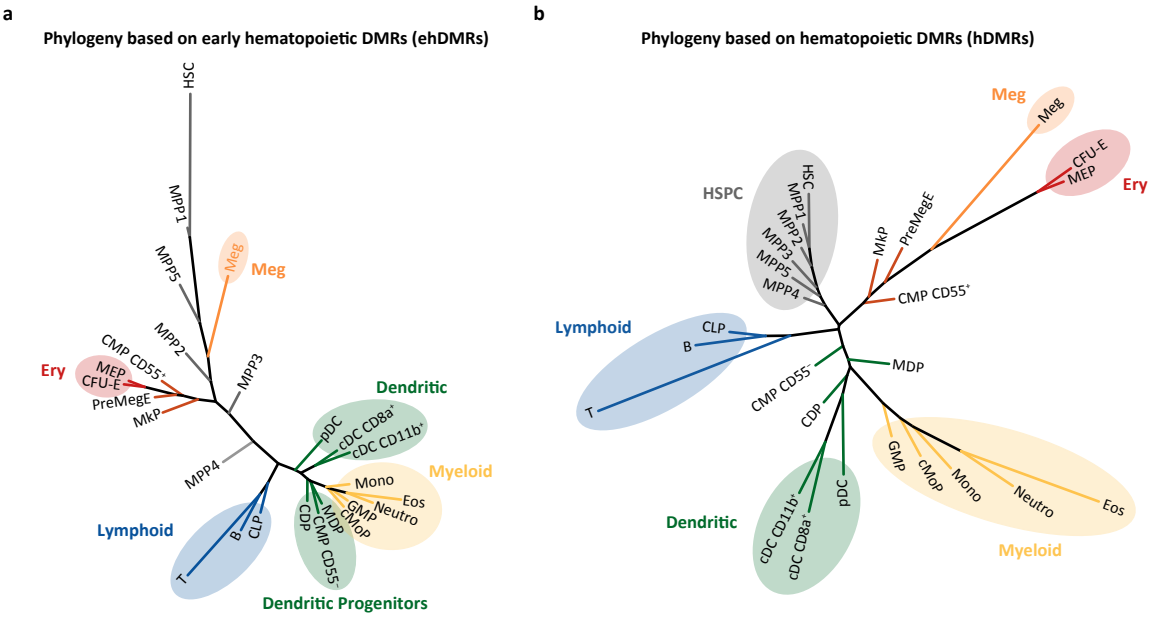

Figure A.10: DMRs identified within the primitive hematopoietic stem and progenitor compartment are sufficient to separate terminally differentiated hematopoietic cell lineages in a classical phylogenetic tree model. Phylogeny of hematopoietic cell populations based on 7,423 ehDMRs **(a)** and 147,232 hDMRs **(b)** using Manhattan distance recapitulates known hematopoietic lineage differentiation trajectories.

References

[Adelman et al., 2019] Adelman, E. R., Huang, H. T., Roisman, A., Olsson, A., Colaprico, A., Qin, T., Lindsley, R. C., Bejar, R., Salomonis, N., Grimes, H. L., and Figueroa, M. E. (2019). Aging Human Hematopoietic Stem Cells Manifest Profound Epigenetic Reprogramming of Enhancers That May Predispose to Leukemia. *Cancer Discov*.

[Adler and Murdoch, 2019] Adler, D. and Murdoch, d. (2019). *rgl: 3D Visualization Using OpenGL*. R package version 0.100.19.

[Adolfsson et al., 2001] Adolfsson, J., Borge, O. J., Bryder, D., Theilgaard-Monch, K., Astrand-Grundstrom, I., Sitnicka, E., Sasaki, Y., and Jacobsen, S. E. (2001). Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity*, 15(4):659–69.

[Adolfsson et al., 2005] Adolfsson, J., Mansson, R., Buza-Vidas, N., Hultquist, A., Liuba, K., Jensen, C. T., Bryder, D., Yang, L., Borge, O. J., Thoren, L. A., Anderson, K., Sitnicka, E., Sasaki, Y., Sigvardsson, M., and Jacobsen, S. E. (2005). Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell*, 121(2):295–306.

[Akashi et al., 2000] Akashi, K., Traver, D., Miyamoto, T., and Weissman, I. L. (2000). A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature*, 404(6774):193–7.

[Akunuru and Geiger, 2016] Akunuru, S. and Geiger, H. (2016). Aging, Clonality, and Rejuvenation of Hematopoietic Stem Cells. *Trends Mol Med*, 22(8):701–712.

[Allis and Jenuwein, 2016] Allis, C. D. and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat Rev Genet*, 17(8):487–500.

[Amabile et al., 2016] Amabile, A., Migliara, A., Capasso, P., Biffi, M., Cittaro, D., Naldini, L., and Lombardo, A. (2016). Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. *Cell*, 167(1):219–232 e14.

[Angerer et al., 2016] Angerer, P., Haghverdi, L., Buttner, M., Theis, F. J., Marr, C., and Buettner, F. (2016). destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics*, 32(8):1241–3.

[Angermueller et al., 2016] Angermueller, C., Clark, S. J., Lee, H. J., Macaulay, I. C., Teng, M. J., Hu, T. X., Krueger, F., Smallwood, S., Ponting, C. P., Voet, T., Kelsey, G., Stegle, O., and Reik, W. (2016). Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat Methods*, 13(3):229–232.

[Baldridge et al., 2010] Baldridge, M. T., King, K. Y., Boles, N. C., Weksberg, D. C., and Goodell, M. A. (2010). Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. *Nature*, 465(7299):793–7.

[Barros-Silva et al., 2018] Barros-Silva, D., Marques, C. J., Henrique, R., and Jeronimo, C. (2018). Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications. *Genes (Basel)*, 9(9).

[Bartholdy et al., 2018] Bartholdy, B., Lajugie, J., Yan, Z., Zhang, S., Mukhopadhyay, R., Greally, J. M., Suzuki, M., and Bouhassira, E. E. (2018). Mechanisms of establishment and functional significance of DNA demethylation during erythroid differentiation. *Blood Adv*, 2(15):1833–1852.

[Beerman et al., 2010] Beerman, I., Bhattacharya, D., Zandi, S., Sigvardsson, M., Weissman, I. L., Bryder, D., and Rossi, D. J. (2010). Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *Proc Natl Acad Sci U S A*, 107(12):5465–70.

[Beerman et al., 2013] Beerman, I., Bock, C., Garrison, B. S., Smith, Z. D., Gu, H., Meissner, A., and Rossi, D. J. (2013). Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. *Cell Stem Cell*, 12(4):413–25.

[Beerman and Rossi, 2014] Beerman, I. and Rossi, D. J. (2014). Epigenetic regulation of hematopoietic stem cell aging. *Experimental Cell Research*, 329(2):192–199.

[Beerman et al., 2014] Beerman, I., Seita, J., Inlay, M., Weissman, I., and Rossi, D. (2014). Quiescent Hematopoietic Stem Cells Accumulate DNA Damage during Aging that Is Repaired upon Entry into Cell Cycle. *Cell Stem Cell*, 15(1):37–50.

[Ben-Hattar and Jiricny, 1988] Ben-Hattar, J. and Jiricny, J. (1988). Methylation of single CpG dinucleotides within a promoter element of the Herpes simplex virus tk gene reduces its transcription in vivo. *Gene*, 65(2):219–27.

[Bernitz et al., 2016] Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H., and Moore, K. (2016). Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell*, 167(5):1296–1309 e10.

[Bianchi et al., 2016] Bianchi, E., Norfo, R., Pennucci, V., Zini, R., and Manfredini, R. (2016). Genomic landscape of megakaryopoiesis and platelet function defects. *Blood*, 127(10):1249–59.

[Bock, 2012] Bock, C. (2012). Analysing and interpreting DNA methylation data. *Nat Rev Genet*, 13(10):705–19.

[Bock et al., 2012] Bock, C., Beerman, I., Lien, W. H., Smith, Z. D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D. J., and Meissner, A. (2012). DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Mol Cell*, 47(4):633–47.

[Bocklandt et al., 2011] Bocklandt, S., Lin, W., Sehl, M. E., Sanchez, F. J., Sinsheimer, J. S., Horvath, S., and Vilain, E. (2011). Epigenetic predictor of age. *PLoS One*, 6(6):e14821.

[Bolger et al., 2014] Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–20.

[Bormann et al., 2018] Bormann, F., Rodriguez-Paredes, M., Lasitschka, F., Edelmann, D., Musch, T., Benner, A., Bergman, Y., Dieter, S. M., Ball, C. R., Glimm, H., Linhart, H. G., and Lyko, F. (2018). Cell-of-Origin DNA Methylation Signatures Are Maintained during Colorectal Carcinogenesis. *Cell Rep*, 23(11):3407–3418.

[Broske et al., 2009] Broske, A. M., Vockentanz, L., Kharazi, S., Huska, M. R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R., Nerlov, C., Leutz, A., Andrade-Navarro, M. A., Jacobsen, S. E., and Rosenbauer, F. (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat Genet*, 41(11):1207–15.

[Buenrostro et al., 2018] Buenrostro, J. D., Corces, M. R., Lareau, C. A., Wu, B., Schep, A. N., Aryee, M. J., Majeti, R., Chang, H. Y., and Greenleaf, W. J. (2018). Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell*, 173(6):1535–1548 e16.

[Buenrostro et al., 2015] Buenrostro, J. D., Wu, B., Chang, H. Y., and Greenleaf, W. J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol*, 109:21 29 1–9.

[Busch et al., 2015] Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., Reth, M., Hofer, T., and Rodewald, H. R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*, 518(7540):542–6.

[Butler et al., 2018] Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol*, 36(5):411–420.

[Cabezas-Wallscheid et al., 2017] Cabezas-Wallscheid, N., Buettner, F., Sommerkamp, P., Klimmeck, D., Ladel, L., Thalheimer, F. B., Pastor-Flores, D., Roma, L. P., Renders, S., Zeisberger, P., Przybylla, A., Schonberger, K., Scognamiglio, R., Altamura, S., Florian, C. M., Fawaz, M., Vonficht, D., Tesio, M., Collier, P., Pavlinic, D., Geiger, H., Schroeder, T., Benes, V., Dick, T. P., Rieger, M. A., Stegle, O., and Trumpp, A. (2017). Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. *Cell*, 169(5):807–823 e19.

