

From Department of Medicine, Huddinge, H7
Karolinska Institutet, Stockholm, Sweden

ABERRANT DNA METHYLATION IN ACUTE MYELOID LEUKEMIA

Ying Qu

屈颖



**Karolinska
Institutet**

Stockholm 2017

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by E-print AB, 2017

© Ying Qu, 2017

ISBN 978-91-7676-645-3

To my family

谨以此论文献给我的家人

ABSTRACT

DNA methylation is an important epigenetic mechanism that influences development and cancer by regulating gene transcription. Aberrant DNA methylation is a feature of cancer including acute myeloid leukemia (AML). It was first established that global DNA hypomethylation combined with hypermethylation of specific gene promoters could often be observed as a DNA methylation signature in cancer. A common set of tumor suppressor genes are found consistently hypermethylated and silenced, suggesting that DNA methylation facilitates tumorigenesis. Lately, the more dynamic DNA methylation at non-CGI regions and CpG sparse regions of the genome has been observed, and it tightly corresponds to gene expression changes. In AML, highly distinctive genome-wide DNA methylation profiles have been linked to different molecular subtypes. It is now suspected that DNA methylation changes play a crucial role in AML development particularly since the identification of frequent somatic mutations in the DNA methylation machinery.

This thesis is focused on characterizing aberrant DNA methylation changes in the subgroup of AML patients identified as cytogenetic normal (CN-AML). We described the mutation associated DNA methylation signatures for *IDH* and *NPM1* in a CGI-focused analysis. We also found that PcG target genes were preferentially targeted by methylation changes and methylation of this group of genes predicted the patient clinical outcomes. In the following studies, we analyzed the DNA methylation in more border regions, and we classified the variably methylated CpG sites in correlations with genetic mutations. We found a predominant impact of *DNMT3A* mutation on determining leukemia-specific methylation patterns and such mutations were associated with a general hypomethylation phenotype, where *HOX* family was primarily affected. We also observed pronounced DNA methylation changes at non-CGI regions, and these changes reflect the regulation of enhancer activity in leukemia. After integrating chromatin accessibility of DHS sequencing data and histone modification marks of H3K27ac, H3K4me1, H3K4me3 and H2A.Z with identified differentially methylated CpG sites, and our results show that DNA methylation alterations preferentially occur in regulatory regions. AML specific DNA methylation changes associated with altered enhancer activities, and these perturbations correlated with transcriptomic changes in CN-AML involving in oncogenesis and associated with patient prognosis.

Our results provide evidence of aberrant DNA methylation in AML linked to patient molecular and genetic characteristics. Studying DNA methylation changes not only contributes to better characterizing subgroups of AML patients but also reveals potentially pathogenic mechanisms for AML development.

LIST OF PUBLICATIONS

- I. **Prognostic DNA methylation patterns in cytogenetically normal acute myeloid leukemia are predefined by stem cell chromatin marks.**
Deneberg S, Guardiola P, Lennartsson A, Qu Y, Gaidzik V, Blanchet O, Karimi M, Bengtzén S, Nahi H, Uggla B, Tidefelt U, Höglund M, Paul C, Ekwall K, Döhner K, Lehmann S. *Blood*, 2011 Nov 17;118(20):5573-82.

- II. **Differential methylation in CN-AML preferentially targets non-CGI regions and is dictated by DNMT3A mutational status and associated with predominant hypomethylation of HOX genes.**
Qu Y, Lennartsson A, Gaidzik V, Deneberg S, Karimi M, Bengtzén S, Höglund M, Bullinger L, Döhner K, and Lehmann S. *Epigenetics*, 2014; (9):8, 1108–1119.

- III. **Cancer-specific changes in DNA methylation reveal aberrant silencing and activation of enhancers in leukemia.**
Qu Y,* Siggens L,* Cordeddu L, Gaidzik V, Karlsson K, Bullinger L, Döhner K, Ekwall K, Lehmann S,† and Lennartsson A.† *Blood*, 2017;129(7):e13-e25.

* shared first authors

† shared senior authors

CONTENTS

1	Introduction.....	1
1.1	Epigenetics.....	1
1.1.1	DNA, Histone and Chromatin	1
1.1.2	DNA methylation and methyltransferases.....	2
1.1.3	DNA demethylation and related enzymes.....	5
1.1.4	Core histones and histone variants.....	6
1.1.5	Histone modifications	7
1.1.6	Epigenetic cross-talks	8
1.1.7	Transcription regulatory sequences	9
1.2	Acute myeloid leukemia	12
1.2.1	Risk factors.....	12
1.2.2	Classification of AML.....	12
1.2.3	Prognostic factors.....	13
1.2.4	Molecular genetic changes in CN-AML	16
1.2.5	Commonly mutated genes in CN-AML	17
1.3	Epigenetic mechanisms in hematopoiesis and AML	22
1.3.1	Hematopoiesis	22
1.3.2	Epigenetic mechanisms in normal hematopoiesis.....	22
1.3.3	Aberrant epigenetic changes in AML.....	25
2	Aim of The Thesis	28
3	Methodological Approaches.....	29
3.1	DNA methylation detection	29
3.1.1	Bisulfite conversion.....	29
3.1.2	Pyrosequencing	30
3.1.3	Illumina methylation arrays	31
3.1.4	Other genome-wide methylation platforms.....	32
3.2	Chromatin accessibility and modification analysis	32
3.2.1	Dnase I hypersensitivity site sequencing.....	33
3.2.2	Chromatin immunoprecipitation and sequencing	33
3.3	Transcriptome profiling.....	34
3.3.1	Gene expression microarray	34
3.3.2	Messenger RNA sequencing.....	35
3.4	Genome editing with CRISPR-Cas9 system	36
4	Results and Discussion	39
4.1	Paper I.....	39
4.2	Paper II.....	40
4.3	Paper III	43
5	Concluding Remarks	45
6	Acknowledgements.....	47
7	References	50

LIST OF ABBREVIATIONS

ASXL1	Additional sex combs like 1
5hmC	5'-hydroxymethylcytosine
5mC	5'-methylcytosine
ALL	Acute lymphoid leukemia
AML	Acute myeloid leukemia
BER	Base excision repair
CEBPA	CCAAT/enhancer binding protein alpha
CGI	CpG island
ChIP-seq	Chromatin immunoprecipitation and sequencing
CN-AML	Cytogenetic normal AML
CRISPR	Clustered regularly interspaced short palindromic repeats
DHS-seq	DNase hypersensitivity sites and sequencing.
DNMT	DNA methyltransferase
EZH2	Enhancer of zeste homolog 2
FLT3	Fms-related tyrosine kinase 3
GADD45	DNA damage inducible protein
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HKMT	Histone lysine methyltransferase
HOX	Homeobox
IDH	Isocitrate dehydrogenase
KDM	Lysine demethylase
MBD	Methyl-CpG binding domain
MDS	Myelodysplastic syndromes
MLL	Mix lineage leukemia
MPD	Myloid proliferative diseases
NPM1	Nucleophosmin 1
NuRD	Nucleosome remodeling deacetylase
PcG	Polycomb group
Pol II	RNA polymerase II
PRC2	Polycomb repressive complex 2
RUNX1	Runt-related transcription factor 1
TET	Ten-eleven-translocation
TSS	Transcription start site
TF	Transcription factor
UHRF1	Ubiquitin-like containing PHD and RING finger domain 1
WT1	Wilms' tumor 1

1.Introduction

1.1 Epigenetics

Every cell in the human body starts with the same genetic information, yet the body produces a variety of distinct cell types, all of which look and function in unique ways. The word "Epigenetics" was first coined by Conrad Waddington(1905-1975), who realized that phenotypic diversity amongst cell types could not be explained by genetics, given that most different cell types are genetically identical(Holliday 2006). Conrad Waddington stated epigenetics is "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being," introducing the idea of an "epigenetic landscape" which related to cell fate decisions during development(Waddington 1957; Waddington 1959).

After Waddington, the following six decades of research into epigenetics has seen considerable developments in what epigenetics represents. Epigenetics has been used to refer to both heritable and non-heritable processes(Bird 2007). A consensus definition proposed at a Cold Spring Harbor meeting in 2008 suggested epigenetics was a "stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence"(Berger, Kouzarides et al. 2009). In a more recent study, the Roadmap Epigenome Project used the following definition: "epigenetics refers to both heritable changes in gene activity and expression (in the progeny of cells or of individuals) and also stable, long-term alterations in the transcriptional potential of a cell that are not necessarily heritable."(Skipper, Eccleston et al. 2015). Nowadays, researchers in the field of epigenetics study biological processes including DNA methylation, histone posttranscriptional modifications (histone modifications in short), chromatin remodeling, and non-coding RNA, which all regulate gene expression during development. These mechanisms play essential roles in normal development and disease, including hematological malignancies.

1.1.1 DNA, histone, and chromatin

There are approximately two meters in length of DNA stored in each human cell nucleus, which in turn is typically only six micrometers in diameter. The DNA, therefore, must be compacted and organized into a functional but extremely efficient space structure. It must also allow active transcription of the relevant genes while at the same time making sure unwanted genes are silent. Chromatin fulfills the requirement to

package, store, and regulate DNA. The basic unit of chromatin is the nucleosome: 146bp of DNA wrapped around an octameric protein complex consisting of histone proteins. Short stretches of linker DNA connect neighboring nucleosomes to one another like beads on a string. Chromatin is further compacted into thicker and thicker fibers. Euchromatin contains the part of the genome that has active genes and is relatively open in structure and accessible to DNA-binding factors such as transcription factors(Allis and Jenuwein 2016). In contrast, heterochromatin is more tightly packaged and contains mostly inactive regions like repetitive sequences or genes that are inactive(Allis and Jenuwein 2016). Chromatin dynamics are tightly regulated by post-translational modifications of the histone proteins as well as methylation of the DNA itself, and the binding of transcription factors. Such regulation is in part what allows for the same genetic material to produce a variety of diverse cell types.

1.1.2 DNA methylation and methyltransferases

DNA methylation is a covalent chemical modification whereby a methyl group is added to the base cytosine or adenine(Weinberg 2014). In lower organisms, such as bacteria or protists, methylation occurs at either the 5' position of cytosine (5mC) or the 6' position of adenine (6mA)(Heyn and Esteller 2015). In vertebrates, 5mC is the predominant form of DNA methylation; it is assumed that 6mA is much less abundant, although recent studies have demonstrated that 6mA does occur in the human genome(Jiang, Wang et al. 2014; Wu, Wang et al. 2016). The modification of DNA through methylation regulates cell behavior and development(Yoder, Walsh et al. 1997; Zhu, Srinivasan et al. 2003; Fujimoto, Kitazawa et al. 2005; Chodavarapu, Feng et al. 2010; Shukla, Kavak et al. 2011; Berman, Weisenberger et al. 2012; ENCODE 2012; Jimenez-Useche, Ke et al. 2013).

In the human genome, 5mC makes up 1.5% to 2% of the total DNA and accounts for the majority (60% to 80%) of total CG sites(Lister, Pelizzola et al. 2009). In mammalian cells, the methyl group is supplied from the metabolite S-Adenosyl-Methionine (SAM) (Takahashi, Wang et al.), and added at the 5' carbon of cytosine's pyrimidine ring to form 5-methylcytosine (5mC) (Figure 1). In eukaryotic organisms, 5mC occurs symmetrically at CG dinucleotides, of which a cytosine nucleotide is located next to a guanidine; this is often referred to as a CpG site(Lister, Pelizzola et al. 2009; Feng, Cokus et al. 2010). Although methylation could occur in CHG and CHH sites at different rates, their functions in the mammalian system remain

unknown(Schultz, He et al. 2015). It is thought that 5mC functions in suppressing transposon activity and regulating gene expression, as well as imprinting and the formation of heterochromatin(Mohandas, Sparkes et al. 1981; Li, Bestor et al. 1992; Okano, Bell et al. 1999; Jones and Liang 2009; Challen, Sun et al. 2014). One consequence of 5mC in the genome is that it favors the spontaneous deamination that results in the conversion of cytosine to uracil (U), which after DNA repair, produces a C>T mutation(Duncan and Miller 1980; Hitchins, Rapkins et al. 2011). Genome-wide studies have revealed that regions enriched with CG's are often gene promoters, of which CpG islands (CGI's) are often clustered(Deaton and Bird 2011; Hernando-Herraez, Garcia-Perez et al. 2015). CGIs are defined as regions longer than 200bp with an expected CpG frequency more than 60% (Gardiner-Garden and Frommer 1987). It is thought that CpG dinucleotides are globally underrepresented in the genome, which may be related to the deamination process in the germline(Law and Jacobsen 2010; He, Chen et al. 2011; Jiang, Wang et al. 2014). While, in general, CG sites are mostly methylated in the mammalian genome, but CGI's usually remain unmethylated(Illingworth and Bird 2009). It may be for the purpose of protection from the spontaneous mutation, and also, allowing access for transcription initiation. In the human genome, approximately 60-70% of genes contain CpG islands in their promoter regions, and many of them are so-called housekeeping genes and genes regulating essential developmental processes(Bird 2009; Deaton and Bird 2011; Smallwood, Tomizawa et al. 2011). These genes are thought to be only transiently or never methylated at the germline in order to ensure the maintenance of pluripotency during embryonic development(Smallwood, Tomizawa et al. 2011). On the contrary, CpG sites located in the gene body are often methylated in highly transcribed genes and positively correlates with gene expression(Oberdoerffer 2012). CpG methylation in the gene body is also related to the regulation of alternative splicing and the transcription of intronic repeat sequences(Lister, Pelizzola et al. 2009; Malousi and Kouidou 2012). The mechanisms that regulate this type of region/content-dependent DNA methylation are not fully understood yet. One recent study revealed that gene body DNA methylation catalyzed by methyltransferase DNMT3B is regulated by local trimethylation at histone H3 lysine 36 (H3K36me3) in highly transcribed genes(Neri, Rapelli et al. 2017). However, an understanding of the functional differences and target preferences of human DNA methyltransferases is still lacking and may lead to a new research focus in the near future.

In mammals, DNA methylation at the 5' position of cytosine is catalyzed by a group of enzymes named **DNA methyltransferases (DNMTs)**. Such DNA methyltransferases from bacteria to human share conserved functional protein domains(Lauster, Trautner et al. 1989). In humans, DNA methylation is carried out in two steps: *de novo* methylation and maintenance methylation during DNA replication. The very first epigenetic reprogramming occurs as early as 4 hours after fertilization, and it is believed that global demethylation functions to "erase" the somatic signatures of the

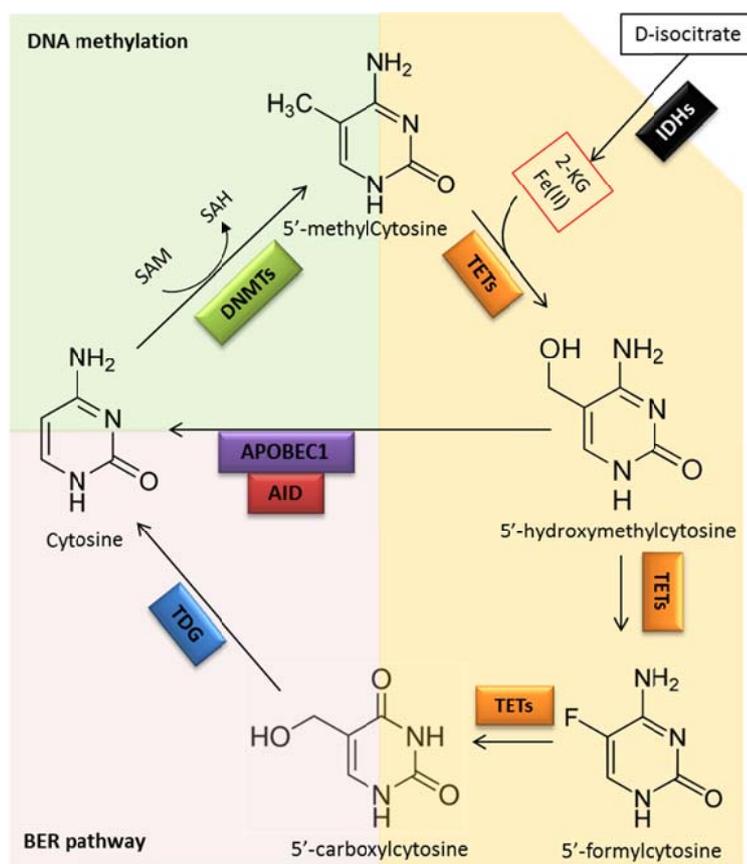


Figure 1. DNA methylation and demethylation. The schematic mechanisms of DNA methylation regulation and involving enzymes are shown: DNMTs, DNA methyltransferases; IDHs, isocitrate dehydrogenases; TETs, ten-eleven-translocation; 2-HG, 2-hydroxyglutarate; AID, activation-induced cytidine deaminase; APOBEC1, apolipoprotein B mRNA-editing enzyme catalytic polypeptide 1; BER, base-excision repair; SAH, S-adenosylhomocysteine; SAM, S-adenosylmethionine; TDG, thymine DNA glycosylase.

paternal genome and reset the differentiation potential of the embryonic stem cells(Smith, Chan et al. 2014). The somatic DNA methylation pattern is started from the time of implantation and can be observed at E10.5 to E13.5 days in mice and from week 4 to week 19 in humans(Guo, Zhu et al. 2014). DNMT3A and DNMT3B are two

de novo methyltransferases involved in the establishment of DNA methylation marks on native DNA strands. Both enzymes contain a methylase catalytic domain and PWWP domain that promote association to heterochromatin. Another member of the DNMT3 subfamily is DNMT3L, which is a catalytic paralog of DNMT3A and DNMT3B involved in the re-establishment of genomic imprinting and methylation of transposon elements at gametogenesis(Bourc'his, Xu et al. 2001) . After somatic methylation is established, DNA methylation is maintained by DNMT1, which is recruited together with ubiquitin-like containing PHD and RING finger domain 1(UHRF1) and methylates the newly synthesized DNA strand during each cell division(Sharif, Muto et al. 2007). In animal models, depletion of DNA methyltransferases influences embryonic development and survival of cells. Previous studies has reported that knocking out of *Dnmt1* or *Dnmt3b* in mice is embryonic lethal, whereas *Dnmt3a* knockout mice are viable after birth but die 21 days postnatally(Li, Bestor et al. 1992; Okano, Bell et al. 1999). The other member of the human DNMT family, DNMT2, also exerts methyltransferase activity but only acts on tRNAs(Goll, Kirpekar et al. 2006).

1.1.3 DNA demethylation and related enzymes

To remove methylation marks from the DNA, there are two conceivable mechanisms: passive demethylation and active demethylation. Disruption to the maintenance of DNA methylation during replication leads to the passive erasure of DNA methylation(Jones 2012). This process can be exemplified by inhibition of DNMT1 activity, for example by using the drug 5'-azacytidine(Issa, Kantarjian et al. 2005). It is a chemical analogue of native nucleoside cytosine and can be incorporated into DNA and RNA that inhibits methyltransferase activity. In lower organisms, the mechanism of active demethylation is through 5-methylcytosine DNA glycosylases (such as DME/ROS1 family in Arabidopsis) by working together with base excision repair (BER) pathway(Penterman, Zilberman et al. 2007). However, the orthologues of DME family is remaining under-discovered in mammals. In vertebrates, active demethylation can occur through cytosine deamination followed by DNA repair. It has been found that activation-induced deaminase (AID) and apolipoprotein B mRNA-editing enzyme catalytic polypeptide 1 (APOBEC1) is able to convert 5mC to uracil (U) resulting in a T-G mismatch(Nabel, Jia et al. 2012). The T-G mismatch can then be removed through BER, nucleotide excision repair (NER), or mismatch repair (MMR). Another suggested mediator of DNA demethylation is the growth arrest and DNA damage-inducible

protein (GADD45) gene family, which may promote locus specific demethylation(Rai, Huggins et al. 2008; Engel, Tront et al. 2009). The discovery of **ten-eleven-translocation(TET)** protein family and 5'-hydroxymethylcytosine (5hmC) suggests that active demethylation may function through "detour" pathways(Iyer, Tahiliani et al. 2009; Tahiliani, Koh et al. 2009).

