

WESTFÄLISCHE WILHELMS-UNIVERSITÄT
INSTITUT FÜR MOLEKULARE TUMORBIOLOGIE

MATTHIAS LUTZ ZEPPER

Epigenetic characterization of murine Dnmt1-deficient MLL-AF9 leukemia

*DNA methylation in large regions and at cis-regulatory elements
dissected in the Dnmt1^{-/-} mouse model*

2020



BIOLOGIE

Epigenetic characterization of murine Dnmt1-deficient MLL-AF9 leukemia

*DNA methylation in large regions and at cis-regulatory elements
dissected in the Dnmt1^{-/chip} mouse model*

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften im Fachbereich Biologie der
Mathematisch-Naturwissenschaftlichen Fakultät der
Westfälischen Wilhelms-Universität Münster

vorgelegt von

MATTHIAS LUTZ ZEPPER
aus LUDWIGSBURG

2020



Dekanin: Prof. Dr.rer.nat. Susanne Fetzner

Erster Gutachter: Prof. Dr.rer.nat. Frank Rosenbauer

Zweiter Gutachter: Prof. Dr.rer.nat. Joachim Kurtz

Tag der mündlichen Prüfung: 12. März 2020

Tag der Promotion:

Abstract

Over the past four decades, little progress was made in the standard therapy of acute myeloid leukemia (AML). Only in recent years, the standard anthracycline/cytarabine combination chemotherapy is increasingly replaced by or supplemented with new substances such as gemtuzumab, ozogamicin, enasidenib or midostaurin. None the less, an unmet need for specific drugs stands, since AML is a heterogeneous disease and many subtypes still lack molecularly targeted therapy options.

To aid the identification of new, specific molecular therapy targets, we utilized a mouse model to elicit acute myeloid leukemia in a DNA hypomethylation background. We proposed that silenced tumor-suppressor genes would become reactivated due to insufficient methylation and the model could thus point us to new molecular targets.

Indeed, we observed that the model required appropriate DNA methylation levels to maintain transcriptional sanity, to avoid senescence and to ultimately preserve its full self-renewal capability. Therefore, we developed a new analysis method for methylation data to elaborate on the underlying causes and consequences. We showed that, contrarily to our initial proposal, no reactivation of genes by promoter hypomethylation occurred. Subsequently, we explored possible alternatives to explain the phenotype.

Since misplaced or anomalous enhancers have emerged as important contributing factors of leukemogenesis, we asked whether enhancers might be sites of therapeutically relevant DNA methylation changes. Here we present a comprehensive characterization of bivalently transcribed active enhancers and their respective methylation status. Our analysis highlights a GC-rich subgroup of regulatory elements, which are unmethylated on DNA level, but characterized by high H3K4me3 in leukemia as well as in various regular hematopoietic lineages. These elements resembled bivalently marked promoters, were presumably bound by Mll2/COMPASS and targeted by the histone demethylase Utx (Kdm6a) for activation. Hence, it is suggested that specific Utx inhibitors upon availability should be investigated with regard to their therapeutic potential in Mll-rearranged leukemia.

Acknowledgments

Firstly, I would like to sincerely thank my graduate advisor Frank Rosenbauer for the opportunity to join his laboratory, to contribute to challenging projects and to become acquainted with a broad range of scientific methods. I gratefully acknowledge the possibility to follow my curiosity and address own hypotheses, while being provided guidance when needed. Furthermore, I much cherish the possibility to attend the SPP1463 meetings in Freiburg (2013) and Menaggio (2016) as well as the hematological symposium in Rotterdam (2015).

I appreciate Joachim Kurtz, Martin Dugas and Carsten Müller-Tidow serving on my thesis committee and providing helpful scientific advice. Sincere thank is given to all collaborators, in particular Michael Rehli, Claudia Gebhard, Frank Lyko, Günther Raddatz and Daniel Lipka. I thankfully acknowledge Ido Amit, Assaf Weiner and Seung-Tae Lee for kindly providing additional data related to their publications upon request.

The helpful and friendly atmosphere at the institute was highly valuable and I would like to thank all former and current colleagues of the IMTB. There are too many to name them all, but Irina Savelyeva, Verena Gröning, Thorsten König and Lena Tepe shall be mentioned for strongly supporting the research presented herein.

Pivotal support to this project was also provided by collaborating bioinformaticians and members of online communities. Many thanks to Carolin Walter, Eduard Szöcs, Christian Rohde for teaching me skills and helping with coding questions. I appreciate the whole online community for contributing to free and open-source software and for writing useful tutorials, manuals, forum posts, blog entries and tweets.

I am fortunate to have been mentored by Elisa Franz while writing this thesis. Many thanks also to Uri Alon, Randy Pausch, Jennifer Heemstra and Ben Barres, whose inspirational talks and encouraging writings on science, motivation and (career) guidance are preeminent. In this context, I also would like to acknowledge Jonathan Haidt and Dan Ariely, whose publications and talks have been extremely helpful to me.

Last but not least, I am deeply grateful for my family, teachers and friends, who supported me and fostered my passion for science.

Contents

Abstract	1
1 Introduction	9
1.1 MLL-AF9 and MLL-rearranged leukemia	9
1.2 DNA methylation in acute myeloid leukemia	10
1.3 Enhancers	11
1.3.1 Enhancers in general	11
1.3.2 Enhancers contribute to leukemogenesis	12
1.4 Previous findings	13
1.5 Aim of this thesis	14
I Methylomes of MLL-AF9 c-Kit+ leukemia	17
2 Methylome data of MLL-AF9 leukemia	19
2.1 Leukemia-related demethylation	20
2.2 Chromatin-state-dependent demethylation	21
2.2.1 Ramifications of lamina-association on methylation	22
2.2.2 Assessment of CpG-Island methylation	23
3 Specification of the compromised regions	27
3.1 Demethylation at single CpG resolution	27
3.2 Increased partial methylation in leukemia	29
3.3 Standard approach failed to discriminate domain borders	31
4 Modeling the methylation probability	33
4.1 A GAM to predict the methylation rate	33
4.1.1 Introduction to GAM models	33
4.1.2 Reasons to choose a GAM	34
4.2 Fitting a GAM on methylome data	35
5 Relationship of chromatin structure and methylation persistency	39
5.1 Chromosomal insulation and interaction	39
5.2 Determining factors of methylation persistency	41
5.3 Summary	44

6	Methylome analysis of matched non-malignant hematopoietic progenitors	45
6.1	WGBS data of the MPP hierarchy	45
6.2	Leukemia-related demethylation revisited	47
II	Deregulated genes and pathways	49
7	Transcriptional analysis	51
7.1	Characterization of Dnmt1-hypomorphic transcription	52
7.1.1	Expression changes linked to promoter hypomethylation	53
7.1.2	Elongation efficiency of reference transcripts	55
7.2	Differential gene expression analysis	56
7.3	Contrast of Dnmt1 ^{-/chip} vs. Dnmt1 ^{+/+}	58
7.3.1	Altered genes	58
7.3.2	Altered pathways	59
7.4	H3K4me3 buffer domains	61
8	Experimental transcriptome	65
8.1	Assembly of non-reference transcripts	65
8.2	Isolated transcriptional initiation events	66
8.3	Summary	68
III	Enhancer delineation	69
9	Enhancer calling and classification	71
9.1	CAGE-seq derived enhancers	72
9.2	Enhancer clustering	73
9.2.1	Major cluster assignment by k-means	73
9.2.2	Minor cluster assignment by hierarchical clustering	75
9.3	Clades accumulating CAGE-enhancers	78
9.3.1	Characteristics in terms of healthy hematopoiesis	78
9.3.2	Characteristics in MLL-AF9 leukemia	80
10	Enhancer motifs, targets and regulation	87
10.1	Motif analysis	87
10.1.1	Basic procedure	87
10.1.2	Motifs enriched in strongly accumulated clades	88
10.1.3	Motifs enriched in strongly depleted clades	90
10.2	Methylation of enhancers and their motifs	91
10.2.1	Methylation mapping at enhancer regions	92
10.2.2	Methylation mapping at isolated motifs	93
10.3	MLL2 (Kmt2b) binding at strongly enriched enhancers	95
10.4	Enhancer target genes	98
10.4.1	Assessment of Mll2 target genes	101

10.5 Summary and outlook	102
IV Discussion	105
11 Synopsis of results	107
12 Ramifications of the Dnmt1 $^{-/chip}$ methylome	109
12.1 Assumptions regarding the Dnmt1 $^{-/chip}$ methylome	110
12.1.1 Inverse relationship of division rate and methylation persistency .	110
12.1.2 Inadvertent reactivation of epigenetically repressed genes	111
12.2 Characteristics of the large-scale compromised regions	112
12.2.1 Do Dnmt1 $^{-/chip}$ methylomes harbor PMDs?	113
12.2.2 Impact of PMD-like compromised regions	114
12.3 Persistent methylation at CpG-Islands and promoters	117
12.3.1 Persistently unmethylated CpG-Islands	117
12.3.2 Methylation transition at CpG-Islands	118
12.3.3 Persistently methylated CpG-Islands	119
12.4 Methylation-independent roles of Dnmt1	119
13 Transcriptional regulation in leukemia	121
13.1 Establishment of our MLL-AF9 enhancer catalog	121
13.2 Notable characteristics of our enhancer catalog	122
13.2.1 Rarity of leukemia-specific enhancers	122
13.2.2 H3K4me3 as hallmark of particular enhancer clades	123
13.3 H3K4me3, Mll2, CXXC-domains and leukemia	125
13.4 Implications for the Dnmt1 $^{-/chip}$ genotype	127
13.4.1 Methylation determines enhancer activity	127
13.4.2 Methylation affects chromatin organization	128
13.5 Outlook	129
Appendices	131
A Sources of reanalyzed third-party datasets	133
B Sample reference	135
C Curriculum vitae	137
Bibliography	139
Abbreviations	163

CONTENTS



- For this thesis, additional figures and more detailed method descriptions are available as supplementary online content.
- All supplements can be found at <https://thesis.matthias-zepper.de>

Text abridged. Supplementary
online information is available.

Chapter 1

Introduction

Contents

1.1	MLL-AF9 and MLL-rearranged leukemia	9
1.2	DNA methylation in acute myeloid leukemia	10
1.3	Enhancers	11
1.3.1	Enhancers in general	11
1.3.2	Enhancers contribute to leukemogenesis	12
1.4	Previous findings	13
1.5	Aim of this thesis	14

1.1 MLL-AF9 and MLL-rearranged leukemia

The MLL-AF9 fusion protein is a recurrent leukemogenic genetic abnormality, which arises from the chromosomal translocation t(9;11)(p21.3;q23.3). This results in a gene fusion of the KMT2A gene (MLL1 protein) with the MLLT3 gene (AF9 protein) [1,2].

MLL-AF9 fusions are part of a larger group, the KMT2A-rearranged leukemia [reviewed in 3]. To date, more than 60 fusion partners [4, reviewed in 5] and multiple breakpoints [6] have been reported for the KMT2A gene / MLL1 protein, which is one of several vertebrate homologues of the *Drosophila* positional identity regulator Trithorax [7,8] and thus implicated in H3K4 methylation.

Although the first KMT2A fusion was reported as additional recombination in chronic myelogenous leukemia (CML) [9], they are much more prevalent in pediatric acute myelogenous leukemia (AML) [10,11]. KMT2A rearrangements in general are found in roughly 40% of the infant¹ cases of acute myeloid leukemia, but constitute to less than 5% of the cases within the AYA group² and are even more scarce in older patients [12]. In elderly patients, KMT2A rearrangements are very rare and occur almost exclusively during relapses as side effect of leukemia treatment with topoisomerase inhibitors [13].

¹ 0 to 3 years of age

² adolescents and young adults, 15 to 39 years old

The MLL-AF9 rearrangement specifically can be found in 9.5% of childhood and in 0.5% of adult acute myeloid leukemia [11]. Therefore, it is recognized as a separate entity by the current WHO classification of myeloid neoplasms and acute leukemia and is considered as one of the eleven subcategories within *AML with recurrent genetic abnormalities*³ [14].

Because the AF9 protein, which is encoded by the MLLT3 gene, is a component of the super elongation complex (SEC) [15, 16], MLL-AF9 leukemia is mostly characterized by a dysregulation of transcriptional elongation [reviewed in 17]. The downregulation of MLLT3 in cultured haematopoietic stem-cells impairs self-renewal and negatively affects engraftment efficiency [18].

Human MLL-AF9 can also transform mouse cells and educe a well characterized leukemia model system [19–21] out of several early hematopoietic lineages [22]. Importantly, it is known that the expression of CD117/c-Kit is expedient to enrich for leukemia stem cells (LSCs) [23] from the leukemic bone marrow [24, 25]. Such LSC-enriched c-Kit⁺ fractions of independently established MLL-AF9 leukemia were investigated in the present study.

1.2 DNA methylation in acute myeloid leukemia

Since 1925, the occurrence of methylated cytosine⁴ in nucleic acid extracts of bacteria is known [26], while its incorporation in regular nucleotides of eukaryotes could be shown in 1951 [27]. Because of the high frequency of 5-methylcytosine (5-mC) in genomes, the term DNA methylation typically refers to this base, although other bases such as adenine can be modified accordingly (N6-methyladenine) [28] [reviewed in 29].

In the subsequent decades, 5-mC DNA methylation was shown to be involved in various, mostly repressive functions such as X chromosome inactivation, imprinting and the silencing of endogenous retroviruses as well as regular genes [reviewed in 30]. Despite its role in long-term silencing, localized DNA methylation is more dynamic than originally believed and can be deposited or removed by various enzymes in a timely manner [reviewed in 31]. In humans and mice, the methylation of 5-mC in genomic DNA is performed by Dnmt1, Dnmt3a and Dnmt3b [reviewed in 32], whereas the TET enzymes [reviewed in 33] mediate the oxidative reversal [reviewed in 34]. Whether these further oxidized bases 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) are just intermediates in the reversal process or have distinct regulatory functions is still disputed [reviewed in 35]. Remarkably, also Dnmt3a and Dnmt3b may be implicated in the active demethylation of DNA [reviewed in 36].

The measurement of DNA methylation is typically performed by high-throughput bisulfite sequencing [37], which can nowadays be scaled down to single cells [38]. Bisulfite treatment will cause a conversion of unmethylated cytosine bases to uracil, which are replaced by thymine during a subsequent PCR amplification. Those bases will appear as

³ AML with t(9;11)(p21.3;q23.3);MLLT3-KMT2A

⁴ also known as 5-methylcytosine (5-mC) and epi-cytosine

mutations with regard to a reference genome and hence the methylation rate can be calculated from the ratio of mutated vs. reference reads after the alignment [reviewed in 39]. Alternatively, enrichment-based methods like MeDIP-seq and MRE-seq are available to measure DNA methylation [40].

In 1983, it was shown that aberrant DNA hypermethylation may cause thalassaemia [41], which was the first pathogenic DNA methylation to be described. Subsequently, an association of altered DNA methylation with several diseases was shown [reviewed in 42]. In cancer, the silencing of tumor-suppressor genes by DNA hypermethylation at promoters is widespread [reviewed in 43], although DNA hypomethylation seems to be the norm on a genome-wide scale [44]. Aberrant DNA methylation is also implicated in the pathogenesis of acute myeloid leukemia (AML), in which epigenetic peculiarities exert a strong influence [45] due to fewer mutations than most other cancers [46]. Yet, mutations in genes related to DNA methylation are quite common [46] and various AML subtypes display distinct methylation profiles [47, 48]. For example, haploinsufficiency of Dnmt3a enhances self-renewal [49] of hematopoietic stem cells and predisposes them to myeloid malignancies [50], possibly by hypomethylation of the intergenic euchromatin space [51]. Contrarily, pathogenic hypermethylation is observed in AML with mutations in genes like IDH1 / IDH2 [52], BCAT1 [53] or WT1 [54] or the TET enzymes [55–57]. Mechanistically, hypermethylation interferes with the binding of transcription factors such as PU.1 [58] or C/EBP α [reviewed in 59], which results in a differentiation block.

1.3 Enhancers

1.3.1 Enhancers in general

Operationally, enhancers augment the activity of a nearby promoter and are enriched in recognition motifs for sequence-specific transcription factors. Since the orientation and to a large extent also the exact position relative to the promoter is insignificant, they are considered to be orientation- and position-independent. Anyhow, most enhancers are thought to reside in the vicinity of the targeted promoter, although long-range contacts are possible [60, 61] as illustrated by the MYC locus [62] or the SHH gene [63]. One gene is typically targeted by several cis-regulatory elements and one enhancer may also be involved in the regulation of different genes [64–66].

The mechanism of enhancer action requires a change of the three-dimensional chromatin structure and the formation of a DNA loop [67]. Initially, it was believed that specific transcription factors initiate the loops directly between enhancers and promoters [68, 69], but more recent models favor preexisting loops that dynamically slide along the chromatin fibre until a specific contact is established [70, 71].

Once enhancer and promoter have converged, a variable cascade of events is triggered that ensures transcriptional initiation or pause release. The most important player in this cascade is the mediator complex [72, 73], but many other coactivators [74] are implicated as well, while regulatory cues are provided by epigenetic marks on histones [75, 76]. Ul-

timately, phosphorylations in the c-terminal tail of RNA polymerase II control the transcriptional activity [77].

Bringing enhancer-bound protein factors close to the promoter-bound preinitiation complex (PIC) is the best characterized function of enhancers, but not the sole. Active enhancers may give rise to bidirectional transcripts termed enhancer RNAs (eRNA) [78–80]. eRNAs were soon shown to be capped on the 5'end, short (<1 kb), bidirectional, unspliced and rapidly degraded by the exosome [81–83], which contested a relevant functional role of the pervasive transcription initiating from enhancers.

Contrarily, eRNAs have been demonstrated to stabilize enhancer promoter association at steroid hormone response genes [84,85], to be of importance for H3K4me1 and H3K4me2 deposition by MLL3 and MLL4 at *de novo* enhancers [86] and to be subject to functional methylation [87]. Thus, they exert meaningful roles at least for a subset of enhancers [reviewed in 88] and their expression generally correlates with their target genes [89].

Intriguingly, the majority of lncRNAs originate from enhancer-like elements [90, reviewed in 91]. Previously the lack of splice donors [92] proximal to enhancers was believed to preclude productive elongation of eRNAs [83], but it was shown that they are mostly actively terminated [93]. The main reason seems to be the prevention of convergent transcription [94,95], which triggers strong DNA-damage signaling [96]. Particularly at super enhancers, which harbor clustered enhancer elements, such RNA polymerase II collisions would inevitably occur upon elongation of eRNAs and thus their timely termination is pivotal [96].

The propensity of an enhancer to generate eRNA transcripts as well the signature of histone marks at the site [97,98] depends on the state: closed, primed, poised or active [reviewed in 99]. Since some of the next-generation sequencing based methods rely on these patterns to identify enhancers genome-wide, sensitivity and specificity of the respective method will vary and sometimes confine itself to enhancers in a particular state [reviewed in 100].



Text abridged. Supplementary online information is available.

- Classes of cis-regulatory elements.
- Mode of action, enhancer states and activation.
- Discovery and function of enhancer RNAs (eRNAs).
- Methods for genome-wide identification of enhancers.
- Principles of pathogenic enhancer aberrations.

1.3.2 Enhancers contribute to leukemogenesis

Hematopoiesis, the development of diverse mature blood cells from hematopoietic stem cells requires an intricate regulation. The appropriate expression of key transcription factors such as PU.1, GATA1, GATA2 or C/EBP α at various stages governs progenitor commitment and differentiation. Ten-thousands of enhancers are presumably involved in hematopoietic regulation in total [101–103].

Generalizations about leukemogenesis are almost futile, given the many different subtypes. MYC and its enhancers, however, are recurrently implicated in various leukemia as well as other cancers [104–106]. In contrast, the downregulation of PU.1 is restricted to hematopoietic cancerogenesis. None the less, it represents a proven route to leukemia [107, 108] and already subtle PU.1 reduction by a heterozygous deletion of an enhancer was sufficient to initiate a myeloid-biased preleukemic state [109].

Especially late-onset leukemia are characterized by the presence of preleukemic hematopoietic stem cells, which have progressively acquired an increasing mutation burden over their lifetime. These cells are not yet leukemic and expansive, but exhibit spurious alterations in their gene expression programs and enhancers, which increase susceptance to uncontrolled cellular expansion [110].

Unsurprisingly, preleukemic states are heavily promoted by aberrant super-enhancers, since they govern the activation of whole gene clusters. The introduction of binding motifs for the MYB transcription factor by somatic mutations forms a novel super enhancer upstream of the TAL1 oncogene and sustains its expression [111, 112] in T cell acute lymphoblastic leukemia (T-ALL). In a particularly dismal ALL subtype driven by TCF3-HLF, the chimeric transcription factor activates an enhancer cluster controlling expression of the MYC gene and instigates the respective transcriptional program [113]. Because hematopoietic MYC expression is intricately regulated by combinatorial and additive activity of individual enhancer modules within this cluster [106], a dysregulation of MYC program can be mediated by various factors or arise from amplifications within the enhancer region [105]. Therefore, the enhancer cluster is complicated in many leukemia subtypes and also pivotal for MLL-AF9-driven AML [106].

A different mode of action has been reported for a distinct subtype of acute myeloid leukemia. In AML with the *inv(3)(q21;q26)* karyotype [14] a genomic rearrangement repositions a distal hematopoietic enhancer of GATA2 in close proximity to the stem-cell regulator EVI1, which is ectopically activated. Concomitantly, GATA2 expression is diminished and both events facilitate leukemic expansion [114, 115].

1.4 Previous findings

The Rosenbauer laboratory has a long-standing interest in the role of DNA-methylation for normal and abnormal hematopoiesis [116, 117]. Lena Vockentanz, a former PhD student [118] and Irina Savelyeva, a previous postdoc in the laboratory, conducted many experiments, which have collectively shown that Dnmt1 expression is essential for cell-autonomous activity of MLL-AF9 leukemia cells.

Using a poly I:C-inducible MLL-AF9 Mx1-Cre × Dnmt1^{fl/fl} mouse model, the rate-limiting impact of diminished Dnmt1 levels on leukemia development was shown. However, non-excised Dnmt1^{fl/fl} cells, which had escaped induction, typically outgrew their rearranged cognates in prolonged experimental settings. Thus, this model appeared non-optimal for studying the function of leukemic stem cells (LSCs) in particular and

MLL-AF9 leukemia with a Dnmt1 hypomorphic background was created using bone marrow cells from Dnmt1^{-/chip} mice [119–121] as donors.

Consistent with the previous MLL-AF9 Mx1-Cre × Dnmt1^{fl/chip} results, animals transplanted with Dnmt1^{-/chip} MLL-AF9 leukemia fell ill significantly later than those of the wild type control group⁵. Animals with end-stage leukemia exhibited massive infiltration of donor MLL-AF9-IRES-GFP cells into the bone marrow and spleens of recipient mice. Like the Dnmt1^{+/+} control, Dnmt1^{-/chip} MLL-AF9 leukemia mimicked CD11b+ granulocyte-macrophage progenitors and in part expressed the precursor marker CD117+ (c-Kit). The latter was useful to enrich LSCs from the leukemic bone marrow [24, 25] and limiting dilution assays showed that the rate of LSCs in leukemia with Dnmt1^{-/chip} genotype was significantly lower [118].

This indicated that the prolonged latency was attributable to cell-autonomous functions of LSCs instead of microenvironment- or engraftment-deficiencies. The latter possibility was firmly ruled out by a short-term (20 h) engraftment assay, which detected that a comparable number of transplanted cells of both genotypes had infiltrated the hematopoietic organs. Furthermore the recipients for the transplantation experiments were all wild-type mice, which corroborated engraftment independence of the observations.

Altogether the results argued for an impaired self-renewal of LSCs, which was confirmed by in vitro serial replating in methyl-cellulose. To determine, if the reduced self-renewal was functionally linked to an altered cell cycle, Hoechst 33342 incorporation was monitored: Dnmt1^{-/chip} MLL-AF9 exhibited a 31 % reduction in cell numbers for LSCs in the S-G2-M phases, but no increase in apoptosis. Hence, a proportion of Dnmt1^{-/chip} LSCs accumulated in the non-cycling G1 phase. By quantitative PCR experiments, this G1 arrest could in part be attributed to an increased expression of the two transcripts p19/Arf and p16/Ink4A at the Cdkn2a locus in Dnmt1^{-/chip} leukemia. Both gene products are known to accompany senescence in murine and human cells. Indeed, variable fractions of senescent cells could be detected in the Dnmt1^{-/chip} group by β-galactosidase (β-gal) staining. While non-leukemic hematopoietic stem/progenitor populations of Dnmt1^{-/chip} were devoid of senescent cells⁶, the leukemia bulk contained up to 7 % (avg. 2.8 %) and the LSC fraction up to 53 % (avg. 9.3 %) senescent cells. Although the rate of senescence was highly variable across replicates, these findings suggested a relevant inherent senescence risk of Dnmt1^{-/chip} MLL-AF9 leukemia and provided a first plausible route to cell cycle exit and self-renewal defects.

1.5 Aim of this thesis

Although the first hypothesis to explain the prolonged latency of Dnmt1^{-/chip} MLL-AF9 leukemia had been drafted, it remained elusive how the senescence program was triggered by reduced Dnmt1 levels in the first place. Since chemical inhibitors of DNA

⁵ median latency 140.8 ± 37.0 days versus 89.8 ± 24.1 days after transplantation

⁶ despite known aberrant hematopoiesis and disordered lymphoid lineage development [116]

methylation, such as decitabine, which has received market authorization by the European Medical Agency, have proven therapeutic efficacy for the treatment of acute myeloid leukemia [122, 123], we assumed a common methylation-dependent mechanism.

The treatment with inhibitors results in an undirected reduction in DNA methylation. However, it is generally presumed that most methylation changes occur silently and therapeutic effects are only conferred, when yet to be characterized key sites have been affected by random. Our mouse model seemed to be suitable to aid the identification of those key sites, as it permitted to reduce DNA methylation by genetic Dnmt1 deficiency instead of inhibitor treatment and thus allowed to circumvent possible side-effects. We utilized the Dnmt1^{-/-} mouse strain to elicit acute myeloid leukemia by transduction of MLL-AF9 and asked, how selective pressure and impaired methylation maintenance would shape the leukemia methylome.

The c-Kit⁺ sorted, leukemic stem cell fractions were subjected to extensive, genome-wide characterization by next-generation sequencing experiments:

- Whole-Genome Bisulfite sequencing (WGBS) to assay DNA methylation
- RNA-seq to study gene expression changes and alternative splicing
- CAGE-seq to detect aberrant transcriptional initiation and call enhancers
- H3K4me3 ChIP-seq to corroborate active transcription and identify broad peaks, which are referred to as buffer domains and mark cell identity genes [124].

The bioinformatic analysis and interpretation of the gathered data from these experiments was the centerpiece of the project. To quantify the methylation persistence across large regions and detect regional trends, a novel method for WGBS data comparison based on Generalized Additive Models was developed. To address anomalous enhancers, which have emerged as important factors in leukemogenesis [↔ subsection 1.3.2], a comprehensive characterization of bivalently transcribed active enhancers and their respective methylation status was performed.

All results were placed in context with published third-party datasets [↔ Appendix A, p.133], which were often reanalyzed from scratch to assure full comparability with our own data. Selected genes and enhancers were also experimentally tested in vitro by shRNA knock-down or CRISPRi for their effect on self-renewal and growth rate.

Part I

Methylomes of MLL-AF9 c-Kit+ leukemia

Chapter 2

Methylome data of MLL-AF9 leukemia

Contents

2.1	Leukemia-related demethylation	20
2.2	Chromatin-state-dependent demethylation	21
2.2.1	Ramifications of lamina-association on methylation	22
2.2.2	Assessment of CpG-Island methylation	23

The data presented above [↔ section 1.4, p.13] indicated that reduced Dnmt1 activity interferes with self-renewal and leukemia stem cell (LSC) function in a cell-autonomous manner. Some cells, in particular LSCs, seemed to undergo G1 arrest and become senescent. Although the Dnmt1^{-/chip} mouse exhibited aberrant hematopoiesis and disordered lymphoid lineage development [116], incidence for senescence was only present in the leukemia. We assumed that oncogenic transformation with MLL-AF9 crucially depends on changes in the methylome, which could no longer be fully maintained in a Dnmt1 hypomorphic setting, being reminiscent of the treatment with chemical inhibitors of DNA methylation [122, 123]. Thus, we asked, how malignant transformation would manifest itself in the methylome of a wild-type cell and in which regard that of a Dnmt1^{-/chip} cell would deviate.

To generate acute myeloid leukemia (AML) in the Dnmt1^{-/chip} background and Dnmt1^{+/+} controls, we transduced respective ex-vivo bone marrow progenitor cells (Lin⁻ Sca-1⁺ c-Kit⁺) with the hematopoietic-specific oncogene MLL-AF9 and transplanted them into sub-lethally irradiated wild-type recipient mice. Transduction of MLL-AF9, a MLL-fusion protein [reviewed in 3], can transform several early hematopoietic lineages [22] in a well characterized manner [19–21]. Importantly, it is known that the expression of CD117/c-kit is expedient to enrich leukemia stem cells (LSCs) from the leukemic bone marrow [24, 25]. Such LSC-enriched c-Kit⁺ fractions (Lin⁻ IL-7R α ⁺ Sca-1⁻ c-Kit⁺ CD34⁺ FC γ R⁺) of three independently established MLL-AF9 Dnmt1^{+/+} and Dnmt1^{-/chip} leukemia were (in collaboration with Frank Lyko and Günter Raddatz) subjected to Whole-Genome Bisulfite Sequencing (WGBS) to assess the DNA methylation. Although RRBS and methylation array data from human leukemia had been published before [46, 48], this were the first WGBS methylomes generated for murine MLL-AF9 leukemia.

2.1 Leukemia-related demethylation

Because of the high cross-sample consistency for the replicates [▷ [supplement](#)], we decided to pool the data of three biological replicates per genotype and generated meta-samples, which simplified further analyses. Instead of averaging over the methylscores of the three biological replicates, which could easily misrepresent sites with highly different coverage, we pooled the aligned samples on a read level before calling the combined methylscore over all reads at a site.

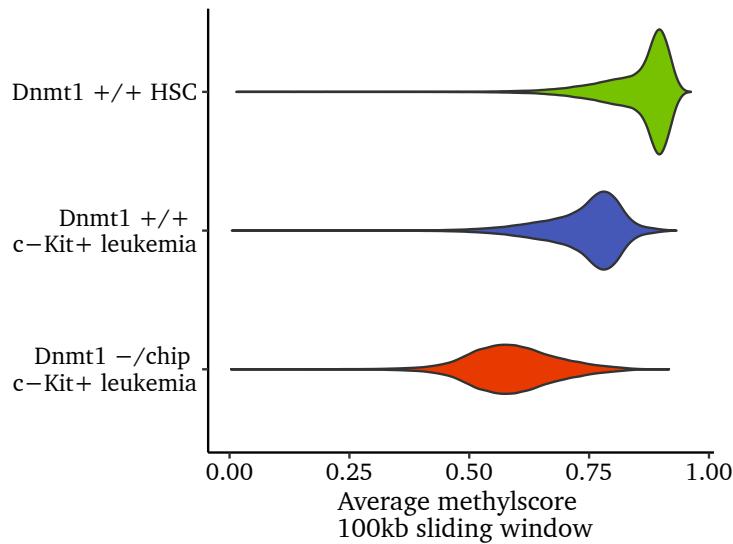


Figure 2.1: Distribution of methylscore averages across 100 kb windows, sliding with a step size of 25 kb along the genome.

In comparison to the previously published methylome of mouse hematopoietic stem cells (HSC) [125], both leukemia were hypomethylated [▷ [Figure 2.1](#)]. The methylscore of $Dnmt1^{+/+}$ HSCs typically averaged at 0.88 within a 100 kb window with little deviation. While the average decreased in both leukemia, the degree of hypomethylation, in accordance with reduced $Dnmt1$ -expression, was more pronounced in $Dnmt1^{-/chip}$. In contrast, the deviation¹, which increased diametrically as indicated by the skewness of the curves, was particularly large in $Dnmt1^{-/chip}$ [▷ [Table 2.1](#)] . Thus, the methylome of $Dnmt1^{-/chip}$ was the most hypomethylated and least uniform in our assay.

A pairwise comparison of $Dnmt1^{+/+}$ HSC vs. $c\text{-}Kit^+$ leukemic cells showed a global reduction of the methylation levels by approximately 10 %, regardless of the original baseline methylscore average within the particular window [▷ [Figure 2.2, left panel](#)].

¹ The median absolute deviation is another measure of spread, however more robust than variance and standard deviation for cases with extremely high or low values and preferred for non-normality. It is defined for univariate data as the median of the absolute deviations from the data's median:

$$\text{MAD}_{x_1, x_2, \dots, x_n} = \text{median}(|x_i - \text{median}(x)|)$$

Sample	Median	Median Absolute Deviation
Dnmt1 ^{+/+} HSC	0.88	0.04
Dnmt1 ^{+/+} c-Kit ⁺ leukemia	0.76	0.06
Dnmt1 ^{-/-chip} c-Kit ⁺ leukemia	0.58	0.09

Table 2.1: Summary for the 100 kb sliding window averages, which are shown in [Figure 2.1](#) as well as [Figure 2.2](#).

In contrast to the uniform hypomethylation, which we observed for the Dnmt1^{+/+} cells, the Dnmt1^{-/-chip} c-Kit⁺ methylome surprisingly divided, albeit globally hypomethylated, into areas of higher and lower methylation persistence [[> Figure 2.2, right panel](#)]. In some areas the hypomethylation mediated by Dnmt1-impairment was confined to further 10 %, whereas other sections of the genome on average lost 25 % of their methylation in comparison to wild-type leukemia. Thus, we named these different sections of the genome *persistent* and *compromised* regions. Such a separation had been shown for solid tumors, but not for leukemia. Importantly, the compromised regions did not overlap with the methylation canyons previously described in hematopoietic stem cells [125] [[> data not shown](#)].

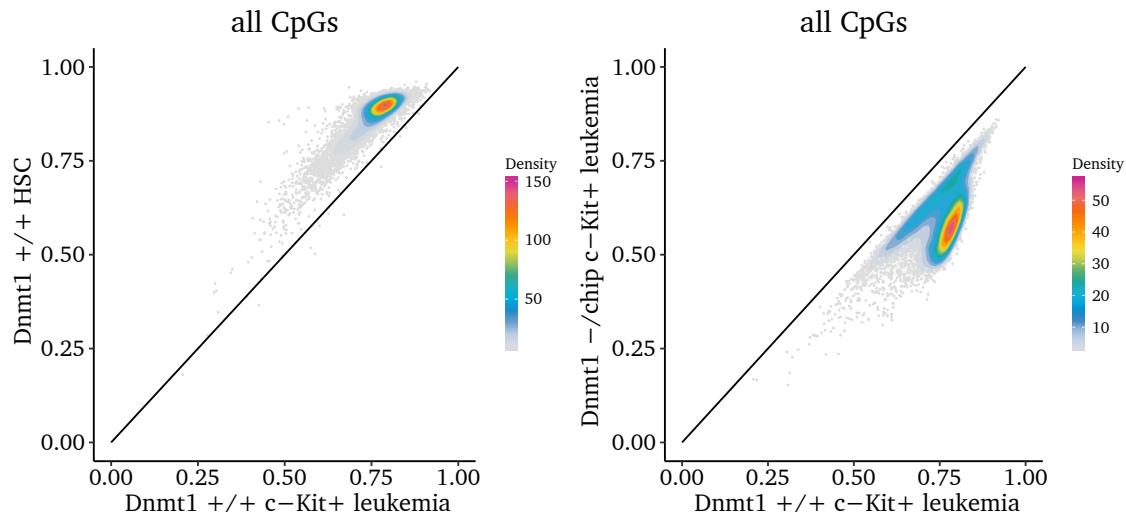


Figure 2.2: Pairwise comparison of two meta-samples per panel. Each dot denotes a particular 100 kb window and its position on the axis is decided by the average methylation score in the respective sample. A color-encoded density scale highlights areas with many individual points.

2.2 Chromatin-state-dependent demethylation

Several previous studies reported hypomethylation in solid tumors, which did not occur uniformly, but seemed to associate with the underlying chromatin structure and to intensify in heterochromatic regions [126–129]. Published WGBS data from hematopoietic malignancies was rare, but with on average just 5% intensification in heterochromatic, lamina-associated domains (LADs), human B-ALL [130] for example was virtually devoid of such patterns. This also applied to the hypomethylation observed in our

murine $Dnmt1^{+/+}$ c-Kit $^+$ dataset, which - as mentioned previously - was weak and uniform [▷ [Figure 2.2, left panel](#)].

The methylome of $Dnmt1^{-/chip}$ c-Kit $^+$ leukemia cells, which exhibited unevenly distributed hypomethylation [▷ [Figure 2.2, right panel](#)], unexpectedly resembled those of aforementioned solid tumors. Therefore, we presumed that those patterns might also be ramifications of the chromatin structure.

2.2.1 Ramifications of lamina-association on methylation

We inferred the chromatin structure of $Dnmt1^{-/chip}$ c-Kit $^+$ leukemia from a published annotation of constitutive lamina-association downloaded from the NCBI Gene Expression Omnibus with the accession GSE36132 [131]. In this dataset, the authors termed regions, which were lamina-associated [132] in all cell types constitutive LADs (cLADs) and called such, which never associated with the lamina, constitutive interLADs (ciLAD). Taken together the two groups comprised 71 % of the genome. The remaining 29 % are variable among cell types and thus termed flexible LADs (fLADs).

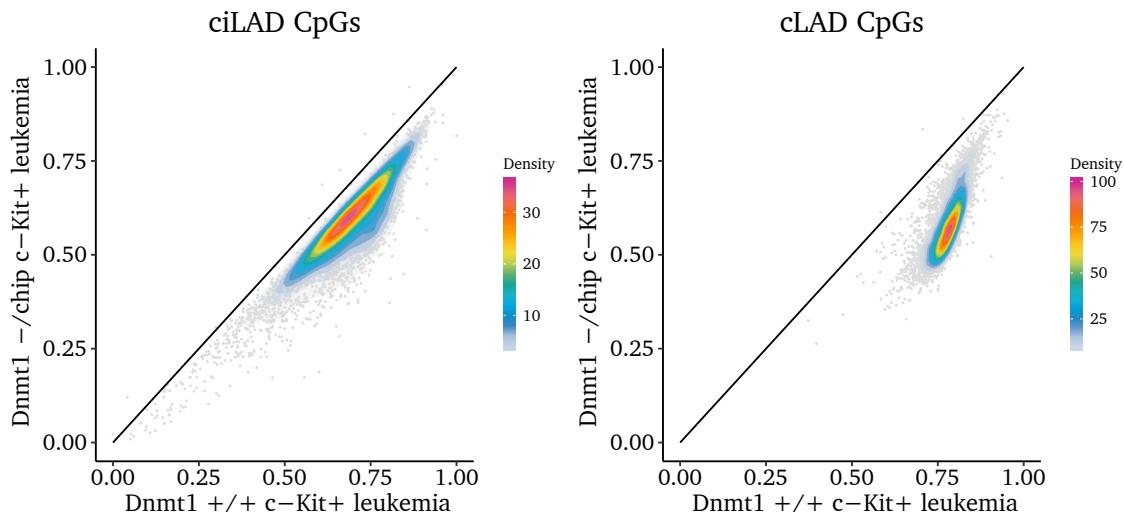


Figure 2.3: For these intra-leukemia contrast plots the methylscore averages within 100 kb windows (slid by 25 kb steps) were calculated after prior separation of the CpGs in ciLAD respectively cLAD sets. Sections without a sufficient coverage were excluded.

It was clearly possible to relate the higher and lower methylation persistence in $Dnmt1^{-/chip}$ to ciLADs and cLADs respectively [▷ [Figure 2.3 vs. right panel of Figure 2.2](#)].

Hence, methylation loss in $Dnmt1^{-/chip}$ c-Kit $^+$ leukemia remarkably intensified in cLAD regions, which were unequivocally methylated in $Dnmt1^{+/+}$ HSC and LSC [▷ [Figure 2.4, right panel](#)]. No difference in methylation persistency between cLADs and ciLADs could be observed for the $Dnmt1^{+/+}$ HSC versus LSC contrast [▷ [Figure 2.4](#)]. Furthermore, it should be noted that areas of relatively low methylation (such as $\leq 65\%$) were confined to the ciLAD areas in $Dnmt1^{+/+}$ samples.

Therefore, this analysis for the first time established an association of compromised regions with a lack of $Dnmt1$. Additionally, the methylome of $Dnmt1^{-/chip}$ MLL-AF9 was

the first leukemia sample to exhibit a drastically variable methylation persistence known from solid tumors [129].

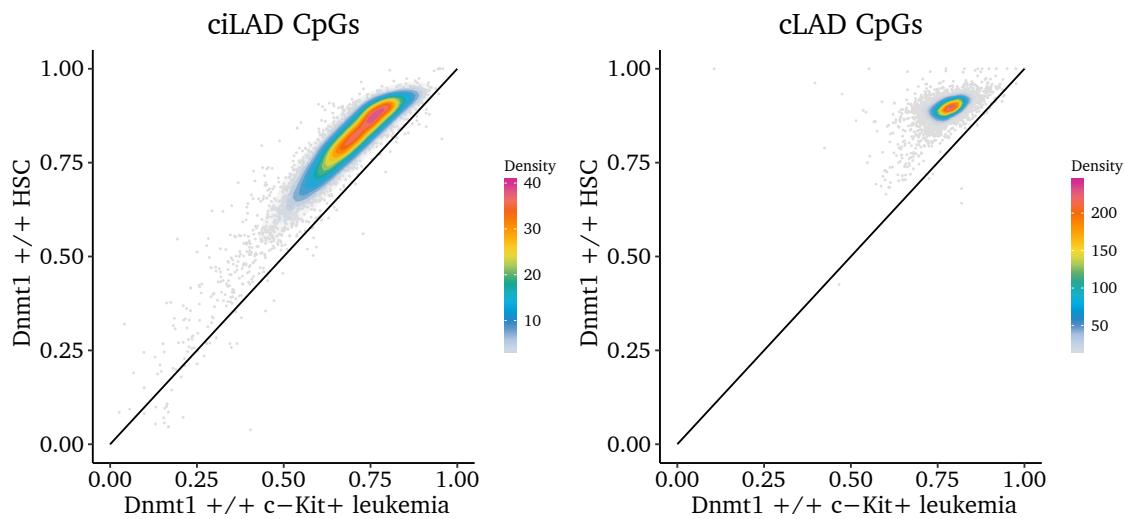


Figure 2.4: CpGs were partitioned into ciLAD and cLAD collections and mapped on 100 kb windows (slid by 25 kb steps) for calculating mean methylation. The comparison of $Dnmt1^{+/+}$ $c\text{-}Kit^+$ leukemia vs. HSCs is shown.

2.2.2 Assessment of CpG-Island methylation

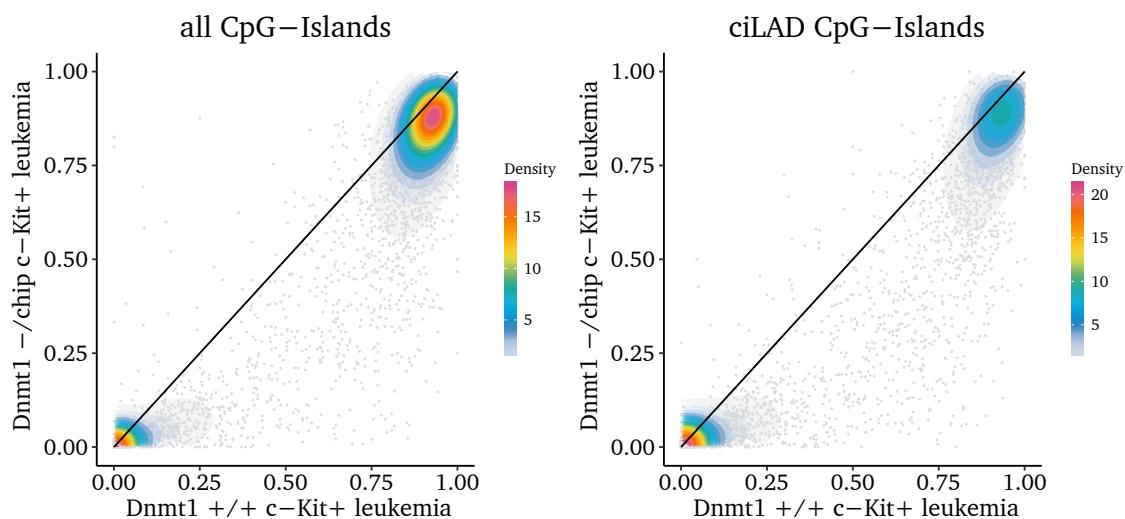


Figure 2.5: Either the CpG entirety or a the ciLAD subset of CpGs has been mapped on a CpG-Island reference and the difference between the $Dnmt1^{-/chip}$ and $Dnmt1^{+/+}$ leukemia is visualized as dotplot.

CpG-Islands, genomic areas with an unusually high frequency of CG-dinucleotide base pairs, are involved in transcriptional regulation and known to be aberrantly methylated in cancer [43, 133]. Furthermore it has been shown that their methylation level is regulated separately from the baseline methylscore of the surrounding sequence in cancer [134]. Therefore, we addressed the methylation level of CpG-Islands in particular by mapping the data on CpG-Island coordinates obtained from <http://www.haowulab.org/software/makeCGI/index.html> [135].

Remarkably, the methylation of CpG-Islands (CGI) was largely unchanged in $Dnmt1^{-/chip}$ c-Kit⁺ versus $Dnmt1^{+/+}$ c-Kit⁺ leukemia [▷ [Figure 2.5](#)]. Just a few hundred of the 74 986 CGIs for the *NCBI37/mm9* reference genome changed their methylation status significantly and basically none was associated with a known promoter [▷ [data not shown](#)].

If a CGI was already unmethylated in the $Dnmt1^{+/+}$ sample, which we only observed in ciLAD space, impairment of faithful propagation by $Dnmt1$ insufficiency would obviously be of no consequence. However, also the highly methylated CGIs in the cLAD areas remained essentially untampered with, although they were situated in the actually more hypomethylating cLADs [▷ [Figure 2.6, left panel](#)].

It remained unknown, whether this resulted from a preferential recruitment of $Dnmt1$ to the CpG-Islands or a de novo methylation by the other methyltransferases and if it reflected a selective pressure. Since those CGIs were also highly methylated in $Dnmt1^{+/+}$ HSCs [▷ [Figure 2.6, right panel](#)], it seemed comprehensible that cells with widespread loss of methylation at these sites had been subject to negative selection.

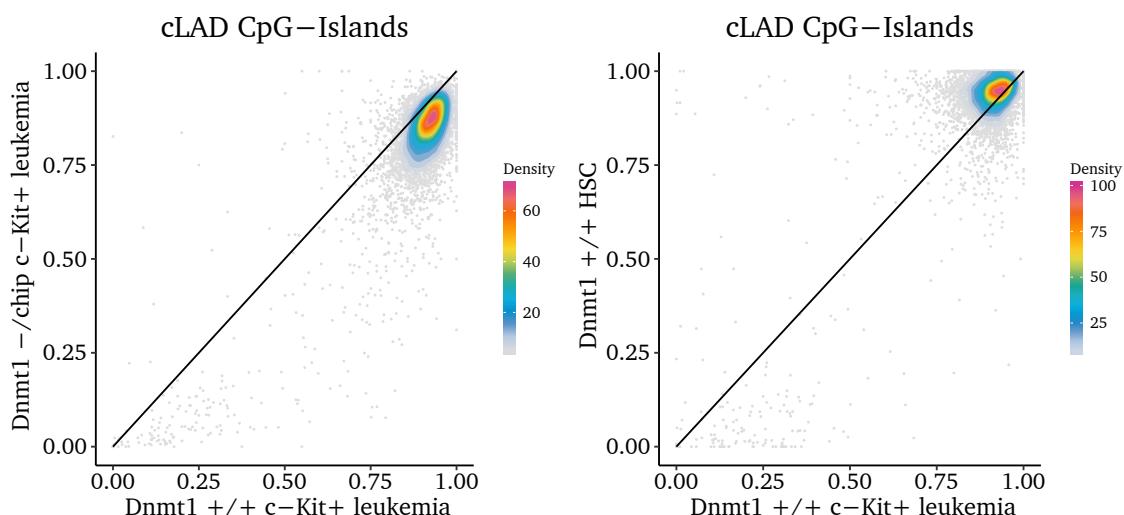


Figure 2.6: A selection of CpGs localized in annotated cLAD regions has been mapped on a CpG-Island reference. Either the contrast $Dnmt1^{+/+}$ c-Kit⁺ vs. HSC or vs. $Dnmt1^{-/chip}$ c-Kit⁺ leukemia is shown.

Since our data confirmed a previous observation, namely that one had to consider CpG-Islands and the methylation of the backbone separately [134], we were now interested, how exclusion of the CGIs would affect our results. Therefore, we repeated the 100 kb sliding window analysis after prior exclusion of all CpGs within islands.

Without distorting CpG-Islands, the large-scale trends in the data became more obvious [▷ [Figure 2.7](#) vs. [Figure 2.2](#)]. While $Dnmt1^{+/+}$ c-Kit⁺ leukemia exhibited a homogeneous demethylation by roughly 10 %, a clear division into regions of better and worse methylation persistence emerged within the $Dnmt1^{-/chip}$ specimen. This bimodal distribution was mainly attributable to the ciLAD/cLAD fractions as we had seen before [▷ [Figure 2.3](#)]. The remaining 29 % flexible LADs (fLADs) appeared as regions of intermediate persistency [▷ [Figure 2.8, left panel](#)], which harbored mostly methylated and

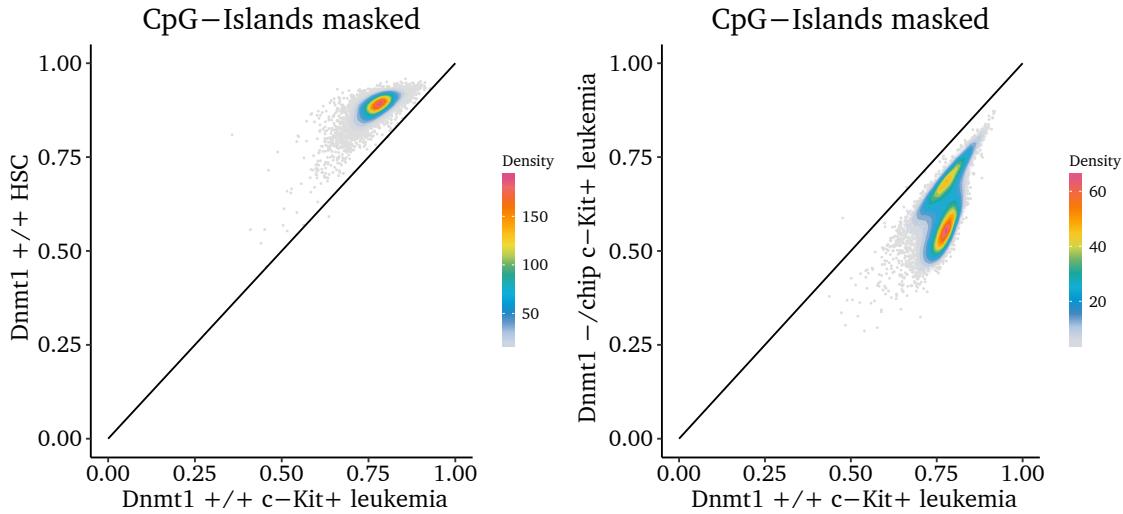


Figure 2.7: Mean methylation for 100 kb windows (slid by 25 kb steps) for all CpGs outside islands. Although some authors regard shores as transition zones influenced by the methylation level of the island [136], we considered shore-CpGs (2 kb margin surrounding an island) as backbone and counted them normally.

stable CpG-Islands [\triangleright Figure 2.8, right panel].

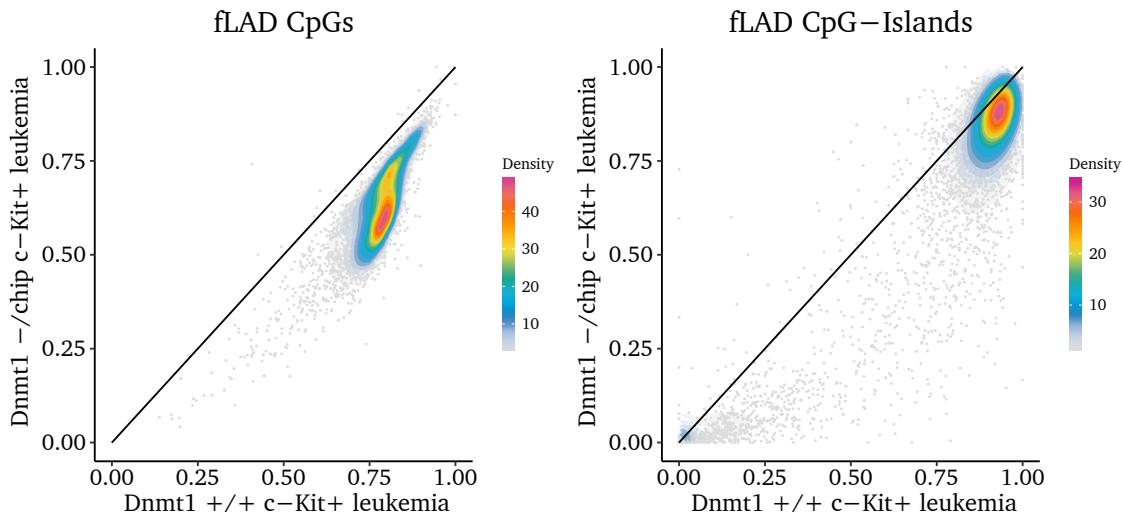


Figure 2.8: Methylscore average within 100 kb windows, sliding with a step size of 25 kb along the genome has been calculated under exclusive consideration of CpGs situated in flexible LADs.

Taken together, this analysis showed that also in MLL-AF9 leukemia, methylation in CpG-Islands and backbone is independently regulated. Therefore, we included this parameter during the GAM modeling [\hookrightarrow section 4.2, p.35]. Furthermore, we were puzzled by the surprisingly good methylation persistency at cLAD CpG-Islands in $Dnmt1^{-/-}$, which later urged us to test, whether they serve as recruitment platforms for $Dnmt1$ [\hookrightarrow section 5.2, p.41].

Chapter 3

Specification of the compromised regions

Contents

3.1 Demethylation at single CpG resolution	27
3.2 Increased partial methylation in leukemia	29
3.3 Standard approach failed to discriminate domain borders	31

In the previous chapter, it was shown that the genome of the $Dnmt1^{-/chip}$ c-Kit $^{+}$ leukemia could be subdivided into areas of variable methylation levels, which mostly associated with the underlying chromatin structure [↔ section 2.2, p.21]. Lamina-associated domains (LADs) seemed to demethylate to a larger extent, whereas open inter-LAD areas tended to preserve methylation better. From the perspective of methylation, we could therefore distinguish compromised and persistent areas in the genome of $Dnmt1^{-/chip}$.

The sliding 100 kb window analysis gave us a first idea of how the cells' methylation would change upon $Dnmt1$ insufficiency. Nevertheless, the accuracy fell short of our requirements in order to understand the implications on leukemia development and self-renewal. Thus, we sought to precisely map the borders of the compromised areas and to quantify the methylation persistency.

3.1 Demethylation at single CpG resolution

In a bottom-up approach we addressed the methylscores of the single CpGs first. One should be aware that methylation is intrinsically binary - a cytosine may be methylated or not. Intermediate methylscores like 0.75 thus essentially reflect an average rate across multiple cells or alleles.

On average, $Dnmt1^{-/chip}$ methylomes were globally hypomethylated by 10 % to 30 %. This meant, either a few CpGs within the 100 kb window changed their methylation status completely from methylated to unmethylated or many CpGs hypomethylated slightly. We detected a large number of slightly hypomethylated CpGs in $Dnmt1^{-/chip}$ c-Kit $^{+}$ leukemia, which suggested a rather random, passive loss of methylation across cell divisions due to insufficient maintenance. In contrast, the incidence of CpGs, which

changed from fully methylated to unmethylated was negligible [▷ [Figure 3.1](#)].