[Cabezas-Wallscheid et al., 2014] Cabezas-Wallscheid, N., Klimmeck, D., Hansson, J., Lipka, D. B., Reyes, A., Wang, Q., Weichenhan, D., Lier, A., von Paleske, L., Renders, S., Wunsche, P., Zeisberger, P., Brocks, D., Gu, L., Herrmann, C., Haas, S., Essers, M. A. G., Brors, B., Eils, R., Huber, W., Milsom, M. D., Plass, C., Krijgsveld, J., and Trumpp, A. (2014). Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell*, 15(4):507–522.

[Cancelas and Williams, 2009] Cancelas, J. A. and Williams, D. A. (2009). Rho GTPases in hematopoietic stem cell functions. *Curr Opin Hematol*, 16(4):249–54.

[Carrelha et al., 2018] Carrelha, J., Meng, Y., Kettyle, L. M., Luis, T. C., Norfo, R., Alcolea, V., Boukarabila, H., Grasso, F., Gambardella, A., Grover, A., Hogstrand, K., Lord, A. M., Sanjuan-Pla, A., Woll, P. S., Nerlov, C., and Jacobsen, S. E. W. (2018). Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. *Nature*, 554(7690):106–111.

[Challen et al., 2010] Challen, G. A., Boles, N. C., Chambers, S. M., and Goodell, M. A. (2010). Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. *Cell Stem Cell*, 6(3):265–78.

[Challen et al., 2011] Challen, G. A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J. S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., Liang, S., Lu, Y., Darlington, G. J., Meissner, A., Issa, J. P., Godley, L. A., Li, W., and Goodell, M. A. (2011). Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet*, 44(1):23–31.

[Challen et al., 2014] Challen, G. A., Sun, D., Mayle, A., Jeong, M., Luo, M., Rodriguez, B., Mallaney, C., Celik, H., Yang, L., Xia, Z., Cullen, S., Berg, J., Zheng, Y., Darlington, G. J., Li, W., and Goodell, M. A. (2014). Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. *Cell Stem Cell*, 15(3):350–364.

[Chen et al., 2018] Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17):i884–i890.

[Cheshier et al., 2007] Cheshier, S. H., Prohaska, S. S., and Weissman, I. L. (2007). The effect of bleeding on hematopoietic stem cell cycling and self-renewal. *Stem Cells Dev*, 16(5):707–17.

[Choudry and Frontini, 2016] Choudry, F. A. and Frontini, M. (2016). Epigenetic Control of Haematopoietic Stem Cell Aging and Its Clinical Implications. *Stem Cells Int*, 2016:5797521.

[Clark et al., 2018] Clark, S. J., Argelaguet, R., Kapourani, C. A., Stubbs, T. M., Lee, H. J., Alda-Catalinas, C., Krueger, F., Sanguinetti, G., Kelsey, G., Marioni, J. C., Stegle, O., and Reik, W. (2018). scNMT-seq enables joint profiling of chromatin accessibility DNA methylation and transcription in single cells. *Nat Commun*, 9(1):781.

[Clark et al., 2017] Clark, S. J., Smallwood, S. A., Lee, H. J., Krueger, F., Reik, W., and Kelsey, G. (2017). Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat Protoc*, 12(3):534–547.

[Corces et al., 2016] Corces, M. R., Buenrostro, J. D., Wu, B., Greenside, P. G., Chan, S. M., Koenig, J. L., Snyder, M. P., Pritchard, J. K., Kundaje, A., Greenleaf, W. J., Majeti, R., and Chang, H. Y. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*, 48(10):1193–203.

[Dahlin et al., 2018] Dahlin, J. S., Hamey, F. K., Pijuan-Sala, B., Shepherd, M., Lau, W. W. Y., Nestorowa, S., Weinreb, C., Wolock, S., Hannah, R., Diamanti, E., Kent, D. G., Gottgens, B., and Wilson, N. K. (2018). A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. *Blood*, 131(21):e1–e11.

[Davis et al., 2018] Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., Hilton, J. A., Jain, K., Baymuradov, U. K., Narayanan, A. K., Onate, K. C., Graham, K., Miyasato, S. R., Dreszer, T. R., Strattan, J. S., Jolanki, O., Tanaka, F. Y., and Cherry, J. M. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*, 46(D1):D794–D801.

[Dawlaty et al., 2014] Dawlaty, M. M., Breiling, A., Le, T., Barrasa, M. I., Raddatz, G., Gao, Q., Powell, B. E., Cheng, A. W., Faull, K. F., Lyko, F., and Jaenisch, R. (2014). Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev Cell*, 29(1):102–11.

[de Graaf et al., 2016] de Graaf, C. A., Choi, J., Baldwin, T. M., Bolden, J. E., Fairfax, K. A., Robinson, A. J., Biben, C., Morgan, C., Ramsay, K., Ng, A. P., Kauppi, M., Kruse, E. A., Sargeant, T. J., Seidenman, N., D'Amico, A., D'Ombrain, M. C., Lucas, E. C., Koernig, S., Baz Morelli, A., Wilson, M. J., Dower, S. K., Williams, B., Heazlewood, S. Y., Hu, Y., Nilsson, S. K., Wu, L., Smyth, G. K., Alexander, W. S., and Hilton, D. J. (2016). Haemopedia: An Expression Atlas of Murine Hematopoietic Cells. *Stem Cell Reports*, 7(3):571–582.

[Delhommeau et al., 2009] Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Masse, A., Kosmider, O., Le Couedic, J. P., Robert, F., Alberdi, A., Lecluse, Y., Plo, I., Dreyfus, F. J., Marzac, C., Casadevall, N., Lacombe, C., Romana, S. P., Dessen, P., Soulier, J., Viguie, F., Fontenay, M., Vainchenker, W., and Bernard, O. A. (2009). Mutation in TET2 in myeloid cancers. *N Engl J Med*, 360(22):2289–301.

[Desper and Gascuel, 2002] Desper, R. and Gascuel, O. (2002). Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J Comput Biol*, 9(5):687–705.

[Doskocil and Sorm, 1962] Doskocil, J. and Sorm, F. (1962). Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. *Biochim Biophys Acta*, 55:953–9.

[Drissen et al., 2016] Drissen, R., Buza-Vidas, N., Woll, P., Thongjuea, S., Gambardella, A., Giustacchini, A., Mancini, E., Zriwil, A., Lutteropp, M., Grover, A., Mead, A., Sitnicka, E., Jacobsen, S. E. W., and Nerlov, C. (2016). Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. *Nat Immunol*, 17(6):666–676.

[Drissen et al., 2019] Drissen, R., Thongjuea, S., Theilgaard-Monch, K., and Nerlov, C. (2019). Identification of two distinct pathways of human myelopoiesis. *Sci Immunol*, 4(35).

[Du et al., 2010] Du, P., Zhang, X., Huang, C. C., Jafari, N., Kibbe, W. A., Hou, L., and Lin, S. M. (2010). Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*, 11:587.

[Dykstra et al., 2007] Dykstra, B., Kent, D., Bowie, M., McCaffrey, L., Hamilton, M., Lyons, K., Lee, S. J., Brinkman, R., and Eaves, C. (2007). Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell*, 1(2):218–29.

[Dykstra et al., 2011] Dykstra, B., Olthof, S., Schreuder, J., Ritsema, M., and de Haan, G. (2011). Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. *J Exp Med*, 208(13):2691–703.

[Ergen et al., 2012] Ergen, A. V., Boles, N. C., and Goodell, M. A. (2012). Rantes/Ccl5 influences hematopoietic stem cell subtypes and causes myeloid skewing. *Blood*, 119(11):2500–9.

[Essers et al., 2009] Essers, M. A., Offner, S., Blanco-Bose, W. E., Waibler, Z., Kalinke, U., Duchosal, M. A., and Trumpp, A. (2009). IFNalpha activates dormant haematopoietic stem cells in vivo. *Nature*, 458(7240):904–8.

[Farlik et al., 2016] Farlik, M., Halbritter, F., Muller, F., Choudry, F. A., Ebert, P., Klughammer, J., Farrow, S., Santoro, A., Ciaurro, V., Mathur, A., Uppal, R., Stunnenberg, H. G., Ouwehand, W. H., Laurenti, E., Lengauer, T., Frontini, M., and Bock, C. (2016). DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell Stem Cell*, 19(6):808–822.

[Farlik et al., 2015] Farlik, M., Sheffield, N. C., Nuzzo, A., Datlinger, P., Schonegger, A., Klughammer, J., and Bock, C. (2015). Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics. *Cell Rep*, 10(8):1386–97.

[Feinberg and Vogelstein, 1983] Feinberg, A. P. and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*, 301(5895):89–92.

[Florian et al., 2012] Florian, M. C., Dorr, K., Niebel, A., Daria, D., Schrezenmeier, H., Rojewski, M., Filippi, M. D., Hasenberg, A., Gunzer, M., Scharffetter-Kochanek, K., Zheng, Y., and Geiger, H. (2012). Cdc42 activity regulates hematopoietic stem cell aging and rejuvenation. *Cell Stem Cell*, 10(5):520–30.

[Frommer et al., 1992] Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A*, 89(5):1827–31.

[Garagnani et al., 2012] Garagnani, P., Bacalini, M. G., Pirazzini, C., Gori, D., Giuliani, C., Mari, D., Di Blasio, A. M., Gentilini, D., Vitale, G., Collino, S., Rezzi, S., Castellani, G., Capri, M., Salvioli, S., and Franceschi, C. (2012). Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*, 11(6):1132–4.

[Gekas and Graf, 2013] Gekas, C. and Graf, T. (2013). CD41 expression marks myeloid-biased adult hematopoietic stem cells and increases with age. *Blood*, 121(22):4463–72.

[Genovese et al., 2014] Genovese, G., Kahler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., Purcell, S. M., Svantesson, O., Landen, M., Hoglund, M., Lehmann, S., Gabriel, S. B., Moran, J. L., Lander, E. S., Sullivan, P. F., Sklar, P., Gronberg, H., Hultman, C. M., and McCarroll, S. A. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*, 371(26):2477–87.

[Giladi et al., 2018] Giladi, A., Paul, F., Herzog, Y., Lubling, Y., Weiner, A., Yofe, I., Jaitin, D., Cabezas-Wallscheid, N., Dress, R., Ginhoux, F., Trumpp, A., Tanay, A., and Amit, I. (2018). Single-cell characterization of haematopoietic progenitors and their trajectories in homeostasis and perturbed haematopoiesis. *Nat Cell Biol*, 20(7):836–846.

[Greenberg and Bourc'his, 2019] Greenberg, M. V. C. and Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol*.

[Gross et al., 2016] Gross, A. M., Jaeger, P. A., Kreisberg, J. F., Licon, K., Jepsen, K. L., Khosroheidari, M., Morsey, B. M., Swindells, S., Shen, H., Ng, C. T., Flagg, K., Chen, D., Zhang, K., Fox, H. S., and Ideker, T. (2016). Methylome-wide Analysis of Chronic HIV Infection Reveals Five-Year Increase in Biological Age and Epigenetic Targeting of HLA. *Mol Cell*, 62(2):157–168.

