DISSERTATION

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

Justyna Anna Wierzbińska, M.Sc. in Biotechnology

born in: Mogilno, Poland

Oral-examination: ……

A new approach to analyze cancer-specific DNA methylation data by considering the cellular origin

Referees:

PD Dr. Odilia Popanda

Prof. Dr. Christoph Plass

Division of Epigenomics and Cancer Risk Factors

Head of Division: Prof. Dr. Christoph Plass

German Cancer Research Center (DKFZ)

Heidelberg, Germany

Dedicated to my friends and family who have supported me throughout my time as a doctoral student

CONTRIBUTIONS

Table 1-1 in the **Section 1.3.3** is partially based on the review 'Epigenetic deregulation in chronic lymphocytic leukemia: Clinical and biological impact' by Mansouri, Wierzbinska, Plass, and Rosenquist published in the journal Seminars in Cancer Biology (2018). Therefore, part of the text in the table might contain suggestions and corrections from co-authors.

Small RNA sequencing data analysis was initially performed by Dr. Thomas Hielscher (Division of Biostatistics, DKFZ, Heidelberg, Germany). This data was downloaded and used for downstream analysis. The initial data source is described in the **Section 3.2.6**.

RNA-seq data analysis was performed using the pipeline written by Dr. Naveed Ishaque (Division of Theoretical Bioinformatics, DKFZ, Heidelberg, Germany). The data processing step is described in the **Section 3.2.7**.

Normal B cells from healthy individuals were sorted and provided by Dr. Marc Seifert (Molecular Genetics Group, Essen University Hospital, Essen, Germany). The sorting protocol is described in detail in the **Section 3.2.2**.

DECLARATIONS

Declarations according to § 8 (3) b) and c) of the doctoral degree regulations:

a) I hereby declare that I have written the submitted dissertation myself and in this process have used no other sources or materials than those expressly indicated,

b) I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this or any other format any other institution as an examination paper, nor submitted it to any other faculty as a dissertation.

Justyna Wierzbinska

TABLE OF CONTENTS

SUMMARY

Current approaches in the cancer field mainly focus on the identification of genetic alterations driving tumors. However, in many tumor types, including chronic lymphocytic leukemia (CLL), no underlying genetic mechanism has been identified in about 30% of the affected patients. Consequently, the focus of research has shifted towards epigenetic modifications, including aberrant DNA methylation as a potential driver contributing to tumorigenesis. Unlike genetic modifications, epigenetic alterations are potentially reversible, making them attractive targets for therapeutic interventions.

Chronic lymphocytic leukemia (CLL) is the most frequent leukemia in adults and originates from rapidly differentiating B cells, which undergo extensive epigenetic reprogramming during normal B cell differentiation. Every differentiation stage of a normal B cell is represented by unique patterns present at the DNA methylation level (methylation footprint), which is maintained and stably propagated in CLL. Consequently, this stable epigenetic patterning can serve as an indicator for the identification of the cell-of-origin for each individual CLL case. For the purpose of this thesis, I define the cancer cell-of-origin as the cell that acquires sufficient oncogenic hits (genetic and/or epigenetic) to initiate its tumorigenic growth defined as a measurable deviation from the normal B cell differentiation trajectory. This means that at least two factors contribute to the epigenetic patterns seen in CLL: first, epigenetic patterns which were present in the tumor-initiating B cell at the time of transformation, and second, CLL-specific epigenetic alterations that occur during leukemogenesis and, which may relate to genetic alterations or to aberrant signaling events that the leukemic cells acquire in response to extrinsic or intrinsic stimuli.

Defining CLL-specific epigenetic events, which are distinct from normal epigenetic B cell programming, is of utmost importance to understand the molecular alterations contributing to CLL. Previous studies in CLL have already proposed aberrant methylation events and attempted to describe their impact on the expression of both, protein-coding genes (e.g. *DAPK1*, *ZAP70*, *ID4*) and microRNAs (e.g. miR-9, miR-181a/b, miR-34a, miR-708). However, all these studies defined aberrant methylation events based on the comparison of CLL methylomes with those of peripheral blood $CD19^+$ B cells as a control. As a result, these studies completely neglected the massive epigenetic programming that occurs during normal B cell differentiation. Therefore, novel approaches aiming at identifying truly CLL-specific methylation changes considering the highly dynamic methylome during normal B cell differentiation were urgently needed.

In this thesis, I used linear modeling to describe the continuum of epigenetic alterations occurring during normal B cell differentiation. DNA methylomes of CLL cells were subsequently precisely positioned into the normal B cell differentiation trajectory to define the DNA methylomes of the cell-of-origin for every CLL patient. Considering this cellular origin, I identified CLL-specific methylation events as well as epigenetic alterations reflecting on normal B cell differentiation. The relevance of this approach was demonstrated by contrasting the number of epigenetically deregulated miRNAs and protein-coding genes to those determined using bulk $CD19^+$ cells from peripheral blood as controls. This analysis highlighted the extent of

overcalling of leukemia-specific methylation changes in previous studies and highlights the importance of the use of proper control cells for the identification of disease-specific DNA methylation events. The analytical approach described in this thesis provides a general framework for the identification of the cancer cell-of-origin that could be applied in the future to other cancer entities.

ZUSAMMENFASSUNG

Die Suche nach den Ursachen von Krebserkrankungen beschränkt sich hauptsächlich auf die Identifizierung von genetischen Veränderungen. Für einige Tumorentitäten, hierunter auch 30% der Patienten mit chronischer lymphatischer Leukämie (CLL), konnten jedoch keine solche Veränderungen beschrieben werden. Somit gelangten anormale epigenetische Modifikationen, wie DNA-Methylierung, als Ursache von Tumoren in den Fokus der Forschung. Im Gegensatz zu genetischen Veränderungen sind epigenetische Modifikationen reversibel und somit attraktive Ansatzpunkte für Therapien.

Die Chronische Lymphatische Leukämie ist die am häufigsten auftretende Form der Leukämie in Erwachsenen. Als Ursprung gelten B-Zellen, welche schnelle differenzieren und eine extensive, epigenetische Reprogrammierung durchlaufen. Jeder Differenzierungszustand einer normalen B-Zelle ist durch ein einzigartiges DNA-Methylierungs-Muster gekennzeichnet, das in der CLL erhalten und stabil propagiert wird. Das Methylom kann daher als Indikator für die Identifizierung der Ursprungszelle für individuelle CLL-Fälle verwendet werden. In der vorliegenden Arbeit wird die Ursprungszelle als jene Zelle definiert, welche ausreichend onkogene, genetische sowohl als auch epigenetische, Veränderungen aufweist, um ein tumoröses Wachstum einzuleiten. Als tumoröses Wachstum wird hierbei eine messbare Abweichung von der normalen B-Zell Differenzierung definiert.

Neben dem epigenetischen Muster der Tumor-initiierenden Zelle zum Zeitpunkt der Transformation definieren CLL-spezifische epigenetische Veränderungen das CLL-Epigenom. Letztere entstehen während der Leukemogenese und resultieren aus genetischen Veränderungen oder Reaktionen auf extrinsische sowohl als auch intrinsische Stimuli. Um die molekularen Veränderungen zu verstehen, die zur Entstehung von CLL beitragen, ist es notwendig CLL-spezifische Veränderungen zu definieren, welche unabhängig von den epigenetischen Veränderungen während einer normalen B-Zell Differenzierung auftreten. Frühere Studien beschrieben bereits anormale Methylierung als Ursache für Veränderungen in der Expression von Protein-Kodierenden Genen sowie microRNAs in CLL. Jedoch beruhen diese Studien auf Vergleichen von CLL-Methylomen mit denen von $CD19^+$ Zellen des peripheren Blutes, wodurch die massiven epigenetischen Veränderungen während einer normalen B-Zell Differenzierung vollkommen vernachlässigt wurden. Neue Herangehensweisen streben nun nach der Identifizierung von wirklichen CLL-spezifischen Methylierungsänderungen unter Berücksichtigung des dynamischen Methyloms differenzierender B-Zellen.

In der vorliegenden Arbeit wurden lineare Modelle angewandt um das Kontinuum epigenetischer Veränderungen während normaler B-Zell Differenzierung mathematisch zu beschreiben. Um das Methylom von dem CLL-Ursprungszellen zu bestimmen, wurden CLL-Methylome anschließend präzise in das Kontinuum der normalen B-Zell Differenzierung eingeordnet. Unter Berücksichtigung des Methyloms der Ursprungszelle wurden epigenetische Veränderungen, die die normale B-Zell Differenzierung widerspiegeln, und CLL-spezifische Veränderungen des Methylierungsmusters, identifiziert. Der Vergleich der epigenetisch deregulierten microRNAs und Protein-Kodierenden Gene, welche in dieser Arbeit identifiziert wurden, mit früheren Studien, in denen $CD19^+$ Zellen aus peripheren Blut als Kontrolle verwendet wurden, unterstreicht die Relevanz dieser Arbeit. Zudem stellt die

vorliegende Arbeit den hohen Anteil von überbewerteten Leukämie-spezifischen Methylirungsänderungen heraus und macht die Bedeutung korrekter Kontrollen für die Identifizierung von erkrankungsbedingten Veränderungen der DNA-Methylierung deutlich. Das analytische Vorgehen, welches in der vorliegenden Arbeit beschrieben wird, dient als Basis für die Identifizierung von Tumor-Ursprungszellen und könnte in Zukunft auch für weitere Tumorentitäten von Bedeutung werden.

LIST OF ABBREVIATIONS

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
Ago-2	Argonaute-2
AID	Activation-Induced Cytidine Deaminase
ALL	Acute Lymphoblastic Leukemia
alloHSCT	Allogeneic Hematopoietic Stem Cell Transplantation
BCR	B-cell Receptor
BER	Base Excision Repair
BT	Bisulfite Treatment
BL	Burkitt Lymphoma
BMF	B-cell Lymphoma 2-modifying Factor
CGI	CpG Islands
CLL	Chronic Lymhocytic Leukemia
CMML	Chronic Myelomonocytic Leukemia
DMSO	Dimethyl Sulfoxide
DTT	Dithiothreitol
DNMTs	DNA Methyltransferases
dNTPs	Deoxynucleoside Triphosphates
ESMO	European Society for Medical Oncology
FCR	Fludarabine, Cyclophosphamide and Rituximab
FISH	Fluorescence *In Situ* Hybridization
FL	Follicular Lymphoma
HE	Housekeeping RNA Expression
HP-CLL	High-programmed CLL
HSCs	Hematopoietic Stem Cells
ICGC	International Cancer Genome Consortium
i-CLL	Intermediate CLL
ICR	Imprinting Control Region
IP-CLL	Intermediate-Programmed CLL
IVT	*In Vitro* Transcription
kb	Kilobases
LINEs	Long Interspersed Nuclear Elements
LP-CLL	Low-Programmed CLL
M. SssI	CpG Methyltransferase
MACS	Magnetic Cell Sorting
MALDI-TOF-MS	Matrix-assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry
Mb	Megabase
MBD	Methyl-CpG-binding Domain
MCL	Mantle Cell Lymphoma
M-CLL	*IGHV* mutated-CLL
m-CLL	MBC-like CLL
ME	miRNA Expression
MM	Multiple Myeloma
NBC	Naïve B cell
GCF	Germinal Center Founder B cell
loMBC	Early Non Class-switched Memory B cell
intMBC	Non Class-switched Memory B cell
ncsMBCs	Non Class-switched Memory B cells
sMGZ	Splenic Marginal Zone B cells
hiMBC	Class-switched Memory B cells

n-CLL	NBC-like CLL
NCT	National Center for Tumor Diseases
PB	Peripheral blood
PBMCs	Peripheral Blood Mononuclear cells
PC	Prinicipal Component
PCA	Prinicipal Component Analysis
PRC2	Polycomb Repressive Complex 2
qRT-PCR	Real-time Quantitative PCR
RPM	Reads Per Million
rlog	Regularized Log Transformation
RNA-seq	RNA-sequencing
RT	Room Temperature
SAP	Shrimp Alkaline Phosphatase
scMBCs	Class-switched Memory B cells
SINEs	Short Interspersed Nuclear Elements
sMGZs	Splenic Marginal Zone B cells
SNVs	Single Nucleotide Variants
TCGA	The Cancer Genome Atlas
TDG	Thymine-DNA Glycosylase
tDMRs	Tissue-specific Differentially Methylated Regions
TET	Ten-eleven Translocation
TFs	Transcription Factor
TRBP	Transactivation-responsive RNA-binding Protein
TSGs	Tumor Suppressor Genes
TSS	Transcriptional Start Site
TTFT	Time-to-first Treatment
U-CLL	Unmutated *IGHV*-CLL
WGBS	Whole-genome Bisulfite Sequencing
ΔMethylation	Methylation Difference

1. INTRODUCTION

1.1 Chronic lymphocytic leukemia

Chronic lymphocytic leukemia (CLL) is the most common type of adult leukemia in the western world (30% of all leukemia cases), and it is characterized by a progressive accumulation of monoclonal $CD19^+/CD5^+/CD23^+$ B cells in the peripheral blood, bone marrow and secondary lymphoid organs. CLL is a cancer of the elderly with a median age at diagnosis of seventy-one years and a clear gender bias, with a male to female ratio of 2:1 [1, 2].

Diagnosis of CLL

The diagnosis of CLL is based on peripheral blood counts, blood smears and immunophenotyping [3] and requires a sustained lymphocytosis (≥5000 B lymphocytes/µL) in the peripheral blood for at least 3 months. In blood smears, the leukemic B cells appear as small, mature lymphocytes with a scant cytoplasm and rounded nuclei that are characterized by a dense chromatin and lack of discernible nucleoli. Frequently, blood smears of CLL patients show the presence of smudge cells, which are abnormally looking, ruptured B lymphocytes [3].

Immunophenotypically, CLL cells show co-expression of the T-cell antigen CD5 together with the B-cell surface antigens CD19 and CD23, low expression of CD20 and CD79b and immunoglobulin light-chain restriction [3, 4].

Prognostic features of CLL

The clinical outcome of CLL patients is highly variable. Some of the patients have an indolent disease course and may survive for decades without any need for therapy, whereas other patients present with aggressive disease and die within a few years despite treatment. Up to now, two staging systems, Rai and Binet, developed more than 30 years ago, have been routinely used in clinical practice [2, 5, 6]. However, these systems fail to faithfully discriminate between aggressive and indolent CLL cases since the majority of patients is asymptomatic at the time of diagnosis and classified as a low-stage disease (80%) [2, 7].

Therefore, recent efforts focused on an improved risk stratification for CLL, implementing a variety of clinical (lymphocyte doubling time, bone marrow infiltration pattern, serum thymidine kinase and β2-microglobulin levels) and expression markers (ZAP70, CD38 and CD49d expression levels) [8]. Likewise, molecular prognostic markers are broadly used. Among them: mutational status of a handful of genes (*IGHV*, *TP53*, *NOTCH1*, *SF3B1*, *BIRC3*, and *EGR2*) [9-22] or the presence of genomic aberrations (del(11q), del(13q), del(17p) and trisomy12) [23]. Currently, the mutational status of immunoglobulin heavy variable (*IGHV*) gene is the most commonly used molecular prognostic marker. It has been shown that patients carrying somatic hypermutations in the *IGHV* genes (M-CLL) have a superior overall survival and longer time to first treatment (TTFT), compared to those with unmutated *IGHV* status (U-CLL) [19, 21].

Currently, fluorescence *in situ* hybridization (FISH) analysis to detect relevant genomic aberrations (i.e. del(11q), del(13q), del(17p) and trisomy12) and *TP53* mutational screening are recommended before the initiation of therapy [2]. The presence of *TP53* aberrations i.e. del(17p) and/or *TP53* mutations has been linked to poor prognosis and resistance to chemoimmunotherapy [2, 24]. Del(13q), which occurs in more than half of all CLL cases, has been associated with a particularly favorable clinical outcome and good response to chemoimmunotherapy with prolonged progression-free and overall survival [2, 25]. Del(11q) and trisomy 12 have been associated with an intermediate disease course [2, 23, 26].

Treatment of CLL

For decades, the front-line therapy for CLL was based on alkylating agents (chlorambucil) [27] and on purine analogs (fludarabine, pentostatin and cladribine) [28-30]. This approach has been challenged recently due to the development of highly active agents, including next-generation inhibitors of B-cell receptor signaling (BCR), Bcl-2 antagonists and anti-CD20 antibodies [4].

Current treatment guidelines, according to the European Society for Medical Oncology (ESMO), suggest a 'watch-and-wait' strategy for early-stage, asymptomatic CLL patients [31], since early treatment with chemotherapeutic agents in asymptomatic patients does not translate into improved survival rates [32]. The presence of one of the following signs should prompt initiation of CLL treatment: significant disease-related symptoms (fever without infection, weight loss or severe fatigue), lymphadenopathy, splenic or hepatic enlargement, lymphocyte doubling count <6 months and progressive anemia and/or thrombocytopenia [31].

In symptomatic patients, the choice of treatment is based on the clinical status of the patient, which includes a history of infections, comorbidities and the presence of del(17p) or *TP53* mutations. Physically fit patients without *TP53* aberrations should receive chemoimmunotherapy consisting of fludarabine, cyclophosphamide and rituximab (FCR). However, in older patients or in case of the presence of significant comorbidities, a combination of chlorambucil and anti-CD20 antibody (rituximab, ofatumumab or obinutuzumab) is recommended. On the contrary, patients carrying *TP53* abnormalities have a poor prognosis with a disease resistant to standard therapeutic regimens. These patients should be treated with next-generation BCR inhibitors (ibrutinib, idelalisib) or with a combination of idelalisib and anti-CD20 antibody, rituximab. Additionally, CLL patients achieving remission should be considered for allogeneic hematopoietic stem cell transplantation (alloHSCT) [31].

Interestingly, several epigenetic compounds including histone deacetylase inhibitors and demethylating agents have been tested in CLL [33-39]. However, lack of response has hindered their broader use in the clinical setting. The low efficacy of these agents is most likely due to the low proliferative rate of CLL cells or low expression levels of hENT1 transporters that are indispensable for the decitabine/azacitidine uptake by the cells [40]. Another aspect is the global hypomethylation pattern detected in CLL [41, 42], suggesting that demethylation therapies may not be optimal for CLL treatment.

1.2 The genome of CLL

Compared to other tumor entities, CLL has been shown to have a rather stable genome [23, 43]. At the time of diagnosis, the vast majority of CLL patients (80%) carry no or very few chromosomal aberrations. The remaining 20% of cases exhibit complex karyotypes, with three or more genomic abnormalities [44]. The most frequently observed chromosomal aberrations are partial losses of chromosomes, such as deletions of 13q14 (50-60%), 11q22-q23 (~20%), 17p (~10%), 6q15-q21 (6%), or gain of chromosome 12 (16%) [23]. The deleted chromosomal segments contain functionally important genes. For instance, both del13q14 and del17p13 disrupt the function of crucial tumor suppressor genes in CLL, miR-15a/16a and *TP53*, respectively [45]. Likewise, del11q22-23 affects the *ataxia telangiectasia mutated* (*ATM*) gene, which is known as a safeguard of genomic integrity, and which leads to genomic instability in this subgroup of CLL [46-49].

In contrast to other B-cell neoplasms, in CLL, recurrent chromosomal translocations are rare, which is most likely related to its cellular origin from post-germinal center B cells [23, 50-52]. For instance, in the case of another translocation-rich B cell cancer entity, non-Hodgkin lymphoma, translocations are the result of aberrant, erroneous germinal center processes, class switch recombination and somatic hypermutation, two mechanisms that are active in the germinal center B cells [53].

Puente and colleagues have comprehensively evaluated the genomic landscape of CLL showing that the frequencies of acquired single nucleotide variants (SNVs) are low relative to other cancer entities, with an average mutation burden of 0.87 mutations per megabase (Mb) [54]. The number of somatic variants was higher in *IGHV*-mutated than in *IGVH*-unmutated tumors. Integrated analysis of mutations in 452 CLL cases identified three main molecular signatures: 1) an age-associated signature with C-to-T transitions affecting CpG dinucleotides, 2) T:A > G:C transversions, and 3) an activation-induced cytidine deaminase (AID) signature. Eight pathways are frequently affected by mutations in CLL: BCR signaling, apoptosis, cell cycle regulation, DNA damage response, NOTCH1 signaling, chromatin remodeling, NF-κB signaling, and RNA metabolism [54] (**Figure 1-1**). The most frequently mutated genes are *NOTCH1* (12.6%), *ATM* (11%), *SF3B1* (~9%), *BIRC3* (~9%), *CHD2* (6%), *TP53* (~5-6%) and *MYD88* (4%). Novel recurrently mutated genes, previously not linked to the disease, have been reported, including *EGR2*, *ZNF292*, *ARID1A*, *SETD2*, *ZMYM3*, *PTPN11*, *KRAS*, *NRAS*, *CDKN1B*, *CDKN2A* and *IKZF3*. Together, these analyses have demonstrated that a number of epigenetic regulators, including *CHD2*, *SETD2*, *ARID1A*, *ASXL1*, *SETD1A*, *MLL2*, *ZMYM3*, and *HISTH1B* are frequently affected by mutations in CLL, highlighting the interplay of genetic and epigenetic mechanisms in CLL [54] (**Figure 1-1**).

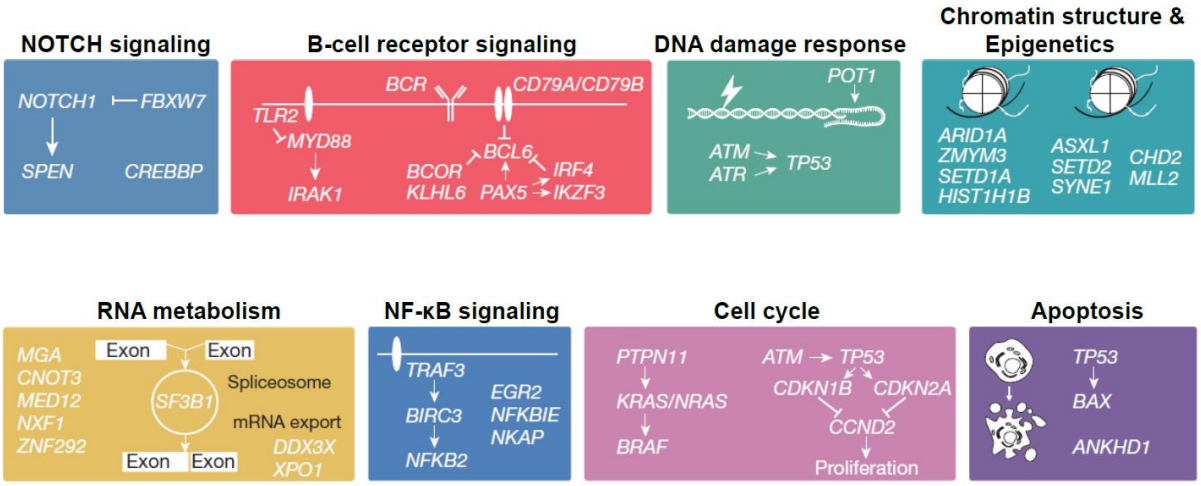


Figure 1-1: Significantly mutated genes and pathways in CLL. Schematic representation of the major molecular pathways altered in CLL. The most recurrently mutated genes in CLL are grouped according to the biological pathways involved. Details are given in the text (modified from [54]).

In the future, it will be important to further investigate the functional role of gene mutations affecting epigenetic regulators, but also to get a more comprehensive view on the mutations affecting non-coding regions in the CLL genome. There are several recent reports highlighting the contribution of mutations affecting enhancers and promoters to establish aberrant gene expression patterns in CLL [2]. For instance, mutations targeting an enhancer region regulating the B-cell specific transcription factor, *PAX5*, have been linked to decreased expression levels of *PAX5* in CLL [54]. The same *PAX5* enhancer is also frequently altered in other B-cell neoplasms, e.g. in follicular lymphoma (23%) and mantle-cell lymphoma (5%), suggesting that this may be a general driver event in B-cell malignancies [2, 54]. Likewise, Kandaswamy and colleagues linked the single nucleotide polymorphism rs539846 to a CLL risk locus on chromosome 15 (15q15.1), and to the RELA binding site within the B-cell super-enhancer region of B cell lymphoma 2-modifying factor (*BMF*). The authors demonstrated that an allelic variant in rs539846 alters the conserved RELA binding site which results in decreased enhancer activity, reduced RELA binding and decreased *BMF* expression and ultimately leads to elevated expression of *BCL-2* [2, 55]. Most recently, non-coding mutations recurrently targeting regulatory regions of genes involved in B-cell development (*BCL6*, *IKZF1*, *PAX5*), NF-κB signaling (*VOPP1*, *TCL1A*, *BACH2*, *BIRC3*), DNA damage response (*BTG2*, *BCL2*) and NOTCH signaling (*HDAC9*) have been reported [56].

1.3 Epigenetics in CLL

1.3.1 Definition of epigenetics

The term “epigenetics” was introduced by Conrad Waddington in 1942 and defined as ‘the branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being’ [57, 58]. Over time, this definition has gradually narrowed to ‘the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence’ [59].

Epigenetic marks are established during development and differentiation and are stably maintained during cell divisions, enabling tissue-specific gene expression patterns despite identical genetic information present in each cell of an organism. To date, several epigenetic mechanisms involved in regulation of gene expression patterns have been identified. These include DNA modifications (cytosine methylation and hydroxymethylation) [60-62], non-coding RNAs (microRNAs, long non-coding RNAs) [63, 64], post-translational modifications of histone tails [65, 66], nucleosome positioning [67], and 3D genome organization [68, 69]. The epigenetic regulation of gene expression has been extensively characterized by Chen *et al.* [70] and Jaenisch *et al.* [71]. In this thesis, I will focus on DNA methylation to characterize epigenetic patterns during normal B cell differentiation and in CLL.

1.3.2 DNA methylation and transcriptional regulation

DNA methylation is by far the most extensively studied epigenetic mark. Although, initially, it was hypothesized that DNA methylation may play a role in the regulation of gene expression, it was not until the 1980s when two groups, Holliday and Pough [72], and Compere and Palmiter [73] demonstrated that DNA methylation is dynamically regulating gene expression patterns and therefore is essential for cellular differentiation. Recent studies have presented a broad spectrum of functions for DNA methylation: programming of DNA methylation patterns play an important role in embryonic development [74, 75] and in genomic imprinting [76], as well as in X chromosome inactivation [77], RNA processing [78, 79], and in the maintenance of genomic integrity by repressing transcription from transposable elements [80-82].

DNA methylation refers to a chemical modification of the pyrimidine ring of cytosine, which in mammalian cells is usually found in the context of CpG dinucleotides [83]. The chemical reaction is catalyzed by a family of enzymes called DNA methyltransferases (DNMTs). DNA methylation is established *de novo* by DNMT3A and DNMT3B enzymes, while DNMT1 plays a role in the maintenance of methylation patterns during cell division, using hemimethylated DNA generated during DNA replication as a template. DNMTs catalyze the addition of a methyl group (-CH3) to the fifth position of the cytosine ring, yielding 5-methylcytosine (5mC). The 5mC mark is faithfully maintained during cell divisions until it is removed by passive or active de-methylation processes. Ten-eleven translocation (TET) proteins are involved in active de-methylation, catalyzing the oxidation of 5mC to 5hmC (5-hydroxymethylcytosine), and further to 5fC (5-formylcytosine) and/or 5caC (5-carboxylcytosine). The final step of such a demethylation would be base-excision facilitated by thymine-DNA glycosylase (TDG), which detects 5fC and 5caC and subsequent repair of the abasic site. Passive DNA de-methylation occurs as a failure to actively maintain DNA methylation marks upon DNA replication [84, 85] (**Figure 1-2**).

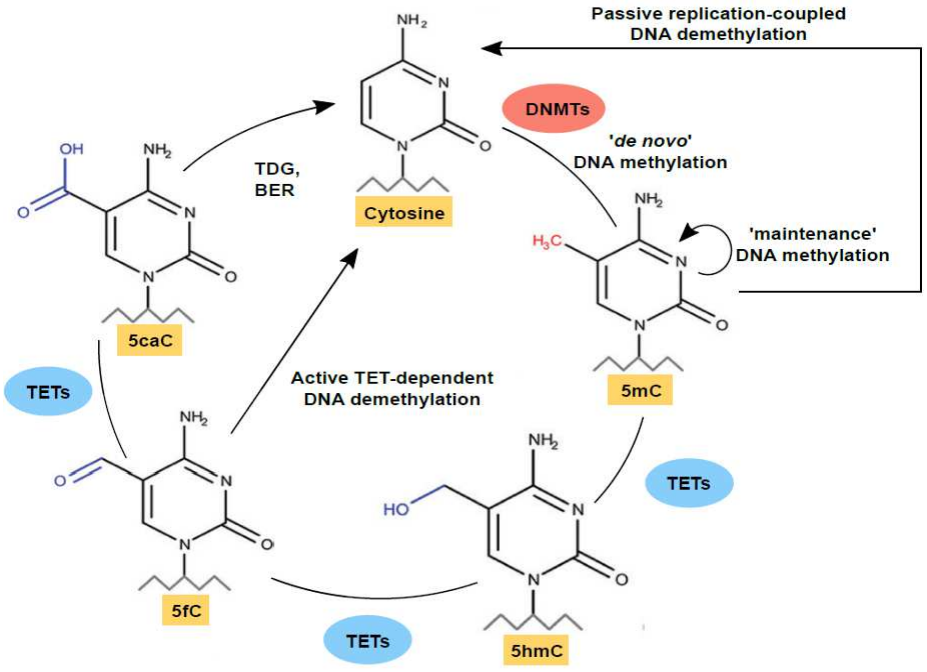


Figure 1-2: DNA methylation and demethylation. DNA methyltransferases (DNMTs) catalyze the addition of a methyl group to the position C5 of the cytosine ring. The process of demethylation is catalyzed by Ten-eleven translocation (TETs) enzymes. At first, TETs catalyze the oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). Subsequent oxidative reactions lead to conversion of 5hmC into 5-formylcytosine (5fC) and 5-carboxylcytosine (5hmC). Further, 5fC and 5caC marks are actively reverted to unmodified cytosine through thymine-DNA glycosylase (TDG)-mediated base excision repair (BER) or passively during DNA replication.

Cytosine methylation is abundant in mammals and is mainly established in the context of CpG dinucleotides, although the presence of non-CpG methylation was demonstrated in neuronal and embryonic stem cells [86-88]. DNA methylation also occurs in invertebrates, where it is described as 'mosaic methylation': highly methylated regions are interspersed with methylation-free domains [89, 90].

In total, the human genome contains about 2.9 x 10^7 CpG dinucleotides. On a global level, this means that the human genome is depleted in CpGs, meaning that the frequency of CpG dinucleotides is lower than what would be expected by chance. CpG dense regions, known as CpG islands (CGI) form an exception from this. CGI are defined, according to Gardiner-Garden and Frommer [91] as DNA sequences longer than 200bp, with a GC content (%GC) greater than 50%, and an observed/expected ratio of CpG to GpC higher than 60% [91]. There are about 29,000 CGI in the human genome, frequently coinciding with transcription start sites (TSS) of annotated genes [92]. More than 60% of protein-coding genes in the human genome associated with CGI. Typically, CGI mark promoters of house-keeping genes but can also drive the expression of tissue-specific and of developmental genes [93-95]. The remaining CGI are located within (intragenic) or between (intergenic) transcripts, and are known as 'orphan CGI'. Interestingly, many of the 'orphan CGI' show characteristics of functional promoters [96]. However, frequently they become methylated during development, losing their transcriptional-initiation properties [96].

In general, the majority of CpG sites in the human genome are methylated. CpG dinucleotides located in CGI are an exception to this, as they are normally unmethylated. The most heavily methylated regions (~85%) are associated with repetitive elements such as short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), and satellite DNA sequences in pericentromeric regions. Furthermore, DNA methylation levels are linked to replication timing [97-99]. In contrast to late-replicating genomic regions, early replicating regions are characterized by higher methylation levels which are faithfully maintained during cell divisions. Hence, repeated cell divisions lead to a gradual loss of methylation in the late replicating, heterochromatic regions [97-99].

The presence of DNA methylation in promoter regions correlates with lack of transcription [100]. The question remains whether this simply reflects the lack of transcription or whether promoter DNA methylation is able to actively repress transcription. According to the first model, promoters of active genes lack DNA methylation: the feature that has been linked to a binding of activating transcription factors (TFs). In this scenario, high DNA methylation levels in the promoters of inactive genes are the consequence of the absence of the binding of transcriptional activator [101-104]. According to the repression model, DNA methylation directly or indirectly interferes with binding or procession of the transcriptional machinery. The latter can occur through the recruitment of methyl-CpG-recognizing transcriptional repressors, either containing methyl-CpG-binding domain (MBD) such as MeCP2, MBD1, MBD2 and MBD4 or zinc-finger domains, e.g. KAISO, ZBTB4, ZBTB38, and ZFP57. These proteins can further recruit chromatin repressors, and therefore mediate transcriptional silencing ([105-107], reviewed in [108]).

Recent studies have shown that methylation levels are widely modulated not only in promoter regions but also in intragenic and intergenic regions. It became evident that DNA methylation within gene bodies is involved in the regulation of multiple processes, e.g. transcript elongation [109, 110], enhancer activation [41, 87, 111, 112], expression of intragenic coding and non-coding transcripts [113-117] as well as alternative splicing [78, 118-124]. Multiple genome-wide epigenetic studies have reported a positive correlation between intragenic DNA methylation and gene expression levels, both in the course of normal development and during cancer progression [41, 87, 110, 125-127].

The importance of DNA methylation is emphasized by the growing number of reports showing deregulated DNA methylation patterns in human diseases, including cancer (reviewed in [128, 129]). In 1983, for the first time, aberrant DNA methylation has been linked to cancer. At that time, two seminal studies by Feinberg and Vogelstein, and Gama-Sosa *et al.* have shown that a substantial proportion of CpGs was hypomethylated in cancer cells as compared to their normal counterparts [130, 131]. These findings have been further complemented by recent genome-wide DNA methylation studies, in which global hypomethylation profiles were reported across different tumor entities, e.g. chronic lymphocytic leukemia, medulloblastoma and colon cancer [41, 132, 133].

Hypomethylation of single genes, such as *S100A4* in colon cancer or *SNCG* in breast and ovarian cancer, leads to their transcriptional activation [134, 135]. Global DNA hypomethylation, as it is often observed in cancer, occurs mainly in repetitive regions and gene bodies, and leads to genomic instability, loss of imprinting and reactivation of

transposable elements (**Figure 1-3**) (reviewed in [128, 129]). Together, DNA hypomethylation in cancer supports the acquisition of genomic aberrations and (re-)activates genes that are essential for tumor cell survival and metastasis [129].

In contrast to aberrant DNA hypomethylation events, gene-specific hypermethylation in cancer occurs predominantly in and around CpG-rich regions including CGI, which are usually unmethylated in normal somatic cells (**Figure 1-3**) [136]. This results in transcriptional silencing of growth regulatory genes, including tumor suppressor genes (TSGs), e.g. *BRCA1* in breast cancer, *RB1* in hereditary retinoblastomas or *APC* in colorectal cancer (reviewed in [137]). Overall, tumor-specific hypermethylation targets genes involved in cell signaling, cell-cycle regulation, chromatin remodeling, DNA repair, transcription and apoptosis [138]. It is very likely that aberrant gain of methylation contributes to transformation as hypermethylation events are already detectable in the earliest precursor lesions [139, 140].

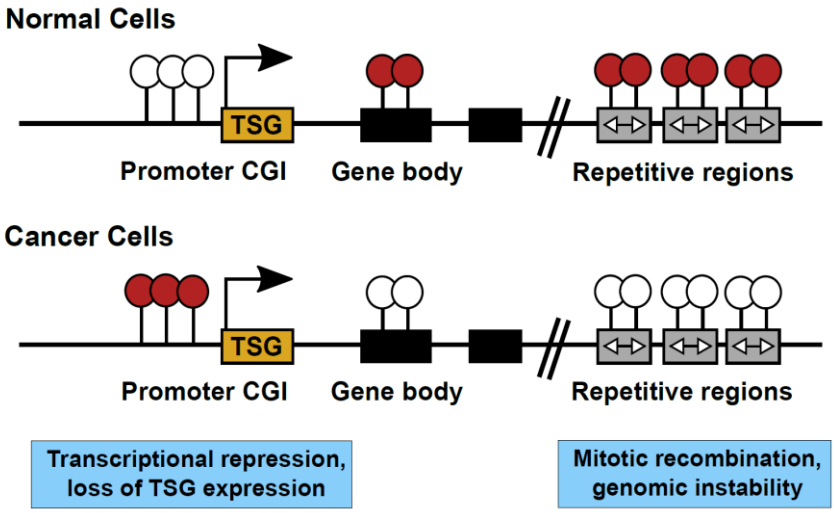


Figure 1-3: DNA methylation in normal and cancer cells. In normal cells, promoter-associated CGIs are lowly methylated (white circles) allowing gene transcription, while CpGs located in gene bodies and repetitive regions are heavily methylated (red circles). In tumor cells, *de novo* methylation of promoter-associated CGI leads to transcriptional silencing of growth-regulatory genes, e.g. tumor suppressor genes, while gene bodies and repeat-rich regions become hypomethylated. TSG, tumor suppressor gene; CGI, CpG island.

1.3.3 CLL methylome

Epigenetic research in CLL started from gene-specific studies. In 1988, Lipsanen *et al.* as first ones reported epigenetic deregulation in CLL. Hypomethylation of the *ornithine decarboxylase* (*ODC*) and the *erb-A1* genes was detected in samples from CLL patients as compared to leukocytes from the healthy donors [2, 141]. Likewise, promoter hypomethylation of *TCL1* was reported to induce higher expression levels of this gene in CLL [142]. Given that *TCL1* overexpression resulted in CLL development in the Eμ-TCL1 mice models [143], this finding was a clear demonstration of the causative role of promoter hypomethylation in CLL development.

With the exception of few reports on promoter hypomethylation, studies on aberrant DNA methylation in CLL have been dominated by the primary focus of understanding promoter hypermethylation. As a result, a number of targets of clinical and biological importance have been identified. Among the genes affected by hypermethylation are tumor suppressors, TFs, genes regulating cell survival and proliferation, and microRNAs, for some of which prognostic relevance has been demonstrated (reviewed in [144]).

In CLL, one of the first descriptions of an aberrantly methylated promoter was published for *E-cadherin* (*CHD1*) in 2000 by Melki and colleagues. Hypermethylation correlated with reduced or absent expression of *CHD1* in CLL as compared to control B cells [145]. Promoter hypermethylation of the catalytic subunit of telomerase, *human telomerase reverse transcriptase* (*hTERT*), was subsequently reported in 2002. The gain of methylation correlated with lower telomerase activity, shorter telomere length and superior overall survival [2, 146]. In 2005, Raval and colleagues demonstrated differential methylation between *IGHV* mutated (M-CLL) and unmutated CLL (U-CLL) patients in the promoter region of the transcription factor *TWIST2*, a regulator of p53. In that study aberrant promoter hypermethylation was mainly detected in M-CLL, correlated with mRNA expression levels. Thus, *TWIST2* hypermethylation was suggested as a potential alternative mechanism for p53 inactivation in CLL [147].

