Integrated analysis of epigenetic and genetic changes during MDS progression

DISSERTATION ZUR ERLANGUNG DES DOKTORGRADES
DER NATURWISSENSCHAFTEN (DR. RER. NAT.) DER FAKULTÄT FÜR
BIOLOGIE UND VORKLINISCHE MEDIZIN DER UNIVERSITÄT REGENSBURG

vorgelegt von

Sandra Pohl (geb. Hoiß)

aus

Kösching

im Jahr

2018

Das Promotionsgesuch wurde eingereicht am:
15.06.2018

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

Unterschrift:

_____

Table of Contents

List of Figures

List of tables

1 Introduction

1.1 Hematopoiesis

The formation of all cellular blood components from hematopoietic stem cells (HSCs), which reside in the bone marrow, is called hematopoiesis. These important cells are capable to self-renew and have the potential to differentiate into diverse precursor cells ensuring continuous sustenance with all blood cells[1]. Hematopoiesis in vertebrates begins in the extraembryonic yolk salc and placenta with erythroid progenitors having no pluripotency and no self-renewal potential. This first so called primitive wave or embryonic hematopoiesis is followed by the definitive wave (adult hematopoiesis), involving multipotent HSCs born in the aorta-gonad-mesonephros (AGM) region and placenta followed by migration to the fetal liver and bone marrow[1,2]. Development of all blood cell lines takes place in a hierarchical manner, with Long-term HSC (LT-HSC) and Short-term HSCs (ST-HSC) on top giving rise to various progenitor cells accompanied by successive loss of self-renewal capacity. The multipotent progenitor (MPP) either commits to the lymphoid or myeloid lineage, generating the common myeloid progenitor (CMP) or the granulocyte-monocyte-lymphoid progenitor (GMLP), respectively. The last progenitors in the hierarchy, involving the megakaryocyte-erythrocyte progenitor (MEP), granulocyte-monocyte progenitor (GMP) and the common lymphoid progenitor (CLP) finally generate functional blood cells[3,4] (see Figure 1-1).

Regulation of hematopoiesis is carried out by two categories of transcription factors, one for formation and function of hematopoietic stem cells (HSCs) and one for cell lineage specific differentiation. The first class includes transcription factors like MLL (mixed lineage-leukemia gene), RUNX1 (Runt-related transcription factor 1), ETV6 (Ets variant 6) or LMO2 (LIM domain only 2). Transcription factors like PU.1 (Spi-1 Proto-Oncogene), GATA (GATA Binding Protein) or C/EBP (CCAAT/Enhancer Binding Protein) are involved in expression of lineage-specific genes and are assigned in the second class of before mentioned TFs. Alterations in this regulatory network of transcription factors, either due to mutations or altered signal transduction result in loss of differentiation potential and uncontrolled cell proliferation. This state can lead to the formation of leukemia, whereas one has to distinguish between acute or chronic ones as well as between myeloid or lymphoid leukemia[4,5].

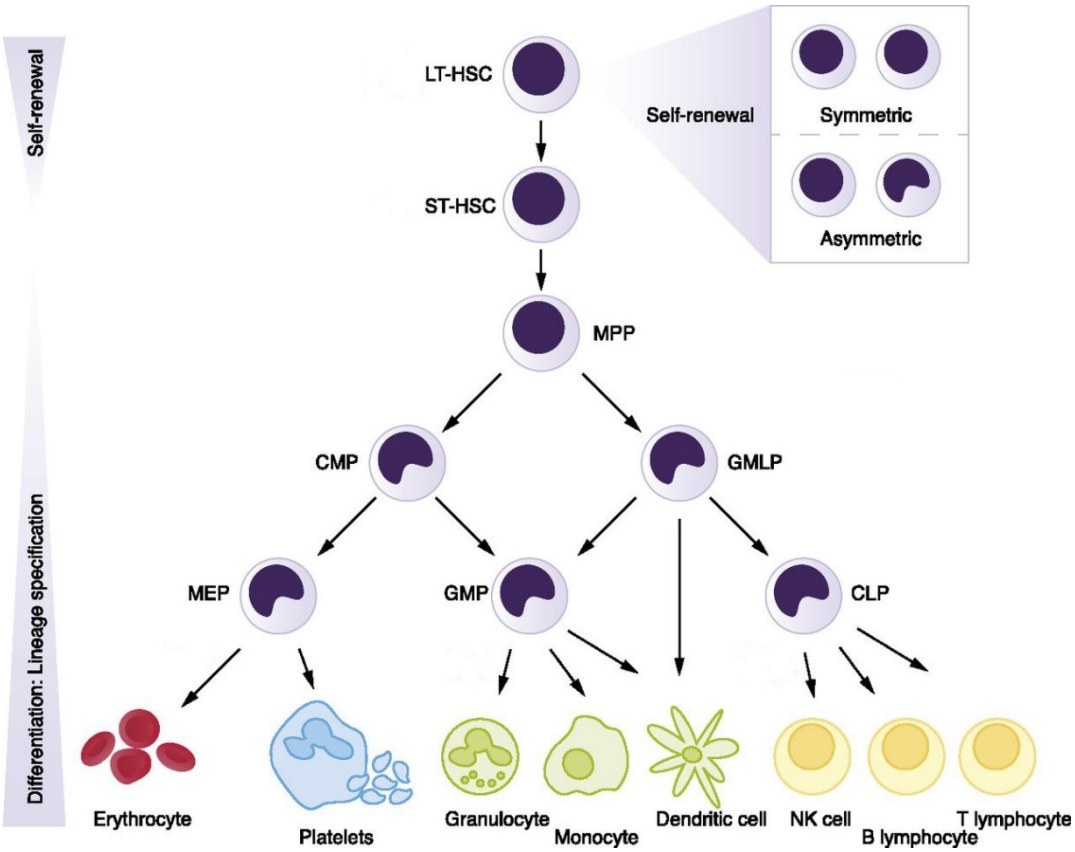


Figure 1-1 - Normal hematopoiesis

LT- and ST-HSCs give rise to various types of progenitor cells, finally leading to functional blood cells. LT-HSC: Long-term HSC, ST-HSC: Short-term HSC, MPP: multipotent progenitor, CMP: common myeloid progenitor, MEP: megakaryocyte-erythrocyte progenitor, GMP: granulocyte-monocyte progenitor, GMLP: granulocyte-monocyte-lymphoid progenitor, CLP: common lymphoid progenitor (modified from Blank et al.[3]).

1.2 Concept of epigenetics

In 1942, Conrad Waddington initially defined the term „epigenetics“ as „the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being”[6]. Nowadays this refers to a variety of processes that have heritable effects on gene expression programs without changes in DNA sequence during mitosis or meiosis. These processes are essential in multiple normal cellular processes, like embryonic development, imprinting or differentiation[7,8]. Gene expression is controlled by four different epigenetic mechanisms, involving posttranslational histone modifications, non-coding RNAs (ncRNAs)[9], chromosomal architecture and DNA methylation[10]. Besides genetic changes, many common diseases such as cancer are caused by alterations in gene expression due to epigenetic changes. Hence, integrated analysis of epigenetic and genetic changes may help to figure out where diseases, especially cancer, originate[11]. The present work mainly focuses on DNA methylation, where cancer cells often show aberrant patterns which can be addressed in some cases by medical treatment.

1.2.1 Histone code

In eukaryotes, DNA is packed into chromatin, which consists of DNA, histone proteins and non-histone proteins[12,13]. The complex of DNA and histones is designated as nucleosome. This basic subunit of chromatin consists of 147 base pairs (bp) of DNA that is ~1.7 times wrapped around a histone octamer comprised of two copies each of histones H2A, H2B, H3 and H4[14]. The arrangement of the nucleosomes resembles a structure similar to beads on a string linked by short stretches of DNA, the so called linker DNA[15]. A flexible and non-structured N-terminal tail that protrudes from the nucleosome is subject to extensive post-translational modifications (PTMs)[16,17]. Those different chemical modifications of histones, also known as histone marks, are important for transcriptional regulation. Depending on the influence of adjacent modifications, the "histone code" can be decoded in different ways, meaning that one specific histone modification can have either activating or repressive consequences[18,19]. There are several different histone modifications, including acetylation, methylation, phosphorylation, deamination, β-N-acetylglucosamination, ADP ribosylation, ubiquitylation and sumoylation, that can be found on over 60 distinct histone positions[20,21] (see Figure 1-2). The two histone modifications, acetylation and methylation, are the most important ones regulating the transcriptional state and are therefore highlighted in the following sections.

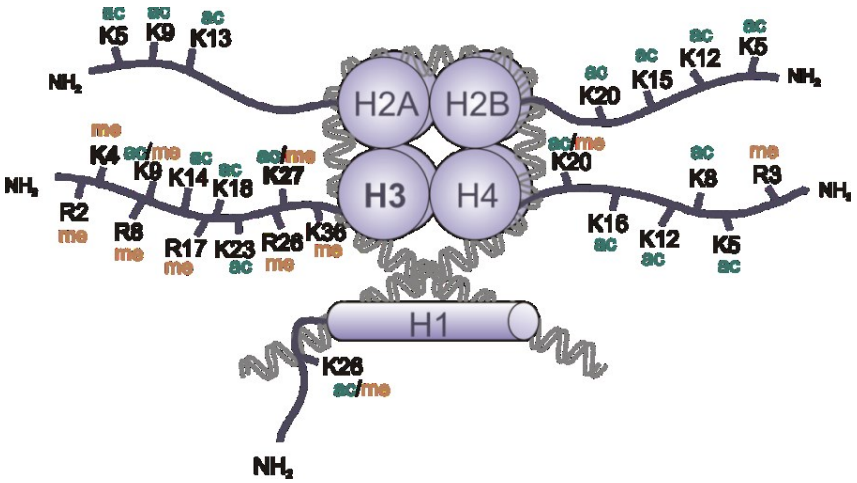


Figure 1-2 - Post-translational histone modifications

The N-terminal tails of the four core histones H2A, H2B, H3 and H4 are subject of acetylation (ac) and methylation (me) on lysine (K) and arginine (R) residues. Moreover, histone tails can be phosphorylated, sumoylated and ubquitinylated, not only at lysine and arginine but also at serine and threonine residues (adopted from Bhartiya et al.[22]).

1.2.1.1 Histone acetylation

Histone acetylation is set and erased by histone acetyltransferases (HATs) and histone deacetylases (HDACs), respectively[7]. This modification occurs on lysine residues and is associated with active transcription (see Figure 1-3). The neutralizing effect of acetylation on the positive charge of histones weakens the interaction between the negatively charged DNA and histones resulting in an open chromatin structure. Due to this accessible state, transcription can be promoted by binding of several bromodomain-containing factors[23].

1.2.1.2 Histone methylation

Histone methylation mainly occurs on the two basic residues lysine and arginine and is controlled by histone methyl transferases (HMTs) and histone demethylases, that possess stronger site specificity than HAT or HDAC[20,24,25]. In contrast to histone acetylation, histone methylation is associated with activation or repression, depending on the influence of the neighborhood. Another thing one has to keep in mind is that histone methyltransferases can methylate their target residues to a different extend. Lysine residues can be mono-, di- and trimethylated, whereas arginine can be monomethylated as well as symmetrically or unsymmetrically dimethylated[21,26]. Histone H3 methylation including H3K4, H3K36 and H3K79 are associated with active transcribed chromatin, while di- as well as trimethylation of H3K9 and H3K27 is linked to transcriptional repression[13,27] (see Figure 1-3). Important examples of both groups are H3K4me3 that marks promoters of actively transcribed genes, whereas H3K27me3 is associated with transcriptional repressed genes[28].

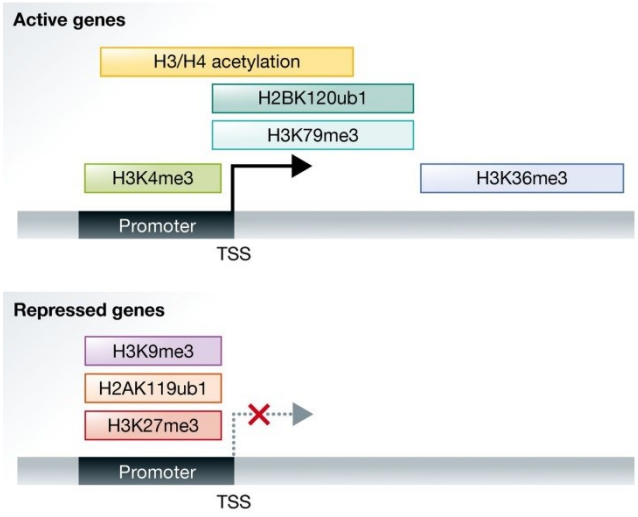


Figure 1-3 - Distribution of active and repressive histone marks

Active genes are associated with H3 and H4 acetylation, trimethylation of H3K4, trimethylation of H3K79, trimethylation of H3K36 and ubiquitylation of H2B on lysine 120. Repressed genes typically carry trimethylation of H3K27, trimethylation of H3K9 and ubiquitylation of H2A on lysine 119 (adopted from Zhang et al.[28]).

1.2.2 DNA methylation

5-methylcytosine, also known as “the fifth base” of DNA is produced by the attachment of a methyl group (CH_3) to the 5’ carbon atom of the base cytosine, catalyzed by DNA methyltransferases (DNMTs)[29]. DNA methylation mainly occurs in CpG dinucleotides (cytosines adjacent to guanines) but can also be observed at non-CpG sites, such as CpA, CpT, and CpC, primarily found in embryonic stem (ES) cells, induced pluripotent stem cells (iPS cells), neurons, oocytes and glial cells[30]. CpG dinucleotides are not equally distributed throughout the genome and are underrepresented compared to other dinucleotides possibly because of the higher tendency of 5-methylcytosine for deamination[31]. Deamination of cytosine gives rise to uracil and is recognized as foreign, which is then repaired subsequently. In contrast, deamination of 5-methylcytosine results in the formation of the naturally occurring base thymidine and is not replaced resulting in loss of cytosines throughout the genome. CpG dinucleotides are enriched in so called CpG islands (CGIs) that are often associated (37%) with promoter regions and are preferentially unmethylated in comparison to CpG dinucleotides located beyond CGIs. Basically, the methylation status of CpGs influences regulation of gene expression and is involved in several biological processes such as embryonic development[32,33], genomic imprinting[34,35], mammalian X-chromosome inactivation[35–37] or tissue-specific gene expression[38–40]. Misregulation of DNA methylation may result in aberrant DNA methylation patterns and can lead to different diseases, particularly hematopoietic malignancies such as acute myeloid leukemia or myelodysplastic syndromes[41–43], underpinning the important role of DNA methylation.

1.2.2.1 DNA methyltransferases and TET enzymes

There are five different DNA methyltransferases in mammals belonging to three distinct families, namely DNMT1, DNMT2 and DNMT3[44–46]. DNMT3a and DNMT3b are *de novo* methyltransferases targeting former unmethylated CpG dinucleotides and are important during embryonic development. DNMT1, also known as maintenance methyltransferase, shows a higher preference for hemimethylated DNA and reestablishes DNA methylation patterns after DNA replication[45,47,48]. Recognition and recruitment of DNMT1 to hemimethylated CpGs during replication is carried out by the cofactor UHRF1 (ubiquitin like with PHD and ring finger domains 1)[47,49]. The last two remaining members of mammalian DNMTs do not methylate DNA. DNMT2 was shown to methylate RNA molecules, while DNMT3L is catalytically inactive and furthermore highly expressed in germ and ES cells probably acting as a cofactor. But all have a common structure consisting of an N-terminal regulatory domain and a C-terminal catalytic domain, with DNMT2 as exception lacking the N-terminal domain. Moreover they share ten characteristic sequence motifs, most of them being highly conserved (see Figure 1-4)[46,50,51].

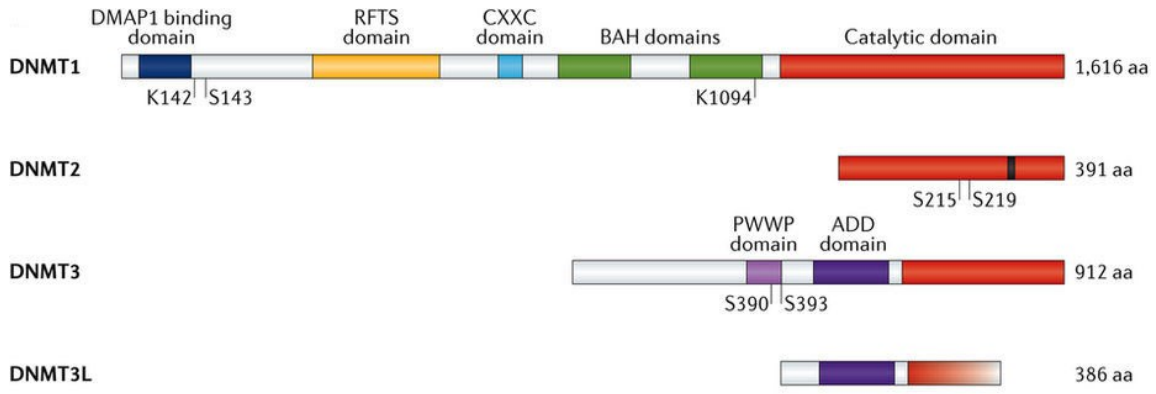


Figure 1-4 – The DNA methyltransferase family

The four different mammalian DNA methyltransferases share a conserved catalytic domain (red), with DNMT3L as exception lacking the C-terminal part of this catalytic domain. The regulatory domain of DNMT1 consists of four different domains, namely the DMAP1 binding (blue), the RFTS (yellow), the CXXC (light blue) domain and two BAH (green) domains. DNMT3 isoforms contain a PWWP (light purple) and ADD (purple) domain, whereas DNMT3L is lacking the latter domain. DMAP1: DNA methyltransferase 1-associated protein 1, RFTS: replication foci targeting sequence, CXXC: cysteine-rich zinc finger, BAH: bromo-adjacent homology, PWWP: proline-tryptophan-tryptophan-proline, ADD: ATRX–DNMT3–DNMT3L (modified from Lyko et al.[46]).

Although in general DNA methylation is a stable and heritable modification, locally, DNA methylation patterns undergo dynamic changes including both, methylation and demethylation in a tissue specific manner[52–55]. Demethylation of 5mC can occur in several ways, either passively by inhibition of the DNA methylation machinery and subsequent dilution of methylation during replication (passive DNA methylation) or actively by removal of methyl groups mediated by Ten-Eleven-Translocation proteins (active DNA methylation). This TET protein family comprises three members, TET1, TET2 and TET3, which have a common core catalytic domain required for oxidation of 5mC (see Figure 1-5). This core catalytic domain consists of a conserved double-stranded β-helix (DSBH) domain, a cysteine-rich domain as well as Fe(II) and 2-oxoglutarate (2-OG) binding sites and is responsible for binding to CpGs[56].

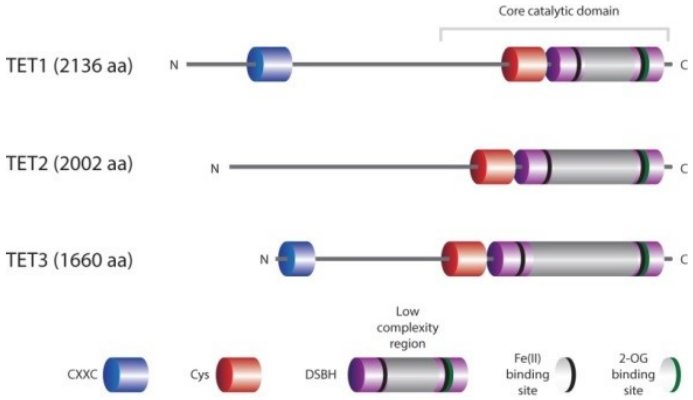


Figure 1-5 – The Ten-Eleven-Translocation (TET) protein family

The core catalytic domain of all TET enzymes consists of a cysteine-rich (Cys) domain, binding sites for the cofactors Fe(II) and 2-oxoglutarate and the DSBH (conserved double-stranded β-helix) domain.

TET1 and TET3 have an additional CXXC domain on their N-terminus important for binding to DNA (adopted from Rasmussen et al.[56]).

TET proteins iteratively oxidize 5mC to 5-hydroxymethylcytosine (5hmC), 5- formylcytosine (5fC) and 5-carboxylcytosine (5caC) (see Figure 1-6). The last two mentioned variants of cytosine can be replaced by an unmodified cytosine via base excision repair (BER) or thymine DNA glycosylase (TDG)-mediated excision.

Alternatively, the two deaminases APOBEC3 and AID can convert 5hmC into 5-hydroxymethyluracil (5hmU) which is then replaced by an unmodified cytosine via TDG/BER[57–61].

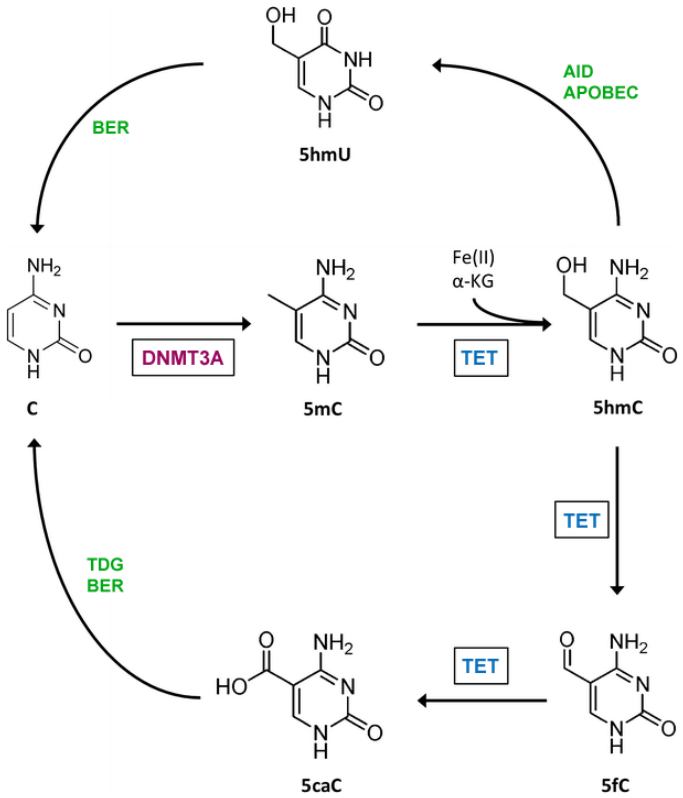


Figure 1-6 - Cycle of DNA methylation and demethylation

Cytosine is methylated via DNA methyltransferases to 5mC and this in turn is demethylated by repeated oxidation through TET oxygenases including the intermediates 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). The two oxidized forms 5fC and 5caC can be removed via TDG (thymine DNA glycosylase) mediated excision or BER and replaced by an unmodified cytosine. Involvement of the two deaminases AID and APOBEC leads to the formation of 5-hydroxymethyluracil (5hmU) from 5hmC and subsequent conversion into cytosine via base excision repair (BER) (modified from Kunimoto et al.[62]).

1.2.2.2 Methyl-CpG binding proteins

Transcriptional repression of genes through DNA methylation is caused by two different mechanisms. Firstly, the methyl group may prevent binding of sequence specific transcription factors (TFs) required for gene expression. The second possibility involves the recruitment of Methyl-CpG-binding proteins (MBPs) which either prevent binding of other factors or directly influence repression[37,48,63].

In the early 90s, the first two proteins possessing affinity for methylated DNA were described. MeCP1 and MeCP2 are both able to bind methylated CpGs, whereas MeCP1 requires a group of at least twelve methylated sites for strong binding and MeCP2 only a single methyl-CpG pair[64,65]. Later it was

shown that MeCP1 is a big complex containing MBD2 as well as all NuRD (Nucleosome Remodeling Deacetylase) components and represses transcription by recruitment of histone deacetylases and corepressor proteins[66–68]. Today, the MBD family consists of seven members: MeCP2, MBD1, MBD2, MBD3, MBD4, MBD5 and MBD6 (see Figure 1-7). Despite their shared conserved Methyl-CpG binding domain, only MeCP2, MBD1, MBD2 and MBD4 are able to bind methylated DNA[69].

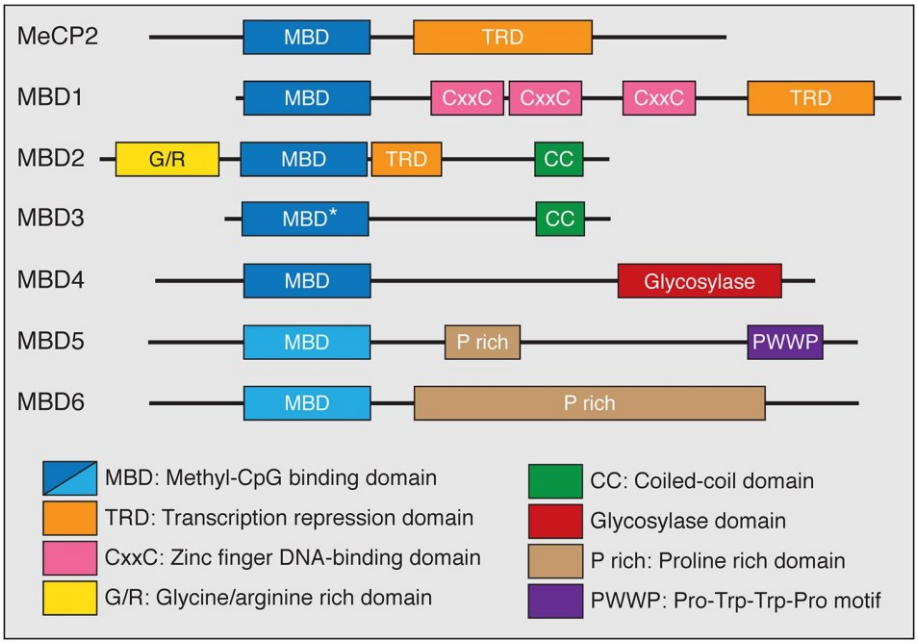


Figure 1-7 - Members of the Methyl-CpG-binding (MBP) protein family

The MBD family comprises MeCP2, MBD1-MBD6 with MeCP2 as founding member. All MBPs contain a highly conserved methyl-binding domain (MBD) for binding methylated DNA, whereas MBD3 has a point mutation (*) in this domain. In addition, MeCP2, MBD1 and MBD2 have a common transcriptional repressor domain (TRD). MBD1 also contains 3 cysteine rich zinc fingers (CXXC) and MBD2 harbors a glycine-arginine (GR) repeat as well as a C-terminal coiled-coil (CC) domain. This CC domain can also be found in MBD3. Besides the conserved MBD, MBD4 contains a C-terminal DNA glycosylase domain important for DNA repair. The last two members, MBD5 and MBD6, contain proline rich domains (P rich) while MBD5 has an additional PWWP motif that binds methylated histones (modified from Wood et al.[70]).

Structural analyses of MeCP2 identified two important domains, the MBD and the transcriptional repressor domain (TRD). The same is true for MBD1, but it contains three additional cysteine-rich domains (CXXC) where one of those is able to bind unmethylated DNA, therefore allowing repression of methylated and unmethylated DNA. MBD2 is a component of the MeCP1 complex and harbors a CpG-density dependent binding affinity. Structurally, it is characterized by two additional domains, the glycine/arginine rich domain (G/R) and the coiled-coil (CC) domain[71,72].

MBD3 is exceptional among the MBD protein family in that harboring a point mutation in the Methyl-CpG binding domain resulting in loss of methyl-CpG binding capacity. Furthermore, MBD3 represents a subunit of the NuRD complex[73]. Another outstanding feature among MBPs is the glycosylase domain of MBD4 allowing involvement in DNA repair[66,69,70]. The last two remaining members, MBD5

and MBD6, are not able to bind methylated DNA and localize preferentially to pericentric heterochromatin[74].

Besides the MBD proteins, the group of the so called DNA methylation readers includes the Kaiso family proteins and the SRA (SET- and Ring finger-associated) domain family. They can be differentiated due to their lack of the MBD domain. In contrast, Kaiso harbors several C-terminal zinc finger domains facilitating the binding to methyl-CpGs. Kaiso is also able to bind to an unmethylated specific consensus sequence, the Kaiso binding sequence (KBS), making it a bifunctional protein[75–77].

The last group of DNA methylation readers, the SET- and Ring finger-associated protein family, comprises the two members UHRF1 and UHRF2 which are known to bind hemimethylated DNA[66].

1.2.2.3 Crosstalk between DNA methylation and histone modifications

Gene expression is regulated by DNA methylation as well as histone modifications, while DNA methylation is a stable long-term repressive mark compared to histone modifications. Since these two mechanisms do not act independently, interaction partners are necessary facilitating the epigenetic crosstalk. Both, Polycomb (PcG) and Trithorax (TrxG) group proteins, are the main interaction partners and are maintaining the balance of stable repression and activation of gene expression[49,78]. They are evolutionarily conserved and working antagonistically to enable the correct expression of genes necessary for cell differentiation and developmental processes[79]. It was shown that recruitment of these two protein groups to chromatin is carried out by regulatory elements, called PcG and TrxG response elements (PREs and TREs)[80].

The Polycomb group proteins (PcGs) are responsible for gene silencing and can be divided into two main complexes, Polycomb repressive complex 1 (PRC1) and PRC2. Polycomb repressive complex 1 consists of chromobox-domain (CBX) proteins and one member each of the PCGF (Polycomb group ring fingers) family, RING1 family and the HPH family. Polycomb repressive complex 2 contains three core members, namely EZH1 or EZH2 (enhancer of zeste), EED (embryonic ectoderm development) and SUZ12 (suppressor of zeste 12)[81]. The SET domain containing methyltransferases, EZH1 and EZH2, are able to catalyze mono-, di- and trimethylation of H3K27, a histone mark associated with transcriptional repression. In turn, H3K27me3 acts as a binding platform for PRC1 that establishes a compact chromatin structure repressing gene transcription[82–84].

The Trithorax group acts like a PcG antagonist by activating gene expression due to different mechanisms. This is carried out by different members of the TrxG proteins: ATP-dependent chromatin remodeling complexes, SET domain containing factors and histone modifying proteins whose interaction is resulting in a more accessible chromatin structure facilitating transcription[85,86].

As mentioned before, histone modifications and DNA methylation are interconnected and one example for such interplay is the PcG-mediated silencing through DNA methylation. The EZH2 subunit of PRC2 catalyzes trimethylation of H3K27 functioning as the first layer of repression and due to interaction of EZH2 with DNA methyltransferases (DNMTs) the second and more stable layer of repression is set[49,87]. This phenomenon is often seen in different cancer types, where PcG targets are more frequently de *novo methylated* leading to aberrant DNA methylation patterns[88–90]. Moreover, the aforementioned Methyl-CpG binding proteins (MBPs) participate on the epigenetic crosstalk between

DNA methylation and histone modifications. MeCP2 for example interacts with histone deacetylases and the Sin3 corepressor complex via its TR domain leading to transcriptional repression. Another example is MBD1 which may act as a transcriptional repressor by binding to methylated DNA. On the other hand, unmethylated DNA is specifically bound by CXXC finger protein 1 (CFP1), a subunit of the Set1 complex and thus leading to H3K4 trimethylation at bound CpG islands representing a chromatin environment favoring transcription[88,91–93].

Another crosstalk between histone modifications and DNA methylation can be observed between H3K36me3 and DNMT3. Trimethylation of histone H3 at lysine residue 36 is preferentially found in gene bodies of actively transcribed genes and is mediated by the histone methyltransferase SETD2 (SET domain containing 2)[94]. This methyltransferase is recruited by RNA polymerase II during transcriptional elongation to maintain a repressive chromatin state preventing spurious transcription of cryptic promoters or transposon remnants[95–97]. The repressive histone mark H3K36me3 is then recognized by DNMT3 enzymes and subsequent DNA methylation of gene bodies take place[95,98].

1.2.3 Non-coding RNAs

Besides DNA methylation and histone modifications, non-coding RNAs (ncRNAs) are regulatory elements of chromatin structure and gene expression, thus providing a third epigenetic mechanism[99]. In contrast to tRNA (transfer RNA) and rRNA (ribosomal RNA), ncRNAs are only transcribed and not translated into proteins. They can be divided into small (< 200 nt) and long ncRNAs (> 200 nt), while small ncRNAs include the most prominent ones like microRNA (miRNA), small nucleolar RNA (snoRNA) and PIWI-interacting RNA (piRNA)[99,100]. The best studied class of ncRNAs, miRNA, are involved in post-transcriptional gene silencing by the RISC (RNA-induced silencing) complex or complementary interaction with mRNA[101,102]. Besides regulation of a variety of biological processes, long non-coding RNAs are also known to function in epigenetic regulation[103]. This epigenetic regulation typically results in transcriptional repression with the lncRNA *Xist* (X-inactivation specific transcripts) as prime example. After transcription from the inactive X chromosome, *XIST* binds to the Polycomb Repressive Complex 2 and trimethylation of histone H3K27 takes place, overall resulting in inactivation of the marked copy[100,104,105].

Nowadays, light was shed on lncRNAs due to their dysregulated expression and consequential role in cancer development or progression. One example is the lncRNA HOTAIR which normally represses homeobox genes (HOX) by recruitment of PRC2 and LSD1 (Lysine-specific histone demethylase 1). Overexpression of HOTAIR has been found in breast cancer, hepatocellular carcinoma or colorectal cancer playing a role in the initiation and progression of these different cancer types[103,104,106,107].

1.3 Myelodysplastic syndromes

Myelodysplastic syndromes (MDS) comprise a heterogeneous group of clonal hematopoietic neoplasms characterized by ineffective hematopoiesis resulting in peripheral blood (PB) cytopenias and an increased risk for leukemic evolution[108–110]. About 30% of patients are progressing towards acute myeloid leukemia (AML)[111]. MDS predominantly occurs in patients older than 65 and shows an incidence in the general population of about 30 – 50 cases per 1.000.000 individuals per year[108,109].
In contrast, myelodysplastic syndromes are rather rare in children with an annual incidence of 0.5 – 4 per 1.000.000 individuals[112].
Based on the WHO classification of 2016, MDS can be divided into six different clinical subtypes: MDS with single lineage dysplasia (MDS-SLD), MDS with multilineage dysplasia (MDS-MLD), MDS with ring sideroblasts (MDS-RS), MDS with isolated del(5q), MDS with excess blasts (MDS-EB) and MDS, unclassifiable (MDS-U)[113]. Concerning the highly variable prognosis the IPSS-R (International Prognostic Scoring System-Revised) represents an useful and important system for prognostication of MDS patients which classifies patients into different groups including very low risk, low risk, intermediate, high risk and very high risk[108]. Regarding the quantitative alterations in these subtypes, low risk and high risk group patients show a remarkable expansion of hematopoietic stem cells with the highest expansion rates seen in high risk MDS patients. In addition, low risk MDS subtypes show a marked increase of common myeloid progenitors (CMPs) and a decrease of megakaryocyte-erythroid progenitors (MEPs) resulting in cytopenias. High risk MDS is furthermore characterized by MEP expansion and a higher risk to develop acute myeloid leukemia[114] (see Figure 1-8).

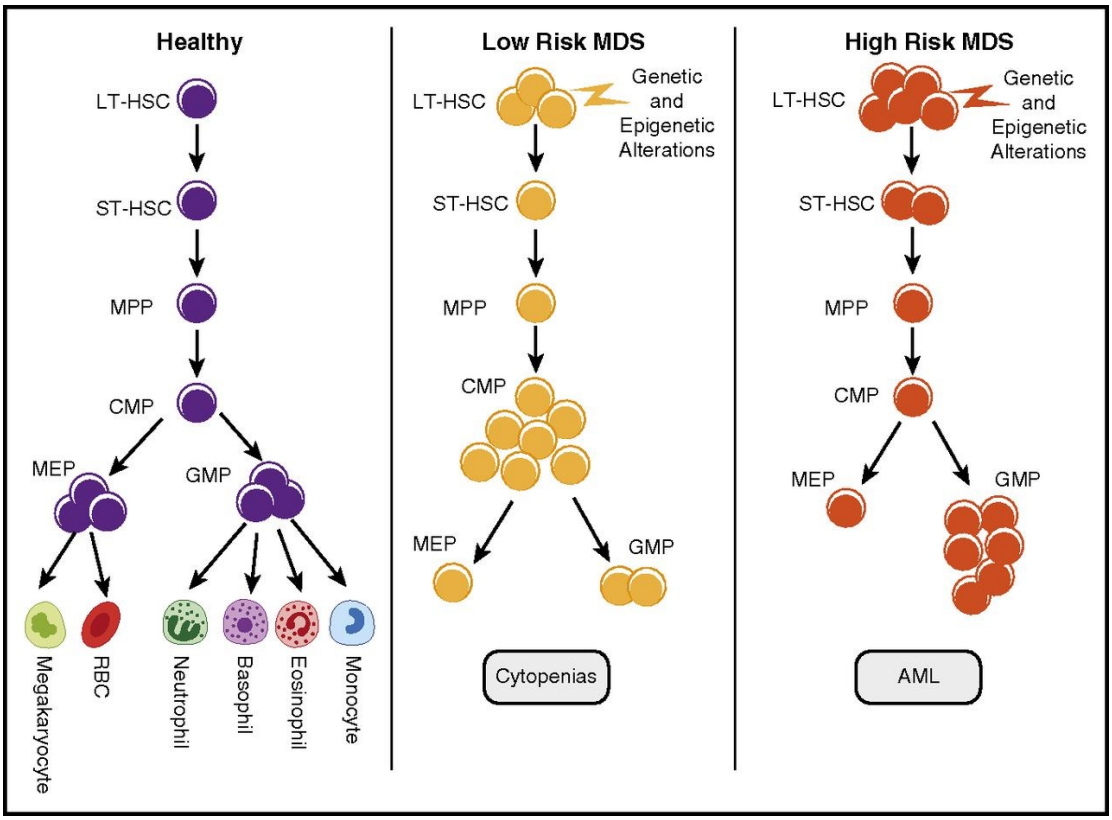


Figure 1-8 - Alterations in stem and progenitor cells in different MDS risk classes

This simplified model compares differentiation in healthy and low risk / high risk MDS cells. In patients with MDS, genetic and epigenetic alterations in long-term hematopoietic stem cells (LT-HSCs) are observed. Expansion of CMPs (common myeloid progenitor) and decrease of MEPs (megakaryocyte-erythroid progenitor), resulting in cytopenias, are characteristic for low risk MDS patients. High risk MDS is associated with GMP (granulocyte-macrophage progenitor) expansion as well as a greater increase of LT-HSCs and ST-HSCs (short-term HSCs), frequently resulting in acute myeloid leukemia (adopted from Shastri et al.[114]).

1.3.1 Pathogenesis of MDS

Cytogenetic and molecular genetics are well studied for myelodysplastic syndromes and are important for progression and prognosis, whereas epigenetic changes are rarely characterized. With the emergence of cost-effective high-throughput sequencing, mutational profiling improved understanding of the heterogeneous disease MDS and is incorporated with increasing frequency into clinical routine investigations[115]. With this approach, it was found that over 90% of patients with MDS show genetic lesions, including mutations, deletions or copy number variations[109,116]. Moreover, about 50% of MDS patients show cytogenetic abnormalities[117]. One has to keep in mind that heterogeneity of MDS is not only induced by genetic aberrations but also by the co-occurrence of cytogenetic and epigenetic alterations and for that reason following chapters will focus on those three underlying “roots” of MDS.

1.3.1.1 Gene mutations

Several studies identified a set of genes frequently mutated in myeloid malignancies that has been further investigated by targeted approaches[116,118,119].

These include genes involved in different cellular processes, like RNA splicing factors, transcription factors, cohesin components, factors important for DNA methylation and histone modification as well as signal transduction molecules[115] (see Figure 1-9).

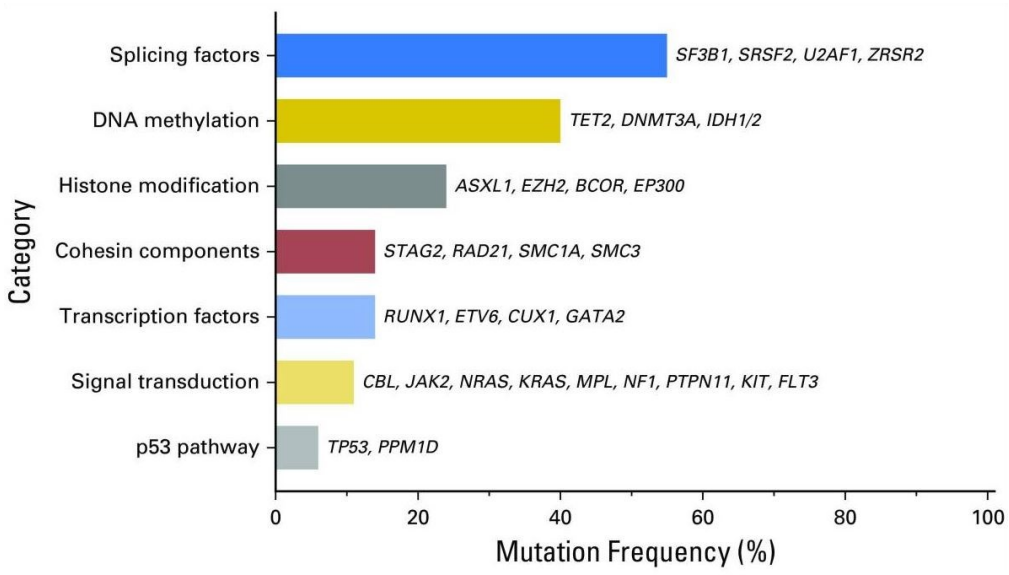


Figure 1-9 – Mutation frequency of genes in myelodysplastic syndromes

Recurrently mutated genes in MDS can be classified into different biological categories with mutations in splicing factors and DNA methylation factors resembling the most common ones with 55% to ~40%, respectively. Examples for every category are listed on the right side of the bar (adopted from Kennedy et al.[115]).

1.3.1.1.1 RNA splicing

Mutations affecting the RNA splicing machinery occur in almost 60% of patients with MDS, among *SF3B1*, *SRSF2*, *U2AF1* and *ZRSR2* being the most common ones. *SF3B1* (Splicing Factor 3b Subunit 1) shows the highest mutational rate with 28% and is strongly associated with ring sideroblasts and a better overall survival (OS)[108,120–124]. In 12 % of patients with MDS mutations in *SRSF2* (Serine And Arginine Rich Splicing Factor 2) can be observed, which are often found together with other mutations such as *RUNX1*, *IDH2* and *ASXL1* and have a poorer OS[125]. Other components of the RNA splicing machinery are mutated at lower frequencies.

In general, mutations in splicing factors were characterized by mutual exclusiveness, that means more than one mutation is almost not seen in a single neoplasia. Furthermore, mutations tend to appear in early stages of the disease and have a heterozygous character. Besides the success of understanding the importance of splicing factor mutations in MDS for pathogenesis and therapy, major questions about their biological consequences remain[108,126–128].

1.3.1.1.2 Epigenetic regulators

Mutations in factors regulating DNA methylation and histone modifications are the second most affected genes in myelodysplastic syndromes. Mutated epigenetic regulators bring along a higher risk for alterations in transcriptional processes that can be retained during cell division and the establishment of a stable MDS clone[125,129]. Mutations in genes involved in DNA methylation processes can be observed in *DNMT3A* and *TET2* with a mutational frequency of 2 – 8% and 21 %, respectively. Mutations in the DNA methyltransferase 3A are known to be loss-of-function, often co-occurring with mutations in *SF3B1* and *U2AF1* but are not the decisive factor alone for development of MDS. On the other hand, these mutations are associated with a poorer overall survival and a higher risk for leukemia development[108,130,131]. TET2 is acting as an opponent of DNMT3A and is the second most mutated gene in patients with MDS[108,132]. Several studies showed that impaired function of TET2 increases HSC self-renewal rate and promotes differentiation into myelomonocytic direction. Furthermore, *TET2* mutations are often found in early stages of the disease suggesting to be one possible initiator of MDS[133,134]. Mutations in the metabolic enzymes IDH1 and IDH2 are directly influencing TET2 activity by production of 2-hydroxyglutarate, inhibiting the hydroxylation of 5mC. Moreover *IDH1/2* and *TET2* mutations are mutually exclusive[115,130,134,135].

The two histone modifying enzymes, ASXL1 and EZH2, are also recurrently mutated in MDS with about 14% and 6%, respectively[108]. ASXL1 is involved in histone methylation via interaction with PRC2 components and is associated with a poor overall survival in MDS[136]. EZH2 is a component of the PRC2 complex and also a predictor of poor prognosis. In addition to mutations of *EZH2*, chromosomal aberrations of chromosome 7 or 7q can lead to the deletion of *EZH2* (located on 7q36.1) and thereby also play a role in the pathogenesis of MDS[130,136,137].

1.3.1.1.3 Transcription factors

With about 18% of mutation frequency, transcription factors (TFs) are a minor class of genes affected in patients with myelodysplastic syndromes. Mutated transcription factors are important for lineage-specific gene expression and mutations are commonly observed in hematologic malignancies, both myeloid and lymphoid ones[115].

One example is RUNX1, which regulates hematopoiesis and is mutated in about 6% of MDS patients. Moreover *RUNX1* mutations are associated with more advanced diseases, a decreased overall survival and chromosome 7 abnormalities (-7 / 7q-)[129,138,139]. Another component mutated in MDS and important for hematopoietic development is the zinc finger transcription factor GATA2 which is highly expressed in hematopoietic stem cells (see Figure 1-10). Important roles are the regulation of HSC survival and self-renewal and therefore disruption of this balance can contribute to leukemogenesis[132,140]. *RUNX1* and *GATA2* can be mutated in both ways, somatic or germline, whereas somatic mutations are only present in 1-2% of MDS patients[116]. (Germline) *GATA2* mutations are related with several diseases including familial MDS/AML, MonoMAC syndrome (monocytopenia and mycobacterial infection), Emberger syndrome, DCML (dendritic cell, monocyte, B and NK lymphoid deficiency) and pediatric MDS[140,141]. In addition to that, familial cases of MDS/AML with *GATA2* mutations show a high incidence of monosomy 7 and trisomy 8[142].

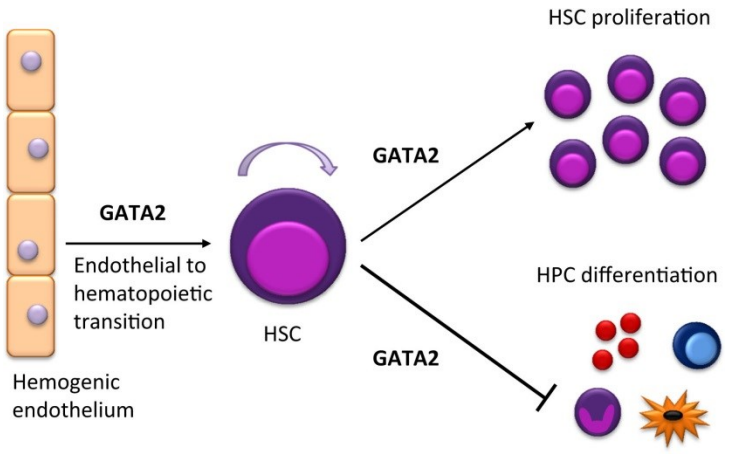


Figure 1-10 - Impact of the transcription factor GATA2

Involvement of GATA2 is essential during transition from hemogenic endothelium to hematopoietic stem cells (HSCs). Later on, GATA2 is important for HSC proliferation and self-renewal as well as for lineage development (adopted from Wlodarski et al.[140]).

1.3.1.2 Cytogenetic aberrations

For prognostication of MDS patients, the karyotype plays an important role and is determined in clinical routine. About 50% of patients were found to have cytogenetic aberrations which are known to be of poor prognosis. The most common ones are isolated deletions of 5q and loss of chromosome 7[129,132,143]. In 10 - 15% of patients with MDS, partial or complete deletions of the long arm of the chromosome 5 are observed and represent a separate MDS subtype according to the WHO classification. Deletion of 5q leads to haploinsufficiency of CSNK1A1 (casein kinase 1 α 1) and results in upregulation of WNT signaling as well as stem cell expansion[132,144,145]. Nevertheless, patients with isolated del(5q) have a good prognosis and can be treated with lenalidomide[129,146].

The underlying mechanism of the immunomodulatory agent lenalidomide includes binding to the $CRL4^{CRBN}$ E3 ubiquitin ligase, altering its substrate affinity and inducing selective degradation of the CSNK1A1 gene product, CK1α. Loss of CK1α results then in activation of p53-mediated apoptosis[115,132,147,148] (see Figure 1-11).

Another frequently observed cytogenetic aberration in MDS is the deletion of 7q and/or monosomy 7 which is associated with poor prognosis. This chromosomal alteration leads to haploinsufficiency of several genes, like *CUX1*, *EZH2* and *MLL3* that may contribute to disease pathogenesis[132,149–152].

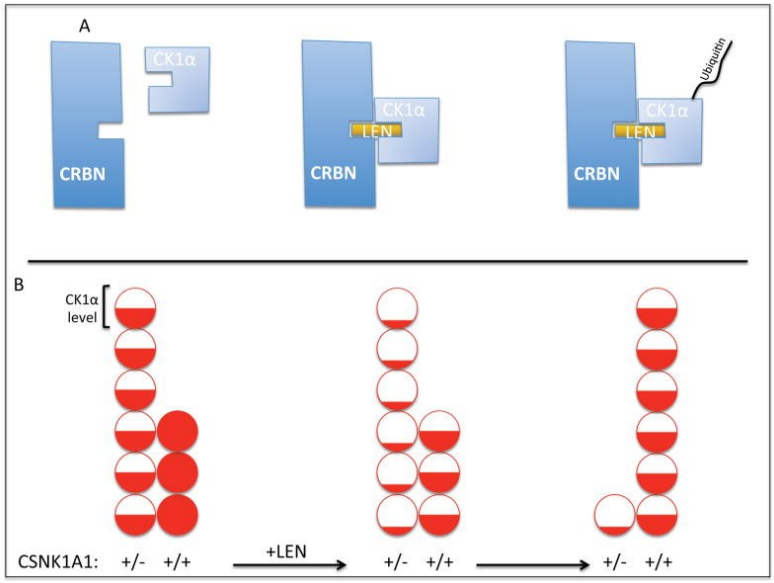


Figure 1-11 - Mechanism of lenalidomide in MDS patients with del5q

(A) CRBN (cereblon), the substrate adaptor of the E3 ubiquitin ligase, shows low affinity for CK1α, whereas presence of lenalidomide increases affinity and therefore catalyzes ubiquitination and degradation of CK1α. (B) Due to the lack of one copy of CSNK1A1 and a resulting lower CK1α level, HSCs have a clonal advantage over wildtype cells at baseline. Lenalidomide treatment selectively depletes CK1α in all HSCs, whereas in 5q- cells levels drop under baseline resulting in apoptosis. On the other side, wildtype cells retain enough CK1α for survival (adopted from Sperling et al.[132]).

1.3.1.3 Role of DNA methylation in MDS

Specific DNA methylation patterns are responsible for sustaining genomic stability and normal gene expression. Several studies showed that MDS and AML are associated with altered DNA methylation pattern suggesting an important role for this epigenetic modification in pathogenesis of MDS[153–157]. Changes in DNA methylation comprise global hypomethylation as well as hypermethylation of CpG islands in promoter regions, such as those of tumor suppressor genes[158].

Application of hypomethylating agents (HMAs), like the nucleoside analogs 5-azacitidine (Vidaza®) and 5-aza-2'-deoxycytidine (Decitabine, Dacogen®), induce clinic response in a distinct subset of MDS patients and delays progression to AML[159,160]. These two azanucleosides show either cytotoxicity due to incorporation into DNA or RNA or hypomethylation of DNA through inhibition of DNA methyltransferases resulting overall in antitumoral effects[161]. Recently published work also suggests immunomodulatory effects of 5-azacitidine[162,163]. In summary, DNA methylation changes somehow play a role in myelodysplastic syndromes, but how this epigenetic mechanism is exactly involved in disease pathogenesis and progression is still unclear.

1.3.1.4 Differences between adult and pediatric MDS

The very heterogeneous group of clonal hematopoietic disorders is observed in adults as well as in children, but several morphological, clinical and cytogenetic differences exist.
First, the clinical incidence in older adults is much higher with 30 – 50 / 10^6 per year than in children with 0.5 – 4 / 10^6 [108,112]. Clinical characteristics in adult MDS patients are isolated anemia and mostly hypercellular bone marrow (BM) compared to bilineage cytopenia and hypocellularity of BM in pediatric MDS. Regarding cytogenetics, MDS with del(5q) are much more frequently found in adults compared to children. Here deletions of chromosome 7 or 7q are the most frequent cytogenetic aberrations[112,164]. On genetic level, mutations in adult MDS patients are mainly observed in *DNMT3A*, *ASXL1*, *TET2* and *SF3B1*, while children often exhibit mutations in the *GATA2*, *SAMD9*[165] or *FANC* (Fanconi anemia) members[166]. Differences between MDS in adults and children are also made in clinical treatment options. In pediatric MDS the treatment of choice represents hematopoietic stem cell transplantation (HSCT), the only curative therapy. Contrary, adult MDS patients are treated due to the severity, low / intermediate risk or high risk, either supportive, immunomodulatory with lenalidomide or with hypomethylating agents, intensive chemotherapy and HSCT[112,164].

1.3.2 Clonal evolution during disease progression

The emergence of next-generation sequencing greatly expanded our knowledge about the pathogenesis of myelodysplastic syndromes and other malignant disorders, whereas mechanisms during disease progression are not well studied at all.
MDS progression is a dynamic event characterized by increasing malignant potential and clonal evolution, whereas the origin of this disease lies within acquired mutations in hematopoietic stem cells (HSCs)[167]. Here one has to differentiate between MDS and clonal hematopoiesis with indeterminate potential (CHIP) which can be distinguished due to the presence or absence of hematopoietic dysplasia in bone marrow, respectively[168,169].
The progression from MDS to secondary AML (sAML) was shown to be associated with the presence of mutations as well as the clonal architecture of these mutations. Founding clones are predominantly associated with mutations of epigenetic modifiers (e.g. ASXL1, TET2) and RNA splicing factors (e.g. SF3B1, SRSF2). Daughter clones tend to acquire mutations in signaling cascade factors, transcription factors or show cytogenetic lesions[170]. In general, it could be shown that sAML clones derive from a MDS founding clone and both entities show the same clonality despite different bone marrow blast counts (Figure 1-12). Each new clone during tumor progression carries all preexisting pathogenic and nonpathogenic mutations[171]. Furthermore, disease progression and thus complex clonal architecture was correlated with phenotype progression into worse WHO categories[170].
Epigenetic changes during disease progression of myelodysplastic syndromes have been scarcely explored so far. One study examined the epigenetic heterogeneity of 138 AML patients in comparison to their genetic landscape during disease progression. They showed that genetic and epigenetic

patterns arise independent during leukemic progression, but both of them playing a unique significant function[172].

In summary, the sequential acquisition of mutations and cytogenetic aberrations result in clonal evolution of MDS and is probably not associated with underlying epigenetic patterns.

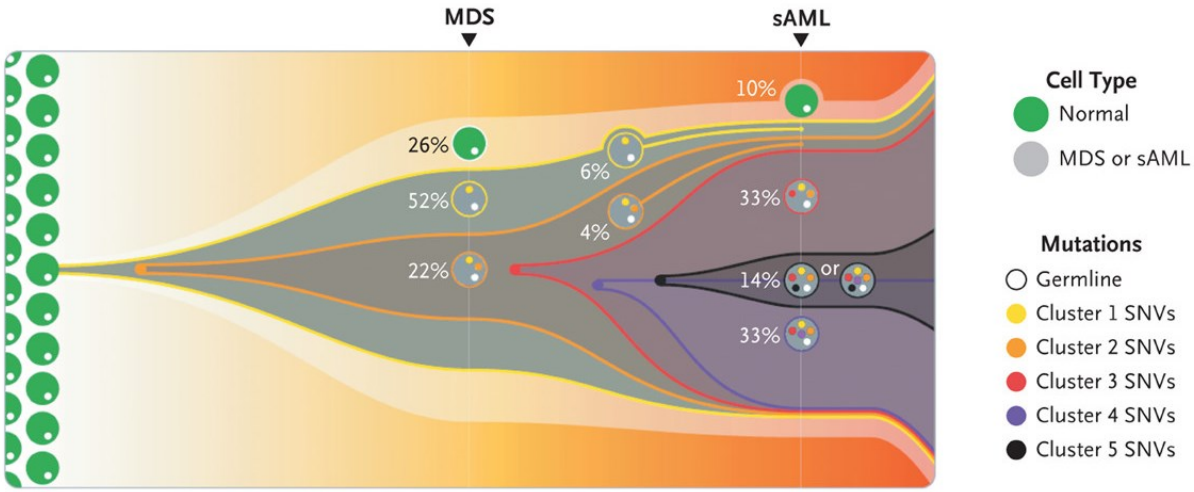


Figure 1-12 - Clonal evolution from MDS to sAML

A model of the clonal evolution from myelodysplastic syndromes (MDS) to secondary acute myeloid leukemia (sAML) is represented by this fish plot. The first clone exhibiting mutations is indicated with yellow and is present in 52 % of the bone marrow cells. This clone is characterized by Cluster 1 somatic single-nucleotide variants (SNVs). Cells indicated in orange originated from clone 1 and are labeled as clone 2 with specific Cluster 2 SNVs. Clone 2 evolved three subsequent subclones (red, purple and black) through serial acquisition of SNVs and is the dominant clone in the sAML sample (adopted from Walter et al.[171]).

2 Research Objectives

DNA methylation and in general all epigenetic mechanisms are important for sustaining genome stability and gene expression. It was shown that alterations of DNA methylation patterns occur in several diseases, like acute myeloid leukemia (AML) or myelodysplastic syndromes (MDS)[154,157], and may be involved in pathogenesis and / or progression. The observation that inhibitors of DNA methyltransferases can delay the progression of MDS points to an additional role of epigenetic mechanisms in disease pathology.

The main goal of this thesis was the integrated analysis of epigenetic and genetic changes during MDS development to identify potential epigenetic target genes that are involved in the progression of MDS to AML. For this purpose, two different approaches were used including methyl-CpG-immunoprecipitation (MCIp) and targeted bisulfite sequencing. The MCIp protocol allows a global DNA methylation analysis and should reveal the inter-individual variability between different patients. To analyze DNA methylation alterations in regions important for myeloid differentiation, a targeted bisulfite approach was used including active regulatory regions as well as promoter regions observed to be involved in development of myeloid cells.

By integrating both, epigenetic and genetic alterations, the analysis of consecutive patient samples may reveal the order of appearance for both types of aberrations. Summarizing all aspects, the analysis of global and specific DNA methylation patterns as well as genetic changes in different patient groups and during disease progression should provide insights into the pathogenesis of MDS and may help to improve clinical treatment.

3 Materials and Equipment

3.1 Equipment

Autoclave	Walter, Geislingen, Germany
Bioanalyzer 2100	Agilent Technologies, Böblingen, Germany
BioPhotometer	Eppendorf, Hamburg, Germany
Caliper LapChip XT	Perkin Elmer, Waltham, USA
Centrifuges	Sigma, Osterode; Heraeus, Hanau; Germany
ChemiDoc XRS + system	Biorad, Munich, Germany
Covaris S220	Covaris, Woburn, USA
Eppendorf centrifuge 5804 R	Eppendorf, Hamburg, Germany
Eppendorf Mastercycler Nexus X2	Eppendorf, Hamburg, Germany
Electrophoresis equipment	Biometra, Göttingen, BioRad, Munich, Germany
Fast blot machine	Biometra, Göttingen, Germany
Heat block	Stuart Scientific, Staffordshire, UK
HiSeq 1000/2000	Illumina, San Diego, USA
HiSeq 3000/4000	Illumina, San Diego, USA
Incubators	Heraeus, Hanau, Germany
Intelli Mixer RM-2L	Elmi-Tech, Riga, Latvia
Lightcycler LC480	Roche, Mannheim, Germany
Luminometer	Sirius Berthold, Oakville, Canada
Magnetic particle concentrator	Thermo Fisher Scientific, Waltham, USA
MassARRAY compact system	Sequenom, San Diego, USA
MassARRAY MATRIX liquid handler	Sequenom, San Diego, USA
MassARRAY Phusio chip module	Sequenom, San Diego, USA
Megafuge 3.0R	Heraeus, Hanau, Germany
Microscopes	Leitz, Heidelberg, Germany
Mr. Frosty™ Freezing Container	Thermo Fisher Scientific, Waltham, USA
Multifuge 3S-R	Heraeus, Hanau, Germany
Multipipette	Eppendorf, Hamburg, Germany
NanoDrop 1000	PeqLab, Erlangen, Germany
PCR-Thermocycler PTC-200	MJ-Research/Biometra, Oldendorf, Germany
pH meter	Knick, Berlin, Germany

Picofuge	Heraeus, Hanau, Germany
Pipetboy	Integra Biosciences, Fernwald, Germany
Pipettes	Eppendorf, Hamburg, Germany
Pipettes	Gilson, Middleton, USA
Power supplies	Biometra, Göttingen, Germany
QIAvac 24 Plus	Qiagen, Hilden, Germany
Qubit 2.0 fluorometer	Thermo Fisher Scientific, Waltham, USA
Realplex Mastercycler epGradientS	Eppendorf, Hamburg, Germany
Heat sealer	Eppendorf, Hamburg, Germany
Sonifier 250	Emerson, St. Louis, USA
Sorvall RC 6 plus	Thermo Fisher Scientific, Waltham, USA
Speed Vac	Christ, Osterode, Germany
Laminar air flow cabinet (Lamin Air: HA 2472)	Heraeus, Osterode, Germany
TapeStation 2200	Agilent Technologies, Böblingen, Germany
Thermomixer	Eppendorf, Hamburg, Germany
Typhoon 9200	Molecular Dynamics, Krefeld, Germany
Vortex-Genie	Scientific Industries Ink., Bohemia, USA
Waterbath	Julabo, Seelstadt, Germany
Water purification system	Millipore, Eschborn, Germany

3.2 Consumables

384-well PCR plates	Thermo Fisher Scientific, Hudson, USA
Adhesive PCR sealing film	Thermo Fisher Scientific, Hudson, USA
Agencourt AMPure XP magnetic beads	Beckman Coulter Genomics, Krefeld, Germany
Cell culture dishes	Greiner, Frickenhausen, Germany
Cell culture dishes	Nunc/Thermo Fisher Scientific, Hudson, USA
Cell culture flasks	Costar, Cambridge, USA
Cryo tubes	Corning, Corning, USA
Heat sealing Film	Eppendorf, Hamburg, Germany
Luminometer vials	Falcon, Heidelberg, Germany
Micro test tubes (0.5, 1.5, 2 ml)	Eppendorf, Hamburg, Germany
Multiwell cell culture plates	Falcon, Heidelberg, Germany
nProteinA Sepharose 4 FastFlow	GE Healthcare, Munich, Germany
PCR plate Twin.tec 96 well	Eppendorf, Hamburg, Germany
PCR plate 384 well (LightCycler)	Roche, Basel, Switzerland
PCR plate 384 well (MassARRAY)	Thermo Fisher Scientific, Hudson, USA
rProteinA Sepharose 4 FastFlow	GE Healthcare, Munich, Germany
Sepharose CL-4 beads	Sigma-Aldrich, Munich, Germany
Sterile combitips for Eppendorf multipette	Eppendorf, Hamburg, Germany
Sterile plastic pipettes	Costar, Cambridge, USA
Syringes and needles	Becton Dickinson, Heidelberg, Germany
Tubes (15 ml, 50 ml, 220 ml)	Falcon, Heidelberg, Germany

3.3 Chemicals

All chemicals were purchased from Sigma-Aldrich (Deisendorf, Germany), Merck Millipore (Darmstadt, Germany) or Carl Roth (Karlsruhe, Germany).