[Grover et al., 2016] Grover, A., Sanjuan-Pla, A., Thongjuea, S., Carrelha, J., Giustacchini, A., Gambardella, A., Macaulay, I., Mancini, E., Luis, T. C., Mead, A., Jacobsen, S. E., and Nerlov, C. (2016). Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. *Nat Commun*, 7:11075.

[Gu et al., 2011] Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., and Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc*, 6(4):468–81.

[Gu et al., 2016] Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18):2847–9.

[Guo et al., 2013] Guo, G., Luc, S., Marco, E., Lin, T. W., Peng, C., Kerenyi, M. A., Beyaz, S., Kim, W., Xu, J., Das, P. P., Neff, T., Zou, K., Yuan, G. C., and Orkin, S. H. (2013). Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell*, 13(4):492–505.

[Haas et al., 2015] Haas, S., Hansson, J., Klimmeck, D., Loeffler, D., Velten, L., Uckelmann, H., Wurzer, S., Prendergast, A. M., Schnell, A., Hexel, K., Santarella-Mellwig, R., Blaszkiewicz, S., Kuck, A., Geiger, H., Milsom, M. D., Steinmetz, L. M., Schroeder, T., Trumpp, A., Krijgsveld, J., and Essers, M. A. (2015). Inflammation-Induced Emergency Megakaryopoiesis Driven by Hematopoietic Stem Cell-like Megakaryocyte Progenitors. *Cell Stem Cell*, 17(4):422–34.

[Haas et al., 2018] Haas, S., Trumpp, A., and Milsom, M. D. (2018). Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. *Cell Stem Cell*, 22(5):627–638.

[Haghverdi et al., 2015] Haghverdi, L., Buettner, F., and Theis, F. J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*, 31(18):2989–98.

[Hahne and Ivanek, 2016] Hahne, F. and Ivanek, R. (2016). Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol*, 1418:335–51.

[Hannum et al., 2013] Hannum, G., Guinney, J., Zhao, L., Zhang, L., Hughes, G., Sadda, S., Klotzle, B., Bibikova, M., Fan, J. B., Gao, Y., Deconde, R., Chen, M., Rajapakse, I., Friend, S., Ideker, T., and Zhang, K. (2013). Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell*, 49(2):359–367.

[Harrison et al., 1993] Harrison, D. E., Jordan, C. T., Zhong, R. K., and Astle, C. M. (1993). Primitive hemopoietic stem cells: direct assay of most productive populations by competitive repopulation with simple binomial, correlation and covariance calculations. *Exp Hematol*, 21(2):206–19.

[He et al., 2011] He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., Sun, Y., Li, X., Dai, Q., Song, C. X., Zhang, K., He, C., and Xu, G. L. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, 333(6047):1303–7.

[Heinz et al., 2010] 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. *Mol Cell*, 38(4):576–89.

[Ho et al., 2019] Ho, Y. H., Del Toro, R., Rivera-Torres, J., Rak, J., Korn, C., Garcia-Garcia, A., Macias, D., Gonzalez-Gomez, C., Del Monte, A., Wittner, M., Waller, A. K., Foster, H. R., Lopez-Otin, C., Johnson, R. S., Nerlov, C., Ghevaert, C., Vainchenker, W., Louache, F., Andres, V., and Mendez-Ferrer, S. (2019). Remodeling of Bone Marrow Hematopoietic Stem Cell Niches Promotes Myeloid Cell Expansion during Premature or Physiological Aging. *Cell Stem Cell*, 25(3):407–418 e6.

[Hodges et al., 2011] Hodges, E., Molaro, A., Dos Santos, C. O., Thekkat, P., Song, Q., Uren, P. J., Park, J., Butler, J., Rafii, S., McCombie, W. R., Smith, A. D., and Hannon, G. J. (2011). Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell*, 44(1):17–28.

[Holliday and Pugh, 1975] Holliday, R. and Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. *Science*, 187(4173):226–32.

[Holoch and Moazed, 2015] Holoch, D. and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet*, 16(2):71–84.

[Horvath, 2013] Horvath, S. (2013). DNA methylation age of human tissues and cell types. *Genome Biol*, 14(10):R115.

[Horvath and Levine, 2015] Horvath, S. and Levine, A. J. (2015). HIV-1 Infection Accelerates Age According to the Epigenetic Clock. *J Infect Dis*, 212(10):1563–73.

[Hotchkiss, 1948] Hotchkiss, R. D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem*, 175(1):315–32.

[Huang et al., 2017] Huang, Y. H., Su, J., Lei, Y., Brunetti, L., Gundry, M. C., Zhang, X., Jeong, M., Li, W., and Goodell, M. A. (2017). DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A. *Genome Biol*, 18(1):176.

[Hui et al., 2018] Hui, T., Cao, Q., Wegrzyn-Woltosz, J., O'Neill, K., Hammond, C. A., Knapp, D., Laks, E., Moksa, M., Aparicio, S., Eaves, C. J., Karsan, A., and Hirst, M. (2018). High-Resolution Single-Cell DNA Methylation Measurements Reveal Epigenetically Distinct Hematopoietic Stem Cell Subpopulations. *Stem Cell Reports*, 11(2):578–592.

[Huveneers and Danen, 2009] Huveneers, S. and Danen, E. H. (2009). Adhesion signaling - crosstalk between integrins, Src and Rho. *J Cell Sci*, 122(Pt 8):1059–69.

[Ito et al., 2011] 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(6047):1300–3.

[Jaiswal et al., 2014] Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B. G., Lindsley, R. C., Mermel, C. H., Burtt, N., Chavez, A., Higgins, J. M., Moltchanov, V., Kuo, F. C., Kluk, M. J., Henderson, B., Kinnunen, L., Koistinen, H. A., Ladenvall, C., Getz,

G., Correa, A., Banahan, B. F., Gabriel, S., Kathiresan, S., Stringham, H. M., McCarthy, M. I., Boehnke, M., Tuomilehto, J., Haiman, C., Groop, L., Atzmon, G., Wilson, J. G., Neuberg, D., Altshuler, D., and Ebert, B. L. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*, 371(26):2488–98.

[Javierre et al., 2016] Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., Cairns, J., Wingett, S. W., Varnai, C., Thiecke, M. J., Burden, F., Farrow, S., Cutler, A. J., Rehnstrom, K., Downes, K., Grassi, L., Kostadima, M., Freire-Pritchett, P., Wang, F., Consortium, B., Stunnenberg, H. G., Todd, J. A., Zerbino, D. R., Stegle, O., Ouwehand, W. H., Frontini, M., Wallace, C., Spivakov, M., and Fraser, P. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*, 167(5):1369–1384 e19.

[Jeong et al., 2018] Jeong, M., Park, H. J., Celik, H., Ostrander, E. L., Reyes, J. M., Guzman, A., Rodriguez, B., Lei, Y., Lee, Y., Ding, L., Guryanova, O. A., Li, W., Goodell, M. A., and Challen, G. A. (2018). Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo. *Cell Rep*, 23(1):1–10.

[Ji et al., 2010] Ji, H., Ehrlich, L. I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M. J., Irizarry, R. A., Kim, K., Rossi, D. J., Inlay, M. A., Serwold, T., Karsunky, H., Ho, L., Daley, G. Q., Weissman, I. L., and Feinberg, A. P. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature*, 467(7313):338–42.

[Jones, 2012] Jones, P. A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet*, 13(7):484–92.

[Karamitros et al., 2018] Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., Quek, L., Otto, G., Repapi, E., Doondeea, J., Usukhbayar, B., Calvo, J., Taylor, S., Goardon, N., Six, E., Pflumio, F., Porcher, C., Majeti, R., Gottgens, B., and Vyas, P. (2018). Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells. *Nat Immunol*, 19(1):85–97.

[Kim and Costello, 2017] Kim, M. and Costello, J. (2017). DNA methylation: an epigenetic mark of cellular memory. *Exp Mol Med*, 49(4):e322.

[Klemm et al., 2019] Klemm, S. L., Shipony, Z., and Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet*, 20(4):207–220.

[Kondo et al., 1997] Kondo, M., Weissman, I. L., and Akashi, K. (1997). Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell*, 91(5):661–72.

[Kovtonyuk et al., 2016] Kovtonyuk, L. V., Fritsch, K., Feng, X., Manz, M. G., and Takizawa, H. (2016). Inflamm-Aging of Hematopoiesis, Hematopoietic Stem Cells, and the Bone Marrow Microenvironment. *Front Immunol*, 7:502.

[Krivtsov et al., 2013] Krivtsov, A. V., Figueroa, M. E., Sinha, A. U., Stubbs, M. C., Feng, Z., Valk, P. J., Delwel, R., Dohner, K., Bullinger, L., Kung, A. L., Melnick, A. M., and Armstrong, S. A. (2013). Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia*, 27(4):852–60.

[Krueger and Andrews, 2011] Krueger, F. and Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27(11):1571–2.

[Kulis et al., 2015] Kulis, M., Merkel, A., Heath, S., Queiros, A. C., Schuyler, R. P., Castellano, G., Beekman, R., Raineri, E., Esteve, A., Clot, G., Verdaguer-Dot, N., Duran-Ferrer, M., Russinol, N., Vilarrasa-Blasi, R., Ecker, S., Pancaldi, V., Rico, D., Agueda, L., Blanc, J., Richardson, D., Clarke, L., Datta, A., Pascual, M., Agirre, X., Prosper, F., Alignani, D., Paiva, B., Caron, G., Fest, T., Muench, M. O., Fomin, M. E., Lee, S. T., Wiemels, J. L., Valencia, A., Gut, M., Flicek, P., Stunnenberg, H. G., Siebert, R., Kuppers, R., Gut, I. G., Campo, E., and Martin-Subero, J. I. (2015). Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat Genet*, 47(7):746–56.

[Landan et al., 2012] Landan, G., Cohen, N. M., Mukamel, Z., Bar, A., Molchadsky, A., Brosh, R., Horn-Saban, S., Zalcenstein, D. A., Goldfinger, N., Zundelevich, A., Gal-Yam, E. N., Rotter, V., and Tanay, A. (2012). Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat Genet*, 44(11):1207–14.

[Langfelder et al., 2016] Langfelder, P., Zhang, B., and Horvath, S. (2016). *dynamicTreeCut: Methods for Detection of Clusters in Hierarchical Clustering Dendrograms*. R package version 1.63-1.