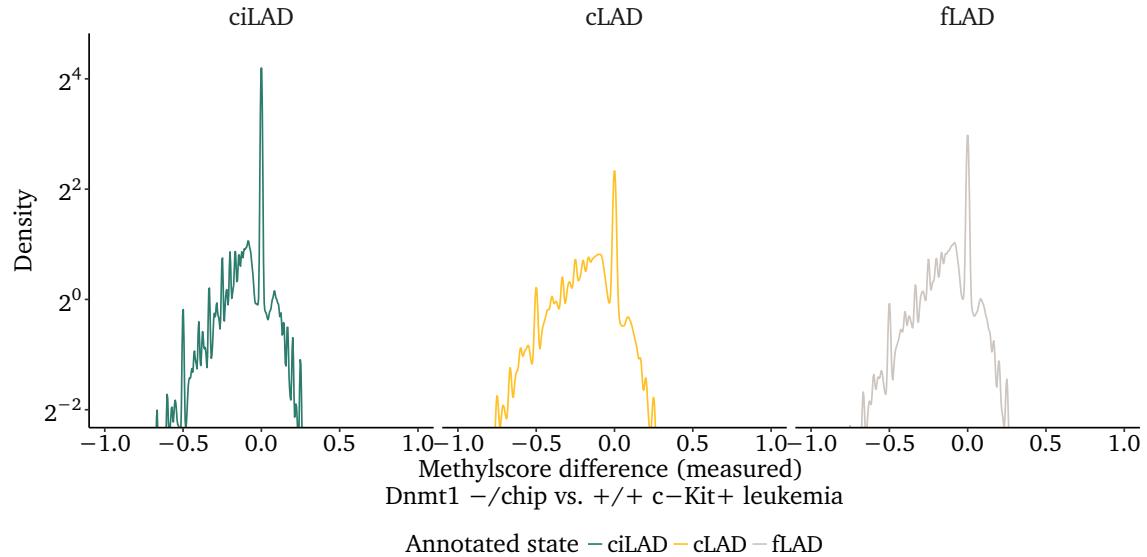


Figure 3.1: Kernel density estimates for methylation score differences on single CpG level, separated by the annotated chromatin state [131]. A high density implies that many CpGs exhibit a particular methylation score difference, which is calculated by subtraction of the Dnmt1 $-/-$ chip c-Kit $^{+}$ methylation score from those of Dnmt1 $+/+$ c-Kit $^{+}$ leukemia.

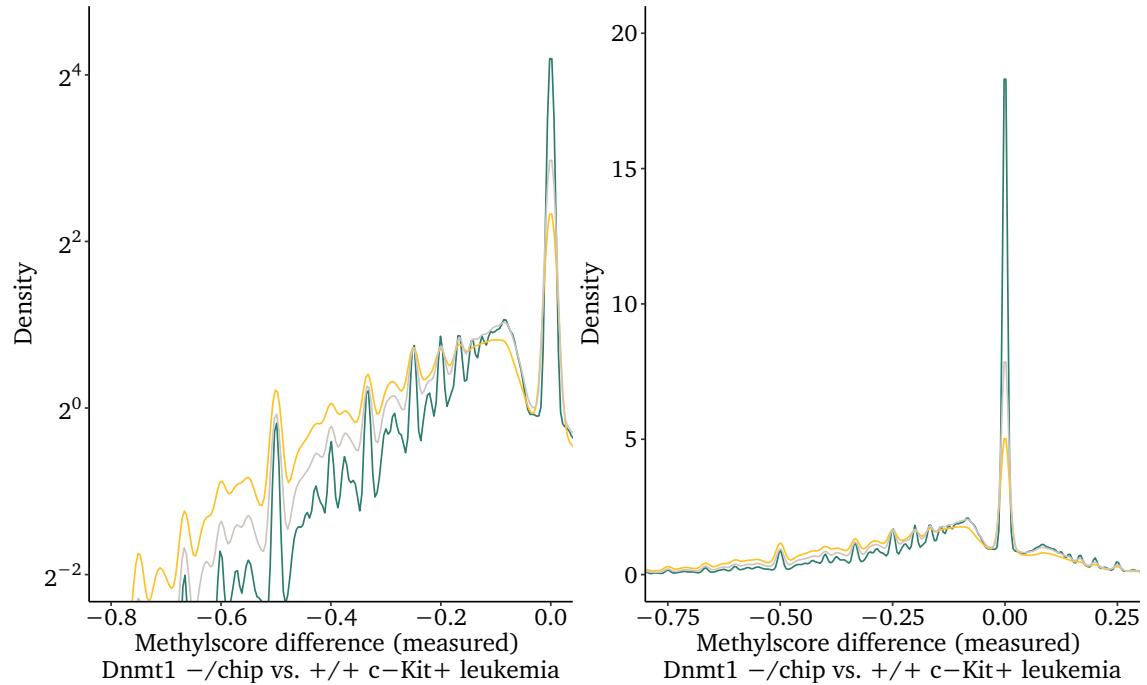


Figure 3.2: Overlay of the density estimates from [Figure 3.1](#) for better comparability. Mind the different scale (logarithmic vs. linear scale) in the two panels to aid visual inspection of different areas of the density estimates. The left panel focuses on the low-density from -0.8 to -0.2 and the right panel on the high density at 0 methylation score difference. Local maxima at $-0.5, -0.33, -0.25$ can be explained by the low cutoff: We permitted all CpGs with a coverage of 3 or more reads.

In terms of magnitude the annotated chromatin state was essentially irrelevant, the slightly greater demethylation in cLADs was only visible in the case of a log-scaled y-axis [▷ [Fig-](#)

ure 3.2, left panel]. The overlay of the density plots however also revealed that the absolute number of persistent CpGs dramatically decreased in fLADs and cLADs compared to ciLADs [▷ Figure 3.2, right panel].

Around the same time, a comprehensive WGBS analysis of human methylomes had suggested, that merely 22 % of CpGs are dynamically regulated and serve regulatory purposes [137]. In light of this study, our own results therefore pointed towards a lack of negative selection in lamina-associated areas. Since these genomic areas are commonly compacted as heterochromatin [138], the need to regulate e.g. transcription factor binding by methylation [139] is not present. Hence, a lack of negative selection was comprehensible.

3.2 Increased partial methylation in leukemia

The distinctive feature between ciLAD, fLAD and cLAD on single CpG resolution was not the magnitude of demethylation but the absolute number of hypomethylated CpGs [▷ Figure 3.2]. Therefore, also persistent and compromised regions mostly differed by the number of CpGs with partial methylation, namely methylscores which were elements of the open interval]0, 1[.

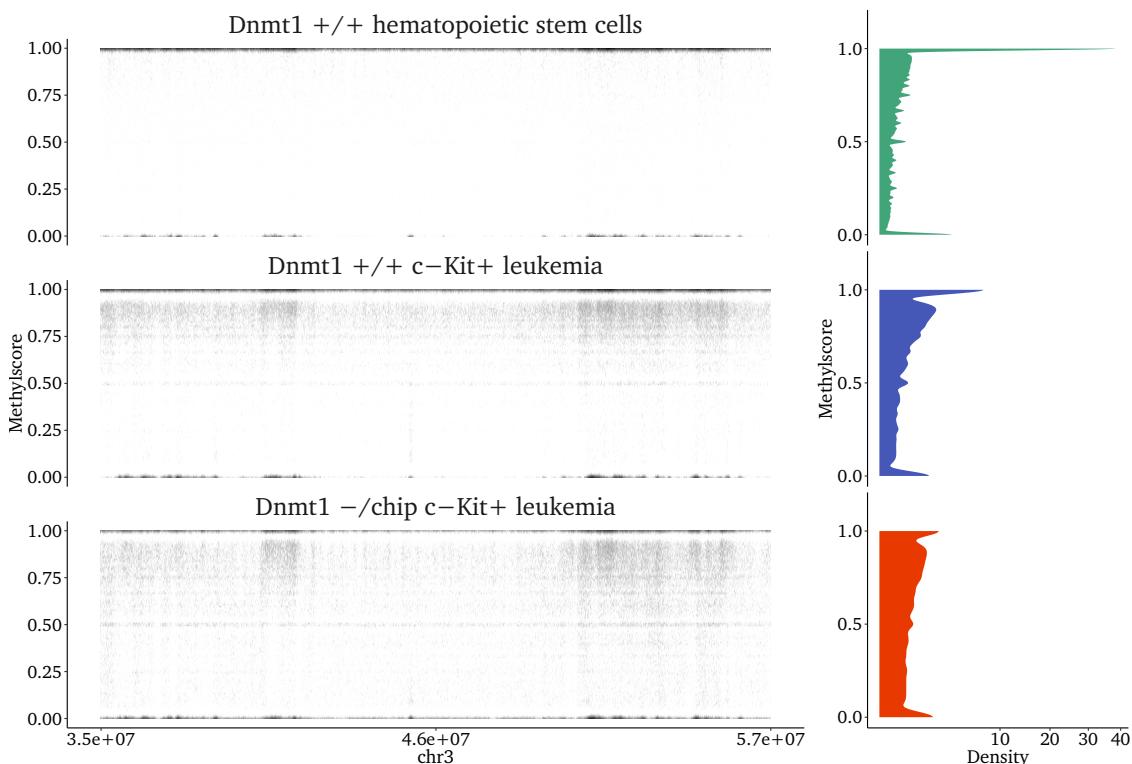


Figure 3.3: Methylation status in a 2.2×10^7 bp region of chromosome 3, the y-axis denotes the CpGs' methylation rate. To avoid overplotting, no single CpGs are shown. Instead tiles represent the underlying data - dark hues indicate a high CpG-density. The overall frequency for the respective percentage values is summarized by density estimates on the right - mind the square-root transformation of the scale.

Although we could confirm this claim, we also measured a surprisingly high amount

of partial methylation in $Dnmt1^{+/+}$ c-Kit $^+$ leukemia. [▷ [Figure 3.3](#)]. In contrast to the published methylome of $Dnmt1^{+/+}$ hematopoietic stem cells [125], both c-Kit $^+$ leukemia methylomes were heavily skewed towards arbitrary partial methylation. A square-root transformation was required to represent the density estimates of the three samples at scale, otherwise it would not have been possible to fit the unambiguously methylated CpGs in the HSC methylome, which by far outnumber the relative amount of such CpGs in the leukemia samples. The hypomethylation, but not the increase in partial methylation, had been described in MLL-AF9 AML before [reviewed in [140](#)].

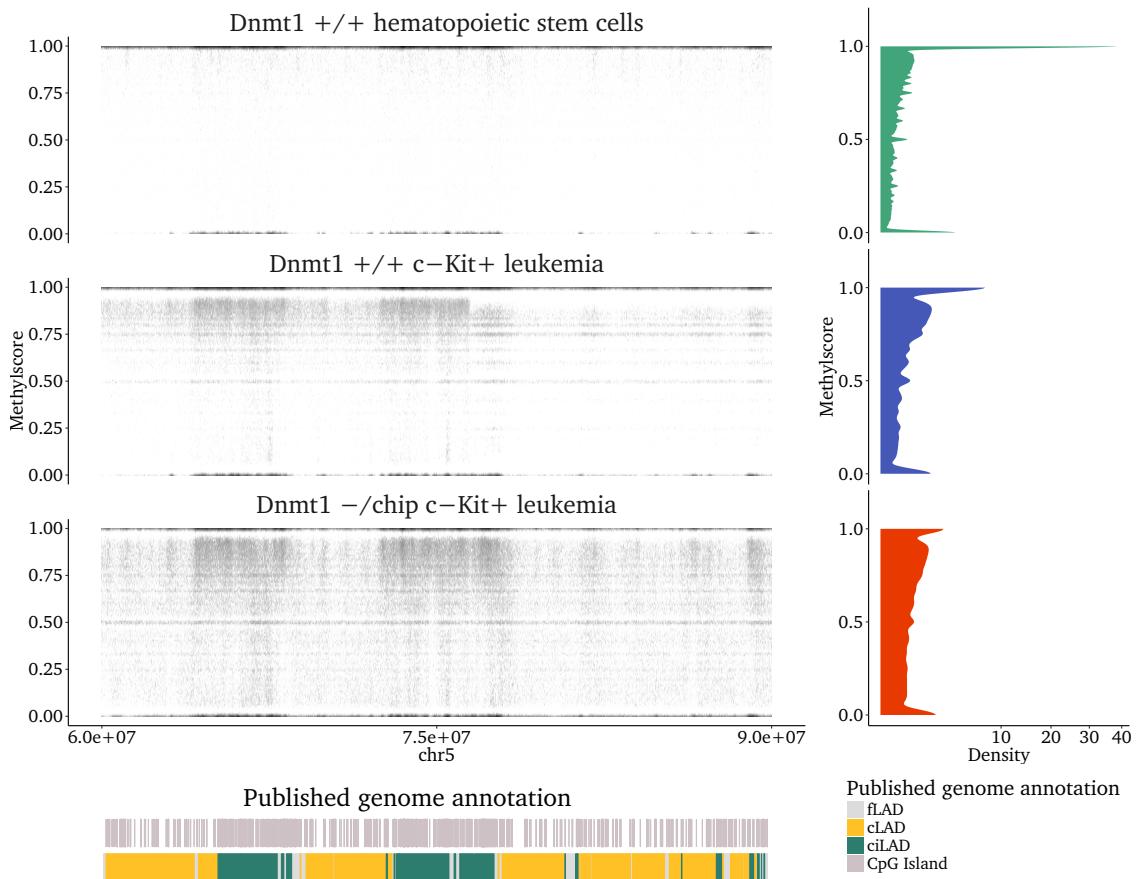


Figure 3.4: Methylation status in a 3×10^7 bp region of chromosome 5. The location of each CpG is determined by its chromosomal position (on x-axis) and methylation rate (on y-axis). To avoid overplotting, tiling was applied such that dark color indicates a high CpG density. Colored blocks below indicate the extent of chromatin or sequence features on the underlying DNA. Overall frequency for the methylation percentage values for the full set of CpGs is summarized by density estimates on the right.

Despite the dramatic, genome-wide increase in partial methylation, it was possible to visually spot areas of higher and lower methylation persistence in leukemia across large, megabase-spanning domains. In some genomic sections the amount of almost fully methylated CpGs (in the range of 0.70 to 0.95) considerably surpassed those of neighboring areas and was reflected in visually darker zones in the figures [▷ [Figure 3.3](#), [Figure 3.4](#)]. Integration with annotated chromatin states [131] corroborated that these darker zones were in coincidence with ciLADs (shown in dark green) and therefore argued for a higher

methylation persistence in open genomic regions. Evidently, the density of well maintained CpG-Islands [\leftrightarrow subsection 2.2.2, p.23] was higher in ciLAD areas [\triangleright Figure 3.4], which may have contributed to the overall better persistency in these areas. However, also when the CGIs had been excluded beforehand from the analysis, the heterogeneity persisted [\triangleright Figure 2.7, p.25].

Given the association of methylation persistence and lamina-association, we proposed that it would be conversely possible to infer the chromatin organization from the observed methylation pattern. Therefore, we aimed to precisely delineate persistent and compromised regions to derive insights into the MLL-AF9 chromatin with the help of the Dnmt1 $^{-/chip}$ mouse.

3.3 Standard approach failed to discriminate domain borders

Shortly after the first comprehensive WGBS datasets became available, a feature termed *Partially Methylated Domain* (PMD) was described [141]: Some methylomes contained large (>150 kb) regions of seemingly disordered methylation harboring many heterogeneously methylated CpGs. Because of the evident similarity of compromised regions and PMDs, we presumed the applicability of a published tool for PMD calling named METHYLSEEKR [142] to precisely determine the domain borders of the compromised regions.

However, we could not derive meaningful and verifiable limits for the compromised regions with this approach. By reverse engineering of the method, we could ascertain why METHYLSEEKR performed badly on Dnmt1 $^{-/chip}$ methylomes [\triangleright supplement]. Briefly, fitting of the mixture model failed, because the resulting density over α for the beta distribution fits was unimodal for the Dnmt1 $^{-/chip}$ methylome. Normally, the density assumes a bimodal distribution split between open and lamina-associated chromatin for methylomes with PMDs, which was illustrated by the control (IMR90 fibroblasts). Subsequently, we tried to replace the beta distribution with the Kumaraswamy distribution, a beta-type distribution with more convenient tractability [143, 144], however, it did not improve the fits.

In conclusion, the compromised regions in Dnmt1 $^{-/chip}$ mice were either very weak PMDs or a distinct feature.



- For this section, supplementary text and figures are available.

Text abridged. Supplementary online information is available.

Chapter 4

Modeling the methylation probability

Contents

4.1 A GAM to predict the methylation rate	33
4.1.1 Introduction to GAM models	33
4.1.2 Reasons to choose a GAM	34
4.2 Fitting a GAM on methylome data	35

4.1 A GAM to predict the methylation rate

To delineate compromised from persistent regions and infer the chromatin structure, we unsuccessfully attempted to use a standard tool [→ [section 3.3, p.31](#)]. Other existing tools mostly focused on CpG-Islands [136, 145, 146] or were not applicable for technical reasons [147]. Hence, we decided to develop an own solution tailored to our needs and data and opted for a *generalized additive model* (GAM).

4.1.1 Introduction to GAM models

The basic assumption of any type of model is that the variable y to be modeled (the response) depends on some combination of other observed variables $x_1, x_2, x_3, \dots, x_p$ (the predictors) and may additionally be confounded by random effects ϵ such as measurement errors.

In the simplest form, the relationship is linear, so the predictors are scalars, which are multiplied with the parameters α to determine the response [▷ [Equation 4.1](#)]. This implies that a constant change in a predictor leads to a constant change in the response variable. Such a model is called *general linear model* and is quite simple, but also not suitable for many types of data. For example it cannot be used to model probabilities, which are bounded on both ends (they are between 0 and 1).

$$\text{LM: } y_i = \alpha + \alpha x_{i1} + \alpha x_{i2} + \dots + \alpha x_{ip} + \epsilon_i \quad (4.1)$$

$$\text{GLM: } g(E(y_i)) = \alpha + \alpha x_{i1} + \alpha x_{i2} + \dots + \alpha x_{ip} + \epsilon_i \quad (4.2)$$

$$\text{GAM: } g(E(y_i)) = \alpha + s_1(x_{i1}) + s_2(x_{i2}) + \dots + s_p(x_{ip}) + \epsilon_i \quad (4.3)$$

However, many additional cases can be covered by the introduction of a link function $g()$ to the term, which is then referred to as *generalized linear model* [▷ [Equation 4.2](#)]. Depending on this function, the response does not need to be continuous and numeric, but could for example also be a categorical yes/no. In this exemplary case, the linear combination of the predictors determines the odds and the categorical response is derived by the link function (e.g. logit or probit).

Many types of link functions can be used and often such models have separate names like *logistic model*, but are technically generalized linear models - in the latter case with the logit link function. Also the beta function, which is often used to model methylome data [[136](#), [148](#), [149](#)], can serve as a link function, which then specifies a beta regression model.

Despite the versatility of GLMs, these models assume that ultimately linear predictors with unknown coefficients exist. However, there might be situations, where the model needs to incorporate categorical input or such that is ordinal. In this case, the predictors themselves need to be connected via linking functions. It could also be, that a predictor needs to be smoothed before it can be used in the model. Both is assured by the introduction of (mostly) nonparametric functions $s_1(\dots) + s_2(\dots) + \dots + s_p(\dots)$, which wrap the predictors [▷ [Equation 4.3](#)] - such a model is then called *generalized additive model* (GAM). A large variety of predictor link functions can be used, which gives GAMs an unprecedented flexibility.

4.1.2 Reasons to choose a GAM

The choice to fit a GAM on our methylome data was mainly guided by two reasons:

Firstly, we needed a method that would bring smoothing capabilities as neighboring CpGs often exhibited sudden shifts, which presumably could be explained in part by low coverage: Given a CpG covered by four reads next to one by three reads, the methylscore of the latter was $f_2 \in \{0, 0.33, 0.66, 1\}$, those of the previous one $f_1 \in \{0, 0.25, 0.5, 0.75, 1\}$. Thus, if both are partially methylated CpGs, they change by >10% due to the different coverage alone. We detected such a relevant fraction of CpGs with low coverage that we could not really premise a continuous scale everywhere, but encountered chromosomal regions, where methylscores rather formed an interval scale [▷ [Figure 3.2, p.28, left panel](#)]. With GAMs, we could nevertheless impose the prior belief that the methylation persistency is inherently smooth in nature and varies only across larger sections of the chromosome. By using a spline $s(\dots)$, as predictor link function and adjusting the level of smoothness, we could effectively deal with noisy measurement data.

Secondly, we (and others [134]) had previously identified CpG-Islands (CGIs) as regions of higher methylation persistency, which may strongly divert from the methylation level of the backbone [\leftrightarrow subsection 2.2.2, p.23]. Therefore, the model needed to account for situations, where a sharp change in the response variable was observed so called “hockey sticks”. It was possible to incorporate those easily, since GAMs can capture common nonlinear patterns that a classic linear parametric regression model would miss. Using a GLM, we would have required binning or extensive use of polynomials to account for such spikes.

Taken together, although the fitting of GAMs demands a high computational cost due to their flexibility, it was a reasonable choice for our complex methylome data.

4.2 Fitting a GAM on methylome data

For fitting the GAM, we considered only CpGs, which were covered with ≥ 4 reads in the respective meta-samples (roughly 11×10^6). All three meta-samples (Dnmt1^{+/+} HSC, Dnmt1^{+/+} c-Kit⁺, Dnmt1^{-/-chip} c-Kit⁺) were smoothed separately. We split the chromosomes into equally-sized fragments of 10 Mbp to allow parallelization of the computation. If a fragment contained more than 5×10^4 CpGs, a random subset was drawn, otherwise all available data was used to fit the GAM.

We relied on the *mgcv*-package [150] in R [151] to fit the Generalized Additive Model (GAM) and optimized its parameters with the restricted maximum likelihood (REML) approach. The GAM to be fitted was restricted to the beta regression family and we used the logit as link function. The model’s formula is given below:

```
library("mgcv")
gam(y ~ s(xi, factor(isCGI), bs="fs", k=n), family=betar(link="logit"), data=...)
```

The response y , equivalent to the probability of a particular site to be methylated, was modeled based on the measured methylation rate x at a given position i , smoothed with a thin plate regression spline [152]. To take persistent CpG-Islands into account, we classified CpGs into the two categories *CGI* or *nonCGI* prior to fitting and smoothing. However, factor smooth interaction fs , which produced a smooth for each level of a single factor variable, ensured the relevant global model parameters were used for both.

We tested the inclusion of other predictors at randomly chosen chromosomal sections, but did not achieve better models. Better in this regard refers to relevant changes (underpinned by the measurement) at the desired level of considerable smoothness. Hence, our finding that these predictors were irrelevant for us, did not contradict published results that they influence local methylation rate at a higher resolution [137, 139, 153]. Furthermore we did not discriminate between expressed and non-expressed transcripts in the predictors, which may have had an impact. The subsequent predictors were tested in the

model term, but ultimately dismissed due to negligible relevance:

- Inside/outside of the promoter region of an annotated transcript.
- Inside/outside an ENCODE regulatory region.
- Distance in base pairs of the CpG to the next referenced transcription start site (TSS).
- Distance in base pairs of the CpG to the next CpG-Island.

We applied the package's default, the restricted maximum likelihood (REML) optimizer for parameter estimation. To gain computational efficiency, we manually set an upper limit on the degrees of freedom k for the smoother. This limit n was chosen relative to the length of the genomic fragment, increasing by one for every 1×10^5 base pairs from a minimum of 4. The actual effective degrees of freedom were controlled by the degree of penalization selected during fitting REML, but the smoother gained computational efficiency by setting a modest upper limit. We ensured by random sampling from our fitted models that the upper limit permitted enough degrees of freedom to represent the underlying truth reasonably well.

Although we herein refer to the modeled response as methylation probability, it is technically a cytosine methylation probability. Even though the purine bases can also be methylated (adenine in bacterial DNA, guanine by chemical mutagenesis), the likelihood of this incidence cannot be modeled from the predictors we used. Thus, if the methylation probability is given for a non-cytosine site, it is a mere mathematical assumption, how likely a methylation at a cytosine at this site would be if present.

The modeling permitted to assign a methylation probability to any site of the genome independently of the actual underlying sequence. Overlay of the samples' methylation probabilities showed that $Dnmt1^{+/+}$ HSC exhibit a constant level of $\approx 90\%$ methylation in backbone, whereas those of the leukemia was lower [\triangleright Figure 4.1]. As we had seen before [\leftarrow section 2.2, p.21], methylation persistency was lower in lamina-associated, heterochromatin areas than in open chromatin.

Subsequently, we validated the model and judged the performance of the fitted GAM against a series of comparisons with 500 bp sliding windows [\triangleright supplement]. We show that the model corresponds well to measurement data. It was generally representative in case the windows comprised 5 to 10 (depending on the sample) or more CpGs [\triangleright supplement].

It outperformed existing tools for this task and provided us with a deeper understanding of the observed methylation. Because of the model's smoothness, which did not capture local deviations, it is however not applicable to small genomic areas, such as single cis-regulatory elements, unless they are also incorporated as separate factor (like the CGIs).

In summary, our custom developed GAM allowed us to derive a quantitative measure of methylation persistency despite challenging data and to account for the difference between CpG-Islands and backbone. By subtracting the backbone response of the $Dnmt1^{-/chip}$

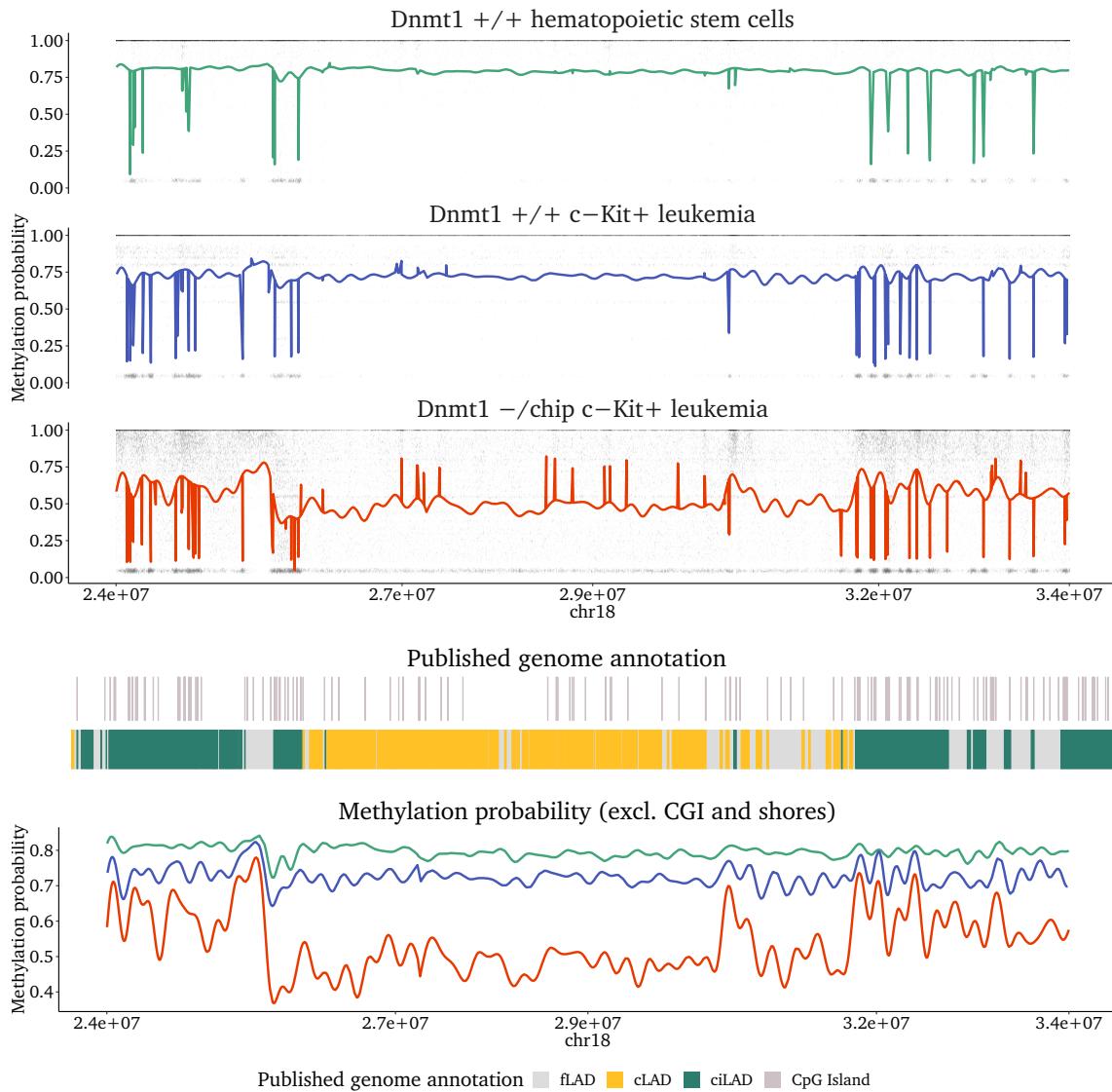


Figure 4.1: Modeled methylation probability (colored lines) and measured methylation rate of single CpGs in a 1×10^7 bp region on chromosome 18. To avoid overplotting, tiling was applied such that dark color indicates a high CpG density. Colored blocks below indicate the extent of chromatin or sequence features on the underlying DNA. Mind the sharp changes in methylation probability at sites of CpG-Islands. When the CGIs are not shown, the drop in backbone methylation probability for *Dnmt1*^{-/-}/chip c-Kit⁺ at cLAD regions is evident in overlay.

model from that of the *Dnmt1*^{+/+} model, we could also clearly distinguish compromised from persistent areas [▷ Figure 4.2]. This allowed us to address potential biological implications.

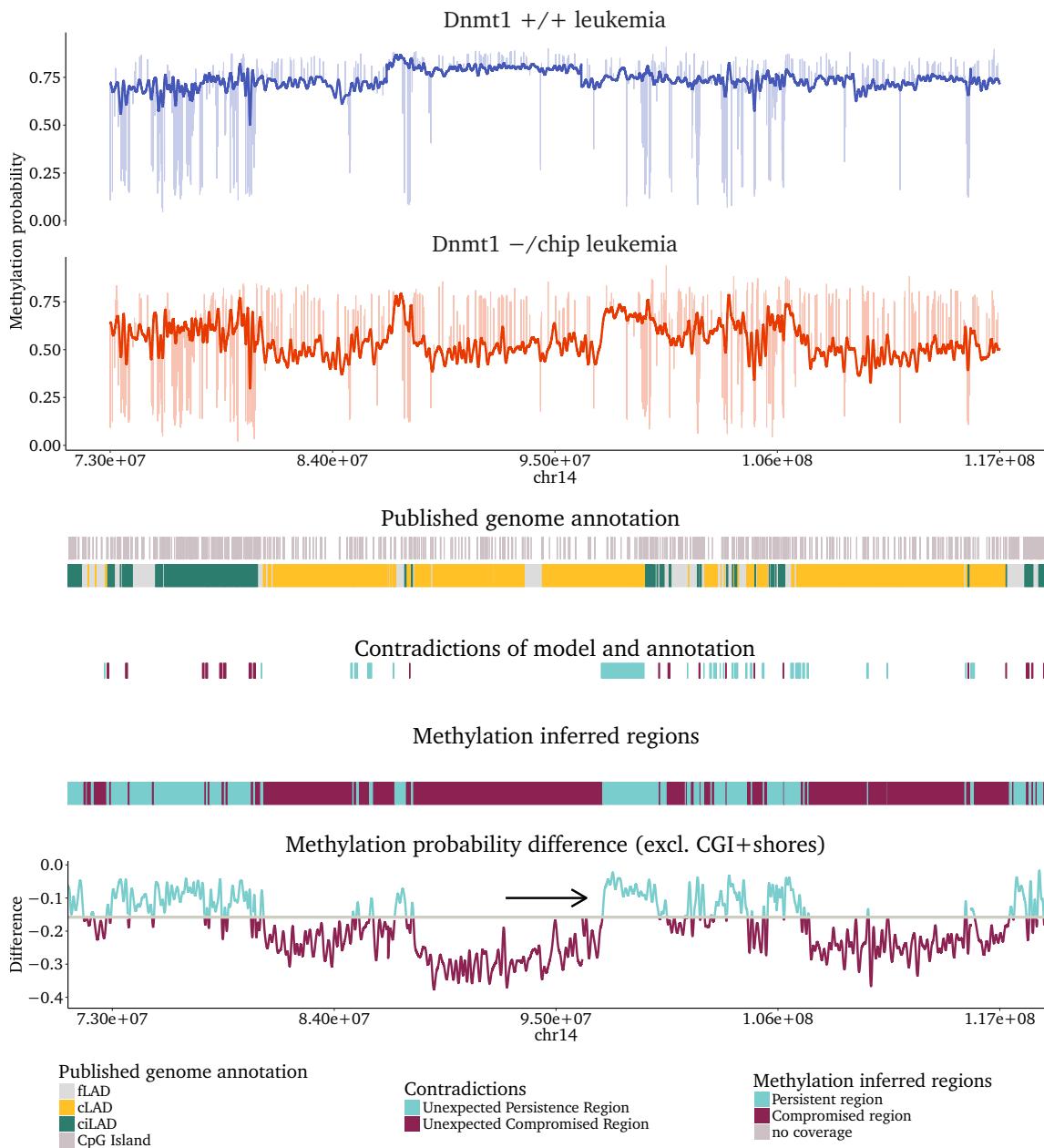


Figure 4.2: Another 4 × 10⁷ bp chromosomal example region. Shown with colored lines is the modeled methylation probability of the backbone (saturated) and CpG-Islands (pastel). Open chromatin ciLAD regions clearly associate with a higher backbone methylation in Dnmt1^{-/-}/chip c-Kit⁺ and thus with persistent regions represented in turquoise. The few areas, where the rule is broken, are considered unexpected and indicated by a separate layer of blocks. A black arrow points at a shifted transition between persistent and compromised regions with regard to the annotated lamina association.

Chapter 5

Relationship of chromatin structure and methylation persistency

Contents

5.1 Chromosomal insulation and interaction	39
5.2 Determining factors of methylation persistency	41
5.3 Summary	44

On the grounds of the association of lamina-association with methylation persistence, we aimed at conversely inferring the chromatin organization from the observed methylation pattern. After we had developed the GAM to precisely delineate persistent and compromised regions, we could observe some inconsistencies with the constitutive annotation [▷ [Figure 4.2, facing page](#)].

We referred to those inconsistencies descriptively as *unexpected*, although it was comprehensible that the annotation was not fully representative of the leukemia situation. Back-testing with measured data proved that the unexpected regions were not incorrect predictions of the model. Instead, we truly observed either small, divergent regions inside contrary larger homogeneous annotations or cases, where the transition between a compromised and a persistent region was shifted relative to those of the annotated lamina-association. [Figure 4.2](#) depicts such an unexpected region of the latter type roughly at 10×10^7 bp, where a large persistent region extends far into a lamina-associated cLAD region (marked by a black arrow).

5.1 Chromosomal insulation and interaction

Acknowledging that the ciLAD/cLAD annotation might not fully correspond with MLL-AF9 leukemia chromatin, we analyzed other third-party data sets for comparison. Since matched methylation/chromatin data had been published for human IMR90 fibroblasts, which were known to exhibit PMDs when senescent [154], we repeated the GAM fitting on this dataset. None the less, a commensurate amount of unexpected regions could be identified [▷ [data not shown](#)].

The presence of unexpected deviant regions in both datasets suggested either a technical limitation or influences other than the chromatin structure, which challenged a purely passive model. To better resolve the chromatin structure and improve comparability, we abandoned the concept of discrete states / regions in favor of a continuous measure.

The insulation score, which can be calculated from Hi-C chromatin interaction data [155], was such a continuous measure of chromatin interaction. We considered the interactivity measured by Hi-C to be a proxy of the openness of the genomic regions, as heterochromatic areas are condensed and typically do not interact dynamically with other chromosomal regions, however they do aggregate with other heterochromatin.

A Hi-C dataset, which seemed promising and applicable our MLL-AF9 c-Kit⁺ leukemia model, was that of the HPC-7 murine blood stem/progenitor cell model [156]. The cell line was used to generate comprehensive binding profiles of key transcription factors [157, 158] and a derivative in-silico model faithfully recapitulates early hematopoiesis as well as its perturbation by leukemogenic TF fusion proteins [159]. Thus, the HPC-7 data likely represented the MLL-AF9 leukemia chromatin much better than the ciLAD/cLAD annotation, which we used before.

We downloaded the aligned reads from the ARRAY EXPRESS repository [[Appendix A, p.134](#)] and proceeded with the software HOMER for analysis: We built tag directories, ran quality control checks and created models of background noise. Ultimately, interaction matrices containing normalized counts at 10 kb resolution were generated, which served as input to TADTOOL [160] for calculating the insulation index/score [155].

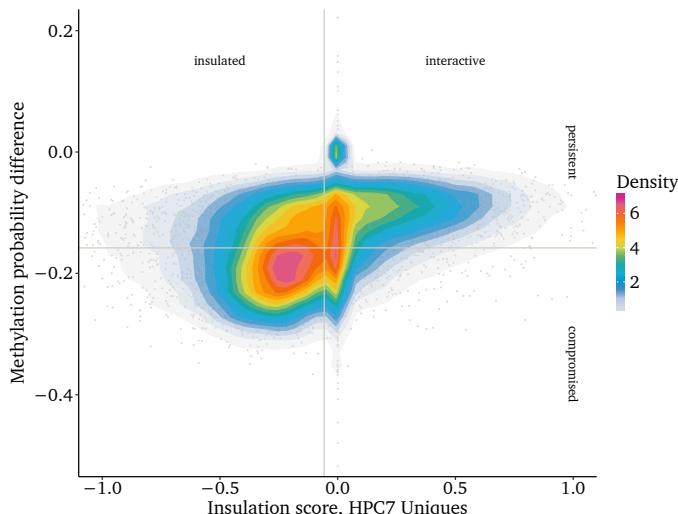


Figure 5.1: Density plots depicting the 2D density of the insulation score vs. the GAM-modeled methylation probability difference between Dnmt1^{-/-} and Dnmt1^{+/+} c-Kit⁺ leukemic cells averaged over the 10 kb windows of the Hi-C dataset. The lines indicate the respective cut offs used for categorical analysis.

At large, compromised regions were typically also insulating, while such that we found to be interacting were very rare. On the other hand among the persistent regions we observed a wide range of possible insulation scores (from very insulating to very interactive) [▷ [Figure 5.1](#)]. It was not yet conclusively shown at the time, but in the meanwhile

it is well established that housekeeping genes or super-enhancers commonly possess an insulating function [161, 162]. This explains, why we observed many insulating, yet persistent areas. An example of such an insulating housekeeping gene was the highly expressed catalytic subunit of the phosphatidylinositol 3-kinase (Pik3c3), which exhibited a demethylated promoter, but otherwise resided in a methylation persistent region on chromosome 18 [▷ Figure 5.2]. A small separate population was formed by CpG-Island rich areas, which exhibited intermediate insulation scores, but were highly persistent [▷ Figure 5.1, small separate population].

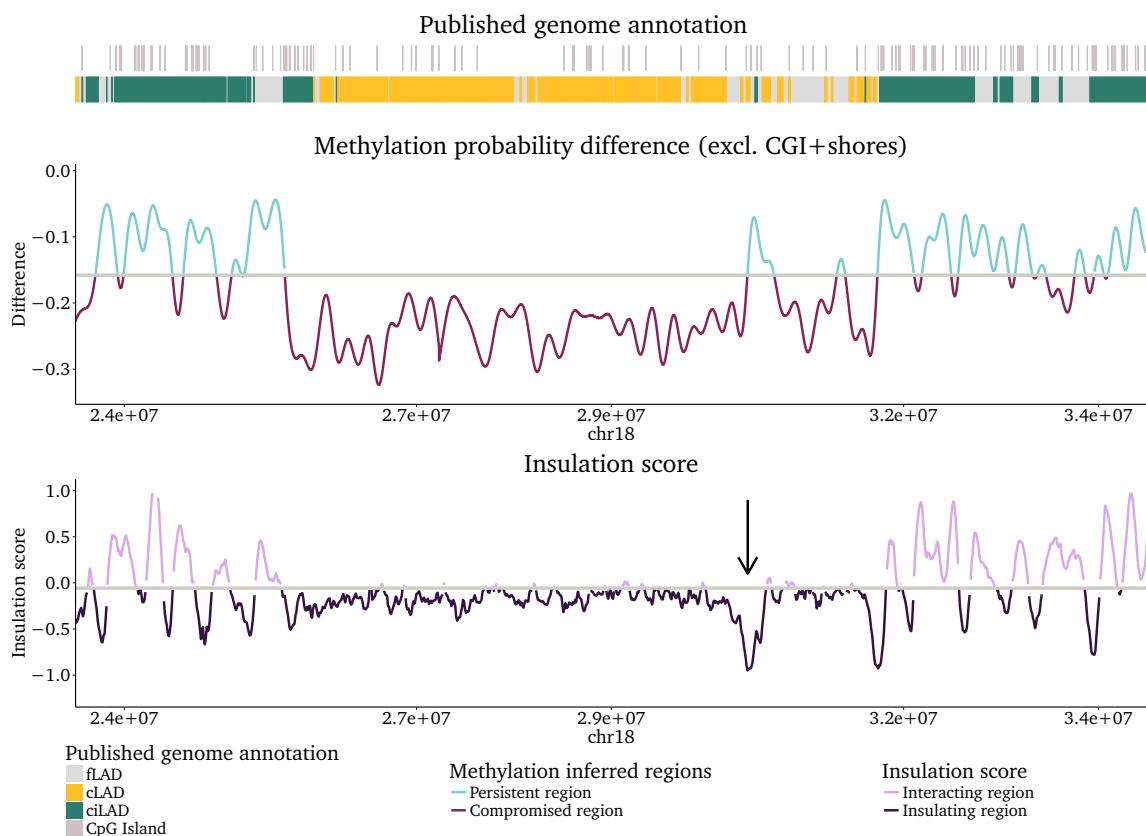


Figure 5.2: GAM-derived methylation probability difference, HPC-7 insulation score and chromatin annotation of a 1×10^7 bp region on chromosome 18. Colored blocks on top indicate the extent of chromatin or sequence features on the underlying DNA (ciLADs, cLADs and CGIs). The middle curve retraces the difference between Dnmt1 $^{-/-}$ and Dnmt1 $^{+/+}$ c-Kit $^+$ leukemic methylation probability, which was used to infer the persistent (turquoise) and compromised region (tyrian purple) categories. The lowermost curve shows the insulation score at 10 kb resolution, also subdivided into insulating (dark pink) and interacting (light pink) areas. A black arrow indicates a rare persistent, but insulating region surrounding the housekeeping gene Pik3c3.

5.2 Determining factors of methylation persistency

On the grounds of this small subpopulation [▷ Figure 5.1], it was tempting to speculate that CpG-Islands might preferably recruit Dnmt1 and thus serve as maintenance initiation centers. While we had ruled out an influence of the surrounding sequence on the

CpG-Island maintenance, we presumed that Dnmt1 and its accessory proteins might assemble as functional complex at the CGI and subsequently extent their activity to the vicinity. This notion was further supported by the observation that the number of CGIs in persistent regions by far surpassed those in compromised areas.

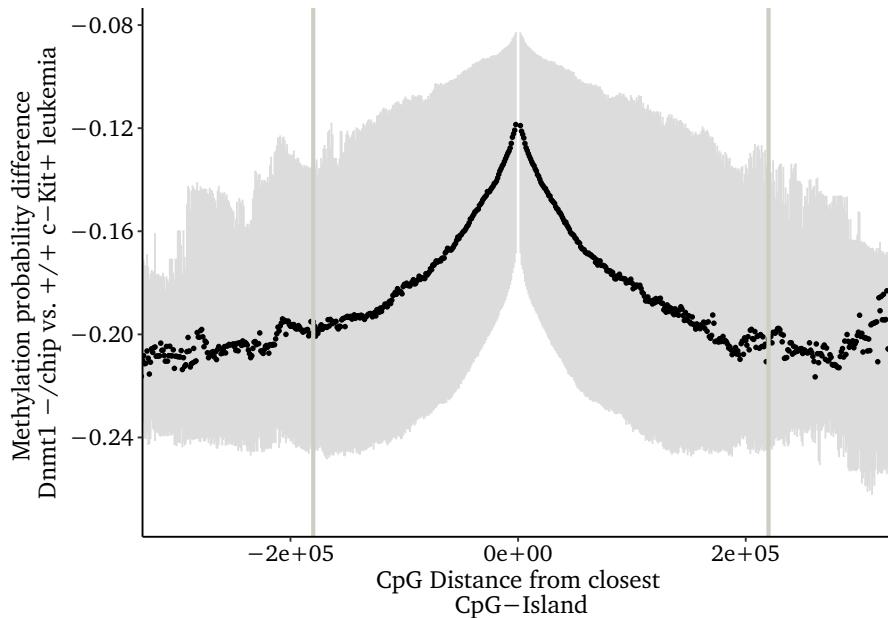


Figure 5.3: Change of the median methylation persistency in $Dnmt1^{-/-} / chip$ vs. $Dnmt1^{+/+}$ c-Kit⁺ MLL-AF9 leukemic cells in relation to the distance of the respective CpG to the next CpG-Island. For this analysis we considered only methylated CpGs (Methylscore > 0.75 in $Dnmt1^{+/+}$). The black dots denote the median and the gray strip marks the interquartile range. Vertical gray lines indicate the range from –180 kb to 220 kb, which we deemed trustworthy based on the fractions shown in the left panel of [Figure 5.4](#).

To test the hypothesis, we conducted an analysis of the spatial persistence relative to the next CGI for methylated CpGs (Methylscore > 0.75 in $Dnmt1^{+/+}$ c-Kit⁺ MLL-AF9 leukemic cells). Indeed, a larger distance to the nearest CpG-Island was related to a decreased methylation probability in $Dnmt1^{-/-} / chip$ mice compared to their wild-type counterparts [> [Figure 5.3](#)]. In comparison to a CpG located directly adjacent to a CGI, the methylation probability of a distant CpG was reduced by up to 10 %. However, the interquartile range was quite notable and therefore a relevant fraction of CpGs diverted from the general pattern.

The inverse relationship of distance to the next methylated CpG-Island and methylation persistency [> [Figure 5.3](#)] supported the hypothesis that Dnmt1 after a cell division would copy methylation at CpG-Islands first and then extent its activity to the vicinity.

To scrutinize this result, we resorted once more to the chromatin categories and specifically to the *unexpected* regions. The density of CpG-Islands can be easily computed for a region and should correlate with the observed methylation persistency within a region. Beforehand, we reassured that the observed distant persistency bias was not an artifact grounded in an unequal consideration of CpGs from the respective categories. Given

the lower frequency of CpG-Islands in compromised regions [▷ [Figure 4.1 and others](#)], it seemed plausible that they would harbor the majority of distant CpGs shown in [Figure 5.3](#). However, we could reassure that the relative fraction of CpGs from each category was constant in a region spanning almost 200 kb in each direction from the closest methylated CGI [▷ [Figure 5.4, left panel](#)].

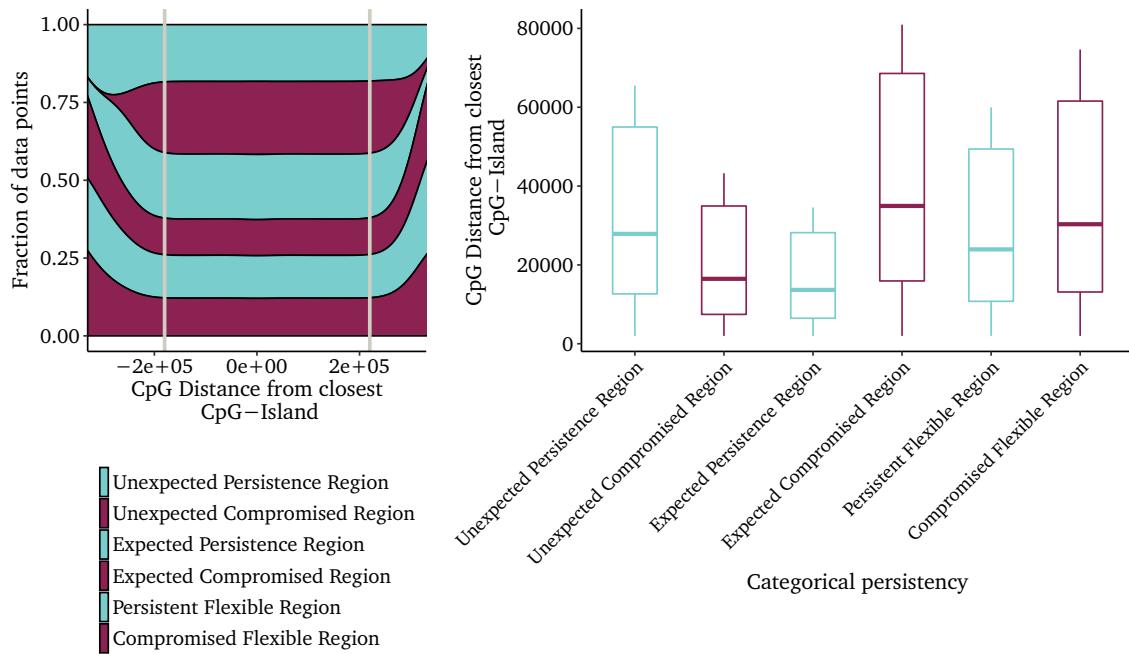


Figure 5.4: The left panel shows the fraction of methylated CpGs in the data used for [Figure 5.3](#), which fall into the respective methylation persistency category [↔ [chapter 5, p.39](#)]. As the relative shares remain almost constant for the -180 kb to 220 kb range, this scope is considered to be trustworthy and marked by vertical gray lines. The right panel depicts the typical distance to the next CGI for a methylated CpG inside each of those regions. In both plots the persistent areas are colored in turquoise and compromised regions in tyrian purple.

After having excluded a possible technical bias, we tested whether the density of CpG-Islands was indeed associated with the observed methylation persistency within a region. The expected compromised regions (ECR) harbored the fewest CGIs and the expected persistent regions (EPR) the most, which was reflected in the average distance of a CpG to the next methylated CGI [▷ [Figure 5.4, right panel](#)].

This model also suggested that the regions of unexpected persistency could potentially be attributable to a remarkable enrichment (UPR) respectively depletion (UCR) of CpG-Islands. However, we could clearly falsify this assumption based on the greater distance in UCRs than in UPRs [▷ [Figure 5.4, right panel](#)]. The gap to the next CGI was the second lowest in UCRs and yet they were compromised with regard to their methylation. The UPRs were almost at par with the compromised flexible regions (CFRs) with regard to the distance. Hence, we could clearly reject a purely passive model of methylation loss based on CpG-Island distance.

Subsequently, we also tested additional/other factors, such as the underlying DNA se-

quence, which had been suggested to determine the methylation within PMDs [163]. By performing an isochore analysis [164], we could confirm that the GC content was correlated with the degree of demethylation. The more CG-rich a sequence was, the lower was its tendency to demethylate [▷ [data not shown](#)]. However, particular flanking sequences or CpG positions were irrelevant for the methylation persistency, such that we could not corroborate the existence of specific sequence motifs beyond the mere CG-content.

Furthermore, the DNA sequence could also not explain the unexpected regions and neither did replication timing [165] [▷ [data not shown](#)]. It seemed that there was - at least in part - some active regulation involved in shaping the compromised regions in $Dnmt1^{-/chip}$.

5.3 Summary

In this chapter, we argue that neither CpG-Island density nor chromosomal insulation, the DNA sequence or replication timing can fully explain the methylation pattern of persistent and compromised regions observed in $Dnmt1^{-/chip}$. All of those factors except CGI density had - in conjunction with a lack of negative selection - previously been shown to exert some influence on DNA methylation.

However, it seemed that more factors contributed to the persistency, as none of the aforementioned ones could explain the deviations of chromatin structure and DNA methylation, which we referred to as *unexpected*. At least in those regions, the effect of passive demethylation must have been marginalized by additional factors shaping the methylome. Such additional factors could be the histone marks H3R2me2a, H3K9me3, H3K18ub or H3K23ub, whose importance for the recruitment of DNA-methyltransferases was recently shown [166–168].

Taken together, we could not fully unveil the epigenetic control mechanisms shaping the $Dnmt1^{-/chip}$ methylome. None the less, the unexpected regions might, in conjunction with ChIP-seqs for e.g. H3R2me2a point to sites harboring relevant genes for leukemia development and maintenance of the self-renewal program.

Chapter 6

Methylome analysis of matched non-malignant hematopoietic progenitors

Contents

6.1 WGBS data of the MPP hierarchy	45
6.2 Leukemia-related demethylation revisited	47

In the previous chapters our efforts to characterize specific methylome aberrations, which confer a cell-autonomous self-renewal deficit on $Dnmt1^{-/chip}$ c-Kit $^+$ leukemia, were detailed out. However, an essential cue was missing, as we did not have matched WGBS data from healthy $Dnmt1^{-/chip}$ or $Dnmt1^{+/+}$ hematopoietic stem/progenitor cell (HSPC) to compare the leukemia data to.

Since we had ascertained disordered lymphoid lineage development [116], it seemed likely that hypomethylation in part occurred already in the non-transformed HSPCs and was not strictly limited to the period after leukemic induction by MLL-AF9 transformation [\hookrightarrow section 12.1, p.110]. Therefore, it was highly desirable to separate the preexisting $Dnmt1^{-/chip}$ methylome alterations from those inflicted after transformation of HSPCs with MLL-AF9-IRES-GFP e.g. due to the much faster cell cycling and rapid expansion.

6.1 WGBS data of the MPP hierarchy

For the longest time of the project, we could not generate WGBS data because the state of the technology did not allow profiling of such small populations. Finally, a cooperation with Daniel Lipka provided us access to his tagmentation-based whole genome bisulfite sequencing (TWGBS) method [169] and permitted us to investigate even the earliest stages of hematopoietic development with some of the rarest populations of the hierarchy. Early stem and multipotent progenitor cells were separated by the canonical six surface makers (Sca1, c-Kit, CD135, CD48, CD150, and CD34) [170] into the four early populations HSC and MPP1 to MPP3.

Those populations are known to differ by the occupied niches and their balance between self-renewal and differentiation activity, which also implies a different cell cycling rate.