TET family proteins were discovered as fusion proteins of MLL translocations in acute myeloid leukemia(Tahiliani, Koh et al. 2009). Since then, there have been three family members (TET1, TET2, TET3) identified in humans, of which all have been found to display oxidase activity converting 5mC to 5-hmC, 5-formylcytosine, 5-carboxylcytosine in a serial of reactions using α -ketoglutarate (α -KG) as substrate and Fe^{2+} as a cofactor(He, Li et al. 2011; Ito, Shen et al. 2011). This chain of reactions leads finally to DNA demethylation via thymine DNA glycosylase (TDG) mediated BER mechanism. It has found that depletion of TET1 in human cells led to increased 5mC and decreased 5hmC globally(Xu, Wu et al. 2011). Other than catalytic activity, TET1 was found to enriched bind to CpG dense regions in mouse embryonic stem cell. Loss of TET expression in mouse ES cells comprised their differentiation capacity by deregulation of gene expressions and global promoter hypermethylation was found(Dawlaty, Breiling et al. 2014). In the cells, α -KG is produced through the tricarboxylic cycle. The reaction is catalyzed by **Isocitrate dehydrogenase (IDH)**, that D-isocitrate undergoes oxidative decarboxylation to α -KG(Medeiros, Fathi et al. 2017). Two isoforms of IDH, IDH1 and IDH2 are found in different cellular compartments. IDH1 is mainly found in cytoplasm and peroxisomes, whereas IDH2 locates in the mitochondrial matrix. Both *IDH* genes are frequently mutated in hematological malignancies, that leads to "gain-of-function" and abnormally produces an "oncometabolite", 2-hydroglutarate (2-HG), instead of α -KG(Dang, White et al. 2009; Icard, Poulain et al. 2012). Therefore, mutations of *IDH* inhibit the TET's function and disrupt TET-mediated demethylation machinery. Moreover, 2-HG also inhibits α -KG-dependent histone demethylases, which leads to consequential increase of repressive chromatin marks, for instance, tri-methylation at histone H3 lysine 9 residue (Lu, Ward et al. 2012).

1.1.4 Core histones and histone variants

Histones, together with the DNA, make the two essential components of chromatin. There are four canonical core histone proteins. In addition to the linker histone H1 that

sits above on each nucleosome at the entry/exit of the linker DNA strand, the nucleosome core particle is made up of histones H2A, H2B, H3, and H4(Kornberg 1974; Kornberg and Thomas 1974). Each octameric nucleosome contains two H2A-H2B dimers and two H3-H4 dimers and their unpacked amino acid tails at both ends of each histone protein extend from complex core. Post-transcription modifications covalently occur on these histone amino acid tails and they are crucial to gene regulatory mechanisms. More than canonical histone proteins, histone variants exist for H3, H2A, and H2B(Buschbeck and Hake 2017). They give rise to diversity amongst nucleosomes, and often, these variant proteins are specified for different functional roles. For example, one major variant form of the H2A core protein, H2A.Z, is more often found in the promoter region of genes and enhancers and antagonizes DNA methylation(Raisner, Hartley et al. 2005; Ku, Jaffe et al. 2012). Another H2A variant, H2A.X, is highly involved in DNA double-strand break repair and undergoes phosphorylation to signal to the DNA repair enzymes(Kuo and Yang 2008; Mah, El-Osta et al. 2010). Another example of histone variant centromere-specific H3 variant (CENP-A) is found in centromeric regions and is associated with repressive chromatin(Molina, Vargiu et al. 2016).

1.1.5 Histone modifications

Histones are subject to at least 15 different post-translational modifications, among which acetylation, methylation, and phosphorylation are the most studied ones(Bannister and Kouzarides 2011). These modifications occur on several amino acid residues including Lysine (K), Arginine(R), Serine(S), Glutamate(E), and Tyrosine(T) and serve in signaling to the transcription regulatory apparatus.

Lysine is the most commonly modified amino acid residue in histone proteins. The **acetylation of lysine** neutralizes its positive charge and weakens the electrostatic association with wrapping DNA and is associated with active "open" chromatin(Bannister and Kouzarides 2011). This covalent change is tightly associated with the cellular factors that require access structure to the DNA. Histone acetylation is enriched at regions of transcription start sites (TSS) (such as H3K9ac) and presented through the gene body (such as H3K12ac) of actively transcribed genes and regulatory elements such as active enhancers (with the presence of H3K27ac, H3K122ac)(Wang, Zang et al. 2008; Tang, An et al. 2014). This modification is catalyzed by a family of

enzymes named histone acetyltransferases (HATs) and can be removed by histone deacetylases (HDACs)(Bannister and Kouzarides 2011).

The methylation of histone lysine residues has diverse impacts on function depending on the state of progressive methylation (since lysine residues can be mono-, di- and trimethylated) with different lysine residues playing distinct roles(Bannister, Schneider et al. 2002). This modification is catalyzed by histone lysine methyltransferases (HKMTs) with a methyl group donated from the metabolite SAM replacing each hydrogen of the lysine NH₃- group(Audia and Campbell 2016). Most HKMTs are highly substrate-specific and contain a highly conserved SET-domain often functioning within protein complex formed with other cofactors(Li, Han et al. 2016). More recently, enzymes without an SET-domain have been found to display similar HKMT activity, such as DOT1L, catalyze methylation of H3K79(Feng, Wang et al. 2002). The methylation of histone is considered a stable mark that helps to epigenetically stabilize chromatin states, yet there are also histone demethylase enzymes capable of removing methyl groups(Bannister, Schneider et al. 2002). Lysine-specific demethylase 1 (LSD1) was the first histone demethylase identified in 2004 that facilitates the removal of mono- and di-methylation of H3K9 and H3K4(Shi, Lan et al. 2004). In addition, a large family of enzymes containing a jumonji-domain was also discovered(Takeuchi, Watanabe et al. 2006). They catalyze histone demethylation by Fe²⁺-and- α -KG-dependent dioxygenase activity. The functional roles of methylation of histones are linked to active, repressive, or bivalent states of transcription. H3K4me₃ has often been identified as the promoter of active gene whereas H3K27me₃ marks repressed transcriptional activity when seen at promoter regions(Bannister, Schneider et al. 2002; Klose and Zhang 2007). However, H3K9me₃ generally associates to heterchromatin states and transcription repression. It is also found that in the stem and progenitor cells, developmental-required genes often associate both H3K4me₃ and H3K27me₃ and are called "bivalent marks" whereby the switching on/off occurs at the appropriate time during lineage commitment(Marks, Kalkan et al. 2012; Vastenhouw and Schier 2012).

1.1.6 Epigenetic cross-talks

Our current understanding of cross-talk between epigenetic mechanisms is not yet completely understood; however, there do appear to be clear examples of such cross-talks. For instance, methylated DNA can be recognized by protein families such as the methyl-CpG binding domain (MBD) protein family and SPA-family (such as

UHRF1)(Bogdanovic and Veenstra 2009). Five members are included in MBD family, MeCP2 and MBD1-4. They are believed to function as a mediator of transcriptional repression by recruiting HDACs and HKMTs. For example, MBD2 and MBD3 participates in the nucleosome remodeling deacetylase (NuRD) complex together with other cofactors, such as HDAC1 and HDAC2, chromo domain3 (CHD3) or CHD4. Overexpression of MBD3 in the NuRD complex jeopardized the reprogramming of IPS cells through the enhancement of heterochromatin establishments and silencing of stem cell genes(Luo, Ling et al. 2013). On the other hand, unmethylated DNA can be recognized by proteins containing CXXC domain(Xu, Bian et al. 2011). Two members of the H3K4 methyltransferases MLL family (MLL1 and MLL2) contain the CXXC domain as well as the CXXC finger protein 1 (CFP1), two TET proteins (TET1, TET3), and H3K36 demethylases (KDM2A/2B)(Long, Blackledge et al. 2013). They are recruited to unmethylated DNA loci and facilitate active chromatin states and often promote gene expression. Moreover, transcription factors or DNA binding factors bind to and prevent methylation of such loci and interact with local histone modifications(Jones 2012). It can be exemplified by polycomb 2 (PRC2) complex occupancy at unmethylated CGI promoter catalyzing regional H3K27me3 and leading to the transcription repression of target genes(Khan, Lee et al. 2015). These interactions of epigenetic mechanisms help in the self-reinforcement of epigenetic states, therefore promoting phenotypic stability.

1.1.7 Transcription regulatory sequences

In eukaryotic cells, protein-coding genes are transcribed by RNA polymerase II (Pol II), and this process is precisely regulated by multiple factors to ensure appropriate transcription. Open reading frames consist of exons and introns to be transcribed into pre-mRNA. Intronic sequences will be later spliced out, and mRNA matures with 5'-cap and 3'-poly-adenylation. The immediate sequences adjacent to the open reading frame are the 5' untranslated region (UTR) upstream and the 3'UTR. **Promoter** sequences are defined as regulatory regions upstream of transcription start site (TSS) and contain binding platforms for Pol II (core promoter) and active transcription factors (proximal promoter). A core promoter serves as the entry site for Pol II complex and often contains TATA box and a B-recognition element(Lagrange, Kapanidis et al. 1998; Smale and Kadonaga 2003). They are recognized by TATA-box binding protein (TBP) and promote the recruitment of general transcription factors (GTFs) to assemble into the transcription pre-initiation complex. The proximal promoter refers to the region

upstream of the core promoter and TSS containing sequence-specific transcription factor binding sites. It is recognized by activated TF proteins, which in turn facilitate the recruitment of coactivators or repressors thus regulating gene expression(Weake and Workman 2010).

Other than gene promoter, regulation of transcription activity is fine-tuned by TSS-distal regions called enhancers and insulators. These are cis-acting regulatory elements at various distances from target gene promoters(Visel, Rubin et al. 2009). **Enhancers** are non-coding DNA sequences containing binding sites for DNA-binding proteins, and range in size from 200 to 1000 bp(Andersson, Gebhard et al. 2014). It is believed that enhancers regulate their target gene's expression by promoting physical interactions with the cognate promoters through DNA looping (Visel, Rubin et al. 2009). It has been shown that enhancers can recruit the transcription pre-initiation complex at its locus(Andersson, Gebhard et al. 2014). Meanwhile, the co-localization of cohesin and mediator complexes, as well as the transcription factor CTCF helps on generating cell-type specific DNA looping to activate gene expression(Wendt, Yoshida et al. 2008; Kagey, Newman et al. 2010; Deng, Lee et al. 2012). In this process, lineage-specific transcription factors are thought to be with a particular importance, for instance, pioneer transcription factors PU.1 and GATA1, can bind to chromatin and initiate cell type-specific histone modification changes during development(Xu, Watts et al. 2009; Heinz, Benner et al. 2010). These pieces of evidence suggest that enhancers deliver functional protein complexes to target promoters and facilitate changes to the local chromatin.

In recent years, growing efforts have been put into identifying putative enhancers and their activities during development and in cell-specific stages. Similar to promoters, enhancers are also found with functional relevant histone modifications(Pennacchio, Bickmore et al. 2013; Heinz, Romanoski et al. 2015). Active enhancers are often marked with the absence of H3K27me3 but a high level of H3K4me1 and H3K27ac together. On the other hand, poised enhancers could display H3K27me3 and H3K4me1 at the same time, both often at lower levels and in the absence of the active chromatin mark H3K27ac(Heintzman, Stuart et al. 2007; Visel, Blow et al. 2009; Kundaje, Meuleman et al. 2015). Moreover, active enhancers have been identified by the binding of the HAT enzyme called P300, or at DNaseI hypersensitive sites(DHSs), and often transcribed into non-coding/enhancer RNAs(Birney, Stamatoyannopoulos et al. 2007;

Li, Notani et al. 2013). Several international research consortiums have focused on identifying genome-wide putative enhancers for tissue/cell type-specific enhancers in large numbers of cell lines and primary human samples. By Sequencing of the 5-Capped end of RNA (CAGE), the Fantom Consortium defined 43,011 putative enhancers cross more than 800 human cell types and reported a strong cell type-specific enhancer activity(Andersson, Gebhard et al. 2014).

1.2 Acute Myeloid Leukemia

Acute myeloid leukemia (AML) is a group of hematological malignancies, whereby abnormal leukemic blast cells derived from the myeloid lineage go through clonal expansion in the bone marrow resulting in impaired normal bone marrow function(Liesveld 2015). Clinical signs are primarily a result of impairment of the production of normal functional blood cells and include pallor and dyspnea due to anemia, hemorrhages due to thrombocytopenia and increased frequency of infections, due to granulocytopenia and other immunocompromising conditions. AML is the most common acute leukemia in adults and but can also occur in pediatric patients(Gamis, Alonzo et al. 2013). Overall incidence of adult AML is in the range of 3–4 cases per 100,000 inhabitants in Sweden and median age at diagnosis is approximately 71 years(Juliusson, Antunovic et al. 2009). Despite treatment with intensive chemotherapy, median survival is less than 1 year and only a minority of the patients obtain a cure and a long-term(Juliusson, Abrahamsson et al. 2017).

1.2.1 Risk factors

In the majority of AML cases, no specific cause of AML development can be identified. However, environmental factors such as high dose radiation and benzene exposure are associated with an increased risk of AML development(Tsushima, Iwanaga et al. 2012; Liesveld 2015). Chemotherapeutic agents, including topoisomerase II inhibitors and alkylating agents, lead to an increased risk of developing AML, caused by exposure to mutagenic DNA damage (Park, Chi et al. 2013). AML cases that develop after treatment for previous malignant diseases are referred as therapy-related. Chronic hematological diseases can evolve into AML as the secondary disease, preceded by antecedent disorders, such as myelodysplastic syndromes(MDS), myeloid proliferative diseases(MPD) and chronic myeloid leukemia(CML). (Behm 2003; Liesveld 2015). AML may also develop from other nonmalignant diseases or inherited or congenital conditions such as Fanconi Anemia and Blooms Syndrome.

1.2.2 Classification of AML

Traditionally and historically, AML was classified according to the French-American-British Classification (FAB) where AML was subdivided into subclasses from M0 to M7 based on the morphological and cytochemical characteristics of bone marrow smears(Behm 2003). The WHO Classification of Myeloid Neoplasms was first

introduced in 2002, then updated in 2008, and very recently in 2016(Vardiman, Harris et al. 2002; Wandt, Haferlach et al. 2010; Arber, Orazi et al. 2016). This new classification system distinguished AML subclasses by their genetic characteristics, morphological features as well as clinical parameters and other background information such as if the patient has an antecedent hematological disorder or therapy-related AML(Table1). Overall, the diagnosis of AML is still primarily based on blast counts in the bone marrow. Cases with myeloid blasts exceeding 20% are sufficient to warrant a diagnosis of AML. However, lower blast counts can be confirmed as AML when the translocation t(15:17), t(8:21), or inv(16) is identified. Other than cytogenetic features, molecular genetic events of nucleophosmin1 (*NPM1*) mutation, biallelic mutation of CCAAT/enhancer-binding protein α (*CEBPA*), and patient with mutated Runt-related transcription factor 1(*RUNX1*) are incorporated as separate entities.

Over the past 15 years, with the development of high-through-put sequencing techniques, the knowledge of the genetic changes in AML has grown significantly(Network 2013; Papaemmanuil, Gerstung et al. 2016). Several further somatic mutations have been discovered as recurrent events in AML and show evidence as important regulators of disease and treatment progression in experimental models.

1.2.3 Prognostic factors

Factors such as age, karyotype, and molecular genetic features are used to assess the patient's prognosis and to choose therapeutic strategies and of AML, especially the decision to perform a hematopoietic stem cell transplantation (HSCT). In general, increasing age and coexisting health conditions are associated with poorer clinical outcomes and often treatment-related early death(Grimwade and Hills 2009; De Kouchkovsky and Abdul-Hay 2016). Based on both cytogenetic and molecular factors, AML patients can be divided into favorable, intermediate, and adverse outcome groups(Dohner, Estey et al. 2017). More than half of adult AML cases carry chromosomal arrangements, which significantly contribute to prognosis and clinical decision-making(De Kouchkovsky and Abdul-Hay 2016).

Table 1. WHO classification of acute myeloid leukemia and related myelodysplasia and neoplasm 2016*

AML with recurrent genetic abnormalities
<p>AML with t(8;21)(q22;q22.1);RUNX1-RUNX1T1</p> <p>AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22);CBFB-MYH11</p> <p>APL with PML-RARA^a</p> <p>AML with t(9;11)(p21.3;q23.3);MLLT3-KMT2A^b</p> <p>AML with t(6;9)(p23;q34.1);DEK-NUP214</p> <p>AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2, MECOM</p> <p>AML (megakaryoblastic) with t(1;22)(p13.3;q13.3);RBM15-MKL1^c</p> <p><i>Provisional entity: AML with BCR-ABL1</i></p> <p>AML with mutated NPM1^d</p> <p>AML with biallelic mutations of CEBPA^d</p> <p><i>Provisional entity: AML with mutated RUNX1</i></p>
AML with myelodysplasia-related changes^e
Therapy-related myeloid neoplasms^f
AML, nonotherwise specified (NOS)
<p>AML with minimal differentiation</p> <p>AML without maturation</p> <p>AML with maturation</p> <p>Acute myelomonocytic leukemia</p> <p>Acute monoblastic/monocytic leukemia</p> <p>Pure erythroid leukemia</p> <p>Acute megakaryoblastic leukemia</p> <p>Acute basophilic leukemia</p> <p>Acute panmyelosis with myelofibrosis</p>
Myeloid sarcoma
Myeloid proliferations related to Down syndrome
<p>Transient abnormal myelopoiesis (TAM)</p> <p>Myeloid leukemia associated with Down syndrome</p>
Blastic plasmacytoid dendritic cell neoplasm
Acute leukemias of ambiguous lineage
<p>Acute undifferentiated leukemia</p> <p>Mixed phenotype acute leukemia (MPAL) with t(9;22)(q34.1;q11.2); BCR-ABL1^h</p> <p>MPAL with t(v;11q23.3); KMT2A rearranged</p> <p>MPAL, B/myeloid, NOS</p> <p>MPAL, T/myeloid, NOS</p>

for a diagnosis of AML, a marrow blast count of $\geq 20\%$ is required, except for AML with the recurrent genetic abnormalities t(15;17), t(8;21), inv(16) or t(16;16).

- a. Other recurring translocations involving RARA should be reported accordingly: e.g., AML with t(11;17)(q23;q12); ZBTB16-RARA; AML with t(11;17)(q13;q12); NUMA1-RARA; AML with t(5;17)(q35;q12); NPM1-RARA; or AML with STAT5B-RARA (the latter having a normal chromosome 17 on conventional cytogenetic analysis).
- b. Other translocations involving KMT2A (MLL) should be reported accordingly: e.g., AML with t(6;11)(q27;q23.3); MLLT4-KMT2A; AML with t(11;19)(q23.3;p13.3); KMT2A-MLLT1; AML with t(11;19)(q23.3;p13.1); KMT2A-ELL; AML with t(10;11)(p12;q23.3); MLLT10-KMT2A.
- c. Rare leukemia most commonly occurring in infants.
- d. Diagnosis is made irrespective of the presence or absence of multilineage dysplasia.
- e. $\geq 20\%$ blood or marrow blasts AND any of the following: previous history of myelodysplastic syndrome (MDS), or myelodysplastic/myeloproliferative neoplasm (MDS/MPN); myelodysplasia-related cytogenetic abnormality (see below); multilineage dysplasia; AND absence of both prior cytotoxic therapy for unrelated disease and aforementioned recurring genetic abnormalities; cytogenetic abnormalities sufficient to diagnose AML with myelodysplasia-related changes are:
 - Complex karyotype (defined as 3 or more chromosomal abnormalities in the absence of one of the WHO-designated recurring translocations or inversions, i.e., t(8;21), inv(16) or t(16;16), t(9;11), t(v;11)(v;q23.3), t(6;9), inv(3) or t(3;3); AML with BCR-ABL1);
 - Unbalanced abnormalities: -7 or del(7q); -5 or del(5q); i(17q) or t(17p); -13 or del(13q); del(11q); del(12p) or t(12p); idic(X)(q13);
 - Balanced abnormalities: t(11;16)(q23.3;p13.3); t(3;21)(q26.2;q22.1); t(1;3)(p36.3;q21.2); t(2;11)(p21;q23.3); t(5;12)(q32;p13.2); t(5;7)(q32;q11.2); t(5;17)(q32;p13.2); t(5;10)(q32;q21.2); t(3;5)(q25.3;q35.1).
- f. Cases should be classified with the related genetic abnormality given in the diagnosis.
- g. The former subgroup of acute erythroid leukemia, erythroid/myeloid type ($\geq 50\%$ bone marrow erythroid precursors and $\geq 20\%$ myeloblasts among non-erythroid cells) was removed; myeloblasts are now always counted as percentage of total marrow cells. The remaining subcategory AML, NOS, pure erythroid leukemia requires the presence of $>80\%$ immature erythroid precursors with $>30\%$ proerythroblasts.
- h. BCR-ABL1 positive leukemia may present as mixed phenotype acute leukemia; treatment should include a tyrosine kinase inhibitor.