Likewise, a gain of methylation in intron 1 of *zeta-chain-associated protein kinase 70* (*ZAP70*) was found to be associated with loss of *ZAP70* expression and a favorable clinical outcome [2, 148-150]. This was an important finding since *ZAP70*, a member of SYK protein tyrosine kinase family, was a known prognostic marker in CLL, used to discriminate between U- and M-CLL subtypes. U-CLLs have an about 6-fold higher *ZAP70* expression as compared to M-CLLs and *ZAP70* expression correctly predicts IGVH mutation status in 93% of patients [2, 151]. *HOXA4*, a member of the HOX family of transcription factors, which is crucial for the regulation of early development, has also been shown to be aberrantly methylated in CLL. Elevated methylation levels were linked to decreased protein levels, *IGHV* mutational status and to inferior clinical outcome [2, 152]. Finally, promoter hypermethylation of the apoptosis-mediating *death-associated protein kinase 1* (*DAPK1*) gene, has been shown to mediate its transcriptional silencing. Interestingly, *DAPK1* expression can be modulated both by somatic and germline events in CLL, suggesting epigenetic deregulation as an alternative mechanism of its inactivation in CLL [2, 153].

In recent years these gene-specific studies were complemented by genome-wide profiling studies. Global methylation profiling in CLL has shown that *IGHV* mutation-defined CLL subgroups are characterized by unique methylation profiles and that CLL methylomes are relatively stable over time and highly similar between the resting (peripheral blood) and the proliferative (lymph node) compartment, implying aberrant methylation as an early leukemogenic event [2, 144]. Very recently, it was proposed that the traditional way of comparing CLL methylomes to normal CD19+ B cells should be replaced by an approach that considers DNA methylation dynamics in the context of normal B-cell differentiation. The main findings of the epigenome profiling studies in CLL are summarized in **Table 1-1**.

Table 1-1: Summary of genome-wide epigenome profiling in CLL

Topic and Main findings	Technology	Reference
DNA methylation profiles of different CLL prognostic subgroups • U-CLL, M-CLL and IGHV3-21 expressing CLLs are carrying different methylation profiles. • In a poor prognostic group, U-CLL, tumor suppressors (e.g. *VHL*, *ABI3*, *IGSF4*) are hypermethylated while genes associated with cell proliferation and invasion (*ADORA3*, *PRF1*) are hypomethylated.	Illumina HumanMethylation 27K BeadChip	Kanduri *et al.*, Blood 2010 [154]
Global profiling of epigenetically deregulated microRNAs in CLL • 128 microRNAs are carrying aberrant DNA methylation in their promoter regions. • Hypermethylated loci include miR-124-2, miR-9-2, miR-129-2, miR-551b and miR-708 while among the hypomethylated ones are miR-21, miR-29a/b-1, miR-34a, miR-155, miR-574 and miR-1204.	Agilent custom-design 244 array, MCIp-array, ChIP-chip	Baer *et al.*, Cancer Research 2012 [155]
Whole-genome bisulfite sequencing in CLLs and normal B cells • Broad epigenetic reprogramming occurs during normal B-cell differentiation/maturation and CLL development. The epigenetic progression involves mainly global hypomethylation affecting enhancers and gene bodies. • An epigenetic signature of CLL is associated with a putative B cell-of-origin. • CLL patients are classified into three epigenetic subgroups, i.e. poor-prognosis naïve-like CLL (n-CLL), favorable-prognosis memory B cell-like CLL (m-CLL) and intermediate CLL (i-CLL) with the intermediate clinical outcome.	WGBS, Illumina HumanMethylation 450K BeadChip	Kulis *et al.*, Nature Genetics 2012 [41]
DNA methylation profiles of paired diagnostic/follow-up samples from M-CLL/untreated, U-CLL/treated and patient-matched blood/lymph node samples • M-CLLs and U-CLLs display differential DNA methylation, affecting mainly regions outside annotated CpG islands. • CLL prognostic genes (e.g. *CLLU1*, *LPL*, *ZAP70* and *NOTCH1*), epigenetic regulators (e.g. *HDAC9*, *HDAC4* and *DNMT3B*) and several signaling pathways (TGFβand NF-κB/TNF) are differentially methylated. • DNA methylation is stable over time and in different hematopoietic compartments (blood, lymph nodes), implying aberrant methylation as an early leukemogenic event.	Illumina HumanMethylation 450K BeadChip, Bisulfite Pyrosequencing	Cahill *et al.*, Leukemia 2013 [156]
Evolution of DNA methylation in CLL • Intratumor methylome variation in CLL is associated with shorter time to treatment and disease aggressiveness. • Increased methylation heterogeneity co-evolves with genetic alterations such as mutations in *TP53*, *SF3B1*, *BRAF* and del(11q) and del(17p). • Independent evolution of epigenetic and genetic events are rare events and may be restricted to specific aberrations only e.g. del(13q).	Illumina HumanMethylation 450K BeadChip, 454-sequencing	Oakes *et al.*, Cancer Discovery 2014 [157]
Epigenetic variability in primary CLL samples and its impact on the patient outcomes • High intratumor epigenetic variability in CLL results from locally	WGBS, RRBS	Landau *et al.*, Cancer Cell 2014 [158]

disordered methylation patterns (variability within DNA fragments) in the malignant B cells. • Disordered methylation serves as an additional mechanism of genetic diversification, resulting in increased survival of CLL cells.		
Genetic and epigenetic changes during progression from indolent to aggressive CLL • Disease progression is in majority of CLL cases not linked to genetic clonal evolution. • Significant DNA methylation changes during disease progression occur at CpGs near PRC2 targets. • The progression-associated CpGs undergo methylation changes in the same direction as those found during normal B-cell differentiation.	Exome Sequencing, Illumina HumanMethylation 450k BeadChip, Illumina Omni 2.5 BeadChip arrays	Smith *et al.,* Blood Cancer Journal 2015 [159]
Whole-genome profiling of DNA methylome during B-cell differentiation and in lymphoid malignancies • Broad epigenetic reprogramming occurs during the B-cell commitment and differentiation. • Early differentiation changes involve the demethylation of enhancer regions whereas late B-cell commitment stages mainly affect heterochromatin (demethylation) and PRC2-repressed genes (gain of methylation). • B-cell neoplasms frequently acquire methylation changes in regions which already undergo dynamic methylation during normal B-cell differentiation.	WGBS, Illumina HumanMethylation 450k BeadChip	Kulis *et al.,* Nature Genetics 2015 [160]
Clinically applicable method to identify epigenetic subgroups in CLL • 5 CpG-based classifier accurately assigns patients into three CLL subgroups i.e. n-CLL, i-CLL and m-CLL. • Epigenetic classification correlates with distinct clinico-biological features, e.g. outcome, Binet stage, CD38 expression levels and *SF3B1* mutations.	Bisulfite Pyrosequencing	Queirós *et al.,* Leukemia 2015 [161]
CLL methylation in the context of normal B-cell differentiation • CLLs derive from a continuum of the maturation stages reflected in the normal B-cell differentiation. • CLL is classified into three subgroups, namely LP-CLL, IP-CLL and HP-CLL, based on the overall levels of methylation programming, relative to normal B-cell differentiation. • Higher CLL maturation stage (HP-CLL) correlates with a favorable clinical outcome and an indolent gene expression pattern. • Large proportions of previously identified methylation events in CLL are observed during the normal B-cell differentiation process. • There is a potential role of aberrant transcription factor programming (EGR, NFAT, EBF, AP-1) in the pathogenesis of CLL.	WGBS, Illumina HumanMethylation 450k BeadChip	Oakes *et al.,* Nature Genetics 2016 [42]
CLL epigenome in the context of normal B-cell differentiation • The epigenetic configuration of CLLs can be divided into three different patterns. • Pattern 1: U-CLLs and M-CLLs show imprints of their cellular origin on the level of DNA methylation and chromatin accessibility, but not for active regulatory regions marked with H3K27ac. • Pattern2: The chromatin landscape in CLL can be linked to complex dynamics during B cell differentiation process that relates U-CLLs and M-CLLs to a variety of combinatorial patterns in B cells, e.g. U-CLLs acquire features of GCBCs that may be associated with higher	WGBS, ATAC-seq, ChIP-seq	Beekman *et al.,* Nature Medicine 2018 [162]

proliferation rate of U-CLLs as compared to M-CLLs.
- Pattern3: CLLs reconfigure their chromatin landscape independent of normal B cell differentiation. This *de novo* chromatin reprogramming is mediated by NFAT, FOX and TCF/LEF transcription factors and is characterized by a transition from inactive regions in normal B cells to super-enhancers in CLLs.
- Although most genetic aberrations are not associated with consistent epigenetic profiles, MYD88 and trisomy 12 have distinct chromatin features.

U-CLL, *IGHV*-unmutated CLL; M-CLL, *IGHV*-mutated CLL; PRC2, polycomb repressive complex 2; LP-CLL, low-programmed CLL; IP-CLL, intermediate-programmed CLL; HP-CLL, high-programmed CLL; ChIP, chromatin immunoprecipitation; WGBS, whole-genome bisulfite sequencing; RRBS, reduced representation bisulfite sequencing. GCBC; germinal center B cell. Adapted from [2].

1.3.4 CLL displays genome-wide hypomethylation

In 1992, Wahlfors and colleagues as first ones demonstrated global DNA hypomethylation in CLL as compared to healthy controls [163]. These findings were confirmed by Stach *et al*. and Fabris *et al.* using capillary electrophoresis and quantitative bisulfite-PCR pyrosequencing, respectively [164-166]. The latter study evaluated the methylation status of repetitive DNA elements, including satellite-α sequences (SAT-α), Alu and LINE-1, in the early-stage CLL cases and thus identified global hypomethylation in CLL. Low SAT-α methylation was an independent prognostic marker and correlated with shorter treatment-free survival [165]. Hypomethylation of repetitive sequences may lead to genomic instability and therefore might contribute to the clonal evolution of CLL by favoring the acquisition of genomic aberrations.

Using whole-genome bisulfite sequencing (WGBS) and DNA methylation arrays, in 2012 Kulis *et al.* have extensively characterized DNA methylomes of 139 CLL (M-CLL and U-CLL) patients and normal B cell subsets, reporting a global loss of methylation in both instances [41]. DNA hypomethylation was shown to primarily target gene bodies and enhancer regions, indicating the functional relevance of methylation changes outside of promoter regions. Kulis *et al.* subclassified CLL samples into three epigenetic subtypes: each of them presenting DNA methylation foot-prints reminiscent of their putative B cell-of-origin [41]. In another seminal study, Oakes *et al.* replicated these findings, further showing that subtype-specific DNA hypomethylation events mainly target enhancers and TF binding sites while hypermethylation events occur primarily in actively transcribed genomic regions [42].

1.3.5 Distinct methylation profiles in CLL subgroups relate to different cellular origins

In the last 10-15 years, the focus of CLL methylome studies was to identify differential methylation patterns between distinct molecular subtypes of CLL patients. Here, the introduction of methylation arrays provided the opportunity to investigate the methylation status of thousands of CpG sites simultaneously. In one of the first studies, using 27K methylation arrays, Kanduri and colleagues have demonstrated global differences between methylation profiles of U-CLL and M-CLLs. Seven candidate tumor suppressor genes (e.g. *ABI3*, *VHL* and *IGSF4*) and eight genes involved in cell proliferation and tumor progression (e.g. *PRF1*, *ADORA3*) were identified as being

differentially methylated in these two major CLL subtypes [2, 154]. Likewise, the differences in global methylation profiles between U-CLL and M-CLL were further characterized with the use of second-generation methylation arrays (450K). As a result, alternative differential methylation patterns between CLL subgroups were found in genes involved in BCR-, TGF-β- and NF-κB/TNF-signaling pathways, in genes encoding epigenetic regulators (e.g. *HDAC9*, *HDAC4* and *DNMT3B*), and in genes known to be of prognostic value in CLL, such as *CLLU1*, *LPL*, *ZAP70* and *NOTCH1* [2, 156].

More recently, considering DNA-methylation dynamics during normal B-cell maturation, new clinico-biological CLL subgroups have been identified, demonstrating that methylation profiles of different CLL subgroups reflect on different cellular origins. In the initial study by Kulis *et al.* in 2012, the methylomes of 139 CLL patients, with and without *IGHV* mutations, and of three normal B cell subtypes, i.e. naïve B cells (NBCs), class-switched memory B cells (scMBCs) and non class-switched memory B cells (ncsMBCs) were characterized. In this study, it was shown that widespread epigenetic reprogramming occurred during normal B cell differentiation which is also reflected in CLL development. The epigenetic changes involved mainly loss of methylation in enhancer regions and gene bodies. Three prognostic CLL subgroups were identified, i.e. NBC-like CLL (n-CLL), largely overlapping with *IGHV*-unmutated CLL, intermediate methylation cluster CLL (i-CLL) with an intermediate clinical outcome, and an MBC-like CLL (m-CLL) subgroup mostly carrying mutated *IGHV* genes. The epigenetic signature of these subtypes is associated with their putative cell-of-origin, meaning that they most likely derive from distinct B-cell differentiation stages. It was proposed that n-CLL originate from more naïve-like B cells, m-CLLs from more mature B cells, whereas i-CLL may derive from a third B cell subtype, which is potentially an antigen-exposed B cell that has not yet completed germinal center maturation (**Figure 1-4**) [2, 41]. Following this study, Quieros and colleagues developed a 5 CpG-based methylation classifier that accurately assigns patients into one of the three subgroups, n-CLL, i-CLL and m-CLL. This classifier is based on CpGs located in the promoter of *SCARF1*, and in the gene bodies of *CTBP2*, *B3GNTL1*, and *TNF*, and in an intergenic region on chromosome 14. This epigenetic classification correlated with distinct clinico-biological features, including patient outcome, Binet stage, *CD38* expression and *SF3B1* mutations [2, 161].

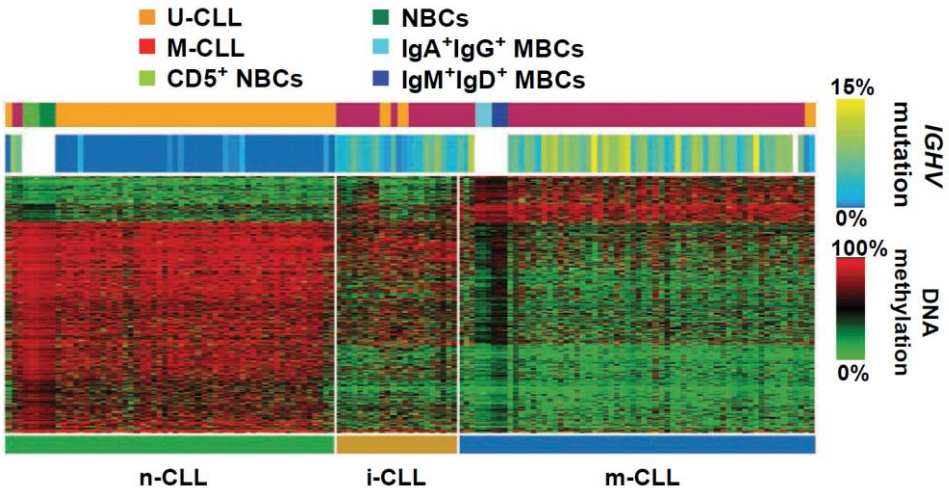


Figure 1-4: Methylation-based clinico-biological subgroups of CLL. The heatmap is based on consensus clustering on 1,649 CpGs identified using differential methylation analysis between U-CLL and M-CLL. U-CLL, *IGHV*-unmutated

CLL; M-CLL, *IGHV*-mutated CLL; IgA, immunoglobulin-α; IgG, immunoglobulin-γ; IgD, immunoglobulin-δ; IgM, immunoglobulin-μ; NBCs, naïve B cells, MBCs, memory B cells; n-CLL, NBC-like CLL; i-CLL, intermediate CLL; m-CLL, MBC-like CLL. Modified from [41].

The presence of three CLL subgroups was confirmed in an independent study by Oakes *et al.* It was shown that CLL derives from a continuum of differentiation stages, rather than being restricted to one specific stage of B-cell differentiation. At the level of DNA methylation, all CLLs were most similar to memory B cells, reaching 70-100% of methylation programming normally observed in memory B cells. Based on the overall level of normal B cell programming achieved, CLL samples were classified into three groups, namely low-programmed (LP-CLL), intermediate-programmed (IP-CLL) and high-programmed CLLs (HP-CLL), with LP-CLL exhibiting the most immature methylation patterns and HP-CLL the most mature. HP-CLLs, in contrast to LP-CLLs, are characterized by a good clinical prognosis and exhibit an indolent gene expression pattern (**Figure 1-5**) [2, 42].

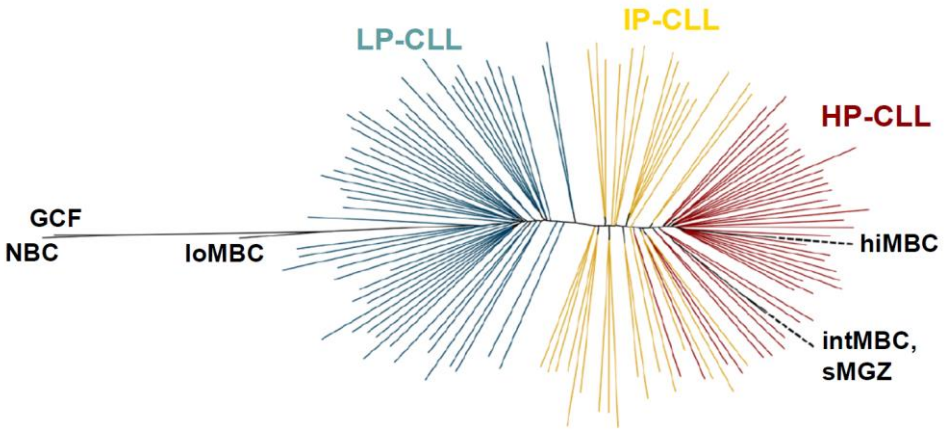


Figure 1-5: CLL derives from a continuum of differentiation stages in normal B cells. DNA methylation-based phylogenetic tree of CLL samples and normal B cell subsets. The tree was generated using dynamic during normal B cell differentiation, high confidence TF binding sites (AP-1, EBF1 and RUNX3) together with CpG sites in hypermethylated transcriptional elongation domains. NBCs, naïve B cells; GCF, germinal center founder B cells; loMBC, early non class-switched memory B cells; intMBC, non class-switched memory B cells; sMGZ, splenic marginal zone B cells; hiMBC, class-switched memory B cells; LP-CLL, low-programmed CLLs; IP-CLL, intermediate-programmed CLLs; HP-CLL, high-programmed CLLs. Modified from [42].

1.3.6 Disease-specific methylation profiles in CLL and normal B-cell differentiation

Previous efforts to define CLL-specific methylation patterns were based on comparisons between CLL samples and $CD19^{+}$ normal B cells, which are a mixture mostly composed of naive and mature memory B cells [167]. As a consequence of the observation that CLL originates from a spectrum of distinct normal B cell subtypes, B-cell differentiation-related and disease-specific methylation changes should be distinguished.

Along those lines, Kulis *et al.* were the first to investigate CLL methylomes in the context of normal B-cell differentiation. They profiled DNA methylomes from ten normal B-cell subpopulations across the entire B-cell

differentiation and compared the dynamic DNA methylation patterns to those of 139 CLL samples. They reported widespread DNA methylation changes during B-cell maturation and found high concordance with methylation patterns previously considered as being CLL-specific [160].

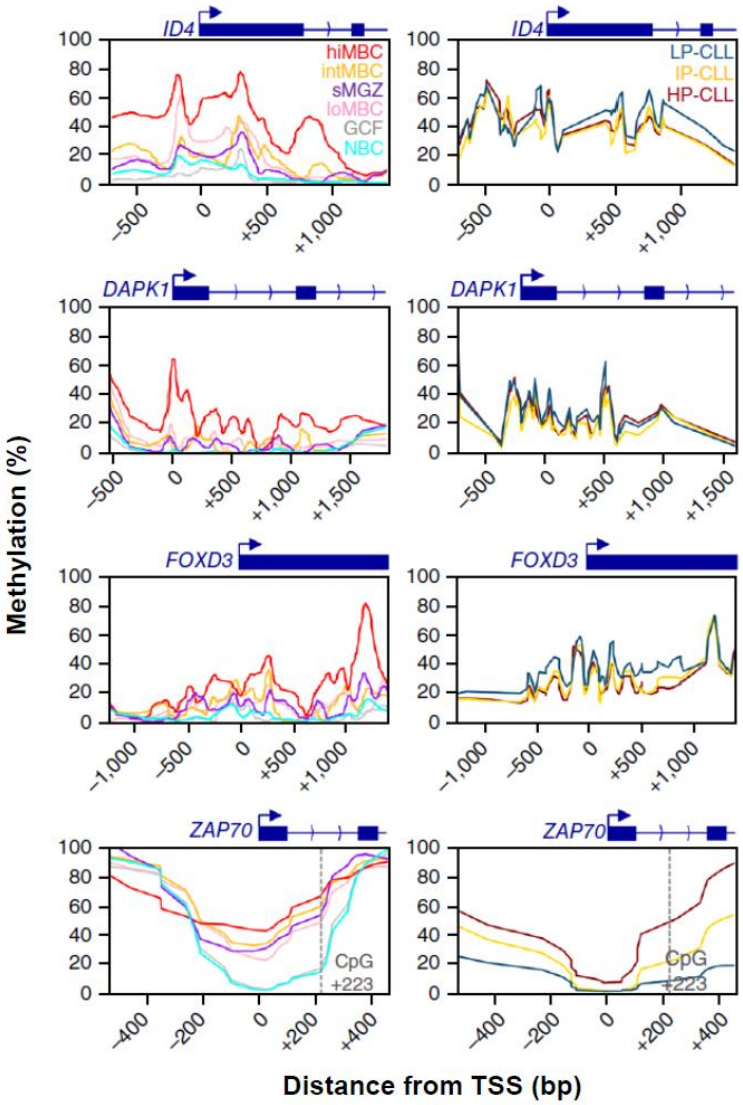


Figure 1-6: Normal B-cell specific methylation confounds CLL-specific findings. DNA methylation profiles in promoters of genes identified previously as being hypermethylated in CLL. Methylation profiles are shown for B cells (left) and CLL (right). TSS, Transcription start site; NBCs, naïve B cells; GCF, germinal center founder B cells; loMBC, early non class-switched memory B cells; intMBC, non class-switched memory B cells; sMGZ, splenic marginal zone B cells; hiMBC, class-switched memory B cells; LP-CLL, low-programmed CLL; IP-CLL, intermediate-programmed CLL; HP-CLL, high-programmed CLL. Modified from [42].

The issue of disease-specific methylation events has been addressed in more depth in a follow-up paper by Oakes *et al.* Investigating genes reported as being hypermethylated in CLL, e.g. *DAPK1*, *ID4*, *FOXD3*, they were able to demonstrate that methylation changes previously thought to be CLL-specific were also found in normal differentiating B cells (**Figure 1-6**). The authors then performed a systematic analysis of DNA methylation changes in CLL as compared to normal B cell differentiation and thus identified truly CLL-specific aberrant methylation

patterns. They further suggested a pathogenic role of aberrant TF programming, involving excessive methylation programming at EGR and NFAT motifs and reduced methylation programming at AP-1 and EBF binding sites [42].

1.3.7 microRNAs – biogenesis and function

MicroRNAs (miRNAs) are defined as small non-coding RNAs of 19-22 nt in length that control gene expression at the post-transcriptional level. Lee and colleagues were pioneering the miRNA field, discovering the first miRNA, *lin-4*, in 1993 in the nematode *Caenorhabditis elegans* [168]. Seven years later the presence of another miRNA, *let-7a*, was reported in *C.elegans* paving the way towards future discoveries [169]. Since then, microRNAs have been identified in the majority of animal and plant species. Currently, the repository of miRNA annotations hosts over 48,885 mature miRNA sequences from 271 species (miRBase version 22, [170]). The general function of miRNAs is to control the expression of protein-coding genes by binding to complementary sequences, mostly present in the 3' untranslated regions (3'UTRs) of their targets that eventually lead to transcript degradation or translational repression (reviewed in [171]). Translational repression is considered the primary mechanism of miRNA-mediated gene silencing [172]. Interestingly, more than 60% of protein-coding genes in the human genome are predicted to contain miRNA-binding sites in their 3'UTRs, suggesting that miRNAs constitute the most abundant class of regulatory molecules [173].

The process of miRNA biogenesis starts in the nucleus where miRNA gene loci are transcribed by RNA Polymerase II (Pol II) or RNA Polymerase III (Pol III), together with their host gene (intragenic miRNAs) or independently of the host gene with the use of their own promoter (intergenic miRNAs). The transcribed sequences, so-called primary miRNA transcripts (pri-miRNAs), have variable lengths ranging from hundred nucleotides to several kilobases (kb) and contain a local hairpin structure. Many miRNAs are 3' polyadenylated and 5'capped – a feature of Pol II-mediated transcription. The pri-miRNA sequences are further trimmed into the 70-100 nt long hairpin intermediates (pre-miRNAs) by the microprocessor complex formed by the RNase III endonuclease Drosha and the DiGeorge syndrome critical region 8 protein (DGCR8). After nuclear processing, pre-miRNAs are exported to the cytoplasm by the nuclear transport receptor exportin-5 (XPO5)/RanGTP complex (reviewed in [174], [175], [176]) (**Figure 1-7**).

In the cytoplasm, pre-miRNAs are subjected to a second processing step, namely the final cleavage by another RNase III enzyme, Dicer, operating in complex with transactivation-responsive RNA-binding protein (TRBP) to generate the final ~22 nt miRNA product. Following the generation of miRNA duplexes, one strand of the duplex is removed and degraded (passenger strand), while the other strand remains as a mature miRNA (guide strand). Multiple studies on miRNA duplexes have shown that strand selection is mainly determined by the thermodynamic stability of the ends of the duplex, with a preferential choice of the strand containing unstable base pairs at the 5' end, e.g. G:U versus G:C pair. Mature miRNAs are incorporated into effector complexes, called miRNA-containing RNA-induced silencing complexes (miRISC) that guides them towards specific mRNA targets. The process of duplex unwinding together with miRISC activation is mediated by the Argonaute-2 (Ago-2) protein (reviewed in [174], [175], [176]) (**Figure 1-7**).

Many biological processes are critically regulated by miRNAs, such as development, differentiation, cell proliferation, apoptosis, immune responses and angiogenesis (reviewed in [177]). Aberrant miRNA expression profiles have been reported in many diseases, including various cancer types [178-186]. Yet, the knowledge about underlying causes of this deregulation is limited. Both, transcriptional and post-transcriptional mechanisms play a role in the regulation of miRNA expression. Among others, transcriptional regulation of miRNA expression can be achieved by differential TF binding, by regulating the expression of the host gene and by epigenetic changes in miRNA regulatory regions. For instance, TP53 and PTEN were shown to regulate the expression of nine miRNAs in renal cell carcinoma [185]. Likewise, it was shown that miR-155 expression is host gene-dependent [187]. DNA methylation was shown to regulate expression of miR-132 in colorectal cancer and miR-34b/c in CLL [188, 189].

With respect to post-transcriptional regulation of miRNA expression, so far two mechanisms have been reported: 1) changes in the miRNA processing and 2) changes in miRNA stability. Defects in the miRNA biogenesis machinery, mostly due to mutations in proteins/enzymes of the miRNA biogenesis pathway (Dicer, TRBP, Drosha or Xpo-5), have been reported in various tumors [190, 191]. Moreover, miRNA stability was shown to be dependent on the tissue-context and on the presence of modifications of the 3' end of miRNAs such as adenylation and uridylation [192]. A lower rate of miRNA degradation and an increased half-life have been reported in overexpression experiments of Argonaute proteins [193].

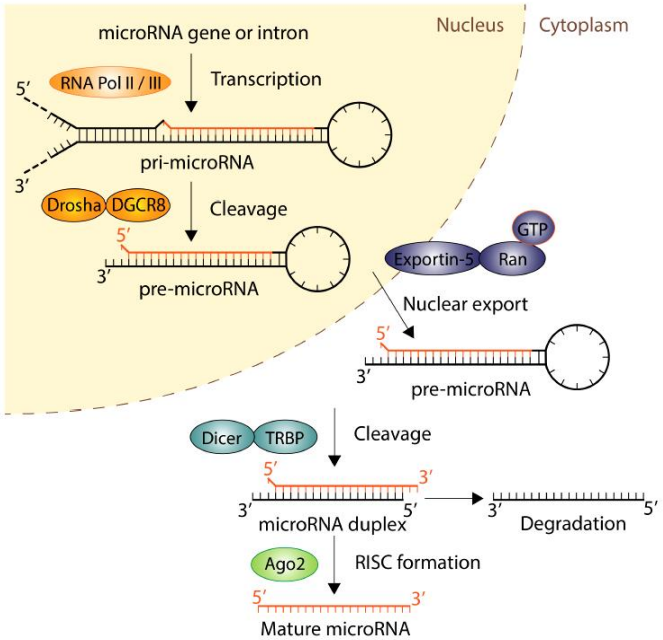


Figure 1-7: Biogenesis of miRNAs. Pri-miRNAs are transcribed by Polymerase II or III in the nucleus and further cleaved by the Drosha/DGCR8 complex generating pre-miRNAs. Exportin-5 transports pre-miRNAs into the cytoplasm, in which Dicer and TRBP cleave the precursor sequences into ~22nt miRNA duplexes. One of the strands of the duplex is degraded and the remaining, functional one is incorporated together with Ago-2 protein into the silencing complex, miRISC that guides the mature miRNA towards target genes. Further details are explained in the text. Modified from [194].

1.3.8 Epigenetic regulation of miRNA expression in CLL

Aberrant epigenetic regulation plays a key role in dysregulated miRNA expression levels in CLL. So far, all published studies have identified aberrant methylation events in CLL by comparison of peripheral blood $CD19^+$ normal B cells to CLL cells. In a first large-scale attempt published in 2009, Pallasch and colleagues used a bead-chip expression platform to delineate populations of aberrantly expressed miRNAs in CLL [195]. Search for a potential mechanism underlying the deregulated expression profiles has shown no obvious association with genetic alterations but pointed towards epigenetic transcriptional silencing of miRNA expression. Several epigenetically deregulated miRNAs have been identified, e.g. miR-181a, miR-181b, miR-126, miR-424 and miR-107 [195]. Similarly, using a custom-designed DNA methylation array (Agilent 244k), the methylation patterns of 939 miRNAs were studied by Baer *et al.* [155]. This study identified 128 miRNAs with aberrant methylation in their promoter regions, e.g. miR-9-2, miR-708, miR-34a, miR-155 or miR-21, of which the majority were found to be hypomethylated [2, 155].

Apart from the genome-wide studies, CLL researchers focused mainly on individual miRNA promoters and the impact of DNA methylation on their aberrant expression levels. For instance, miR-129-2, a known tumor suppressor gene, was identified as being aberrantly hypermethylated and downregulated in CLL and miR-129-2 hypermethylation was associated with poor survival of CLL patients [2, 196]. A member of the miR-9 family, miR-9-3, was reported as being lowly methylated in normal $CD19^+$ cells but hypermethylated in CLL and miR-9-3 promoter hypermethylation was associated with advanced Rai stage [2, 197]. 5-Aza-2'-deoxycytidine treatment resulted in promoter demethylation and miR-9-3 upregulation that led to enhanced apoptosis combined with downregulation of NFκB1. Wang and Deneberg both have shown that the miR-34b/c promoter, located within a commonly deleted segment on chromosome 11 (11q23), is completely methylated in CLL. Inverse correlation between miR-34b/c promoter methylation and expression was demonstrated. The presence of 11q deletion and increased miR-34b/c methylation were reported as being mutually exclusive, indicating that these are two alternative silencing events for these miRNAs. Overexpression of miR-34b/c in a CLL cell line, HG3, significantly increased apoptosis, suggesting a tumor suppressive function of miR-34b/c [2, 189, 198]. In another study, Baer *et al.* reported aberrant methylation of an enhancer regulating miR-708 and showed that miR-708 targets IKKβ, a key kinase in the NF-κB signaling pathway [2, 199]. Wang and colleagues found promoter hypermethylation of miR-3151 in CLL, but not in normal $CD19^+$ controls. Promoter methylation of miR-3151 inversely correlated its expression levels. Restoration of miR-3151 expression in CLL resulted in enhanced apoptosis, presumably through downregulation of direct miR-3151 targets. These include MCL1, MADD and PIK3R2, which are the essential components of MEK/ERK and PI3K/AKT signaling networks [2, 200]. Likewise, Kopparapu and colleagues linked elevated methylation levels of miR-26A1 to U-CLL and poor survival of patients. Overexpression of miR-26A1 led to reduced transcript and protein levels of an enzymatic subunit of the polycomb repressive complex 2 (PRC2), EZH2 [2, 201].

1.3.9 The cell-of-origin of CLL

From recent epigenome profiling studies, it has become apparent that most of the aberrant DNA methylation events reported in CLL in the past are in reality changes that are related to normal B-cell differentiation rather than to CLL biology. This problem can be explained by the use of inadequate control samples, which usually consisted of $CD19^+$ B cells, representing a mixture of different B cell subtypes spanning the entire B cell differentiation axis. Therefore, the current challenge in the CLL field is to determine the appropriate control B cell, which would be the CLL cell-of-origin.

Although tracing the cellular origin of CLL is of conceptual importance to understand CLL pathogenesis, the knowledge of the precise stage of B cell differentiation at which the final transformation events occurs is still limited. Initial studies by Seifert *et al.* based on immunophenotyping and transcription profiling have shown that unmutated IGHV CLLs derive from unmutated $CD5^+$ B cells while mutated IGHV CLLs originate from a distinct $CD5^+CD27^+$ post-germinal center B-cell subtype [202]. However, recent findings by Kikushige *et al.* have challenged the concept of CLL as a disease arising from mature B cells [203]. It has been demonstrated that hematopoietic stem and progenitor cells from patients with CLL engraft into immunodeficient mice and display cell-intrinsic propensity to generate mono- or oligo-clonal CLL-like B-cells [203]. The work by Damm *et al.* suggests that pre-leukemic multipotent progenitors may carry somatic mutations in genes found to be mutated in CLL, like e.g. *BRAF*, *NOTCH1*, *SF3B1*, *NFKBIE* and *EGR2* [16]. Despite the existence of preleukemic clones in the HSPC compartment of CLL patients, at least one additional driver event, either genetic or epigenetic, is required in order to develop overt CLL (Figure 1.8). This transforming event can be identified by using DNA methylation as a footprint of the B cell differentiation stage achieved in CLL founder cell (i.e. the „cell-of-origin“). The DNA methylome landscape of the CLL cell-of-origin is ‘frozen’ and thus will be stably propagated in the leukemic cells [42].

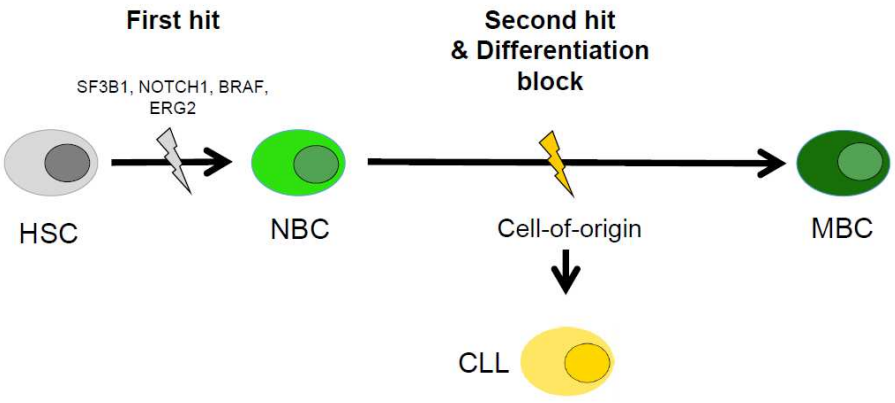


Figure 1-8: Cell-of-origin of CLL. The CLL cell-of-origin is defined as a cell that has acquired a first oncogenic hit (*SF3B1*, *NOTCH1*, *BRAF*, *ERG2*) and which will initiate tumorigenic growth if one or more additional hits have been acquired [202]. This transforming event ultimately blocks normal B cell differentiation and/or resistance to apoptotic cell death.

2. AIM OF THE THESIS

Chronic lymphocytic leukemia (CLL) originates from differentiating B cells which undergo extensive epigenetic reprogramming during normal B cell differentiation [42, 160]. Depending on the differentiation stage of the B cell acquiring the leukemia-initiating event, a distinct epigenome is 'frozen' and stably propagated in the leukemic cells [42]. This stable epigenetic patterning can serve as an indicator for the identification of the cell-of-origin for each individual CLL case. The cancer cell-of-origin is defined as a cell that has acquired a first oncogenic hit and which will initiate tumorigenic growth if one or more additional hits have been acquired [202]. Consequently, this means that two factors contribute to the epigenetic profile of CLL cells: first, epigenetic profiles which were present in the founder B cell at the time of malignant transformation, and second, CLL-specific epigenetic alterations that occur during leukemogenesis and which relate to genetic alterations and/or to aberrant signaling events acquired in the leukemic cells via either extrinsic or intrinsic stimuli.

Defining CLL-specific epigenetic events, which are distinct of B cell-specific epigenetic reprogramming events, is of utmost importance to understand molecular alterations contributing to CLL. Previous studies in CLL have focused on aberrant methylation events and their impact on aberrant gene expression of both, protein-coding genes (e.g. *DAPK1*, *ZAP70*, *ID4*) and microRNAs (e.g. miR-9, miR-181a/b, miR-34a, miR-708). However, all aberrant DNA methylation events have been identified using peripheral blood $CD19^+$ B cells as controls, neglecting the massive epigenetic programming during normal B cell differentiation. Therefore, novel approaches aiming at identifying truly CLL-specific methylation changes in the context of normal B cell differentiation are urgently needed.

Accordingly, the **first aim** of this thesis was to **define the epigenome of the cell-of-origin, unique for every CLL patient and to identify CLL-specific methylation profiles.** To do so, linear regression was used to model the epigenome dynamics occurring during normal B cell differentiation. DNA methylomes of CLL cells were subsequently precisely positioned onto the normal B cell differentiation trajectory to define the DNA methylomes of the cell-of-origin for every CLL patient. By considering the cellular origin, epigenetic alterations reflecting on normal B cell differentiation were substracted and CLL-specific methylation events were defined. Further, the molecular programs behind CLL-specific methylation profiles were investigated using chromatin states and transcriptional factor binding profiles from B cell immortalized cell lines.

The second aim of this thesis was to **determine epigenetically deregulated miRNAs in CLL.** For this purpose, an unbiased strategy for the identification of miRNA regulatory regions was developed. The impact of CLL-specific aberrant DNA methylation on miRNA expression was assessed using a correlation of DNA methylation in the promoter regions with miRNA expression. CLL-specific candidates were further validated using MassARRAY assay and qRT-PCR for the quantification of methylation and expression profiles, respectively. Finally, miRNA target genes were determined using publicly available experimental datasets.

The third aim of this thesis was to **demonstrate the relevance of our approach by contrasting the number of epigenetically deregulated miRNAs and protein-coding genes** to numbers determined using bulk $CD19^+$ cells from

peripheral blood as controls **and to validate selected candidate regions identified in our cell-of-origin model using targeted methylation analysis and qRT-PCR**.