3.4 Enzymes, kits and products for molecular biology

Agilent DNA 1000 Kit	Agilent Technologies, Santa Clara, USA
Agilent High Sensitivity DNA Kit	Agilent Technologies, Santa Clara, USA
Agilent RNA 6000 Nano Kit	Agilent Technologies, Santa Clara, USA
Alkaline Phosphatase	Roche, Basel, Switzerland
Amersham ECL Prime Western Blotting Detection Reagent	GE Healthcare, Freiburg, Germany
Amersham Hyperfilm ECL	GE Healthcare, Freiburg, Germany
Beetle-Juice BIG KIT	PJK, Kleinblittersdorf, Germany
CpG Methyltransferase M.SssI	NEB, Frankfurt, Germany
D1000 Screen Tape and reagents	Agilent Technologies, Santa Clara, USA
DNA ladder 1 kb plus	Invitrogen, Karlsruhe, Germany
DNA 50 bp ladder	NEB, Frankfurt, Germany
DNeasy Blood & Tissue Kit	Qiagen, Hilden, Germany
dNTPs	GE Healthcare, Buckinghamshire, UK
Effectene® transfection reagent	Qiagen, Hilden, Germany
EpiMark® Methylated DNA Enrichment Kit	NEB, Frankfurt, Germany
Ethidium bromide	Sigma-Aldrich, Munich, Germany
EZ DNA Methylation-Lightning Kit	Zymo Research, Orange, USA
Fermentas DNA loading dye (6x)	Thermo Fisher Scientific, Hudson, USA
Gibson Assembly Master Mix	NEB, Frankfurt, Germany
Glycogen	Ambion/Life Technologies, Carlsbad, USA
High Sensitivity D1000 Screen Tape and reagents	Agilent Technologies, Santa Clara, USA
Kapa Library Preparation Kit	Kapa Biosystems, Wilmington, USA
Klenow exo- (3'-5' exo minus)	Enzymatics, Beverly, USA
Klenow fragment	Enzymatics, Beverly, USA
λ DNA/Hind III Fragments	Invitrogen, Karlsruhe, Germany
LapChip XT DNA 300/750 Kit and Chips	Perkin Elmer, Waltham, USA
M-MLV Reverse Transcriptase	Promega, Madison, USA
MinElute Gel Extraction Kit	Qiagen, Hilden, Germany
MinElute PCR Purification Kit	Qiagen, Hilden, Germany

Monarch® PCR & DNA Cleanup Kit	NEB, Frankfurt, Germany
NEBNext Multiplex Oligos for Illumina	NEB, Frankfurt, Germany
NEBNext Ultra II DNA Library Prep Kit for Illumina	NEB, Frankfurt, Germany
NEXTflex DNA Barcodes	Bioo Scientific, Austin, USA
Nuclease-free water	Gibco/Life Technologies, Carlsbad, USA
NucleoSpin Plasmid Quick Pure Kit	Macherey-Nagel, Duren, Germany
Pancoll	Pan Biotech, Aidenbach, Germany
Phusion High-Fidelity DNA Polymerase	NEB, Frankfurt, Germany
Protease Inhibitor Cocktail Tablets, EDTA-free	Roche, Basel, Switzerland
Proteinase K	Roche, Basel, Switzerland
QIAGEN Plasmid Plus Midi Kit	Qiagen, Hilden, Germany
QIAquick Gel Extraction Kit	Qiagen, Hilden, Germany
QIAquick PCR Purification Kit	Qiagen, Hilden, Germany
QuantiFast SYBR green	Qiagen, Hilden, Germany
Qubit DNA HS Kit	Life Technologies, Carlsbad, USA
Random Decamers	Ambion/Life Technologies, Carlsbad, USA
Renilla-Juice BIG KIT	PJK, Kleinblittersdorf, Germany
Restriction endonucleases	NEB, Frankfurt, Germany; Roche, Basel, Switzerland
RNase A	Sequenom, San Diego, USA
S-adenosyl-methionine (SAM)	NEB, Frankfurt, Germany
ScriptSeq™ Complete Kit	Epicenter (an Illumina company), Madison, USA
SeqCap Adapter Kit A	Roche, Basel, Switzerland
SeqCap Epi Accessory Kit	Roche, Basel, Switzerland
SeqCap Hybridization and Wash Kit	Roche, Basel, Switzerland
SeqCap Pure Capture Bead Kit	Roche, Basel, Switzerland
(S)-JQ1	Dana-Farber Cancer Institute, Boston, USA
TaqDNA Polymerase	Roche, Basel, Switzerland
T4 DNA Polymerase	Enzymatics, Beverly, USA
T4 DNA Ligase	Promega, Madison, USA
T4 DNA Ligase (Rapid)	Enzymatics, Beverly, USA
T4 Polynucleotide Kinase	Enzymatics, Beverly, USA

3.5 Antibodies

Table 3-1 - List of antibodies

Antigene	Application	Origin	Company	Order#
BRD4	ChIP	rabbit	Bethyl Laboratories	A301-985A100-3
BRD4	ChIP	rabbit	Bethyl Laboratories	
Goat anti-IgG F(c), HRP conjugated	Western Blot	goat	Rockland	
H3K27ac	ChIP	rabbit	abcam	ab4729
H3K27me3	ChIP	rabbit	diagenode	C15410069
IgG	ChIP	rabbit	Millipore	12-370

3.6 Antibiotics

Ampicillin	Ratiopharm, Ulm, Germany
Hygromycin	Clontech, Mountain View, USA
Zeocin	Invitrogen, Karlsruhe, Germany

3.7 Cell lines

Drosophila Schneider S2 cells	macrophage-like insect cell line (provided by Prof. Werner Falk)
THP-1	human acute monocytic leukemia (DSMZ no: ACC 16)
U937	human histiocytic lymphoma (DSMZ no: ACC 5)

3.8 *E. coli* strains

DH10B	Invitrogen / Life Technologies, Darmstadt, Germany
One Shot PIR1	Invitrogen / Life Technologies, Darmstadt, Germany

3.9 Plasmids

pCoHygro	Invitrogen, Karlsruhe, Germany
phRL-TK	Promega, Madison, USA
pMT/BiP/V5-His B	Invitrogen, Karlsruhe, Germany
pCpGL-basic (#861)	AG Rehli, Regensburg, Germany
pCpGL-CMV/T.EF1A (#1341)	AG Rehli, Regensburg, Germany
pCpGL-CMV1/T.A1FE (#1366)	AG Rehli, Regensburg, Germany

3.10 Oligonucleotides

3.10.1 PCR primers

BiooPrimer1	5'-AATGATACGGCGACCACCGAGATCTACAC-3'
Bioo Primer 2	5'-CAAGCAGAAGACGGCATACGAGAT-3'
KDM2B_NheI_for	5'-ATCGCTAGCAACCGGACAACGGCAGGAGC-3'
KDM2B_XbaI_rev	5'-ACTTTCTAGACTCCTTGAGCAGGGAGCCGG-3'
SIX5_PstI_1_for	5'- CTTCCTGCAGTTTCACAACAAAGGCAGAAGACGG-3'
SIX5_BglII_1_rev	5'- ATGCAGATCTGTCGCTCACGCAGGTCAG-3'
SIX5_PstI_2_for	5'- CCTTCCTGCAGGGGTGGGAGGAGAAGGGTTTG-3'
SIX5_BglII_2_rev	5'- ATGCAGATCTTGTCGCTCACGCAGGTCA-3'
SIX5_PstI_3_for	5'- CCTTCCTGCAGTTGGGTTACAGGGAAACCGGAG-3'
SIX5_BglII_3_rev	5'- ACTCAGATCTGCTCACGCAGGTCAGCAA-3'
Ventx_PstI_1_for	5'- CCTTCCTGCAGCCTGACTCTCCCAGCCTGAA-3'
Ventx_BglII_1_rev	5'- ATGCAGATCTCCAGTCCACGGAGCCAAAG-3'
Ventx_PstI_2_for	5'- CCTTCCTGCAGGGAGAGGAGAGGTCGCCC-3'
Ventx_BglII_2_rev	5'- ATGCAGATCTCAGGGAGAAGTCGGCAG-3'
Ventx_PstI_3_for	5'- CCTTCCTGCAGCTGAAGCCTGCTCGCCCT-3'
Ventx_BglII_3_rev	5'-ATGCAGATCTGGAGAGCCAGTCCACGGAG-3'
Ventx_PstI_4_for	5'- CCTTCCTGCAGGCCCTGACTCTCCCAGCC-3'
Ventx_BglII_4_rev	5'- ATGCAGATCTCGCAGGCAGATTTGAGGACCC-3'
Ventx_PstI_5_for	5'- CCTTCCTGCAGCCAACCTAAGTCCGCCCTGA-3'
Ventx_BglII_5_rev	5'- ATGCAGATCTGCAGGCAGATTTGAGGACCC-3'

3.10.2 qPCR primers

bAct_for	5’-GCAACTTTCGGAACGGCGCA-3’
bAct_rev	5’-GCGAAGCCGGTGAGTGAGC-3’
BCL3_1_for	5’-ACTTCTCTGTGCCTGTTTCCTCCT-3’
BCL3_1_rev	5’-GCAACTAAGAATTGGCCCACTTAATCCT-3’
BCL3_2_for	5’-CTTCTCTGTGCCTGTTTCCTCCTC-3’
BCL3_2_rev	5’-AATTGGCCCACTTAATCCTCTTAACACC-3’
BCL6_1 for	5’-TTCCCTTTGCCACTCCACTCTTAGC-3’
BCL6_1 rev	5’-TCTGGCTGGTTCTACTACTGCCTTAGAG-3’
BHLHE40_1_for	5’-GGTGAGGAGCAAAGAAATTGACATACGA-3’
BHLHE40_1_rev	5’-ACAATTCCTACGTGGCATAGATCTTTCC-3’
BHLHE40_2_for	5’-GGGCTCTAAATTCTTGCACAGGAAGG-3’
BHLHE40_2_rev	5’-CCTATTGGGACTCAGCATCTTGGTGG-3’
ChIP-GAPDH_1 for	5’-AGGCTGGATGGAATGAAAGGCAC-3’
ChIP-GAPDH_1 rev	5’-CTCCCACAAAGGCACTCCTG-3’
empty6.2_for	5’-GAAACCCTCACCCAGGAGATACAC-3’
empty6.2_rev	5’-TGCAGTGGGACTTTATTCCATAGAAGAG-3’
GAPDH_up75 for	5’-GCTTCTCACAGGACTTCCCTTGTCTC-3’
GAPDH_up75 rev	5’-ACTGCCTATGGATCTGGAACTCCC-3’
GSX2_2F	5’-GAAGGTCTATCTAATCCCTGCTGCGT-3’
GSX2_2R	5’-CATTCCAGGGCAATCCTACAAACTCCA-3’
ID2_1 for	5’-CCTGTCAGTGAGATGATGCGTATACC-3’
ID2_1 rev	5’-TGGCTCTGGCGAGGGTACAAA-3’
ID2_2 for	5’-AGGAAACACCTGTAATGGCTGCTG-3’
ID2_2 rev	5’-CTACCGGATCACCCTAACAACCCT-3’
KDM2B_NheI_for	5’-GATCGCTAGCAACCGGACAACGGCAGGAGC-3’
KDM2B_XbaI_rev	5’-ACTTTCTAGACTCCTTGAGCAGGGAGCCGG-3’
MBNL1_1 for	5’-ATTAACGCTGGGAAGAGGCACCAT-3’
MBNL1_1 rev	5’-TCCAGAGGCACTAATAGCCGACT-3’
MBNL1_2 for	5’-ATGGCAAGGTTGTGATACTGTAGGACAT-3’
MBNL1_2 rev	5’-ACGTCAGGAAAGGCACTTCTTAGTACCA-3’
MBNL1_3_for	5’-GAGCTCGACGAGTCCGCC-3’
MBNL1_3_rev	5’-GGTCTGGCTTCCGCTGCTG-3’
MBNL1_4 for	5’-CGCGTGCATTAGGAGCTCGAC-3’
MBNL1_4 rev	5’-AGGTCTGGCTTCCGCTGCT-3’
MBNL1_5 for	5’-CATTAGGAGCTCGACGAGTCCGC-3’
MBNL1_5 rev	5’-AGGTCTGGCTTCCGCTGCT-3’
MBNL1_neg for	5’-GCACCCAGAACACAAGCCCA-3’
MBNL1_neg rev	5’-CCATCAAGCCTCTAGCTGCCTTT-3’
Myc-1 for	5’-GGACCCGCTTCTCTGAAAGGCT-3’

Myc-1 rev	5’-AGGCAAGTGGACTTCGGTGCTTA-3’
Myc-2 for	5’-TTGCGGTCACACCCTTCTCCC-3’
Myc-2 rev	5’-GTTCACCATGTCTCCTCCCAGCAG-3’
Myc-3 for	5’-CAGCAAACCTCCTCACAGCCC-3’
Myc-3 rev	5’-ACACTGTCCAACTTGACCCTCTTG-3’
NCL_1 for	5’-CTGCCCAGTAATCGCCTGTGGAA-3’
NCL_1 rev	5’-GCCGCGAGCTTTGGTTGGT-3’
PTPN6 for	5’-TCCGCCTTCCTTGTGACTTGAG-3’
PTPN6 rev	5’-ACCAGAGGCAAAGAGAAACGCAG-3’
SNRPN_for	5’-TACATCAGGGTGATTGCAGTTCC-3’
SNRPN_rev	5’-TACCGATCACTTCACGTACCTTCG-3’
SRFP2_2F	5’-GGAGGGCGAAGTTCTTTCATATGTAAGG-3’
SRFP2_2R	5’-TCTGAGCCTGTGAATGACTCTTAAGTGG-3’
Suv420h1_for	5’-CTCACCTCGCCTCGCGCA-3’
Suv420h1_rev	5’-CCTGGGACGCGGAGTCCT-3’

3.10.3 RT-qPCR primers

ACTB_1 for	5’-CGAGAAGATGACCCAGATCATGTTTGAG-3’
ACTB_1 rev	5’-CAGAGGCGTACAGGGATAGCACAG-3’
GAPDH for	5’-CCACATCGCTCAGACACCAT-3’
GAPDH rev	5’-GCCACCAATATCCACTTTACCAGAGT-3’
HPRT1_1 for	5’-AATTATGGACAGGACTGAACGTCTTGC-3’
HPRT1_1 rev	5’-GCTTTGATGTAATCCAGCAGGTCAGC-3’
Myc_RTPCR_1 for	5’-GTCCTCGGATTCTCTGCTCTCCT-3’
Myc_RTPCR_1 rev	5’-CTCATCTTCTTGTTCCTCCTCAGAGTCG-3’
Myc_RTPCR_2 for	5’-CGACTCTGAGGAGGAACAAGAAGATGAG-3’
Myc_RTPCR_2 rev	5’-GGCTGTGAGGAGGTTTGCTGTG-3’

3.10.4 Sequencing primers

MT forward	5’-CATCTCAGTGCAACTAAA-3’
pcDNA3.1/BGH rev	5’-TAGAAGGCACAGTCGAGG-3’
pCG sense	5’-TAAATCTCTTTGTTCAGCTCTCTG-3’
pMT_rev	5’-GCATTCTAGTTGTGGTTTGTCC-3’

3.11 gblocks® gene fragments

CREB3L

5’-CAAACAGCAGATTAAAAGGAATTCCTGCAGGAAGGGTCTCCGTTCCCGTTCCACCCCTCGGCACCCGTCC
TCGCGGCGCGCCTGGGCCCCTAGAAGGACCCGACTACACATCACTGGGCAGGAGCCGGGGAGAGGGTTCA
GCGCAGGGTCTCCAGGAGCGACATGTGTTTGGAGCTAAGCGCCCCTCCTGGGGGCTCAATCTTTGAATACA
CTGGCCTCCACCTTCTAGGGGGAAGGGGCCATCGAGGTATCGGGTCCGTCCGATCCTCGGATCTGAGCTCA
AGAGACACCAATCCCGCCCTCCACCCACGTCTTCTAGCCGTCCCCAACCCACCCCAGGTTCCAGGACCAGA
CTGGGCGCAAAGTGGCAGCGCCCTTTCCTGCGCCTCCCTTGGGGCTTCGGTGGCTCATAGGCTGGATCTCC
GCTGGGGGGCGCACCGGGGAACGTTCAGAGGGCCTGAGACCCGACCCCTCGGGGAAGCCAGGCCCAGAA
ATTTCAGGGTGCTACCCTGGCAGCCCCAGGGAATCAGGCCCAGAGACCCCCCACCCCAGGGAGGGACCTG
AGGCTGGGGGCTGGGAAGAGGCGGTCAGGGCAACATAGGGTGATGGCAGATCTTCATAATAAAATATCTTT
ATTTTC-3‘

FOXO3_1

5‘-CAAACAGCAGATTAAAAGGAATTCCTGCAGCAGGGGACGCCGGCGAGGGAAGGGCGAACGGACAGGAG
TACATTTGCTGGATTCTCCGGACAGCACCGAGGAGGTAGGTCCGCAGCCAACTTTGGAGTAGAGTTTACCGT
AGTGGGGGTCTGGTGCGGGCGCTGTGGCTGGAGTGAGAGGTGTGGGGGTGTTTGGGACGCAGTGTCTGGA
GGAGGGGTTGGATGTGTGGTTTCCAGTCTGTCGCAGGAGCATGTTGCGTCGTCACTAGCTGAATGAGAACC
TTCGGGTCCAAGTTTCAGCTTGTGGGTGTTAACACCTACAGGCACATCGATCCGATTAGAAAAAGCAGTGGT
TGCAAACCTTTTCCTGGACGGCTTCCTTTCCTTGCCTATATTGATACCTTTTCTTCTCGGAGATGTCGCTCCA
GTAAACCTGCTTCTGACTAGCTGCTTCTGAAATGTTCTGGGGCCTCGAACCGGCCGGTCTGGCCACCTCAAT
CCAGACTGGCTGCACCCGCTGCTCCAGATCTTCATAATAAAATATCTTTATTTTC-3‘

FOXO3_2

5’-CAAACAGCAGATTAAAAGGAATTCCTGCAGAGGGGATCCCGCTGCTCCCGCCGAGGCCTGGATTCAGGC
ACTGTTTGAGGGAAGGGCTCTGGTGTGGGGGACGCGGGGCCTGCGCTTCATTCACTCTGGCGTTTGGGAG
GAGGTTTGGGGGCTGGGCCGGGCCGTAATGGCTTCTACCTGTAAATGAACCCGAAGAGAGACTCTCTACCG
TTCCTTATCATCCTCTTGGGGCACGGGACACGTGGTAGGGAGAGAGGTGTGGAAAGGCCTATGATGTAATG
ACTTCCTGCAGTTGGGTATCCAAAGAGATTAACTCCTCTGCCCCTGCTGGGCTGCTGATATCCAGCTGCTTT
CCTCTTCCTAGGCACCTCCCTGGCCCCTTCCCCCAAGACAGTGTGTGTCAGTACCAGGCCAAAGAGTCCTGT
GCGATTGGAGATATACATCTCTTATCAGCTGCCCATATTAATCCAGACTGGCTGCACCCGCTGCTCCAGATCT
TCATAATAAAATATCTTTATTTTC-3’

GADD45B

5‘-CAAACAGCAGATTAAAAGGAATTCCTGCAGGCCACACAGTGGGAAGGCCAAGGACCTGCTCCCCAGGAC
GCTGGGCCAGGGAGACCGCAGGCACACGGGACAGCCACACCAGAGCTGCATCTGGAGAAGGGGTGGGGG
GTTTCCTCTCGCCGCTCCTCGCTATTCTTTCCCTACATTGTTCCGTTCCGTGCTCCCAGTTCCCACCCCACTC
CCTCCACCCGGACGCCGGCGGTTCCCAGGCCCCGGAGCCGGGGTGCATTGGGGGTGGGGGTGGGGTAGA
GTGAGGGGCCTCGTGAAGAGCGCTGCGGGCCAAGACCCTCTTGCGCATGTGACGAGAGCGCACGAGTGTG
GGAGCTTCGGCCTGCGCTCACAGGCGTGTGCGTGGGGAAGTGTTTGGGCCGCCGTGTACACAGCAGGCAC
GCCATAAACGCTTCTTGGATGGCTCCATGCGAGGGAAGGGAGGGAGGGCAGAGTTCAGGCCCTACAGGGA
GGAAAAGGGAGGGGCGGAGGCCTGGCTTTAGGTTTGAATCTCCGCGGCTTACTAGCTCCAGCAGCTCAGCA

AAAGCGTGACTCAGTTTCCCTCGTCCATAAAAGGGAAAGCCGAATCGCCCTCCCACGTGTCGAGGTGAAGAT
CTTCATAATAAAATATCTTTATTTTC-3‘

ZFPM1

5’-CAAACAGCAGATTAAAAGGAATTCCTGCAGAGGGGATCCCCGGGGATGCAGGACGCATCCACCGTCATG
CGGCTGGGAGGTGGTGGAGCTGGGTTTAGGCCTGGAGTTCATAGACACTGGTTTGATCGCCACCCCAAGGG
TCCCTGTCCATCCCTGCGGTGTTGAGAGGGGTTAGATCATCCTCCTGTTCTCCAGACGGAGAAGGCCCTGG
GGGATGGAGCCTGGCGCCGGGAAGGGCCGGGCCCCTGAGGCCGGCCGTGGTTTGGGGCTGGGGCACAC
CCCCGAGCTTCCTGTGCTGTGTGCATCCTCAGGACCAGGCGGTGCCCTATCGGCTCCCGCAGATATGAATT
TCTGTGCAGTGGGCGGTGATGGGCTGATACGCGGCCGCTGGCCAACCGCTGCAGTGGCTACGGCCTATCA
GACCCACACAGGCTGAGCTATCACCCGCGGGGAAGCCAAGGAGGCTCTGTGGGCTGTTATCTCTCGGGGC
GGCCCGTGCGTCGATGGCCAGCAGGGTCAGTGGCCCCAAGGGCAGTGCCCAGCCAGGTGACCGGTCCAG
ATCTTCATAATAAAATATCTTTATTTTC-3’

1341_RUNX1

5‘-CAAACAGCAGATTAAAAGGAATTCCTGCAGAGGGAAACCTTTTCGCCTGGTCTCCAATGCATTTCCCCGA
GATCCCACCCAGGGCTCCTGGGGCCACCCCCACGTGCATCCCCCGGAACCCCCGAGATGCGGGAGGGAG
CACGAGGGTGTGGCGGCTCCAAAAGTAGGCTTTTGACTCCAGGGGAAATAGCAGACTCGGGTGATTTGCCC
CTCGGAAAGGTCCAGGGAGCTCCTCTGGGTCTCGGGCCGCTTGCCTAAAACCCTAAACCCCGCGACGGGG
GCTGCGAGTCGGACTCGGGCTGCGGTCTCCCAGGAGGGAGTCAAGTTCCTTTATCGAGTAAGGAAAGTTGG
TCCCAGCCTTGCATGCACCGAGTTTAGCCGTCAGAGGCAGCGTCGTGGGAGCTGCTCAGCTAGGAGTTTCA
ACCGATAAACCCCGAGTTTGAAGCCCGACAAAAAGCTGATAGCAATCACAGCTTTTGCTCCTTGACTCGATG
GGATCGCGGGACATTTGGGTTTCCCCGGAGCGGCGCAGGCTGTTAACTGCGCAGCGCGGTGCCCTCTTGA
AAAGAAGAAACAGACCAACCTCTGCCCTTCCTTACTGAGGATCTAAAATGAATGGAAAGAGGCAGGGGCTCC
GGGGAAAGGGAACCCCTTAGTCGGCCGGGAGATCTTCATAATAAAATATCTTTATTTTC-3‘

861_VENTX

5‘-CAAACAGCAGATTAAAAGGAATTCCTGCAGAGGGAAACCTTTTCGCCTGGTCTCCAATGCATTTCCCCGA
GATCCCACCCAGGGCTCCTGGGGCCACCCCCACGTGCATCCCCCGGAACCCCCGAGATGCGGGAGGGAG
CACGAGGGTGTGGCGGCTCCAAAAGTAGGCTTTTGACTCCAGGGGAAATAGCAGACTCGGGTGATTTGCCC
CTCGGAAAGGTCCAGGGAGCTCCTCTGGGTCTCGGGCCGCTTGCCTAAAACCCTAAACCCCGCGACGGGG
GCTGCGAGTCGGACTCGGGCTGCGGTCTCCCAGGAGGGAGTCAAGTTCCTTTATCGAGTAAGGAAAGTTGG
TCCCAGCCTTGCATGCACCGAGTTTAGCCGTCAGAGGCAGCGTCGTGGGAGCTGCTCAGCTAGGAGTTTCA
ACCGATAAACCCCGAGTTTGAAGCCCGACAAAAAGCTGATAGCAATCACAGCTTTTGCTCCTTGACTCGATG
GGATCGCGGGACATTTGGGTTTCCCCGGAGCGGCGCAGGCTGTTAACTGCGCAGCGCGGTGCCCTCTTGA
AAAGAAGAAACAGACCAACCTCTGCCCTTCCTTACTGAGGATCTAAAATGAATGGAAAGAGGCAGGGGCTCC
GGGGAAAGGGAACCCCTTAGTCGGCCGGGAGATCTTAAGCTTAGTCCATGGAGGATGCC-3‘

3.12 Databases and software

Bioconductor	https://bioconductor.org/
BioEdit	http://www.mbio.ncsu.edu/bioedit/bioedit.html
BioLayout	https://kajeka.com/biolayout-express-upgrade/
Blueprint Epigenome	http://www.blueprint-epigenome.eu/
Bowtie2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
BSMAP	https://code.google.com/archive/p/bsmap/
EGA	https://www.ebi.ac.uk/ega/home
Enrichr	http://amp.pharm.mssm.edu/Enrichr/
EpiTYPER 1.2	
Fastqc	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
GeneRunner	http://generunner.net/
GENTle	http://gentle.magnusmanske.de/
G.R.E.A.T	http://bejerano.stanford.edu/great/public/html/
Homer	http://homer.salk.edu/homer/
IGV	https://www.broadinstitute.org/igv/
ImageLab v4.0	
Integrative Genomics Viewer	http://software.broadinstitute.org/software/igv/
Metascape	http://metascape.org/gp/index.html#/main/step1
metilene	https://www.bioinf.uni-leipzig.de/Software/metilene/
PerlPrimer	http://perlprimer.sourceforge.net/
Picard	https://broadinstitute.github.io/picard/
R	https://bioconductor.org/biocLite.R
UCSC Genome Browser	http://genome.ucsc.edu/

4 Methods

4.1 General cell and bacteria culture methods

4.1.1 Cell line culture

4.1.1.1 Assessing cell number and vitality

The total number of cells and their vitality can be determined by Trypan blue exclusion. The cell suspension was diluted suitably with Trypan blue solution and counted in a Neubauer hemocytometer. Dead cells appear blue since the blue stain is able to enter the cytoplasm. The concentration of viable cells was then calculated using the following equation:

$$\text{Number of viable cells/ml} = \text{N} \cdot \text{D} \cdot 10^4$$

N: average of unstained cells per corner square (1 mm containing 16 sub-squares)
D: dilution factor

Required solutions and materials:
Trypan blue solution: 0.2 % (w/v) Trypan blue in 0.9 % NaCl solution
Neubauer hemocytometer

4.1.1.2 Culture conditions and passaging

Cells were cultured in RPMI 1640 (w/o L-glutamine; Gibco) routinely supplemented with 10% inactivated FCS, L-glutamine (2 mM), sodium pyruvate (1 mM), antibiotics (50 U/ml penicillin and 50 U/ml streptomycin), 2 ml vitamins (100x), 5 ml non-essential amino acids (100x) and 0.05 mM β-mercaptoethanol. Media supplements were purchased from Gibco, Biochrome or Sigma respectively. FCS was heat inactivated for 30 min at 56°C before use. Each batch of FCS as well as each RPMI batch was tested before use.
Culturing of cells was performed at 37°C, with 5% CO_2 and 95% relative humidity in an incubator. U937 and THP-1 cells grow in suspension and were split 1:2 to 1:4 in fresh medium every 2-4 days.
For washing and harvesting, mammalian cells were centrifuged using the general cell program: 8 min, 300xg, 4°C.

4.1.1.3 Freezing and thawing cells

Cells were harvested and resuspended at 10×10^6 cells/ml in 900 µl ice cold medium including 10% FCS. After inverting the mix and transferring it into cryo-vials 100 µl DMSO (10% final) were added and the tubes were rapidly inverted to mix cells properly. To allow gradual freezing at a rate of 1°C/min, the cryo-vials were placed in isopropanol-filled cryo-containers (Nalgene). For long-term storage, samples were transferred in liquid nitrogen (-196°C).

4.1.1.4 Transfection of THP-1 cells with DEAE dextran

One day prior transfection, a sufficient number of THP-1 cells was seeded into tissue culture flasks with a density of $0.5x10^6$ cells / ml. After o/n culture at normal conditions, 6 ml cell suspension per transfection (equals $3x10^6$ cells per transfection) was transferred into a Falcon tube and centrifuged (1200rpm, 10min, 4°C). Supernatant was discarded and 5 ml STBS buffer added to wash remaining cells. A second centrifugation round was performed and supernatant discarded. For each transfection, 200 ng of the luciferase reporter plasmid and 20 ng of the renilla control plasmid (pHRL-TK) in 70 µl STBS are combined with 70 µl DEAE-Dextran (800 µg/ml), mixed and immediately added drop-wise directly onto the THP-1 cell pellets. The cells were shaken shortly and incubated for 20 min at 37°C, washed twice with 5 ml STBS each, resuspended in 6 ml RPMI (with 10% FCS) and cultured in 60 mm tissue culture dishes. Cells are harvested after 48h, the culture dishes washed once with 5 ml PBS at RT and the cells are pelleted (10 min, 400xg, 4°C). Pellets are washed once with 10 ml PBS and the PBS removed completely by decanting and briefly inverting the centrifuge tubes onto Kleenex paper towels. After 48h cells were lysed and assayed (see section 4.1.1.5).

Required materials:

STBS buffer:	3.029 g	(25 mM)	Tris/HCl, pH 7.5
	8.01 g	(137 mM)	NaCl
	0.373 g	(5 mM)	KCl
	0.084 g	(0.6 mM)	Na_2HPO_4
	0.103 g	(0.7 mM)	$CaCl_2{\cdot}2H_2O$
	0.102 g	(0.5 mM)	$MgCl_2{\cdot}6H_2O$
	Ad 1000 ml ddH_2O, autoclave		

DEAE dextran: 10 mg/ml in STBS

4.1.1.5 Measuring Luciferase activity

Luciferase activity was tested with the Dual-Luciferase Reporter Assay System (Promega) following the manufacturer's instructions.48 hours after transfection, cells were transferred to 14 ml polystyrene round-bottom tubes, centrifuged at 300×g for 10 minutes and washed with PBS. After discarding the supernatant, cells were lysed by adding 300 µl diluted passive lysis buffer and incubation for 15 min at RT. The lysate was cleared and Firefly as well as Renilla luciferase activities were measured on a Sirius photometer. Firefly luciferase activity of individual transfections was normalized against Renilla luciferase activity.

4.1.2 Bacterial culture

4.1.2.1 Bacterial growth medium

E .coli strains were streaked out on solid LB agar with required selection antibiotics and grown for 12-36 h. DH10B, PIR1 and TOP10 strains were grown at 37 °C. Single colonies were picked and inoculated in LB medium containing the appropriate selection antibiotics and grown overnight at 37 °C on a shaker at 200 rpm.

Table 4-1 - Antibiotics for selective bacterial culture

Antibiotic	Stock concentration	Final concentration	Dilution
Ampicillin	100 mg/ml	100 µg/ml	1:1000
Zeocin	100 mg/ml	25 µg/ml	1:4000

Required reagents and materials:

LB medium:	10 g	1%	Bacto Tryptone (BD, #211705)
	10 g	170 mM	NaCl (AppliChem, #A3597)
	5 g	0.5%	Bacto yeast extract (BD, #212720)
	Ad 1000 ml ddH_2O, adjust to pH 7.5, autoclave		
LB agar plates:	10 g		Bacto Tryptone (BD, #211705)
	10 g		NaCl (AppliChem, #A3597)
	5 g		Bacto yeast extract (BD, #212720)
	15 g		Bacto agar (BD, #214010)
	Ad 1000 ml ddH_2O, adjust to pH 7.5, autoclave, cool to 50°C and add the required antibiotics. Pour solution into 10 cm Petri dishes and store at 4°C.		
Petri dishes:	94 mm diameter	Greiner Bio-one, #633102	

4.1.2.2 Glycerol stocks

For long-term storage, bacteria were stored in autoclaved LB-Glycerol medium (60%). 200 µl of bacterial suspension was added to 1000 µl of LB-Glycerol media, mixed, and stored at -80 °C.

4.1.2.3 Transformation of chemically competent *E.coli*

Chemically competent *E.coli* (50 µl) were thawed on ice, 1-25 ng plasmid DNA was added and the suspension was mixed gently and incubated on ice for 30 min. Cells were then heat-shocked at 42°C for 30 s, immediately cooled on ice for 2 min and 250 µl SOC medium was added. To express the resistance, bacteria were incubated for 1 h at 37°C with shaking. Afterwards 50-150 µl of the transformation reaction were plated and incubated at 37°C on LB agar containing the antibiotic necessary for selection of transformed cells overnight.

Required reagents:

SOC medium:	20 g	2%	Bacto Tryptone (BD, #211705)
	0.6 g	10 mM	NaCl (AppliChem, #A3597)
	5 g	0.5%	Bacto yeast extract (BD, #212720)
	0.2 g	3 mM	KCl
	Ad 1000 ml ddH_2O, autoclave and		
	Add to the cooled solution:		
	10 ml	10 mM	$MgCl_2$ (1 M), sterile filtered
	10 ml	10 mM	$MgSO_4$ (1 M), sterile filtered
	10 ml	20 mM	Glucose (2 M), sterile filtered

4.1.2.4 Isolation of human white blood cells

Peripheral white blood cells of humans were obtained after ammonium-chloride-potassium lysis of red blood cells (RBCs). RBCs are permeable to ammonium chloride and the uptake results in the accumulation of cytoplasmic ions and subsequent osmotic lysis through uptake of water.
For ACK lysis 5-10 ml peripheral blood (EDTA or heparin treated) were mixed with 25 ml ACK lysis buffer, vortexed and incubated for 5 min at room temperature. After centrifugation (10 min, 1300rpm, 4-20°C) supernatant was removed carefully with a pipette and cell pellet was resuspended with 25 ml fresh ACK lysis buffer. Incubation for 5 min at room temperature was followed by centrifugation and removal of supernatant. The cell pellet was washed three times with PBS (phosphate buffered saline) and is then ready for further treatment or can be stored at -20°C.

Required reagents:

6x ACK lysis buffer:	49.64 g	0.155 M	NH_4Cl
	6.00 g	0.1 M	$KHCO_3$
	0.222 g	0.1 mM	EDTA-disodium dihydrate
	Ad 1000 ml ddH_2O, adjust to pH 7.4, sterile filtered		

4.1.2.5 Isolation of human mononuclear cells by density gradient centrifugation

Mononuclear cells (MNCs) from peripheral blood (PB) or bone marrow (BM) were isolated by density gradient centrifugation over Pancoll (Pan Biotech). Bone marrow samples or anticoagulant-treated blood samples were diluted 1:2 with PBS and layered carefully over Pancoll (10 ml). After centrifugation for 25 min at 2000 rpm and with soft break, several layers with separate cell types according to the differential migration can be observed. The bottom layer contains erythrocytes and erythrocytes followed by a Pancoll layer. Between the Pancoll layer and the top layer containing the plasma, the mononuclear cells can be found in the interface. This interface layer is aspirated using a 5 ml transfer pipette and washed three times with PBS. Cells were then counted using Tuerk solution and either frozen in FCS including 10% DMSO or subsequently further processed.

4.2 General molecular biological methods

4.2.1 Preparation and analysis of DNA

4.2.1.1 Isolation of plasmid DNA from *E. coli*

Plasmid preparation for up to 50 µg plasmid DNA was done using the NucleoSpin® Plasmid Quick Pure Kit (Macherey-Nagel) following the manufacturer´s instructions. For higher plasmid DNA content up to 100 µg plasmid preparation was done with the Plasmid Midi Kit (Qiagen) following the manufacturer´s instructions. After preparation plasmids were verified by restriction digest with suitable restriction enzymes (1.5 h at 37°C). The resulting fragments were separated by gel electrophoresis together with a suitable DNA ladder to test for insert presence and size.

4.2.1.2 Isolation of genomic DNA from mammalian cells

Genomic DNA from human blood cells was isolated after ACK lysis (see No. x) using the Blood & Tissue Culture Kit (Qiagen) according to the manufacturer´s instructions. DNA concentration was determined with the NanoDrop 1000 spectrophotometer.

4.2.1.3 Fragmentation of genomic DNA and chromatin

4.2.1.3.1 Sonication of genomic DNA

For Methyl-CpG-immunoprecipitation (MCIp) or targeted bisulfite sequencing genomic DNA was fragmented to a mean size of 200-300 bp or 180-220 bp respectively using the Covaris S2 system. This system works with focused ultrasonication and leads to high reproducibility.
Ultrasonication was done in microTUBEs (6 mm x 16 mm) either screw-capped or snap-capped with a maximum volume of 130 µl. Following settings were used for ultrasonication:

Table 4-2 - Parameters for focused ultrasonication with Covaris

Parameters	Settings for MCIp	Settings for targeted BS-seq
Intensity	5	5
Duty Cycle	10%	10%
Cycles per burst	200	200
Treatment time [s]	180	120
Temperature [°C]	7	7
Water Level S2	12	12
Mode	Frequency sweeping	Frequency sweeping

Fragment range was controlled by agarose gel electrophoresis (2%, 110 V, 1h, see section 4.2.1.4).

4.2.1.3.2 Sonication of chromatin

Crosslinked chromatin for Chromatin immunoprecipitation (ChIP) was fragmented by sonication using the Branson Sonifier 250. The following settings were used:

Output	3
Sonication time	5 x 10 sec

After each sonication step the sample was placed for 30 sec on ice to cool again. The fragmented lysates were centrifuged at 4°C for 5 min at 13000 g to pellet remaining cell debris. Supernatant was transferred to a new 1.5 ml tube. Fragment range was controlled by agarose gel electrophoresis (2%, 110 V, 1h, see section 4.2.1.4).

4.2.1.4 Agarose gel electrophoresis

The required amount of agarose according to Table 4-3 was added to the corresponding amount of TAE (1x). The slurry was heated in a microwave oven until the agarose was completely dissolved. Ethidium bromide was added after cooling the solution to 50-60 °C. The gel was cast, mounted in the electrophoresis tank and covered with TAE (1x). DNA-containing samples were diluted 4:1 with DNA

loading dye (5x), mixed, and loaded into the slots of the submerged gel. Depending on the size and the desired resolution, gels were run at 35-120 V for 30 min to 3 h.

Table 4-3 - Agarose concentrations for different separation ranges

% agarose in gel	Efficient range of separation (kb)
0.5	Genomic DNA
1	0.4 – 6
1.5	0.2 – 3
2.0	0.1 – 2

Required reagents:

TAE (50x)	252.3 g	2 M	Tris
	250 ml	250 mM	NaOAc/HOAc, pH 7.8
	18.5 g	50 mM	EDTA
	Ad 1000 ml ddH_2O, autoclave		
DNA loading dye (5x)	500 µl	50 mM	Tris/HCl, pH 8.0
	500 µl	1%	SDS (20%)
	1 ml	50 mM	EDTA (0.5 M), pH 8.0
	4 ml	40%	Glycerol
	10 mg	1%	Bromophenol blue
	Ad 10 ml nuclease-free ddH_2O, store at 4°C		

4.2.1.5 Molecular cloning of vectors

Subcloning of PCR products into a sequencing or expression vector was done by digestion of both, PCR product and vector, with the appropriate restriction endonuclease (see section 4.2.1.8) or using the Gibson assembly (see section 4.2.1.7). For directional cloning, restriction sites were introduced by adding the correct recognition sequences to the primer sequences. Following ligation reaction and transformation in chemically competent cells was done as described in section 4.2.1.9 and 4.1.2.3, respectively. Successful insertion of the fragment into the vector was controlled by preparing plasmid DNA from liquid cultures (see section 4.2.1.1) and subsequent sequencing using vector-specific primers.

4.2.1.6 Reporter gene assays

Differentially methylated regions (DMRs) ranging from 400 – 800 bp were ordered as gblocks® gene fragments at the IDT company and solved with TE buffer according the manufacturer's instructions (final concentration: 10 ng / µl). Reporter plasmids pCpGL-basic (empty reporter vector, CpG-free),

pCpGL-CMV/T.EF1A (heterologous EF1A promoter with upstream terminator, CpG-free) and pCpGL-CMV/T.A1FE (inverted EF1A promoter with upstream terminator, CpG-free) were linearized (see 4.2.1.8) and assembled with the appropriate gblocks® using the Gibson assembly reaction (see 4.2.1.7). Ligated constructs were transformed into competent PIR1 E.coli cells (see4.1.2.3) and plasmids were isolated as described in 4.2.1.1. Inserts were verified by sequencing and luciferase reporter constructs were either mock-treated or methylated in vitro with SssI methylase and purified using PEG precipitation. Transfection of THP-1 with DEAE and Luciferase measurement was done according to sections 4.1.1.4 and 4.1.1.5.

4.2.1.7 Gibson assembly

DMRs (differentially methylated regions) were cloned using the isothermal single tube Gibson assembly reaction. After linearization of the plasmid with the appropriate restriction endonuclease (see section 4.2.1.8) following reaction was set up:

Table 4-4 - Reaction composition for Gibson assembly

Component	Volume
50 ng linearized plasmid	x µl
Insert (3-fold molar excess)	z µl
Gibson Assembly Master Mix (2x)	10 µl
H_2O	20 µl – (x +z) µl
Final Volume	20 µl

Reaction mix was incubated for 1h at 50°C in a thermal cycler and subsequently used for transformation into bacteria (see section 4.1.2.3). The recommended 3-fold molar excess of insert was calculated using following formula:

$$\mathrm{m_{Insert}(ng)} = \frac{3 \text{ x } \mathrm{m_{Plasmid}(ng)} \text{ x } \mathrm{length_{Insert}(bp)}}{\mathrm{length_{Plasmid}(bp)}}$$

4.2.1.8 Restriction endonuclease digestion

Control digest from Mini/Midi preparations was done using 1-2 µl Plasmid DNA and the appropriate restriction endonucleases. The Plasmid DNA was digested at 37°C for 1.5 h and controlled with agarose gel electrophoresis.

To avoid relegation of digested vector ends, dephosphorylation of 5' end was done using 1 Unit alkaline phosphatase (Roche). Reaction mixture was incubated at 37°C for 30 min and subsequent purification of the linearized vector was done using gel extraction (see section 4.2.1.11).

4.2.1.9 Ligation reaction

Restriction enzyme treated vectors and PCR products were ligated in a 10 µl reaction at a 3-fold molar excess of insert to vector, using 25-50 ng of vector. Ligation was carried out overnight at 16°C with 1U T4 DNA ligase.

4.2.1.10 Quantification of DNA

The exact DNA concentration was determined either by using the PicoGreen dsDNA Quantitation Reagent (Molecular Probes) or by using the NanoDrop spectrophotometer.

4.2.1.11 Purification of DNA fragments by gel extraction

DNA fragments were size-separated by agarose gel electrophoresis (see section 4.2.1.4) and the desired fragment was excised under UV illumination. DNA fragments were purified by gel extraction using QIAquick Gel Extraction Kit (Qiagen) following the manufacturer´s instructions.

4.2.1.12 Polyethylene glycol precipitation of DNA

For precipitation of DNA from small volumes, e.g. PCR reactions or endonuclease digestions, one volume of PEG 8000 Mix was added to the DNA-containing solution, vortexed and incubated for 10 min at RT. After centrifugation (10 min, 13000 rpm, RT), the supernatant was discarded and the precipitated DNA was washed by carefully adding 200 µl EtOH (100%) to the tube wall opposite of the pellet. Following centrifugation (10 min, 13000 rpm, RT), the supernatant was carefully removed. The pellet was dried and resuspended in H_2O in 50-75% of the initial volume.

Required reagents:

PEG 8000 Mix	26.2 g	26.2 %	PEG 8000
	20 ml	0.67 M	NaOAc (3 M) pH 5.2
	660µl	0.67 mM	$MgCl_2$ (1 M)
	Ad 250 ml ddH_2O		

4.2.1.13 Polymerase chain reaction

Polymerase chain reaction (PCR) allows in vitro amplification of specific DNA segments. A thermo stable DNA polymerase synthesizes the sister strand of a heat-denatured single-stranded DNA when deoxynucleotide triphosphates are added under appropriate conditions[173]. The polymerization reaction is “primed” with small oligonucleotides that anneal to the template DNA strand through base pairing giving the reaction its specificity by defining the borders of the amplified segment. Standard applications of PCR reactions are explained in the following and are used unless otherwise mentioned. More specialized applications are explained in more detail within the specific method.

Primer design

Unless otherwise mentioned sequences for generating primers were extracted using the UCSC Genome Browser. In general primers were designed using PerlPrimer Software and controlled using PCR and BLAT functions of the UCSC Genome Browser. Following settings were used to design primers:

Primer T_m: 65 – 68°C
Primer length: 18 – 28 bp
Amplicon size: 80 – 150 bp

Standard PCR

PCRs were generally performed in 0.5 ml PCR tubes in a MJ research PTC 200 thermocycler (Biozym). The "calculated temperature" feature was used to decrease temperature hold times and additionally the lid was heated to 105°C to prevent vaporization.
Standard PCR to amplify specific target regions was performed with the PhusionTM Hot Start II High Fidelity DNA Polymerase (Thermo Fisher Scientific) according to the following protocol.

Table 4-5 - Reaction composition for standard PCR

Component	Volume	Final concentration
H_2O	Add 50.00 µl	
5 x Phusion HF buffer	10.00 µl	1 x
10 mM dNTPs	1.00 µl	200 µM each
Primer for	1.00 µl	0.5 µM
Primer rev	1.00 µl	0.5 µM
Template DNA	x µl	
Phusion Hot Start High Fidelity DNA Polymerase (2 U/µl)	0.50 µl	0.02 U/µl

General parameter settings for analytical PCR are summarized in Table 4-6.

Table 4-6 - Cycling protocol for standard PCR

Cycle step	2-step protocol		Cycles
	Temperature	Time	
Initial denaturation	98°C	30 sec	1
Denaturation	98°C	7 sec	
Annealing	-	-	32
Extension	72°C	30 sec	
Final extension	72°C	10 min	1
	4°C	hold	

Quantitative real-time PCR

Quantitative real-time PCR (qPCR) was used to test for specific DNA enrichment by Methyl-CpG immunoprecipitation (MCIp, see section 4.2.1.16), Chromatin immunoprecipitation (ChIP, see section 4.2.1.17) as well as for local expression analysis after reverse transcription quantitative PCR (see section 4.2.1.13). QPCR was performed with the QuantiFast SYBR Green Kit (Qiagen) in 96-well format adapted to the Eppendorf Realplex Mastercycler EpGradient S system (Eppendorf) and 384-well format adapted to the LightCycler480 system (Roche). The relative amount of amplified DNA was measured through the emission of light by the SYBR green dye after each extension step. Specificity of the amplification product was determined by a melting curve.

The reaction was composed of following reagents with conditions shown in Table 4-7 and Table 4-8.

Table 4-7 - Reaction composition for qPCR

Component	Volume	Final concentration
SYBR Green mix (2x)	5.00 µl	1 x
Nuclease-free ddH_2O	2.00 µl	
Primer forward (10 µM)	0.50 µl	0.5 µM
Primer reverse (10 µM)	0.50 µl	0.5 µM
Template DNA	2.00 µl	
Final Volume	10.00 µl	

Table 4-8 - Cycling protocol for qPCR

Cycle step	Temperature	Time	Number of cycles
Initial denaturation	95°C	5 min	1 x
Denaturation	95°C	8 sec	45 x
Annealing & Extension	60°C	20 sec	
Final denaturation	95°C	15 sec	1 x
Final extension	60°C	15 sec	1 x
Melting Curve	60 → 95°C	10 min	1 x
Cooling	4°C	hold	

To calculate amplification efficiency, a dilution series (e.g. 1:10; 1:50; 1:100, 1:1000) of a suitable DNA sample was additionally measured for each amplicon. The software automatically calculated relative DNA amounts based on the generated slope and intercept. Specific amplification was determined by melting curve analyses. Data were imported and processed in Microsoft Excel 2010. All samples were generally measured in duplicates.

4.2.1.14 *In vitro* methylation of DNA

As a control for methylation analysis fully methylated DNA was generated using SssI methyltransferase (New England Biolabs). S-adenosyl methionine (SAM) was used as a methyl donor. In general the following reaction conditions were used:

Table 4-9 - Reaction composition for *in vitro* methylation

Component	Volume	Final concentration
DNA	x µl	2 µg
Nuclease-free ddH_2O	39.1 – x µl	
Alternative buffer (5x)	10.00 µl	1 x
S-adenosyl methionine (32 mM)	0.40 µl	160 µM
SssI (20.000 U / ml)	0.50 µl	10 U
Final Volume	50.00 µl	

After 2 hours incubation at 37°C the reaction was supplied with additional 0.5 µl SAM and 0.2 µl M.SssI followed by incubation of another two hours.
After methylation reaction in vitro methylated genomic DNA was purified with PEG precipitation and finally quantified using a NanoDrop spectrophotometer. Completeness of methylation was controlled by digesting both methylated and unmethylated DNA using the methylation-sensitive restriction enzyme *HpaII*.

Alternative buffer:

0.029 g	50 mM	NaCl
100 µl	10 mM	Tris-HCl (1 M) pH 7.9
0.029 g	10 mM	EDTA
50 µl	160 µM	S-adenosylmethionine (32 mM)
Ad 10 ml ddH_2O		

4.2.1.15 Sanger Sequencing

Sanger sequencing was done by Geneart/Life Technologies (Regensburg, Germany) with the Applied BiosystemsTM system. The obtained sequence was analyzed with the BioEdit or GeneRunner Software or with the Blat function from the UCSC genome browser.

4.2.1.16 Methyl-CpG-immunoprecipitation (MCIp)

MCIp is based on a fusion protein consisting of the methyl-CpG-binding domain of MBD2 fused to the Fc-tail of human IgG1. The method enables rapid enrichment of methylated CpG rich DNA as well as fractionation of DNA fragments according to their methylation level. DNA affinity of the antibody-like protein is dependent on the density of methylated CpGs and salt concentrations in the buffer. MCIp was performed with the EpiMark® Methylated DNA Enrichment Kit (NEB), which is based on a method established in our laboratory. The method allows for the fractionation of DNA fragments according to their level of methylated CpGs as well as the enrichment of methylated DNA fragments. Fractionation is achieved by washing the immobilized MBD2-Fc-bound DNA fragments with increasing NaCl concentration in the wash buffer. Genomic DNA fragments that exhibit low to intermediate methylation levels reside in fractions obtained from washing with buffer of 150-500 mM NaCl concentration. Remaining bound DNA fragments, which exhibit high methylation levels, can be totally recovered by washing with buffer of ≥ 1000 mM NaCl concentration. To enrich in general for methylated DNA fragments, all remaining bound DNA fragments after washing with buffer of 150 mM NaCl concentration can be recovered by incubating the immobilized DNA in nuclease-free ddH_2O at 65 °C for 15 min. Both protocols can be combined.

MCIp was performed according to the manufacturer's instructions. Fragments with low to intermediate methylation level were eluted by washing with 150 mM and 500 mM NaCl wash buffer. Highly methylated fragments were recovered by incubation with 52 µl nuclease-free ddH_2O at 65 °C for 15 min. Successful fractionation was confirmed by quantitative PCR with control loci listed in the following table.

Table 4-10 - qPCR control loci for MCIp

Gene symbol	Expected outcome in H_2O fraction
'*empty 6.2'*	CpG-free region → depletion of mCpGs (no signal)
Snrpn (promoter)	Imprinted region → 50 % enrichment of mCpGs

In somatic cells, imprinted regions exhibit allele-specific DNA methylation depending on the parental origin. For *Snrpn*, the paternal allele is, in contrast to the completely methylated maternal allele, free of DNA methylation. The paternal allele resides in the lower salt fractions (50% recovery) whereas the maternal allele was enriched in the fraction with the highest salt concentration (1000 mM) (50% recovery). In contrast, wash fractions with intermediate salt concentration are depleted of fragments from the Snrpn locus.
The '*empty 6.2'* control locus comprises a region without any CpGs, resulting in no binding to the fusion protein and therefore no signal in the qPCR.

4.2.1.17 Chromatin immunoprecipitation (ChIP)

20 Mio cells were harvested by centrifugation at 4°C for 10 min at 300xg and subsequently resuspended in 20 ml RPMI inclusive 10 % FCS. Crosslink reaction was started by addition of 1.33 ml p-Formaldehyde (16 %) per 20 ml cell suspension and incubation for 10 min at RT. Reaction was stopped by addition of 1/20 Vol Glycine (2.625 M) followed by two washing steps with cold PBS/PMSF (1 mM). Pellets can be stored at -80 °C up to months after snap freezing in N_2 (l).
Cell lysis of crosslinked cells was carried out in 500 µl Suspension buffer and 500 µl Cell lysis buffer for 10 min on ice. The lysate was transferred to a new 1.5 ml tube and centrifuged at 4 °C for 5 min at 700 xg to pellet the nuclei. The nuclei were then lysed by resuspension in 450 µl Nuclear Lysis Buffer (for Branson Sonifier) or 800 µl (for Covaris). The lysed nuclei can be stored in Nuclear Lysis Buffer at -80 °C up to months after snap freezing in N_2 (l).
After sonication according to section 4.2.1.3.2, lysate was cleared from remaining cell debris by centrifugation at 4 °C for 5 min at 13000xg.
To evaluate the fragmentation size, 30 µl of the sonicated and cleared lysate was mixed with 14 µl NaCl (5 M) and 26 µl L2 (without PMSF and inhibitors) and incubated o/n at 65 °C to reverse the crosslinking. After digestion of RNAs with 10 µg RNase A for 1 h at 37 °C, the sample was purified with the NEB Monarch® PCR & DNA Cleanup Kit according the manufacturer's instructions. The sample was eluted in 30 µl EB buffer and size-separated by agarose gel electrophoresis (2% agarose gel, 50 bp ladder) according to section 4.2.1.4. The remaining lysate can be stored at -80 °C up to months.
For each immunoprecipitation (IP), 80 µl of cleared lysate (chromatin of ~2 Mio cells) were used. For normalization purposes, per sample 4 µl of lysate (5% of IP) were retained as input.
The first step included preparation of Sepharose CL-4B beads for preclearing of the chromatin samples. 50 µl beads per immunoprecipitation (IP) were washed twice with TE buffer (pH 8.0) and centrifuged for 1 min at 600xg at RT. Dilution buffer was added to the previous volume as well as 25 µl 20 % BSA /1 ml beadsand 4 µl Glycogen/1 ml beads. Mixture was rotated for 2 h at room temperature.
For pre-clearing, the chromatin batch was diluted with 1.5x volume Dilution buffer (with inhibitors, results in 200 µl diluted lysate per IP) and incubated with 50 µl pre-clearing beads per IP on a rotor at 4 °C for 2 h. To recover the pre-cleared lysate, the mix was centrifuged at 4 °C for 5 min at 13000xg and the supernatant (200 µl) transferred to a new 0.5 ml tube. In the next step, appropriate antibodies (2.5 µg per IP) were added to the supernatant and incubated over night at 4°C on a rotor.

In parallel to the o/n incubation of the chromatin antibody mixture, 40 µl Protein A-coated beads per individual IP for the subsequent pull down of the antibody-bound chromatin fragments were prepared. The beads were washed twice by repeated centrifugation at RT for 1 min at 600 xg and resuspension in TE buffer. The washed beads were diluted to the initial volume with Dilution buffer (DB, w/o inhibitors). After addition of 25 µl BSA (20%) and 4 µl Glycogen per ml Protein A-coated beads, the beads were incubated o/n on a rotor at 4 °C.

To pull down the chromatin-antibody complexes, 50 µl of blocked beads were added to each IP sample and incubated on a rotor for 3 h at 4 °C. The beads with the bound chromatin-antibody complex were centrifuged at 4 °C for 5 min at 500xg. Supernatant was discarded and unspecifically bound chromatin was washed away by repeated washing steps with increasing stringency.

The beads were resuspended twice in 400 µl of each of the three washing buffers (WBI-III) and incubated on a rotor for 5 min at RT after each resuspension. Final washing step was performed by three times resuspension in 400 µl TE buffer. To elute the bound chromatin-antibody complexes, the beads were incubated with 110 µl Elution Buffer (EB) for 20 min at RT (shaking every 5 min). After centrifugation at 4 °C for 1 min at 500xg, 100 µl supernatant were transferred to a new 1.5 ml tube. Elution was repeated with additional 110 µl Elution Buffer and incubation for 10 min and after centrifugation, another 100 µl supernatant was combined with the first 100 µl eluate. Crosslinking of the eluted chromatin fragments was reversed by incubation with 5 µl Proteinase K (20 µg/µl) o/n at 65 °C. In parallel, 4 µl input were diluted to 200 µl with Elution Buffer and also subject to Proteinase K treatment.

The DNA fragments released from the recovered chromatin as well as from the input sample by Proteinase K treatment were purified by RNase A digestion (0.33 µg/µl, 2 h at 37 °C) and subsequent clean up with the NEB Monarch® PCR & DNA Cleanup Kit according to the manufacturer's instructions. DNA is loaded onto the column by centrifugation at 16000xg and eluted after 2 min incubation at RT with 53 µl pre-warmed Elution Buffer (55 °C, part of the kit).

To validate specific enrichment by immunoprecipitation, 10 µl of the eluate were diluted with Elution Buffer for subsequent qPCR according to section 4.2.1.13.

Required materials:

Formaldehyde	16%		Formaldehyde (ThermoFisher, # 28906)
Glycine	2.625 M	9.85 g	Glycine (Roth, #3908.1) → Ad 50 ml nuclease-free ddH_2O
BSA	20%	2 g	Bovine serum albumin (Sigma-Aldrich, #A4503) → Ad 10 ml nuclease-free ddH_2O

Glycogen		5 mg/ml	Glycogen (Life Techn., #AM9510)
HEPES/KOH (pH 7.9)	1 M	238.3 g	4-(2-hydroxyethyl)-1-Piperazineethanesulfonic acid (Sigma-Aldrich, #H3357) → Adjust with KOH to pH 7.9 → Ad 1 l nuclease-free ddH_2O
KCl	2 M	149.1 g	Potassium chloride (Merck Millipore, #1049360250) → Ad 1 l nuclease-free ddH_2O
EDTA/NaOH (pH 8)	0.5 M	186.1 g	Ethylenediaminotetraacetic acid disodium salt dehydrate (AppliChem, #A2937) → Dissolve in 500 ml nuclease-free ddH_2O → Adjust with NaOH to pH 8.0 → Ad 1 l nuclease-free ddH_2O
PMSF	100 mM	1.74 g	Phenylmethanesulfonyl fluoride (Sigma-Aldrich, #P7626) → Ad 100 ml Isopropanol (p.a.) (Merck Millipore, #1096342511)
NaCl	5 M	29.2 g	Sodium chloride (AppliChem, #A3597) → Ad 100 ml nuclease-free ddH_2O
Tris/HCl (pH 7.4)	1 M	121.1 g	Tris ultrapure (AppliChem, #A1086) → Adjust with HCl to pH 7.4 → Ad 1 l nuclease-free ddH_2O
TE (1x, pH 8.0)	10 mM 1 mM	10 ml 2 ml	Tris/HCl (1 M, pH 8.0) EDTA/NaOH (0.5 M, pH 8.0) → Ad 1 l nuclease-free ddH_2O

SDS	20%	20 g	Sodium dodecyl sulfate (Sigma-Aldrich, #L4390) → Ad 100 ml nuclease-free ddH_2O
Empigen BB	30%	3 ml 7 ml	Empigen BB detergent (100%) (Sigma-Aldrich, #45165) nuclease-free ddH_2O
Triton X-100	10%	1 ml 9 ml	Triton X-100 (100%) (Sigma-Aldrich, #T8787) nuclease-free ddH_2O
LiCl	2.5 M	10.6 g	Lithium chloride (Merck Millipore, # 438002) → Ad 100 ml nuclease-free ddH_2O
Deoxycholate	10%	1g	Sodium deoxycholate (Sigma-Aldrich, #D6750) → Ad 10 ml nuclease-free ddH_2O
$NaHCO_3$	1 M	84 g	Sodium hydrogen carbonate (Merck Millipore, #1063290500) → Ad 1 l nuclease-free ddH_2O
Cell Buffer Mix	10 mM 85 mM 1 mM	1 ml 4.25 ml 200 µl 91.55 ml	HEPES/KOH (1 M, pH 7.9) KCl (2 M) EDTA/NaOH (500 mM, pH 8.0) nuclease-free ddH_2O
Immediately before use, add the following inhibitors per ml CBM:			
	1 mM	10 µl 20 µl	PMSF (100 mM) Inhibitor Mix (50x) (Roche, #04693132001)
Suspension Buffer		900 µl 100 µl	CBM nuclease-free ddH_2O
Cell Lysis Buffer	 1%	900 µl 100 µl	CBM NP-40 (10%) (Roche, #11332473001)

Nuclear Lysis Buffer	50 mM	5 ml	Tris/HCl (1 M, pH 7.4)
	1%	5 ml	SDS (20%)
	0.5%	1.667 ml	Empigen BB (30%)
	10 mM	2 ml	EDTA/NaOH (500 mM, pH 8.0)
	83.33 ml		nuclease-free ddH_2O

Immediately before use, add the following inhibitors per ml Nuclear Lysis Buffer:

	1 mM	10 µl	PMSF (100 mM)
		20 µl	Inhibitor Mix (50x)
			(Roche, #04693132001)

Dilution Buffer (DB)	20 mM	2 ml	Tris/HCl (1 M, pH 7.4)
	100 mM	2 ml	NaCl (5 M)
	2 mM	400 µl	EDTA/NaOH (500 mM, pH 8.0)
	0.5%	5 ml	Triton X-100 (10%)
		87.6 ml	nuclease-free ddH_2O

Immediately before use, add the following inhibitors per ml Dilution Buffer:

	1 mM	10 µl	PMSF (100 mM)
		20 µl	Inhibitor Mix (50x)
			(Roche, #04693132001)

Wash Buffer

WBI	20 mM	2 ml	Tris/HCl (1 M, pH 7.4)
	150 mM	3 ml	NaCl (5 M)
	0.1%	500 µl	SDS (20%)
	1%	10 ml	Triton X-100 (10%)
	2 mM	400 µl	EDTA/NaOH (500 mM, pH 8.0)
		84.1 ml	nuclease-free ddH_2O
WBII	20 mM	2 ml	Tris/HCl (1 M, pH 7.4)
	500 mM	10 ml	NaCl (5 M)
	1%	10 ml	Triton X-100 (10%)
	2 mM	400 µl	EDTA/NaOH (500 mM, pH 8.0)
		84.1 ml	nuclease-free ddH_2O
WBIII	10 mM	1 ml	Tris/HCl (1 M, pH 7.4)
(protect from light!)	250 mM	10 ml	LiCl (2.5 M)
	1%	10 ml	NP-40 (10%)
	1%	10 ml	Deoxycholat (10%)
	1 mM	200 µl	EDTA/NaOH (500 mM, pH 8.0)
		68.8 ml	nuclease-free ddH_2O

Elution Buffer	0.1 M	500 µl	$NaHCO_3$ (1 M)
	1%	250 µl	SDS (20%)
		4.25 ml	nuclease-free ddH_2O

4.2.1.18 Library preparation for next generation sequencing

Next generation sequencing (NGS) libraries were prepared to be sequenced on Illumina platforms (HiSeq1000). In general, DNA low binding tubes (Sarstedt) were used. Library preparation was subsequently enhanced during the time of this thesis, resulting in four different preparation protocols (NGS library protocol I, II, III & NEBNext Ultra II DNA Library Prep Kit for Illumina).

NGS library protocol I

The first library preparation step is an end repair to convert overhangs of double stranded DNA fragments into phosphorylated blunt ends. Following reaction set up was used:

Table 4-11 - Reaction composition for end repair (NGS library protocol I)

Component	Volume [µl]
DNA sample	40.00
H2O	3.34
T4 DNA ligase buffer (10x)	5.00
dNTP mix (10 mM each)	1.00
T4 DNA polymerase	0.30
Klenow fragment	0.06
T4 Polynucleotide kinase	0.30
Total Volume	50.00

The reaction was incubated in a thermal cycler at 20°C for 30 min (without heated lid) and cleaned up with AMPure XP magnetic beads (equilibrated for 30 min at RT on a rotor) at a ratio 1:1.8.
90 µl beads were added to the end repair mix (50 µl), vortexed (speed 6) two times, inverted and vortexed again and then incubated at RT for 5 min. The tube was placed on a magnetic particle concentrator (MPC) for 2 min. After removal of the supernatant, the trapped beads were washed twice with 500 µl Ethanol (70%, prepared freshly) for 30 sec each. To remove residual Ethanol, the beads were air-dried at 37 °C for 2-5 min. The bound DNA fragments were eluted by incubation with 44.4 µl nuclease-free ddH2O at RT for 5 min. After centrifugation at RT for 1 sec at 13000xg, the tubes were placed on a MPC for 2 min and the supernatant was transferred to a new 1.5 ml tube for a second round of clean up.
For introduction of A-overhangs to the 3' end of the phosphorylated blunt ends of DNA fragments following reaction was done:

Table 4-12 - Reaction composition for A-overhang introduction (NGS library protocol I)

Component	Volume [µl]
Blue buffer (10x)	5.00
dATP (10 mM)	0.10
Klenow (3'-5' exo-)	0.50
Total Volume	5.60

The reaction mix was added to the end-repaired sample and incubated at 37°C for 30 min in a thermal cycler. After a cleanup with AMPure XP beads (1:1.8 ratio) as described above, adapters were ligated to the end of the DNA fragments allowing later hybridization to an Illumina Next Generation Sequencing flow cell.