*Reprint with permission from original publication by Arber D. *et al.* Blood,2016.

Recently, European LeukemiaNet has revised the risk stratification of adult AML in which six well-studied genes (*NPM1*, *FLT3-ITD*, *RUNX1*, *CEBPA*, *ASXL1*, *TP53*) have been taken into consideration in clinical practice for prognosis (Table 2)(Dohner, Estey et al. 2017). Notably, among the risk group proposed by LeukemiaNet, a large proportion of the patients have a so-called **cytogenetically normal AML (CN-AML)**. CN-AML is a subgroup that constitutes about 40% of adult AML cases and where a karyotypic analysis of the chromosomes of the leukemia cells do not show any abnormalities(Klepin, Rao et al. 2014). Although, with the most recent updates of risk assessments, where new mutations have been added that can help to prognostically stratify some additional CN-AML patients, there is still a lack of prognostic markers for intermediate-risk and CN-AML patients. This indicates a further need for information and factors that can help to accurately diagnose and prognostically assess AML patients.

Table 2. Risk Assessment of Acute Myeloid Leukemia according to ELN 2016*

Risk Category	Genetic Abnormality
Favorable	t(8;21)(q22;q22.1); RUNX1-RUNX1T1
	inv(16)(p13.1;q22) or t(16;16)(p13.1;q22); CBFB-MYH11
	Mutated NPM1 without FLT3-ITD or with FLT3-ITD ^{low(c)}
	Biallelic mutated CEBPA
Intermediate	Mutated NPM1 and FLT3-ITD ^{high(c)}
	Wild type NPM1 without FLT3-ITD or with FLT3-ITD ^{low(c)} (w/o adverse risk genetic lesions)
	t(9;11)(p21.3;q23.3); MLLT3-KMT2A ^d
	Cytogenetic abnormalities not classified as favorable or adverse
Adverse	t(6;9)(p23;q34.1); DEK-NUP214
	t(v;11q23.3); KMT2A rearranged
	t(9;22)(q34.1;q11.2); BCR-ABL1
	inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); GATA2,MECOM(EV11)
	-5 or del(5q); -7; -17/abn(17p)
	Complex karyotyp ^e ,e monosomal karyotype ^f
	Wild type NPM1 and FLT3-ITD ^{high(c)}
	Mutated RUNX1 ^g
	Mutated ASXL1 ^g
	Mutated TP53 ^h

- Frequencies, response rates and outcome measures should be reported by risk category, and, if sufficient numbers are available, by specific genetic lesions indicated.
- Prognostic impact of a marker is treatment-dependent and may change with new therapies.
- Low, low allelic ratio (0.5); semi-quantitative assessment of FLT3-ITD allelic ratio (using DNA fragment analysis) is determined as ratio of the area under the curve (AUC) “FLT3-ITD” divided by AUC “FLT3-wild type”; recent studies indicate that acute myeloid leukemia with NPM1 mutation and FLT3-ITD low allelic ratio may also have a more favorable prognosis and patients should not routinely be assigned to allogeneic hematopoietic-cell transplantation.
- The presence of t(9;11)(p21.3;q23.3) takes precedence over rare, concurrent adverse-risk gene mutations.
- Three or more unrelated chromosome abnormalities in the absence of one of the World Health Organization-designated recurring translocations or inversions, i.e., t(8;21), inv(16) or t(16;16), t(9;11), t(v;11)(v;q23.3), t(6;9), inv(3) or t(3;3); AML with BCR-ABL1.
- Defined by the presence of one single monosomy (excluding loss of X or Y) in association with at least one additional monosomy or structural chromosome abnormality (excluding core-binding factor AML).
- These markers should not be used as an adverse prognostic marker if they co-occur with favorable-risk AML subtypes.
- TP53 mutations are significantly associated with AML with complex and monosomal karyotype.

* Reprint with permission from original publication by Döhner *et al.* Blood, 2016.

1.2.4 Molecular genetic changes in CN-AML

Although CN-AML is considered as negative for cytogenetic abnormalities by a clinical definition based on the karyotype, it displays a number of somatic mutations that have a role in the development of the disease (Welch, Ley et al. 2012; Miller, Wilson et al. 2013). During disease progression, founding cancerous clones may acquire additional mutations, forming subclones that contribute to secondary progression leading to relapses of AML (Genovese, Kahler et al. 2014; Yoshizato, Dumitriu et al. 2015). In one recent study, 76 somatic mutations were found to

recurrently occur in an AML cohort of over 1500 patients(Papaemmanuil, Gerstung et al. 2016). Notably, compared to other most other types of malignancies, the AML genome contain significantly fewer mutations in general(Miller, Wilson et al. 2013; Vogelstein, Papadopoulos et al. 2013). This highlights the importance of such mutations in relation to leukemic transformation and clonal evolution.

1.2.5 Commonly mutated genes in CN-AML

Nucleophosmin 1 (NPM1)

The *NPM1* gene encodes for a histone chaperone, located on chromosome 5q23. Mutations in the C-terminal of *NPM1* result in an impaired DNA binding function, therefore aberrantly exporting and translocating NPM1 into the cytoplasm(Grisendi, Mecucci et al. 2006). *NPM1* mutations are often found in association with mutations in the *DNMT3A* and *FLT3-ITD* genes in CN-AML(Papaemmanuil, Gerstung et al. 2016). The mutation predicts a favorable outcome in CN-AML in most of the age groups in the absence of *FLT3-ITD*(Dohner, Schlenk et al. 2005; Verhaak, Goudswaard et al. 2005; Becker, Marcucci et al. 2010; Schnittger, Bacher et al. 2011). The clinical value of *NPM1* mutations for detection of minimal residual disease has recently been validated and shown to be the only independent molecular factor for predicting death in this group of patients(Hills, Ivey et al. 2016).

Fms-related Tyrosine Kinase 3 (FLT3)

The *FLT3* gene encodes for a Class III tyrosine kinase receptor, expressed on the cell surfaces of hematopoietic progenitors. There are two types of mutations affecting the *FLT3* gene with distinctive functional implications. An internal tandem duplication of *FLT3* (*FLT3-ITD*) involves the juxtamembrane domain and occurs in nearly one-third of CN-AML patients(Rombouts, Lowenberg et al. 2001). It results in the constitutive activation of the tyrosine kinase, which consequentially leads to enhanced signaling through the RAS and STAT5 pathways(Neben, Schnittger et al. 2005; Chen, Drakos et al. 2010). There is evidence that the prognosis in patients with *FLT3-ITD* shows a dosage dependency of the mutated allele, where the presence of a high allelic burden (ratio of ITD/WT>0.5) is linked to a poorer prognosis of CN-AML. The other type of mutation affects *FLT3* at D835 and I836 of the tyrosine kinase domain (TKD) and this is referred to as *FLT3-TKD*. *FLT3-TKD* is found in 11-14% of CN-AML, but its presence remains debatable with regard to its prognostic impact(Whitman, Ruppert et al. 2008; Santos, Jones et al. 2011).

DNA Methyltransferase 3A (DNMT3A)

Somatic mutation in the *DNMT3A* gene in hematological malignancies was discovered and reported by several groups in 2010 and 2011 (Ley, Ding et al. 2010; Yan, Xu et al. 2011). It is found in 20-22% of total AML and with higher frequency in the normal karyotype group (Ley, Ding et al. 2010; Yan, Xu et al. 2011; Gaidzik, Schlenk et al. 2013). In total, 35 point mutations have been found in the *DNMT3A* gene up to the date (Yang, Rau et al. 2015). Among these mutations, 58% harbor a mutation at arginine position 882 (R882). The R882 mutation is predominantly heterozygous in the most of the hematological malignancies except in T-cell acute lymphoblastic leukemia (T-ALL), in which biallelic mutations frequently occur (Grossmann, Haferlach et al. 2013; Roller, Grossmann et al. 2013). Moreover, *DNMT3A* mutations are age related with increased frequency in elderly and associated with premalignant clonal expansion (Xie, Lu et al. 2014). Biochemical studies of the human DNMT3A with a mutation at position R882 as well as the mutated mouse equivalent R878 display an impaired catalytic activity and reduced DNA-binding efficiency comparing to wild type DNMT3A (Holz-Schietinger, Matje et al. 2012; Russler-Germain, Spencer et al. 2014). However, comparing to R882, much less attention has been drawn to non-R882 mutations. Few studies have found decreased methylation capacity in non-R882 mutations, which most likely leads to the loss of function of DNMT3A (Gowher, Loutchanwoot et al. 2006; Holz-Schietinger, Matje et al. 2011). However, despite the agreement of the high prevalence of *DNMT3A* mutations in AML, their impact on patients' clinical outcomes remains surprisingly inconclusive. It was first reported associated with an adverse prognosis in AML by Ley and his colleague, however, contradictory results have been published by Patel et al. and Gaidzik et al. (Ley, Ding et al. 2010; Patel, Gonen et al. 2012; Gaidzik, Schlenk et al. 2013). Very recently, a more comprehensive characterization of a large AML cohort suggests a more complex prognostic interaction between *NPM1*, *DNMT3A*, and *FLT3-ITD* mutations, where mutations in all three genes confer a poor survival (Papaemmanuil, Gerstung et al. 2016).

CCAAT/enhancer Binding Protein Alpha (CEBPA)

CEBPA is a transcription factor belongs to leucine zipper family that is essential for lineage specification and granulopoiesis (Radomska, Huettner et al. 1998). Mutation of *CEBPA* leads to insufficient activation of granulocytic specific genes and a maturation

arrest in the myeloid lineage. Among all sequence variants reported, two types of mutations occur more frequently and that often involve one allele each, the out-of-frame mutation at the N-terminal leading to a truncated protein that is dominant negative and the in-frame insertion/deletion at the bZip domain resulting in DNA binding defects(Pabst, Mueller et al. 2001; Carnicer, Lasa et al. 2008). Germline *CEBPA* mutations have been reported for familiar AML and somatic mutations are found in approximately 15% of CN-AML(Smith, Cavenagh et al. 2004; Green, Koo et al. 2010; Taskesen, Bullinger et al. 2011). Biallelic mutation in *CEBPA* confers a favorable prognosis for AML patients and has been incorporated into WHO classification since 2008.

Runx-related Transcription Factor 1(RUNX1)

RUNX1 is a transcription factor that regulates both primitive hematopoiesis during embryonic development and differentiation of blood cells in adults(de Bruijn and Dzierzak 2017). Animal studies of the *Runx1* knockout model revealed embryonic lethality due to inadequate fetal liver hematopoiesis. Mutations in *RUNX1* are reported in 6% to 18% of CN-AML with increasing frequency by age(Gaidzik, Bullinger et al. 2011; Greif, Konstandin et al. 2012; Gaidzik, Teleanu et al. 2016). In contrast to core binding factor leukemias that are often characterized by a translocation involving the *RUNX1* gene, non-translocation mutations in *RUNX1* has been found to be associated with a negative prognostic impact in AML patients(Schnittger, Dicker et al. 2011; Greif, Konstandin et al. 2012; Mandler, Maharry et al. 2012).

Isocitrate Dehydrogenase (IDH)

Mutations in the *IDH* gene family were discovered in 2009, and both family members, *IDH1* and *IDH2*, can be mutated in AML(Mardis, Ding et al. 2009; Marcucci, Maharry et al. 2010). The mutations are most frequently affecting *IDH1* at the arginine residue 132 (R132) while arginine 140 (R140) and arginine 172 (R172) are commonly mutated in the *IDH2* gene. In contrast to *IDH2*^{R140}, *IDH2*^{R172} is not associated with *NPM1* mutations and is found to have a distinct gene expression profile(Marcucci, Maharry et al. 2010). Mutations in *IDH1* and *IDH2* are often mutually exclusive as well as mutually exclusive with *TET2* mutations which suggest a functional convergence among these genes(Figueroa, Abdel-Wahab et al. 2010; Patel, Gonen et al. 2012).

Ten-Eleven-Translocation 2 (TET2)

Mutations of *TET2* have been found broadly associated with different myeloid malignant diseases including MDS, myeloproliferative neoplasm(MPN) as well as AML(Tefferi, Lim et al. 2009; Tefferi, Lim et al. 2009; Bowman and Levine 2017). Mutations of *TET2* are detected in between 23% to 28% of AML patients with a slightly higher frequency in CN-AML(Tefferi, Lim et al. 2009; Papaemmanuil, Gerstung et al. 2016). They affect multiple exons and hotspots have not been clearly observed. *TET2* mutation is age related and is associated with clonal hematopoiesis in elderly individuals(Xie, Lu et al. 2014; Bowman and Levine 2017). No prognostic impact has been reported for AML patients in more recent publications despite lower rates of complete remission and shorter overall survival found by some earlier studies(Chou, Chou et al. 2011; Gaidzik, Paschka et al. 2012; Weissmann, Alpermann et al. 2012).

Wilms' Tumor 1 (WT1)

The *WT1* gene is located on chromosome 11p13 and encodes for a transcription factor that is essential for urogenital development(Yang, Han et al. 2007). Overexpression of WT1 has been known since long to be overexpressed in hematological malignancies including MDS, acute lymphoid and myeloid leukemia, as well as CML with blast crisis(Miyagi, Ahuja et al. 1993; Menssen, Renkl et al. 1995; Tamaki, Ogawa et al. 1999; Barragan, Cervera et al. 2004). In MDS, the elevated WT1 expression is associated with a higher blast count and an increased risk of progression to secondary AML. It is also associated with a poor overall survival and a higher incidence of relapses in AML patients(Miyagi, Ahuja et al. 1993; Tamaki, Ogawa et al. 1999). Interestingly, along with increased gene expression, mutations in *WT1* was initially discovered in nephroblastoma in pediatric patients and as first identified in AML in 1996(King-Underwood, Renshaw et al. 1996). Somatic mutations of *WT1* recurrently occur in approximately 10% of AML patients with a slightly higher incidence in CN-AML(Barragan, Cervera et al. 2004; Network 2013; Papaemmanuil, Gerstung et al. 2016). The mutations of the *WT1* gene involve primarily exons 1, 7, and 9 and mainly results in a loss of function caused by either a truncated protein, lacking zinc finger domains, or a premature stop codon(Hou, Huang et al. 2010). Yet frequently overexpressed, *WT1* may function both as a tumor suppressor and an oncogene(Yang, Han et al. 2007). The mechanisms and the role of the paradoxical aberrations in WT1,

including both overexpression of the wild type protein as well as loss of function mutations remains to be elucidated.

Additional Sex Combs Like 1 (ASXL1)

ASXL1 is the human homolog of *Drosophila* *Asx*, which is a polycomb group (PcG) associated protein that acts on transcriptional regulation (Fisher, Berger et al. 2003). By interacting with PcG complexes, it plays an important role in regulating histone modifications and homeotic gene expression (Abdel-Wahab and Dey 2013). Mutations of *ASXL1* are more frequently seen in myelomonocytic leukemia and MDS but are found in approximately 6–16% of AML patients with increasing frequency in older patients (Boultonwood, Perry et al. 2010). The vast majority of the *ASXL1* mutations involve exon 12, leading to a truncated C-terminal, losing the NHR binding domain and the PHD domain. It is often heterozygous and probably dominantly negative when forming interacting complexes. The studies of *ASXL1* mutations in myeloid malignancies have shown that the mutations are mediating *HOX* gene repression by H3K27me3 through cooperation with PRC2 complex (Gelsi-Boyer, Trouplin et al. 2009; Abdel-Wahab, Adli et al. 2012). MDS patients with *ASXL1* mutations have a poorer clinical outcome and shorter time to progression to AML (Thol, Friesen et al. 2011). Among AML patients, *ASXL1* mutations are associated to an adverse prognosis (Metzeler, Becker et al. 2011; Paschka, Schlenk et al. 2015).

1.3 Epigenetic mechanisms in hematopoiesis and AML

1.3.1 Hematopoiesis

Hematopoiesis is the developmental process by which hematopoietic stem cells produce differentiated blood cells (Jagannathan-Bogdan and Zon 2013). Two major lineages exist: the myeloid and lymphoid (Figure 2). Myelopoiesis starts with the common myeloid progenitor (CMP) and generates megakaryocytes, erythrocytes, mast cells, and mature granulocytes including neutrophils, basophils, and eosinophils. Meanwhile, T cells, B cells, natural killer cells, and lymphoid dendritic cells are produced from the common lymphoid progenitor (CLP). Lineage choices are thought to depend on growth factor signals, which lead to the upregulation of cell type-specific genes in tandem with the repression of paternal pluripotent genes. In hematopoiesis during fetal development, GATA1 and PU.1 are the two key transcription factors that regulate erythroid-myeloid fates by cross-inhibitory mechanisms (Ferreira, Ohneda et al. 2005; Chou, Khandros et al. 2009). In adult life, RUNX1 is known for its essential role in the regulation of hematopoietic stem cells (Zhu and Emerson 2002). Early decisions during myeloid/lymphoid commitment are also regulated by transcription factors. For instance, C/EBP α , GATA1, and PU.1 are crucial for generating CMP and support further myeloid differentiation, whereas IL-7 receptor is highly expressed in CLP but absent in CMPs (Schlenner, Madan et al. 2010; Ohlsson, Schuster et al. 2016).

1.3.2 Epigenetic mechanisms in normal hematopoiesis

During these developmental stages of hematopoiesis, DNA methylation levels change dynamically. The lymphoid lineage somewhat gains methylation during differentiation, but myeloid and erythroid development is associated with a significant DNA demethylation globally (Ji, Ehrlich et al. 2010; Hodges, Molaro et al. 2011; Shearstone, Pop et al. 2011). At promoter level, methylation changes of lineage-specific genes lead to transcriptional activation during blood cell differentiation (Calvanese, Fernandez et al. 2012). During the myeloid-lymphoid lineage choice, DNA methylation was found to regulate the activation of lineage-specific genes and the repression of transcription factors from other lineages (Hodges, Molaro et al. 2011). Meanwhile, the increased DNA methylation of myeloid transcription factor binding sites of GATA1, RUNX1, and LMO2 was found in CLP cells, suggesting that DNA methylation also facilitates the modulating sensitivity to differentiation signaling. It can be further exemplified by DNA methylation of enhancers during granulopoiesis, that major difference is found

corresponding to lineage restriction from GMP to PMC step(Ostronoff, Othus et al. 2015).