3. MATERIAL AND METHODS

3.1 Materials

The laboratory equipment, chemicals, kits, buffers, enzymes, plasmid vectors, computer software, and databases used for experiments or analyses in this thesis are summarized in **Table 3-1** to **Table 3-7**.

Table 3-1: Laboratory equipment

Equipment	Manufacturer
Agilent 2100 Bioanalyzer	Agilent Technologies, Santa Clara, USA
Benchtop centrifuge	Eppendorf, Hamburg, Germany
Centrifuge 5415 R	Eppendorf, Hamburg, Germany
Centrifuge 5810 R	Eppendorf, Hamburg, Germany
Electrophoresis power supply EPS 300	Amersham Pharmacia Biotech, Maryland, USA
Gel documentation system EASY Doc plus	Herolab, Wiesloch, Germany
Gel chambers (horizontal and vertical)	Bio-Rad Laboratories, Munich, Germany
LightCycler® 480 Real-Time PCR System	Roche Diagnostics, Mannheim, Germany
MassARRAY nanodispensor	Agena Bioscience, San Diego, USA
Mastercycler® pro 384	Eppendorf, Hamburg, Germany
Mastercycler® epgradient S	Eppendorf, Hamburg, Germany
Matrix pipettes (30 and 125 µl)	Matrix Technologies, Feldkirchen, Germany
Nanodrop ND-1000 spectrophotometer	PeqLab, Erlangen, Germany
Pipettes (10, 20, 100, and 1000 µl)	Eppendorf, Hamburg, Germany
Plate centrifuge 5430	Eppendorf, Hamburg, Germany
Qubit® 2.0 Fluorometer	Invitrogen, Life Technologies, Carlsbad, USA
Rotating oven (OV2)	Biometra, Göttingen, Germany
Sequenom MALDI-TOF Mass Spectrometer	Agena Bioscience, San Diego, USA
UV lamp	NeoLab GmbH, Heidelberg Germany
Water bath Julabo TW-12	Eppendorf, Hamburg, Germany

Table 3-2: General disposables

Material	Manufacturer
Adhesive plate seals	Steinbrenner, Wiesenbach, Germany
8- and 12-well strips with lids	Greiner Bio-One, Frickenhausen, Germany
96- and 384-well PCR plates	Thermo Fischer Scientific, Waltham, USA
Combitips (1, 5 and 10 ml)	Eppendorf, Hamburg, Germany
Falcon tubes (15 and 50 ml)	BD Biosciences, San Jose, USA
Filter and normal tips for pipettes (10, 20, 200 and 1000 µl)	Sarstedt, Nümbrecht, Germany / Steinbrenner, Wiesenbach, Germany
Low Binding tubes (1.5 ml)	Eppendorf, Hamburg, Germany
Sterican® needles	B. Braun, Melsungen, Germany
Syringes	Terumo, Leuven, Belgium

Table 3-3: General chemicals and reagents

Chemical/reagent	Manufacturer
100 bp DNA ladder	Invitrogen, Carlsbad, USA
1 Kb plus DNA ladder	Invitrogen, Carlsbad, USA
Chloroform	VWR, Radnor, USA
Dimethyl sulfoxide (DMSO)	Merck, Darmstadt, Germany
Dithiothreitol (DTT)	GERBU, Wieblingen, Germany
dNTP mix	Thermo Fischer Scientific, Waltham, USA
Ethanol	Sigma-Aldrich, Taufkrichen, Germany
Ethidium Bromide	Sigma-Aldrich, Taufkirchen, Germany
Formaldehyde solution	Sigma-Aldrich, Taufkirchen, Germany
Isopropanol	Sigma-Aldrich, Taufkirchen, Germany
Gel Loading Dye, Purple (6X)	New England Biolabs, Frankfurt am Main, Germany
PeqGOLD universal agarose	PeqLab, Erlangen, Germany
Resin	Agena Bioscience, San Diego, USA
RNase-Free water (Ultrapure DNase/Rnase-Free Distilled Water)	Invitrogen, Carlsbad, USA
Tris-Borat-EDTA (TBE, 10X)	Carl Roth GmbH, Karlsruhe, Germany
Triton-X-100	Sigma-Aldrich, Taufkirchen, Germany
TRIzol®	Invitrogen, Carlsbad, USA
β-mercaptoethanol	AppliChem GmbH, Darmstadt, Germany

Table 3-4: Commercial kits

Kit	Manufacturer
AllPrep DNA/RNA Mini kit	Qiagen, Hilden, Germany
Agilent high sensitivity DNA kit	Agilent Technologies, Santa Clara, USA
Agilent RNA 6000 nano kit	Agilent Technologies, Santa Clara, USA
Agilent RNA 6000 pico kit	Agilent Technologies, Santa Clara, USA
EZ DNA methylation kit	Zymo Research, Orange, USA
MassCLEAVE T7 kit (T Cleavage)	Agena Bioscience, San Diego, USA
miScript II RT kit	Qiagen, Hilden, Germany
miScript SYBR Green PCR kit	Qiagen, Hilden, Germany
QuantiTect SYBR Green PCR kit	Qiagen, Hilden, Germany
Qubit RNA BR Assay Kit	Invitrogen, Carlsbad, USA
Qubit dsDNA HS assay kit	Invitrogen, Carlsbad, USA
RNase-free DNase set	Qiagen, Hilden, Germany
RNeasy mini kit	Qiagen, Hilden, Germany

Table 3-5: Enzymes and buffers

Enzyme	Manufacturer
Buffer RWT concentrate	Qiagen, Hilden, Germany
CpG methyltransferase (*M. SssI*)	New England Biolabs, Massachusetts, USA
First stand reaction buffer (5X)	New England Biolabs, Ipswich, USA
PCR buffer (10X)	Qiagen, Hilden, Germany
Puregene Proteinase K	Qiagen, Hilden, Germany
Shrimp alkaline phosphatase (SAP)	Agena Bioscience, San Diego, USA
RNaseA	Agena Bioscience, San Diego, USA

Table 3-6: Additional material

Material	Manufacturer
Spectro CHIP arrays	Agena Bioscience, San Diego, USA

Table 3-7: Computer software, web-based tools and databases

Name	Manufacturer/web link
Software/Packages	
2100 Expert Software	Agilent Technologies, Santa Clara, USA
DESeq2 v.1.20.0	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
EpiTYPER-TM 1.0	Agena Bioscience, San Diego, USA
FastQC v.0.11.6	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
FeatureCounts (Subread package v.1.5.1)	http://bioinf.wehi.edu.au/featureCounts/
ggplot2 v.2.2.1	http://ggplot2.tidyverse.org/
GraphPad Prism 5	GraphPad Software Inc., La Jolla, USA
HOMER v.4.5	http://biowhat.ucsd.edu/homer/
IGV 2.1	http://www.broadinstitute.org/igv/home
LightCycler Software 4	Roche Applied Science, Penzberg, Germany
LOLA v.1.10.0	https://bioconductor.org/packages/release/bioc/html/LOLA.html
MS Office 2010	Microsoft, Redmond, USA
Pheatmap v.1.0.10	https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf
R statistical computing environment v.3.5.0	http://www.r-project.org
RNA-seQC v.1.1.8	http://archive.broadinstitute.org/cancer/cga/rna-seqc
RnBeads v.1.12.1	https://rnbeads.org/
Sambamba v.0.6.5	https://github.com/biod/sambamba
SAMTools v.1.6	http://samtools.sourceforge.net/
STAR v.2.5.2b	https://github.com/alexdobin/STAR
Web based tools	
EpiDesigner beta	http://www.epidesigner.com
Primer3 v.0.4.0	http://frodo.wi.mit.edu/
Databases/Resources	
Bioconductor	http://bioconductor.org/
ENCODE	https://www.encodeproject.org/
DIANA-TarBase v7.0	http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/
Ubuntu 12.04 LTS	Canonical Ltd., Ubuntu community, London, UK
Human genome (hg19, GRCh37)	http://genome.ucsc.edu/downloads.html
microRNA.org	http://www.microrna.org
miRBase v.18.0	http://www.mirbase.org
miRTarBase release 7.0	http://mirtarbase.mbc.nctu.edu.tw/php/index.php
NCBI database, NCBI blast search	http://ncbi.nlm.nih.gov
UCSC Genome Browser	http://genome.ucsc.edu

3.2 Methods

3.2.1 CLL samples

Peripheral blood (PB) samples from the patients with CLL were obtained from two clinical centers in Germany: the National Center for Tumor Diseases (NCT) in Heidelberg, and from the University Hospital in Ulm. The informed consent was obtained by the procedure approved by the Ethics Committee of the University Hospital of Heidelberg and of Ulm University. All samples were purified using magnetic cell sorting (MACS) with selection for $CD19^+$ cells, and then cultured overnight in 10% autoserum. DNA and RNA were isolated according to the Method section 'DNA & RNA isolation' (**Section 3.2.3**).

3.2.2 Normal B cell isolation

Naive and class-switched B cells were isolated from full-blood donations from healthy donors (54, 56 and 60 years of age), recruited at the Medical School in Essen. The study protocol was approved by the Internal Review Board of the Medical School in Essen. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Paque density centrifugation (Amersham, Freiburg, Germany) from 500 ml of PB. $CD19^+$ B cells were enriched to >98% by magnetic cell separation using the MACS system (Miltenyi Biotech, Bergisch Gladbach, Germany). The B cell suspension of each donor was stained with anti-CD27-APC (MT271), anti-IgD-PECy7 (IA6-2), anti-CD23-PE (M-L233) and anti-IgG-FITC (G18-145) antibodies (Becton Dickinson Biosciences, Heidelberg, Germany) and sorted with a FACSAria cell sorter (Becton Dickinson) as naive ($IgD^{high}CD27^-CD23^+$) and class-switched memory ($IgD^-IgG^+CD23^-CD27^+$) B cells. The purity was >99% for each B cell population as determined by reanalysis on a FACSCanto flow cytometer (Becton Dickinson Biosciences) with FACSDiva software.

The normal B cell isolation was performed by Dr. Marc Seifert from Molecular Genetics Group, Essen University Hospital, Essen, Germany.

3.2.3 RNA and DNA isolation

DNA and RNA from B cells and CLLs were isolated using a combined protocol with Trizol and AllPrep DNA/RNA kit (Qiagen, Hilden, Germany). Cell pellets were lysed in 500 µl RLT-Plus lysis buffer containing β-Mercaptoethanol and homogenized for 5 min at room temperature (RT). The lysates were then transferred into AllPrep DNA spin column and centrifuged (1 min; 10,000 rpm). Genomic DNA was bound to a column, and further DNA isolation was performed according to the Qiagen AllPrep genomic DNA isolation protocol. The flow-through was used for RNA isolation, using a combined protocol with Trizol and RNeasy Mini columns (Qiagen). Briefly, 1 ml of Trizol was added to the flow-through. Samples were mixed, vortexed and incubated for 5 min at RT. 200 µl of chloroform was added to each of the samples. Samples were vortexed for 20 s, incubated for 15 min and centrifuged (4°C at 13,000 x rpm; 15 min). 500 µL of isopropanol was added to the aqueous layer of each of the samples, followed by 30 min-incubation at RT and Proteinase K treatment. 400 µL of RWT buffer was added to each of the samples

followed by mixing and spin down (1 min; 9,000 rpm at 21°C). The residual lysate was transferred to the column, followed by centrifugation (1 min; 9000 rpm at 21°C). 400 µL of RWT Buffer was added to each column and centrifugation was performed (30 s; 9,000 rpm at 21°C). 80 µL of DNAseI mix (RNase-free DNase set) was applied to the column, followed by 15 min-incubation at RT. 500 µL of Buffer RWT was added to the column, followed by spin down (30 s; 9,000 rpm at 21°C). The flow-through was re-applied to the same columns and spin down (30 s; 9,000 rpm at 21°C). This time flow-through was discarded and 500 µL of Buffer RPE (Qiagen) was added the column, followed by a centrifugation for 30 s at 9,000 rpm at 21°C. The washing step with RPE buffer was repeated twice (3x washes in total). 500 µL of 75% EtOH was applied to the RNeasy Mini Spin column, incubated for 1 min, and centrifuged (30 s; 9,000 rpm at 21°C). Additional dry-centrifugation step was performed (2-3 min; 9.000 rpm at 21°C). RNA was eluted in RNase-free water, incubated for 1 min and centrifuged (1 min; 10,000 rpm at 21°C). The re-eluate was re-applied to the column and centrifuged for 1 min at 10,000 rpm at 21°C.

3.2.4 Quantitative DNA methylation analysis using MassARRAY

The quantitative analysis of DNA methylation profiles was performed using MassARRAY system and EpiTYPER software (Agena Bioscience, San Diego, USA). The method starts with bisulfite treatment (BT) of genomic DNA and is followed by PCR amplification of the region of interest, *in vitro* transcription (IVT) generating single-stranded RNA, and RNase A cleavage. The endoribonuclease-digested products are then subjected to matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) (**Figure 3-1**). In the following sections, the MassARRAY procedure is described more in details.

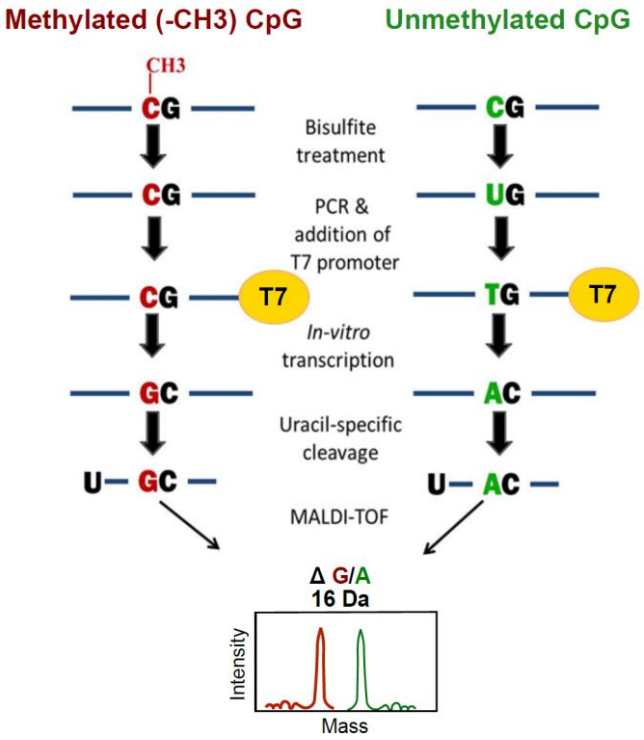


Figure 3-1: MassARRAY scheme. CpG dinucleotides with a methylated cytosine are represented in red and unmethylated one in green. The MassArray procedure starts with a sodium bisulfite treatment reaction (BT). Then the amplicon of interest is PCR-amplified with primers incorporating T7 polymerase tag necessary for the *in vitro* transcription (IVT). Subsequently, PCR products are *in vitro* transcribed into single-stranded RNA and U-specifically

cleaved. Afterwards, desalting step is performed and desalted sequences are analyzed by mass spectrometry. The mass shift of 16Da is observed between a G (originally methylated cytosine) and an A (originally unmethylated cytosine). Adapted from [203].

3.2.4.1 Primer design and PCR optimization

Primers specific for the sequence of interest were designed with the EpiDesigner Software (http://www.epidesigner.com). Several aspects were taken into consideration: 1) the length of the primers (21 to 25bp); 2) number of consecutive thymidines (not more than 6Ts); 3) size of genomic region to be amplified (200 to 500bp); 4) annealing temperature for the primers (55 to 65°C). The temperature should be similar for the forward and reverse primer. In addition, T7 promoter tag (cagtaatacgactcactatagggagaaggct) was added to the 5' end of the reverse primer which enables *in vitro* transcription by the T7 polymerase. Likewise, a 10-mer tag (aggaagagag) was added to the 5'end of the forward primer to balance the length of the PCR primers. The fragmentation patterns of the amplicons and the coverage of CpGs of interest were checked using *in silico* fragmentation tool in R (RSeqMeth package, version 1.0.2.) [204]. The primers used for MassARRAY are listed in the **Appendix** (**Table 7-1**).

Primer optimization was performed using bisulfite-treated DNA (BT-DNA) as a template (see **Methods 3.2.4.2**.) Three different annealing temperatures (52, 56 and 58°C) were used to test the annealing efficiency of the primers. The amplification step was performed for 37 cycles. The PCR reaction was carried in a final volume of 5 µL, following the procedure that is shown in **Table 3-8** and **Table 3-9**. The PCR products were examined for the correctness of their length using agarose gel electrophoresis (2% agarose gels).

Table 3-8: TD PCR reaction

Reagent	Amount (1X)
BT DNA	1 µL
dNTPs (10 mM)	0.1 µL
Primer mix (10 µM)	1 µL
PCR buffer (10X)	0.5 µL
HotStar Taq (5 U/µl)	0.04 µL
H_2O (RNase/DNase free)	2.36 µL
Total volume	5 µL

Table 3-9: TD PCR cycling conditions

Cycle step	Temperature (°C)	Time (min)	Cycle number
Initial denaturation	95	15	1
Denaturation	94	0:30	37
Annealing	52, 56, 58°C	0:30	
Extension	72	1	
Final elongation	72	5	1
Hold	4	∞	

3.2.4.2 Bisulfite conversion

The methylation status of the locus of interest is determined using sodium bisulfite conversion (bisulfite treatment, BT). This method helps to differentiate and detect unmethylated and methylated CpGs. The principle of the method is the following: the sodium bisulfite treatment leads to deamination of the unmethylated cytosines to uracils while the methylated cytosines (5mC) remain unchanged (**Figure 3-2**). The uracils are then amplified in the subsequent PCR reactions as thymines, while the methylated cytosines remain the same.

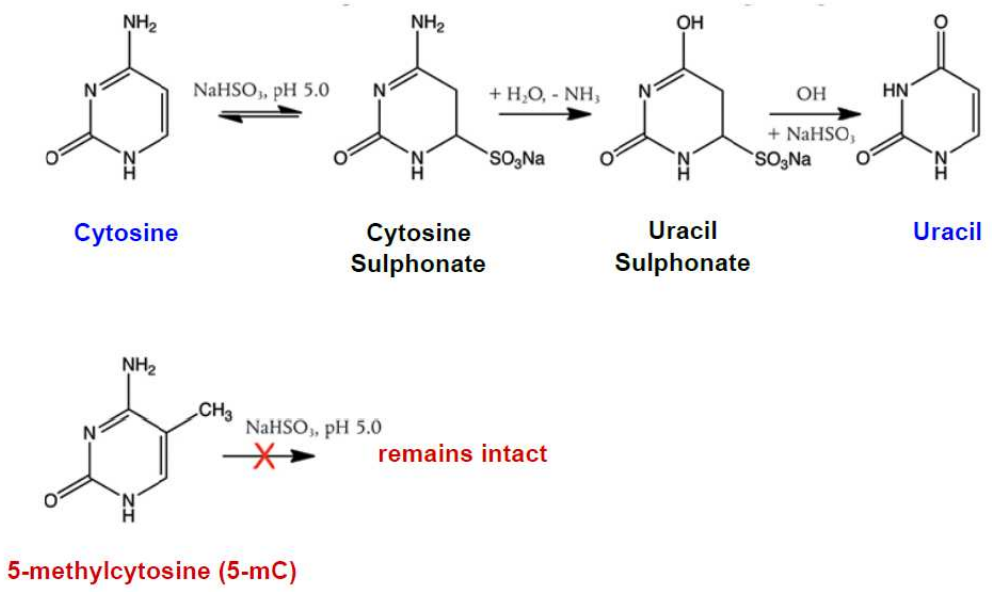


Figure 3-2: Bisulfite conversion in MassArray. After the thermal DNA denaturation, DNA sequence of interest is incubated with high sodium bisulfite concentrations at high temperature and low pH (5-6) generating cytosine sulphonate. Next, in the cytosine deamination step, cytosine sulphonate is converted into uracil sulphonate. In the final step of desulphonation, the sulfite moiety is removed from the ring, generating uracil. 5-methylcytosines (5-mC) are not susceptible to bisulfite conversion and remain intact.

In brief, genomic DNA (0.5 to 1µg) was sodium bisulfite treated using EZ DNA methylation kit (Zymo Research, Irvine, USA) according to the manufacturer's instructions. The elution process of BT-DNA was performed twice, with 30 µL of M-elution buffer. Afterwards, BT-DNA was immediately stored at -20°C.

3.2.4.3 MassARRAY

MassARRAY was performed as described previously [203, 205]. Briefly, BT-DNA was PCR amplified and the free, unincorporated deoxynucleoside triphosphates (dNTPs) were removed by shrimp alkaline phosphatase (SAP) treatment. The SAP treatment was performed in a final volume of 5 µl (3 µl of PCR product, 0.3 µl of SAP and 1.7 µl

of H_20) at 37°C for 20 min, followed by heat inactivation at 85°C for 5 min. Subsequently, 2 µl of the dephosphorylated PCR products were used for the *in vitro* transcription (IVT), in which PCR products are i*n vitro* transcribed into single-stranded RNA (37°C for 3 h, according to the protocol shown in **Table 3-10)**. During the IVT reaction, the T7 polymerase incorporates ribonucleotides (rUTP, rGTP and rATP) and the non-cleavable dCTP deoxynucleotide into the transcript. This is followed by RNase A-mediated uracil-specific cleavage at the 3´end of cytosine (rCTP) and uracil (rUTP) ribonucleotides. This allows the quantification of methylation patterns based on a mass difference of 16Da between a guanine (corresponds to originally methylated cytosine) and adenine (unmethylated cytosine).

Table 3-10: *In vitro* transcription reagents

Reagent	Amount (1X)
T7 polymerase buffer (5X)	0.89 µL
T cleavage mix	0.22 µL
DTT	0.22 µL
T7 polymerase	0.4 µL
RNase A	0.06 µL
H_2O (RNase/DNase free)	3.21 µL
PCR template	2 µL
Total volume	7 µL

After the IVT reaction and RNaseA cleavage, desalting process takes place. During this step, 6 mg of clean resin (Agena Bioscience) and 20 µl of ddH_2O are added to the samples, followed by 30 min-incubation in a rotator (at room temperature). 15 nL of each desalted sample is then spotted by the MassARRAY dispenser onto a 384-format SpectroCHIP (Agena Bioscience) and the fragments are analyzed by matrix-assisted light desorption/ionization time-of-flight mass spectrometry (MALDI-TOF). The mass spectra containing methylation profiles are generated with the EpiTYPER-TM v.1.0 software (Agena Bioscience). The EpiTYPER software identifies individual peaks by comparing the obtained mass spectra with the expected ones. The difference between methylated and unmethylated cytosine corresponds to a mass shift of 16Da per CpG in the fragment. The software automatically calculates the relative methylation values, which is the ratio between peaks of methylated and unmethylated fragments. The spectra containing low or high mass fragments were excluded from the analysis.

3.2.4.4 MassARRAY standards

To correct for a possible bias introduced during the PCR amplification, six BT-DNA standards carrying known methylation patterns (0%, 20%, 40%, 60%, 80%, and 100%) were PCR amplified together with each amplicon. MassARRAY standards were generated by whole genome amplification of buffy coat DNA (REPLI-g Mini Kit, Qiagen) according to the manufacturer's instructions [206]. 50% of DNA was *in vitro* methylated using the *M.SssI* CpG methyltransferase. *M.SssI* CpG methyltransferase synthesizes methyl groups using SAM as a methyl donor. Subsequently, DNA was purified (QIAamp DNA Mini Kit, Qiagen) following the manufacturer's recommendations [155]. The methylated and unmethylated DNAs were then mixed in different ratios to obtain six different

methylation standards (0%, 20%, 40%, 60%, 80%, and 100%). The standards were then bisulfite treated as described in the **Methods 3.2.4.1**.

3.2.4.5 Statistical analysis of MassARRAY data

For each amplicon of interest, CpG units having low spectra quality (bad intensity values) and two or more overlapping peaks were excluded from the final analysis. Statistical significance of the methylation changes between normal B cells and CLLs was determined using Wilcoxon signed-rank test in R. Mean methylation across samples (B cells or CLL) per CpG was used.

3.2.5 Real-time quantitative PCR analysis of miRNA expression

3.2.5.1 cDNA synthesis

100 ng total RNA was reversely transcribed using miScript II Reverse Transcription kit (Qiagen) following the provider's instructions [207]. In general, the miScript RT mix is a mixture composed of poly(A) polymerase and reverse transcriptase, which allows the reverse transcription by oligo(dT) priming. HiFlex buffer was used to meet the need of conversion of all RNA species (mature miRNAs, precursor miRNAs, noncoding RNAs, and mRNAs). Briefly, 4µL of the miScript HiFlex buffer, 2 µL of nucleic acids and miScript RT mix and the variable volume of RNase free water was added to RNA template (100 ng) to reach the final reaction volume of 20 µL. The cDNA synthesis was performed at 37°C for 1h. Subsequently, the reaction was heat inactivated at 95°C for 5 min. The reaction components are shown in **Table 3-11**.

To check for potential contamination of genomic DNA in RNA isolates, a negative control (non-RT) was used. Here the same reagents were used with the exception of the reverse transcriptase, which was substituted by RNase free water.

Table 3-11: Reverse transcription reaction components

Reagent	Amount (1X)
miScript HiFlex buffer (5X)	4 µl
Nucleic acids (10X)	2 µl
miScript RT mix	2 µl
RNase free water	Variable
RNA template (100 ng)	Variable
Total volume	20 µl

3.2.5.2 Real-time quantitative PCR (qRT-PCR)

The cDNA was further diluted 1:8 and used as a template for qPCR reaction with the use of miScript SYBR Green PCR kit (Qiagen) and the commercial forward primer (Qiagen). The miScript forward primers used for the quantification are listed in **Table 7-2** (**Appendix**). Universal poly(A) primer, which serves as a reverse primer, binds

to all the cDNAs synthesized using oligo(dT) primers. The detailed protocol for the miScript qPCR reaction is shown in **Table 3-12**. The cycling conditions for real-time PCR were modified according to Qiagen's instructions [207] (**Table 3-13**). The qPCR reaction was performed in Roche Light Cycler 480.

Table 3-12: Reagents for qPCR

Reagent	Amount (1X)
SYBR green	5 µl
miScript primer assay (forward primer)	1 µl
Universal poly-A (reverse primer)	1 µl
cDNA (1:8)	1 µl
RNase Free water	2 µl
Total volume	10 µl

Table 3-13: Cycling conditions for qPCR

Cycle step	Temperature (°C)	Time (min)	Cycle number
Initial activation step	95	15	1
3-step cycling*			
Denaturation	94	0:15	45
Annealing	55	0:30	
Extension	70	0:30	

*the ramp rate was adjusted to 1°C/s.

The small nuclear and small nucleolar RNAs (SNORD72 and SNORD61) were used as control housekeeping RNAs to normalize the relative amount of expression of miRNAs of interest. The broad list of reference small RNAs is described in the miScript PCR System Handbook [204]. The normalized expression was determined using $2^{-\Delta Ct}$ method [208]. In brief, the average of all Ct values for the housekeeping RNAs (HE) and the miRNA of interest (ME) was calculated. The difference between averaged ME and HE values was then determined (ΔCt) and $2^{-\Delta Ct}$ were calculated**.** Wilcoxon signed-rank test in R was then used to determine the significance of expression changes between normal B cells and CLLs. All samples were measured in triplicates. The technical replicate was excluded from the analysis only if no expression signal was being detected by the LightCycler 480 Software, suggesting a pipetting error.

3.2.6 Small RNA sequencing data

Small RNA sequencing data was obtained from the published study by Ferreira *et al.* [209]. Normalized miRNA read counts (reads per million, RPM) were obtained and used in the downstream analysis to determine CLL-specific miRNAs.

The data was processed by Dr. Thomas Hielscher from Division of Biostatistics, DKFZ, Heidelberg, Germany.

3.2.7 RNA sequencing data

Raw expression data from CLLs was obtained from the published study by Dietrichs *et al.* [210], in which libraries were prepared with Illumina TruSeq RNA sample preparation v2 kit and 76bp-pair-end RNA sequencing (RNA-seq) was performed on Illumina HiSeq 2000 platform with 2-3 samples multiplexed per lane (sequencing chemistry v4; average number of uniquely aligned reads 86,490,476, range 45,057,886-114,248,236). The access to raw sequencing data from normal B cells, generated by Ferreira *et al*. [211], was granted by International Cancer Genome Consortium (ICGC). Each library was prepared with mRNA-seq Illumina TruSeq protocol and pair-end sequenced (76bp read length) on one lane (one sample per lane) of a HiSeq 2000 sequencer (average number of uniquely aligned reads 109,506,086, range 80,223,144-146,462,691.

The two datasets were processed and analyzed using pipeline written by Dr. Naveed Ishaque from Division of Theoretical Bioinformatics at DKFZ, Heidelberg. RNAseq reads from CLLs were demultiplexed and the quality of the raw sequencing data for both datasets was assessed using FastQC version 0.11.5. Overall, all samples had good per-base quality scores (>Q30), therefore they were further processed. RNAseq reads were further aligned to the human reference genome (GRCh 37.1/hg19) using STAR version 2.5.2b [212] with default parameters. RNA-seQC (version 1.1.8) [213] was used to assess the quality of the alignment with the following metrics, i.e. number of alignable reads, duplication rates and rRNA contamination. The statistics of the alignment are presented for normal B cells in **Table 7-3** (**Appendix**) and for CLLs in **Table 7-4** (**Appendix**). Aligned reads were further sorted by coordinates using SAMTools [214] version 1.6. Duplicates were marked using Sambamba [215] version 0.6.5 and kept for further analysis. Read counts per transcript were obtained with featureCounts using Genecode19 as a gene model [216] and a strand specificity parameter, -s 2, meaning that reversely stranded read counting was performed.

The obtained read counts were further loaded into DESeq2 [217], in which both datasets were processed together. Data was adjusted for differing library sizes using size factor estimation. The size factor was estimated by 'median ratio method' by Anders and Huber [217]. In brief, the virtual reference sample (pseudo sample) is calculated by taking for each gene the geometric mean of counts across all the samples. Each sample is then normalized to the pseudo sample, to obtain one scaling factor ('size factor') per sample [217].

In the final step, the regularized log (rlog) transformation was applied. rlog function transforms count data to the log2 scale, minimizing differences between samples for transcripts with small counts and normalizing with respect to library size. Data quality was assessed using Principal Component Analysis (PCA), checking for the reproducibility among biological replicates and for possible batch effects. It was expected that biological replicates of the same condition (normal B cells or CLLs) will cluster together. Indeed, Principal Component (PC)1 and PC2 nicely separated CLLs from normal B cells and biological replicates were clustering together (**Figure 3-3**). No other PC was able to separate B cells or CLLs into subtypes (**Figure 7-1, Figure 7-2** in the **Appendix**).

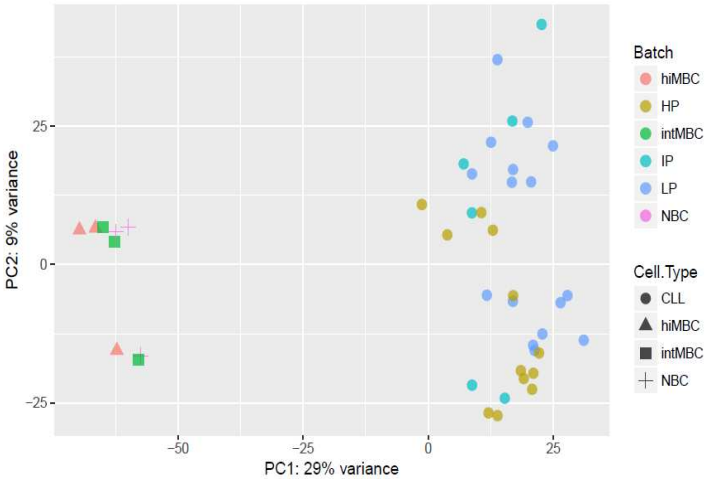


Figure 3-3: Principal Component Analysis (PCA) on rlog transformed expression counts from normal B cells and CLL. Normal B cell subtypes are depicted with crosses (NBC), squares (intMBC) and triangles (hiMBC). CLLs are depicted with dots.

The rlog transformed counts were further used in my analysis as normalized counts and depicted graphically as 'log2 normalized counts'.

3.2.8 450K methylome data analysis

450K data from B cells was obtained from the study by Oakes *et al.* [42]. Similarly, 450K data from CLL patient samples was obtained from the study by Dietrich *et al.* [210]. The analysis of the methylation data was performed using RnBeads software [218]. Both datasets (normal B cells and CLLs) were processed simultaneously. Raw 450K data for both CLL and healthy B cell samples were normalized by the BMIQ method [219] without the background subtraction. The probes overlapping SNPs and the sex chromosomes (X and Y) were removed during the data processing. The remaining probes were considered for the downstream analysis, in which CLL-specific methylation events were identified.

3.2.9 Promoters of miRNAs

To identify promoters of miRNAs, the promoter segmentation data from CLLs, generated by DKFZ PRECiSe consortium (promoter segmentation data is deposited under GSE113336; raw ChIP-seq data can be found in the European Genome-phenome Archive under the accession number EGAS00001002518) and normal cell lines (Encyclopedia of DNA Elements – ENCODE; ENCODE Mar 2012 Freeze, UCSC accession numbers: wgEncodeEH000784, wgEncodeEH000785, wgEncodeEH000790, wgEncodeEH000789, wgEncodeEH000788, wgEncodeEH000786, wgEncodeEH000787, wgEncodeEH000791, wgEncodeEH000792) was used.

3.2.10 TF enrichment analysis

To determine transcription factor binding sites (class A, class C and class E) or recognition motifs of TFs (class D) present within disease-specific methylation events, LOLA package [220] or HOMER software v4.5 [221] were used, respectively. In the former case, TF ChIP-seq data from GM12878 cell line from ENCODE consortium was

downloaded [222] and loaded into LOLA as a database of regions (regionDB). The TF enrichment was subsequently calculated using runLOLA function providing all set of 450K CpG probes as a universe (userUniverse). LOLA statistical model is based on Fisher's exact test with false discovery rate correction for each pairwise comparison. The output of the analysis is presented as a rank score for each region set and is based on three measures: *P*-value, log odds ratio and a number of overlapping regions. In the case of HOMER analysis, class D events were screened against a selected background of CpG sites (all CpGs from the 450k array) that were adjusted to have an equal GC content and the same number of CpGs. For the final conclusion, only 'HOMER known motifs' search algorithm was considered.

3.2.11 Data visualization

All data visualization was performed using R. Consensus clustering and heatmaps were generated using pheatmap package. All the other graphs were generated either using the modified functions present in the graphics or in the ggplot2 package. Custom schematics of the data were prepared in Inkscape 0.91.

4. RESULTS

4.1 Modelling the epigenome of the cell-of-origin

The aim of this work was to identify biologically meaningful CLL-specific aberrant methylation events. In order to achieve this goal, I conceived an analytical framework that consists of the following steps:

1) **To establish an *in silico* model of DNA methylation dynamics during normal B cell differentiation** published Illumina 450K datasets from six normal discrete B cell subpopulations were used [160]. Raw methylation data (.idat files) for both, normal B cells [160] and CLLs [209], was analyzed using the RnBeads R-package and processed simultaneously (see **Methods 3.2.8**) [217]. The normal B cell methylomes were used to define a set of CpG sites (B cell-specific CpGs) with >20% methylation difference between naïve and mature B cells. The hierarchical relationship was inferred between normal B cell subsets ranging from naïve to mature B cells across the entire differentiation spectrum based on their DNA methylation patterns on B cell-specific CpGs. Pairwise Manhattan distances on these methylation profiles were used to determine the mode of methylation progression during B cell differentiation and to infer differentiation stage for every B cell across the entire differentiation spectrum. Each node in the phylogenetic tree corresponds to a certain differentiation stage reached by the normal B cell. A linear, non-branched progression of methylation profiles was observed during B cell differentiation (minimum evolution method in the ape package and F-test). This allowed me to build a linear model of normal B cell differentiation based on DNA methylation patterns at B cell-specific CpGs using data from all B cell-specific CpG sites (see **Results 4.1.1** for more details).

2) **To model the putative cell-of-origin for each CLL sample in the present patient cohort**, published DNA methylome data sets from the normal B cells [160] and 34 CLL samples [209] were used. The differentiation stage of the closest cell-of-origin for every patient was determined using phylogenetic analysis with minimum evolution method implemented in the ape package. Linear regression modeling was further used to infer the DNA methylation levels for each CpG site in the putative cell-of-origin for every patient (see **Results 4.1.1** for more details).

The DNA methylome of the cell-of-origin had to be modeled for every CLL patient individually. This is a necessary step since CLL methylomes are composed of at least two major methylation signatures: the former signature stemming from the leukemia-initiating cell, and the latter reflects disease-specific methylation events which could be further subdivided in driver and passenger events shaped by epigenetic alterations occurring only during leukemogenesis (CLL-specific profiles).

3) **To identify disease (CLL-) specific methylation events**, the epigenome of the cell-of-origin was used as a reference for aberrant DNA methylation calls across all CpG sites in a given sample. CLL-specific CpGs were defined as sites with >20% deviation from the computed DNA methylation levels in the corresponding cell-of-origin. Two categories of CLL-specific methylation events; those occurring at sites undergoing epigenetic programming during

B cell differentiation and those that normally do not change during B cell differentiation were identified (see **Results 4.2** for more details).

4.1.1 Modelling of normal B cell differentiation

The first task was to identify a function, which precisely models the DNA methylation dynamics observed during normal B cell maturation (differentiation axis).

To do so, DNA methylation programming during normal B cell differentiation was studied using six discrete B cell subpopulations, namely naïve B cells (NBCs, $CD23^{+}IgD^{high}CD27^{-}$), germinal center founder B cells (GCFs, $CD20^{high}CD38^{int}IgM^{+}IgD^{+}CD80^{high}$), early non class-switched memory B cells (loMBCs, $IgM^{+}IgD^{low}CD27^{-}CD23^{-}Rhodamine123^{+}$), non class-switched memory B cells (intMBC, $IgM^{+}IgD^{+}CD27^{+}$), splenic marginal zone B cells (sMGZs, $IgM^{+}CD21^{high}CD27^{+}$), and class-switched memory B cells (hiMBCs, $IgG^{+}CD27^{+}$) using Illumina 450k array on 2-4 donors per each healthy B cell subpopulation [160]. The normal B cell methylomes were used to define a set of the CpG sites (B cell-specific CpGs) with a prominent gain (>20%) or loss of methylation (<20%) during the B cell differentiation process, between naïve and mature B cells. In total, 75,669 CpGs were found to be dynamically regulated during normal B cell differentiation.

I assumed that similar to DNA nucleotide changes reflecting the evolutionary time, DNA methylation dynamics observed during cellular differentiation can be used to infer the differentiation stage achieved. Initially, I hypothesized that DNA methylation dynamics during the B cell differentiation can progress linearly or non-linearly. In the first approach, the methylation profiles of normal B cell subpopulations represent a continuum of differentiation stages, and this can be described using a simple linear function. In the second model, I would expect a non-linear mode of methylation progression during normal B cell differentiation, allowing a branched evolution of the various B cell subtypes. To determine the mode of methylation progression, the hierarchical (phylogenetic) relationship of normal B cell subsets, from NBCs to hiMBCs, was inferred based on their DNA methylation profiles. For this, DNA methylation profiles of B cell-specific CpGs were used to generate matrices of pairwise distances (Manhattan distance) which were subsequently used to draw phylogenetic trees. For the tree construction, the minimum evolution method by Desper and Gascuel [223], implemented in the ape package (fastme.bal function), was used. Briefly, this method computes the sum (S) of estimates of the length of tree branches for all possible topologies, and the topology having the smallest S value is chosen as the best tree [223]. Using this approach, a linear, non-branched progression of methylation profiles was observed during normal B cell differentiation (**Figure 4-1A**).