Table 4-13 - Reaction composition for adapter ligation (NGS library protocol I)

Component	Volume [µl]
Rapid Ligation buffer (2x)	16.00
Bioo NEXTflex adapter (1:2 dilution)	1.00
T4 DNA Ligase	1.00
Sample with A-overhangs	14.50
Total Volume	32.50

After incubation at 30°C for 10 min in a thermal cycler reaction mix was filled up with 1 µl of 0.5 M EDTA (pH 8.0) and 16.5 µl water to a total volume of 50 µl. To remove unligated adapters or adapter dimers purification with AMPure XP beads was done using a ratio of 1:1.8.
Size selection of DNA fragments was performed by gel electrophoresis and excision of the desired fragment size. The eluate (12 µl) was mixed with 2 µl loading dye (6x) and loaded together with 1 µg 50 bp DNA ladder (NEB) (separated by one lane) on a 2% agarose gel (50 ml, 1x TAE). The gel was run for 60 min at 100 V. Using a clean razor, fragments in the range of 200-380 bp were excised and subject to DNA extraction with the QIAquick Gel Extraction Kit. The gel piece was weighted and dissolved in 6 Volumes QG buffer and incubated at RT on a rotor until gel is dissolved completely (approx.10 min). The dissolved gel sample was mixed with 2 Volumes Isopropanol and loaded twice onto the column by centrifugation at RT for 30 sec at 13000xg. The flow-through was discarded and the column washed with 500 µl QG buffer. The column was washed twice with 750 µl PE buffer and centrifuged at RT for 45 sec at 13000xg. The empty column was centrifuged again at RT for 2 min at 13000xg and transferred to a new 1.5 ml tube. After residual PE buffer was aspirated from the purple ring, the column was air-dried for 1 min. To elute the bound DNA fragments, the column was incubated with 27 µl prewarmed EB buffer (50 °C) for 5 min at RT and subsequently centrifuged at RT for 1 min at 13000xg. 25 µl elute were recovered.

Following excess adapter removal and size selection, the adapter-modified DNA fragments were enriched by PCR according to Table 4-14 and Table 4-15.

Table 4-14 - Reaction composition for PCR enrichment (NGS library protocol I)

Component	Volume [µl]
Phusion HF reaction buffer (5x)	10.00
Betaine (5 M)	13.00
BiooPrimer1 (100 µM)	0.25
BiooPrimer2 (100 µM)	0.25
dNTPs (10 mM)	1.00
Phusion DNA polymerase	0.50
Total Volume	25.00

Table 4-15 – Cycling protocol for PCR enrichment (NGS library protocol I)

Cycle step	Temperature	Time	Number of cycles
Initial denaturation	98°C	2 min	1 x
Denaturation	98°C	30 sec	18 x
Annealing & Extension	65°C	30 sec	
Final denaturation	72°C	45 sec	1 x
Final extension	72°C	5 min	1 x
Cooling	4°C	hold	

The PCR-enriched DNA fragments were purified with AMPure XP beads (ration 1:1.8) as described above and eluted with 13 µl water.

To validate the successful enrichment of the size-selected DNA fragments, 1 µl of the library sample was analyzed on a Bioanalyzer DNA HS assay on a TapeStation High Sensitivity D1000 ScreenTape. The library samples were further processed for NGS data acquisition by the KFB (Biopark, Regensburg).

NGS library protocol II

The first library preparation step is an end repair to convert overhangs of double stranded DNA fragments into phosphorylated blunt ends. Following reaction set up was used:

Table 4-16 - Reaction composition for end repair (NGS library protocol II)

Component	Volume [µl]
DNA sample	40.00
H_2O	3.34
T4 DNA ligase buffer (10x)	5.00
dNTP mix (10 mM each)	1.00
T4 DNA polymerase	0.30
Klenow fragment	0.06
T4 Polynucleotide kinase	0.30
Total Volume	50.00

The reaction was incubated in a thermal cycler at 20°C for 30 min (without heated lid) and cleaned up with AMPure XP magnetic beads (equilibrated for 30 min at RT on a rotor) at a ratio 1:1.8.
90 µl beads were added to the end repair mix (50 µl), vortexed (speed 6) two times, inverted and vortexed again and then incubated at RT for 5 min. The tube was placed on a magnetic particle concentrator (MPC) for 2 min. After removal of the supernatant, the trapped beads were washed twice with 500 µl Ethanol (70%, prepared freshly) for 30 sec each. To remove residual Ethanol, the beads were air-dried at 37 °C for 2-5 min. The bound DNA fragments were eluted by incubation with 44.4 µl nuclease-free ddH2O at RT for 5 min. After centrifugation at RT for 1 sec at 13000xg, the tubes were placed on a MPC for 2 min and the supernatant was transferred to a new 1.5 ml tube for a second round of clean up.
For introduction of A-overhangs to the 3’ end of the phosphorylated blunt ends of DNA fragments following reaction was done:

Table 4-17 - Reaction composition for A-overhang introduction (NGS library protocol II)

Component	Volume [µl]
Blue buffer (10x)	5.00
dATP (10 mM)	0.10
Klenow (3’-5’ exo-)	0.50
Total Volume	5.60

The reaction mix was added to the end-repaired sample and incubated at 37°C for 30 min in a thermal cycler. After a cleanup with AMPure XP beads (1:1.8 ratio) as described above, adapters were ligated

to the end of the DNA fragments allowing later hybridization to an Illumina Next Generation Sequencing flow cell.

Table 4-18 - Reaction composition for adapter ligation (NGS library protocol II)

Component	Volume [µl]
Rapid Ligation buffer (2x)	16.00
Bioo NEXTflex adapter (1:2 dilution)	1.00
T4 DNA Ligase	1.00
Sample with A-overhangs	14.50
Total Volume	32.50

After incubation at 30°C for 10 min in a thermal cycler excess adapter removal was done by purification with AMPure XP beads using a ratio of 1:1.1.
Adapter-modified DNA fragments were then enriched by PCR according to Table 4-19 and Table 4-20.

Table 4-19 - Reaction composition for PCR enrichment (NGS library protocol II)

Component	Volume [µl]
Phusion HF reaction buffer (5x)	10.00
Betaine (5 M)	13.00
BiooPrimer1 (100 µM)	0.25
BiooPrimer2 (100 µM)	0.25
dNTPs (10 mM)	1.00
Phusion DNA polymerase	0.50
Total Volume	25.00

Table 4-20 - Cycling protocol for PCR enrichment (NGS library protocol II)

Cycle step	Temperature	Time	Number of cycles
Initial denaturation	98°C	2 min	1 x
Denaturation	98°C	30 sec	18 x
Annealing & Extension	65°C	30 sec	
Final denaturation	72°C	45 sec	1 x
Final extension	72°C	5 min	1 x
Cooling	4°C	hold	

The PCR-enriched DNA fragments were purified with AMPure XP beads (ration 1:1.1) as described above and eluted with 12 µl water, followed by size selection by gel electrophoresis and excision of the desired fragment size. The eluate (12 µl) was mixed with 2 µl loading dye (6x) and loaded together with 1 µg 50 bp DNA ladder (NEB) (separated by one lane) on a 2% agarose gel (50 ml, 1x TAE). The gel was run for 60 min at 100 V. Using a clean razor, fragments in the range of 200-380 bp were excised and subject to DNA extraction with the QIAquick Gel Extraction Kit. The gel piece was weighted and dissolved in 6 Volumes QG buffer and incubated at RT on a rotor until gel is dissolved completely (approx.10 min). The dissolved gel sample was mixed with 2 Volumes Isopropanol and loaded twice onto the column by centrifugation at RT for 30 sec at 13000xg. The flow-through was discarded and the column washed with 500 µl QG buffer. The column was washed twice with 750 µl PE buffer and centrifuged at RT for 45 sec at 13000xg. The empty column was centrifuged again at RT for 2 min at 13000xg and transferred to a new 1.5 ml tube. After residual PE buffer was aspirated from the purple ring, the column was air-dried for 1 min. To elute the bound DNA fragments, the column was incubated with 13 µl prewarmed EB buffer (50 °C) for 5 min at RT and subsequently centrifuged at RT for 1 min at 13000xg. To validate the successful enrichment of the size-selected DNA fragments, 1 µl of the library sample was analyzed on a Bioanalyzer DNA HS assay on a TapeStation High Sensitivity D1000 ScreenTape. The library samples were further processed for NGS data acquisition by the KFB (Biopark, Regensburg).

NGS library protocol III

This protocol is similar to NGS protocol II except for the PCR enrichment and size selection procedure. Relevant alterations to the established protocol are stated accordingly.

Following excess adapter removal, the adapter-modified DNA fragments were enriched by PCR in two rounds; round one was performed according to the following tables. In between the two PCR reactions, purification with AMPure XP beads (ration 1:1.8) and size selection using the Caliper LabChip approach (PerkinElmer) was performed.

Table 4-21 - Reaction composition for PCR enrichment (NGS library protocol III)

Component	Volume [µl]
Phusion HF reaction buffer (5x)	10.00
Betaine (5 M)	13.00
BiooPrimer1 (100 µM)	0.25
BiooPrimer2 (100 µM)	0.25
dNTPs (10 mM)	1.00
Phusion DNA polymerase	0.50
Total Volume	25.00

Table 4-22 – Cycling protocol for PCR enrichment (NGS library protocol III)

Cycle step	Temperature	Time	Number of cycles
Initial denaturation	98°C	2 min	1 x
Denaturation	98°C	30 sec	4 x
Annealing & Extension	65°C	30 sec	
Final denaturation	72°C	45 sec	1 x
Final extension	72°C	5 min	1 x
Cooling	4°C	hold	

Excess primers were removed by clean up with AMPure XP magnetic beads (equilibrated for 30 min at RT on a rotor) at a ratio 1:1.8. 90 µl beads were added to the PCR reaction mix (50 µl) and processed as before. The bound DNA fragments were eluted with 10 µl nuclease-free ddH2O and transferred to a new tube for the Caliper approach. Normally, 2 µl dye was added to the eluted fragments and fragments of 275 bp ± 15% (234-316 bp) using the LabChip XT DNA 700 Chip were recovered.

After adjusting the sample volume with nuclease-free ddH2O to 25 µl, the second round of PCR enrichment of the size selected DNA fragments was performed. For round two, reaction composition is the same as for round one (see Table 4-21), but the cycle number was increased (see Table 4-23).

Table 4-23 - Cycling protocol part 2 for PCR enrichment (NGS library protocol III)

Cycle step	Temperature	Time	Number of cycles
Initial denaturation	98°C	2 min	1 x
Denaturation	98°C	30 sec	12 x
Annealing & Extension	65°C	30 sec	
Final denaturation	72°C	45 sec	1 x
Final extension	72°C	5 min	1 x
Cooling	4°C	hold	

The PCR product was again depleted of excessive primers with AMPure XP beads (twice) as described above (1:1.1 ratio). Elution volume of the first round was 50 µl nuclease-free ddH2O, of the second 11 µl. 1 µl of the final eluate was run on a Bioanalyzer DNA HS Chip or on a TapeStation High Sensitivity D1000 ScreenTape to validate successful enrichment of size-selected DNA fragments.
The library samples were further processed for NGS data acquisition by the KFB (Biopark, Regensburg).

NGS library protocol IV

Using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) and NEBNext Multiplex Oligos for Illumina (NEB) the sample was first subject to end repair converting overhangs of dsDNA into phosphorylated blunt ends and subsequent dA-tailing according to the manufacturer´s instructions (see Table 4-24).

Table 4-24 - Reaction composition for end repair reaction (NGS library protocol IV)

Component	Volume [µl]
NEBNext Ultra II End Prep Enzyme Mix	3.00
NEBNext Ultra II End Prep Reaction Buffer	7.00
Fragmented DNA	50.00
Total Volume	60.00

Reaction was mixed by pipetting up and down at least 10 times and then incubated in a thermocycler (heated lid ≥ 75°C) for 30 minutes at 20°C followed by 30 minutes at 65°C and subsequent cooling to 4°C.

In the next step, adapters were ligated to the end of the DNA fragments allowing later hybridization to an Illumina Next Generation Sequencing flow cell (see Table 4-26). Adapter dilution depends on input concentration and is provided in the manual (see Table 4-25).

Table 4-25 - Recommended adapter working concentrations

Input	Adapter dilution	Working adapter concentration
1 µg – 101 ng	No dilution	15 µM
100 ng – 5 ng	1:10	1.5 µM
Less than 5 ng	1:25	0.6 µM

Table 4-26 - Reaction composition for adapter ligation (NGS library protocol IV)

Component	Volume [µl]
End Prep Reaction Mixture	60.00
NEBNext Ultra II Ligation Master Mix	60.00
NEBNext Ligation Enhancer	1.00
Adapter for Illumina	2.50
Total Volume	93.50

The reaction mix was incubated at 20 °C for 15 min in a thermal cycler with the heated lid off. After addition of 3 µl USER™ Enzyme the mix was incubated for 15 min at 37°C with the heated lid set to ≥ 47°C.

Subsequent size selection was done using AMPure XP magnetic beads and recommended condition for 200 bp inserts.
40 µl beads were added to the ligation mix (96.5 µl), mixed well by pipetting up and down 10 times and then incubated at RT for 5 min. The tube was placed on a magnetic particle concentrator (MPC) for 2 min. The supernatant containing the DNA fragments of interest is carefully transferred to a new tube and beads with the unwanted large fragments are discarded.
Another 20 µl of AMPure XP beads are added to the supernatant and mixed well by pipetting up and down 10 times. After separation of the beads on a magnetic particle concentrator, supernatant is removed carefully and trapped beads were washed twice with 200 µl Ethanol for 30 seconds (80%, prepared freshly). To remove residual Ethanol, the beads were air-dried for 2-5 min. The bound DNA fragments were eluted by incubation with 17 µl of 0.1 x TE buffer at RT for 5 min. After short centrifugation, the tubes were placed on a MPC for 2 min and 15 µl of the supernatant was transferred to a new 1.5 ml tube for PCR amplification (see Table 4-27 and Table 4-28).

Table 4-27 - Reaction composition for PCR enrichment (NGS library protocol IV)

Component	Volume [µl]
Adaptor ligated DNA fragments	15.00
Index Primer / i7 Primer*	2.00
Universal PCR Primer / i5 Primer	2.00
NEBNext Ultra II Q5 Master Mix	25.00
Sterile H2O	6.00
Total Volume	50.00

* Index Primer are provided in the NEBNext Multiplex Oligos for Illumina (#E7335, #E7500)

Table 4-28 - Cycling protocol for PCR enrichment (NGS library protocol IV)

Cycle step	Temperature	Time	Number of cycles
Initial denaturation	98°C	30 sec	1 x
Denaturation	98°C	10 sec	10-15 x
Annealing & Extension	65°C	75 sec	
Final extension	65°C	5 min	1 x
Cooling	4°C	hold	

The PCR product was again depleted of excessive primers with AMPure XP beads as described above (1:0.9 ratio). Elution volume was 17 µl 0.1 x TE. 1 µl of the final eluate was run on a Bioanalyzer DNA HS Chip or on a TapeStation High Sensitivity D1000 ScreenTape to validate successful enrichment of size-selected DNA fragments.
The library samples were further processed for NGS data acquisition by the KFB (Biopark, Regensburg) or BSF (Vienna).

4.2.1.19 Next generation sequencing on the Illumina platform

High-throughput DNA sequencing with the Sequencing-by-Synthesis (SBS) Technology (Illumina) is based on iterative sequencing using reversible terminator chemistry. Labeling the modified nucleotides with different fluorescent dyes permits parallel readout of all immobilized library molecules. The sequencing-by-synthesis workflow is illustrated in Figure 4-1. The advantage to Sanger sequencing is the massive parallel sequencing of millions of DNA fragments. Due to the high sequencing depth, up to 4 billion reads per lane, multiplexing of samples by using barcoded libraries dramatically reduces costs. Samples analyzed by NGS in this work included ChIP, MCIp and targeted bisulfite enrichment.

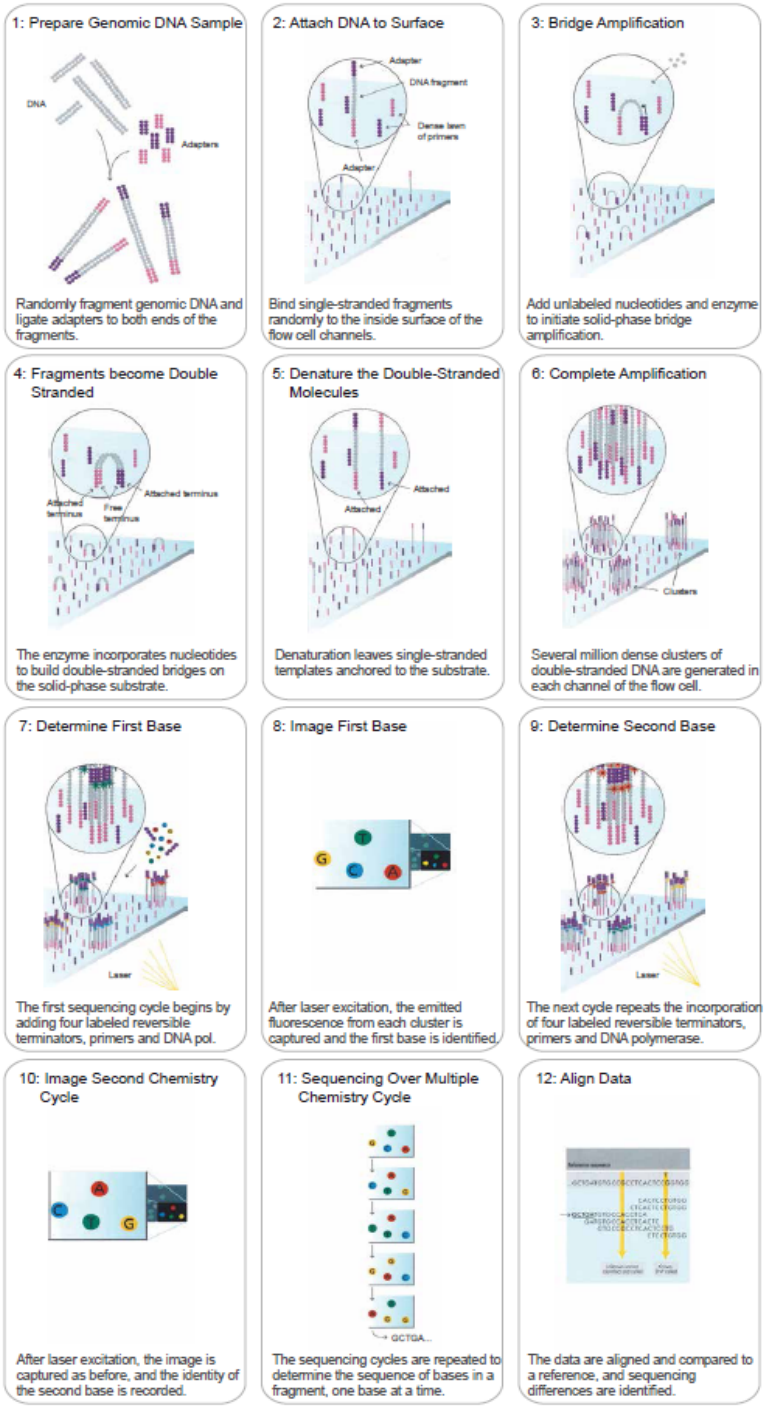


Figure 4-1 - Next generation sequencing technology

1, 2 – Single strands of the NGS library generated according to the NGS library protocols listed in 4.2.1.18 bind to complementary adapters. 3-6 – Clusters with more than 100 copies of the starting library molecule are generated by bridge amplification. 7 – Sequencing primer anneal to the adapter sequence of the cluster molecule and 3'-terminated, fluorescent-labeled nucleotides are incorporated at the 3' end of the sequencing primer by a modified polymerase (due to the 3' terminating modification, only one nucleotide per cycle is incorporated). 8 – Using different lasers for excitation and filters, an image is generated of the attached fluorophores, which reflects the incorporation positions of the individual nucleotides. 9-11 – Following each imaging cycle, the 3' terminators as well as the fluorescent labels are chemically removed permitting for a new cycle of incorporation and imaging. The read length for all clusters is determined by the numbers of repeated incorporation and imaging cycles. Barcodes of pooled libraries are encoded in the adapter sequence and assessed accordingly. Different sequences of left and right adapter allows for paired-end sequencing. 12 – Following sequencing, the Illumina base caller analyzes the images. The collected sequence data are aligned and compared with a reference sequence. (Taken and modified from Illumina, www.illumina.com)

4.2.1.20 Targeted bisulfite sequencing

Investigation of DNA methylation in selected regulatory regions was done using hybridization-based target enrichment (NimbleGen SeqCap® Epi Enrichment System) followed by next-generation sequencing. These customer chosen regions are active in human AML and healthy monocytic cells and comprise 71K regions corresponding to 86 Mb of the human genome.
In this approach pools of biotinylated oligonucleotide probes specific for target regions of interests are hybridized to a sequencing library in solution and pulled down by streptavidin-coated magnetic beads.
The first step of the protocol requires fragmentation of the input DNA using the Covaris S2 system (see 4.2.1.3). 400 ng of input DNA were diluted with 10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA to a final volume of 53 µl and sonicated in a microTUBE AFA snap-cap (6 mm x 16 mm). Following settings were used 4 times per sample whereas in between sample was centrifuged shortly.

Table 4-29 - Parameters for focused ultrasonication with Covaris for targeted Bis-Seq

Parameters	Settings
Intensity	5
Duty Cycle	10%
Cycles per burst	200
Treatment time [s]	120
Temperature [°C]	7
Water Level S2	12
Mode	Frequency sweeping

Dilute 1 µl of fragmented DNA 1:5 with 10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA and run sample on a Bioanalyzer High Sensitivity DNA 1000 chip to check average fragment size.

The following step comprises end repair of the fragmented DNA to convert overhangs of double stranded DNA fragments into phosphorylated blunt ends and was set up as followed:

Table 4-30 - Reaction composition for end repair reaction (Targeted bisulfite sequencing)

Reagent	Volume
Water	8 µl
10x KAPA End Repair Buffer	7 µl
KAPA End Repair Enzyme	5 µl
Fragmented DNA	50 µl
Total	70 µl

The reaction mix was incubated at 20°C for 30 min in a thermocycler and subsequently purified using AMPure XP beads (ratio 1:1.7). 70 µl beads were added to the end repair mix and incubated for 15 min at room temperature. After separation on a magnetic particle concentrator remove and discard supernatant and wash beads twice with 200 µl of 80% ethanol for 30 sec. Beads were air-dried at 37°C for 3-5 min to remove residual ethanol and A-tailing master mix for elution of beads was prepared(see Table 4-31).

Table 4-31 - Reaction composition for A-tailing reaction (Targeted bisulfite sequencing)

Reagent	Volume
Water	42 µl
10x KAPA A-Tailing Buffer	7 µl
KAPA A-Tailing Enzyme	5 µl
Total	50 µl

A-tailing master mix was added to the beads, resuspend thoroughly by pipetting up and down and incubated at 30°C for 30 min. After this, 90 µl of PEG/NaCl SPRI solution (adapted to room temperature) were added to each 50 µl A-tailing reaction with beads. Solution was mixed thoroughly, incubated for 15 min at room temperature and placed on a magnet to capture beads. Supernatant was removed and discarded and trapped beads were washed twice with 200 µl of 80% ethanol for 30 seconds. Beads were dried at room temperature to remove residual ethanol.
Resuspend beads with the Ligation Master Mix (see Table 4-32) and incubate at 20°C for 15 min in a thermocycler.

Table 4-32 - Reaction composition for ligation reaction (Targeted bisulfite sequencing)

Reagent	Volume
Water	30 µl
5x KAPA Ligation Buffer	10 µl
KAPA T4 DNA Ligase	5 µl
Indexed Adapter	5 µl
Total	50 µl

In the next step, 50 µl PEG/NaCl SPRI solution was added to each 50 µl ligation reaction with beads and incubated for 15 min at room temperature. Following bead procedure was done as described above. Beads were resuspended in 100 µl elution buffer (10 mM Tris-HCl, pH 8.0) and incubated for 2 min at room temperature to allow the DNA to elute of the beads. Afterwards 60 µl PEG/NaCl SPRI solution was added and mixed thoroughly by pipetting up and down. After incubation for 15 min at room temperature beads were put on a magnetic particle concentrator and 155 µl of the supernatant were transferred to a new tube. Next, 20 µl of AMPure XP beads were added to the supernatant and again incubated for 15 min at room temperature. After concentration of magnetic particles, supernatant is removed and discarded and trapped beads were washed twice with 200 µl 80% ethanol. Beads were then air-dried at room temperature and eluted in 25 µl elution buffer (10 mM Tris-HCl, pH 8.0).

To differentiate between unmethylated and methylated cytosines, 20 µl of the eluate were treated with sodium bisulfite using the EZ DNA Methylation-Lightning Kit according to the manufacturer´s instructions. Cytosine derivates undergo reversible reactions with sodium bisulfite, yielding a 5,6-Dihydro-6-sulfonate, which deaminates under alkaline conditions leaving uracil whereas 5-methylcytosine is not affected by this reaction. After PCR amplification, unmethylated cytosine appears as thymine in contrast to 5-methylcytosine, which remains a cytosine[174]. The bisulfite-treated DNA was recovered by adding 21.5 µl of PCR-grade water directly to the column matrix and centrifugation for 30 sec at 10.000 x g.

For amplification of the bisulfite-converted DNA the following LM-PCR Master Mix was prepared on ice (see Table 4-33).

Table 4-33 - Reaction composition for PCR enrichment (Targeted bisulfite sequencing)

Reagent	Volume
KAPA HiFi HotStart Uracil + Ready Mix (2x)	25 µl
PCR grade water	2 µl
Pre LM-PCR Oligos 1 & 2 (5 µM)	3 µl
Total	30 µl

This master mix was pipetted to 20 µl of bisulfite-converted DNA and incubated in a thermocycler using the Pre-Capture LM-PCR program (see Table 4-34).

Table 4-34 - Cycling protocol for PCR enrichment (Targeted bisulfite sequencing)

Cycle step	Temperature	Time	Number of cycles
1	95°C	2 min	1 x
2	98°C	30 sec	12 x
3	60°C	30 sec	
4	72°C	4 min	
5	72°C	10 min	1 x
6	4°C	hold	

PCR reaction was cleaned up with SeqCap EZ Purification beads by adding 90 µl to 50 µl amplified sample library and incubation for 15 min at room temperature. After concentrating the beads on a magnetic rack, supernatant was removed and discarded. Trapped beads were then washed twice with 200 µl freshly prepared 80 % ethanol and air dried at room temperature to remove all residual ethanol. DNA was recovered by resuspending the beads with 52 µl of PCR-grade water and incubation for 2 min at room temperature. After putting the tube in a magnetic particle concentrator 50 µl of the supernatant were transferred into a new 1.5 ml tube. Next, A_{260}/A_{280} ratio was measured on a NanoDrop spectrophotometer and checked via automated electrophoresis using the Bioanalyzer instrument. The average fragment size should be between 150 and 500 bp and library yield at least 1 µg.

The next step comprised hybridization of the sample library to a pool of biotinylated oligonucleotide probes to enrich for target regions of interest. Here 10 µl of Bisulfite Capture Enhancer and 1 µg of amplified bisulfite-converted DNA were put in a 1.5 ml tube. 1000 pmol (1 µl) of SeqCap HE Universal Oligo 1 and 1000 pmol (1 µl) of the appropriate SeqCap HE Index Oligo were added to this mix.

Tube was closed and a hole was made in the top of the tube´s cap with an 18 – 20 gauge needle to suppress contamination in the DNA vacuum concentrator.

Multiplexed DNA sample library pool was dried in a DNA vacuum concentrator at high heat (> 60°C) and the hole was covered with a sticker. The following components were added to the reaction (see Table 4-35), mixed and incubated for 10 min at 95°C.

Table 4-35 - Reaction composition for hybridization reaction (Targeted bisulfite sequencing)

Reagent	Volume
2x SC Hybridization Buffer	7.5 µl
SC Hybridization Component A	3 µl

Transfer 4.5 µl SeqCap Epi Library aliquot to the reaction mix and vortex for 3 sec followed by centrifugation for 10 sec at full speed. The mixture was incubated in a thermocycler at 47°C for 72 h. Recovery of captured sample was done using 100 µl Capture Beads per reaction which have to be washed twice with 200 µl 1x Bead Wash Buffer before using. After resuspending the Capture Beads in 100 µl 1x Bead Wash Buffer and addition of the hybridization sample mixture is incubated at 47°C for 45 min. Next, beads were washed according to the table below (Table 4-36).

Table 4-36 - Guide for washing beads (Targeted bisulfite sequencing)

Wash Buffer	Wash Volume	Mixture Time	Time in Water Bath	Temperature
1x Wash Buffer I	100 µl	Vortex for 10 sec		47°C
1x Stringent Wash Buffer	200 µl	Pipette up and down 10 times	5 min	47°C
1x Stringent Wash Buffer	200 µl	Pipette up and down 10 times	5 min	47°C
1x Wash Buffer I	200 µl	Vortex for 2 min	-	RT
1x Wash Buffer II	200 µl	Vortex for 1 min	-	RT
1x Wash Buffer III	200 µl	Vortex for 30 sec		RT

Recovering of captured sample was done by adding 50 µl PCR grade water to the tube and subsequently these captured fragments were amplified by adding the Post-Capture LM-PCR Master Mix (see table x) and incubation as described in the following table (Table 4-37).

Table 4-37 - Reaction composition for Post-Capture LM-PCR (Targeted bisulfite sequencing)

Reagent	Volume
KAPA HiFi HotStart Ready Mix	50 µl
Post-LM-PCR Oligos 1 & 2 (5 µM)	10 µl
Total	60 µl

Table 4-38 - Cycling protocol for Post-Capture LM-PCR (Targeted bisulfite sequencing)

Cycle step	Temperature	Time	Number of cycles
1	98°C	45 sec	1 x
2	98°C	15 sec	16 x
3	60°C	30 sec	
4	72°C	30 sec	
5	72°C	1 min	1 x
6	4°C	hold	

Purification of PCR reaction mixture was done with 180 µl (ratio 1:1.8) SeqCap EZ Purification Beads (equilibrated for 30 min at RT) like described above. The bound DNA fragments were eluted with 52 µl of PCR-grade water and 50 µl transferred to a new tube.
Quality and quantity of the library samples were checked with the Bioanalyzer device and NanoDrop spectrophotometer, respectively, followed by further processing of the library samples for NGS data acquisition by the KFB (Biopark, Regensburg).

Selection of the myeloid regulome

To design a region set for targeted enrichment that would capture as many regulatory regions that may be relevant for myeloid cell biology and acute myeloid leukemias, we used both publicly available data sets as well as own data. From the FANTOM5 CAGE atlas data sets we used both transcriptionally active promoters and enhancers from available myeloid cell types including cell lines (Kasumi-3, KG-1, HYT-1, Kasumi-1, Kasumi-6, NKM-1, HL60, FKH-1, HNT-34, EoL-1, EoL-3, NOMO-1, P31_FUJ, THP-1, U-937_DE-4, EEB, F-36E, F-36P, M-MOK, MKPL-1) and primary cells (CD14+ monocytes (0h and 2h), monocyte subsets (classical, intermediate and non-classical), CD34+ progenitors (blood and bone marrow), CD133+ stem cells (bone marrow and cord blood), common myeloid progenitors (CMP), monocyte-derived immature dendritic cells, plasmacytoid dendritic cells, eosinophils, granulocyte macrophage progenitor, mast cells, neutrophils). For promoters we counted power law-normalized read counts for the 180K CAGE promoter set across all samples and kept all promoter regions (-300 to +50) with a normalized read count of at least four, resulting in 54K promoters. From the atlas of bidirectionally expressed enhancers, we kept those with a score above three with detectable expression in any of the samples and within a lower size limit of 100 bp. Enhancers were symmetrically extended to at least 400bp. The remaining myeloid enhancer set comprised 15K enhancers. For CD34 cells we used available ENCODE data, including DNase hypersensitivity (DHS) sequencing, ChIP sequencing for H3K4me3 and H3K4me1 to determine active and poised regulatory elements. H3K4me3 peak regions were defined using the findPeaks script provided in the HOMER package and the following options: -region -size 250 -L 0 -F 5 -minDist 350 -fdr 0.00001 -ntagThreshold 10, resulting in 24K H3K4me3 peak regions. Peaks in DHS regions were called using findPeaks and options -fragLength 1 -region -size 10 and extended by 50 bp on each side. Strongest DHS peaks (peak score > 40) were selected (14K CD34+ open chromatin regions). We also identified all DHS regions overlapping with H3K4me1 peaks, extended them by 100 bp on either side and merged them to a resulting set of 79K DHS regions with proximal H3K4me1 enrichment. Lastly, we identified H3K27ac peak regions in normal primary cells (monocytes, neutrophils, CD34+ progenitors) as well as data from 44 AML patients using findPeaks and -region -size 250 -L 0 -F 5 -minDist 350 -fdr 0.00001 -ntagThreshold 20 generating a merged peak set of 43K regions. The resulting six region sets (myeloid promotors and enhancers from CAGE, CD34+ H3K4me3 peaks, CD34+ wide open chromatin, CD34+ DHS regions overlapping H3K4me1 and myeloid H3K27ac peaks) were merged and further extended by any overlapping DHS region in CD34+ cells. Regions that were previously determined as problematic in ChIP seq experiments were excluded, remaining regions were further extended by 50bp on each side, and merged within a

distance of 400 bp. The final myeloid regulome comprised 71K regions and 86 Mb of the human genome.

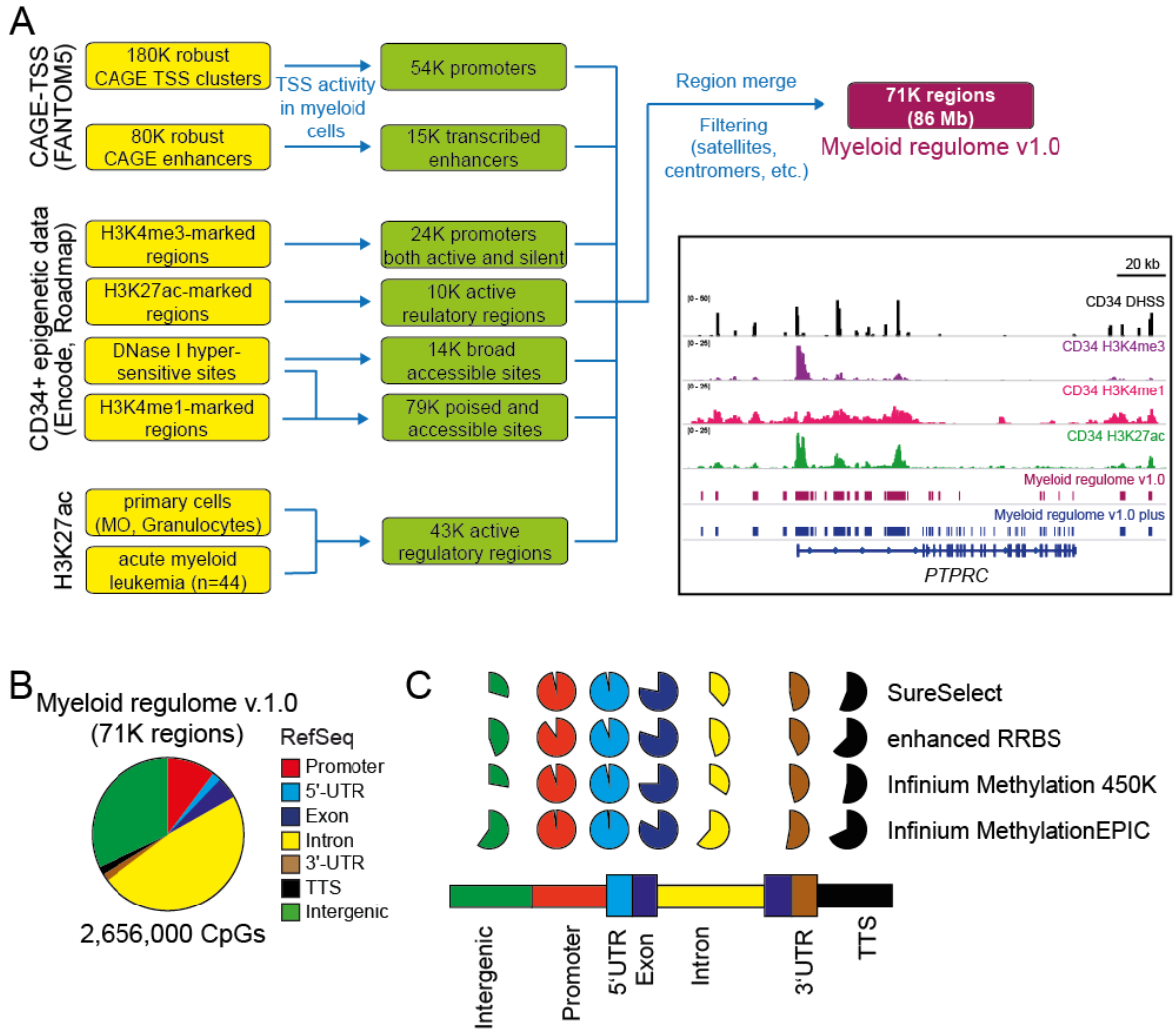


Figure 4-2 - Definition of the active myeloid regulome for targeted enrichment

(A) Workflow for the selection of regulatory regions that are active or poised in myeloid cell types, including monocytes, granulocytes, CD34+ hematopoietic progenitor cells, and acute myeloid leukemia cells from 44 patients. For targeted enrichment on the genomic DNA level, the selected myeloid regulatory regions were merged with the human exome (myeloid regulome v1.0 plus) (B) Pie chart showing the genomic distribution of the selected myeloid regulome. (C). Comparison of coverage with currently available methods, including Agilent's targeted Methylseq Kit (SureSelect), Illumina´s 450K and EPIC (860K) methylation bead arrays, and a recently described enhanced RRBS protocol (Akalin et al., 2012). While existing methods cover the large majority of promoter and 5'UTR regions that are likely relevant for myeloid cells, the coverage of distal myeloid regulatory sites (intron/intergenic) is relatively poor, suggesting that the custom design will cover myeloid enhancers more comprehensively.

4.2.2 Preparation and analysis of RNA

4.2.2.1 Isolation of total RNA

Total RNA was isolated using the RNeasy Midi, mini or Micro Kit (Qiagen) depending on the available number of cells according to the manufacturer's instructions. To remove potential DNA contaminations, on-column DNA digestion with the RNase-free DNase Set (Qiagen) was integrated according to the protocol. RNA concentration was then determined with the NanoDrop spectrophotometer and quality assessed with the Agilent TapeStation (Agilent Technologies) according to the manufacturer's instructions.

4.2.2.2 Reverse transcription quantitative real-time PCR (RT-qPCR)

To analyse transcription levels, total RNA was transcribed into complementary DNA (cDNA) with the M-MLV RT (H- point mutant) enzyme (Promega) combined with random decamers (Promega) to prime cDNA synthesis. Individual expression levels were subsequently assessed by quantitative PCR (see section 4.2.1.13) on cDNA level. Reaction setup is depicted in the following table (Table 4-39).

Table 4-39 - Reaction composition for RT-qPCR

Component	Volume
RNA (1 µg)	x µl
Random decamers (50 µM)	1 µl
dNTPs (10 mM)	1 µl
Water	(15 – x) µl
Total	15 µl

Incubate reaction mix for 5 min at 65°C in a thermal cycler to dissolve secondary structures followed by immediate incubation on ice for 1 min. After centrifugation, 4 µl M-MLV Buffer (5x) were added, mixed and incubated for 2 min at 42°C.
Reverse transcription started upon addition of 1 µl M-MLV Reverse transcriptase at 42 °C for 50 min and was stopped by heat inactivation of the enzyme (15 min, 70 °C).

4.2.3 Analyses of NGS data sets

Bioinformatic analyses of next generation sequencing (NGS) data were performed on an Apple Mac Pro workstation with the indicated programs and versions. Raw read files were delivered in fastq or bam format from the sequencing facilities.

4.2.3.1 Targeted bisulfite sequencing data

DNA methylation of the myeloid regulome either in bone marrow or blood tissues from adult MDS patients was analyzed using the targeted bisulfite sequencing approach (NimbleGen). The following script is representative for all analyzed samples, but will be illustrated for one example. If not stated otherwise in the command line, default settings were applied. For this bisulfite sequencing data, special programs were required, e.g. bsmap or metilene.

```
### Conversion of unaligned bam file into fastq file using Picard tools
# converting a SAMPLE.bam file into paired end SAMPLE_r1.fastq and SAMPLE_r2.fastq
java -jar $PICARD SamToFastq I= /path-to-raw-data-archive/MDS53_ATGTCA_S16114.bam \
F= /path-to-read1-fastq-file/MDS53_timepoint1_ATGTCA_S16114_r1.fastq \
F2= /path-to-read2-fastq-file/MDS53_ timepoint1_ATGTCA_S16114_r2.fastq

### Quality assessment with Fastqc

### Alignment of raw reads to GRCh38 plus Lambda reference genome with BSMAP
bsmap -r 0 -s 16 -n 1 -a \
/path-to-fastq-file/ MDS53_ timepoint1_ATGTCA_S16114_r1.fastq -b \
/path-to-fastq-file/ MDS53_ timepoint1_ATGTCA_S16114_r2.fastq -d \
/path-to-genome-file/GRCh38.plusLambda.genome.fa -p <#CPU> -o \
/path-to-output-folder/MDS53_ timepoint1.sam

### convert sam to bam file format
java -Xmx8G -Xms8G -jar $PICARD AddOrReplaceReadGroups \
VALIDATION_STRINGENCY=LENIENT INPUT=/path-to-sam-file/MDS53_ timepoint1.sam \
OUTPUT=/output-path-for-bam-file/MDS53_timepoint1.bam CREATE_INDEX=TRUE \
RGID=MDS53_ timepoint1 RGLB=MDS53_timepoint1 RGPL=illumina RGSM=MDS53_timepoint1\
RGPU= ATGTCA
### split bam file into top and bottom strands using bamtools
cd /path-for-output-folder/ bamtools split -tag ZS -in MDS53_timepoint1.bam

### merge strand bam files
bamtools merge -in MDS53_timepoint1.TAG_ZS_++.bam -in \
MDS53_timepoint1.TAG_ZS_+-.bam -out MDS53_timepoint1.top.bam; \
bamtools merge -in MDS53_timepoint1.TAG_ZS_-+.bam -in \
MDS53_timepoint1.TAG_ZS_--.bam -out MDS53_timepoint1.bottom.bam
```

```
### sort bam files
samtools sort MDS53_timepoint1.top.bam MDS53_timepoint1.top.sorted
samtools sort MDS53_timepoint1.bottom.bam MDS53_timepoint1.bottom.sorted

### remove duplicates
java -Xmx8G -Xms8G -jar $PICARD MarkDuplicates VALIDATION_STRINGENCY=LENIENT \
INPUT= MDS53_timepoint1.bottom.sorted.bam UTPUT=MDS53_timepoint1.bottom.rmdups.bam\
METRICS_FILE=MDS53_timepoint1.bottom.rmdups_metrics.txt REMOVE_DUPLICATES=true \
ASSUME_SORTED=true CREATE_INDEX=true

### Merge duplicate removed bam files
cd /path-to-output-folder/bamtools merge -in MDS53_timepoint1.top.rmdups.bam \
-in MDS53_timepoint1.rmdups.bam -out MDS53_timepoint1.rmdups.bam

### filter and clip overhang (removing overlapping reads)
bamtools filter -isMapped true -isPaired true -isProperPair true –forceCompression\
-in /path-to-input-file/MDS53_timepoint1.rmdups.bam \
-out /path-to-output-file/MDS53_timepoint1.filtered.bam

bam clipOverlap --stats --in \
/path-to-filtered-bam-file/MDS53_timepoint1.filtered.bam \
--out /path-to-output-file/MDS53_timepoint1.clipped.bam

samtools sort -f /path-to-input-file/MDS53_timepoint1.bam \
/path-to-output-file/MDS53_timepoint1.sorted.bam

### indexing bam files
samtools index /path-to-input-file/MDS53_timepoint1.sorted.bam
samtools index /path-to-input-file/MDS53_timepoint1.clipped.bam

### basic mapping metrics
java -Xmx8G -Xms8G -jar $PICARD CollectAlignmentSummaryMetrics \
METRIC_ACCUMULATION_LEVEL=ALL_READS \
INPUT=/path-to-input-file/MDS53_timepoint1.sorted.bam \
OUTPUT=/path-to-output-file/MDS53_timepoint1_picard_alignment_metrics.txt \
REFERENCE_SEQUENCE=/path-to-genome-file/GRCh38.plusLambda.genome.fa \
VALIDATION_STRINGENCY=LENIENT
samtools view -H /path-to-input-file/MDS53_timepoint1.sorted.bam \
> /path-for-output-file/MDS53_timepoint1_bam_header.txt

### Hybrid Selection (HS) Analysis Metrics
java -Xmx16G -Xms16G -jar $PICARD CalculateHsMetrics \
BAIT_INTERVALS=/path-to-capture-design-file/Regulome_bait_intervals.txt \
TARGET_INTERVALS=/path-to-capture-design/Regulome_target_intervals.txt \
INPUT=/path-to-input-file/MDS53_timepoint1.clipped.bam \
```

```
OUTPUT=/path-for-output-file/MDS53_timepoint1_picard_hs_metrics.txt \
METRIC_ACCUMULATION_LEVEL=ALL_READS \
REFERENCE_SEQUENCE=/path-to-genome-file/GRCh38.plusLambda.genome.fa \
VALIDATION_STRINGENCY=LENIENT TMP_DIR=.

### Estimate insert size
java -Xmx16g -jar $PICARD CollectInsertSizeMetrics VALIDATION_STRINGENCY=LENIENT \
HISTOGRAM_FILE=/path-to-input-file/MDS53_timepoint1_picard_insert_size_plot.pdf \
INPUT=/path-to-input-file/MDS53_timepoint1.filtered.bam \
OUTPUT=/path-for-output-file/MDS53_timepoint1_picard_insert_size_metrics.txt

### Count on-target reads and calculate depth of coverage
bedtools intersect -bed -abam /path-to-bam-file/MDS53_timepoint1.sorted.bam -b \
/path-to-capture-design/Regulome_primary_targets.bed | wc -l \
> /path-for-output-file/MDS53_timepoint1.on_primary_targets.txt ;
bedtools intersect -bed -abam /path-to-input-file/MDS53_timepoint1.sorted.bam -b \
/path-to-capture-design/Regulome_capture_targets.bed | wc -l \
> /path-for-output-file/MDS53_timepoint1.on_capture_targets.txt ;
java -Xmx8G -Xms8G -jar $GATK -T DepthOfCoverage -R \
/path-to-genome-file/GRCh38.plusLambda.genome.fa -I \
/path-to-input-file/MDS53_timepoint1.clipped.bam -o \
/path-for-output-file/MDS53_timepoint1_gatk_capture_target_coverage \
-L /path-to-capture-design/Regulome_capture_targets.bed -ct 1 -ct 10 -ct 20

### Determine methylation percentage using BSMAP
python /path-to-bsmap-software/methratio.py -d \
/path-to-genome-file/GRCh38.plusLambda.genome.fa -m 1 -z -i skip \
-o /path-for-output-file/MDS53_timepoint1.methylation_results.txt \
/path-to-bam-file/MDS53_timepoint1.clipped.bam

### Determine bisulfite conversion efficiency using BSMAP
python /path-to-bsmap-software/methratio.py -d \
/path-to-genome-file/GRCh38.plusLambda.genome.fa -m 1 -z -i skip --chr=lambda -o \
/path-for-output-file/MDS53_timepoint1.NC_001416.methylation_results.txt \
/path-to-bam-file/MDS53_timepoint1.bam

### Combined SNP/methylation calling using BisSNP
java -Xmx20g -jar $BisSNP -R /path-to-genome-file/GRCh38.plusLambda.genome.fa \
-I /path-to-input-file/MDS53_timepoint1.clipped.bam -T BisulfiteCountCovariates \
-cov ReadGroupCovariate -cov QualityScoreCovariate -cov CycleCovariate –recalFile \
/path-for-output-file/MDS53_timepoint1.recalFile_before.csv -nt 12 -knownSites \
/path-to-genome-file/human_9606_b142_GRCh38_reformated.vcf

java -Xmx20g -jar $BisSNP -R /path-to-genome-file/GRCh38.plusLambda.genome.fa
-I /path-to-input-file/EPI_MDS_Goett_timepointA.clipped.bam –o \
```

```
/path-for-output-file/ MDS53_timepoint1.recal.bam -T BisulfiteTableRecalibration \
-recalFile /path-to-csv-file/EPI_MDS_Goett_timepointA.recalFile_before.csv -maxQ 40

java -Xmx20g -jar $BisSNP -R /path-to-genome-file/GRCh38.plusLambda.genome.fa -I \
/path-to-input-file/MDS53_timepoint1.recal.bam -T BisulfiteGenotyper -D \
/path-to-genome-file/human_9606_b142_GRCh38_reformated.vcf -vfn1 \
/path-for-output-file1/EPI_MDS_Goett_timepointA.cpg.raw.vcf -vfn2 \
/path-for-output-file2/EPI_MDS_Goett_timepointA.snp.raw.vcf -L \
/path-to-capture-design/Regulome_capture_targets.bed -stand_call_conf 20 \
-stand_emit_conf 0 -mmq 30 -mbq 0 -nt 12

sortByRefAndCor.pl --k 1 --c 2 /path-to-input-file/MDS53_timepoint1.snp.raw.vcf \
/path-to-genome-file/GRCh38.plusLambda.genome.fa.fai \
> /path-for-output-file/MDS53_timepoint1.snp.raw.sorted.vcf

sortByRefAndCor.pl --k 1 --c 2 /path-to-input-file/MDS53_timepoint1.cpg.raw.vcf \
/path-to-genome-file/GRCh38.plusLambda.genome.fa.fai \
> /path-for-output-file/MDS53_timepoint1.cpg.raw.sorted.vcf

myVcf2bedGraph.pl /path-to-input-file/MDS53_timepoint1.cpg.raw.sorted.vcf CG

vcf2coverage.pl /path-to-input-file/MDS53_timepoint1.cpg.raw.sorted.vcf CG

myFixBedGraph.pl /path-to-input-file/MDS53_timepoint1.cpg.raw.sorted.CG.bedgraph \
> /path-for-output-file/MDS53_timepoint1.cpg.cleared.CG.bedgraph

myFixBedGraph.pl \
/path-to-input-file/MDS53_timepoint1.cpg.raw.sorted.CG.coverage.bedgraph \
> /path-for-output-file/MDS53_timepoint1.cpg.cleared.CG.coverage.bedgraph

bedGraphToBigWig /path-to-input-file/MDS53_timepoint1.cpg.cleared.CG.bedgraph \
/path-to-genome-file/GRCh38.chromosome.sizes \
/path-for-output-file/MDS53_timepoint1.cpg.CG.bigWig

bedGraphToBigWig
/path-to-input-file/MDS53_timepoint1.cpg.cleared.CG.coverage.bedgraph \
/path-to-genome-file/GRCh38.chromosome.sizes \
/path-for-output-file/MDS53_timepoint1.cpg.CG.coverage.bigWig

### Convert results for HOMER
myVcf2allC.pl /path-to-input-file/MDS53_timepoint1.cpg.raw.sorted.vcf ; sed -f \
/path-to-genome-file/GRCh38.hg38.sed < \
/path-to-txt-file/MDS53_timepoint1.cpg.raw.sorted.allC.txt > \
/path-for-output-file/MDS53_timepoint1.cpg.raw.sorted.allC.hg38.txt
```

```
makeTagDirectory /path-to-tag-directory/MDS53_timepoint1 -format allC \
-minCounts 5 -genome hg38 -checkGC \
/path-for-output-file/MDS53_timepoint1.cpg.raw.sorted.allC.hg38.txt

makeUCSCfile /path-to-tag-directory/MDS53_timepoint1 -style unmethylated \
-bigWig /path-to-hg38-genome-file/chrom.sizes -o \
/path-for-output-file/MDS53_timepoint1.unmethylated.bigwig ;
makeUCSCfile /path-to-tag-directory/ MDS53_timepoint1 -style methylated \
-bigWig /path-to-hg38-genome-file/chrom.sizes -o \
/path-for-output-file/MDS53_timepoint1.methylated.bigwig

### Creating scatter plots using methylKit
library(methylKit)
setwd("/path-to-working-directory /")
file.list = list("/path-to-input-
file/CD34_589.cpg.methylKitFormat.hg38.txt","/path-to-input-
file/MDS53_timepoint1.cpg.methylKitFormat.hg38.txt","/path-to-input-
file/MDS53_timepoint2.cpg.methylKitFormat.hg38.txt","/path-to-input-file/
MDS53_timepoint3.cpg.methylKitFormat.hg38.txt","/path-to-input-
file/MDS53_timepoint4.cpg.methylKitFormat.hg38.txt","/Vpath-to-input-
file/MDS53_timepoint5.cpg.methylKitFormat.hg38.txt")
sample.list = list("CD34","P53_1","P53_2","P53_3","P53_4","P53_5")
methData <-read(
file.list,
sample.id=sample.list,
treatment=c(0,1,1,1,1,1),
assembly="hg38",\
resolution="base",
context="CpG",
)
filtered.methData=filterByCoverage(methData,lo.count=5,lo.perc=NULL,hi.count=NULL,\
hi.perc=NULL)
meth=unite(filtered.methData,destrand=TRUE)
png(filename="correlationMDS53.png", height=4000, width=4000)
getCorrelation(meth,plot=T)
dev.off()
```

The R package "methylKit"[175] was used for several bioinformatic analyses, including principal component analysis (PCA). Vcf input files obtained with BisSNP need to be converted into a special methylKit format.

```
### Converting vcf file into methylkit format file
myVcf2methylKit.pl /path-to-input-file/EPI_THP1_1_749.cpg.raw.sorted.vcf CG ; \
sed -f /path-to-genome-indices/GRCh38.hg38.sed < \
/path-of-methylkit-format-file/EPI_THP1_1_749.cpg.raw.sorted.methylKitFormat.txt \
> /path-for-output-file/EPI_THP1_1_749.cpg.methylKitFormat.hg38.txt ;
```

```
### Correlation and principal component analysis (PCA)
library(methylKit)
setwd("/path-to-working-directory /")
file.list = list("<all desired methylKit input files in a comma separated list>”)
sample.list = list("<comma separated list of corresponding sample names>")
methData <-read(
file.list,
sample.id=sample.list,
treatment=c(<vector of numbers \ indicating the different patient samples>),
assembly="hg38", \
resolution="base",
context="CpG",
)
filtered.methData=filterByCoverage(methData,lo.count=5,lo.perc=NULL,hi.count=NULL,\
hi.perc=NULL)
meth=unite(filtered.methData,destrand=TRUE)
pdf(file="clusterSamples.pdf", height=8, width=8)
clusterSamples(meth, dist="correlation", method="ward", plot=TRUE)
dev.off()
png(filename="correlation.png", height=10000, width=10000)
getCorrelation(meth,plot=T)
dev.off()
pdf(file="PCASamples.pdf", height=8, width=8)
PCASamples(meth)
dev.off()
```

The identification of differentially methylated regions (DMRs) was done with the software tool metilene (v0.2-6)[176]. Input files have to be divided by 100 because metilene needs values between 0 and 1. DMRs from pairwise comparison of sorted tab separated files were output in a bed-like format.

```
### Reformat input file for metilene
myReformatCpGBedgraph100to1.pl \
/path-to-input-file/MDS53_timepoint1.cpg.cleared.CG.bedgraph \
> /path-for-output-file/conv_MDS53_timepoint1.cpg.cleared.CG.bedgraph

### Pairwise comparison of patients with metilene to find DMRs
metilene_input.pl -in1 /path-to-input-file1/conv_CD34.cpg.cleared.CG.bedgraph \
-in2 /path-to-input-file2/convMDS53_timepoint1.cpg.cleared.CG.bedgraph -out \
/path-for-output-file/conv_CD34vsMDS53_timepoint1_metilene.input -h1 CD34 -h2 P53_1

metilene -M 100 -f 1 -t 12 -a CD34 -b P53_1 \
/path-to-input-file/conv_CD34vsMDS53_timepoint1_metilene.input \
> /path-for-output-file/conv_CD34vsMDS53_timepoint1_metilene.output
```

```
sort -k1,1 -k2,2n /path-to-input-file/conv_CD34vsMDS53_timepoiont1_metilene.output\
> /path-for-output-file/conv_CD34vsMDS53_timepoint1_metilene.sorted.output

metilene_output.pl -q \
/path-to-input-file/conv_CD34vsMDS53_timepoint1_metilene.sorted.output \
-o /path-for-output-folder/CD34_vs_MDS53_timepoint1/ -c 5 -d 0.1 -a CD34 -b P53_1
```

To distinguish between DMRs that lost or gained DNA methylation in contrast to CD34+ cells, metilene output files were filtered for both categories. Then all DMRs between patient time points were put into one file, sorted and then merged. Merging of the files gives you the total number of DMRs with information about the distribution of these regions in the single patients. Gene ontology analysis with metascape was done after annotation of these regions to gencode transcripts. For this annotation, input files have to be converted to the genome hg19 since gencode data is only available for this genome version.

```
### Annotation of DMRs
getAllGeneEnhancerAssociations.pl \
/path-to-input-file/CD34vsMDSall_commonDMRs_loss_merged_liftOverhg19.txt \
/path-to-gencode-file/gencode.v19.transcripts.txt D34vsMDSall_commonDMRs_loss_hg19\
/path-to-output-folder/commonDMRs -gtex /path-to-GTEx-file/WBA.snpgenes.1000.bed
```

Patient specific DMRs were analyzed by putting together all found DMRs of one patient, merging and annotating these regions. *K*-means clustering and annotation of epigenetic data in CD34+ cells was done with R as described above (section 4.2.3.2.1), while the later one was drawn as histogram with 95% confidence interval:

```
library(ggplot2)
library(reshape)
setwd("/path-to-working-directory/")
data <- read.table("input-file.txt", header=T, sep="\t", stringsAsFactors=FALSE,
check.names = FALSE)
tdata <- t(data[,-1])
m <- apply(tdata, 1, mean)
s <- apply(tdata, 1, sd)
n <- nrow(as.matrix(tdata))
error <- qnorm(0.975)*s/sqrt(n)
ci975 <- m+error
ci025 <- m-error
d <- as.numeric(as.character(rownames(tdata)))
result <- data.frame(cbind(d,m, s,ci975,ci025))
colband <- "gray"
colline <- "blue"
p <- ggplot(result, aes(d,m)) +
```

```
      geom_ribbon(aes(x=d, ymax=ci975, ymin=ci025), fill=colband, alpha=.25) +
      geom_line(colour = colline) +
      theme_bw(base_size=12) +
      xlab("Distance from peak center") +
      ylab("ChIPseq coverage") +
      xlim(-3000, 3000)
plot(p)
pdf(file="output-file.pdf", height=4, width=4)
plot(p)
dev.off()
```

To visualize changes in the subclonal architecture of the patients, fish plots were drawn using the fishplot package for R (exemplary script for one patient).

```
library(gplots)
library(RColorBrewer)
library(fishplot)
setwd("\path-to-working-directory \")
timepoints=c(1,2,3,4,5)
frac.table = matrix(
      c( 95.7, 30.4, 0, 0, 0,
         45, 14.5, 0, 0, 0,
         0.1, 0.05, 33.3, 0, 0,
        0.1, 0.05, 8, 0, 0,
        86.7, 0, 0, 73.3, 66.7),
      ncol=length(timepoints))
parents = c(0,1,0,1,4)
fish = createFishObject(frac.table,parents,timepoints=timepoints, col =
c("violetred4", "ivory4", "darkorange", "darkgreen", "royalblue4"))
fish = layoutClones(fish)
pdf(file="fishplot.pdf", height=3, width=3)
fishPlot(fish,shape="spline",title.btm="MDS",
             cex.title=0.5, vlines=c(1,2,3,4,5), col.vline="grey",
             vlab=c("1","2","3","4","5"), cex.vlab=1.0, border = 0.5, col.border =
"midnightblue",
           bg.type = "solid", bg.col = "white")
dev.off()
```

Motif analysis for adult and pediatric MDS patients was done with the software HOMER[177] and could be done either with or without subtraction of a specific background (option “-bg”). The following command shows one example with background subtraction.

```
LAYOUTfindMotifsGenome.pl \
/path-to-input-file/MDS25_DMRs_kmeans5_cluster1forMotifs.txt hg38r \
/path-to-output-folder/motifs -size given -len 7,8,9,10,11,12,13,14 -p <#CPU> \
-h -sample Name-for-output-folder -sampleID Name-for-output-folder_bg -bg \
/path-to-background-file/Regulome_capture_targets.txt -chopify
```

4.2.3.2 MCIp-seq data

MCIp-seq data sets were generated by indexed single 50 bp sequencing (~ 25 million reads per sample) on a HiSeq 1000 (Illumina) or HiSeq 3000/4000 (Illumina). The general NGS workflow is stated in section 4.2.1.19.

4.2.3.2.1 Analyzing global DNA methylation in pediatric MDS patients using MCIp-seq

Global DNA methylation in purified bone marrow granulocytes from pediatric MDS patients was analyzed using the MCIp-seq approach. The following script is representative for all analyzed samples, but will be illustrated for one example. If not stated otherwise in the command line, default settings were applied.