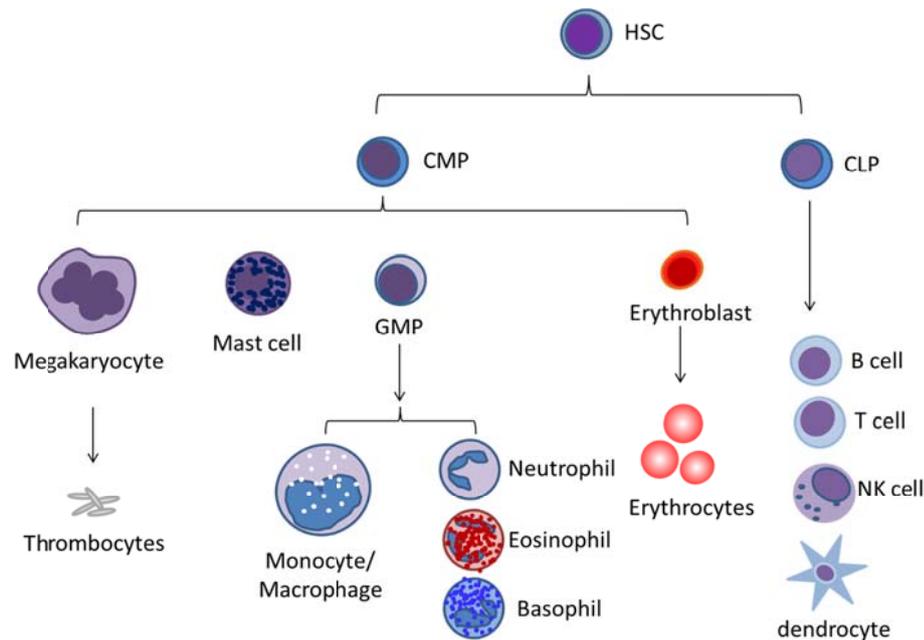


Figure 2. Schematic differentiation steps of hematopoiesis. All blood cells derive from a hematopoietic stem cell (HSC) and generate two major lineages. It gives rise to both common myeloid progenitor (CMP) cells and common lymphoid progenitor (CLP) cells. Mature granulocytes, including neutrophils, eosinophils, and basophils are differentiated from granulocyte/macrophage progenitor (GMP) cells, which also gives rise to monocytes and macrophages. During granulopoiesis, GMP differentiates into myeloblasts, promyelocytes, myelocytes, metamyelocytes, band cell, and to the end mature to neutrophils.

The importance of DNA methylation in regulating hematopoiesis is demonstrated by experimental models with genetically modified DNMTs in HSCs. All three DNMTs are expressed constitutively through hematopoietic development. DNMT3B is highly expressed in CD34⁺ stem/progenitor cells and downregulated following cell differentiation, meanwhile, DNMT3A is more ubiquitously expressed throughout the stages of differentiation(Mizuno, Chijiwa et al. 2001). Loss of DNMT3B in HSCs produces a mild phenotype; however, loss of DNMT3A in HSC results in enhanced self-renewal and impaired differentiation capacity(Challen, Sun et al. 2014; Mayle, Yang et al. 2015). Whereas loss of de methyltransferase DNMT1 has shown to disrupt long-term stemness and the multipotency of HSCs(Broske, Vockentanz et al. 2009; Trowbridge, Snow et al. 2009). Notably, in these studies, reduced expression of DNMT1 in hypomorphic mice blocked differentiation of the lymphoid lineage but not myeloid cells. Moreover, HSCs from the *Dnmt1*-hypomorphic model displayed less

methylation and elevated expression of myeloerthyroid signature genes such as *GATA1* and *CEBPA*. This represents an excellent example of how DNA methylation is involved in lineage-specific regulation during hematopoiesis.

Other epigenetic mechanisms, such as histone modifications, also correspond to these types of lineage differential signals. The differentiation from CMP to erythroid or myeloid cells is coupled with HDAC1 expression by upstream signaling of GATA1 and CEBPA(Wada, Kikuchi et al. 2009). Lineage-specific genes such as *PAX5* and *GATA3* are poised with bivalent histone marks (H3K4me3 and H3K27me3) in hematopoietic progenitor cells and are associated with increased levels of H3K4me1 and H2A.Z upon differentiation(Cui, Zang et al. 2009; Abraham, Cui et al. 2013).

More recent genome-wide analyses of DNA methylation have suggested that lineage- and cell type-specific methylation changes occur more frequently in CGI proximal regions (CGI shores) rather than in CGIs themselves. Also, a stronger correlation to gene expression has been suggested as a result of methylation changes CGI shores(Irizarry, Ladd-Acosta et al. 2009; Shearstone, Pop et al. 2011). Since then, the focus of DNA methylation studies has expanded from a previous focus on methylation changes in CGIs to other genomic areas such regions distal to TSSs and in gene bodies. Notably, DNMT3B may contribute to changes in intragenic methylation and the interaction with other epigenetic modifiers(Weisenberger, Velicescu et al. 2004; Duymich, Charlet et al. 2016). It has been shown that one isoform of DNMT3B lacks the catalytic domain but that is able to recruit DNMT3A, mediating gene body methylation in relation to H3K36me3.

On the other hand, demethylation is also crucial for hematopoietic development. Disruption of *Tet2* by Cre-mediated deletion of exon 3 resulted in enhanced proliferation and self-renewal of HSC and differentiation towards the myeloid lineage(Moran-Crusio, Reavie et al. 2011). *Tet*^{-/-} mice developed multiple myeloid malignancies that resemble conditions with recurrent mutations of *TET2* in humans(Li, Cai et al. 2011). All of these consequences coincide with a substantial loss of 5hmC and an increase of 5mC, especially at lineage-specific genes. It has been suggested that TET2 may also respond to modulation of enhancer activity of key lineage-specific genes. A recent study of *Dnmt3a* and *Tet2* double-knockout mice suggests that these genes cooperate in repressing HSC genes and promote erythroid-specific genes (such as

Klf4) during hematopoiesis(Zhang, Su et al. 2016). Moreover, the TET2 protein also interacts with transcription factors such as PU.1 together and DNMT3B, regulating the differentiation of monocyte to osteoclast (de la Rica, Rodriguez-Ubreva et al. 2013).

1.3.3 Aberrant epigenetic changes in AML

In hematological malignancies such as AML, whole genome and exome sequencing have revealed several classes of recurrently mutated genes, of which mutations in epigenetic modulators have attracted a special interest. Recurrent mutations have been identified in DNA methylation regulators (*DNMTs*, *TET2*, *IDHs*), chromatin modification regulators (*MLL*, *ASXL1*, *EZH2* etc.) as well as in cohesion complex components(Network 2013; Papaemmanuil, Gerstung et al. 2016). It should be noted that compared to other frequently mutated genes such as *FLT3*, mutations affecting epigenetic mechanisms occur significantly earlier during clonal evolution of AML and are stable during relapse. Mutations in epigenetic regulators are often mutually exclusive with gene fusions involving transcription factors(Network 2013; Faber, Chen et al. 2016). This suggests that mutations in epigenetic factors may constitute distinct pathogenic events that are complementary to the direct disturbance of lineage transcription factor signaling.

In line with the frequent mutations found in DNA methylation regulating genes, aberrant DNA methylation has been extensively studied and reported in AML. In general, the AML methylome shows the decreased level of methylation globally but also hypermethylation at CGI containing promoter regions, typically affecting tumor suppressor genes, as such it follows a similar pattern as compared to other cancer types(Deneberg, Grovdal et al. 2010; You and Jones 2012). Genome-wide methylation signatures correlate to the patients' cytogenetic and genetic subtypes suggesting a biological and pathological relevance. The first methylation profiling of a large AML cohort was published by Figueroa et al. in 2010 and demonstrated an important link between aberrant DNA methylation and known genetic lesions that drive leukemogenesis(Figueroa, Lugthart et al. 2010). It is worth mentioning that the recognition of such methylation alterations was described before the discovery of recurrent mutations in DNA methylation regulators.

Although some chromosomal rearrangements are discordant with mutations involving epigenetic mechanisms, AML with recurrent gene fusions such as *RUNX1-RUNX1T1*,

PML-RARA, *CBF-MYH11*, as well as fusions involving *MLL*, still show distinct methylation patterns (Figueroa, Lugthart et al. 2010; Network 2013; Conway O'Brien, Prideaux et al. 2014). The mechanisms of how these translocations alter DNA methylation remain unclear. A recent study shows that *DNMT3A* is dispensable for leukemic blasts with *RUNX1-RUNX1T1* or *MLL-AF9* but that it is required for APL transformation due to *PML-RARA* rearrangements (Cole, Verdoni et al. 2016). *MLL* is a histone methyltransferase and frequently affected by translocations and partial tandem duplication (PTD) (Krivtsov and Armstrong 2007). It is known that *MLL* fusion proteins, compared to the wild type *MLL*, are misguided to aberrant target regions, such as *HOXA9* and *MEIS1* loci which lead to transcriptional activation by catalyzing local H3K4me2 and preventing CpG methylation, resulting in impairment of normal hematopoietic differentiation (Wang, Lin et al. 2009).

Notably, almost all regulators of DNA methylation are recurrently mutated in AML. Mutation of *DNMT3A* results in a dominant negative protein that out-competes the formation of the most catalytically active homotetramer of wild-type *DNMT3A* and this change decreases its methylation activity by 80% on DNA strands (Holz-Schietinger, Matje et al. 2012; Kim, Zhao et al. 2013; Russler-Germain, Spencer et al. 2014). Logically, a hypomethylated signature has been reported in hematological malignancies with mutated *DNMT3A* (Hajkova, Markova et al. 2012; Russler-Germain, Spencer et al. 2014; Xu, Wang et al. 2014). It is known that loss of *DNMT3A* results in large hypomethylated canyons close to stem cell specific genes such as the Homeobox family (Jeong, Sun et al. 2014). Interestingly, *Dnmt3a* null mice develop both myeloid and lymphoid malignancies, however, mice transplanted with *Dnmt3a*^{+/-} HSC, generates only myeloid cancers (Challen, Sun et al. 2014; Mayle, Yang et al. 2015). It is in line with the observation that biallelic *DNMT3A* mutations are found in T-cell acute leukemia but not in AML (Grossmann, Haferlach et al. 2013). These findings highlight the pathological relevance of impaired DNA methylation as a transforming event in leukemogenesis. Nevertheless, demethylation pathways are also frequently disrupted in AML. Increased 5mC together with decreased 5hmC are significant findings in both *IDH* and *TET2* mutated AML (Figueroa, Abdel-Wahab et al. 2010). Although *TET2* and *DNMT3A* functions are thought to counteract each other, a recent study suggests that they may work in cooperation in regulating HSC homeostasis and mutations of both genes induce T-ALL (Scourzic, Couronne et al. 2016; Zhang, Su et al. 2016).

As discussed in the previous sections, DNA methylation interacts with chromatin modifications, and thus, mutations in chromatin modifiers crosstalk with DNA methylation and histone modifications in inducing aberrant gene transcription patterns. For example, the PRC2 core component EZH2 is known to interact with DNMTs, mediating promoter DNA methylation of polycomb targets(Vire, Brenner et al. 2006). Mutations of *EZH2* can be either gain-of-function or inactivating mutations and interestingly, both these types of mutations can be found in AML. It has been shown that *AXSL1* mutations are associated with a decrease in H3K27me3 and may act through an EZH2-dependent mechanism, resulting myeloid expansion(LaFave, Beguelin et al. 2015). A recent study showed that AXSL1 is an important functional partner to the cohesion complex by frequently sharing binding sites, regulating telophase chromatid disjunction in hematopoietic cells(Li, Zhang et al. 2017). To note, mutations in cohesin complex also recurrently occur in AML(Network 2013).

Conspicuously, these gene mutations are often found in hematologically healthy elderly individuals. Mutations of *DNMT3A*, *TET2*, and *AXSL1* are among the most frequent somatic mutation events reported in age-related clonal hematopoiesis and this is significantly associated with an increased risk of developing hematological malignancies including AML(Genovese, Kahler et al. 2014; Xie, Lu et al. 2014; Kwok, Hall et al. 2015). All these findings suggest that alterations in epigenetic mechanisms are of particular importance in hematological malignancies and that they likely contribute to the transformation of normal hematopoietic progenitors, leading to the development of leukemia.

2. Aim of The Thesis

The aim of this thesis is to explore epigenetic changes in cytogenetically normal AML with a particular focus on alterations of DNA methylation. In addition, to correlate the aberrations to clinical and biological characteristics.

Paper I

To characterize CGI focused aberrant DNA methylation in cytogenetically normal AML and study its role in relation to prognostic outcome.

Paper II

To study the genome-wide aberrant DNA methylation in CGI as well as in CGI distal regions in cytogenetically normal AML in relation to genetic mutations and gene expression.

Paper III

To identify AML specific changes in DNA methylation of enhancer elements and characterize their associations with changes in histone modification and gene expression.

3. Methodological Approaches

In this thesis, we have used multidisciplinary approaches to address the aberrant epigenome in CN-AML and their potential impact on the transcriptome and the clinical outcome of the patients. Methods including Illumina® Methylation Arrays, pyrosequencing, DNase I hypersensitivity site sequencing (DHS-seq), chromatin immunoprecipitation and sequencing (ChIP-seq), RNA sequencing and CRISPR-Cas9 for genomic editing, are discussed in this chapter. Comprehensive experimental protocols are described in detail in the methods and material sections for each publication.

3.1 DNA methylation detection

Papers included in this thesis are primarily focusing on aberrant DNA methylation in CN-AML. The discovery of bisulfite treatment has opened the possibility of profiling methylation patterns in targeted regions and on a genome-wide level. In recent years, with advances of microarray platforms such as the Illumina arrays and next generation sequencing technologies, methylation patterns can now be detected at single CpG level throughout the whole genome. In this thesis, two versions of the Illumina human methylation arrays (27K and 450K) have been used. In addition, the pyrosequencing of bisulfite-converted DNA has also been extensively used for locus-specific methylation analyses.

3.1.1 Bisulfite conversion

To detect methylation in the genome, technologies must allow recognition of cytosine modifications and quantification of their frequency either globally or site specifically. The degree of methylation in the DNA sequence can be detected by sequencing after bisulfite conversion treatment (Hayatsu 2008). Sodium bisulfite chemically modifies cytosine (C), converting it to uracil (U) through deamination. However, during bisulfite exposure, cytosine with 5' modifications, including methylation and hydroxymethylation, remain unchanged. Based on this process, the methylation level of the given cytosine position can be detected by analyzing single nucleotide polymorphisms between C and T after PCR amplification. This provides a technical base for several methods including pyrosequencing, methylation specific PCR, hybridization-based microarray methods, etc. (Shapiro R. 1970; Sasaki, Anast et al. 2003). The common disadvantage of bisulfite conversion-based methods is its inability

to distinguish 5hmC from 5mC signals with a standard protocol. With recent advances, additional oxidation treatment can successfully overcome this limitation (Song, Szulwach et al. 2011).

3.1.2 Pyrosequencing

Based on the "sequencing by synthesis" principle, pyrosequencing detects the order of nucleotides of the desired DNA sequence (Nyren 2015). It uses single-stranded DNA as the template after hybridizing with the sequencing primer at the target site. A correct incorporation of one of the four deoxynucleoside triphosphates (dNTPs) of the complementary strand will release pyrophosphate (PPi) catalyzed by the DNA polymerase. This PPi will generate an ATP molecule that gives a visible light under catalyzation of a luciferase. Light signals from this "synthesis" procedure will be

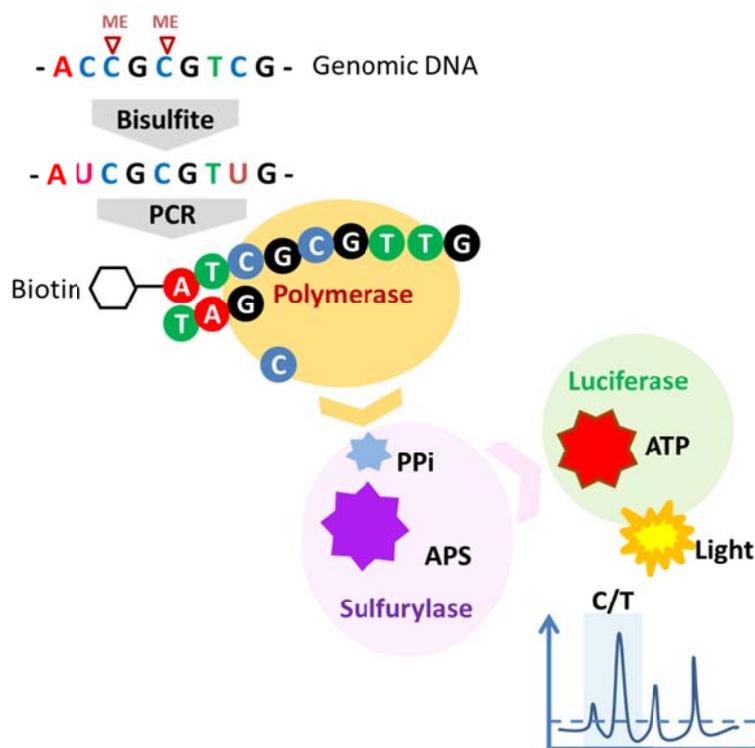


Figure 3. DNA methylation analysis by Pyrosequencing. Unmethylated cytosine(C) is converted to uracil(U)through bisulfite conversion and further translate to thymidine(T) by PCR. Sequencing DNA template is purified and annealed with pyrosequencing primer. DNA polymerase catalyzes the elongation of a synthetic strand and releases pyrophosphate (PPi) molecule. ATP Sulfurylase uses this PPi together with adenosine 5'-phosphosulfate (APS) generates ATP molecule that further facilitates light signals by a luciferase-catalyzed reaction. The light signals are proportional to the ATP productions, therefore used to estimate the incorporated dNTPs. The pyrogram represents the signal histogram at each position. For CpG site, incorporation of C or T indicates the methylated or unmethylated cytosine.

captured by a camera that will give the final readout in a pyrogram. For detection of DNA methylation, PCR amplification of the target of interest after bisulfite conversion produces a single-stranded DNA template linked to biotin. At the site of the CpG dinucleotide, the ratio of light signals is proportional to the incorporated C/T (Figure3). The method of bisulfite conversion followed by pyrosequencing has been extensively used in all three papers for locus-specific methylation detection.

3.1.3 Illumina methylation arrays

The Illumina Methylation Arrays, including the 27K and the 450K array (referred to as 27K and 450K), are probe-based array platforms designed to cover genome-wide CpG sites with two different resolutions (Bibikova, Lin et al. 2006). Illumina 27K is an earlier version that contains only Infinium I assay for more than 27,000k probes, exclusively targets CGIs. Its update, the Illumina 450K, contains more than 480,000 probes of both Infinium I and Infinium II types of assays extending to more CpG-sparse regions and regulatory elements in the human genome (Roessler, Ammerpohl et al. 2012). Both assays require bisulfite conversion of genomic DNA followed by whole genomic amplification. Successful single nucleotide extension with labeled dideoxynucleotides results in incorporation of fluorescence signals that are captured and methylation levels which are computed from fluorescence intensities. For Infinium I, a pair of probes is designed to target the same locus for either methylated (end with CG) or unmethylated allele (ended with CA). In Infinium II, a single probe ending with an open position (ending with C) is targeting the cytosine of the CpG site of interest. Therefore, incorporation of either G or A at the next base determines the methylation status for the designated locus.

Two types of values from the Illumina Methylation Arrays have been used in various publications, including the papers in this thesis, the β -value and the M-value (Marabita, Almgren et al. 2013). The β -value of each probe is computed as a methylation signal versus the sum of methylated and unmethylated signals. It ranges from 0, for a completely unmethylated site, to 1 for a fully methylated site. The M-value is the log₂ transformed ratio of signal from a methylated site versus unmethylated site. There are pros and cons for these two types of methylation estimations. For example, compared to the M-value, the β -value presents a more intuitive interpretation the β -value may provide an easier solution for direct comparisons between studies. Due to its logarithmic scale, the M-value presents as a typical bimodal distribution which is

difficult to directly link to the degree of methylation but gives a better statistical applicability(Bibikova, Lin et al. 2006). Bioinformatic validations have shown that the profound homoscedasticity of the M-value provides a better base for statistical modeling of a differential methylation analysis. Both of these two types of estimations are generally accepted, and they are both widely used in various studies. In this thesis, methylation profiles patients samples and of normal counterparts were analyzed by the Illumina arrays and in the paper I and paper II the estimations are based on β -value whereas, in the paper III, the M-value was used.