[Lara-Astiaso et al., 2014] Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D. A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., Friedman, N., and Amit, I. (2014). Immunogenetics. Chromatin state dynamics during blood formation. *Science*, 345(6199):943–9.

[Laszlo et al., 2013] Laszlo, A. H., Derrington, I. M., Brinkerhoff, H., Langford, K. W., Nova, I. C., Samson, J. M., Bartlett, J. J., Pavlenok, M., and Gundlach, J. H. (2013). Detection and mapping of 5-methylcytosine and 5-hydroxymethylcytosine with nanopore MspA. *Proc Natl Acad Sci U S A*, 110(47):18904–9.

[Laurenti and Gottgens, 2018] Laurenti, E. and Gottgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. *Nature*, 553(7689):418–426.

[Lê, S. and Josse, J. and Husson, F., 2008] Lê, S. and Josse, J. and Husson, F. (2008). FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, 25(1):1–18.

[Levo and Segal, 2014] Levo, M. and Segal, E. (2014). In pursuit of design principles of regulatory sequences. *Nat Rev Genet*, 15(7):453–68.

[Ley et al., 2010] Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. E., Kandoth, C., Payton, J. E., Baty, J., Welch, J., Harris, C. C., Lichti, C. F., Townsend, R. R., Fulton, R. S., Dooling, D. J., Koboldt, D. C., Schmidt, H., Zhang, Q., Osborne, J. R., Lin, L., O'Laughlin, M., McMichael, J. F., Delehaunty, K. D., McGrath, S. D., Fulton, L. A., Magrini, V. J., Vickery, T. L., Hundal, J., Cook, L. L., Conyers, J. J., Swift, G. W., Reed, J. P., Alldredge, P. A., Wylie, T., Walker, J., Kalicki, J., Watson, M. A., Heath, S., Shannon, W. D., Varghese, N., Nagarajan, R., Westervelt, P., Tomasson, M. H., Link, D. C., Graubert, T. A., DiPersio, J. F., Mardis, E. R., and Wilson, R. K. (2010). DNMT3A mutations in acute myeloid leukemia. *N Engl J Med*, 363(25):2424–33.

[Li et al., 1993] Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature*, 366(6453):362–5.

[Li et al., 1992] Li, E., Bestor, T. H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6):915–26.

[Li, 2013] Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

[Li et al., 2009] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–9.

[Lieberman-Aiden et al., 2009] Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., and Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950):289–93.

[Lipka et al., 2014] Lipka, D. B., Wang, Q., Cabezas-Wallscheid, N., Klimmeck, D., Weichenhan, D., Herrmann, C., Lier, A., Brocks, D., von Paleske, L., Renders, S., Wunsche, P., Zeisberger, P., Gu, L., Haas, S., Essers, M. A., Brors, B., Eils, R., Trumpp, A., Milsom, M. D., and Plass, C. (2014). Identification of DNA methylation changes at cis-regulatory elements during early steps of HSC differentiation using tagmentation-based whole genome bisulfite sequencing. *Cell Cycle*, 13(22):3476–87.

[Lister et al., 2009] 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., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., and Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271):315–22.

[Lopez-Otin et al., 2013] Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. *Cell*, 153(6):1194–217.

[Luo et al., 2018] Luo, C., Hajkova, P., and Ecker, J. R. (2018). Dynamic DNA methylation: In the right place at the right time. *Science*, 361(6409):1336–1340.

[Macaulay et al., 2017] Macaulay, I. C., Ponting, C. P., and Voet, T. (2017). Single-Cell Multiomics: Multiple Measurements from Single Cells. *Trends Genet*, 33(2):155–168.

[Macaulay et al., 2016] Macaulay, I. C., Svensson, V., Labalette, C., Ferreira, L., Hamey, F., Voet, T., Teichmann, S. A., and Cvejic, A. (2016). Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells. *Cell Rep*, 14(4):966–977.

[Mann et al., 2018] Mann, M., Mehta, A., de Boer, C. G., Kowalczyk, M. S., Lee, K., Haldeman, P., Rogel, N., Knecht, A. R., Farouq, D., Regev, A., and Baltimore, D. (2018). Heterogeneous Responses of Hematopoietic Stem Cells to Inflammatory Stimuli Are Altered with Age. *Cell Rep*, 25(11):2992–3005 e5.

[McGhee and Ginder, 1979] McGhee, J. D. and Ginder, G. D. (1979). Specific DNA methylation sites in the vicinity of the chicken beta-globin genes. *Nature*, 280(5721):419–20.

[McLean et al., 2010] 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. *Nat Biotechnol*, 28(5):495–501.

[Meissner et al., 2008] Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., and Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, 454(7205):766–70.

[Michalak et al., 2019] Michalak, E. M., Burr, M. L., Bannister, A. J., and Dawson, M. A. (2019). The roles of DNA, RNA and histone methylation in ageing and cancer. *Nat Rev Mol Cell Biol*, 20(10):573–589.

[Miura et al., 2012] Miura, F., Enomoto, Y., Dairiki, R., and Ito, T. (2012). Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res*, 40(17):e136.

[Mohandas et al., 1981] Mohandas, T., Sparkes, R. S., and Shapiro, L. J. (1981). Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science*, 211(4480):393–6.

[Monk et al., 1987] Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development*, 99(3):371–82.

[Moran et al., 2016] Moran, S., Arribas, C., and Esteller, M. (2016). Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics*, 8(3):389–99.

[Moran-Crusio et al., 2011] Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M. E., Vasanthakumar, A., Patel, J., Zhao, X., Perna, F., Pandey, S., Madzo, J., Song, C., Dai, Q., He, C., Ibrahim, S., Beran, M., Zavadil, J., Nimer, S. D., Melnick, A., Godley, L. A., Aifantis, I., and Levine, R. L. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*, 20(1):11–24.

[Morita et al., 2010] Morita, Y., Ema, H., and Nakauchi, H. (2010). Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. *J Exp Med*, 207(6):1173–82.

[Morrison et al., 1996] Morrison, S. J., Wandycz, A. M., Akashi, K., Globerson, A., and Weissman, I. L. (1996). The aging of hematopoietic stem cells. *Nat Med*, 2(9):1011–6.

[Morrison et al., 1997] Morrison, S. J., Wandycz, A. M., Hemmati, H. D., Wright, D. E., and Weissman, I. L. (1997). Identification of a lineage of multipotent hematopoietic progenitors. *Development*, 124(10):1929–39.

[Morrison and Weissman, 1994] Morrison, S. J. and Weissman, I. L. (1994). The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity*, 1(8):661–73.

[Muller-Sieburg et al., 2004] Muller-Sieburg, C. E., Cho, R. H., Karlsson, L., Huang, J. F., and Sieburg, H. B. (2004). Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness. *Blood*, 103(11):4111–8.

[Muller-Sieburg et al., 2002] Muller-Sieburg, C. E., Cho, R. H., Thoman, M., Adkins, B., and Sieburg, H. B. (2002). Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. *Blood*, 100(4):1302–9.

[Mulqueen et al., 2018] Mulqueen, R. M., Pokholok, D., Norberg, S. J., Torkenczy, K. A., Fields, A. J., Sun, D., Sinnamon, J. R., Shendure, J., Trapnell, C., O'Roak, B. J., Xia, Z., Steemers, F. J., and Adey, A. C. (2018). Highly scalable generation of DNA methylation profiles in single cells. *Nat Biotechnol*, 36(5):428–431.

[Na Nakorn et al., 2002] Na Nakorn, T., Traver, D., Weissman, I. L., and Akashi, K. (2002). Myeloerythroid-restricted progenitors are sufficient to confer radioprotection and provide the majority of day 8 CFU-S. *J Clin Invest*, 109(12):1579–85.

[Naik et al., 2013] Naik, S. H., Perie, L., Swart, E., Gerlach, C., van Rooij, N., de Boer, R. J., and Schumacher, T. N. (2013). Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature*, 496(7444):229–32.

[Nestorowa et al., 2016] Nestorowa, S., Hamey, F. K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N. K., Kent, D. G., and Gottgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood*, 128(8):e20–31.

[Notta et al., 2016] Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O. I., Wilson, G., Kaufmann, K. B., McLeod, J., Laurenti, E., Dunant, C. F., McPherson, J. D., Stein, L. D., Dror, Y., and Dick, J. E. (2016). Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science*, 351(6269):aab2116.

[Oakes et al., 2014] Oakes, C. C., Claus, R., Gu, L., Assenov, Y., Hullein, J., Zucknick, M., Bieg, M., Brocks, D., Bogatyrova, O., Schmidt, C. R., Rassenti, L., Kipps, T. J., Mertens, D., Lichter, P., Dohner, H., Stilgenbauer, S., Byrd, J. C., Zenz, T., and Plass, C. (2014). Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer Discov*, 4(3):348–61.

[Oguro et al., 2013] Oguro, H., Ding, L., and Morrison, S. J. (2013). SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell*, 13(1):102–16.

[Okano et al., 1999] Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3):247–57.

[O'Keeffe et al., 2002] O'Keeffe, M., Hochrein, H., Vremec, D., Caminschi, I., Miller, J. L., Anders, E. M., Wu, L., Lahoud, M. H., Henri, S., Scott, B., Hertzog, P., Tatarczuch, L., and Shortman, K. (2002). Mouse plasmacytoid cells: long-lived cells, heterogeneous in surface phenotype and function, that differentiate into CD8(+) dendritic cells only after microbial stimulus. *J Exp Med*, 196(10):1307–19.

[Orkin and Zon, 2008] Orkin, S. H. and Zon, L. I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. *Cell*, 132(4):631–44.

[Pang et al., 2011] Pang, W. W., Price, E. A., Sahoo, D., Beerman, I., Maloney, W. J., Rossi, D. J., Schrier, S. L., and Weissman, I. L. (2011). Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc Natl Acad Sci U S A*, 108(50):20012–7.

[Paradis and Schliep, 2019] Paradis, E. and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3):526–528.

[Park and Wu, 2016] Park, Y. and Wu, H. (2016). Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics*, 32(10):1446–53.

[Paul et al., 2015] Paul, F., Arkin, Y., Giladi, A., Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., Winter, D., Lara-Astiaso, D., Gury, M., Weiner, A., David, E., Cohen, N., Lauridsen, F. K., Haas, S., Schlitzer, A., Mildner, A., Ginhoux, F., Jung, S., Trumpp, A., Porse, B. T., Tanay, A., and Amit, I. (2015). Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell*, 163(7):1663–77.

[Perie et al., 2015] Perie, L., Duffy, K. R., Kok, L., de Boer, R. J., and Schumacher, T. N. (2015). The Branching Point in Erythro-Myeloid Differentiation. *Cell*, 163(7):1655–62.