The first steps, from conversion of bam to fastq to alignment with bowtie2 are identical like described above (section 4.2.3.2.2). To remove clonality from the sequencing data, we cleaned up downstream analysis by forcing tag counts at each position to have a maximum of 1 (indicated with “tbp1”). Furthermore, we normalized the data to the 99th percentile to compensate different sequencing depth levels of the samples. This is done using a norm factor that results from division of total tag positions with the 99th percentile value.

```
###Creating tbp1 Tag Directories with the homer tool pipeline
makeTagDirectory /path-to-output-folder/MCIp_D770_186406_19 \
/path-to-tag-directory/MCIp_D770_186406_19 -tbp1

###Creating USCC custom tracks to visualize aligned sequence tags
makeUCSCfile /path-to-tbp1-tag-directory/KM_MDS01550292_62 -norm <norm factor> \
-bigWig /path-to-genome-file/chrom.sizes -fsize 1e20 -o \
/path-to-output-file/MDS.1.0.bigWig

### normalization to 99th percentile
normalizeMCIpToXpercentile.pl \
/path-to-region-file/MCIp_detected_AutoNonRepeatRegionsMappable.txt \
/Volumes/path-to-tbp1-tag-directory/MCIp_D770_186406_19 output-name 0.99

### pasting all normalized patient files into one file

### t-SNE (t-distributed Stochastic Neighbor Embedding) approach with R
```

```
library(ggplot2)
library(Rtsne)
library(edgeR)
setwd("/path-to-working-directory /")
data <- read.delim("McIpPeaks_allWoTimecourses_ann_tbp1.counts.txt", header=TRUE,
row.names="ID")
mydata <- data.matrix(t(data))
set.seed(35)
rtsne_out <- Rtsne(mydata, check_duplicates = FALSE, pca = TRUE, perplexity=5,
theta=0.125, dims=2, max_iter = 50000)
group <- factor(c(rep("TL_208021_1",1), rep("D801_203226_2",1), <list of all
patients> )))
embedding <- as.data.frame(rtsne_out$Y)
embedding$Class <- as.factor(group)
p <- ggplot(embedding, aes(x=V1, y=V2, color=Class)) +
     geom_point(size=5) +
     guides(colour = guide_legend(override.aes = list(size=5))) +
     xlab("") + ylab("") +
     theme_light(base_size=10) +
     theme(strip.background = element_blank(),
          panel.background = element_rect(fill = NA, colour = "black"),
          legend.position  = "none",
            strip.text.x     = element_text(size = 3),
            axis.text.x      = element_text(size = 20, color="black"),
            axis.text.y      = element_text(size = 20, color="black"),
            panel.border     = element_blank())
pdf(file="tSNE_allWoTimecourses_tbp1.counts_seed35_2.pdf", height=6, width=6)
plot(p, labels=TRUE)
dev.off()

### Wilcoxon test for calculation of significance levels of distinct clinical
features
library(ggplot2)
library(plyr)
setwd("/path-to-working-directory /")
data <- read.table("completeCohort_woTC_tSNEclustering_GATA2.txt", header=TRUE,
sep=" ", row.names="ID")
colnames(data) <- c("GATA2","cluster")
df <- data.frame(data)
p1 <- ggplot(df, aes(factor(cluster), GATA2, fill=factor(cluster)))
p2 <- p1 + geom_boxplot(position=position_dodge(0.8))
p3 <- p2 + geom_dotplot(binaxis='y',
stackdir='center',position=position_dodge(0.8), binwidth = .02, stackratio=1.5) +
labs(x="Cluster", y = "GATA2")
p <- p3 + scale_fill_brewer(palette="Blues") + theme_minimal()
pdf(file="ScatterPlot_allPatientswoTC_tSNE_GATA2.pdf", height=3, width=4)
```

```
plot(p)
dev.off()
wilcox.test(GATA2 ~ cluster, data=df)

### identifying DMRs between different t-SNE region clusters
#Paste all specific peaks from patients belonging to the same cluster into one file
#and merge peaks. Do peak annotation using a script that overlaps regions with
#promoters and then filters out GTEx associated peaks. The resulting gene list was
#analyzed regarding gene ontology using metascape.
getAllGeneEnhancerAssociations.pl \
/path-to-peak-file-folder/McIpPeaks_tSNE_DMRs_merged_cluster1.txt \
/path-to-gencode-file/gencode.v19.transcripts.txt <output name> \
/path-to-output-folder -gtex /path-to-GTEx-file/WBA.snpgenes.1000.bed

### Creating scatter plots with total and patient specific regions
library(ggplot2)
setwd("/path-to-working-directory /")
data <- read.table("input-file.txt", header=T, sep="\t")
attach(log10 (data + 0.1))
d <- data.frame(log10 (data + 0.1))
lm_eqn = function(d){
    m = lm(total regions control ~ total regions patient_sample1, d);
    eq <- substitute(italic(r)^2~"="~r2,
         list(r2 = format(summary(m)$r.squared, digits = 3)))
    as.character(as.expression(eq)); }
p1 <- ggplot(d,aes(x= total regions control, y= total regions patient_sample1)) +
theme_bw(base_size = 8, base_family = "Helvetica") +
coord_cartesian(xlim=c(0,4),ylim=c(-0,4))  +
geom_point(size=.20,colour="black",alpha=0.25) + annotate("text", x = 3, y = .25,
label = lm_eqn(d), size = 3, colour="black", parse = TRUE) +
geom_point(aes(x=specific regions control, y=specific regions
patient_sample1),size=.20,colour="firebrick2")
pdf(file="Scatter.control_patient_sample1.pdf", height=2, width=2)
plot(p1)
dev.off()

### k-means clustering of patient specific regions
x <- read.table("input-file.txt", header=TRUE, sep="     ", row.names="ID")
data <- as.matrix(log2(x+0.01))
breaks = col_breaks, na.rm=TRUE, scale="none", margins=c(10,10), #cexRow=0.5,
cexCol=1.0, key=TRUE, density.info="none", trace="none")
#Kmeans cluster number:
wss <- (nrow(data)-1)*sum(apply(data,2,var))
for (i in 2:15) wss[i] <- sum(kmeans(data,
       centers=i)$withinss)
plot(1:15, wss, type="b", xlab="Number of Clusters",
```

```
  ylab="Within groups sum of squares")
fit <- kmeans(data, <preferred cluster number>)
aggregate(data,by=list(fit$cluster),FUN=mean)
mydata <- data.frame(data, fit$cluster)
attach(mydata)
mysorteddata <- mydata[order(-fit.cluster), ]
write.table(as.matrix(mysorteddata),file="kmeans-clustered-file.txt",sep="\t",
col.names=NA)
mysorteddata <- read.table("kmeans-clustered-file.txt ", header=TRUE, sep=" ",
row.names="ID")
data <- as.matrix(mysorteddata[,c("<time point1>","<time point2>")])
mycol <- colorRampPalette(c("blue","white","red"))(299)
col_breaks = c(seq(-6,-2,length=100), seq(-2,2,length=100), seq(2,6,length=100))
pdf(file="kmeans-clustered-file.pdf", height=8, width=8)
heatmap.2(data, Rowv=NA, Colv=NA, col = mycol, breaks = col_breaks, na.rm=TRUE,
scale="none", margins=c(10,10), cexRow=0.5, cexCol=1.0, key=TRUE,
density.info="none", trace="none")
dev.off()

### annotate epigenetic data from CD34+ cells in k-means region clusters and make
heatmap
annotatePeaks.pl \
/path-to-kmeans-output-position-file/kmeans-clustered-position-file.txt hg19 \
-size 6000 -hist 100 -ghist -d /path-to-tag-directory/epigeneticDataInCD34 \
/path-to-output-folder/kmeans-clustered-file-with-epigenetic-data-in-CD34.txt

library(gplots)
library(RColorBrewer)
setwd("/path-to-working-directory /")
data <- read.delim("kmeans-clustered-file-with-epigenetic-data-in-CD34.txt.txt",
row.names="Gene")
d <- data.matrix(data)
mycol <- colorRampPalette(c("white","blue"))(199)
col_breaks = c(seq(0,5,length=100), seq(5,10,length=100))
png(filename=" kmeans-clustered-file-with-epigenetic-data-in-CD34.png",
height=8000, width=8000)
heatmap.2(d, scale="none", Rowv=NA, Colv=NA, col = mycol, breaks = col_breaks,
dendrogram = "none", margins=c(0,0), cexRow=0.5, cexCol=1.0, key=TRUE,
density.info="none", trace="none", symm=FALSE,symkey=F,symbreaks=TRUE)
dev.off()

### find DMRs between patient time points
setwd("/path-to-working-directory /")
library(edgeR)
data <- read.delim("patient-specific-peaks.tbp1txt", row.names="ID")
group <- factor(c(rep("time point1",1), rep("time point 2",1)))
```

```
d <- DGEList(counts=data,group=group)
colnames(d) <- c("time point 1", "time point 2")
d <- calcNormFactors(d)
d$common.dispersion <- 0.05
de.com <- exactTest(d, pair=c("time point 1","time point 2"))
results <- topTags(de.com,n = Inf)
write.table (as.matrix(results$table), file = "patient-specific-peaks-tbp1-
timepoint1VStimepoint2.txt", sep = "  ", col.names=NA)
counts.per.m <- cpm(d, normalized.lib.sizes=TRUE)
write.table (counts.per.m, file = "patient-specific-peaks-tbp1-counts.txt", sep = "
      ", col.names=NA)
```

4.2.3.2.2 Comparison of MCIp-seq data sets from peripheral blood and bone marrow

MCIp-seq data from peripheral blood and bone marrow allowed analysis of global DNA methylation in these tissues. The following scripts were used to compare MCIp-seq data from those different tissues. It is shown exemplary for sample “KM_62” but is representative for all analyzed samples. If not stated otherwise in the command line, default settings were applied.

```
### Conversion from bam to fastq file using bedtools
bedtools bamtofastq -i /path-to-file/01550292_KM_62_TGACCA_L002_R1_001.bam\ -fq
/path-to-file/01550292_KM_62_TGACCA_L002_R1_001.fastq

### Quality assessment with Fastqc

### Alignment of raw reads to hg19 reference genome with bowtie2
myMap-bowtie2.pl -x hg19 -p <#CPU> -name /path-to-output-folder/name-of-sample\
/path-to-file/01550292_KM_62_TGACCA_L002_R1_001.fastq.gz

### Creating Tag Directories with the homer tool pipeline
makeTagDirectory /path-to-output-folder/KM_MDS01550292_62 \
/path-to-file/KM_MDS01550292_62.sam -keepOne -genome hg19 -checkGC

### Creating USCC custom tracks to visualize aligned sequence tags
makeUCSCfile /path-to-tag-directory/KM_MDS01550292_62 -bigWig \
/path-to-genome-file/chrom.sizes -o /path-to-output-file/ KM_MDS01550292_62 .bigWig

### Annotating region files
annotatePeaks.pl /path-to-region-file \
/MCIp_detected_AutoNonRepeatRegionsMappable.txt hg19 -size given –d
<space separated list of tag directories> -nogene -noadj > \
/path-to-output-file/MCIp_detected_AutoNonRepeatRegionsMappable_tagAnn.txt
```

```
### Clustering of samples and drawing correlation heat map with R
library(gplots)
library(RColorBrewer)
setwd("/path-to-working-directory/")
data <- read.table("input-filename.txt", header=TRUE, sep="    ", row.names="ID")
x  <- data.matrix(data)
#logx <- log10(x)
y <- cor(x, use="complete.obs", method="pearson")
h <- hclust(dist(y, method = "manhattan"), method="ward.D2")
mycol <- colorRampPalette(c("white","red"))(199)
col_breaks = c(seq(0.6,0.8,length=100), seq(0.81,1,length=100))
pdf(file="output-filename.pdf", height=6, width=6)
heatmap.2(y, Rowv=as.dendrogram(h), Colv=as.dendrogram(h), col = mycol,
breaks=col_breaks, scale="none", margins=c(5,5), cexRow=0.5, cexCol=1.0, key=TRUE,
density.info="none", trace="none", symm=FALSE,symkey=FALSE,symbreaks=TRUE)
dev.off()
```

5 Results

5.1 Adult MDS

The maintenance of genomic stability and normal gene expression is ensured via specific DNA methylation patterns in the cell, while aberrant DNA methylation may cause a number of diseases, like AML or MDS. Several studies showed that patients with MDS or AML display alterations in DNA methylation compared to healthy donors[154,156,178]. Aberrations comprise global hypomethylation as well as hypermethylation of CpG islands (CGIs) in promoter regions. Hypermethylated promoters of tumor suppressor genes lead to inactivation of the affected gene and could promote tumorigenesis and / or progression to secondary AML.
Targeted bisulfite sequencing of the myeloid regulome in longitudinal samples from MDS patients was performed in order to identify promising biomarkers that could be useful for diagnosis and prognosis. This is a cost effective alternative to study evolution of DNA methylation patterns in comparison to genetic events and biomarkers. Our region set ("myeloid regulome") for targeted enrichment comprises a customized set of regulatory regions important for myeloid differentiation and AML pathogenesis and its selection is described in section 4.2.3.1.
One limitation of targeted bisulfite sequencing is the interpretation of resulting data. It has to be taken into account that the extent of DNA methylation depends on the number of altered cells. Since we were analyzing whole bone marrow samples from patients, consecutive samples are always varying in the number of altered cells. Furthermore myelodysplastic syndromes are known to be a very heterogeneous group of diseases making it difficult to compare different patients amongst each other. But here, the clonal architecture of the bone marrow defined by the variant allele frequency can be very helpful for data interpretation.

To illuminate disease progression of myelodysplastic syndromes in adults, DNA methylation of the myeloid regulome was studied in a cohort of 6 patients (2 females and 4 males) with a median age of 69.5 (50 – 79) years (see Table 5-1). Bone marrow samples from patients P02 – P20 were kindly provided by Dr. Daniel Nowak (Department of Hematology and Oncology, Medical Faculty Mannheim) and the peripheral blood sample from patient P53 by Dr. Detlef Haase (Department of Hematology and Oncology, Goettingen). Longitudinal samples for patients P02 – P20 and P53 were available for 4 time points or 5 time points, respectively. As we aimed for an integrated analysis regarding epigenetic, cytogenetic and genetic changes, we focused on more advanced subtypes due to the increased occurrence of alterations in high-risk MDS patients and on the other hand the higher probability to progress into acute myeloid leukemia.

Recurrent somatic mutations encountered in MDS, such as SF3B1 or TET2, were observed in all patients, ranging from two to seven mutations per patient. Out of six patients, two had normal karyotypes (P19 and P20), while three patients were found with del(5q) (P02, P13 and P15) and one patient with complex karyotype (P53). The percentage of bone marrow blasts ranged from 0 % to 13 %. Since patients underwent different clinical treatment regimes, DNA methylation changes due to distinct medication needed to be considered additionally.

Table 5-1 - Clinical data of six adult MDS patients

Patient	Sex	Age at Dx	WHO at Dx	Cytogenetics	Gene mutations	BM blast count (%)	Therapy
P02	f	50	MDS with del(5q)	46,XX[8]/46 XX del(5)(q14q33)[12]	DNMT3A, CHRM2, RAET1G	0	Lenalidomide, phlebotomy
P13	f	66	MDS with del(5q)	46,XX,del(5)(q13q33) [20]/ 46, XX [3]	TET2, SF3B1, TP53, PML, C7, TNIK	<5	-
P15	m	73	MDS with del(5q)	46,XY,del(5)(q14q34) [20]	ASXL1, EZH2 (2x), ETV6, RUNX1, NF1, CSNK1A1	0	Lenalidomide
P19	m	74	MDS-RS-SLD	46,XY [22]	ASXL1, SF3B1, TTBK1	1-2	APG101
P20	m	66	MDS-EB1	46,XY [20]	IDH2, SRSF2, SPEG2, BRCC3, NF1	NA	Lenalidomide, APG101
P53	m	79	MDS-EB2	46,XY,del(20) (q12q13.2?) [15] /47,idem, +12, i(12)(q10)[7]/46,XY[1]	RUNX1, SRSF2	13	Azacitidine

Abbreviations: Dx: diagnosis; BM: bone marrow

5.1.1 Comprehensive analysis of DNA methylation data of all patients

5.1.1.1 Identification of DMRs in patients during disease progression

To get an overview of differences and commonalities we looked for differentially methylated regions (DMRs) during disease progression among all six patients. DMRs that were found in comparison between consecutive samples of individual patients were merged and resulted in a total of 1740 DMRs.

The number of overlapping DMRs within samples of single patients is shown in Figure 5-1 A. The Venn diagram depicted shows that there are no common differentially methylated regions between all six patients. Most of the identified DMRs were patient specific, meaning that there was only a small amount of overlapping DMRs between the patients. Figure 5-1 B illustrates this phenomenon, showing that the amount of common DMRs between patients decreased with number of patients. In summary, 1352 DMRs were found only in one patient, 271 DMRs were common in two patients, while the number of common DMRs further decreased with higher patient numbers (85 DMRs in three patients, 30 DMRs in four patients, 2 DMRs in five patients). This demonstrated that most progression associated changes in DNA methylation were individual and only few commonalities were observed in this patient cohort.

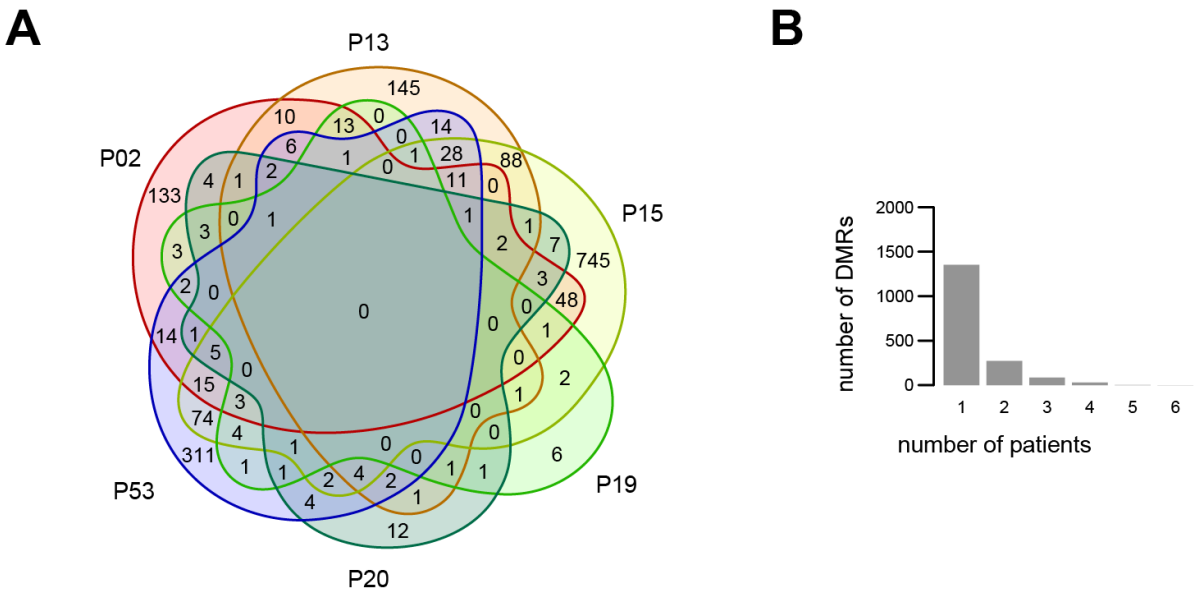


Figure 5-1 – Distribution of differentially methylated regions between all patients

(A) Venn diagram of differentially methylated regions (DMRs) showing the number of overlapping DMRs between distinct patients. (B) Bar plot which depicts the amount of DMRs for a distinct number of patients.

5.1.1.2 Identification of DMRs in comparison to CD34+ cells

Next, we compared methylation profiles of patients with CD34+ cells. These cells serve as a control group representing a normal state of CGI methylome. Furthermore, comparison with these cells allows the identification of disease specific alterations and differentiation dependent changes.

Additionally we separated the DMRs into regions that gained or lost DNA methylation in comparison to CD34+ cells. In detail, differentially methylated regions between the single time points of the patients were filtered for gain and loss of methylation and then merged into one file. These DMRs were further analyzed regarding their overlap with other patients, their gene ontology and epigenetic landscapes in HSC (Figure 5-2 A – D).

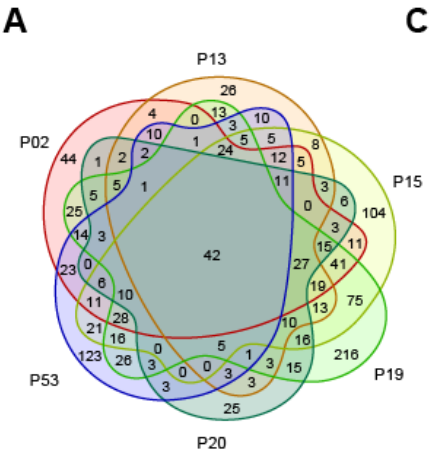
The upper part of Figure 5-2 shows regions that lost DNA methylation in patients in comparison to CD34+ cells. The Venn diagram depicts that 42 DMRs out of 1165 DMRs are commonly demethylated in patients compared to hematopoietic stem cells (Figure 5-2 A). Proportional distribution of DMRs over the different number of patients is depicted as bar plot and showed a negative correlation (Figure 5-2 B). The higher the number of patients incorporated, the lower was the amount of overlapping DMRs. For instance, 538 regions were found to be differentially methylated in one patient only and therefore altered individually, whereas 78 DMRs were altered in five patients. In order to functionally annotate DMRs, we assigned them to neighboring genes using an approach similar to GREAT which also considered eQTL associations of closeby SNPs provided by GTEx[179].

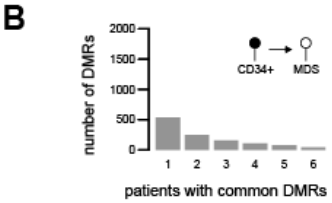
Significantly enriched GO terms found with the online tool Metascape[180] were for instance “hematopoietic or lymphoid organ development”, “leukocyte activation involved in immune response”, “pathways in cancer” or “myeloid differentiation” (Figure 5-2 C). By integrating analysis of epigenetic marks associated in these regions, more details could be gathered regarding the characteristics of these DMRs. Histograms depicted in Figure 5-2 D show that these hypomethylated regions are rarely associated with the repressive histone mark H3K27me3 (blue), while active marks like H3K4me3 (orange) and H3K27ac (purple) are highly enriched. Furthermore regions are accessible for DNase I, indicated by the signal of DHSs (DNase I hypersensitive site, green).

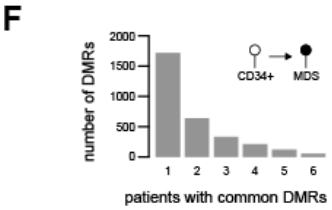
Using the approach mentioned above, DMRs hypermethylated in patients compared to CD34+ cells included 53 regions that were commonly more methylated in all patients than in stem cells (Figure 5-2 E). In total, 3068 DMRs were observed, while 1720 DMRs showed again individual DNA methylation alterations. The negative correlation between abundance of DMRs and number of patients is depicted in the bar plot in Figure 5-2 F.

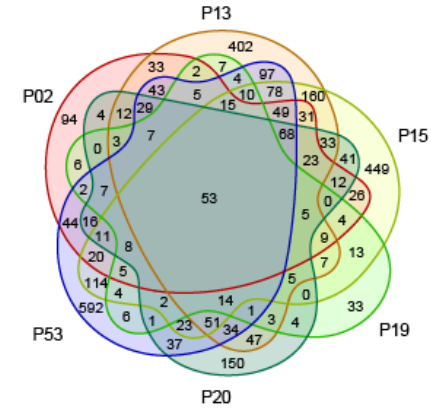
Gene ontology analysis of hypermethylated regions depicted in Figure 5-2 G, generally revealed terms with higher significance compared to hypomethylated ones. Associated interesting pathways were “hematopoietic or lymphoid organ development”, “negative regulation of cell proliferation”, “negative regulation of cell differentiation”, “regulation of cell migration”, or “regulation of hemopoiesis”.

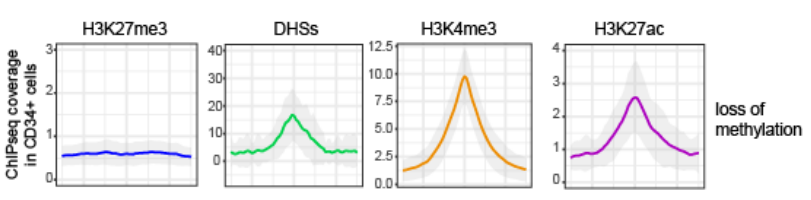
Histone H3K27 trimethylation (blue) showed a high signal in DMRs which gained DNA methylation in comparison to CD34+ cells, but also active marks like H3K4me3 (orange), H3K27ac (purple) and DNase I hypersensitive sites (green). Due to this results regions seemed to be bivalent ones (Figure 5-2 H). The phenomenon of bivalent promoters, that become hypermethylated in cancer cells was described in several studies[181–183].

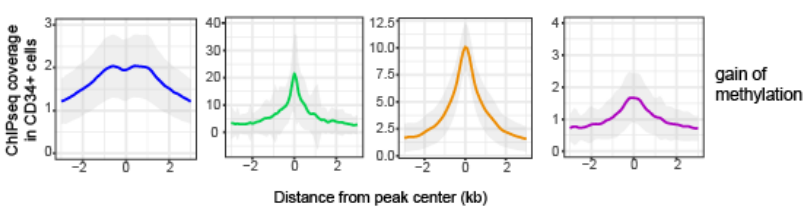












C

GO term analysis for DMRs hypomethylated compared to CD34+ cells	Log(q-value)
hematopoietic or lymphoid organ development	-7.8
T cell activation	-7.8
leukocyte activation involved in immune response	-7.8
Phosphorylation of CD3 and TCR zeta chains	-6.3
Pathways in cancer	-5.0
Hemostasis	-4.6
Focal adhesion	-4.5
myeloid cell differentiation	-4.0
regulation of nuclear-transcribed mRNA catabolic process, deadenylation-dependent decay	-3.5
endocytosis	-3.0
positive T cell selection	-2.9
negative regulation of transcription from RNA polymerase II promoter	-2.8
regulation of catabolic process	-2.7
Adherens junction	-2.6
Clathrin-mediated endocytosis	-2.6
cellular response to nitrogen compound	-2.6
negative regulation of phosphate metabolic process	-2.6
multicellular organismal homeostasis	-2.3
response to drug	-2.3
CtBP complex	-2.3

G

GO term analysis for DMRs hypermethylated compared to CD34+ cells	Log(q-value)
embryonic morphogenesis	-17.3
hematopoietic or lymphoid organ development	-16.1
negative regulation of transcription from RNA polymerase II promoter	-12.8
tube development	-12.5
regionalization	-12.1
sensory organ development	-11.4
skeletal system development	-10.6
negative regulation of cell proliferation	-10.1
blood vessel development	-10.0
negative regulation of cell differentiation	-9.2
regulation of cell development	-8.6
response to wounding	-8.3
cell surface receptor signaling pathway involved in cell-cell signaling	-7.7
regulation of cell migration	-7.5
muscle structure development	-7.2
cell fate commitment	-7.2
regulation of hemopoiesis	-7.1
appendage development	-7.1
cell morphogenesis involved in differentiation	-6.6
head development	-6.6

Figure 5-2 - Analysis of DMRs between CD34+ cells and adult MDS patients

Differentially methylated regions in comparison to hematopoietic stem cells were distributed into regions either losing (A – D) or gaining (E – H) DNA methylation compared to CD34+ cells. (A & E) Venn diagram of DMRs illustrates the overlapping regions between the different patients. (B & F) Proportional distribution of DMRs in the different number of patients is depicted as bar plot. (C & G) Table of enriched gene ontology terms in analyzed DMRs obtained with Metascape and associated q-values. (D & H) Histograms show the average coverage of different epigenetic marks in the analyzed regions done with ChIPseq (Chromatin immunoprecipitation sequencing), while the 95% confidence interval is depicted in gray.

5.1.1.2.1 Detailed characterization of common DMRs in comparison to CD34+ cells

In order to characterize the identified differentially methylated regions in greater detail, we analyzed the localization across the genome of these DMRs. This was done by annotating the DMRs to the reference genome using the software HOMER.

Analysis was restricted to DMRs that were common in all patients in comparison to CD34+ cells, since we were looking for potential common target regions that could serve as biomarkers.

Distribution of DMRs that are losing or gaining DNA methylation in comparison to CD34+ cells is depicted in the two pie charts below (Figure 5-3). In both cases, the majority of DMRs were assigned to intronic and intergenic regions followed by transcription start sites and exons. The only difference is that hypomethylated DMRs could be observed with a higher percentage in intergenic regions and less in introns than hypermethylated DMRs compared to CD34+ cells.

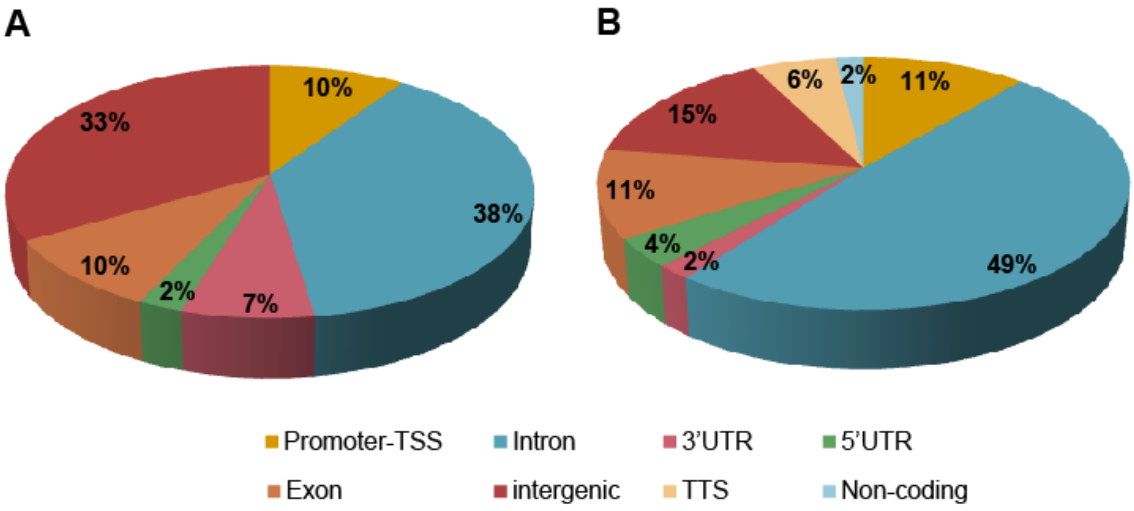


Figure 5-3 – Genomic distribution of common DMRs in comparison to CD34+ cells

Pie chart showing the percentage of DMRs localized in different genomic parts with DMRs losing (A) or gaining (B) DNA methylation in patients compared to hematopoietic stem cells. Annotation of identified differentially methylated regions to the human genome was done using the software HOMER.

Due to their localization, some DMRs may have important effects on gene expression, especially on tumor suppressor genes or oncogenes, and could play a role in the pathogenesis of MDS or during progression to acute myeloid leukemia. For every subset of DMRs, hypomethylated and hypermethylated ones respectively, results of GO term analysis were used to identify interesting affected genes.

Focusing on DMRs which were common in all patients and lost DNA methylation in comparison to CD34+ cells (in total 42 DMRs), GO term analysis revealed several interesting genes (see Table 5-2). They were found in the term “chromatin organization”, including for instance *PRDM2*, *PRDM14* and *YEATS4*. PRDM2 and PRDM14 (PR/SET Domain 2) are members of the nuclear histone/protein

methyltransferase superfamily and are acting as tumor suppressor genes in solid tumors and leukemia[184–187]. Normally, YEATS4 (YEATS Domain-Containing Protein 4) is a component of the NuA4 histone acetyltransferase (HAT) complex and therefore involved in transcriptional activation, whereas overexpression of this oncogene is involved in the development of various tumors[188,189].

Table 5-2 - GO analysis of common DMRs with loss of DNA methylation in comparison to CD34+ cells

GO term	Involved genes
Chromatin organization	*HNRNPC, PRDM2, YEATS4, HIST1H2BL, HDAC4, EP400, PRDM14, MYSM1*
actin cytoskeleton reorganization	*PRKCZ, CAPN10, MICALL2*
negative regulation of phosphorylation	*GFRA2, PRKCZ, HDAC4, RPTOR*
HATs acetylate histones	*YEATS4, HIST1H2BL, EP400*

Table 5-3 summarizes the results of gene ontology analysis obtained with common DMRs which gained DNA methylation in contrast to hematopoietic stem cells (53 DMRs in total). With the aim to identify potential epigenetic target genes that are involved in pathogenesis of myelodysplastic syndromes and / or progression to acute myeloid leukemia, genes involved in the GO terms “positive regulation of hemopoiesis” and “methylation” seemed to be promising.

The important transcription factor RUNX1 is involved in generation of hematopoietic stem cells and lineage-specific differentiation[190]. Therefore the observed hypermethylation in the promoter region could repress RUNX1 expression and misbalance hematopoiesis. Another affected and interesting transcription factor is FOXO3, which plays important roles in cellular processes, including differentiation, proliferation and apoptosis. Dysregulation of FOXO3 expression was found to be involved in tumorigenesis and progression[191,192]. ZFPM1 (Zinc Finger Protein, FOG Family Member 1) or also known as FOG1 (friend of GATA protein 1) that was found in hypermethylated DMRs acts as a cofactor of GATA1 in regulating erythroid and megakaryocytic cell differentiation[193,194].

Table 5-3 - GO analysis of common DMRs with gain of DNA methylation in comparison to CD34+ cells

GO term	Involved genes
Vibrio cholera infection	*GNAS, PLCG2, ZFPM1*
Muscle cell development	*ENG, NFATC2, CTDP1, BIN3, EGR3, ZFPM1*
Positive regulation of hemopoiesis	*RUNX1, EGR3, FOXO3, GNAS, ZFPM1, PLCG2*
Positive regulation of transporter activity	*PLCG2, SYNGR3, CRACR2A, COX7A1, KCNN4*
Transforming growth factor β receptor signaling pathway	*LDLRAD4, ENG, DUSP22, PRDM16, EGR3, FOXO3*
Methylation	*AMT, GNAS, PRDM8, PRDM16, ZFP57*
Regulation of fat cell differentiation	*PRDM16, ZFPM1, ZADH2, LDLRAD4, FOXE3, FOXO3*
Negative regulation of transcription from RNA polymerase II promoter	*FOXO3, NFATC2, BHLHE40, DUSP22, PRDM16, ZFPM1, ZFP57*
Positive regulation of angiogenesis	*RUNX1, ENG, HSPB6, EGR3, FBXW8, EGFL7*
Antigen receptor-mediated signaling pathway	*KCNN4, NFATC2, PLCG2, DUSP22*
Oxidative phosphorylation	*COX7A1, ATP6V1H, NDUFA11*
Regulation of homeostatic process	*FOXO3, PLCG2, XRCC3, SMG5, ZFPM1*
Positive regulation of epithelial cell proliferation	*EGR3, FOXE3, EGFL7*

5.1.1.2.2 Functional analysis of DNA methylation dependence in DMRs

To address DNA methylation sensitivity of these differentially methylated regions (RUNX1, FOXO3 and ZFPM1), affected regulatory genomic regions were analyzed in reporter gene assays (see section 4.2.1.6). Genome browser tracks of selected regions were listed in Appendix (Figure 11-1 - Figure 11-4), where chosen sequences were highlighted in red.

All three regions were tested for their enhancer activity and the RUNX1 region additionally regarding the promoter activity. For this approach, a CpG-free luciferase reporter plasmid (pCpGL-basic, #861) was used[195], where effects of DNA methylation in promoter constructs can be tested. Based on this method, further plasmids were developed to test enhancer activity depending on DNA methylation status. This reporter plasmid consists of the E1AF-promoter/CMV-enhancer cassette with an upstream located terminator (pCpGL-CMV/T.E1AF, #1341), where the CMV enhancer was replaced by the DMR region via cloning. Luciferase reporter constructs were either mock-treated or methylated *in vitro* with SssI methylase (suffix “methyl”).

The bar plot in Figure 5-4 shows that only one DMR (FOXO3_2) showed activity in the unmethylated state above the one of the control plasmid pCpGL-CMV/T.E1AF (#1341). In this case, *in vitro* methylation showed a loss of enhancer activity at a significant level suggesting that differentially methylation in the FOXO3_2 region could have an effect on gene expression. As mentioned above, alterations in FOXO3 expression were shown to be involved in tumorigenesis and progression[191,192] and therefore may also play a role in development of myelodysplastic syndromes.

All other tested DMRs showed no activity above the corresponding control reporter plasmid indicating to be not active and therefore we were not able to make a statement regarding DNA methylation dependence of these regions.

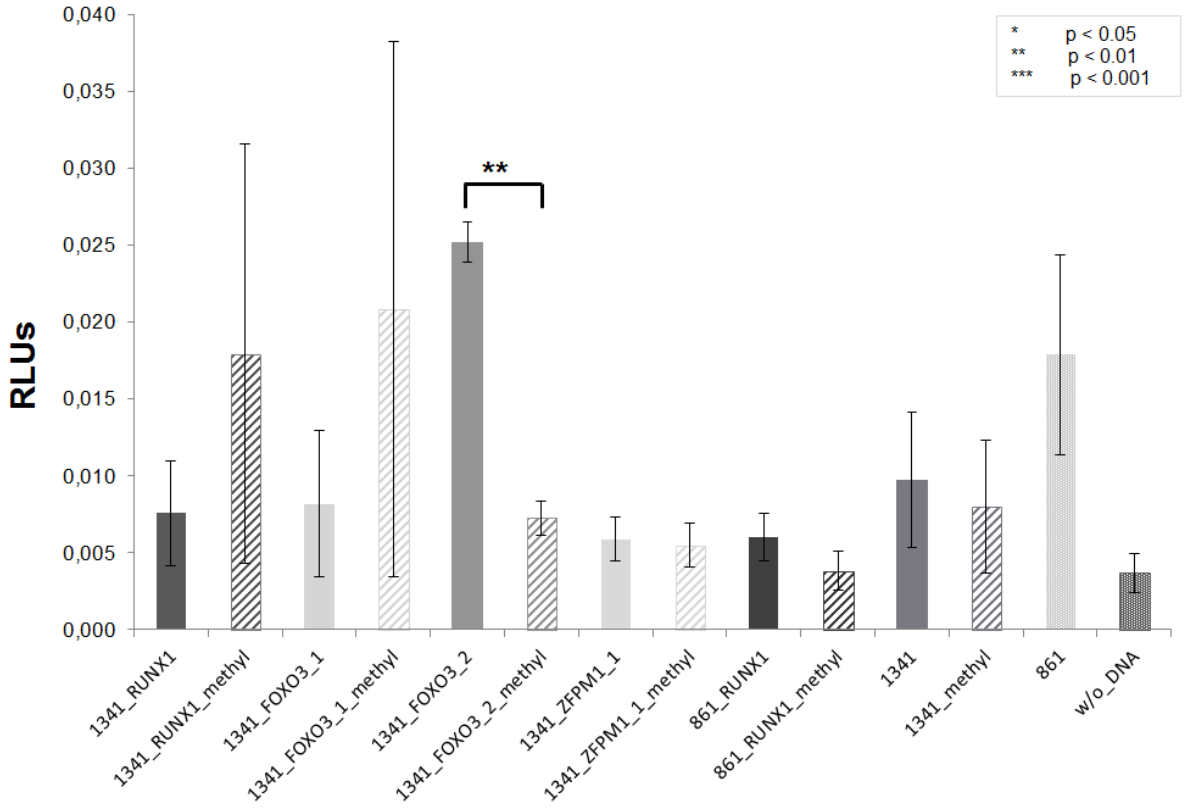


Figure 5-4 - Bar plot of gene reporter assays in adult MDS patients

Differentially methylated regions of adult MDS patients that were hypermethylated in contrast to CD34+ cells were cloned into the reporter gene vector #1341 (pCpGL-CMV/T.E1AF) or #861 (pCpGL-basic). The indicated plasmids were either unmethylated or *in vitro* SssI-methylated ("methyl") and transiently transfected into THP-1 cells. Luciferase activity was normalized against the activity of a co-transfected Renilla construct and mean values of RLUs (relative luminometer units) +/- standard deviation are shown. Significance levels were calculated using the Student´s t-test (one-tailed, unequal variance).

5.1.2 Comprehensive analysis of DNA methylation in consecutive samples

The data above demonstrated that DNA methylation changes in MDS patients are mostly private. Only few similarities between all patients could be found. In order to get further insights into disease progression, we analyzed DNA methylation in consecutive samples of individual patients. Furthermore, existing cytogenetic and genetic data were included allowing integrated analyses of all three possible causes of myelodysplastic syndromes.

The initial question how related the MDS samples are, was addressed using principal component analysis (PCA) (see Figure 5-5). This unsupervised technique reduces high dimensionality data sets to fewer dimensions allowing easier interpretation of the data[196,197].

In detail, we used the R package "methylKit"[198], that allows analysis and annotation of DNA methylation information obtained by high-throughput bisulfite sequencing. The control group comprised bisulfite sequencing data from CD34+ cells (hematopoietic progenitors), CD14+ cells (monocytes) and CD15+ cells (granulocytes). Principal component analysis included methylation data from all measured CpGs from all patients and control cells (command see section 4.2.3.1, "Correlation and principal component analysis (PCA)").

In general, the two control sets of granulocytes and monocytes were found to be in close proximity to each other illustrating the similarity between these two cell types regarding DNA methylation. In contrast, CD34+ cells were localized further away, which was not surprising since hematopoietic stem cells exhibit a different landscape of DNA methylation[199]. With one exception, the different time points of the single patients were very similar represented by their close vicinity. Patient P15 was the outlier, where the sample of time point one (P15.1) was far away from the other three samples, suggesting a bigger difference in DNA methylation. Moreover, the individual longitudinal samples of patient P53 were not as close together as consecutive samples of the other patients, suggesting more differences. In summary, longitudinal samples of single patients showed a high degree of similarity regarding DNA methylation, with one exception (time point one of patient P15). Furthermore different patients showed also a distinct DNA methylation pattern due to the different localizations over the plot.

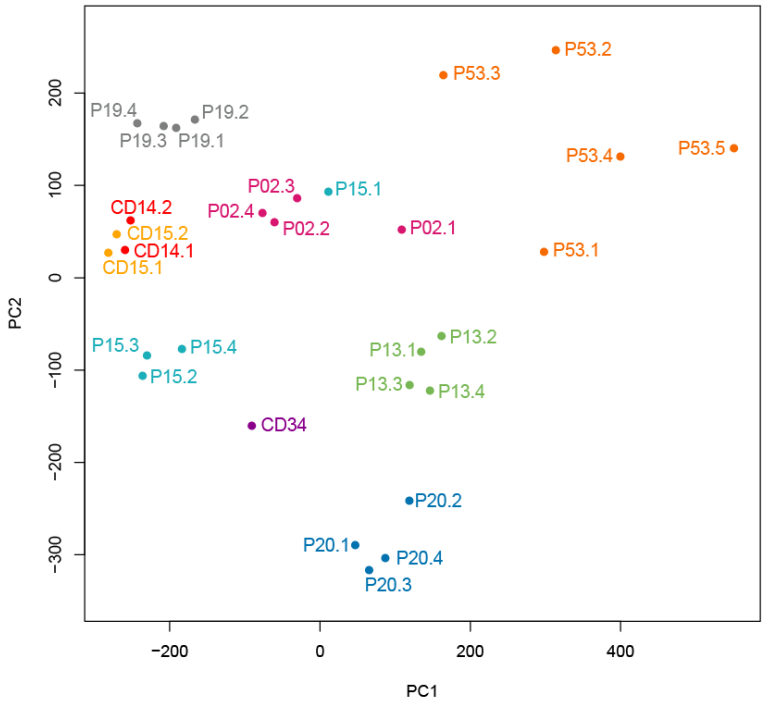


Figure 5-5 - PCA of all adult MDS patients with consecutive samples and controls

Principal component analysis (PCA) of methylation data in all measured CpGs in six different MDS patients with four or five time points each as well as CD34+ cells, CD14+ cells (monocytes) and CD15+ cells (granulocytes). Single patients were colored differently and numbers at the end of the sample labeling indicates consecutive samples of individual patients in a time-dependent manner.

5.1.2.1 Detailed analysis of patient P02

5.1.2.1.1 Analysis in comparison to CD34+ cells

Since the principal component analysis showed little correlation between all patients, we performed a detailed analysis for every single patient to shed more light on disease progression. This was implemented by two different approaches, first by comparing DNA methylation profiles of consecutive patient samples with CD34+ cells as a reference. And secondly, we compared the data across longitudinal samples of each individual patient to reveal DNA methylation changes over time.

In order to assign the similarity between CD34+ cells and samples of MDS patients, Pearson correlation of CpG methylation was calculated using the R software package “methylKit”. Percentage of CpG methylation was depicted as a scatter plot (see Figure 5-6 A) and was generated with the code described in section 4.2.3.1 (“Correlation and principal component analysis (PCA)”).

Regarding the relationship between hematopoietic stem cells and the single time points, we saw a high correlation and an even higher one between the longitudinal samples among each other.

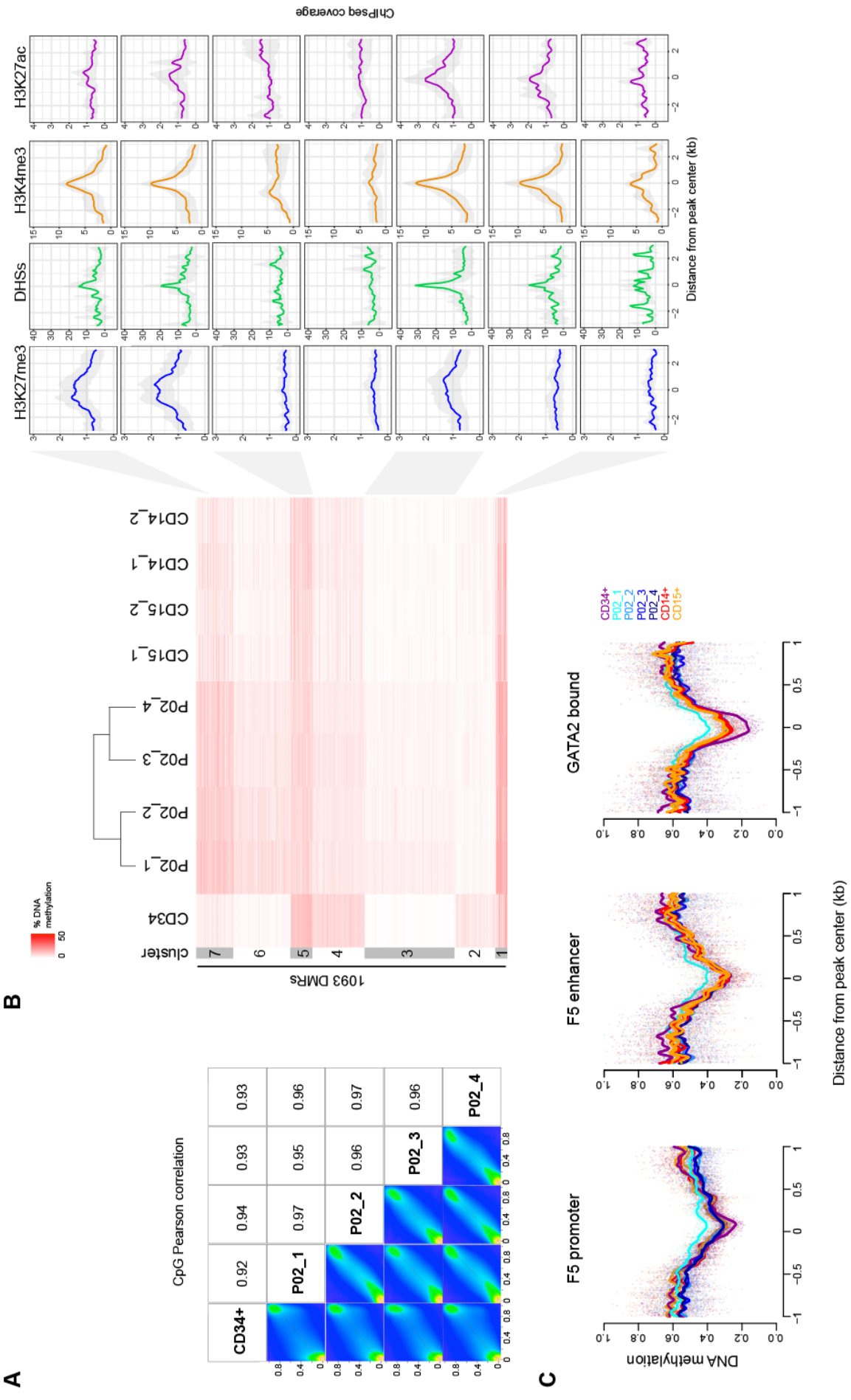

A
CpG Pearson correlation
CD34+
P02_1
P02_2
P02_3
P02_4
0.92
0.94
0.93
0.93
0.97
0.96
0.95
0.96
0.97
0.96
B
% DNA methylation
0
50
CD34
P02_1
P02_2
P02_3
P02_4
CD15_1
CD15_2
CD14_1
CD14_2
cluster
1093 DMRs
H3K27me3
DHSs
H3K4me3
H3K27ac
ChIPseq coverage
Distance from peak center (kb)
C
F5 promoter
F5 enhancer
GATA2 bound
DNA methylation
CD34+
P02_1
P02_2
P02_3
P02_4
CD14+
CD15+
Distance from peak center (kb)

Figure 5-6 – DNA methylation analysis and associated epigenetics of patient P02 in comparison to CD34+ cells

(A) Scatter plots of % CpG methylation for each pair of samples. Signal density is represented in different colors ranging from yellow to blue indicating a high and low density, respectively. Numbers on the upper right corner denote pair-wise Pearson correlation coefficients. (B) Heat map depicting *K*-means clustering of 1093 DMRs obtained in comparison to CD34+ cells as well as corresponding data of CD14+ and CD15+ cells. Dendrogram shows clustering of the patients into two groups. Each horizontal line represents a single DMR. Epigenetic marks in CD34+ cells in these DMRs are depicted on the right histograms for every cluster of regions. Average ChIPseq coverage for every cluster was drawn over a 4 kb region of the analyzed DMRs. (C) Histogram of average DNA methylation over 2 kb regions of FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions in CD34+ cells (purple), different time points of patients (different blue shades) as well as CD14+ (red) and CD15+ (yellow) cells. DNA methylation data in all observed 1093 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

In order to determine the similarities and differences between DNA methylation of the individual longitudinal samples in comparison to CD34+ cells, we identified differentially methylated regions (DMRs) between CD34+ and each time point using the software "metilene"[176]. DMRs obtained between each pair were merged into one file, annotated to the human reference genome (GRCh37/ hg19) and clustered using the *K*-means algorithm. Figure 5-6 B depicts this *K*-means clustering of 1093 DMRs regarding their DNA methylation degree.

Corresponding data of granulocytes (CD15+) and monocytes (CD14+) were also annotated across these regions and are shown in panels 6-9 of the heat map in Figure 5-6. To get some information about the analyzed DMRs, epigenetic data of CD34+ cells were annotated and depicted as histograms on the right side. Here, average ChIPseq coverage for every cluster was drawn over a 4 kb region of the analyzed DMRs. Data were available for the repressive histone mark H3K27me3, for DHSs (DNase I hypersensitive sites), H3K4me3 and H3K27ac, whereby H3K4me3 marks promoters and H3K27ac promoters and enhancers.

K-means clusters 1, 3, 6 and 7 were hypermethylated in contrast to stem cells, whereas the remaining clusters 2, 4 and 5 show a demethylated state. Considering epigenetic data of clusters 2, 4 and 5 (loss of DNA methylation, gain of H3K27ac and loss of H3K27me3 modifications), they likely correspond to regulatory elements associated with differentiation.

This was supported by the identification of distinct TF binding motifs in the analyzed clusters using HOMER. In cluster 4 we observed an enrichment of binding motifs for the transcription factors CEBP (q-value = 0.0057) and AP-1 (q-value = 0.094), both being involved in myeloid differentiation. In contrast, regions of clusters 1, 3, 6 and 7 were specific for MDS and are not associated with maturation. This statement was made due to their hypermethylation and H3K27me3 signature (except for some regions in cluster 1 showing no H3K27me3 and already marked levels of DNA methylation in CD34+ cells).

Interestingly, motif analysis in DMRs belonging to cluster 3 revealed a high frequency of GATA binding motifs (q-value = 0.0076), whereas remaining clusters 1, 6 and 7 were not significantly enriched for TF binding motifs.

Regarding DNA methylation degree in this patient, the first sample exhibited a higher extent of DNA methylation than the other time points. This was also seen in our next approach, where we looked at distribution of DNA methylation in subtypes of regions, including promoter regions (defined in the FANTOM5 promoter atlas), potential enhancer regions (defined in the FANTOM5 enhancer atlas) or those regions bound by GATA2 in HSCs. The reason for the selection of these promoter and enhancer regions is that DNA methylation influences gene expression and aberrations in this epigenetic mark may lead to malfunctions in this system. The GATA2 transcription factor is known to play a role in normal and malignant hematopoiesis[132,140] and therefore analysis of DNA methylation in regions bound by this TF could be very informative.
Regarding DNA methylation in promoter regions (see Figure 5-6 C), we could see lowest level of DNA methylation in CD34+ cells followed by CD14+ and CD15+ cells. Longitudinal samples 2, 3 and 4 displayed only a slight increase compared to monocytes and granulocytes in DNA methylation, whereas the sample from time point 1 showed the highest degree of methylation. Almost the same pattern could be observed in enhancer regions, where all samples with exception of time point 1 displayed the same degree of DNA methylation. Regions preferentially bound by GATA2 in HSCs were methylated at low levels in hematopoietic stem cells, increasing in CD14+ / CD15+ control cells and patient samples 2-4, while time point 1 exhibited again the highest DNA methylation. Moreover, average DNA methylation differences across the three region subsets were similar.

In addition to our targeted bisulfite sequencing data, we analyzed publicly available data sets obtained by whole genome bisulfite sequencing (Blueprint) from several precursor cells as well as more differentiated cells in the three region compartments (Figure 5-7). Here, we used the same regions as before (Figure 5-6 C), only adding public data instead of our own data sets. With this approach we intended to somehow class MDS patients to a certain state of differentiation. Unfortunately, data sets were obtained with two different methods (whole genome bisulfite sequencing versus targeted bisulfite sequencing) and comparison of equivalent samples between these two approaches showed strong likely platform dependent differences. Nevertheless, annotation of WGBS data gave an indication of the behavior of DNA methylation in our analyzed regions on distinct levels of differentiation.
Generally, we observed B- and T-lymphocyte samples to be methylated the most (Treg, CD8+ T cell, precursor lymphocyte of B lineage and naïve B cell), while myeloid precursor and mature cells (neutrophilic metamyelocyte and myelocyte, band form neutrophil and classical monocyte) showed a lower level of DNA methylation. Hematopoietic multipotent progenitor cells exhibited DNA methylation ratios in between these two groups. The same was seen for FANTOM5 enhancer and GATA2 bound regions, whereby regulatory T cells (Treg) and CD8+ T cells showed even higher DNA methylation levels compared to promoter regions.

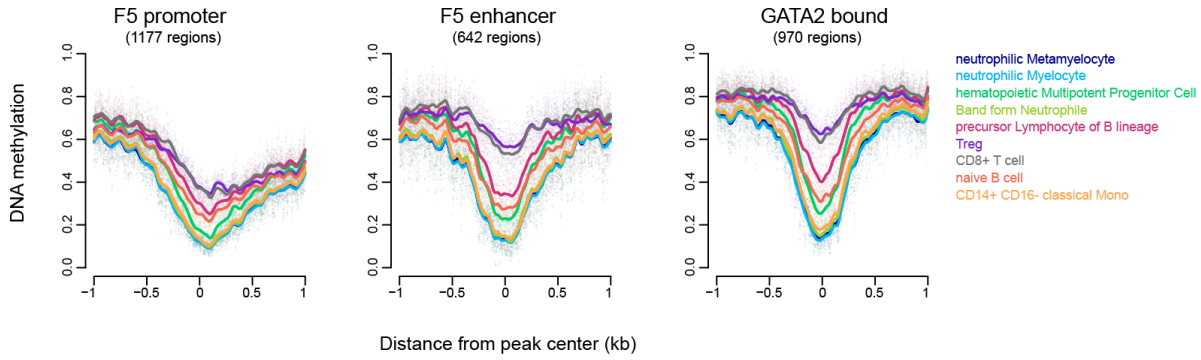


Figure 5-7 - Histogram of DNA methylation in different genomic regions using publicly available data sets (P02)

Histogram of DNA methylation ratios over 2 kb regions across FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions overlapping with DMRs found in P02 in comparison to CD34+ cells in hematopoietic precursor and more differentiated cells. Publicly available whole genome bisulfite sequencing data in the patient specific 1093 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

5.1.2.1.2 Integrated analysis of DNA methylation and genetics / cytogenetics of patient P02

So far, we analyzed differences regarding DNA methylation in comparison to hematopoietic stem cells (CD34+) resulting in a general overview. Variance between the distinct longitudinal samples could be better analyzed when comparing the samples among each other.

Without CD34+ cells as the reference, differentially methylated regions are restricted only to the longitudinal samples of the patient and changes during disease progression could be easier addressed. Pairwise comparison of samples with the software “metilene” revealed subsets of DMRs which were merged into one DMR set. For patient P02 we obtained 287 differentially methylated regions which were split into 5 clusters using *K*-means. DNA methylation data from control cells were annotated in these regions and plotted next to the patient samples (Figure 5-8). The corresponding heat map showed that consecutive samples of one patient exhibit individual DNA methylation patterns, while time point 1 and 2 look similar as well as time point 3 and 4. In order to visualize the relationships to genetic and cytogenetic data, fish plots were drawn (command see section 4.2.3.1) with corresponding clinical characteristics and therapy added below (lower part of Figure 5-8). Fish plots were used to illustrate the clonal architecture over time and were created using the R package “fishplots”. The first two consecutive samples of patient P02 were classified as MDS with del(5q) and exhibit the same clonal architecture with two coexisting clones, a *CHRM2*/del(5q) clone (green) and a *DNMT3A/RAET1G* mutated clone (blue/orange). Upon treatment with lenalidomide (LEN) the patient achieved hematologic remission and the CHRM2/del(5q) clone completely disappeared. In addition, the clone with mutated *DNMT3A* and *RAET1G* fully expanded and the patient received a phlebotomy at the same time.

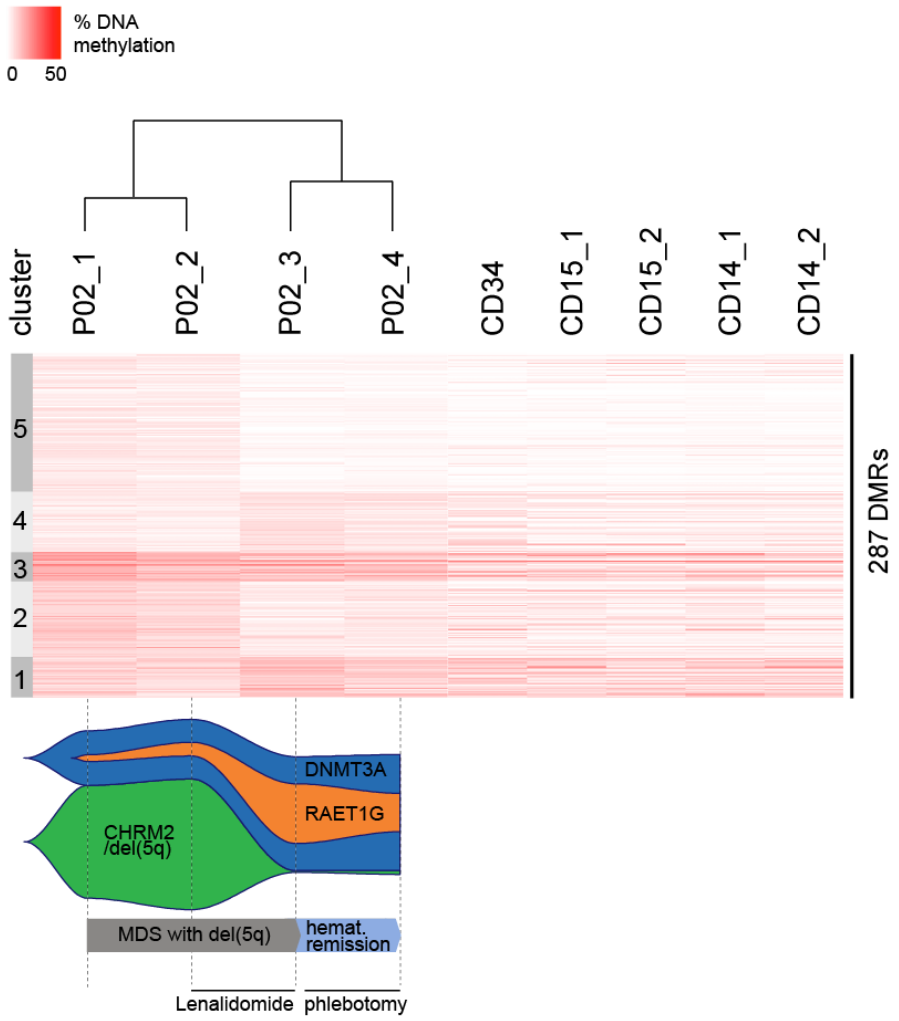


Figure 5-8 - DNA methylation analysis between time points and comparison with genetic and clinical data
The upper part depicts a heat map of *K*-means clustering of DNA methylation data from 287 DMRs between the single time points and control cells (CD34+, CD15+ and CD14+). Dendrogram shows clustering of the patients into two groups. Each horizontal line represents a single DMR while DNA methylation degree is indicated in different shades of red, ranging from 0 % methylation (white) to 50 % methylation (red). The bar on the left side shows classification of DMRs into the five different clusters. Below, fish plot displays mutational evolution during clinical follow-up at analyzed time points (marked with a dashed line). The height of the "fishes" are reflecting the variant allele frequencies of depicted clones, the higher the more alleles are mutated. Different (sub)-clones were colored separately. Clinical classification as well as therapy is drawn at the bottom of the figure.

In summary, samples of patient P02 exhibited two different genetic architectures, where the first one was characterized by the coexistence of a *CHRM2*/del(5q) clone and a *DNMT3A/RAET1G* mutated clone. At time point 3, the *CHRM2*/del(5q) clone completely disappeared due to treatment with lenalidomide and the *DNMT3A/RAET1G* clone expanded which was generally accompanied by DNA methylation changes. Also in line with the clustering in Figure 5-6 B, DNA methylation changes between consecutive samples correlated with changes in the clonal composition.

5.1.2.2 Detailed analysis of patient P13

5.1.2.2.1 Analysis in comparison to CD34+ cells

The same analyses were performed with all other patients. Patient P13 showed high Pearson correlation coefficients of CpG methylation compared to CD34+. Correlation values between the single time points of patient P13 were even higher (see Figure 5-9 A). In total, we obtained 1675 differentially methylated regions compared to CD34+ cells, which were comparted in six different clusters using *K*-means algorithm (see Figure 5-9 B). The heat map showed that patient samples at time point 1 and 2 were very similar and showed a lower DNA methylation degree than longitudinal samples 3 and 4, while the methylation pattern remained constant indicating a lower amount of aberrant cells in samples 1 and 2. Increasing DNA methylation in comparison to hematopoietic stem cells and an H3K27me3 signature could be observed in clusters 1, 2, 4 and 6. The remaining cluster 3 and 5 showed demethylation compared to CD34+ cells. The corresponding epigenetic signature, like loss of DNA methylation and H3K27me3 as well as gain of H3K27ac suggested an association with differentiation. Motif analysis in those six clusters of DMRs only revealed a GATA signature (q-value = 1e-9) in cluster 6, while clusters 1, 2 and 4 didn´t show significant enrichment for TF binding motifs.
The histogram of DNA methylation data in FANTOM5 promoter regions showed an overlap of all control cells (CD34+, CD14+ and CD15+) and a higher DNA methylation degree in the patient (Figure 5-9 C). There, time point 1 and 2 displayed a lower methylation compared to time point 3 and 4, which was already pointed out in the heat map of Figure 5-9 B. A similar picture could be seen in F5 enhancer regions, while in GATA2 bound regions CD34+ cells were preferentially demethylated, as already observed with patient P02.

Figure 5-10 depicts DNA methylation of publicly available data sets of precursor and mature cells in regions described above (Figure 5-9 C). As observed for patient P02 in F5 promoter regions, lymphoid samples (precursor lymphocyte of B lineage, Treg, CD8+ T cell) showed the highest DNA methylation degree, followed by naïve B cells and hematopoietic multipotent progenitor cells. The three myeloid samples (neutrophilic metamyelocyte and myelocyte, band from neutrophil) were the least methylated. The same picture was obtained for FANTOM5 enhancer regions and GATA2 bound regions, whereas only DNA methylation extent is a bit higher in these two compartments than in promoter regions.