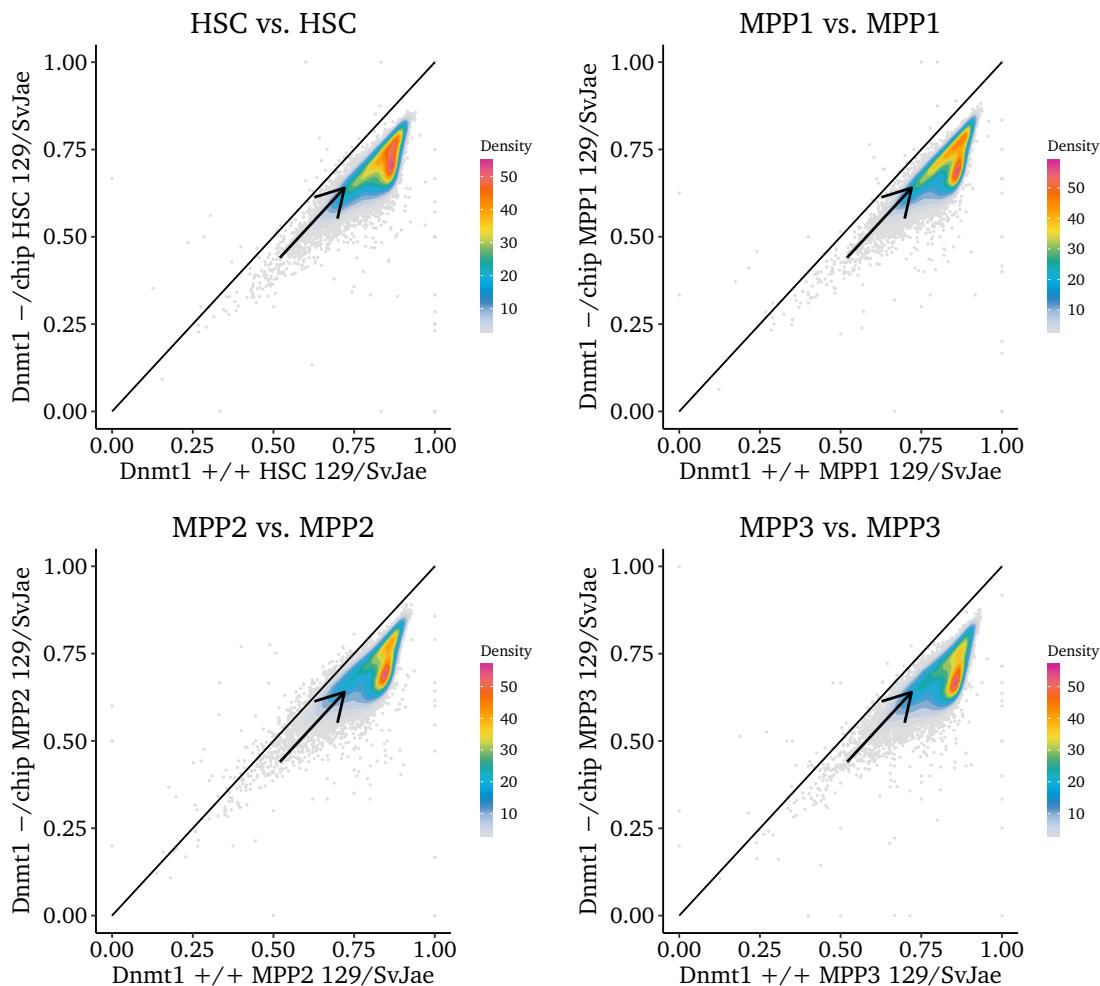


Figure 6.1: Scatterplot of the global methylation levels in the early hematopoietic hierarchy in $Dnmt1^{+/+}$ and $Dnmt1^{-/-}$ mouse strains. All CpGs have been mapped on 100 kb windows (slid by 25 kb) and the difference between the wild-type and hypomorphic hematopoietic stem/progenitor cells (HSPC) is visualized as dotplot. A black arrow indicates relatively persistent regions with roughly 75 % methylation in normal HSCs, whose frequency is gradually decreasing in MPPs.

Based on a presumed strong influence of passive methylation loss in $Dnmt1^{-/-}$, we were expecting to see an aggravated demethylation when the relatively dormant HSC gives way to the faster cycling MPP populations. However, a direct comparison between the matched populations of $Dnmt1^{-/-}$ and $Dnmt1^{+/+}$ [▷ Figure 6.1] showed that the majority of methylation loss had already occurred on HSC level. The $Dnmt1^{-/-}$ vs. $Dnmt1^{+/+}$ comparison in all four cell populations exhibited the bimodal pattern of persistent and compromised regions described previously for the leukemic cells [↔ section 2.1, p.20] [▷ Figure 4.2, p.38].

However, the degree of separation was variable and a particular group of relatively persistent regions (with roughly 75 % methylation in normal HSCs) seemed to abate upon differentiation [▷ Figure 6.1, black arrows]. Based on these plots it appeared possible that those sections gradually hypermethylated in wild-type MPPs, but failed to do so in $Dnmt1^{-/-}$. Yet, a direct comparison of the MPP3 population with the matched HSC

data showed only marginal changes at the 100 kb resolution [▷ [Figure 6.2](#)].

Initially, we had hypothesized that the formation of compromised regions was causally linked to senescence and cell cycle exit, which we had predominantly observed in leukemia the Dnmt1^{-/chip} background [↔ [section 1.4, p.13](#)]. However, HSCs from Dnmt1^{-/chip} were mostly negative in β -galactosidase (β -gal) staining. Thus, based on this finding, we had to challenge close ties between the presence of compromised regions and senescence. For a detailed discussion see [subsection 12.2.2](#) on page [114](#).

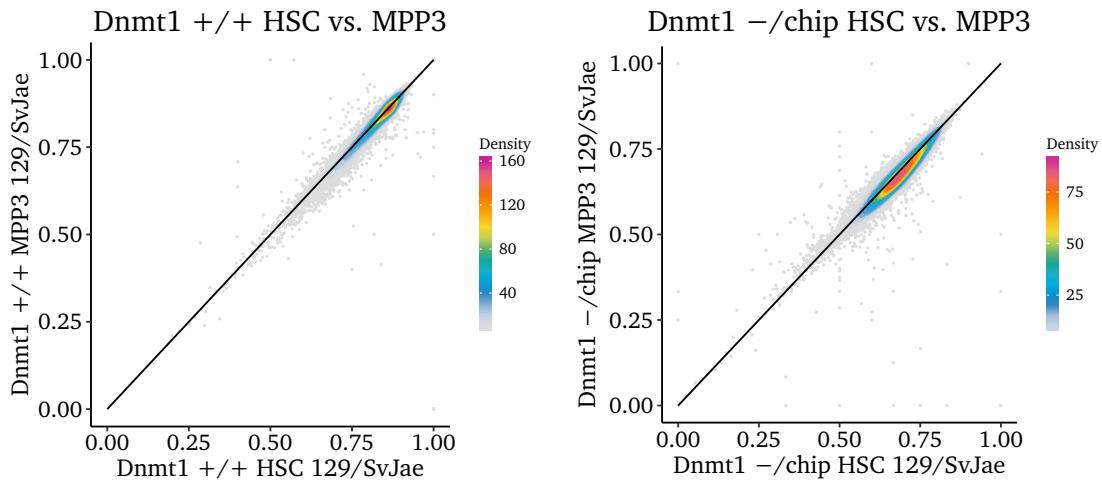


Figure 6.2: Scatterplot of the HSC/MPP3 contrast for Dnmt1^{+/+} (left panel) and Dnmt1^{-/-/chip} (right panel) hematopoietic progenitors. The CpGs have been mapped on 100 kb windows (slid by 25 kb steps) to generate these graphs of average methylation.

6.2 Leukemia-related demethylation revisited

The new data permitted us to revisit the extent of demethylation upon transformation by MLL-AF9 [↔ [section 2.1, p.20](#)]. For most of the project we were reliant on third-party methylome data of HSCs [↔ [Appendix A, p.134](#)], which was generated from C57BL/6 mice [125]. However, the new data showed that the HSCs of Dnmt1^{+/+} 129/SvJae were hypomethylated by approximately 10 % in comparison to those from C57BL/6 [▷ [supplement](#)]. Thus, we may have overestimated the methylation loss accompanying leukemic transformation by MLL-AF9 [▷ [Figure 2.1, p.20](#)], particularly in the case of Dnmt1^{+/+}.

By applying the formula $x_{c/chip} \leq f(x) = 0.7 \cdot \sin((x_{wt} + 0.26)^5)$, which we developed to separate compromised regions in a sliding window analysis, we surprisingly could also detect a few compromised regions in Dnmt1^{+/+}. A set of lowly methylated regions in leukemia (25 % to 45 % methylation) was 80 % methylated in Dnmt1^{+/+} healthy controls [▷ [supplement](#)]. It seemed plausible that their rather arbitrary level of methylation was permitted by a lack of negative selection.

Part II

Deregulated genes and pathways

Chapter 7

Transcriptional analysis

Contents

7.1	Characterization of Dnmt1-hypomorphic transcription	52
7.1.1	Expression changes linked to promoter hypomethylation	53
7.1.2	Elongation efficiency of reference transcripts	55
7.2	Differential gene expression analysis	56
7.3	Contrast of Dnmt1^{-/chip} vs. Dnmt1^{+/+}	58
7.3.1	Altered genes	58
7.3.2	Altered pathways	59
7.4	H3K4me3 buffer domains	61

DNA methylation was shown to be involved in important functions like transcriptional silencing of genes or endogenous retroviruses, imprinting, X chromosome inactivation and is relevant for genomic stability [30, 171]. Therefore, we suspected to observe pronounced aberrant transcription in Dnmt1^{-/chip} MLL-AF9 leukemia as a consequence of the progressive methylation loss.

Widespread aberrant alternative splicing across tumors was documented by other researchers [172] at the time of writing this thesis, the cause of which however was still disputed. Loss of methylation in gene bodies seemed to be a plausible mechanism to affect splicing [173].

An alternative explanation was provided by the discovery of epigenetically repressed cryptic promoters, which became activated upon hypomethylation and seriously perturbed regular splicing [174]. These promoters were referred to as *treatment-induced non-annotated transcription start sites* (TINATs), because they were identified in the context of therapeutic treatment with DNA methyltransferase inhibitors (DNMTi) for cancer therapy. Previously the therapeutic effect of hypomethylating drugs for cancer therapy had been widely attributed to the reversal of focal hypermethylation silencing tumour-suppressor genes [175], but the derepression of cryptic promoters represented an appealing different mechanism.

7.1 Characterization of Dnmt1-hypomorphic transcription

In collaboration with Gangcai Xie from the group of Wei Chen, Lena Vockentanz had generated RNA-seq data from MLL-AF9 c-Kit^{high} and c-Kit^{low} fractions [24, 25] as well as a H3K4me3 ChIP-seq from c-Kit^{high} cells for both genotypes (Dnmt1^{-/-chip}, Dnmt1^{+/+}). In-depth analysis of this data was another task, which was addressed as part of this thesis work.

Briefly, we aligned the FASTQ quality-trimmed single-end reads to the NCBI37/mm9 reference genome with BBMAP [176] and used the RSUBREAD/EDGER-pipeline for downstream analysis [↔ section 7.2, p.56]. We updated the transcript reference multiple times over the course of the project, but the data freeze for this thesis was set to release 84 of the NCBI REFERENCE SEQUENCE DATABASE (REFSEQ) published on September 11, 2017. We considered any transcript with less than 2 CPM to be not expressed in that particular sample as recommended by the authors of the EDGER-pipeline [177]. Depending on the level of summarization, we therefore included either 11 996 (of 24 588) annotated genes or 9505 (of 38 006) transcripts. That the number of genes surpassed that of single transcripts was attributable to genes with multiple transcripts, which only met the inclusion criteria when combined.

While reviewing the aligned RNA-seq data, we had observed that a surprisingly large number of genes seemed to be incompletely annotated in the reference genome, since the canonical transcripts could sometimes not explain the distribution of sequencing reads. Often, we could spot peaks of aligned reads in the vicinity of a known gene, despite no reference exon was annotated at those loci. Furthermore, the readcounts measured at the exons of a particular transcript were sometimes unbalanced, arguing against a full-length transcription of the transcript in question.

Since fusion proteins of MLL (MLL-FP) are known to impact crucial regulators of the transcriptional machinery [5], we presumed to detect new transcripts and splice variants. To generate an experimentally determined transcriptome of MLL-AF9 cells, Irina Savelyeva in collaboration with Claudia Gebhard from the laboratory of Michael Rehli performed CAGE-seq [178–180], a technique to sequence capped 5'-ends of mRNA to determine transcription start sites (TSS). We united the CAGE-seq and RNA-seq datasets to determine an experimental transcriptome [↔ section 8.1, p.65].

Correct assignment of the RNA-seq samples to the respective genotypes was assured by detection of the neomycin phosphotransferase II (EC 2.7.1.95) transcript, which originated from the Dnmt1^c-allele [▷ supplementary figure]. The genotype of the CAGE-seq samples could not be validated, because the promoter of the PGK-NeoR-PGKpoly(A)-cassette of the Dnmt1^c-allele could not be distinguished from the housekeeping-gene phosphoglycerate kinase 1 (PGK). Typically, there are weak signals at exon starts in addition to the TSS clusters in CAGE-seq. Thus, we tried an alternative approach for genotype validation based on the deleted exon 35 in the Dnmt1^c-allele. However, we could detect a signal from this exon in all samples and the levels were inconclusive. Since the

Dnmt1^{chip}-allele is an unspliced cDNA, CAGE-seq proved that active transcription and post-transcriptional splicing is taking place in Dnmt1^{-/chip} at the locus of the Dnmt1^c-allele despite the c-terminal truncation.

7.1.1 Expression changes linked to promoter hypomethylation

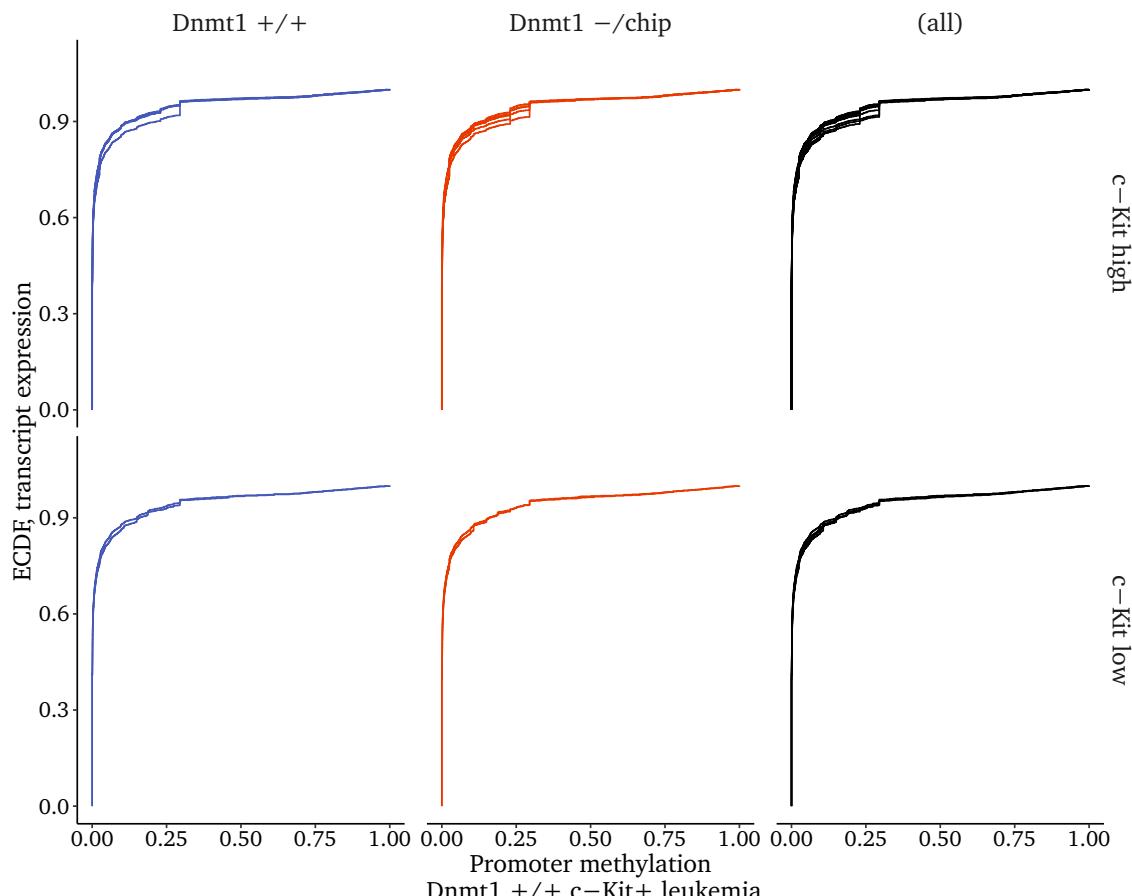


Figure 7.1: Empirical Cumulative Distribution Functions of the transcript expression measured by RNA-seq plotted relative to the ranked promoter methylation of the respective transcript. For each genotype, seven individual samples are shown, which are divided into c-Kit^{high} ($n=5$) and c-Kit^{low} ($n=2$). The populations distinguishable by c-Kit-expression are split in rows, while the columns depict the separate genotypes. The rightmost column features the overlay of both genotypes for direct comparison.

Since the Dnmt1^{-/chip} mouse exhibited an impaired methylation maintenance in our preceding WGBS analysis, we sought to investigate the extent of promoter hypomethylation and its effect on transcription in Dnmt1^{-/chip}. In accordance with the commonly accepted notion that promoter methylation is a repressive mark, more than 90 % of the total transcripts originated from completely methylation-free promoters (-500 bp to 100 bp around the TSS) [▷ Figure 7.1]. The ECDF plots relative to Dnmt1^{-/chip} promoter methylation as well as those based on CAGE-seq expression data were highly similar to Figure 7.1 and thus corroborated that predominantly unmethylated promoters initiated RNA transcription [▷ data not shown]. Nevertheless promoter-methylation-linked transcriptional changes seemed to be rare at the first glance, given that the ECDF plots of the

cross-controls ($Dnmt1^{-/-}chip$ expression combined with $Dnmt1^{+/+}$ promoter methylation or vice versa) were highly similar to the regular pairs [▷ [Figure 7.1 & data not shown](#)].

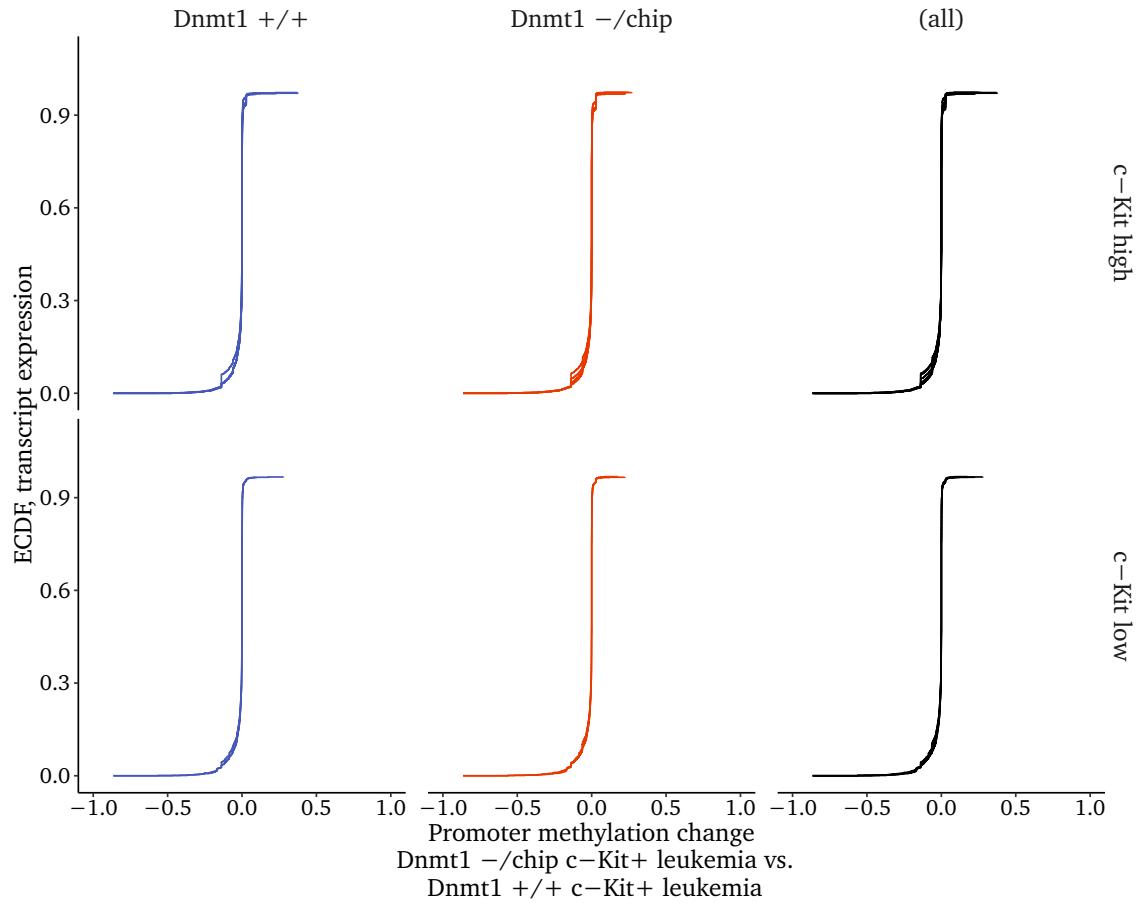


Figure 7.2: ECDF plots of the cumulated transcript expression relative to the methylation change observed at the respective promoters. The columns comprise the genotypes, while populations are spread out in rows. The methylation change at some promoters of expressed transcripts could not be determined due to insufficient coverage in WGBS, thus the cumulated expression in the plot is less than 1.

It was tempting to speculate that the $Dnmt1^{-/-}chip$ genotype would facilitate spontaneous hypomethylation of promoters and subsequent initiation of transcription. However, the RNA-seq/WGBS data yielded just a few dozen affected reference transcripts [↔ [subsection 7.3.1, p.58](#)], whose contribution to the overall expression was marginal [▷ [Figure 7.2](#)].

Because this finding would have profound consequences for our hypothesis, we sought to confirm it with another data set. While constitutive lamina-associated domains (cLADs), which contact the nuclear lamina with high cell-to-cell consistency are gene-poor, the flexible lamina-associated domains (fLADs) harbor a notable fraction of transcripts [138]. We conjectured that the fLADs should be a hot spot of stray transcription, since they combined relevant hypomethylation with the chromatin accessibility necessary to initiate transcription [181]. However, mapping transcript expression on the first principal component [182] of Hi-C data from HPC-7 cells clearly indicated the absence of stray transcripts [▷ [supplement](#)].

Possibly three effects could explain this finding:

1. Most methylated promoters harbored a CpG-Island, which we have already shown to be particularly persistent in the $Dnmt1^{-/chip}$ [↔ section 5.2, p.41].
2. Assuming demethylation in $Dnmt1^{-/chip}$ was mostly passive and random, then recurrent and consequential promoter demethylation of the same transcripts in biological replicates would be unlikely. Thus, the respective genes would hardly be considered as differentially expressed.
3. Pronounced methylation loss occurred in lamina-associated domains, typically packed in dense heterochromatin during G1/G2 phase. Productive elongation is mostly precluded for transcripts located in LADs [181,183], so even when methylation loss had been inflicted on their promoter, a reactivation was unlikely. On top of that, active demethylation recruits additional epigenetic marks to ensure transcriptional activation, which were missing in the case of passive demethylation [184–186].

Taken together, we deemed the reactivation of transcripts by spurious passive promoter hypomethylation to be a rare event in $Dnmt1^{-/chip}$.

7.1.2 Elongation efficiency of reference transcripts

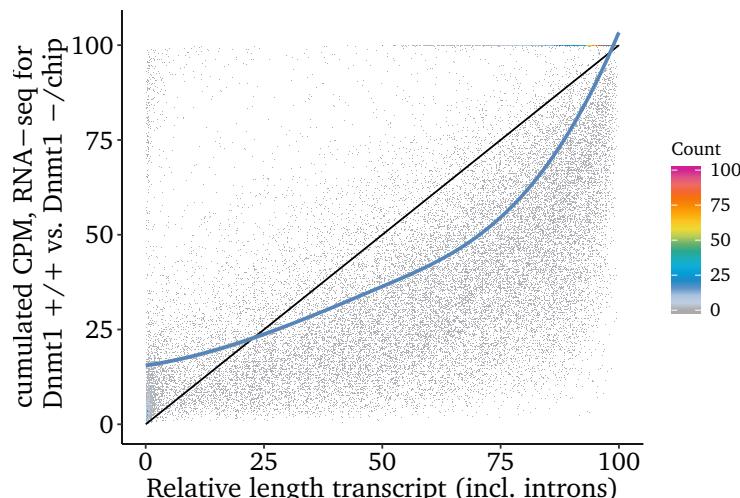


Figure 7.3: Spatial distribution of read counts along the relative length of expressed reference transcripts. Grey dots represent single data points, while color indicates areas with many overlapping points. The general trend is visualized by a blueish polynomial B-spline, whose deviation from the original straight line of fixed points (black) constitutes a bias.

While the repressive regulatory function of methylation at promoters is well established, the mechanistic importance of methylated cytosines in gene bodies mostly remains elusive. It was suggested that they may regulate splicing [173] or increase transcriptional efficiency in active genes [165]. Although an in-vitro transcription model had challenged those functions and rather pointed towards a solely H3K36me3-mediated mechanism [187], we investigated potential impacts of methylation loss in $Dnmt1^{-/chip}$ on transcriptional elongation.

Taking into account that the predominant MLL fusion partners function in transcriptional elongation [16, 17, 188] and MLL-AF9 is known to interact with the super elongation complex (SEC) [5], we asked if the $Dnmt1^{-/-}$ genotype would have an impact on elongation. For this purpose we quantified the expression of every exon separately in RNA-seq and assigned the measurements to reference transcripts. Counts for exons shared among transcripts were split according to the RPKM ratios of the full transcripts. For all exons we determined the position within the respective transcript relative to the transcription start site and thus obtained a spatial view of the expression.

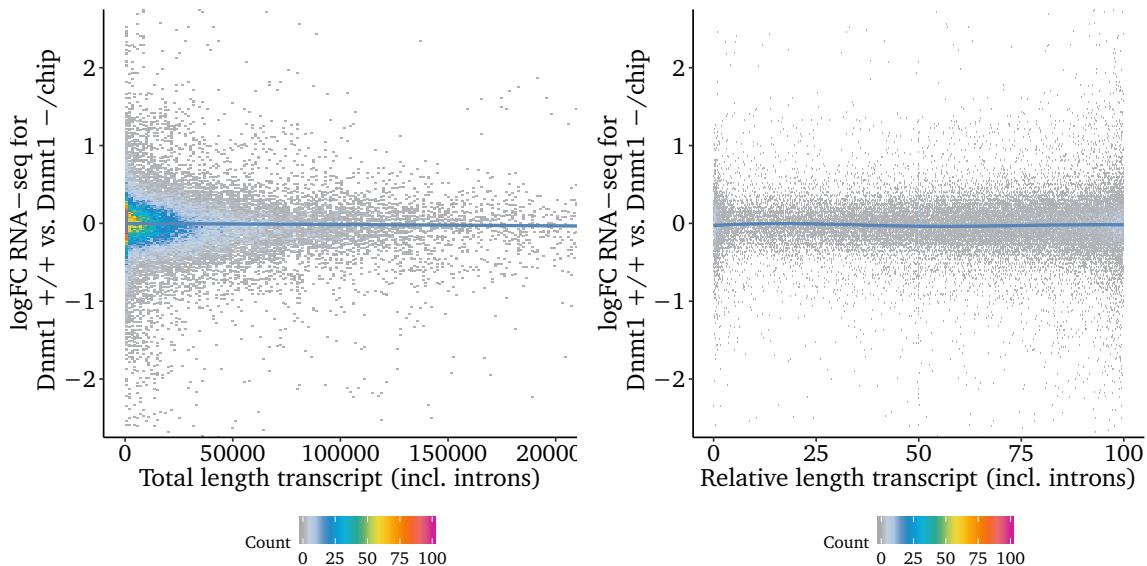


Figure 7.4: Summarized expression fold change of single exons is shown in dependence to absolute (left panel) and relative (right panel) transcript length. Genes with negative logFC were downregulated in $Dnmt1^{-/-}$, while positive items exhibited increased expression. A smoothed representation (blue line) of the data was achieved by fitting a polynomial B-spline with 4 knots.

Ideally, reads should be distributed equally along the whole transcript, unless they are influenced i.e. by the mapability. Despite a relatively high variance, we could observe a significant bias in our expression data. We detected disproportionate read fractions at the distal ends of the transcripts (in particular 3'), whereas the middle sections fell short of the expected representation [▷ Figure 7.3], which was probably an artifact of a poly(A)-enrichment performed during sequencing library prep. Separate fits for the genotypes were virtually indifferent [▷ data not shown], arguing against a specific elongation bias of $Dnmt1^{-/-}$. This was corroborated by subsequent fold change calculations for all exons with a sufficient expression (>2 CPM). We did not observe an increase or decrease in differential expression, neither with regard to absolute nor relative transcript length [▷ Figure 7.4].

7.2 Differential gene expression analysis

Changes in the biological processes of a cell are typically associated with an adaption on the genetic as well as the proteomic level. Stability, post-translational modification

or cellular localization of proteins may change and the transcription of previously unexpressed genes may be initiated, while that of other genes ceases. Often, it is of interest which genes are particularly relevant for a functional alteration such as differentiation or cell cycle progression. Starting from measurements of transcript abundance in different populations or under varied conditions, such genes can be identified by means of an analysis for differential expression.

The determination of genes, which are differentially expressed between two datasets, is therefore a common, yet not trivial task. A gene expression study will typically comprise just a few replicates for each condition, but assay thousands of genes in parallel. Often hundreds of genetic changes can be observed in parallel owing to the complex regulatory circuitry of genomic pathways as well as the cellular heterogeneity within cell populations. To designate a gene as differently expressed between two groups, whose true expression is confounded by experimental or biological variability, the observed expression averages must be sufficiently distinct in relation to the observed variance. Inaccurate estimation of the variance can heavily skew the test statistics and either erroneously reject or confirm the candidate gene as differentially expressed [▷ [supplement](#)].

Lena Vockentanz in collaboration with Gangcai Xie from the group of Wei Chen had performed RNA-seq from MLL-AF9 c-Kit^{high} ($n = 5$) and c-Kit^{low} ($n = 2$) fractions for each genotype (Dnmt1^{-/-}/*chip*, Dnmt1^{+/+}). The analysis was mostly carried out according to the RSUBREAD/EDGER pipeline [177], however we deviantly chose the *Genewise Negative Binomial Generalized Linear Models* approach [189] to test for differentially expressed genes. The alternative method applied an χ^2 -approximation to the likelihood ratio statistic instead of the pipeline's default quasi-likelihood F-test, which had a more rigorous type I error rate control and was thus more conservative.

We defined two contrasts, Dnmt1^{-/-}/*chip* vs. Dnmt1^{+/+} and c-Kit^{high} vs. c-Kit^{low}, in the test's design matrix. Since we had sequenced more c-Kit^{high} ($n = 5$) than c-Kit^{low} ($n = 2$) samples per genotype, each group consisted of three individual ex-vivo leukemia c-Kit^{high} fractions and two paired samples of both, the c-Kit^{high} and c-Kit^{low} populations. This experimental design required to account for batch effects, therefore we used a set-up with blocking in the specification of the GLM formula.

Ultimately, we could identify 4581 differentially expressed genes (3261 individual differential transcripts) at a significance level of 0.05. For some genes, it was not possible to assign a change to a specific transcript, thus the number of differentially expressed genes surpassed that of the transcripts. Considering the respective contrasts individually, a total of 730 genes (477 transcripts) were differentially expressed in Dnmt1^{-/-}/*chip* vs. Dnmt1^{+/+}. In comparison, the changes for the c-Kit^{high} vs. c-Kit^{low} contrast were significantly larger as 4393 gene (3109 transcripts) differed.

Evidently consequences of Dnmt1-reduction were mild compared to the distinctions between self-renewing leukemia stem cells to leukemic bulk. However, since the genetic basis of self-renewal and cancer cell stemness in LSCs had been addressed previously

[19, 21, 110, 190–193], we focused primarily on the effects of Dnmt1-insufficiency.

7.3 Contrast of Dnmt1^{-/chip} vs. Dnmt1^{+/+}

Experiments by former scientists of the Rosenbauer laboratory had collectively shown that Dnmt1 expression is essential for the cell-autonomous activity of MLL-AF9 leukemia cells [↔ section 1.4, p.13] as well was normal hematopoiesis [116]. This was in accordance with the beneficial use of DNA methyltransferase inhibitors (DNMTi) in hematological cancer therapy [194, 195]. Nevertheless the exact mechanism remained elusive, even though the Orkin laboratory had already investigated this issue with a different mouse model. Their results suggested a mechanism by which bivalent chromatin domains can no longer be suppressed by methylation in Dnmt1 -hypomorphic mice [196]. However, this was in contrast to our own finding that there was no significant increase in transcription by hypomethylated promoters [↔ subsection 7.1.1]. Furthermore, the study also did not measure methylation and instead just claimed that all upregulated genes would respond due to promoter hypomethylation [196]. Concordant with published results in mesenchymal stromal cells [197], research by Irina Savelyeva from our laboratory had instead suggested a senescence-mediated mechanism triggered by insufficient Dnmt1 levels. However, there was still uncertainty as to how Dnmt1 / methylation levels in a cell might affect the onset of senescence.

7.3.1 Altered genes

We could identify 730 significantly altered genes associated with the Dnmt1^{-/chip} genotype. Not surprisingly Dnmt1 was among the top hits in the table, but on average the reduction was only 50 % and not 70 % to 90 %, as suggested by previous studies of our laboratory [118]. Just 40 of 455 covered upregulated transcripts exhibited a hypomethylated promoter and were mostly lowly expressed. Given the organization in pathways, direct regulation by promoter hypomethylation was not considered to be imperative for all differentially expressed genes, but the extremely low number once more corroborated the absence of noteworthy transcript upregulation in Dnmt1^{-/chip} due to promoter hypomethylation. On top of that, in vitro experimental validation (qPCR, shRNA-mediated knock-down) of selected candidate genes with a hypomethylated promoter (e.g. Plekhg4, Nov, Rnf17) failed [▷ data not shown].

The most significantly upregulated gene in Dnmt1^{-/chip} was *nuclear protein in testis 1* (Nut1). This gene is namesake of the NUT-midline carcinomas (NMC), a rare, but aggressive disease often affecting visceral tissue. This group of tumors is characterized by translocations that often fuse NUT to bromodomain-containing proteins like BRD4 or BRD3 and result in an abnormal strictly nuclear localization of the fusion protein [198, 199]. Thus, Bromodomain and extraterminal domain inhibitors (BETis) show therapeutic efficacy on NUT midline carcinoma (NMC), just as they do on MLL-AF9 leukemia [200, 201]. Taking into account that similar pathways in both tumor entities are to blame for a lack of response to treatment with BETis [202, 203], it was tempting to speculate

that Nut1 upregulation represented some kind of reciprocal compensation mechanism in Dnmt1^{-/chip} MLL-AF9 leukemia. However, we did not pursue this approach, because the promoter failed to meet the hypomethylation cutoff for candidate genes.

Remarkably, several highly ranked genes were related to the cytoskeleton, among them Iqcd, Myom1, Arc, Amotl2, Shank1 or Pard6b. The latter was studied in detail by Irina Savelyeva because we suspected that a n-terminally truncated variant would be expressed from a cryptic promoter. Although this could be confirmed by 5'-RACE-PCR, an artificial expression of the shorter fragment had no inhibitory biological effect on Dnmt1^{+/+} MLL-AF9 leukemic cells in various in-vitro experiments [▷ [data not shown](#)]. Nevertheless, we could not completely rule out the importance of varied microtubule and filament formation, modified cytoskeletal transport or altered motility as well as cell polarization, especially with regard to a presenescent state.

Furthermore a variety of genes were linked to various signaling pathways, in particular their second messengers. The adenylate cyclases Adcy6 and Adcy7 are crucial for cAMP synthesis in both B and T cells [204], while the GTP binding proteins Gbp2b, Ifi47, Igtp, Irgm1, Irgm2 all exhibit GTPase activity and are linked to cellular interferon response [205, 206]. Last but not least, we also identified several, mostly cGMP-specific, phosphodiesterases (Pde3b, Pde5a, Pik3r6) in the top 100 genes.

7.3.2 Altered pathways

Several implementations to test gene sets for enrichment exist, most of which rely on a hypergeometric test [207]. The hypergeometric distribution describes the probability to obtain a number of successes in a sequence of n draws from a finite population without replacement. Thus, it is being tested, if the count of successes ¹ surpasses the number expectable by random. We used the KEGGPROFILER [208] package to perform the tests against the KYOTO ENCYCLOPEDIA OF GENES AND GENOMES (KEGG) database. It contains manually curated cellular pathways as well as various metabolic and disease related information.

In total, we considered 13 pathways to be valid enrichments for the Dnmt1^{-/chip} vs. Dnmt1^{+/+} contrast. The enrichment of the sets *Focal Adhesion* (mmu04510), *Regulation of Actin Cytoskeleton* (mmu04810) and *Tight Junctions* (mmu04530) did not come by surprise [↔ [subsection 7.3.1](#)]. We had also anticipated the presence of the *PI3K-Akt signaling pathway* (mmu04151) or *Calcium signaling pathway* (mmu04020) based on the manual review of differentially expressed genes.

Although of highest significance, we initially did not expect the enrichment of *Histidine Metabolism* (mmu00340) to be a valid finding. Upon closer inspection however, the result was sound and fitted into the overall picture. The key enzyme of the pathway, Histidine decarboxylase (Hdc, EC 4.1.1.22), was downregulated in Dnmt1^{-/chip} and thus the biosynthesis of histamine from histidine was impaired [▷ [Figure 7.5, top row](#)]. Further-

¹ Success is defined here as a differentially expressed gene, which is element of the pathway.

more we observed a mild increase of Histamine N-methyltransferase (Hnmt, EC 2.1.1.8) as well as a strong overexpression of the subsequent Amine oxidase flavin-containing A (Maoa, EC 1.4.3.4). These two enzymes catalyze the breakdown of histamine via N-methylhistamine to methylimidazole acetaldehyde, one of two possible routes². Taken together, the biosynthesis of histamine seemed to be impaired and its breakdown accelerated in Dnmt1^{-/chip} MLL-AF9 leukemia. Thus, we proposed markedly reduced histamine levels, although we did not verify that in an experiment.

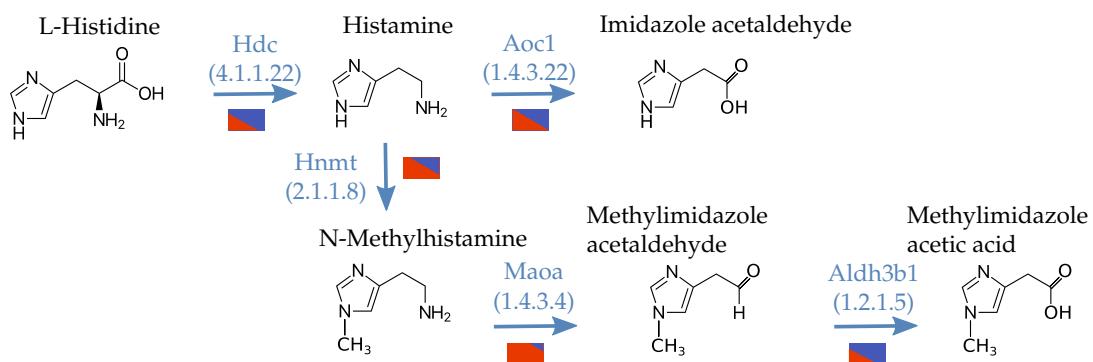


Figure 7.5: Key enzymes and catalyzed reactions of the histamine metabolism. Metabolites are shown in black, the enzymes' gene symbols and EC numbers in blue. The red/blue colored rectangles indicate the approximate expression ratio in Dnmt1^{-/chip} (red) and Dnmt1^{+/+} (blue).

Investigations regarding the role of histamine in leukemia date back to the 1940s and have typically reported elevated blood serum levels of histamine in myeloid but not lymphoid leukemia [209]. Although the H₂ histamine receptor is commonly expressed on AML of the subtypes M4 and M5 and secretion of histamine by leukemic blasts is frequent, it increases the susceptibility to elimination by NK and cytotoxic T cells [210, 211]. Furthermore sustained signaling through H₂ receptors is able to differentiate leukemia-derived cell lines [212]. In this regard, the reduction of autocrine histamine stimulation should actually benefit the leukemia - at least in vivo, where the evasion of anti-tumor immunity is required.

The cell-autonomous functions of histamine signaling are multifarious, including changes to the cytoskeleton. H₂-signaling triggers a dual response through phospholipase C stimulation (Gq/G11 family) and adenylate cyclase stimulation (Gs family) [213]. Thus, a decrease in autocrine histamine stimulation in Dnmt1^{-/chip} might explain the enrichment of a variety of second-messenger related accessory proteins as well as cytoskeleton components within the differentially expressed genes.

Leukemia of both genotypes expressed the H₂ histamine receptor, but none of the altered genes in the *Histidine Metabolism* (mmu00340) pathway exhibited a methylation change at the promoter. Therefore, the differential expression of the enzymes in Dnmt1^{-/chip} leukemia likely occurred in a regulated manner and was not an immediate consequence of promoter hypomethylation.

² The second route via Diamine oxidase (Aoc1, EC 1.4.3.22) was likely irrelevant in leukemia due to almost absent expression of the enzyme.

7.4 H3K4me3 buffer domains

Broad H3K4me3 domains spreading far into the bodies of genes (*buffer domains*) are believed to promote transcriptional consistency at key lineage genes [124]. We used SICER [214,215] to call buffer domains in Dnmt1^{+/+} and Dnmt1^{-/chip} leukemia [▷ [supplement](#)]. Surprisingly, just 553 (19.1 %) of the Dnmt1^{+/+} and 418 (17.0 %) of the Dnmt1^{-/chip} buffer domains overlapped any annotated gene at all. A majority of buffered genes was shared among the genotypes and seemed to exhibit an elevated median expression. The genotype disparity in buffer domains was higher than that of regular H3K4me3 peaks. Therefore, we anticipated that the impairment of Dnmt1^{-/chip} cells to acquire and maintain malignant self-renewal properties arose at least in part from the deviant buffer domains.

Although it was suggested that the primary purpose of broad H3K4me3 domains would be the stringent stabilization of gene expression for key regulatory genes [124], our data did show otherwise.

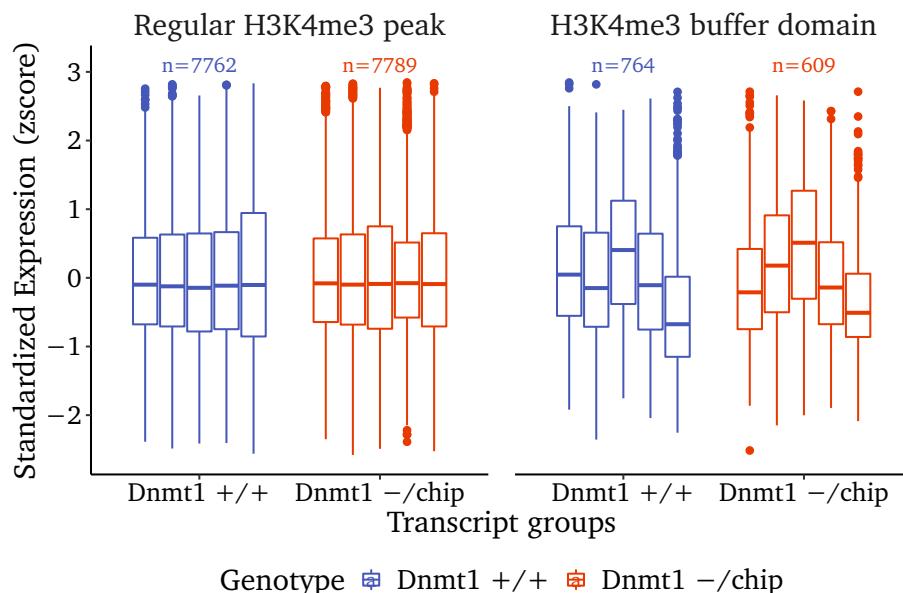


Figure 7.6: Boxplots of the standardized expression, which was calculated for each transcript individually. Raw values were centered by having the corresponding mean expression of the particular gene subtracted from them. Scaling was performed afterwards and refers to dividing the centered values by the sample standard deviation of the set. As the sum of all standardized values per transcript equals 0, purely random deviations should ultimately cancel out over hundreds of genes, which was not the case for those marked by buffer domains.

While comparing the centered and scaled (standardized) expression values, we observed directional deviations for single replicates in the buffer domain category [▷ [Figure 7.6, right panel](#)]. The reason for this rather consistent expression bias of some replicates however remained elusive - it was neither correlated with the strength of the H3K4me3 signal [▷ [data not shown](#)] nor with the expression of Kdm5b [216]. It should be noted that some buffered genes like Foxc1 were not expressed at all, although that gene in particular had previously been reported to play an important role for AML homeostasis [217].

Subsequently, we evaluated the genotype-specific allocation of the broad H3K4me3 domains. To assess the functional ramifications, we used the union as well as the genotype-specific gene sets as input and performed KEGG pathway enrichment analyses as described before [↔ subsection 7.3.2]. In total, we identified 27 enriched pathways, the genes of which were buffered in the genotypes to a varying extend. No entirely genotype-specific pathways could be determined.

The most enriched pathway was *Ribosome* (mmu03010), which comprised 30 genes encoding for various ribosomal proteins and RNAs. However, the result was likely house-keeping-gene-related, since we detected only minor differences between the genotypes and many of the distinct genes were still marked by regular H3K4me3 peaks in the other genotype. The second hit *Systemic lupus erythematosus* (mmu05322) was more promising, considering that our collaborator Melinda Czeh had identified an altered pathology of that particular disease in a different Dnmt1-hypomorphic mouse strain [▷ manuscript in preparation]. Yet, only the commonly buffered proteases Cathepsin G (Ctsg) and Neutrophil elastase (Elane) were highly expressed from this set in the MLL-AF9 LSCs. Further findings comprised several cancer-linked or viral-infection-related pathways, recurrently enriching due to the same genes shown in Figure 7.7. All genes from the top 10 enriched pathways are listed in the supplement. However, we did not pursue the experimental validation of any of those candidate genes.

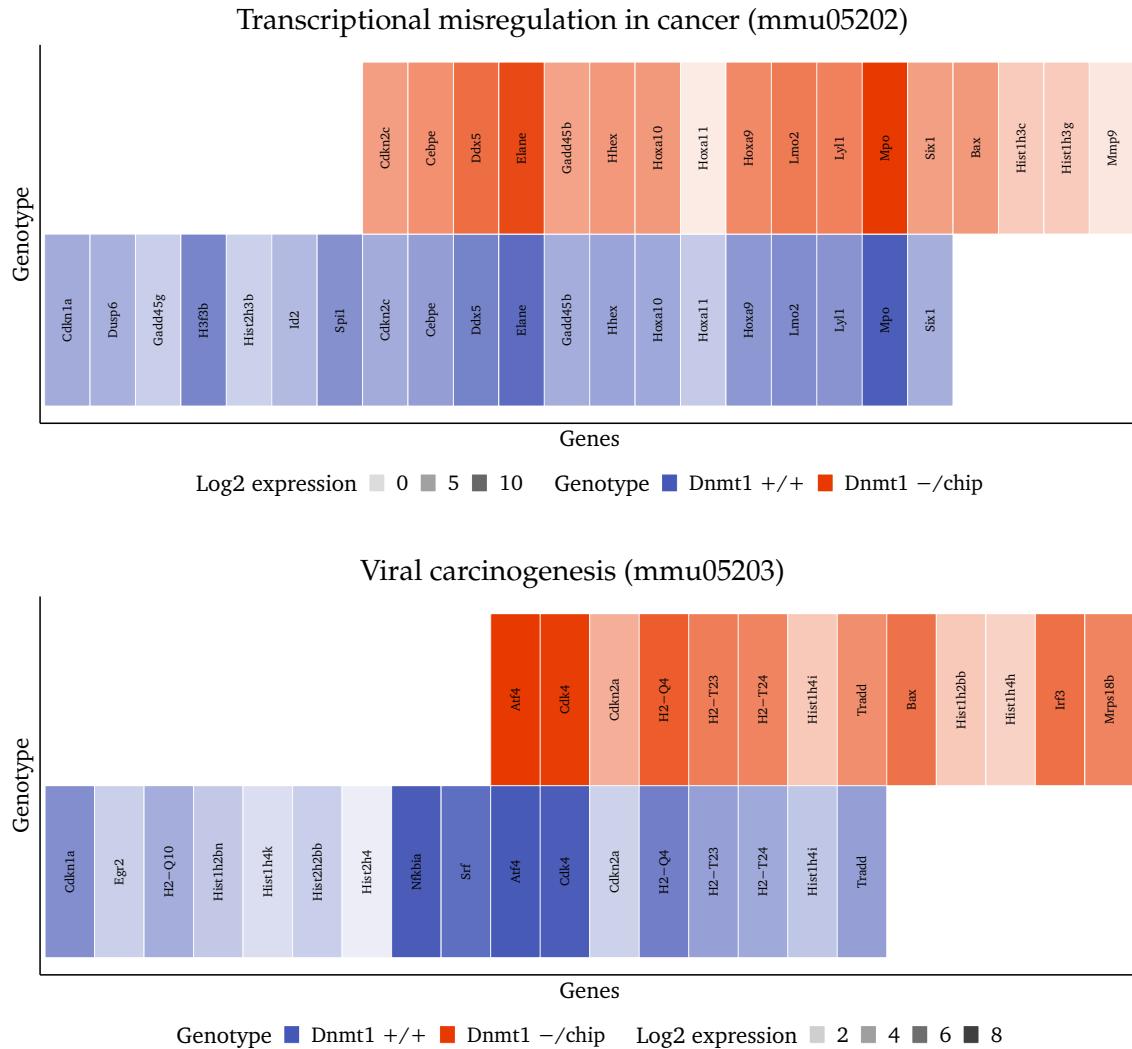


Figure 7.7: Heatmap for buffered genes, which are element of the respective pathways. Presence of a box indicates coverage of at least one transcript of the gene by a broad H3K4me3 mark in the color-coded genotype. Saturation of the color represents the expression of the gene in log2-scaled RPKM.

Chapter 8

Experimental transcriptome

Contents

8.1	Assembly of non-reference transcripts	65
8.2	Isolated transcriptional initiation events	66
8.3	Summary	68

The previous chapter addressed the altered genes and pathways in the Dnmt1^{-/-} genotype on the grounds of annotated genes and transcripts. Source of this annotation was release 84 of the NCBI REFERENCE SEQUENCE DATABASE (REFSEQ) published on September 11, 2017. Although the REFSEQ collection does include alternatively spliced transcripts, pseudogenes and alternative haplotypes as well as provisional entries, its scope of being a non-redundant database conflicts with its aptitude to represent the true transcriptional complexity of every cell type.

The FANTOM projects had first shed light on widespread cell-type specific promoter usage [218, 219] and later frequent aberrant alternative splicing across tumors was proven [172]. Furthermore fusion proteins of MLL (MLL-FP) affect the transcriptional machinery [5] and epigenetically repressed cryptic promoters can be activated upon hypomethylation [174]. Therefore, we assumed that the REFSEQ annotation might not adequately reflect the situation in MLL-AF9 cells and set out to generate an experimentally determined transcriptome.

8.1 Assembly of non-reference transcripts

This section is printed in condensed form. Optionally, a version with additional information regarding methodical details and quality control is available as online supplement.

In principle, there are two approaches to establish a custom transcriptome from RNA-seq data. One can opt for a true de novo assembly [220], which combines overlapping sequencing reads into longer continuous genomic sequences (so called contigs). However, for most bioinformatic approaches to the problem (like *De Bruijn graphs*), there is a trade-off between ambiguity as well as efficiency and computational demands such as memory consumption. Thus, for common model organisms, for which reference genomes exist,

it is generally preferable and more precise to align the reads first to the genome and reconstruct the transcriptome from those alignments [221–223]. After alignment [[▷ supplement](#)], we employed STRINGTIE [224] to reconstruct the transcripts and retained only those assemblies, which were supported by a 5'-prime CAGE-seq signal with a custom script.

We were able to reconstruct 43 597 elongated transcripts from CAGE-seq confirmed transcription start sites. 12 850 (29.47 %) were ultimately considered for downstream analyses, others were e.g. artificial mergers of reads originating from different samples. As before with reference transcripts, considerably more transcripts were differentially expressed between c-Kit^{high} and c-Kit^{low} cells ($n = 3686$) than between the genotypes Dnmt1^{-/-chip} and Dnmt1^{+/+} ($n = 519$).

25 % respectively 17 % were non-reference transcripts. The most common non-reference transcripts were unique, intergenic transcripts with no direct relation to annotated genes. These were mostly short (<2 kb), unspliced fragments typically framed by SINEs or LINEs, possibly pseudogenes, sequences of viral origin or transposons. Their expression or number did not increase in Dnmt1^{-/-chip}, therefore they were probably not attributable for the impairment of self-renewal, although their exact role remained elusive [[▷ supplement](#)].

Published literature had suggested that DNA hypomethylation is capable of reactivating cryptic, dormant promoters, which were referred to as *treatment-induced non-annotated transcription start sites* (TINATs) [174]. We detected only few transcripts with novel splice junctions (j), none of them was differentially expressed and thus of interest for this project. We also could not identify a non-reference transcript class, whose transcripts were predominantly initiated from hypomethylated, reactivated promoters in Dnmt1^{-/-chip} [[▷ supplement](#)]. Not even the promoters of the few truly differentially expressed non-reference transcripts exhibited a pronounced hypomethylation in Dnmt1^{-/-chip} [[▷ supplement](#)]. Our data therefore did not corroborate widespread joining of RNAs originating from cryptic promoters to regular reference transcripts during splicing as described before [174].

8.2 Isolated transcriptional initiation events

Because we observed many isolated transcription initiation events in CAGE-seq, which could not be extended to full length transcripts by RNA-seq, we were concerned to underestimate the amount of TINATs by the combined CAGE-seq/RNA-seq approach. For example, we observed a strong CAGE-seq signal (and hypomethylation) right at the site of the cryptic promoter in the Dapk1 gene, which had been described in the original TINAT publication [174], but the RNA-seq data did not allow for a successful elongation into a de novo assembled transcript. It seemed that our de novo assembled transcriptome of MLL-AF9 leukemia comprised mostly recurrent, faithfully reconstructed transcripts at the expense of most random, rare RNAs originating from TINAT-like initiation events.

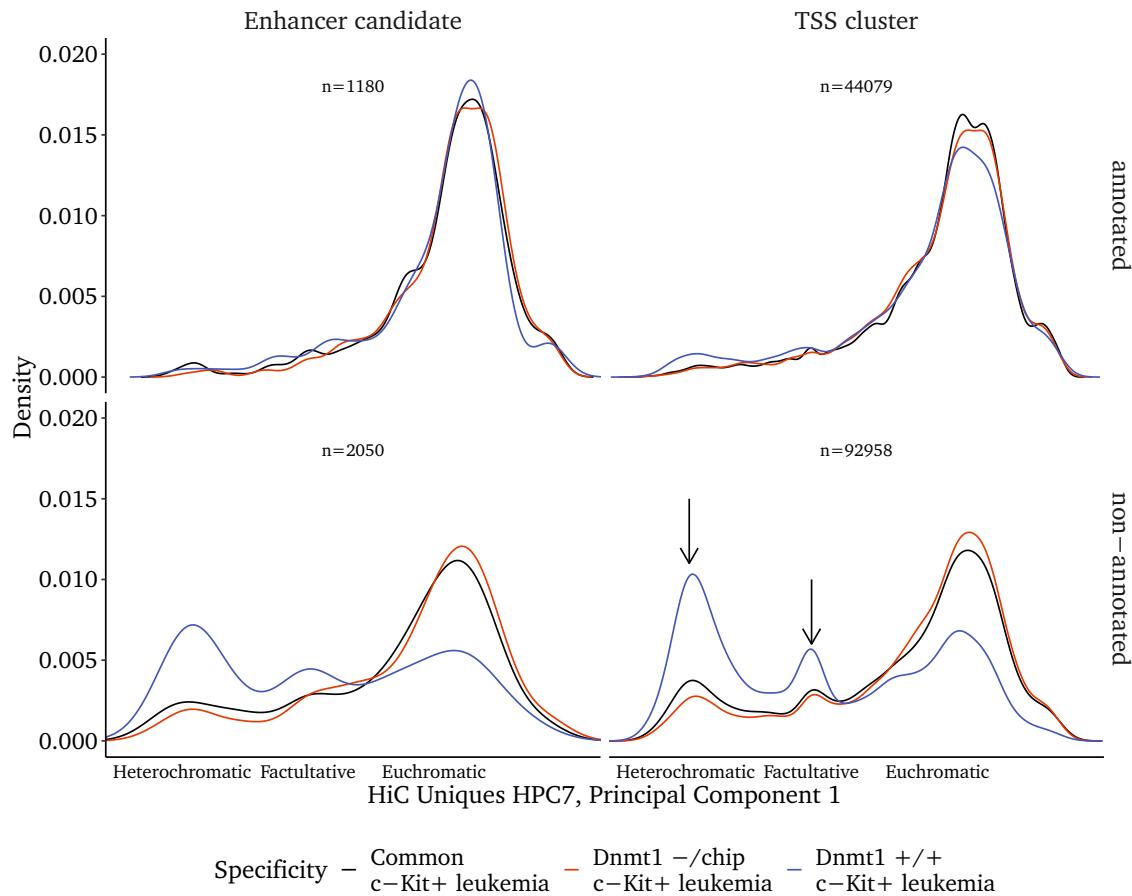


Figure 8.1: Genomic localization of transcriptional initiation. The tag clusters were assigned the respective first principal component of HPC-7 murine blood stem/progenitor cell Hi-C uniquely mapping interaction data. TSS were separated according to specific occurrence, overlap with Fantom 5 reference as well was classification as enhancer or promoter. Black arrows emphasize the unusual enrichment of robust or facultative heterochromatic localizations in the wild-type specific, unannotated clusters.