[Petkovich et al., 2017] Petkovich, D. A., Podolskiy, D. I., Lobanov, A. V., Lee, S. G., Miller, R. A., and Gladyshev, V. N. (2017). Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. *Cell Metab*, 25(4):954–960.e6.

[Pietras et al., 2015] Pietras, E. M., Reynaud, D., Kang, Y. A., Carlin, D., Calero-Nieto, F. J., Leavitt, A. D., Stuart, J. M., Gottgens, B., and Passegue, E. (2015). Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell*, 17(1):35–46.

[Qiu et al., 2014] Qiu, J., Papatsenko, D., Niu, X., Schaniel, C., and Moore, K. (2014). Divisional history and hematopoietic stem cell function during homeostasis. *Stem Cell Reports*, 2(4):473–90.

[Quinlan and Hall, 2010] Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–2.

[Rakyan et al., 2010] Rakyan, V. K., Down, T. A., Maslau, S., Andrew, T., Yang, T. P., Beyan, H., Whittaker, P., McCann, O. T., Finer, S., Valdes, A. M., Leslie, R. D., Deloukas, P., and Spector, T. D. (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res*, 20(4):434–9.

[Reisinger et al., 2017] Reisinger, E., Genthner, L., Kerssemakers, J., Kensche, P., Borufka, S., Jugold, A., Kling, A., Prinz, M., Scholz, I., Zipprich, G., Eils, R., Lawerenz, C., and Eils, J. (2017). OTP: An automatized system for managing and processing NGS data. *J Biotechnol*, 261:53–62.

[Reizis, 2019] Reizis, B. (2019). Plasmacytoid Dendritic Cells: Development, Regulation, and Function. *Immunity*, 50(1):37–50.

[Riggs, 1975] Riggs, A. D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet*, 14(1):9–25.

[Rodrigues et al., 2018] Rodrigues, P. F., Alberti-Servera, L., Eremin, A., Grajales-Reyes, G. E., Ivanek, R., and Tussiwand, R. (2018). Distinct progenitor lineages contribute to the heterogeneity of plasmacytoid dendritic cells. *Nat Immunol*, 19(7):711–722.

[Rodriguez-Fraticelli et al., 2018] Rodriguez-Fraticelli, A. E., Wolock, S. L., Weinreb, C. S., Panero, R., Patel, S. H., Jankovic, M., Sun, J., Calogero, R. A., Klein, A. M., and Camargo, F. D. (2018). Clonal analysis of lineage fate in native haematopoiesis. *Nature*, 553(7687):212–216.

[Rossi et al., 2005] Rossi, D. J., Bryder, D., Zahn, J. M., Ahlenius, H., Sonu, R., Wagers, A. J., and Weissman, I. L. (2005). Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proc Natl Acad Sci U S A*, 102(26):9194–9.

[Rossi et al., 2008] Rossi, D. J., Jamieson, C. H., and Weissman, I. L. (2008). Stems cells and the pathways to aging and cancer. *Cell*, 132(4):681–96.

[Rundberg Nilsson et al., 2016] Rundberg Nilsson, A., Soneji, S., Adolfsson, S., Bryder, D., and Pronk, C. J. (2016). Human and Murine Hematopoietic Stem Cell Aging Is Associated with Functional Impairments and Intrinsic Megakaryocytic/Erythroid Bias. *PLoS One*, 11(7):e0158369.

[Sanchez-Castillo et al., 2015] Sanchez-Castillo, M., Ruau, D., Wilkinson, A. C., Ng, F. S., Hannah, R., Diamanti, E., Lombard, P., Wilson, N. K., and Gottgens, B. (2015). CODEX: a next-generation sequencing experiment database for the haematopoietic and embryonic stem cell communities. *Nucleic Acids Res*, 43(Database issue):D1117–23.

[Sanford et al., 1987] Sanford, J. P., Clark, H. J., Chapman, V. M., and Rossant, J. (1987). Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. *Genes Dev*, 1(10):1039–46.

[Sanjuan-Pla et al., 2013] Sanjuan-Pla, A., Macaulay, I. C., Jensen, C. T., Woll, P. S., Luis, T. C., Mead, A., Moore, S., Carella, C., Matsuoka, S., Bouriez Jones, T., Chowdhury, O., Stenson, L., Lutteropp, M., Green, J. C., Facchini, R., Boukarabila, H., Grover, A., Gambardella, A., Thongjuea, S., Carrelha, J., Tarrant, P., Atkinson, D., Clark, S. A., Nerlov, C., and Jacobsen, S. E. (2013). Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature*, 502(7470):232–6.

[Sathe et al., 2013] Sathe, P., Vremec, D., Wu, L., Corcoran, L., and Shortman, K. (2013). Convergent differentiation: myeloid and lymphoid pathways to murine plasmacytoid dendritic cells. *Blood*, 121(1):11–9.

[Schatz, 2017] Schatz, M. C. (2017). Nanopore sequencing meets epigenetics. *Nat Methods*, 14(4):347–348.

[Seita and Weissman, 2010] Seita, J. and Weissman, I. L. (2010). Hematopoietic stem cell: self-renewal versus differentiation. *Wiley Interdiscip Rev Syst Biol Med*, 2(6):640–53.

[Shearstone et al., 2011] 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(6057):799–802.

[Shema et al., 2019] Shema, E., Bernstein, B. E., and Buenrostro, J. D. (2019). Single-cell and single-molecule epigenomics to uncover genome regulation at unprecedented resolution. *Nat Genet*, 51(1):19–25.

[Shigematsu et al., 2004] Shigematsu, H., Reizis, B., Iwasaki, H., Mizuno, S., Hu, D., Traver, D., Leder, P., Sakaguchi, N., and Akashi, K. (2004). Plasmacytoid dendritic cells activate lymphoid-specific genetic programs irrespective of their cellular origin. *Immunity*, 21(1):43–53.

[Smallwood et al., 2014] Smallwood, S. A., Lee, H. J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S. R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods*, 11(8):817–820.

[Smith and Meissner, 2013] Smith, Z. D. and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat Rev Genet*, 14(3):204–20.

[Stadler et al., 2011] Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., and Schubeler, D. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*, 480(7378):490–5.

[Steensma et al., 2015] Steensma, D. P., Bejar, R., Jaiswal, S., Lindsley, R. C., Sekeres, M. A., Hasserjian, R. P., and Ebert, B. L. (2015). Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*, 126(1):9–16.

[Stelzer et al., 2015] Stelzer, Y., Shivalila, C. S., Soldner, F., Markoulaki, S., and Jaenisch, R. (2015). Tracing dynamic changes of DNA methylation at single-cell resolution. *Cell*, 163(1):218–29.

[Stepper et al., 2017] Stepper, P., Kungulovski, G., Jurkowska, R. Z., Chandra, T., Krueger, F., Reinhardt, R., Reik, W., Jeltsch, A., and Jurkowski, T. P. (2017). Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. *Nucleic Acids Res*, 45(4):1703–1713.

[Stricker et al., 2017] Stricker, S. H., Koferle, A., and Beck, S. (2017). From profiles to function in epigenomics. *Nat Rev Genet*, 18(1):51–66.

[Stuart et al., 2019] Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., r., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902 e21.

[Stubbs et al., 2017] Stubbs, T. M., Bonder, M. J., Stark, A. K., Krueger, F., Team, B. I. A. C., von Meyenn, F., Stegle, O., and Reik, W. (2017). Multi-tissue DNA methylation age predictor in mouse. *Genome Biol*, 18(1):68.

[Sudo et al., 2000] Sudo, K., Ema, H., Morita, Y., and Nakauchi, H. (2000). Age-associated characteristics of murine hematopoietic stem cells. *J Exp Med*, 192(9):1273–80.

[Sun et al., 2014] Sun, D., Luo, M., Jeong, M., Rodriguez, B., Xia, Z., Hannah, R., Wang, H., Le, T., Faull, K., Chen, R., Gu, H., Bock, C., Meissner, A., Göttgens, B., Darlington, G., Li, W., and Goodell, M. (2014). Epigenomic Profiling of Young and Aged HSCs Reveals Concerted Changes during Aging that Reinforce Self-Renewal. *Cell Stem Cell*, 14(5):673–688.

[Taiwo et al., 2013] Taiwo, O., Wilson, G. A., Emmett, W., Morris, T., Bonnet, D., Schuster, E., Adejumo, T., Beck, S., and Pearce, D. J. (2013). DNA methylation analysis of murine hematopoietic side population cells during aging. *Epigenetics*, 8(10):1114–22.

[Takizawa et al., 2012] Takizawa, H., Boettcher, S., and Manz, M. G. (2012). Demand-adapted regulation of early hematopoiesis in infection and inflammation. *Blood*, 119(13):2991–3002.

[Takizawa et al., 2011] Takizawa, H., Regoes, R. R., Boddupalli, C. S., Bonhoeffer, S., and Manz, M. G. (2011). Dynamic variation in cycling of hematopoietic stem cells in steady state and inflammation. *J Exp Med*, 208(2):273–84.

[Tejedor et al., 2018] Tejedor, J. R., Bueno, C., Cobo, I., Bayon, G. F., Prieto, C., Mangas, C., Perez, R. F., Santamarina, P., Urdinguio, R. G., Menendez, P., Fraga, M. F., and Fernandez, A. F. (2018). Epigenome-wide analysis reveals specific DNA hypermethylation of T cells during human hematopoietic differentiation. *Epigenomics*, 10(7):903–923.

[Teschendorff et al., 2010] Teschendorff, A. E., Menon, U., Gentry-Maharaj, A., Ramus, S. J., Weisenberger, D. J., Shen, H., Campan, M., Noushmehr, H., Bell, C. G., Maxwell, A. P., Savage, D. A., Mueller-Holzner, E., Marth, C., Kocjan, G., Gayther, S. A., Jones, A., Beck, S., Wagner, W., Laird, P. W., Jacobs, I. J., and Widschwendter, M. (2010). Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res*, 20(4):440–6.

[Unnikrishnan et al., 2018] Unnikrishnan, A., Hadad, N., Masser, D. R., Jackson, J., Freeman, W. M., and Richardson, A. (2018). Revisiting the genomic hypomethylation hypothesis of aging. *Ann N Y Acad Sci*, 1418(1):69–79.

[Velten et al., 2017] Velten, L., Haas, S. F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B. P., Hirche, C., Lutz, C., Buss, E. C., Nowak, D., Boch, T., Hofmann, W. K., Ho, A. D., Huber, W., Trumpp, A., Essers, M. A., and Steinmetz, L. M. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*, 19(4):271–281.