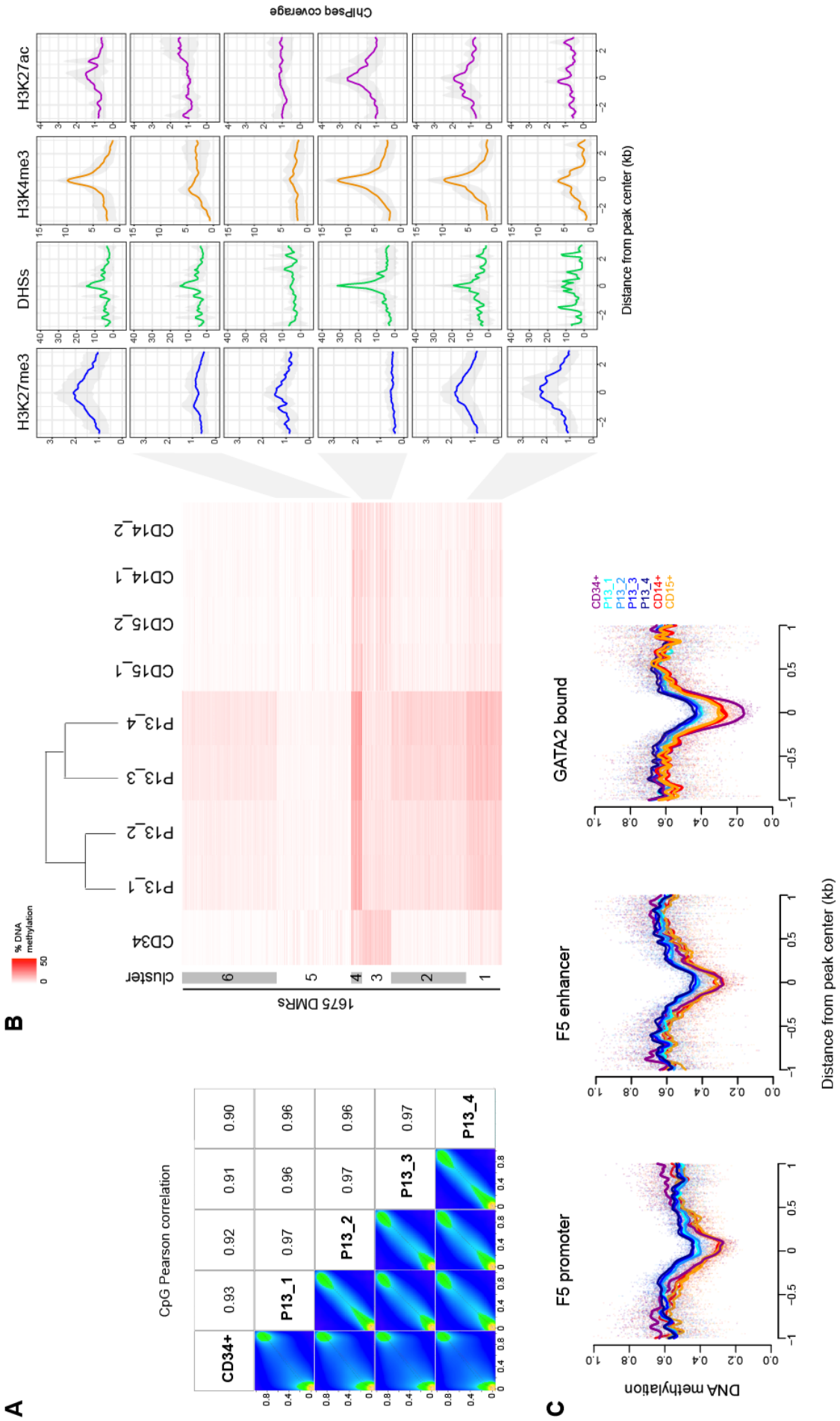

A
CpG Pearson correlation
CD34+
P13_1
P13_2
P13_3
P13_4
0.93
0.92
0.91
0.90
0.97
0.96
0.96
0.97
0.96
0.97
B
% DNA methylation
0 50
cluster
CD34
P13_1
P13_2
P13_3
P13_4
CD15_1
CD15_2
CD14_1
CD14_2
1675 DMRs
H3K27me3
DHSs
H3K4me3
H3K27ac
ChIPseq coverage
Distance from peak center (kb)
C
F5 promoter
F5 enhancer
GATA2 bound
CD34+
P13_1
P13_2
P13_3
P13_4
CD14+
CD15+
DNA methylation
Distance from peak center (kb)

Figure 5-9 - DNA methylation analysis and associated epigenetics of patient P13 in comparison to CD34+ cells

(A) Scatter plots of % CpG methylation for each pair of samples. Signal density is represented in different colors ranging from yellow to blue indicating a high and low density, respectively. Numbers on the upper right corner denote pair-wise Pearson correlation coefficients. (B) Heat map depicting *K*-means clustering of 1675 DMRs obtained in comparison to CD34+ cells as well as corresponding data of CD14+ and CD15+ cells. Dendrogram shows clustering of the patients into two groups. Each horizontal line represents a single DMR. Epigenetic marks in CD34+ cells in these DMRs are depicted on the right histograms for every cluster of regions. Average ChIPseq coverage for every cluster was drawn over a 4 kb region of the analyzed DMRs. (C) Histogram of average DNA methylation over 2 kb regions of FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions in CD34+ cells (purple), different time points of patients (different blue shades) as well as CD14+ (red) and CD15+ (yellow) cells. DNA methylation data in all observed 1675 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

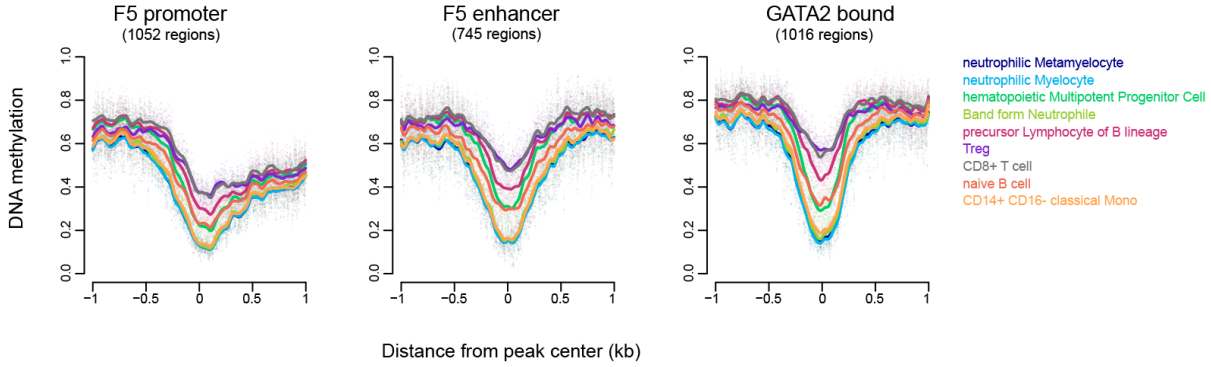


Figure 5-10 - Histogram of DNA methylation in different genomic regions using publicly available data sets (P13)

Histogram of DNA methylation ratios over 2 kb regions across FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions overlapping with DMRs found in patient P13 in comparison to CD34+ cells in hematopoietic precursor and more differentiated cells. Publicly available whole genome bisulfite sequencing data in the patient specific 1675 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

Patient P13 exhibited 348 regions differentially methylated among all longitudinal samples, which were clustered into four groups using the *K*-means algorithm. Overall, the intensities of DNA methylation signals were weaker for time points 1 and 2 compared to 3 and 4 (Figure 5-11).

Regarding clinical classifications during the analyzed time period, progression from MDS with del(5q) (time point 1) over MDS with excess blast (time points 2 and 3) until transition into secondary AML (time point 4) was observed. Genetic data provided from the group in Mannheim made it possible to draw a fish plot with variant allele frequencies of mutations. Cells with PML lesions were shown to be the founder clone and VAFs were very stable over all 4 time points. A *TET2* mutated subclone arose from this *PML* clone, which was more or less stable during progression. This *TET2* clone exhibited acquisition of two different subclones with *TNIK* or del(5q) lesions, respectively, while *SF3B1*/*TP53*

mutations appeared in aberrant del(5q) cells as a further subclone. Over time, VAFs for SF3B1/TP53 and del(5q) increased with climax at time point 3, while frequency of cells with *TNIK* lesions slightly decreased, indicating a marked clonal shift between samples 2 and 3.

In summary, patient P13 did not acquire novel mutations during disease progression, only VAFs of the different lesions varied on a small scale, which could be seen on DNA methylation level.

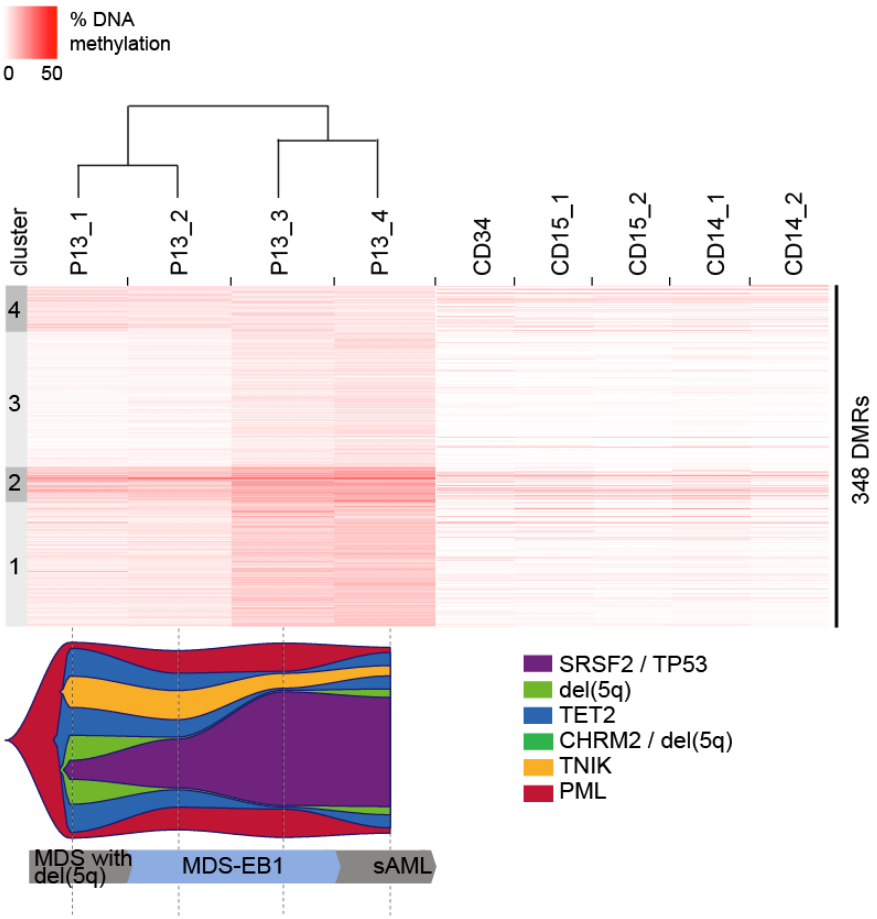


Figure 5-11 - DNA methylation analysis of patient P13 and comparison with genetic and clinical data

The upper part depicts a heat map of *K*-means clustering of DNA methylation data from 348 DMRs between the single time points and control cells (CD34+, CD15+ and CD14+). Dendrogram shows clustering of the patients into two groups. Each horizontal line represents a single DMR while DNA methylation degree is indicated in different shades of red, ranging from 0 % methylation (white) to 50 % methylation (red). The bar on the left side shows classification of DMRs into the five different clusters. Below, fish plot displays mutational evolution during clinical follow-up at the analyzed time points (marked with a dashed line). The height of the “fishes” are reflecting the variant allele frequencies of depicted clones, the higher the more alleles are mutated. Different (sub)-clones were colored separately. Clinical classification as well as therapy is drawn at the bottom of the figure.

5.1.2.3 Detailed analysis of patient P15

5.1.2.3.1 Analysis in comparison to CD34+ cells

Comparison of Pearson correlation coefficients of CpG methylation between CD34+ cells and longitudinal samples of patient P15 revealed lower values compared to patients P02 and P13. A generally higher correlation was obtained by comparison of the patient samples amongst each other (Figure 5-12 A). Interestingly we noticed a lower correlation between time point 1 and all other time points suggesting larger differences in CpG methylation. This phenomenon was also seen in the *K*-means clustering of 2208 differentially methylated regions that were found between all consecutive samples of patient P15 in comparison to CD34+ cells (see Figure B). The sample of time point 1 showed a clear decrease of DNA methylation in contrast to longitudinal samples 2-4, which exhibited almost the same pattern. The first sample was taken under treatment with lenalidomide and therefore lower DNA methylation degree was probably due to the reduction of aberrant cells.

The six different *K*-means clusters of DMRs could be divided into clusters associated with differentiation (cluster number 3 and 4) and those specific for MDS (clusters 1, 2, 5 and 6).

Differentiation specific clusters were demethylated in patients compared to CD34+ cells and showed low signals of H3K27 trimethylation. In addition, cluster 4 exhibited histone marks associated with regulatory activity, namely H3K4me3 and H3K27ac. The remaining DMR clusters showed higher DNA methylation than in CD34+ cells and were already repressed in hematopoietic stem cells by H3K27me3. Furthermore these regions showed active histone H3K4 trimethylation supposing preferentially primer regions that were hypermethylated. Findings of motif analysis in these *K*-means clusters showed that cluster 3 was linked with binding motifs for CEBP (q-value = 1e-8) and AP-1 (q-value = 0.009) and also cluster 4 was linked with CEBP motifs (q-value = 1e-8), both transcription factors necessary for normal differentiation processes.

Analysis of DNA methylation degree in F5 promoter regions (Figure 5-12 C) revealed a generally higher DNA methylation of patient samples than control cells (CD34+, CD14+ and CD15+). Remarkably, the first sample time point showed a lower DNA methylation than all other time points, but still a slightly higher one than control cells. Enhancer regions were methylated to a similar extent regarding patient samples as well as control cells. Regions which were preferentially bound by GATA2 exhibited the lowest DNA methylation in CD34+ cells, whereas signals of all other samples and controls (CD14+ and CD15+) were found at almost the same level, but generally higher than in stem cells.

In summary, this suggested that aberrant MDS cells in this specific patient were similar to monocytes and granulocytes regarding their differentiation level apparently shown by complete overlap of DNA methylation signals in GATA2 bound regions.

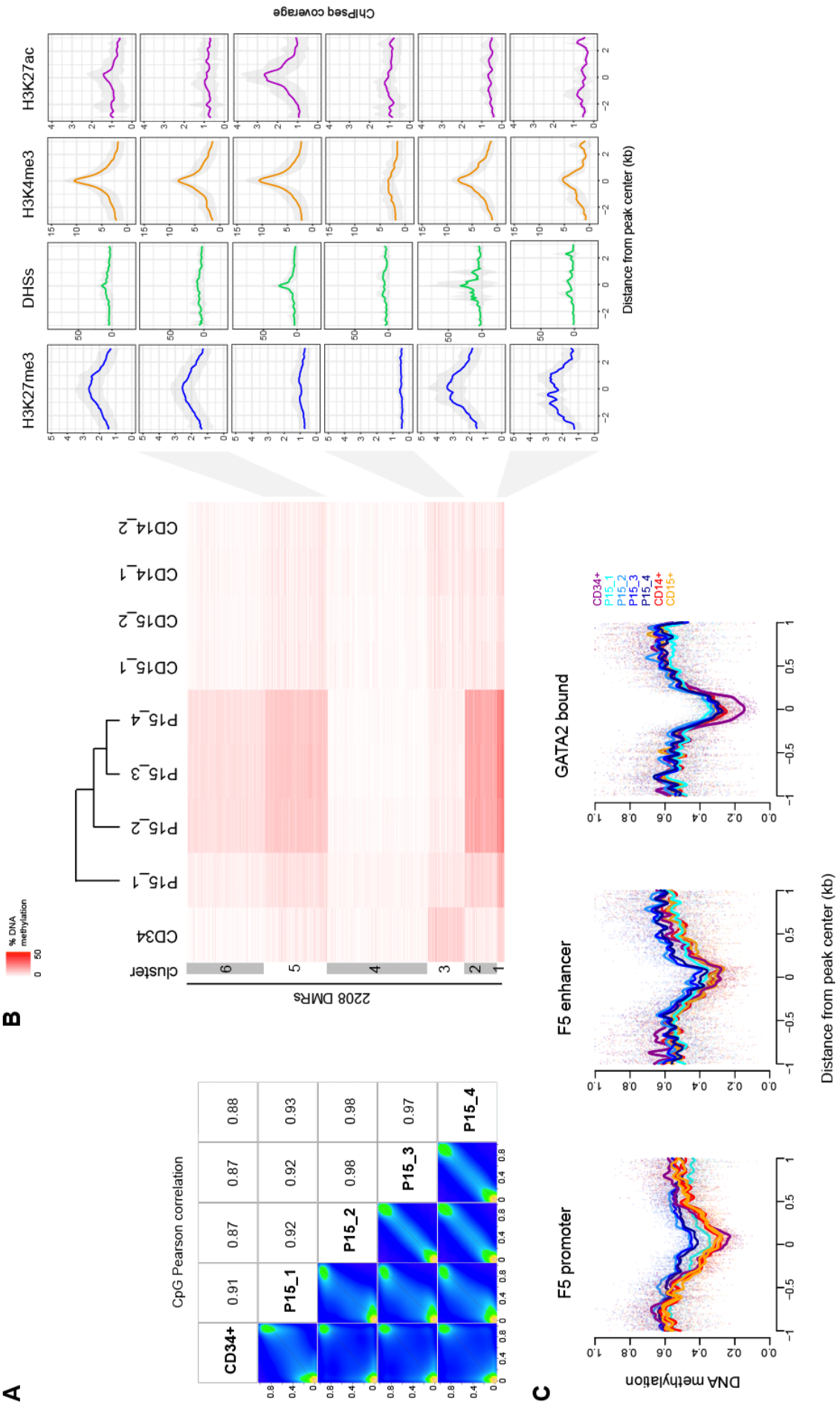

A
CD34+
P15_1
P15_2
P15_3
P15_4
CpG Pearson correlation
0.91
0.87
0.87
0.92
0.92
0.98
0.98
0.97
0.88
0.93
0.98
B
% DNA methylation
0
50
cluster
CD34
P15_1
P15_2
P15_3
P15_4
CD15_1
CD15_2
CD14_1
CD14_2
2208 DMRs
H3K27me3
DHSs
H3K4me3
H3K27ac
ChIPseq coverage
Distance from peak center (kb)
C
F5 promoter
F5 enhancer
GATA2 bound
DNA methylation
CD34+
P15_1
P15_2
P15_3
P15_4
CD14+
CD15+
Distance from peak center (kb)

Figure 5-12 - DNA methylation analysis and associated epigenetics of patient P15 in comparison to CD34+ cells

(A) Scatter plots of % CpG methylation for each pair of sample. Signal density is represented in different colors ranging from yellow to blue indicating a high and low density, respectively. Numbers on the upper right corner denote pair-wise Pearson correlation coefficients. (B) Heat map depicting *K*-means clustering of 2208 DMRs obtained in comparison to CD34+ cells as well as corresponding data of CD14+ and CD15+ cells. Dendrogram shows clustering of the patients into two groups, while sample 1 forms a separate group. Each horizontal line represents a single DMR. Epigenetic marks in CD34+ cells in these DMRs are depicted on the right histograms for every cluster of regions. Average ChIPseq coverage for every cluster was drawn over a 4 kb region of the analyzed DMRs. (C) Histogram of average DNA methylation over 2 kb regions of FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions in CD34+ cells (purple), different time points of patients (different blue shades) as well as CD14+ (red) and CD15+ (yellow) cells. DNA methylation data in all observed 2208 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

In those differentially methylated regions obtained for patient P15 which overlapped with regulatory features (F5 promoter, F5 enhancer, GATA2 binding), we noticed again a higher DNA methylation degree in T cells (Treg and precursor lymphocyte of B lineage) (see Figure 5-13). Regarding F5 promoter regions, the remaining data sets showed a similar extent, while CD8+ T cells and naïve B cells were slightly more methylated. In FANTOM5 enhancer and GATA2 bound regions, the signals were more separate, but the order due to DNA methylation degree was similar to patients shown before.

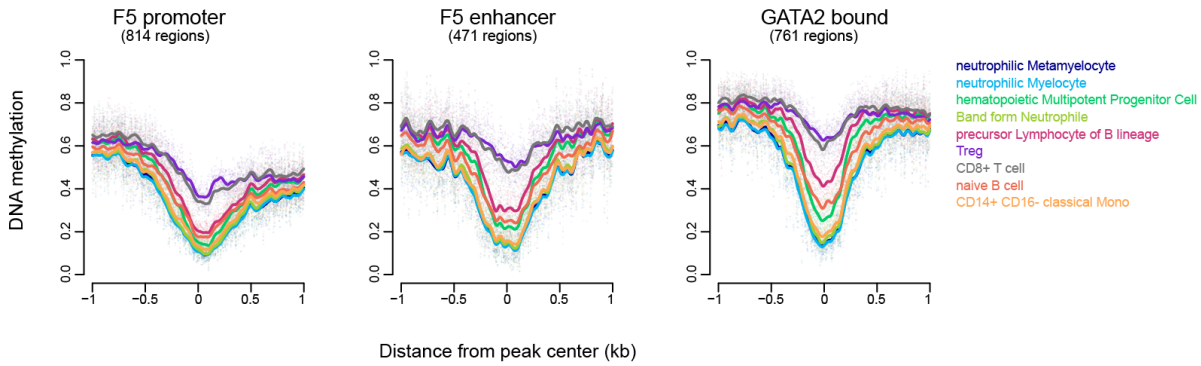


Figure 5-13 - Histogram of DNA methylation in different genomic regions using publicly available data sets (P15)

Histogram of DNA methylation ratios over 2 kb regions across FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions overlapping with DMRs found in P02 in comparison to CD34+ cells in hematopoietic precursor and more differentiated cells. Publicly available whole genome bisulfite sequencing data in the patient specific 1675 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

5.1.2.3.2 Integrated analysis of DNA methylation and genetics / cytogenetics of patient P15

All merged DMRs of patient P15 obtained by pairwise comparison of the time points, were divided into five clusters using *K*-means algorithm (Figure 5-14).

Consecutive samples 2-4 showed a very similar DNA methylation pattern across those regions, while time point 1 was methylated to a lower extent in all clusters. This might be due to the fact that the first sample was taken under treatment with lenalidomide, where a reduction of aberrant cells was already observed.

The following time points were clinically classified as MDS with multilineage dysplasia. Genetic evolution during disease progression was depicted in the fish plot, where *ASXL1* was identified to be the founder clone. Sequential acquisition of subclonal lesions could be noticed, while time point 1 was characterized by the presence of *ASXL1*, *EZH2*/del(5q), *CSNK1A1*/*NF1* and *EZH2*/5qUPD. Lesions in *ETV6* were arising between the first and second time point. After the second longitudinal sample, the *ETV6* mutated subclone was fully substituted by an independent branching subclone carrying monosomy 7/*RUNX1*/*ETV6* lesion.

Altogether, the composition of genetic lesions varied between time point 1 and the other consecutive samples of the patient, whereas an additional subclone arose at time point 3. But apparently this new subclone seemed to have no effect on DNA methylation patterns perhaps due to low variant allele frequency.

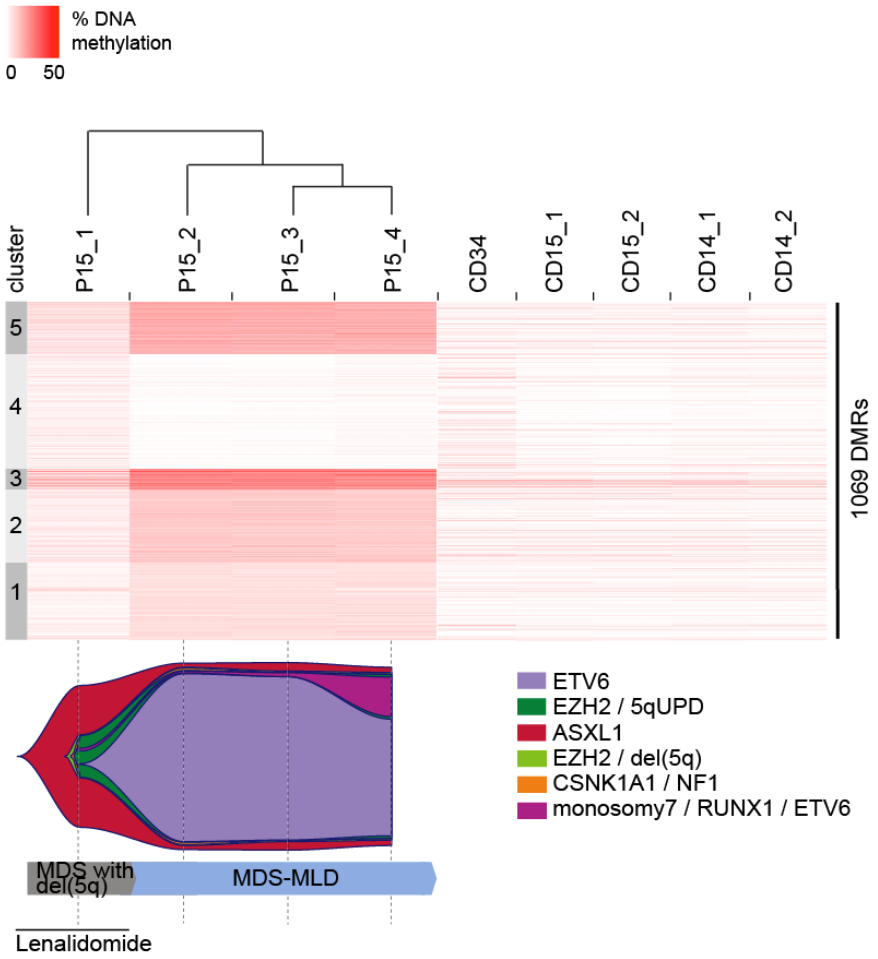


Figure 5-14 - DNA methylation analysis of patient P15 and comparison with genetic and clinical data

The upper part depicts a heat map of *K*-means clustering of DNA methylation data from 1069 DMRs between the single time points and control cells (CD34+, CD15+ and CD14+). Dendrogram shows clustering of the patients into two groups, while sample one represents a single group. Each horizontal line represents a single DMR while DNA methylation degree is indicated in different shades of red, ranging from 0 % methylation (white) to 50 % methylation (red). The bar on the left side shows classification of DMRs into the five different clusters. Below, fish plot displays mutational evolution during clinical follow-up at analyzed time points (marked with a dashed line). The height of the "fishes" are reflecting the variant allele frequencies of depicted clones, the higher the more alleles are mutated. Different (sub)-clones were colored separately. Clinical classification as well as therapy is drawn at the bottom of the figure.

5.1.2.4 Detailed analysis of patient P19

5.1.2.4.1 Analysis in comparison to CD34+ cells

CpG methylation between hematopoietic stem cells and consecutive samples of patient P19 showed correlation values between 0.91 and 0.92. Almost perfect consistency could be observed between the single time points of the patient (Figure 5-15 A).

Pairwise comparison of DNA methylation degree between CD34+ cells and longitudinal samples resulted in 940 merged DMRs dividing into 8 different clusters using *K*-means algorithm (Figure 5-15 B). Clusters 1, 3, 4 and 7 were epigenetically characterized with signals for H3K27me3, H3K4me3 and H3K27ac, while H3K27 acetylation was low in cluster 1 and 7. Furthermore DNA methylation increased in these clusters in contrast to CD34+ cells suggesting those regions to be disease specific. The opposite could be found in clusters 2, 5, 6 and 8, where regions were hypomethylated and epigenetic signature reveals a differentiation associated phenotype (loss of H3K27me3 and gain of H3K27ac). Significantly enriched binding motifs for transcription factors important for differentiation (CEBP, AP-1) could be found via motif analysis with HOMER in clusters 2, 5, 6 and 8. Furthermore, cluster 1 comprised regions which were found to be preferentially bound by HOX transcription factors.

As described above, DNA methylation degree and pattern between the single time points of the patient were very similar. This could also be shown by analyzing DNA methylation in the three different region compartments, F5 promoter, F5 enhancer and GATA2 bound regions.

Here, signals for the longitudinal samples lay perfectly on top of each other, while located a little bit higher than control cells (Figure 5-15 C). This is not the case for GATA2 bound regions, where patient samples show the same DNA methylation compared to CD14+ and CD15+ cells, but CD34+ cells exhibit a lower DNA methylation extent. Taken together, there were almost no differences in DNA methylation between the single time points of the patient. However, changes could be observed in comparison to CD34+ cells, but to a lesser extent due to the small number of differentially methylated regions.

Histograms in Figure 5-16 depict DNA methylation data in the above mentioned region compartments (F5 promoter/enhancer, GATA2 bound regions) of precursor and mature blood cells obtained from publicly available databases. FANTOM5 promoter regions overlapping with DMRs found between CD34+ cells and patient samples showed a similar DNA methylation degree in different cell types with regulatory T cells (Treg) and CD8+ T cells having the highest one. Signals in F5 enhancer regions were somewhat scattered probably due to the small number of regions.

Nevertheless, we could observe the same picture like in promoter regions. GATA2 bound regions were generally higher methylated than the other two region compartments and displayed three cell types exhibiting the highest DNA methylation extent (Treg, CD8+ T cell and precursor lymphocyte of B lineage). Naïve B cells were a little less methylated followed by the remaining data sets (neutrophilic metamyelocyte and myelocyte, band form neutrophil, classical monocyte and hematopoietic multipotent progenitor cell).

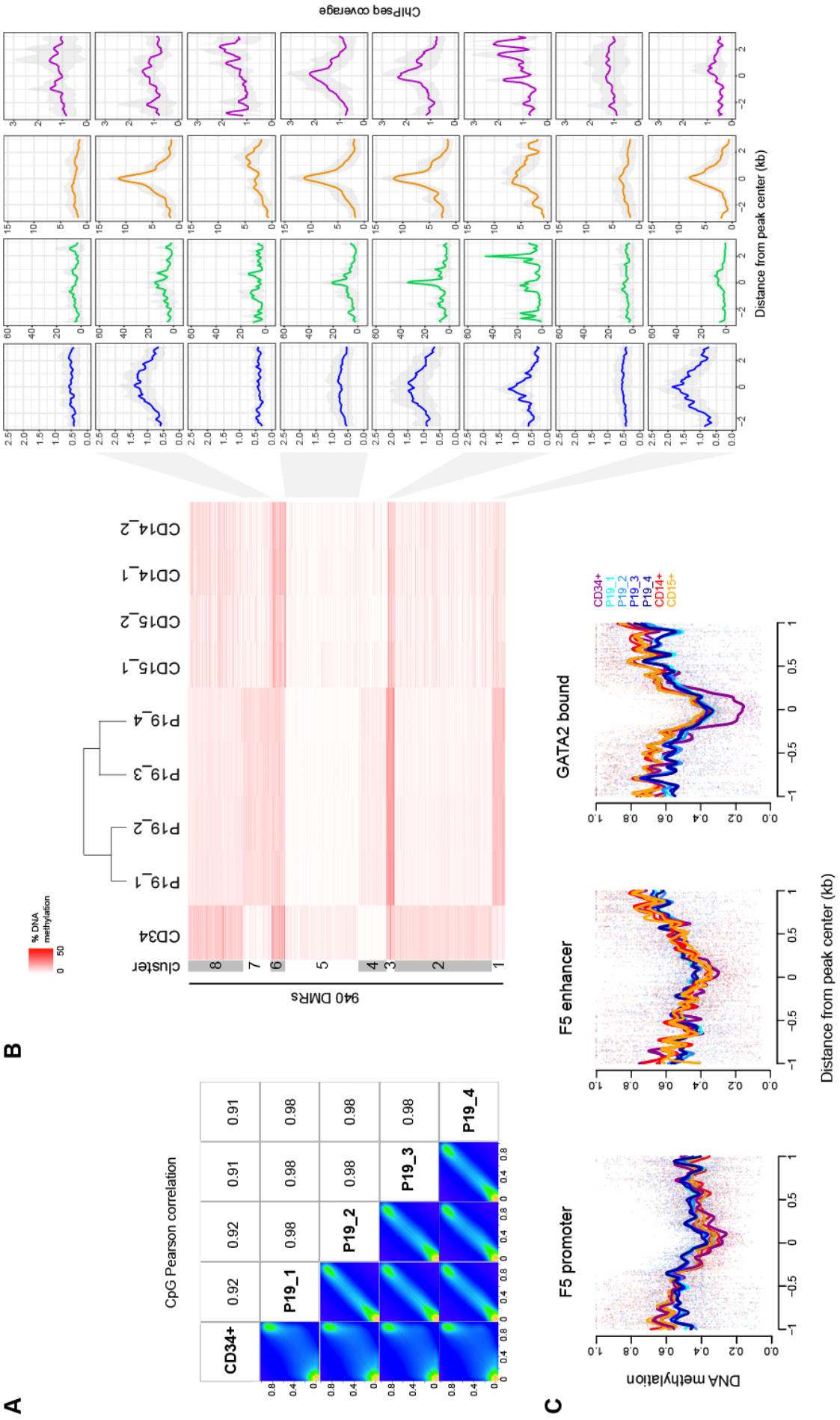

A
CpG Pearson correlation
CD34+
P19_1
P19_2
P19_3
P19_4
0.92
0.91
0.98
B
% DNA methylation
0
50
940 DMRs
cluster
CD34
P19_1
P19_2
P19_3
P19_4
CD15_1
CD15_2
CD14_1
CD14_2
H3K27me3
DHSs
H3K4me3
H3K27ac
ChIPseq coverage
Distance from peak center (kb)
C
F5 promoter
F5 enhancer
GATA2 bound
DNA methylation
CD34+
P19_1
P19_2
P19_3
P19_4
CD14+
CD15+
Distance from peak center (kb)

Figure 5-15 - A methylation analysis and associated epigenetics of patient P19 in comparison to CD34+ cells

(A) Scatter plots of % CpG methylation for each pair of samples. Signal density is represented in different colors ranging from yellow to blue indicating a high and low density, respectively. Numbers on the upper right corner denote pair-wise Pearson correlation coefficients. (B) Heat map depicting *K*-means clustering of 940 DMRs obtained in comparison to CD34+ cells as well as corresponding data of CD14+ and CD15+ cells. Dendrogram shows clustering of the patients into two groups. Each horizontal line represents a single DMR. Epigenetic marks in CD34+ cells in these DMRs are depicted on the right histograms for every cluster of regions. Average ChIPseq coverage for every cluster was drawn over a 4 kb region of the analyzed DMRs. (C) Histogram of average DNA methylation over 2 kb regions of FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions in CD34+ cells (purple), different time points of patients (different blue shades) as well as CD14+ (red) and CD15+ (yellow) cells. DNA methylation data in all observed 940 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

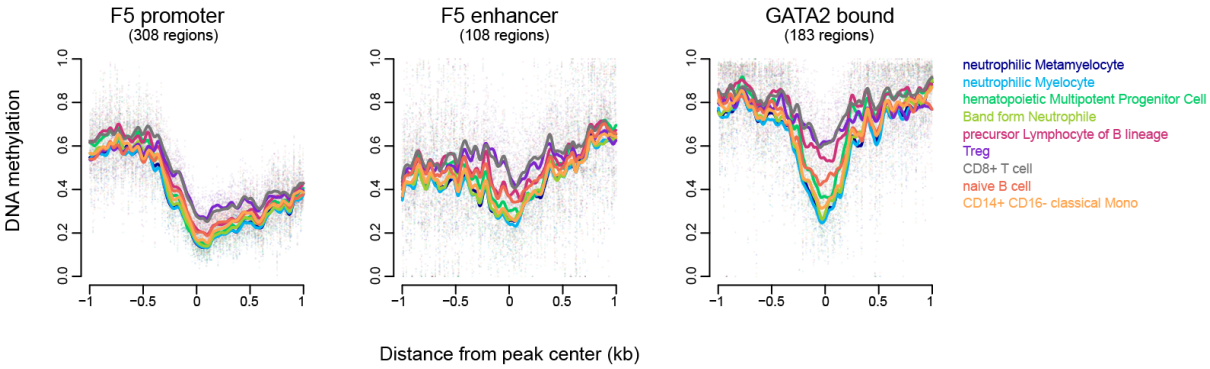


Figure 5-16 -Histogram of DNA methylation in different genomic regions using publicly available data sets (P19)

Histogram of DNA methylation ratios over 2 kb regions across FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions overlapping with DMRs found in P02 in comparison to CD34+ cells in hematopoietic precursor and more differentiated cells. Publicly available whole genome bisulfite sequencing data in the patient specific 1675 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

5.1.2.4.2 Integrated analysis of DNA methylation and genetics/cytogenetics of patient P19

In order to get a more detailed view of the differences between the consecutive samples, we identified DMRs between the time points and did a comparative analysis with genetic data (Figure 5-17). DMRs were clustered into 4 compartments using the *K*-means algorithm, while regions displayed few changes regarding DNA methylation over time. At each time point, the patient was classified as MDS with single lineage dysplasia and ring sideroblasts (MDS-RS-SLD) and the only therapy obtained was APG101 application at time point 1. APG101 was found to rescue erythropoiesis in low risk MDS patients with severe impairment of hematopoiesis[200]. This fusion protein consisting of the extracellular

domain of human CD95 (Fas receptor) and the Fc region of IgG1 binds to CD95L on target cells and in solution, thus inhibiting activation of CD95 mediated apoptosis.

The fish plot displayed the clonal evolution during disease progression, where *ASXL1*/*SF3B1* was determined as founder clone. Two different subclones emerged before sampling of time point 1 with lesions in *TTBK1* and del(ETV6). Over time, the four different patient samples showed almost the same VAF. Taken together, constancy of both, clinical classification and genetic landscape, probably coincide the low variance of DNA methylation between the single sample time points of the patient.

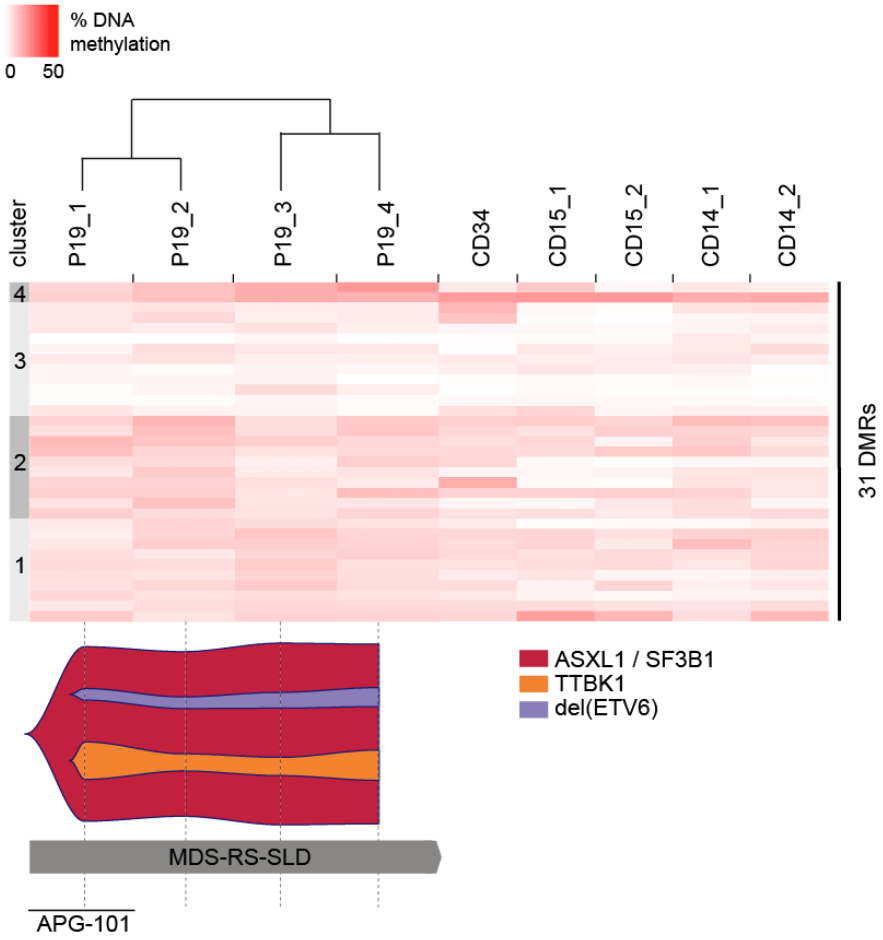


Figure 5-17 - DNA methylation analysis of patient P19 and comparison with genetic and clinical data

The upper part depicts a heat map of *K*-means clustering of DNA methylation data from 31 DMRs between the single time points and control cells (CD34+, CD15+ and CD14+). Dendrogram shows clustering of the patients into two groups. Each horizontal line represents a single DMR while DNA methylation degree is indicated in different shades of red, ranging from 0 % methylation (white) to 50 % methylation (red). The bar on the left side shows classification of DMRs into the five different clusters. Below, fish plot displays mutational evolution during clinical follow-up at analyzed time points (marked with a dashed line). The height of the “fishes” are reflecting the variant allele frequencies of depicted clones, the higher the more alleles are mutated. Different (sub)-clones were colored separately. Clinical classification as well as therapy is drawn at the bottom of the figure.

5.1.2.5 Detailed analysis of patient P20

5.1.2.5.1 Analysis in comparison to CD34+ cells

Similar patient characteristics as described above (patient P19), were observed in the analyses of patient P20. When comparing the single samples of different time points amongst each other, almost no differences in DNA methylation could be observed (see Figure 5-18 A). However, correlation values between CD34+ cells and patient samples were lower. The heat map in Figure 5-18 B represents DNA methylation degree in 975 differentially methylated regions in comparison to hematopoietic stem cells. There, the similarity between the single patient samples is well illustrated, while some slight differences were observed in sample of time point 2. Using the *K*-means algorithm, regions were split into 8 different clusters. Similar region clusters based on DNA methylation and epigenetic signature in HSCs were number 1, 5 and 7 as well as 2, 3, 4, 6 and 8.

The first group is characterized by loss of DNA methylation in the patient together with loss of H3K27me3. Clusters 1 and 5 additionally gained H3K27ac, all in all supposing to be regions affected by differentiation. Results of motif finding only revealed significant enrichment of CEBP (q-value = 0.005) as transcription factor important for differentiation processes. The second group of region clusters (2, 3, 4, 6, 8) became methylated in patient samples in comparison to CD34+ cells and showed an H3K27me3 signature. Some of those clusters also exhibited a signal for the activating histone mark H3K4me3, indicating that these regions were specific for MDS. Moreover, cluster 4 was significantly enriched for GATA binding motifs.

The high correlation between the four samples of the patient was also reflected in DNA methylation analysis in three different region compartments (Figure 5-18 C). DMRs overlapping with FANTOM5 promoter or enhancer regions displayed higher DNA methylation degree in all patient samples compared to control cells (CD34+, CD14+ and CD15+), whereas signals of the consecutive time points lied on top of each other. Overlapping GATA2 bound regions were found to be less methylated in CD34+ cells, followed by CD14+/CD15+ cells and patient samples.

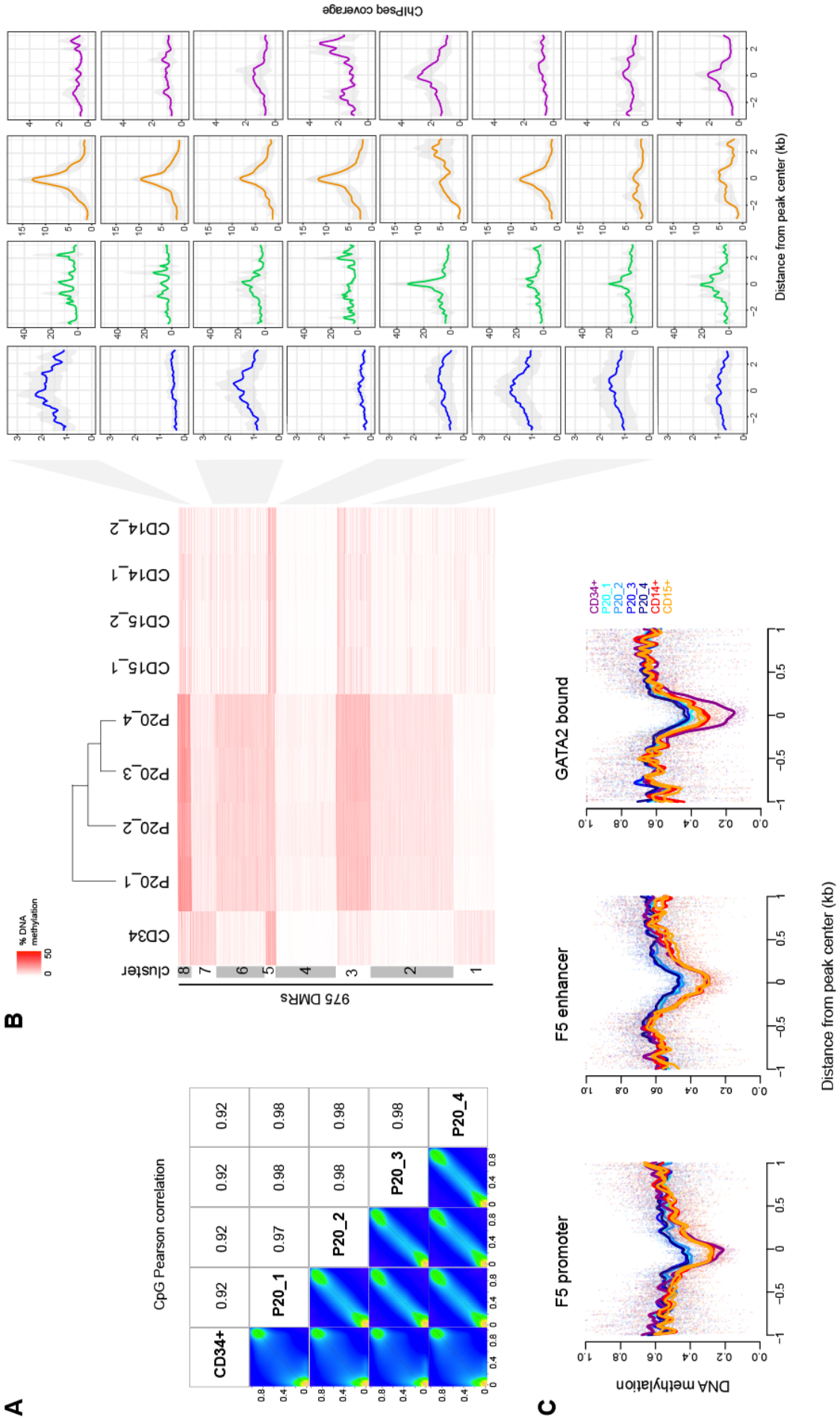

A
CpG Pearson correlation
CD34+
P20_1
P20_2
P20_3
P20_4
B
% DNA methylation
cluster
975 DMRs
CD34
P20_1
P20_2
P20_3
P20_4
CD15_1
CD15_2
CD14_1
CD14_2
H3K27me3
DHSs
H3K4me3
H3K27ac
ChIPseq coverage
Distance from peak center (kb)
C
F5 promoter
F5 enhancer
GATA2 bound
CD34+
P20_1
P20_2
P20_3
P20_4
CD14+
CD15+
DNA methylation
Distance from peak center (kb)

Figure 5-18 - DNA methylation analysis and associated epigenetics of patient P20 in comparison to CD34+ cells

(A) Scatter plots of % CpG methylation for each pair of samples. Signal density is represented in different colors ranging from yellow to blue indicating a high and low density, respectively. Numbers on the upper right corner denote pair-wise Pearson correlation coefficients. (B) Heat map depicting *K*-means clustering of 975 DMRs obtained in comparison to CD34+ cells as well as corresponding data of CD14+ and CD15+ cells. Dendrogram shows clustering of the patients into two groups with sample 1 in the first group and samples 2,3 and 4 in the second group. Each horizontal line represents a single DMR. Epigenetic marks in CD34+ cells in these DMRs are depicted on the right histograms for every cluster of regions. Average ChIPseq coverage for every cluster was drawn over a 4 kb region of the analyzed DMRs. (C) Histogram of average DNA methylation over 2 kb regions of FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions in CD34+ cells (purple), different time points of patients (different blue shades) as well as CD14+ (red) and CD15+ (yellow) cells. DNA methylation data in all observed 975 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

A very uniform DNA methylation degree is observed in F5 promoter regions which were overlapping with DMRs found between CD34+ cells and longitudinal patient samples of P20 (Figure 5-19).
Different maturation states as well as cells from distinct hematopoietic lineages showed almost the same DNA methylation signal in those regions. Regarding F5 enhancer regions, we saw the same characteristic picture as in the samples before. Regulatory T cells (Treg), CD8+ T cells and precursor of B lineage were methylated the most followed by naïve B cells and hematopoietic multipotent progenitor cells. Signals from classical monocytes, neutrophilic metamyelocytes and myelocytes as well as band form neutrophils were found to be on top of each other and methylated about one half of Tregs. The same is true for GATA2 bound regions with the difference that all data sets showed a generally higher DNA methylation degree.

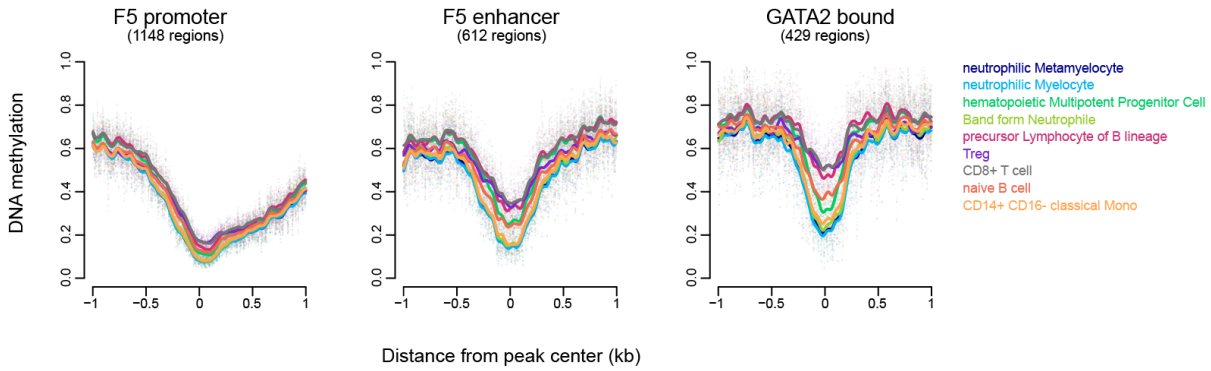


Figure 5-19 -Histogram of DNA methylation in different genomic regions using publicly available data sets (P20)

Histogram of DNA methylation ratios over 2 kb regions across FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions overlapping with DMRs found in P02 in comparison to CD34+ cells in hematopoietic precursor and more differentiated cells. Publicly available whole genome bisulfite sequencing data in the patient specific 1675 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

5.1.2.5.2 Integrated analysis of DNA methylation and genetics/cytogenetics of patient P20

Comparison of DNA methylation in DMRs between the longitudinal samples allowed *K*-means clustering into six compartments (Figure 5-20). Patterns across these six clusters were, with exception of time point 2, very similar. The second sample showed a lower DNA methylation extent in clusters 5 and 6 and a higher one in clusters 1 and 2. Considering clinical classification and genetic landscape over time, we saw a very stable disease in patient P20. Classification of MDS-EB1 did not change. The existing genetic lesions *IDH2*/*SRSF2* were identified as founder clone with subsequent acquisition of *SPEG2*, *BRCC3* and *NF1* mutated subclones. However, regarding the clonal evolution during disease progression, only small changes in subclonal VAFs were observed, while composition was the same. To sum up, DNA methylation during disease progression was consistent and accompanied by a stable clinical subtype and almost invariable genetic landscape, despite treatment with lenalidomide and APG101 at two different time points.

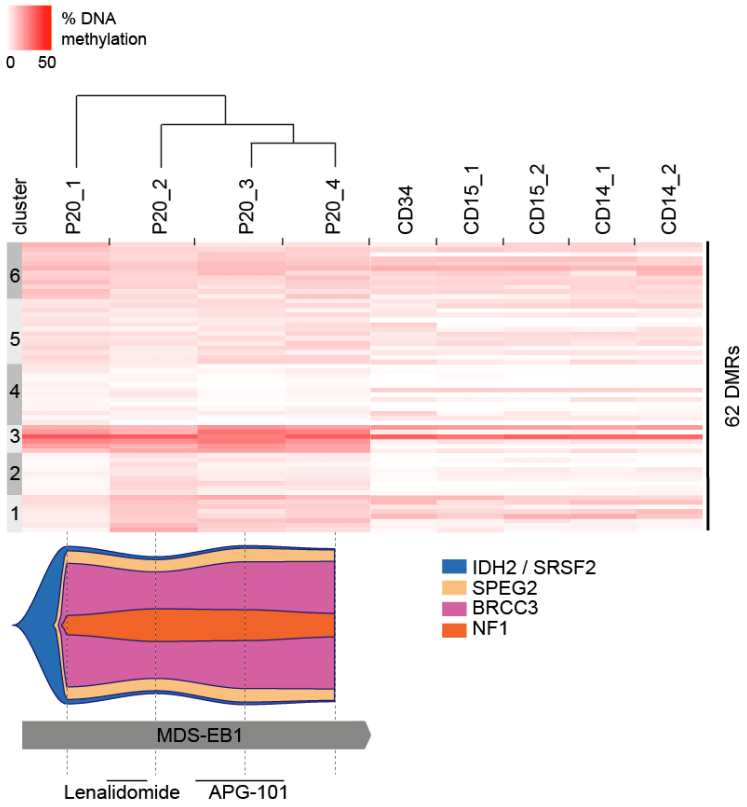


Figure 5-20 - DNA methylation analysis of patient P20 and comparison with genetic and clinical data

The upper part depicts a heat map of *K*-means clustering of DNA methylation data from 62 DMRs between the single time points and control cells (CD34+, CD15+ and CD14+). Dendrogram shows clustering of the patients into two groups. Each horizontal line represents a single DMR while DNA methylation degree is indicated in different shades of red, ranging from 0 % methylation (white) to 50 % methylation (red). The bar on the left side shows classification of DMRs into the five different clusters. Below, fish plot displays mutational evolution during clinical follow-up at analyzed time points (marked with a dashed line). The height of the “fishes” are reflecting the variant allele frequencies of depicted clones, the higher the more alleles are mutated. Different (sub)-clones were colored separately. Clinical classification as well as therapy is drawn at the bottom of the figure.

5.1.2.6 Detailed analysis of patient P53

5.1.2.6.1 Analysis in comparison to CD34+ cells

Calculation of CpG Pearson correlation values (Figure 5-21 A) between hematopoietic stem cells and each of the longitudinal patient samples revealed high similarities. Correlation values were increasing when comparing the different patient samples amongst each other. After identifying and merging differentially methylated regions between every single sample and CD34+ cells, 1925 DMRs were split into 8 clusters using *K*-means algorithm (Figure 5-21 B). DNA methylation patterns in those 8 clusters generally looked similar between consecutive samples, but the DNA methylation degree varied. Time points 1 and 2 displayed a greater methylation in cluster 7 compared to the other samples and time point 3 was observed to be the time point with lowest DNA methylation. Overall, clusters could be divided in those which were hypermethylated in contrast to CD34+ cells (2, 3, 4, 7 and 8) and those which were hypomethylated (5 and 6). The only cluster with stable DNA methylation degree over time was cluster number 1 with almost no differences between CD34+ and patient samples. Hypermethylated regions in the clusters mentioned before displayed an H3K27me3 signature in hematopoietic stem cells, whereas this effect was strongest in clusters 7 and 8. The additional activating mark H3K4me3 suggested those regions to be bivalent and disease specific. Furthermore, motif analysis showed a significant occurrence of binding motifs for GATA transcription factors in clusters 2 and 3. Regions of cluster 8 were enriched for RUNX1 (q-value = 1e-8), MEF2C (q-value = 1e-7) and SPIB (q-value = 0.002) binding sequences. RUNX1 is known to be important for hematopoietic differentiation[138,138]. MEF2C (Myocyte Enhancer Factor 2C) is normally highly expressed in common myeloid progenitors and decreases with further differentiation of the cell [201]. Spi-B transcription factor (SPIB) plays an important role for differentiation of mature B-cells into plasma cells and plasmacytoid dendritic cells[202].

Clusters 5 and 6 were associated with hypomethylation in comparison to CD34+ cells and showed only slight signals for H3K4me3 and H3K27ac. Enriched motifs for transcription factors BATF (q-value = 0.0034), AP-1 (q-value = 0.0034) and FRA-1 (q-value = 0.0065) were found in cluster 5, while these TFs are known to play a role in several differentiation processes. All in all, these regions were probably “opened” during differentiation processes[203,204].

When analyzing DNA methylation in FANTOM5 promoter and enhancer as well as in GATA2 bound regions, variation between the longitudinal patient samples was also clearly evident (Figure 5-21 C). Promoter regions showed the lowest DNA methylation degree in CD34+ cells, followed by CD14+ and CD15+ cells. Patient P53 was generally higher methylated, while time point 5 exhibited the highest and time point 3 the lowest DNA methylation degrees. Samples from the other time points were located in between.

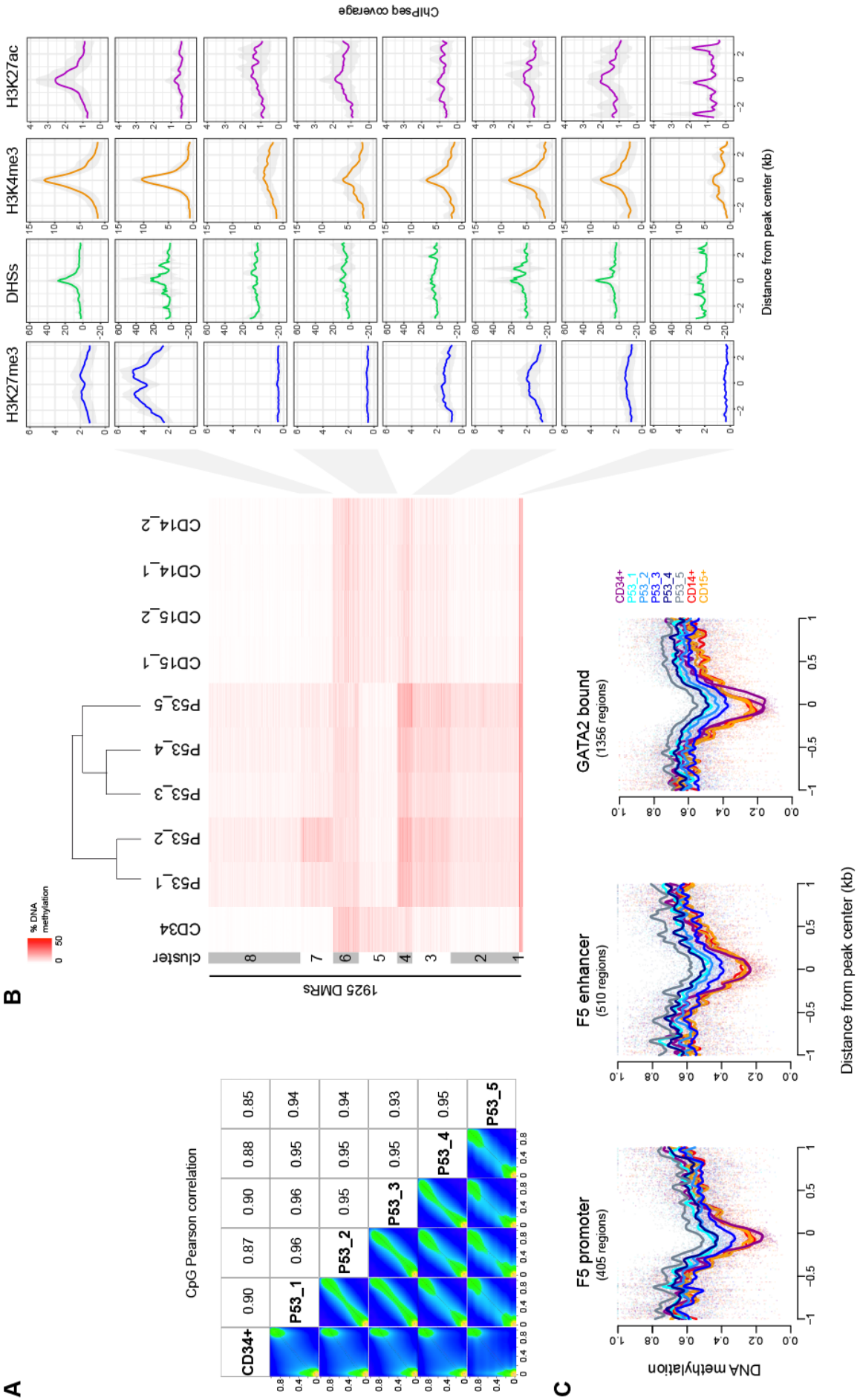

A
CpG Pearson correlation
CD34+
P53_1
P53_2
P53_3
P53_4
P53_5
B
% DNA methylation
cluster
CD34
P53_1
P53_2
P53_3
P53_4
P53_5
CD15_1
CD15_2
CD14_1
CD14_2
1925 DMRs
H3K27me3
DHSs
H3K4me3
H3K27ac
ChIPseq coverage
Distance from peak center (kb)
C
F5 promoter
(405 regions)
F5 enhancer
(510 regions)
GATA2 bound
(1356 regions)
DNA methylation
CD34+
P53_1
P53_2
P53_3
P53_4
P53_5
CD14+
CD15+
Distance from peak center (kb)

Figure 5-21 - DNA methylation analysis and associated epigenetics of patient P53 in comparison to CD34+ cells

(A) Scatter plots of % CpG methylation for each pair of samples. Signal density is represented in different colors ranging from yellow to blue indicating a high and low density, respectively. Numbers on the upper right corner denote pair-wise Pearson correlation coefficients. (B) Heat map depicting *K*-means clustering of 1925 DMRs obtained in comparison to CD34+ cells as well as corresponding data of CD14+ and CD15+ cells. Dendrogram shows clustering of the patients into two groups, while samples 1 and 2 represent one group and samples 3, 4 and 5 the second group. Each horizontal line represents a single DMR. Epigenetic marks in CD34+ cells in these DMRs are depicted on the right histograms for every cluster of regions. Average ChIPseq coverage for every cluster was drawn over a 4 kb region of the analyzed DMRs. (C) Histogram of average DNA methylation over 2 kb regions of FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions in CD34+ cells (purple), different time points of patients (different blue shades) as well as CD14+ (red) and CD15+ (yellow) cells. DNA methylation data in all observed 1925 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

The same scenario could be observed for FANTOM5 enhancer regions and for GATA2 bound regions. Publicly available data from different hematopoietic cells were used to compare DNA methylation degrees in regions analyzed above (Figure 5-21 C). The lowest DNA methylation was displayed by classical monocytes, band form neutrophils, neutrophilic metamyelocytes and neutrophilic myelocytes. A higher extent was observed in the two B lineage derived cells (naïve B cells and precursor lymphocyte of B lineage) followed by Treg and CD8+ T cells. The same applied for FANTOM5 enhancer and GATA2 bound regions (Figure 5-22).

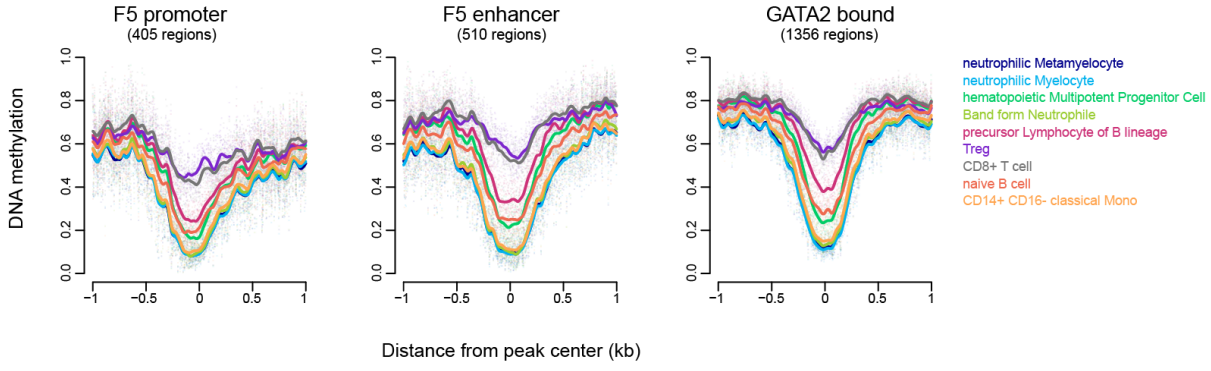


Figure 5-22 -Histogram of DNA methylation in different genomic regions using publicly available data sets (P53)

Histogram of DNA methylation ratios over 2 kb regions across FANTOM5 (F5) promoter, F5 enhancer and GATA2 bound regions overlapping with DMRs found in P02 in comparison to CD34+ cells in hematopoietic precursor and more differentiated cells. Publicly available whole genome bisulfite sequencing data in the patient specific 1675 DMRs was overlapped with publicly available data of the FANTOM5 consortium for promoter regions or enhancer regions as well as with ChIPseq data of GATA2 binding motifs.

5.1.2.6.2 Integrated analysis of DNA methylation and genetics/cytogenetics of patient P53

The above described variation among the longitudinal samples became more obvious when looking for DMRs between the different sample time points. 415 differentially methylated regions were identified and were split into 8 *K*-means clusters. DNA methylation patterns over time were similar with exceptions at time point 2 and 5, where clusters 6 and 4/7 showed higher methylation degree than other samples, respectively (see Figure 5-23).

Differences regarding DNA methylation during disease progression were linked to genetic and cytogenetic changes, which are displayed in the corresponding fish plots (Figure 5-23). Since DNA methylation analysis was done from peripheral blood, analysis of molecular genetics was performed with the same material (first fish plot, labeled with “PB”).