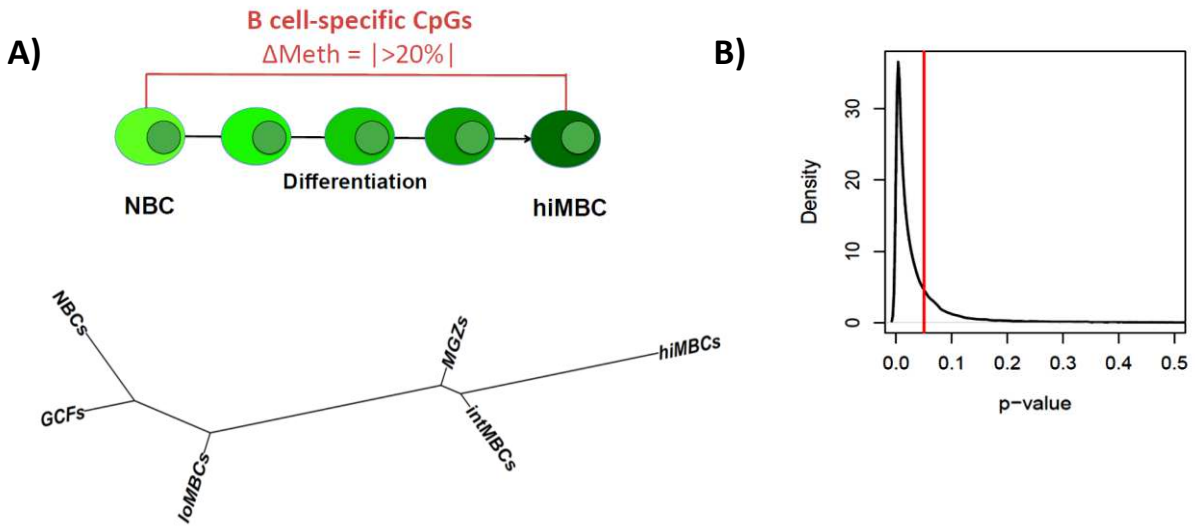


Figure 4-1: Linear progression of methylation profiles during normal B cell differentiation. A) DNA methylation-based phylogenetic tree of normal B cell differentiation. The phylogenetic tree was generated using a set of CpG sites that show dynamic methylation changes during normal B cell differentiation (B cell-specific CpGs). Pairwise Manhattan distances between DNA methylation profiles of normal B cells at B cell-specific CpG sites were used to model methylation dynamics from naïve to mature B cells. NBCs - Naïve B cells; GCFs – Germinal center founder B cells; loMBCs – Early non class-switched memory B cells; intMBCs – Non class-switched memory B cells; sMGZs – Splenic marginal zone B cells; hiMBCs – Class-switched memory B cells (mature B cells). **B) Linear relationship between the differentiation stage of B cell subtypes and the methylation profiles at B cell-specific CpGs.** F-test statistics were used to test for a linear relationship between the assigned differentiation stage for every B cell and the methylation values at B cell-specific sites, at the single CpG level. The majority of B cell-specific CpGs (78.8%, 59,660 CpGs) undergo a linear progression of DNA methylation (either gain or loss) during normal B cell differentiation. The y-axis represents the density of B cell-specific CpG sites. The x-axis represents p-values from the F-test. The red line indicates a p-value=0.05.

Furthermore, the pairwise distances between the individual B cell methylation profiles at B cell-specific CpGs were used to define a differentiation stage for every B cell across the differentiation spectrum. Each node in the phylogenetic tree corresponds to a differentiation stage reached by a hypothetic B cell. The linear relationship between the differentiation stage achieved and the DNA methylation profiles at B cell-specific CpGs was tested at the single CpG level (F-test, **Figure 4-1B**). For the vast majority of B cell-specific CpGs (78.8%, 59,660 CpGs), a linear relationship between DNA methylation dynamics and differentiation stage achieved was observed across six distinct B cell differentiation stages. This allowed me to build a linear model of normal B cell differentiation based on DNA methylation profiles in B cell subtypes.

4.1.2 Identification of the cell-of-origin and its epigenome

My goal was to identify a unique cell-of-origin for every CLL patient. The cancer cell-of-origin is defined as the cell that acquires the first oncogenic hit and, once additional hits have been acquired, this cell initiates tumorigenic growth [202]. Although several mutations creating a preleukemic clone including *SF3B1*, *NOTCH1* or *TP53* have

been identified in the HSC pool of CLLs, the second driver event, either genetic or epigenetic, is required for the transformation [16]. This driver event blocks the B cell differentiation, providing the cells with a proliferative advantage.

To identify cell-of-origin for every patient, I modeled all potential B cell subpopulations present in the B cell differentiation axis and assigned the closest virtual normal B cell methylome (cell-of-origin) to every CLL case in the patient cohort. Firstly, a phylogenetic analysis of all CLL samples in relation to normal B cell differentiation was performed using DNA methylation levels at B cell-specific CpG sites only (ape package, minimum evolution method) [223]. Briefly, DNA methylation patterns were used to calculate a pairwise distance matrix for normal B cells and CLLs (using Manhattan distance). Subsequently, phylogenetic trees were inferred with plot.phylo function incorporated into ape package. Unrooted and phylogram topologies were used. Each CLL case was precisely positioned onto the normal B cell differentiation trajectory defining the closest normal B cell on the normal differentiation axis that underwent transformation event. In other words, the differentiation stage of the cell-of-origin was assigned as the position of the closest node in the phylogenetic tree for each CLL case (**Figure 4-2**). The position of each node was determined using edge.length function from the Ape package.

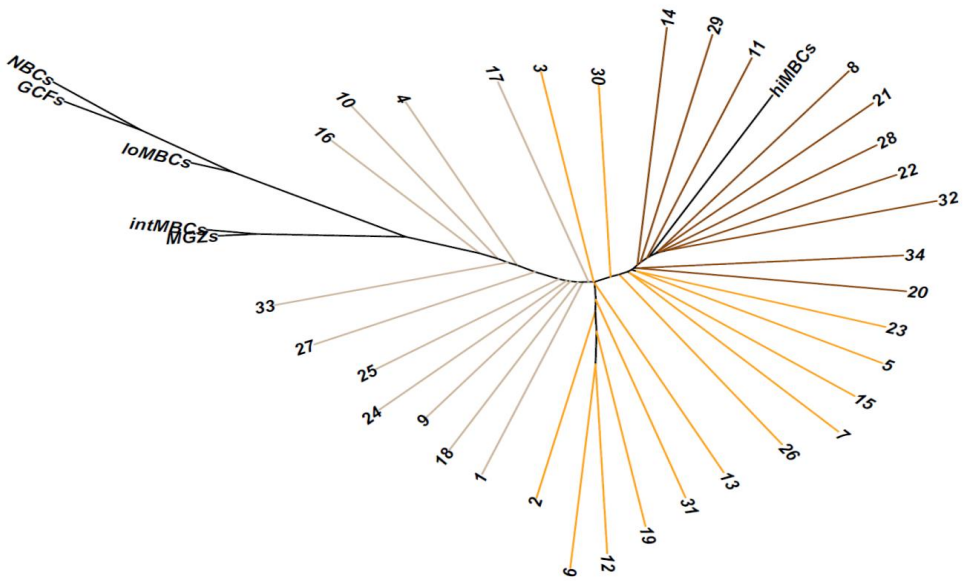


Figure 4-2: Identification of the cell-of-origin in CLL samples. A phylogenetic tree was generated using a set of CpG sites that show dynamic methylation changes during normal B cell differentiation (B cell-specific CpGs). Pairwise Manhattan distances were calculated between DNA methylation profiles of normal B cells and CLL samples at B cell-specific CpGs and were subsequently used to assign the closest normal (virtual) B cell methylome (differentiation stage of the cell-of-origin) to each CLL case. NBCs - Naïve B cells; GCFs – Germinal center founder B cells; loMBCs – Early non class-switched memory B cells; intMBCs – Non class-switched memory B cells; sMGZs – Splenic marginal zone B cells; hiMBCs – Class-switched memory B cells (mature B cells); 1:34 – CLL patients. The gradient color code of CLL samples corresponds to different levels of maturity reached by the cell-of-origin during the transformation event. CLL samples with a relatively immature cell-of-origin that are reprogrammed early during the differentiation process are represented in bisque color. CLLs with a cell-of-origin reprogrammed at a late stage of differentiation and resembling more mature B cells are depicted in dark orange color.

In the next step, the virtual DNA methylome of the cell-of-origin was calculated. The linear regression model was applied to normal B cells and CLLs, and was used to determine the DNA methylation profiles for every CpG site of the cell-of-origin for each of the CLLs according to the following equation:

$$M = \alpha + \beta * d.s.$$

, where

M denotes the calculated beta methylation value for a CpG site of cell-of-origin,
d.s. denotes the differentiation stage (obtained from the phylogenetic analysis),
β denotes the slope of the regression line,
α denotes the vertical (y-axis) intercept.

4.2 Identification of disease-specific methylation events in CLL

During the initial modeling step, the closest virtual normal B cell methylome was assigned to every CLL case, which allowed the calculation of the DNA methylation levels that should be present in the cell-of-origin. This virtual DNA methylome was used to infer disease-specific methylation events in each CLL sample (**Figure 4-3**).

Substituting for $CD19^+$ B cells that are used as controls in most studies, for each CLL sample, the DNA methylation levels present in the cell-of-origin were used as the individual control cell population to identify aberrant CLL-specific methylation events across all CpG sites. Disease-specific CpGs were defined as sites with significant deviation from the expected methylation levels as compared to the corresponding cell-of-origin (**Figure 4-3A**). Two categories of CLL-specific methylation events were identified: 1) CpG sites with methylation programming during normal B cell differentiation (i.e. B cell-specific CpGs) at which CLL samples display inadequate methylation programming as compared to the predicted cell-of-origin ('sites with epigenetic B-cell programming'); 2) CpG sites without methylation programming during normal B cell differentiation at which only CLL samples show aberrant methylation events ('sites without epigenetic B-cell programming'; see **Figure 4-3C** for schematic representation).

The majority of disease-specific methylation events within the category of 'sites with epigenetic B-cell programming' (class A & B) were required to either show a minimum of 20% methylation loss or gain relative to the calculated cell-of-origin methylation value (M value) in at least 95% of the CLL patients. Additionally, hierarchical clustering on B cell-specific CpGs identified a set of CpGs with divergent methylation states across CLL samples (class C; **Figure 4-3C**).

For the identification of CpG sites belonging to the category of 'sites without epigenetic B-cell programming', all non-B cell-specific CpGs were used. The average methylation changes on these sites during normal B cell differentiation, from naive (NBCs) to mature B cells (hiMBCs), were determined. Likewise, the mean methylation profiles were calculated for the CLL cases. Next, the methylation difference between the averaged methylation values for the CLLs and the normal B cell differentiation axis was determined. Each disease-specific methylation

event at 'sites without epigenetic B-cell programming' in CLL was then categorized as either methylation loss (class D) or gain (class E) if there was at least 30% methylation change observed relative to the normal B cell differentiation axis (**Figure 4-3C**).

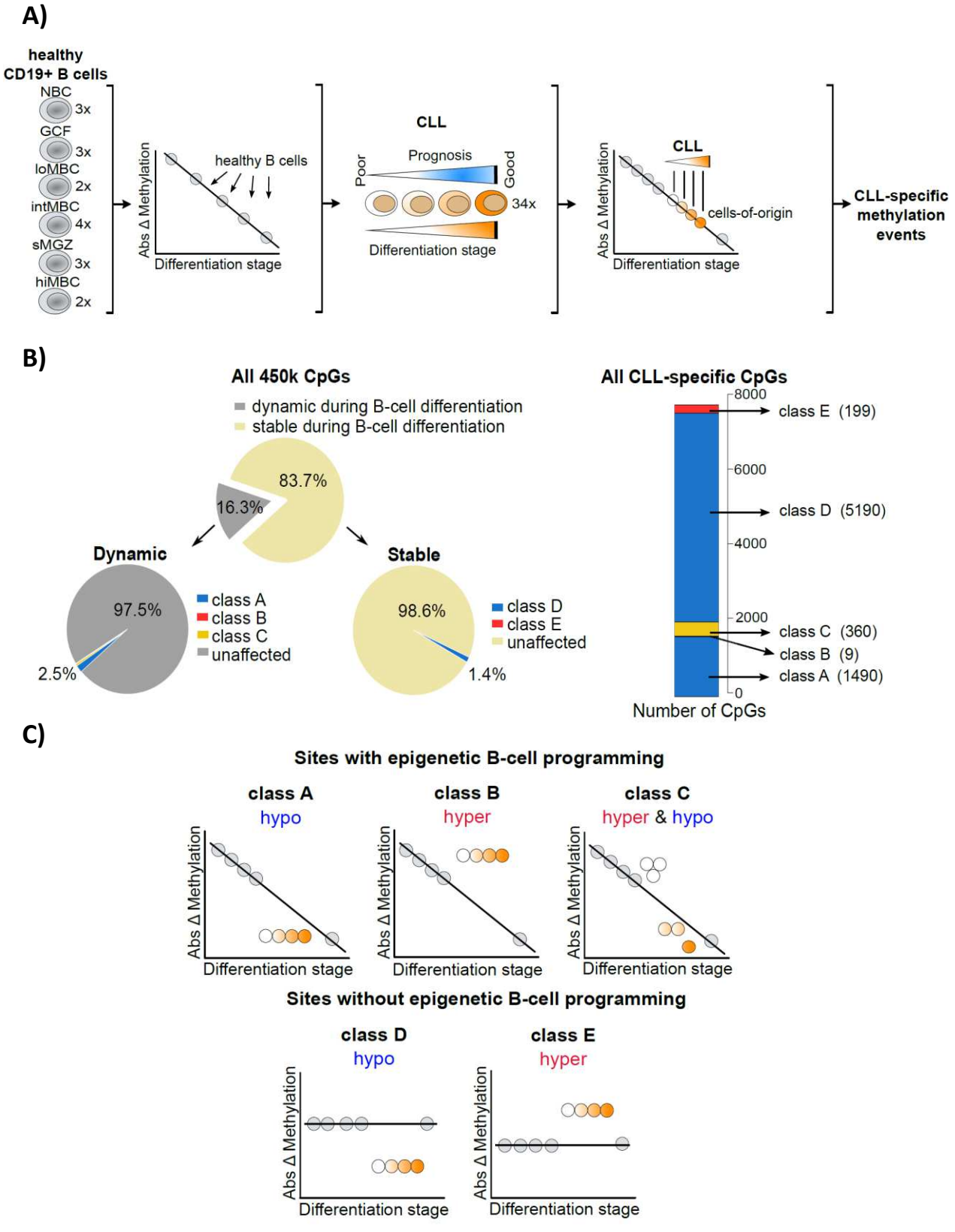


Figure 4-3: Identification of CLL-specific methylation events. A) Schematic outline of the analysis pipeline used for the identification of CLL-specific methylation events. Methylome data on six distinct B-cell subpopulations, representing different stages of terminal B cell differentiation were used for linear regression modeling. DNA methylomes of 34 primary CLL samples of different differentiation stage and prognosis (immature CLLs – bad prognosis and mature CLLs – good prognosis) were used for identification of cell-of-origin and disease-specific methylation events in the context of normal B cell differentiation. Linear regression modeling on DNA methylation patterns on B cell-specific CpG sites facilitated the prediction of the methylomes for the cell-of-origin for each CLL patient in the cohort. The cell-of-origin was defined based on the closest virtual normal B cell methylome present on the regression line of the model. CLL-specific methylation events were inferred based on the deviation from the expected DNA methylation patterning as compared to the assigned cell-of-origin (>20% for 'sites with epigenetic B-cell programming'; >30% for 'sites without epigenetic B-cell programming'). NBC - Naïve B cells; GCF – Germinal center founder B cells; loMBC – Early non class-switched memory B cells; intMBC – Non class-switched memory B cells; sMGZ – Splenic marginal zone B cells; hiMBC – Class-switched memory B cells. **B) Summary of CLL-specific methylation events**. The pie chart displays the frequency of stable and dynamic CpGs, respectively, during the B-cell differentiation (B cell-specific CpGs), the frequency of CLL-specific CpG methylation events (class A, B, C, D, and E), and the number of unaffected CpG sites. The number of CpGs affected by disease-specific methylation events is indicated on the cumulative bar chart. **C) Categorization of CLL-specific methylation events.** CLL-specific methylomes were categorized into two groups: 'sites with epigenetic B cell programming' defined as CpG sites with methylation programming during normal B cell differentiation at which CLLs display aberrant methylation patterns as compared to the predicted cell-of-origin; 'sites without epigenetic B cell programming' defined as CpG sites without methylation programming during normal B cell differentiation at which only CLL genomes display aberrant methylation events. Both categories were further subdivided according to the direction of methylation change as compared to the expected methylation patterns in the cell-of-origin. Class A and D are characterized by hypomethylation relative to normal B cell differentiation, while class B and E show hypermethylation. Class C comprises CpG sites with both hyper- and hypomethylation across CLL samples.

Globally, looking at disease-specific methylation patterns, prominent loss-of-methylation was observed (**Figure 4-4**). Interestingly, only 1.6% of the CpG-sites (7,248 CpGs) represented on the 450k array are affected by disease-specific methylation programming, the majority of which were 'sites without epigenetic B cell programming' (5,389 CpG sites). The majority of CLL-specific methylation events were characterized by hypomethylation (1,490 CpGs in class A and 5,190 CpGs in class D), while only a small proportion of CpGs was hypermethylated compared to their cell-of-origin (9 CpGs in class B and 199 CpGs in class E) (**Figure 4-3B**; **Figure 4-4**). The CpG sites in class C (360 CpGs) displayed either hypermethylation (class C hyper; 307 CpGs) or hypomethylation (class C hypo; 53 CpGs) in immature CLL samples but showed no or opposing methylation changes in mature CLL cases as compared to normal B-cell programming (**Figure 4-3C**). Surprisingly, this phenomenon was not observed at sites without epigenetic B-cell programming'. Overall, this data indicate that loss of DNA methylation is the most frequent CLL-specific finding irrespective of the sites being involved in normal B cell programming or not.

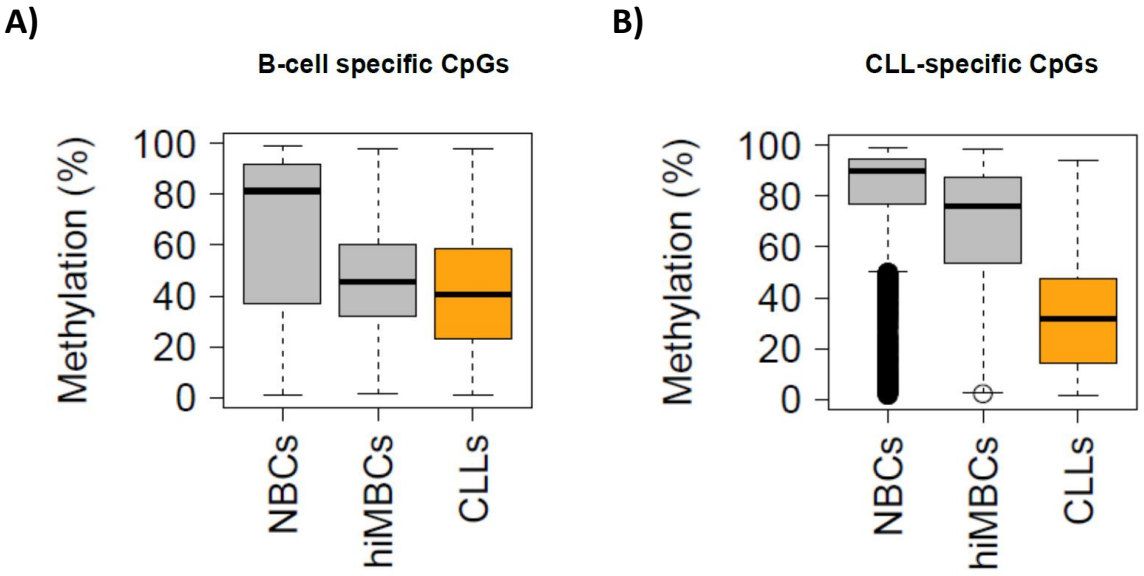


Figure 4-4: Net methylation changes on B cell- and CLL-specific CpGs. A) Net methylation change on B cell-specific CpGs for B cells (grey) and CLLs (orange). The x-axis represents different cell types: NBCs – Naïve B cells; hiMBCs - Class-switched memory B cells; CLL – chronic lymphocytic leukemia cells. The y-axis denotes absolute methylation (%). **B) Net methylation change on CLL-specific CpGs for B cells (grey) and CLLs (orange).**

4.3 Immature CLLs fail to activate normal B cell programming

To further characterize disease-specific methylation patterns in CLL, I quantified the methylation changes for each disease-specific CpG site in each sample as compared to the cell-of-origin and inspected the methylation profiles by unsupervised hierarchical clustering (**Figure 4-5**). For classes A and B, a consistent pattern of either loss or gain of methylation relative to the cell-of-origin was observed, which was uniformly distributed, irrespective of the differentiation state of the CLL cell-of-origin (**Figure 4-5A**, left panel). Hypomethylation at class A sites results from exaggerated loss of methylation at sites which show loss of methylation during normal B-cell differentiation (**Figure 4-5B**, 'class A hypo', 1490 CpGs). Aberrant hypermethylation observed at class B sites occurs mostly on top of hypermethylation normally present during B-cell differentiation or, alternatively, results from failed hypomethylation during normal B cell programming (**Figure 4-5B**, 'class B hyper', 9 CpGs).

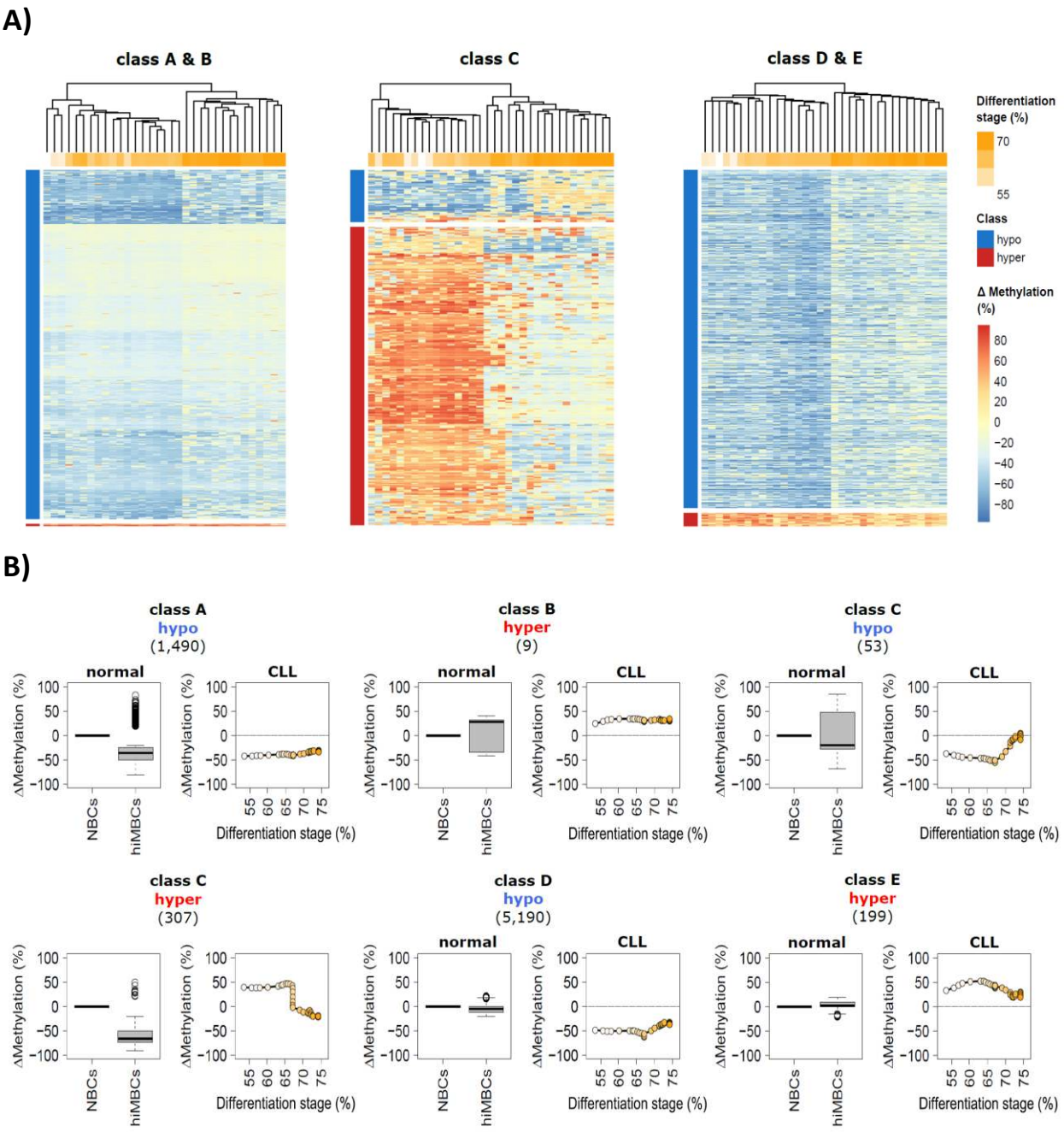


Figure 4-5: Programming of disease-specific methylation patterns in CLL. A) Heatmap showing relative methylation levels at CLL-specific CpGs. Unsupervised hierarchical clustering of CLL-specific methylation events of the subgroups, class A, B (left), class C (middle) and class D and E (right). ΔMethylation represents average methylation change (%) relative to the cell-of-origin. Hypomethylation is denoted as a blue bar, hypermethylation as a red bar on the left-hand side of the heatmaps. Differentiation stages are denoted with the gradient color (white-orange), where CLLs with immature cell-of-origin are depicted in white and the mature one in orange. **B) Box plots displaying average methylation change for each class of CLL-specific alterations across normal B cells and CLL.** Total numbers of the individual CpGs are given. Left (normal), average methylation change (ΔMethylation) of CLL-specific CpGs during normal B cell differentiation from naïve B cells (NBCs) to class-switched memory B cells (hiMBC) plotted for all subclasses (class A, B, C, D, and E). Right (CLL), methylation change (ΔMethylation) for CLL-specific CpGs in CLL. ΔMethylation is represented as methylation change relative to the expected methylation level of the cell-of-origin (dotted line).

For class C events (**Figure 4-5A, Figure 4-5B**, 'class C hyper', 'class C hypo'; 360 CpGs), a pattern of divergent methylation changes was observed across all CLL samples and the majority of these sites result from failure to loose methylation during normal B cell programming and this was more pronounced in more immature CLL samples while the most mature CLL samples presented methylation levels comparable to those seen in normal B cells (**Figure 4-5B**, 'class C hyper'). Of note, the average methylation change was the highest (40%) for class C sites in immature CLLs (**Figure 4-5B**). This indicated a failure to activate normal differentiation programming at these CpGs sites, specifically in the subset of the more aggressive, immature CLL cases. Loss or gain of methylation relative to the cell-of-origin was observed at class D (5,190 CpGs) and E sites (199 CpGs), respectively (**Figure 4-5A and Figure 4-5B**, 'class D', 'class E'). Interestingly, sites from both of these classes of events did not undergo any significant methylation programming during normal B cell differentiation, indicating the potential importance of these sites for CLL pathogenesis.

4.4 Annotation of sequences demonstrating CLL-specific methylation differences identifies aberrant transcription factor programming

Next, I investigated the molecular programs affected by CLL-specific aberrant methylation. Therefore, CLL-specific CpG sites were annotated with chromatin states derived from immortalized B cells (GM12878) [224]. Aberrantly methylated CpGs from class A, C & D were enriched for enhancer elements (class A & C & D), while the CLL-specific gain of methylation in regions not related to B cell maturation (class E) was enriched for weak and poised promoters, repetitive elements and insulator regions (**Figure 4-6A)**. Overall, these results indicate that disease-specific methylation events target transcriptionally relevant sequences in CLL.

As normal B cell differentiation is associated with coordinated expression of many B cell-specific transcription factors (TFs) [225], I further tested which transcription factor binding sites were enriched in CLL-specific differentially methylated regions (see **Methods 3.2.10**). Significant enrichment was observed for B cell-specific TFs, including IKZF1 and BATF (in class A), EBF1, IKZF1, BATF, ATF2 (in class C), and NFATC1 and EGR1 (in class D). Additionally, the enrichment of non-B cell related TFs was observed, NFIC and ATF2 in class C and for TFs related to genome architecture: CTCF, SMC3 and RAD21 in class E CpG sites (**Figure 4-6B**). The latter might explain the observed enrichment of class E CpGs in the insulator regions as CTCF and the cohesin complex composed of e.g. RAD21, SMC3, SMC1, and STAG1/STAG2 are working together to mediate long-range interactions [226].

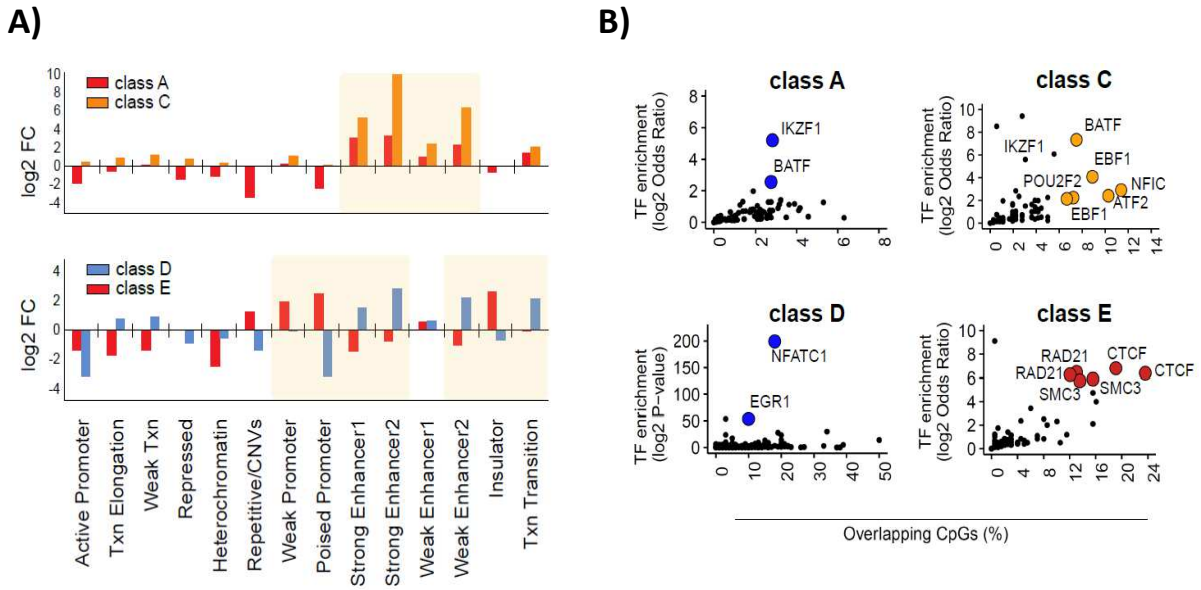


Figure 4-6: A) Enrichment of chromatin states in sequences representing CLL-specific methylation. Chromatin states were annotated to CLL-specific methylation sites of the classes A, C, D and E using the 15-state ChromHMM model derived from immortalized B cells [224]. The enrichment in two chromatin states under the category 'Repetitive/CNV' was represented as the averaged value. Log2 fold change (log2 FC) was calculated using all 450k probes as a background. **B) Transcription factor binding to CLL-specific methylation sites.** Bubble scatterplots represent either transcription factor ChIP-seq peaks from the GM12878 cell line that overlap CLL-specific events (class A and C and E) or transcription factor recognition motifs present at CLL-specific methylation sites (class D). The bubbles are colored according to the class assignment. The x-axis represents the percentage of CLL-specific CpGs overlapping transcription factor binding sites. The y-axis represents the outcome of the enrichment analysis, either as log2 odds ratio or log2 p-values (Fisher's exact test).

Next, I hypothesized that the aberrant methylation phenotype in CLL is associated with different TF expression levels, whose sites were enriched in differentially methylated regions. At first, I investigated TF expression levels in the context of normal B cell differentiation (**Figure 4-7**). The analysis revealed (two sample t-test p.value<0.05 & FC >2) that the degree of expression changes in CLL is equal to that observed in normal B cells for most of the TFs, with the exception of SMC3, POU2F2 and EBF1. No significant expression changes were detected for IKZF1, BATF, ATF2, NFIC, NFATC1, EGR1, CTCF and RAD21. Further, I related the expression levels of TFs to differentiation stages of CLLs. Homogenous patterns of overexpression of POU2F2 and downregulation of EBF1 and SMC3 were observed for all CLLs, irrespective of the differentiation stage of their cell-of-origin (**Figure 4-8**). Interestingly, almost complete loss of EBF1 expression was observed in CLLs as compared to normal B cells, suggesting its importance for CLL pathogenesis. This is in line with the previous reports showing that EBF1 is essential for normal B cell differentiation, and that genetic disruption of EBF1 contributes to leukemogenesis [227, 228].

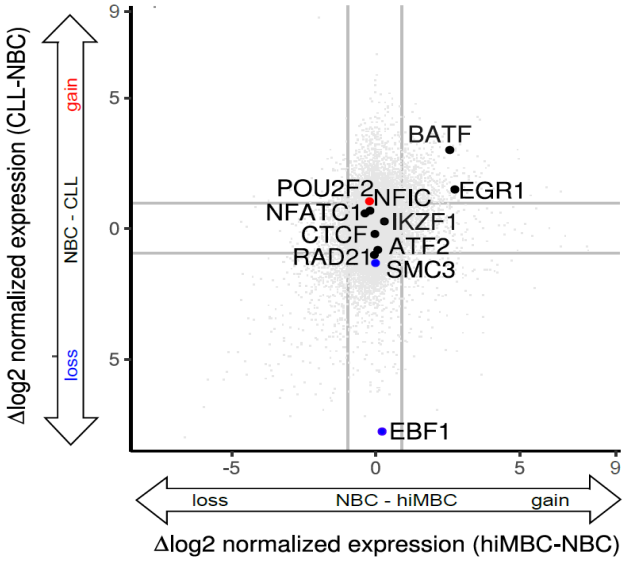


Figure 4-7: Differential expression of TFs between normal B cells and CLLs. Differences in the expression levels of TFs from naive B cells (NBCs) to high-maturity memory B cells (hiMBCs) and to all CLLs using RNA-seq data. TF expression levels were averaged across replicates for each subtype (naïve B cells, *n* = 3; high-maturity memory B cells, *n* = 3; CLLs, *n* = 34). Differentially expressed TFs (abs FC>2 and two-sample t-test p.value <0.05) are either downregulated (dark blue) or upregulated (red) in CLLs as compared to normal B cells.

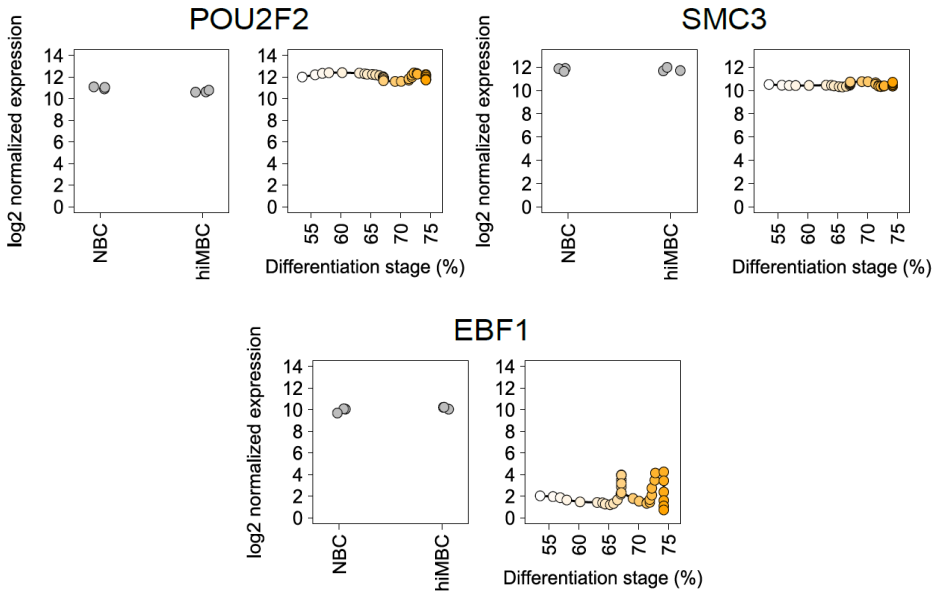


Figure 4-8: Expression data of transcription factors enriched in CLL-specific methylation sites. Left, expression levels (rlog normalized) of transcription factors in healthy B cells ('normal'), representing terminal stages of B-cell differentiation (NBC - Naïve B cell; hiMBC – Class-switched memory B cells). Right, expression levels of transcription factors in CLL. The y-axis represents normalized expression levels for TFs (rlog normalized values). Differentiation stage is denoted with the gradient color (white-orange), where immature CLLs are represented in white and more mature ones in orange.

Next, the potential impact of aberrant promoter methylation on TF expression was tested. Here only *EBF1* promoter hypermethylation correlated with transcriptional *EBF1* downregulation in CLL (**Figure 4-9)**. Other TFs did not show any significant changes in promoter methylation, which could, at least in part, be attributed to the limited CpG coverage of the 450K array (**Figure 7-3** in the **Appendix)**.

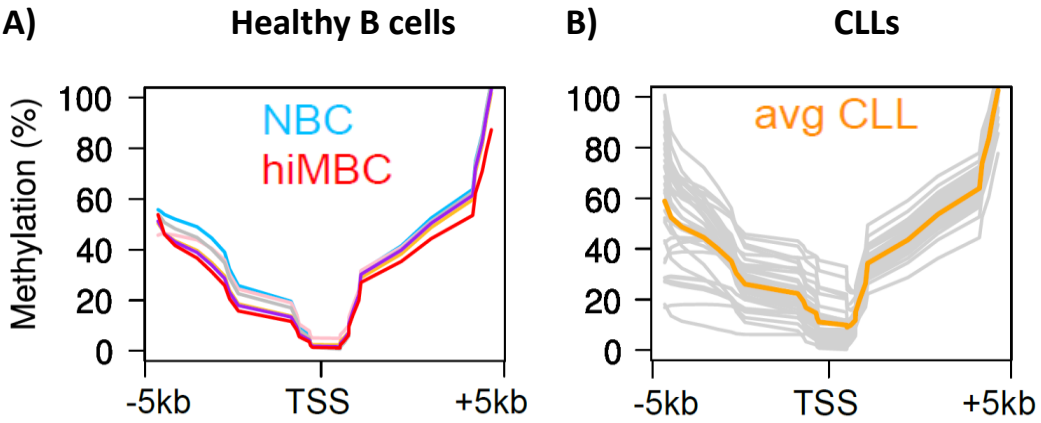


Figure 4-9: DNA methylation profiles of the *EBF1* promoter. DNA methylation in the promoter region was shown for two normal B cell subsets, representing advanced stages of B-cell differentiation (left) and the CLLs (right). NBC - Naïve B cells; hiMBC – Class-switched memory B cells; avg CLL – average methylation change in CLLs. Y-axis represents methylation levels (%). The x-axis represents the distance to transcription start site (TSS) in kb.