Therefore, we once loosened the criteria and focused solely on the CAGE-seq data to elaborate on aberrant transcriptional initiation, although MLL-FP are known to rather affect the elongation of transcripts than their initiation [17]. In total we could identify 140 267 tag clusters, of which 45 259 overlapped known promoters from the FANTOM 5 reference, while 95 008 were unique. Two-thirds of the unannotated sites were specific to either leukemic genotype.

To explore the genomic localization of the tag clusters, we mapped the first principal component, which distinguishes *active/permissive* from *inactive/inert* chromatin compartments [182], of Hi-C chromatin interaction data generated in the HPC-7 murine blood stem/progenitor cell model [\leftrightarrow section 5.1, p.39] [156]. While basically all annotated tag clusters irrespective of their specificity were exclusively located in the open chromatin regions [\triangleright Figure 8.1, top row], we noticed an abnormal enrichment of Dnmt1 $^{+/+}$ -specific, unannotated TSS clusters in typically inert heterochromatic regions. Since the decompaction of chromatin is a prerequisite of active transcription, this either pointed towards

an increased flexibility of the chromatin structure or a more readily initiated transcription [▷ [Figure 8.1, bottom row](#)].

8.3 Summary

This and the previous chapter focused on the transcriptome of MLL-AF9 leukemia and in particular on the changes induced by hypomethylation as a consequence of Dnmt1 reduction. On the grounds of the reference transcripts however, neither general promoter hypomethylation [↔ [subsection 7.1.1, p.53](#)] nor a putative elongation bias [↔ [subsection 7.1.2, p.55](#)] had profound effects on the transcriptome.

Just 730 genes were consistently and significantly differentially expressed between the two genotypes and only 40 exhibited a promoter hypomethylation in conjunction with an upregulation [↔ [section 7.3, p.58](#)]. Nevertheless we saw enrichment of the genes in some interesting pathways [↔ [subsection 7.3.2, p.59](#)], which could eventually explain the observed differences in self-renewal, tumor growth and senescence. Yet, a direct link to methylation could not be established, neither directly nor with the help of a comprehensive buffer domain analysis, which aimed for the identification of crucial regulatory genes linked to cell identity [↔ [section 7.4, p.61](#)].

Subsequently, we hoped that the reconstruction of non-reference transcripts might shed light on the ramifications of DNA hypomethylation in the Dnmt1^{-/-chip} mouse model. Although we detected a relevant number of non-reference transcripts [↔ [section 8.1](#)], they were not regulated by methylation [▷ [supplement](#)] and no superordinate mechanism could be elucidated.

Importantly, indications were weak that TINAT transcripts [174] would commonly be spliced to reference RNAs¹, although technical limitations might have caused us to underestimate the true extent. While stray transcription of full length transcripts was virtually absent [↔ [subsection 7.1.1, p.53](#)], the purely CAGE-seq based approach resulted in the identification of thousands of non-annotated Dnmt1^{+/+} specific TSS clusters in heterochromatin areas of the genome [↔ [section 8.2](#)].

Although the latter was definitively a finding of great interest, its interpretation remained challenging. Possibly, the samples were switched and these clusters were in fact confined to the Dnmt1^{-/-chip} genotype, which would resonate well with the pronounced hypomethylation of LADs and a explanatory model involving TINATs. Alternatively, these clusters were indeed absent in Dnmt1^{-/-chip} and reflected a diminished cellular plasticity. In the latter case, hypomorphic MLL-AF9 LSCs would, according to a model by the Feinberg lab [226], poorly respond when challenged by variable conditions resulting in a survival and self-renewal bias.

¹ just 11 transcripts in class j [225], none of which differentially expressed

Part III

Enhancer delineation

Chapter 9

Enhancer calling and classification

Contents

9.1	CAGE-seq derived enhancers	72
9.2	Enhancer clustering	73
9.2.1	Major cluster assignment by k-means	73
9.2.2	Minor cluster assignment by hierarchical clustering	75
9.3	Clades accumulating CAGE-enhancers	78
9.3.1	Characteristics in terms of healthy hematopoiesis	78
9.3.2	Characteristics in MLL-AF9 leukemia	80

In the previous chapters, the results of our studies regarding the immediate effects of Dnmt1 reduction in MLL-AF9 leukemia on transcription were presented. We addressed several proposed mechanisms, how abnormal DNA methylation can impact cancerogenesis at the site of the transcript: The potential reversal of abnormal promoter DNA hypermethylation and associated gene silencing of key regulatory genes [175], the derepression of cryptic promoters and a perturbation of regular splicing [174] and an elongation bias due to diminished gene body methylation [173]; none of which could be singly held responsible for the striking impairment of self-renewal and decrease in LSCs observed in Dnmt1^{-/-} leukemia.

In the course of our work, a growing body of papers stressed the importance of cis-regulatory elements as sites of pathogenic mutations and influential methylation changes [139,227–231], while emerging comprehensive WGBS datasets around the same time suggested that just 20 % of the CpG methylation changes under physiological conditions at all [137]. Taken together, we hypothesized that not the large scale demethylation observed in the lamina-associated domains [↔ section 2.2, p.21], but possibly small, yet decisive methylation changes in regulatory regions might be the long sought answer explaining our phenotype. Comparable mechanisms had been described in the cancerogenesis of other tumors [232,233] and inflammation [234], but data for MLL-AF9 leukemia was lacking at that point in early 2014.

9.1 CAGE-seq derived enhancers

This section is printed in condensed form. Optionally, a version with additional information regarding methodical details and quality control is available as online supplement.

Several methods exist, which can be used to identify putative enhancers in genome-wide datasets [reviewed in 235, 236]. Because we had already generated cap analysis of gene expression (CAGE-seq) [178–180] datasets from ex-vivo sorted the c-Kit⁺ fractions of four independently established leukemia to characterize TINATs [↔ section 8.2, p.66], we ultimately settled for that approach.

We called the enhancers according to a published protocol [82], which detects bidirectional eRNA transcription in CAGE-seq data [▷ [supplement](#)]. Additionally, we filtered such sites, whose cumulated expression did not exceed 0.5 TPM in total and 0.2 TPM in at least two replicates, which eliminated poorly supported locations with very weak signals, which were more frequent in Dnmt1^{+/+} and thus corroborated the higher transcriptional noise in these samples [▷ [supplementary figure](#)].

Ultimately, we retained 6386 and 6662 putative enhancers in Dnmt1^{+/+} and Dnmt1^{-/chip} respectively. Surprisingly, the majority of them (82.45 %) was specific for either of the genotypes [▷ [Figure 9.1](#)]. The large number of unconnected sites suggested a relevant share of false positive sites and called for extra caution in handling the data.

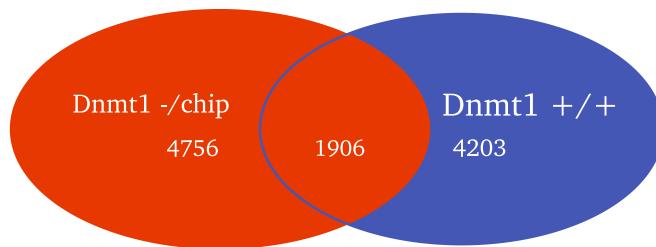


Figure 9.1: Venn diagram of the enhancer candidates, which passed the initial filtering step.

Among the characterized enhancers, we expected to find hundreds of sites, which were inherited from the cell of origin and had no direct relevance for the leukemia. To segregate pathogenic from other enhancers, we intersected the identified coordinates with the 48 415 enhancers contained in a comprehensive reference catalog of murine hematopoietic enhancers [101]. We could identify 889 CAGE-based enhancer candidates, which had been characterized before by the group of Ido Amit in healthy hematopoiesis.

Substantial fractions (37 % to 50 %) originated from the *Myeloid + Progenitors* (II) and *Myeloid* (VI) clusters [▷ [Figure 9.2](#)], thus corroborated the known relatedness of the MLL-AF9 Lin⁻ Sca-1⁺ c-Kit⁺ cells with granulocyte macrophage progenitors (GMPs). Yet, a notable number of enhancers (up to 28 %) also originated from progenitor clusters of the other lineages (III,IV,V), which indicated that either lineage commitment had not been

finalized or that the c-Kit⁺-fraction consisted of a rather heterogenous mixture of various cellular stages. Lymphoid-priming in AML has been known for some years [110,237] and was very recently characterized in great detail on single-cell level [238].

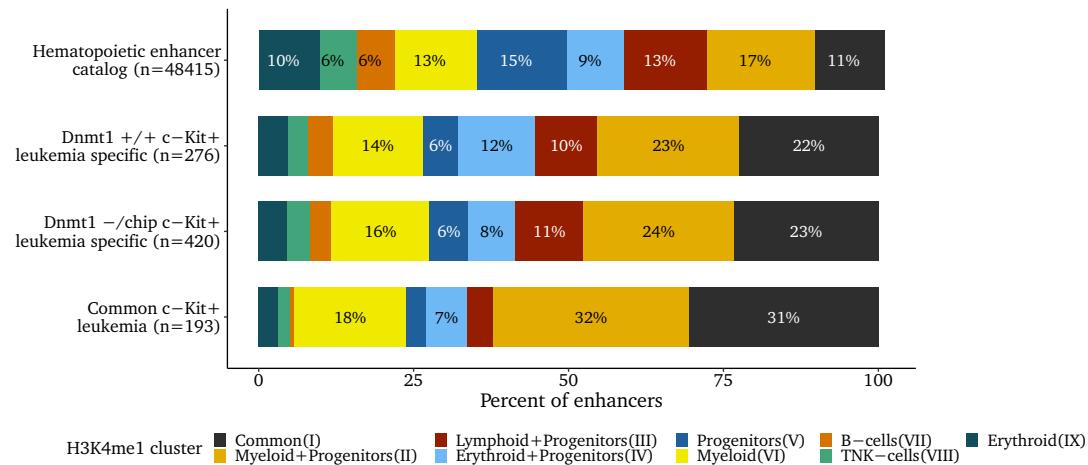


Figure 9.2: Bar graph showing the H3K4me1-cluster assignments of the overlapped hematopoietic enhancers.

The percentage of known hematopoietic enhancers in the set (8.2 % of 10 865) was surprisingly low. Since the largest part of the putative enhancers was not recorded in the hematopoietic enhancer catalog, it eluded a direct functional characterization or validation.

However, we presumed that the catalog was incomplete and we would be able to identify more putative enhancers, because no method can claim to have exactly identified all enhancers of a particular cell type as exemplified by a recent comparative study [239]. All techniques will preferably pick up elements in a specific state or rely on features not exclusive to enhancers. While the CAGE-seq method may have a particular bias towards other cis-regulatory elements [240], its validation rate of 70 % [82] outperforms H3K4me1/ H3K27ac based approaches with \approx 30 % conformation [241, 242]. None the less, the large number of unconnected sites suggested a relevant share of false positive sites and called for extra caution in handling the data.

9.2 Enhancer clustering

9.2.1 Major cluster assignment by k-means

To increase the reliability as well as to validate additional candidate sites, we repeated the clustering performed in the original study [101]. The rationale behind this approach was, that a direct overlap with the catalog was not imperative. If a candidate site faithfully recapitulated a known enhancer chromatin signature across multiple lineages, then one could assume it to be a valid call.

Therefore, we downloaded the aligned reads of all samples in the study [\rightarrow Appendix A, p.133] and probed the ChIP-seq signal at all 10 865 candidate sites as well as the reference

sites in every cell type. Prior to mapping, we resized the CAGE-seq derived enhancer sites uniformly to 2 kb to match the 47 526 non-overlapping sites of the catalog, which we remapped as background reference. This background data was used to calculate the initial values defining the centers for the subsequent repetition of the H3K4me1 *k*-means-clustering. Since *k*-means-clustering allocates elements to the most related cluster, even when hardly matching at all, we introduced a tenth cluster *Undefined* (X) to account for candidate sites, which were not marked by H3K4me1 in any cell type. This cluster was supposed to comprise sites with little or no H3K4me1-signal in the healthy hematopoietic system. Subsequently, we repeated the *k*-means-clustering with the joint datasets using the initial centers precalculated from the hematopoietic enhancer catalog [▷ [supplementary figure](#)].

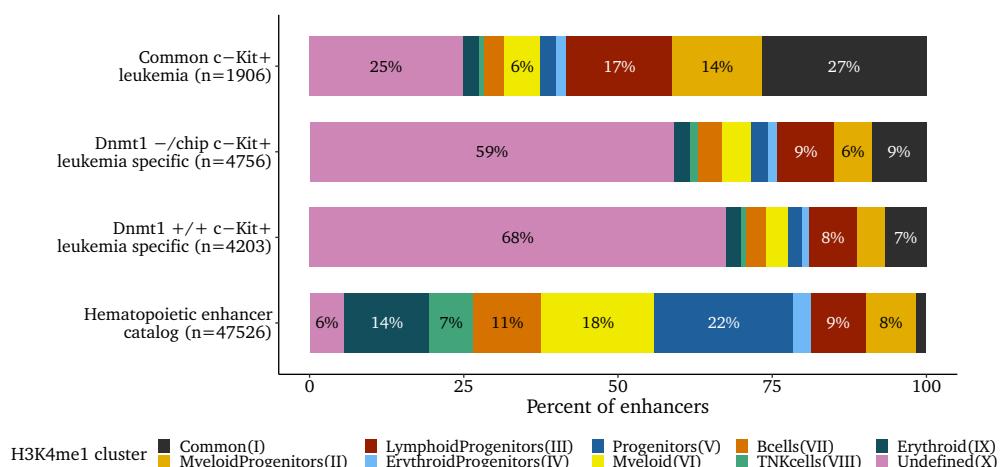


Figure 9.3: Bar graph showing the H3K4me1-cluster assignments as obtained by the repetition of the *k*-means-clustering. Rows represent the respective enhancer sets and colors the ten major H3K4me1-based functional clusters.

The reanalysis affirmed our conjecture that the hematopoietic enhancer catalog was incomplete, since 4742 bidirectionally transcribed sites were assigned to the clusters I to IX. Thus, 3853 non-recorded candidates exhibited a H3K4me1-signature reminiscent of regular hematopoietic enhancers [▷ [Figure 9.3, top three sets](#)]. The proportion of which however varied widely between the three enhancer sets: While 75 % of the common leukemic candidate sites were assigned to the healthy clusters, the specific sets were dominated by the *Undefined* (X) cluster (59 % and 68 % respectively) [▷ [Figure 9.3](#)]. In total 6123 putative MLL-AF9 leukemic enhancers were attributed to the newly introduced cluster X, which means that there was little evidence of involvement in regular hematopoiesis. Yet, the heatmap representation [▷ [supplement](#)] showed that at least some members of the *Undefined* (X) cluster were not entirely devoid of H3K4me1 modifications. Frequently, putative enhancers confined to one specific cell type clustered together with the *Undefined* (X) group, which also explained, why 6 % of the regular hematopoietic enhancer catalog were correspondingly reassigned [▷ [Figure 9.3, bottom row](#)].

In contrast to the immediate overlap with the enhancer catalog, consideration of non-overlapping bidirectionally transcribed sites introduced a remarkable shift towards the

lymphoid lineage. While myeloid enhancers of the clusters *Myeloid + Progenitors* (II) and *Myeloid* (VI) had dominated the direct intersections [▷ [Figure 9.2](#)], their share now essentially halved. This was exemplified by the common leukemic set, where the myeloid fraction dropped from 50 % to 20 %, which would correspond to an actual share of 26 %, ignoring cluster X for reasons of comparability. The *Lymphoid + Progenitors* (III) cluster in particular recorded a disproportionate increase at the expense of the myeloid ones [▷ [Figure 9.3, top three sets](#)]. This finding was intriguing, since it was suggested that cells with a functional similarity to lymphoid-primed multipotential progenitors (LMPPs) exist in human acute myeloid leukemia (AML) and represent a distinct, less mature subpopulation of leukemic stem cells (LSCs) [237].

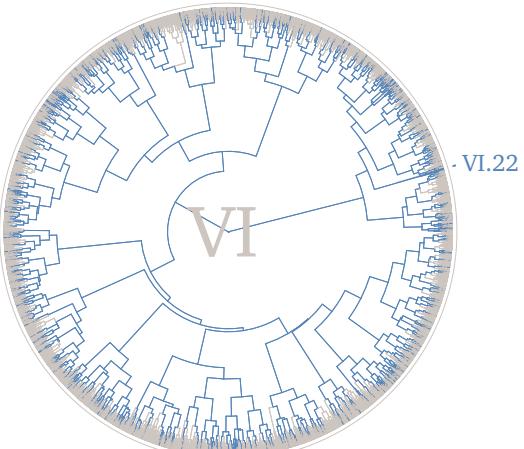
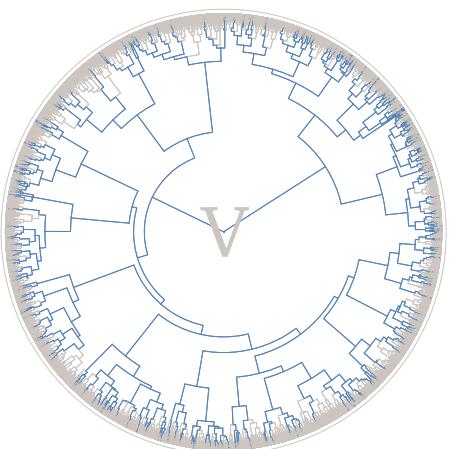
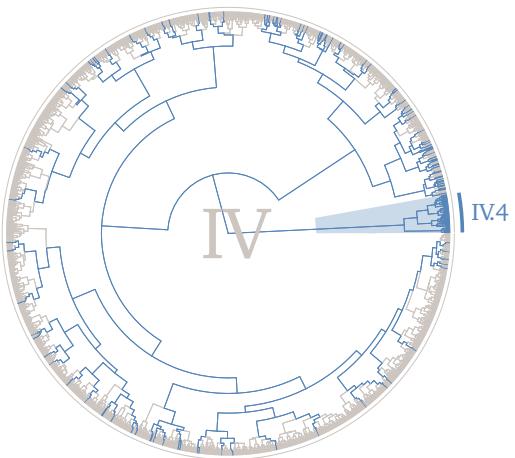
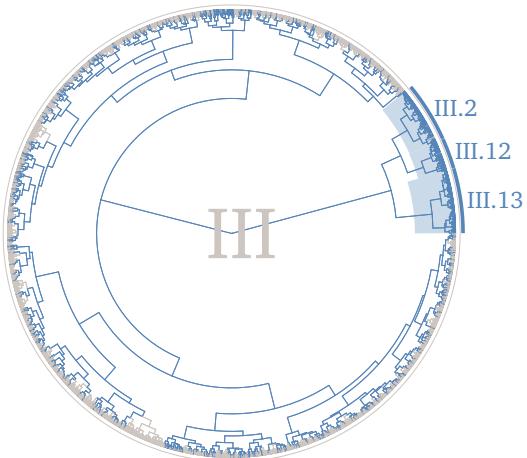
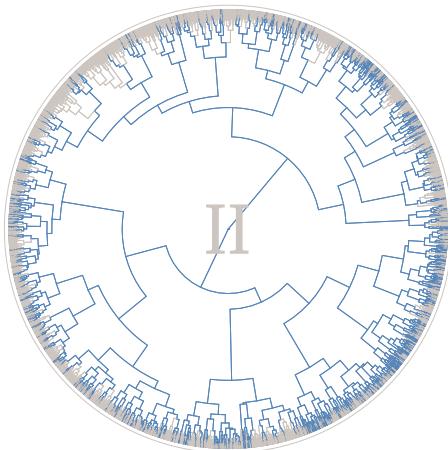
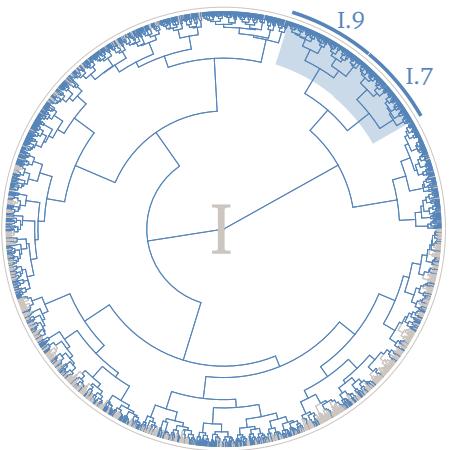
9.2.2 Minor cluster assignment by hierarchical clustering

Even though the division into the ten major clusters was suggestive of which enhancers might be relevant for the development and support of leukemia, it was only a coarse clustering. Importantly, although H3K4me1 extensively covers cis-regulatory elements, it is impossible to distinguish between active and poised enhancers based on that mark alone [▷ [supplement](#)]. Therefore, we resorted H3K27ac to obtain a detailed picture of our candidate sites. We mapped the H3K27ac as described before with the H3K4me1 ChIP-seqs and employed hierarchical clustering¹ to delineate the putative enhancers.

The rationale behind this approach was that sites, which shared similar patterns of activation throughout the hematopoietic hierarchy would likely be targeted by the same set of transcription factors. Thus, if we were able to identify groups of putative enhancers acting in a congeneric manner, the case for a purposeful activation in MLL-AF9 leukemia would be strengthened.

Ultimately, we obtained 151 H3K27ac-based subclusters, which will subsequently be referred to as clades, analogous to the branches of phylogenetic trees. Visual inspection of the resulting trees [▷ [Figure 9.4](#)] already confirmed preferential enrichment of CAGE-defined enhancer candidates in particular clades. A total of seven clades, one each from clusters I, III, IV, VI, VII, VIII and IX comprised solely actively transcribed sites. The remaining clades harbored some currently not active sites from the hematopoietic enhancer catalog as well as putative enhancers detected in MLL-AF9 leukemia. Odds ratios of the latter varied greatly from 0.06 to 116.38, therefore corroborating an uneven distribution likely reflecting functional disparity. Such great variation violated the assumption of homogeneously distributed odds ratios, which is a prerequisite for applying the Cochran-Mantel-Haenszel χ^2 -squared test, therefore we opted for a Woolf test ($p < 1 \times 10^{-16}$) to formally confirm the clade heterogeneity. Additionally, clades were singly tested by a regular χ^2 -squared test for enrichment [▷ [supplement](#)].

¹ Ward's minimum variance method based on euclidean distance



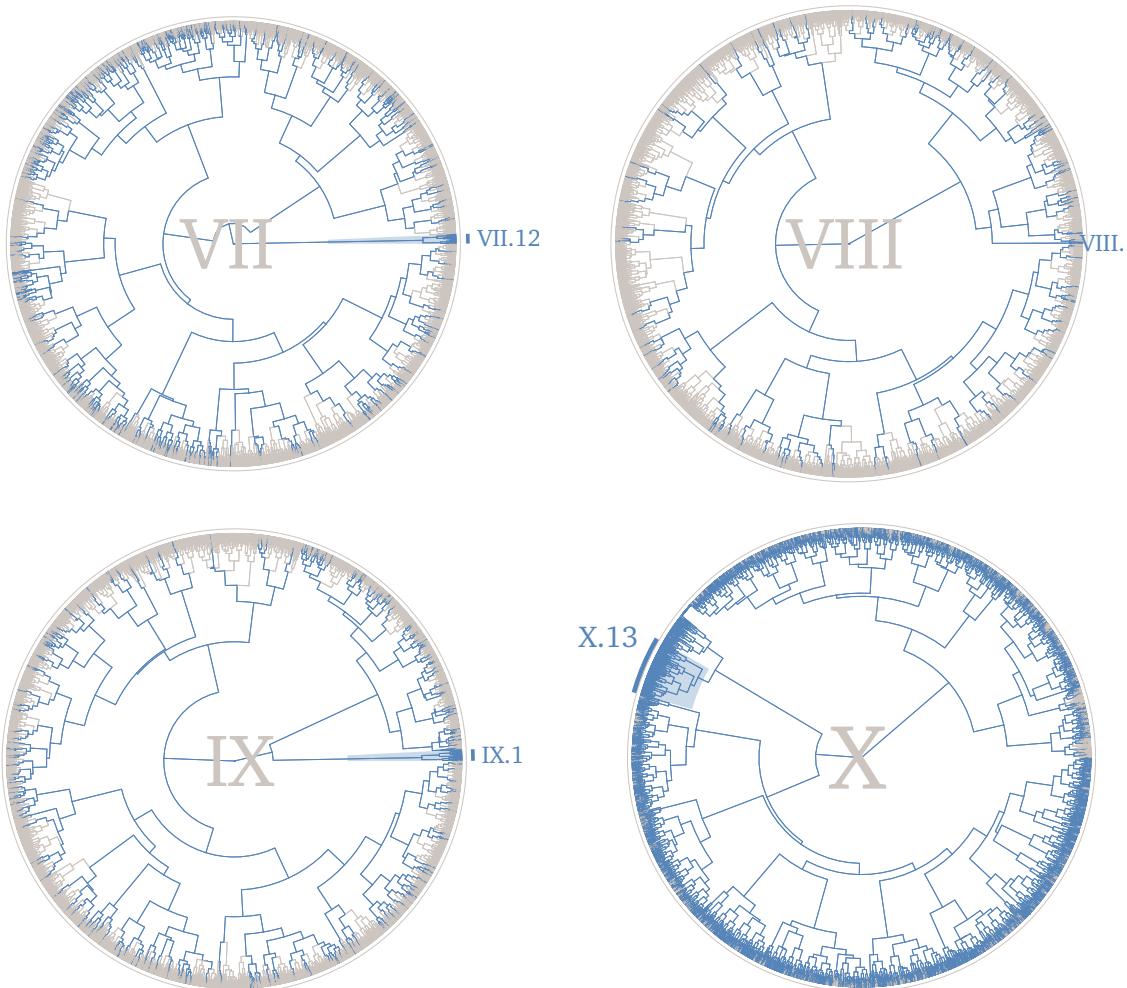


Figure 9.4: Visualization of the ten major H3K4me1 clusters as hierarchical trees. Clades highly enriched for sites bidirectionally transcribed in MLL-AF9 (positive χ^2 -squared test, odds ratio > 10) are specifically highlighted. For clarity, genotype specificity is not shown, instead blue color denotes any bidirectionally transcribed putative enhancer, gray items are recorded in the hematopoietic enhancer catalog, but seemingly inactive in leukemia.

9.3 Clades accumulating CAGE-enhancers

9.3.1 Characteristics in terms of healthy hematopoiesis

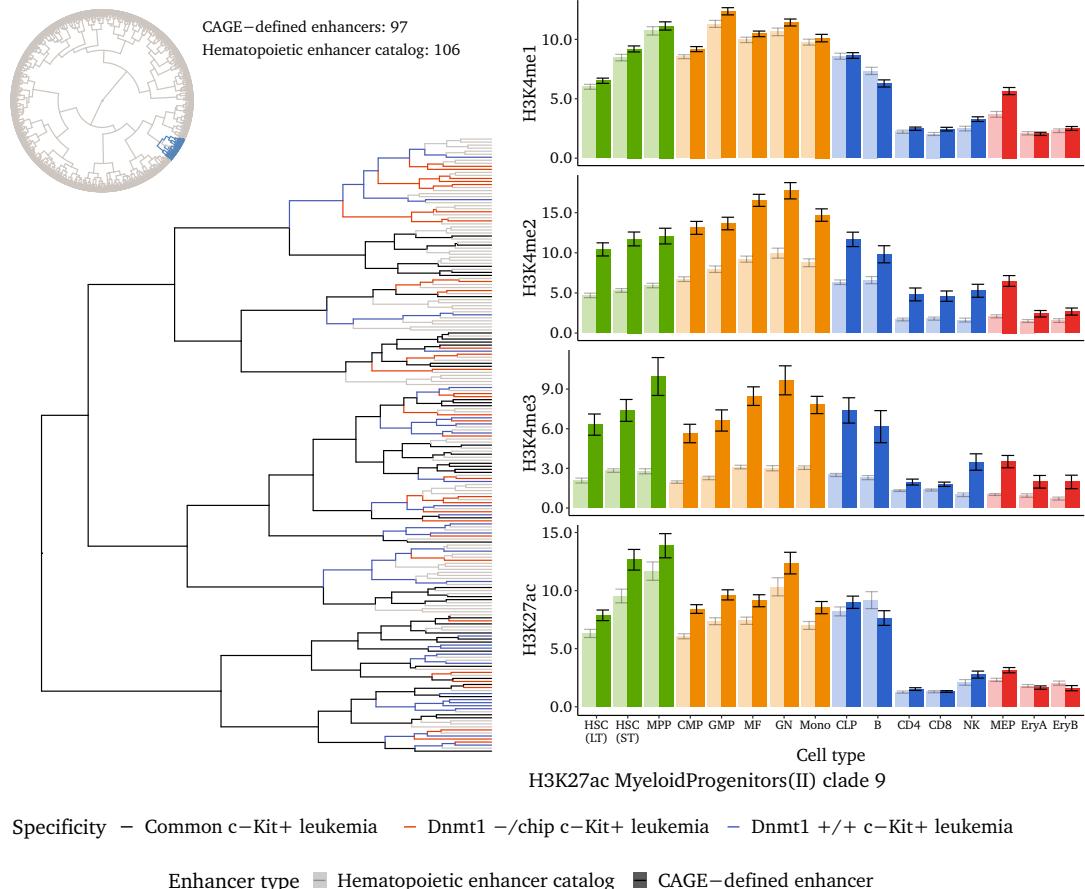


Figure 9.5: Visual representation of the single clade from the *Myeloid + Progenitors (II)* cluster. The left panel details the composition of the clade: Light gray represents inactive enhancers from the hematopoietic enhancer catalog, black marks putative enhancers active in MLL-AF9 leukemia of both genotypes, while those specific for Dnmt1^{+/+} and ^{-/-}chip are colored distinctively in blue and red. The right panel depicts the average ChIP-seq signal of all enhancers in that clade and its standard error of the mean. The darker bars summarize the average of all CAGE-seq detected sites and the light bars those of other enhancers from the hematopoietic enhancer catalog.

By design, enhancers within a clade featured similar patterns of H3K4me1 and H3K27ac throughout the healthy hematopoietic hierarchy. Furthermore, if they were also detected by CAGE-seq, their bidirectional transcription suggested cis-regulatory function, likely that of an enhancer in MLL-AF9 c-Kit⁺ leukemia. Both shall be exemplified by clade II.9 [▷ Figure 9.5]: It comprised 97 CAGE-defined putative and 106 further enhancers from the hematopoietic enhancer catalog. An odds ratio of 5.24 meant significant but not excessive accumulation of CAGE enhancers, so it was not separately highlighted in Figure 9.4.

Genotype-specific as well as commonly active ones were found among the associated CAGE-defined bidirectionally transcribed cis-regulatory elements [▷ Figure 9.5, left panel],

a pattern, which was reflected in all other clades. We did not find any clades, which contained entirely or almost exclusively genotype-specific elements [▷ [data not shown](#)]. Taking into account that a congenic activation pattern presumably argued for shared transcription factor motifs, we could conclude that the same set of transcription factors likely governed the transcriptional programs in $Dnmt1^{+/+}$ as well as $Dnmt1^{-/chip}$ MLL-AF9 leukemia.

Consequently, the detection of accumulated CAGE-defined enhancers within a clade likely indicated an ordered, non-random activity in leukemia. This was corroborated by the homogeneous H3K4me1 and H3K27ac signals within, which were typically akin to that of non-transcribed, cataloged enhancers of a clade [▷ [Figure 9.5, right panel](#)]. However, as exemplified by clade II.9, this did not necessarily apply to the H3K4me2 and H3K4me3 signals [▷ [Figure 9.6, right panel](#)].

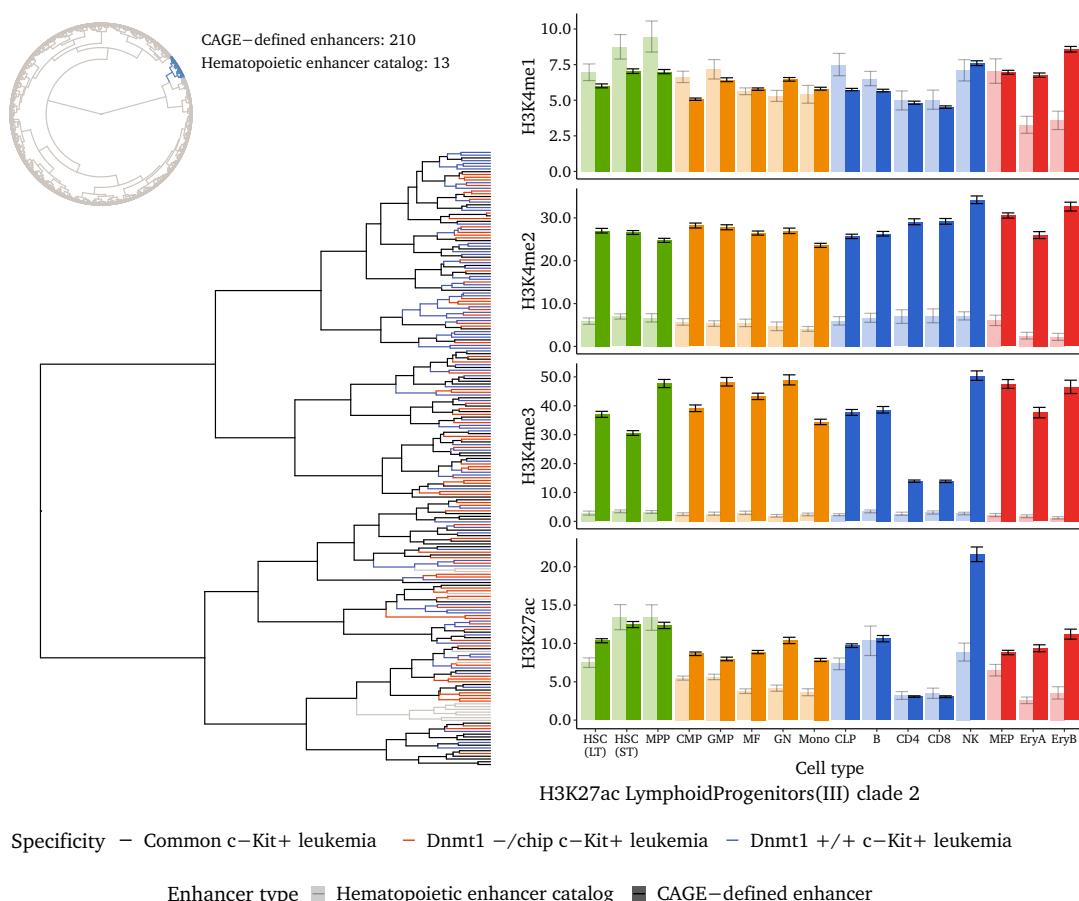


Figure 9.6: Details of an exemplary clade with a very high enrichment for CAGE-defined enhancers (odds ratio 76.77). Note the uniform presence of common (black), $Dnmt1^{+/+}$ specific (blue) and $Dnmt1^{-/chip}$ specific (red) enhancer candidates. The right panel displays the mean ChIP-seq signal of all enhancers within the clade for the CAGE-defined (rich) vs. inactive catalog enhancers (pale). Mind the strong differences in H3K4me2 and H3K4me3.

CAGE-defined enhancers typically exhibited a particularly strong H3K4me3 signal, which often doubled that of the non-detected controls from the catalog [▷ [Figure 9.5, right](#)

panel, rich vs. pale colored bars]. Although we did not use the H3K4me3 signal to build the cluster and clade assignments, many clades accumulating bidirectionally transcribed elements were remarkably strongly marked by H3K4me2 and in particular by H3K4me3 in many healthy hematopoietic lineages [▷ [Figure 9.6](#)].

Furthermore, we observed a notably strong H3K27ac signal in natural killer cells (NK cells), which typically marked CAGE-defined enhancers in clades with extraordinary high odds ratios (>10) [▷ [Figure 9.6, bottom bar graph](#)]. Since these clades also stood out clearly in terms of their absolute values, we subsumed an activation of stretch enhancers (super enhancers) linked to the NK cell lineage in MLL-AF9 leukemia. In contrast, clades with a clear yet moderate accumulation of CAGE-defined enhancers ($2 < \text{odds ratio} < 10$) exhibited only an average H3K27ac signal in NK cells [▷ [supplementary figure](#)].

Taken together, two intriguing starting points for further studies could be identified: Firstly, the strong H3K4me3 marks found throughout the hematopoietic hierarchy, which link those enhancers to histone K4 methyltransferases like the MLL / COMPASS or SET1 / COMPASS complexes. Secondly, the possible activation of super-enhancers (SEs) in MLL-AF9 leukemia, which under physiological conditions likely impart or contribute to NK cell lineage identity.

9.3.2 Characteristics in MLL-AF9 leukemia

The clustering strategy based on the histone marks from healthy hematopoiesis [↔ [section 9.2](#)] was mainly devised to provide validation and functional categorization, despite a lack of suitable datasets to integrate the CAGE-seq data with. About a year after we performed the initial clustering analysis², the laboratory of Michael Cleary published comprehensive ChIP-seq data from MLL-AF9 c-Kit^{high} and c-Kit^{low} cells [216]. This dataset allowed to streamline the clustering results in terms of eliminating false positives and to narrow down candidates for experimental validation.

We downloaded the data from the repository and reanalyzed it with regard to our CAGE-defined enhancer candidates. As exemplified by the *Lymphoid + Progenitors* (III) cluster, the accumulation of said enhancers within a clade typically coincided with higher average signals for H3K18ac and H3K27ac in leukemia [▷ [Figure 9.7, center and bottom row](#)]. As both marks are redundantly generated by CBP/p300 [243] and tag active genetic elements, we could thus infer an increased activity (and possibly importance) in leukemia of those enhancers and presumed binding of MLL-AF9 [244].

The strength of this effect varied depending to the H3K4me1 major clusters, but basically held true for the regular clusters I - IX. Enhancers within the *Common* (I) cluster were irrespectively of the clades' enrichment status marked strongly by H3K18ac and H3K27ac, which was in accordance to their presumed role of housekeeping gene support. Also in the second cluster *Myeloid + Progenitors* (II), the average signal of normal and enriched clades was quite indifferent, yet strong [▷ [Figure 9.8](#)].

² at that point just for the Dnmt1^{+/+} genotype, since we originally promoted this as a separate project.

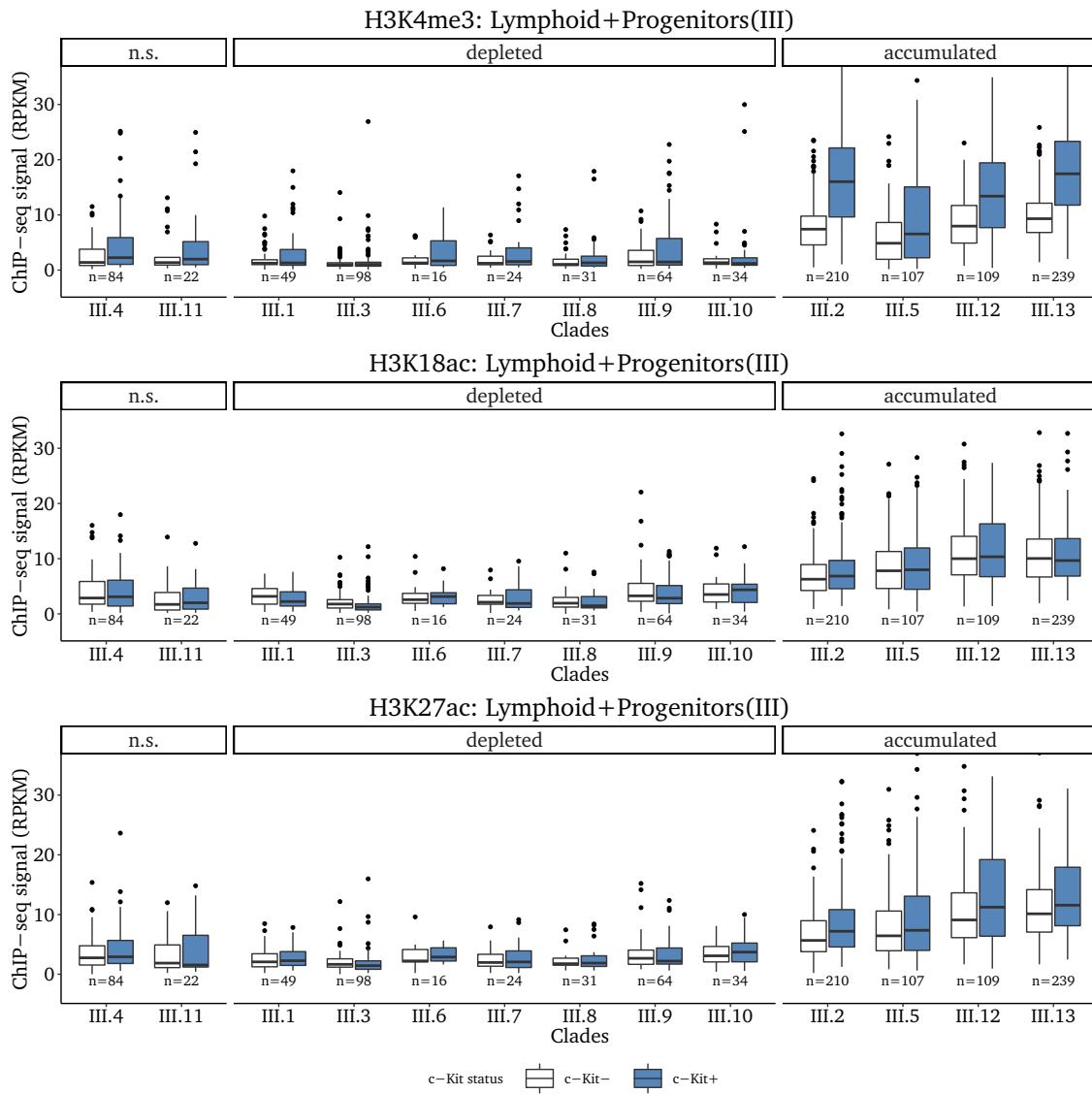


Figure 9.7: H3K4me3, H3K18ac and H3K27ac ChIP-seqs in MLL-AF9 c-Kit^{low} and c-Kit^{high} leukemic cells mapped to the CAGE-defined enhancers within the clades of the *Lymphoid + Progenitors* (III) cluster. Counts were normalized to sequence length and number of mapped reads in sample. In the plot, the clades are ordered by accumulation of putative leukemic enhancers.

In contrast, most enhancers of the cluster *B-cells* (VII) and, importantly, also of the *Undefined* (X) cluster lacked notable activity in leukemia. While the lack of B-cell-related enhancers seemed plausible, surprisingly just a few dozen out of thousands of CAGE-defined enhancers in the *Undefined* (X) cluster were marked by H3K27ac in MLL-AF9 cells [▷ Figure 9.8, *Undefined* (X) cluster]. Therefore, we conjectured that these represented mostly false positive calls and that the MLL-AF9 leukemic phenotype was almost exclusively sustained by hijacked physiological enhancers.

H3K4me3 was another histone mark that varied depending on the major cluster. While it was consistently low in insignificant clades [▷ Figure 9.8, top tow], accumulated clades of the clusters *Lymphoid + Progenitors* (III), *Progenitors* (V), *TNK-cells* (VIII) and *Ery-*

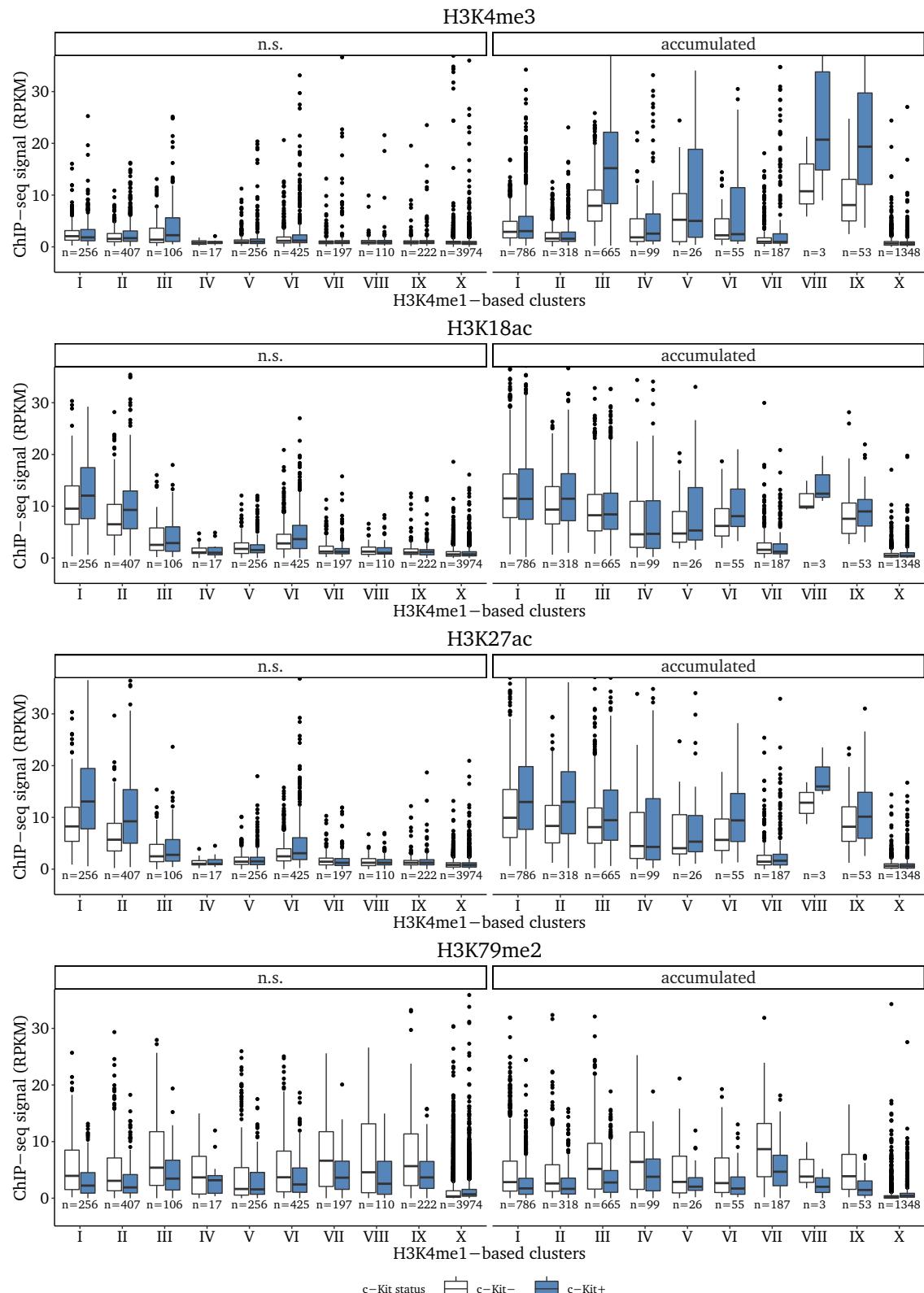


Figure 9.8: Boxplots of normalized H3K4me3, H3K18ac and H3K27ac and H3K79me2 ChIP-seqs generated from sorted MLL-AF9 c-Kit^{low} and c-Kit^{high} leukemic cells. The data is mapped to the putative leukemic enhancers called from the CAGE data and split according to the H3K4me1-based major clusters. Enhancers that were assigned to clades with depletion are not shown.

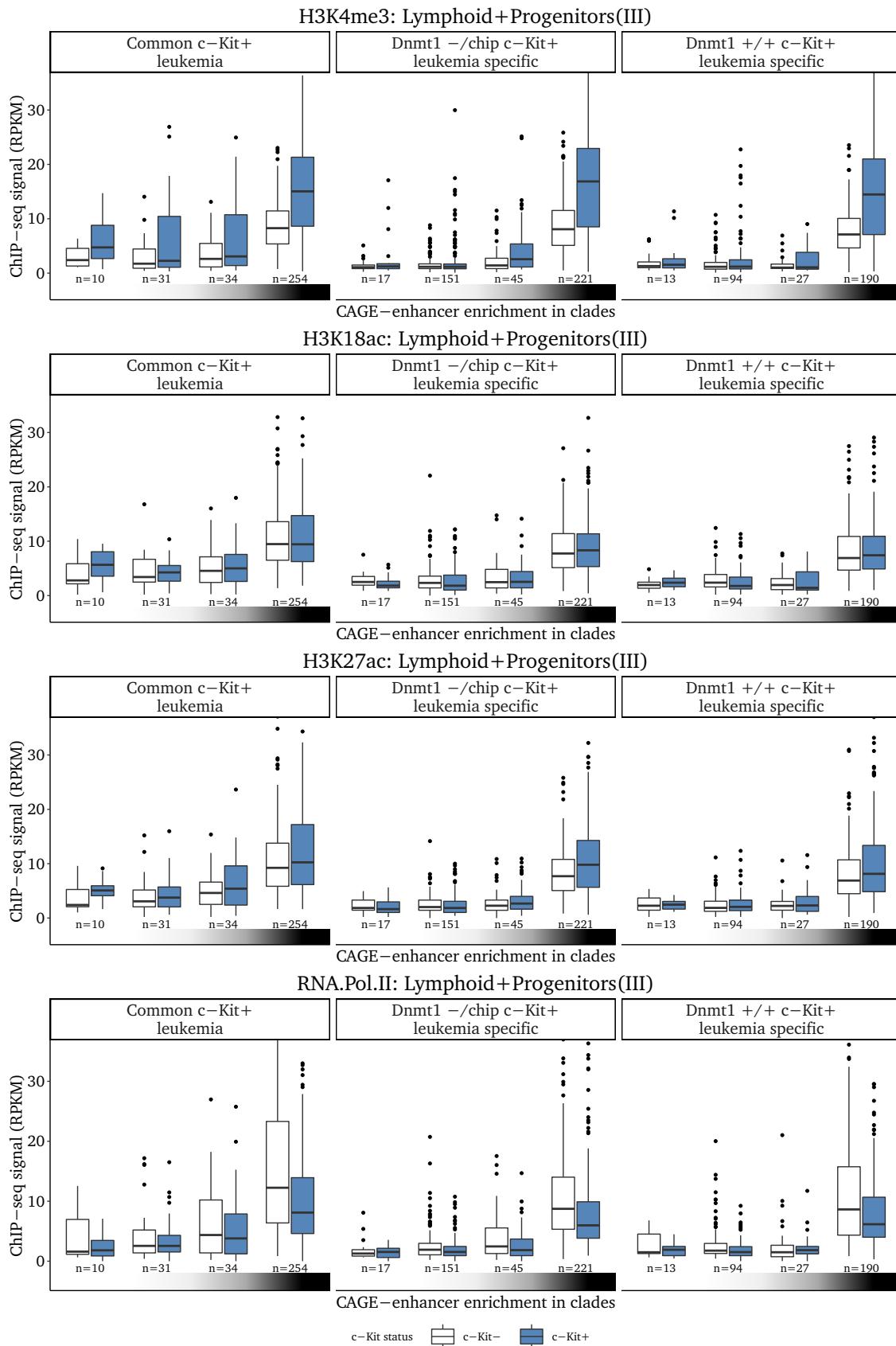


Figure 9.9: Clades within the cluster *Lymphoid + Progenitors (III)* were collapsed into four categories from left to right: strong depletion, mild depletion, no significance and strong enrichment depending on the clade's odds ratio (≤ 0.25 ; $]0.25, 0.75]$; $]0.75, 1.25]$; >4) and test significance (FDR < 0.01). Genotype specificity with regard to Dnmt1 $+/+$ and $-/\text{chip}$ is shown separately.

throid (IX) exhibited a propensity to acquire strong H3K4me3 marks. This finding is discussed later in greater detail [\leftrightarrow subsection 13.2.2, p.123]. Yet, by absolute numbers, the majority of the H3K4me3-positive enhancers were contained within clades of cluster III [\triangleright Figure 9.7, top row]. Given the pronounced H3K4me3 signal recorded at those sites in virtually all healthy hematopoietic cell types [\leftrightarrow subsection 9.3.1], we furthermore suspected that the allocation to separate clusters was misleading in that case. Instead, we favored the view that all of these elements were functionally one group that had been split erroneously into the different clusters due to slight individual variations in the H3K4me1 signature.

Contrary to the already mentioned histone marks, the DOT1L-mark H3K79me2 [245] was not linked to any particular cluster or clade enrichment. Since it had been conclusively shown that DOT1L and H3K79me2 are required to protect target genes of MLL fusion proteins from a repressive complex composed of Sirt1 and Suv39h1 [246], we conjectured that the open chromatin state of at least some enhancers from the insignificant clades must nevertheless be maintained to uphold the leukemic differentiation block [247]. Remarkably the number of H3K79me2 positive CAGE-enhancer candidates in the *Undefined* (X) cluster was much higher than the number of H3K27ac or H3K18ac positive ones. This suggested that cluster X comprised relevant amounts of KEEs [248] and might have contained less false positives than initially conceived by us based on the H3K27ac mapping. It should however not go unnoticed that the normalized signal tended to be stronger in c-Kit^{low} than in c-Kit^{high} cells [\triangleright Figure 9.8, bottom row], which could imply that said cis-regulatory sites were less relevant for the leukemic stem cells (LSC) than for the bulk leukemia.

Separate mapping with regard to genotype specificity corroborated previous findings that putative enhancers specific for Dnmt1^{+/+} and Dnmt1^{-/chip} did not differ with regard to their biological properties: Although the clades were exclusively assigned based on healthy hematopoietic data, specific enhancers within the same clade also exhibited similar histone modification patterns in leukemia [\triangleright Figure 9.9, columnwise comparisons]. Therefore, it seemed likely that a separate consideration of the enhancers in terms of genotype specificity was not required in a functional context. Instead, active Dnmt1^{+/+} and Dnmt1^{-/chip} specific enhancers seemed to be targeted by the same set of transcription factors. Furthermore, the vast majority of genotype specific putative enhancers had been assigned to the *Undefined* (X) cluster, whose cis-regulatory elements mostly lacked H3K18ac or H3K27ac marks in leukemia and only occasionally exhibited H3K79me2 marks. Therefore, the validity of those specific calls was anyway questionable.

ATAC-seq data [\leftrightarrow Appendix A, p.133] generated by the group of Jennifer Trowbridge [22] also challenged the validity of most enhancer calls of the *Undefined* (X) cluster, as most of them were located in supposedly closed chromatin [\triangleright Figure 9.10, bottom row].