Complete cytogenetic analysis was not feasible in peripheral blood of the patient. A karyotype analysis with chromosome banding relies on proliferating cells. In peripheral blood only a very small amount of those cells can be found and therefore banding analysis is done by default with bone marrow.

Therefore chromosome banding analysis in bone marrow of corresponding samples was displayed as additional fish plot (labeled with “BM”). Aberrations in peripheral blood comprised lesions in *RUNX1* and *SRSF2*, while *RUNX1* was identified to be the founder clone with later acquisition of *SRSF2* mutations. Variant allele frequency of these lesions rapidly dropped at time point 3 and remained more or less stable until end of monitoring. Only a small shift in the VAF could be observed at the last time point. A more complex situation was seen in the bone marrow of the patient.

At diagnosis, most of the cells exhibited 20q deletions harboring a subclone with lower frequency for i(12q) aberrations. The same composition, but with lower frequency for both, was observed at time point 2. At time point 3, changes in clinical classification, cytogenetics, molecular genetics and DNA methylation occurred. Transition into acute myeloid leukemia came along with disappearance of the 20q-/i(20q) clone, shrinkage of *RUNX1* / *SRSF2* lesions and upcoming of a new clone with trisomy 12. Starting with time point 4 through to time point 5, this +12 clone was fading out until complete disappearance. This came along with the recurrence of the 20q- clone having new subclones with +12 and t(X;14) aberrations, but unlike before this trisomy 12 is a new alteration.

Summarizing, it could not be determined which aberrations were responsible for the changes in DNA methylation patterns in the different patient samples. In order to clarify this, cytogenetic and genetic analysis has to be done with the same material. But it is clear that several independent subclones existed during disease progression. This is reflected by the observed variation of DNA methylation between consecutive samples.

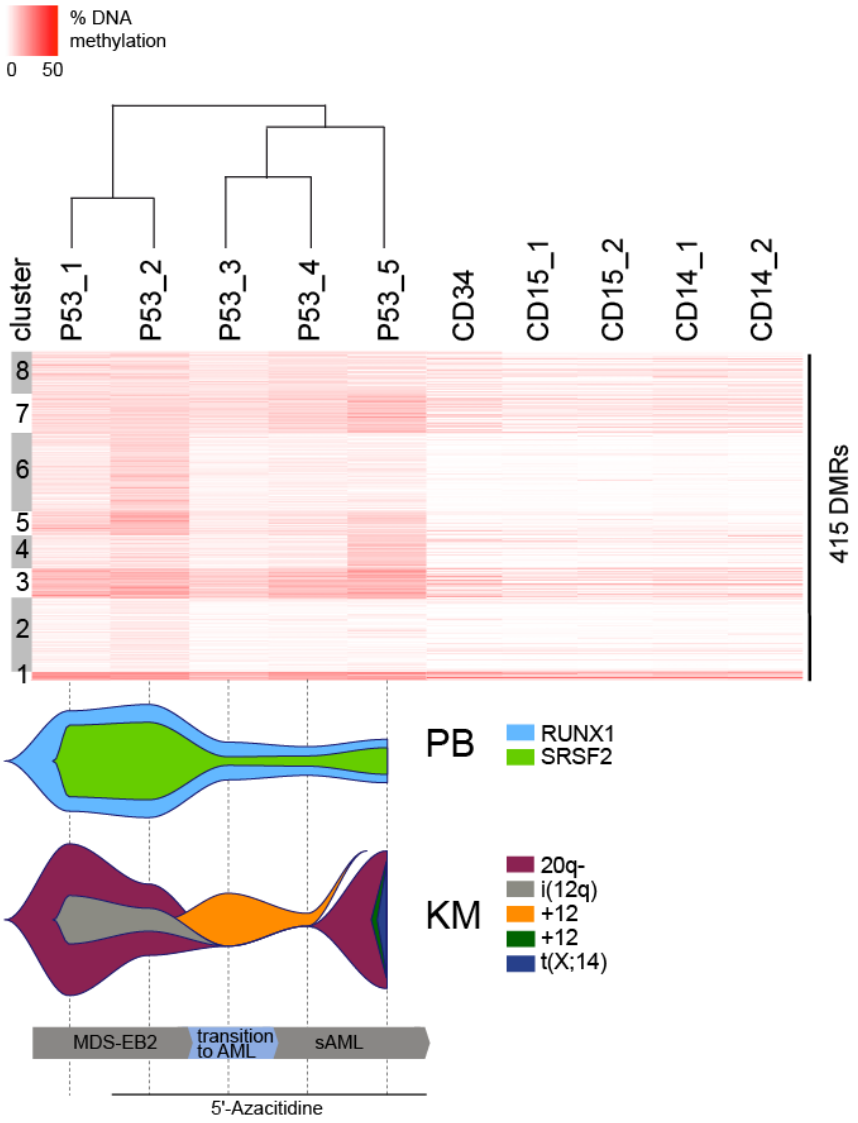


Figure 5-23 - DNA methylation analysis of patient P53 and comparison with genetic and clinical data

The upper part depicts a heat map of *K*-means clustering of DNA methylation data from 415 DMRs between the single time points and control cells (CD34+, CD15+ and CD14+). Each horizontal line represents a single DMR while DNA methylation degree is indicated in different shades of red, ranging from 0 % methylation (white) to 50 % methylation (red). The bar on the left side shows classification of DMRs into the five different clusters. Below, fish plot displays mutational evolution during clinical follow-up at analyzed time points (marked with a dashed line) in peripheral blood (PB) and bone marrow (BM). The height of the “fishes” are reflecting the variant allele frequencies of depicted clones, the higher the more alleles are mutated. Different (sub)-clones were colored separately. Clinical classification as well as therapy is drawn at the bottom of the figure.

5.2 Pediatric MDS

5.2.1 DNA methylation analysis in pediatric MDS

With an incidence of 0.5 – 4 / 10^6 per year pediatric myelodysplastic syndromes (MDS) are very rare hematopoietic disorders[112]. Therefore, studies of genetic and epigenetic alterations in pediatric population are challenging and thus there is less knowledge compared to adult MDS. Since MDS in children and adults show evident differences in clinics, cytogenetics, molecular genetics and probably in epigenetics, an integrated analysis of the different aberrations is interesting and might shed light on disease pathogenesis and / or progression.

For this purpose we studied a cohort of 42 children (18 females and 24 males) with a median age of 10.5 years (Table 5-4). Since the group of Dr. Wlodarski, who kindly provided the samples, was interested whether germline *GATA2* mutations in familial MDS cases were linked to aberrant DNA methylation, our study cohort was enriched for those cases (59.5 %). Wlodarski et al. showed that germline *GATA2* mutations occur in 15 % of advanced and 7 % of all primary MDS cases in children. Furthermore they pointed out that *GATA2* mutations were preferentially associated with monosomy 7[205], explaining the high percentage of this cytogenetic alteration (71.4 %) in our patient cohort. The majority of the patients were classified with refractory cytopenia (RC, 42.9 %) or RAEB (33.3 %), while a smaller portion was diagnosed with advanced subtypes (RAEB-t, 14.3 % and MDR-AML, 9.5 %).

Besides *GATA2* mutations, lesions in *ASXL1* and *SETBP1* could be observed in 19.0 % and 28.6 %, respectively. *ASXL1* mutations were shown to significantly co-occur with GATA2 deficiency in MDS/AML[206] and *SETBP1* lesions were enriched among MDS patients with *ASXL1* mutations[135]. Furthermore, available longitudinal samples from eight out of the 42 patients allowed analysis during disease progression. DNA from purified bone marrow granulocytes was kindly provided from Dr. Wlodarski´s group (Pediatric Hematology and Oncology, University of Freiburg) and further processed in our lab using the Methyl-CpG immunoprecipitation (MCIp). This method allows genome-wide profiling of aberrant DNA methylation of CpG-rich regions, including CpG islands (CGIs). Those regions are prone to disease-related changes, since their DNA methylation status rarely changes during differentiation[207].

Subsequent library preparation and next-generation sequencing on the Illumina HiSeq 1000/2000 platform were performed. Sequences were mapped to the human reference genome (hg18) using Bowtie 2[208]. Downstream analysis of uniquely mapped tags including quality control, annotation, normalization and motif analysis were done using HOMER[177]. To compensate variations in clonality, data was normalized to one tag per base pair (tbp) and analysis was done after removal of sex chromosomes. The free software “R”[209] was used for the calculation of Pearson correlation coefficients, scatter plots, t-SNE visualization (t-Distributed Stochastic Neighbor Embedding), *K*-means clustering and drawing of corresponding heat maps as well as Wilcoxon sign tests.

Table 5-4 - Clinical data and mutational frequencies of 42 pediatric MDS patients

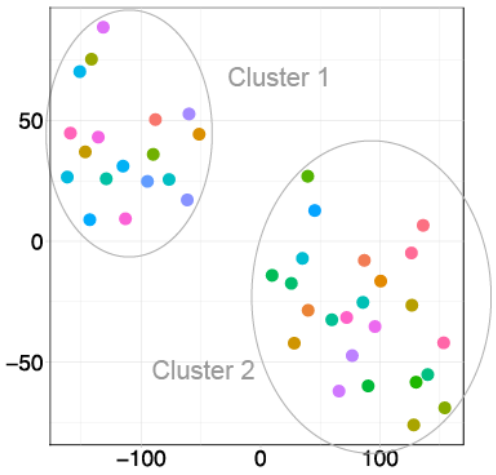
N = 42	**Number (percentage)**
gender	
female	18 (42.9 %)
male	24 (57.1 %)
Median age at diagnosis	10.5 years
WHO subtype	
RC	18 (42.9 %)
RAEB	14 (33.3 %)
RAEB-t	6 (14.3 %)
MDR-AML	4 (9.5 %)
Mutations	
GATA2 germline	25 (59.5 %)
ASXL1	8 (19.0 %)
SETBP1	12 (28.6 %)
Karyotype	
Normal	4 (9.5 %)
Monosomy 7*	30 (71.4 %)
Trisomy 8	7 (16.7 %)
Monosomy 7 & Trisomy 8	1 (0.024 %)

Abbreviations: RC: Refractory Cytopenia; RAEB: Refractory Anemia with Excess Blasts; RAEB-t: RAEB in transformation; MDR-AML: MDS-related AML; *Includes monosomy 7 with one additional aberration.

The first subset analysis was done with all available patients at first diagnosis, but without consecutive samples of the progressed patients to check for commonalities between all pediatric MDS patients. To visualize the complex next-generation sequencing data of global DNA methylation in our patient cohort, we used the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique (see Figure 5-24 A). As mentioned above, data was first normalized to one tag per base pair (tbp1) and sex chromosomes were removed. Furthermore, data were normalized to the 99th percentile to compensate different sequencing depth levels of the samples. In this step, regions were restricted to those that could be detected with the MCIp approach and moreover to non-repeat regions.

In the t-SNE we saw that patients formed two different clusters due to their global DNA methylation pattern. Cluster 1 comprised 18 patients and 24 patients were found in the bigger cluster 2. To test whether clustering occurred due to the significant occurrence of genetic, cytogenetic or clinical differences, we applied the non-parametric Wilcoxon test (Figure 5-24 B). Interestingly, only 2 events were significantly enriched in cluster 2, namely *GATA2* mutations (p value = 0.0205) and refractory cytopenia (RC, p value = 0.022).

A



B

Variable	p value	significant
GATA2^{mut}	0.0205	* in cluster 2
Monosomy 7	0.0629	n.s.
Trisomy 22	0.106	n.s.
Trisomy 8	0.225	n.s.
SETBP1^{mut}	0.211	n.s.
ASXL1^{mut}	0.225	n.s.
RC	0.022	* in cluster 2
RAEB	0.524	n.s.
RAEB-t	0.216	n.s.
MDR-AML	0.185	n.s.

C

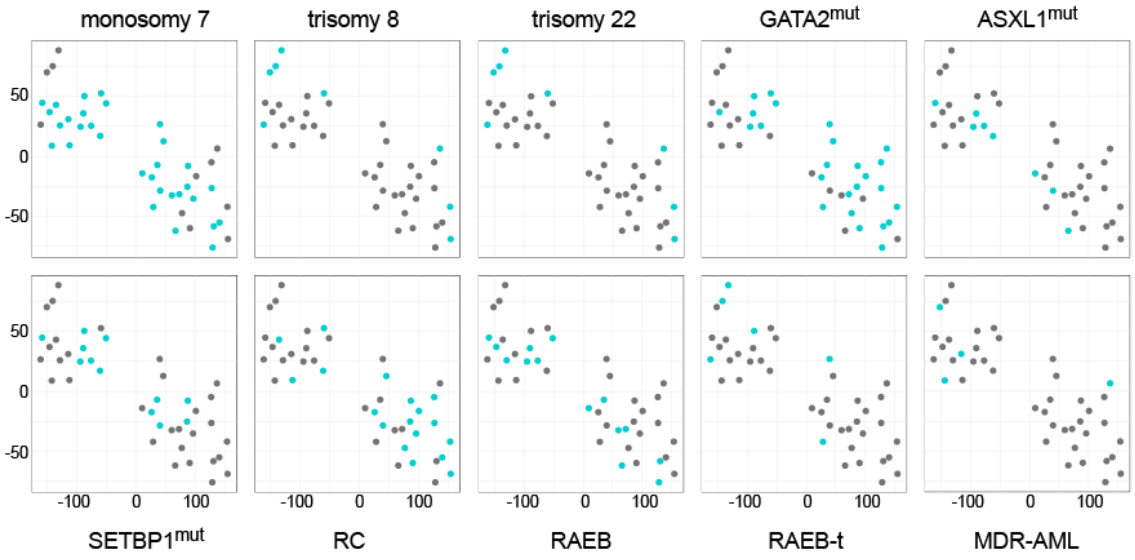


Figure 5-24 – T-SNE with global DNA methylation data in 42 pediatric MDS patients and Wilcoxon test

(A) Global DNA methylation data obtained with Methyl-CpG-immunoprecipitation (MCIp) followed by NGS was analyzed using the t-SNE (t-Distributed Stochastic Neighbor Embedding) approach. Each of the 42 patients was colored differently and the two obvious clusters were encircled. (B) Table with results from Wilcoxon test applied for different mutations, WHO subtypes and cytogenetic aberrations. (C) T-SNE plots with highlighting the occurrence of different mutations, cytogenetic aberrations and WHO subtypes (turquoise) in the two clusters. Patients without the analyzed feature were plotted in gray.

To illustrate the results of the above described Wilcoxon test, we highlighted patients with the analyzed events in turquoise and patients without this feature in gray (Figure 5-24 C). This figure visualizes that the different events (with exception of GATA2^{mut} and RC) were distributed equally over the two patient clusters. The observation of Wlodarski et al.[205] that *GATA2* mutations were preferentially associated with monosomy 7 could be confirmed, but this was probably due to the bias of these two abnormalities in our analyzed patient cohort. Furthermore, the study from Inoue et al.[210] in 2015 showed that MDS patients with *SETBP1* mutations were enriched in those with *ASXL1*

mutations. Comparing these two mutations in our t-SNE plots, we could also see the occurrence between these two lesions. Moreover, these two mutations tend to be more frequent in cluster 1, which is associated with the more advanced clinical subtypes (RAEB, RAEBt and MDR-AML). These two findings suggested patients in cluster 1 to be more advanced and associated with poorer overall survival, while those in cluster 2 seemed to be low risk patients.

Next, we wanted to figure out regions that were differentially methylated between patients of the two clusters. In the first step, patient specific MCIp peaks had to be identified by comparison of DNA methylation data of patients with control cells (granulocytes, monocytes and hematopoietic stem cells).These patient specific MCIp peaks are specifically methylated or demethylated in contrast to all three sets of control cells. In the next step, patient specific MCIp peaks from all patients were merged into one peak set and overlapped with promoter regions. DNA methylation data of patients was then annotated in these identified DMRs between cluster 1 and 2 and furthermore analysed regarding their gene ontology using Metascape[180]. Interesting significantly enriched GO terms were for instance “cell fate commitment”, “negative regulation of cell differentiation” and “connective tissue development” (see Table 5-5). To limit the amount of DMRs, regions were further restricted whether they are associated with a transcription factor (TF) or chromatin factor (CF). Via comparison of mean methylation values in control monocytes, patients of cluster 1 and patients of cluster 2 we obtained interesting DMRs that were specifically methylated in cluster 1.
One example is ZIC5, a zinc finger protein with transcriptional repressor function. Elevated expression has been observed in various human cancers and it may contribute to cancer progression[211–213]. DNA methylation data in this *ZIC5* region (chr13:100624101-100624465) as well as ChIPseq data (e.g. H3K4me1, H3K27ac…) in hematopoietic stem cells are depicted in Figure 5-25. Patients were colored according to their cluster membership (orange for cluster 1 and blue for cluster 2), while control cells were drawn in green. Patients in cluster 1 were specifically methylated in this promoter region of *ZIC5*, whereas patients in cluster 2 showed either no or only small DNA methylation signals.

Table 5-5 - GO term analysis of DMRs between pediatric MDS patients of the two clusters

GO term	Log(q-value)
Embryonic organ development	-31.4
Sensory organ development	-26.7
Tube development	-22.9
Negative regulation of transcription from RNA polymerase II promoter	-21.4
Head development	-20.8
Cell fate commitment	-18.8
Central nervous system neuron differentiation	-14.9
Negative regulation of cell differentiation	-13.9
Gland development	-13.7
Neuron fate commitment	-13.4
Behavior	-13.1
regulation of nervous system development	-12.2
connective tissue development	-10.2

We found two further DMRs associated with cancer, namely *VILL* and *TRIM45*, which were methylated to a greater extent in patients belonging to cluster 1. Only few publications are available regarding the link of *VILL* with tumors, but Senchenko et al.[214] pointed out the possible role of VILL as tumor suppressor gene in cervical cancer.

TRIM45 is a transcriptional repressor of the mitogen-activated protein kinase (MAPK) pathway and was shown to function as tumor suppressor in the brain[215,216].

Unfortunately, functional analysis (DNA methylation sensitivity) of the three mentioned DMRs was not possible due to lack of patient material. Therefore we could only suppose that differential methylation of ZIC5, VILL and TRIM45 may contribute to disease pathogenesis or progression.

Nevertheless we could divide the patient cohort according to different characteristics. Patients belonging to cluster 1 were associated with more advanced clinical subtypes, mutations of *ASXL1* and *SETBP1* as well as higher DNA methylation in regions potentially relevant in cancer (ZIC5, VILL, TRIM45). Cluster 2 comprised patients which were significantly associated with *GATA2* mutations, refractory cytopenia and fewer DNA methylation aberrations in the three DMRs (ZIC5, VILL, TRIM45).

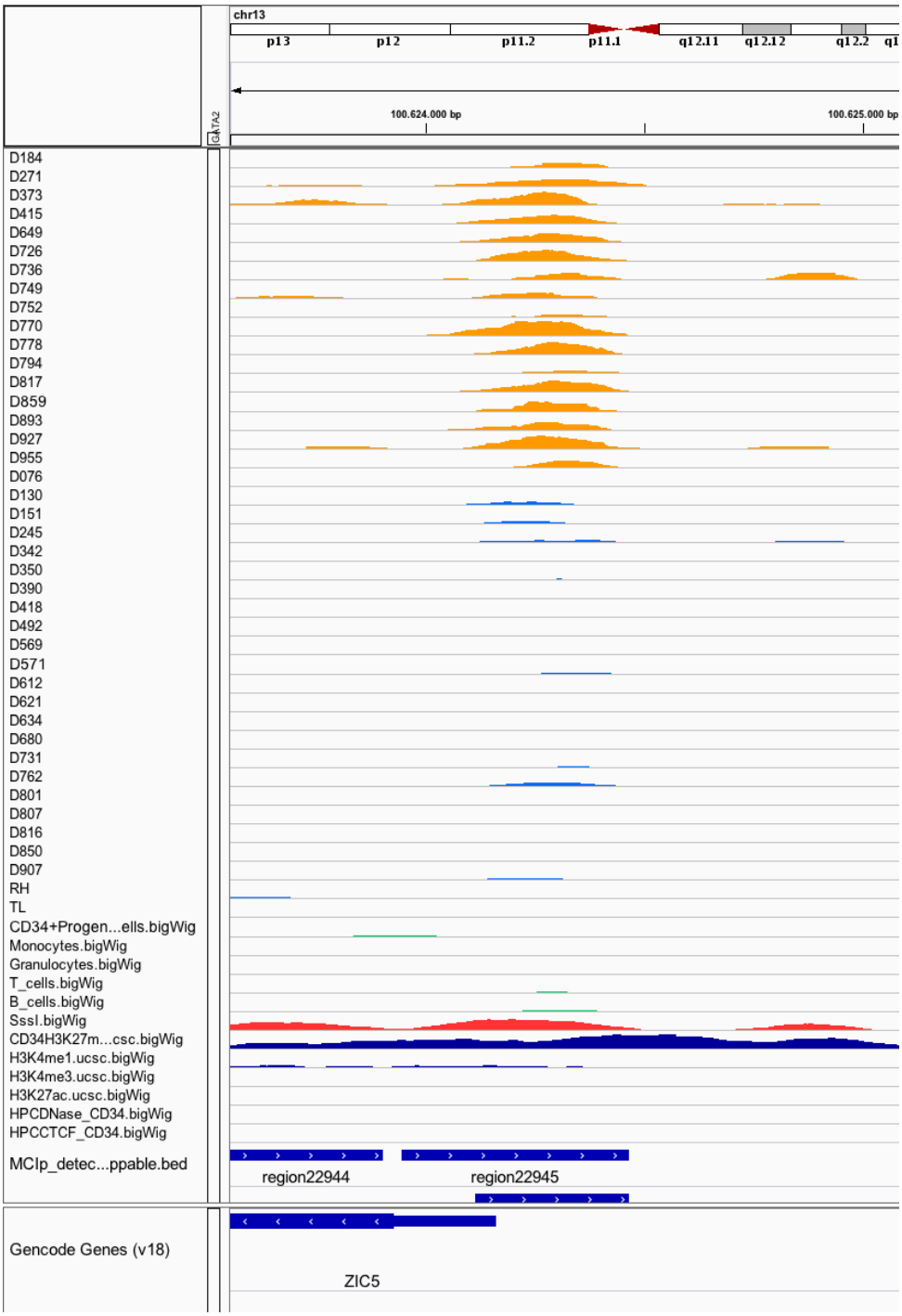


Figure 5-25 - Genome browser snapshot of the DMR located in *ZIC5*

Integrated Genome Viewer (IGV) track of DNA methylation data (MCIp-seq) from MDS patients in the ZIC5 promoter region (hg19_chr13:100624101-100624465). Patients were ordered according to obtained t-SNE clusters 1 and 2, whereas patients belonging to cluster 1 were depicted in orange (upper part) and patients from cluster 2 were displayed in blue (middle part). In the lower part MCIp-seq data from control cells were depicted in green as well as epigenetic marks in stem cells (dark blue).

5.2.2 Changes of DNA methylation during disease progression in pediatric MDS

One of the major goals in our project was to identify potential epigenetic marker genes that are involved in progression from MDS to acute myeloid leukemia. For this purpose, we analyzed consecutive patient samples using the MCIp-seq approach as mentioned in section 4.2.1.16. The study cohort comprised eight patients (3 females, 5 males) with 2-5 consecutive samples of each patient (see Table 5-6). Most of the patients (6 out of 8) were diagnosed with refractory cytopenia (RC) and the other two patients with RAEB. All MDS patients carry germline *GATA2* mutations, with one patient having additional *SETBP1* and *ASXL1* lesions. 50 % of the patients exhibit monosomy 7 aberrations and the same percentage received a hematopoietic stem cell transplant.

Table 5-6 - Clinical data of pediatric MDS patients with consecutive samples

ID	# of samples	Sex	Age at Dx	WHO at DX	Cytogenetics	Gene mutation	BM blast count (%)	Therapy
D151	2	f	16.0	RC	normal	GATA2	1	-
D271	2	m	12.5	RAEB	45,XY,-7	GATA2	9	HSCT
D342	2	m	16.6	RC	45,XY,-7	GATA2	0	HSCT
D569	2	f	10.3	RC	normal	GATA2	0	-
D762	3	m	9.7	RC	45,XY,-7	GATA2	1	HSCT
D770	5	f	10.2	RAEB	46,XX,-7,+22	GATA2; ASXL1; SETBP1; CBL	3	HSCT
D801	2	m	15.6	RC	normal	GATA2	0	-
D807	4	m	13.7	RC	normal	GATA2	0	-

Global DNA methylation data from consecutive samples was visualized using the t-Distributed Stochastic Neighbor Embedding (t-SNE) approach to get a general idea of similarities and differences between the longitudinal samples of the patients (see Figure 5-26). In general, consecutive samples from the different patients showed a high similarity due to their close vicinity in the t-SNE plot.
The only exception was patient D271, where the two time points were not in close proximity to each other.
In order to get more details regarding DNA methylation changes during disease progression, we analyzed longitudinal patients separately (see following section).

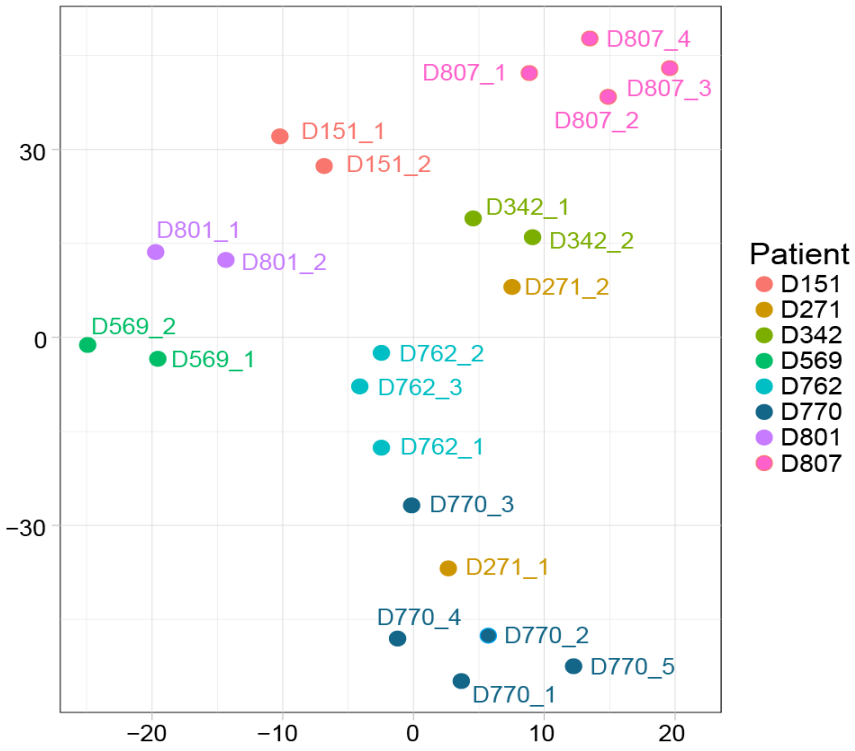


Figure 5-26 - T-SNE with global DNA methylation data in 8 pediatric MDS patients with longitudinal samples

Global DNA methylation data obtained with Methyl-CpG-immunoprecipitation (MCIp) followed by NGS was analyzed using the t-SNE (t-Distributed Stochastic Neighbor Embedding) approach. Each of the longitudinal patients with their corresponding consecutive samples was colored differently and sample time points were marked with increasing numbers.

5.2.2.1 Detailed analysis of patient D770 – a case with progression

Of all patients, only patient D770 showed significant changes of DNA methylation during disease progression. This patient was analyzed at five consecutive time points with a total of 18 months between first and last sample (see Figure 5-27). The patient exhibited a germline *GATA2* mutation, *ASXL1*, *SETBP1* and *CBL* mutations as well as an aberrant karyotype with monosomy 7 and trisomy 22. The WHO classification (RAEB) did not change over time but increasing bone marrow blast counts could be observed from 3 % to 52 %. Treatments included hypomethylating agents (Azacitidine) and hematopoietic stem cell transplantation.

It is important to mention, that variant allele frequencies (VAFs) of gene mutations changed during disease progression. At the beginning, the patient revealed mutations in *GATA2*, *SETBP1*, *ASXL1* and *CBL* with VAFs of 46 %, 44 %, 38 % and 4 %, respectively. Then, time point 3 showed altered distribution of gene mutations, with decrease of *SETBP1* and *ASXL1* to VAFs of about 25 %, whereas *CBL* mutations increased significantly to about 23 %. After disease progression, *ASXL1* mutations completely disappeared at time point 5, while *CBL* mutations dramatically increased to VAFs of about 90 % and VAFs for *SETBP1* returned to levels similarly to initial ones.

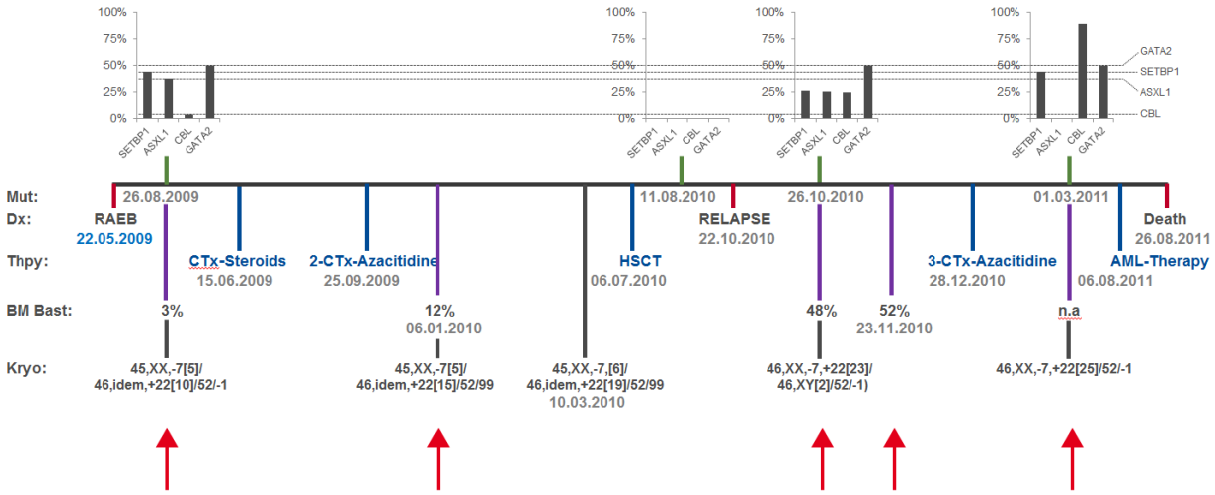


Figure 5-27 – Time line with clinical data of patient D770 during disease progression

The upper part depicts bar plots of variant allele frequencies (VAFs) regarding mutations in *SETBP1*, *ASXL1*, *CBL* and *GATA2* for various time points. Below, information about therapy, BM blast count and karyotype is shown, while analyzed samples were marked with a red arrow.

Global DNA methylation data of the MCIp-seq approach was compared between the five time points of the patient and between control monocytes (MO) using scatter plots (Figure 5-28). In these scatter plots, DNA methylation signals in all mappable and detectable regions were depicted in black and DNA methylation signals in regions specific for the patient were drawn in red. Correlation values calculated with regard to all mappable and detectable regions were shown in the bottom right corner of each plot.

Correlation values between monocytes and the five samples were found between 75.1 % and 81.4 %, indicating several differences in DNA methylation between control and patient cells. What could be seen very nicely is the high number of regions (red dots), that were specifically methylated in patient D770 compared to monocytes. These patient specific regions were mainly hypermethylated compared to control monocytes. Comparing DNA methylation between the single consecutive patient samples, correlation was observed at higher values ranging from 86.6 % to 94.4 %. With a few exceptions, specific regions of the patient were also methylated differentially in the consecutive samples indicated by the scattering of these dots, while DNA methylation pattern over time tend to be stable. This hypothesis was supported by the similar signal cloud between the patient time points, only the shift of the signals demonstrated some changes in DNA methylation during disease progression.

Interestingly, one could see a connection between the hematopoietic stem cell transplantation before time point 3 and DNA methylation pattern. This pair showed the highest Pearson correlation with 81.4 % and exhibited the least changes in patient specific regions. Here we could see that the transplanted patient showed features similar to healthy control monocytes indicating a successful treatment.

But other than expected, treatment with Azacitidine didn´t influence DNA methylation pattern significantly[217].

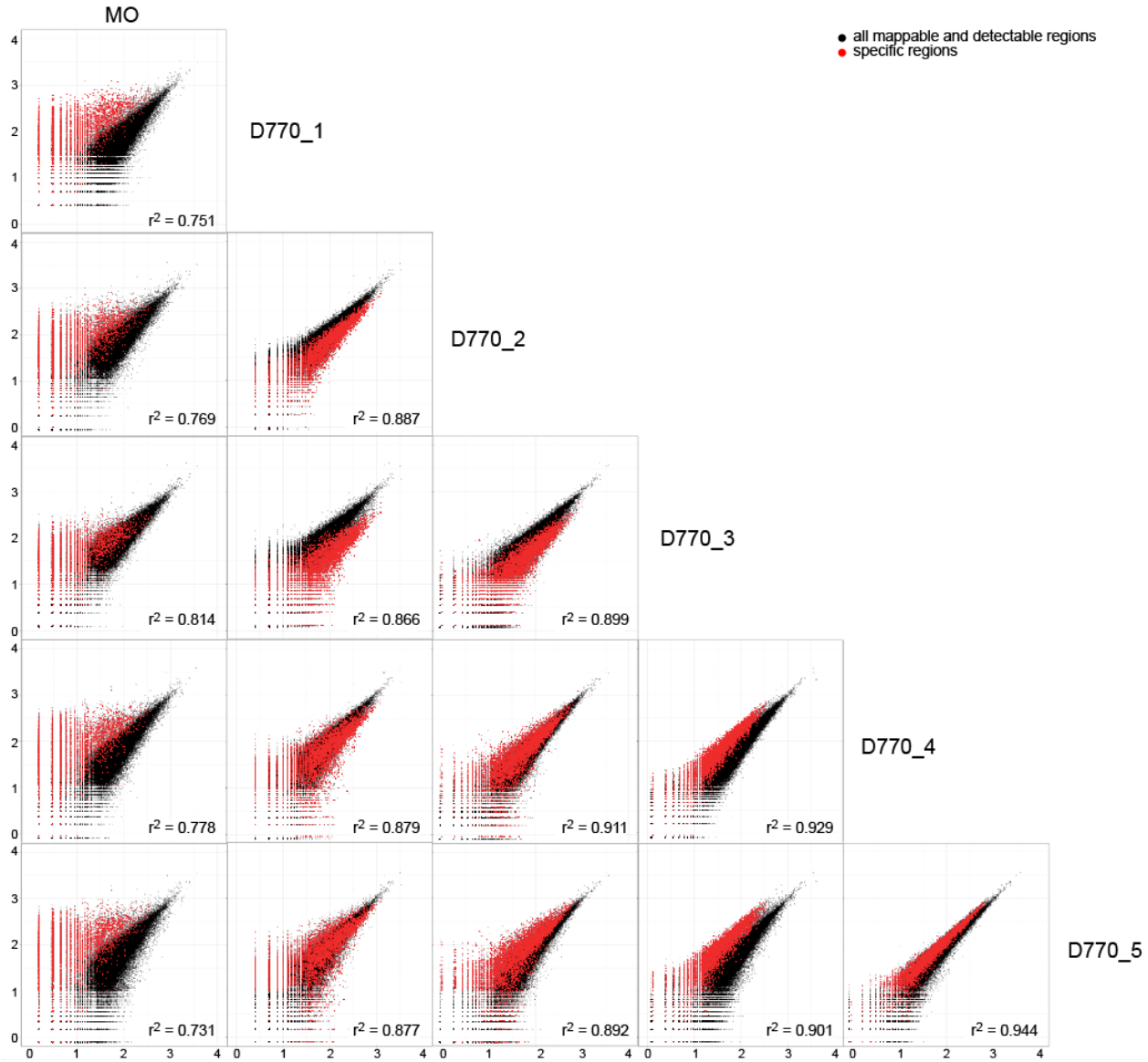


Figure 5-28 - Scatter plots of global DNA methylation data from patient D770 and control monocytes

Comparison of DNA methylation patterns between different time points of patient D770 and control monocytes depicted as scatterplots. In black all mappable and detectable regions were depicted, while patient specific regions were displayed in red. Correlation values could be found on the bottom right corner of the plots.

To get more information about DNA methylation changes over time, patient specific regions were clustered using the *K*-means algorithm. The six obtained region clusters were compared with epigenetic data (H3K27me3, H3K4me3, H3K27ac and DHSs) in hematopoietic stem cells providing insight into activity states of the regions (see Figure 5-29). This heat map revealed clusters of regions that gained DNA methylation during progression (clusters 1-3) and those that lost this epigenetic mark (clusters 4-6). Regions with an increase of DNA methylation over time were already marked with H3K27me3 in CD34+ progenitor cells indicating an early repressive state of these regions. Therefore additional DNA methylation likely plays a less important role and thus we focused on regions which were hypomethylated during disease progression and exhibited no repressive trimethylation of H3K27 (clusters 4-6).

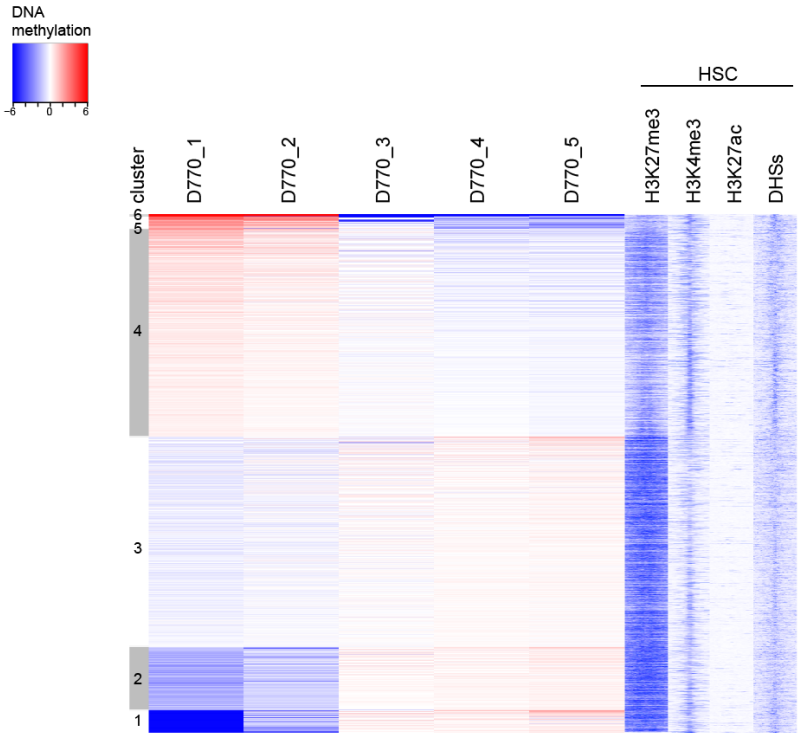


Figure 5-29 - *K*-means clustering of patient specific regions and annotation of epigenetic data in HSCs
Heat map depicts clustering of specific regions from patient D770 using the *K*-means algorithm with DNA methylation degrees ranging from high (indicated with red) to low levels (indicated with blue). The obtained six region clusters were compared with epigenetic data in hematopoietic stem cells (HSC), while intensity ranged from white to blue.

After checking associations of regions with transcription or chromatin factors, we obtained interesting regions potentially important for disease progression, including VENTX, SIX5, CREB3L1, GADD45B, ETV5, ETV2 and FOXO6. The influence of DNA methylation in these regions was tested with the above described reporter gene assay (see 4.2.1.6), either regarding promoter activity (VENTX) or enhancer activity. Figure 5-30 A depicts box plots for gene reporter assays with SIX5 and VENTX, the only two regions with significant changes in luciferase activity (other data see Appendix 11.2). As controls we used the empty vector pCpGL-basic, either unmethylated (pCpGL) or treated with CpG methyltransferase (pCpGL-methyl). All analyzed constructs were also tested in both states, unmethylated and fully methylated.

Reporter gene activity in the unmethylated state above the one of the control plasmid pCpGL-CMV/T.E1AF (#1341) was observed for almost all DMRs with exception of CREB3L1 and ETV5.

The box plot showed that methylation of the two regions VENTX and SIX5 resulted in a significant lower reporter gene activity compared to control gene activity. This result suggested that differential methylation in those regions could have an effect on gene expression and therefore contribute to disease progression or pathogenesis. In the literature, VENTX was shown to play a role in cancer, to promote human erythroid differentiation, and to also highly expressed in acute myeloid leukemia[218]. To illustrate the differences in DNA methylation in the promoter region of VENTX, we had a look at the data with the integrated genome viewer (IGV) (Figure 5-30 B). The upper part showed DNA

methylation signals in all five consecutive patient samples with decreasing signals in the promoter region. At the same time, DNA methylation signals in control cells (CD34+ cells, granulocytes, monocytes, T- and B-cells) showed very similar pattern amongst each other. In summary, this demethylation of the promoter region of VENTX and the resulting elevated expression could contribute to disease progression, specifically in the case of patient D770.

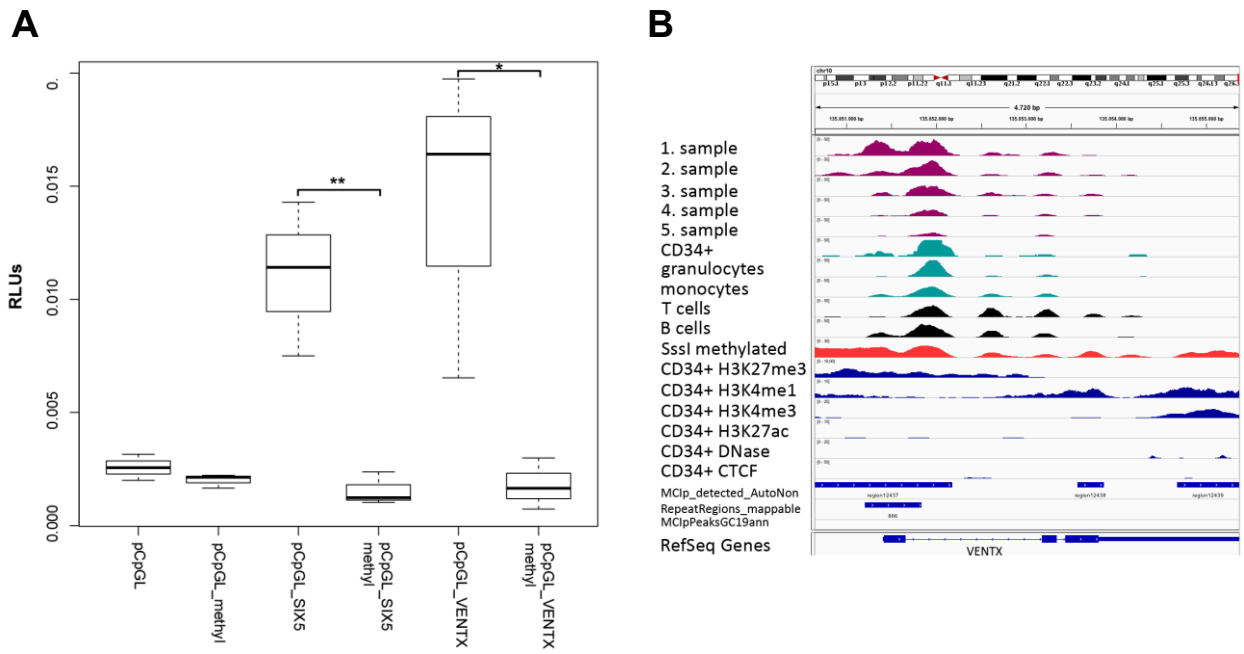


Figure 5-30 - Box plot of gene reporter assays in patient D770 and genome browser of VENTX region

(A) Relative luminometer units (RLUs) were shown in a box plot for pCpGL as control and for SIX5 and VENTX. All constructs were additionally tested in a fully methylated state and marked with the suffix “methyl”. The bold black line denotes medians, boxes the interquartile ranges and whiskers the 5th and 95th percentiles. Significance between corresponding pairs was tested via one-tailed t-test (** p<0.001, * p<0.01). (B) IGV snapshot from analyzed VENTX promoter region with DNA methylation signals in all five samples from patient D770 (purple) as well as in CD34+ cells, granulocytes, monocytes, T- and B-cells (turquoise).

5.2.2.2 Detailed analysis of patients with stable disease

The remaining seven patients of this cohort were analyzed the same way. We observed that all of them did not show big differences in DNA methylation over disease progression. Consecutive samples of each single patient were very similar, thus identifying DMRs between the time points was very challenging.

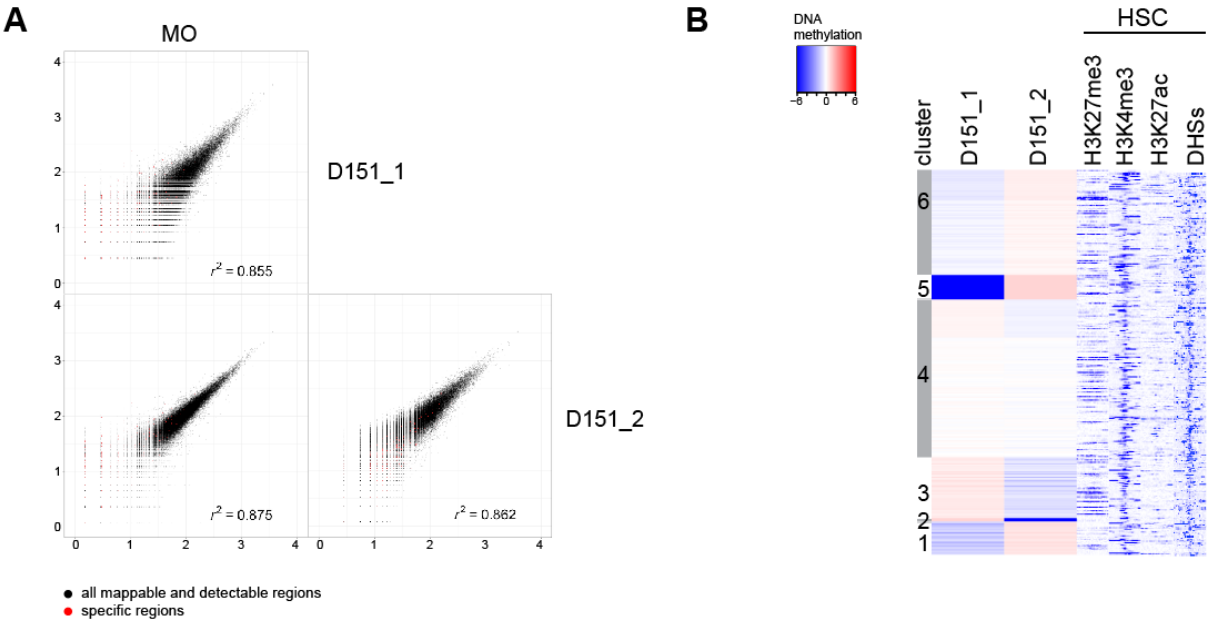
In this section I will describe one example patient (D151) with stable disease since all others displayed the same behavior. Figures from additional patients can be found in the Appendix 11.2.

DNA methylation from patient D151 was analyzed at two different time points 26 months apart. The patient was diagnosed with refractory cytopenia, germline *GATA2* mutation and normal karyotype. The only difference between these two time points was the percentage of bone marrow blasts with slight increase from 1 % to 3 % (see Figure 5-31 C). Scatter plots in Figure 5-31 A depict correlation values between control monocytes and patient samples from 85.5 % to 87.5 %, whereas correlation between the two longitudinal patient samples was little bit higher with 88.2 %. Only 241 patient specific regions (red dots) were observed for patient D151 and they were methylated similarly in both samples.

The patient specific regions were clustered using the *K*-means algorithm to get more information about DNA methylation during disease progression (Figure 5-31 B). In addition, the six obtained region clusters were compared with epigenetic data (H3K27me3, H3K4me3, H3K27ac and DHSs) in hematopoietic stem cells (Figure 5-31 B) to check association with active or repressive marks. Clusters 1, 5 and 6 gained DNA methylation during disease progression, whereas clusters 2, 3 and 4 lost DNA methylation. We couldn´t see any correlation of the regions with active or repressive histone marks, probably due to the low number of DMRs.

To sum up, the patient D151 did not show strong DNA methylation changes over the time period of 26 months and DMRs contributing to disease progression (same approach as described in 5.2.2.1) couldn't be found either. This could be associated with the fact that there were no changes in WHO classification, genetics or cytogenetics between the consecutive samples which could result in changes of DNA methylation pattern and degree. As mentioned above, the other patient samples behaved similarly in the way that DNA methylation during disease progression was very constant and furthermore they had stable genetic and cytogenetic patterns.

Thus, these results led us to the conclusion that DNA methylation changes over time were associated with changes in genetic or cytogenetic landscape of the patient and were not the cause for progression of the disease.



C

Number	Cumulative time [m]	WHO	Treatment	BM blasts	Mutations	Karyotype
D151_1	0	RC	-	1 %	GATA2	Normal
D151_2	26	RC	-	3 %	GATA2	Normal

Figure 5-31 - DNA methylation analysis and clinical data of patient D151

(A) Comparison of DNA methylation patterns between different time points of patient D151 and control monocytes depicted as scatterplots. In black all mappable and detectable regions were depicted, while patient specific regions were displayed in red. Correlation values can be found on the bottom right corner of the plots. (B) *K*-means clustering of DNA methylation data in patient specific regions and corresponding epigenetic data in hematopoietic stem cells (HSCs). (C) Overview of clinical data from consecutive samples of patient D151 with cumulative time between diagnosis and follow up sample, WHO classification, treatment, BM blast count, mutational and cytogenetic status.

5.3 Comparison PB and KM

The possibility to diagnose and monitor myelodysplastic syndromes in peripheral blood (PB) would be very advantageous for both, patients and clinical practicability. In contrast to bone marrow (BM), peripheral blood is easily obtained during routine check-ups and disease-related alterations in DNA methylation are maintained, allowing detection in mature blood cells[207]. To study the behavior of both materials, DNA methylation patterns were compared using the Methyl-CpG-immunoprecipitation followed by next-generation sequencing (MCIp-seq).

We studied a cohort of eight donors of whom four were patients diagnosed with MDS and four healthy donors served as control group. From MDS patients both materials were acquired and analyzed, whereas only peripheral blood samples were available from healthy donors. Samples in both groups were matched regarding age and sex. For all healthy donors a laboratory workup including peripheral blast count was performed with no abnormalities found. Furthermore, MDS patients showed a good mixture regarding WHO subtypes and cytogenetic characteristics showed a wide variety ensuring no bias in clinical data (see Table 5-7).

Table 5-7 - Clinical data of healthy donors and MDS patients

no	Sample type	sex	Age (y)	WHO	Cytogenetic	BM blasts (%)
1	PB & BM MDS	f	49	MDS with del(5q)	del(5q)	<2 / 0 (BM / PB)
2	PB & BM MDS	m	67	MDS-MLD	46 XY	<5 / 3 (BM / PB)
3	PB & BM MDS	m	64	MDS-EB1	t(3;8)	7 / 0 (BM / PB)
4	PB & BM MDS	m	73	MDS-MLD	Complex karyotype	<5 / 0 (BM / PB)
6	PB healthy	m	64	-	-	-
7	PB healthy	m	64	-	-	-
8	PB healthy	f	58	-	-	-
9	PB healthy	m	87	-	-	-

Abbreviations: PB, peripheral blood; BM, bone marrow; MDS, myelodysplastic syndrome; f, female; m, male; MDS-MLD, MDS with multilineage dysplasia, MDS-EB1, MDS with excess blasts 1;del(5q), deletion of 5q

Peripheral blood and bone marrow samples were equally treated, first separating the mononuclear cells (MNCs) via density gradient centrifugation (see section 4.1.2.5) followed by isolation of genomic DNA. 200 ng of patient DNA was sonicated and methylated fragments were enriched using the MCIp protocol. Subsequent steps included library preparation and next generation sequencing on the Illumina HiSeq 3000/4000 platform.

Bowtie 2[208] was used to map the obtained sequencing data to the human reference genome (GRCh37/hg19). Downstream analysis (peak annotation, creation of tag directories, genome browser

visualization) of uniquely mapped tags were done using the HOMER software (Hypergeometric Optimization of Motif EnRichment)[177]. To compensate variations in sequencing depth, data was normalized to one tag per base pair (tbp) and analysis was done regardless of gender by removal of sex chromosomes. The free software "R"[209] was used for calculation of Pearson correlation coefficients, t-SNE visualization (t-Distributed Stochastic Neighbor Embedding) and drawing of corresponding heat maps. The Pearson correlation coefficient is a measure of the linear correlation between two variables, ranging from values of -1 to +1, the latter indicating a perfect correlation.

All commands used for this analysis can be found in section 4.2.3.2.2.

The resulting heat map of correlation coefficients showed a strong correlation between the paired samples (see Figure 5-32 A). The corresponding samples, BM and PB, of one patient were always located in the same cluster and are highly correlated. Comparison of clinical data with clustered patients showed no correlation regarding WHO subtype or sex.

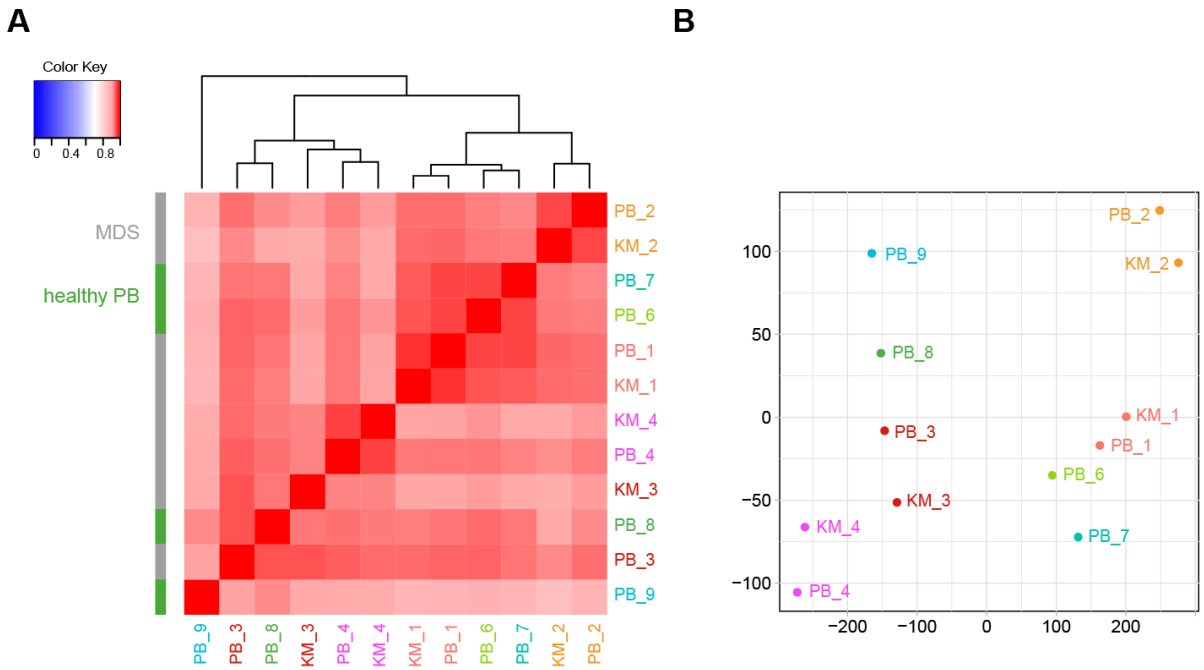


Figure 5-32 - Comparison of global DNA methylation in peripheral blood and bone marrow samples

(A) Heat map representing pairwise correlations using global DNA methylation data. Ordering of patients is based on hierarchical clustering using Pearson correlation, which results in three clusters with similar correlation illustrated by the dendrogram on top. The colors of the cells depict a high correlation (red) and a low correlation (blue). Same numbers were used for corresponding samples of one patient. The bar on the left shows sample category with MDS samples in grey and healthy donors in green. (B) t-SNE (t-Distributed Stochastic Neighbor Embedding) visualization of the global DNA methylation data. MDS patients were colored in red shades and healthy peripheral blood samples in green/blue ones.

A second approach to illustrate similarity of PB and BM samples was made with the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique. It allows the visualization of high-dimensional data sets by reducing the dimensions to a two-dimensional map (see Figure 5-32 B).

The t-SNE plot also represented similarity of peripheral blood and bone marrow samples due to close localization of the corresponding data points of one patient to each other. Moreover, we could see that

MDS patients exhibit only little aberrations on DNA methylation level because the distance to healthy PB samples is not very high. The only patient exhibiting more alterations in DNA methylation was patient #4, which could be explained by several cytogenetic lesions.

As depicted in Table 5-7, patient #4 shows a complex karyotype, whereas the other MDS patients show only one or even no cytogenetic aberration. Figure 5-33 exemplarily shows a section of the HOXA cluster, where differentially methylated regions were highlighted with black boxes. The left box shows a region where patient #4 shows less DNA methylation and the right box with more DNA methylation in comparison to other MDS samples.

In summary, two independent techniques suggested close resemblance of the two different sample types, peripheral blood and bone marrow, regarding global DNA methylation at CpG-rich regions. Nevertheless, a clear classification of healthy and diseased samples was not possible. This could due to the fact that the restriction to CpG rich regions using the MCIp-Seq approach leads to some loss of information regarding DNA methylation changes in CpG poor regions or maybe patients were not methylated differentially.

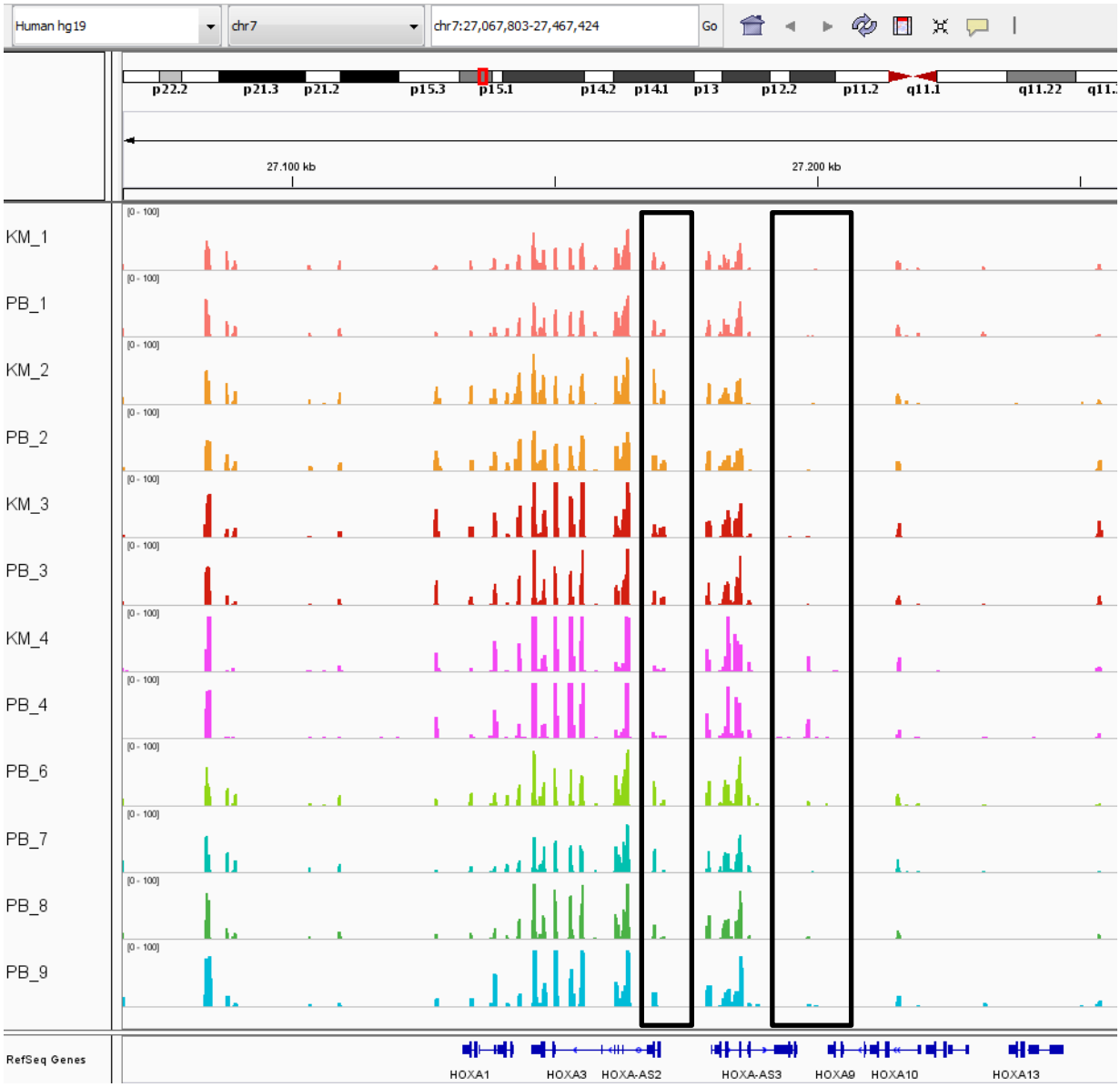


Figure 5-33 - Genome browser track of HOXA cluster for MDS patients and healthy donors

IGV snapshot of the HOXA cluster (hg19_chr7:27067803-27467424) with DNA methylation data derived from MCIp-seq for MDS patients and healthy donors. The first rows represent paired peripheral blood and bone marrow samples (shades of red and pink), whereas the bottom rows comprises all healthy peripheral blood samples (shades of green and turquoise). Reference genes are depicted at the bottom of the figure (blue). Black boxes represent differentially methylated regions regarding KM_4 and PB_4.

6 Discussion & Perspectives

Epigenetic mechanisms are known to control gene expression without changing the underlying genomic sequence. Besides genetic and cytogenetic alterations, DNA methylation changes were shown to play important roles in the development of several diseases[153–156,219,220]. Predominantly, hypermethylation in CpG-rich areas of the genome could be observed as well as global hypomethylation[158]. Myelodysplastic syndromes are one of these diseases showing DNA methylation alterations where extensive studies regarding the role of this epigenetic modification for disease development or progression are still missing.

With this thesis, I aimed to identify potential epigenetic marker genes that are involved in the progression from myelodysplastic syndromes (MDS) to acute myeloid leukemia (AML) or those playing a role in the pathogenesis of this heterogeneous disease. With an integrated analysis of epigenetic and genetic changes, we investigated the relationship between these two types of alterations. The studies were performed in cohorts of pediatric and adult MDS patients, which feature different genetic changes. For DNA methylation analysis we used two different approaches, methyl-CpG-immunoprecipitation-sequencing (MCIp-seq) for global analyses and bisulfite-sequencing for targeted analyses.

Global DNA methylation analyses in pediatric MDS patients allowed the separation of patients into two groups featuring consistent changes in DNA methylation at several interesting loci. However, analysis of consecutive samples in adults and in children did not show common patterns of differentially methylated regions (DMRs) which could serve as biomarkers for disease progression. Instead, progression associated changes were mostly private to individual patients and reflected their clonal evolution.

6.1 Epigenetic and genetic changes in MDS

Beside cytogenetic and genetic changes, epigenetic alterations, especially DNA methylation changes, may play a role in the pathogenesis and disease progression of myelodysplastic syndromes[154–156]. Genetic lesions were found in over 90 % and cytogenetic aberrations in about 50 % of MDS patients, while the latter is contributing to the classification into prognostic subgroups[109,116,117]. In contrast, DNA methylation changes are not routinely implemented for prognostication or classification, although several studies have shown an association of DNA methylation changes at specific genes with prognosis or pathogenesis[157]. DNA methylation is used as a therapeutic target during MDS treatment[157,221], while studies regarding the correlation of DNA methylation pattern and response to distinct treatments are still missing. Since one could possibly improve the diagnosis and risk

stratification of MDS by combining both, the analysis of genetic and epigenetic changes, we studied these two features in adult and pediatric patients suffering from MDS.

6.1.1 Comparison of DNA methylation data in adult MDS patients in order to identify common DMRs

As already mentioned, changes in DNA methylation patterns can alter gene expression and contribute to the development of different diseases. Frequent abnormal DNA methylation has been described in acute myeloid leukemia (AML)[153,178,222] and few studies revealed changes in specific genes in myelodysplastic syndromes (MDS)[223–225]. The list of aberrant methylated targets in MDS contains for instance p15INK4b, HIC1, CD1, ER[226], DLX4[224] or GPX3[223]. Few studies have investigated DNA methylation changes during the progression of myelodysplastic syndromes. These studies suggest that MDS stem cells generally exhibit DNA methylation changes and that methylation changes at gene promoters correlate with the progress of MDS[156,227].