4.5 CLL-specific microRNAs

Previous studies identified hundreds of epigenetic events deregulating the expression of miRNAs in CLL cells [155, 189, 195-200]. However, all published studies used $CD19^{+}$ B cells as controls. By doing so, they neglected the broad epigenetic reprogramming that occurs during normal B cell differentiation [155, 189, 195-200]. To demonstrate that the knowledge about CLL-specific aberrant methylation patterns is indispensable for the identification of important pathogenic events in CLL, the contribution of epigenetically deregulated miRNAs in CLL was re-analyzed.

To do so, a strategy for an unbiased identification of miRNA regulatory regions was developed. Functional genome segmentation data derived from ChIP-seq experiments performed on CLL samples and on all available ENCODE cell lines (see **Methods 3.2.9** for the data source) was used to identify genomic segments that show chromatin characteristics indicative of promoter activity. All available active and weak promoter segmentation tracks were used for this analysis. Briefly, promoter tracks from CLLs and normal cell lines were put together and the function reduce from GenomicRanges package was used to define constant promoter regions. The reduce method takes into consideration all promoter segments, and merges them together to produce a simplified, joined promoter region (**Figure 4-10**). Potential promoters of pri-miRNAs were then assigned based on the distance of promoters to the annotated transcription start sites (TSS) of the pri-miRNAs. The annotation of pri-miRNAs/miRNAs was downloaded from miRBase (version 20) [229]. Every promoter located within 100kb from TSS of the pri-miRNA was considered as a putative promoter of the pri-miRNA. The distance of 100kb was chosen based on similar distance

constraints used in the past by Corcoran *et al.*, Fujita *et al*. and Fukao *et al*. [230-232]. The larger distance of putative promoters to the pri-miRNA TSS is especially important in case of intergenic miRNAs, which are originating from intronic sequences and are believed to be transcribed together with their host gene.

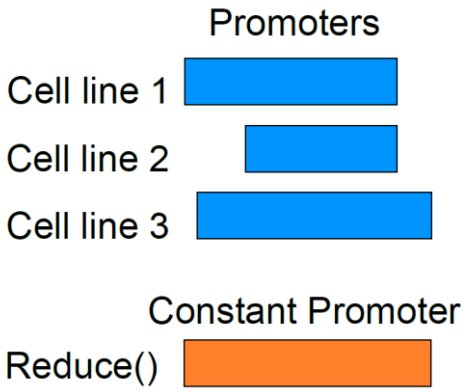
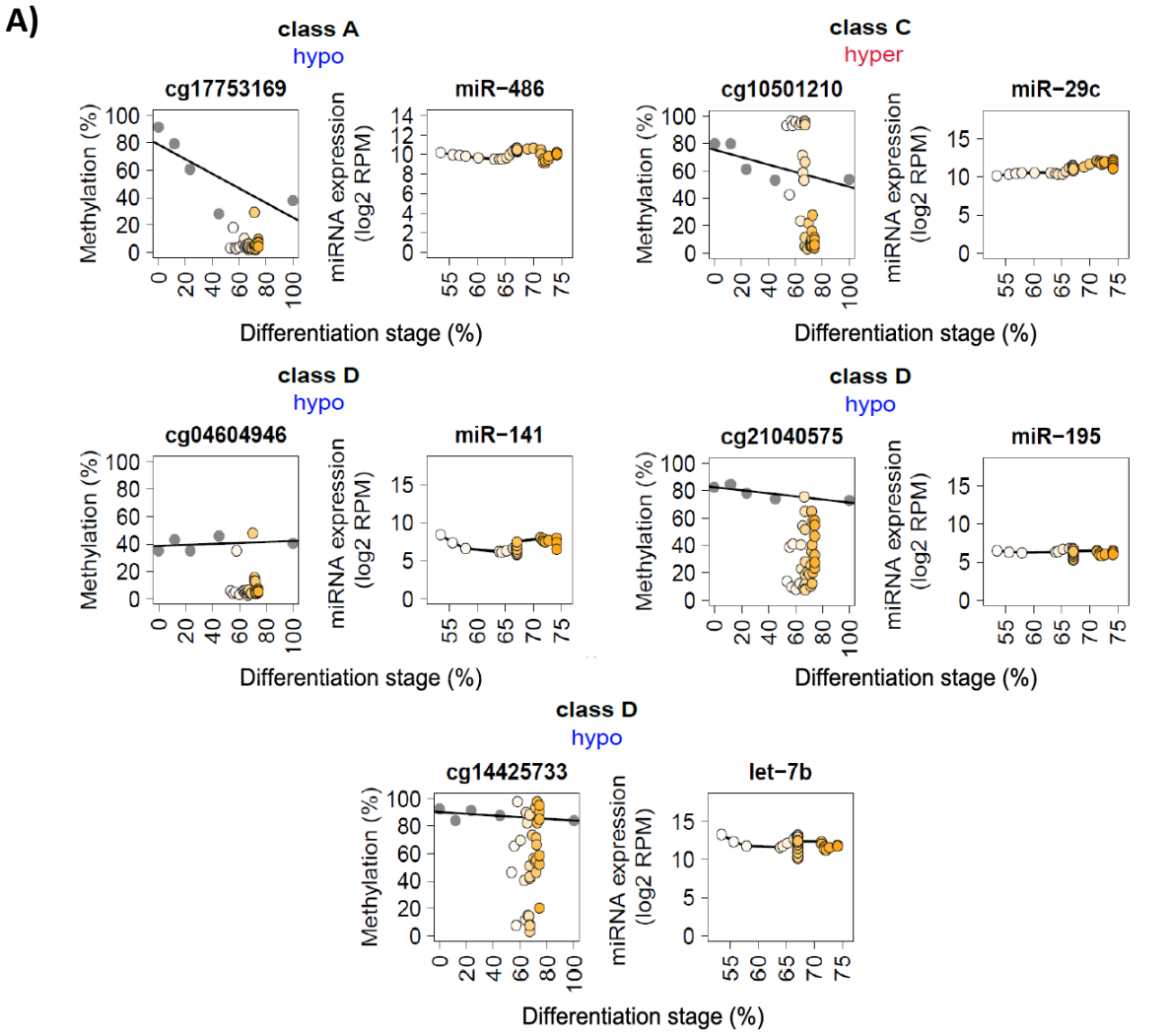


Figure 4-10: **Schematic representation of definition of the miRNA promoter region**. The reduce function from GenomicRanges package was used to create a joint, simplified promoter region, present in all cell lines.

To identify CLL-specific miRNAs, CLL-specific methylation events were overlapped with potential promoters of pri-miRNAs and the correlation (Spearman correlation test, p-value > 0.05; abs(correlation coefficient ρ)>0.4) between disease-specific DNA methylation and the expression of pri-miRNA in CLLs was computed. Since many mature miRNAs are derived from the same pri-miRNAs, correlations were calculated using small RNA sequencing data for pri-miRNAs. Testing the impact of CLL-specific methylation in the promoter regions on miRNA expression identified seven candidate miRNAs that showed a correlation of DNA methylation with miRNA expression: miR-486 and miR-3688-2 in class A; miR-29c in class C hyper; miR-141, miR-195, let-7b and miR-3605 in class D (**Figure 4-11A**). Hypomethylation accounted for the majority of events (6 microRNA candidates), and only one class C hypermethylation event correlated with miR-29c expression in CLL (**Figure 4-11A**). Although there were reports showing aberrant expression of miR-29c and miR-195 in CLL [233-235], none of the candidate regions identified by us had been described previously as being deregulated by aberrant DNA methylation in CLL.



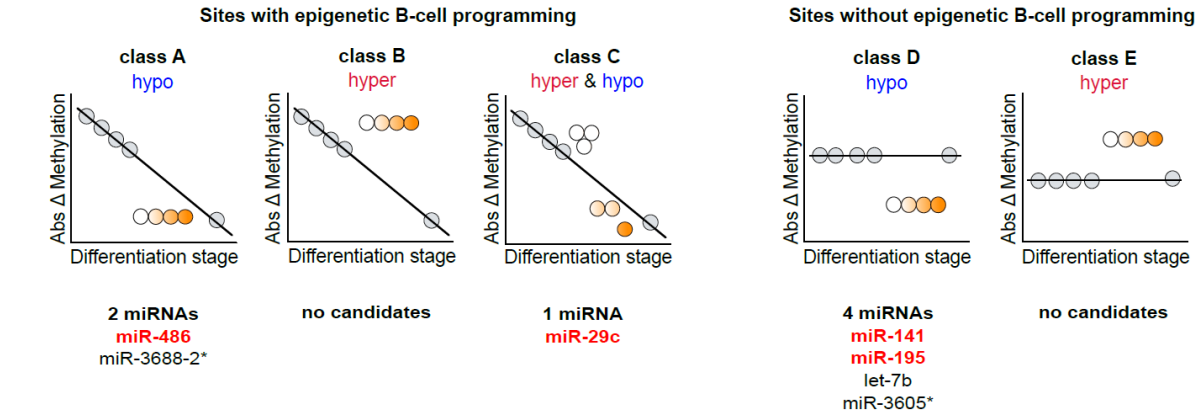


Figure 4-11: Candidate miRNAs associated with CLL-specific aberrant DNA methylation. A) <u>Left</u>, CLL-specific differentially methylated CpGs identified in microRNA promoters grouped by subclasses (class A, C, D and E). Epigenetic programming during normal B cell differentiation is represented as a grey line. Average methylation values are represented as dots; normal B cell subpopulations (grey dots); CLL samples (white-orange dots). Y-axis represents methylation levels (%), while X-axis assigns differentiation stage either to normal B cell or CLLs. **<u>Right</u>, expression levels of candidate microRNAs in CLL (log2 RPM)**. Y-axis represents log2 normalized expression values of microRNAs (log2 RPM), while X-axis assigns differentiation stage either to normal B cell or CLLs. **B) Schematic**

representation of CLL-specific microRNAs. Validated CLL-specific microRNAs are represented in red. Lowly expressed candidate CLL-specific microRNAs, which were excluded from the validation studies, are marked with asterisks.

Aberrant CLL-specific methylation and expression patterns of candidate miRNAs were confirmed in independent CLL and normal B cell samples from two differentiation stages (naive and mature B cells). Two candidate miRNAs (miR-3688-2 & miR-3605) were excluded from the validation experiments since their expression levels were close to the detection limit. From the remaining 5 candidate miRNAs I was able to validate 4 as being specifically aberrantly expressed in CLL: miR-486 (miR-486-5p), miR-29c (miR-29c-3p), miR-141 (miR-141-3p) and miR-195 (miR-195-5p) (**Figure 4-11B, Figure 4-12, Table 7-5** in the **Appendix**). CLL-specific promoter hypomethylation of miR-486 is paralleled by increased expression. Normal B cells also show loss of methylation in the miR-486 promoter, although to a much lesser extent as observed in CLL, and expression of miR-486 remains stable during B cell differentiation. The divergent miR-29c promoter methylation pattern seen in immature versus mature CLLs (class C sites), with strong hypermethylation in immature CLLs and slight hypomethylation in more mature CLLs, was confirmed in the validation experiments. Nevertheless, we observed higher expression levels of miR-29c in all CLLs as compared to normal B cells. For all miRNA candidates with promoter methylation patterns belonging to class D, I validated CLL-specific promoter hypomethylation that was paralleled by the increased miRNA expression in CLL, while we did not observe methylation programming in their promoters during normal B cell differentiation (**Figure 4-12**).

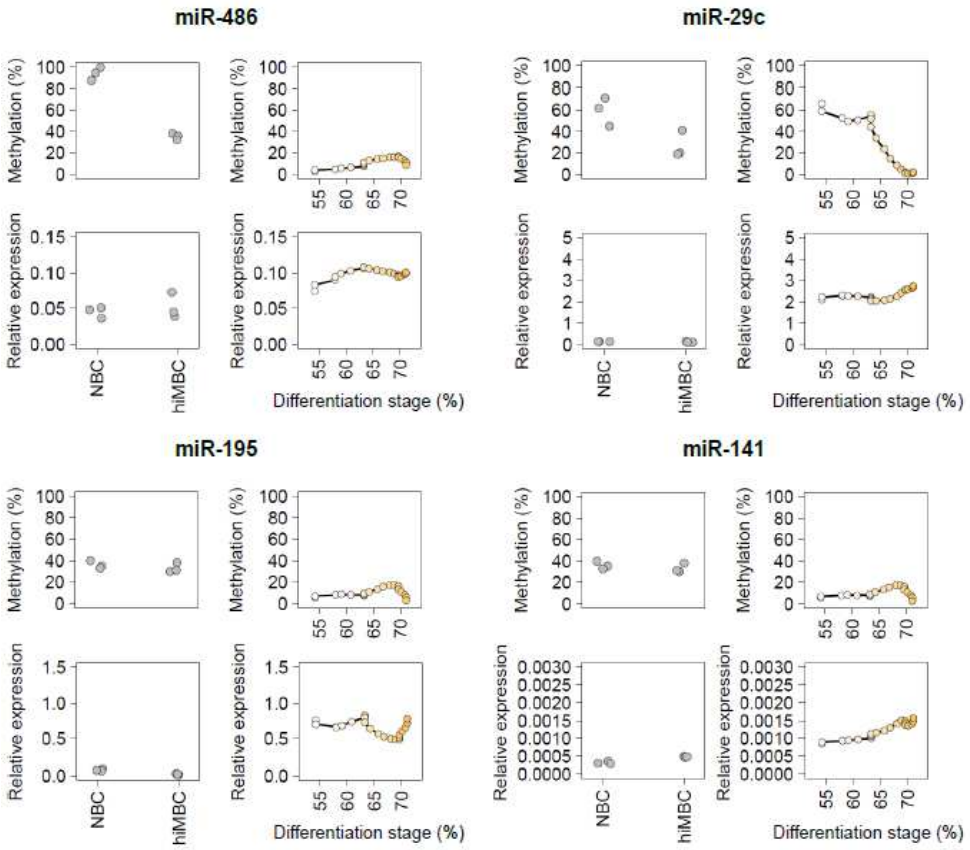


Figure 4-12: Validation experiments for CLL-specific miRNAs. Top, MassArray validation of CLL-specific methylation events in miRNA promoters. Average methylation values are represented as dots; normal B cell subpopulations (grey dots; left panel); CLL samples (white-orange dots; right panel). Bottom, Relative expression levels of CLL-specific microRNAs in B cells and CLLs. The relative expression levels are represented as dots; normal B cell subpopulations (grey dots; left panel); CLL samples (white-orange dots; right panel). Differentiation stage of CLLs is denoted with the gradient color (white-orange), where immature CLLs are represented with the white color and more mature ones with the orange color.

To further link CLL-specific microRNAs with their pathogenetic effects, I searched two databases of experimentally validated microRNA-target gene interactions (TarBase v7.0, miRTarBase). Among the validated targets of the newly identified disease-specific miRNAs are well known recurrently mutated epigenetic regulators in CLL: miR-486 interacts with *ARID1A*, miR-195 & miR-141 interact with *CHD2* and with *ASXL1* (**Figure 4-13; Table 7-6: Table 7-13** in the **Appendix**) [54]. Of note, many other epigenetic regulators are predicted targets of CLL-specific microRNAs, e.g. *EZH2*, *DNMT3A*, *SIRT1*, *TET2*, *SETD1A* and *H3F3B* (**Figure 4-13; Table 7-6: Table 7-12** in the **Appendix**). These findings suggest epigenetic dysregulation of microRNAs as an alternative mechanism for the inactivation of these epigenetic regulators in CLL.

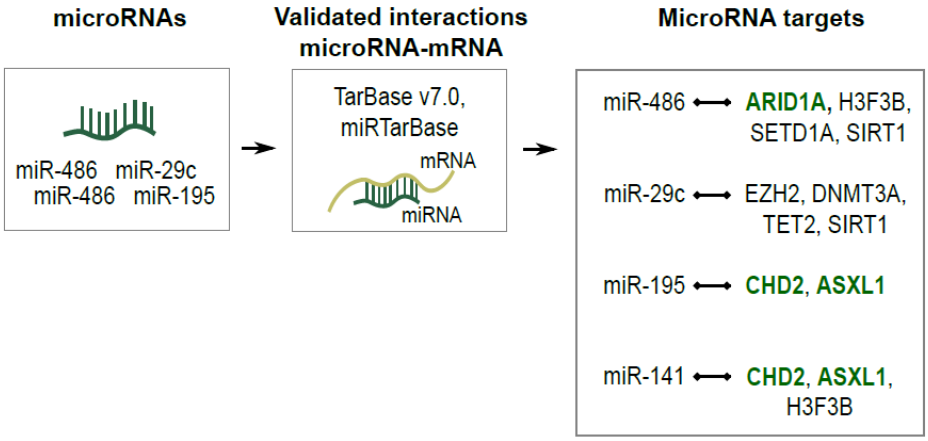


Figure 4-13: Schematic outline of results for miRNA target gene prediction. Two databases of experimentally validated targets of microRNAs, TarBase v7.0 and miRTarBase, were used to define a set of CLL-specific microRNA targets. Recurrently mutated genes in CLL are presented in green. The full list of targets is presented in **Table 7-6 to Table 7-13** in the **Appendix**.

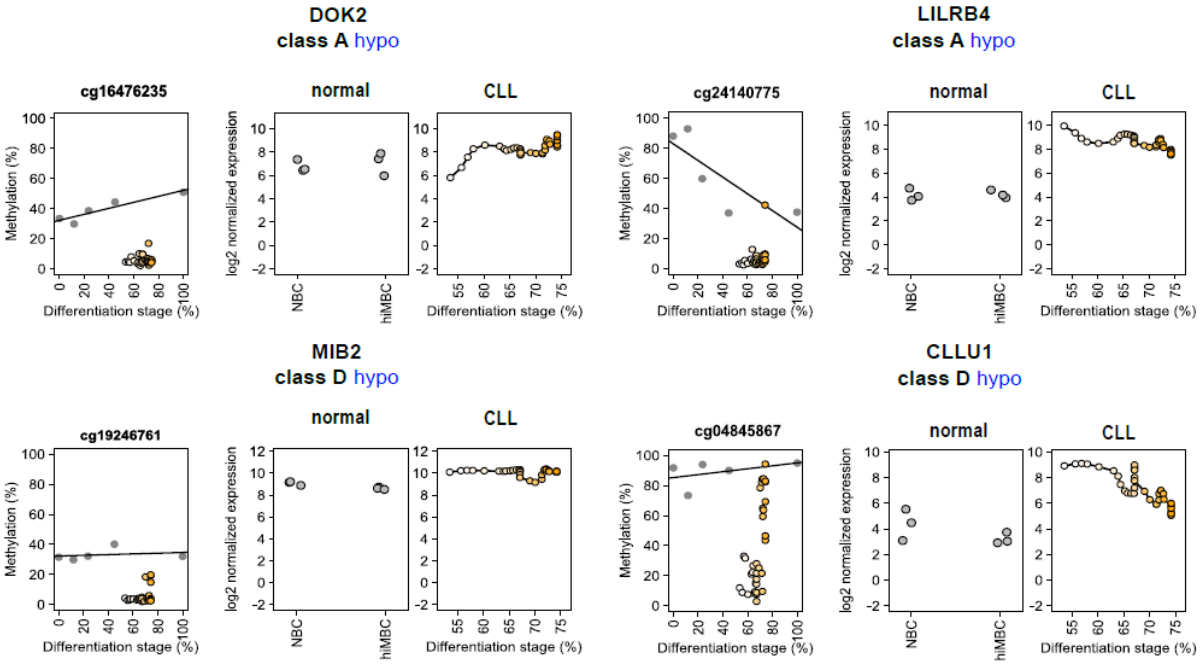
4.6 CLL-specific protein-coding genes

Previous studies identified many epigenetic events deregulating the expression of protein-coding genes in CLL cells (**Section 1.3.3**). However, as in the case of miRNAs, all of the published work used $CD19^{+}$ B cells as controls to call aberrant DNA methylation [142-153]. To demonstrate that identifying CLL-specific DNA methylation events is

indispensable for the identification of pathogenic events in CLL, the contribution of epigenetically deregulated protein-coding genes in CLL was re-analyzed.

I used a similar strategy as described above for miRNAs. First, the impact of CLL-specific DNA methylation in promoter regions (-1.5kb, +0.5kb to TSS) of protein-coding genes was assessed. This was achieved by correlating DNA methylation with gene expression (Spearman correlation test). Using this approach, I identified 20 CLL-specific protein coding-genes (p-value<0.05; abs(correlation coefficient)>0.4)(**Table 7-14** in the **Appendix**). The expression levels of these genes were compared to those observed in normal B cells and only those genes with a significant expression change (log2FC>1) between CLLs and normal B cells were considered as CLL-specific. Using this strategy, I was able to identify 11 CLL-specifically deregulated protein-coding genes (**Figure 4-14, Figure 7-4** in the **Appendix**).

A)



B)

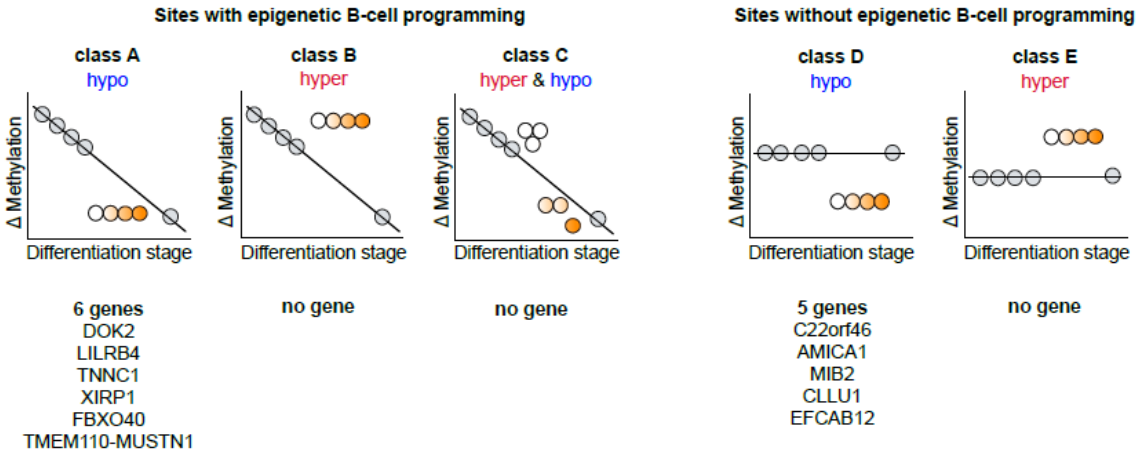


Figure 4-14: Candidate protein-coding genes associated with CLL-specific aberrant DNA methylation (exemplary data). A) Left panel, CLL-specific differentially methylated CpGs identified in gene promoters grouped by subclasses (class A, C, D and E). Epigenetic programming during normal B cell differentiation is represented as a grey line. Average methylation values are represented as dots; normal B cell subpopulations (grey dots); CLL samples (white-orange dots). Y-axis represents methylation levels (%), while X-axis differentiation stage assigned either to normal B cell or CLLs. **Middle panel, expression levels of protein-coding genes in normal B cells** (log2 normalized expression values, NBC; naive B cells, hiMBC; high maturity memory B cells). **Right panel, expression levels of protein-coding genes in CLL** (log2 normalized expression values). The y-axis represents log2 normalized expression values of protein-coding genes, and the x-axis assigns differentiation stage either to normal B cells or to CLLs. The remaining plots can be found in **Figure 7-4** in the **Appendix. B) Schematic representation of CLL-specific protein-coding genes identified.**

Further exploration of the function of CLL-specific protein-coding genes by functional enrichment analysis (MSigDB, GO analysis) was not successful due to the small size of the gene set. Therefore, for functional annotation, I decided to search the existing literature using PubMed. Two genes among the identified CLL-specific protein-coding genes, *DOK2* and *CLLU1*, were previously reported as differentially methylated in AML and CLL, respectively [144, 236] (**Figure 4-15**). Interestingly, they were also linked to prognostic features of patients with gastric adenocarcinoma, AML and CLL [237].

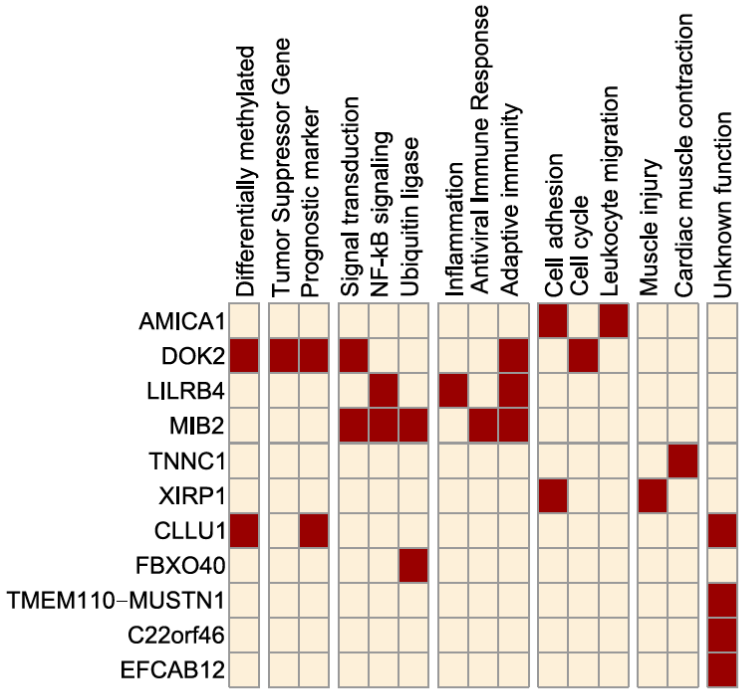


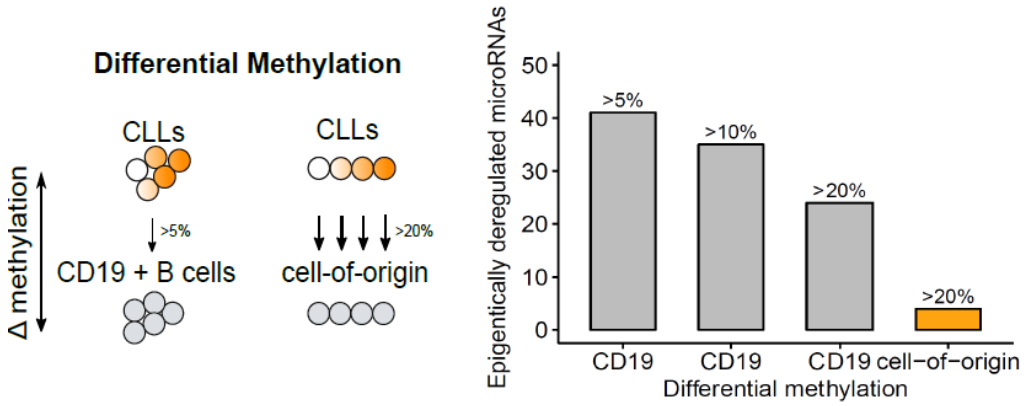
Figure 4-15: Functional annotation of candidate genes. The annotations were assigned based on thorough literature screen (PubMed). Genes are presented in rows. Annotations are presented in columns. The source of annotation is presented in **Table 4-1**.

Table 4-1: CLL-specific annotation – literature

Gene	Function	Publication
AMICA1	Cell adhesion, Leukocyte migration	Weber *et al.* [238], Verdino P *et al.* [239], Luissint AC *et al.* [240], Moog-Lutz C *et al.* [241]
DOK2	Tumor suppressor gene	Berger AH *et al.* [242], Niki M *et al.* [243], Yasuda T *et al.* [244], Ann CH *et al.* [245], Coppin E *et al.* [246]
	Differentially methylated	Fang F *et al.* [247], Lum E *et al.* [248], He PF *et al.* [236]
	Prognostic marker	Fang F *et al.* [247], He PF *et al.* [236], Huang J *et al.* [249], Miyagaki H *et al. [237]*
	Signal transduction	Laroche-Lefebvre C *et al.* [250], Downer EJ *et al.* [251], Mihrshahi R *et al.* [252, 253], Yasuda T *et al.* [254], Van Slyke P *et al.* [255], Abramson J *et al.* [256], Suzu S *et al.* [257], Jones N *et al.* [258]
	Adaptive immunity	Celis-Gutierrez J *et al.* [259]
	Cell Cycle	Coppin E *et al.* [260]
LILRB4	NF-κB signaling	Vlad G *et al.* [261], Buckland M *et al.* [262], Chang CC *et al.* [263]
	Inflammation	Fanning LB *et al.* [264], Chang CC *et al.* [263]
	Adaptive immunity	Cella M *et al.* [265], Bankey PE *et al.* [266], Fanning LB *et al.* [264], Innui M *et al.* [267], Chang CC *et al.* [263], Ju XS *et al.* [268]
MIB2	Signal transduction	Davis ME *et al.* [269], Hu H *et al.* [270], Shi JH and Sun SC [271], Kwon DY *et al.* [272], Ossipova O *et al.* [273]
	NF-κB signaling	Hu H *et al.* [270], Jiang X and Chen ZJ [274], Shi JH and Sun SC *et al.* [271], Chen ZJ *et al.* [275]
	Ubiquitin ligase	Davis ME *et al.* [269], Hu H *et al.* [270], Jiang X and Chen ZJ [274], Shi JH and Sun SC [271], Chen ZJ *et al.* [275]
	Antiviral immune response	Davis ME *et al.* [269], Jiang X and Chen ZJ [274], Wang L *et al.* [276]
	Adaptive immunity	Hu H *et al.* [270], Jiang X and Chen ZJ [274], Shi JH and Sun SC *et al.* [271]
TNNC1	Cardiac muscle contraction	Li MX and Hwang PM [277], Parvatiyar MS *et al.* [278]
XIRP1	Cell adhesion	Wang Q *et al.* [279]
	Muscle injury/damage	Rebalka IA and Hawke TJ [280], Nilsson MI *et al.* [281]
CLLU1	Differentially methylated	Cahill N *et al.* [144]
	Prognostic marker	Buhl AM *et al.* [282], Gonzalez D *et al.* [283], Rosenwald A [284], Abur U *et al.* [285], Kaderi MA *et al.* [286]
FBXO40	Ubiquitin ligase	Deshaies RJ [287]
TMEM110-MUSTN1	Unknown function	
C22orf46		
EFCAB12		

4.7 The importance of usage of proper control cell for aberrant DNA methylation calls

To illustrate the importance of using a proper control B cell population when calling pathogenic disease-specific methylation events, I compared our results with those obtained when using $CD19^+$ B cells as controls. I determined differentially methylated CpG sites between $CD19^+$ B cells and CLLs using different thresholds (5%, 10% and 20%) for the DNA methylation. The use of different thresholds was necessary since previous publications used different cut-offs for calling differential methylation [155, 195, 199]. The differential methylation events identified were overlapped with microRNA promoters and inverse correlation analysis between methylation and expression was applied to identify potentially epigenetically deregulated microRNAs. In this analysis, I identified previously reported microRNAs (e.g. miR-708, miR-23a or miR-10b). An almost 10-fold higher number of CLL-specific miRNAs was detected when applying the „classic" analytical approach (5% threshold: from 41 to 4 microRNAs, 10% threshold: from 35 to 4 microRNAs) (**Figure 4-16**), highlighting the importance of the usage of a proper control B cell in the epigenetic studies in CLL.



Differential methylation (%)	Nr of candidate microRNAs	Previously reported	Examples
5	41	7	miR-708, miR-23a, miR-143,
10	35	7	miR-429, miR-10b, miR-200b,
20	24	4	miR-4326

Figure 4-16: Usage of $CD19^+$ B cells overestimates the proportion of CLL-specific microRNAs. Differential methylation between control and tumor B cells was calculated using different methylation thresholds (5%, 10% or 20%). The bar plot illustrates the proportion of microRNAs defined as CLL-specific using different control B cell sources ($CD19^+$ B cells are represented in grey, individual cell-of-origin is represented in orange). The table depicts the number of candidate CLL-specific microRNAs defined using $CD19^+$ B cells as a control B cell population. The column "previously reported" reflects the number of microRNAs previously reported in the literature as CLL-specific.

Similarly, going beyond the miRNA example, I again applied this comparison using protein-coding genes. The same strategy as described for miRNAs was used and CD19$^+$ B cells were used as controls. Differentially methylated CpG sites between CD19$^+$ B cells and CLLs were calculated using different thresholds (5%, 10% and 20%) for the minimum DNA methylation required. The differential methylation events were overlapped with promoter regions of protein-coding genes (-1.5kb, +0.5kb to TSS) and a correlation analysis between methylation and expression was applied to identify epigenetically deregulated genes (Spearman correlation test, p-value<0.05; abs(correlation coefficient)>0.4)). An almost 30-fold difference in the numbers CLL-specific protein-coding genes were identified between the two different approaches (5% threshold: from 549 to 11 genes, 10% threshold: from 405 to 11 genes), again highlighting the overcalling of differential methylation events in previous studies of CLL using CD19$^+$ B cells as controls (**Figure 4-17**). Interestingly, previously identified differentially methylated promoters of *ZAP70*, *DAPK1*, *TWIST2* or *HOXA4* did not pass the stringent filtering criteria of my correlation analysis (p-value<0.05, abs(correlation coefficient)>0.4)(**Table 7-15** in the **Appendix**). Similarly to miRNAs, these findings stress the importance of the use of appropriate controls for any epigenome study.

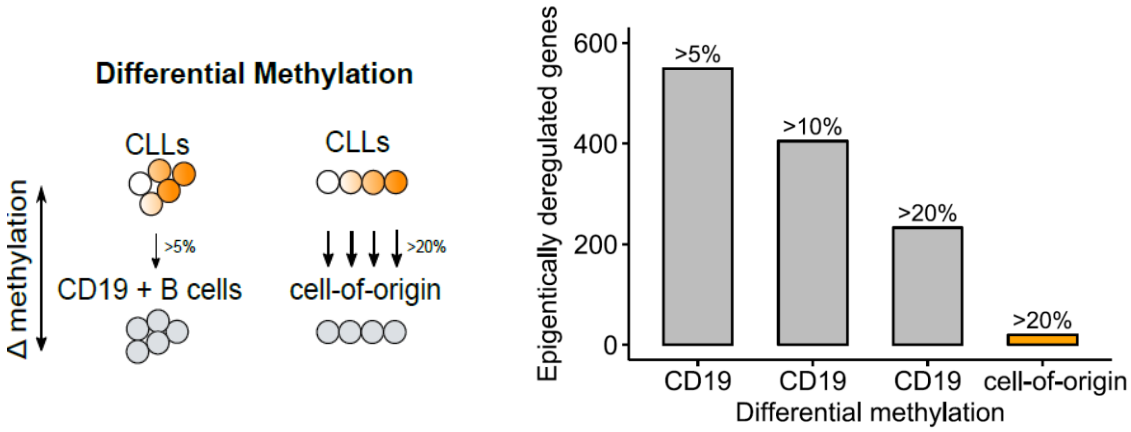


Figure 4-17: Usage of CD19$^+$ B cells overestimates the proportion of CLL-specific protein-coding genes. Differential methylation between control and tumor B cells was calculated using different methylation thresholds (5%, 10% or 20%). The bar plot illustrates the proportion of genes defined as CLL-specific using different control B cell sources (CD19$^+$ B cells are represented in grey, individual cell-of-origin is represented in orange).

5. DISCUSSION

In this thesis, I described a new analytical method to identify cancer-specific DNA methylation patterns using CLL as a model. I showed that knowledge of the epigenome of the cell-of-origin is indispensable to accurately resolve disease-specific methylation events. I determined and dissected two main factors, which together form the epigenetic patterns observed in CLL: 1) DNA methylation patterns derived from the tumor-initiating B cell and 2) epigenetic alterations that are acquired in CLL cells during disease pathogenesis. Using DNA methylomes of normal B cells along the differentiation trajectory, I was able to precisely identify the unique virtual epigenome of the cell-of-origin for every CLL patient, which enabled the delineation of disease-specific methylation events. CLL-specific differentially methylated regions showed evidence for deregulated signaling pathways that are known to be important in normal B cell differentiation as well as for other mechanisms (e.g. CTCF) implicated in CLL pathogenesis. The identification of CLL-specific, epigenetically deregulated miRNAs and protein-coding genes demonstrated that previously used conventional analysis approaches overestimated the proportion of epigenetically deregulated transcripts in CLL. This suggests a need for a proper control B cell for the identification of truly disease-specific methylation events. Furthermore, I found that epigenetically deregulated miRNAs might play a role in epigenetic remodeling in CLL pathogenesis by altering the expression of epigenetic regulators. Finally, based on the assumption that observed DNA methylation occurs in a CLL-specific manner, I propose new candidate genes that might play a role in CLL pathogenesis.

5.1 Modeling of normal B cell differentiation and of the epigenome of the cell-of-origin – CLL as a model disease

The aim of my thesis was to develop a method that allows for the identification of truly aberrant DNA methylation profiles in CLL on the background of the highly dynamic epigenetic landscape of normal B cell differentiation. Normal B cells exhibit distinct DNA methylation patterns reflecting their differentiation stage achieved along the B cell differentiation trajectory starting from naive B cells towards fully mature memory B cells. As a consequence, the DNA methylation landscape in any given CLL sample results from a combination of epigenetic changes acquired during normal B cell differentiation and those acquired during leukemogenesis. Previous efforts to uncover CLL-specific epigenetic alterations were based on the comparison of CLL epigenomes to peripheral blood $CD19^{+}$ B cells, which mostly reflect fully mature B cells. Such an approach completely neglects the massive epigenetic programming that occurs during normal B cell differentiation [8, 9]. Therefore, the dissection of cancer-specific epigenetic events requires the selection of a proper reference B cell epigenome: in this case the epigenome of the B cell differentiation stage that has acquired the transforming hit (cell-of-origin). Although several mutations creating a preleukemic clone including *SF3B1*, *NOTCH1* or *TP53* have been identified in the HSC pool of CLL patients, a second (and maybe third) driver event, which could be genetic or epigenetic, is required for the transformation [16]. This transforming event ultimately blocks normal B cell differentiation and/or resistance to apoptotic cell death.

When developing the cell-of-origin modeling method, I found that a simple linear regression model most precisely reflected the DNA methylation dynamics observed during normal B cell differentiation. Hence, I employed this linear model to describe the continuum of epigenetic alterations occurring during normal B cell development. For each CLL sample, the DNA methylome was subsequently precisely mapped on the normal B cell differentiation trajectory to define the closest virtual normal DNA methylome which was then defined its cell-of-origin. Using this cell-of-origin as an internal control, I then identified CLL-specific methylation events.

Taking into consideration that the vast majority of cancer DNA methylome data, i.e. The Cancer Genome Atlas (TCGA), or International Cancer Genome Consortium (ICGC) data is currently derived from Illumina 450K Bead Arrays I decided to develop the analysis pipeline for this type of data. The limited CpG-coverage of 450K arrays is, of course, a limitation of this model and could be overcome by the use of whole-genome bisulfite sequencing (WGBS) data once cost issues and computational challenges associated with the substantial increase in the number of CpG sites that are necessary to build the model (~44-fold change in the CpG coverage, from ~4,5*10^5 to ~20*10^6) have been addressed.