Apart from cluster X, the ATAC-seq however supported the notion that the clade enrichment was indicative of relevant enhancers. Typically, clades with a strong enrichment

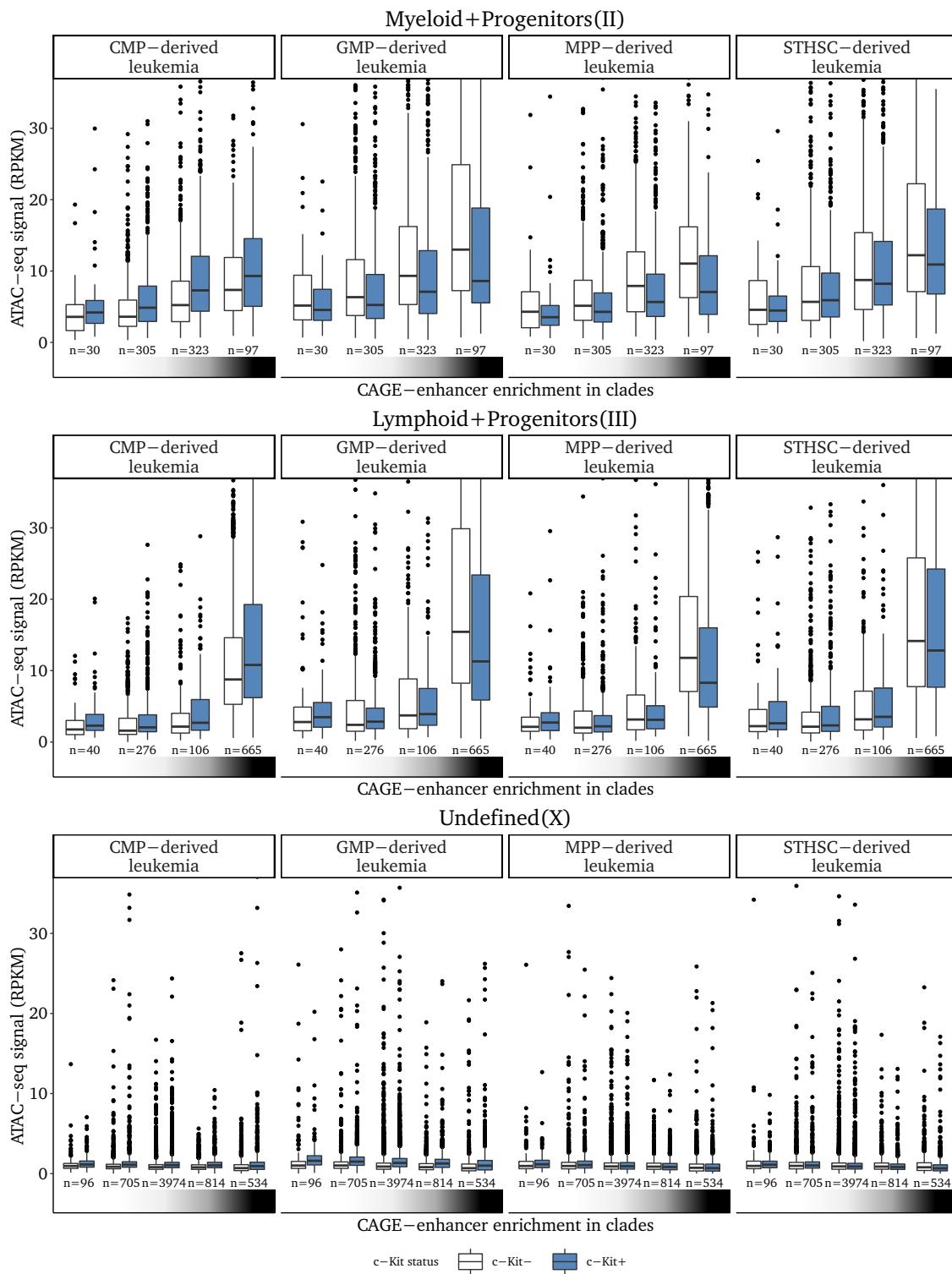


Figure 9.10: Open chromatin profiling by ATAC-seq at sites of putative enhancers in MLL-AF9 leukemia. Clades from the clusters *Myeloid + Progenitors* (II), *Lymphoid + Progenitors* (III) and *Undefined* (X) were separated depending on their odds ratio (≤ 0.25 ; $]0.25, 0.75]$; $]0.75, 1.25]$; $]1.25, 4]$; >4) as well as their test significance ($FDR < 0.01$) into five categories: strong depletion, mild depletion, no significance, mild enrichment and strong enrichment. The clusters II and III however lack the mild enrichment category. The four initial cell types, which the Trowbridge group transduced with the MLL-AF9 fusion gene to generate leukemia are shown as distinct columns.

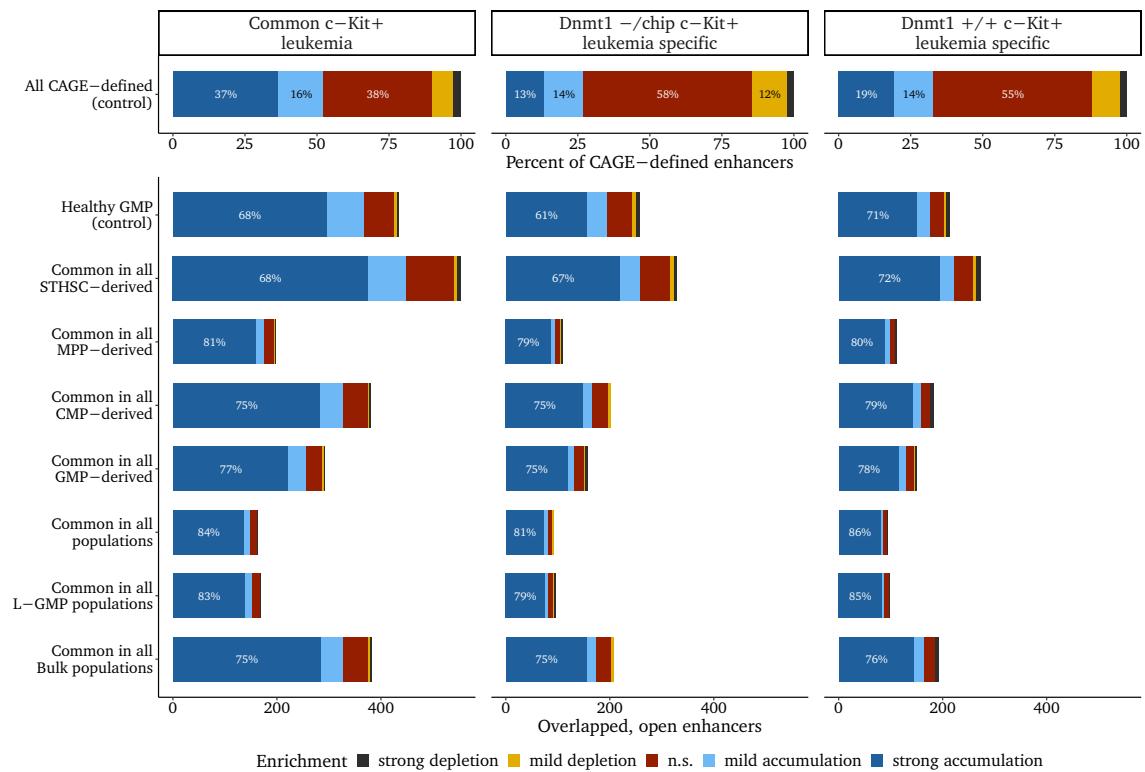


Figure 9.11: Absolute number of CAGE-defined candidate enhancers and their respective assignments to clades. Odds ratio (≤ 0.25 ;]0.25, 0.75];]0.75, 1.25];]1.25, 4]; >4) as well as test significance (FDR < 0.01 determined the five categories ranging from depletion to enrichment.

also featured elevated ATAC-seq signals. Furthermore the activity of such enhancers was ubiquitous in the sense that the openness of these chromatin regions did not depend on the cell of origin [▷ Figure 9.10, columns]. Notwithstanding a cell type specificity reflected in the ATAC-seq data [22], the enhancers contained within the highly enriched clade were uniformly marked by strong signals and thus could be attributed to the core signature of MLL-AF9 leukemia [▷ Figure 9.10, columns].

When intersecting our candidates with the ATAC-seq defined enhancers, the majority of confirmed putative CAGE-defined enhancers were assigned to clades exhibiting strong accumulation. Apart of the healthy GMP control, where it was lower, between 67% to 86% of the confirmed enhancers were constituted by strongly enriched clades [▷ Figure 9.11]. Importantly, it was not the very same group of enhancers, which affected the results over and over again in various contexts. Instead, as illustrated on the heatmap [▷ supplementary figure], we observed a quite heterogeneous mixture of ubiquitous as well as more specific enhancers, yet commonly assigned to accumulated clades, which were apparently involved in the leukemic transformation.

Chapter 10

Enhancer motifs, targets and regulation

Contents

10.1 Motif analysis	87
10.1.1 Basic procedure	87
10.1.2 Motifs enriched in strongly accumulated clades	88
10.1.3 Motifs enriched in strongly depleted clades	90
10.2 Methylation of enhancers and their motifs	91
10.2.1 Methylation mapping at enhancer regions	92
10.2.2 Methylation mapping at isolated motifs	93
10.3 MLL2 (Kmt2b) binding at strongly enriched enhancers	95
10.4 Enhancer target genes	98
10.4.1 Assessment of Mll2 target genes	101
10.5 Summary and outlook	102

In the previous chapter, it was shown that between 67 % to 86 % of the putative CAGE-defined enhancers confirmed by ATAC-seq in MLL-AF9 leukemic cells were constituted by strongly enriched clades. Since the clades are composed of congeneric enhancers [↔ subsection 9.2.2, p.75], we hypothesized that the transcription factors mediating their activation, would be specifically relevant in MLL-AF9, albeit possibly not specific to leukemia.

10.1 Motif analysis

10.1.1 Basic procedure

Transcription factors, which bind directly to DNA exhibit affinity for a particular geometry of the DNA-helix, typically determined by a combination of the underlying sequence, its methylation, its coil and bound accessory proteins [249, 250]. Although binding is often considered binary (e.g. ChIP peaks), transcription factors in reality bind in proportion to their affinity and weak interactions actually confer most of the regulatory activity [251]. While the higher structure of the DNA is hard to predict¹ [253], the sequence of a genomic segment is well tangible and typically informative on its own.

¹ exemplified by the c-Kit promoter [252]

To analyze, which transcription factors might be involved in the regulation of our candidate enhancers, we derived de novo sequence motifs associated with the strongly enriched clades with the HOMER software. To do so, we built ten separate contrasts, each for every major H3K4me1-derived cluster I - X [↔ subsection 9.2.1, p.73]. Within each cluster, we used the active CAGE-defined enhancers from the clades with strong accumulation as positive set and the CAGE-defined sequences from the depleted clades as control. This ensured that only transcribed enhancers were compared to transcribed enhancers. Thus, a possible bias due to enhancer vs. random sequence contrast or due to CAGE-defined vs. histone-mark-defined enhancer comparison was avoided. Subsequently, we united the enriched motifs for the ten cluster-specific sets and merged highly similar de novo motifs. The consolidated new motifs were united with the HOMER enhancer motif reference into one unified, curated library.

Aforementioned library was used to screen the CAGE-defined candidate enhancers² for presence and spatial location of the motifs (if present) under investigation. Obviously, relevant motifs should exhibit a clear enrichment over the background. Additionally, the highest frequency of relevant motifs should be focused around the enhancers' centers. We came up with the term centrality to refer to the latter property. We subsumed that sequence motifs, which are targeted by pioneering factors should exhibit the highest centrality, because their binding recruits secondary transcription factors and ultimately determines the position of the nucleosome-free chromatin region.

10.1.2 Motifs enriched in strongly accumulated clades

Except one notable case, all relevant motifs in the clades with strong accumulation were also characterized by a high centrality [▷ Figure 10.1, left panel].

The motif *DeNovo.CACTTCCTGT* was representative of this general trend [▷ Figure 10.1, rufous color]. Its mean frequency was among the top five and its maximum the second highest. Furthermore, its centrality was high, the maximum was located just a few bases off the center. Such striking resemblance to second most relevant motif *PU.1.ThioMac-PU.1.ChIP.Seq.Homer*, a known pioneering factor with high relevance for acute myeloid leukemia [107, 254], was clearly no coincidence. Both motifs comprised the core sequence **CACTTCC**. In this respect, it is very likely that the de novo motif was ultimately also a PU.1 binding motif with slightly different flanking bases, since the importance of PU.1 for MLL-rearranged leukemia was already known [255].

Since we did not experimentally validate the motifs (e.g. by ChIP experiments), no definitive assignments could be made, but the second-tier match for *DeNovo.CACTTCCTGT* was *Ets1.like.CD4.PoII.ChIP.Seq.Homer* [▷ Figure 10.1, ultramarine color], with which it shared a core sequence of **TTCCT**. ETS is a large family of transcription factors [256] and comprises 28 genes in the mouse. Appearance of this motif in a set of enhancers putatively linked to leukemogenesis was not surprising, as the founding member of this transcription factor family was initially identified as a leukemia oncogene transduced by the virus E26 [257].

² 2621 from the strongly enriched as well as the 2500 from the strongly depleted clades

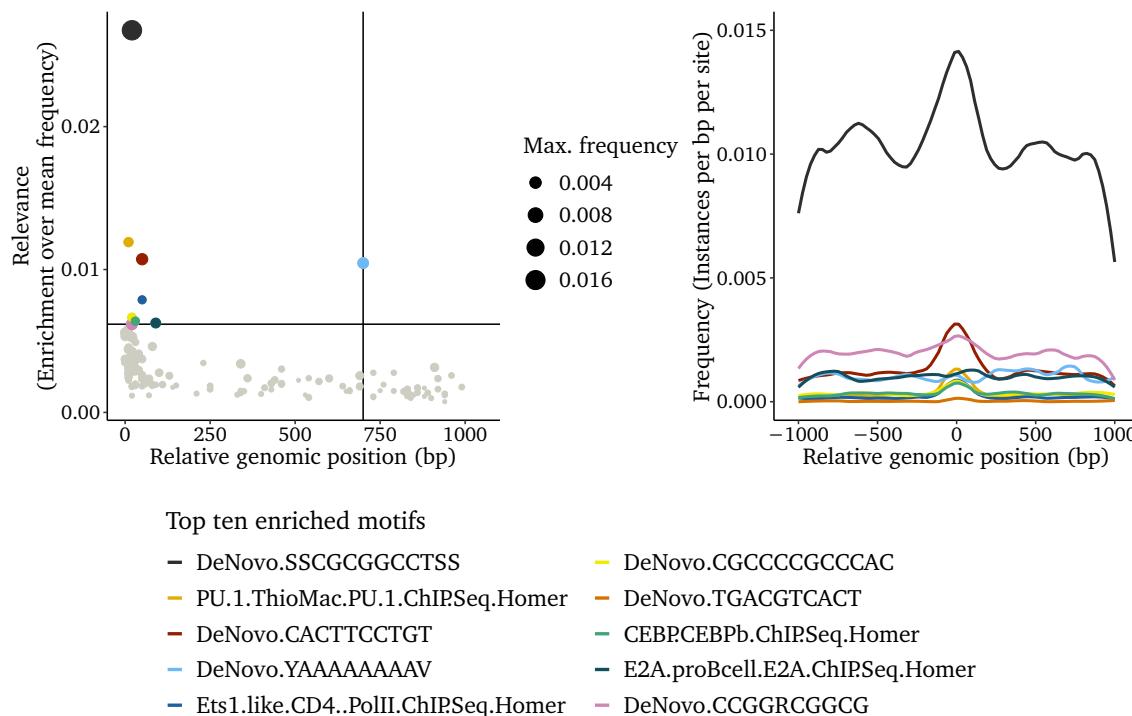


Figure 10.1: Details of the top ten enriched motifs, which were associated with the 2621 putative enhancers in clades with strong accumulation. In the left panel each motif is represented as a point in a coordinate system with the axes relevance and centrality. The relevance of the motif corresponds to the fraction of maximum (shown as dot size) divided by average frequency. The centrality refers to the genomic location of said maximum relative to the center of the enhancer. The latter is more clearly depicted in the right panel, which shows the course of the aggregated frequency within a 2 kb genomic segment around the enhancers' centers.

For this reason, *DeNovo.CACTTCCTGT* could be considered to belong to a ETS family transcription factor, very likely PU.1.

Another enriched de novo motif was *DeNovo.TGACGTCACT*, which however was appreciably rare [▷ Figure 10.1, fulvous color]. This motif strongly resembled the recognition sequence **TGACGTCA** of the basic leucine zipper domain (bZIP domain), which is found in many eukaryotic DNA binding proteins. bZIP transcription factors dimerize when binding to DNA and represent an extremely old class of transcription factors dating back more than a billion years in evolution [258]. Therefore, many transcription factors, such as the activator protein 1 (AP-1) could potentially bind there [259]. However, the motif is most likely recognized by CEPB in our cells, since also a similar reference motif *CEBP.CEBPb.ChIP.Seq.Homer* was enriched. Furthermore it was already shown that C/EBP α is frequently mutated [260] and co-occupies open chromatin regions with PU.1 in MLL-AF9 leukemia [247].

In this respect, the two de novo motifs with the most tangible binding sequences could be straightforwardly assigned to two transcription factors with well known involvement in leukemia and hematopoiesis. While this could be taken as a confirmation that we had

actually identified enhancers relevant to leukemia, it was of course disappointing at the same time, since it left little room for new discoveries.

The motif **DeNovo.YAAAAAAAV** was exceptional, as it was the sole top ten motif with a low centrality and a quite uniform distribution over the whole 2 kb range. Yet, its presence in the vicinity of enhancers was reasonable, since polyadenylation of eRNAs or related small RNAs does occur in some cases [reviewed in [261](#)].

The remaining motifs consisted of predominantly CG-rich sequences³, rarely interspersed with adenines or thymines. So, we puzzled over whether we could regard all three as basically identical. Undoubtedly, however, one of the three motifs was significantly more frequent than the other two [\triangleright [Figure 10.1, black color](#)], challenging complete equivalence. It also dominated all others by a clear margin in terms of relevance and centrality, which was quite remarkable given its somewhat uncommon sequence composition. It should be noted at this point that the motif matches were just short stretches of CpGs, which did not meet the usual length requirements to be considered as regular CpG-Islands (CGIs).

Because of the CG-rich sequences, we suspected that the motif **DeNovo.SSCGGGCCCTSS** as well as the two less frequent cognate candidates (**DeNovo.CGCCCGCCAC**, **DeNovo.CCGGR-CGGCG**) could be recognized by a protein comprising a CXXC zinc finger domain. Although subgroups of different DNA-binding specificities exist [[264,265](#)], this domain generally binds to unmethylated CpG-dinucleotides and is found in a variety of chromatin-associated proteins, such as MLL1 [[266](#)]. Because it is also retained in all known MLL fusion proteins [reviewed in [5](#)], we suspected that those motifs might be directly bound by MLL-AF9. However, when we reanalyzed a published MLL-AF9 ChIP-seq [[245](#)], we could not observe direct binding [\triangleright [data not shown](#)]. Later, we could identify Kmt2b (Mll2) as the key methyltransferase to cling to **DeNovo.SSCGGGCCCTSS** and **DeNovo.CCGGRCG-GCC**, but not to **DeNovo.CGCCCGCCAC** [\leftrightarrow [section 10.3](#)].

10.1.3 Motifs enriched in strongly depleted clades

We also ran a similar analysis for the strongly depleted enhancers, hoping to identify motifs relevant to enhancer decommissioning in MLL-AF9 leukemia. However, no known transcription factor motif arose in the top ten motifs and even the identified de novo candidates were extremely rare [\triangleright [Figure 10.2, right panel drawn at scale with Figure 10.1](#)]. Because of the very low average frequencies, the absolute values for the relevance score were comparably high (as they are the ratio of the maximum divided by the average frequency). None the less, the motifs were without any practical biological significance.

³ Shorter nucleotide tuples rich in CG are a general feature of transcribed enhancers in humans [[262](#)]. The overall CG content of enhancers, however, varies depending on cell type [[263](#)].

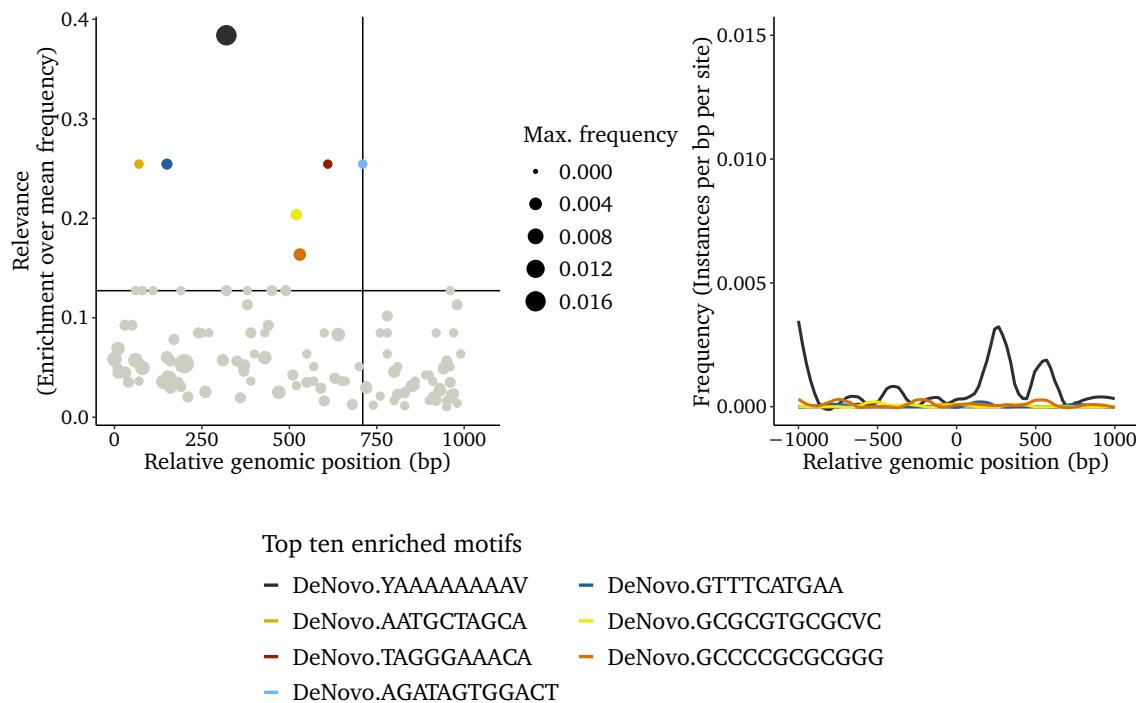


Figure 10.2: Ten most enriched sequence motifs in the 2500 putative enhancers originating from strongly depleted clades. Analogous to Figure 10.1, the left panel depicts centrality and relevance as well as the maximum frequency of the motif. The centrality refers to the genomic location relative to the center of the enhancer, where the maximum frequency was recorded, whereas the relevance expresses the ratio of said maximum relative to the average frequency. The right panel shows the change of the aggregated average frequency along the genomic region surrounding the enhancer.

10.2 Methylation of enhancers and their motifs

DNA methylation is a crucial regulatory layer for normal and malignant hematopoiesis [reviewed in 30, 169] and a growing body of papers stresses the importance of influential methylation changes at cis-regulatory elements in health and disease [139, 227–231].

Identification of a potential CXXC zinc finger motif within the strongly accumulating enhancers suggested an investigation of the methylation status, since CXXC binds exclusively to unmethylated CpG-dinucleotides [266]. Also many other transcription factors are known to bind in a methylation sensitive manner [267, 268]. Decisive regulatory methylgroups do not necessarily have to be located directly at the binding site of the transcription factor: A particularly interesting paper had shown how methylation at distant sites facilitates the efficiency of Egr1 target search process [269].

Therefore, we hoped that decisive methylation changes in those regulatory regions might be the long sought answer to explain the Dnmt1^{-/-} phenotype, in particular its self-renewal bias observed in leukemic stem cells (LSC) [118].

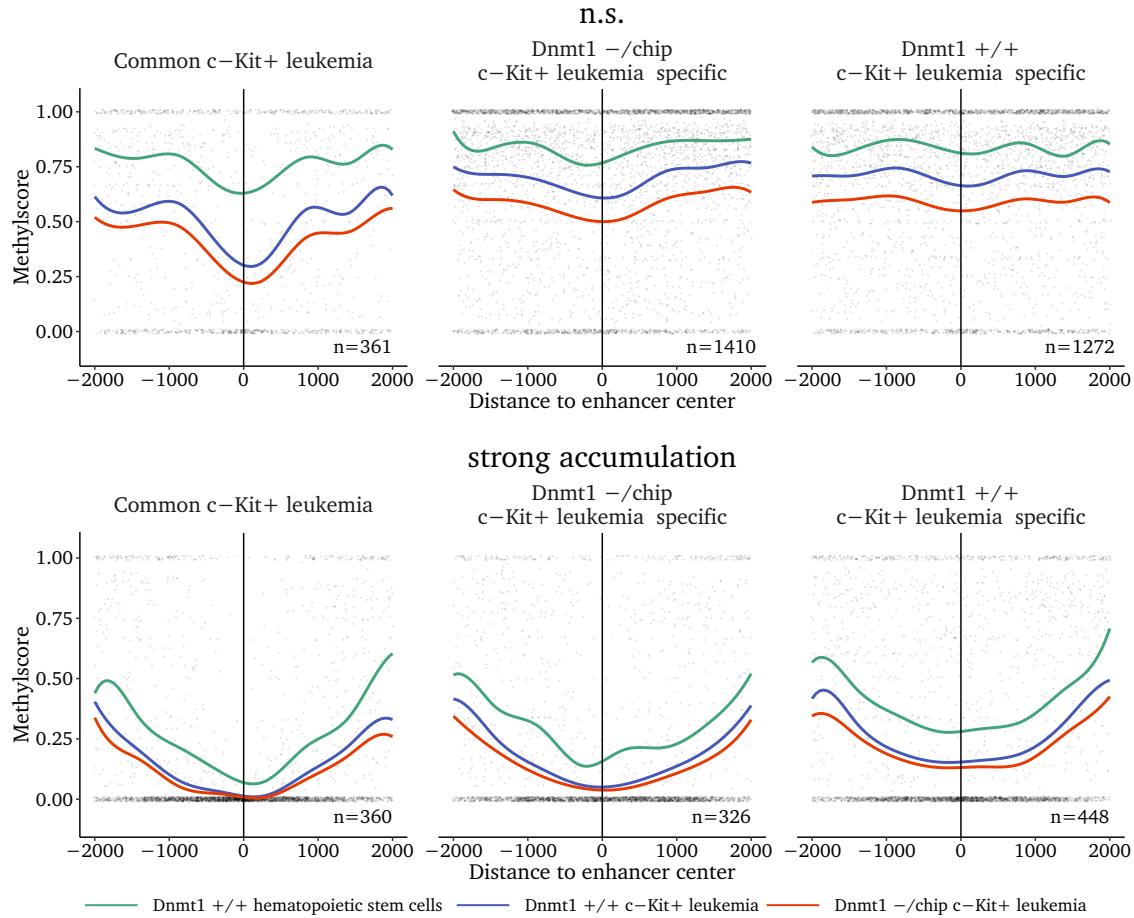


Figure 10.3: Methylation in a 4 kb window surrounding CAGE-defined putative enhancers from two clade groups. The top row depicts Colored lines represent the smoothed average methylscore in the three meta-samples, which are displayed on top of the measured methylation rate of single CpGs (black dots). CpGs without sufficient WGBS coverage (3 reads) are not shown. Furthermore, only candidate enhancers are considered, which feature at least one covered CpG within the window under investigation.

10.2.1 Methylation mapping at enhancer regions

When we mapped the WGBS meta-samples [↔ section 2.1, p.20] on the putative enhancer regions, we found that they were generally hypomethylated compared to the surrounding backbone regions in accordance with published literature [139]. However, there were notable differences between the various enhancer groups [▷ Figure 10.3].

Enhancers, which were assigned to clades with depletion [▷ data not shown] exhibited a methylation pattern similar to enhancers from the non-significant clades [▷ Figure 10.3, top row]. Among those, enhancers active in Dnmt1^{+/+} as well as Dnmt1^{-/-chip} leukemia exhibited a local methylation minimum located right over the center of the cis-regulatory element. Said minimum was seen in both leukemia and also the normal hematopoietic stem cell (HSC). Yet, in comparison to the stem cell, these sites in leukemia showed the highest degree of demethylation observed for any enhancer. Remarkably, this did not apply to the genotype-specific sets, which were insignificantly hypomethylated at all

[> Figure 10.3, top middle and top right panel]. Since Dnmt1^{-/chip} leukemia exhibited the least methylation in any group, we could rule out that genotype-specificity arose from differential methylation.

A slightly different methylation pattern could be observed in clades with strong accumulation of CAGE-defined enhancers. Here, the dent was rather wide, shallow and overall methylation levels were extremely low [> Figure 10.3, bottom row]. In leukemia, the common sites were completely devoid of methylation and only sparsely methylated in the healthy hematopoietic stem cell [> Figure 10.3, bottom left panel]. The genotype-specific sites were characterized by marginally higher levels of methylation, with Dnmt1^{+/+} specific sites exhibiting the highest. Nevertheless, all three were still lower methylated than the enhancers of the non-significant clades.

Taken together, we could observe differential methylation in various groups of putative, CAGE-defined enhancers. However, within the most relevant group, the clades with strong accumulation, the differences were small. Here, the sites were typically unmethylated in leukemia and seldom methylated in the HSC. However, we presumed that the methylation of single motifs might exhibit more distinct patterns.

10.2.2 Methylation mapping at isolated motifs

This section is printed in condensed form. Optionally, a version with additional information regarding methodical details and quality control is available as online supplement.

Contrary to whole enhancers, motifs are small and never occur isolated in the genome. A recent study found, that the average random 80-mer nucleotide sequence has about 138 binding sites for 68 different transcription factors [251]. To account for confounding motifs during methylation analysis, we first searched for frequently co-occurring motifs and then tried to model their reciprocal influence on methylation. For this purpose, we used item set mining with the APRIORI algorithm [270] to derive frequent motif patterns. Several frequent combinations could be identified, but low coverage in our WGBS data (2 % to 21 %) mostly hindered proper consideration of co-occurring motifs, because seldom all motifs of a set were covered in a particular instance. Ultimately, a different approach based on fitting a Kumaraswamy distribution separately enabled the identification of motifs with dynamic methylation [> supplement].

The most dynamic methylation was found for the motifs *Ets1.like.CD4.PoII.ChIP.Seq.Homer* as well as *DeNovo.TTCCCCWTTYG* [> Figure 10.4, top and middle row]. However, both motifs were sparsely covered in the WGBS data, plus *DeNovo.TTCCCCWTTYG* was relatively rare ($n_{\text{CAGE-defined}} = 227$, $n_{\text{control}} = 5463$). Nevertheless, the counts were commensurate, since *Ets1.like.CD4.PoII.ChIP.Seq.Homer* was about three times more frequent in both categories ($n_{\text{CAGE-defined}} = 651$, $n_{\text{control}} = 16663$). Assuming that the few covered instances of the two motifs are representative, we observed an almost complete demethylation in CAGE-defined enhancers, but an ambiguous methylation in the controls.

DeNovo.sscggggctss, the most frequent motif in CAGE-defined putative enhancers as-

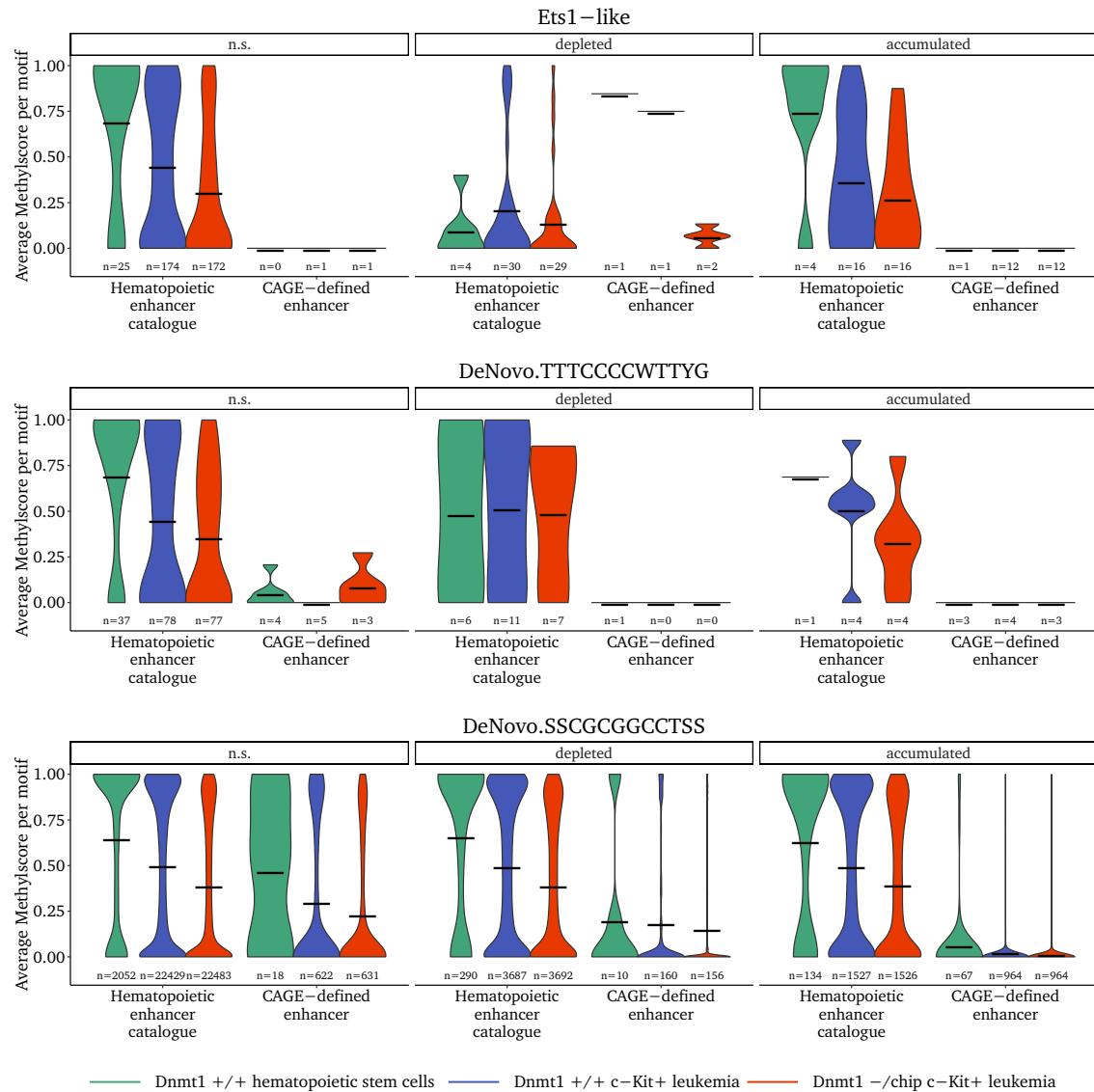


Figure 10.4: Detailed representation of three selected motifs and their methylation dynamics. For this plot, motif instances had been split among the enhancer groups and the WGBS meta-samples were mapped. The average methylation was calculated per motif instance and all covered sites (see counts below) were included in the violin plots. Methylscore distributions are depicted as vertical density plots and the methylation mean of the respective motifs as horizontal black bar.

signed to strongly accumulated clades [→ subsection 10.1.2], was covered in 21.67 % of said enhancers in leukemia. At large, all instances within active sites were unmethylated in MLL-AF9 leukemia, but still partially methylated in hematopoietic stem cells (HSCs) [▷ Figure 10.4, bottom row]. Remarkably, the degree of methylation HSCs, but not in leukemia, varied depending on the clades. The motif was mostly demethylated if found in the accumulated clade enhancers and ambiguously methylated elsewhere in CAGE-defined enhancers. In the control set derived from the hematopoietic enhancer catalog, the motif was ambiguously methylated, too. Generally, it exhibited the highest average methylation in HSCs and the lowest in Dnmt1^{-/chip} leukemia. Clade enrichment did not matter for the methylation of control enhancers.

This pattern suggested an active regulation of the motif's methylation in the hematopoietic system. Therefore, it was intriguing to speculate that the variable methylation might alter the binding of a CXXC protein and we aimed to identify said protein.

10.3 MLL2 (Kmt2b) binding at strongly enriched enhancers

Because of the CG-rich sequence, we speculated that the motif *DeNovo*.SSCGGGCTSS could be recognized by a CXXC zinc finger domain, which binds to various motifs of unmethylated CpG-dinucleotides [264]. Since this domain is contained in a variety of chromatin-associated proteins, such as MLL1 [266], we initially suspected that *DeNovo*.SSCGGGCTSS might be directly bound by MLL-AF9. However, when we reanalyzed a published MLL-AF9 ChIP-seq [245], we could not observe direct binding [▷ [data not shown](#)]. Subsequently, we tested further published ChIP-seq datasets of other CXXC proteins like the CXXC-type zinc finger protein 1 (Cfp1), which, despite being crucially involved in hematopoietic regulation [271] and H3K4me3 deposition [272], also did not bind to the motif [▷ [data not shown](#)]. Ultimately, the search for the correct binding partner stalled.

Fortunately, in 2017, the laboratory of Ali Shilatifard published the results of a study aimed at deepening knowledge about the COMPASS family protein Kmt2b (Mll2) and its role in embryonal stem cells (ES cells) [273]. It was already known to be implicated in the regulation of bivalent promoters in the stem cells [274], but now the group showed that it also implements H3K4me3 at a subset of non-TSS regulatory elements [273]. Because the properties of these sites ⁴ strikingly resembled CAGE-defined putative enhancers from the strongly accumulated clades III.2, III.12, III.13 [▷ [Figure 9.7, p.81](#)], we assumed that the motif *DeNovo*.SSCGGGCTSS might be the target of Kmt2b/ Mll2.

Therefore, we downloaded the comprehensive datasets [↔ [Appendix A, p.133](#)], which accompanied the study [273] and mapped them on the CAGE-defined putative enhancer regions. Although the data originated from murine embryonic stem cells instead of MLL-AF9 leukemic cells, we hypothesized that the dataset could still be informative based on the mostly universal H3K4me3 methylation in many hematopoietic cell types [▷ [Figure 9.6, p.79](#)]. Indeed, clades, which comprised many CG-rich *DeNovo*.SSCGGGCTSS enhancers, were typically H3K4me3-positive in ES cells [▷ [Figure 10.5](#).] While enhancers of the strongly accumulated clades within the clusters *Lymphoid + Progenitors* (III), *Erythroid + Progenitors* (IV) or *B-cells* (VII) exhibited H3K4me3 marked nucleosomes, no association could be observed in the other clusters [▷ [data not shown](#)]. This was in accordance with the frequency of the motifs *DeNovo*.SSCGGGCTSS and *DeNovo*.CCGRCGGCG in those clades and suggested that the mark is implemented by Mll2 at those sites.

However, the results were contradictory in this respect, since closer inspection revealed that H3K4me3 was also present in the sample *Mll2* -/- +*Mll2-Y2604A*, which solely expresses a Mll2-variant with a dysfunctional catalytic subunit [▷ [Figure 10.5, medium sea green bar](#)]. Also a disrupted targeting (*Mll2* -/- +*Mll2ΔCXXC*) did not impact H3K4me3

⁴ CG-rich, high H3K4me3, but low H3K4me1, high H3K27ac

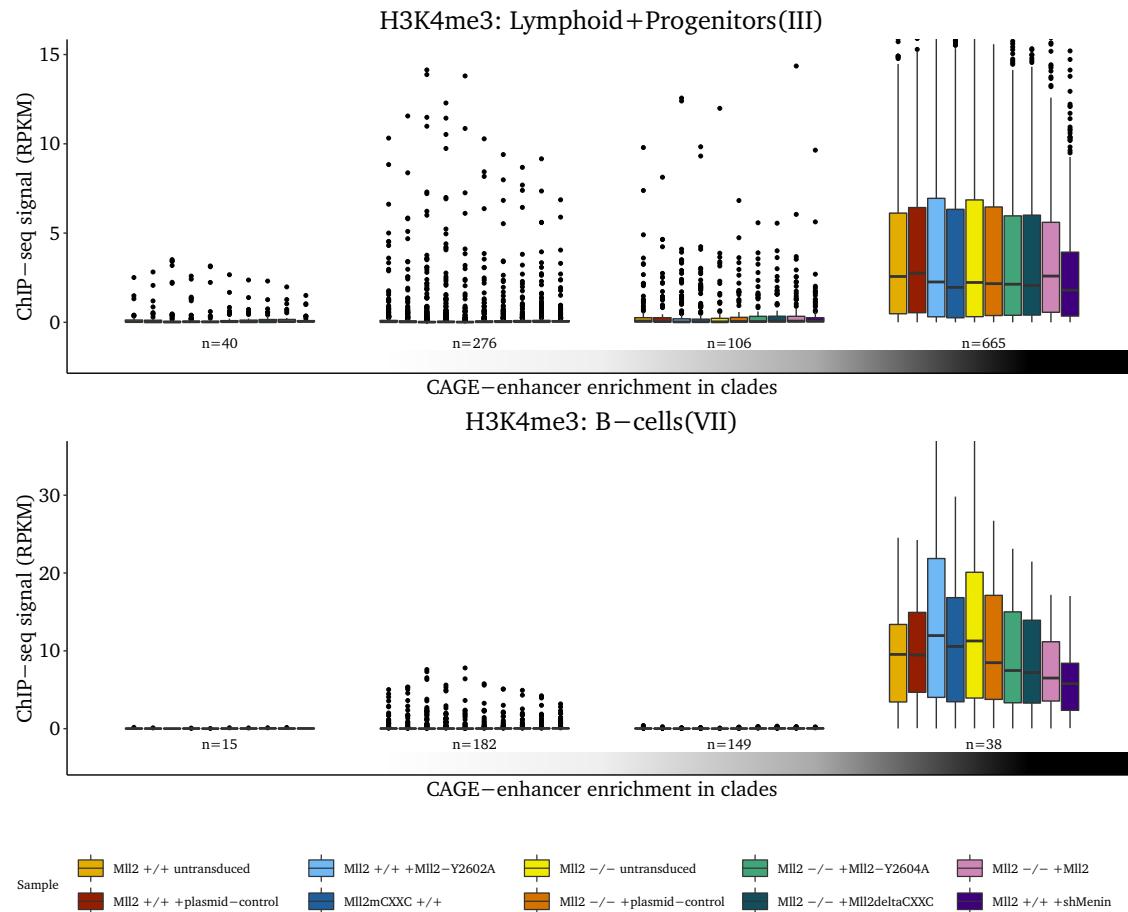


Figure 10.5: H3K4me3 ChIP-seqs in murine embryonic stem cells of different genotypes and under various experimental conditions. Data was mapped to the CAGE-defined enhancers of the *Lymphoid + Progenitors* (III) and *B-cells* (VII) clusters and counts were normalized to the number of mapped reads per sample. Clades are shown ordered by the accumulation of putative leukemic enhancers.

deposition noticeably. Thus, Mll2 could not be the sole histone methyltransferase targeting those sites and its function is possibly safeguarded by one of the other five Trithorax group (TrxG) proteins in mice [275].

While the H3K4me3 results were still ambiguous, the Mll2-ChIP-seqs clearly substantiated our hypothesis. Binding of Mll2 was preferably detectable in the strongly accumulated clades of the clusters *Lymphoid + Progenitors* (III), *Erythroid + Progenitors* (IV) or *B-cells* (VII) but not elsewhere [> Figure 10.6]. Furthermore, the binding was definitely mediated by the CXXC zinc finger domain, since the occupancy was strongly diminished for the *Mll2ΔCXXC* samples.

Nevertheless, we corroborated the role of the *DeNovo*.*SSCGGGCTSS* as Mll2 recognition site by a reciprocal analysis. Of the 6418 Mll2-positive non-TSS sites identified in embryonic stem cells by the group of Ali Shilatifard [273], we considered 374 (5.83 %) as transcribed enhancers in MLL-AF9 leukemia, the majority of which (313, 83.68 %) belonged to strongly accumulated clades in the clusters III, IV or VII. We compared this to a control

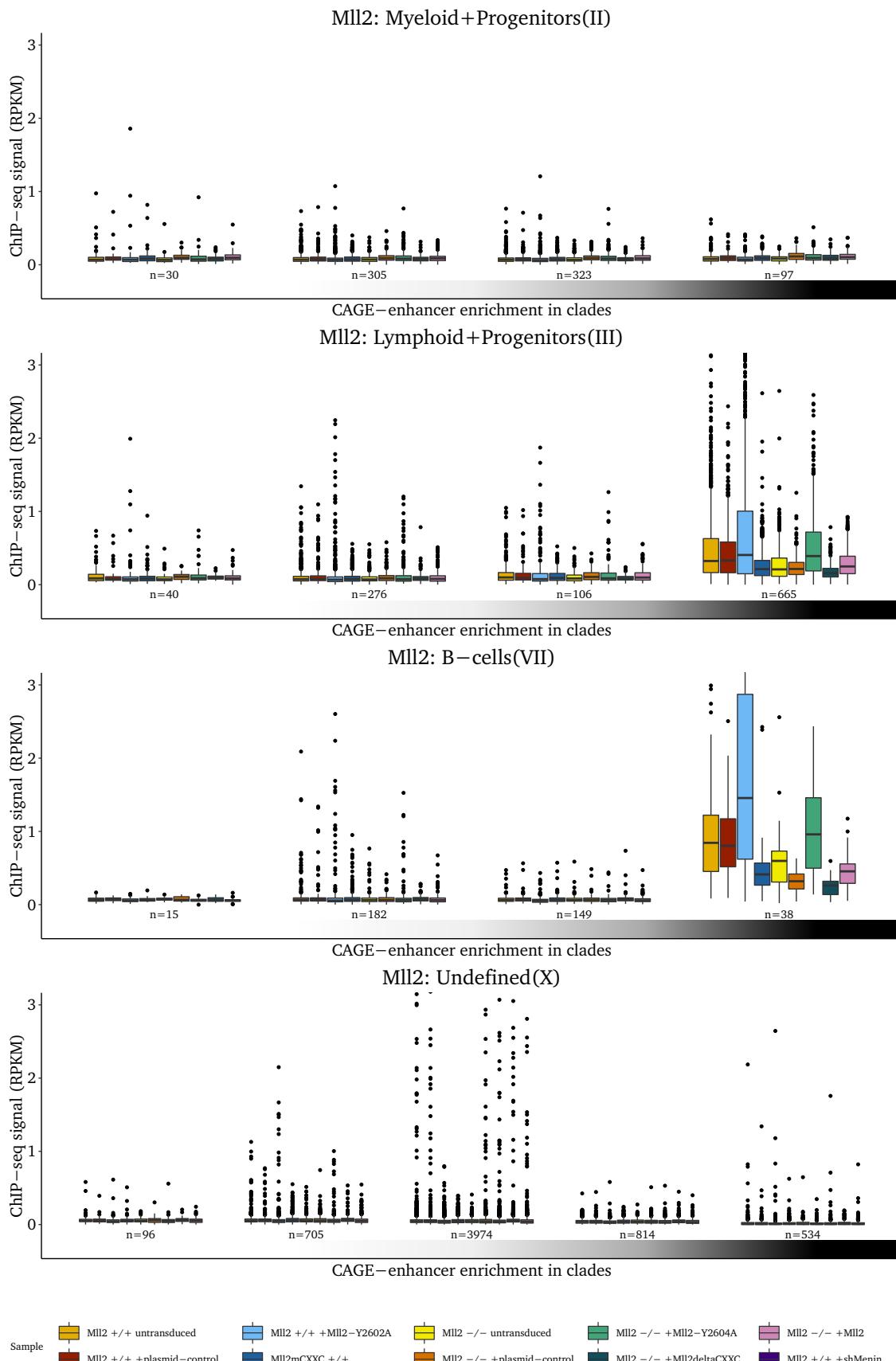


Figure 10.6: Normalized Mll2 occupancy at CAGE-defined enhancers in murine embryonic stem cells. Putative enhancers are split according to cluster and clade enrichment as explained in the previous chapter.

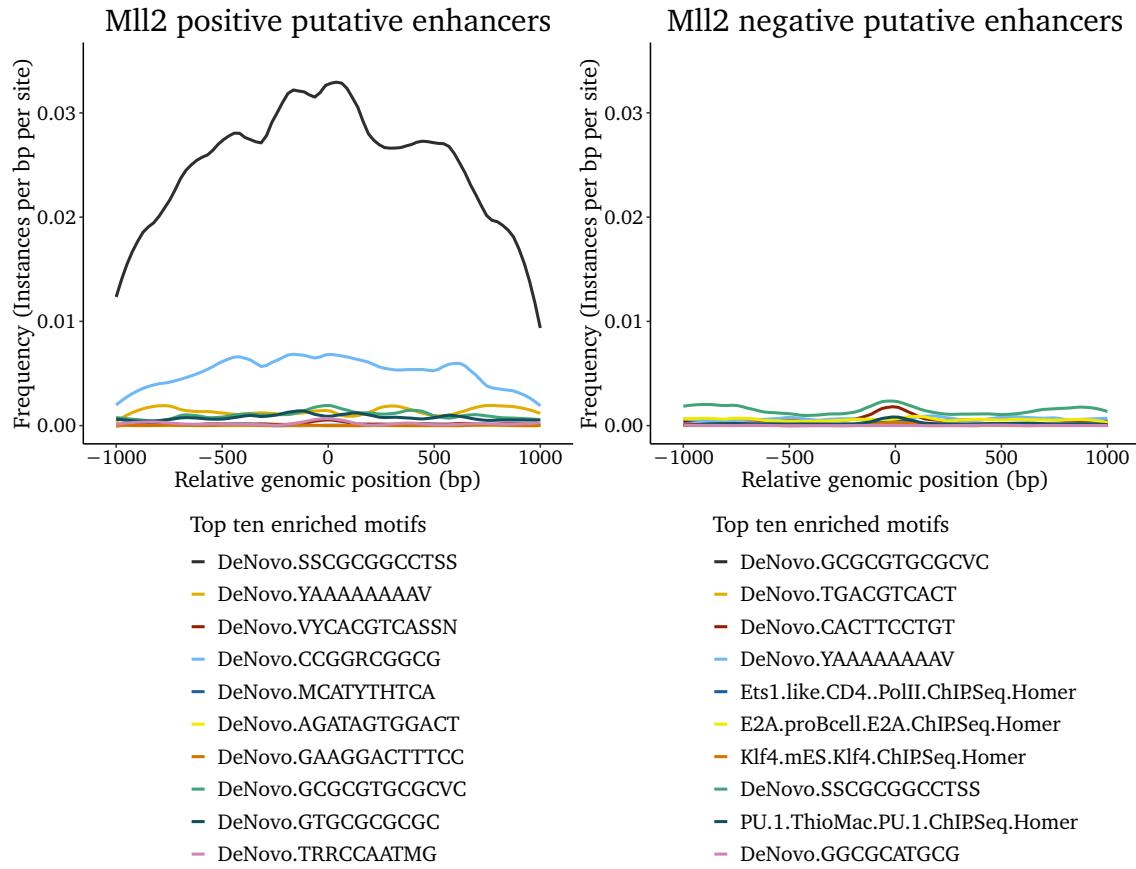


Figure 10.7: Aggregated frequencies of the top ten enriched motifs within a 2 kb genomic segment around the enhancers' centers. Shown are 374 transcribed enhancers, which are bound by Mll2 in embryonic stem cells and a equal number of matched unbound control enhancers.

set consisting of an equal number of randomly chosen of Mll2-negative CAGE-defined enhancers.

Then, we derived the top ten most frequent motifs from both sets and were able to establish a clear association between the motif *DeNovo.sscgccccctss* and the binding of Mll2 [▷ Figure 10.7]. To a much lesser extent, also *DeNovo.ccggrggcgg* seemed to be involved in the binding.

10.4 Enhancer target genes

Despite the clear enrichment, Mll2 seemed to be an unlikely candidate, since subtle differences between the CXXC-domains of Mll1 and Mll2 preclude an oncogenic potential of the latter in the context of fusion proteins [276]. While we were still double-checking the results and pondering, whether we should experimentally follow up on the topic, the laboratory of Patricia Ernst published a detailed study, which highlighted the importance of Mll2 for MLL-AF9 leukemia [277]. The study measured the effects of Mll2 knock-out by RNA-seq, but did not provide a mechanism. We were intrigued to see, if some of the genes were responding due to abridged binding at enhancers.

This section is printed in condensed form. Optionally, a version with additional information regarding methodical details and quality control is available as online supplement.

Therefore, we aimed at the identification of the enhancers' presumable target genes. For the establishment of reliable enhancer-promoter interactions, we utilized promoter-capture Hi-C data rather than just assigning the closest transcription start site. In total, we could derive 11 534 potential pairs, comprising 3103 putative enhancers and 4317 genes (6728 transcripts). After ordering the connections by score, many renowned hematopoietic regulators appeared in the top ranks, which suggested that the derived pairing scores accurately reflected the biology.

Among the top100 enhancer promoter interactions we could identify many renowned hematopoietic regulators with known involvement in leukemia (e.g. Irf2bp2, Pten, Fosl2, Spred1). Furthermore, the majority of involved enhancers in the top 100 originated from strongly accumulated clades (73 %), which corroborated their importance and thus validated our clustering strategy [→ [section 9.2, p.73](#)].

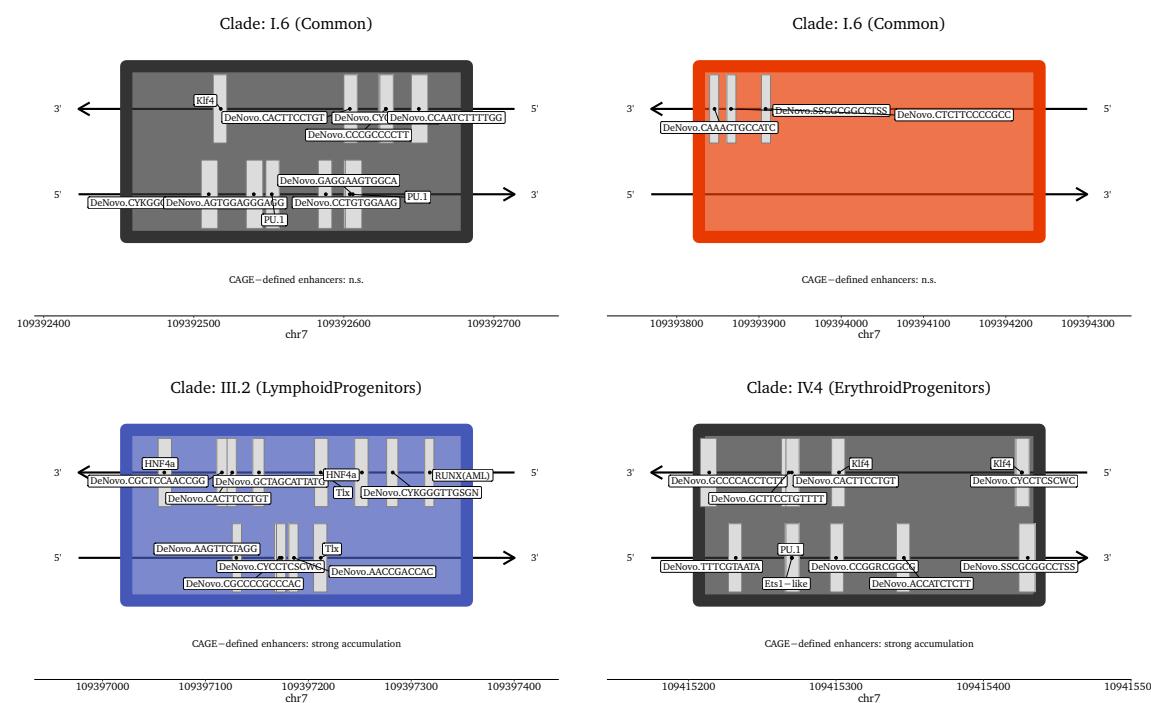


Figure 10.8: Schematic representation of the four enhancers presumably involved in regulating the expression of Rhog and Nup98 in MLL-AF9 leukemia. Genotype-specificity is indicated by the colors red (-/chip), blue (+/+) and black (common). Gray boxes symbolize approximate positions of transcription factor binding motifs. Note the **DeNovo.SSCCGGGCCTSS** motif in two of the four enhancers.