[Waddington, 1942] Waddington, C. H. (1942). The Epigenotype. *Endeavor*, 1:18–20.

[Waddington, 1957] Waddington, C. H. (1957). The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology. *George Allen & Unwin*.

[Wagers et al., 2002] Wagers, A. J., Allsopp, R. C., and Weissman, I. L. (2002). Changes in integrin expression are associated with altered homing properties of Lin(-/lo)Thy1.1(lo)Sca-1(+)c-kit(+) hematopoietic stem cells following mobilization by cyclophosphamide/granulocyte colony-stimulating factor. *Exp Hematol*, 30(2):176–85.

[Wahlestedt et al., 2017] Wahlestedt, M., Erlandsson, E., Kristiansen, T., Lu, R., Brakebusch, C., Weissman, I. L., Yuan, J., Martin-Gonzalez, J., and Bryder, D. (2017). Clonal reversal of ageing-associated stem cell lineage bias via a pluripotent intermediate. *Nat Commun*, 8:14533.

[Walter et al., 2015] Walter, D., Lier, A., Geiselhart, A., Thalheimer, F. B., Huntscha, S., Sobotta, M. C., Moehrle, B., Brocks, D., Bayindir, I., Kaschutnig, P., Muedder, K., Klein, C., Jauch, A., Schroeder, T., Geiger, H., Dick, T. P., Holland-Letz, T., Schmezer, P., Lane, S. W., Rieger, M. A., Essers, M. A., Williams, D. A., Trumpp, A., and Milsom, M. D. (2015). Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature*, 520(7548):549–52.

[Walter et al., 2011] Walter, M. J., Ding, L., Shen, D., Shao, J., Grillot, M., McLellan, M., Fulton, R., Schmidt, H., Kalicki-Veizer, J., O'Laughlin, M., Kandoth, C., Baty, J., Westervelt, P., DiPersio, J. F., Mardis, E. R., Wilson, R. K., Ley, T. J., and Graubert, T. A. (2011). Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia*, 25(7):1153–8.

[Wang et al., 2013] Wang, Q., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., Bahr, M., Wolf, S., Shendure, J., Eils, R., Plass, C., and Weichenhan, D. (2013). Tagmentation-based whole-genome bisulfite sequencing. *Nat Protoc*, 8(10):2022–32.

[Wang et al., 2017] Wang, T., Tsui, B., Kreisberg, J. F., Robertson, N. A., Gross, A. M., Yu, M. K., Carter, H., Brown-Borg, H. M., Adams, P. D., and Ideker, T. (2017). Epigenetic aging signatures in mice livers are slowed by dwarfism, calorie restriction and rapamycin treatment. *Genome Biol*, 18(1):57.

[Weidner et al., 2014] Weidner, C. I., Lin, Q., Koch, C. M., Eisele, L., Beier, F., Ziegler, P., Bauerschlag, D. O., Jockel, K. H., Erbel, R., Muhleisen, T. W., Zenke, M., Brummendorf, T. H., and Wagner, W. (2014). Aging of blood can be tracked by DNA methylation changes at just three CpG sites. *Genome Biol*, 15(2):R24.

[Wickham, 2016] Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

[Wickham et al., 2019] Wickham, H., François, R., Henry, L., and Müller, K. (2019). *dplyr: A Grammar of Data Manipulation*. R package version 0.8.0.1.

[Williams et al., 2008] Williams, D. A., Zheng, Y., and Cancelas, J. A. (2008). Rho GTPases and regulation of hematopoietic stem cell localization. *Methods Enzymol*, 439:365–93.

[Wilson et al., 2008] Wilson, A., Laurenti, E., Oser, G., van der Wath, R. C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C. F., Eshkind, L., Bockamp, E., Lio, P., Macdonald, H. R., and Trumpp, A. (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*, 135(6):1118–29.

[Wilson and Jones, 1983] Wilson, V. L. and Jones, P. A. (1983). DNA methylation decreases in aging but not in immortal cells. *Science*, 220(4601):1055–7.

[Wu and Morris, 2001] Wu, C. and Morris, J. R. (2001). Genes, genetics, and epigenetics: a correspondence. *Science*, 293(5532):1103–1105.

[Wu and Zhang, 2017] Wu, X. and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet*, 18(9):517–534.

[Xie et al., 2014] Xie, M., Lu, C., Wang, J., McLellan, M. D., Johnson, K. J., Wendl, M. C., McMichael, J. F., Schmidt, H. K., Yellapantula, V., Miller, C. A., Ozenberger, B. A., Welch, J. S., Link, D. C., Walter, M. J., Mardis, E. R., Dipersio, J. F., Chen, F., Wilson, R. K., Ley, T. J., and Ding, L. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med*, 20(12):1472–8.

[Yamamoto et al., 2013] Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K. L., Ema, H., and Nakauchi, H. (2013). Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell*, 154(5):1112–1126.

[Yamamoto et al., 2018] Yamamoto, R., Wilkinson, A. C., Ooehara, J., Lan, X., Lai, C. Y., Nakauchi, Y., Pritchard, J. K., and Nakauchi, H. (2018). Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment. *Cell Stem Cell*, 22(4):600–607 e4.

[Yang et al., 2017] Yang, J., Tanaka, Y., Seay, M., Li, Z., Jin, J., Garmire, L. X., Zhu, X., Taylor, A., Li, W., Euskirchen, G., Halene, S., Kluger, Y., Snyder, M. P., Park, I. H., Pan, X., and Weissman, S. M. (2017). Single cell transcriptomics reveals unanticipated features of early hematopoietic precursors. *Nucleic Acids Res*, 45(3):1281–1296.

[Yang et al., 2005] Yang, L., Bryder, D., Adolfsson, J., Nygren, J., Mansson, R., Sigvardsson, M., and Jacobsen, S. E. (2005). Identification of Lin(-)Sca1(+)kit(+)CD34(+)Flt3- short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. *Blood*, 105(7):2717–23.

[Yu et al., 2017] Yu, V. W. C., Yusuf, R. Z., Oki, T., Wu, J., Saez, B., Wang, X., Cook, C., Baryawno, N., Ziller, M. J., Lee, E., Gu, H., Meissner, A., Lin, C. P., Kharchenko, P. V., and Scadden, D. T. (2017). Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells. *Cell*, 168(5):944–945.

[Zerbino et al., 2015] Zerbino, D. R., Wilder, S. P., Johnson, N., Juettemann, T., and Flicek, P. R. (2015). The ensembl regulatory build. *Genome Biol*, 16:56.

[Zheng and Xie, 2019] Zheng, H. and Xie, W. (2019). The role of 3D genome organization in development and cell differentiation. *Nat Rev Mol Cell Biol*, 20(9):535–550.

[Zhou et al., 2011] Zhou, V. W., Goren, A., and Bernstein, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet*, 12(1):7–18.

[Zhu, 2013] Zhu, L. J. (2013). Integrative analysis of chip-chip and chip-seq dataset. *Methods Mol Biol*, 1067:105–24.

[Zhu et al., 2010] Zhu, L. J., Gazin, C., Lawson, N. D., Pages, H., Lin, S. M., Lapointe, D. S., and Green, M. R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, 11:237.

[Ziller et al., 2013] Ziller, M. J., Gu, H., Muller, F., Donaghey, J., Tsai, L. T., Kohlbacher, O., De Jager, P. L., Rosen, E. D., Bennett, D. A., Bernstein, B. E., Gnirke, A., and Meissner, A. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature*, 500(7463):477–81.

List of Figures

List of Tables

Abbreviations

Table A.3: List of abbreviations.

	Abbreviation	Definition
A	A	Adenosine
	ACK	Ammonium-chloride-potassium
	aDMR	Aging DMR
	AML	Acute myeloid leukemia
	APC	Allophycocyanin
	APC-Cy7	APC-cyanine 7
	ARCH	Age-related clonal hematopoiesis
	ATAC-seq	Assay for transposase-accessible chromatin using sequencing
B	B	B lymphoid cell
	bp	Base pair
C	C	Cytosine
	Cas	CRISPR-associated protein
	Cas9	Native Cas9 nuclease
	CD	Cluster of differentiation
	cDC	Conventional dendritic cell
	cDNA	Complementary DNA
	CDS	Coding sequence
	CDP	Common dendritic cell progenitor
	CFU-E	Colony forming uni-erythroid
	CGI	CpG island
	CH	Non-CpG methylation
	ChIP-seq	Chromatin immunoprecipitation sequencing
	chr	Chromosome
	CLP	Common lymphoid progenitor
	cMoP	Common monocyte progenitor
	CMP	Common myeloid progenitor
	CpG	Cytosine phosphate guanine
	CRISPR	Clustered regularly interspaced short palindromic repeats
	CTCF	CCCTC-binding factor
D	DC	Dendritic cell
	DCRD	Distant *cis*-regulartory domain
	dCas9	Deactivated Cas9
	DKFZ	German Cancer Research Center
	DML	Differentially methylated loci
	DMR	Differentially methylated region
	DNA	Deoxyribonucleic acid
	DNMT	DNA methyltransferase
	dNTP	Deoxynucleoside triphosphate
	dsDNA	double-stranded DNA
E	EDTA	Ethylenediaminetetraacetic acid
	eGFP	Enhanced green fluorescent protein
	ehDMR	Early hematopoietic DMR
	Eosino	Eosinophil
	Ery	Erythrocyte
	et al.	Et alii
F	FACS	Fluorescence-activated cell sorting
	FCS	Fetal calf serum
	FH	Fachhochschule
	FITC	Fluorescein isothiocyanate
	FSC	Forward Scatter
G	g	Gravity
	G	Guanosine
	GEM	Gel Bead-In Emulsion
	GMP	Granulocyte-monocyte progenitor
	GO	Gene Ontology
	Granu	Granulocyte
	GREAT	Genomic Regions Enrichment of Annotation Tool
	GTP	Nucleotide guanosine triphosphate
H	h	Hour
	H_2O	Water
	H2B	Histone 2B