DNA methylation is considered as a stable epigenetic mark which allows the precise mapping of tissue types and even differentiation stage assignments in differentiating tissues such as the hematopoietic system. Hence, DNA methylation can be used to identify the founder cell of virtually any cancer entity. As shown in this thesis, tracing the cell-of-origin is of conceptual importance beyond the CLL field and, in my opinion, will serve as a model for similar studies in other cancer entities. The identification of the cell-of-origin of any cancer type will be crucial to dissect the mechanisms leading to tumor initiation and progression. In most cancers, so far, the precise cell-of-origin remains elusive. However, there are initial hints in the current literature indicating that the cell-of-origin influences tumor biology. For instance, studies in MLL-rearranged AML have shown that the cell-of-origin affects the phenotype and the clinical behavior of resulting leukemia [288]. Depending on the targeted cell type (cell-of-origin), hematopoietic stem cells (HSCs) or granulocyte-macrophage progenitors (GMPs), different epigenetic and transcriptional programs were orchestrated. Leukemias derived from HSCs were characterized by higher resistance to chemotherapy, had a stem-cell-like expression signature and had elevated global 5'-mC levels as compared to GMP-derived leukemias [288]. Likewise, mouse models of glioblastoma pointed towards neural stem cells and glial progenitors as glioma founder cells [289]. It was shown that several glioma subtypes share molecular signatures with different normal neural lineages, and vary in terms of their response to therapy, which suggested different cellular origins [289-291]. Similarly, using genetic lineage-tracing experiments in mice, Blanpain and colleagues nicely demonstrated the presence of distinct cells-of-origin for two types of skin cancer, good prognosis basal cell carcinoma, and invasive squamous cell carcinoma [164]. Likewise, the potential cell-of-origin in colorectal cancer was studied intensively, pointing towards three potential cell types as founder cells: intestinal stem cells [292-296], transit amplifying cells [292, 297], and differentiated villus cells [297]. Also, DNA methylation-based classification of central nervous system tumors revealed differences in cellular origin of these tumors. Gibson P *et al.* reported distinct cellular origins for two medulloblastoma subtypes [298]. The activated Sonic Hedgehog type (SHH type) most likely originates from the cells inside cerebellum, while WNT subtype arises from the cells of the dorsal brainstem, located outside cerebellum [298]. In a more global approach, Capper and colleagues used DNA

methylation to show distinct epigenome profiles in each of the examined 82 central nervous tumors [299], reflecting on the distinct origin of these tumors.

There might be caveats which have to be considered when implementing my approach to other tumor types. Unlike CLL, many solid tumors are very heterogeneous with respect to their methylation profiles [300]. Such intratumoral heterogeneity has been detected in spatially distinct regions of solid tumors, e.g. lung, prostate or hepatocellular carcinoma [201, 301, 302]. Additionally, mutations in epigenetic regulators can affect methylation profiles of tumors, creating genotype-specific methylation patterns. For instance, *DNMT3A* mutations in AML and other hematopoietic tumors induce characteristic patterns of hypomethylation [303, 304]. Likewise, the presence of *IDH1* mutation (R132) was linked to CpG island methylator phenotype in gliomas [305]. Along the same lines, mutations in genes encoding histones can influence DNA methylation profiles of the tumors, e.g. cell lines and primary samples harboring K27M mutation of histone H3.3 displayed global DNA hypomethylation [306-308]. Therefore, the performance of my method would have to be tested thoroughly in other tumors.

5.2 Hypomethylation is the most common CLL-specific event

Using the epigenome of the cell-of-origin as a control for aberrant DNA methylation calls, two categories of disease-specific methylation events were identified: 1) CpGs sites undergoing epigenetic programming during normal B cell differentiation, at which CLLs display inadequate levels of DNA methylation programming (‘Sites with epigenetic B cell programming‘) , and 2) CpG sites at which B cells are normally not programmed but CLL samples display aberrant DNA methylation patterns (‘Sites without epigenetic B cell programming‘). Applying my method to CLL, I found that disease-specific methylation events are actually rare. Only 1.6% of the CpG-sites (7,248 CpGs) represented on the 450K array are affected by disease-specific methylation programming in CLL. Moreover, the majority of the identified CLL-specific methylation events were characterized by hypomethylation (6,680 CpGs) as compared to the virtual cell-of-origin.

Remarkably, hypomethylation is a common feature for all of the B-cell malignancies [160, 309-313]. Also, it is commonly observed in EBV-transformed lymphoblastoid cell lines [314, 315] and in solid tumors [316, 317]. Interestingly, in B cell malignancies the degree of hypomethylation programming correlates with the differentiation stage achieved by the tumor-initiating B cell at the time of transformation. Acute lymphoblastic leukemias (ALLs) which are arrested somewhere at the pre-B cell stage show only minor hypomethylation. In contrast, Multiple Myelomas (MM) originating from plasma cells exhibit the most profound hypomethylation programming [160, 309-313]. Although global methylation loss occurs throughout the normal B cell differentiation process, this pattern is exaggerated in tumor cells.

DNA hypomethylation in CLL has been associated with genome-wide disordered methylation states enriched in gene-poor regions [318]. Previous studies linked genome-wide hypomethylation to elevated mutation rates and increased genomic instability [319]. Since global hypomethylation is a feature of both normal B cells and CLL samples, loss of methylation in B cell malignancies, in general, does not seem to contribute to genomic instability.

Yet, the question of the impact of this process on the pathogenesis of B cell malignancies remains unclear and has to be addressed in future studies.

Along the same lines, the mechanisms driving tumor-specific hypomethylation in B cell malignancies remain still elusive. While loss of DNA methylation at heterochromatic regions may occur through a passive, replication-dependent process, targeted hypomethylation at regulatory regions most likely requires an active demethylation process. Caron *et al.* intensively studied terminal differentiation of B cells into plasma B cells. This study linked the plasma cell differentiation process with an extensive DNA demethylation and with the local acquisition of 5-hydroxymethylcytosine at regulatory regions, pointing towards an active DNA demethylation [320]. Although still highly speculative, this event could require TF binding that results in the recruitment of TET family enzymes mediating active DNA demethylation. Additionally, one should also consider the possibility of active DNA demethylation guided by long noncoding RNAs (lncRNAs). It was shown that lncRNA, TARID, recruits DNA demethylation machinery to the promoter of *TCF21*, leading to increased *TCF21* expression [321]. Another scenario would involve active deamination of 5mC by the AICDA/APOBEC family of cytosine deaminases. Dominguez and colleagues have elegantly shown that the transit of differentiating B cells through the germinal center (GC) is associated with prominent locus-specific DNA methylation loss and increased methylation diversity. Both of these patterns were lost in $Aicda^{(-/-)}$ mice [322]. Moreover, AIDCA binding sites were enriched in hypomethylated loci in GC B cells [323]. Interestingly, AID overexpression in the mouse resulted in aggressive lymphomas, a phenotype that was associated with increased DNA methylation heterogeneity [324]. Altogether, this suggests the need for further studies on the demethylation mechanism in B cell malignancies.

5.3 Programming of CLL-specific methylation events

In this thesis, using chromatin states derived from an immortalized B cell cell line, I showed that disease-specific methylation events are mostly enriched for enhancer sequences (class A & B & C), weak and poised promoters and insulator regions (class E), suggesting the impact of aberrant DNA methylation on deregulated CLL transcriptome.

However, this profile is not unique to CLL *per se*. Recently, large-scale epigenomic studies on a broad range of tumor and tissue types have emphasized the role of DNA methylation in regulatory regions, particularly outside of promoters [325-327]. Enhancer activation, typically through DNA hypomethylation or H3K4me1/H3K27ac gain, was linked to specific gene expression profiles, and to other regulatory functions, e.g. response to intrinsic signaling or RNA splicing [328, 329]. Interestingly, aberrant DNA methylation programming at enhancer regions is a common feature for other B-cell malignancies, i.e. mantle cell lymphoma (MCL), acute lymphocytic leukemia (ALL), follicular lymphoma (FL), Burkitt lymphoma (BL), and multiple myeloma (MM) [309-312, 330]. In most instances, hypomethylation affects enhancer loci, with the exception of MM which is characterized by enhancer hypermethylation and decommissioning [310]. In MCL, *de novo* hypomethylation of the enhancer located ~650 kb downstream *SOX11* was associated with overexpression of the *SOX11* oncogene-mediated by *de novo* 3D chromatin interactions [312].

Normal B cell differentiation is associated with the coordinated expression of many B cell-specific transcription factors (TFs) [225]. It was shown that DNA hypomethylation is predominantly enriched at enhancer loci bound by these B cell-specific TFs in differentiating B cells [160, 331-333]. Tumor cells may employ aberrant TF programming to modify epigenetic states in order to establish a deregulated B cell differentiation program. ALL, for instance, is classified into different subtypes based on genetic fusions involving one or more TFs [334]. Remarkably, these genetically defined subtypes are characterized by distinct methylation profiles [335, 336]. Several TF fusions present in AML have also been linked to distinct DNA methylation patterns [337]. Together, these data demonstrate that the disruption of TF programs affects DNA methylation profiles.

In this thesis, I have found that CLL-specific methylation events are enriched for B-cell specific TF binding sites, including IKZF1, BATF (in class A), EBF1, IKZF1, BATF, ATF2 (in class C), NFATC1 and EGR1 (in class D). This B-cell specific TF signature may suggest an underlying defect in the normal B-cell differentiation program. For instance, activation of NFAT and EGR is known to occur downstream of B cell receptor signaling (via calcium influx and MAPK induction, respectively), linking the aberrant DNA methylation profiles to hyperactive BCR activity [338]. In my thesis statistically significantly deregulated TF expression patterns in CLL were observed only for EBF1. Interestingly, there are reports showing the pioneering activity of EBF1. It has been shown to induce the expression of a component of the B-cell receptor complex, Cd79a, facilitating DNA demethylation and chromatin remodeling at the Cd79a promoter region [339]. This pioneering activity is dependent on its C-terminal domain, which is indispensable for EBF1 to mediate chromatin accessibility and DNA demethylation in previously inactive chromatin regions [340]. The ability of EBF1 to demethylate DNA is most likely associated with TET2 recruitment as both of these factors were shown to interact in many tumor types, e.g. AML, low-grade glioma, cholangiocarcinoma and chondrosarcoma [341]. Despite the lack of drastic effects on the transcript levels of other TFs identified in my analysis, one should consider the scenario, in which slight expression changes below the sensitivity of the RNAseq assay are sufficient to mediate differential binding at some sites. Potentially subtle changes in the expression of co-factors may play a role, too. For instance, IKZF1, a member of IKAROS family is known to form heterodimers with other IKAROS zinc-finger proteins that results in the enhanced DNA affinity and transcriptional activation [342]. Likewise, an AP-1 transcription factor, ATF2, is known to form stable dimers with c-Jun [343, 344]. Another possibility would be that post-translational modifications of TFs might affect their protein expression levels and/or binding patterns. Phosphorylation of serine-43 in the DNA binding-domain of BATF, for instance, converts BATF from DNA binding inhibitor into a non-DNA binding form, which is able to form heterodimers but loses the ability to bind target DNA sequences [345]. ATF2, on the other hand, is phosphorylated at threonine-69 and threonine-71 and, following this event, it forms heterodimers with other AP-1 proteins that leads to transcriptional activation of target genes [346-348]. Similarly, phosphorylated form of EGR1 was shown to bind DNA more efficiently than the unphosphorylated protein [349].

Interestingly, I found enrichment of CTCF, RAD21 and SMC3 binding sites among CLL-specific gain of methylation events (class E). This explains the observed enrichment of class E CpGs in the insulator regions as CTCF and the cohesin complex composed of e.g. RAD21, SMC3, SMC1, and STAG1/STAG2 are working together to mediate long-range interactions [226]. Remarkably, during normal B cell differentiation insulators and CTCF binding sites are

protected from DNA hypermethylation, suggesting this as a strictly CLL-specific phenomenon. Insulator elements have two main functions [350, 351]. First, they are blocking the action of a distal enhancer on promoters of genes located within the insulator sequences. Second, they prevent the spread of nearby condensed chromatin that might otherwise silence the expression of genes within the insulator sequences. Some reports show that DNA hypermethylation abolishes CTCF binding and as a result deregulates locus insulation. Bell and Felsenfeld, for instance, have shown that CTCF binds to the imprinting control region (ICR) of Igf2 and H19 [352]. H19 is expressed only from the maternal allele, and Igf2 only from the paternal allele. Under normal circumstances, CTCF binds to the unmethylated maternal ICR, creating a locus insulation that prevents distant enhancers from the activation of Igf2 expression. However, methylation of CpGs within the CTCF-binding sites eliminates binding of CTCF *in vitro*, and results in loss of enhancer-blocking activity, thereby allowing Igf2 expression [352]. Likewise, Wang and colleagues have analyzed genome-wide occupancy of CTCF binding in 19 human cell types (normal primary cells and immortalized cell lines) showing that 41% of variable CTCF binding is linked to differential DNA methylation [353]. Disruption of CTCF binding was associated with increased methylation at CTCF binding sites [353]. Therefore, in CLL, DNA hypermethylation at class E sites most likely results in a block of CTCF binding and loss of locus insulation. This potentially leads to new enhancer-promoter interactions and therefore links loss of CTCF binding with aberrant transcriptional activity in CLL. With the current technologies, it would be tempting to perform either a deletion of CTCF binding sites within class E sites or to employ the CRISPR/dCas9 system for targeted DNA methylation (dCas9-DNMT3A) at class E loci [354]. The new, aberrant chromatin interactions could then be identified using chromosome conformation capture techniques [355], either from a global perspective with HiC-seq or from a CTCF viewpoint with CTCF ChIA-PET technology.

5.4 CLL-specific microRNAs

In the past, the majority of the identified candidate miRNAs were defined using the epigenome of $CD19^+$ B cells as controls, resulting in long lists of potentially epigenetically deregulated miRNAs [155, 189, 195, 198]. The results presented in my thesis show that this was likely an overestimation and that only a small proportion of miRNAs (miR-141, miR-195, miR-486 and miR-29c) can be associated with disease-specific DNA methylation events in CLL. To illustrate the difference in the proportion of CLL-specific microRNAs identified using the different approaches, I additionally performed the aberrant miRNA analysis on bulk $CD19^+$ B cells as controls. The results of this analysis show that virtually all previously reported 'CLL-specific' differences in miRNA promoter methylation are in fact related to methylation dynamics during normal B cell differentiation and thus may be used as CLL biomarkers but they do not contribute to our understanding of CLL leukemogenesis.

Previous studies reported overexpression of miR-29c and miR-195 in CLL. miR-29c regulates the expression of two important proto-oncogenes in CLL, namely *TCL1* and *MCL1* [235, 356-358]. Interestingly, so far, there are no reports on miR-486 or miR-141 neither in CLL nor in B cell biology, making them interesting candidates for future studies. Of note, epigenetic dysregulation has not been described for any of the above-mentioned microRNAs.

A potential functional impact of these microRNAs in tumorigenesis is suggested from studies in other cancer entities. miR-29c, for instance, was reported to have a tumor suppressor role in a wide range of malignancies, e.g. hepatocellular carcinoma, lung cancer, AML, and bladder and pancreatic cancer [359-363]. In hepatocellular carcinoma, miR-29c suppresses the oncogenic histone deacetylase, *SIRT1* [359]. Members of the miR-29 family (miR-29a, miR-29b, miR-29c) are directly targetting *DNMT3a* and *DNMT3b* in lung cancer and AML [360, 361]. In bladder cancer, miR-29c inhibits the proliferation and invasion of cancer cells via targeting *CDK6* [362, 363]. A broad functional spectrum was also demonstrated for miR-141. It was reported as a biomarker in non-small cell lung carcinoma [364, 365]. In colorectal cancer, miR-141 acts as an oncogene by targeting the tumor suppressor gene *DLC1* [366]. It was also shown that miR-141 is a major regulator of brain metastasis from breast cancer [367]. Its overexpression was linked to migratory and invasive properties in triple negative breast cancer through the activation of PI3K/AKT pathway [368]. Likewise, miR-486 overexpression in cervical cancer was associated with downregulation of *PTEN* that resulted in activation of PI3K/AKT pathway [369]. Similarly, upregulated miR-486 drives tumorigenesis in prostate cancer by direct targeting of components of PTEN/PI3K/AKT, FOXO, and TGF-b/Smad2 signaling pathways [370]. miR-195, on the other hand, targets *CCND3* and *BIRC5* (survivin) to inhibit the tumorigenesis of non-small cell lung cancer [371]. Similarly, it suppresses cyclin D1 and cyclin E1 in glioma cells [372].

In this thesis, I also showed that recurrently mutated epigenetic regulators in CLL: *ARID1A*, *CHD2* and *ASXL1* are among the validated targets of the newly identified disease-specific miRNAs. Of note, many other epigenetic regulators were predicted targets of CLL-specific microRNAs, e.g. *EZH2*, *DNMT3A*, *SIRT1*, *TET2*, *SETD1A* and *H3F3B*. These findings suggest epigenetic dysregulation of microRNAs as an alternative mechanism for the inactivation of these genes in CLL. The question that remains to be addressed is whether these CLL-specific miRNAs are exerting their function through feedback and/or feedforward loops. Tsang *et al.* and Martinez *et al.* have shown that miRNAs are overrepresented in gene regulatory networks, linking them with their regulatory functions [373, 374]. miR-486, for instance, is able to simultaneously disrupt multiple NF-κB negative feedback loops resulting in sustained NF-κB activity [375]. Similarly, in MM, miR-29 family was shown to be a part of negative feedback loop, in which c-myc inhibits the expression of the miR-29 family. miR-29 then suppresses DNMT expression, which in turn is responsible for the inhibition of miR-34 expression through promoter DNA hypermethylation [376].

5.5 CLL-specific protein-coding genes

Previous studies identified CLL-specific protein-coding genes using the epigenome of $CD19^{+}$ B cells as a reference, resulting in a long list of potentially epigenetically deregulated transcripts [142-153]. In my thesis I showed that this was a drastic overestimation (~30-fold) of epigenetically dysregulated target-genes and that only a small proportion of protein-coding genes (11 genes: *AMICA1, DOK2, LILRB4, MIB2, TNNC1, XIRP1, CLLU1, FBXO40, TMEM110-MUSTN1, C22orf46, EFCAB12*) can undoubtedly be associated with disease-specific DNA methylation events in CLL. Almost all reported 'CLL-specific' differences in gene promoters are in fact related to developmental dynamics occurring during normal B cell differentiation and thus do not contribute to disease biology.

Previous studies reported overexpression of *CLLU1* gene in CLLs [282]. Interestingly, this gene is neither expressed in other hematopoietic malignancies nor during normal hematopoiesis or in normal B cells during their terminal differentiation [282, 377]. This suggests a strictly CLL-specific mechanism of gene activation. *CLLU1* was also reported as a prognostic marker in CLL, predicting time to first treatment and overall survival [282]. DNA methylation in the promoter region was shown to correlate with decreased expression levels in M-CLL [144, 156]. Higher promoter DNA methylation was present in M-CLL but not in U-CLLs [156]. However, with the exception of its prognostic impact, nothing is known about its molecular function.

Another CLL-specific transcript, *DOK2*, was previously reported as being differentially methylated in AML and ovarian cancer [236, 248]. Lum and colleagues have shown that *DOK2* promoter hypermethylation results in decreased protein levels and in an impaired response to carboplatin in ovarian cancer [248]. Interestingly, *DOK2* was also linked to prognostic features of patients with gastric adenocarcinoma, ovarian cancer and AML [237, 247]. It has been also reported as a tumor suppressor gene in CMML and in ovarian carcinoma [248, 378]. Coppin *et al.* demonstrated that *DOK2* controls cell cycle regulation in hematopoietic stem cells [260]. Dok2 gene inactivation induced myeloproliferative disorders in aging mice. The precise function of *DOK2* in CLL has yet to be determined.

Among the identified epigenetically deregulated genes I also found the E3 ubiquitin ligase, *MIB2*, a component of the NF-κB pathway which is frequently activated in CLL [379]. Moreover, *MIB2* was reported as being indispensable for antiviral responses [276, 380].

Using my method, I found CLL-specifically deregulated transcripts with yet unknown functions, e.g. *FBXO40*, *TMEM110-MUSTN1*, *C22orf46*, *EFCAB12*, as well as genes for which there is only a little knowledge about their physiological or pathophysiological roles, i.e. *XIRP1*, *AMICA1* or *FBXO40*. Therefore, future studies are required to determine their function in the healthy organism and in the context of CLL.

Conclusions and Outlook

In this thesis, I described a new analytical approach to identify cancer-specific DNA methylation patterns using CLL as a model disease. This model illustrates that appropriate controls are crucial for any epigenome study. Using my newly developed method, I was able to identify truly CLL-specific methylation changes by considering DNA methylation dynamics present in the context of normal B cell differentiation. This involved the definition of a virtual cell-of-origin DNA methylome and using it as a reference methylome to call aberrant DNA methylation in CLL samples. I defined and dissected the two main factors that together form the epigenetic landscape of CLL: 1) DNA methylation patterns derived from the tumor-initiating B cell at a specific differentiation stage and 2) epigenetic alterations that are acquired in CLL cells during leukemogenesis. Applying this method, I showed that disease-specific methylation events are rare. Only 1.6% of the CpGs from the 450K array are affected by CLL-specific methylation programming. Therefore, all previous comparisons of CLLs to bulk $CD19^{+}$ B cells should be interpreted very cautiously as the vast majority of aberrant DNA methylation events identified in CLL using this

classical approach are actually occurring during normal B cell differentiation and, thus, are not related to disease biology.

Although the mechanisms driving CLL-specific DNA methylation programming are still elusive, my study provides evidence for a role of deregulated B cell-related signaling pathways and, at the same time, implicates novel mechanisms (e.g. deregulated CTCF binding) in CLL pathogenesis. Further, I demonstrated the pathogenic impact of CLL-specific methylation events by linking them to transcriptional changes observed in CLL (miRNAs and protein-coding genes). The identification of CLL-specific, epigenetically deregulated miRNAs and protein-coding genes demonstrated that previous conventional analysis approaches overestimated the proportion of epigenetically deregulated genes in CLL and suggested a need for using a proper control B cell for identification of truly disease-specific methylation events. Only 4 miRNAs and 11 protein-coding genes were identified as CLL-specific. Furthermore, a potential alternative mechanism for silencing of epigenetic regulators through miRNAs in CLL pathogenesis was identified.

One open question that remains to be addressed in future studies is that of the mechanism underlying the aberrant epigenome programming in CLL. First of all, the involvement of TFs into programming of CLL-specific methylation patterns (aberrant TF programming) has to be explored in more details. My TF binding sites enrichment analysis used a simplified approach using ChIP-seq data derived from an immortalized B cell line, GM12878. In an ideal setting, one would have to investigate TF programming during normal B cell differentiation and in primary CLL samples, and to perform a similar integrated analysis as presented in this thesis. By doing so, one could potentially infer the dynamics of TF binding during normal B cell differentiation and *de novo* TF binding present in CLL. To achieve this goal, it would be necessary to sort different B cell types spanning the entire differentiation axis and to perform ChIP-seq experiments on various TFs, e.g. BATF, EBF1, IKZF1, EGR1, EGR2, NFATC1, CTCF etc. The same experiment would have to be performed in a number of CLL samples from different maturation stages. These ChIP-seq data could then be integrated with the CLL-specific methylome data to define the gains/losses of TF binding that are crucial for disease pathogenesis.

Another open question refers to the effects of DNA hypermethylation at class E CLL-specific sites. The hypothesis would be that DNA hypermethylation at class E sites results in a block of CTCF binding and in a subsequent loss of locus insulation. This would further lead to new enhancer-promoter interactions and, hence, it might link loss of CTCF binding with aberrant CLL transcriptomes. With the current technologies, it would be tempting to perform either a deletion of CTCF binding sites within class E sites or to employ the CRISPR/dCas9 system for targeted DNA methylation (dCas9-DNMT3A) at class E loci [354]. New, aberrant chromatin interactions could then be identified using chromosome conformation capture techniques [355], either from a global perspective using HiC-seq or from a CTCF viewpoint using the ChIA-PET assay. Next, to link aberrant CTCF binding to the CLL transcriptome, the intensity of these chromatin interactions should be correlated with the expression levels of genes within the interaction loops.

Furthermore, my study should be performed in a truly genome-wide manner using WGBS data from normal B cells and CLL samples. Based on these data, a similar data analysis approach as demonstrated in this thesis could be used.

6. REFERENCES

1. Smith, A., et al., *Incidence of haematological malignancy by sub-type: a report from the Haematological Malignancy Research Network.* Br J Cancer, 2011. **105**(11): p. 1684-92.
2. Mansouri, L., et al., *Epigenetic deregulation in chronic lymphocytic leukemia: Clinical and biological impact.* Semin Cancer Biol, 2018.
3. Hallek, M., et al., *Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines.* Blood, 2008. **111**(12): p. 5446-56.
4. Hallek, M., *Chronic lymphocytic leukemia: 2017 update on diagnosis, risk stratification, and treatment.* Am J Hematol, 2017. **92**(9): p. 946-965.
5. Rai, K.R., et al., *Clinical staging of chronic lymphocytic leukemia.* Blood, 1975. **46**(2): p. 219-34.
6. Binet, J.L., et al., *A clinical staging system for chronic lymphocytic leukemia: prognostic significance.* Cancer, 1977. **40**(2): p. 855-64.
7. Baliakas, P., et al., *Prognostic indices in chronic lymphocytic leukaemia: where do we stand how do we proceed?* J Intern Med, 2016. **279**(4): p. 347-57.
8. Stilgenbauer, S., *Prognostic markers and standard management of chronic lymphocytic leukemia.* Hematology Am Soc Hematol Educ Program, 2015. **2015**: p. 368-77.
9. Puente, X.S., et al., *Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia.* Nature, 2011. **475**(7354): p. 101-5.
10. Fabbri, G., et al., *Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation.* J Exp Med, 2011. **208**(7): p. 1389-401.
11. Wang, L., et al., *SF3B1 and other novel cancer genes in chronic lymphocytic leukemia.* N Engl J Med, 2011. **365**(26): p. 2497-506.
12. Quesada, V., et al., *Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia.* Nat Genet, 2011. **44**(1): p. 47-52.
13. Rossi, D., et al., *Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia.* Blood, 2012. **119**(12): p. 2854-62.
14. Mansouri, L., et al., *Functional loss of IkappaBepsilon leads to NF-kappaB deregulation in aggressive chronic lymphocytic leukemia.* J Exp Med, 2015. **212**(6): p. 833-43.
15. Ljungstrom, V., et al., *Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent RPS15 mutations.* Blood, 2016. **127**(8): p. 1007-16.
16. Damm, F., et al., *Acquired initiating mutations in early hematopoietic cells of CLL patients.* Cancer Discov, 2014. **4**(9): p. 1088-101.
17. Young, E., et al., *EGR2 mutations define a new clinically aggressive subgroup of chronic lymphocytic leukemia.* Leukemia, 2017. **31**(7): p. 1547-1554.
18. Damle, R.N., et al., *Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia.* Blood, 1999. **94**(6): p. 1840-7.
19. Oscier, D.G., et al., *Multivariate analysis of prognostic factors in CLL: clinical stage, IGVH gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors.* Blood, 2002. **100**(4): p. 1177-84.
20. Hamblin, T.J., et al., *Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia.* Blood, 1999. **94**(6): p. 1848-54.
21. Ghia, P., et al., *ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia.* Leukemia, 2007. **21**(1): p. 1-3.
22. Pospisilova, S., et al., *ERIC recommendations on TP53 mutation analysis in chronic lymphocytic leukemia.* Leukemia, 2012. **26**(7): p. 1458-61.
23. Dohner, H., et al., *Genomic aberrations and survival in chronic lymphocytic leukemia.* N Engl J Med, 2000. **343**(26): p. 1910-6.

24. Hallek, M., et al., *Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial.* Lancet, 2010. **376**(9747): p. 1164-74.
25. Fischer, K., et al., *Long-term remissions after FCR chemoimmunotherapy in previously untreated patients with CLL: updated results of the CLL8 trial.* Blood, 2016. **127**(2): p. 208-15.
26. Gunnarsson, R., et al., *Large but not small copy-number alterations correlate to high-risk genomic aberrations and survival in chronic lymphocytic leukemia: a high-resolution genomic screening of newly diagnosed patients.* Leukemia, 2010. **24**(1): p. 211-5.
27. Silvennoinen, R., et al., *Pharmacokinetics of chlorambucil in patients with chronic lymphocytic leukaemia: comparison of different days, cycles and doses.* Pharmacol Toxicol, 2000. **87**(5): p. 223-8.
28. Plunkett, W., et al., *Fludarabine: pharmacokinetics, mechanisms of action, and rationales for combination therapies.* Semin Oncol, 1993. **20**(5 Suppl 7): p. 2-12.
29. Rai, K.R., et al., *Fludarabine compared with chlorambucil as primary therapy for chronic lymphocytic leukemia.* N Engl J Med, 2000. **343**(24): p. 1750-7.
30. Steurer, M., et al., *Purine antagonists for chronic lymphocytic leukaemia.* Cochrane Database Syst Rev, 2006(3): p. Cd004270.
31. Eichhorst, B., et al., *Chronic lymphocytic leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.* Ann Oncol, 2015. **26 Suppl 5**: p. v78-84.
32. Dighiero, G., et al., *Chlorambucil in indolent chronic lymphocytic leukemia. French Cooperative Group on Chronic Lymphocytic Leukemia.* N Engl J Med, 1998. **338**(21): p. 1506-14.
33. Malik, A., et al., *Azacitidine in fludarabine-refractory chronic lymphocytic leukemia: a phase II study.* Clin Lymphoma Myeloma Leuk, 2013. **13**(3): p. 292-5.
34. Blum, K.A., et al., *Phase I trial of low dose decitabine targeting DNA hypermethylation in patients with chronic lymphocytic leukaemia and non-Hodgkin lymphoma: dose-limiting myelosuppression without evidence of DNA hypomethylation.* Br J Haematol, 2010. **150**(2): p. 189-95.
35. Blum, K.A., et al., *Phase II study of the histone deacetylase inhibitor MGCD0103 in patients with previously treated chronic lymphocytic leukaemia.* Br J Haematol, 2009. **147**(4): p. 507-14.
36. El-Khoury, V., et al., *The histone deacetylase inhibitor MGCD0103 induces apoptosis in B-cell chronic lymphocytic leukemia cells through a mitochondria-mediated caspase activation cascade.* Mol Cancer Ther, 2010. **9**(5): p. 1349-60.
37. Fournel, M., et al., *MGCD0103, a novel isotype-selective histone deacetylase inhibitor, has broad spectrum antitumor activity in vitro and in vivo.* Mol Cancer Ther, 2008. **7**(4): p. 759-68.
38. Byrd, J.C., et al., *Depsipeptide (FR901228): a novel therapeutic agent with selective, in vitro activity against human B-cell chronic lymphocytic leukemia cells.* Blood, 1999. **94**(4): p. 1401-8.
39. Aron, J.L., et al., *Depsipeptide (FR901228) induces histone acetylation and inhibition of histone deacetylase in chronic lymphocytic leukemia cells concurrent with activation of caspase 8-mediated apoptosis and down-regulation of c-FLIP protein.* Blood, 2003. **102**(2): p. 652-8.
40. Fernandez Calotti, P., et al., *Modulation of the human equilibrative nucleoside transporter1 (hENT1) activity by IL-4 and PMA in B cells from chronic lymphocytic leukemia.* Biochem Pharmacol, 2008. **75**(4): p. 857-65.
41. Kulis, M., et al., *Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia.* Nat Genet, 2012. **44**(11): p. 1236-42.
42. Oakes, C.C., et al., *DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia.* Nat Genet, 2016. **48**(3): p. 253-64.
43. Zack, T.I., et al., *Pan-cancer patterns of somatic copy-number alteration.* Nature genetics, 2013. **45**(10): p. 1134-1140.
44. Malek, S.N., *The biology and clinical significance of acquired genomic copy number aberrations and recurrent gene mutations in chronic lymphocytic leukemia.* Oncogene, 2013. **32**(23): p. 2805-17.
45. Gaidano, G., R. Foa, and R. Dalla-Favera, *Molecular pathogenesis of chronic lymphocytic leukemia.* J Clin Invest, 2012. **122**(10): p. 3432-8.

46. Rossi, D. and G. Gaidano, *ATM and chronic lymphocytic leukemia: mutations, and not only deletions, matter.* Haematologica, 2012. **97**(1): p. 5-8.
47. Schaffner, C., et al., *Somatic ATM mutations indicate a pathogenic role of ATM in B-cell chronic lymphocytic leukemia.* Blood, 1999. **94**(2): p. 748-53.
48. Stankovic, T., et al., *Inactivation of ataxia telangiectasia mutated gene in B-cell chronic lymphocytic leukaemia.* Lancet, 1999. **353**(9146): p. 26-9.
49. Guarini, A., et al., *ATM gene alterations in chronic lymphocytic leukemia patients induce a distinct gene expression profile and predict disease progression.* Haematologica, 2012. **97**(1): p. 47-55.
50. Chiorazzi, N., K.R. Rai, and M. Ferrarini, *Chronic lymphocytic leukemia.* N Engl J Med, 2005. **352**(8): p. 804-15.
51. Caporaso, N., et al., *Chronic lymphocytic leukaemia genetics overview.* Br J Haematol, 2007. **139**(5): p. 630-4.
52. Pasqualucci, L., et al., *Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas.* Nature, 2001. **412**(6844): p. 341-6.
53. Klein, U. and R. Dalla-Favera, *Germinal centres: role in B-cell physiology and malignancy.* Nat Rev Immunol, 2008. **8**(1): p. 22-33.
54. Puente, X.S., et al., *Non-coding recurrent mutations in chronic lymphocytic leukaemia.* Nature, 2015. **526**(7574): p. 519-24.
55. Kandaswamy, R., et al., *Genetic Predisposition to Chronic Lymphocytic Leukemia Is Mediated by a BMF Super-Enhancer Polymorphism.* Cell Rep, 2016. **16**(8): p. 2061-2067.
56. Burns, A., et al., *Whole-genome sequencing of chronic lymphocytic leukaemia reveals distinct differences in the mutational landscape between IgHVmut and IgHVunmut subgroups.* Leukemia, 2017. Nov 21. doi: 10.1038/leu.2017.311. [Epub ahead of print].
57. Waddington, C.H., *The epigenotype. 1942.* Int J Epidemiol, 2012. **41**(1): p. 10-3.
58. Waddington, C.H., *Towards a theoretical biology.* Nature, 1968. **218**(5141): p. 525-7.
59. Wu, C. and J.R. Morris, *Genes, genetics, and epigenetics: a correspondence.* Science, 2001. **293**(5532): p. 1103-5.
60. Ito, S., et al., *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.* Science, 2011. **333**(6047): p. 1300-3.
61. Breiling, A. and F. Lyko, *Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond.* Epigenetics & Chromatin, 2015. **8**: p. 24.
62. Richa, R. and R.P. Sinha, *Hydroxymethylation of DNA: an epigenetic marker.* EXCLI Journal, 2014. **13**: p. 592-610.
63. Costa, F.F., *Non-coding RNAs: lost in translation?* Gene, 2007. **386**(1-2): p. 1-10.
64. Costa, F.F., *Non-coding RNAs, epigenetics and complexity.* Gene, 2008. **410**(1): p. 9-17.
65. Lawrence, M., S. Daujat, and R. Schneider, *Lateral Thinking: How Histone Modifications Regulate Gene Expression.* Trends Genet, 2016. **32**(1): p. 42-56.
66. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Research, 2011. **21**(3): p. 381-395.
67. Gaffney, D.J., et al., *Controls of nucleosome positioning in the human genome.* PLoS Genet, 2012. **8**(11): p. e1003036.
68. Matharu, N. and N. Ahituv, *Minor Loops in Major Folds: Enhancer–Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease.* PLoS Genetics, 2015. **11**(12): p. e1005640.
69. Long, H.K., S.L. Prescott, and J. Wysocka, *Ever-changing landscapes: transcriptional enhancers in development and evolution.* Cell, 2016. **167**(5): p. 1170-1187.
70. Chen, Z., et al., *Epigenetic Regulation: A New Frontier for Biomedical Engineers.* Annu Rev Biomed Eng, 2017. **19**: p. 195-219.
71. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals.* Nature Genetics, 2003. **33**: p. 245.
72. Holliday, R. and J.E. Pugh, *DNA modification mechanisms and gene activity during development.* Science, 1975. **187**(4173): p. 226-32.

73. Compere, S.J. and R.D. Palmiter, *DNA methylation controls the inducibility of the mouse metallothionein-I gene lymphoid cells.* Cell, 1981. **25**(1): p. 233-40.
74. Santos, F., et al., *Dynamic reprogramming of DNA methylation in the early mouse embryo.* Dev Biol, 2002. **241**(1): p. 172-82.
75. Monk, M., M. Boubelik, and S. Lehnert, *Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development.* Development, 1987. **99**(3): p. 371-82.
76. Li, E., C. Beard, and R. Jaenisch, *Role for DNA methylation in genomic imprinting.* Nature, 1993. **366**(6453): p. 362-5.
77. Mohandas, T., R.S. Sparkes, and L.J. Shapiro, *Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation.* Science, 1981. **211**(4480): p. 393-6.
78. Anastasiadou, C., et al., *Human epigenome data reveal increased CpG methylation in alternatively spliced sites and putative exonic splicing enhancers.* DNA Cell Biol, 2011. **30**(5): p. 267-75.
79. Zhou, Y., Y. Lu, and W. Tian, *Epigenetic features are significantly associated with alternative splicing.* BMC Genomics, 2012. **13**: p. 123.
80. Walsh, C.P., J.R. Chaillet, and T.H. Bestor, *Transcription of IAP endogenous retroviruses is constrained by cytosine methylation.* Nat Genet, 1998. **20**(2): p. 116-7.
81. Bourc'his, D. and T.H. Bestor, *Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L.* Nature, 2004. **431**(7004): p. 96-9.
82. Dodge, J.E., et al., *Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization.* J Biol Chem, 2005. **280**(18): p. 17986-91.
83. Goll, M.G. and T.H. Bestor, *EUKARYOTIC CYTOSINE METHYLTRANSFERASES.* Annual Review of Biochemistry, 2005. **74**(1): p. 481-514.
84. Tahiliani, M., et al., *Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1.* Science (New York, N.Y.), 2009. **324**(5929): p. 930-935.
85. He, Y.-F., et al., *Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA.* Science (New York, N.Y.), 2011. **333**(6047): p. 1303-1307.
86. Lister, R., et al., *Global epigenomic reconfiguration during mammalian brain development.* Science, 2013. **341**(6146): p. 1237905.
87. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-22.
88. Ramsahoye, B.H., et al., *Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a.* Proceedings of the National Academy of Sciences, 2000. **97**(10): p. 5237.
89. Tweedie, S., et al., *Methylation of genomes and genes at the invertebrate-vertebrate boundary.* Mol Cell Biol, 1997. **17**(3): p. 1469-75.
90. Bird, A.P., M.H. Taggart, and B.A. Smith, *Methylated and unmethylated DNA compartments in the sea urchin genome.* Cell, 1979. **17**(4): p. 889-901.
91. Gardiner-Garden, M. and M. Frommer, *CpG islands in vertebrate genomes.* J Mol Biol, 1987. **196**(2): p. 261-82.
92. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
93. Edwards, Y.H., *CpG islands in genes showing tissue-specific expression.* Philos Trans R Soc Lond B Biol Sci, 1990. **326**(1235): p. 207-15.
94. Larsen, F., et al., *CpG islands as gene markers in the human genome.* Genomics, 1992. **13**(4): p. 1095-107.
95. Liu, H., et al., *Quantitative epigenetic co-variation in CpG islands and co-regulation of developmental genes.* Sci Rep, 2013. **3**: p. 2576.
96. Illingworth, R.S., et al., *Orphan CpG islands identify numerous conserved promoters in the mammalian genome.* PLoS Genet, 2010. **6**(9): p. e1001134.