Accordingly, two of the four enhancers presumably involved in the regulation of Rhog + Nup98 contained a motif for Mll2 binding [▷ [Figure 10.8](#)]. Only the Hi-C promoter capture data allowed to decipher the complex regulation at this gene locus [▷ [Figure 10.9](#)],

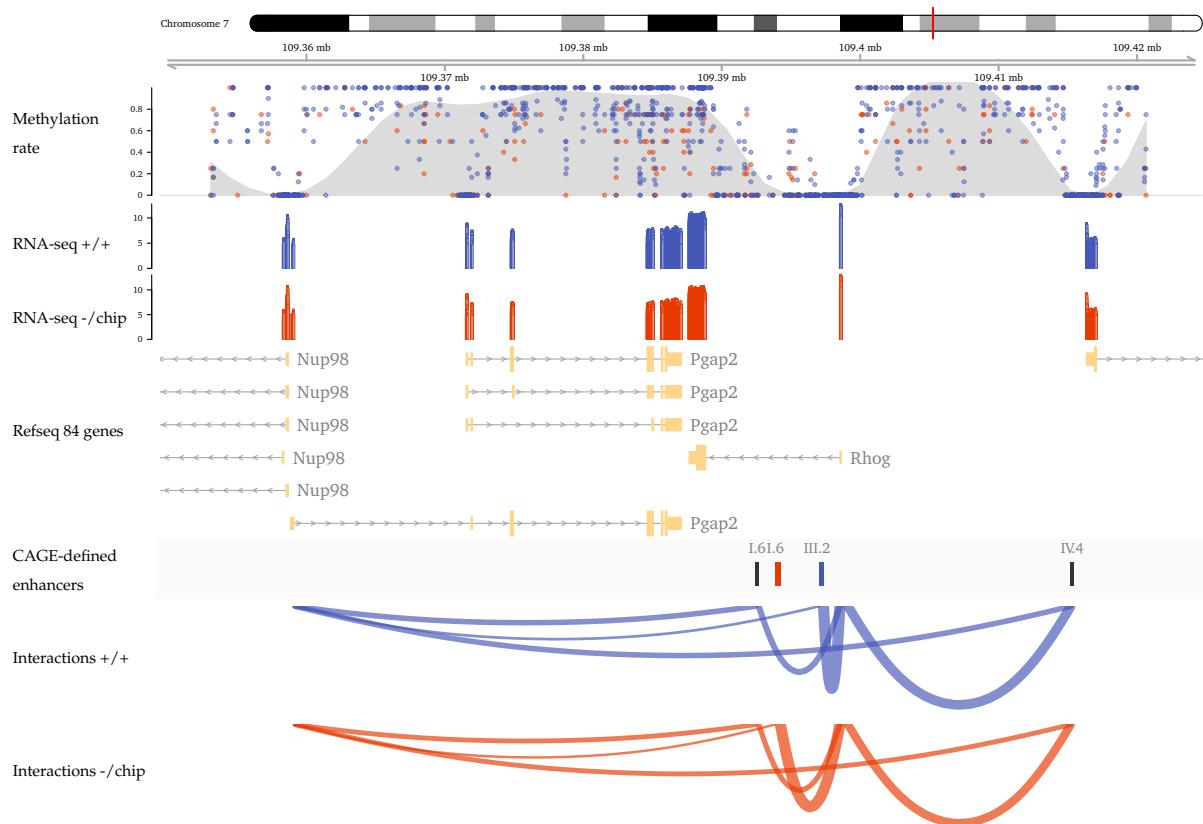


Figure 10.9: Representation of the second ranking gene locus comprising the promoters of Nup98, Pgap2, Rhog and Stim1 (from left to right). The top track contains a scatterplot representation of single CpG methylation rates as well as a LOESS smooth thereof depicted in gray. RNA-seq data is log-scaled to base 2 and (like all other relevant items) colored by genotype: Red is used for data referring to $Dnmt1^{-/-}chip$ and blue constitutes $Dnmt1^{+/+}$ items. Thickness of the arcs conveys the frequency of the respective interaction.

which deserved a particular attention, because both genes are notoriously implicated in leukemic development [278–281] [282–287]. Although further important hematopoietic regulators like Irf2bp2 or Ikzf2 were putatively targeted by several enhancers in MLL-AF9.

We corroborated the enhancer assignments with our RNA-seq data. As expected, the expression of transcripts that were assigned to enhancer(s) was significantly higher than that of expressed transcripts without an enhancer assignment [[supplementary figure](#)]. Intriguingly, we observed consistent downregulation of enhancer-assigned transcripts in $Dnmt1^{-/-}chip$ [[supplementary figure](#)]. As it was shown that hemi-methylation flanking CTCF motifs is highly relevant for the directionality of the binding [288], the downregulation suggested that regulatory enhancer promoter interactions in $Dnmt1^{-/-}chip$ might have been perturbed in select cases by differential methylation. For an in-depth discussion see [[section 13.4, p.127](#)].

10.4.1 Assessment of Mll2 target genes

This section is printed in condensed form. Optionally, a version with additional information regarding methodical details and quality control is available as online supplement.

Having established the enhancer promoter pairs in general, we specifically focused on the regulatory effects of Mll2. A detailed study from the laboratory of Patricia Ernst had highlighted the importance of Mll2 for MLL-AF9 leukemia [277], but no mechanism had been established. We reanalyzed the data from that study [→ [Appendix A, p.134](#)] to test, if some of the transcriptional effects observed after Mll2 deletion in MLL-AF9 leukemia could be attributed to enhancers rather than promoters.

15% of the genes, whose expression was altered as a result of Mll2 deletion likely responded in an enhancer-mediated manner. Most of these genes were involved in crucial cellular functions - among them e.g. Ras homolog family member G (Rhog) [▷ [Figure 10.9](#)], which was already described above.

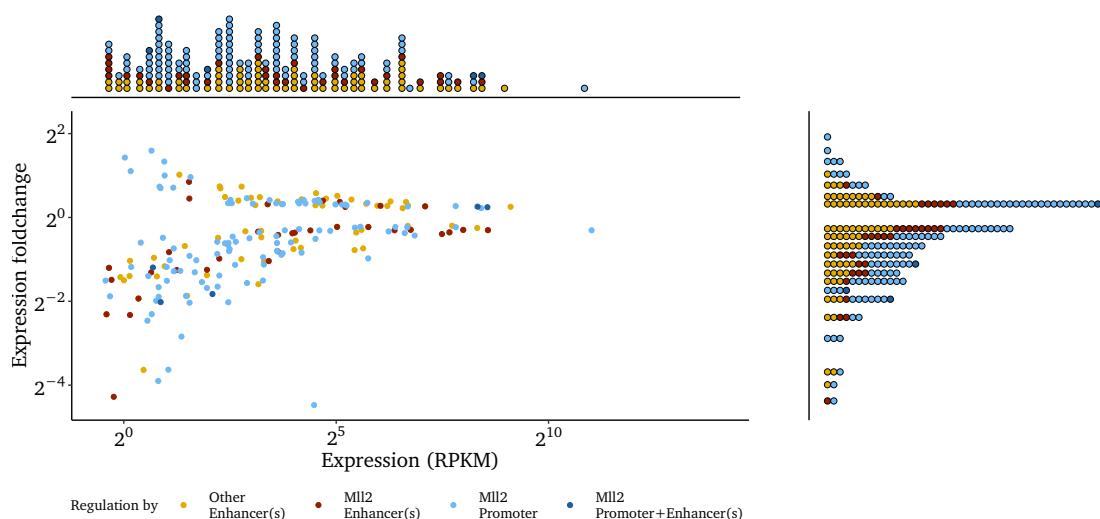


Figure 10.10: Dot plot of the expression pattern after knock-out of Mll2 in MLL-AF9 leukemic cells. Only significantly differentially expressed transcripts are shown. Colors indicate the regulatory assignment. On top and to the right, one-dimensional pile-up plots provide visual aids to assess the absolute number of the significantly differentially expressed transcripts and their respective assignment.

Approximately two-thirds of the differentially expressed transcripts responded to Mll2 loss by downregulation (151 down, 73 up). In terms of effect size, the observed expression change (particularly downregulation) was typically more prominent in the promoter category than in the enhancer category [▷ [Figure 10.10, pile-up graph to the right](#)]. This finding was likely attributable to potential redundant enhancers and clearly not related to a prior expression bias, since the transcripts could be found in the full range of the spectrum [▷ [Figure 10.10, top pile-up graph](#)].

Despite being small in relation to promoter-mediated regulation, both in terms of magnitude and number of affected transcripts, there was a noticeable effect of Mll2-enhancer deficiency. Functionally, some of the respondent genes were involved in crucial cellular

functions such that an effect on self-renewal and leukemogenesis seemed plausible. Nevertheless, none of the candidate genes was experimentally tested anymore. Also if there was a methylation-dependent impairment at those particular enhancers in *Dnmt1*^{c/chip} still warrants investigation.

10.5 Summary and outlook

This chapter describes the common mechanisms, which presumably govern the recruitment of the congeneric enhancers in the strongly accumulated clades [↔ subsection 9.2.2, p.75]. We derived motifs and inferred PU.1 (or another ETS transcription factor), C/EBP α (or another bZIP dimer) as well as Mll2 (or another CXXC protein) as the key transcription factors involved [↔ subsection 10.1.2] [↔ section 10.3].

Since the roles of PU.1 and C/EBP α in leukemia are well established, we were most excited by the identification of Mll2, particularly because the data suggested an dynamic regulation of the corresponding motif *DeNovo.sscggggctss* by DNA methylation [↔ subsection 10.2.2]. Indeed, the laboratory of Patricia Ernst shortly thereafter published a detailed study, which highlighted the importance of Mll2 for MLL-AF9 leukemia [277].

Since the study did not propose a mechanism, we consulted the STRINGDB database for possible interaction partners of Mll2 [▷ Figure 10.11]. Especially Kdm6a/UTX, which demethylates H3K27me2/ H3K27me3 and is implicated in the differentiation of natural killer cells [289], caught our attention due to the enhancers' strong H3K27ac signal in natural killer cells [↔ subsection 9.3.1, p.78]. Therefore, we conjectured that Mll2 recruits Kdm6a/UTX in NK cells (and possibly also in MLL-AF9 leukemia) to enable subsequent acetylation of lysine 27.

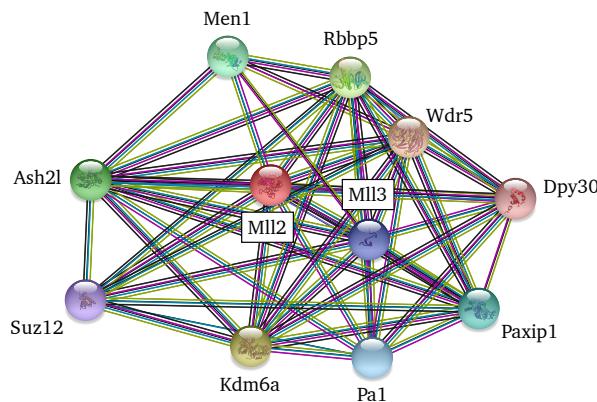


Figure 10.11: High confidence interactions of Mll2 with other proteins.

To test this, we treated MLL-AF9 leukemic cells in vitro with the inhibitor GSK-J4, a pro-drug of GSK-J1, which inhibits Kdm6a and Kdm6b effectively [290], but may also show activity against the Kdm5-family [291]. It emerged as a potential inhibitor for prostate cancer [292] and various hematological malignancies [293,294], however, a positive effect was questionable since inhibition of Kdm5c would have counteracted the effect [216].

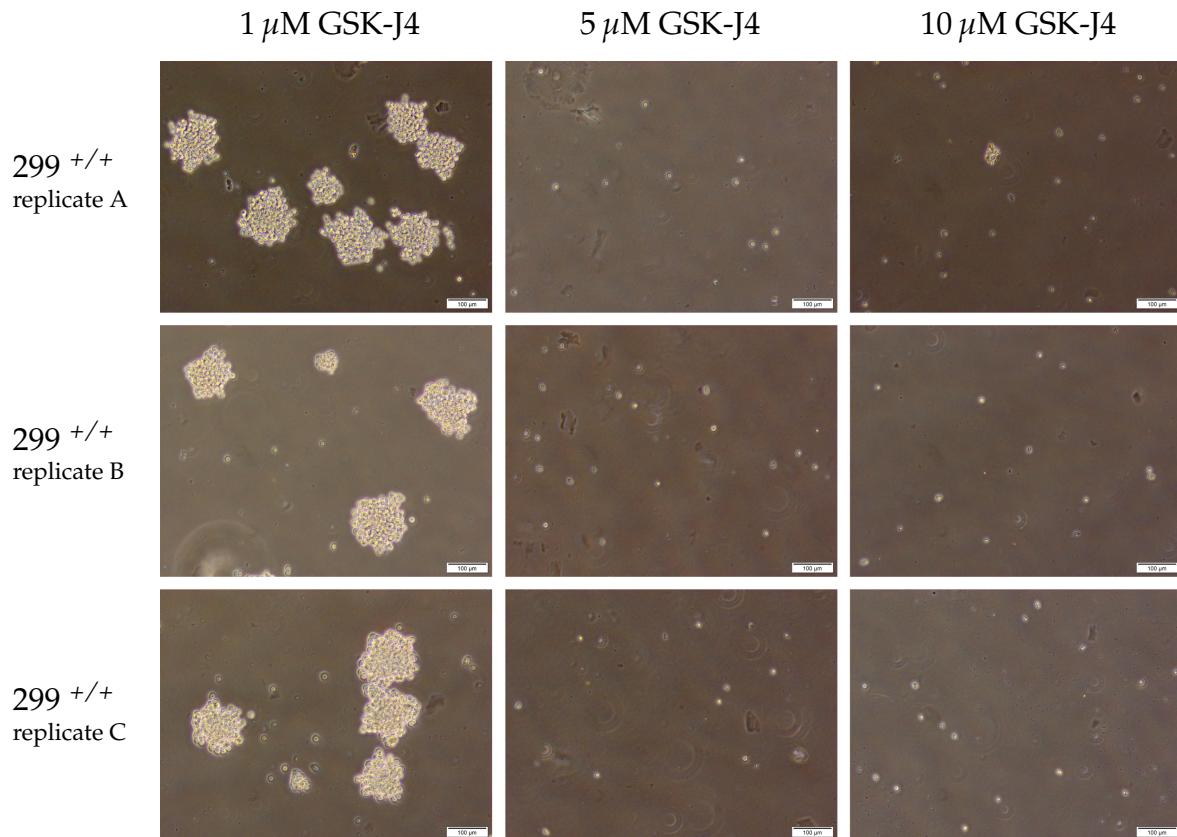


Figure 10.12: Bright-field micrographs of MLL-AF9 leukemic cells cultured on methyl-cellulose semisolid medium. Before starting the experiment, two serial replatings were performed in semisolid medium to enrich for leukemic stem cells. At day 0, 1000 cells per plate were seeded in methyl-cellulose/IMDM medium supplemented with 10 %FCS and cytokines IL-3,IL-6 and SCF as described before [118]. Appropriate amounts of GSK-J4 dissolved in DMSO or pure DMSO were added to the medium while seeding. Micrographs depict the status at day 3, DMSO control (not shown) corresponded to 1 μ M.

GSK-J4 inhibited the growth of MLL-AF9 leukemic cells at concentrations of 2 μ M to 5 μ M depending on the leukemic clone. The effects were consistent in liquid culture as well as on methyl-cellulose semisolid medium [▷ Figure 10.12]. However, it also noticeably affected normal hematopoietic control cells at just 10 μ M, casting doubts on its suitability for therapeutic application. Furthermore, another study around the same time reported efficacy of GSK-J4 for the treatment of AML (including MLL-AF9). Based on overexpression of Kdm6b and a greatly exaggerated specificity of GSK-J4, the authors proposed a Kdm6b-dependent mechanism without further proof [295].

Therefore, the next step upon continuation of the project would be to address this issue and to design experiments, which allow for the discrimination of Kdm6a- and Kdm6b-mediated effects.

Part IV

Discussion

Chapter 11

Synopsis of results

Using the *Dnmt1^{-/chip}* and other hypomorphic *Dnmt1* mouse models, the former PhD student Lena Vockentanz had demonstrated the importance of DNA methylation for self-renewal in leukemia stem cells (LSCs) [118]. However, it remained unresolved, which gene programs were affected and if other mechanisms such as senescence or chromatin instability would be involved, too.

This doctoral project was part of a joint effort with Irina Savelyeva, a postdoc in our laboratory, to identify epigenetically regulated genes and mechanisms with high relevance for leukemia self-renewal. The respective objectives of the project were the bioinformatic analysis and interpretation of comprehensive massive parallel sequencing datasets as well as the experimental validation of the findings in vitro:

- For the first time, whole-genome bisulfite sequencing was carried out for a mouse model of *Dnmt1^{+/+}* as well as *Dnmt1^{-/chip}* MLL-AF9 leukemia. An in-depth characterization of the methylomes was performed.
- A novel method for methylome analyses based on generalized additive models was developed. The new technique allowed to accurately quantify the methylation persistency across large regions and simultaneously to account for distinctive deviations within spatially constrained regions such as CpG-Islands.
- To assess the consequences of the *Dnmt1^{-/chip}* genotype, sophisticated bioinformatic procedures, such as reference-guided transcriptome assembly or analysis of Hi-C data, were applied.
- Since anomalous enhancers emerged as important factors in leukemogenesis, bivalently transcribed active enhancers were called and comprehensively characterized, including motifs and methylation status.
- Numerous third-party datasets were integrated with our own results to put them into context and to pinpoint the most relevant enhancers for MLL-AF9 leukemia.
- More than 110 000 lines of code were written for all analyses combined.
- Although these results are not presented herein, selected genes and enhancers were experimentally tested in vitro by shRNA knock-down or CRISPRi for their effect on self-renewal and growth rate.

Chapter 12

Ramifications of the Dnmt1^{-/chip} methylome

Contents

12.1 Assumptions regarding the Dnmt1^{-/chip} methylome	110
12.1.1 Inverse relationship of division rate and methylation persistency	110
12.1.2 Inadvertent reactivation of epigenetically repressed genes	111
12.2 Characteristics of the large-scale compromised regions	112
12.2.1 Do Dnmt1 ^{-/chip} methylomes harbor PMDs?	113
12.2.2 Impact of PMD-like compromised regions	114
12.3 Persistent methylation at CpG-Islands and promoters	117
12.3.1 Persistently unmethylated CpG-Islands	117
12.3.2 Methylation transition at CpG-Islands	118
12.3.3 Persistently methylated CpG-Islands	119
12.4 Methylation-independent roles of Dnmt1	119

During the doctoral project, the Dnmt1^{-/chip} mouse model [121] was used to elaborate on the role of DNA methylation in the context of MLL-AF9 acute myelogenous leukemia. We sought to explore, why proper Dnmt1 expression is essential for self-renewal of MLL-AF9 leukemia stem cells in a cell-autonomous manner [118].

Two key observations characterized the Dnmt1^{-/chip} methylome:

1. Large hypomethylated regions in lamina-associated domains [\leftrightarrow section 12.2].
2. Stable methylation at CpG-Islands [\leftrightarrow section 12.3].

The observed degree of hypomethylation suggested noteworthy deleterious effects on cellular homeostasis and transcriptional regulation in particular and we wanted to elucidate the most relevant mechanisms. This chapter of the discussion shall therefore address our findings in the context of general knowledge about DNA methylation and provide possible interpretations.

12.1 Assumptions regarding the Dnmt1^{-/chip} methylome

It is well established that an interplay of active and passive processes shapes a methylome. Methylation may be actively deposited or removed and is either faithfully copied or passively lost during replication [reviewed in 153]. In case of the Dnmt1^{-/chip} methylome, one can safely assume that the reduced Dnmt1 levels will to some extent impair copying of methylation marks to the newly synthesized strand during replication and thus increase the passive methylation loss. The experimental and analytical plan of the project was conceived based on two assumptions:

1. Our main working hypothesis was the inverse relationship between methylation persistency and the cells' division rate. We proposed that methylation loss would intensify in fast-cycling cells such as the leukemic bulk compared to slowly dividing cells such as HSCs or LSCs. In accordance with previous work [165], we conjectured that compromised regions would be prone to methylation loss as a consequence of constrained time for methylation maintenance during replication.
2. Numerous reports had indicated, that cancer cells can repress tumor suppressor genes by promoter hypermethylation [reviewed in 43]. We assumed that the fast-cycling Dnmt1^{-/chip} leukemic cells would be unable to maintain (or reestablish) all relevant repressive promoter methylation. Thus, tumor suppressor genes would inadvertently become reactivated when the methylation capacity of Dnmt1^{-/chip} was exhausted. Therefore, we aimed to pinpoint epigenetically regulated genes, which were crucial for cancer self-renewal by intersecting the reactivated genes in all replicates.

Both assumptions were refuted by our own results as well as published studies over the course of the project:

12.1.1 Inverse relationship of division rate and methylation persistency

Although there was, at large, an association of late-replication with profound methylation loss, not all compromised regions were late replicating [↔ section 5.2, p.41]. The most important challenge, however, was that the separation into persistent and compromised regions in Dnmt1^{-/chip} did not arise after leukemic transformation, but was already effectuated on the level of slowly propagating HSCs [↔ section 6.2, p.47]. Neither did we observe an intensification of the methylation loss from MPP1 to MPP3 stage in healthy Dnmt1^{-/chip} hematopoiesis [↔ section 6.1, p.45], which would be expected if the proliferation rate would predominantly determine the methylation loss.

Along this line, a recent publication elucidated new details about the methylation copying process, such as the inheritance of most methylation marks within 20 minutes of replication [288]. For the first time, genome-wide quantitative data became available and the authors also developed a new computational method called *in silico Strand Annealing*(iSA), which permitted to assemble the original DNA double strands from book-ended reads. By applying iSA, the authors characterized surprisingly large fractions of

hemimethylation¹, which accounted for 4 % to 18 % of the DNA methylome. Remarkably, their analysis challenged the prevailing view that hemimethylation is purely transient, since they found hemimethylated sites that are stably inherited over several cell divisions dependent on Dnmt3b. These sites serve as directional binding site for CTCF and enable orientation-specific co-localization of methyl-binding proteins [288].

While these results introduced another crucial regulatory function of methylation, which could hypothetically be affected in Dnmt1^{-/chip}, the study also demonstrated that de novo methylation can ably compensate for critical passive loss, if required. For repetitive sequences, this had been suggested before [296].

None the less, Dnmt1^{-/chip} methylomes clearly exhibited pronounced compromised regions, although the propagation rate of MLL-AF9 cells did not seem high enough to enforce critical methylation loss in the light of the typically quick restoration within minutes after replication. Taking into account that genome-wide just 20 % of methylation serves a regulatory purpose [137], it appeared likely that methylation in those compromised regions was predominantly dispensable for the MLL-AF9 cells.

The discovery that cells can actively mark genomic regions where 5-methylcytosine is dispensable with H3R2me2a for renouncement of methylation maintenance [166] substantiated this model. Strikingly, the responsible enzyme Prmt6 was significantly upregulated in Dnmt1^{-/chip} MLL-AF9 leukemia in RNA-seq. Therefore, a future H3R2me2a ChIP-seq in Dnmt1^{-/chip} is suggested to integrate the mark with the location and strength of the compromised regions.

12.1.2 Inadvertent reactivation of epigenetically repressed genes

When the previous PhD student, Lena Vockentanz, handed the project over, the global mRNA sequencing of hypomethylated and control leukemia had not yet been evaluated. Therefore, it was unresolved, which gene programs promoting leukemic self-renewal might be perturbed in Dnmt1^{-/chip}.

Additionally, she suggested to focus on new surface markers present on hypomethylated stem cells, since such a discovery would be of interest from a therapeutic point of view. Distinct surface markers appearing upon hypomethylation would be a valuable target for anti-cancer therapy to combine demethylating drugs with specific antibodies [118].

Accordingly, we proceeded with the project, analyzed the RNA-seq and determined significantly differentially expressed genes [→ subsection 7.3.1, p.58]. We also identified divergent transcriptional programs and pathways, which clearly discriminated Dnmt1^{+/+} and Dnmt1^{-/chip} leukemia [→ subsection 7.3.2, p.59]. Our in-depth characterization also involved a H3K4me3 buffer domain analysis [→ section 7.4, p.61] and elaborated on a possible elongation bias [→ subsection 7.1.2, p.55] as well as perturbed splicing [→ section 8.1, p.65].

¹ CpGs-dyads containing just one methylated cytosine

Although we could detect notable differences between the genotypes, none of them could be straightforwardly linked to hypomethylated promoters or derepressed genes. Since methylation at CpG-Islands was predominantly persistent in Dnmt1^{-/chip} [↔ section 12.3], few promoter CpG-Islands hypomethylated. Differential pathways typically comprised none or just a few downstream genes with altered promoter methylation. On a global level, reactivation of repressed genes was almost completely absent [↔ subsection 7.1.1, p.53] and Dnmt1^{-/chip} exhibited dramatically fewer active promoters than the Dnmt1^{+/+} [↔ section 8.2, p.66].

While several, also recent studies, stressed reactivation of silenced promoters as possible mechanism of Dnmt inhibitors [174, 175] or Dnmt1 reduction [196, 297], it was never discriminated between active and passive demethylation. The importance of this discrimination is discussed later [↔ subsection 12.3.2].

Furthermore, the new study from Chenhuan Xu and Victor Corces [288] conclusively showed that hemimethylation was virtually absent around transcription start sites in a variety of different mouse embryonic stages. While the degree of hemimethylation in the gene body varied, it was consistently depleted at the TSS, which suggested stringent regulation. Considering that the methylation status of single CpGs within an island is tightly spatially correlated [136], little room was left for gradual, passive demethylation at promoter CGIs.

Taken together, we now favor the interpretation that the few transcripts, which exhibited hypomethylation at the promoter and concomitant transcriptional upregulation, were likely upregulated on purpose by active demethylation. This view is also backed by published literature [298, 299]. Deleterious methylation loss in Dnmt1^{-/chip} rather focused on cis-regulatory elements than promoters, which will be discussed later [↔ section 13.4, p.127]. On top of this complexity, Dnmt1 is involved in a variety of methylation-independent functions, which might have contributed to the phenotype [↔ section 12.4].

12.2 Characteristics of the large-scale compromised regions

Certainly, the most distinct features of the Dnmt1^{-/chip} methylome were the large compromised regions. Based on our assumptions [↔ section 12.1], we proposed that the hypomethylated regions would explain the self-renewal deficit observed in Dnmt1^{-/chip} MLL-AF9 leukemia. However, a derepression of epigenetically silenced genes was not detectable and Irina Savelyeva, a former postdoc of our laboratory, also ruled out deficits in genomic stability [↔ section 1.4, p.13].

On the other hand, she noticed evidence for senescence in Dnmt1^{-/chip}. Because a paper around the same time linked senescence with a methylome harboring compromised regions [154], we were intrigued to explore a potential causal relationship.

To do so, we conducted an in-depth investigation regarding the properties of those regions. We ascertained that the compromised regions differed from the persistent areas

mostly by the number of unchanged CpGs and not by the degree of demethylation. Furthermore, we made every effort to exactly localize the regions and quantify the degree of demethylation by fitting a custom generalized additive model [→ section 4.1, p.33], because a standard approach had failed to discriminate domain borders [→ section 3.3, p.31]. We could clearly show that the compromised regions were distinct from the methylation canyons described in HSCs [125], which are located in intergenic open chromatin and related to H3K36me2 [51].

Subsequently, we reanalyzed several third-party WGBS datasets for which other hypomethylation features had been described. We sought to explore, if the compromised regions were mechanistically identical to the *partially methylated domains* (PMDs) [141] or large-scale hypomethylations in cancer [129] [→ subsection 12.2.1]. This was required to understand, how these regions could potentially relate to the $Dnmt1^{-/chip}$ phenotype [→ subsection 12.2.2].

12.2.1 Do $Dnmt1^{-/chip}$ methylomes harbor PMDs?

In a purely descriptive sense, such a statement is correct, since the compromised regions exhibited a significant increase in partial methylation. However, we have extensively reviewed the scientific literature describing methylomes with PMDs or PMD-like features (including reanalysis of some third-party datasets) and advocate that there are at least two distinct mechanisms.

Thus, it may be helpful to delineate the possible methylation patterns that could give rise to a partially methylated domain: The CpG dyad is the smallest unit that can exhibit partial methylation, which is then preferably called hemimethylation [▷ Figure 12.1, panel A]. Indeed, pulse chase bisulfite sequencing experiments suggested that pronounced hemimethylation is strongly associated with partial methylation seen in conventional WGBS [288]. However, a similar pattern would also emerge, if smoothing was applied to fully methylated and unmethylated CpGs alternating on the same DNA strand [▷ Figure 12.1, panel B]. It could also be that methylation is predominantly limited to one allele [▷ Figure 12.1, panel C], a scenario which occurs for example in breast cancer [128]. Lastly, the pooling of hundreds and thousands of cells for a WGBS library implies that such patterns could also arise as an averaged value of absolute, but mixed methylation states in different cells [▷ Figure 12.1, panel D]. On top of this complexity, these states may change dynamically [31].

The mechanism underlying a specific PMD is often impossible to determine, unless the experiment has been designed to allow for it: In a dataset generated from a pulse chase experiment, one can distinguish newly synthesized strands and thus detect hemimethylation [288]. Other than that, it is sometimes possible to separate the alleles based on SNPs, if primary tumor material is used [128].

Evidently, the latter approach is not feasible when inbred mouse strains with little genomic variation are used like in our case. Eventually, we opted for compromised regions, since a lack of $Dnmt1$ strongly suggested an impaired methylation inheritance across cell

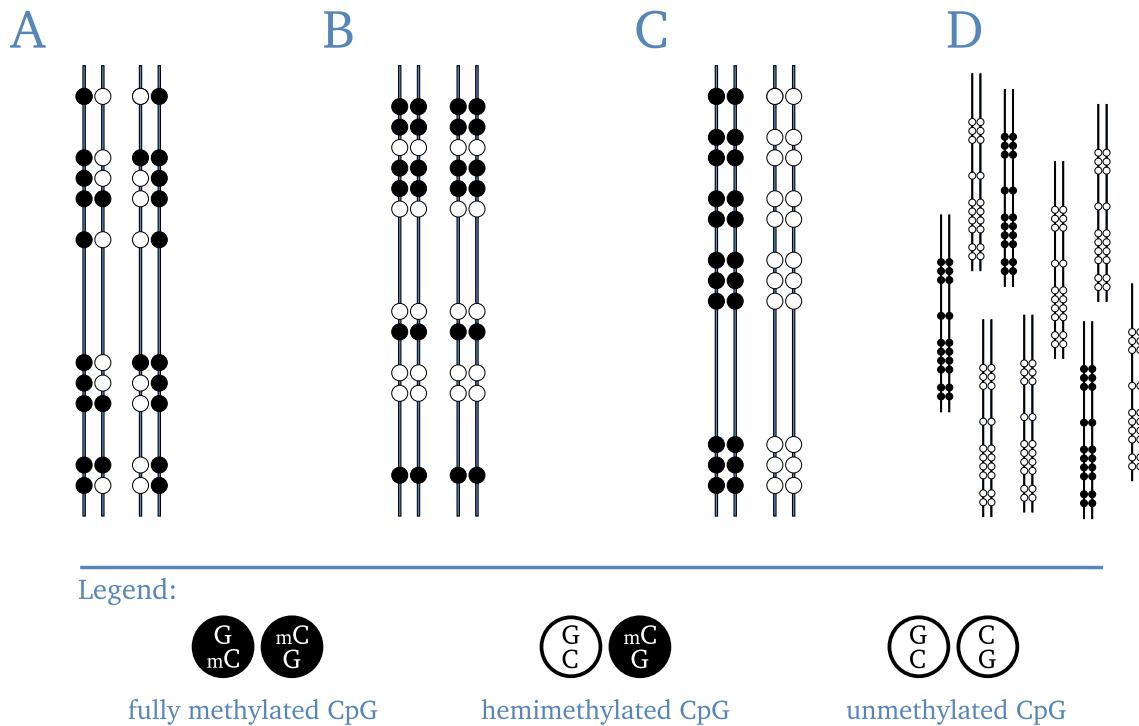


Figure 12.1: Schematic representation of the four principal methylation patterns that could resemble a partially methylated domain in a WGBS dataset. Both alleles of one cell are shown in panels **A** to **C**, whereas **D** depicts DNA from several cells. For the sake of simplicity, a CpG dyad on a strand is just symbolized as one circle. White color indicates no modification and black fill represents a methylated cytosine residue on the particular strand. **A**, **C** and **D** will appear as partial methylation even at base resolution, while **B** required smoothing to be applied (which is typically the case in WGBS analysis [39, 147]).

divisions. Therefore, the large-scale PMD-like features in $Dnmt1^{-/chip}$ probably emerged as increased hemimethylation [▷ Figure 12.1, panel A], which after subsequent cell divisions deteriorated into a heterogeneous methylation pattern [▷ Figure 12.1, panel B].

12.2.2 Impact of PMD-like compromised regions

One important question was, whether and how the compromised regions can provide an explanation for the self-renewal deficits of $Dnmt1^{-/chip}$ MLL-AF9 c-Kit^{high} cells. As discussed previously [↔ subsection 12.1.2], promoter hypomethylation linked to gene dysregulation was absent in $Dnmt1^{-/chip}$ leukemia [↔ subsection 7.1.1, p.53]. However, because of the relationship of PMDs to cancer and senescence, we suspected that there might be other effects, which shall be discussed herein.

PMDs or PMD-like features were described in early embryonal cell types [300] as well as the placenta [301], in pancreatic cells [302], in senescent cells [154], in long-term cultured cells [137] and particularly in cancer cells [44, 126, 127, 129].

In cancer: In tumors, successive loss of DNA methylation during tumorigenesis seems to be the norm: From healthy tissues to the primary tumors and their associated metastases, a clear trend of hypomethylation in lamina-associated, late-replicating regions is

observable [44]. The authors further conclude based on the loss of association between methylation levels of neighboring CpG sites that hypomethylation in PMDs occurs randomly rather than at distinct consecutive CpG sites [44]. Therefore, most cancer-related PMDs probably arise from pooling of heterogeneously methylated chromosomal sections in WGBS [▷ [Figure 12.1, panels B+D](#)].

Intriguingly, presence and intensity of the PMDs does not reflect the expression level of methyltransferases. In a comprehensive study, human post-mortem samples of 18 tissue types from four individuals were investigated and no systematic expression difference of Dnmt1, Dnmt3a, Dnmt3b and Dnmt3l between samples with and without PMDs were found [302].

However, this observation does not preclude temporary enzyme insufficiencies in fast cycling cancer cells. Intriguingly, it has been shown that the arginine methyltransferase Prmt6, which mediates H3R2me2a deposition, is upregulated in many cancers and its depletion or inhibition restores DNA methylation in hypomethylated breast cancer cells [166]. Since Prmt6 was also significantly upregulated in Dnmt1^{-/chip} MLL-AF9 leukemia, it was tempting to speculate that compromised areas might be characterized by H3R2me2a deposition as a means to prioritize methylation maintenance.

Prioritizing might be relevant to Dnmt1^{-/chip}, since a lack of Dnmt1 could facilitate stochastic epigenetic silencing by laying down repressive histone marks at sites of fork stalling [303, 304]. This hypothesis provides a rationale, how hypomethylation could be linked to the formation of long-range repressive chromatin [128], a process which appears to play role in several cancers. The respective repressive chromatin domains were termed LOCKs [305] or LRES [306].

In this context, it should be noted that LRES exhibit hypermethylation of consecutive CGIs [306], which is reminiscent of the methylation pattern of the largest compromised regions in Dnmt1^{-/chip} [▷ [supplement](#)]. In that sense, the compromised regions in Dnmt1^{-/chip} may even emerge in different ways: Regions of 150 kb rather by hemimethylation [▷ [Figure 12.1, A → B+D](#)], while the larger areas preferably located near the distal ends of chromosomes might be linked to LOCKs respectively LRES [▷ [Figure 12.1, C+D](#)].

In senescence: Repressive chromatin domains akin to the LOCKs/LRES in cancer were also described in senescent cells, where they are known as senescence-associated heterochromatic foci (SAHF) [307]. Analogous to the process triggered by a lack of Dnmt1 at the replication forks [↔ [section 12.4, p.119](#)], oncogenes may induce DNA replication stress and trigger an ATR (ataxia telangiectasia and Rad3-related)-mediated senescence response involving SAHF formation [308].

Like the LOCKs/LRES domains in cancer, SAHF coincide with large regions of partial methylation [154]. This interwoven nature of repressive heterochromatin and DNA hypomethylation [▷ [Figure 12.1, panel C](#)] was already noted in a remarkable forward-looking review by Bruce H. Howard published in 1996:

"Interestingly, the above results fit very well with a model of cellular senescence in which an interdependence exists between DNA methylation and maintenance of heterochromatin domains. [...] Errors in maintenance [...] are postulated to accumulate during the proliferative life span, ultimately triggering a cell cycle checkpoint and consequent irreversible cell cycle exit. Such a heterochromatin-linked model of senescence is directly coupled to DNA replication, because maintenance of heterochromatin-like structures requires that these structures be reformed in conjunction with each traverse of the cell cycle. [...] A semistochastic character also follows simply by assuming that recession of heterochromatin-like domains is progressive and widespread, but that not all domains, when lost, trigger a cell cycle checkpoint with equal efficiency." [309]

Although this proposal was made years before the era of genome-wide sequencing, it was substantiated two decades later: In single cells, chromosomal compartmentalization may be abrogated [310], and lamina-associated heterochromatic domains may dissociate from the nuclear lamina in a possibly incidental manner [311]. For aging hematopoietic stem cells it was shown that the dissociation and other effects are related to the altered expression of LaminA/C [312, 313]. Accordingly, Lamin B1 binding is redistributed in senescent cells [314] and its depletion provokes fundamental chromatin reorganization that consolidates cell-cycle exit [311, 315].

The exact influence of DNA methylation on this detachment remains elusive up to date, but it was shown that proper nuclear organization during terminal differentiation is dependent on methyl-binding proteins [316, 317] connecting the chromatin fiber to the nuclear lamina [318]. Moreover, a recent comprehensive study profiled the methylomes of 39 diverse primary tumors and analyzed them alongside 343 additional human and 206 mouse WGBS datasets [319]. By studying PMDs in cancer, they derived a local CpG sequence context associated with preferential hypomethylation and thereby noted a previously undetected methylation loss in almost all healthy tissue types. The degree of hypomethylation reflected the cell division history and suggested that senescence is the regular endpoint of normal differentiation [311]. In contrast, other authors emphasize that terminal differentiation and senescence are two distinct processes [320], since only the latter is associated with a secretory phenotype [321]. None the less, methylation in (some?) PMDs apparently serves as a mitotic clock in healthy cells, which is implicated in terminal differentiation/senescence [319].

Intriguingly, altered expression of Dnmt1 can figuratively change the clock to a different time in a methylation-dependent manner [197]. However, this mitotic clock barrier is generally overcome entirely during tumorigenesis for example by inactivation of the kinase ATM (ataxia telangiectasia mutated) or loss of p53 [322]. Accordingly, remethylation of the PMDs is not required for senescence bypass instigated by the SV40-T antigen [154]. Thus, the increased heterochromatin induction observed in premalignant cells is typically retained in the tumors [reviewed in 323].

In summary, most methylation loss occurring in the compromised regions of Dnmt1^{-/-}/*chip* has probably no impact on gene expression or any other gene oriented regulatory function. If the proposed mitotic clock function is correct, cells in the Dnmt1^{-/-}/*chip* mice could age faster and exhibit an altered formation of repressive chromatin domains as well as divergent lamina association harboring an inherent senescence risk. Eventually, the cells would become senescent sooner, unless MLL-AF9 transformation or contributory random mutations undermine the clocking mechanism completely.

Ultimately, it remained elusive, if the on average 2.8 % / 9.3 % senescent cells (in leukemic bulk and LSC respectively), which we observed by β -galactosidase staining, were sufficient to justify Dnmt1^{-/-}/*chip* leukemia phenotype in its entirety. It should be noted, however, that this method does not suffice for the proper detection of senescent cells [320].

12.3 Persistent methylation at CpG-Islands and promoters

The second noteworthy features of the Dnmt1^{-/-}/*chip* methylome were the remarkably persistent CpG-Islands (CGIs). Because we sought to characterize the epigenetic mechanisms causing the self-renewal impairment of Dnmt1^{-/-}/*chip* leukemia, we mostly focused on compromised sections of the genome.

Even though no evident hypomethylation at CpG-Islands could be determined in WGBS, they would be an interesting subject for further studies due to their considerable regulatory functions. However, single-cell methylome analysis would be required to substantiate incidences of stochastic aberrant methylation at CpG-Islands, because subclones with a truly deleterious epigenetic aberration would be quickly marginalized due to their competitive disadvantage [324–326].

12.3.1 Persistently unmethylated CpG-Islands

In both leukemic methylomes, unmethylated CpG-Islands were confined to the open chromatin / interLAD areas [↔ subsection 2.2.2, p.23]. At first glance, one might be tempted to dismiss unmethylated CGIs as irrelevant to the Dnmt1^{-/-}/*chip* phenotype, since further passive demethylation is impossible.

However, it may be helpful to keep in mind that an unmethylated CGI is a peculiarity. Eukaryotic genomes are typically dominated by AT [327], which can be explained by the spontaneous hydrolytic desamination of unmethylated cytosine to uracil. Most CpGs in genomes are methylated to better preserve them [137], which is the actively enforced default [328]. Therefore, a CpG-Island, which inadvertently hypomethylated in Dnmt1^{-/-}/*chip* would not just happen to remain unmethylated ever since.

So the unmethylated rather than the methylated state of a CpG-Island demands explanation and active regulation. Recent studies have shown, that DNA secondary structures are heavily implicated in maintaining regulatory CpG-Islands in an unmethylated state: G-quadruplex (G4) structures tightly bind and sequester Dnmt1 away from certain

CGIs and prevent their methylation [329]. However, it would be inaccurate to consider G-quadruplex structures solely as decoys for Dnmt1, since they exert a wealth of different regulatory activity [reviewed in 330, 331]. The expression of c-MYC [332] as well as c-Kit [252] is for example regulated by G-quadruplex structures at the respective promoters. Another secondary structure diverting from the regular double-helix strand is the i-motif, which can be formed in cytosine-rich DNA and might also serve regulatory purposes in vivo [333].

Although true structural complexity of DNA in vivo is just being revealed, unmethylated CpG-Islands are clearly hot spots of such uncommon formations. Particularly in the light of the so far underappreciated precise temporal orchestration of DNA methylation, which was lately uncovered by single-cell techniques [reviewed in 31], it is therefore recommended to revisit the DNA methylation of CpG-Islands in Dnmt1^{-/chip} MLL-AF9 leukemia with single-cell techniques.

12.3.2 Methylation transition at CpG-Islands

Transitions of the methylation state of a CGI are tightly controlled. For promoter CpG-Islands, it has been shown that EZH2 and PRC2/3 recruit DNA methyltransferases to cease the expression of genes [334]. Lineage commitment and differentiation are for example typically accompanied by a successive gain in methylation [335, 336]. During commitment, the cells achieve the permanent silencing of genes mediating stemness or governing alternative lineage development by promoter methylation and by the placement of repressive histone marks. This process has been proven for ES-cells [337, 338] as well as hematopoiesis [339].

On the other hand, differentiation also requires the activation of previously silenced genes. Active demethylation is mostly attributed to oxidative reversal [reviewed in 34] by the TET proteins [reviewed in 33]. However, also Dnmt3a and Dnmt3b are capable of active demethylation [340] [reviewed in 36].

Remarkably, TET proteins play an important role in the regulation of hematopoietic malignancies [reviewed in 341] and MLL-rearranged leukemia. Through coordination with MLL-fusion proteins, TET1 acts and reactivates critical co-targets such as Hoxa9, Meis1 and Pbx3 [342, 343]. Rarely, leukemia cases are reported that even exhibit a direct fusion of TET1 to Mll1 [55, 344].

It should be stressed, that active demethylation fosters the deposition of epigenetic marks, which are important for the integrity of the H3K4me3 depositing SET1/COMPASS complex. TET proteins associate with the O-GlcNAc transferase (EC 2.4.1.255, OGT), which glycosylates histone 2B (H2B), host cell factor 1 (HCF1) and other proteins [184, 186]. Since these glycosylations promote the deposition of histone 2 K120 monoubiquitination [185] and ultimately H3K4me3, it is ensured that transcription is initiated as a result of the active demethylation of a promoter.

A lack of those additional epigenetic cues was probably responsible for the absence of

functional gene reactivation after passive demethylation in $Dnmt1^{-/chip}$ [↔ [subsection 12.1.2](#)]. This interpretation is strongly backed by a study in fibroblasts, where passive promoter demethylation after shRNA-mediated knockdown of $Dnmt1$ intensified in areas of lower chromatin accessibility and did seldom translate into direct expression changes [299]. Instead, a lion's share of the reactivation could be clearly attributed to active DNA demethylation by the Ten-eleven translocation methylcytosine dioxygenase 1(TET1) [299].

12.3.3 Persistently methylated CpG-Islands

Methylated CpG-Islands could be observed in the open as well as lamina-associated domains of the MLL-AF9 genome. At promoters, methylated CpG-Islands are clearly associated with transcriptional repression [reviewed in 345]. However, CGIs are not limited to the promoter of genes, but may also occur in introns or be found within the coding sequence itself. Methylation of such CGIs reduces physical interaction with promoters, abates bivalent chromatin and results in transcriptional activation of key regulatory genes such as PAXs, HOXs and WNTs [346]. Therefore, the persistent methylation of CpG-Islands in $Dnmt1^{-/chip}$ could not only serve as repressive mark but also support the expression of other genes.

To identify CpG-Islands, which specifically need to be preserved in a methylated state, ubiquitylation of H3 at lysines K18 and K23 could be monitored in $Dnmt1^{-/chip}$. If H3R2me2a is absent, but H3K9me3 present, Uhrf1 establishes H3K18ub and H3K23ub, which promote DNA methylation inheritance by $Dnmt1$ [167, 168, 347]. Further histone marks are likely also implicated [reviewed in 348]. It is conceivable that such marks will allow to pinpoint critical regulatory CpG-Islands in $Dnmt1^{-/chip}$.

12.4 Methylation-independent roles of Dnmt1

Lastly, it should be pointed out that there are methylation-independent functions of $Dnmt1$. These need to be considered, when DNA hypomethylation is induced by hypomorphic $Dnmt1$ mouse strains as well as the established $Dnmt1$ inhibitors.

Studies investigating the cellular transcriptome after treatment with DNA methyltransferase inhibitors (DNMTis) have repeatedly reported effects unrelated to direct promoter DNA hypomethylation [349, 350]. DNMTis are typically cytosine nucleoside analogues, which become integrated into the DNA and form stable adducts with $Dnmt1$ [351]. The enzyme becomes irreversibly bound to 5-Aza-2'-deoxycytidine (Decitabine) residues in the DNA, which eventually confers cytotoxicity. Therefore, the therapeutic effect of DNMTis in cancer therapy is possibly methylation-independent [352].

Normally, DNA (cytosine-5)-methyltransferase 1 is recruited to replication foci by interacting with Pcpa and Uhrf1 [353, 354] and is loaded onto hemiCpGs to methylate the nascent cytosines during DNA replication. It is well established that $Dnmt1$ knockdown triggers intra-S-phase arrests [355] or activates stress response checkpoints such as ataxia

telangiectasia mutated-Rad3-related (ATR) [356]. Apparently, removal of Dnmt1 from replication forks is the trigger for these responses, since ectopic expression of Dnmt1 lacking a functional catalytic domain alleviated the stress response [356]. This notion is also backed by the observation that senescence in IMR90 fibroblasts can be overcome by the SV40 T-antigen, while the characteristic hypomethylated methylome is retained [154].

Although we observed on average 2.8 % senescent cells in the hypomorphic leukemic bulk (9.3 % in LSCs), it remained unclear to what extent such a stress response has relevance for the Dnmt1^{-/chip} strain. The negative γH2AX-stainings performed by Irina Savelyeva challenged an ATR-mediated response and most Dnmt1^{-/chip} cells were arrested in G1-phase [▷ [data not shown](#)], whereas a lack of Dnmt1 typically causes an intra-S-phase cell cycle arrest [355]. Both observations seemed to argue against a pronounced stress response in Dnmt1^{-/chip}.

However, acute replication stress challenges proper chromatin restoration even below the threshold that results in a cell cycle arrest [reviewed in 304]. It facilitates stochastic epigenetic silencing by laying down repressive histone marks at sites of fork stalling [303]. Unfortunately, we did not assay repressive chromatin in Dnmt1^{-/chip}, but the striking absence of unannotated TSS clusters in corresponding areas [▷ [Figure 8.1, p.67, bottom row](#)] might reflect increased compaction of chromatin and diminished cellular plasticity.

Additionally, the persistently methylated CpG Islands in compromised, mostly heterochromatic regions [↔ [section 5.2, p.41](#)] could be indicative of long-range epigenetic silencing (LRES) in the hypomorphic mice, a process with particular relevance for carcinogenesis [306]. Because of the excessive heterochromatin formation, Dnmt1^{-/chip} MLL-AF9 leukemic stem cells would, according to a model by the Feinberg lab [226], respond poorly when challenged by variable conditions. Therefore, a lack of Dnmt1 might cause improper chromatin restoration after cell divisions resulting in a survival and self-renewal bias in Dnmt1^{-/chip} independently of the methylation levels.

Apart from a possible heterochromatin-spreading in Dnmt1^{-/chip}, the manifold interactions of Dnmt1 with histone-modifying enzymes demands attention [reviewed in 357].

These predominantly repressive chromatin modifiers allow Dnmt1 to alter transcription of target genes independent of DNA methylation in a direct manner [358, 359]. Given that many upregulated transcripts did not feature a hypomethylated promoter [↔ [subsection 7.3.1, p.58](#)], it was therefore conceivable that also methylation-independent functions of Dnmt1 constituted to the phenotypic alterations of the Dnmt1^{-/chip} strain.

Chapter 13

Transcriptional regulation in leukemia

Contents

13.1 Establishment of our MLL-AF9 enhancer catalog	121
13.2 Notable characteristics of our enhancer catalog	122
13.2.1 Rarity of leukemia-specific enhancers	122
13.2.2 H3K4me3 as hallmark of particular enhancer clades	123
13.3 H3K4me3, Mll2, CXXC-domains and leukemia	125
13.4 Implications for the Dnmt1 ^{-/chip} genotype	127
13.4.1 Methylation determines enhancer activity	127
13.4.2 Methylation affects chromatin organization	128
13.5 Outlook	129

The intricate balance between transcriptional control and noise is pivotal for carcinogenesis [reviewed in 226]. Undoubtedly, enhancers contribute strongly to this process and thus to cellular homeostasis in general. By identification of enhancers with relevance for leukemia, we presumed to find possible cues explaining the impairment of Dnmt1^{-/chip} leukemic stem cells. Since enhancers may be subject to carcinogenic methylation changes [232, 233, 360], the Dnmt1^{-/chip} genotype might complicate establishment and maintenance of such pathogenic alterations [reviewed in 361, 362].

For this reason, we investigated the presumably active enhancers in MLL-AF9 leukemia [\leftrightarrow chapter 9, p.71] and derived important commonalities [\leftrightarrow chapter 10, p.87]. Herein, we will discuss the most relevant aspects of this endeavor and implications for leukemogenesis in a Dnmt1^{-/chip} background.

13.1 Establishment of our MLL-AF9 enhancer catalog

A few years ago, it was not yet well known that enhancers contribute to leukemogenesis [\leftrightarrow subsection 1.3.2, p.12] and that malignant transformation often involves perturbation of enhancer activity [reviewed in 363]. By now, hematopoietic enhancers have been characterized at single-cell and single-variant resolution [102, 110], but the corresponding datasets were not yet available when we conducted our project.

Back in 2014, the laboratory of Ido Amit had just published the first comprehensive enhancer study in healthy hematopoiesis [101], but similar data for leukemia was still missing. Therefore, we compiled our own catalog for MLL-AF9 leukemia [\leftrightarrow chapter 9, p.71] and identified 6386 and 6662 putative enhancers in $Dnmt1^{+/+}$ and $Dnmt1^{-/chip}$ respectively. Surprisingly, the majority of them (82.45 %) was specific for either of the genotypes [\triangleright Figure 9.1, p.72].

The sheer abundance of distinct sites suggested a relevant fraction of false positive enhancers and called for extra caution in handling the data. Therefore, we hierarchically clustered the sites to derive repeatedly occurring characteristic signatures [\leftrightarrow section 9.2, p.73]. We presumed that a common chromatin signature across healthy hematopoietic cell types [\leftrightarrow subsection 9.3.1, p.78] mirrored a concordant regulation and indicated close functional ties.

Particularly, such subclusters (referred to as *clades*) caught our attention, which consisted of significantly more CAGE-defined enhancers than was expectable by chance [\leftrightarrow subsection 9.3.2, p.80]. Such clades exhibited a considerable enrichment of H3K4me3, H3K18ac and H3K27ac as well as RNA.Pol.II in MLL-AF9 leukemia [\triangleright Figure 9.7, p.81] [\triangleright Figure 9.9, p.83] and were frequently confirmed by ATAC-seq [\triangleright Figure 9.11, p.86]. In particular, the H3K4me3 mark was quite unusual and shall be discussed below in greater detail [\leftrightarrow subsection 13.2.2].

As a downside of this strategy, we mostly disregarded putative enhancers in the *Undefined* (X) cluster [\leftrightarrow subsection 9.2.1, p.73]. It comprised CAGE-defined enhancers that were mostly devoid of chromatin modifications in healthy hematopoiesis. Thus, we considered them as a mixture of mostly false positive calls with a minority of leukemia-specific enhancers, which we could not address adequately.

13.2 Notable characteristics of our enhancer catalog

13.2.1 Rarity of leukemia-specific enhancers

As stated in the previous section, the high disagreement between the putative enhancer sets of the two genotypes prompted us to incorporate data from the healthy hematopoiesis to reduce the number of false positive sites under consideration. In return, we deliberately accepted the consequences of missing out on some leukemia-specific enhancers.

The rationale behind this approach was that leukemia-specific enhancers in human leukemia typically arise from somatic mutations. For instance, such mutations may introduce a novel super enhancer upstream of the TAL1 oncogene and sustain its expression [111, 112].

Thus, it normally takes a preleukemic hematopoietic stem cell, which has progressively acquired an increasing mutation burden over its lifetime, to identify a relevant number of aberrant enhancers [110].

Littermates of a mouse model, on the other hand, are by design as similar as possible on the genetic level. Accordingly, the search for somatic mutations is unpromising, but has none the less been attempted for MLL-AF9 leukemia [22]. On the other hand, some leukemic clones are propagated several times in recipient mice or kept for longer periods in vitro and therefore might exhibit significant somatic mutations. Additionally, long-term culture is associated with pronounced genomic demethylation [137], which could reactivate dormant enhancers [232,233].

Despite the limited applicability of a mouse model to investigate leukemia-specific enhancers, we noted a relevant set of H3K79me2 positive sites within the tenth cluster. This chromatin modification is deposited by DOT1L and is required to protect elements from a repressive protein complex composed of Sirt1 and Suv39h1 in leukemia [246]. Since binding of the fusion proteins MLL-AF4 or MLL-AF9 recruits DOT1L with great efficiency [364], strong H3K79me2 marks typically suggest oncprotein binding [365]. Such MLL-targeted enhancers have been termed KEEs [248] and preservation of their open chromatin state is pivotal to uphold the leukemic differentiation block [247].

However, reanalysis of published data [245] showed that neither H3K79me2 nor direct binding by MLL-AF9 [\leftrightarrow section 10.3, p.95] were associated with or enriched in any particular cluster or clade. Furthermore, the normalized H3K79me2 signal tended to be stronger in c-Kit^{low} than in c-Kit^{high} cells [\triangleright Figure 9.8, p.82, bottom row], which implied that KEEs might be less relevant for the leukemic stem cells (LSC) than for the bulk leukemia.

KEEs also eluded a closer examination, since the chromatin interaction data used to assign the targeted promoters originated from the HPC-7 murine blood stem/progenitor cell model [156]. Therefore, leukemia-specific putative enhancers were typically not assigned to target transcripts [\leftrightarrow section 10.4, p.98], which is why hardly any putative enhancers from the *Undefined* (X) cluster were recorded in the top interactions despite their large number [\triangleright supplement].

In summary, the rarity of leukemia-specific enhancers in our catalog was attributable to usage of a mouse model instead of patient material as well as technical limitations such as the lack of chromatin interaction data from leukemia.

13.2.2 H3K4me3 as hallmark of particular enhancer clades

The classical histone mark signature linked to active enhancers is the combination of H3K4me1 and H3K27ac [98,366,367]. Therefore, it is the most used pattern to screen ChIP-seq datasets for the presence of enhancers and was used by the laboratory of Ido Amit to cluster their hematopoietic enhancer catalog [101]. Consequently, we modeled our clustering strategy on the same data and approach [\leftrightarrow section 9.2, p.73].

Hence, we were surprised to find H3K4me3 to be characteristic of a variety of highly accumulated clades of the clusters *Lymphoid + Progenitors* (III), *Progenitors* (V), *TNK-cells* (VIII) and *Erythroid* (IX) in MLL-AF9 leukemia [\leftrightarrow subsection 9.3.2, p.80]. By ab-

solute numbers, the majority of H3K4me3-positive enhancers were assigned to clades of cluster III [▷ [Figure 9.7, p.81, top row](#)], whereas enhancers from insignificant clades were consistently low in H3K4me3 [▷ [Figure 9.8, p.82, top tow](#)]. We were inclined to elaborate on this finding, since H3K4me3 plays an important role in leukemogenesis [→ [section 13.3](#)].