3.1.4 Other genome-wide methylation platforms

Whole genome bisulfite sequencing(WGBS) is the most comprehensive method for methylome profiling, however, the method requires extensive efforts. Due to the still high cost of next generation sequencing, the theoretical "whole genome" is often represented by methylation analyses of enriched sequences, namely reduced representative bisulfite sequencing (RRBS)(Meissner, Gnirke et al. 2005). This technique uses methylation-insensitive restriction enzymes(such as MspI) with a combination of fragments size selection(often 40bp-200bp) to yield CpG containing sequences that are sequenced in the next step. Thus, RRBS is effective for moderate/high CpG content regions but less informative for CpG sparse regions such as regions distal to promoters. Similarly, methylated DNA immunoprecipitation(MeDIP), which utilizes an antibody against 5mC to pull down DNA fragments with methylated cytosine that are then subjected to sequencing, also results in an uneven coverage of pulled-down genomic regions due to differences in CpG density and antibody affinity(Jacinto, Ballestar et al. 2008).

Although the 450K array has limitations with respect to the coverage of the genome, due to the easier bioinformatic pipeline, the lower cost and the probe-based design, it offers some advantages and can determine DNA methylation that occurs not only in CpG dense regions but also in sites distal to CGIs and promoters.

3.2 Chromatin accessibility and modification analysis

The chromatin is centered with the histone core complex and the DNA that is wrapped around it. The density of nucleosomes and the modifications on the histone tails affect the availability of particular DNA sequences that results in different levels of binding accessibility and recruitment of DNA binding proteins and protein

complexes(Marabita, Almgren et al. 2013). DNase I hypersensitivity sites (DHSs) refer to accessible regions that are sensitive to DNase I catalyzed cleavage in the genome. DHSs often relate to cis-regulatory elements such as promoters, enhancers and insulators(Sabo, Kuehn et al. 2006). Other than being sensitive for DNase I mediated cleavage, these regions are also characterized by specific histone modifications(Aparicio, Geisberg et al. 2004). As a result of the advances in the technique of chromatin immunoprecipitation followed by massive parallel sequencing(ChIP-seq), this now allows for global analyses of these histone codes, and together with detection of DHSs, it maps the chromatin status and status of gene regulation process.

3.2.1 DNase I hypersensitivity site and sequencing

DNase hypersensitivity site sequencing (DHS-seq) is a tool that combines DNase digestion and next generation sequencing to map highly accessible regions at a genome-wide level(Berger 2007). After digesting the chromatin with DNase I, exposed DNA strands are cleaved into small fragments, whereas tightly packed nucleosomes will remain intact. These sites can be detected with a PCR-based method for a specific locus of interests, or map DHSs on the genome-wide scale by sequencing. A short read sequencing library is generated after a biotinylated linker sequence has been added to the DNase I digested ends of DNA fragments. In order to describe the degree of accessibility of a sequenced regions, the reads per million (RPM) value is commonly applied for this estimation(Li, Carey et al. 2007). In paper III, DNase-seq data from the ENCODE project and a previously published paper have also been integrated(Birney, Stamatoyannopoulos et al. 2007) (Rosenbloom, Dreszer et al. 2012).

3.2.2 Chromatin immunoprecipitation and sequencing

Chromatin immunoprecipitation and sequencing(ChIP-Seq) is a method to study the interactions between a DNA binding protein and the DNA sequence that is bound to the protein(Kumar, Muratani et al. 2013). Using an antibody against the protein of interests, affinity-based chromatin pulling-down will extract the DNA regions that are bound to the protein. This DNA is then identified by massive parallel sequencing after being transformed into a cDNA sequencing library. In ChIP-Seq analysis, protein-DNA interactions are preserved by fixation, usually by paraformaldehyde, resulting in reversible cross-linking. Following the cross-linking, chromatin will be sheared by sonication in order to yield the fragments. Often, magnetic beads are linked to the

antibody against the protein of interest, such as a transcription factor or a histone modification and this is used to separate the target chromatin fragments. After denaturation, DNA sequences are released and purified. These DNA fragments are ligated to an adapter sequence and sometimes together with an additional barcode, and then amplified before being analyzed by a sequencer. ChIP-Seq generates a reading that covers the genomic regions that interact with the protein or histone modification at interest. ChIP-Seq represents advancement from the ChIP-on-chip technology, which is a microarray-based methodology that also provides a genome-wide view of protein-DNA interactions. To analyze datasets generated from ChIP-seq experiments, peak-calling is often needed. This uses computational algorithms to define the regions with signals (reads) over the whole background. In ChIP-Seq, sensitivity can be determined by sequencing depth, whereas the specificity of ChIP results critically depends on the chosen antibody. Often an isotype control is needed. In paper III, histone modifications of H3K27ac, H3K4me, and H2A.Z have been analyzed by ChIP-Seq. After genomic alignment by bowtie2, representative peaks of histone modifications were defined by MACS peaks builder, using the SeqMonk program.

3.3 Transcriptome profiling

Analysis of the gene transcription on a genome-wide scale is referred to as transcriptome profiling and several high-throughput methods have been developed for this purpose. Based on a DNA chip, a gene expression microarray has been widely applied in numerous studies during the past twenty years. In recent time, next generation sequencing has taken over the role as the most used technique in transcriptome profiling. In the current thesis, both microarrays and RNA sequencing have been applied for transcriptome profiling.

3.3.1 Gene expression microarray

Gene expression microarrays are one of the most popular applications of DNA chips, which use microscopic probes fixed to a solid surface in order to capture nucleotide sequences for a target (Rosenbloom, Dreszer et al. 2012). These probes are designed to target known gene transcripts, and genome-wide gene expression levels can be measured simultaneously. Depending on the platforms, one or several probes may target same genes/transcripts and will hybridize to input cDNA, reverse transcribed from an RNA sample. The hybridization generates signals due to the complete complementation of the designed probe and the fluorescently-labeled cDNA input

sequence. The strength of these signals provides a quantitative result that represents the level of transcription for each gene/transcript. However, gene expression microarrays can only detect known gene transcripts. Moreover, unavoidable background signals and batch effects may lead to the requirement for additional data normalization (Ramsay 1998). Nevertheless, gene expression microarray provides a cost-effective platform and can provide meaningful and reproducible results. In paper II, the Human Genome U133 Plus2.0 Array from Affymetrix was applied to analyze the global gene expression in CN-AML patients.

3.3.2 Messenger RNA sequencing

Messenger RNA (mRNA) sequencing, in other words, whole transcriptome shotgun sequencing, is a high-throughput technique to characterize the transcriptome at a given time point (Holt and Jones 2008). In brief, mRNA is purified from total RNA by the removal of ribosomal RNA and then reverse-transcribed to cDNA. These cDNAs are used as the template to generate a DNA library after covalently adding synthetic adaptor sequences at the end of the cDNA fragments by DNA ligase. Essentially, library construction is amplification based, which ensures sufficient signal intensity at the sequencing step. For each platform, the adaptor sequence is specific. It ensures the fixation of sequencing templates onto a solid surface (such as a flowcell of Illumina HiSeq 2000) and allows for the parallel reaction of extension for every fragment. The sequencing step is carried out by cycles of adding labeled single nucleotides followed by washing, then scanning. The camera captures the signals at each cycle and translates this into a nucleotide code. These massive simultaneous reactions generate millions of reads at desired length and, after alignment to the reference genome, it will produce information of gene expression at a genome-wide level.

The bioinformatic analysis is required to quantify gene expression in RNAseq experiments. Output reads are trimmed and aligned to the genome and the reads mapped to repetitive regions and those with ambiguous mapping are often eliminated at this step. Naturally, the number of reads that is mapped to a given gene corresponds to the amount of mRNA that is subjected to sequencing. However, due to the fact that the number of reads per gene is affected by the sequencing depth and initially determined by the total length of the gene, instead of the raw read counting per gene, reads (for single end sequencing)/fragments (for paired ends) per kilo base per million reads (RPKM/FPKM) are often computed to normalize the gene expression (Holt and Jones

2008). In Paper III, mRNA sequencing was performed on the Illumina HiSeq2000 platform. mRNA was purified from total RNA that was extracted by TRIzol[®], then reversely transcribed to cDNA and constructed into the sequencing library by the TruSeq RNA Library Preparation Kit v2. Library from 12 samples (7 CN-AML and 5 NBM CD34⁺ cells) was barcoded and pooled in six lanes. A total of 1.4 billion reads were produced with 97% mapping efficacy and an average of 114 million reads per sample. In this paper, gene expression levels were estimated in FPKM values after being aligned first by Tophat to human genome version GRCh37, and then analyzed in the Cufflinks and R program.

3.4 Genome editing with the CRISPR-Cas9 system

The adaptation of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-Associated (Cas) system provides a powerful tool to introduce targeted editing into an established genome(Burgess 2013). Before the CRISPR-Cas system, this purpose is often taken with protein-based targeting methods, such as transcription activator-like effector nuclease (TALEN) systems and zinc finger nuclease (ZFN) technology, which often have lower efficiency, long experimental protocols and high off-target rates(Veres, Gosis et al. 2014; Koo, Lee et al. 2015). The CRISPR-Cas system is carried out with a simplified cloning protocol and DNA sequence-based complimentary targeting ensures a more specific knocking out at the targeted site. In Paper III, we applied CRISPR-Cas9 system onto KG1a leukemic cell line to introduce the site-specific knocking-out of selected enhancer elements in order to study the resulting effects on their putative target genes.

CRISPR was first discovered in bacterial genomes and later in archaea as acquired sequences. The *Cas* genes, often located at neighboring sites of CRISPR sequences in bacteria genome, possesses helicase and nucleases activity(Burgess 2013). Based on these observations, the CRISPR-Cas system was developed for mammalian genome engineering(Figure 4). Two major components constitute the basis of the CRISPR-Cas9 system, the single guider RNA (sgRNA) together with the CRISPR scaffold RNA sequence and the human codon optimized endonuclease Cas9 protein. sgRNA is a synthetic short nucleotide sequence that is often 20nt in length and is complementary to the sequence of the target as a "seed"(Ran, Hsu et al. 2013). At the following position to the target sequence, it must contain a species-specific protospacer motif (PAM) sequence. After introducing a CRISPR-Cas vector into the experimental model,

a riboprotein complex is formed by the Cas9 protein together with hairpin-folded scaffold RNA, and the sgRNA sequence is continued at the "tail." Cas9 recognizes its PAM sequence (5'-NGG-3' for spCas9) in the target genome and executes endonuclease activity to cut the target upon complementary matches of sgRNA. The Cas9 protein contains two nuclease domains that cut each DNA strand simultaneously. Once the cutting occurs in the host genome, a double-strand break is introduced when using wild type Cas9 protein. In mammalian cells, two mechanisms mainly respond by repairing the double-strand break, homologous recombination (HR) and non-homologous end joining (NHEJ). When a homologous template is present in the cells, such as another intact allele or repair template sequence, the HR mechanism often results in an accurate repair. However, when a homologous template is absent, a more rapid NHEJ reaction is activated to ensure for genomic integrity. Moreover, to able to join the point of unmatching ends, a small deletion or insertion is often formed.

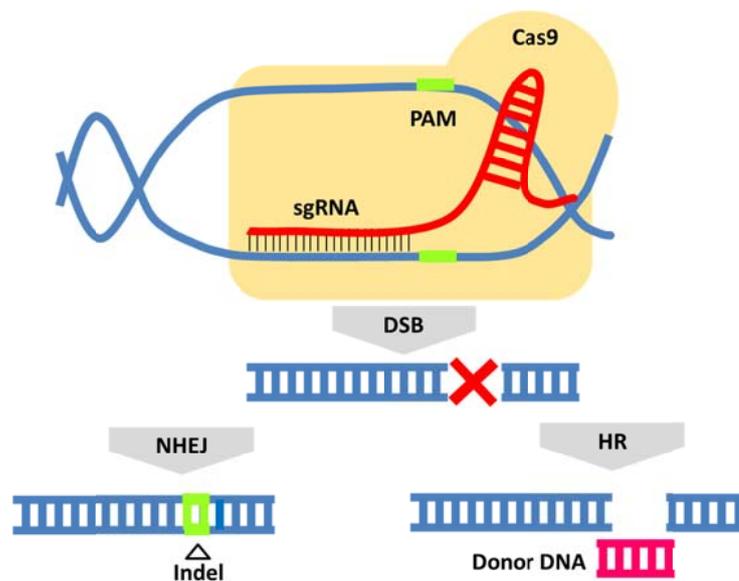


Figure 4. Genome editing with CRISPR-Cas9 system. Cas9 recognizes the protospacer motif (PAM) sequence and mediates double strands break upon a complementary match of sgRNA to genomic target regions. Non-homologous end joining (NHEJ) rapidly repairs the DNA breaks and often forms small indel at the joining site. Homologous recombination (HR) utilizes the donor DNA template to mediate precise repair.

To study the aberrantly activated enhancers in the leukemic system (in paper III), we used the CRISPR-Cas9 system to introduce site-specific deletion of selected enhancers into KG1a cells that were marked with H3K27ac and that were hypomethylated in CN-AML. To ensure the complete removal, two gRNA sequences were designed to target the start and end of each enhancer separately and they were cloned into a plasmid

vector with the same backbone but with different fluoresce reporters (PLKO5-sgRNA-EFS vector with eGFP/RFP, Addgene #57823/57822). The vector carrying the human codon optimized S.Pyrogenes Cas9 (PX458, Addgene #48138), was co-transfected by electroporation into KG1a cells by the Neon[®] Transfection system (Invitrogen). Cells were cultivated for 48 hours before harvested for FACS sorting(BD, Aira II). Double positive cells were sorted into both a bulk population and into 96 well plates with single cells in order to generate clones. Single cell colonies were propagated for 3 weeks under controlled conditions and then expanded for genotyping and RNA extraction. Four clones of each, either with a double allelic deletion or wild-type enhancers were selected. Expressions of putative target genes were tested using q-PCR.

4. Results and Discussion

In the three studies presented in this thesis, we characterized aberrant DNA methylation in CN-AML, correlated the methylation signatures with patient genetic features for CGIs (paper I), promoter-distal regions (enhancers, paper III) as well as in genome-wide level (paper II). We investigated the relationships between alternated DNA methylation and gene expression changes, chromatin states, and prognostic values for patient clinical outcomes. The gene mutation associated methylation changes in CN-AML were specifically addressed for *NPM1*, *IDH* (Paper I), and *DNMT3A* (Paper II). Prognostic values of DNA methylation PcG target genes were discussed (Paper I). Moreover, we addressed the AML associated aberrant enhancer activity in Paper III.

4.1 Paper I

In Paper I, promoter-focused DNA methylation was profiled on genome level for 58 AML patients with normal karyotype by Illumina methylation27k arrays (27k, in a total of 27,578 probes, 72.5% annotated to CGI) in test cohort, and 60 additional CN-AML patient samples were analyzed for validation by 27k and 450k arrays. In this paper, we found a significant increase in methylation associating to *IDH* mutations after comparing global average methylation level of each mutation with its wild type counterpart. We observed possible differences between *IDH1* and *IDH2* mutations where an elevated methylation level was found only in CGI regions of *IDH1* mutated but among non-CGI sites for *IDH2*. We also found that mutations of *IDH* and *NPM1* were significantly associated with aberrant DNA methylation signatures as defined by unsupervised clustering. This study provides additional evidence for and highlights the role of specific mutations in relation to methylation changes in cytogenetic normal AML. In later studies, the hypermethylation phenotype related to *IDH* mutations has been well defined by both genetically manipulated cell models and other tumor types harboring *IDH* mutations (Figueroa, Abdel-Wahab et al. 2010; Duncan, Barwick et al. 2012; Lu, Ward et al. 2012). This hypermethylation phenotype has been linked to TET-mediated oxidative carboxylation of 5mC and recurrent mutations of *TET2*, *IDH1* and *IDH2* have been shown to be mutually exclusive in AML. The difference in the functional impact of *IDH1* and *IDH2* mutations are unknown. Along with a difference in the localization in different cellular compartments, a more robust 2-HG production has been found for mutated *IDH2* compared to *IDH1* (Ward, Lu et al. 2013). These

findings suggest a difference in the dysregulating mechanisms between the two IDH mutations.

By hierarchical clustering, we also found that *HOX* genes are enriched for differential methylation changes in AML. Among four differentially methylated CpG clusters (DMCs), *HOX* genes were particularly represented in the groups with methylation changes predominantly in CGI regions. The homeobox gene family is a well-defined polycomb target which plays an important function during the developmental process (Bird 2007; Khan, Lee et al. 2015). Moreover, the key developmental genes are marked with both H3K27me3 and H3K4me1 at the same time (Bernstein, Mikkelsen et al. 2006). Based on this knowledge, we further integrated PRC2 ChIP-seq data and genes with bivalent histone marks with the results from our Illumina 27K array. Both genesets were found to have significantly increased CGI methylation. Moreover, PRC2 targets had greater changes than non-PRC2 targets. After addressing the prognostic value of methylation level of the PRC2 targets gene in CN-AML patients, we found that increased PRC2 target methylation was associated with a better clinical outcome. Overexpression of *HOX* genes is well documented in AML with mixed lineage leukemia and results suggest a more unfavorable clinical outcome for these patients (Krivtsov and Armstrong 2007). Our study showed that elevated *HOX* expression may associate with low DNA methylation, and the DNA methylation signature of *HOX* genes can be an independent predictor for patient prognosis.

4.2 Paper II

In the past decades, methylation studies have been focused on CpG rich promoter regions. However, more recent study by Feinberg *et.al.*, have shown a higher degree of methylation variation in cancer cells in regions adjacent to CGIs (named CGI shores) and even further distal regions (Irizarry, Ladd-Acosta et al. 2009). This finding has raised questions about the functional role of genome-wide DNA methylation changes outside CGIs and their relation to gene expression regulation and chromatin organization. The updated Illumina Methylation array platform, Illumina 450K, is composed of over 480,000 probes covering genome-wide CGIs, CGI adjacent regions (shores, <2kb), near proximal regions (shelves, 2-4kb) and distal regions (open sea >4kb) (Dedeurwaerder, Defrance et al. 2011). In paper II, by using the Illumina 450K array, we characterized methylation of 62 CN-AML patients on genome-wide level together with CD34+ cells from the bone marrow of healthy donors. Indeed,

methylation changes were observed in regions ranging from CGIs to open seas and differentially methylated CpG residues (DMCs) were found to be enriched in regions farther from CGIs (shelves and open seas, $p < 10^{-3}$). Although promoter hypermethylation has been well recognized in most type of malignant diseases, DNA methylation changes in non-CGI regions in AML have not previously been well defined. Our study, among others, provides an overlook of altered methylation in non-CGI regions. We demonstrated that methylation changes affect regions distal to promoters more frequently compared to promoter CGIs, and moreover, we found enhancer probes, as defined in the Illumina array, to be over-represented for DMCs. Meanwhile, we have also addressed dynamic methylation changes during normal granulopoiesis in collaboration with Rönnerblad et al.(212). Therefore, we hypothesized that differential methylation in CGI distal regions affects enhancer activity and that this can have a role in the leukemogenic process in AML. This question was further addressed in **paper III**.

Transcription factors (TFs) drive the cell differentiation process and typically, the normal function of lineage-specific TFs is disrupted in hematological malignancies, especially in AML. We hypothesize that transcription factors are preferentially targeted by DNA methylation changes. We analyzed the methylation changes in a list of 1620 TFs obtained from Fantome project (Forrest, Kawaji et al. 2014) and found that aberrant methylation was preferentially located in TF genes compared to the genes at a global level ($p < 10^{-4}$). Among TFs, *WT1* gene was most enriched for DMCs with consistent hypermethylation in the intragenic region but not main promoter ($p < 10^{-4}$). *WT1* is known to be overexpressed and yet recurrently mutated in myeloid malignancies (see introduction). Its intragenic site has previously shown allelic-specific methylation and is believed to regulate both an alternative transcript of *WT1* and an antisense transcript(Hancock, Brown et al. 2007; Brown, Power et al. 2008). In the hematological system, overexpression of *WT1* leads to both G1 arrest of progenitor cells and promoted differentiation towards the myeloid monocyte lineage(Ellisen, Carlesso et al. 2001). In AML, overexpression of *WT1* has been linked to hypermethylation of an intragenic CTCF binding site(Zitzmann, Mayr et al. 2014). Taken together with our findings, the *WT1* gene serves an example of how dysregulation at several levels may occur. Interestingly, it is also found that the occurrence of *WT1* mutation not only anti-correlate with *TET2* or *IDH* mutations, but that it also is regulating 5hmC through interacting with *TET2* and *TET3*, inducing

target gene expression(Rampal, Alkalin et al. 2014; Wang, Xiao et al. 2015). Therefore, WT1 may mediate an oncogenic mechanism that is the result of the loss of imprinting and that induces leukemogenesis via TET-mediated epigenetic reprogramming.