97. Aran, D., et al., *Replication timing-related and gene body-specific methylation of active human genes.* Hum Mol Genet, 2011. **20**(4): p. 670-80.
98. Suzuki, M., et al., *Late-replicating heterochromatin is characterized by decreased cytosine methylation in the human genome.* Genome Res, 2011. **21**(11): p. 1833-40.
99. Zhang, J., et al., *Establishment of transcriptional competence in early and late S phase.* Nature, 2002. **420**(6912): p. 198-202.
100. Siegfried, Z., et al., *DNA methylation represses transcription in vivo.* Nat Genet, 1999. **22**(2): p. 203-6.
101. Stadler, M.B., et al., *DNA-binding factors shape the mouse methylome at distal regulatory regions.* Nature, 2011. **480**(7378): p. 490-5.
102. Macleod, D., et al., *Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island.* Genes Dev, 1994. **8**(19): p. 2282-92.
103. Brandeis, M., et al., *Sp1 elements protect a CpG island from de novo methylation.* Nature, 1994. **371**(6496): p. 435-8.
104. Wiench, M., et al., *DNA methylation status predicts cell type-specific enhancer activity.* Embo j, 2011. **30**(15): p. 3028-39.
105. Quenneville, S., et al., *In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions.* Mol Cell, 2011. **44**(3): p. 361-72.
106. Prokhortchouk, A., et al., *The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor.* Genes Dev, 2001. **15**(13): p. 1613-8.
107. Filion, G.J., et al., *A family of human zinc finger proteins that bind methylated DNA and repress transcription.* Mol Cell Biol, 2006. **26**(1): p. 169-81.
108. Klose, R.J. and A.P. Bird, *Genomic DNA methylation: the mark and its mediators.* Trends Biochem Sci, 2006. **31**(2): p. 89-97.
109. Hahn, M.A., et al., *Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks.* PLoS One, 2011. **6**(4): p. e18844.
110. Ball, M.P., et al., *Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells.* Nat Biotechnol, 2009. **27**(4): p. 361-8.
111. Varley, K.E., et al., *Dynamic DNA methylation across diverse human cell lines and tissues.* Genome Res, 2013. **23**(3): p. 555-67.
112. Schmidl, C., et al., *Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity.* Genome Res, 2009. **19**(7): p. 1165-74.
113. Saito, Y., et al., *Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells.* Cancer Cell, 2006. **9**(6): p. 435-43.
114. Cheung, H.H., et al., *Methylation of an intronic region regulates miR-199a in testicular tumor malignancy.* Oncogene, 2011. **30**(31): p. 3404-15.
115. Lyle, R., et al., *The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1.* Nat Genet, 2000. **25**(1): p. 19-21.
116. Stoger, R., et al., *Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal.* Cell, 1993. **73**(1): p. 61-71.
117. Wu, W., et al., *Hypomethylation of noncoding DNA regions and overexpression of the long noncoding RNA, AFAP1-AS1, in Barrett's esophagus and esophageal adenocarcinoma.* Gastroenterology, 2013. **144**(5): p. 956-966.e4.
118. Maunakea, A.K., et al., *Conserved role of intragenic DNA methylation in regulating alternative promoters.* Nature, 2010. **466**(7303): p. 253-7.
119. Archey, W.B., et al., *Methylation of CpGs as a determinant of transcriptional activation at alternative promoters for transforming growth factor-beta3.* Cancer Res, 1999. **59**(10): p. 2292-6.
120. Shmelkov, S.V., et al., *Alternative promoters regulate transcription of the gene that encodes stem cell surface protein AC133.* Blood, 2004. **103**(6): p. 2055-61.
121. Cheong, J., et al., *Diverse DNA methylation statuses at alternative promoters of human genes in various tissues.* DNA Res, 2006. **13**(4): p. 155-67.

122. Malousi, A., N. Maglaveras, and S. Kouidou, *Intronic CpG content and alternative splicing in human genes containing a single cassette exon.* Epigenetics, 2008. **3**(2): p. 69-73.
123. Shukla, S., et al., *CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.* Nature, 2011. **479**(7371): p. 74-9.
124. Cowley, M., et al., *Epigenetic control of alternative mRNA processing at the imprinted Herc3/Nap1l5 locus.* Nucleic Acids Res, 2012. **40**(18): p. 8917-26.
125. Hon, G.C., et al., *Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer.* Genome Res, 2012. **22**(2): p. 246-58.
126. Rakyan, V.K., et al., *An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs).* Genome Res, 2008. **18**(9): p. 1518-29.
127. Rauch, T.A., et al., *A human B cell methylome at 100-base pair resolution.* Proc Natl Acad Sci U S A, 2009. **106**(3): p. 671-8.
128. Esteller, M., *Epigenetics in cancer.* N Engl J Med, 2008. **358**(11): p. 1148-59.
129. Robertson, K.D., *DNA methylation and human disease.* Nat Rev Genet, 2005. **6**(8): p. 597-610.
130. Feinberg, A.P. and B. Vogelstein, *Hypomethylation distinguishes genes of some human cancers from their normal counterparts.* Nature, 1983. **301**(5895): p. 89-92.
131. Gama-Sosa, M.A., et al., *The 5-methylcytosine content of DNA from human tumors.* Nucleic Acids Res, 1983. **11**(19): p. 6883-94.
132. Hovestadt, V., et al., *Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing.* Nature, 2014. **510**(7506): p. 537-41.
133. Hansen, K.D., et al., *Increased methylation variation in epigenetic domains across cancer types.* Nat Genet, 2011. **43**(8): p. 768-75.
134. Nakamura, N. and K. Takenaga, *Hypomethylation of the metastasis-associated S100A4 gene correlates with gene activation in human colon adenocarcinoma cell lines.* Clin Exp Metastasis, 1998. **16**(5): p. 471-9.
135. Gupta, A., et al., *Hypomethylation of the synuclein gamma gene CpG island promotes its aberrant expression in breast carcinoma and ovarian carcinoma.* Cancer Res, 2003. **63**(3): p. 664-73.
136. Strichman-Almashanu, L.Z., et al., *A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes.* Genome Res, 2002. **12**(4): p. 543-54.
137. Feinberg, A.P. and B. Tycko, *The history of cancer epigenetics.* Nat Rev Cancer, 2004. **4**(2): p. 143-53.
138. Gronbaek, K., C. Hother, and P.A. Jones, *Epigenetic changes in cancer.* Apmis, 2007. **115**(10): p. 1039-59.
139. Valo, S., et al., *DNA hypermethylation appears early and shows increased frequency with dysplasia in Lynch syndrome-associated colorectal adenomas and carcinomas.* Clinical Epigenetics, 2015. **7**(1): p. 71.
140. Chan, A.O., et al., *CpG island methylation in aberrant crypt foci of the colorectum.* Am J Pathol, 2002. **160**(5): p. 1823-30.
141. Lipsanen, V., et al., *Hypomethylation of ornithine decarboxylase gene and erb-A1 oncogene in human chronic lymphatic leukemia.* Blood, 1988. **72**(6): p. 2042-4.
142. Yuille, M.R., et al., *TCL1 is activated by chromosomal rearrangement or by hypomethylation.* Genes Chromosomes Cancer, 2001. **30**(4): p. 336-41.
143. Bichi, R., et al., *Human chronic lymphocytic leukemia modeled in mouse by targeted TCL1 expression.* Proc Natl Acad Sci U S A, 2002. **99**(10): p. 6955-60.
144. Cahill, N. and R. Rosenquist, *Uncovering the DNA methylome in chronic lymphocytic leukemia.* Epigenetics, 2013. **8**(2): p. 138-48.
145. Melki, J.R., et al., *Hypermethylation of E-cadherin in leukemia.* Blood, 2000. **95**(10): p. 3208-13.
146. Bechter, O.E., et al., *CpG island methylation of the hTERT promoter is associated with lower telomerase activity in B-cell lymphocytic leukemia.* Exp Hematol, 2002. **30**(1): p. 26-33.
147. Raval, A., et al., *TWIST2 demonstrates differential methylation in immunoglobulin variable heavy chain mutated and unmutated chronic lymphocytic leukemia.* J Clin Oncol, 2005. **23**(17): p. 3877-85.

148. Corcoran, M., et al., *ZAP-70 methylation status is associated with ZAP-70 expression status in chronic lymphocytic leukemia.* Haematologica, 2005. **90**(8): p. 1078-88.
149. Claus, R., et al., *Quantitative DNA methylation analysis identifies a single CpG dinucleotide important for ZAP-70 expression and predictive of prognosis in chronic lymphocytic leukemia.* J Clin Oncol, 2012. **30**(20): p. 2483-91.
150. Chantepie, S.P., et al., *ZAP-70 intron1 DNA methylation status: determination by pyrosequencing in B chronic lymphocytic leukemia.* Leuk Res, 2010. **34**(6): p. 800-8.
151. Wiestner, A., et al., *ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile.* Blood, 2003. **101**(12): p. 4944-51.
152. Strathdee, G., et al., *Promoter hypermethylation silences expression of the HoxA4 gene and correlates with IgVh mutational status in CLL.* Leukemia, 2006. **20**(7): p. 1326-9.
153. Raval, A., et al., *Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia.* Cell, 2007. **129**(5): p. 879-90.
154. Kanduri, M., et al., *Differential genome-wide array-based methylation profiles in prognostic subsets of chronic lymphocytic leukemia.* Blood, 2010. **115**(2): p. 296-305.
155. Baer, C., et al., *Extensive promoter DNA hypermethylation and hypomethylation is associated with aberrant microRNA expression in chronic lymphocytic leukemia.* Cancer Res, 2012. **72**(15): p. 3775-85.
156. Cahill, N., et al., *450K-array analysis of chronic lymphocytic leukemia cells reveals global DNA methylation to be relatively stable over time and similar in resting and proliferative compartments.* Leukemia, 2013. **27**(1): p. 150-8.
157. Oakes, C.C., et al., *Evolution of DNA methylation is linked to genetic aberrations in chronic lymphocytic leukemia.* Cancer Discov, 2014. **4**(3): p. 348-61.
158. Landau, D.A., et al., *Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia.* Cancer Cell, 2014. **26**(6): p. 813-825.
159. Smith, E.N., et al., *Genetic and epigenetic profiling of CLL disease progression reveals limited somatic evolution and suggests a relationship to memory-cell development.* Blood Cancer J, 2015. **5**: p. e303.
160. Kulis, M., et al., *Whole-genome fingerprint of the DNA methylome during human B cell differentiation.* Nat Genet, 2015. **47**(7): p. 746-56.
161. Queiros, A.C., et al., *A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact.* Leukemia, 2015. **29**(3): p. 598-605.
162. Beekman, R., et al., *The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia.* Nature Medicine, 2018. **24**(6): p. 868-880.
163. Wahlfors, J., et al., *Genomic hypomethylation in human chronic lymphocytic leukemia.* Blood, 1992. **80**(8): p. 2074-80.
164. Blanpain, C., *Tracing the cellular origin of cancer.* Nat Cell Biol, 2013. **15**(2): p. 126-34.
165. Fabris, S., et al., *Biological and clinical relevance of quantitative global methylation of repetitive DNA sequences in chronic lymphocytic leukemia.* Epigenetics, 2011. **6**(2): p. 188-94.
166. Stach, D., et al., *Capillary electrophoretic analysis of genomic DNA methylation levels.* Nucleic Acids Res, 2003. **31**(2): p. E2.
167. Klein, U., K. Rajewsky, and R. Kuppers, *Human immunoglobulin (Ig)M+IgD+ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells.* J Exp Med, 1998. **188**(9): p. 1679-89.
168. Lee, R.C., R.L. Feinbaum, and V. Ambros, *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.* Cell, 1993. **75**(5): p. 843-54.
169. Reinhart, B.J., et al., *The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.* Nature, 2000. **403**(6772): p. 901-6.
170. Griffiths-Jones, S., et al., *miRBase: tools for microRNA genomics.* Nucleic Acids Research, 2008. **36**(Database issue): p. D154-D158.

171. Pillai, R.S., S.N. Bhattacharyya, and W. Filipowicz, *Repression of protein synthesis by miRNAs: how many mechanisms?* Trends Cell Biol, 2007. **17**(3): p. 118-26.
172. Wilczynska, A. and M. Bushell, *The complexity of miRNA-mediated repression.* Cell Death and Differentiation, 2015. **22**(1): p. 22-33.
173. Friedman, R.C., et al., *Most mammalian mRNAs are conserved targets of microRNAs.* Genome Res, 2009. **19**(1): p. 92-105.
174. Kim, V.N., *MicroRNA biogenesis: coordinated cropping and dicing.* Nat Rev Mol Cell Biol, 2005. **6**(5): p. 376-85.
175. Winter, J., et al., *Many roads to maturity: microRNA biogenesis pathways and their regulation.* Nat Cell Biol, 2009. **11**(3): p. 228-34.
176. Gulyaeva, L.F. and N.E. Kushlinskiy, *Regulatory mechanisms of microRNA expression.* J Transl Med, 2016. **14**(1): p. 143.
177. Tufekci, K.U., R.L. Meuwissen, and S. Genc, *The role of microRNAs in biological processes.* Methods Mol Biol, 2014. **1107**: p. 15-31.
178. Hawkes, J.E., et al., *microRNAs in Psoriasis.* J Invest Dermatol, 2016. **136**(2): p. 365-71.
179. Hu, Y.B., et al., *Diagnostic Value of microRNA for Alzheimer's Disease: A Systematic Review and Meta-Analysis.* Front Aging Neurosci, 2016. **8**: p. 13.
180. O'Reilly, S., *MicroRNAs in fibrosis: opportunities and challenges.* Arthritis Res Ther, 2016. **18**: p. 11.
181. Wang, N., et al., *Role of microRNAs in cardiac hypertrophy and heart failure.* IUBMB Life, 2009. **61**(6): p. 566-71.
182. Dugo, M., et al., *MicroRNA co-expression patterns unravel the relevance of extra cellular matrix and immunity in breast cancer.* Breast, 2018. **39**: p. 46-52.
183. Giordano, M., et al., *Differential microRNA expression profiles between young and old lung adenocarcinoma patients.* Am J Transl Res, 2018. **10**(3): p. 892-900.
184. Pogribny, I.P., F.A. Beland, and I. Rusyn, *The role of microRNAs in the development and progression of chemical-associated cancers.* Toxicol Appl Pharmacol, 2016. **312**: p. 3-10.
185. Song, C., et al., *The network of microRNAs, transcription factors, target genes and host genes in human renal cell carcinoma.* Oncol Lett, 2015. **9**(1): p. 498-506.
186. Tantawy, M., et al., *Identification of microRNA signature in different pediatric brain tumors.* Genet Mol Biol, 2018. **41**(1): p. 27-34.
187. Eis, P.S., et al., *Accumulation of miR-155 and BIC RNA in human B cell lymphomas.* Proc Natl Acad Sci U S A, 2005. **102**(10): p. 3627-32.
188. Qin, J., et al., *Downregulation of microRNA-132 by DNA hypermethylation is associated with cell invasion in colorectal cancer.* Onco Targets Ther, 2015. **8**: p. 3639-48.
189. Deneberg, S., et al., *microRNA-34b/c on chromosome 11q23 is aberrantly methylated in chronic lymphocytic leukemia.* Epigenetics, 2014. **9**(6): p. 910-7.
190. Kawahara, Y., *Human diseases caused by germline and somatic abnormalities in microRNA and microRNA-related genes.* Congenit Anom (Kyoto), 2014. **54**(1): p. 12-21.
191. Mehraein, Y., et al., *DICER1 syndrome can mimic different genetic tumor predispositions.* Cancer Lett, 2016. **370**(2): p. 275-8.
192. Wyman, S.K., et al., *Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity.* Genome Res, 2011. **21**(9): p. 1450-61.
193. Winter, J. and S. Diederichs, *Argonaute proteins regulate microRNA stability: Increased microRNA abundance by Argonaute proteins is due to microRNA stabilization.* RNA Biol, 2011. **8**(6): p. 1149-57.
194. www.vectorbiolabs.com.
195. Pallasch, C.P., et al., *miRNA deregulation by epigenetic silencing disrupts suppression of the oncogene PLAG1 in chronic lymphocytic leukemia.* Blood, 2009. **114**(15): p. 3255-64.
196. Wong, K.Y., et al., *Epigenetic inactivation of the MIR129-2 in hematological malignancies.* J Hematol Oncol, 2013. **6**: p. 16.

197. Wang, L.Q., et al., *Epigenetic inactivation of miR-9 family microRNAs in chronic lymphocytic leukemia--implications on constitutive activation of NFkappaB pathway.* Mol Cancer, 2013. **12**: p. 173.
198. Wang, L.Q., et al., *Epigenetic inactivation of mir-34b/c in addition to mir-34a and DAPK1 in chronic lymphocytic leukemia.* J Transl Med, 2014. **12**: p. 52.
199. Baer, C., et al., *Epigenetic silencing of miR-708 enhances NF-kappaB signaling in chronic lymphocytic leukemia.* Int J Cancer, 2015. **137**(6): p. 1352-61.
200. Wang, L.Q., et al., *Epigenetic silencing of tumor suppressor miR-3151 contributes to Chinese chronic lymphocytic leukemia by constitutive activation of MADD/ERK and PIK3R2/AKT signaling pathways.* Oncotarget, 2015. **6**(42): p. 44422-36.
201. Boutros, P.C., et al., *Spatial genomic heterogeneity within localized, multifocal prostate cancer.* Nat Genet, 2015. **47**(7): p. 736-45.
202. Visvader, J.E., *Cells of origin in cancer.* Nature, 2011. **469**(7330): p. 314-22.
203. Ehrich, M., et al., *Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry.* Proc Natl Acad Sci U S A, 2005. **102**(44): p. 15785-90.
204. *RSeqMeth 1.0.2*. Available from: https://cran.r-project.org/src/contrib/Archive/RSeqMeth/.
205. Ehrich, M., S. Bocker, and D. van den Boom, *Multiplexed discovery of sequence polymorphisms using base-specific cleavage and MALDI-TOF MS.* Nucleic Acids Res, 2005. **33**(4): p. e38.
206. Qiagen, *REPLI-g Midi/Mini Handbook* 2011: Qiagen Resources
207. Qiagen, *miScript PCR System Handbook.* Qiagen Resources, 2011.
208. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.* Methods, 2001. **25**(4): p. 402-8.
209. Blume, C.J., et al., *p53-dependent non-coding RNA networks in chronic lymphocytic leukemia.* Leukemia, 2015. **29**(10): p. 2015-23.
210. Dietrich, S., et al., *Drug-perturbation-based stratification of blood cancer.* J Clin Invest, 2018. **128**(1): p. 427-445.
211. Ferreira, P.G., et al., *Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia.* Genome Res, 2014. **24**(2): p. 212-26.
212. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
213. DeLuca, D.S., et al., *RNA-SeQC: RNA-seq metrics for quality control and process optimization.* Bioinformatics, 2012. **28**(11): p. 1530-1532.
214. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
215. Tarasov, A., et al., *Sambamba: fast processing of NGS alignment formats.* Bioinformatics, 2015. **31**(12): p. 2032-4.
216. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-30.
217. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.
218. Assenov, Y., et al., *Comprehensive analysis of DNA methylation data with RnBeads.* Nat Methods, 2014. **11**(11): p. 1138-1140.
219. Teschendorff, A.E., et al., *A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data.* Bioinformatics, 2013. **29**(2): p. 189-96.
220. Sheffield, N.C. and C. Bock, *LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor.* Bioinformatics, 2016. **32**(4): p. 587-9.
221. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-89.
222. ENCODE. *http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeAwgTfbsUniform/*.

Available from: http://hgdownload.cse.ucsc.edu/goldenpath/hg19/encodeDCC/wgEncodeAwgTfbsUniform/.
223. Desper, R. and O. Gascuel, *Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle.* J Comput Biol, 2002. **9**(5): p. 687-705.
224. Ernst, J., et al., *Mapping and analysis of chromatin state dynamics in nine human cell types.* Nature, 2011. **473**(7345): p. 43-9.
225. Smith, E. and M. Sigvardsson, *The roles of transcription factors in B lymphocyte commitment, development, and transformation.* J Leukoc Biol, 2004. **75**(6): p. 973-81.
226. Phillips, J.E. and V.G. Corces, *CTCF: master weaver of the genome.* Cell, 2009. **137**(7): p. 1194-211.
227. Zandi, S., et al., *EBF1 Is Essential for B-Lineage Priming and Establishment of a Transcription Factor Network in Common Lymphoid Progenitors.* The Journal of Immunology, 2008. **181**(5): p. 3364-3372.
228. Heltemes-Harris, L.M., et al., *Ebf1 or Pax5 haploinsufficiency synergizes with STAT5 activation to initiate acute lymphoblastic leukemia.* J Exp Med, 2011. **208**(6): p. 1135-49.
229. miRBase. *miRBase download page*. Available from: http://www.mirbase.org/ftp.shtml.
230. Fujita, S. and H. Iba, *Putative promoter regions of miRNA genes involved in evolutionarily conserved regulatory systems among vertebrates.* Bioinformatics, 2008. **24**(3): p. 303-308.
231. Corcoran, D.L., et al., *Features of mammalian microRNA promoters emerge from polymerase II chromatin immunoprecipitation data.* PLoS One, 2009. **4**(4): p. e5279.
232. Fukao, T., et al., *An evolutionarily conserved mechanism for microRNA-223 expression revealed by microRNA gene profiling.* Cell, 2007. **129**(3): p. 617-31.
233. Visone, R., et al., *Karyotype-specific microRNA signature in chronic lymphocytic leukemia.* Blood, 2009. **114**(18): p. 3872-9.
234. Mraz, M., et al., *miR-34a, miR-29c and miR-17-5p are downregulated in CLL patients with TP53 abnormalities.* Leukemia, 2009. **23**(6): p. 1159-63.
235. Zanette, D.L., et al., *miRNA expression profiles in chronic lymphocytic and acute lymphocytic leukemia.* Braz J Med Biol Res, 2007. **40**(11): p. 1435-40.
236. He, P.F., et al., *Methylation-associated DOK1 and DOK2 down-regulation: Potential biomarkers for predicting adverse prognosis in acute myeloid leukemia.* J Cell Physiol, 2018. **233**(9): p. 6604-6614.
237. Miyagaki, H., et al., *DOK2 as a marker of poor prognosis of patients with gastric adenocarcinoma after curative resection.* Ann Surg Oncol, 2012. **19**(5): p. 1560-7.
238. Weber, D.A., et al., *Neutrophil-derived JAML inhibits repair of intestinal epithelial injury during acute inflammation.* Mucosal Immunol, 2014. **7**(5): p. 1221-32.
239. Verdino, P., et al., *Molecular insights into gammadelta T cell costimulation by an anti-JAML antibody.* Structure, 2011. **19**(1): p. 80-9.
240. Luissint, A.C., et al., *JAM-L-mediated leukocyte adhesion to endothelial cells is regulated in cis by alpha4beta1 integrin activation.* J Cell Biol, 2008. **183**(6): p. 1159-73.
241. Moog-Lutz, C., et al., *JAML, a novel protein with characteristics of a junctional adhesion molecule, is induced during differentiation of myeloid leukemia cells.* Blood, 2003. **102**(9): p. 3371-8.
242. Berger, A.H., et al., *Identification of DOK genes as lung tumor suppressors.* Nat Genet, 2010. **42**(3): p. 216-23.
243. Niki, M., et al., *Role of Dok-1 and Dok-2 in leukemia suppression.* J Exp Med, 2004. **200**(12): p. 1689-95.
244. Yasuda, T., et al., *Role of Dok-1 and Dok-2 in myeloid homeostasis and suppression of leukemia.* J Exp Med, 2004. **200**(12): p. 1681-7.
245. An, C.H., et al., *Mutational and expressional analysis of a haploinsufficient tumor suppressor gene DOK2 in gastric and colorectal cancers.* Apmis, 2011. **119**(8): p. 562-4.
246. Coppin, E., et al., *Mutational analysis of the DOK2 haploinsufficient tumor suppressor gene in chronic myelomonocytic leukemia (CMML).* Leukemia, 2015. **29**(2): p. 500-2.
247. Fang, F., et al., *Genomic and Epigenomic Signatures in Ovarian Cancer Associated with Resensitization to Platinum Drugs.* Cancer Res, 2018. **78**(3): p. 631-644.

248. Lum, E., et al., *Loss of DOK2 induces carboplatin resistance in ovarian cancer via suppression of apoptosis.* Gynecol Oncol, 2013. **130**(2): p. 369-76.
249. Huang, J., et al., *Co-expression and significance of Dok2 and Ras p21 protein activator 1 in breast cancer.* Oncol Lett, 2017. **14**(5): p. 5386-5392.
250. Laroche-Lefebvre, C., et al., *Dok-1 and Dok-2 Regulate the Formation of Memory CD8+ T Cells.* J Immunol, 2016. **197**(9): p. 3618-3627.
251. Downer, E.J., D.G. Johnston, and M.A. Lynch, *Differential role of Dok1 and Dok2 in TLR2-induced inflammatory signaling in glia.* Mol Cell Neurosci, 2013. **56**: p. 148-58.
252. Mihrshahi, R., A.N. Barclay, and M.H. Brown, *Essential roles for Dok2 and RasGAP in CD200 receptor-mediated regulation of human myeloid cells.* J Immunol, 2009. **183**(8): p. 4879-86.
253. Mihrshahi, R. and M.H. Brown, *Downstream of tyrosine kinase 1 and 2 play opposing roles in CD200 receptor signaling.* J Immunol, 2010. **185**(12): p. 7216-22.
254. Yasuda, T., et al., *Dok-1 and Dok-2 are negative regulators of T cell receptor signaling.* Int Immunol, 2007. **19**(4): p. 487-95.
255. Van Slyke, P., et al., *Dok-R mediates attenuation of epidermal growth factor-dependent mitogen-activated protein kinase and Akt activation through processive recruitment of c-Src and Csk.* Mol Cell Biol, 2005. **25**(9): p. 3831-41.
256. Abramson, J., G. Rozenblum, and I. Pecht, *Dok protein family members are involved in signaling mediated by the type 1 Fcepsilon receptor.* Eur J Immunol, 2003. **33**(1): p. 85-91.
257. Suzu, S., et al., *p56(dok-2) as a cytokine-inducible inhibitor of cell proliferation and signal transduction.* Embo j, 2000. **19**(19): p. 5114-22.
258. Jones, N. and D.J. Dumont, *Recruitment of Dok-R to the EGF receptor through its PTB domain is required for attenuation of Erk MAP kinase activation.* Curr Biol, 1999. **9**(18): p. 1057-60.
259. Celis-Gutierrez, J., et al., *Dok1 and Dok2 proteins regulate natural killer cell development and function.* Embo j, 2014. **33**(17): p. 1928-40.
260. Coppin, E., et al., *Dok1 and Dok2 Proteins Regulate Cell Cycle in Hematopoietic Stem and Progenitor Cells.* J Immunol, 2016. **196**(10): p. 4110-21.
261. Vlad, G., et al., *Membrane and soluble ILT3 are critical to the generation of T suppressor cells and induction of immunological tolerance.* Int Rev Immunol, 2010. **29**(2): p. 119-32.
262. Buckland, M. and G. Lombardi, *Aspirin and the induction of tolerance by dendritic cells.* Handb Exp Pharmacol, 2009(188): p. 197-213.
263. Chang, C.C., et al., *Ig-like transcript 3 regulates expression of proinflammatory cytokines and migration of activated T cells.* J Immunol, 2009. **182**(9): p. 5208-16.
264. Fanning, L.B., et al., *Downregulation of key early events in the mobilization of antigen-bearing dendritic cells by leukocyte immunoglobulin-like Receptor B4 in a mouse model of allergic pulmonary inflammation.* PLoS One, 2013. **8**(2): p. e57007.
265. Cella, M., et al., *A novel inhibitory receptor (ILT3) expressed on monocytes, macrophages, and dendritic cells involved in antigen processing.* J Exp Med, 1997. **185**(10): p. 1743-51.
266. Bankey, P.E., et al., *Cytokine induced expression of programmed death ligands in human neutrophils.* Immunol Lett, 2010. **129**(2): p. 100-7.
267. Inui, M., et al., *Human CD43+ B cells are closely related not only to memory B cells phenotypically but also to plasmablasts developmentally in healthy individuals.* Int Immunol, 2015. **27**(7): p. 345-55.
268. Ju, X.S., et al., *Immunoglobulin-like transcripts ILT2, ILT3 and ILT7 are expressed by human dendritic cells and down-regulated following activation.* Gene, 2004. **331**: p. 159-64.
269. Davis, M.E. and M.U. Gack, *Ubiquitination in the antiviral immune response.* Virology, 2015. **479-480**: p. 52-65.
270. Hu, H. and S.-C. Sun, *Ubiquitin signaling in immune responses.* Cell Research, 2016. **26**: p. 457.
271. Shi, J.H. and S.C. Sun, *TCR signaling to NF-kappaB and mTORC1: Expanding roles of the CARMA1 complex.* Mol Immunol, 2015. **68**(2 Pt C): p. 546-57.
272. Kwon, D.Y., et al., *The E3 ubiquitin ligase mind bomb 1 ubiquitinates and promotes the degradation of survival of motor neuron protein.* Mol Biol Cell, 2013. **24**(12): p. 1863-71.

273. Ossipova, O., J. Ezan, and S.Y. Sokol, *PAR-1 phosphorylates Mind bomb to promote vertebrate neurogenesis.* Dev Cell, 2009. **17**(2): p. 222-33.
274. Jiang, X. and Z.J. Chen, *The role of ubiquitylation in immune defence and pathogen evasion.* Nat Rev Immunol, 2011. **12**(1): p. 35-48.
275. Chen, Z.J., *Ubiquitination in signaling to and activation of IKK.* Immunol Rev, 2012. **246**(1): p. 95-106.
276. Wang, L., S. Li, and M.E. Dorf, *NEMO binds ubiquitinated TANK-binding kinase 1 (TBK1) to regulate innate immune responses to RNA viruses.* PLoS One, 2012. **7**(9): p. e43756.
277. Li, M.X. and P.M. Hwang, *Structure and function of cardiac troponin C (TNNC1): Implications for heart failure, cardiomyopathies, and troponin modulating drugs.* Gene, 2015. **571**(2): p. 153-66.
278. Parvatiyar, M.S., et al., *A mutation in TNNC1-encoded cardiac troponin C, TNNC1-A31S, predisposes to hypertrophic cardiomyopathy and ventricular fibrillation.* J Biol Chem, 2012. **287**(38): p. 31845-55.
279. Wang, Q., et al., *New insights into the roles of Xin repeat-containing proteins in cardiac development, function, and disease.* Int Rev Cell Mol Biol, 2014. **310**: p. 89-128.
280. Rebalka, I.A. and T.J. Hawke, *Potential biomarkers of skeletal muscle damage.* Biomark Med, 2014. **8**(3): p. 375-8.
281. Nilsson, M.I., et al., *Xin is a marker of skeletal muscle damage severity in myopathies.* Am J Pathol, 2013. **183**(6): p. 1703-1709.
282. Buhl, A.M., et al., *Analysis of CLLU1 expression levels before and after therapy in patients with chronic lymphocytic leukemia.* Eur J Haematol, 2011. **86**(5): p. 405-411.
283. Gonzalez, D., et al., *CLLU1 expression has prognostic value in chronic lymphocytic leukemia after first-line therapy in younger patients and in those with mutated IGHV genes.* Haematologica, 2013. **98**(2): p. 274-8.
284. Rosenwald, A., *CLLU1 expression: the latest risk predictor in chronic lymphocytic leukemia.* Leuk Lymphoma, 2007. **48**(9): p. 1665-6.
285. Abur, U., et al., *Impact of Fluorescent In Situ Hybridization Aberrations and CLLU1 Expression on the Prognosis of Chronic Lymphocytic Leukemia: Presentation of 156 Patients from Turkey.* Turk J Haematol, 2018. **35**(1): p. 61-65.
286. Kaderi, M.A., et al., *LPL is the strongest prognostic factor in a comparative analysis of RNA-based markers in early chronic lymphocytic leukemia.* Haematologica, 2011. **96**(8): p. 1153-60.
287. Deshaies, R.J., *SCF and Cullin/Ring H2-based ubiquitin ligases.* Annu Rev Cell Dev Biol, 1999. **15**: p. 435-67.
288. Krivtsov, A.V., et al., *Cell of origin determines clinically relevant subtypes of MLL-rearranged AML.* Leukemia, 2013. **27**(4): p. 852-60.
289. Lai, A., et al., *Evidence for sequenced molecular evolution of IDH1 mutant glioblastoma from a distinct cell of origin.* J Clin Oncol, 2011. **29**(34): p. 4482-90.
290. Alcantara Llaguno, S., et al., *Malignant Astrocytomas Originate from Neural Stem/Progenitor Cells in a Somatic Tumor Suppressor Mouse Model.* Cancer Cell. **15**(1): p. 45-56.
291. Verhaak, R.G., et al., *Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1.* Cancer Cell, 2010. **17**(1): p. 98-110.
292. Barker, N., et al., *Crypt stem cells as the cells-of-origin of intestinal cancer.* Nature, 2009. **457**(7229): p. 608-11.
293. Zhu, L., et al., *Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation.* Nature, 2009. **457**(7229): p. 603-7.
294. Sangiorgi, E. and M.R. Capecchi, *Bmi1 is expressed in vivo in intestinal stem cells.* Nat Genet, 2008. **40**(7): p. 915-20.
295. Westphalen, C.B., et al., *Long-lived intestinal tuft cells serve as colon cancer-initiating cells.* J Clin Invest, 2014. **124**(3): p. 1283-95.
296. Powell, A.E., et al., *The pan-ErbB negative regulator Lrig1 is an intestinal stem cell marker that functions as a tumor suppressor.* Cell, 2012. **149**(1): p. 146-58.

297. Schwitalla, S., et al., *Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties.* Cell, 2013. **152**(1-2): p. 25-38.
298. Gibson, P., et al., *Subtypes of medulloblastoma have distinct developmental origins.* Nature, 2010. **468**(7327): p. 1095-9.
299. Capper, D., et al., *DNA methylation-based classification of central nervous system tumours.* Nature, 2018. **555**(7697): p. 469-474.
300. Brocks, D., et al., *Intratumor DNA methylation heterogeneity reflects clonal evolution in aggressive prostate cancer.* Cell Rep, 2014. **8**(3): p. 798-806.
301. Quek, K., et al., *DNA methylation intratumor heterogeneity in localized lung adenocarcinomas.* Oncotarget, 2017. **8**(13): p. 21994-22002.
302. Lin, D.C., et al., *Genomic and Epigenomic Heterogeneity of Hepatocellular Carcinoma.* Cancer Res, 2017. **77**(9): p. 2255-2265.
303. Feinberg, A.P., M.A. Koldobskiy, and A. Gondor, *Epigenetic modulators, modifiers and mediators in cancer aetiology and progression.* Nat Rev Genet, 2016. **17**(5): p. 284-99.
304. Spencer, D.H., et al., *CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression.* Cell, 2017. **168**(5): p. 801-816.e13.
305. Turcan, S., et al., *IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype.* Nature, 2012. **483**: p. 479.
306. Chan, K.M., et al., *The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression.* Genes Dev, 2013. **27**(9): p. 985-90.
307. Bender, S., et al., *Reduced H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas.* Cancer Cell, 2013. **24**(5): p. 660-72.
308. Lewis, P.W., et al., *Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma.* Science, 2013. **340**(6134): p. 857-61.
309. Kulis, M., et al., *Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia.* Nature Genetics, 2012. **44**(11): p. 1236-1242.
310. Agirre, X., et al., *Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers.* Genome Res, 2015. **25**(4): p. 478-87.
311. Kretzmer, H., et al., *DNA methylome analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control.* Nat Genet, 2015. **47**(11): p. 1316-25.
312. Queiros, A.C., et al., *Decoding the DNA Methylome of Mantle Cell Lymphoma in the Light of the Entire B Cell Lineage.* Cancer Cell, 2016. **30**(5): p. 806-821.
313. Wedge, E., et al., *Global hypomethylation is an independent prognostic factor in diffuse large B cell lymphoma.* Am J Hematol, 2017. **92**(7): p. 689-694.
314. Hansen, K.D., et al., *Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization.* Genome Res, 2014. **24**(2): p. 177-84.
315. Hernando, H., et al., *The B cell transcription program mediates hypomethylation and overexpression of key genes in Epstein-Barr virus-associated proliferative conversion.* Genome Biol, 2013. **14**(1): p. R3.
316. Berman, B.P., et al., *Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains.* Nat Genet, 2011. **44**(1): p. 40-6.
317. Hansen, K.D., et al., *Increased methylation variation in epigenetic domains across cancer types.* Nature Genetics, 2011. **43**(8): p. 768-U77.
318. Landau, D.A., et al., *Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia.* Cancer Cell, 2014. **26**(6): p. 813-25.
319. Chen, R.Z., et al., *DNA hypomethylation leads to elevated mutation rates.* Nature, 1998. **395**(6697): p. 89-93.
320. Caron, G., et al., *Cell-Cycle-Dependent Reconfiguration of the DNA Methylome during Terminal Differentiation of Human B Cells into Plasma Cells.* Cell Rep, 2015. **13**(5): p. 1059-71.