While it was irritating in the first place, the presence of H3K4me3 and H3K4me1 was not contradicting as every nucleosome consists of two histones H3 and may thus carry both marks simultaneously. Yet, H3K4me3 is commonly believed to be restricted to gene promoters instead of enhancers. This belief is put in question by unified models seeking to overcome the distinction between promoters and enhancers [368, 369]. Furthermore, H3K4me3 has been regarded as the best indicator of active enhancers in lymphoid lineages [81, 370], while H3K4me1 is not required for correct enhancer function in *Drosophila* [371].

Taken together, both H3K4me1 as well as H3K4me3 mark active enhancers, but H3K4me1 is relatively universal while H3K4me3 only characterizes a subset of enhancers. About a third of those H3K4me3 enhancers is also bound by CTCF as illustrated by a comprehensive study of murine chromatin states [▷ [Figure 13.1](#)] [372].

Aforementioned study distinguished two major enhancer classes mostly based on the occurrence of H3K4me3. Conversely, those enhancers might be targeted by different histone-lysine N-methyltransferases. While *Drosophila* has just three K4 methyltransferases¹, mammals possess six COMPASS-like complexes: Setd1a and Setd1b, Mll1 (Kmt2a) and Mll2 (Kmt2b) as well as Mll3 (Kmt2c) and Mll4 (Kmt2d) [reviewed in 275, 373].

At regular gene promoters the trimethylation is typically established by the SET1/COMPASS complex, of which two variants with either Setd1a or Setd1b exist [374]. The non-overlapping nuclear localization of the two variants suggests that both exert non-redundant functions [374].

Targeting of the SET1/COMPASS complex is mainly mediated by Wdr82, which recognizes the Ser5-phosphorylated C-terminal domain of RNA polymerase II [375] and thereby directs it to sites of active transcription in a histone H2B ubiquitination-dependent manner [376]. Intriguingly, Wdr82 has also been shown to be responsible for the active termination of enhancer RNAs (eRNAs). Upon Wdr82 depletion, enhancers abundantly spawned long and non-coding RNAs due to termination defects [93].

Because we identified our enhancers based on active bidirectional eRNA transcription [→ [section 9.1, p.72](#)], a Wdr82-mediated recruitment of a histone-lysine N-methyltransferase complex appeared to be likely. Since Wdr82 is missing in the MLL/COMPASS complexes [376] and is limited to the SET1/COMPASS complexes, the latter seemed to be responsible for the H3K4me3 marks at the respective enhancers. However, CXCC-type zinc finger protein 1 (Cfp1), which is also part of SET1/COMPASS [272, 377, 378] did not bind to the motif in ChIP-seq [▷ [data not shown](#)].

¹ dSet1, Trithorax (Trx) and Trithorax-related (Trr)

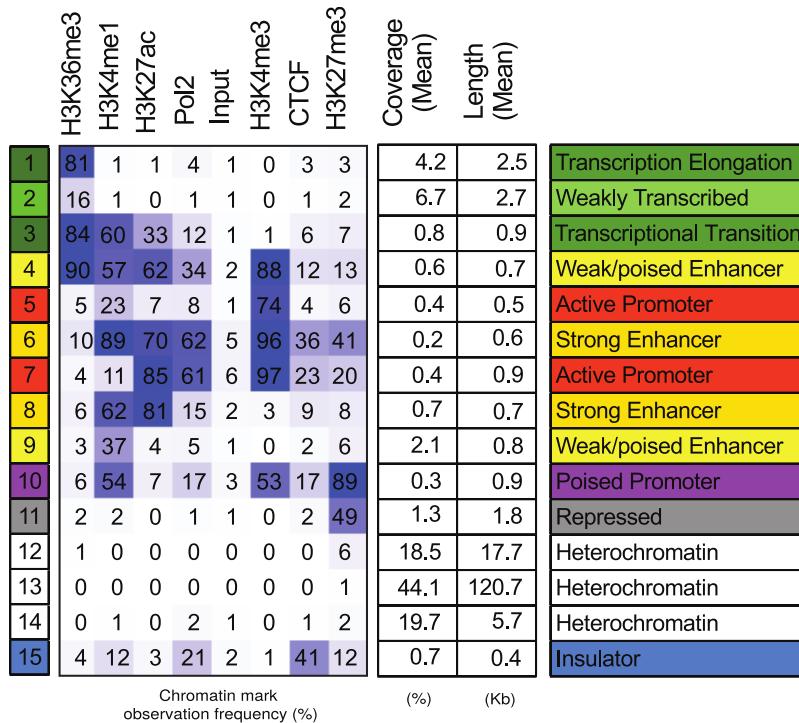


Figure 13.1: Using both RNA-seq and ChIP-seq data from eight murine tissues (brain, heart, liver, kidney, spleen, small intestine, testes and thymus) as well as mouse embryonic stem cells, a comprehensive chromatin state map was computed by the group of Marc Marti-Renom [372]. Reprinted here is Figure 3A of this publication. Each cell in the table denotes the percentage of cases in which a given ChIP-seq peak is found at genomic positions corresponding to a specific chromatin state.

None the less, the H3K4me3 deposition was indeed mediated by a CXXC domain, albeit that of Kmt2b (Mll2 [[section 10.3, p.95](#)]). Hence, the CG-rich motifs [[subsection 10.1.2, p.88](#)] recruited MLL2/COMPASS to deposit H3K4me3 at those enhancers.

13.3 H3K4me3, Mll2, CXXC-domains and leukemia

H3K4me3: Around the same time when we first observed the abnormal H3K4me3 pattern at the accumulated enhancers, the group of Micheal Cleary reported that the oncogenic potential of MLL-AF9 LSCs was mainly regulated by high-level H3K4me3 marks at promoters [216]. In contrast, H3K79me2 was low in LSCs, but increased upon differentiation into c-Kit^{low} blast cells. At this step, also the KEEs come into play [[subsection 13.2.1](#)].

Intriguingly, H3K4me3 is also implicated in the leukemic potential of Nup98 fusions, in which an extrinsic plant homeodomain (PHD) finger targets the joint protein to H3K4me3 marked regions of the genome [286]. When mutations in the PHD fingers abrogate H3K4me3 binding, the leukemic transforming capability is lost, since the differentiation-associated polycomb-mediated removal of the mark can no longer be prevented [286]. Later, it was also shown that Nup98 not only blocks removal of H3K4me3, but can also recruit the Wdr82-Set1A/COMPASS complex to mediate deposition of H3K4me3 [287].

Subsequently, H3K4me3 triggers the H3K27ac modification [379].

Mll2: Hence, the mechanism for deposition of the mark was of great interest to us. MLL2/COMPASS could not be considered likely, as it is dispensable for self-renewal in embryonic stem cells [380]. However, its deletion impairs the differentiation of ES cells into primordial germ cells (PGCs) [381], the precursors for the oocytes and spermatozoa. Seemingly unrelated at the first glance, leukemia and germ cells in reality are closely connected [reviewed in 382]. Germ cells may give rise to leukemia [383] and AML and CML cells express sex hormone receptors and respond to stimulation with gonadotrophins [384].

Therefore, we deemed the ChIP-seq data from ES cells / PCGs [273] applicable to infer binding of Mll2 at the accumulated enhancers in MLL-AF9 [\leftrightarrow section 10.3, p.95]. In oocytes, MLL2/COMPASS primarily targets distal cis-regulatory elements for H3K4me3 deposition [385], but may also aim at some bivalent promoters in PGCs [273].

In MLL-AF9 leukemia, Mll2 is indeed pivotal for self-renewal and maintenance of the leukemic stem cell [277]. It is expressed at least as abundantly as Mll1 in AML, including both, MLL-rearranged and other subtypes [277]. The knock-out of Mll2 is deleterious to the leukemia *in vivo* [277] by promoter as well as enhancer-mediated effects [\leftrightarrow subsection 10.4.1, p.101].

As illustrated by a different response to menin inhibition [386], Mll2 does not jointly act with MLL-AF9 at the same sites. Furthermore, an artificial Mll2 fusion protein is unable to transform hematopoietic cells *in vitro* [276]. Despite structural conservation, no other Mll homolog can replace Mll1 in such leukemogenic fusion proteins, which is attributed to differences in their CXXC-domain.

CXXC-domain: This domain generally binds to unmethylated CpG-dinucleotides and is found in a variety of chromatin-associated proteins [264] with different chromatin binding properties and functions [reviewed in 387]. The CXXC-domain is retained in all known Mll1 fusion proteins (MLL-FP) [reviewed in 5], but lacked entirely by Mll3 and Mll4 [264]. Subtle differences between the CXXC-domain of Mll1 and Mll2 preclude an oncogenic potential of the latter in the context of fusion proteins [276].

In vitro, gel shift experiments indicate almost indistinguishable DNA-binding properties [276], but *in vivo* a divergent nuclear localization and function of Mll1 and Mll2 is evident [276]. Experimentally, mutagenesis of the CXXC-domain of Mll2 to mimic the binding properties of Mll1 has been attempted [388]. Yet, domain-swapping with the CXXC-domains of other proteins has shown that solely the CXXC-domain of Dnmt1 is functionally equivalent to that of Mll1 and elicits leukemia in the context of a MLL-FP [389].

Remarkably, a MLL-FP is not necessary to confer leukemogenic capacity, if the CXXC-domain of Mll1 is disfigured. In roughly 5 % to 10 % of AML or ALL cases, Mll1 is mutated by chromosomal translocations with other proteins, while partial tandem dupli-

tion (PTD) of Mll1 constitutes another 5 % to 10 % [390]. In PDT leukemia, either a part of the CXXC-domain or the full region is duplicated [reviewed in 391] and the HOX gene cluster becomes activated due to intensified binding [392]. Because of the two CXXC-domains, this type of leukemia resembles a lot the subgroup of Mll1 fusions with other CXXC-proteins such as LCX [393]. PDT leukemia has an extremely poor outcome [394], but critically depends on contributory mutations for leukemogenesis [395].

Early additional mutations in PTD leukemia commonly affect the methylome of cells. For example *Dnmt3a* is frequently affected [396], but also *IDH1*, *IDH2* and *TET2* [397]. Mutations in the latter three typically cause a genomic hypermethylation phenotype [reviewed in 52]. Since the CXXC-domain is methylation-sensitive, this emphasizes the importance of abnormal binding to CG-motifs in the pathogenesis of such leukemia.

13.4 Implications for the $Dnmt1^{-/chip}$ genotype

Most aspects of the $Dnmt1^{-/chip}$ genotype have been discussed previously [\leftrightarrow chapter 12, p.109]. However, two topics with respect to enhancers are still missing. Firstly, how methylation may affect enhancer activity itself; and secondly, how demethylation may perturb enhancer-promoter pairs.

13.4.1 Methylation determines enhancer activity

In hematopoiesis, enhancers prime a cell far before the actual lineage separation takes place during differentiation [398–400]. These findings are in accordance with the notion that enhancer signatures characterize cell types superiorly and foreshadow future gene expression programs [89, 101, 110].

Also cancer predisposition is to a good extent reflected by pre-neoblastic alterations at the enhancer sites. These can be either genetic such as mutations [106, 111, 401, 402] and translocations [114, 115]) or epigenetic. Because active enhancers are typically unmethylated or lowly methylated [139], the prevalent pre-malignant or malignant epigenetic alteration is hypermethylation [232, 233, 360, 403]). In a way, hypermethylation represents a restoration of the default methylated state, whereby de novo methyltransferases outcompete methylation-sensitive transcription factors such as NRF1 [328] and decommission the enhancer.

However, it would be oversimplified to assume that a methylated enhancer can not contribute to the regulation of gene expression, since there are also many transcription factors that preferably bind to methylated sites [268]. Furthermore, unmethylated degenerate binding sites sequester the transcription factor EGR1 away to non-functional locations [269].

The aforementioned mechanism is of great interest with regard to the $Dnmt1^{-/chip}$ genotype, since Irina Savelyeva noticed an intriguing enrichment of EGR1 motifs within the promoters of genes that were downregulated in $Dnmt1^{-/chip}$ MLL-AF9 leukemia. Be-

cause of the rather random demethylation, it is much easier to rationalize such a decoy-based mechanism in Dnmt1^{-/-} than a spurious reactivation of a specific decommissioned enhancer. The latter requires a precise loss of methylation at the enhancer site, whereas degenerate EGR1 motifs are quite common in the genome such that an arbitrary loss of MECP2 binding would suffice for an impairment. However, we could not explain the EGR1 motif enrichment in the data back then and mainly focused on upregulated genes.

Further results also suggested that methylation levels of the motif *DeNovo*.SSCGGGCTSS were subject to specific regulation, possibly to govern MLL2/COMPASS binding [[↔ section 10.2, p.91](#)]. It was tempting to speculate that also for this quite frequent and degenerate motif, a sequestration takes place, since we observed a hypomethylation of the motif and a significant redistribution of the H3K4me3 mark in Dnmt1^{-/-} [[↔ section 7.4, p.61](#)].

13.4.2 Methylation affects chromatin organization

By and large, enhancers either act through RNAs [reviewed in [91](#)] or by increasing the concentration of transcriptional activators near a gene promoter. This mechanism of enhancer action typically requires a change of the three-dimensional chromatin structure and the formation of a DNA loop - a process that is still incompletely understood. The discovery of preexisting chromatin looping, which precedes the actual signaling [[70](#)] has challenged the previous model [[67](#)] that solely specific transcription factors govern the looping [[68,69](#)]. Importantly, the size of chromatin loops may be altered without dissociation of the cohesin complex [[71](#)].

In any case, it is evident that the formation and dissociation of loops must undergo tight sequential coordination, since a gene is typically targeted by several cis-regulatory elements and one element may also be involved in the regulation of different genes [[64–66](#)]. This is further illustrated by chromatin loop collisions in cells lacking the cohesin-unloading factor WAPL [[404](#)].

Since CTCF-mediated chromatin interactions are influenced by stably inherited hemimethylation that flanks CTCF motifs [[288](#)], a perturbation of the latter in Dnmt1^{-/-} seems possible. In gliomas with IDH mutations and a hypermethylation phenotype, insulator dysfunction due to abrogated CTCF-binding allows for an enhancer-driven increased expression of the receptor tyrosine kinase PDGFRA, a prominent glioma oncogene [[405](#)]. Lacking actual Hi-C data from MLL-AF9 leukemia or the Dnmt1^{-/-} genotype, we could not verify if enhancer-promoter pairs were truly malformed in Dnmt1^{-/-}.

In general, CTCF is not considered to be relevant for chromatin compartmentalization, which is rather attributed to cohesin [[71](#)]. Specifically for the HOX gene cluster, however, it has been shown that CTCF subdivides it into euchromatin and facultative heterochromatin [[406](#)] and that some CTCF-binding sites serve as promoters for functional lncRNAs, which impact chromosomal interactions [[407](#)]. Considering the tremendous importance of the HOX gene cluster, particularly HOXA9, for leukemia [[408–410](#)], a mis-

regulation in $Dnmt1^{-/-} / chip$ might have deleterious effects. The GAM-derived methylation persistency suggested a breakdown of the chromatin compartmentalization at some positions [▷ [Figure 4.2, p.38, black arrow](#)], but this needs to be verified by additional data².

13.5 Outlook

Currently, the continuation of the project is not scheduled. Upon advancement, the readjusted scope of the project would determine the subsequent steps and experiments.

Role of methylation and its relationship to senescence: The PMD-like compromised regions in $Dnmt1^{-/-} / chip$ might affect the chromatin association with the nuclear lamina and interfere with the mitotic clock [↔ [subsection 12.2.2](#)]. Eventually, chromatin in aging hematopoietic stem cells dissociates from the nuclear lamina and the higher-order chromatin architecture collapses [313]. The effect of the altered lamina composition has already been shown [313], none the less the $Dnmt1^{-/-} / chip$ strain could represent an interesting model system to study a prematurely aging hematopoiesis.

To better assess the scientific potential of this project, comprehensive WGBS datasets of mouse embryonic fibroblasts (MEFs) could be reanalyzed and integrated. By now, methylome data of MEFs after $Dnmt1$ knockdown (GSE93058) and throughout the regular cell cycle (GSE92903) have been published [297, 299]. Therefore, after integration with previously published lamina-association data from MEFs [131], it would be possible to elaborate on the variability and fluctuation of methylation marks in the context of LADs and interLADs.

Ultimately, proper chromatin maps and matched methylome data of aged $Dnmt1^{-/-} / chip$ HSCs would be required for an authoritative study. Considering that the Vaquerizas group in Münster possesses comprehensive skills to elucidate the chromatin architecture from challenging cell types and little source material [161], a local collaboration would be possible.

² such as Lamin-DamID [411]

Methylation in MLL-AF9 leukemia and novel therapeutic targets: The initial focus of the project was the identification of novel, potentially druggable targets for the treatment of AML. We proposed that hypomethylation at promoters in Dnmt1^{-/-} would lead to a reactivation of tumor suppressor genes, which we would be able to single out and target specifically for therapeutic purposes. However, it turned out that inadvertent reactivation of genes was implausible considering the absence of hemimethylation at gene promoters and lack of supportive epigenetic marks required for transcription [\leftrightarrow section 12.3]. The Dnmt1^{-/-} mouse strain is therefore probably an unsuitable model system for this approach.

On top of that, methylation-independent effects due to a stress response are to be expected in Dnmt1^{-/-} leukemia. Hence, the ectopic expression of Dnmt1 lacking a functional catalytic domain would be recommended to alleviate a potential stress response [\leftrightarrow section 12.4, p.119]. The same applies to the classical DNA methylation inhibitors 5-Azacytidine and 5-aza-2'-deoxycytidine, which are by no means more suitable, since they trap Dnmt1 irreversibly at the DNA [352] and also trigger a stress response [351]. Therefore, either approach harbors the risk to wrongly attribute effects to DNA demethylation that are actually related to other functions of Dnmt1 [352].

To study the effect of demethylation in MLL-AF9 specifically, another approach is hence needed. Targeting the adenosylhomocysteinase Ahcy [412, 413] instead of Dnmt1 might be such an alternative. It is the only enzyme capable of hydrolyzing S-adenosyl-L-homocysteine³, which is generated during the DNA methylation process [415]. Since S-adenosyl-L-homocysteine is a strong inhibitor of Dnmt1, the knockdown or inhibition of Ahcy in MLL-AF9 leukemia should impair DNA methylation without triggering a stress response, since the replication fork complex can still be faithfully assembled.

However, manipulation of Ahcy may also have unwanted side effects. A study suggested that the downregulation of adenosylhomocysteinase might actually promote tumorigenesis [416], while strong overexpression can trigger apoptosis due to unphysiologic accumulation of adenosine [417]. Besides DNA methylation, it is also implicated in mRNA cap methylation [418]. The latter mechanism may predominantly underlie the efficacy of Ahcy inhibitors, which are in preclinical development for the treatment of c-Myc-driven tumors [419].

Taken together, knockdown or inhibition of Ahcy may be a different approach to investigate hypomethylation on a genome-wide scale with fewer side-effects. In contrast, CRISPR-dCas9 based epigenome-editing tools could be used to add or remove methyl-residues in a site-specific manner [reviewed in 420], if the effects of hypo- or hypermethylation at specific promoters or cis-regulatory element are of interest.

³ See Gene Ontology category GO:0004013. In spite of its similarity with Ahcy, recombinant S-adenosylhomocysteine hydrolase-like protein 1 Ahcyl1 ectopically expressed in bacteria neither affects the enzyme activity of Ahcy nor does it itself exhibit hydrolase activity [414].

Appendices

Appendix A

Sources of reanalyzed third-party datasets

ATAC-seq

Source	Identifier	Data type	referred at
[101]	GSM1463173	ATAC-seq profile of granulocyte-macrophage progenitors as healthy control sample	Pages 84-86
[22]	GSE74688	Profiles of MLL-AF9 bulk AML samples	Pages 84-86
[22]	GSE81805	Profiles of MLL-AF9 L-GMP samples	Pages 84-86

ChIP-seq

Source	Identifier	Data type	referred at
[245]	GSE29130	H3K4me3, H3K27me3, H3K36me3 and H3K79me2 as well as MLL-AF9 ChIP-seq from HSCs, MLL-AF9 c-Kit ^{high} + c-Kit ^{low}	Page 90
[101]	GSE60103	H3K4me1, H3K4me2, H3K4me3 and H3K27ac ChIP-seq from healthy hematopoietic cells	Pages 73-80
[216]	GSE60193	H3K4me2, H3K4me3, H3K18ac, H3K27ac, H3K27me3, H3K36me3, H3K79me2 and RNA polymerase II ChIP-seqs from MLL-AF9 c-Kit ^{high} as well as c-Kit ^{low} cells	Pages 80-84
[273]	GSE78708	Mll2, Menin and H3K4me3 ChIP-seq from murine embryonal stem cells (cell line V6.5)	Pages 95-102

Hi-C

Source	Identifier	Data type	referred at
[156]	E-MTAB-3954	Hi-C and Capture Hi-C in HPC-7 cells	Pages 39 , 67 , 98

RNA-seq

Source	Identifier	Data type	referred at
[277]	GSE93622	RNA-seq of MLL-AF9-transformed c-Kit ⁺ bone marrow cells. Either Mll1, Mll2 or both were deleted	Page 101

WGBS

Source	Identifier	Data type	referred at
[125]	GSE49714	Bisulfite sequencing of secondarily transplanted wild-type hematopoietic stem cells	Pages 20-37 , 93-95

Appendix B

Sample reference



Figure B.1: Donor and recipient mice IDs and transplant hierarchy for the WGBS and CAGE-seq samples. The original donors of the untransduced bone marrow for both genotypes could no longer be determined. The mouse IDs of the primary recipients, which were transplanted by Lena Vockentanz are listed in the colored boxes. The secondary transplants of frozen ex-vivo leukemic bone marrow from the primary recipients were performed by Irina Savylyeva, who also sorted the ex-vivo bone marrow after the animals succumbed to their disease.

Appendix C

Lebenslauf

The CV was redacted from the online version of the thesis.

Bibliography

- [1] Ziemin-van der Poel S, McCabe NR, Gill HJ, Espinosa R, Patel Y, Harden A, et al. Identification of a gene, MLL, that spans the breakpoint in 11q23 translocations associated with human leukemias. *Proceedings of the National Academy of Sciences*. 1991;88(23):10735–10739.
- [2] Thirman MJ, Gill HJ, Burnett RC, Mbangkollo D, McCabe NR, Kobayashi H, et al. Rearrangement of the MLL gene in acute lymphoblastic and acute myeloid leukemias with 11q23 chromosomal translocations. *New England Journal of Medicine*. 1993;329(13):909–914.
- [3] Li BE, Ernst P. Two decades of leukemia oncprotein epistasis: the MLL1 paradigm for epigenetic deregulation in leukemia. *Exp Hematol*. 2014 Dec;42(12):995–1012. Available from: <http://dx.doi.org/10.1016/j.exphem.2014.09.006>.
- [4] Van der Burg M, Beverloo H, Langerak A, Wijsman J, Van Drunen E, Slater R, et al. Rapid and sensitive detection of all types of MLL gene translocations with a single FISH probe set. *Leukemia* (08876924). 1999;13(12).
- [5] Slany RK. The molecular mechanics of mixed lineage leukemia. *Oncogene*. 2016;35(40):5215–5223.
- [6] Kobayashi H, Thirman M, Gill H, Fernald A, Diaz M, Le Beau M, et al. Heterogeneity of breakpoints of 11q23 rearrangements in hematologic malignancies identified with fluorescence in situ hybridization. *Blood*. 1993;82(2):547–551.
- [7] Tkachuk DC, Kohler S, Cleary ML. Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias. *Cell*. 1992;71(4):691–700.
- [8] Djabali M, Selleri L, Parry P, Bower M, Young BD, Evans GA. A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias. *Nature genetics*. 1992;2(2):113–118.
- [9] Rowley JD. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*. 1973;243(5405):290–293.
- [10] Sorensen P, Chen Cs, Smith FO, Arthur DC, Domer PH, Bernstein ID, et al. Molecular rearrangements of the MLL gene are present in most cases of infant acute myeloid leukemia and are strongly correlated with monocytic or myelomonocytic phenotypes. *Journal of Clinical Investigation*. 1994;93(1):429.
- [11] Balgobind BV, Raimondi SC, Harbott J, Zimmermann M, Alonzo TA, Aufrignon A, et al. Novel prognostic subgroups in childhood 11q23/MLL-rearranged acute myeloid leukemia: results of an international retrospective study. *Blood*. 2009 Sep;114:2489–2496.
- [12] Bolouri H, Farrar JE, Triche Jr T, Ries RE, Lim EL, Alonzo TA, et al. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. *Nature medicine*. 2017;
- [13] Super H, McCabe NR, Thirman MJ, Larson RA, Le Beau MM, Pedersen-Bjergaard J, et al. Rearrangements of the MLL gene in therapy-related acute myeloid leukemia in patients previously treated with agents targeting DNA-topoisomerase II. *Blood*. 1993;82(12):3705–3711.
- [14] Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016 May;127:2391–2405.
- [15] He N, Liu M, Hsu J, Xue Y, Chou S, Burlingame A, et al. HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of HIV-1 transcription. *Molecular cell*. 2010 May;38:428–438.

- [16] Lin C, Smith ER, Takahashi H, Lai KC, Martin-Brown S, Florens L, et al. AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Molecular cell*. 2010;37(3):429–437.
- [17] Mohan M, Lin C, Guest E, Shilatifard A. Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nature reviews Cancer*. 2010 Oct;10:721–728.
- [18] Calvanese V, Nguyen AT, Bolan TJ, Vavolina A, Su T, Lee LK, et al. MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. *Nature*. 2019 Nov;.
- [19] Somervaille TCP, Matheny CJ, Spencer GJ, Iwasaki M, Rinn JL, Witten DM, et al. Hierarchical maintenance of MLL myeloid leukemia stem cells employs a transcriptional program shared with embryonic rather than adult stem cells. *Cell Stem Cell*. 2009 Feb;4(2):129–140. Available from: <http://dx.doi.org/10.1016/j.stem.2008.11.015>.
- [20] Krivtsov AV, Figueroa ME, Sinha AU, Stubbs MC, Feng Z, Valk PJM, et al. Cell of origin determines clinically relevant subtypes of MLL-rearranged AML. *Leukemia*. 2013 Apr;27(4):852–860. Available from: <http://dx.doi.org/10.1038/leu.2012.363>.
- [21] Stavropoulou V, Kaspar S, Brault L, Sanders MA, Juge S, Morettini S, et al. MLL-AF9 Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome. *Cancer Cell*. 2016 Jun;Available from: <http://dx.doi.org/10.1016/j.ccr.2016.05.011>.
- [22] George J, Uyar A, Young K, Kuffler L, Waldron-Francis K, Marquez E, et al. Leukaemia cell of origin identified by chromatin landscape of bulk tumour cells. *Nature communications*. 2016 Jul;7:12166.
- [23] Dick JE, Lapidot T. Biology of normal and acute myeloid leukemia stem cells. *Int J Hematol*. 2005 Dec;82(5):389–396. Available from: <http://dx.doi.org/10.1532/IJH97.05144>.
- [24] Krivtsov AV, Twomey D, Feng Z, Stubbs MC, Wang Y, Faber J, et al. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature*. 2006 Aug;442(7104):818–822. Available from: <http://dx.doi.org/10.1038/nature04980>.
- [25] Somervaille TCP, Cleary ML. Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. *Cancer Cell*. 2006 Oct;10(4):257–68. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17045204>.
- [26] Johnson TB, Coghill RD. Researches on pyrimidines C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. *Journal of the American Chemical Society*. 1925;47(11):2838–2844. Available from: <https://doi.org/10.1021/ja01688a030>.
- [27] Cohn WE. The Isolation and Identification of Desoxy-5-methylcytidylic Acid from Thymus Nucleic Acid1. *J Am Chem Soc*. 1951 Apr;73(4):1539–1541. Available from: <https://doi.org/10.1021/ja01148a039>.
- [28] Dominissini D, Moshitch-Moshkovitz S, Schwartz S, Salmon-Divon M, Ungar L, Osenberg S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*. 2012 Apr;485:201–206.
- [29] Ji P, Wang X, Xie N, Li Y. N6-Methyladenosine in RNA and DNA: An Epitranscriptomic and Epigenetic Player Implicated in Determination of Stem Cell Fate. *Stem cells international*. 2018;2018:3256524.
- [30] Schübeler D. Function and information content of DNA methylation. *Nature*. 2015 Jan;517:321–326.
- [31] Luo C, Hajkova P, Ecker JR. Dynamic DNA methylation: In the right place at the right time. *Science (New York, NY)*. 2018 Sep;361:1336–1340.
- [32] Lyko F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature reviews Genetics*. 2018 Feb;19:81–92.
- [33] Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature reviews Molecular cell biology*. 2013 Jun;14:341–356.
- [34] Shen L, Song CX, He C, Zhang Y. Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annual review of biochemistry*. 2014;83:585–614.

- [35] Plongthongkum N, Diep DH, Zhang K. Advances in the profiling of DNA modifications: cytosine methylation and beyond. *Nature reviews Genetics*. 2014 Oct;15:647–661.
- [36] van der Wijst MGP, Venkiteswaran M, Chen H, Xu GL, Plösch T, Rots MG. Local chromatin microenvironment determines DNMT activity: from DNA methyltransferase to DNA demethylase or DNA dehydroxymethylase. *Epigenetics*. 2015;10:671–676.
- [37] Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic acids research*. 2005 Jan;33(18):5868–77. Available from: <http://www.ncbi.nlm.nih.gov/article/abstract.fcgi?artid=1258174&tool=pmcentrez&rendertype=abstract>.
- [38] Mulqueen RM, Pokholok D, Norberg SJ, Torkenczy KA, Fields AJ, Sun D, et al. Highly scalable generation of DNA methylation profiles in single cells. *Nature biotechnology*. 2018 Apr;
- [39] Wreczycka K, Gosdschan A, Yusuf D, Grüning B, Assenov Y, Akalin A. Strategies for analyzing bisulfite sequencing data. *Journal of biotechnology*. 2017 Nov;261:105–115.
- [40] Xing X, Zhang B, Li D, Wang T. Comprehensive Whole DNA Methylome Analysis by Integrating MeDIP-seq and MRE-seq. *Methods in molecular biology* (Clifton, NJ). 2018;1708:209–246.
- [41] Kioussis D, Vanin E, deLange T, Flavell RA, Grosveld FG. Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. *Nature*. 1983;306:662–666.
- [42] Bergman Y, Cedar H. DNA methylation dynamics in health and disease. *Nat Struct Mol Biol*. 2013 Mar;20(3):274–281. Available from: <http://dx.doi.org/10.1038/nsmb.2518>.
- [43] Rodríguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. *Nature medicine*. 2011 Mar;17(3):330–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21386836>.
- [44] Vidal E, Sayols S, Moran S, Guillaumet-Adkins A, Schroeder MP, Royo R, et al. A DNA methylation map of human cancer at single base-pair resolution. *Oncogene*. 2017 Oct;36:5648–5657.
- [45] Eriksson A, Lennartsson A, Lehmann S. Epigenetic aberrations in acute myeloid leukemia: Early key events during leukemogenesis. *Exp Hematol*. 2015 Aug;43(8):609–624. Available from: <http://dx.doi.org/10.1016/j.exphem.2015.05.009>.
- [46] Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013 May;368(22):2059–2074. Available from: <http://dx.doi.org/10.1056/NEJMoa1301689>.
- [47] Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer cell*. 2010 Jan;17:13–27.
- [48] Akalin A, Garrett-Bakelman FE, Kormaksson M, Busuttil J, Zhang L, Khrebtukova I, et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. *PLoS genetics*. 2012;8:e1002781.
- [49] Challen GA, Sun D, Mayle A, Jeong M, Luo M, Rodriguez B, et al. Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells. *Cell stem cell*. 2014 Sep;15:350–364.
- [50] Cole CB, Russler-Germain DA, Ketkar S, Verdoni AM, Smith AM, Bangert CV, et al. Haploinsufficiency for DNA methyltransferase 3A predisposes hematopoietic cells to myeloid malignancies. *The Journal of clinical investigation*. 2017 Oct;127:3657–3674.
- [51] Weinberg DN, Papillon-Cavanagh S, Chen H, Yue Y, Chen X, Rajagopalan KN, et al. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature*. 2019 Sep;:
- [52] Im AP, Sehgal AR, Carroll MP, Smith BD, Tefferi A, Johnson DE, et al. DNMT3A and IDH mutations in acute myeloid leukemia and other myeloid malignancies: associations with prognosis and potential treatment strategies. *Leukemia*. 2014 Sep;28:1774–1783.
- [53] Raffel S, Falcone M, Kneisel N, Hansson J, Wang W, Lutz C, et al. BCAT1 restricts KG levels in AML stem cells leading to IDHmut-like DNA hypermethylation. *Nature*. 2017 Nov;551:384–388.

- [54] Rampal R, Alkalin A, Madzo J, Vasanthakumar A, Pronier E, Patel J, et al. DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia. *Cell reports*. 2014 Dec;9:1841–1855.
- [55] Lorsbach R, Moore J, Mathew S, Raimondi S, Mukatira S, Downing J. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t (10; 11)(q22; q23). *Leukemia*. 2003;17(3):637–641.
- [56] Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer cell*. 2011 Jul;20:11–24.
- [57] Solary E, Bernard OA, Tefferi A, Fuks F, Vainchenker W. The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia*. 2014 Mar;28:485–496.
- [58] Sonnet M, Claus R, Becker N, Zucknick M, Petersen J, Lipka DB, et al. Early aberrant DNA methylation events in a mouse model of acute myeloid leukemia. *Genome medicine*. 2014;6:34.
- [59] Schoofs T, Berdel WE, Müller-Tidow C. Origins of aberrant DNA methylation in acute myeloid leukemia. *Leukemia*. 2013 Aug;Available from: <http://dx.doi.org/10.1038/leu.2013.242>.
- [60] Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature*. 2012;489(7414):109.
- [61] Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nature genetics*. 2015 Jun;47:598–606.
- [62] Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, Perez EM, et al. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science (New York, NY)*. 2016 Nov;354:769–773.
- [63] Lettice LA, Heaney SJH, Purdie LA, Li L, de Beer P, Oostra BA, et al. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human molecular genetics*. 2003 Jul;12:1725–1735.
- [64] Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, Lynch M, et al. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature genetics*. 2014;46(2):205.
- [65] Bertolino E, Reinitz J, Manu. The analysis of novel distal Cebpa enhancers and silencers using a transcriptional model reveals the complex regulatory logic of hematopoietic lineage specification. *Developmental biology*. 2016 May;413:128–144.
- [66] Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*. 2016 Nov;167:1369–1384.e19.
- [67] van Arensbergen J, van Steensel B, Bussemaker HJ. In search of the determinants of enhancer-promoter interaction specificity. *Trends in cell biology*. 2014;24(11):695–702.
- [68] Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, Philipsen S, et al. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes & development*. 2004 Oct;18:2485–2490.
- [69] Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, et al. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Molecular cell*. 2005 Feb;17:453–462.
- [70] Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*. 2013 Nov;503:290–294.
- [71] Haarhuis JHI, van der Weide RH, Blomen VA, Yáñez-Cuna JO, Amendola M, van Ruiten MS, et al. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell*. 2017 May;169:693–707.e14.

- [72] Kornberg RD. Mediator and the mechanism of transcriptional activation. *Trends in biochemical sciences*. 2005 May;30:235–239.
- [73] Robinson PJ, Trnka MJ, Bushnell DA, Davis RE, Mattei PJ, Burlingame AL, et al. Structure of a Complete Mediator-RNA Polymerase II Pre-Initiation Complex. *Cell*. 2016 Sep;166:1411–1422.e16.
- [74] Kaiser K, Meisterernst M. The human general co-factors. *Trends in biochemical sciences*. 1996 Sep;21:342–345.
- [75] Ogryzko VV, Schiltz RL, Russanova V, Howard BH, Nakatani Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell*. 1996 Nov;87:953–959.
- [76] Spencer TE, Jenster G, Burcin MM, Allis CD, Zhou J, Mizzen CA, et al. Steroid receptor coactivator-1 is a histone acetyltransferase. *Nature*. 1997 Sep;389:194–198.
- [77] Hirose Y, Ohkuma Y. Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression. *Journal of biochemistry*. 2007 May;141:601–608.
- [78] Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature*. 2010 May;465(7295):182–187. Available from: <http://dx.doi.org/10.1038/nature09033>.
- [79] De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, et al. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol*. 2010 May;8(5):e1000384. Available from: <http://dx.doi.org/10.1371/journal.pbio.1000384>.
- [80] Zhu Y, Sun L, Chen Z, Whitaker JW, Wang T, Wang W. Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res*. 2013 Dec;41(22):10032–10043. Available from: <http://dx.doi.org/10.1093/nar/gkt826>.
- [81] Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, et al. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature structural & molecular biology*. 2011 Jul;18:956–963.
- [82] Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014 Mar;507(7493):455–461. Available from: <http://dx.doi.org/10.1038/nature12787>.
- [83] Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature genetics*. 2014 Dec;46:1311–1320.
- [84] Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*. 2011 May;474:390–394.
- [85] Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*. 2013 Jun;498(7455):516–520. Available from: <http://dx.doi.org/10.1038/nature12210>.
- [86] Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, et al. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Molecular cell*. 2013 Aug;51:310–325.
- [87] Aguiló F, Li S, Balasubramaniyan N, Sancho A, Benko S, Zhang F, et al. Deposition of 5-Methylcytosine on Enhancer RNAs Enables the Coactivator Function of PGC-1 α . *Cell reports*. 2016 Jan;14:479–492.
- [88] Lam MTY, Li W, Rosenfeld MG, Glass CK. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci*. 2014 Apr;39(4):170–182. Available from: <http://dx.doi.org/10.1016/j.tibs.2014.02.007>.
- [89] Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drabløs F, et al. Gene regulation. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science*. 2015 Feb;347(6225):1010–1014. Available from: <http://dx.doi.org/10.1126/science.1259418>.

- [90] Su X, Malouf GG, Chen Y, Zhang J, Yao H, Valero V, et al. Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. *Oncotarget*. 2014 Oct;5:9864–9876.
- [91] Chen H, Du G, Song X, Li L. Non-coding Transcripts from Enhancers: New Insights into Enhancer Activity and Gene Expression Regulation. *Genomics, proteomics & bioinformatics*. 2017 Jun;15:201–207.
- [92] Fong YW, Zhou Q. Stimulatory effect of splicing factors on transcriptional elongation. *Nature*. 2001;414:929–933.
- [93] Austenaa LMI, Barozzi I, Simonatto M, Masella S, Della Chiara G, Ghisletti S, et al. Transcription of Mammalian cis-Regulatory Elements Is Restrained by Actively Enforced Early Termination. *Mol Cell*. 2015 Nov;60(3):460–474. Available from: <http://dx.doi.org/10.1016/j.molcel.2015.09.018>.
- [94] C Quaresma AJ, Bugai A, Barboric M. Cracking the control of RNA polymerase II elongation by 7SK snRNP and P-TEFb. *Nucleic acids research*. 2016 Sep;44:7527–7539.
- [95] Flynn RA, Do BT, Rubin AJ, Calo E, Lee B, Kuchelmeister H, et al. 7SK-BAF axis controls pervasive transcription at enhancers. *Nature structural & molecular biology*. 2016 Mar;23:231–238.
- [96] Meng FL, Du Z, Federation A, Hu J, Wang Q, Kieffer-Kwon KR, et al. Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell*. 2014 Dec;159:1538–1548.
- [97] Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*. 2010 Aug;28(8):817–825. Available from: <http://dx.doi.org/10.1038/nbt.1662>.
- [98] Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? *Mol Cell*. 2013 Mar;49(5):825–837. Available from: <http://dx.doi.org/10.1016/j.molcel.2013.01.038>.
- [99] Heinz S, Romanoski CE, Benner C, Glass CK. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol*. 2015 Feb;Available from: <http://dx.doi.org/10.1038/nrm3949>.
- [100] Lim LWK, Chung HH, Chong YL, Lee NK. A survey of recently emerged genome-wide computational enhancer predictor tools. *Computational biology and chemistry*. 2018 Jun;74:132–141.
- [101] Lara-Astiaso D, Weiner A, Lorenzo-Vivas E, Zaretsky I, Jaitin DA, David E, et al. Chromatin state dynamics during blood formation. *Science*. 2014 Aug;345(6199):943–949. Available from: <http://dx.doi.org/10.1126/science.1256271>.
- [102] Ulirsch JC, Lareau CA, Bao EL, Ludwig LS, Guo MH, Benner C, et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nature genetics*. 2019 Mar;.
- [103] Bresnick EH, Johnson KD. Blood disease-causing and -suppressing transcriptional enhancers: general principles and GATA2 mechanisms. *Blood advances*. 2019 Jul;3:2045–2056.
- [104] Zuber J, Shi J, Wang E, Rappaport AR, Herrmann H, Sison Ea, et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature*. 2011 Oct;478(7370):524–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21814200>.
- [105] Shi J, Whyte WA, Zepeda-Mendoza CJ, Milazzo JP, Shen C, Roe JS, et al. Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes & development*. 2013 Dec;27:2648–2662.
- [106] Bahr C, von Paleske L, Uslu VV, Remeseiro S, Takayama N, Ng SW, et al. A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature*. 2018 Jan;553:515–520.
- [107] Rosenbauer F, Wagner K, Kutok JL, Iwasaki H, Le Beau MM, Okuno Y, et al. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU. 1. *Nature genetics*. 2004;36(6):624.
- [108] Metcalf D, Dakic A, Mifsud S, Di Rago L, Wu L, Nutt S. Inactivation of PU.1 in adult mice leads to the development of myeloid leukemia. *Proceedings of the National Academy of Sciences of the United States of America*. 2006 Jan;103:1486–1491.
- [109] Will B, Vogler TO, Narayanagari S, Bartholdy B, Todorova TI, da Silva Ferreira M, et al. Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia. *Nat Med*. 2015 Oct;21(10):1172–1181. Available from: <http://dx.doi.org/10.1038/nm.3936>.

- [110] Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*. 2016 Aug; Available from: <http://dx.doi.org/10.1038/ng.3646>.
- [111] Mansour MR, Abraham BJ, Anders L, Berezovskaya A, Gutierrez A, Durbin AD, et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science*. 2014 Dec;346(6215):1373–1377. Available from: <http://dx.doi.org/10.1126/science.1259037>.
- [112] Vahedi G, Kanno Y, Furumoto Y, Jiang K, Parker SCJ, Erdos MR, et al. Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature*. 2015 Feb; Available from: <http://dx.doi.org/10.1038/nature14154>.
- [113] Huang Y, Mouttet B, Warnatz HJ, Risch T, Rietmann F, Frommelt F, et al. The Leukemogenic TCF3-HLF Complex Rewires Enhancers Driving Cellular Identity and Self-Renewal Conferring EP300 Vulnerability. *Cancer cell*. 2019 Nov;.
- [114] Yamazaki H, Suzuki M, Otsuki A, Shimizu R, Bresnick EH, Engel JD, et al. A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. *Cancer Cell*. 2014 Apr;25(4):415–427. Available from: <http://dx.doi.org/10.1016/j.ccr.2014.02.008>.
- [115] Gröschel S, Sanders MA, Hoogenboezem R, de Wit E, Bouwman BAM, Erpelink C, et al. A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell*. 2014 Apr;157(2):369–381. Available from: <http://dx.doi.org/10.1016/j.cell.2014.02.019>.
- [116] Bröske AM, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M, et al. DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat Genet*. 2009 Nov;41(11):1207–1215. Available from: <http://dx.doi.org/10.1038/ng.463>.
- [117] Vockentanz L, Bröske AM, Rosenbauer F. Uncovering a unique role for DNA methylation in hematopoietic and leukemic stem cells. *Cell cycle (Georgetown, Tex)*. 2010 Feb;9(4):640–1. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20107327>.
- [118] Vockentanz L. Leukemia Stem Cell Fates are Determined by DNA Methylation Levels [phdthesis]. Mathematisch-Naturwissenschaftlichen Fakultät I der Humboldt-Universität zu Berlin; 2011. Available from: <https://edoc.hu-berlin.de/bitstream/handle/18452/16973/voekentanz.pdf>.
- [119] Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*. 1992 Jun;69(6):915–926.
- [120] Tucker KL, Talbot D, Lee MA, Leonhardt H, Jaenisch R. Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene. *Proc Natl Acad Sci U S A*. 1996 Nov;93(23):12920–12925.
- [121] Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, et al. Induction of tumors in mice by genomic hypomethylation. *Science*. 2003 Apr;300(5618):489–492. Available from: <http://dx.doi.org/10.1126/science.1083558>.
- [122] Strelsemann C, Brueckner B, Musch T, Stopper H, Lyko F. Functional diversity of DNA methyltransferase inhibitors in human cancer cell lines. *Cancer research*. 2006 Mar;66:2794–2800.
- [123] Hollenbach PW, Nguyen AN, Brady H, Williams M, Ning Y, Richard N, et al. A comparison of azacitidine and decitabine activities in acute myeloid leukemia cell lines. *PloS one*. 2010 Feb;5:e9001.
- [124] Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, et al. H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell*. 2014 Jul;158(3):673–688. Available from: <http://dx.doi.org/10.1016/j.cell.2014.06.027>.
- [125] Jeong M, Sun D, Luo M, Huang Y, Challen GA, Rodriguez B, et al. Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat Genet*. 2014 Jan;46(1):17–23. Available from: <http://dx.doi.org/10.1038/ng.2836>.
- [126] Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, et al. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*. 2011 Aug;43(8):768–775. Available from: <http://dx.doi.org/10.1038/ng.865>.

- [127] Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet*. 2012 Jan;44(1):40–46. Available from: <http://dx.doi.org/10.1038/ng.969>.
- [128] Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, et al. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res*. 2012 Feb;22(2):246–258. Available from: <http://dx.doi.org/10.1101/gr.125872.111>.
- [129] Timp W, Bravo HC, McDonald OG, Goggins M, Umbrecht C, Zeiger M, et al. Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med*. 2014;6(8):61. Available from: <http://dx.doi.org/10.1186/s13073-014-0061-y>.
- [130] Lee ST, Muench MO, Fomin ME, Xiao J, Zhou M, de Smith A, et al. Epigenetic remodeling in B-cell acute lymphoblastic leukemia occurs in two tracks and employs embryonic stem cell-like signatures. *Nucleic Acids Res*. 2015 Mar;43(5):2590–2602. Available from: <http://dx.doi.org/10.1093/nar/gkv103>.
- [131] Meuleman W, Peric-Hupkes D, Kind J, Beaudry JB, Pagie L, Kellis M, et al. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res*. 2013 Feb;23(2):270–280. Available from: <http://dx.doi.org/10.1101/gr.141028.112>.
- [132] Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SWM, Solovei I, Brugman W, et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell*. 2010 May;38(4):603–613. Available from: <http://dx.doi.org/10.1016/j.molcel.2010.03.016>.
- [133] Issa JP. CpG island methylator phenotype in cancer. *Nat Rev Cancer*. 2004 Dec;4(12):988–993. Available from: <http://dx.doi.org/10.1038/nrc1507>.
- [134] Lee ST, Wiemels JL. Genome-wide CpG island methylation and intergenic demethylation propensities vary among different tumor sites. *Nucleic Acids Res*. 2015 Oct;Available from: <http://dx.doi.org/10.1093/nar/gkv1038>.
- [135] Wu H, Caffo B, Jaffee HA, Irizarry RA, Feinberg AP. Redefining CpG islands using hidden Markov models. *Biostatistics*. 2010;p. kxq005. Available from: <http://rafalab.jhsph.edu/CGI/>.
- [136] Hebestreit K, Dugas M, Klein HU. Detection of significantly differentially methylated regions in targeted bisulfite sequencing data. *Bioinformatics*. 2013 Jul;29(13):1647–1653. Available from: <http://dx.doi.org/10.1093/bioinformatics/btt263>.
- [137] Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LTY, Kohlbacher O, et al. Charting a dynamic DNA methylation landscape of the human genome. *Nature*. 2013 Aug;500(7463):477–481. Available from: <http://dx.doi.org/10.1038/nature12433>.
- [138] Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature*. 2008 Jun;453(7197):948–951. Available from: <http://dx.doi.org/10.1038/nature06947>.
- [139] Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*. 2011 Dec;480(7378):490–495. Available from: <http://dx.doi.org/10.1038/nature10716>.
- [140] Schoofs T, Rohde C, Hebestreit K, Klein HU, Göllner S, Schulze I, et al. DNA methylation changes are a late event in acute promyelocytic leukemia and coincide with loss of transcription factor binding. *Blood*. 2013 Jan;121(1):178–187. Available from: <http://dx.doi.org/10.1182/blood-2012-08-448860>.
- [141] Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009 Nov;462(7271):315–322. Available from: <http://dx.doi.org/10.1038/nature08514>.
- [142] Burger L, Gaidatzis D, Schübeler D, Stadler MB. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res*. 2013 Sep;41(16):e155. Available from: <http://dx.doi.org/10.1093/nar/gkt599>.

- [143] Kumaraswamy P. A generalized probability density function for double-bounded random processes. *Journal of Hydrology*. 1980;46(1-2):79–88. Available from: <http://www.sciencedirect.com/science/article/pii/0022169480900360>.
- [144] Jones M. Kumaraswamys distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*. 2009;6(1):70–81.
- [145] Fan S, Zhang MQ, Zhang X. Histone methylation marks play important roles in predicting the methylation status of CpG islands. *Biochemical and biophysical research communications*. 2008;374(3):559–564.
- [146] Zheng H, Wu H, Li J, Jiang SW. CpGIMethPred: computational model for predicting methylation status of CpG islands in human genome. *BMC medical genomics*. 2013;6(1):S13.
- [147] Hansen KD, Langmead B, Irizarry RA. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol*. 2012 Oct;13(10):R83. Available from: <http://dx.doi.org/10.1186/gb-2012-13-10-r83>.
- [148] Kuan PF, Wang S, Zhou X, Chu H. A statistical framework for Illumina DNA methylation arrays. *Bioinformatics*. 2010 Nov;26(22):2849–2855. Available from: <http://dx.doi.org/10.1093/bioinformatics/btq553>.
- [149] Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, et al. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One*. 2013;8(12):e81148. Available from: <http://dx.doi.org/10.1371/journal.pone.0081148>.
- [150] Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semi-parametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2011;73(1):3–36.
- [151] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2015. Available from: <http://www.R-project.org/>.
- [152] Wood SN. Thin-plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2003;65(1):95–114.
- [153] Iurlaro M, von Meyenn F, Reik W. DNA methylation homeostasis in human and mouse development. *Current opinion in genetics & development*. 2017 Apr;43:101–109.
- [154] Cruickshanks HA, McBryan T, Nelson DM, Vanderkraats ND, Shah PP, van Tuyn J, et al. Senescent cells harbour features of the cancer epigenome. *Nat Cell Biol*. 2013 Dec;15(12):1495–1506. Available from: <http://dx.doi.org/10.1038/ncb2879>.
- [155] Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*. 2015 Jul;523:240–244.
- [156] Wilson NK, Schoenfelder S, Hannah R, Sánchez Castillo M, Schütte J, Ladopoulos V, et al. Integrated genome-scale analysis of the transcriptional regulatory landscape in a blood stem/progenitor cell model. *Blood*. 2016 Mar;127(13):e12–e23. Available from: <http://dx.doi.org/10.1182/blood-2015-10-677393>.
- [157] Wilson NK, Foster SD, Wang X, Knezevic K, Schütte J, Kaimakis P, et al. Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell stem cell*. 2010 Oct;7:532–544.
- [158] Calero-Nieto FJ, Ng FS, Wilson NK, Hannah R, Moignard V, Leal-Cervantes AI, et al. Key regulators control distinct transcriptional programmes in blood progenitor and mast cells. *The EMBO journal*. 2014 Jun;33:1212–1226.
- [159] Schütte J, Wang H, Antoniou S, Jarratt A, Wilson NK, Riepsaame J, et al. An experimentally validated network of nine haematopoietic transcription factors reveals mechanisms of cell state stability. *eLife*. 2016 Feb;5:e11469.
- [160] Kruse K, Hug CB, Hernández-Rodríguez B, Vaquerizas JM. TADtool: visual parameter identification for TAD-calling algorithms. *Bioinformatics (Oxford, England)*. 2016 Oct;32:3190–3192.

- [161] Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. *Cell*. 2017 Apr;169:216–228.e19.
- [162] Flyamer IM, Gassler J, Imakaev M, Brandão HB, Ulianov SV, Abdennur N, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*. 2017 Apr;544:110–114.
- [163] Gaidatzis D, Burger L, Murr R, Lerch A, Dessus-Babus S, Schübeler D, et al. DNA sequence explains seemingly disordered methylation levels in partially methylated domains of Mammalian genomes. *PLoS Genet*. 2014 Feb;10(2):e1004143. Available from: <http://dx.doi.org/10.1371/journal.pgen.1004143>.
- [164] Thiery JP, Macaya G, Bernardi G. An analysis of eukaryotic genomes by density gradient centrifugation. *Journal of molecular biology*. 1976 Nov;108:219–235.
- [165] Aran D, Toporoff G, Rosenberg M, Hellman A. Replication timing-related and gene body-specific methylation of active human genes. *Human molecular genetics*. 2011 Feb;20:670–680.
- [166] Veland N, Hardikar S, Zhong Y, Gayatri S, Dan J, Strahl BD, et al. The Arginine Methyltransferase PRMT6 Regulates DNA Methylation and Contributes to Global DNA Hypomethylation in Cancer. *Cell reports*. 2017 Dec;21:3390–3397.
- [167] DaRosa PA, Harrison JS, Zelter A, Davis TN, Brzovic P, Kuhlman B, et al. A Bifunctional Role for the UHRF1-UBL Domain in the Control of Hemi-methylated DNA-Dependent Histone Ubiquitylation. *Molecular cell*. 2018 Nov;72:753–765.e6.
- [168] Li T, Wang L, Du Y, Xie S, Yang X, Lian F, et al. Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. *Nucleic acids research*. 2018 Apr;46:3218–3231.
- [169] Lipka DB, Wang Q, Cabezas-Wallscheid N, Klimmeck D, Weichenhan D, Herrmann C, et al. Identification of DNA methylation changes at cis-regulatory elements during early steps of HSC differentiation using tagmentation-based whole genome bisulfite sequencing. *Cell Cycle*. 2014;13(22):3476–3487. Available from: <http://dx.doi.org/10.4161/15384101.2014.973334>.
- [170] Wilson A, Oser GM, Jaworski M, Blanco-Bose WE, Laurenti E, Adolphe C, et al. Dormant and self-renewing hematopoietic stem cells and their niches. *Annals of the New York Academy of Sciences*. 2007 Jun;1106:64–75.
- [171] Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. *Annual review of biochemistry*. 2005;74:481–514.
- [172] Kahles A, Lehmann KV, Toussaint NC, Hüser M, Stark SG, Sachsenberg T, et al. Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. *Cancer cell*. 2018 Aug;34:211–224.e6.
- [173] Mendizabal I, Zeng J, Keller TE, Yi SV. Body-hypomethylated human genes harbor extensive intragenic transcriptional activity and are prone to cancer-associated dysregulation. *Nucleic acids research*. 2017 May;45:4390–4400.
- [174] Brocks D, Schmidt CR, Daskalakis M, Jang HS, Shah NM, Li D, et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nature genetics*. 2017 Jun;
- [175] Cai Y, Tsai HC, Yen RWC, Zhang YW, Kong X, Wang W, et al. Critical threshold levels of DNA methyltransferase 1 are required to maintain DNA methylation across the genome in human cancer cells. *Genome research*. 2017 Apr;27:533–544.
- [176] Bushnell B. BBMap short read aligner and other bioinformatic tools. Lawrence Berkeley National Laboratory; 2014. Available from: <https://sourceforge.net/projects/bbmap/>.
- [177] Chen Y, Lun ATL, Smyth GK. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research*. 2016;5:1438.
- [178] Carninci P, Kvam C, Kitamura A, Ohsumi T, Okazaki Y, Itoh M, et al. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics*. 1996 Nov;37(3):327–336. Available from: <http://dx.doi.org/10.1006/geno.1996.0567>.