	H3K27ac	Histone H3 lysine 27 acetylation
	H3K4me1	Histone H3 lysine 4 monomethylation
	H3K4me3	Histone H3 lysine 4 trimethylation
	hDMR	Hematopoietic DMR
	Hi-C	Chromosome conformation capture
	HIV	Human immunodeficiency virus
	HOMER	Hypergeometric Optimization of Motif EnRichment
	HSC	Hematopoietic stem cell
	HSPC	Hematopoietic stem and progenitor cell
I	iDMR	Inflammation DMR
	IMDM	Iscove's modified Dulbecco's medium
	i.p.	Intraperitoneal
	IRF	Interferon regulatory factor
	IVC	Individually ventilated cages
K	kb	Kilobase
	kg	Kilogram
L	LDMNC	Low density mononuclear cells
	LK	Lineage$^-$ Sca1$^-$ ckit$^+$ cells
	LMPP	Lymphoid-primed MPP
	LRC	Label retaining cell
	LSK	Lineage$^-$ Sca1$^+$ ckit$^+$ cells
	LT-HSC	Long-term HSC
M	M	Molar
	MDP	Monocyte-dendritic cell progenitor
	MDS	Myelodysplastic syndrome
	MeDIP	Methylated DNA immunoprecipitation
	Meg a	Megakaryocyte
	MEP	Megakaryocyte–erythroid progenitor
	mg	Milligram
	min	Minute
	mL	Milliliter
	mm	Mus musculus
	Mono	Monocyte
	MPP	Multipotent progenitor
	MRE	Methylation-sensitive restriction enzyme
	MSigDB	Molecular Signatures Database
N	ncRNA	Non-coding RNA
	Neutro	Neutrophil
	ng	Nanogram
	NGS	Next-generation sequencing
	NK cell	Natural killer cell
	NMF	Non-negative matrix factorization
O	ODCF	Omics IT and Data Management Core Facility
	OTP	One Touch Pipeline
P	PBAT	Post-bisulfite adapter tagging
	PBS	Phosphate buffered saline
	PC	Principal Component
	PCA	Principal Component Analysis
	PCR	Polymerase chain reaction
	pDC	Plasmacytoid dendritic cell
	PE	Phycoerythrin
	PE-Cy5	PE-cyanine 5
	PE-Cy7	PE-cyanine 7
	PerCP	Peridinin-chlorophyll protein
	PerCP-Cy5.5	PerCP-cyanine 5.5
	pI:pC	Polyinosinic-polycytidylic acid
	preGM	Pre–granulocyte-macrophage progenitor
	preMegE	Pre-megakaryocyte-erythroid progenitor
R	RefSeq	Reference Sequence
	RNA	Ribonucleic acid
	ROS	Reactive oxygen species
	RRBS	Reduced representation bisulfite sequencing
S	s	Second
	SAM	S-Adenosyl methionine
	SC	Standard deviation
	scOpenLab	Single-cell open lab

	Sca1	Stem cells antigen 1
	scRNA-seq	Single-cell RNA sequencing
	Seq	Sequencing
	SPF	Specific pathogen free
	SSC	Side scatter
	ST-HSC	Short-term HSC
T	T	Thymidine
	T	T lymphoid cell
	TAD	Topologically associating domain
	TET	Ten-eleven translocation methylcytosine dioxygenase
	TF	Transcription factor
	TSS	Transcription start site
	TWGBS	Tagmentation-based whole genome bisulfite sequencing
U	U	Unit
	UCSC	UCSC University of California, Santa Cruz
	UMAP	Uniform manifold approximation and projection
	UMI	Unique molecular identifier
	UTR	Untranslated region
	Vwf	Von Willebrand factor
V	vol	Volume
W	WBC	White blood counts
	WGBS	Whole genome bisulfite sequencing
	%	Percentage
	°C	Degree Celsius
	3D	Three dimensional
	5-caC	5-Carboxylcytosine
	5-fC	5-Formylcytosine
	5-FU	5-Fluorouracil
	5-hmC	5-Hydroxymethylcytosine
	5-mC	5-Methylcytosine
	7AAD	7-Aminoactinomycin
	µg	Microgram
	µL	Microliter
	µm	Micrometer
	µM	Micromolar

Conference Talks and Poster Presentations

Conference talks

DNA methylation dynamics identify lineage-specific regulation patterns of hematopoietic differentiation.
Sina Stäble, Stephen Krämer, Jens Langstein, Ruzhica Bogeska, Mark Hartmann, Maximilian Schönung, Natasha Anstee, Simon Haas, Dieter Weichenhan, Melinda Czeh, Julius Gräsel, Daniel Hübschmann, Frank Rosenbauer, Christoph Plass, Matthias Schlesner, Michael Milsom, Daniel Lipka. *24^{th} European Hematology Association Congress*, Amsterdam, 07/2019.

DNA methylation changes define hematopoietic lineage commitment.
Sina Staeble, Stephen Kraemer, Jens Langstein, Ruzhica Bogeska, Dieter Weichenhan, Julia Knoch, Oliver Muecke, Monika Helf, Melinda Czeh, Julius Graesel, Daniel Huebschmann, Frank Rosenbauer, Roland Eils, Christoph Plass, Matthias Schlesner, Michael D. Milsom, Daniel B. Lipka. *DKFZ-Weizmann Cancer-TRAX Symposium*, Heidelberg, 02/2019.

DNA methylation patterns define hematopoietic differentiation.
Sina Staeble, Stephen Kraemer, Jens Langstein, Ruzhica Bogeska, Dieter Weichenhan, Julia Knoch, Oliver Mücke, Monika Helf, Melinda Czeh, Julius Gräsel, Frank Rosenbauer, Christoph Plass, Matthias Schlesner, Michael D. Milsom, Daniel B. Lipka. *10^{th} International Heinrich F. C. Behr-Symposium*, Heidelberg, 09/2018. (Lightning talk with poster presentation)

Unidirectional and progressive DNA methylation changes define hematopoietic lineage commitment.
Sina Staeble, Stephen Kraemer, Jens Langstein, Ruzhica Bogeska, Dieter Weichenhan, Julia Knoch, Oliver Mücke, Monika Helf, Melinda Czeh, Julius Gräsel, Frank Rosenbauer, Christoph Plass, Matthias Schlesner, Michael D. Milsom, Daniel B. Lipka. *6^{th} Annual German Stem Cell Network Conference*, Heidelberg, 09/2018.

Poster presentations

Deconvolution of Hematopoietic Commitment Decisions by Genome-wide Analysis of Progressive DNA Methylation Changes.
Sina Staeble, Stephen Kraemer, Jens Langstein, Ruzhica Bogeska, Mark Hartmann, Maximilian Schoenung, Melinda Czeh, Natasha Anstee, Simon Haas, Abdelrahman Mahmoud, Julius Graesel, Daniel Huebschmann, Lars Feuerbach, Weichenhan Dieter, Benedikt Brors, Karsten Rippe, Jan-Philipp Mallm, Frank Rosenbauer, Christoph Plass, Matthias Schlesner, Michael D. Milsom, Daniel B. Lipka. *61^{th} American Society of Hematolog Annual Meeting*, Orland, 12/2019.

DNA methylation patterns define hematopoietic differentiation.
Sina Staeble, Stephen Kraemer, Jens Langstein, Ruzhica Bogeska, Dieter Weichenhan, Julia Knoch, Oliver Mücke, Monika Helf, Melinda Czeh, Julius Gräsel, Frank Rosenbauer, Christoph Plass, Matthias Schlesner, Michael D. Milsom, Daniel B. Lipka. *10^{th} International Heinrich F. C. Behr-Symposium*, Heidelberg, 09/2018.

DNA methylation in normal and aged hematopoiesis.
Sina Staeble, Stephen Kraemer, Ruzhica Bogeska, Jens Langstein, Julius Graesel, Julia Knoch, Christoph Plass, Matthias Schlesner, Michael Milsom, Daniel Lipka. *PhD poster presentation*, Heidelberg, 11/2017.

Manuscripts in preparation

Deconvolution of Hematopoietic Commitment Decisions By Genome-Wide Analysis of Progressive DNA Methylation Changes.
Sina Stäble, Stephen Kraemer, Jens Langstein, Ruzhica Bogeska, Mark Hartmann, Maximilian Schoenung, Melinda Czeh, Julia Knoch, Natasha Anstee, Simon Haas, Abdelrahman Mahmoud, Julius Graesel, Daniel Huebschmann, Lars Feuerbach, Weichenhan Dieter, Benedikt Brors, Karsten Rippe, Jan-Philipp Mallm, Frank Rosenbauer, Christoph Plass, Matthias Schlesner, Michael D. Milsom, Daniel B. Lipka. (manuscript in preparation)

Genome-wide methylation dynamics drive dendritic cell development from multipotent precursors.
Melinda Czeh, Sina Stäble, Lena Tepe, Sweta Talyan, Joana Carrelha, Yiran Meng, Bar-

bara Heitplatz, Sten Eirik W. Jacobsen, Claus Nerlov, Miguel A Andrade-Navarro, Daniel Lipka, Frank Rosenbauer. (manuscript in preparation)

Hypomethylating drugs specifically target *DNMT3A* mutation associated clonal hematopoiesis and acute myeloid leukemia.
Marina Scheller-Wendorff, Anne Kathrin Ludwig, Christian Rohde, Sina Staeble, Stephen Kraemer, Nicole Bäumer, Christian Arnold, Judith Zaugg, Daniel B. Lipka, Carsten Müller-Tidow. (manuscript in preparation)

No Evidence for Hematopoietic Stem Cell Self-Renewal *in-Vivo* Following Inflammatory Challenge.
Ruzhica Bogeska, Paul Kaschutnig, Stella V Paffenholz, Julia Knoch, Jan-Philipp Mallm, Malak Fawaz, Florian Buettner, Dagmar Walter, Felix Frauhammer, Tim Holland-Letz, Noboru Asada, Ana-Matea Mikecin Drazic, Marleen Buechler-Schaff, Martha Correno-Gonzalez, Melanie Ball, Sina Staeble, Julius Gräsel, Simon Haas, Daniel B. Lipka, Karsten Rippe, Benedikt Brors, Paul S. Frenette, Michael A. Rieger, Marieke A. G. Essers, Michael D. Milsom. (manuscript in preparation)

Contributions

The work presented in this dissertation would not have been possible without contributions from the following people:

This PhD thesis was jointly supervised by **PD Dr. Daniel Lipka** (Cancer Epigenomics, DKFZ) and **Dr. Michael Milsom** (Experimental Hematology, DKFZ and Hi-STEM).

This PhD thesis is part of a collaboration project between the research groups of **PD Dr. Daniel Lipka**, **Dr. Michael Milsom**, and **Dr. Matthias Schlesner** (Bioinformatics and Omics Data Analytics, DKFZ).

Stephen Krämer (Bioinformatics and Omics Data Analytics, DKFZ) developed pipelines for DNA methylation calling, genomic region annotations, and enrichment analyses. He identified the hDMRs, performed the hierarchical clustering of hDMRs, and generated DMR-centric line plots. Furthermore, he provided advice for additional bioinformatic analyses.