In order to extend the set of marker genes and to identify potential target genes that are involved in the progression from MDS to AML, we analyzed a cohort of six adult MDS patients with a targeted bisulfite sequencing approach. This method analyzes DNA methylation of regions that are relevant for myeloid cell biology and acute myeloid leukemia. Our customized “myeloid regulome” comprises 85 mega bases (95K regions) equivalent to approximately 2.8 % of the human genome and includes 2,600,000 CpGs (see section 4.2.3.1).

Using Principal Component Analysis (PCA), segregation of the single patients could be observed, whereas the longitudinal samples from one patient were found to be in close proximity indicating high similarity between these samples. In addition, we included control data sets from CD34+ cells, CD14+ and CD15+ cells. Regarding these three healthy cell types, PCA segregated monocytes and granulocytes from CD34+ cells, reflecting the distinct DNA methylation landscape during differentiation that was demonstrated in a recent study. There, Farlik et al.[199] studied lineage-specific DNA methylation using WGBS (whole genome bisulfite sequencing) of 10 hematopoietic cell populations sorted by FACS (fluorescence-activated cell sorting). They observed lower average DNA methylation levels in differentiated cells of the myeloid lineage compared to all stem and progenitor cell types.

This first view on our DNA methylation data with PCA suggests that DNA methylation patterns between longitudinal samples from individual patients show few differences. To figure out similarities and differences between patients, we looked for differentially methylated regions (DMRs). Comparison of DMRs between the six patients revealed no common DMRs, but many DMRs in single patients during disease progression. This indicated that each patient shows private patterns of DNA methylation and almost no overlapping DMRs with other patients.

This point was analyzed with another approach by comparing differentially methylated regions in patients with those in CD34+ cells. Since hematopoietic stem cells show a different DNA methylation landscape than differentiated cells, we expected to get more DMRs that are common between all patients. In total, we observed 95 common DMRs in comparison to CD34+ cells and those could be

divided into 42 DMRs and 53 DMRs, which loose or gain DNA methylation in contrast to CD34+ cells, respectively. Nevertheless, most of DMRs found in contrast to CD34+ cells are again individually altered in each patient reflecting the heterogeneity of the disease.

Previous studies suggested that DNA methylation in cancer occurs at regions marked by the repressive mark H3K27me3 as well as at bivalent regions marked with both, H3K27me3 and H3K4me3[181–183,228]. Furthermore it was shown by Wong et al.[229] that DNA methylation preferentially occurs at H3K27me3 marked regions that regulate developmental processes. This hypothesis could also be confirmed with our data, where hypermethylated DMRs in contrast to CD34+ cells show an association with the PcG mark H3K27me3 and trimethylation of H3K4 indicating bivalent regions. Those DMRs were furthermore enriched in gene ontology terms affecting different developmental processes including “hematopoietic or lymphoid organ development”, “blood vessel development” or “regulation of cell development”, to name important ones for hematologic malignancies.

Moreover it was shown that tumors undergo global hypomethylation[158] as well as demethylation at specific sites. This demethylation process is normally associated with transcriptional reactivation and could lead to activation of proto-oncogenes in several tumors[36,230–232]. In gastric cancer, the group around Nam-Soon Kim[230] identified *ZNF312b* as a cell proliferation-associated oncogene and that its promoter demethylation promotes gastric tumorigenesis. Another study from Watt et al.[231] demonstrated that the HOX11 proto-oncogene is frequently activated in T-cell acute lymphoblastic leukemia (T-ALL) due to promoter demethylation.

For this reason, we also analyzed DMRs loosing DNA methylation in contrast to CD34+ cells. The regions affected by hypomethylation in this work were involved for instance in gene pathways responsible for “hematopoietic or lymphoid organ development”, “pathways in cancer”, “myeloid cell differentiation” or “regulation of catabolic processes”. We also observed an association of these DMRs in hematopoietic stem cells with the active histone marks H3K4me3 and H3K27ac. But we were not able to assign regions either specific for differentiation or cancer. Summarizing, we could say that hypomethylation of DMRs in patients may activate oncogenes that play a role for development or progress of the disease.

In both cases, hypermethylation and hypomethylation, DMRs were preferentially located in introns, while only 10 % were located close to transcription start sites. Hypomethylated DMRs were additionally located in intergenic regions, which is typical for global hypomethylation in cancer[43].

Since we aimed to find common regions that are differentially methylated in all patients, we focused on the 42 DMRs getting hypomethylated and on the 53 DMRs getting hypermethylated in contrast to CD34+ cells. The latter mentioned group contained DMRs involved in interesting gene pathways, like “positive regulation of hemopoiesis” and “methylation”. Especially genes enriched in the first term attracted our attention with RUNX1, FOXO3 and ZFPM1 as examples. The transcription factor RUNX1 is involved in lineage-specific differentiation and in the generation of hematopoietic stem cells[190]. Hypermethylation of the promoter region of RUNX1 could therefore lead to altered gene expression and misbalance proper hematopoiesis. Furthermore, it is known that mutations in *RUNX1* are described in about 10 % of acute myeloid leukemias emphasizing the important role of this transcription factor for normal hematopoiesis[233,234].

Another important transcription factor for the regulation of several cellular processes, including differentiation, proliferation and apoptosis, is the Forkhead box O 3 factor (FOXO3). Alterations of FOXO3 expression were found to be involved in tumorigenesis and progression[191,192]. As demonstrated by Ticchioni et al., transcriptional activation of FOXO3 is required to prevent B-chronic lymphocytic leukemia[235].
Another transcription factor enriched in gene ontology analysis in our hypermethylated DMRs was ZFPM1. This TF is regulating erythroid and megakaryocytic cell differentiation[193,194].
When tested in luciferase reporter assays, FOXO3 showed enhancer function that was abrogated by *in vitro* methylation of the reporter construct. This observation shows that hypermethylation of this DMR could play a role in the pathogenesis or progress of myelodysplastic syndromes. The other regions showed no significant alterations of reporter gene expression due to *in vitro* methylation. One reason for this may be the wrong selection of the analysed region or of the used cell line. To prove this, one has to choose another region around the target DMR and/or use another cell line.

In the last years, molecular profiling of myelodysplastic syndromes has been done to a large extent contributing to our understanding of the pathogenesis of this disease[116,236,237].
In myelodysplastic syndromes genetic lesions are found in over 90 % of patients[109,116] and mutational hierarchies dynamically change upon disease progression and treatment[170,238]. The clonal architecture plays an important role for therapy response[170] and may be associated with DNA methylation changes. To investigate this theory, we analyzed the interplay between genetic and epigenetic alterations at different stages of the disease.
All analyzed adult MDS patients were genetically and cytogenetically characterized by the group of Prof. Dr. Nowak (Mannheim) and Prof. Dr. Haase (Goettingen) allowing integrated analysis with DNA methylation data. All patients exhibit recurrent mutations affecting epigenetic modifiers (e.g. *TET2*, *ASXL1*, *DNMT3A*, *EZH2*), RNA splicing factors (e.g. *SF3B1*, *SRSF2*) or transcription factors (e.g. *RUNX1*, *ETV6*) and underwent different epigenetic therapy regimes. Although our DNA methylation study was limited to six patients with four to five consecutive samples each, we observed on the one hand common differences compared to CD34+ cells and on the other hand consistent inter-individual changes in DNA methylation correlating with the genetic landscape.
Pairwise comparison of DNA methylation data revealed a higher Pearson correlation between single samples from one patient than between the patient samples and CD34+ cells. This phenomenon is expected since hematopoietic stem cells exhibit a different DNA methylation pattern than more differentiated cells as described above[199]. Detailed analyses revealed clusters of differentially methylated regions in patients in comparison to CD34+ cells that can be divided into regions associated with differentiation and those specific for MDS. Differentiation specific clusters of regions were typically demethylated in contrast to CD34+ cells and showed less trimethylation of H3K27 but a high signal for H3K4 trimethylation. This indicates that those regions are probably transcriptionally active and promote transcription programs important for cellular maturation. Analysis of common transcription factor binding motifs in those demethylated regions showed enrichment of binding motifs for CEBPA, AP-1, BATF and FRA-1.

It was shown that these transcription factors regulate several differentiation processes[201,203,204] underpinning the hypothesis that those hypomethylated regions are involved in maturation of cells.
One enriched binding motif in our demethylated regions is CEBPA, where the group around Subramaniam Agatheeswaran[201] demonstrated that elevated expression of CEBPA leads to arrest of CML blasts and enforcement of granulocytic differentiation.
The function of the AP-1 transcription factor Fra1 was analysed by Grötsch et al.[204] and they demonstrated that it negatively regulates B cell function. This in turn inhibits plasma cell differentiation.
Another TF motif enriched in hypomethylated regions of our patients was BATF (basic leucine zipper transcription factor). The study from Kurachi et al.[203] revealed that BATF is required for CD8+ T cell effector differentiation. Therefore the enrichment of binding sites for CEBP, FRA-1 and BATF in these hypomethylated regions indicates a role of those regions for differentiation processes.

Clusters of regions which are hypermethylated in contrast to CD34+ cells showed high signals of H3K27 trimethylation indicating to be methylated in a disease-dependent manner. Moreover these regions were enriched for binding motifs for GATA and HOX, which are important for cellular differentiation and identity. Due to DNA methylation of these binding sites, processes regulated by GATA and HOX transcription factors may be disrupted and could contribute to MDS development or progression. On the other hand, these regions were already marked with H3K27me3 in HSCs and it is also possible that these regions were hypermethylated in the course of differentiation.
Comparison of DMRs with publicly available data of the FANTOM5 (F5) consortium showed a generally higher CpG methylation degree in F5 promoter, F5 enhancer or GATA2 bound regions in patient samples than in control cells (CD14+, CD15+ and CD34+ cells). As mentioned above, this is due to different DNA methylation landscapes occurring during hematopoiesis.
In order to see how different hematopoietic precursor cells are methylated in the patient specific DMRs, we annotated publicly available data sets into those regions. We generally observed the highest DNA methylation degree in all 3 subsets (F5 promoter, F5 enhancer and GATA2 bound regions) in Treg and CD8+ T cells, followed by naïve B cells and precursor lymphocyte of B lineage. Myeloid precursor cells (neutrophilic metamyelocyte and neutrophilic myelocyte) and mature myeloid cells (band form neutrophil and CD14+ CD16- classical monocyte) exhibit a similar or lower DNA methylation degree than hematopoietic multipotent progenitor cells. This phenomenon was previously demonstrated by Farlik et al.[199] where they analyzed single-cell methylomes in 10 different hematopoietic cell types revealing many regions with lower DNA methylation levels in myeloid cells than in lymphoid cells. To sum up, comparison of DNA methylation data of patients and CD34+ cells revealed differences in DNA methylation patterns with DMRs showing hypo- or hypermethylation due to differentiation processes or disease specific methylation processes, respectively.

Integrated analysis of genetics and epigenetics on the basis of individual patients indicated an association of DNA methylation changes with alterations in the genetic landscape.
In our cohort, 3 out of 6 patients showed disease progression with regard to the WHO subtype, while two of them showed a stable disease and one turned into hematologic remission. Mossner et al.[170] could demonstrate that clinical progress regarding WHO subtypes is associated with complex

mutational hierarchies. 52 MDS patients at several time points were analysed using whole exome sequencing and targeted deep-sequencing to resolve molecular dynamics during drug treatment and progression. They demonstrated that clonal architecture changes upon treatment with lenalidomide and other drugs as well as with disease progression into worse WHO categories.

In our study, upon treatment with lenalidomide (LEN), patients P02 and P15 showed a loss of their del(5q) bearing subclones, while earlier founding clones carrying *DNMT3A*/*RAET1G* (P02) or *ASXL1* (P15) strongly expanded. Clinically, P02 turned into hematologic remission after treatment with LEN accompanied by changes in DNA methylation pattern. In patient P15 we observed a switch from MDS with del(5q) into MDS-MLD and simultaneous shifts in clonal architecture between sample one and two. DNA methylation data nicely correlates with these changes, where sample one shows completely different DNA methylation patterns than samples two to four with stable genetic landscape and WHO subtype.

Patient P20 underwent also treatment with lenalidomide despite the absence of the target lesion, del(5q). Here, only small changes in mutational burden were observed and furthermore WHO subtype (MDS-EB1) remained stable over time. At time point two, VAFs of all mutations were a little bit lower and there we observed a slight different DNA methylation pattern tending to be less methylated than the other samples. This phenomenon could possibly attribute to the preceding treatment with LEN that is known to have also cytotoxic effects[239] resulting in a lower cell frequency included in analysis.

Patient P19 was a case observed with MDS-RS-SLD over all four samples and was treated with an experimental anti-CD95 antibody (APG101). DNA methylation pattern showed very low variance between the different samples of this patient, consistent with the molecularly and clinically stable disease. Serial follow up bone marrow samples from patient P13 demonstrated the existence of the same mutations over time but with varying VAFs. After sample 2 variant allele frequencies for *SF3B1*/*TP53* increased rapidly with climax at sample 3. In parallel, a slight decrease of the *TNIK* subclone could be observed. The patient was classified initially with MDS with del(5q), progressed towards MDS-EB1 and finally turned into a secondary AML. Comparison with DNA methylation data presented a link between this epigenetic mark and changes in VAFs for mutations.

A more complex situation was found in patient P53, where DNA methylation analysis was done in peripheral blood samples, but genetic and cytogenetic assays with both, peripheral blood (PB) and bone marrow (BM) samples. This was due to implementation problems of cytogenetic analysis in peripheral blood and therefore switching to bone marrow was necessary. Clinically, the patient progressed from MDS-EB2 to secondary AML (sAML) despite treatment with 5'-Azacitidine. Results of molecular genetics in peripheral blood showed decreasing VAFs for the *RUNX1* founder clone and the *SRSF2* subclone over time. Changes of cytogenetic aberrations in the bone marrow were more complex with complicated clonal hierarchies for every single sample. Regarding DNA methylation data, which show various patterns over all time points, we could not identify an association between those patterns and cytogenetic / molecular changes.

A recent study of epigenetic and genetic heterogeneity in AML suggests that epigenetic heterogeneity is associated with inferior outcome and shows significant variation during disease progression[172]. Herein epigenetic heterogeneity represents the pattern of DNA methylation at defined genomic loci with four CpGs, the so called epigenetic alleles. Furthermore they claimed that epigenetic and genetic

heterogeneity develops independently during disease progression. Since sequencing depth of our DNA methylation data was not sufficient for epigenetic heterogeneity analysis like described here, we were not able to assign the same phenomenon in myelodysplastic syndromes.
But we could demonstrate a patient specific and globally stable DNA methylation pattern during disease progression. There was very little progressive DNA hypermethylation observed. However, changes in DNA methylation patterns correlate with selection of genetically different subclones.
But it has to be considered that every single sample of the patient is an individual mixture of healthy and diseased cells with different ratios. Therefore some changes in DNA methylation patterns could also be due to the heterogeneity of the analyzed samples.
However, it remains still unclear what came first – changes in genetics or DNA methylation. There is some evidence that genetic changes trigger DNA methylation alterations. On the one hand, DNA methylation profiles correlate with the selection of genetically distinct subclones indicating DNA methylation changes to be more "passenger" like events. On the other hand, observed mutations occur mainly in genes coding for epigenetic modifiers. Therefore it is postulated, that alterations in molecular genetics triggers the development of MDS[240].
Nevertheless, the "chicken-egg-problem" needs to be solved by further integrating analysis of MDS patients with disease progression, perhaps using single-cell technologies.

6.2 Integrated analysis of epigenetic and genetic changes in pediatric MDS patients

In contrast to adult MDS, pediatric MDS is rather rare with an incidence of 0.5 – 4 / 10^6 per year[108,112] and extensive studies regarding epigenetics and genetics are still missing. Another difference between the two entities is the mutational pattern where adult MDS patients preferentially exhibit acquired mutations in *DNMT3A*, *ASXL1*, *TET2* and *SF3B1* and children often inherited mutations in *GATA2*, *SAMD9*[165] or *FANC* members[166]. Furthermore, adult MDS patients were preferentially found with del(5q) lesions, whereas children often carry deletions of chromosome 7 or 7q[110,164]. Due to the discrepancy regarding pathogenesis of pediatric and adult MDS, we wanted to investigate DNA methylation in childhood MDS separately. Despite the rarity of pediatric MDS, our cohort comprises 42 patients with a median age of 10.5 years and a balanced female to male ratio. From eight out of these 42 patients, longitudinal sample were available allowing DNA methylation analysis during disease progression. The patient cohort was furthermore enriched for germline *GATA2* mutations (59.5 %) and monosomy 7 (71.4 %). It was previously demonstrated by Marcin et al.[205], that monosomy 7 lesions are preferentially associated with *GATA2* mutations. Therefore we were interested in potentially different DNA methylation patterns in *GATA2* mutated and non-*GATA2* mutated patients.
Global DNA methylation data was obtained from purified bone marrow granulocytes using the Methyl-CpG-immunoprecipitation (MCIp) approach followed by next-generation sequencing. In a first subset analysis with all 42 patients, but without consecutive samples, t-SNE segregated the patients into two

clusters according to their DNA methylation data. One of those clusters could be observed with a significant correlation of *GATA2* mutation status (p-value = 0.0205) and WHO subtype "refractory cytopenia" (p-value = 0.022). Moreover we could confirm the findings by Marcin et al.[205] that *GATA2* mutations were preferentially associated with monosomy 7 lesions. But here one has to keep in mind that this patient cohort was biased towards these two aberrations.

Another connection in childhood MDS was found between mutations of *SETBP1* and *ASXL1*. Inoue et al.[210] demonstrated that in MDS patients *SETBP1* mutations were enriched in those with mutated *ASXL1*. Our results of t-SNE clustering revealed that 6 out of 8 patients (75 %) with mutations in *ASXL1* carry mutations in *SETBP1*. This correlation is preferentially found in the cluster of patients with more advanced subtypes. Summarizing the above, we could identify two groups of patients according to their DNA methylation data. The first cluster predominantly comprises MDS patients with *GATA2* mutations and refractory cytopenia (RC) indicating to include patients with low risk. Patients in the second cluster tend to be more advanced due to their mutations in *SETBP1* and *ASXL1* as well as their WHO classification into RAEB, RAEB-t and MDR-AML.

In order to get more details about differences in DNA methylation between the two clusters of patients, we identified differentially methylated regions (DMRs) restricted to promoter regions associated with transcription or chromatin factors. Three interesting DMRs specifically methylated in cluster 1 compared to cluster 2 could be identified, namely ZIC5, VILL and TRIM45. Zinc finger protein of the cerebellum 5 (ZIC5) exhibits transcriptional repressor function and was shown to be higher expressed in several human cancers. Studies in melanoma[211], non-small cell lung cancer[213] as well as in prostate and colorectal cancer[212] demonstrated that elevated expression of ZIC5 could contribute to cancer progression. DNA methylation data in the ZIC5 promoter region revealed a DNA methylation signal in patients belonging to cluster 1 and a loss of this epigenetic mark in cluster 2.

Methylation and subsequent transcriptional repression of the potential tumor suppressors VILL and TRIM45 in patients belonging to cluster 1 could also contribute to disease development or progression. Senchenko et al.[214] identified VILL as novel tumor suppressor in cervical cancer and TRIM45 was shown to play important roles in the brain regarding tumor suppression[215,216].

Altogether, those demethylation events probably contribute to higher transcription rates and therefore might contribute for disease progression in this patient group.

A major aim of this thesis was to identify potential epigenetic target genes that are involved in the progression from MDS to AML. The above mentioned results focused on global DNA methylation analysis in the complete patient cohort, but not in longitudinal samples and thus do not provide information regarding alterations during disease progression. In order to investigate this topic, we analyzed global DNA methylation in consecutive patient samples from eight children suffering from MDS. To our knowledge, analysis of DNA methylation in longitudinal pediatric MDS samples does not exist until now thus representing very valuable results. All eight patients in this study cohort carried germline *GATA2* mutations and were classified as RC or RAEB.

Interestingly, global DNA methylation patterns of consecutive samples from each patient remained very stable over time. Quantitative changes correlate with alterations of clonal sizes. Contrary to the literature[172], our findings suggest that there is an association between kinetics and patterns in the

genetic compartment with DNA methylation changes. However, the groups around Mason and Melnick demonstrated that epigenetic and genetic heterogeneity in AML show different kinetics and patterns during disease progression. It has to be mentioned, that they studied the interplay of genetic and epigenetic heterogeneity in acute myeloid leukemia and that these two features could be linked differently in other types of tumors.

We divided our patient cohort into those with stable disease and those with disease progression. A case with progression was found in patient D770 that was analyzed at five different stages of the disease and DNA methylation changes could be observed in every single sample. In addition to germline *GATA2* mutations, the patient exhibited mutations in *ASXL1*, *SETBP1* and *CBL* with varying VAFs of these lesions during progression. Moreover, the patient received several cycles of 5'-Azacitidine and hematopoietic stem cell transplantation (HSCT). The first two samples were taken before HSCT and show the most similar DNA methylation patterns. After HSCT, clonal architecture changed completely which was accompanied by abundant changes in DNA methylation data. We furthermore observed that 5'-Azacitidine did not have an impact on DNA methylation pattern.
In the past, 5'-Azacitidine (Aza) was believed to be cytostatic at higher doses[241]. Nowadays, 5'-Azacitidine is known to have inhibitory roles regarding DNA methylation when used at lower doses[161]. Recently, a structured study from Tobiasson et al.[217] suggested Aza to have several mechanisms of actions including immunomodulatory effects. They showed a general increase in gene expression but no correlation with DNA methylation changes upon treatment with Azacitidine. In detail, they analyzed gene expression, DNA methylation and histone modifications H3K18ac and H3K9me3 (via ChIP-seq) in primary bone marrow MDS cells from 11 patients. Treatment with Aza did not influence H3K18ac modifications and only weak changes could be observed regarding the repressive H3K9 trimethylation. Regarding DNA methylation, only a very modest degree of demethylation was observed, especially in regions annotated as heterochromatin. Therefore demethylation of these regions was probably not responsible for increased gene expression. Furthermore, 5`-Azacitidine was observed to have immunomodulatory functions on natural killer cells, T cells and dendritic cells indicating Aza to have anti-tumor benefits[242].
Beside these effects, Aza induced expression of several endogenous retroviruses (ERVs) which in turn is correlated with embryonal cell differentiation. This was confirmed by a recent study by the group around Christoph Plass, which demonstrated that Aza activates endogenous retroviral elements and promoters from long terminal repeats (LTRs)[243]. ERV reactivation comes along with synthesis of double-stranded RNA which in turn activates antiviral response pathways leading to apoptosis of the cell. So overall, it is not surprising that DNA methylation changes upon treatment with 5'-Azacitidine were not observed.

A more detailed view on differentially methylated regions of the distinct time points revealed seven interesting regions potentially important for disease progression (VENTX, SIX5, CREB3L1, GADD45B, ETV5, ETV2 and FOXO6). Tested in luciferase reporter assays, only VENTX and SIX5 showed significant abrogation by *in vitro* methylation of the reporter construct suggesting to be important methylation sensitive regions for disease progression. Literature research yielded interesting

information about the role of VENTX in cancer. Gentner et al.[218] published a report in 2016 showing that VENTX is highly expressed in acute myeloid leukemia (AML) and is furthermore involved in human erythroid differentiation. Strikingly, DNA methylation signals in the promoter region of VENTX significantly dropped during disease progression. This demethylation and the resulting elevated transcription of VENTX could be involved in the development of MDS or contribute to disease progression.

So far, SIX5 is only poorly investigated in leukemia. There are two studies in solid tumors demonstrating that SIX5 is involved in tumor initiation and progression. Xu et al. showed in breast cancer that patients with overexpression of SIX factors are associated with poorer clinical outcome[244]. In non-small cell lung cancer, Liu et al. obtained similar results with greater possibility of tumorigenesis due to higher expression of Sineoculis homeobox homolog (SIX) family proteins[245]. Observed methylation dependent downregulation of the SIX5 reporter gene suggests that demethylation of this region in patients may play a role in development of myelodysplastic syndromes.

The enhancer region of GADD45B was also observed to be demethylated between the different time points. This tumor suppressor gene was shown to be overexpressed in colorectal cancer by Wang et al.[246]. Since the enhancer region of GADD45B didn´t show significant abrogation due to *in vitro* methylation, it is possible that expression of this gene is not reduced via DNA methylation and is therefore not as important as expected.

The ETS Variant 2 transcription factor is required for tumor angiogenesis, which was recently demonstrated by the group around Saulius Sumanas[247]. But intensive studies regarding the role of ETV2 are still missing. Elevated expression of FOXO6 were associated with progressive gastric cancer, respectively[248]. Analysis of DNA methylation dependence of these last 2 mentioned enhancer regions did not show an abrogation due to *in vitro* methylation. This leads to the conclusion that these cancer relevant genes probably do not play a role during MDS progression.

The remaining seven patients in our pediatric MDS cohort showed a stable disease with no progression regarding WHO subtype, molecular genetics or cytogenetics. All patients exhibited germline *GATA2* mutations. Scatter plots of DNA methylation data revealed a very high correlation between the single time points of each patient. Generally, lower correlation values between patient samples were observed in patients D271 and D762 that received hematopoietic stem cell transplants (HSCT) and suffered a relapse. The decrease in correlation can be explained by the presence of donor cells as well as recipient cells resulting in a different DNA methylation pattern. Patients with a stable disease did not show many regions specifically methylated compared to control monocytes and analysis of DMRs which could be different between the samples of one single patient revealed no results. We only see changes in DNA methylation data to a distinct degree after stem cell transplantation. There we have to mention that all analyzed patients relapsed after HSCT and therefore DNA methylation changes could be on the one hand due to a distinct upcoming diseased cell population or on the other hand due to the different donor hematopoiesis.

In summary, DNA methylation patterns correlate with changes in genetics, cytogenetics and WHO subtype but remain stable if no progression in these parameters is observed. One has to keep in mind that this result could be due to detection limitation of the method. Since all patient samples consist of healthy and diseased cells to a different extent, with very few bone marrow blasts in most cases, DNA

methylation data always represents a mean of these two entities. The problem could be circumvented by using single cell analysis or samples with sorted population of cells.

6.3 Comparability of DNA methylation data in paired samples of peripheral blood and bone marrow

Routine diagnosis for myelodysplastic syndromes still comprises morphologic and cytogenetic analysis of bone marrow aspirates. In contrast to peripheral blood, which is readily accessible, bone marrow samples are hard to obtain. Despite differences in cell composition in peripheral blood and bone marrow, it was shown that disease-related alterations in DNA methylation are maintained in mature blood cells[249]. Zhou et al. analysed the methylome of MDS peripheral blood cells using the HELP assay (HpaII tiny fragment enrichment by ligation-mediated PCR assay). The epigenetic pattern found in MDS leukocytes was globally distinct from age-matched controls and characterized by numerous, aberrant hypermethylated marks. By an integrative epigenetic-genetic approach combining HELP assay and aCGH (comparative genomic hybridization) Dock4 could be found as a candidate pathogenic gene on chromosome 7q. To sum up, this study showed that disease-specific DNA methylation alterations - relating here to MDS - are maintained in mature blood cells. Affected regions in the course of the disease are mainly CpG islands (CGI), which are not regulated during hematopoietic differentiation[250]. Those CGIs have a high CpG content and are therefore predominantly enriched with our methyl-CpG-immunoprecipitation (MCIp) approach. Altogether, this indicates that peripheral blood could also serve as a material source for analyzing disease specific DNA methylation changes.

In this work, analysis of MCIp-seq data from paired samples of peripheral blood and bone marrow demonstrated a high similarity of DNA methylation patterns in both materials. This was shown with two different bioinformatic approaches, hierarchical clustering of Pearson correlation values and t-Distributed Stochastic Neighbor Embedding (t-SNE).

Pearson correlation values were overall greater than 0.8 for comparison of peripheral blood and bone marrow samples indicating a high degree of similarity in two different materials from one patient. The reasons why there is no perfect correlation could be on the one hand experimental variation or on the other hand the different amount of aberrant cells analyzed in both tissues. Since every patient sample is an individual mixture of healthy and aberrant cells, one will never obtain the same composition of cells in two samples of the same origin. To circumvent this problem, one would have to perform single-cell-analysis, which however is expensive, time consuming and not well established for DNA methylation so far and therefore not practicable for routine MDS diagnosis.

Furthermore, we see that healthy control peripheral blood samples did not form their own cluster demonstrating that healthy donors show individual DNA methylation patterns to some degree. But it must be taken into account that we used aged-matched healthy controls which already may exhibit age-related alterations in DNA methylation patterns. Until now, there is no literature available

regarding DNA methylation and age-related hematopoiesis. But there are several studies about the connection of molecular genetics and age-related clonal hematopoiesis. In general it was observed that somatic mutations in genes associated with hematological neoplasia (e.g. DNMT3A, TET2, JAK2, SF3B1, ASXL1, TP53…) are frequently acquired during human aging without evidence of a blood disorder[169,251–253]. This phenomenon has been termed clonal hematopoiesis of indeterminate potential (CHIP) and appears in 10 % among persons between 70 and 80 years[168]. CHIP is furthermore associated with an increased risk to develop hematologic cancer and all-cause mortality[254].

So these studies illustrated that molecular genetics can be changed during aging representing a pre-malignant state. The same may apply for alterations in DNA methylation patterns, which would explain our observed results. And as described above, we detected a close relationship between the genetic architecture and epigenetic changes during MDS progression underpinning the theory that DNA methylation changes may occur together with aging and emergence of somatic mutations.

With regard to MDS patients, similarities were hardly detected which is probably due to the small cohort and therefore almost no statistical significance.

Altogether, the data provide strong evidence that DNA methylation data from the two different tissue samples, peripheral blood and bone marrow of the same patient, show high similarity. Therefore peripheral blood samples could be used instead of bone marrow samples at least for selected methods or questioning.

6.4 Perspectives

Analysis of DNA methylation in adult and pediatric MDS patients revealed the co-occurrence of DNA methylation changes with alterations in clonal architecture. Global patterns of DNA methylation were very consistent over time and furthermore specific for every single patient.

The individuality of DNA methylation patterns in patients led to identification of only few commonly differentially methylated regions that may be involved in disease pathogenesis or progression. Here additional studies in a greater patient cohort will be necessary to find an optimal set of epigenetic markers. This is true for both, adult and pediatric MDS patients, where the focus should be on patients undergoing disease progression since we could demonstrate that patients with a stable disease also show consistent DNA methylation patterns. After identification of a potential set of epigenetic marker genes, validation of these regions should be done with another method that monitors exactly epigenetic alterations during disease progression. This could be realized with the Epityper® MassARRAY, that uses bisulfite treated DNA to measure quantitatively DNA methylation at single CpG levels.

Our integrated analysis of epigenetic and genetic changes during disease progression did not reveal a clear chronological order of these two events. There are two possible scenarios explaining the emergence of DNA methylation differences. On the one hand, there could be accumulation of DNA methylation changes over time leading to genomic instability and finally results in genomic or cytogenetic aberrations[255,256]. On the other hand, it could be the other way round, that a genetic hit

triggers changes in DNA methylation[257]. In order to clarify this, one could do DNA methylation and mutational analysis in single cells. Furthermore it would be beneficial to increase the number of analyzed samples from one patient during observation period, to get a close monitoring of changes in epigenetics and genetics. This would facilitate correlation of these two changes and maybe dissolve the hierarchy of chronological appearance.

Another interesting study could be the characterization of DNA methylation pattern in hematopoietic precursor cells. Comparison of diseased patient samples could then reveal the similarity with distinct stages of differentiation and maybe improve diagnosis and prognostication.

Since DNA methylation changes could occur across the whole genome and our methods were focused on specific targets (myeloid regulome) or CGIs (MCIp), it would be more informative to do whole genome bisulfite sequencing (WGBS). However, this method is very expensive and is thus not applicable for clinical routine investigations.

7 Summary

DNA methylation is important during development of vertebrate organisms as well as for sustaining genome integrity and gene expression. Alterations of DNA methylation patterns are often associated with different diseases, for instance myelodysplastic syndromes (MDS) or acute myeloid leukemia (AML). Consequences of aberrant DNA methylation are the silencing of tumor suppressor genes due to hypermethylation as well as the hypomethylation-mediated weakening of transcriptional repression, reactivation of retrotransposons and genomic instability.

The major aim of this thesis was the integrated analysis of epigenetic and genetic changes during disease progression to identify target genes that could be involved in development or progression of myelodysplastic syndromes. To address this issue, DNA methylation analysis was performed in pediatric and adult MDS patients using the methyl-CpG-immunoprecipitation sequencing approach (MCIp-seq) and the targeted bisulfite sequencing of the myeloid regulome, respectively.

It could be demonstrated that adult MDS patients show largely private DNA methylation changes and almost no common differentially methylated regions. Identified DMRs include RUNX1, FOXO3 and ZFPM1, which show methylation sensitivity in *in vitro* reporter gene assays. Another observation was made in this patient cohort in which DNA methylation changes only occur with alterations in clonal architecture. In cases of a genetically stable disease, no differences in DNA methylation patterns were observed over time.

In pediatric MDS patients global DNA methylation analysis revealed a correlation of DNA methylation changes with germline *GATA2* mutations and refractory cytopenia (RC). In detail, a patient cluster with lower DNA methylation degree exhibited the mentioned two features, while the other cluster of patients was associated with more advanced subtypes and higher DNA methylation. DMRs identified between these two patient groups are ZIC5, VILL and TRIM45, possibly playing a role in cancer. Methylation sensitivity of these regions has to be tested with *in vitro* reporter gene assays and will give information about a possible role for the development or progress of MDS.

Regarding longitudinal studies in pediatric MDS patients, we could show the same result like in adult patients where DNA methylation changes correlate with alterations in genetic landscape. One potential epigenetic target gene found to be methylation sensitive and already described to play an important role in AML is the *VENTX* promoter region.

In summary, our data suggest a tight correlation of epigenetic changes with clonal architecture of the diseased hematopoiesis, but the chronological order of appearance is still an open issue.

8 Zusammenfassung

DNA-Methylierung ist ein wichtiger Prozess bei der Entwicklung von Organismen und spielt ebenfalls eine wichtige Rolle bei der Aufrechterhaltung der genomischen Integrität sowie der Genexpression. Eine Assoziation von Veränderungen im DNA-Methylierungsmuster konnte für verschiedene Erkrankungen festgestellt werden, unter anderen bei den myelodysplastischen Syndromen (MDS) und der akuten myeloischen Leukämie (AML). Durch die aberrante DNA-Methylierung kommt es zur Stilllegung von Tumorsuppressorgenen, die durch eine Hypermethylierung hervorgerufen wird und auf der anderen Seite zu einer Abschwächung der transkriptionellen Repression, Reaktivierung von Retrotransposons sowie genomischer Instabilität aufgrund einer Hypomethylierung.

Das Ziel der vorliegenden Arbeit war die integrierte Analyse epigenetischer und genetischer Veränderungen während dem Progress des MDS, um potentielle Zielgenen zu identifizieren, die bei der Entwicklung oder dem Voranschreiten der Erkrankung involviert sind.

Im Zuge dessen wurden globale DNA-Methylierungsanalysen in einer kindlichen MDS-Kohorte mittels Sequenzierung von Methyl-CpG-Immunopräzipitationen (MCIp-Seq) durchgeführt. Bei den adulten MDS-Patienten wurde gezielt die DNA-Methylierung in myeloischen, regulatorischen Komponenten mittels Bisulfit-Sequenzierung analysiert, wodurch der Fokus auf die für die Entwicklung myeloischer Zelltypen wichtigen Regionen gelegt wurde.

Die globale DNA-Methylierungsanalyse der 42 kindlichen MDS-Patienten zeigte, dass diese ein differenzielles DNA-Methylierungsmuster im Vergleich zu den Kontrollen sowie eine signifikante Korrelation der DNA-Methylierung mit dem *GATA2*-Mutationsstatus und dem WHO-Subtyp „Refraktäre Zytopenie“ aufweisen. Im Detail sind die Patienten mit einer geringeren DNA-Methylierung mit den beiden genannten Merkmalen assoziiert, während die andere Patientengruppe eine Korrelation mit höheren DNA-Methylierungen und fortgeschritteneren WHO-Subtypen aufweist. Zwischen diesen beiden Entitäten konnten differentiell methylierte Regionen (DMRs) ausfindig gemacht werden, die eine Rolle bei Tumoren spielen. Diese Liste beinhaltet ZIC5, VILL und TRIM45. Die Methylierungssensitivität dieser Regionen mithilfe von *in-vitro* Reportergenassays bleibt noch zu testen, um Aussagen über ihre mögliche Rolle bei der Entwicklung und dem Progress von MDS treffen zu können. Longitudinale Analysen in kindlichen MDS-Patienten zeigten ein analoges Ergebnis wie bei den adulten Patienten. Das DNA-Methylierungsmuster bleibt während des Krankheitsprogresses gleich, die quantitativen Veränderungen korrelieren jedoch mit der Klongröße. Es konnte eine differentiell methylierte Region gefunden werden, die Promoterregion von *VENTX*, welche methylierungssensitiv ist und bereits bei der AML beschrieben wurde.

Zusammenfassend können wir sagen, dass unsere Daten eine enge Korrelation der epigenetischen Veränderungen mit der klonalen Architektur der erkrankten Hämatopoese aufzeigt, jedoch bleibt die chronologische Abfolge dieser Veränderungen noch zu klären.

9 References

1. Orkin SH. Diversification of haematopoietic stem cells to specific lineages. *Nat. Rev. Genet.* 2000;1(1):57-64.
2. Jagannathan-Bogdan M, Zon LI. Hematopoiesis. *Development (Cambridge, England)*. 2013;140(12):2463-2467.
3. Blank U, Karlsson S. TGF-β signaling in the control of hematopoietic stem cells. *Blood*. 2015;125(23):3542-3550.
4. Orkin SH, Zon LI. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell*. 2008;132(4):631-644.
5. Steffen B, Müller-Tidow C, Schwäble J, Berdel WE, Serve H. The molecular pathogenesis of acute myeloid leukemia. *Crit. Rev. Oncol. Hematol.* 2005;56(2):195-221.
6. Waddington CH. The epigenotype. 1942. *International journal of epidemiology*. 2012;41(1):10-13.
7. Rodríguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. *Nat Med*. 2011:330-339.
8. Zhang G, Pradhan S. Mammalian epigenetic mechanisms. *IUBMB life*. 2014;66(4):240-256.
9. Marchese FP, Huarte M. Long non-coding RNAs and chromatin modifiers: their place in the epigenetic code. *Epigenetics*. 2014;9(1):21-26.
10. Bheda P, Schneider R. Epigenetics reloaded: the single-cell revolution. *Trends in cell biology*. 2014;24(11):712-723.
11. Lechner M, Boshoff C, Beck S. Cancer epigenome. *Advances in genetics*. 2010;70:247-276.
12. Bell O, Tiwari VK, Thomä NH, Schübeler D. Determinants and dynamics of genome accessibility. *Nature reviews. Genetics*. 2011;12(8):554-564.
13. Goyama S, Kitamura T. Epigenetics in normal and malignant hematopoiesis: An overview and update 2017. *Cancer science*. 2017;108(4):553-562.
14. Conaway JW. Introduction to Theme “Chromatin, Epigenetics, and Transcription”. *Annu. Rev. Biochem.* 2012;81(1):61-64.
15. Szerlong HJ, Hansen JC. Nucleosome distribution and linker DNA: Connecting nuclear function to dynamic chromatin structure. *Biochemistry and cell biology = Biochimie et biologie cellulaire*. 2011;89(1):24-34.
16. Nikolov M, Fischle W. Systematic analysis of histone modification readout. *Molecular bioSystems*. 2013;9(2):182-194.
17. Mariño-Ramírez L, Kann MG, Shoemaker BA, Landsman D. Histone structure and nucleosome stability. *Expert review of proteomics*. 2005;2(5):719-729.
18. Li G-M. Decoding the histone code: Role of H3K36me3 in mismatch repair and implications for cancer susceptibility and therapy. *Cancer research*. 2013;73(21):6379-6383.

19. La Cruz X de, Lois S, Sánchez-Molina S, Martínez-Balbás MA. Do protein motifs read the histone code? *BioEssays : news and reviews in molecular, cellular and developmental biology*. 2005;27(2):164-175.
20. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell research*. 2011;21(3):381-395.
21. Kouzarides T. Chromatin modifications and their function. *Cell*. 2007;128(4):693-705.
22. Bhartiya D, Lenka N, eds. *Pluripotent stem cells.* Edition 2014. Croatia: InTech; 2013.
23. Marmorstein R, Zhou M-M. Writers and readers of histone acetylation: Structure, mechanism, and inhibition. *Cold Spring Harbor perspectives in biology*. 2014;6(7):a018762.
24. D'Oto A, Tian Q-W, Davidoff AM, Yang J. Histone demethylases and their roles in cancer epigenetics. *Journal of medical oncology and therapeutics*. 2016;1(2):34-40.
25. Tsukada Y-i, Fang J, Erdjument-Bromage H, et al. Histone demethylation by a family of JmjC domain-containing proteins. *Nature*. 2006;439(7078):811-816.
26. Greer EL, Shi Y. Histone methylation: A dynamic mark in health, disease and inheritance. *Nat Rev Genet*. 2012;13(5):343-357.
27. Bernstein BE, Meissner A, Lander ES. The Mammalian Epigenome. *Cell*. 2007;128(4):669-681.
28. Zhang T, Cooper S, Brockdorff N. The interplay of histone modifications - writers that read. *EMBO reports*. 2015;16(11):1467-1481.
29. Singal R, Ginder DG. DNA Methylation. *Blood*. 1993;(93):4059-4070.
30. Jang HS, Shin WJ, Lee JE, Do JT. CpG and Non-CpG Methylation in Epigenetic Gene Regulation and Brain Function. *Genes*. 2017;8(6).
31. Fazzari MJ, Greally JM. Epigenomics: beyond CpG islands. *Nat Rev Genet*. 2004;5(6):446-455.
32. Hackett JA, Surani MA. DNA methylation dynamics during the mammalian life cycle. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. 2013;368(1609):20110328.
33. Messerschmidt DM, Knowles BB, Solter D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev.* 2014;28(8):812-828.
34. Hanna CW, Kelsey G. Genomic imprinting beyond DNA methylation: A role for maternal histones. *Genome biology*. 2017;18(1):177.
35. Li E, Beard C, Jaenisch R. Role for DNA methylation in genomic imprinting. *Nature*. 1993;366(6453):362-365.
36. Singal R, Ginder GD. DNA methylation. *Blood*. 1999;93(12):4059-4070.
37. Moore LD, Le T, Fan G. DNA methylation and its basic function. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*. 2013;38(1):23-38.
38. Huang K, Fan G. DNA methylation in cell differentiation and reprogramming: An emerging systematic view. *Regenerative medicine*. 2010;5(4):531-544.
39. Lokk K, Modhukur V, Rajashekar B, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome biology*. 2014;15(4):r54.
40. Ehrlich M, Lacey M. DNA methylation and differentiation: Silencing, upregulation and modulation of gene expression. *Epigenomics*. 2013;5(5):553-568.

41. Baylin SB. DNA methylation and gene silencing in cancer. *Nature clinical practice. Oncology*. 2005;2 Suppl 1:S4-11.
42. Lee C-J, Evans J, Kim K, Chae H, Kim S. Determining the effect of DNA methylation on gene expression in cancer cells. *Methods in molecular biology (Clifton, N.J.)*. 2014;1101:161-178.
43. Klutstein M, Nejman D, Greenfield R, Cedar H. DNA Methylation in Cancer and Aging. *Cancer research*. 2016;76(12):3446-3450.
44. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet*. 1998;19(3):219-220.
45. Lan J, Hua S, He X, Zhang Y. DNA methyltransferases and methyl-binding proteins of mammals. *Acta Biochimica et Biophysica Sinica*. 2010;42(4):243-252.
46. Lyko F. The DNA methyltransferase family: A versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* 2018;19(2):81-92.
47. Meng H, Cao Y, Qin J, et al. DNA methylation, its mediators and genome integrity. *International journal of biological sciences*. 2015;11(5):604-617.
48. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*. 2006;31(2):89-97.
49. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet*. 2009;10(5):295-304.
50. Lauster R, Trautner TA, Noyer-Weidner M. Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *Journal of molecular biology*. 1989;206(2):305-312.
51. Pósfai J, Bhagwat AS, Pósfai G, Roberts RJ. Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Research*. 1989;17(7):2421-2435.
52. Orlanski S, Labi V, Reizel Y, et al. Tissue-specific DNA demethylation is required for proper B-cell differentiation and function. *Proc. Natl. Acad. Sci. U.S.A.* 2016;113(18):5018-5023.
53. Cheng X. nar-02599-f-2011-File003.
54. Klug M, Schmidhofer S, Gebhard C, Andreesen R, Rehli M. 5-Hydroxymethylcytosine is an essential intermediate of active DNA demethylation processes in primary human monocytes. *Genome biology*. 2013;14(5):R46.
55. Schmidl C, Klug M, Boeld TJ, et al. Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Research*. 2009;19(7):1165-1174.
56. Rasmussen KD, Helin K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev.* 2016;30(7):733-750.
57. Wu X, Zhang Y. TET-mediated active DNA demethylation: Mechanism, function and beyond. *Nat. Rev. Genet.* 2017;18(9):517-534.
58. Tan L, Shi YG. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development*. 2012;139(11):1895-1902.
59. Gehring M, Reik W, Henikoff S. DNA demethylation by DNA repair. *Trends in Genetics*. 2009;25(2):82-90.
60. Ito S, Shen L, Dai Q, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (New York, N.Y.)*. 2011;333(6047):1300-1303.

61. Tahiliani M, Koh KP, Shen Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, N.Y.)*. 2009;324(5929):930-935.
62. Kunimoto H, Nakajima H. Epigenetic dysregulation of hematopoietic stem cells and preleukemic state. *International journal of hematology*. 2017;106(1):34-44.
63. Esteller M. ABERRANT DNA METHYLATION AS A CANCER-INDUCING MECHANISM. *Annu. Rev. Pharmacol. Toxicol.* 2005;45(1):629-656.
64. Lewis JD, Meehan RR, Henzel WJ, et al. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to Methylated DNA. *Cell*. 1992;69(6):905-914.
65. Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell*. 1989;58(3):499-507.
66. Gigek CO, Chen ES, Smith MAC. Methyl-CpG-Binding Protein (MBD) Family: Epigenomic Read-Outs Functions and Roles in Tumorigenesis and Psychiatric Diseases. *Journal of cellular biochemistry*. 2016;117(1):29-38.
67. Feng Q, Zhang Y. The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. *Genes Dev.* 2001;15(7):827-832.
68. Ballestar E, Wolffe AP. Methyl-CpG-binding proteins. Targeting specific gene repression. *European journal of biochemistry*. 2001;268(1):1-6.
69. Fournier A, Sasai N, Nakao M, Defossez P-A. The role of methyl-binding proteins in chromatin organization and epigenome maintenance. *Briefings in functional genomics*. 2012;11(3):251-264.
70. Wood KH, Zhou Z. Emerging Molecular and Biological Functions of MBD2, a Reader of DNA Methylation. *Frontiers in genetics*. 2016;7:93.
71. Hendrich B, Tweedie S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in Genetics*. 2003;19(5):269-277.
72. Menafra R, Brinkman AB, Matarese F, et al. Genome-wide binding of MBD2 reveals strong preference for highly methylated loci. *PloS one*. 2014;9(6):e99603.
73. Du Q, Luu P-L, Stirzaker C, Clark SJ. Methyl-CpG-binding domain proteins: Readers of the epigenome. *Epigenomics*. 2015;7(6):1051-1073.
74. Laget S, Joulie M, Le Masson F, et al. The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PloS one*. 2010;5(8):e11982.
75. Sasai N, Nakao M, Defossez P-A. Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Research*. 2010;38(15):5015-5022.
76. Clouaire T, Stancheva I. Methyl-CpG binding proteins: Specialized transcriptional repressors or structural components of chromatin? *Cellular and molecular life sciences : CMLS*. 2008;65(10):1509-1522.
77. Buck-Koehntop BA, Defossez P-A. On how mammalian transcription factors recognize methylated DNA. *Epigenetics*. 2013;8(2):131-137.
78. Geisler SJ, Paro R. Trithorax and Polycomb group-dependent regulation: A tale of opposing activities. *Development (Cambridge, England)*. 2015;142(17):2876-2887.
79. Schuettengruber B, Bourbon H-M, Di Croce L, Cavalli G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell*. 2017;171(1):34-57.

80. Pirrotta V. Polycombing the Genome: PcG, trxG, and Chromatin Silencing. *Cell*. 1998;93(3):333-336.
81. Di Croce L, Helin K. Transcriptional regulation by Polycomb group proteins. *Nature structural & molecular biology*. 2013;20(10):1147-1155.
82. Ringrose L. Polycomb comes of age: Genome-wide profiling of target sites. *Current opinion in cell biology*. 2007;19(3):290-297.
83. Cao R, Zhang Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Molecular cell*. 2004;15(1):57-67.
84. Shao Z, Raible F, Mollaaghababa R, et al. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell*. 1999;98(1):37-46.
85. Kingston RE, Tamkun JW. Transcriptional regulation by trithorax-group proteins. *Cold Spring Harbor perspectives in biology*. 2014;6(10):a019349.
86. Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. Genome regulation by polycomb and trithorax proteins. *Cell*. 2007;128(4):735-745.
87. Viré E, Brenner C, Deplus R, et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature*. 2006;439(7078):871-874.
88. Jin B, Li Y, Robertson KD. DNA methylation: Superior or subordinate in the epigenetic hierarchy? *Genes & cancer*. 2011;2(6):607-617.
89. Mohn F, Weber M, Rebhan M, et al. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular cell*. 2008;30(6):755-766.
90. Schlesinger Y, Straussman R, Keshet I, et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet*. 2006;39(2):232-236.
91. Clouaire T, Webb S, Bird A. Cfp1 is required for gene expression-dependent H3K4 trimethylation and H3K9 acetylation in embryonic stem cells. *Genome biology*. 2014;15(9):451.
92. Deaton AM, Bird A. CpG islands and the regulation of transcription. *Genes Dev.* 2011;25(10):1010-1022.
93. Thomson JP, Skene PJ, Selfridge J, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature*. 2010;464(7291):1082-1086.
94. Kimura H. Histone modifications for human epigenome analysis. *Journal of human genetics*. 2013;58(7):439-445.
95. Teissandier A, Bourc'his D. Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. *The EMBO journal*. 2017;36(11):1471-1473.
96. Pu M, Ni Z, Wang M, et al. Trimethylation of Lys36 on H3 restricts gene expression change during aging and impacts life span. *Genes Dev.* 2015;29(7):718-731.
97. Hahn MA, Wu X, Li AX, Hahn T, Pfeifer GP. Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. *PloS one*. 2011;6(4):e18844.
98. Neri F, Rapelli S, Krepelova A, et al. Intragenic DNA methylation prevents spurious transcription initiation. *Nature*. 2017;543(7643):72-77.
99. Kanwal R, Gupta K, Gupta S. Cancer epigenetics: an introduction. *Methods in molecular biology (Clifton, N.J.)*. 2015;1238:3-25.

100. Cao J. The functional role of long non-coding RNAs and epigenetics. *Biological procedures online*. 2014;16:11.
101. Fellmann C, Lowe SW. Stable RNA interference rules for silencing. *Nature cell biology*. 2014;16(1):10-18.
102. Shabalina SA, Koonin EV. Origins and evolution of eukaryotic RNA interference. *Trends in ecology & evolution*. 2008;23(10):578-587.
103. Prensner JR, Chinnaiyan AM. The emergence of lncRNAs in cancer biology. *Cancer discovery*. 2011;1(5):391-407.
104. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Molecular cell*. 2011;43(6):904-914.
105. Peschansky VJ, Wahlestedt C. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. *Epigenetics*. 2014;9(1):3-12.
106. Hajjari M, Salavaty A. HOTAIR: An oncogenic long non-coding RNA in different cancers. *Cancer biology & medicine*. 2015;12(1):1-9.
107. Zhang H, Chen Z, Wang X, Huang Z, He Z, Chen Y. Long non-coding RNA: A new player in cancer. *Journal of hematology & oncology*. 2013;6:37.
108. Montalban-Bravo G, Garcia-Manero G. Myelodysplastic syndromes: 2018 update on diagnosis, risk-stratification and management. *American journal of hematology*. 2018;93(1):129-147.
109. Meers S. The myelodysplastic syndromes: The era of understanding. *European journal of haematology*. 2015;94(5):379-390.
110. Greenberg PL. The Multifaceted Nature of Myelodysplastic Syndromes: Clinical, Molecular, and Biological Prognostic Features. *J Natl Compr Canc Netw*. 2013;11(7):877-885.
111. da Silva-Coelho P, Kroeze LI, Yoshida K, et al. Clonal evolution in myelodysplastic syndromes. *Nature communications*. 2017;8:15099.
112. Vasekova P, Plank L. The Differencies in Adult and Pediatric Myelodysplastic Syndrome: A Review. *Acta Medica Martiniana*. 2016;16(2):172.
113. Zini G. Diagnostics and Prognostication of Myelodysplastic Syndromes. *Annals of laboratory medicine*. 2017;37(6):465-474.
114. Shastri A, Will B, Steidl U, Verma A. Stem and progenitor cell alterations in myelodysplastic syndromes. *Blood*. 2017;129(12):1586-1594.
115. Kennedy JA, Ebert BL. Clinical Implications of Genetic Mutations in Myelodysplastic Syndrome. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2017;35(9):968-974.
116. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of Genetic Lesions in 944 Patients with Myelodysplastic Syndromes. *Leukemia*. 2013.
117. Zahid MF, Malik UA, Sohail M, Hassan IN, Ali S, Shaukat MHS. Cytogenetic Abnormalities in Myelodysplastic Syndromes: An Overview. *International journal of hematology-oncology and stem cell research*. 2017;11(3):231-239.
118. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478(7367):64-69.

119. Papaemmanuil E, Cazzola M, Boultwood J, et al. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *The New England journal of medicine*. 2011;365(15):1384-1395.
120. Malcovati L, Papaemmanuil E, Bowen DT, et al. Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative neoplasms. *Blood*. 2011;118(24):6239-6246.
121. Damm F, Thol F, Kosmider O, et al. SF3B1 mutations in myelodysplastic syndromes: Clinical associations and prognostic implications. *Leukemia*. 2012;26(5):1137-1140.
122. Patnaik MM, Lasho TL, Hodnefield JM, et al. SF3B1 mutations are prevalent in myelodysplastic syndromes with ring sideroblasts but do not hold independent prognostic value. *Blood*. 2012;119(2):569-572.
123. Bejar R, Stevenson KE, Caughey BA, et al. Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2012;30(27):3376-3382.
124. Bravo GM, Lee E, Merchan B, Kantarjian HM, García-Manero G. Integrating genetics and epigenetics in myelodysplastic syndromes: Advances in pathogenesis and disease evolution. *Br. J. Haematol.* 2014;166(5):646-659.
125. Bejar R, Steensma DP. Recent developments in myelodysplastic syndromes. *Blood*. 2014;124(18):2793-2803.
126. Joshi P, Halene S, Abdel-Wahab O. How do messenger RNA splicing alterations drive myelodysplasia? *Blood*. 2017;129(18):2465-2470.
127. Saez B, Walter MJ, Graubert TA. Splicing factor gene mutations in hematologic malignancies. *Blood*. 2017;129(10):1260-1269.
128. Murati A, Brecqueville M, Devillier R, Mozziconacci M-J, Gelsi-Boyer V, Birnbaum D. Myeloid malignancies: Mutations, models and management. *BMC cancer*. 2012;12:304.
129. Lindsley RC, Ebert BL. Molecular pathophysiology of myelodysplastic syndromes. *Annual review of pathology*. 2013;8:21-47.
130. Shahrabi S, Khosravi A, Shahjahani M, Rahim F, Saki N. Genetics and Epigenetics of Myelodysplastic Syndromes and Response to Drug Therapy: New Insights. *Oncology reviews*. 2016;10(2):311.
131. Walter MJ, Ding L, Shen D, et al. Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. *Leukemia*. 2011;25(7):1153-1158.
132. Sperling AS, Gibson CJ, Ebert BL. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nature reviews. Cancer*. 2016.
133. Li Z, Cai X, Cai C-L, et al. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood*. 2011;118(17):4509-4518.
134. Moran-Crusio K, Reavie L, Shih A, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell*. 2011;20(1):11-24.
135. Inoue S, Lemonnier F, Mak TW. Roles of IDH1/2 and TET2 mutations in myeloid disorders. *International journal of hematology*. 2016;103(6):627-633.
136. Thol F, Friesen I, Damm F, et al. Prognostic Significance of ASXL1 Mutations in Patients With Myelodysplastic Syndromes. *Journal of Clinical Oncology*. 2011;29(18):2499-2506.

137. Rinke J, Müller JP, Blaess MF, et al. Molecular characterization of EZH2 mutant patients with myelodysplastic/myeloproliferative neoplasms. *Leukemia*. 2017;31(9):1936-1943.
138. Sood R, Kamikubo Y, Liu P. Role of RUNX1 in hematological malignancies. *Blood*. 2017;129(15):2070-2082.
139. Chen C-Y, Lin L-I, Tang J-L, et al. RUNX1 gene mutation in primary myelodysplastic syndrome--the mutation can be detected early at diagnosis or acquired during disease progression and is associated with poor outcome. *Br. J. Haematol.* 2007;139(3):405-414.
140. Wlodarski MW, Collin M, Horwitz MS. GATA2 deficiency and related myeloid neoplasms. *Seminars in hematology*. 2017;54(2):81-86.
141. Wang X, Muramatsu H, Okuno Y, et al. GATA2 and secondary mutations in familial myelodysplastic syndromes and pediatric myeloid malignancies. *Haematologica*. 2015;100(10):401.
142. Spinner MA, Sanchez LA, Hsu AP, et al. GATA2 deficiency: A protean disorder of hematopoiesis, lymphatics, and immunity. *Blood*. 2014;123(6):809-821.
143. Pellagatti A, Boultwood J. The molecular pathogenesis of the myelodysplastic syndromes. *European journal of haematology*. 2015;95(1):3-15.
144. Krönke J, Fink EC, Hollenbach PW, et al. Lenalidomide induces ubiquitination and degradation of CK1α in del(5q) MDS. *Nature*. 2015;523(7559):183-188.
145. Schneider RK, Ademà V, Heckl D, et al. Role of casein kinase 1A1 in the biology and targeted therapy of del(5q) MDS. *Cancer Cell*. 2014;26(4):509-520.
146. Fenaux P, Giagounidis A, Selleslag D, et al. A randomized phase 3 study of lenalidomide versus placebo in RBC transfusion-dependent patients with Low-/Intermediate-1-risk myelodysplastic syndromes with del5q. *Blood*. 2011;118(14):3765-3776.
147. Gaballa MR, Besa EC. Myelodysplastic syndromes with 5q deletion: Pathophysiology and role of lenalidomide. *Annals of hematology*. 2014;93(5):723-733.
148. Wei S, Chen X, McGraw K, et al. Lenalidomide promotes p53 degradation by inhibiting MDM2 auto-ubiquitination in myelodysplastic syndrome with chromosome 5q deletion. *Oncogene*. 2013;32(9):1110-1120.
149. Pezeshki A, Podder S, Kamel R, Corey SJ. Monosomy 7/del (7q) in inherited bone marrow failure syndromes: A systematic review. *Pediatric blood & cancer*. 2017;64(12).
150. Chen C, Liu Y, Rappaport AR, et al. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell*. 2014;25(5):652-665.
151. McNerney ME, Brown CD, Wang X, et al. CUX1 is a haploinsufficient tumor suppressor gene on chromosome 7 frequently inactivated in acute myeloid leukemia. *Blood*. 2013;121(6):975-983.
152. Nikoloski G, Langemeijer SMC, Kuiper RP, et al. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nature genetics*. 2010;42(8):665-667.
153. Figueroa ME, Lugthart S, Li Y, et al. DNA Methylation Signatures Identify Biologically Distinct Subtypes in Acute Myeloid Leukemia. *Cancer Cell*. 2010;17(1):13-27.
154. Figueroa ME, Skrabanek L, Li Y, et al. MDS and secondary AML display unique patterns and abundance of aberrant DNA methylation. *Blood*. 2009;114(16):3448-3458.

155. Shen L, Kantarjian H, Guo Y, et al. DNA methylation predicts survival and response to therapy in patients with myelodysplastic syndromes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2010;28(4):605-613.
156. Jiang Y, Dunbar A, Gondek LP, et al. Aberrant DNA methylation is a dominant mechanism in MDS progression to AML. *Blood*. 2009;113(6):1315-1325.
157. Dexheimer GM, Alves J, Reckziegel L, Lazzaretti G, Abujamra AL. DNA Methylation Events as Markers for Diagnosis and Management of Acute Myeloid Leukemia and Myelodysplastic Syndrome. *Disease markers*. 2017;2017:5472893.
158. Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. *Nat Rev Genet*. 2002;3(6):415-428.
159. Fenaux P, Mufti GJ, Hellstrom-Lindberg E, et al. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: A randomised, open-label, phase III study. *The Lancet Oncology*. 2009;10(3):223-232.
160. Kantarjian H, Issa J-PJ, Rosenfeld CS, et al. Decitabine improves patient outcomes in myelodysplastic syndromes: Results of a phase III randomized study. *Cancer*. 2006;106(8):1794-1803.
161. Diesch J, Zwick A, Garz A-K, Palau A, Buschbeck M, Götze KS. A clinical-molecular update on azanucleoside-based therapy for the treatment of hematologic cancers. *Clinical epigenetics*. 2016;8:71.
162. Sánchez-Abarca LI, Gutierrez-Cosio S, Santamaría C, et al. Immunomodulatory effect of 5-azacytidine (5-azaC): Potential role in the transplantation setting. *Blood*. 2010;115(1):107-121.
163. Goodyear OC, Dennis M, Jilani NY, et al. Azacitidine augments expansion of regulatory T cells after allogeneic stem cell transplantation in patients with acute myeloid leukemia (AML). *Blood*. 2012;119(14):3361-3369.
164. Glaubach T, Robinson LJ, Corey SJ. Pediatric myelodysplastic syndromes: They do exist! *Journal of pediatric hematology/oncology*. 2014;36(1):1-7.
165. Narumi S, Amano N, Ishii T, et al. SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. *Nature genetics*. 2016;48(7):792-797.
166. Schwartz JR, Ma J, Lamprecht T, et al. The genomic landscape of pediatric myelodysplastic syndromes. *Nature communications*. 2017;8(1):1557.
167. Schanz J, Cevik N, Fonatsch C, et al. Detailed analysis of clonal evolution and cytogenetic evolution patterns in patients with myelodysplastic syndromes (MDS) and related myeloid disorders. *Blood cancer journal*. 2018;8(3):28.
168. Heuser M, Thol F, Ganser A. Clonal Hematopoiesis of Indeterminate Potential. *Deutsches Arzteblatt international*. 2016;113(18):317-322.
169. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015;126(1):9-16.
170. Mossner M, Jann J-C, Wittig J, et al. Mutational hierarchies in myelodysplastic syndromes dynamically adapt and evolve upon therapy response and failure. *Blood*. 2016.
171. Walter MJ, Shen D, Ding L, et al. Clonal architecture of secondary acute myeloid leukemia. *The New England journal of medicine*. 2012;366(12):1090-1098.