- [179] Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A*. 2003 Dec;100(26):15776–15781. Available from: <http://dx.doi.org/10.1073/pnas.2136655100>.
- [180] Takahashi H, Lassmann T, Murata M, Carninci P. 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat Protoc*. 2012 Mar;7(3):542–561. Available from: <http://dx.doi.org/10.1038/nprot.2012.005>.
- [181] Leemans C, van der Zwalm MCH, Brueckner L, Comoglio F, van Schaik T, Pagie L, et al. Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs. *Cell*. 2019 May;177:852–864.e14.
- [182] Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, NY)*. 2009 Oct;326:289–293.
- [183] Reddy KL, Zullo JM, Bertolino E, Singh H. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature*. 2008 Mar;452:243–247.
- [184] Chen Q, Chen Y, Bian C, Fujiki R, Yu X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature*. 2013 Jan;493:561–564.
- [185] Fujiki R, Hashiba W, Sekine H, Yokoyama A, Chikanishi T, Ito S, et al. GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature*. 2011 Nov;480:557–560.
- [186] Deplus R, Delatte B, Schwinn MK, Defrance M, Méndez J, Murphy N, et al. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *The EMBO journal*. 2013 Mar;32:645–655.
- [187] Nanan KK, Ocheltree C, Sturgill D, Mandlir MD, Prigge M, Varma G, et al. Independence between pre-mRNA splicing and DNA methylation in an isogenic minigene resource. *Nucleic acids research*. 2017 Dec;45:12780–12797.
- [188] Slany RK. The molecular biology of mixed lineage leukemia. *Haematologica*. 2009;94(7):984–993.
- [189] McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research*. 2012 May;40:4288–4297.
- [190] Chen W, Kumar AR, Hudson Wa, Li Q, Wu B, Staggs Ra, et al. Malignant transformation initiated by Mll-AF9: gene dosage and critical target cells. *Cancer cell*. 2008 May;13(5):432–40. Available from: <http://www.ncbi.nlm.nih.gov/article/abstract.fcgi?artid=2430522&tool=pmcentrez&rendertype=abstract>.
- [191] Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA*. 2010 Dec;304(24):2706–2715. Available from: <http://dx.doi.org/10.1001/jama.2010.1862>.
- [192] Eppert K, Takenaka K, Lechman ER, Waldron L, Nilsson B, van Galen P, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med*. 2011 Sep;17(9):1086–1093. Available from: <http://dx.doi.org/10.1038/nm.2415>.
- [193] Cabezas-Wallscheid N, Eichwald V, de Graaf J, Löwer M, Lehr HA, Kreft A, et al. Instruction of hematopoietic lineage choices, evolution of transcriptional landscapes and cancer stem cell hierarchies derived from an AML1-ETO mouse model. *EMBO Mol Med*. 2013 Dec;5(12):1804–1820. Available from: <http://dx.doi.org/10.1002/emmm.201302661>.
- [194] Navada SC, Steinmann J, Lübbert M, Silverman LR. Clinical development of demethylating agents in hematology. *The Journal of clinical investigation*. 2014 Jan;124:40–46.
- [195] Jones PA. At the tipping point for epigenetic therapies in cancer. *The Journal of clinical investigation*. 2014 Jan;124:14–16.
- [196] Trowbridge JJ, Sinha AU, Zhu N, Li M, Armstrong Sa, Orkin SH. Haploinsufficiency of Dnmt1 impairs leukemia stem cell function through derepression of bivalent chromatin domains. *Genes & development*. 2012 Feb;26(4):344–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22345515>.

- [197] Lin SP, Chiu FY, Wang Y, Yen ML, Kao SY, Hung SC. RB maintains quiescence and prevents premature senescence through upregulation of DNMT1 in mesenchymal stromal cells. *Stem Cell Reports*. 2014 Dec;3(6):975–986. Available from: <http://dx.doi.org/10.1016/j.stemcr.2014.10.002>.
- [198] French CA, Ramirez CL, Kolmakova J, Hickman TT, Cameron MJ, Thyne ME, et al. BRD-NUT onco-proteins: a family of closely related nuclear proteins that block epithelial differentiation and maintain the growth of carcinoma cells. *Oncogene*. 2008 Apr;27:2237–2242.
- [199] Thompson-Wicking K, Francis RW, Stirnweiss A, Ferrari E, Welch MD, Baker E, et al. Novel BRD4-NUT fusion isoforms increase the pathogenic complexity in NUT midline carcinoma. *Oncogene*. 2013 Sep;32:4664–4674.
- [200] Dawson MA, Prinjha RK, Dittmann A, Girotopoulos G, Bantscheff M, Chan WI, et al. Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*. 2011;478(7370):529.
- [201] Roe JS, Mercan F, Rivera K, Pappin DJ, Vakoc CR. BET Bromodomain Inhibition Suppresses the Function of Hematopoietic Transcription Factors in Acute Myeloid Leukemia. *Mol Cell*. 2015 Jun;58(6):1028–1039. Available from: <http://dx.doi.org/10.1016/j.molcel.2015.04.011>.
- [202] Fong CY, Gilan O, Lam EYN, Rubin AF, Ftouni S, Tyler D, et al. BET inhibitor resistance emerges from leukaemia stem cells. *Nature*. 2015 Sep;525(7570):538–542. Available from: <http://dx.doi.org/10.1038/nature14888>.
- [203] Liao S, Maertens O, Cichowski K, Elledge SJ. Genetic modifiers of the BRD4-NUT dependency of NUT midline carcinoma uncovers a synergism between BETis and CDK4/6is. *Genes & development*. 2018 Sep;32:1188–1200.
- [204] Duan B, Davis R, Sadat EL, Collins J, Sternweis PC, Yuan D, et al. Distinct roles of adenylyl cyclase VII in regulating the immune responses in mice. *Journal of immunology (Baltimore, Md : 1950)*. 2010 Jul;185:335–344.
- [205] Collazo CM, Yap GS, Sempowski GD, Lusby KC, Tessarollo L, Vande Woude GF, et al. Inactivation of LRG-47 and IRG-47 reveals a family of interferon gamma-inducible genes with essential, pathogen-specific roles in resistance to infection. *The Journal of experimental medicine*. 2001 Jul;194:181–188.
- [206] Bafica A, Feng CG, Santiago HC, Aliberti J, Cheever A, Thomas KE, et al. The IFN-inducible GT-Pase LRG47 (Irgm1) negatively regulates TLR4-triggered proinflammatory cytokine production and prevents endotoxemia. *Journal of immunology (Baltimore, Md : 1950)*. 2007 Oct;179:5514–5522.
- [207] Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*. 2009 Jan;37:1–13.
- [208] Zhao S, Guo Y, Shyr Y. KEGGprofile: An annotation and visualization package for multi-types and multi-groups expression data in KEGG pathway; 2015. R package version 1.16.0.
- [209] Valentine WN, Pearce ML, Lawrence JS. Studies on the histamine content of blood, with special reference to leukemia, leukemoid reactions and leukocytoses. *Blood*. 1950 Jul;5:623–647.
- [210] Mellqvist UH, Hansson M, Brune M, Dahlgren C, Hermodsson S, Hellstrand K. Natural killer cell dysfunction and apoptosis induced by chronic myelogenous leukemia cells: role of reactive oxygen species and regulation by histamine. *Blood*. 2000 Sep;96:1961–1968.
- [211] Aurelius J, Martner A, Brune M, Palmqvist L, Hansson M, Hellstrand K, et al. Remission maintenance in acute myeloid leukemia: impact of functional histamine H2 receptors expressed by leukemic cells. *Haematologica*. 2012 Dec;97:1904–1908.
- [212] Monczor F, Copsel S, Fernandez N, Davio C, Shayo C. Histamine H2 receptor in blood cells: a suitable target for the treatment of acute myeloid leukemia. In: *Histamine and Histamine Receptors in Health and Disease*. Springer; 2016. p. 141–160.
- [213] Kühn B, Schmid A, Harteneck C, Gudermann T, Schultz G. G proteins of the Gq family couple the H2 histamine receptor to phospholipase C. *Molecular endocrinology (Baltimore, Md)*. 1996 Dec;10:1697–1707.

- [214] Zang C, Schones DE, Zeng C, Cui K, Zhao K, Peng W. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. *Bioinformatics*. 2009 Aug;25(15):1952–1958. Available from: <http://dx.doi.org/10.1093/bioinformatics/btp340>.
- [215] Xu S, Grullon S, Ge K, Peng W. Spatial clustering for identification of ChIP-enriched regions (SICER) to map regions of histone methylation patterns in embryonic stem cells. *Methods in molecular biology* (Clifton, NJ). 2014;1150:97–111.
- [216] Wong SHK, Goode DL, Iwasaki M, Wei MC, Kuo HP, Zhu L, et al. The H3K4-Methyl Epigenome Regulates Leukemia Stem Cell Oncogenic Potential. *Cancer Cell*. 2015 Aug;28(2):198–209. Available from: <http://dx.doi.org/10.1016/j.ccr.2015.06.003>.
- [217] Somerville TDD, Wiseman DH, Spencer GJ, Huang X, Lynch JT, Leong HS, et al. Frequent Derepression of the Mesenchymal Transcription Factor Gene FOXC1 in Acute Myeloid Leukemia. *Cancer cell*. 2015 Sep;28:329–342.
- [218] Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, et al. The transcriptional landscape of the mammalian genome. *Science* (New York, NY). 2005 Sep;309:1559–1563.
- [219] Faulkner GJ, Forrest ARR, Chalk AM, Schroder K, Hayashizaki Y, Carninci P, et al. A rescue strategy for multimapping short sequence tags refines surveys of transcriptional activity by CAGE. *Genomics*. 2008 Mar;91(3):281–288. Available from: <http://dx.doi.org/10.1016/j.ygeno.2007.11.003>.
- [220] Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013 Aug;8(8):1494–1512. Available from: <http://dx.doi.org/10.1038/nprot.2013.084>.
- [221] Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research*. 2013 May;41:e108.
- [222] Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* (Oxford, England). 2014 Apr;30:923–930.
- [223] Maretty L, Sibbesen JA, Krogh A. Bayesian transcriptome assembly. *Genome biology*. 2014;15:501.
- [224] Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology*. 2015 Mar;33:290–295.
- [225] Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols*. 2016 Sep;11:1650–1667.
- [226] Pujadas E, Feinberg AP. Regulated noise in the epigenetic landscape of development and disease. *Cell*. 2012 Mar;148(6):1123–1131. Available from: <http://dx.doi.org/10.1016/j.cell.2012.02.045>.
- [227] Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet*. 2013 Oct;45(10):1198–1206. Available from: <http://dx.doi.org/10.1038/ng.2746>.
- [228] Kieffer-Kwon KR, Tang Z, Mathe E, Qian J, Sung MH, Li G, et al. Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell*. 2013 Dec;155(7):1507–1520. Available from: <http://dx.doi.org/10.1016/j.cell.2013.11.039>.
- [229] Schlesinger F, Smith AD, Gingeras TR, Hannon GJ, Hodges E. De novo DNA demethylation and noncoding transcription define active intergenic regulatory elements. *Genome Res*. 2013 Oct;23(10):1601–1614. Available from: <http://dx.doi.org/10.1101/gr.157271.113>.
- [230] Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, et al. Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res*. 2013 Mar;23(3):555–567. Available from: <http://dx.doi.org/10.1101/gr.147942.112>.
- [231] Sheaffer KL, Kim R, Aoki R, Elliott EN, Schug J, Burger L, et al. DNA methylation is required for the control of stem cell differentiation in the small intestine. *Genes Dev*. 2014 Mar;28(6):652–664. Available from: <http://dx.doi.org/10.1101/gad.230318.113>.
- [232] Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol*. 2013;14(3):R21. Available from: <http://dx.doi.org/10.1186/gb-2013-14-3-r21>.

- [233] Aran D, Hellman A. Unmasking risk loci: DNA methylation illuminates the biology of cancer predisposition: analyzing DNA methylation of transcriptional enhancers reveals missed regulatory links between cancer risk loci and genes. *Bioessays*. 2014 Feb;36(2):184–190. Available from: <http://dx.doi.org/10.1002/bies.201300119>.
- [234] Pacis A, Tailleux L, Morin AM, Lambourne J, MacIsaac JL, Yotova V, et al. Bacterial infection remodels the DNA methylation landscape of human dendritic cells. *Genome research*. 2015 Dec;25:1801–1811.
- [235] Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*. 2014 Apr;15(4):272–286. Available from: <http://dx.doi.org/10.1038/nrg3682>.
- [236] Whitaker JW, Nguyen TT, Zhu Y, Wildberg A, Wang W. Computational schemes for the prediction and annotation of enhancers from epigenomic assays. *Methods*. 2015 Jan;72:86–94. Available from: <http://dx.doi.org/10.1016/j.ymeth.2014.10.008>.
- [237] Goardon N, Marchi E, Atzberger A, Quek L, Schuh A, Soneji S, et al. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell*. 2011 Jan;19(1):138–152. Available from: <http://dx.doi.org/10.1016/j.ccr.2010.12.012>.
- [238] Granja JM, Klemm S, McGinnis LM, Kathiria AS, Mezger A, Corces MR, et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nature biotechnology*. 2019 Dec;37:1458–1465.
- [239] Benton ML, Talipineni SC, Kostka D, Capra JA. Genome-wide enhancer annotations differ significantly in genomic distribution, evolution, and function. *BMC genomics*. 2019 Jun;20:511.
- [240] Young RS, Kumar Y, Bickmore WA, Taylor MS. Bidirectional transcription initiation marks accessible chromatin and is not specific to enhancers. *Genome biology*. 2017 Dec;18:242.
- [241] Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, et al. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res*. 2013 May;23(5):800–811. Available from: <http://dx.doi.org/10.1101/gr.144899.112>.
- [242] Kwasnieski JC, Fiore C, Chaudhari HG, Cohen BA. High-throughput functional testing of ENCODE segmentation predictions. *Genome research*. 2014 Oct;24:1595–1602.
- [243] Jin Q, Yu LR, Wang L, Zhang Z, Kasper LH, Lee JE, et al. Distinct roles of GCN5/PCAF-mediated H3K9ac and CBP/p300-mediated H3K18/27ac in nuclear receptor transactivation. *The EMBO journal*. 2011 Jan;30:249–262.
- [244] Li Y, Wen H, Xi Y, Tanaka K, Wang H, Peng D, et al. AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. *Cell*. 2014 Oct;159(3):558–571. Available from: <http://dx.doi.org/10.1016/j.cell.2014.09.049>.
- [245] Bernt KM, Zhu N, Sinha AU, Vempati S, Faber J, Krivtsov AV, et al. MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell*. 2011 Jul;20(1):66–78. Available from: <http://dx.doi.org/10.1016/j.ccr.2011.06.010>.
- [246] Chen CW, Koche RP, Sinha AU, Deshpande AJ, Zhu N, Eng R, et al. DOT1L inhibits SIRT1-mediated epigenetic silencing to maintain leukemic gene expression in MLL-rearranged leukemia. *Nature medicine*. 2015 Apr;21:335–343.
- [247] Cusan M, Cai SF, Mohammad HP, Krivtsov A, Chramiec A, Loizou E, et al. LSD1 inhibition exerts its antileukemic effect by recommissioning PU.1- and C/EBP α -dependent enhancers in AML. *Blood*. 2018 Apr;131:1730–1742.
- [248] Godfrey L, Crump NT, Thorne R, Lau IJ, Repapi E, Dimou D, et al. DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. *Nature communications*. 2019 Jun;10:2803.
- [249] Belmont P, Constant JF, Demeunynck M. Nucleic acid conformation diversity: from structure to function and regulation. *Chemical Society Reviews*. 2001;30(1):70–81.
- [250] Harteis S, Schneider S. Making the bend: DNA tertiary structure and protein-DNA interactions. *International journal of molecular sciences*. 2014 Jul;15:12335–12363.

- [251] de Boer CG, Vaishnav ED, Sadeh R, Abeyta EL, Friedman N, Regev A. Deciphering eukaryotic gene-regulatory logic with 100 million random promoters. *Nature biotechnology*. 2019 Dec;.
- [252] Phan AT, Kuryavyi V, Burge S, Neidle S, Patel DJ. Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *Journal of the American Chemical Society*. 2007 Apr;129:4386–4392.
- [253] SantaLucia J, Hicks D. The thermodynamics of DNA structural motifs. *Annual review of biophysics and biomolecular structure*. 2004;33:415–440.
- [254] Rosenbauer F, Owens BM, Yu L, Tumang JR, Steidl U, Kutok JL, et al. Lymphoid cell growth and transformation are suppressed by a key regulatory element of the gene encoding PU.1. *Nature genetics*. 2006 Jan;38:27–37.
- [255] Aikawa Y, Yamagata K, Katsumoto T, Shima Y, Shino M, Stanley ER, et al. Essential role of PU.1 in maintenance of mixed lineage leukemia-associated leukemic stem cells. *Cancer science*. 2015 Mar;106:227–236.
- [256] Sharrocks AD. The ETS-domain transcription factor family. *Nature reviews Molecular cell biology*. 2001 Nov;2:827–837.
- [257] Leprince D, Gegonne A, Coll J, de Taisne C, Schneeberger A, Lagrou C, et al. A putative second cell-derived oncogene of the avian leukaemia retrovirus E26. *Nature*. 1983;306:395–397.
- [258] Amoutzias GD, Veron AS, Weiner J, Robinson-Rechavi M, Bornberg-Bauer E, Oliver SG, et al. One billion years of bZIP transcription factor evolution: conservation and change in dimerization and DNA-binding site specificity. *Molecular biology and evolution*. 2007 Mar;24:827–835.
- [259] Chaudhari HG, Cohen BA. Local sequence features that influence AP-1 , javax.xml.bind.JAXBElement@962aa40 , -regulatory activity. *Genome research*. 2018 Feb;28:171–181.
- [260] Braun TP, Okhovat M, Coblenz C, Carratt SA, Foley A, Schonrock Z, et al. Myeloid Lineage Enhancers Drive Oncogene Synergy in CEBPA/CSF3R Mutant Acute Myeloid Leukemia. *bioRxiv*. 2019; Available from: <https://www.biorxiv.org/content/early/2019/05/15/639617>.
- [261] Li W, Notani D, Rosenfeld MG. Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nature reviews Genetics*. 2016 Apr;17:207–223.
- [262] Kleftogiannis D, Ashoor H, Bajic VB. TELS: A Novel Computational Framework for Identifying Motif Signatures of Transcribed Enhancers. *Genomics, proteomics & bioinformatics*. 2018 Oct;16:332–341.
- [263] Maricque BB, Dougherty JD, Cohen BA. A genome-integrated massively parallel reporter assay reveals DNA sequence determinants of cis-regulatory activity in neural cells. *Nucleic acids research*. 2017 Feb;45:e16.
- [264] Xu C, Liu K, Lei M, Yang A, Li Y, Hughes TR, et al. DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. *Structure (London, England : 1993)*. 2018 Jan;26:85–95.e3.
- [265] Xue H, Yao T, Cao M, Zhu G, Li Y, Yuan G, et al. Structural basis of nucleosome recognition and modification by MLL methyltransferases. *Nature*. 2019 Sep;.
- [266] Allen MD, Grummitt CG, Hilcenko C, Min SY, Tonkin LM, Johnson CM, et al. Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *The EMBO journal*. 2006 Oct;25:4503–4512.
- [267] Hu S, Wan J, Su Y, Song Q, Zeng Y, Nguyen HN, et al. DNA methylation presents distinct binding sites for human transcription factors. *Elife*. 2013;2:e00726. Available from: <http://dx.doi.org/10.7554/eLife.00726>.
- [268] Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science (New York, NY)*. 2017 May;356.
- [269] Kemme CA, Marquez R, Luu RH, Iwahara J. Potential role of DNA methylation as a facilitator of target search processes for transcription factors through interplay with methyl-CpG-binding proteins. *Nucleic acids research*. 2017 Jul;45:7751–7759.

- [270] Agrawal R, Srikant R, et al. Fast algorithms for mining association rules. In: Proc. 20th int. conf. very large data bases, VLDB. vol. 1215; 1994. p. 487–499.
- [271] Chun KT, Li B, Dobrota E, Tate C, Lee JH, Khan S, et al. The epigenetic regulator CXXC finger protein 1 is essential for murine hematopoiesis. *PLoS One*. 2014;9(12):e113745. Available from: <http://dx.doi.org/10.1371/journal.pone.0113745>.
- [272] Cao W, Guo J, Wen X, Miao L, Lin F, Xu G, et al. CXXC finger protein 1 is critical for T-cell intrathymic development through regulating H3K4 trimethylation. *Nature Communications*. 2016 May;7:11687-. Available from: <http://dx.doi.org/10.1038/ncomms11687>.
- [273] Hu D, Gao X, Cao K, Morgan MA, Mas G, Smith ER, et al. Not all H3K4 methylations are created equal: Mll2/COMPASS dependency in primordial germ cell specification. *Molecular Cell*. 2017;65(3):460–475.
- [274] Hu D, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, et al. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nature structural & molecular biology*. 2013;20(9):1093–1097.
- [275] Piunti A, Shilatifard A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science (New York, NY)*. 2016 Jun;352:aad9780.
- [276] Bach C, Mueller D, Buhl S, Garcia-Cuellar M, Slany R. Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. *Oncogene*. 2009;28(6):815–823.
- [277] Chen Y, Anastassiadis K, Kranz A, Stewart AF, Arndt K, Waskow C, et al. MLL2, Not MLL1, Plays a Major Role in Sustaining MLL-Rearranged Acute Myeloid Leukemia. *Cancer Cell*. 2017;31(6):755–770.
- [278] Tybulewicz VLJ. Vav-family proteins in T-cell signalling. *Current opinion in immunology*. 2005 Jun;17:267–274.
- [279] Jackson BC, Ivanova IA, Dagnino L. An ELMO2-RhoG-ILK network modulates microtubule dynamics. *Molecular biology of the cell*. 2015 Jul;26:2712–2725.
- [280] Wang JY, Yu P, Chen S, Xing H, Chen Y, Wang M, et al. Activation of Rac1 GTPase promotes leukemia cell chemotherapy resistance, quiescence and niche interaction. *Molecular oncology*. 2013 Oct;7:907–916.
- [281] Nimmagadda SC, Frey S, Edelmann B, Hellmich C, Zaitseva L, König GM, et al. Bruton's tyrosine kinase and RAC1 promote cell survival in MLL-rearranged acute myeloid leukemia. *Leukemia*. 2018 Mar;32:846–849.
- [282] Nakamura T, Largaespada DA, Lee MP, Johnson LA, Ohyashiki K, Toyama K, et al. Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. *Nature genetics*. 1996 Feb;12:154–158.
- [283] Borrow J, Shearman AM, Stanton VP, Becher R, Collins T, Williams AJ, et al. The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9. *Nature genetics*. 1996 Feb;12:159–167.
- [284] van Zutven LJCM, Onen E, Velthuizen SCJM, van Drunen E, von Bergh ARM, van den Heuvel-Eibrink MM, et al. Identification of NUP98 abnormalities in acute leukemia: JARID1A (12p13) as a new partner gene. *Genes, chromosomes & cancer*. 2006 May;45:437–446.
- [285] Wang GG, Cai L, Pasillas MP, Kamps MP. NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nature cell biology*. 2007 Jul;9:804–812.
- [286] Wang GG, Song J, Wang Z, Dormann HL, Casadio F, Li H, et al. Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature*. 2009 Jun;459:847–851.
- [287] Franks TM, McCloskey A, Shokirev MN, Benner C, Rathore A, Hetzer MW. Nup98 recruits the Wdr82-Set1A/COMPASS complex to promoters to regulate H3K4 trimethylation in hematopoietic progenitor cells. *Genes & development*. 2017 Nov;31:2222–2234.
- [288] Xu C, Corces VG. Nascent DNA methylome mapping reveals inheritance of hemimethylation at CTCF/cohesin sites. *Science (New York, NY)*. 2018 Mar;359:1166–1170.

- [289] Beyaz S, Kim JH, Pinello L, Xifaras ME, Hu Y, Huang J, et al. The histone demethylase UTX regulates the lineage-specific epigenetic program of invariant natural killer T cells. *Nature immunology*. 2017;18(2):184–195.
- [290] Kruidenier L, Chung Cw, Cheng Z, Liddle J, Che K, Joberty G, et al. A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature*. 2012;488(7411):404–408.
- [291] Heinemann B, Nielsen JM, Hudlebusch HR, Lees MJ, Larsen DV, Boesen T, et al. Inhibition of demethylases by GSK-J1/J4. *Nature*. 2014;514(7520):E1–E2.
- [292] Morozov VM, Li Y, Clowers MM, Ishov AM. Inhibitor of H3K27 demethylase JMJD3/UTX GSK-J4 is a potential therapeutic option for castration resistant prostate cancer. *Oncotarget*. 2017;8(37):62131.
- [293] Mathur R, Sehgal L, Havranek O, Köhrer S, Khashab T, Jain N, et al. Inhibition of demethylase KDM6B sensitizes diffuse large B-cell lymphoma to chemotherapeutic drugs. *haematologica*. 2017;102(2):373–380.
- [294] Boila LD, Chatterjee SS, Banerjee D, Sengupta A. KDM6 and KDM4 histone lysine demethylases emerge as molecular therapeutic targets in human acute myeloid leukemia. *Experimental hematology*. 2017;.
- [295] Li Y, Zhang M, Sheng M, Zhang P, Chen Z, Xing W, et al. Therapeutic potential of GSK-J4, a histone demethylase KDM6B/JMJD3 inhibitor, for acute myeloid leukemia. *Journal of cancer research and clinical oncology*. 2018 Mar;.
- [296] Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, et al. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Molecular and cellular biology*. 2002 Jan;22:480–491.
- [297] He S, Sun H, Lin L, Zhang Y, Chen J, Liang L, et al. Passive DNA demethylation preferentially up-regulates pluripotency-related genes and facilitates the generation of induced pluripotent stem cells. *The Journal of biological chemistry*. 2017 Nov;292:18542–18555.
- [298] De Smet C, Loriot A. DNA hypomethylation in cancer: epigenetic scars of a neoplastic journey. *Epi-genetics*. 2010 Apr;5:206–213.
- [299] He S, Wang F, Zhang Y, Chen J, Liang L, Li Y, et al. Hemi-methylated CpG sites connect Dnmt1-knockdown-induced and Tet1-induced DNA demethylation during somatic cell reprogramming. *Cell discovery*. 2019;5:11.
- [300] Shirane K, Kurimoto K, Yabuta Y, Yamaji M, Satoh J, Ito S, et al. Global Landscape and Regulatory Principles of DNA Methylation Reprogramming for Germ Cell Specification by Mouse Pluripotent Stem Cells. *Developmental cell*. 2016 Oct;39:87–103.
- [301] Schroeder DI, Blair JD, Lott P, Yu HOK, Hong D, Crary F, et al. The human placenta methylome. *Proceedings of the National Academy of Sciences of the United States of America*. 2013 Apr;110:6037–6042.
- [302] Schultz MD, He Y, Whitaker JW, Hariharan M, Mukamel EA, Leung D, et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature*. 2015 Jul;523:212–216.
- [303] Jasencakova Z, Groth A. Replication stress, a source of epigenetic aberrations in cancer? *BioEssays : news and reviews in molecular, cellular and developmental biology*. 2010 Oct;32:847–855.
- [304] Nikolov I, Taddei A. Linking replication stress with heterochromatin formation. *Chromosoma*. 2016 Jun;125:523–533.
- [305] Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. *Nat Rev Cancer*. 2013 Jul;13(7):497–510. Available from: <http://dx.doi.org/10.1038/nrc3486>.
- [306] Clark SJ. Action at a distance: epigenetic silencing of large chromosomal regions in carcinogenesis. *Human molecular genetics*. 2007 Apr;16 Spec No 1:R88–R95.

- [307] Narita M, Núñez S, Heard E, Narita M, Lin AW, Hearn SA, et al. Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. *Cell*. 2003 Jun;113(6):703–716.
- [308] Bartkova J, Rezaei N, Liontos M, Karakaidos P, Kletsas D, Issaeva N, et al. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. *Nature*. 2006 Nov;444(7119):633–637. Available from: <http://dx.doi.org/10.1038/nature05268>.
- [309] Howard BH. Replicative senescence: considerations relating to the stability of heterochromatin domains. *Exp Gerontol*. 1996;31(1-2):281–293.
- [310] Kind J, Pagie L, de Vries SS, Nahidazar L, Dey SS, Bienko M, et al. Genome-wide maps of nuclear lamina interactions in single human cells. *Cell*. 2015 Sep;163(1):134–147. Available from: <http://dx.doi.org/10.1016/j.cell.2015.08.040>.
- [311] Chandra T, Ewels PA, Schoenfelder S, Furlan-Magaril M, Wingett SW, Kirschner K, et al. Global reorganization of the nuclear landscape in senescent cells. *Cell Rep*. 2015 Feb;10(4):471–483. Available from: <http://dx.doi.org/10.1016/j.celrep.2014.12.055>.
- [312] Espada J, Varela I, Flores I, Ugalde AP, Cadiñanos J, Pendás AM, et al. Nuclear envelope defects cause stem cell dysfunction in premature-aging mice. *J Cell Biol*. 2008 Apr;181(1):27–35. Available from: <http://dx.doi.org/10.1083/jcb.200801096>.
- [313] Grigoryan A, Guidi N, Senger K, Liehr T, Soller K, Marka G, et al. LaminA/C regulates epigenetic and chromatin architecture changes upon aging of hematopoietic stem cells. *Genome biology*. 2018 Nov;19:189.
- [314] Sadaie M, Salama R, Carroll T, Tomimatsu K, Chandra T, Young ARJ, et al. Redistribution of the Lamin B1 genomic binding profile affects rearrangement of heterochromatic domains and SAHF formation during senescence. *Genes Dev*. 2013 Aug;27(16):1800–1808. Available from: <http://dx.doi.org/10.1101/gad.217281.113>.
- [315] Shah PP, Donahue G, Otte GL, Capell BC, Nelson DM, Cao K, et al. Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev*. 2013 Aug;27(16):1787–1799. Available from: <http://dx.doi.org/10.1101/gad.223834.113>.
- [316] Brero A, Easwaran HP, Nowak D, Grunewald I, Cremer T, Leonhardt H, et al. Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J Cell Biol*. 2005 Jun;169(5):733–743. Available from: <http://dx.doi.org/10.1083/jcb.200502062>.
- [317] Linhoff MW, Garg SK, Mandel G. A high-resolution imaging approach to investigate chromatin architecture in complex tissues. *Cell*. 2015 Sep;163(1):246–255. Available from: <http://dx.doi.org/10.1016/j.cell.2015.09.002>.
- [318] Guarda A, Bolognese F, Bonapace IM, Badaracco G. Interaction between the inner nuclear membrane lamin B receptor and the heterochromatic methyl binding protein, MeCP2. *Exp Cell Res*. 2009 Jul;315(11):1895–1903. Available from: <http://dx.doi.org/10.1016/j.yexcr.2009.01.019>.
- [319] Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, et al. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nature genetics*. 2018 Apr;50:591–602.
- [320] Gorgoulis V, Adams PD, Alimonti A, Bennett DC, Bischof O, Bishop C, et al. Cellular Senescence: Defining a Path Forward. *Cell*. 2019 Oct;179:813–827.
- [321] Ito T, Teo YV, Evans SA, Neretti N, Sedivy JM. Regulation of Cellular Senescence by Polycomb Chromatin Modifiers through Distinct DNA Damage- and Histone Methylation-Dependent Pathways. *Cell reports*. 2018 Mar;22:3480–3492.
- [322] Collado M, Gil J, Efeyan A, Guerra C, Schuhmacher AJ, Barradas M, et al. Tumour biology: senescence in premalignant tumours. *Nature*. 2005 Aug;436(7051):642. Available from: <http://dx.doi.org/10.1038/436642a>.
- [323] Di Micco R, Sulli G, Dobreva M, Liontos M, Botrugno OA, Gargiulo G, et al. Interplay between oncogene-induced DNA damage response and heterochromatin in senescence and cancer. *Nature cell biology*. 2011;13(3):292–302.

- [324] Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell*. 2012 Jul;150(2):264–278. Available from: <http://dx.doi.org/10.1016/j.cell.2012.06.023>.
- [325] Jan M, Majeti R. Clonal evolution of acute leukemia genomes. *Oncogene*. 2013 Jan;32(2):135–140. Available from: <http://dx.doi.org/10.1038/onc.2012.48>.
- [326] Assenov Y, Brocks D, Gerhäuser C. Intratumor heterogeneity in epigenetic patterns. *Seminars in cancer biology*. 2018 Jan;;
- [327] Chargaff E, Lipshitz R, Green C, Hodes ME. The composition of the deoxyribonucleic acid of salmon sperm. *The Journal of biological chemistry*. 1951 Sep;192:223–230.
- [328] Domcke S, Bardet AF, Adrian Ginno P, Hartl D, Burger L, Schübeler D. Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature*. 2015 Dec;528:575–579.
- [329] Mao SQ, Ghanbarian AT, Spiegel J, Martínez Cuesta S, Beraldi D, Di Antonio M, et al. DNA G-quadruplex structures mold the DNA methylome. *Nature structural & molecular biology*. 2018 Oct;25:951–957.
- [330] Tian T, Chen YQ, Wang SR, Zhou X. G-Quadruplex: a regulator of gene expression and its chemical targeting. *Chem*. 2018;4(6):1314–1344.
- [331] Mukherjee AK, Sharma S, Chowdhury S. Non-duplex G-Quadruplex Structures Emerge as Mediators of Epigenetic Modifications. *Trends in genetics : TIG*. 2019 Feb;35:129–144.
- [332] Panda D, Debnath M, Mandal S, Bessi I, Schwalbe H, Dash J. A Nucleus-Imaging Probe That Selectively Stabilizes a Minor Conformation of c-MYC G-quadruplex and Down-regulates c-MYC Transcription in Human Cancer Cells. *Scientific reports*. 2015 Aug;5:13183.
- [333] Zeraati M, Langley DB, Schofield P, Moye AL, Rouet R, Hughes WE, et al. I-motif DNA structures are formed in the nuclei of human cells. *Nature Chemistry*. 2018 Apr;Available from: <https://doi.org/10.1038/s41557-018-0046-3>.
- [334] Viré E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature*. 2006 Feb;439:871–874.
- [335] Luu PL, Schöler HR, Araúzo-Bravo MJ. Disclosing the crosstalk among DNA methylation, transcription factors, and histone marks in human pluripotent cells through discovery of DNA methylation motifs. *Genome Res*. 2013 Dec;23(12):2013–2029. Available from: <http://dx.doi.org/10.1101/gr.155960.113>.
- [336] Rulands S, Lee HJ, Clark SJ, Angermueller C, Smallwood SA, Krueger F, et al. Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency. *Cell systems*. 2018 Jul;7:63–76.e12.
- [337] Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell*. 2013 May;153(5):1134–1148. Available from: <http://dx.doi.org/10.1016/j.cell.2013.04.022>.
- [338] Lee HJ, Hore TA, Reik W. Reprogramming the methylome: erasing memory and creating diversity. *Cell stem cell*. 2014 Jun;14:710–719.
- [339] Cabezas-Wallscheid N, Klimmeck D, Hansson J, Lipka DB, Reyes A, Wang Q, et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA Methylome analysis. *Cell Stem Cell*. 2014 Oct;15(4):507–522. Available from: <http://dx.doi.org/10.1016/j.stem.2014.07.005>.
- [340] Chen CC, Wang KY, Shen CKJ. The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *J Biol Chem*. 2012 Sep;287(40):33116–33121. Available from: <http://dx.doi.org/10.1074/jbc.C112.406975>.
- [341] Han JA, An J, Ko M. Functions of TET Proteins in Hematopoietic Transformation. *Molecules and cells*. 2015 Nov;38:925–935.
- [342] Huang H, Jiang X, Li Z, Li Y, Song CX, He C, et al. TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proceedings of the National Academy of Sciences of the United States of America*. 2013 Jul;110:11994–11999.

- [343] Huang H, Jiang X, Wang J, Li Y, Song CX, Chen P, et al. Identification of MLL-fusion/MYCmiR-26TET1 signaling circuit in MLL-rearranged leukemia. *Cancer letters*. 2016 Mar;372:157–165.
- [344] Lee SG, Cho SY, Kim MJ, Oh SH, Cho EH, Lee S, et al. Genomic breakpoints and clinical features of MLL-TET1 rearrangement in acute leukemias. *Haematologica*. 2013 Apr;98:e55–e57.
- [345] Bouras E, Karakioulaki M, Bougioukas KI, Aivaliotis M, Tzimogiorgis G, Chourdakis M. Gene promoter methylation and cancer: An umbrella review. *Gene*. 2019 Aug;710:333–340.
- [346] Lee SM, Lee J, Noh KM, Choi WY, Jeon S, Oh GT, et al. Infragenic CpG islands play important roles in bivalent chromatin assembly of developmental genes. *Proceedings of the National Academy of Sciences of the United States of America*. 2017 Mar;114:E1885–E1894.
- [347] Harrison JS, Cornett EM, Goldfarb D, DaRosa PA, Li ZM, Yan F, et al. Hemi-methylated DNA regulates DNA methylation inheritance through allosteric activation of H3 ubiquitylation by UHRF1. *eLife*. 2016 Sep;5.
- [348] Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochimica et biophysica acta*. 2014 Dec;1839:1362–1372.
- [349] Schmelz K, Sattler N, Wagner M, Lübbert M, Dörken B, Tamm I. Induction of gene expression by 5-Aza-2'-deoxycytidine in acute myeloid leukemia (AML) and myelodysplastic syndrome (MDS) but not epithelial cells by DNA-methylation-dependent and -independent mechanisms. *Leukemia*. 2005 Jan;19:103–111.
- [350] Flotho C, Claus R, Batz C, Schneider M, Sandrock I, Ihde S, et al. The DNA methyltransferase inhibitors azacitidine, decitabine and zebularine exert differential effects on cancer gene expression in acute myeloid leukemia cells. *Leukemia*. 2009 Jun;23(6):1019–1028. Available from: <http://dx.doi.org/10.1038/leu.2008.397>.
- [351] Christman JK. 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene*. 2002 Aug;21:5483–5495.
- [352] Jüttermann R, Li E, Jaenisch R. Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proceedings of the National Academy of Sciences of the United States of America*. 1994 Dec;91:11797–11801.
- [353] Easwaran HP, Schermelleh L, Leonhardt H, Cardoso MC. Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO reports*. 2004 Dec;5:1181–1186.
- [354] Estève PO, Chin HG, Smallwood A, Feehery GR, Gangisetty O, Karpf AR, et al. Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. *Genes & development*. 2006 Nov;20:3089–3103.
- [355] Milutinovic S, Zhuang Q, Niveleau A, Szyf M. Epigenetic stress response. Knockdown of DNA methyltransferase 1 triggers an intra-S-phase arrest of DNA replication and induction of stress response genes. *The Journal of biological chemistry*. 2003 Apr;278:14985–14995.
- [356] Unterberger A, Andrews SD, Weaver ICG, Szyf M. DNA methyltransferase 1 knockdown activates a replication stress checkpoint. *Molecular and cellular biology*. 2006 Oct;26:7575–7586.
- [357] Qin W, Leonhardt H, Pichler G. Regulation of DNA methyltransferase 1 by interactions and modifications. *Nucleus*. 2011;2(5):392–402. Available from: <http://dx.doi.org/10.4161/nuc.2.5.17928>.
- [358] Espada J, Peinado H, Lopez-Serra L, Setién F, Lopez-Serra P, Portela A, et al. Regulation of SNAIL1 and E-cadherin function by DNMT1 in a DNA methylation-independent context. *Nucleic acids research*. 2011 Nov;39:9194–9205.
- [359] Clements EG, Mohammad HP, Leadem BR, Easwaran H, Cai Y, Van Neste L, et al. DNMT1 modulates gene expression without its catalytic activity partially through its interactions with histone-modifying enzymes. *Nucleic acids research*. 2012 Jan;p. 1–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22278882>.
- [360] Rasmussen KD, Jia G, Johansen JV, Pedersen MT, Rapin N, Bagger FO, et al. Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev*. 2015 May;29(9):910–922. Available from: <http://dx.doi.org/10.1101/gad.260174.115>.

- [361] Herz HM, Hu D, Shilatifard A. Enhancer malfunction in cancer. *Molecular cell*. 2014;53(6):859–866.
- [362] Smith E, Shilatifard A. Enhancer biology and enhanceropathies. *Nat Struct Mol Biol*. 2014 Mar;21(3):210–219. Available from: <http://dx.doi.org/10.1038/nsmb.2784>.
- [363] Sur I, Taipale J. The role of enhancers in cancer. *Nature reviews Cancer*. 2016 Aug;16:483–493.
- [364] Kuntimaddi A, Achille NJ, Thorpe J, Lokken AA, Singh R, Hemenway CS, et al. Degree of recruitment of DOT1L to MLL-AF9 defines level of H3K79 Di- and tri-methylation on target genes and transformation potential. *Cell reports*. 2015 May;11:808–820.
- [365] Prange KHM, Mandoli A, Kuznetsova T, Wang SY, Sotoca AM, Marneth AE, et al. MLL-AF9 and MLL-AF4 oncofusion proteins bind a distinct enhancer repertoire and target the RUNX1 program in 11q23 acute myeloid leukemia. *Oncogene*. 2017 Jun;36:3346–3356.
- [366] Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 2007 Mar;39(3):311–318. Available from: <http://dx.doi.org/10.1038/ng1966>.
- [367] Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. 2011 May;473(7345):43–49. Available from: <http://dx.doi.org/10.1038/nature09906>.
- [368] Andersson R. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *Bioessays*. 2015 Mar;37(3):314–323. Available from: <http://dx.doi.org/10.1002/bies.201400162>.
- [369] Andersson R, Sandelin A, Danko CG. A unified architecture of transcriptional regulatory elements. *Trends Genet*. 2015 Aug;31(8):426–433. Available from: <http://dx.doi.org/10.1016/j.tig.2015.05.007>.
- [370] Koch F, Andrau JC. Initiating RNA polymerase II and TIPs as hallmarks of enhancer activity and tissue-specificity. *Transcription*. 2011;2(6):263–268. Available from: <http://dx.doi.org/10.4161/trns.2.6.18747>.
- [371] Rickels R, Herz HM, Sze CC, Cao K, Morgan MA, Collings CK, et al. Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nature genetics*. 2017;49(11):1647.
- [372] Bogu GK, Vizán P, Stanton LW, Beato M, Di Croce L, Martí-Renom MA. Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. *Molecular and cellular biology*. 2015 Dec;36:809–819.
- [373] Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annual review of biochemistry*. 2012;81:65–95.
- [374] Lee JH, Tate CM, You JS, Skalnik DG. Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *The Journal of biological chemistry*. 2007 May;282:13419–13428.
- [375] Lee JH, Skalnik DG. Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Molecular and cellular biology*. 2008 Jan;28:609–618.
- [376] Wu M, Wang PF, Lee JS, Martin-Brown S, Florens L, Washburn M, et al. Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. *Mol Cell Biol*. 2008 Dec;28(24):7337–7344. Available from: <http://dx.doi.org/10.1128/MCB.00976-08>.
- [377] Lee JH, Skalnik DG. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J Biol Chem*. 2005 Dec;280(50):41725–41731. Available from: <http://dx.doi.org/10.1074/jbc.M508312200>.
- [378] Clouaire T, Webb S, Skene P, Illingworth R, Kerr A, Andrews R, et al. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev*. 2012 Aug;26(15):1714–1728. Available from: <http://dx.doi.org/10.1101/gad.194209.112>.
- [379] Morgan MAJ, Rickels RA, Collings CK, He X, Cao K, Herz HM, et al. A cryptic Tudor domain links BRWD2/PHIP to COMPASS-mediated histone H3K4 methylation. *Genes & development*. 2017 Oct;31:2003–2014.

- [380] Glaser S, Schaft J, Lubitz S, Vintersten K, van der Hoeven F, Tufteland KR, et al. Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development* (Cambridge, England). 2006 Apr;133:1423–1432.
- [381] Lubitz S, Glaser S, Schaft J, Stewart AF, Anastassiadis K. Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. *Molecular biology of the cell*. 2007 Jun;18:2356–2366.
- [382] Ratajczak MZ. Why are hematopoietic stem cells so ‘sexy’? on a search for developmental explanation. *Leukemia*. 2017 Aug;31:1671–1677.
- [383] Ladanyi M, Samaniego F, Reuter VE, Motzer RJ, Jhanwar SC, Bosl GJ, et al. Cytogenetic and immunohistochemical evidence for the germ cell origin of a subset of acute leukemias associated with mediastinal germ cell tumors. *Journal of the National Cancer Institute*. 1990 Feb;82:221–227.
- [384] Abdelbasset-Ismail A, Borkowska S, Janowska-Wieczorek A, Tonn T, Rodriguez C, Moniuszko M, et al. Novel evidence that pituitary gonadotropins directly stimulate human leukemic cells-studies of myeloid cell lines and primary patient AML and CML cells. *Oncotarget*. 2016 Jan;7:3033–3046.
- [385] Hanna CW, Taudt A, Huang J, Gahurova L, Kranz A, Andrews S, et al. MLL2 conveys transcription-independent H3K4 trimethylation in oocytes. *Nature Structural & Molecular Biology*. 2018;p. 1.
- [386] Chen Y, Jones KL, Anastassiadis K, Kranz A, Stewart AF, Grembecka J, et al. Distinct pathways affected by menin versus MLL1/MLL2 in MLL-rearranged acute myeloid leukemia. *Experimental hematology*. 2019 Jan;69:37–42.
- [387] Long HK, Blackledge NP, Klose RJ. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochemical Society transactions*. 2013 Jun;41:727–740.
- [388] Birch NW. Critical Functions Specified by the MLL CXXC Domain Determine Leukemogenic Capacity [phdthesis]. Loyola University Chicago; 2013. Available from: https://ecommons.luc.edu/luc_diss/504.
- [389] Risner LE, Kuntimaddi A, Lokken AA, Achille NJ, Birch NW, Schoenfelt K, et al. Functional specificity of CpG DNA-binding CXXC domains in mixed lineage leukemia. *The Journal of biological chemistry*. 2013 Oct;288:29901–29910.
- [390] Somervaille TC, Cleary ML. Grist for the MLL: how do MLL oncogenic fusion proteins generate leukemia stem cells? *International journal of hematology*. 2010;91(5):735–741.
- [391] Basecke J, Whelan JT, Griesinger F, Bertrand FE. The MLL partial tandem duplication in acute myeloid leukaemia. *British journal of haematology*. 2006;135(4):438–449.
- [392] Dorrance AM, Liu S, Yuan W, Becknell B, Arnoczky KJ, Guimond M, et al. Mll partial tandem duplication induces aberrant Hox expression in vivo via specific epigenetic alterations. *The Journal of clinical investigation*. 2006 Oct;116:2707–2716.
- [393] Ono R, Taki T, Taketani T, Taniwaki M, Kobayashi H, Hayashi Y. LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t (10; 11)(q22; q23). *Cancer research*. 2002;62(14):4075–4080.
- [394] Choi SM, Dewar R, Burke PW, Shao L. Partial tandem duplication of KMT2A (MLL) may predict a subset of myelodysplastic syndrome with unique characteristics and poor outcome. *Haematologica*. 2018 Mar;103:e131–e134.
- [395] Yip BH, Tsai CT, Rane JK, Vetharoy W, Anguita E, Dong S, et al. Amplification of mixed lineage leukemia gene perturbs hematopoiesis and cooperates with partial tandem duplication to induce acute myeloid leukemia. *Haematologica*. 2017 Aug;102:e300–e304.
- [396] Kao HW, Liang DC, Kuo MC, Wu JH, Dunn P, Wang PN, et al. High frequency of additional gene mutations in acute myeloid leukemia with MLL partial tandem duplication: DNMT3A mutation is associated with poor prognosis. *Oncotarget*. 2015 Oct;6:33217–33225.
- [397] Sun QY, Ding LW, Tan KT, Chien W, Mayakonda A, Lin DC, et al. Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD). *Leukemia*. 2017 Jan;31:1–10.

- [398] Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, et al. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell*. 2015 Dec;163(7):1663–1677. Available from: <http://dx.doi.org/10.1016/j.cell.2015.11.013>.
- [399] Drissen R, Buza-Vidas N, Woll P, Thongjuea S, Gambardella A, Giustacchini A, et al. Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing. *Nature immunology*. 2016 Jun;17:666–676.
- [400] Buenrostro JD, Corces MR, Lareau CA, Wu B, Schep AN, Aryee MJ, et al. Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. *Cell*. 2018 May;173:1535–1548.e16.
- [401] Pomerantz MM, Ahmadiyah N, Jia L, Herman P, Verzi MP, Doddapaneni H, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nature genetics*. 2009 Aug;41:882–884.
- [402] Wasserman NF, Aneas I, Nobrega MA. An 8q24 gene desert variant associated with prostate cancer risk confers differential in vivo activity to a MYC enhancer. *Genome research*. 2010 Sep;20:1191–1197.
- [403] Planello AC, Singhania R, Kron KJ, Bailey SD, Roulois D, Lupien M, et al. Pre-neoplastic epigenetic disruption of transcriptional enhancers in chronic inflammation. *Oncotarget*. 2016 Feb;Available from: <http://dx.doi.org/10.18632/oncotarget.7513>.
- [404] Allahyar A, Vermeulen C, Bouwman BAM, Krijger PHL, Verstegen MJAM, Geeven G, et al. Enhancer hubs and loop collisions identified from single-allele topologies. *Nature genetics*. 2018 Jul;.
- [405] Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 2016 Jan;529:110–114.
- [406] Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, et al. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science (New York, NY)*. 2015 Feb;347:1017–1021.
- [407] Nwigwe IJ, Kim YJ, Wacker DA, Kim TH. Boundary Associated Long Noncoding RNA Mediates Long-Range Chromosomal Interactions. *PloS one*. 2015;10:e0136104.
- [408] Fujino T, Yamazaki Y, Largaespada DA, Jenkins NA, Copeland NG, Hirokawa K, et al. Inhibition of myeloid differentiation by Hoxa9, Hoxb8, and Meis homeobox genes. *Experimental hematology*. 2001;29(7):856–863.
- [409] Adamaki M, Lambrou GI, Athanasiadou A, Vlahopoulos S, Papavassiliou AG, Moschovi M. HOXA9 and MEIS1 gene overexpression in the diagnosis of childhood acute leukemias: Significant correlation with relapse and overall survival. *Leukemia research*. 2015;39(8):874–882.
- [410] Mohr S, Doebele C, Comoglio F, Berg T, Beck J, Bohnenberger H, et al. Hoxa9 and Meis1 Cooperatively Induce Addiction to Syk Signaling by Suppressing miR-146a in Acute Myeloid Leukemia. *Cancer Cell*. 2017;31(4):549–562.
- [411] Vogel MJ, Peric-Hupkes D, van Steensel B. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nature protocols*. 2007;2:1467–1478.
- [412] Turner MA, Yuan CS, Borchardt RT, Hershfield MS, Smith GD, Howell PL. Structure determination of selenomethionyl S-adenosylhomocysteine hydrolase using data at a single wavelength. *Nature structural biology*. 1998 May;5:369–376.
- [413] Fumi K, Beluzi R, Cuk M, Pavkov T, Kloor D, Bari I, et al. Functional analysis of human S-adenosylhomocysteine hydrolase isoforms SAHH-2 and SAHH-3. *European journal of human genetics : EJHG*. 2007 Mar;15:347–351.
- [414] Ando H, Mizutani A, Matsu-ura T, Mikoshiba K. IRBIT, a novel inositol 1,4,5-trisphosphate (IP3) receptor-binding protein, is released from the IP3 receptor upon IP3 binding to the receptor. *The Journal of biological chemistry*. 2003 Mar;278:10602–10612.
- [415] Yang X, Hu Y, Yin DH, Turner MA, Wang M, Borchardt RT, et al. Catalytic strategy of S-adenosyl-L-homocysteine hydrolase: transition-state stabilization and the avoidance of abortive reactions. *Biochemistry*. 2003 Feb;42:1900–1909.

- [416] Leal JF, Ferrer I, Blanco-Aparicio C, Hernández-Losa J, Ramón Y Cajal S, Carnero A, et al. S-adenosylhomocysteine hydrolase downregulation contributes to tumorigenesis. *Carcinogenesis*. 2008 Nov;29:2089–2095.
- [417] Hermes M, Osswald H, Riehle R, Piesch C, Kloor D. S-Adenosylhomocysteine hydrolase overexpression in HEK-293 cells: effect on intracellular adenosine levels, cell viability, and DNA methylation. *Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology*. 2008;22:223–236.
- [418] Fernandez-Sánchez ME, Gonatopoulos-Pournatzis T, Preston G, Lawlor MA, Cowling VH. S-adenosyl homocysteine hydrolase is required for Myc-induced mRNA cap methylation, protein synthesis, and cell proliferation. *Molecular and cellular biology*. 2009 Dec;29:6182–6191.
- [419] Uchiyama N, Dougan DR, Lawson JD, Kimura H, Matsumoto SI, Tanaka Y, et al. Identification of AHCY inhibitors using novel high-throughput mass spectrometry. *Biochemical and biophysical research communications*. 2017 Sep;491:1–7.
- [420] Pulecio J, Verma N, Mejía-Ramírez E, Huangfu D, Raya A. CRISPR/Cas9-Based Engineering of the Epigenome. *Cell stem cell*. 2017 Oct;21:431–447.
- [421] Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013 Dec;10(12):1213–1218. Available from: <http://dx.doi.org/10.1038/nmeth.2688>.
- [422] Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007 May;129(4):823–837. Available from: <http://dx.doi.org/10.1016/j.cell.2007.05.009>.
- [423] Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. *Science (New York, NY)*. 2007 Jun;316:1497–1502.
- [424] Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007 Aug;448:553–560.
- [425] Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome research*. 2007 Jun;17:877–885.
- [426] Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science (New York, NY)*. 2008 Dec;322:1845–1848.
- [427] Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*. 2006 Aug;126:663–676.
- [428] Todaro GJ, Green H. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *The Journal of cell biology*. 1963 May;17:299–313.
- [429] Isakova A, Groux R, Imbeault M, Rainer P, Alpern D, Dainese R, et al. SMiLE-seq identifies binding motifs of single and dimeric transcription factors. *Nature methods*. 2017 Mar;14:316–322.

Abbreviations

Reference	Explanation
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia / Acute myelogenous leukemia
ATAC-seq	Assay for Transposase Accessible Chromatin) [421]
CDF	Cumulative distribution function
ChIP-seq	Chromatin Immunoprecipitation Sequencing [422–424]
CGI	CpG-Island, areas of genomic sequence rich in CG-dinucleotides
cLAD	constitutive inter-Lamina-associated Domain [131]
cLAD	constitutive Lamina-associated Domain [131]
CML	Chronic myelogenous leukemia
CPM	Counts per Million, a measure of gene expression
DMR	Differentially methylated region
ECDF	Empirical Cumulative Distribution Function
eRNA	enhancer RNA
ES cell	Embryonic stem cell
FAIRE-seq	Formaldehyde Assisted Isolation of Regulatory Elements [425]
fLAD	flexible Lamina-associated Domain [131]
GAM	Generalized additive model [↔ subsection 4.1.1, p.33]
GLM	Generalized linear model [↔ subsection 4.1.1, p.33]
GLMM	Generalized linear mixed model
GRO-seq	Global Run-on Sequencing [426]
GWAS	Genome-wide Association Studies
HSC	Hematopoietic Stem Cell

Reference	Explanation
HSPC	Hematopoietic Stem/Progenitor Cell
iPSC / iPS cell	Induced pluripotent stem cell [427]
IQR	Interquartile range, a measure of statistical dispersion
KEEs	H3K79me2/ H3K79me3 positive enhancers with relevance for leukemia [248]
LAD	Lamina-associated Domain [138]
LOCKs	Large organized chromatin lysine modifications [305]
LRES	Long range epigenetic silencing [306]
LSC	Leukemic Stem Cell
MEF	Mouse embryonic fibroblast [428]
PCA	Principal Component Analysis, a method for dimensionality reduction
PDF	Probability Density Function
PMD	Partially Methylated Domain [141]
poly I:C	Polyinosinic:polycytidylic acid, an immunostimulant
RPKM	Reads per kilobase per million mapped reads, a unit to compare expression among transcripts
SAHF	Senescence-associated heterochromatic foci [307]
SMiLE-seq	Selective Microfluidics-based Ligand Enrichment followed by sequencing [429]
TINAT	Treatment-induced non-annotated transcription start sites [174]
TSS	Transcription start site, the 5' starting point for pre-mRNA production
WGBS	Whole-Genome Bisulfite Sequencing
WHO	World Health Organization