Dr. Ruzhica Bogeska (Experimental Hematology, DKFZ) sorted the first set of hematopoietic cell populations for this project. **Dr. Meldina Czeh** (Rosenbauer lab, University of Münster) sorted the dendritic cell populations.

Hematopoietic cell layers for scRNA-seq were jointly isolated with **Dr. Mark Hartmann** (Cancer Epigenomics, DKFZ) and **Jens Langstein** (Cancer Epigenomics, DKFZ). **Dr. Mark Hartmann** and **Katharina Bauer** (scOpenLab, DKFZ) generated scRNA-seq libraries. **Abdelrahman Mahmoud** (Applied Bioinformatics, DKFZ) performed the pre-processing of the scRNA-seq data. **Dr. Simon Haas** (Stem Cells at the Interface of Hematopoiesis, Immunity and Cancer, Hi-STEM), **Dr. Mark Hartmann**, and **Jens Langstein** annotated the single-cell RNA clusters.

ScBS-seq libraries were jointly generated with **Dr. Mark Hartmann** and **Maximilian Schönung** (Cancer Epigenomics, DKFZ). **Anand Mayakonda** (Cancer Epigenomics, DKFZ) performed the read alignment and methylation calling.

Technical support, including bone marrow isolations, FACS-sorting, injection of mice, and TWGBS library preparations, was provided by **Julia Knoch** (Experimental Hematology, DKFZ and Hi-STEM), **Oliver Mücke** (Cancer Epigenomics, DKFZ), **Dr. Julius**

Gräsel (Experimental Hematology, DKFZ), **Jens Langstein**, **Dr. Natasha Anstee** (Experimental Hematology, DKFZ), **Megan Druce** (Experimental Hematology, DKFZ), **Marleen Büchler-Schäff** (Experimental Hematology, DKFZ), **Dr. Ruzhica Bogeska**, **Dr. Paul Kaschutnig** (Experimental Hematology, DKFZ), and **Jeyan Jayarayan** (Experimental Hematology, DKFZ).

The **Imaging and Cytometry Core Facility** (DKFZ) supported several FACS sorting experiments.

The **Genomics and Proteomics Core Facility** (DKFZ) performed sequencing of TWGBS libraries, scRNA-seq libraries, and scBS-seq libraries.

Prof. Hartmut Geiger (University of Ulm) provided aged C57BL/6J mice.

Acknowledgments

This PhD thesis was only possible with the support of many people. In particular, I want to thank:

My supervisors, **PD Dr. Daniel Lipka** and **Dr. Michael Milsom**, for the opportunity to pursue my doctoral studies in two laboratories in parallel. I am sincerely grateful for the great co-supervision, the scientific input from two research environments, and the opportunity to present my work at international conferences. Thank you both for always believing in me.

PD Dr. Daniel Lipka for your endless support in every situation during my doctoral studies if it was scientific or not at all. Thank you for the open doors throughout the years, for many fruitful discussions, and especially for all your input and guidance to the 'Hierarchy' project. During the last four years, I have learned a lot from you, especially with regards to epigenetics, data analysis, and single-cell sequencing techniques. I am very proud to have been the first member of your group, which developed towards a wonderful team of scientists.

Dr. Michael Milsom for your support and scientific guidance throughout my doctoral studies. Without your smart ideas and scientific expertise, my work would not have been successful. I highly appreciate your advice and encouragement to focus on the 'Hierarchy' project. During my doctoral studies, I learned a lot from you that will make my scientific and personal life a lot easier in the future.

Prof. Dr. Christoph Plass for your wise advice at the right time and for motivating me to apply for the Cancer-TRAX PhD-to-Postdoc transition fellowship. Furthermore, I want to thank you for all sequencing that would not have been possible without you.

Prof. Dr. Jan Lohmann, **PD Dr. Odilia Popanda**, **Dr. Michael Milsom**, and **Prof. Dr. Peter Angel** for being members of my defense committee.

PD Dr. Daniel Lipka, **Dr. Mark Hartmann**, and **Dr. Natasha Anstee** for taking the time to proofread my dissertation.

Stephen Krämer for collaborating on the 'Hierarchy' project in the last few years, for many night shifts before poster presentations and conference talks, for private bioinformatic

lessons, but especially for being a very helpful and pleasant colleague. Many thanks for all your bright ideas and your bioinformatic contributions that moved our project forward.

PD Dr. Dieter Weichenhan and **Marion Bähr** for their technical advice and tricks for the preparations of TWGBS libraries.

Clarissa Feuerstein for sharing a desk with me for the last four years, discussing R issues, joining my addiction to chewing gums, and answering the phone when special people called.

The present and past members of **the Lipka and the Milsom group** for the support and help over the years, many constructive discussions, and the great atmosphere in the lab: Natasha Anstee, Melanie Ball, Milena Block, Ruzhica Bogeska, Marleen Büchler-Schäff, Megan Druce, Julius Gräsel, Mariam Hakobyan, Mark Hartmann, Jeyan Jayarajan, Paul Kaschutnig, Julia Knoch, Jens Langstein, Ana-Matea Mikecin, Oliver Mücke, Esther Rodriguez, and Maximilian Schönung.

Julia Knoch for the technical support during mouse preps, sorts, library preps, genotyping, mouse breedings, and many more. But more importantly, thank you for your positive nature and for motivating me in every situation. Without your help and organizational talent I would have never been able to finish my doctoral studies.

Oliver Mücke for hundreds of DNA isolations, library preparations, and genotyping. Thanks for the awesome rhythms next door (not disturbing at all). Thank you for your humorous character, which made me laugh in every situation during the last four years.

Dr. Julius Gräsel for your support, especially at the beginning of our doctoral studies. We have started this journey together, learned a lot from each other and became really good friends over the years. I don't know how I would have managed some phases of my doctoral studies without you.

Dr. Natasha (Tash) Anstee for taking over the Dnmt3a project and all the support in the lab. But more importantly, thank you so much for your friendship. We experienced so many good times together over the last years and Heidelberg will change a lot without you.

Dr. Mark Hartmann for all your support and help in the lab. I am incredibly grateful that you joined the Lipka group in the second year of my PhD. Thank you for so many helpful advice, if it was work-related or not all. You have always encouraged me to be

proud of my own achievements.

Jens Langstein for all your help during early morning mouse preps and late night sorts. Thank you for countless coffee breaks and especially for being the most relaxed scientist on earth. I wish I could be like you in some situations.

Dr. Ruzhica (Ruzhi) Bogeska for starting the 'Hierarchy' project and answering all my questions at the beginning of my PhD. You taught me the most important method for my PhD project - how to prep a mouse.

Megan Druce for your friendship, your good mood in lab, and so many great moments outside the lab. It is always so much fun to have you around.

Maximilian (Maxi) Schönung for your help in the lab and for discussing many R programming-related issues. I always enjoyed our conversations about random stuff during coffee breaks. You are one of the smartest and most enthusiastic PhD student I know.

My Bachelor and Masters students for the great work during their internships and Bachelor/Master theses: **Primin Killinger** (Intern and Bachelor student), **Can Sönmezer** (Intern), **Jeyan Jayarajan** (Intern and Master student), **Julia Messmer** (Intern), and **Nil Üsün** (Master student). By supervising you, I have learned a lot for my future career. Thank you!

All present and former **members of the division "Cancer Epigenomics" and of Hi-STEM** for the supportive atmosphere in lab, the helpful input for my project at department meetings, and many enjoyable hours during social events. Special thanks to **Carsten Bahr**, **Lisa Becker**, **David Brocks**, **Kristin Decker**, **Alexander Kühn**, **Reinhard Liebers**, **Simon Haas**, **Joschka Hey**, **Chris Hirche**, **Elisa Noll**, **Pavlo Lutsik**, **Simin Öz**, **Manuel Reitberger**, **Simon Renders**, **Marc Their**, and **Paula Werner**. I also want to thank **Prof. Dr. Christoph Plass** and **Prof. Dr. Andreas Trumpp** for establishing the great research environment in both divisions.

Stewart (Mac) Mein, the unofficial Hi-STEM member, for your friendship and so many nice events that we experienced together.

The **Cytometry Core Facility** including **Dr. Steffen Schmitt**, **Klaus Hexel**, **Dr. Marcus Eich**, **Tobias Rubner**, and **Florian Blum** for their great technical support

during innumerable sort appointments. I always enjoyed the time at the 7^{th} floor.

The **Genomics and Proteomics Core Facility** for sequencing my libraries and the **Animal Core Facility** for taking care of mice.

The **Cancer-TRAX program** for funding my position and the opportunity to do research at the Weizmann Institute in Israel after my PhD defense.

Danksagung

Ein besonderer Dank geht an all die Menschen, die mich außerhalb des Deutschen Krebsforschungszentrums auf unterschiedlichste Weise unterstützt haben. Insbesondere danke ich:

Meinem Bruder Tim Zimmermann und seiner wundervollen Familie, für viele schöne Momente, die mir gezeigt haben, dass auch andere Dinge im Leben wichtig sind. Danke für euer Verständnis, dass ich in den letzten Jahren viel zu wenig Zeit für euch hatte.

Meinen Schwiegereltern Beate und Rolf Stäble, für eure Unterstützung und Bewunderung für meinen Beruf.

Meinem Chor proVocal für die gemeinsame Musik, die mir jeden Mittwoch den perfekten Ausgleich gegeben hat. Aber vor allem für die vielen Gänsehautmomente in Konzerten, die musikalische Bildung und die tolle Gemeinschaft, die mein Leben unendlich bereichert.

Meinen Freunden Sabrina, **Nico**, **Laura**, **Martha**, **Anna-Lena**, **Marc**, **David** und **Basti** für eure Unterstützung in einer für mich sehr schwierigen Zeit.
Ein besonderer Dank geht an **meine besten Freundinnen Sabrina**, **Laura** und **Martha**, dass ihr immer für mich da gewesen seid und ein offenes Ohr hattet, wenn ich von meinen Problemen erzählt habe. Danke, dass es euch gibt.

Meinem Mann Markus Stäble, der die letzten Jahre oft auf mich verzichten musste. Danke, für deine Unterstützung in den letzten Monaten, die Motivation zum Sport und viele unvergessliche Reisen, die mir immer wieder neue Kraft gegeben haben. Aber vor allem danke ich Dir dafür, dass du noch an unsere Liebe geglaubt hast, als ich sie schon längst aufgegeben hatte.

Meinen Eltern Carmen und Uwe Zimmermann für ihre Liebe, ihre Unterstützung in jeglichen Lebenslagen und dafür, dass ihr immer an mich glaubt. Danke, dass ihr mich zu der Person gemacht habt, die ich heute bin. Ohne euren Rückhalt wäre ich niemals so weit gekommen.