Our previous study showed methylation patterns in CN-AML associated with *IDH* or *NPM1* mutations, using a strongly CGI biased global methylation analysis (Illumina's 27K methylation array). In this study, we extended the window and investigated DNA methylation with better coverage also including methylation outside of CGIs using the 450K array. We then found mutations of *DNMT3A* to have the most pronounced impact in defining methylation-based clusters by consensus clustering on the most variably methylated CpG residues (overall SD>0.15)(Senbabaoglu, Michailidis et al. 2014; Kiselev, Kirschner et al. 2017). Clusters in the lower hierarchy were linked to the mutations of *NPM1* and *IDH* and in the least step to *FLT3* mutations. Unsupervised consensus clustering not only determines the hierarchy of the dataset but also assesses the stability of the discovered clusters. Our result highlighted the importance of *DNMT3A* mutations in determining the leukemia-specific DNA methylation signatures in relation to other gene mutations. By further analysis, we found that *DNMT3A* mutations resulted in a clear genome-wide hypomethylated pattern where the *HOX* gene family was the only protein family that was significantly affected.

Before us, *DNMT3A* mutation associated hypomethylation was observed in specific gene promoters (Hajkova, Markova et al. 2012). It is also found that *HOX* loci are located in hypomethylated *DNMT3A* dependent canyons(Jeong, Sun et al. 2014). Exogenous expression of mutated *DNMT3A* in mice results in a dominant negative effect on establishing DNA methylation(Kim, Zhao et al. 2013). In addition, mutations in the *DNMT3A* gene result in reactivation of *MEIS1* and *HOX* in hematological malignancies(Ferreira, Heyn et al. 2016; Tan, Sun et al. 2016). One study pointed out that dysregulation of *HOX* genes is a result from *BMI1* dysfunction due to *DNMT3A* mutations. Interestingly, a specific interaction between the *PRC1* complex and *DNMT3A*^{R882} mutations has been also reported(Koya, Kataoka et al. 2016). These findings also suggested a leukemogenic mechanism where *PRC1* mediates suppression of differentiation-associated genes while dysregulation of genes such as the *HOX* family leads to enhanced HSC proliferation in *DNMT3A* mutated leukemia.

4.3 Paper III

In this study, we explored AML specific methylation induced dysregulation of enhancer elements as well as its effects on chromatin status and putative target genes. We integrated information on enhancers, as defined by the FANTOM consortium based on CAGE analysis, with our Illumina 450K data for 57 CN-AML patients and normal bone myeloid cells representing 4 differentiation stages (CMP to PMN). Our results revealed pronounced cell type-specific signatures of CAGE-defined enhancers by the methylation status. This suggests that DNA methylation potentially could be involved regulating hematopoietic cell states through methylation of enhancers. We defined changes in DNA methylation located in enhancers as belonging to one of three groups: either changes specific for normal myeloid development, specific for AML compared to any normal myeloid cells or finally as changes that were shared between normal myeloid development and those found in AML. Our observations showed that enhancer methylation in AML could represent both changes the parental epigenome as well as cancer specific changes. Moreover, recurrent mutations occurring in AML were found influence DNA methylation to a lesser degree in CAGE enhancer compared to the rest of the genome. This may suggest that different pathogenetic mechanisms converge to altered enhancer activities. A similar observation has been made in leukemic stem cells where mutation-independent methylation signatures have been defined (Jung, Dai et al. 2015). However, it has also been shown that *Tet2* mutation leads to enhancer hypermethylation in mice (Rasmussen, Jia et al. 2015). Furthermore, it would be of interest to analyze 5hmC levels of differentially methylated enhancers and integrate this with information of the samples' *TET2* mutation status.

Moreover, we linked DNA methylation changes to chromatin states and histone marks. Firstly, we found that differential methylation primarily occurred at regulatory regions of marked by DHSs in CD34⁺ cells. By further analyses of DHS-seq data from two leukemia cell lines and 4 AML patient samples together with the analysis of the histone marks H3K4me1, H3K4me3, H3K27ac, and H2A.Z. We found that hypermethylated DMCs are simultaneously associated with a reduced accessibility and a loss of active histone marks. Among hypomethylated DMCs, although no overall pattern of chromatin changes was observed, we found that a subset of enhancers undergone DNA hypomethylation was aberrantly activated accommodating with an increase in active histone marks (H3K27ac and H2A.Z). Moreover, we also found that DNA hypomethylation occurred in both poised enhancers marked with H3K4me1 in normal

CD34⁺ and the newly formed enhancer sites in leukemia. These aberrantly activated enhancers were associated with an increased transcription of their target genes, and this activation of transcription could be modified using CRISPR-mediated knock out of the enhancer sequences.

In this study, we could demonstrate a link between aberrant DNA methylation and acquired enhancer activity changes in AML. We could show that gaining methylation at enhancers corresponds to a repressed activity, however, losing methylation alone is not sufficient to induce enhancer activation. Silencing of active enhancers has been previously shown in multiple myeloma where enhancers required for B-cell differentiation can be targeted and silenced(Agirre, Castellano et al. 2015). Moreover, a tumor-related gain of DNA methylation in super-enhancers has been linked to transcriptionally repressive effects on corresponding genes in a recent study(Heyn, Vidal et al. 2016). It is noteworthy that hypermethylation can be found at the same enhancer loci across several cancer types in this study. It is not known how these enhancers acquire aberrant DNA methylation changes and how this related to tissue specificity. Transcription factors are likely to coordinate with DNA methylation mechanisms when establishing activate enhancers(Stadler, Murr et al. 2011; Feldmann, Ivanek et al. 2013). For example, hypomethylated super-enhancers co-occur with upregulated TFs in colon cancer(Heyn, Vidal et al. 2016). To note, we observed distinct enrichments of transcription factor binding motifs in hyper- and hypomethylated CAGE-enhancers, respectively. PU.1 and AP1 binding sites were found significantly enriched in hypomethylated sites. Further studies should address the question what transcription factors mediate additional signals to activate hypomethylated enhancer in AML.

5. Concluding Remarks

These three studies represent an effort to characterize aberrant DNA methylation in cytogenetically normal AML in relation to genetic mutations, transcriptomic alterations, histone modification marks as well as clinical outcome. The research focus has extended from primarily an interest in methylation of promoters, and CGI enriched regions to TSS-distal regulatory elements and this development is also reflected in this thesis.

A CGI dominated differential methylation pattern in CN-AML has been linked to mutations in the *IDH* and *NPM1* genes in the paper I. Moreover, clusters of differential methylation revealed preferential methylation of genes in the HOX family, which lead to a more focused analysis on PRC2 targets. The prognostic value of using DNA methylation of PRC2 targets as a predictor of patient outcome was particularly discussed in this work. The data proposed DNA methylation as a novel marker for predicting patient outcome in CN-AML, and potentially this could indicate that new subgroups of AML could be discovered based on DNA methylation.

With the extended view on DNA methylation in paper II, we found an enrichment of leukemia-specific DNA methylation changes in regions distal to CGIs in CN-AML. We also found that transcription factors are a preferred target of aberrant DNA methylation. Moreover, we described a pattern of global hypomethylation of AML samples with *DNMT3A* mutations as we further classified the hierarchy of the relationships between gene mutations and DNA methylation profiles.

The study of leukemic specific enhancer methylation changes in paper III revealed an interplay between aberrant DNA methylation and altered enhancer activity in relation to the leukemic transcriptome, highlighting the pathogenic dysregulation of enhancer elements in leukemia. This study provided novel observations regarding the regulatory function of DNA methylation in the AML epigenome and provided evidence of acquired aberrant enhancer activities in CN-AML.

Nevertheless, technical limitation of Illumina platforms as the restricted coverage and their inability to distinguish 5hmC from 5mC leaves remaining questions. By

considering *TET* mutations and integrating this with a 5hmC analysis, this would give a more comprehensive of AML specific DNA methylation changes in the future.

To conclude, leukemia-specific changes of DNA methylation correlates with patient genotypes and the clinical outcome of CN-AML patients. It is coherent with pathogenic mechanisms and the changes contribute to a certain degree of transcriptomic alterations. Characterization of genome-wide DNA methylation profiles does not only provide a more comprehensive view of the AML epigenome and aberrant molecular events but also serves to identify potentially novel mechanisms of leukemia transformation and may also identify targets for novel AML treatments. The exact contribution of DNA methylation changes and other epigenetic mechanisms to the leukemogenic process needs to be further studied and so do their relation to genetic mutations. If epigenetic changes can precede and pave the way for genetic changes is cancer is still an unknown question. Also, as all novel prognostic markers in AML, they need to be further validated in relation to treatment decisions such as how hematopoietic stem cell transplantation can impact the prognostic predictions of epigenetic marks. Such studies are needed before they could be implemented as new prognostic factors in the clinical routine.

6. Acknowledgements

During the years of my Ph.D. study, life has gone up and down in the lab and also for personal life. In the days when things didn't go as according to the plan, it has been so helpful to have so many people around, will to listen, share their knowledge and experiences and offer their supports. I am honored to work with these people, and I would like to express my sincere gratitude to everyone that I have worked, met and discussed with at HERM, Lennartsson's group, KEK group members and more.

First of all, I would like to express my gratitude to the bone marrow **donors**. Thank you for your big love. Without your contributions, none of the works in this thesis would have happened.

I would like to express my gratitude to my principle supervisor **Sören Lehmann**. Thank you for giving me an opportunity to pursue the postgraduate study in your group. Thank you for being a supportive and generous tutor, sharing your knowledge and experiences, gave me opportunities to explore and develop my ideas. I appreciate your calmness and optimism, admire your enthusiasm and scientific perspectives.

To my co-supervisor, **Andreas Lennartsson**, thank you for guiding me through those learning curves, being continuously engaged with my projects. You have always been patient and encouraging. I endorse your persistence and thoughtfulness. Foremost, "No Science, No Fun."

A special mention goes to **Lehmann group** at HERM and Uppsala. To **Stefan**, for those good discussions. You have generously shared valuable experiences and knowledge with me, introduced me to the first study and helped me through all the projects. To **Huthayfa**, it is so nice to have you joined the group and make a great company to me. Thank you for all of your help in the lab, fun chats and shared the important duty of backing-up for biobank. To **Sofia**, thank you for organizing the samples and keeping our cell room and lab benches in good orders. To **Christer**, you solved the "normal" problem and helped me so many times on the sampling and getting the clinical data. To **Anna, Anne** and **My**, you make a great Uppsala team. Thank you for sharing those interesting discussions and inputs in the meetings. **Sylvain**, the true bioinformatician, you raised the bar of discussions. To **Minna**, for sharing the positivity and efforts on organizing the documentation system. To, Professor **Christer Paul**, for showing me the passion and devotion to science as a senior scientist. To former Lehmann group members, **Dina** and **Kerstin**, thank you for introducing me to the lab and generous offered your help.

I would like to thank all collaborators, for your contributions to my projects and my scientific developments. Special thanks of mine go to **Mohsen Karimi**. You taught me all fundamental knowledge about DNA methylation when I started and guided me through several important techniques. It changed my view and benefited on all of my research works. To **Lee Siggens**, thank you for everything you have helped on. You

have offered much more than you should. I won't be able to make the last manuscript in time if without your contributions. More, I cannot survive the thesis writing :). To collaborators at Ulm, **Verena I. Gaidzik**, **Lars Bullinger** and **Konstanze Döhner**, thank you for your help on mutation analysis and your inputs on the projects and manuscripts. To **Lina Cordeddu** for your help on ChIPping the patient samples for the third project; **Karl Ekwall**, for all the interesting discussions at KAW meetings and your inputs on the projects. To **Kasper Karlsson** and **Sten Linnarsson**, for your help on performing RNA-sequencing. To **Marios Dimitriou**, for introducing the basics of flow cytometry and helped me with cell sorting. I enjoyed a lot of our scientific discussions in those long lab days. I would also like to thank all the other **co-authors** for my manuscripts, for your contributions on providing samples and performing the analysis.

I would like to thank Professor **Margaret Goodell** and **the awesome Goodell's lab members**, for your warmest hosting during my visit to BCM, giving me the opportunity to experience a different research environment and letting me join those interesting discussions and meetings.

I want to express my gratitude to **all HERM members**. HERM is of particular importance to me. I am so grateful to be able to work at such helpful environment with so many kind people. It has made my days much easier and cheerful. I will always remember my days that spent here, with each one of you.

In particular to

Eva Hellström-Lindberg, for your efforts on promoting an active and collaborative research environment at HERM. **Peter Höglund**, for acceptance, to be the chairman of my defense. **Julian Walfridsson**, for all scientific inputs on my projects and discussions. I appreciate those constructive feedbacks you gave me and every Ph.D should have a mentor like you. **Hong Qian**, for giving me suggestions for my lab works and future career. **Robert Månsson** and **Evren Alici**, for sharing your knowledge, experiences and demonstrating critical thinking at SAP. **Iyadh Douagi**, for keeping the FACS facility well-functioning.

Monika Jansson, for your generous help on solving my lab emergencies, your efforts to keep everything up-and-running in orders yet facilitate every person's needs. **Mari Gilljam**, **Brigitta Stellan**, **Inger Vedin**, **Asmaa Ben Azenkoud**, **Charlotte Gustafsson** for your help on my day-by-day lab life. **Lili Andersson**, **Annette Engström**, **Sri Sahlin**, for your help on the administrative works.

Erle Refsum, **Caroline Gavin**, **Nadir Kadri**, **Stephan Meinke**, **Srinivasa Rao Nagubothu**, **Carin Dahlberg**, **Makoto Kondo**, **Ayla De Paepe**, **Thibault Boudierlique**, **Monika Donlinska**, for sharing fun lunch time and coffees together. **Sridharan Ganesan**, **Shabnam Kharazi**, **Thuy Luu Thanh**, for giving me advice and cheering me up all the time. **Xiaoze Li** and **Pingnan Xiao**, for our "secret" Chinese

conversations. **Edda Maria Elvarsdottir, Teresa Mortera-Blanco, Hani Abdulkadir, Deepika Nair, Yaser Heshmati, Aditya Harisankar, Gözde Türköz, Michael Chrobok, Hero Nikdin**, for the time that we have spent together. Thank you for being ready to listen to my constructive but mostly impulsive talks and give me honest opinions. Your supports are priceless for me. All the fun moments at lunches and fikas just make my day becoming a bit brighter and more cheerful every time. **Simona Conte**, for those laughs and tears. They mean a great deal to me. I am glad to share the period of Ph.D life with you together.

Thanks to my friends that have given me the encouragement and support in these years.

Michelle Ljungmark, for all the discussions we had for data analysis and science, but more for those conversations for life and motherhood. I am glad to get to know you since we both work in the lab.

Jia Sun, Xiaozhen Li, Simei Yu, Bin Li, Tianwei Gu for the Saturdays that we spent together. Your friendships make an important part of my life in Sweden. I am sure that we will meet more back home. **Qinzi Yan&Shuo Liu, Xin Xia& Sike Huang, Qian Wang&Yao Yao**, for giving a lot of supports for my family life. **Xinming Wang**, you came to pick me up from the airport on my very first arrival. **Nina Ning Wang**, for being such a fun person and those wonderful fikas.

My old friends back home in China. **Bin Shan**, for all the Wechat messages. You are the one be there for me no matter when I need. To **Guang Yang**, for knowing me more than myself and sharing the beautiful teenagerhood together.

I would like to thank my **lovely family** deeply.

衷心感谢我的**爸爸妈妈**，多年来给我的支持和爱。感谢您们一直都相信我对生活的决定和对理想的追求，一直无私的给予我最大程度的理解和帮助。您们用自己对事业的努力和对家庭的责任感为我做了最真实的榜样，成为我追求自己目标最大的动力。**戴威**，因为有你所以我的童年从来都不孤单。**爷爷，外婆**，谢谢您们对我的爱和谅解多年来我不能经常去探望。**老伯，姑姑，舅舅**和我的大家庭，感谢您们多年的关注和鼓励。

To **Andrew**, I love you. Thank you for loving me even more.

To my little princess **Jade**, thank you for joining my life and choosing me to be your mother.

7. References

- Abdel-Wahab, O., M. Adli, et al. (2012). "ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression." Cancer Cell **22**(2): 180-193.
- Abdel-Wahab, O. and A. Dey (2013). "The ASXL-BAP1 axis: new factors in myelopoiesis, cancer and epigenetics." Leukemia **27**(1): 10-15.
- Abraham, B. J., K. Cui, et al. (2013). "Dynamic regulation of epigenomic landscapes during hematopoiesis." BMC Genomics **14**: 193.
- Agirre, X., G. Castellano, et al. (2015). "Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers." Genome Res **25**(4): 478-487.
- Allis, C. D. and T. Jenuwein (2016). "The molecular hallmarks of epigenetic control." Nat Rev Genet **17**(8): 487-500.
- Andersson, R., C. Gebhard, et al. (2014). "An atlas of active enhancers across human cell types and tissues." Nature **507**(7493): 455-461.
- Aparicio, O., J. V. Geisberg, et al. (2004). "Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences in vivo." Curr Protoc Cell Biol **Chapter 17**: Unit 17 17.
- Arber, D. A., A. Orazi, et al. (2016). "The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia." Blood **127**(20): 2391-2405.
- Audia, J. E. and R. M. Campbell (2016). "Histone Modifications and Cancer." Cold Spring Harb Perspect Biol **8**(4): a019521.
- Bannister, A. J. and T. Kouzarides (2011). "Regulation of chromatin by histone modifications." Cell Res **21**(3): 381-395.
- Bannister, A. J., R. Schneider, et al. (2002). "Histone methylation: dynamic or static?" Cell **109**(7): 801-806.
- Barragan, E., J. Cervera, et al. (2004). "Prognostic implications of Wilms' tumor gene (WT1) expression in patients with de novo acute myeloid leukemia." Haematologica **89**(8): 926-933.
- Becker, H., G. Marcucci, et al. (2010). "Favorable prognostic impact of NPM1 mutations in older patients with cytogenetically normal de novo acute myeloid leukemia and associated gene- and microRNA-expression signatures: a Cancer and Leukemia Group B study." J Clin Oncol **28**(4): 596-604.
- Behm, F. G. (2003). "Chapter 2: Classification of acute leukemias." Treatment of Acute Leukemias: New Directions for Clinical Research.
- Berger, S. L. (2007). "The complex language of chromatin regulation during transcription." Nature **447**(7143): 407-412.
- Berger, S. L., T. Kouzarides, et al. (2009). "An operational definition of epigenetics." Genes Dev **23**(7): 781-783.
- Berman, B. P., D. J. Weisenberger, et al. (2012). "Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains." Nat Genet **44**(1): 40-46.
- Bernstein, B. E., T. S. Mikkelsen, et al. (2006). "A bivalent chromatin structure marks key developmental genes in embryonic stem cells." Cell **125**(2): 315-326.
- Bibikova, M., Z. Lin, et al. (2006). "High-throughput DNA methylation profiling using universal bead arrays." Genome Res **16**(3): 383-393.
- Bird, A. (2007). "Perceptions of epigenetics." Nature **447**(7143): 396-398.
- Bird, A. (2009). "On the track of DNA methylation: an interview with Adrian Bird by Jane Gitschier." PLoS Genet **5**(10): e1000667.