172. Li S, Garrett-Bakelman FE, Chung SS, et al. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nature medicine*. 2016;22(7):792-799.
173. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology*. 1986;1986.
174. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PloS one*. 2010;5(1):e8888.
175. *methylKit:* Bioconductor; 2017.
176. Jühling F, Kretzmer H, Bernhart SH, Otto C, Stadler PF, Hoffmann S. metilene: Fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. *Genome Research*. 2016;26(2):256-262.
177. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell*. 2010;38(4):576-589.
178. Schoofs T, Müller-Tidow C. DNA methylation as a pathogenic event and as a therapeutic target in AML. *Cancer Treatment Reviews*. 2011;37:S13-S18.
179. Lonsdale J, Thomas J. The Genotype-Tissue Expression (GTEx) project. *Nature genetics*. 2013;45(6):580-585.
180. Tripathi S, Pohl MO, Zhou Y, et al. Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding. *Cell host & microbe*. 2015;18(6):723-735.
181. Hinoue T, Weisenberger DJ, Lange CPE, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Research*. 2012;22(2):271-282.
182. Rodriguez J, Muñoz M, Vives L, Frangou CG, Groudine M, Peinado MA. Bivalent domains enforce transcriptional memory of DNA methylated genes in cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* 2008;105(50):19809-19814.
183. Easwaran H, Johnstone SE, van Neste L, et al. A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Research*. 2012;22(5):837-849.
184. Mori N, Ohwashi-Miyazaki M, Yoshinaga K, et al. Tumor suppressor gene methylation on the short arm of chromosome 1 in chronic myelogenous leukemia. *European journal of haematology*. 2017;98(5):467-477.
185. Sasaki O, Meguro K, Tohmiya Y, Funato T, Shibahara S, Sasaki T. Altered expression of retinoblastoma protein-interacting zinc finger gene, RIZ, in human leukaemia. *Br J Haematol*. 2002;119(4):940-948.
186. Mori N, Yoshinaga K, Tomita K, et al. Aberrant methylation of the RIZ1 gene in myelodysplastic syndrome and acute myeloid leukemia. *Leuk. Res.* 2011;35(4):516-521.
187. Nishikawa N, Toyota M, Suzuki H, et al. Gene amplification and overexpression of PRDM14 in breast cancers. *Cancer research*. 2007;67(20):9649-9657.
188. Pikor LA, Lockwood WW, Thu KL, et al. YEATS4 is a novel oncogene amplified in non-small cell lung cancer that regulates the p53 pathway. *Cancer research*. 2013;73(24):7301-7312.
189. Tao K, Yang J, Hu Y, Deng A. Knockdown of YEATS4 inhibits colorectal cancer cell proliferation and induces apoptosis. *American journal of translational research*. 2015;7(3):616-623.

190. Goyama S, Huang G, Kurokawa M, Mulloy JC. Posttranslational modifications of RUNX1 as potential anticancer targets. *Oncogene*. 2015;34(27):3483-3492.
191. Farhan M, Wang H, Gaur U, Little PJ, Xu J, Zheng W. FOXO Signaling Pathways as Therapeutic Targets in Cancer. *International journal of biological sciences*. 2017;13(7):815-827.
192. Kim CG, Lee H, Gupta N, et al. Role of Forkhead Box Class O proteins in cancer progression and metastasis. *Seminars in cancer biology*. 2017.
193. Gregory GD, Miccio A, Bersenev A, et al. FOG1 requires NuRD to promote hematopoiesis and maintain lineage fidelity within the megakaryocytic-erythroid compartment. *Blood*. 2010;115(11):2156-2166.
194. Fujiwara T, Sasaki K, Saito K, et al. Forced FOG1 expression in erythroleukemia cells: Induction of erythroid genes and repression of myelo-lymphoid transcription factor PU.1. *Biochemical and biophysical research communications*. 2017;485(2):380-387.
195. Klug M, Rehli M. Functional analysis of promoter CpG methylation using a CpG-free luciferase reporter vector. *Epigenetics*. 2006;1(3):127-130.
196. Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*. 2016;374(2065):20150202.
197. Lever J, Krzywinski M, Altman N. Points of Significance: Principal component analysis. *Nat Meth*. 2017;14(7):641-642.
198. Akalin A, Kormaksson M, Li S, et al. methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome biology*. 2012;13(10):R87.
199. Farlik M, Halbritter F, Müller F, et al. DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation. *Cell stem cell*. 2016;19(6):808-822.
200. Raimbault A, Pierre-Eugene C, Rouquette A, et al. APG101 efficiently rescues erythropoiesis in lower risk myelodysplastic syndromes with severe impairment of hematopoiesis. *Oncotarget*. 2016;7(12):14898-14911.
201. Agatheeswaran S, Chakraborty S. MEF2C and CEBPA: Possible co-regulators in chronic myeloid leukemia disease progression. *The international journal of biochemistry & cell biology*. 2016;77(Pt A):165-170.
202. Takagi Y, Shimada K, Shimada S, et al. SPIB is a novel prognostic factor in diffuse large B-cell lymphoma that mediates apoptosis via the PI3K-AKT pathway. *Cancer science*. 2016;107(9):1270-1280.
203. Kurachi M, Barnitz RA, Yosef N, et al. The transcription factor BATF operates as an essential differentiation checkpoint in early effector CD8+ T cells. *Nature immunology*. 2014;15(4):373-383.
204. Grötsch B, Brachs S, Lang C, et al. The AP-1 transcription factor Fra1 inhibits follicular B cell differentiation into plasma cells. *The Journal of experimental medicine*. 2014;211(11):2199-2212.
205. Wlodarski MW, Hirabayashi S, Pastor V, et al. Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. *Blood*. 2016;127(11):1387-1397.

206. Micol J-B, Abdel-Wahab O. Collaborating constitutive and somatic genetic events in myeloid malignancies: ASXL1 mutations in patients with germline GATA2 mutations. *Haematologica*. 2014;99(2):201-203.
207. Deaton AM, Webb S, Kerr ARW, et al. Cell type-specific DNA methylation at intragenic CpG islands in the immune system. *Genome Research*. 2011;21(7):1074-1086.
208. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012;9(4):357-359.
209. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*. 1996;5(3):299.
210. Inoue D, Kitaura J, Matsui H, et al. SETBP1 mutations drive leukemic transformation in ASXL1-mutated MDS. *Leukemia*. 2015;29(4):847-857.
211. Satow R, Nakamura T, Kato C, et al. ZIC5 Drives Melanoma Aggressiveness by PDGFD-Mediated Activation of FAK and STAT3. *Cancer research*. 2017;77(2):366-377.
212. Satow R, Inagaki S, Kato C, Shimozawa M, Fukami K. Identification of zinc finger protein of the cerebellum 5 as a survival factor of prostate and colorectal cancer cells. *Cancer science*. 2017;108(12):2405-2412.
213. Sun Q, Shi R, Wang X, Li D, Wu H, Ren B. Overexpression of ZIC5 promotes proliferation in non-small cell lung cancer. *Biochemical and biophysical research communications*. 2016;479(3):502-509.
214. Senchenko VN, Kisseljova NP, Ivanova TA, et al. Novel tumor suppressor candidates on chromosome 3 revealed by NotI-microarrays in cervical cancer. *epigenetics*. 2013;8(4):409-420.
215. Zhang J, Zhang C, Cui J, et al. TRIM45 functions as a tumor suppressor in the brain via its E3 ligase activity by stabilizing p53 through K63-linked ubiquitination. *Cell death & disease*. 2017;8(5):e2831.
216. Shibata M, Sato T, Nukiwa R, Ariga T, Hatakeyama S. TRIM45 negatively regulates NF-κB-mediated transcription and suppresses cell proliferation. *Biochemical and biophysical research communications*. 2012;423(1):104-109.
217. Tobiasson M, Abdulkadir H, Lennartsson A, et al. Comprehensive mapping of the effects of azacitidine on DNA methylation, repressive/permissive histone marks and gene expression in primary cells from patients with MDS and MDS-related disease. *Oncotarget*. 2017;8(17):28812-28825.
218. Gentner E, Vegi NM, Mulaw MA, et al. VENTX induces expansion of primitive erythroid cells and contributes to the development of acute myeloid leukemia in mice. *Oncotarget*. 2016;7(52):86889-86901.
219. Tahara T, Arisawa T. DNA methylation as a molecular biomarker in gastric cancer. *Epigenomics*. 2015;7(3):475-486.
220. Woo HD, Fernandez-Jimenez N, Ghantous A, et al. Genome-wide profiling of normal gastric mucosa identifies Helicobacter pylori- and cancer-associated DNA methylome changes. *International journal of cancer*. 2018.
221. Khan H, Vale C, Bhagat T, Verma A. Role of DNA methylation in the pathogenesis and treatment of myelodysplastic syndromes. *Seminars in hematology*. 2013;50(1):16-37.

222. Li Y, Zhao H, Xu Q, et al. Detection of prognostic methylation markers by methylC-capture sequencing in acute myeloid leukemia. *Oncotarget*. 2017;8(66):110444-110459.
223. Zhou J-D, Lin J, Zhang T-J, et al. GPX3 methylation in bone marrow predicts adverse prognosis and leukemia transformation in myelodysplastic syndrome. *Cancer medicine*. 2017;6(1):267-274.
224. Zhang T-J, Zhou J-D, Yang D-Q, et al. Hypermethylation of DLX4 predicts poor clinical outcome in patients with myelodysplastic syndrome. *Clinical chemistry and laboratory medicine*. 2016;54(5):865-871.
225. Fu HY, Zhou HR, Yan JG, Chen CJ, Shen JZ. Clinical significance of hypermethylation of DLC-1 gene in myelodysplastic syndrome patients and effects of decitabine on DLC-1 gene expression. *Zhonghua yi xue za zhi*. 2017;97(6):412-417.
226. Aggerholm A, Holm MS, Guldberg P, Olesen LH, Hokland P. Promoter hypermethylation of p15INK4B, HIC1, CDH1, and ER is frequent in myelodysplastic syndrome and predicts poor prognosis in early-stage patients. *Eur J Haematol*. 2006;76(1):23-32.
227. Will B, Zhou L, Vogler TO, et al. Stem and progenitor cells in myelodysplastic syndromes show aberrant stage-specific expansion and harbor genetic and epigenetic alterations. *Blood*. 2012;120(10):2076-2086.
228. Lee TI, Jenner RG, Boyer LA, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell*. 2006;125(2):301-313.
229. Wong Y-F, Micklem CN, Taguchi M, et al. Longitudinal Analysis of DNA Methylation in CD34+ Hematopoietic Progenitors in Myelodysplastic Syndrome. *Stem cells translational medicine*. 2014;3(10):1188-1198.
230. Song I-S, Ha G-H, Kim J-M, et al. Human ZNF312b oncogene is regulated by Sp1 binding to its promoter region through DNA demethylation and histone acetylation in gastric cancer. *International journal of cancer*. 2011;129(9):2124-2133.
231. Watt PM, Kumar R, Kees UR. Promoter demethylation accompanies reactivation of the HOX11 proto-oncogene in leukemia. *Genes, chromosomes & cancer*. 2000;29(4):371-377.
232. Shao C, Sun W, Tan M, et al. Integrated, genome-wide screening for hypomethylated oncogenes in salivary gland adenoid cystic carcinoma. *Clinical Cancer Research*. 2011;17(13):4320-4330.
233. Gaidzik VI, Teleanu V, Papaemmanuil E, et al. RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features. *Leukemia*. 2016;30(11):2160-2168.
234. Haferlach T, Stengel A, Eckstein S, et al. The new provisional WHO entity 'RUNX1 mutated AML' shows specific genetics but no prognostic influence of dysplasia. *Leukemia*. 2016;30(10):2109-2112.
235. Ticchioni M, Essafi M, Jeandel PY, et al. Homeostatic chemokines increase survival of B-chronic lymphocytic leukemia cells through inactivation of transcription factor FOXO3a. *Oncogene*. 2007;26(50):7081-7091.
236. Walter MJ, Shen D, Shao J, et al. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia*. 2013;27(6):1275-1282.
237. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013.

238. Uy GL, Duncavage EJ, Chang GS, et al. Dynamic changes in the clonal structure of MDS and AML in response to epigenetic therapy. *Leukemia*. 2017;31(4):872-881.
239. Kotla V, Goel S, Nischal S, et al. Mechanism of action of lenalidomide in hematological malignancies. *Journal of hematology & oncology*. 2009;2:36.
240. Langstein J, Milsom MD, Lipka DB. Impact of DNA methylation programming on normal and pre-leukemic hematopoiesis. *Seminars in cancer biology*. 2017.
241. Sorm F, Píškala A, Cihák A, Veselý J. 5-Azacytidine, a new, highly effective cancerostatic. *Experientia*. 1964;20(4):202-203.
242. Lindblad KE, Goswami M, Hourigan CS, Oetjen KA. Immunological effects of hypomethylating agents. *Expert review of hematology*. 2017;10(8):745-752.
243. Daskalakis M, Brocks D, Sheng Y-H, et al. Reactivation of endogenous retroviral elements via treatment with DNMT- and HDAC-inhibitors. *Cell cycle (Georgetown, Tex.)*. 2018:1-12.
244. Xu H-X, Wu K-J, Tian Y-J, et al. Expression profile of SIX family members correlates with clinic-pathological features and prognosis of breast cancer: A systematic review and meta-analysis. *Medicine*. 2016;95(27):e4085.
245. Liu Q, Li A, Tian Y, et al. The expression profile and clinic significance of the SIX family in non-small cell lung cancer. *Journal of hematology & oncology*. 2016;9(1):119.
246. Wang L, Xiao X, Li D, et al. Abnormal expression of GADD45B in human colorectal carcinoma. *Journal of translational medicine*. 2012;10:215.
247. Baltrunaite K, Craig MP, Palencia Desai S, et al. ETS transcription factors Etv2 and Fli1b are required for tumor angiogenesis. *Angiogenesis*. 2017;20(3):307-323.
248. Wang J-H, Tang H-S, Li X-S, et al. Elevated FOXO6 expression correlates with progression and prognosis in gastric cancer. *Oncotarget*. 2017;8(19):31682-31691.
249. Zhou L, Opalinska J, Sohal D, et al. Aberrant epigenetic and genetic marks are seen in myelodysplastic leukocytes and reveal Dock4 as a candidate pathogenic gene on chromosome 7q. *The Journal of biological chemistry*. 2011;286(28):25211-25223.
250. Jeong M, Goodell MA. New answers to old questions from genome-wide maps of DNA methylation in hematopoietic cells. *Experimental hematology*. 2014;42(8):609-617.
251. McKerrell T, Park N, Moreno T, et al. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell reports*. 2015;10(8):1239-1245.
252. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *The New England journal of medicine*. 2014;371(26):2488-2498.
253. Malcovati L, Gallì A, Travaglino E, et al. Clinical significance of somatic mutation in unexplained blood cytopenia. *Blood*. 2017;129(25):3371-3378.
254. Malcovati L, Cazzola M. The shadowlands of MDS: Idiopathic cytopenias of undetermined significance (ICUS) and clonal hematopoiesis of indeterminate potential (CHIP). *Hematology. American Society of Hematology. Education Program*. 2015;2015:299-307.
255. Jones PA, Gonzalgo ML. Altered DNA methylation and genome instability: A new pathway to cancer? *Proc. Natl. Acad. Sci. U.S.A.* 1997;94(6):2103-2105.

256. Sheaffer KL, Elliott EN, Kaestner KH. DNA Hypomethylation Contributes to Genomic Instability and Intestinal Cancer Initiation. *Cancer prevention research (Philadelphia, Pa.)*. 2016;9(7):534-546.
257. Di Ruscio A, Welner RS, Tenen DG, Amabile G. The second hit of DNA methylation. *Molecular & cellular oncology*. 2016;3(3):e1093690.

10 Abbreviations

%	Percent
°C	Degree Celsius
µg	Microgram
µl	Microliter
µM	Micromolar
3‘	3-prime
5‘	5-prime
5caC	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
A	Adenine
aa	Amino acids
ac	Acetylation
ACK	ammonium-chloride-potassium
AcOH	Acetic Acid
AGM	aorta-gonad-mesonephros
AML	Acute myeloid leukemia
Amp	Ampicillin
ATP	Adenosine triphosphate
Aza	Azacitidine
BER	Base excision repair
BM	bone marrow
bp	Base pair
BS	Bisulfite sequencing
BSA	Bovine Serum Albumin
BS-Seq	bisulfite-sequencing
C	Cytosine
C/EBP	CCAAT/Enhancer Binding Protein
CAGE	Cap analyses gene expression
CBM	Cell Buffer Mix
CD	cluster of differentiation
cDNA	Complementary DNA
CF	Chromatin factor
CFP1	CXXC zinc finger protein 1
CGI	CpG island
ChIP	Chromatin immunoprecipitation
Chr	Chromosome
CLP	common lymphoid progenitor
cm	Centimeter

cM	Centimorgan
CM	Chloramphenicol
cm^2	Square centimeter
CMP	common myeloid progenitor
CMV	cytomegalie virus
CO_2	Carbon dioxide
CpA	CpA dinucleotide
CpC	CpC dinucleotide
CpG	CpG dinucleotide
CpT	CpT dinucleotide
CTX	Clinical trial exemption
CXXC	Cysteine-rich zinc finger domain
Da	Dalton
dATP	Deoxyadenosine triphosphate
DB	Dilution Buffer
ddH_2O	Double-distilled H2O
ddNTP	Dideoxyribonucleotide triphosphate
DEAE	Diethylaminoethyl
del	deletion
DHS	DNase I hypersensitive site
DMR	Differentially methylated region
DMSO	dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	Deoxyribonucleotide triphosphate
DOB	date of birth
DTT	Dithiothreitol
dUTP	Deoxyuridine triphosphate
Dx	diagnosis
E. coli	Escherichia coli
EDTA/Na_2EDTA	Ethylenediaminotetraacetic acid disodium salt dehydrate
ESC / ES cell	Embryonic stem cell
EtBr	Ethidium bromide
EtOH	Ethanol
ETV6	Ets variant 6
Ezh2	Enhancer of zeste homolog 2
f	female
Fc	fragment crystallizable
FCS	Fetal calf serum
for	forward
G	Guanine
g	Gram
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase
GATA	GATA Binding Protein
GC	Guanine/Cytosine
gDNA	Genomic DNA

GMLP	granulocyte-monocyte-lymphoid progenitor
GMP	granulocyte-monocyte progenitor
GO	gene ontology
GRCh38	Genome Reference Consortium Human Build 38
H	Histone
h	Hour
HAT	Histone acetyltransferase
HCl	Hydrochloric acid
HDAC	Histone deacetylase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HF	high fidelity
hg18	Human genome assembly 18
hg19	Human genome assembly 19
hMeDIP	Hydroxymethylcytosine methylated DNA immunoprecipitation
HMT	Histone methyltransferase
HOAc	acetic acid
Hpa	*Haemophilus parainfluenzae*
HS	high sensitivity
HSC	hematopoietic stem cells
HSCT	hematopoietic stem cell transplantation
Hz	Hertz
IDT	Integrated DNA Technologies
IgG	Immunoglobulin G
IGV	Integrative Genomics Viewer
IP	Immunoprecipitation
iPS	induced pluripotent stem cells
K	Lysine / Kilo
kb	Kilo base
KCl	Potassium chloride
kDa	Kilodalton
KFB	Kompetenzzentrum für fluoreszente Bioanalytik
KOH	Potassium hydroxide
KZF	KRAB-containing zinc finger
l	Liter
LB	Lysogeny broth
LEN	Lenalidomide
LMO2	LIM domain only 2
LM-PCR	ligation-mediated PCR
lncRNA	Long non-coding RNA
LT-HSC	Long-term hematopoietic stem cell
m	mass / male
M	Molar
M.SssI	Methyltransferase Spiroplasma sp. Strain MQ1
mA	Milliampere
MALDI-TOF MS	Matrix-assisted laser desorption/ionization coupled with mass spectrometry analyses by Time-of-flight
Mb	Mega base

MBD	Methyl-binding domain
MBP	Methyl-CpG binding protein family
MCIp	Methyl-CpG immunoprecipitation
mCpG	Methylated CpG
MDR-AML	MDS-derived AML
MDS	myelodysplastic syndromes
MDS-EB	MDS with excess blasts
MDS-MLD	MDS with multilineage dysplasia
MDS-RS	MDS with ring sideroblasts
MDS-SLD	MDS with single lineage dysplasia
MDS-U	unclassifiable MDS
me1	Monomethylation
me3	Trimethylation
MeCP	Methyl-CpG binding protein
MEP	megakaryocyte-erythrocyte progenitor
mg	Milligram
MgCl2	Magnesium chloride
min	Minute
Mio	Million
miRNA	Micro RNA
ml	Milliliter
MLL	megakaryocyte-erythrocyte progenitor
mm	millimeter
mM	Millimolar
M-MLV RT	Moloney Murine Leukemia Virus Reverse Transcriptase
MNC	mononuclear cells
MPC	magnetic particle concentrator
MPP	multipotent progenitor
mRNA	Messenger RNA
msec	Millisecond
NaCl	Sodium chloride
NaHCO3	Sodium hydrogen carbonate
NaOAc	Sodium acetate
ncRNA	Non-coding RNA
ng	Nanogram
NGS	Next generation sequencing
NH_2	Amino
nm	Nanometer
nM	nanomolar
nt	Nucleotide
NuRD	Nucleosome remodeling deacetylase
o/n	Overnight
Oligo	Oligonucleotide
ORF	Open reading frame
OS	Overall survival
PB	peripheral blood
PBS	Phosphate-buffered saline

PCA	Principal component analysis
Pc	Polycomb
PcG	Polycomb-Group
PCR	Polymerase chain reaction
PEG	Polyethylenglycol
PMSF	Phenylmethanesulfonyl fluoride
PRC(1/2)	Polycomb repressor complex (1/2)
qPCR	Quantitative Real-time PCR
RAEB (-t)	refractory anemia with excess blasts (in transformation)
RBC	red blood cell
RC	refractory cytopenia
rev	reverse
RLU	Relative luminometer unit
RNA	Ribonucleic acid
RNAi	RNA-mediated interference
Rnase	Ribonuclease
rpm	Rounds per minute
RPMI	Roswell Park Memorial Institute medium
rRNA	Ribosomal RNA
RT	Room temperature
RT-qPCR	Reverse-transcription quantitative real-time PCR
RUNX1	Runt-related transcription factor 1
SAM	S-adenosylmethionin
SAP	Shrimp alkaline phosphatase
SBS	sequencing by synthesis
SC	SeqCap
SDS	Sodium dodecyl sulfate
sec	Second
SET	Su(var)3-9 and ‘Enhancer of zeste’ domain
SINE	Short interspersed elements
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNA
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
Snrpn	small nuclear ribonucleoprotein N
SOC	super optimal broth
SPRI	solid phase reversible immobilization
ST-HSC	Short-term hematopoietic stem cell
t	translocation
T	Thymine
T_a	Annealing temperature
TAE	Tris/Acetate/EDTA
T-ALL	T-cell acute lymphoblastic leukemia
tbp	tags per base pair
TE	Tris/EDTA
TET	Ten-eleven translocation
TF	Transcription factor

T_m	Melting temperature
TRD	Transcriptional repressor domain
Tris	Tris(hydroxymethyl)-aminomethan
tRNA	Transfer RNA
t-SNE	t-Distributed Stochastic Neighbor Embedding
TSS	transcription start site
TTS	transcription termination site
Tyr	Tyrosinase
U	Unit
ub	Ubiquitinylation
UCSC	University of California, Santa Cruz
UTR	untranslated region
UV	Ultra violet
V	Volt
Vol	Volume
W	Watt
w/o	Without
WBI-III	Wash buffer I-III
WHO	world health organization
WT	Wild type
xg	G-force acceleration

11 Appendix

Appendix I – Gene reporter assays with DMRs obtained in adult MDS patients (section 11.1, page 176)

Figure 11-1 to Figure 11-4 display genome browser snapshots from genomic regions, which were used for gene reporter assays in adult patients suffering from MDS. The precise position of the chosen region is indicated with a red horizontal bar. Targeted bisulfite data sets are shown in the upper part with different colors for distinct samples. Additional ChIP-seq data sets are colored in blue and depicted at the bottom of the figure.

Appendix II – Gene reporter assays in pediatric MDS patient D770 (section 11.2, page 180)

Differentially methylated regions of pediatric MDS patient D770 that were hypomethylated during disease progression were cloned into the reporter gene vector #1341 (pCpGL-CMV/T.E1AF) or #861 (pCpGL-basic). The indicated plasmids were either unmethylated or *in vitro* SssI-methylated ("methylated") and transiently transfected into THP-1 cells. Luciferase activity was normalized against the activity of a co-transfected Renilla construct and mean values of RLUs (relative luminometer units) +/- standard deviation are shown.

Appendix III – DNA methylation analyses and clinical data from longitudinal MDS patients (section 11.3, page 181)

(A) Comparison of DNA methylation patterns between different time points of longitudinal patients and control monocytes depicted as scatterplots. All mappable and detectable regions were depicted in black, while patient specific regions were displayed in red. Correlation values could be found on the bottom right corner of the plots. (B) *K*-means clustering of DNA methylation data in patient specific regions and corresponding epigenetic data in hematopoietic stem cells (HSCs). (C) Overview table of clinical data from consecutive samples of the respective patient with cumulative time between diagnosis and follow up sample, WHO classification, treatment, BM blast count, mutational and cytogenetic status.

11.1 Appendix I – Gene reporter assays with DMRs obtained in adult MDS patients

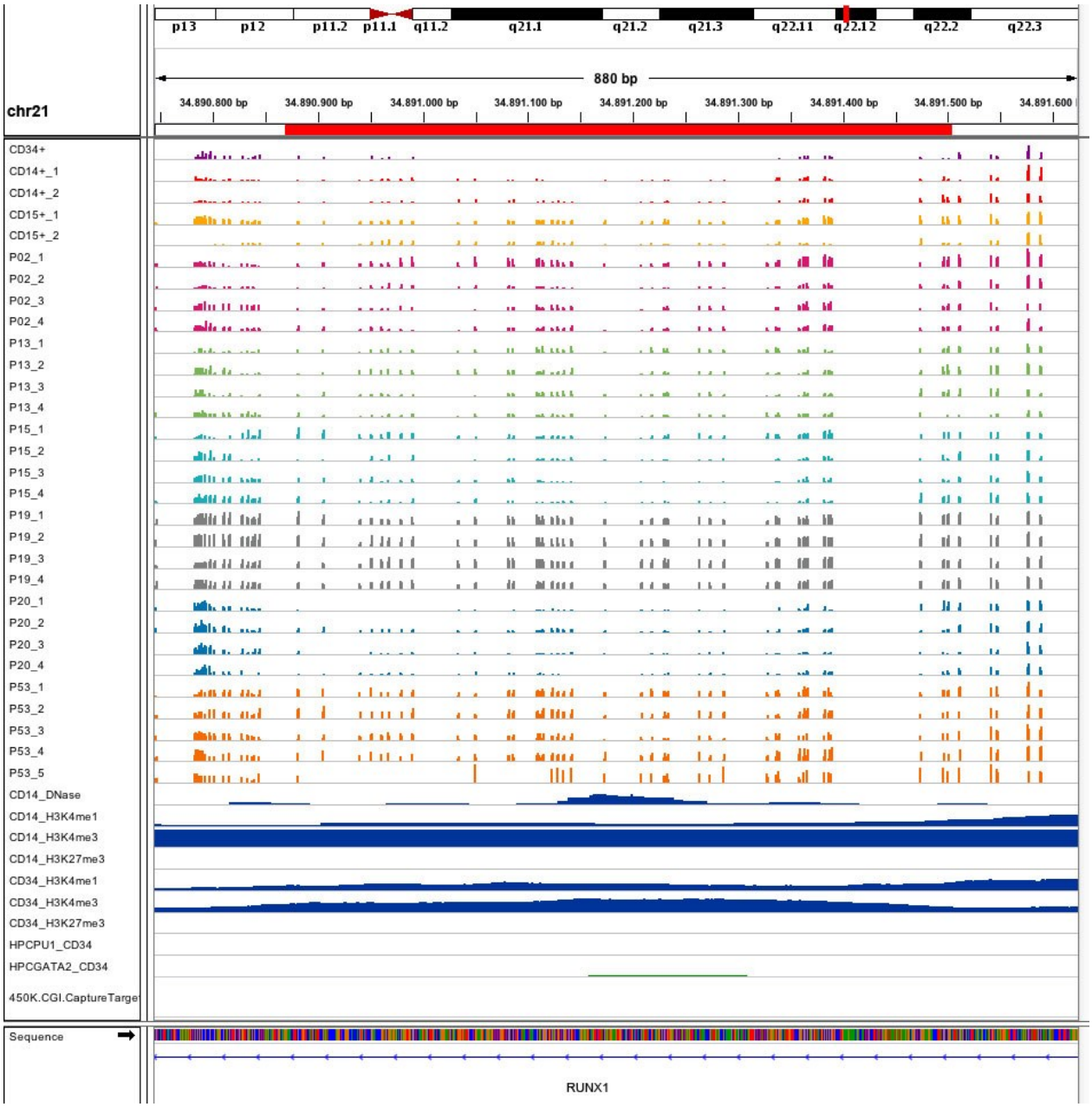


Figure 11-1 - Genome browser track of selected region for RUNX1 gene reporter assay

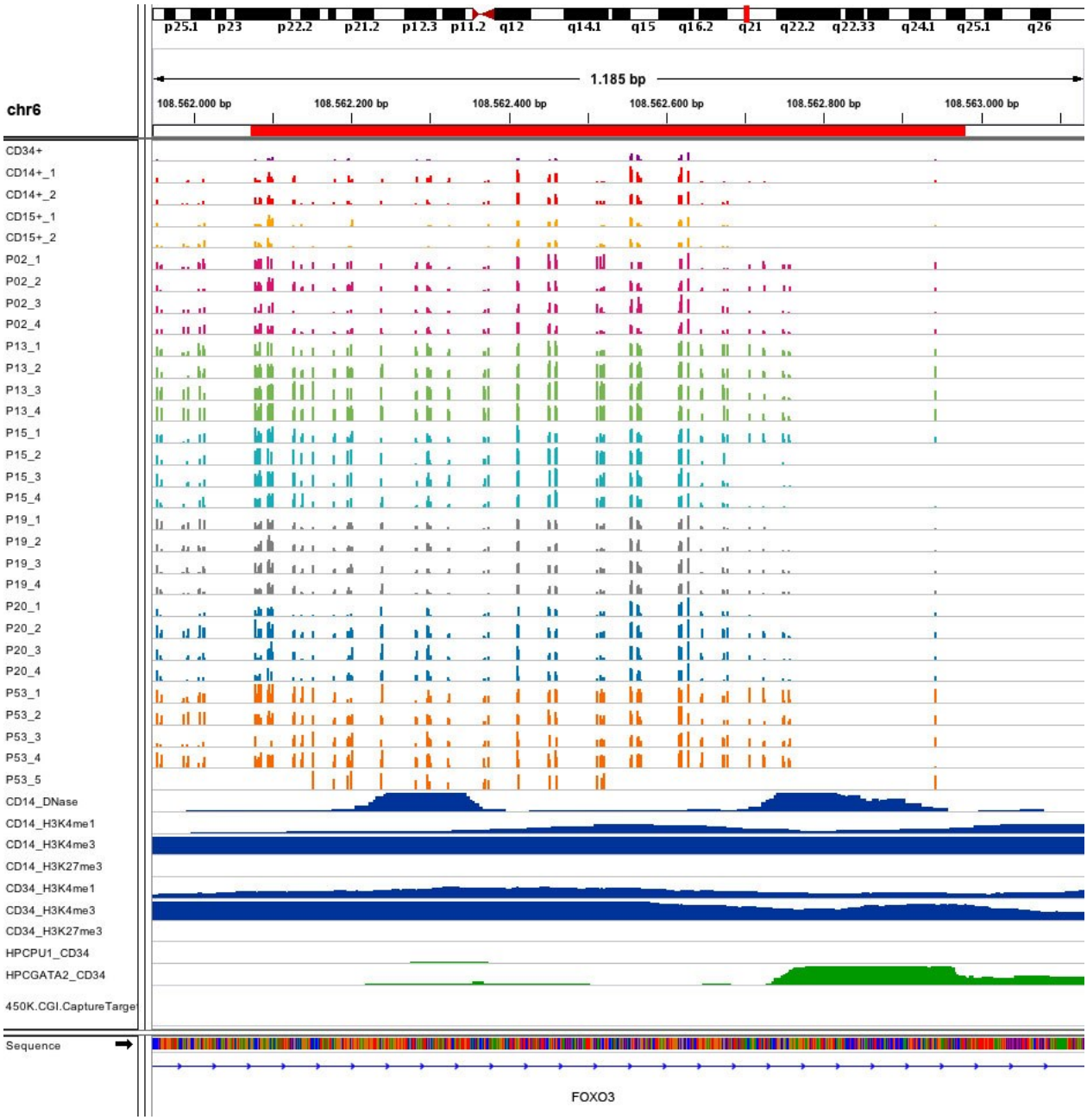


Figure 11-2 - Genome browser track of selected first region for FOXO3 gene reporter assay

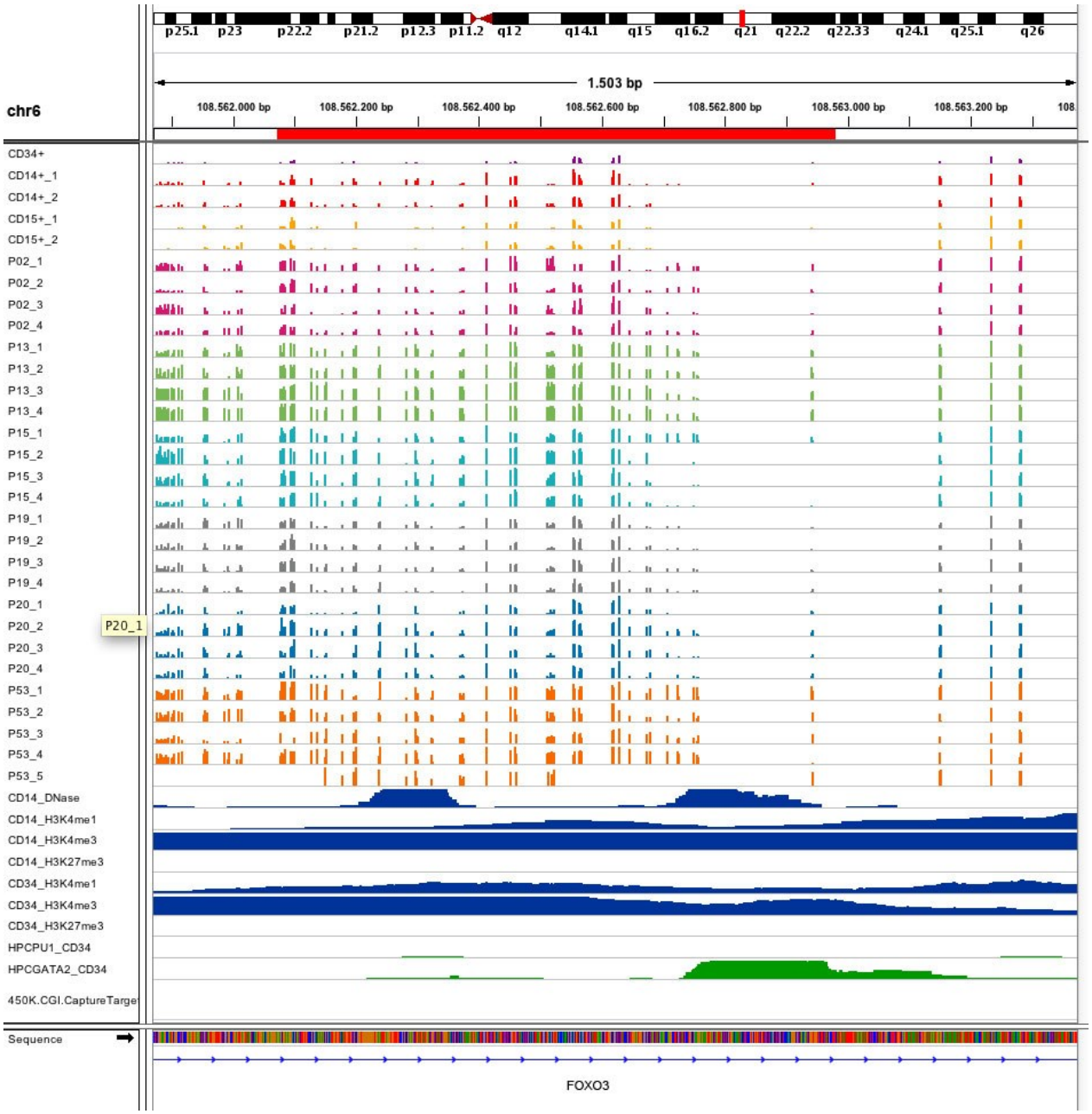


Figure 11-3 - Genome browser track of selected second region for FOXO3 gene reporter assay

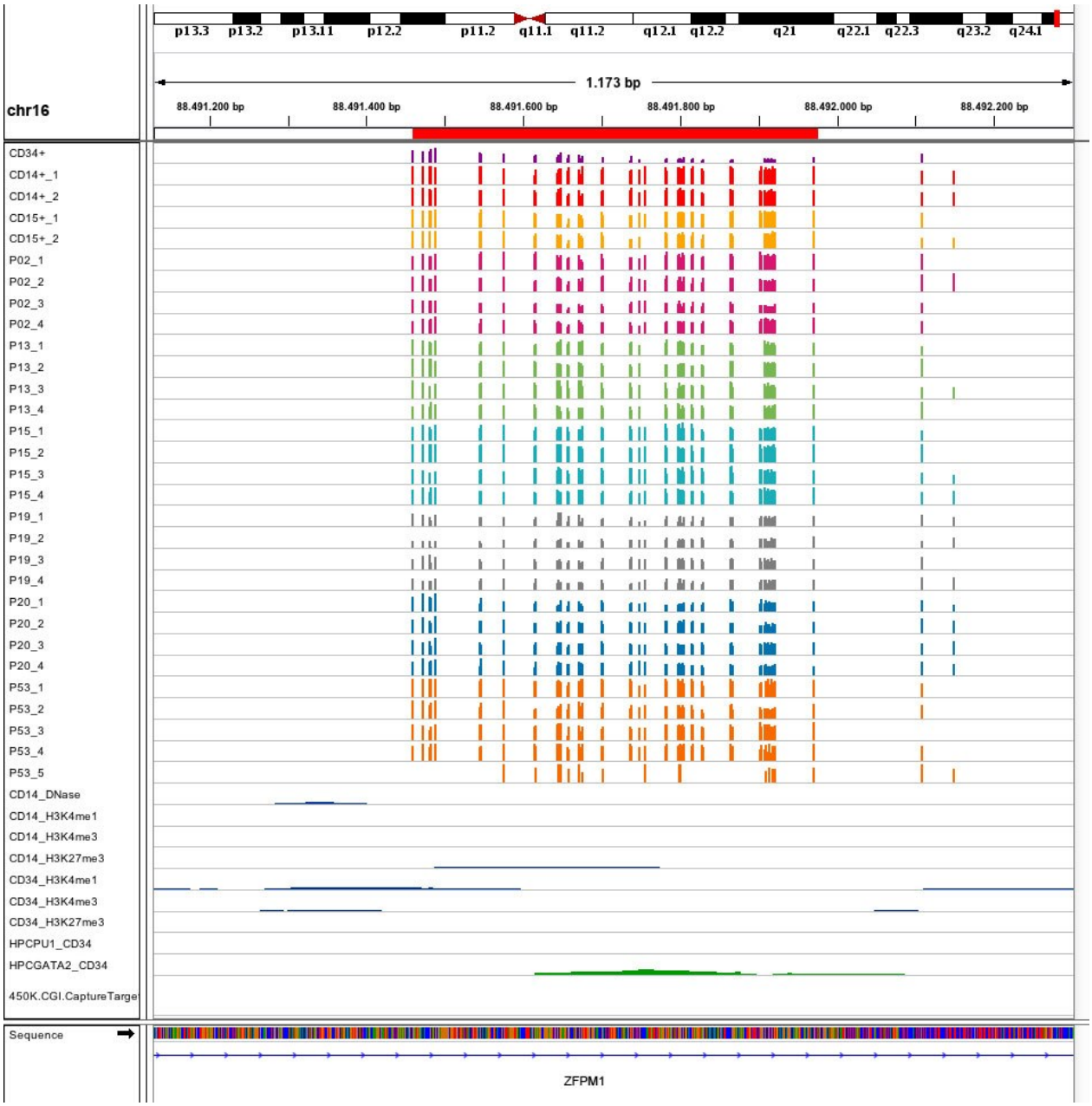


Figure 11-4 - Genome browser track of selected region for ZFPM1 gene reporter assay

11.2 Appendix II – Reporter gene assay from pediatric MDS patient D770

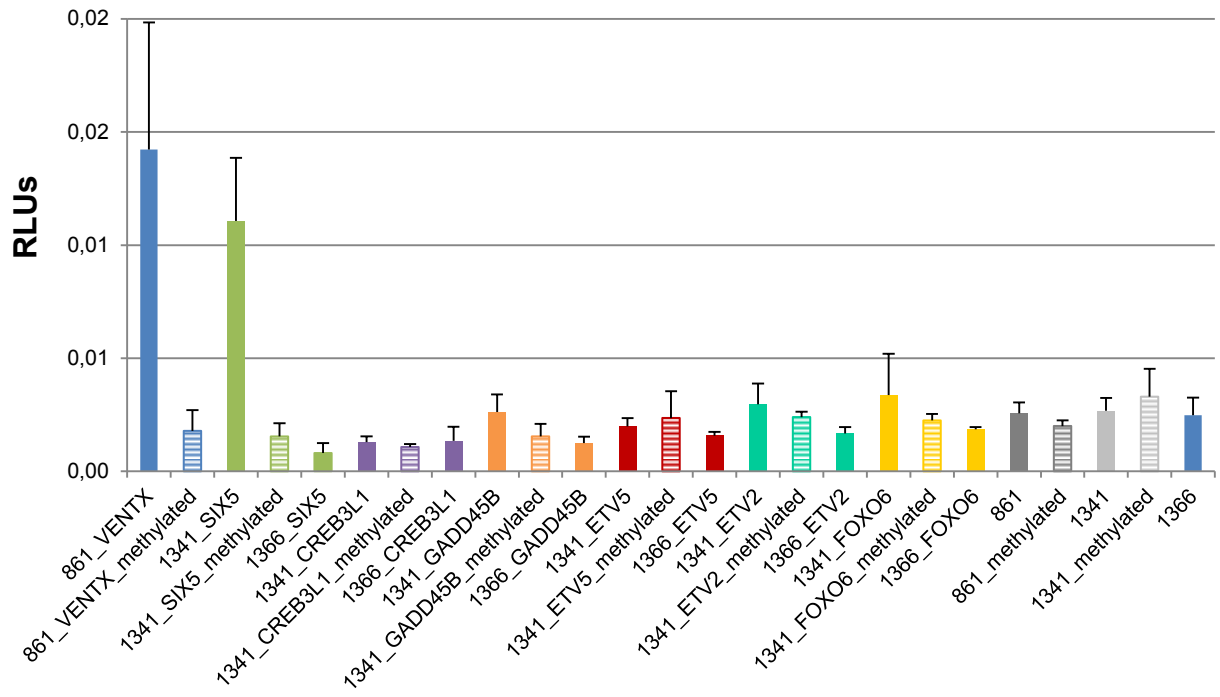
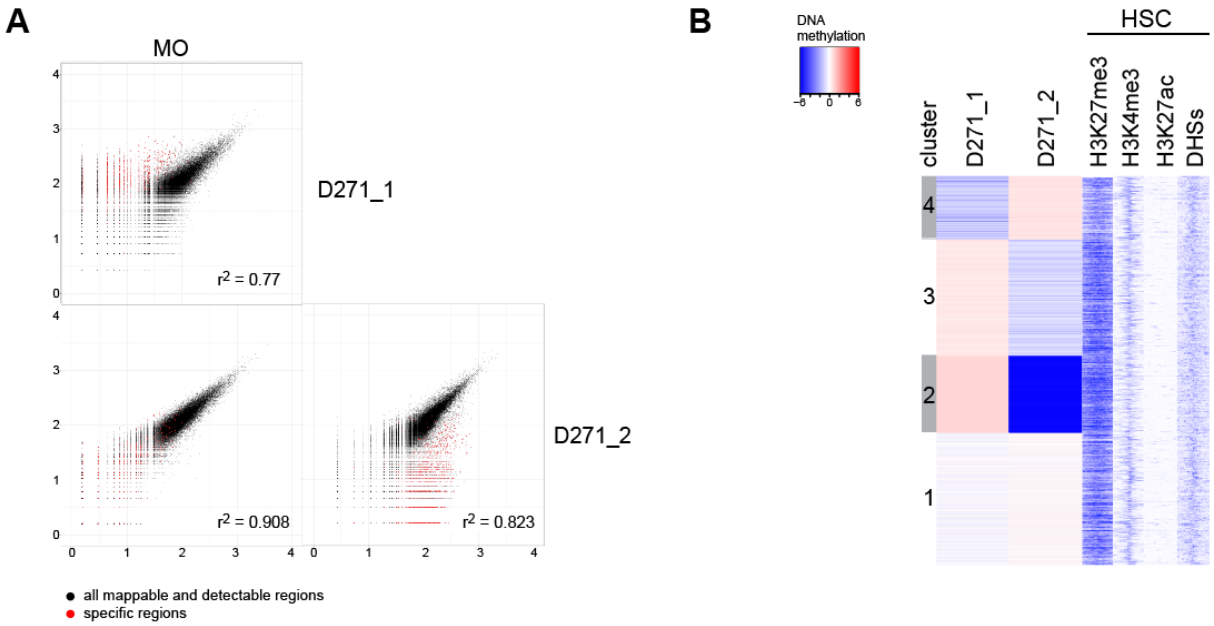


Figure 11-5 – Bar plot of gene reporter assay in pediatric MDS patient D770

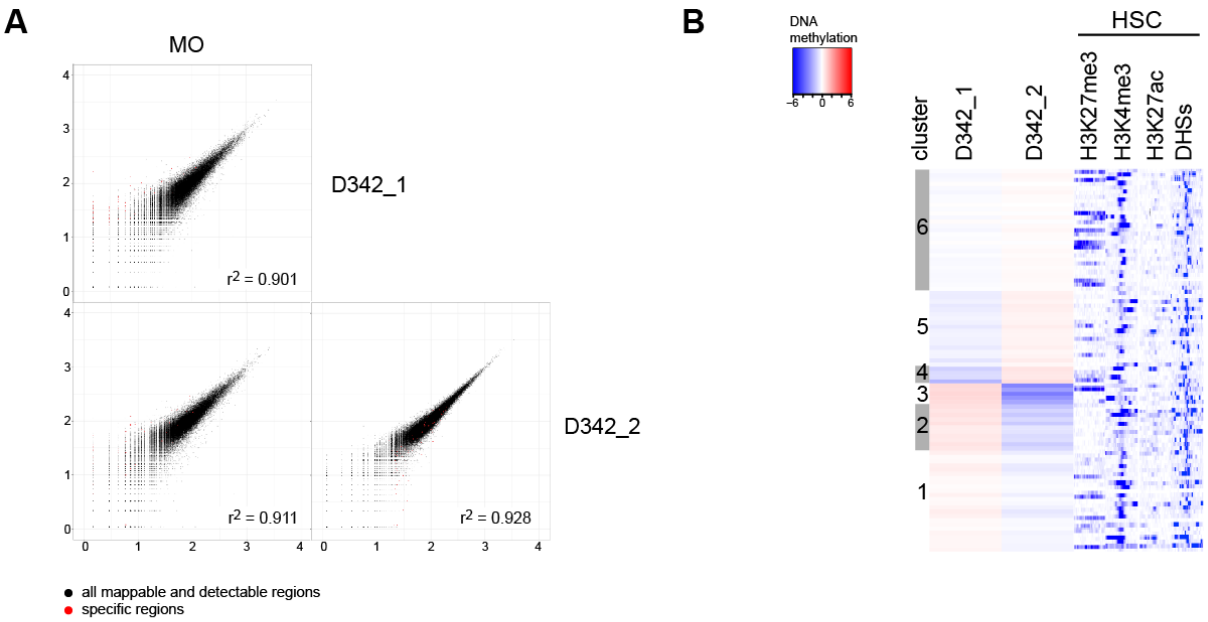
11.3 Appendix III – DNA methylation analyses and clinical data from longitudinal MDS patients



C

Number	Cumulative time [m]	WHO	Treatment	BM blasts	Mutations	Karyotype
D271_1	0	RAEB	-	9 %	GATA2	45,XY,-7
D271_2	24	RAEB	After HSCT	14 %	GATA2	45,XY,-7

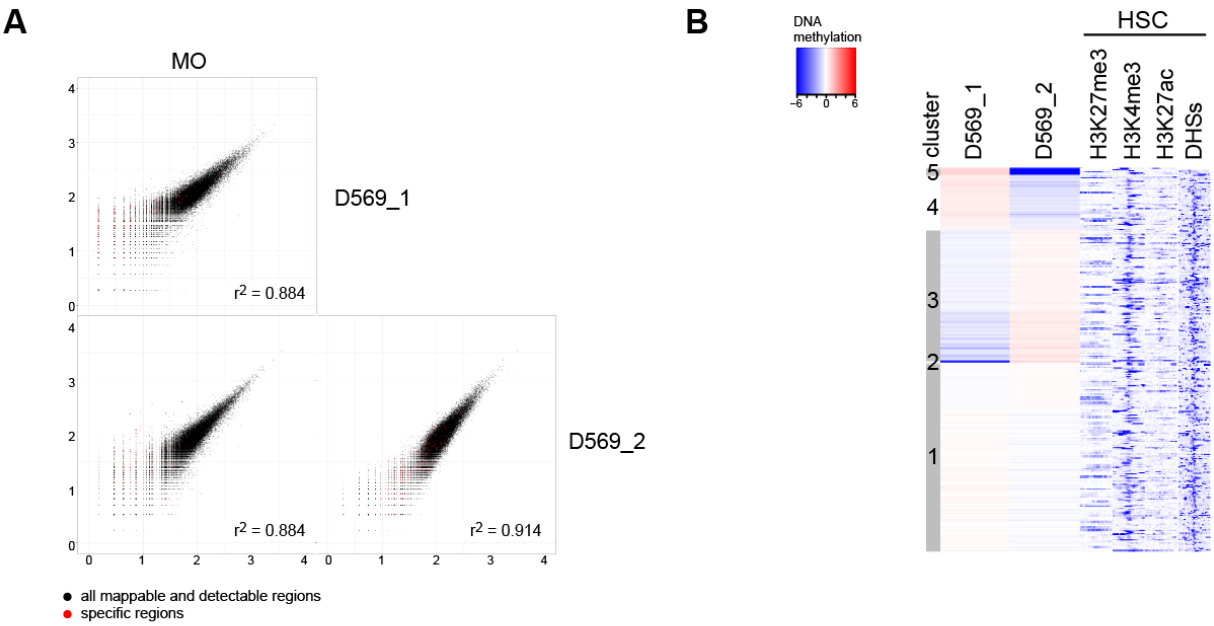
Figure 11-6 - DNA methylation analyses and clinical data from patient D271



C

Number	Cumulative time [m]	WHO	Treatment	BM blasts	Mutations	Karyotype
D342_1	0	RC	-	2 %	GATA2	45,XY,-7
D342_2	8	RC	-	0 %	GATA2	45,XY,-7

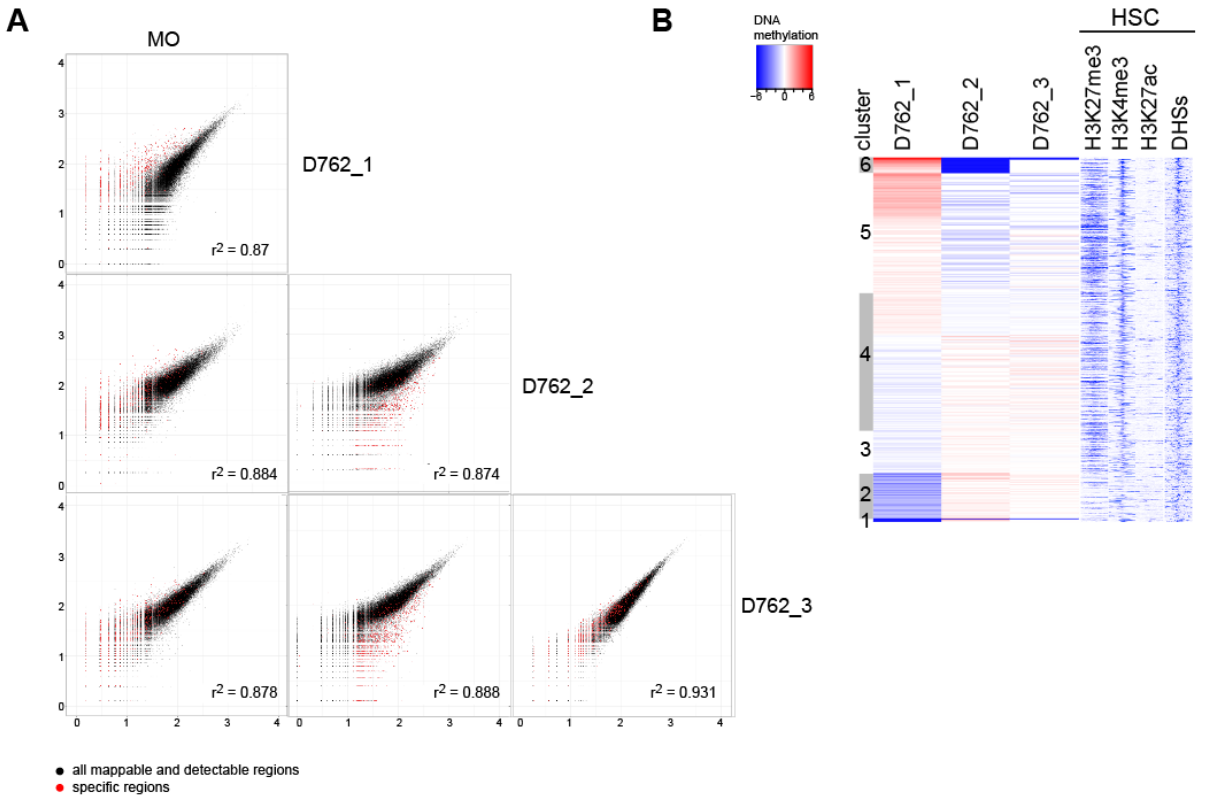
Figure 11-7 - DNA methylation analyses and clinical data from patient D342



C

Number	Cumulative time [m]	WHO	Treatment	BM blasts	Mutations	Karyotype
D569_1	0	RC	-	0 %	GATA2	Normal
D569_2	37	RC	-	0 %	GATA2	Normal

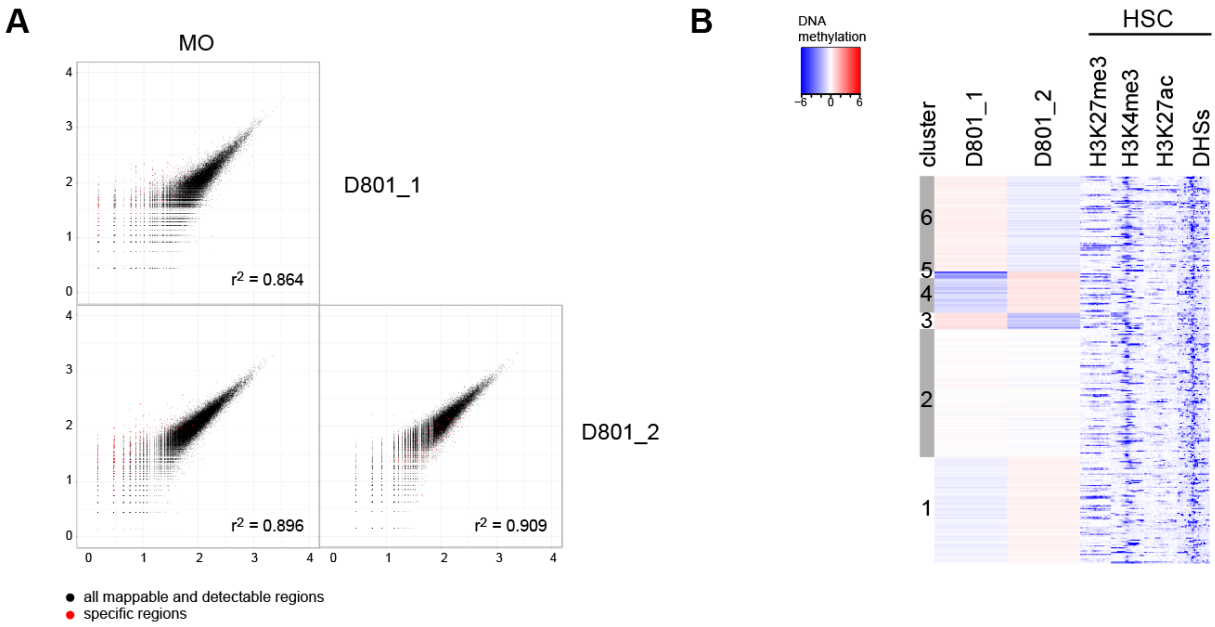
Figure 11-8 - DNA methylation analyses and clinical data from patient D569



C

Number	Cumulative time [m]	WHO	Treatment	BM blasts	Mutations	Karyotype
D762_1	0	RC	-	1 %	GATA2	45,XY,-7
D762_2	22	RC	After HSCT	NA	GATA2	45,XY,-7
D762_3	25	RC	After Vidaza	NA	GATA2	45,XY,-7

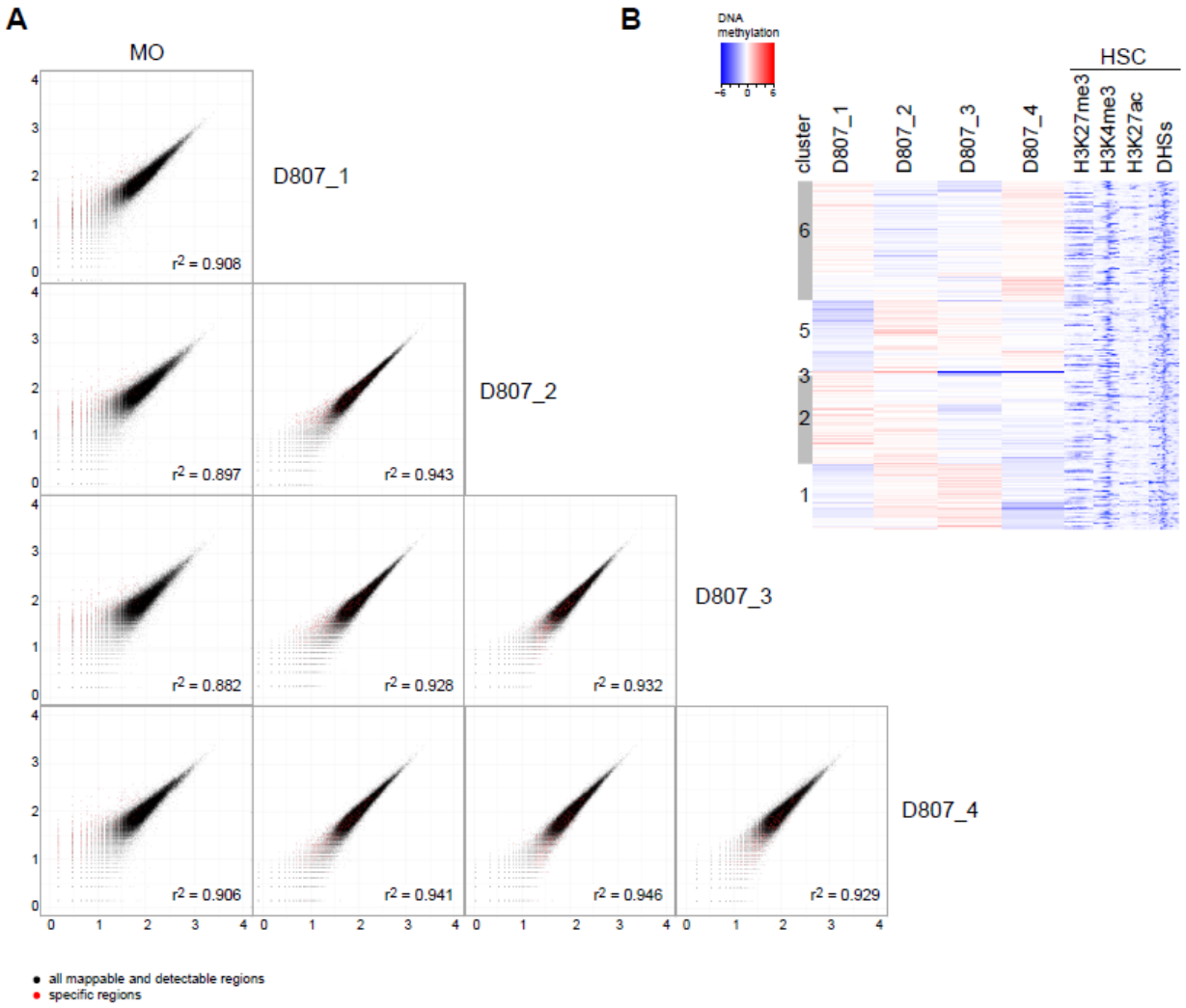
Figure 11-9 - DNA methylation analyses and clinical data from patient D762



C

Number	Cumulative time [m]	WHO	Treatment	BM blasts	Mutations	Karyotype
D801_1	0	RC	-	0 %	GATA2	Normal
D801_2	24	RC	-	0 %	GATA2	Normal

Figure 11-10 - DNA methylation analyses and clinical data from patient D801



C

Number	Cumulative time [m]	WHO	Treatment	BM blasts	Mutations	Karyotype
D807_1	0	RC	-	0 %	GATA2	Normal
D807_2	21	RC	-	0 %	GATA2	Normal
D807_3	33	RC	-	0 %	GATA2	Normal
D807_4	45	RC	-	0 %	GATA2	Normal

Figure 11-11 - DNA methylation analyses and clinical data from patient D807

Acknowledgment

Das Werk ist vollbracht und somit möchte ich mich hier nochmal bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben ☺.

Ich danke dem Direktor der Klinik und Poliklinik für Innere Med III, Prof. Dr. med Wolfgang Herr, für die Möglichkeit die Dissertation in seiner Abteilung anfertigen zu dürfen und für die tollen Weihnachtsfeiern, Betriebsausflüge und Retreats in Hirschberg.

Mein besonderer Dank gilt Prof. Dr. Michael Rehli für die gute Betreuung meiner Arbeit und sein immer offenes Ohr. Danke, dass du mir deinen Sabretooth überlassen hast ☺ und mir bei meinen bioinformatischen Problemen zur Seite standest. Danke auch für die vielen Kongressen und die tollen Forschungsretreats.

Vielen herzlichen Dank auch an Dr. med. Daniel Heudobler, meinem Co-PI, der mir durch seine positive Denkweise und Begeisterung für das Thema immer wieder neuen Mut gemacht hat. Danke auch für dein medizinisches Hintergrundwissen, welches du mir immer gerne erklärt hast. Und ohne dich wären die meisten Kongresse nur halb so lustig gewesen!

Ganz herzlich möchte ich mich bei der ganzen AG Rehli und allen Ehemaligen für das super Arbeitsklima bedanken. Mit euch konnte man den manchmal schwierigen Laboralltag mit Spaß meistern und die große Hilfsbereitschaft fand ich einfach super! Lucia – danke, dass du immer Zeit für mich hattest, wenn ich mal eine Frage hatte. Julia und Johanna, ich danke euch für die lustige und tolle Zeit, die wir hatten und für eure Freundschaft. Und danke Johanna für die tatkräftige Unterstützung bei den ganzen Transfektionen!
Claudi, danke für deinen Zuspruch und Austausch zu jeder Zeit und auch für die gute Einarbeitung am Anfang – das hat vieles erleichtert! Und natürlich danke ich allen anderen der Mannschaft für die tolle Zeit im Labor und auch außerhalb!

Danke auch an die anderen „H1-Bewohner“, vor allem AG Kreutz und AG Edinger/Hoffmann, für den schönen Laboralltag. Die immer stete Hilfsbereitschaft und den Austausch weis ich sehr zu schätzen. Ihr habt unseren „tollen“ Container zu einem schönen Arbeitsplatz gemacht!

Nicht zuletzt geht mein Dank an all meine Freunde, die immer für mich da waren und mich aufgebaut haben! Vor allem Marion danke ich für das Korrigieren meiner Arbeit – Danke für dein Durchhaltevermögen! ☺

Für die Unterstützung und den Zuspruch möchte ich ein riesiges Dankeschön an unsere Familien richten. Danke, dass ihr immer für mich da seid und mir auf meinen Weg von der Schule bis zur Promotion geholfen habt! Vor allem meinen Eltern und meiner Schwiegermama möchte ich für ihren Zuspruch und die aufbauenden Worte sowie den immerwährenden Rückhalt danken.

Zum Schluss möchte ich mich beim besten Menschen bedanken –Martin! Danke für deine unendliche Geduld, dein Vertrauen in mich und deine Liebe! Ohne dich wäre ich auf diesem steinigen Weg schon manche Male liegen geblieben. Danke, dass du immer für mich da bist! Ich möchte auch meinem Sohn Jonathan danken, der mich durch sein Rumblödeln mit mir immer wieder zum Lachen gebracht hat! Ohne euch wäre das Leben nicht mal halb so schön!! Ich liebe euch!