- Birney, E., J. A. Stamatoyannopoulos, et al. (2007). "Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project." Nature **447**(7146): 799-816.
- Bogdanovic, O. and G. J. Veenstra (2009). "DNA methylation and methyl-CpG binding proteins: developmental requirements and function." Chromosoma **118**(5): 549-565.
- Boultonwood, J., J. Perry, et al. (2010). "Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia." Leukemia **24**(5): 1062-1065.
- Bourc'his, D., G. L. Xu, et al. (2001). "Dnmt3L and the establishment of maternal genomic imprints." Science **294**(5551): 2536-2539.
- Bowman, R. L. and R. L. Levine (2017). "TET2 in Normal and Malignant Hematopoiesis." Cold Spring Harb Perspect Med.
- Broske, A. M., L. Vockentanz, et al. (2009). "DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction." Nat Genet **41**(11): 1207-1215.
- Brown, K. W., F. Power, et al. (2008). "Frequency and timing of loss of imprinting at 11p13 and 11p15 in Wilms' tumor development." Mol Cancer Res **6**(7): 1114-1123.
- Burgess, D. J. (2013). "Technology: a CRISPR genome-editing tool." Nat Rev Genet **14**(2): 80.
- Buschbeck, M. and S. B. Hake (2017). "Variants of core histones and their roles in cell fate decisions, development and cancer." Nat Rev Mol Cell Biol.
- Calvanese, V., A. F. Fernandez, et al. (2012). "A promoter DNA demethylation landscape of human hematopoietic differentiation." Nucleic Acids Res **40**(1): 116-131.
- Carnicer, M. J., A. Lasa, et al. (2008). "K313dup is a recurrent CEBPA mutation in de novo acute myeloid leukemia (AML)." Ann Hematol **87**(10): 819-827.
- Challen, G. A., D. Sun, et al. (2014). "Dnmt3a and Dnmt3b have overlapping and distinct functions in hematopoietic stem cells." Cell Stem Cell **15**(3): 350-364.
- Chen, W., E. Drakos, et al. (2010). "mTOR signaling is activated by FLT3 kinase and promotes survival of FLT3-mutated acute myeloid leukemia cells." Mol Cancer **9**: 292.
- Chodavarapu, R. K., S. Feng, et al. (2010). "Relationship between nucleosome positioning and DNA methylation." Nature **466**(7304): 388-392.
- Chou, S. T., E. Khandros, et al. (2009). "Graded repression of PU.1/Sfp1 gene transcription by GATA factors regulates hematopoietic cell fate." Blood **114**(5): 983-994.
- Chou, W. C., S. C. Chou, et al. (2011). "TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics." Blood **118**(14): 3803-3810.
- Cole, C. B., A. M. Verdoni, et al. (2016). "PML-RARA requires DNA methyltransferase 3A to initiate acute promyelocytic leukemia." J Clin Invest **126**(1): 85-98.
- Conway O'Brien, E., S. Prideaux, et al. (2014). "The epigenetic landscape of acute myeloid leukemia." Adv Hematol **2014**: 103175.
- Cui, K., C. Zang, et al. (2009). "Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation." Cell Stem Cell **4**(1): 80-93.
- Dang, L., D. W. White, et al. (2009). "Cancer-associated IDH1 mutations produce 2-hydroxyglutarate." Nature **462**(7274): 739-744.

- Dawlaty, M. M., A. Breiling, et al. (2014). "Loss of Tet enzymes compromises proper differentiation of embryonic stem cells." *Dev Cell* **29**(1): 102-111.
- de Bruijn, M. and E. Dzierzak (2017). "Runx transcription factors in the development and function of the definitive hematopoietic system." *Blood*.
- De Kouchkovsky, I. and M. Abdul-Hay (2016). "'Acute myeloid leukemia: a comprehensive review and 2016 update'." *Blood Cancer J* **6**(7): e441.
- de la Rica, L., J. Rodriguez-Ubreva, et al. (2013). "PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation." *Genome Biol* **14**(9): R99.
- Deaton, A. M. and A. Bird (2011). "CpG islands and the regulation of transcription." *Genes Dev* **25**(10): 1010-1022.
- Dedeurwaerder, S., M. Defrance, et al. (2011). "Evaluation of the Infinium Methylation 450K technology." *Epigenomics* **3**(6): 771-784.
- Deneberg, S., M. Grovdal, et al. (2010). "Gene-specific and global methylation patterns predict outcome in patients with acute myeloid leukemia." *Leukemia* **24**(5): 932-941.
- Deng, W., J. Lee, et al. (2012). "Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor." *Cell* **149**(6): 1233-1244.
- Dohner, H., E. Estey, et al. (2017). "Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel." *Blood* **129**(4): 424-447.
- Dohner, K., R. F. Schlenk, et al. (2005). "Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations." *Blood* **106**(12): 3740-3746.
- Duncan, B. K. and J. H. Miller (1980). "Mutagenic deamination of cytosine residues in DNA." *Nature* **287**(5782): 560-561.
- Duncan, C. G., B. G. Barwick, et al. (2012). "A heterozygous IDH1R132H/WT mutation induces genome-wide alterations in DNA methylation." *Genome Res* **22**(12): 2339-2355.
- Duymich, C. E., J. Charlet, et al. (2016). "DNMT3B isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells." *Nat Commun* **7**: 11453.
- Ellisen, L. W., N. Carlesso, et al. (2001). "The Wilms tumor suppressor WT1 directs stage-specific quiescence and differentiation of human hematopoietic progenitor cells." *EMBO J* **20**(8): 1897-1909.
- ENCODE (2012). "An integrated encyclopedia of DNA elements in the human genome." *Nature* **489**(7414): 57-74.
- Engel, N., J. S. Tront, et al. (2009). "Conserved DNA methylation in Gadd45a(-/-) mice." *Epigenetics* **4**(2): 98-99.
- Faber, Z. J., X. Chen, et al. (2016). "The genomic landscape of core-binding factor acute myeloid leukemias." *Nat Genet* **48**(12): 1551-1556.
- Feldmann, A., R. Ivanek, et al. (2013). "Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions." *PLoS Genet* **9**(12): e1003994.
- Feng, Q., H. Wang, et al. (2002). "Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain." *Curr Biol* **12**(12): 1052-1058.
- Feng, S., S. J. Cokus, et al. (2010). "Conservation and divergence of methylation patterning in plants and animals." *Proc Natl Acad Sci U S A* **107**(19): 8689-8694.

- Ferreira, H. J., H. Heyn, et al. (2016). "DNMT3A mutations mediate the epigenetic reactivation of the leukemogenic factor MEIS1 in acute myeloid leukemia." *Oncogene* **35**(23): 3079-3082.
- Ferreira, R., K. Ohneda, et al. (2005). "GATA1 function, a paradigm for transcription factors in hematopoiesis." *Mol Cell Biol* **25**(4): 1215-1227.
- Figueroa, M. E., O. Abdel-Wahab, et al. (2010). "Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation." *Cancer Cell* **18**(6): 553-567.
- Figueroa, M. E., S. Lugthart, et al. (2010). "DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia." *Cancer Cell* **17**(1): 13-27.
- Fisher, C. L., J. Berger, et al. (2003). "A human homolog of Additional sex combs, ADDITIONAL SEX COMBS-LIKE 1, maps to chromosome 20q11." *Gene* **306**: 115-126.
- Forrest, A. R., H. Kawaji, et al. (2014). "A promoter-level mammalian expression atlas." *Nature* **507**(7493): 462-470.
- Fujimoto, M., R. Kitazawa, et al. (2005). "Methylation adjacent to negatively regulating AP-1 site reactivates TrkA gene expression during cancer progression." *Oncogene* **24**(32): 5108-5118.
- Gaidzik, V. I., L. Bullinger, et al. (2011). "RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group." *J Clin Oncol* **29**(10): 1364-1372.
- Gaidzik, V. I., P. Paschka, et al. (2012). "TET2 mutations in acute myeloid leukemia (AML): results from a comprehensive genetic and clinical analysis of the AML study group." *J Clin Oncol* **30**(12): 1350-1357.
- Gaidzik, V. I., R. F. Schlenk, et al. (2013). "Clinical impact of DNMT3A mutations in younger adult patients with acute myeloid leukemia: results of the AML Study Group (AMLSSG)." *Blood* **121**(23): 4769-4777.
- Gaidzik, V. I., V. Teleanu, et al. (2016). "RUNX1 mutations in acute myeloid leukemia are associated with distinct clinico-pathologic and genetic features." *Leukemia* **30**(11): 2282.
- Gamis, A. S., T. A. Alonzo, et al. (2013). "Children's Oncology Group's 2013 blueprint for research: acute myeloid leukemia." *Pediatr Blood Cancer* **60**(6): 964-971.
- Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." *J Mol Biol* **196**(2): 261-282.
- Gelsi-Boyer, V., V. Trouplin, et al. (2009). "Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia." *Br J Haematol* **145**(6): 788-800.
- Genovese, G., A. K. Kahler, et al. (2014). "Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence." *N Engl J Med* **371**(26): 2477-2487.
- Goll, M. G., F. Kirpekar, et al. (2006). "Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2." *Science* **311**(5759): 395-398.
- Gowher, H., P. Loutchanwoot, et al. (2006). "Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase." *J Mol Biol* **357**(3): 928-941.
- Green, C. L., K. K. Koo, et al. (2010). "Prognostic significance of CEBPA mutations in a large cohort of younger adult patients with acute myeloid leukemia: impact of double CEBPA mutations and the interaction with FLT3 and NPM1 mutations." *J Clin Oncol* **28**(16): 2739-2747.
- Greif, P. A., N. P. Konstandin, et al. (2012). "RUNX1 mutations in cytogenetically normal acute myeloid leukemia are associated with a poor prognosis and up-regulation of lymphoid genes." *Haematologica* **97**(12): 1909-1915.

- Grimwade, D. and R. K. Hills (2009). "Independent prognostic factors for AML outcome." Hematology Am Soc Hematol Educ Program: 385-395.
- Grisendi, S., C. Mecucci, et al. (2006). "Nucleophosmin and cancer." Nat Rev Cancer **6**(7): 493-505.
- Grossmann, V., C. Haferlach, et al. (2013). "The molecular profile of adult T-cell acute lymphoblastic leukemia: mutations in RUNX1 and DNMT3A are associated with poor prognosis in T-ALL." Genes Chromosomes Cancer **52**(4): 410-422.
- Guo, H., P. Zhu, et al. (2014). "The DNA methylation landscape of human early embryos." Nature **511**(7511): 606-610.
- Hajkova, H., J. Markova, et al. (2012). "Decreased DNA methylation in acute myeloid leukemia patients with DNMT3A mutations and prognostic implications of DNA methylation." Leuk Res **36**(9): 1128-1133.
- Hancock, A. L., K. W. Brown, et al. (2007). "A CTCF-binding silencer regulates the imprinted genes AWT1 and WT1-AS and exhibits sequential epigenetic defects during Wilms' tumorigenesis." Hum Mol Genet **16**(3): 343-354.
- Hayatsu, H. (2008). "Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis--a personal account." Proc Jpn Acad Ser B Phys Biol Sci **84**(8): 321-330.
- He, X. J., T. Chen, et al. (2011). "Regulation and function of DNA methylation in plants and animals." Cell Res **21**(3): 442-465.
- He, Y. F., B. Z. Li, et al. (2011). "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA." Science **333**(6047): 1303-1307.
- Heintzman, N. D., R. K. Stuart, et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome." Nat Genet **39**(3): 311-318.
- Heinz, S., C. Benner, et al. (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." Mol Cell **38**(4): 576-589.
- Heinz, S., C. E. Romanoski, et al. (2015). "The selection and function of cell type-specific enhancers." Nat Rev Mol Cell Biol **16**(3): 144-154.
- Hernando-Herraez, I., R. Garcia-Perez, et al. (2015). "DNA Methylation: Insights into Human Evolution." PLoS Genet **11**(12): e1005661.
- Heyn, H. and M. Esteller (2015). "An Adenine Code for DNA: A Second Life for N6-Methyladenine." Cell **161**(4): 710-713.
- Heyn, H., E. Vidal, et al. (2016). "Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer." Genome Biol **17**: 11.
- Hills, R. K., A. Ivey, et al. (2016). "Assessment of Minimal Residual Disease in Standard-Risk AML." N Engl J Med **375**(6): e9.
- Hitchins, M. P., R. W. Rapkins, et al. (2011). "Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR." Cancer Cell **20**(2): 200-213.
- Hodges, E., A. Molaro, et al. (2011). "Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment." Mol Cell **44**(1): 17-28.
- Holliday, R. (2006). "Epigenetics: a historical overview." Epigenetics **1**(2): 76-80.
- Holt, R. A. and S. J. Jones (2008). "The new paradigm of flow cell sequencing." Genome Res **18**(6): 839-846.
- Holz-Schietinger, C., D. M. Matje, et al. (2011). "Oligomerization of DNMT3A controls the mechanism of de novo DNA methylation." J Biol Chem **286**(48): 41479-41488.

- Holz-Schietinger, C., D. M. Matje, et al. (2012). "Mutations in DNA methyltransferase (DNMT3A) observed in acute myeloid leukemia patients disrupt processive methylation." *J Biol Chem* **287**(37): 30941-30951.
- Hou, H. A., T. C. Huang, et al. (2010). "WT1 mutation in 470 adult patients with acute myeloid leukemia: stability during disease evolution and implication of its incorporation into a survival scoring system." *Blood* **115**(25): 5222-5231.
- Icard, P., L. Poulain, et al. (2012). "Understanding the central role of citrate in the metabolism of cancer cells." *Biochim Biophys Acta* **1825**(1): 111-116.
- Illingworth, R. S. and A. P. Bird (2009). "CpG islands--'a rough guide'." *FEBS Lett* **583**(11): 1713-1720.
- Irizarry, R. A., C. Ladd-Acosta, et al. (2009). "The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores." *Nat Genet* **41**(2): 178-186.
- Issa, J. P., H. M. Kantarjian, et al. (2005). "Azacitidine." *Nat Rev Drug Discov* **4**(4): 275-276.
- Ito, S., L. Shen, et al. (2011). "Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine." *Science* **333**(6047): 1300-1303.
- Iyer, L. M., M. Tahiliani, et al. (2009). "Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids." *Cell Cycle* **8**(11): 1698-1710.
- Jacinto, F. V., E. Ballestar, et al. (2008). "Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome." *Biotechniques* **44**(1): 35, 37, 39 passim.
- Jagannathan-Bogdan, M. and L. I. Zon (2013). "Hematopoiesis." *Development* **140**(12): 2463-2467.
- Jeong, M., D. Sun, et al. (2014). "Large conserved domains of low DNA methylation maintained by Dnmt3a." *Nat Genet* **46**(1): 17-23.
- Ji, H., L. I. Ehrlich, et al. (2010). "Comprehensive methylome map of lineage commitment from haematopoietic progenitors." *Nature* **467**(7313): 338-342.
- Jiang, N., L. Wang, et al. (2014). "Conserved and divergent patterns of DNA methylation in higher vertebrates." *Genome Biol Evol* **6**(11): 2998-3014.
- Jimenez-Useche, I., J. Ke, et al. (2013). "DNA methylation regulated nucleosome dynamics." *Sci Rep* **3**: 2121.
- Jones, P. A. (2012). "Functions of DNA methylation: islands, start sites, gene bodies and beyond." *Nat Rev Genet* **13**(7): 484-492.
- Jones, P. A. and G. Liang (2009). "Rethinking how DNA methylation patterns are maintained." *Nat Rev Genet* **10**(11): 805-811.
- Juliusson, G., J. Abrahamsson, et al. (2017). "Prevalence and characteristics of survivors from acute myeloid leukemia in Sweden." *Leukemia* **31**(3): 728-731.
- Juliusson, G., P. Antunovic, et al. (2009). "Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry." *Blood* **113**(18): 4179-4187.
- Jung, N., B. Dai, et al. (2015). "An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis." *Nat Commun* **6**: 8489.
- Kagey, M. H., J. J. Newman, et al. (2010). "Mediator and cohesin connect gene expression and chromatin architecture." *Nature* **467**(7314): 430-435.
- Khan, A. A., A. J. Lee, et al. (2015). "Polycomb group protein-mediated histone modifications during cell differentiation." *Epigenomics* **7**(1): 75-84.
- Kim, S. J., H. Zhao, et al. (2013). "A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells." *Blood* **122**(25): 4086-4089.
- King-Underwood, L., J. Renshaw, et al. (1996). "Mutations in the Wilms' tumor gene WT1 in leukemias." *Blood* **87**(6): 2171-2179.

- Kiselev, V. Y., K. Kirschner, et al. (2017). "SC3: consensus clustering of single-cell RNA-seq data." *Nat Methods*.
- Klepin, H. D., A. V. Rao, et al. (2014). "Acute myeloid leukemia and myelodysplastic syndromes in older adults." *J Clin Oncol* **32**(24): 2541-2552.
- Klose, R. J. and Y. Zhang (2007). "Regulation of histone methylation by demethylination and demethylation." *Nat Rev Mol Cell Biol* **8**(4): 307-318.
- Koo, T., J. Lee, et al. (2015). "Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9." *Mol Cells* **38**(6): 475-481.
- Kornberg, R. D. (1974). "Chromatin structure: a repeating unit of histones and DNA." *Science* **184**(4139): 868-871.
- Kornberg, R. D. and J. O. Thomas (1974). "Chromatin structure; oligomers of the histones." *Science* **184**(4139): 865-868.
- Koya, J., K. Kataoka, et al. (2016). "DNMT3A R882 mutants interact with polycomb proteins to block haematopoietic stem and leukaemic cell differentiation." *Nat Commun* **7**: 10924.
- Krivtsov, A. V. and S. A. Armstrong (2007). "MLL translocations, histone modifications and leukaemia stem-cell development." *Nat Rev Cancer* **7**(11): 823-833.
- Ku, M., J. D. Jaffe, et al. (2012). "H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions." *Genome Biol* **13**(10): R85.
- Kumar, V., M. Muratani, et al. (2013). "Uniform, optimal signal processing of mapped deep-sequencing data." *Nat Biotechnol* **31**(7): 615-622.
- Kundaje, A., W. Meuleman, et al. (2015). "Integrative analysis of 111 reference human epigenomes." *Nature* **518**(7539): 317-330.
- Kuo, L. J. and L. X. Yang (2008). "Gamma-H2AX - a novel biomarker for DNA double-strand breaks." *In Vivo* **22**(3): 305-309.
- Kwok, B., J. M. Hall, et al. (2015). "MDS-associated somatic mutations and clonal hematopoiesis are common in idiopathic cytopenias of undetermined significance." *Blood* **126**(21): 2355-2361.
- LaFave, L. M., W. Beguelin, et al. (2015). "Loss of BAP1 function leads to EZH2-dependent transformation." *Nat Med* **21**(11): 1344-1349.
- Lagrange, T., A. N. Kapanidis, et al. (1998). "New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB." *Genes Dev* **12**(1): 34-44.
- Lauster, R., T. A. Trautner, et al. (1989). "Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains." *J Mol Biol* **206**(2): 305-312.
- Law, J. A. and S. E. Jacobsen (2010). "Establishing, maintaining and modifying DNA methylation patterns in plants and animals." *Nat Rev Genet* **11**(3): 204-220.
- Ley, T. J., L. Ding, et al. (2010). "DNMT3A mutations in acute myeloid leukemia." *N Engl J Med* **363**(25): 2424-2433.
- Li, B., M. Carey, et al. (2007). "The role of chromatin during transcription." *Cell* **128**(4): 707-719.
- Li, E., T. H. Bestor, et al. (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." *Cell* **69**(6): 915-926.
- Li, W., D. Notani, et al. (2013). "Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation." *Nature* **498**(7455): 516-520.
- Li, Y., J. Han, et al. (2016). "Structural basis for activity regulation of MLL family methyltransferases." *Nature* **530**(7591): 447-452.

- Li, Z., X. Cai, et al. (2011). "Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies." *Blood* **118**(17): 4509-4518.
- Li, Z., P. Zhang, et al. (2017). "ASXL1 interacts with the cohesin complex to maintain chromatid separation and gene expression for normal hematopoiesis." *Sci Adv* **3**(1): e1601602.
- Liesveld, J. L., Lichtman, M.A. (2015). "Chapter 88: Acute Myelogenous Leukemia " *Williams Hematology*.
- Lister, R., M. Pelizzola, et al. (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences." *Nature* **462**(7271): 315-322.
- Long, H. K., N. P. Blackledge, et al. (2013). "ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection." *Biochem Soc Trans* **41**(3): 727-740.
- Lu, C., P. S. Ward, et al. (2012). "IDH mutation impairs histone demethylation and results in a block to cell differentiation." *Nature* **483**(7390): 474-478.
- Luo, M., T. Ling, et al. (2013). "NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells." *Stem Cells* **31**(7): 1278-1286.
- Mah, L. J., A. El-Osta, et al. (2010). "gammaH2AX: a sensitive molecular marker of DNA damage and repair." *Leukemia* **24**(4): 679-686.
- Malousi, A. and S. Koidou (2012). "DNA hypermethylation of alternatively spliced and repeat sequences in humans." *Mol Genet Genomics* **287**(8): 631-642.
- Marabita, F., M. Almgren, et al. (2013). "An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform." *Epigenetics* **8**(3): 333-346.
- Marcucci, G., K. Maharry, et al. (2010). "IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study." *J Clin Oncol* **28**(14): 2348-2355.
- Mardis, E. R., L. Ding, et al. (2009). "Recurring mutations found by sequencing an acute myeloid leukemia genome." *N Engl J Med* **361**(11): 1058-1066.
- Marks, H., T. Kalkan, et al. (2012). "The transcriptional and epigenomic foundations of ground state pluripotency." *Cell* **149**(3): 590-604.
- Mayle, A., L. Yang, et al. (2015). "Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation." *Blood* **125**(4): 629-638.
- Medeiros, B. C., A. T. Fathi, et al. (2017). "Isocitrate dehydrogenase mutations in myeloid malignancies." *Leukemia* **31**(2): 272-281.
- Meissner, A., A. Gnirke, et al. (2005). "Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis." *Nucleic Acids Res* **33**(18): 5868-5877.
- Mendler, J. H., K. Maharry, et al. (2012). "RUNX1 mutations are associated with poor outcome in younger and older patients with cytogenetically normal acute myeloid leukemia and with distinct gene and MicroRNA expression signatures." *J Clin Oncol* **30**(25): 3109-3118.
- Menssen, H. D., H. J. Renkl, et al. (1995). "Presence of Wilms' tumor gene (wt1) transcripts and the WT1 nuclear protein in the majority of human acute leukemias." *Leukemia* **9**(6): 1060-1067.
- Metzeler, K. H., H. Becker, et al. (2011). "ASXL1 mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category." *Blood* **118**(26): 6920-6929.
- Miller, C. A., R. K. Wilson, et al. (2013). "Genomic landscapes and clonality of de novo AML." *N Engl J Med* **369**(15): 1473.

- Miyagi, T., H. Ahuja, et al. (1993). "Expression of the candidate Wilm's tumor gene, WT1, in human leukemia cells." *Leukemia* **7**(7): 970-977.
- Mizuno, S., T. Chijiwa, et al. (2001). "Expression of DNA methyltransferases DNMT1, 3A, and 3B in normal hematopoiesis and in acute and chronic myelogenous leukemia." *Blood* **97**(5): 1172-1179.
- Mohandas, T., R. S. Sparkes, et al. (1981). "Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation." *Science* **211**(4480): 393-396.
- Molina, O., G. Vargiu, et al. (2016). "Epigenetic engineering reveals a balance between histone modifications and transcription in kinetochore maintenance." *Nat Commun* **7**: 13334.
- Moran-Crusio, K., L. Reavie, et al. (2011). "Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation." *Cancer Cell* **20**(1): 11-24.
- Nabel, C. S., H. Jia, et al. (2012). "AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation." *Nat Chem Biol* **8**(9): 751-758.
- Neben, K., S. Schnittger, et al. (2005). "Distinct gene expression patterns associated with FLT3- and NRAS-activating mutations in acute myeloid leukemia with normal karyotype." *Oncogene* **24**(9): 1580-1588.
- Neri, F., S. Rapelli, et al. (2017). "Intragenic DNA methylation prevents spurious transcription initiation." *Nature* **543**(7643): 72-77.
- Network, T. C. G. A. R. (2013). "Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia." *N Engl J Med* **368**(22): 2059-2074.
- Nyren, P. (2015). "The History of Pyrosequencing((R))." *Methods Mol Biol* **1315**: 3-15.
- Oberdoerffer, S. (2012). "A conserved role for intragenic DNA methylation in alternative pre-mRNA splicing." *Transcription* **3**(3): 106-109.
- Ohlsson, E., M. B. Schuster, et al. (2016). "The multifaceted functions of C/EBPalpha in normal and malignant haematopoiesis." *Leukemia* **30**(4): 767-775.
- Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." *Cell* **99**(3): 247-257.
- Ostronoff, F., M. Othus, et al. (2015). "Prognostic significance of NPM1 mutations in the absence of FLT3-internal tandem duplication in older patients with acute myeloid leukemia: a SWOG and UK National Cancer Research Institute/Medical Research Council report." *J Clin Oncol* **33**(10): 1157-1164.
- Pabst, T., B. U. Mueller, et al. (2001). "Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia." *Nat Genet* **27**(3): 263-270.
- Papaemmanuil, E., M. Gerstung, et al. (2016). "Genomic Classification and Prognosis in Acute Myeloid Leukemia." *N Engl J Med* **374**(23): 2209-2221.
- Park, S. H., H. S. Chi, et al. (2013). "Evaluation of prognostic factors in patients with therapy-related acute myeloid leukemia." *Blood Res* **48**(3): 185-192.
- Paschka, P., R. F. Schlenk, et al. (2015). "ASXL1 mutations in younger adult patients with acute myeloid leukemia: a study by the German-Austrian Acute Myeloid Leukemia Study Group." *Haematologica* **100**(3): 324-330.
- Patel, J. P., M. Gonen, et al. (2012). "Prognostic relevance of integrated genetic profiling in acute myeloid leukemia." *N Engl J Med* **366**(12): 1079-1089.
- Pennacchio, L. A., W. Bickmore, et al. (2013). "Enhancers: five essential questions." *Nat Rev Genet* **14**(4): 288-295.
- Penterman, J., D. Zilberman, et al. (2007). "DNA demethylation in the Arabidopsis genome." *Proc Natl Acad Sci U S A* **104**(16): 6752-6757.

- Radomska, H. S., C. S. Huettner, et al. (1998). "CCAAT/enhancer binding protein alpha is a regulatory switch sufficient for induction of granulocytic development from bipotential myeloid progenitors." *Mol Cell Biol* **18**(7): 4301-4314.
- Rai, K., I. J. Huggins, et al. (2008). "DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45." *Cell* **135**(7): 1201-1212.
- Raisner, R. M., P. D. Hartley, et al. (2005). "Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin." *Cell* **123**(2): 233-248.
- Rampal, R., A. Alkalin, et al. (2014). "DNA hydroxymethylation profiling reveals that WT1 mutations result in loss of TET2 function in acute myeloid leukemia." *Cell Rep* **9**(5): 1841-1855.
- Ramsay, G. (1998). "DNA chips: state-of-the art." *Nat Biotechnol* **16**(1): 40-44.
- Ran, F. A., P. D. Hsu, et al. (2013). "Genome engineering using the CRISPR-Cas9 system." *Nat Protoc* **8**(11): 2281-2308.
- Rasmussen, K. D., G. Jia, et al. (2015). "Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis." *Genes Dev* **29**(9): 910-922.
- Roessler, J., O. Ammerpohl, et al. (2012). "Quantitative cross-validation and content analysis of the 450k DNA methylation array from Illumina, Inc." *BMC Res Notes* **5**: 210.
- Roller, A., V. Grossmann, et al. (2013). "Landmark analysis of DNMT3A mutations in hematological malignancies." *Leukemia* **27**(7): 1573-1578.
- Rombouts, W. J., B. Lowenberg, et al. (2001). "Improved prognostic significance of cytokine-induced proliferation in vitro in patients with de novo acute myeloid leukemia of intermediate risk: impact of internal tandem duplications in the Flt3 gene." *Leukemia* **15**(7): 1046-1053.
- Rosenbloom, K. R., T. R. Dreszer, et al. (2012). "ENCODE whole-genome data in the UCSC Genome Browser: update 2012." *Nucleic Acids Res* **40**(Database issue): D912-917.
- Russler-Germain, D. A., D. H. Spencer, et al. (2014). "The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers." *Cancer Cell* **25**(4): 442-454.
- Sabo, P. J., M. S. Kuehn, et al. (2006). "Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays." *Nat Methods* **3**(7): 511-518.
- Santos, F. P., D. Jones, et al. (2011). "Prognostic value of FLT3 mutations among different cytogenetic subgroups in acute myeloid leukemia." *Cancer* **117**(10): 2145-2155.
- Sasaki, M., J. Anast, et al. (2003). "Bisulfite conversion-specific and methylation-specific PCR: a sensitive technique for accurate evaluation of CpG methylation." *Biochem Biophys Res Commun* **309**(2): 305-309.
- Schlenner, S. M., V. Madan, et al. (2010). "Fate mapping reveals separate origins of T cells and myeloid lineages in the thymus." *Immunity* **32**(3): 426-436.
- Schnittger, S., U. Bacher, et al. (2011). "Prognostic impact of FLT3-ITD load in NPM1 mutated acute myeloid leukemia." *Leukemia* **25**(8): 1297-1304.
- Schnittger, S., F. Dicker, et al. (2011). "RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis." *Blood* **117**(8): 2348-2357.
- Schultz, M. D., Y. He, et al. (2015). "Human body epigenome maps reveal noncanonical DNA methylation variation." *Nature* **523**(7559): 212-216.
- Scourzic, L., L. Couronne, et al. (2016). "DNMT3A(R882H) mutant and Tet2 inactivation cooperate in the deregulation of DNA methylation control to induce lymphoid malignancies in mice." *Leukemia* **30**(6): 1388-1398.

- Senbabaoglu, Y., G. Michailidis, et al. (2014). "Critical limitations of consensus clustering in class discovery." *Sci Rep* **4**: 6207.
- Shapiro R., S. R. E. a. W. M. (1970). "Reactions of uracil and cytosine derivatives with sodium bisulfite. A specific deamination method." *J. Am. Chem. Soc.* **92**: 3.
- Sharif, J., M. Muto, et al. (2007). "The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA." *Nature* **450**(7171): 908-912.
- Shearstone, J. R., R. Pop, et al. (2011). "Global DNA demethylation during mouse erythropoiesis in vivo." *Science* **334**(6057): 799-802.
- Shi, Y., F. Lan, et al. (2004). "Histone demethylation mediated by the nuclear amine oxidase homolog LSD1." *Cell* **119**(7): 941-953.
- Shukla, S., E. Kavak, et al. (2011). "CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing." *Nature* **479**(7371): 74-79.
- Skipper, M., A. Eccleston, et al. (2015). "Presenting the epigenome roadmap." *Nature* **518**(7539): 313.
- Smale, S. T. and J. T. Kadonaga (2003). "The RNA polymerase II core promoter." *Annu Rev Biochem* **72**: 449-479.
- Smallwood, S. A., S. Tomizawa, et al. (2011). "Dynamic CpG island methylation landscape in oocytes and preimplantation embryos." *Nat Genet* **43**(8): 811-814.
- Smith, M. L., J. D. Cavenagh, et al. (2004). "Mutation of CEBPA in familial acute myeloid leukemia." *N Engl J Med* **351**(23): 2403-2407.
- Smith, Z. D., M. M. Chan, et al. (2014). "DNA methylation dynamics of the human preimplantation embryo." *Nature* **511**(7511): 611-615.
- Song, C. X., K. E. Szulwach, et al. (2011). "Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine." *Nat Biotechnol* **29**(1): 68-72.
- Stadler, M. B., R. Murr, et al. (2011). "DNA-binding factors shape the mouse methylome at distal regulatory regions." *Nature* **480**(7378): 490-495.
- Tahiliani, M., K. P. Koh, et al. (2009). "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1." *Science* **324**(5929): 930-935.
- Takahashi, K., F. Wang, et al. (2017). "Preleukaemic clonal haemopoiesis and risk of therapy-related myeloid neoplasms: a case-control study." *Lancet Oncol* **18**(1): 100-111.
- Takeuchi, T., Y. Watanabe, et al. (2006). "Roles of jumonji and jumonji family genes in chromatin regulation and development." *Dev Dyn* **235**(9): 2449-2459.
- Tamaki, H., H. Ogawa, et al. (1999). "The Wilms' tumor gene WT1 is a good marker for diagnosis of disease progression of myelodysplastic syndromes." *Leukemia* **13**(3): 393-399.
- Tan, Y. T., Y. Sun, et al. (2016). "Deregulation of HOX genes by DNMT3A and MLL mutations converges on BMI1." *Leukemia* **30**(7): 1609-1612.
- Tang, H., S. An, et al. (2014). "Characterization of combinatorial histone modifications on lineage-affiliated genes during hematopoietic stem cell myeloid commitment." *Acta Biochim Biophys Sin (Shanghai)* **46**(10): 894-901.
- Taskesen, E., L. Bullinger, et al. (2011). "Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity." *Blood* **117**(8): 2469-2475.
- Tefferi, A., K. H. Lim, et al. (2009). "Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML." *Leukemia* **23**(7): 1343-1345.

- Tefferi, A., K. H. Lim, et al. (2009). "Mutation in TET2 in myeloid cancers." N Engl J Med **361**(11): 1117; author reply 1117-1118.
- Thol, F., I. Friesen, et al. (2011). "Prognostic significance of ASXL1 mutations in patients with myelodysplastic syndromes." J Clin Oncol **29**(18): 2499-2506.
- Trowbridge, J. J., J. W. Snow, et al. (2009). "DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells." Cell Stem Cell **5**(4): 442-449.
- Tsushima, H., M. Iwanaga, et al. (2012). "Late effect of atomic bomb radiation on myeloid disorders: leukemia and myelodysplastic syndromes." Int J Hematol **95**(3): 232-238.
- Wada, T., J. Kikuchi, et al. (2009). "Expression levels of histone deacetylases determine the cell fate of hematopoietic progenitors." J Biol Chem **284**(44): 30673-30683.
- Waddington, C. H. (1957). "The Strategy of the Genes." London: Allen and Unwin. Reprinted 2014.
- Waddington, C. H. (1959). "Canalization of development and genetic assimilation of acquired characters." Nature **183**(4676): 1654-1655.
- Wandt, H., T. Haferlach, et al. (2010). "WHO classification of myeloid neoplasms and leukemia." Blood **115**(3): 748-749; author reply 749-750.
- Wang, P., C. Lin, et al. (2009). "Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II." Mol Cell Biol **29**(22): 6074-6085.
- Wang, Y., M. Xiao, et al. (2015). "WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation." Mol Cell **57**(4): 662-673.
- Wang, Z., C. Zang, et al. (2008). "Combinatorial patterns of histone acetylations and methylations in the human genome." Nat Genet **40**(7): 897-903.
- Ward, P. S., C. Lu, et al. (2013). "The potential for isocitrate dehydrogenase mutations to produce 2-hydroxyglutarate depends on allele specificity and subcellular compartmentalization." J Biol Chem **288**(6): 3804-3815.
- Vardiman, J. W., N. L. Harris, et al. (2002). "The World Health Organization (WHO) classification of the myeloid neoplasms." Blood **100**(7): 2292-2302.
- Vastenhouw, N. L. and A. F. Schier (2012). "Bivalent histone modifications in early embryogenesis." Curr Opin Cell Biol **24**(3): 374-386.
- Weake, V. M. and J. L. Workman (2010). "Inducible gene expression: diverse regulatory mechanisms." Nat Rev Genet **11**(6): 426-437.
- Weinberg, R. A. (2014). "The Biology of Cancer " New York: Garland Science 2ed.
- Weisenberger, D. J., M. Velicescu, et al. (2004). "Role of the DNA methyltransferase variant DNMT3b3 in DNA methylation." Mol Cancer Res **2**(1): 62-72.
- Weissmann, S., T. Alpermann, et al. (2012). "Landscape of TET2 mutations in acute myeloid leukemia." Leukemia **26**(5): 934-942.
- Welch, J. S., T. J. Ley, et al. (2012). "The origin and evolution of mutations in acute myeloid leukemia." Cell **150**(2): 264-278.
- Wendt, K. S., K. Yoshida, et al. (2008). "Cohesin mediates transcriptional insulation by CCCTC-binding factor." Nature **451**(7180): 796-801.
- Veres, A., B. S. Gosis, et al. (2014). "Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing." Cell Stem Cell **15**(1): 27-30.
- Verhaak, R. G., C. S. Goudswaard, et al. (2005). "Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance." Blood **106**(12): 3747-3754.

- Whitman, S. P., A. S. Ruppert, et al. (2008). "FLT3 D835/I836 mutations are associated with poor disease-free survival and a distinct gene-expression signature among younger adults with de novo cytogenetically normal acute myeloid leukemia lacking FLT3 internal tandem duplications." *Blood* **111**(3): 1552-1559.
- Vire, E., C. Brenner, et al. (2006). "The Polycomb group protein EZH2 directly controls DNA methylation." *Nature* **439**(7078): 871-874.
- Visel, A., M. J. Blow, et al. (2009). "ChIP-seq accurately predicts tissue-specific activity of enhancers." *Nature* **457**(7231): 854-858.
- Visel, A., E. M. Rubin, et al. (2009). "Genomic views of distant-acting enhancers." *Nature* **461**(7261): 199-205.
- Vogelstein, B., N. Papadopoulos, et al. (2013). "Cancer genome landscapes." *Science* **339**(6127): 1546-1558.
- Wu, T. P., T. Wang, et al. (2016). "DNA methylation on N(6)-adenine in mammalian embryonic stem cells." *Nature* **532**(7599): 329-333.
- Xie, M., C. Lu, et al. (2014). "Age-related mutations associated with clonal hematopoietic expansion and malignancies." *Nat Med* **20**(12): 1472-1478.
- Xu, C., C. Bian, et al. (2011). "The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain." *Nat Commun* **2**: 227.
- Xu, J., Y. Y. Wang, et al. (2014). "DNMT3A Arg882 mutation drives chronic myelomonocytic leukemia through disturbing gene expression/DNA methylation in hematopoietic cells." *Proc Natl Acad Sci U S A* **111**(7): 2620-2625.
- Xu, J., J. A. Watts, et al. (2009). "Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells." *Genes Dev* **23**(24): 2824-2838.
- Xu, Y., F. Wu, et al. (2011). "Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells." *Mol Cell* **42**(4): 451-464.
- Yan, X. J., J. Xu, et al. (2011). "Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia." *Nat Genet* **43**(4): 309-315.
- Yang, L., Y. Han, et al. (2007). "A tumor suppressor and oncogene: the WT1 story." *Leukemia* **21**(5): 868-876.
- Yang, L., R. Rau, et al. (2015). "DNMT3A in haematological malignancies." *Nat Rev Cancer* **15**(3): 152-165.
- Yoder, J. A., C. P. Walsh, et al. (1997). "Cytosine methylation and the ecology of intragenomic parasites." *Trends Genet* **13**(8): 335-340.
- Yoshizato, T., B. Dumitriu, et al. (2015). "Somatic Mutations and Clonal Hematopoiesis in Aplastic Anemia." *N Engl J Med* **373**(1): 35-47.
- You, J. S. and P. A. Jones (2012). "Cancer genetics and epigenetics: two sides of the same coin?" *Cancer Cell* **22**(1): 9-20.
- Zhang, X., J. Su, et al. (2016). "DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells." *Nat Genet* **48**(9): 1014-1023.
- Zhu, J. and S. G. Emerson (2002). "Hematopoietic cytokines, transcription factors and lineage commitment." *Oncogene* **21**(21): 3295-3313.
- Zhu, W. G., K. Srinivasan, et al. (2003). "Methylation of adjacent CpG sites affects Sp1/Sp3 binding and activity in the p21(Cip1) promoter." *Mol Cell Biol* **23**(12): 4056-4065.

Zitzmann, F., D. Mayr, et al. (2014). "Frequent hypermethylation of a CTCF binding site influences Wilms tumor 1 expression in Wilms tumors." Oncol Rep **31**(4): 1871-1876.