321. Arab, K., et al., *Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A.* Mol Cell, 2014. **55**(4): p. 604-14.
322. Dominguez, P.M., et al., *DNA Methylation Dynamics of Germinal Center B Cells Are Mediated by AID.* Cell Rep, 2015. **12**(12): p. 2086-98.
323. Shaknovich, R., et al., *DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation.* Blood, 2011. **118**(13): p. 3559-3569.
324. Teater, M., et al., *AICDA drives epigenetic heterogeneity and accelerates germinal center-derived lymphomagenesis.* Nat Commun, 2018. **9**(1): p. 222.
325. Andersson, R., et al., *An atlas of active enhancers across human cell types and tissues.* Nature, 2014. **507**(7493): p. 455-61.
326. Roadmap Epigenomics, C., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30.
327. Ziller, M.J., et al., *Charting a dynamic DNA methylation landscape of the human genome.* Nature, 2013. **500**(7463): p. 477-81.
328. Martin-Subero, J.I., C. Lopez-Otin, and E. Campo, *Genetic and epigenetic basis of chronic lymphocytic leukemia.* Curr Opin Hematol, 2013. **20**(4): p. 362-8.
329. Ntziachristos, P., O. Abdel-Wahab, and I. Aifantis, *Emerging concepts of epigenetic dysregulation in hematological malignancies.* Nat Immunol, 2016. **17**(9): p. 1016-24.
330. Almamun, M., et al., *Integrated methylome and transcriptome analysis reveals novel regulatory elements in pediatric acute lymphoblastic leukemia.* Epigenetics, 2015. **10**(9): p. 882-90.
331. Lai, A.Y., et al., *DNA methylation profiling in human B cells reveals immune regulatory elements and epigenetic plasticity at Alu elements during B-cell activation.* Genome Res, 2013. **23**(12): p. 2030-41.
332. Lee, S.T., et al., *A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network.* Nucleic Acids Res, 2012. **40**(22): p. 11339-51.
333. Oakes, C.C., et al., *DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia.* Nat Genet, 2016.
334. Pui, C.H., et al., *Biology, risk stratification, and therapy of pediatric acute leukemias: an update.* J Clin Oncol, 2011. **29**(5): p. 551-65.
335. Nordlund, J., et al., *Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia.* Genome Biol, 2013. **14**(9): p. r105.
336. Lee, S.T., et al., *Epigenetic remodeling in B-cell acute lymphoblastic leukemia occurs in two tracks and employs embryonic stem cell-like signatures.* Nucleic Acids Res, 2015. **43**(5): p. 2590-602.
337. Network, C.G.A.R., *Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia.* N Engl J Med, 2013. **368**(22): p. 2059-74.
338. Mockridge, C.I., et al., *Reversible anergy of sIgM-mediated signaling in the two subsets of CLL defined by VH-gene mutational status.* Blood, 2007. **109**(10): p. 4424-31.
339. Maier, H., et al., *Early B cell factor cooperates with Runx1 and mediates epigenetic changes associated with mb-1 transcription.* Nat Immunol, 2004. **5**(10): p. 1069-77.
340. Boller, S., et al., *Pioneering Activity of the C-Terminal Domain of EBF1 Shapes the Chromatin Landscape for B Cell Programming.* Immunity, 2016. **44**(3): p. 527-541.
341. Guilhamon, P., et al., *Meta-analysis of IDH-mutant cancers identifies EBF1 as an interaction partner for TET2.* Nat Commun, 2013. **4**: p. 2166.
342. Marke, R., F.N. van Leeuwen, and B. Scheijen, *The many faces of IKZF1 in B-cell precursor acute lymphoblastic leukemia.* Haematologica, 2018. **103**(4): p. 565-574.
343. Papavassiliou, A.G., M. Treier, and D. Bohmann, *Intramolecular signal transduction in c-Jun.* Embo j, 1995. **14**(9): p. 2014-9.
344. Li, X.Y. and M.R. Green, *Intramolecular inhibition of activating transcription factor-2 function by its DNA-binding domain.* Genes Dev, 1996. **10**(5): p. 517-27.
345. Deppmann, C.D., et al., *Phosphorylation of BATF regulates DNA binding: a novel mechanism for AP-1 (activator protein-1) regulation.* Biochem J, 2003. **374**(Pt 2): p. 423-31.

346. Lau, E. and Z.A. Ronai, *ATF2 - at the crossroad of nuclear and cytosolic functions.* J Cell Sci, 2012. **125**(Pt 12): p. 2815-24.
347. Bailey, J. and G.N. Europe-Finner, *Identification of human myometrial target genes of the c-Jun NH2-terminal kinase (JNK) pathway: the role of activating transcription factor 2 (ATF2) and a novel spliced isoform ATF2-small.* J Mol Endocrinol, 2005. **34**(1): p. 19-35.
348. Hayakawa, J., et al., *Identification of promoters bound by c-Jun/ATF2 during rapid large-scale gene activation following genotoxic stress.* Mol Cell, 2004. **16**(4): p. 521-35.
349. Huang, R.P. and E.D. Adamson, *The phosphorylated forms of the transcription factor, Egr-1, bind to DNA more efficiently than non-phosphorylated.* Biochem Biophys Res Commun, 1994. **200**(3): p. 1271-6.
350. Kellum, R. and P. Schedl, *A position-effect assay for boundaries of higher order chromosomal domains.* Cell, 1991. **64**(5): p. 941-50.
351. Geyer, P.K. and V.G. Corces, *DNA position-specific repression of transcription by a Drosophila zinc finger protein.* Genes Dev, 1992. **6**(10): p. 1865-73.
352. Bell, A.C. and G. Felsenfeld, *Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene.* Nature, 2000. **405**(6785): p. 482-5.
353. Wang, H., et al., *Widespread plasticity in CTCF occupancy linked to DNA methylation.* Genome Res, 2012. **22**(9): p. 1680-8.
354. Vojta, A., et al., *Repurposing the CRISPR-Cas9 system for targeted DNA methylation.* Nucleic Acids Res, 2016. **44**(12): p. 5615-28.
355. Davies, J.O.J., et al., *How best to identify chromosomal interactions: a comparison of approaches.* Nature Methods, 2017. **14**: p. 125.
356. Mott, J.L., et al., *mir-29 regulates Mcl-1 protein expression and apoptosis.* Oncogene, 2007. **26**: p. 6133.
357. Pekarsky, Y., et al., *Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181.* Cancer Res, 2006. **66**(24): p. 11590-3.
358. Maura, F., et al., *Association between gene and miRNA expression profiles and stereotyped subset #4 B-cell receptor in chronic lymphocytic leukemia.* Leuk Lymphoma, 2015. **56**(11): p. 3150-8.
359. Bae, H.J., et al., *MicroRNA-29c functions as a tumor suppressor by direct targeting oncogenic SIRT1 in hepatocellular carcinoma.* Oncogene, 2014. **33**(20): p. 2557-67.
360. Fabbri, M., et al., *MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B.* Proc Natl Acad Sci U S A, 2007. **104**(40): p. 15805-10.
361. Garzon, R., et al., *MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1.* Blood, 2009. **113**(25): p. 6411-8.
362. Zhao, X., et al., *MiRNA-29c regulates cell growth and invasion by targeting CDK6 in bladder cancer.* Am J Transl Res, 2015. **7**(8): p. 1382-9.
363. Lu, Y., et al., *MiR-29c inhibits cell growth, invasion, and migration of pancreatic cancer by targeting ITGB1.* Onco Targets Ther, 2016. **9**: p. 99-109.
364. Liu, X.G., et al., *High expression of serum miR-21 and tumor miR-200c associated with poor prognosis in patients with lung cancer.* Med Oncol, 2012. **29**(2): p. 618-26.
365. Tejero, R., et al., *miR-141 and miR-200c as markers of overall survival in early stage non-small cell lung cancer adenocarcinoma.* PLoS One, 2014. **9**(7): p. e101899.
366. Wu, P.P., et al., *MicroRNA-141 regulates the tumour suppressor DLC1 in colorectal cancer.* Neoplasma, 2015. **62**(5): p. 705-12.
367. Debeb, B.G., et al., *miR-141-Mediated Regulation of Brain Metastasis From Breast Cancer.* J Natl Cancer Inst, 2016. **108**(8).
368. Choi, S.K., et al., *Overexpression of the miR-141/200c cluster promotes the migratory and invasive ability of triple-negative breast cancer cells through the activation of the FAK and PI3K/AKT signaling pathways by secreting VEGF-A.* BMC Cancer, 2016. **16**: p. 570.
369. Li, C., et al., *Serum miR-486-5p as a diagnostic marker in cervical cancer: with investigation of potential mechanisms.* BMC Cancer, 2018. **18**(1): p. 61.

370. Yang, Y., et al., *The miR-486-5p plays a causative role in prostate cancer through negative regulation of multiple tumor suppressor pathways.* Oncotarget, 2017. **8**(42): p. 72835-72846.
371. Yu, X., et al., *miR-195 targets cyclin D3 and survivin to modulate the tumorigenesis of non-small cell lung cancer.* Cell Death Dis, 2018. **9**(2): p. 193.
372. Hui, W., et al., *MicroRNA-195 inhibits the proliferation of human glioma cells by directly targeting cyclin D1 and cyclin E1.* PLoS One, 2013. **8**(1): p. e54932.
373. Tsang, J., J. Zhu, and A. van Oudenaarden, *MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals.* Mol Cell, 2007. **26**(5): p. 753-67.
374. Martinez, N.J., et al., *A C. elegans genome-scale microRNA network contains composite feedback motifs with high flux capacity.* Genes Dev, 2008. **22**(18): p. 2535-49.
375. Song, L., et al., *miR-486 sustains NF-kappaB activity by disrupting multiple NF-kappaB-negative feedback loops.* Cell Res, 2013. **23**(2): p. 274-89.
376. Kimura, K., et al., *Loop Regulation Between microRNAs and Epigenetics Underlie microRNA Dysregulation in Multiple Myeloma and Is Associated with the Disease Progression.* Blood, 2015. **126**(23): p. 3013-3013.
377. Oppliger Leibundgut, E., et al., *CLLU1 expression distinguishes chronic lymphocytic leukemia from other mature B-cell neoplasms.* Leuk Res, 2012. **36**(9): p. 1204-7.
378. Coppin, E., et al., *Mutational analysis of the DOK2 haploinsufficient tumor suppressor gene in chronic myelomonocytic leukemia (CMML).* Leukemia, 2014. **29**: p. 500.
379. Stempin, C.C., et al., *The E3 ubiquitin ligase mind bomb-2 (MIB2) protein controls B-cell CLL/lymphoma 10 (BCL10)-dependent NF-kappaB activation.* J Biol Chem, 2011. **286**(43): p. 37147-57.
380. Li, S., et al., *Mapping a dynamic innate immunity protein interaction network regulating type I interferon production.* Immunity, 2011. **35**(3): p. 426-40.

7. APPENDIX

7.1 Primer sequences

Table 7-1: Primers for MassARRAY

Primer Name	Forward	Reverse	AT (°C)	Length (bp)
miR-486	aggaagagagGTTTTGTTTAGAGATTTGGAAGTGG	cagtaatacgactcactatagggagaaggctCCACCACCTATACTAAACCCCTACT	58	314
miR-29c	aggaagagagGGTTTGAGTAGTTGATATTAGGGAAAA	cagtaatacgactcactatagggagaaggctAACACCAAATTCTACCAACTCACTT	58	467
miR-141	aggaagagagGGGAAGGGTGGGGGTATTTA	cagtaatacgactcactatagggagaaggctTTACCCCAAAACTAAAAATTCATCA	58	432
miR-195	aggaagagagTTTTTTAGGATGGTTTTTTAGGGTT	cagtaatacgactcactatagggagaaggctAACAATATATTAACCCCTCCAATCA	58	425
let-7b	aggaagagagGATAGATGTGGTTTGTATTTGTGGTT	cagtaatacgactcactatagggagaaggctCCTAAATTAATCCTATCTATCCCAAAA	58	449

forward primers were 5' tagged with 'aggaagagag' sequence and reverse primers with 'cagtaatacgactcactatagggagaaggct' sequence; AT, annealing temperature for PCR.

Table 7-2: miScript qPCR primers from Qiagen

Assay name	Catalog number	miRNA
Hs_let-7b*_1	MS00008281	let-7b-3p
Hs_let-7b_1	MS00003122	let-7b-5p
Hs_miR-141_1	MS00003507	miR-141-3p
Hs_miR-141*_1	MS00008680	miR-141-5p
Hs_miR-195*_1	MS00008953	miR-195-3p
Hs_miR-195_1	MS00003703	miR-195-5p
Hs_miR-29c*_1	MS00009303	miR-29c-5p
Hs_miR-486-3p_2	MS00031892	miR-486-3p
Hs_miR-486_1	MS00004284	miR-486-5p
Hs_SNORD61_11	MS00033705	SNORD61
Hs_SNORD72_11	MS00033719	SNORD72

7.2 RNAseq data – alignment statistics

Table 7-3: RNAseq alignment statistics – normal B cells

Patient ID	CLL-943-01-1R	CLL-943-01-2R	CLL-943-01-3R
Number of input reads	120741342	117753748	142911872
UNIQUE READS:			
Uniquely mapped reads number	**104646255**	**106031958**	**124919571**
Uniquely mapped reads %	**86,67%**	**90,05%**	**87,41%**
MULTI-MAPPING READS:			
Number of reads mapped to multiple loci	0	0	0

% of reads mapped to multiple loci	0,00%	0,00%	0,00%
Number of reads mapped to too many loci	5643396	7033178	6838432
% of reads mapped to too many loci	4,67%	5,97%	4,79%
UNMAPPED READS:			
% of reads unmapped: too many mismatches	0,00%	0,00%	0,00%
% of reads unmapped: too short	3,92%	3,31%	3,64%
% of reads unmapped: other	4,74%	0,67%	4,17%
CHIMERIC READS:			
Number of chimeric reads	1567699	1839893	1560301
% of chimeric reads	1,30%	1,56%	1,09%

Table 7-3 continuation

CLL-944-01-1R	CLL-944-01-2R	CLL-944-01-3R	CLL-948-01-1R	CLL-948-01-2R	CLL-948-01-3R
109819252	87814697	95372186	128504834	160122739	132315885
99906169	**80223144**	**86343174**	**117197150**	**146462691**	**119824662**
90,97%	**91,36%**	**90,53%**	**91,20%**	**91,47%**	**90,56%**
0	0	0	0	0	0
0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
5386257	4242232	4792750	5587890	7560174	6949709
4,90%	4,83%	5,03%	4,35%	4,72%	5,25%
0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
3,87%	3,70%	4,09%	4,11%	3,52%	3,89%
0,25%	0,11%	0,36%	0,34%	0,29%	0,29%
1481457	1390336	1619746	2043121	1932889	1782381
1,35%	1,58%	1,70%	1,59%	1,21%	1,35%

Table 7-4: RNAseq alignment statistics – CLLs

	H005-001N	H005-0TRN	H005-19N0	H005-2LSW
Number of input reads |	47688538	94162171	97157104	70597681
UNIQUE READS:				
Uniquely mapped reads number |	**45057886**	**87278675**	**90826531**	**65981581**
Uniquely mapped reads % |	**94,48%**	**92,69%**	**93,48%**	**93,46%**
MULTI-MAPPING READS:				
Number of reads mapped to multiple loci |	0	0	0	0
% of reads mapped to multiple loci |	0,00%	0,00%	0,00%	0,00%
Number of reads mapped to too many loci |	1742963	4021853	4182122	2955553
% of reads mapped to too many loci |	3,65%	4,27%	4,30%	4,19%

UNMAPPED READS:				
% of reads unmapped: too many mismatches |	0,00%	0,00%	0,00%	0,00%
% of reads unmapped: too short |	1,81%	2,88%	2,13%	2,29%
% of reads unmapped: other |	0,05%	0,16%	0,08%	0,06%
CHIMERIC READS:				
Number of chimeric reads |	655461	1187866	1221986	1141074
% of chimeric reads |	1,37%	1,26%	1,26%	1,62%

Table 7-4 continuation

H005-33WM	**H005-46LY**	**H005-5S0Z**	**H005-9S37**	**H005-A72M**	**H005-CI7B**	**H005-FBE8**	**H005-GN55**
92788445	82827984	121357432	86649223	84243687	81758497	101780776	76967496
86596639	**77340650**	**114248236**	**81395807**	**79152822**	**77292127**	**95847508**	**72435122**
93,33%	**93,38%**	**94,14%**	**93,94%**	**93,96%**	**94,54%**	**94,17%**	**94,11%**
0	0	0	0	0	0	0	0
0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
4311512	3375726	4744199	3309657	3399965	2813320	3599966	2925815
4,65%	4,08%	3,91%	3,82%	4,04%	3,44%	3,54%	3,80%
0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2,00%	2,50%	1,90%	2,17%	1,95%	1,96%	2,22%	2,03%
0,03%	0,05%	0,04%	0,07%	0,05%	0,06%	0,08%	0,06%
1087054	1337615	1348440	1101659	1173904	1108709	1399302	1215882
1,17%	1,61%	1,11%	1,27%	1,39%	1,36%	1,37%	1,58%

Table 7-4 continuation

H005-LBKM	**H005-M7OV**	**H005-MOWV**	**H005-PK0F**	**H005-PQR8**	**H005-Q145**	**H005-Q9ZM**	**H005-QOI5**
107368615	84192412	112750817	82942045	81981181	98034545	109543792	102492500
100990320	**78781811**	**106503668**	**78278867**	**77389165**	**91879580**	**102167727**	**96791680**
94,06%	**93,57%**	**94,46%**	**94,38%**	**94,40%**	**93,72%**	**93,27%**	**94,44%**
0	0	0	0	0	0	0	0
0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
4147152	3767555	3708273	2756010	2828102	3811234	4864130	3487247
3,86%	4,47%	3,29%	3,32%	3,45%	3,89%	4,44%	3,40%
0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2,01%	1,90%	2,18%	2,24%	2,08%	2,34%	2,03%	2,10%

0,06%	0,05%	0,07%	0,06%	0,07%	0,05%	0,26%	0,06%
1461762	933802	1467386	1324089	1089659	1219618	1466292	1315035
1,36%	1,11%	1,30%	1,60%	1,33%	1,24%	1,34%	1,28%

Table 7-4 continuation

H005-RXC4	H005-SCQR	H005-SS98	H005-U8HQ	H005-UCF9	H005-UJI6	H005-X5Z3	H005-XMEH
96140721	97856660	98799773	105875840	105232496	77103110	90600158	81301114
89826236	**91536076**	**93491454**	**98362067**	**98806106**	**73013316**	**84917328**	**76127932**
93,43%	**93,54%**	**94,63%**	**92,90%**	**93,89%**	**94,70%**	**93,73%**	**93,64%**
0	0	0	0	0	0	0	0
0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
4241123	3836258	3656948	5070085	4235225	2740512	3450669	3727535
4,41%	3,92%	3,70%	4,79%	4,02%	3,55%	3,81%	4,58%
0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2,00%	2,47%	1,61%	2,18%	2,03%	1,71%	2,40%	1,72%
0,16%	0,07%	0,06%	0,13%	0,05%	0,04%	0,06%	0,06%
982907	1465170	1024572	1491914	1314964	969985	944934	958614
1,02%	1,50%	1,04%	1,41%	1,25%	1,26%	1,04%	1,18%

Table 7-4 continuation

H005-Y67L	H005-YNIY	H005-YV7I	H005-Z0HX	H005-ZF32	H005-ZRWN
107326927	106016882	87816774	111087895	65636358	86888936
101605564	**99038754**	**81855065**	**103409004**	**61105480**	**81345409**
94,67%	**93,42%**	**93,21%**	**93,09%**	**93,10%**	**93,62%**
0	0	0	0	0	0
0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
4058352	4991109	4047313	4946519	2605671	3712906
3,78%	4,71%	4,61%	4,45%	3,97%	4,27%
0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
1,48%	1,84%	2,13%	2,34%	2,86%	2,06%
0,07%	0,03%	0,05%	0,11%	0,07%	0,05%
785100	1057838	1217734	1598171	793737	962985
0,73%	1,00%	1,39%	1,44%	1,21%	1,11%

7.3 RNAseq data – PCA analysis

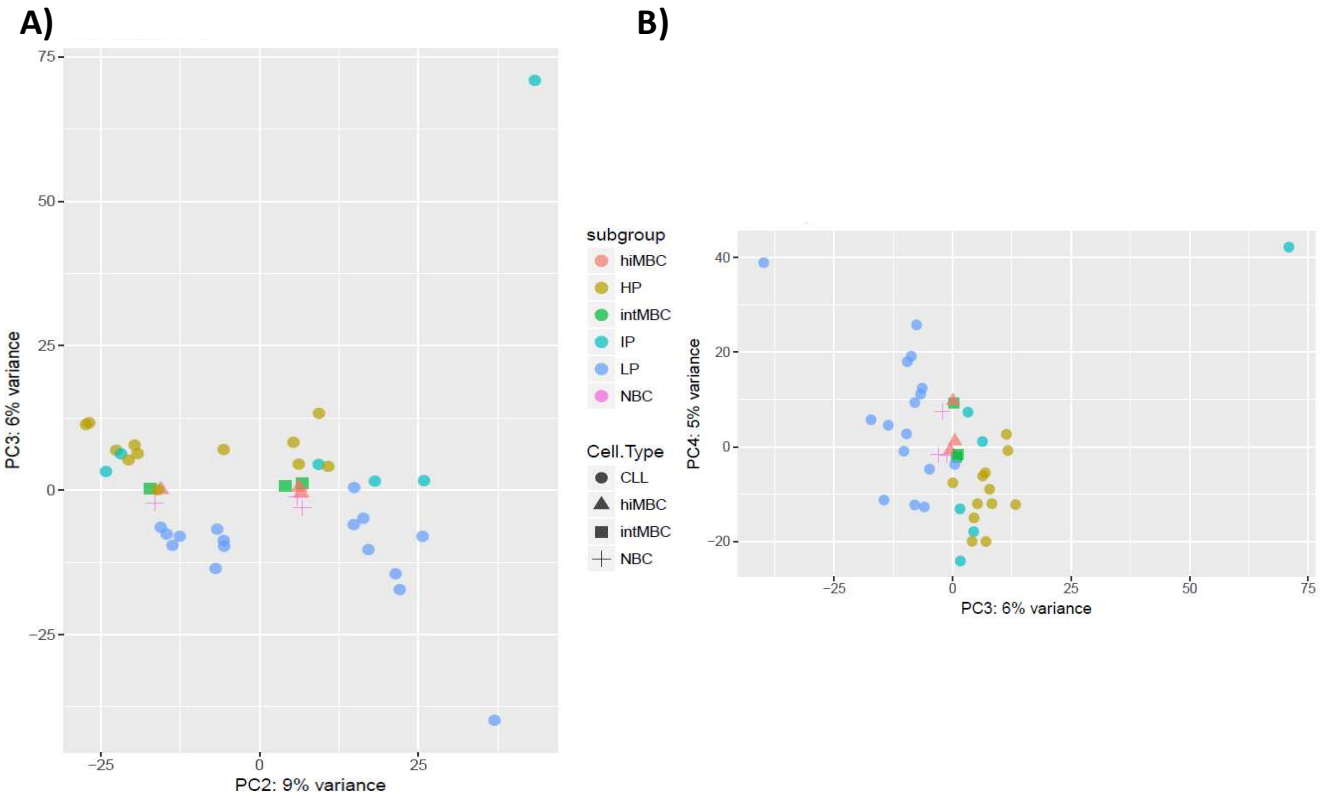


Figure 7-1: Principal Component Analysis (PCA) on rlog transformed expression counts from normal B cells and CLL. A) PC2 and PC3. B) PC3 and PC4. Normal B cell subtypes are depicted with crosses (NBC), squares (intMBC) and triangles (hiMBC). CLLs are depicted with dots.

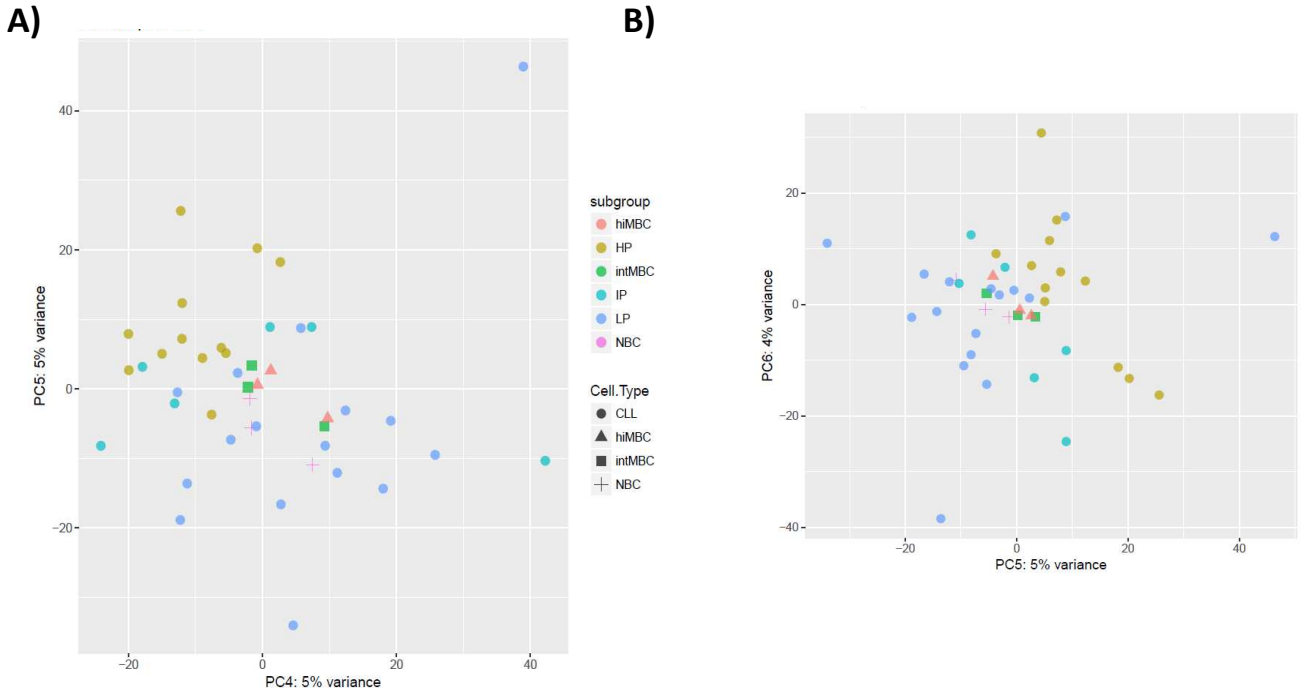


Figure 7-2: Principal Component Analysis (PCA) on rlog transformed expression counts from normal B cells and CLL. A) PC4 and PC5. B) PC5 and PC6. Normal B cell subtypes are depicted with crosses (NBC), squares (intMBC) and triangles (hiMBC). CLLs are depicted with dots

7.4 TF methylation profiles

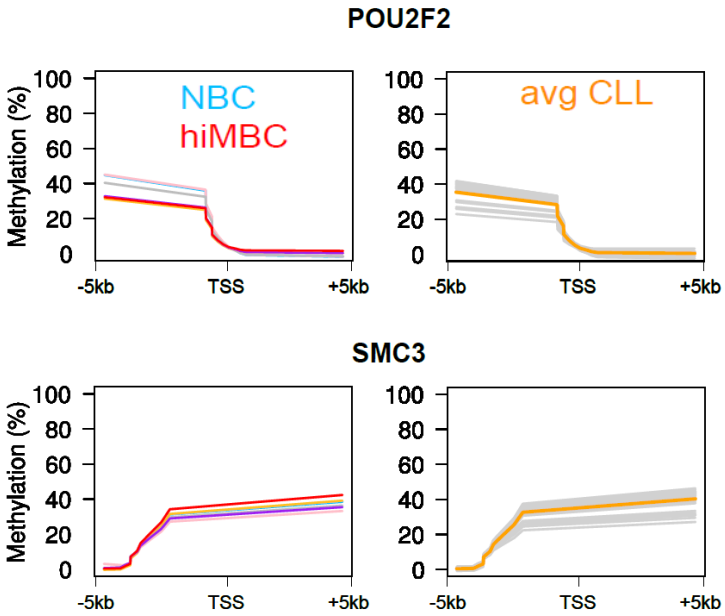


Figure 7-3: DNA methylation profiles of the promoter of POU2F2 and SMC3. Methylation in the promoter regions was shown for two normal B cell subsets, representing advanced stages of B-cell differentiation (left) and the CLLs (right). NBC: Naïve B cell; hiMBC – Class-switched memory B cell; avg CLL – average methylation change in CLLs. Y-axis represents methylation levels (%). The x-axis represents the distance to transcription start site (TSS).

7.5 CLL-specific miRNAs

Table 7-5: List of microRNAs identified as CLL-specific using cell-of-origin as a control

Class	microRNA	Region	CpGs	Correlation coefficient (rho)
classA	mir-486	chr8:41520600-41524000	cg17753169	-0,41
classC	miR-29c	chr1:207992977-207998000	cg10501210	-0,20
classD	mir-141	chr12:7022139-7025739	cg04604946	0,44
classD	mir-195	chr17:6921076-6923876	cg21040575	0,46

Methylation cut-off >20%; p-val < 0.05 (Wilcoxon rank sum test)

Table 7-6: List of experimentally validated targets of miR-29-3p from DIANA Tarbase

Target Gene	miRNA
HBP1, FEM1B, DPYSL5, BRWD3, ZBTB34, ELOVL4, FBXW7, ANKRD52, CDK16, INSIG1, MCL1, CBX6, PTP4A1, SNIP1, BTG2, TAF5, USP37, RMND5A, ATP11B, CBX5, CTDSPL2, HMGCS1, TRA2B, GPCPD1, SP1, MAZ, C1orf52, TOP2A, C1orf131, ATP5C1, CA2, SPRTN, AKAP10, CCNE1, GATAD2B, NUDT16L1, CAPRIN1, HOXA10, GCN1L1, YTHDF1, ADO, RANBP1, NUP62, CSNK1E, FAT1, DDX6, TP53BP1, UHMK1, SBK3, HAUS8, FARP2, SUFU, HACL1, ZBTB37, RABL6, WHSC1, DICER1, CEP41, LENG8, ZFYVE27, WNK1, NME7, TNKS1BP1, PILRB, TMF1, MAP1B, SARS, MDM2, DEF8, ZBTB47, ERRFI1, TUB, WDR4, LAPTM4B, PARP1, LCLAT1, ZC3H14, TMPO, PRR12, DCUN1D5, SOS1, TAS2R14, KIAA2013, HIST1H2BG, SNAPC4, AHCYL1, RGMB, RPL36, MSH6, HNRNPU, NARS, PTPRF, IBTK, HIST1H4C, ADSS, GTF3C4, SOX11, UBE3B, SLC38A1, PDIK1L, ATXN1, CPSF7, LOXL2, MBTD1, MSL3, HDGF, SFPQ, ZSCAN25, CDC23, PLXNA1, RALA, CELF1, WWTR1, JARID2, CANX, ZNF614, SNX11, LASP1, SQLE, SYT7, TMEM181, IL10RB, CSDE1, SLC25A36, SGTB, NUCKS1, LAMP2, TNFRSF10B, KRT80, EVL, CHDH, ASTE1, RPS15A, ZFP36L1, KIF26B, UBC, FAM76B, RANGAP1, HDAC4, USO1, ZNF696, MYCBP2, WASF3, ERICH1, GNB2L1, AP2B1, RREB1, RBM33, TRAM1, SLC7A2, WASF2, ZNF200, FAM135A, RUFY3, MYBL2, KIAA1614, UBA1, REV3L, OBSL1, COX15, HEBP2, IQCH, SLC25A13, DHX33, MGAT4A, BTBD7, POLR3B, NTN4, CCDC88C, FUT11, MADD, KDM1A, C16orf72, NOVA2, GRINA, TNFRSF21, DDX5, PARG, RAD51C, COG8, SPAG9, GAL3ST3, NCAPD2, ANAPC4, ACVR2B, SLFN11, SNN, AMOT, SLC25A5, RABGAP1, ZNF346, DHCR24, POLDIP2, TMEM159, HNRNPA1, GGCT, RBCK1, PHLPP2, FASN, COL5A3, RNF19A, FOXJ2, BAK1, LRP6, SIKE1, CPS1, RORA, SCML1, WDFY1, DDX3Y, TAF11, FAM13B, EHD2, CCNT2, CEACAM6, NDST1, PDS5B, TRAM2, REST, LAMC2, ACER3, ZRANB1, SMARCE1, PIK3C3, SEPHS1, ANKRD13A, SLC2A3, GSK3B, SPEN, FNDC3B, RAB21, AFF4, ZCCHC6, POMGNT1, MAPK6, ISOC1, OXCT1, C12orf4, MTHFD2, LIMA1, ATXN3, VCAN, PIGV, PRDM1, MPHOSPH9, MARK3, ZC3H11A, SLC39A9, ELN, SPDL1, MAP4K4, ERC1, MFAP3, NOP58, NAV3, ARID1B, SMC1A, PRDM11, COL19A1, KIAA2022, TNFRSF1A, GNB1, ATG16L1, CNN2, CLEC2D, ABCB1, YBX3, USH2A, MEF2C, FAM214A, ATP2B4, PIAS2, CASP8, POU2F2, ARID4B, CDK13, CDC42, PAG1, ATP1B3, HADHA, IARS2, NLE1, CLPTM1L, C16orf80, SREBF1, ITGB5, QSER1, TIMP2, VCL, PSME4, BOD1L1, U2AF2, CA12, ZFYVE26, TM9SF3, FERMT2, FAM120A, CCNJ, STMN2, ADAMTS2, SH3PXD2A, MEST, KIAA1432, CAV2, COL1A1, TFAP2C, RGS1, DOT1L, SEC23A, CDK6, ZMIZ1, PPP1R13B, GABPB1, SMS, MIB1, GPI, USP31, DPYSL2, YY1, NUFIP2, JOSD1, GID8, PDRG1, YWHAE, BRAP, G2E3, MTMR9, MED13, HERC1, DNAJB11, CASP2, MMP2, HNRNPH3, MAVS, PIM2, EIF4H, MTPN, EMC2, CHIC2, DDX17, SPTLC2, ALG13, CEMIP, XPNPEP1, GLG1, TRIM37, HIF1A, CBLL1, BLMH, URGCP, SNX5, PCSK5, SLC25A17, CCNB1IP1, SH3GLB1, TSPAN14, XBP1, PMP22, PATZ1, BMF, LARP4B, CPEB3, RPLP0, ITGB1, TNRC6B, USP10, TUBD1, HNRNPUL1, TNRC6A, ARCN1, DLD, TGFB2, USP42, SMARCA1, MAPK8, CDV3, CD276, FAM208B, PIGS, HECTD1, ISYNA1, EZH2, ADNP2, TRPS1, DNMT3B, EFTUD2, BAX, PTPN4, RBM28, IMPDH1, RPL3, CRKL, TMED2, VCPKMT, SRPX, ATAD2B, COL7A1, RLF, SPARC, CRISPLD1, METAP2, KDM5B, ARRDC3, ARSB, KPNA1, KIAA1549, NMI, RNF141, GPX7, HINT3, BCL11A, KCTD20, DTX4, LEPRE1, ATP6V1A, SUB1, COMMD2, SUV420H1, COX7A2L, CCND2, ID3, MED28, C5orf15, MED13L, NR3C1, TRAK2, HMGCR, AIM1, STRN, CAMSAP2, ZNF639, KLF12, EDEM3, FAM126A, NRBP1, LMNB1, SC5D, BBX, CAND1, RNF11, SOX4, TSPAN1, PIKFYVE, PANK3, SLC41A3, SERPINB6, TCP1, CORO1C, MAGOHB, KLHDC3, ACO1, TNFAIP3, MORF4L2, VEGFA, GDA, SELPLG, THADA, H2AFY, SET, ELF2, FBXW2, PIK3R3, GCNT2, SGK1, SASH1, RASSF8, CDKN1A, ENPP5, TBX18, FOXO3, ABCB6, GNAI2, PEPD, STX12, DNMT3A, PRDX1, QKI, PPP6R3, STX16, RCAN3, COL21A1, ASH1L, PRRC2C, CHST10, NCOA3, MOB1A, NKTR, STK38, GAPDH, IREB2, PXDN, PAN2, KDM6B, PI15, TMOD3, COL5A1, LAMC1, RAB30, TDG, DBT, ULBP2, TMTC3, HRK, COL4A2, RNF138, KLF4, EIF2S2, NIPSNAP3A, CD93, EIF2S1, TSPAN8, ITGA11, NOTCH2, LATS1, ANKEF1, PSMF1, MTSS1L, CHSY1, EMP1, KLHDC10, LIF, AMPD3, BCL2L2, ANKRD17, PLAA, MCCC2, THBS1, RNF122, HIVEP3, ATP5G2, ATP13A3, SWAP70, SLC31A1, AHNAK, TFCP2, YWHAH, MRPL50, REV1, DCAF7, EPS15L1, FBN2, ABCB7, ZFP36, PRMT1, RALGPS1, APC, POSTN, TUBGCP3, CNOT1, CALU, SMPD4, TMEM127, BST2, MYCN,	miR-29c-3p

DHX9, SUV420H2, MBD4, SCAF11, EIF4E2, LARS, THEMIS2, TMEM2, PDCL, PRKAA1, SEC31A, NOL11, GLO1, MKLN1, HOXD1, KIDINS220, LRIG3, NAV1, ISLR, PRCP, ETNK1, NREP, RAB1A, EEFSEC, PAPSS1, EIF5A, 42989, AGO3, HP1BP3, RNASEL, EPRS, RAB2B, MMP16, PXYLP1, ISG20L2, FAM167A, OTULIN, ZDHHC5, PIK3R1, RIT1, AMMECR1L, DYNLT1, ADAM12, SERBP1, CREB5, PTPRD, MRPL17, SLC16A1, OSBPL11, ARRDC4, ARNT, LAMTOR1, SBNO1, PALM2-AKAP2, CPT2, RBAK, SETD7, COL6A1, ABHD13, RPS27A, ZCCHC14, DUSP2, HIAT1, SKP2, SOGA1, LDLRAP1, AGPAT5, GGPS1, SH3GL1, CACUL1, ZNF827, G3BP1, ASXL2, CBX4, C5orf28, SKI, FEM1C, IGFBP3, NOVA1, FAM126B, GIT2, PFKM, CLTC, CCNA2, 42795, CNOT8, TBC1D7, SERPINH1, NFIB, DIAPH2, COL6A2, NTMT1, KCNJ6, FAM193B, RNF217, PGAP2, SYT16, ASAP2, BNIP2, SGSM2, AASDHPPT, MKI67, VPS26B, EMP3, TIAL1, EIF4A2, C21orf91, ELMSAN1, LRP8, RQCD1, SACS, PM20D2, ATM, CCSAP, NYAP2, TP53, SLC16A3, TMEM123, ASGR1, ZFHX3, NPC1, ATAD2, TSC22D3, TPM1, C1orf43, HMGCLL1, TACC1, PDCD4, SEC16A, NFIA, COL6A3, NSD1, COL3A1, LYSMD1, C7orf60, COL1A2, FSTL1, KCTD5, WBP1L, CLDN1, OTUD4, ZBTB5, UBTD2, CALM3, VPS37C, KIAA0355, PPP1R15B, AP1G1, WDR41, MPZL3, FZD5, NNMT, PTX3, CCDC117, SPTBN4, F11R, NKIRAS2, FAM208A, ARL5B, TAOK1, CHTOP, C11orf30, DVL3, IRGQ, LRRC58, ERAP1, NXF1, SON, BMPER, WDR26, SPINT1, ITGA5, KIAA1586, RUSC1, FRS2, RNF168, CTNNB1, MED8, MAPK1IP1L, CLMN, IFI16, C19orf55, SPRY1, FBXO41, DDX21, TKT, RNF123, PRMT2, ATP5G1, SLC43A2, ARPC2, CKB, TET2, PIGA, PDHB, KIAA0895, AGL, HIPK1, EIF5A2, TRIM44, RSPRY1, SUN1, RRM1, PEA15, GMPS, REL, TP53INP1, ACOT11, FBN1, CCDC127, ARHGAP35, RICTOR, ANXA5, PDP1, SETD5, WEE1, CCNF, SNRK, REEP3, KIAA1522, LZIC, CASP7, ANKRD11, E2F7, PTEN, LPL, PPIC, CALCB, GNG12, CUEDC1, PLAG1, SS18L1, C2orf68, SERPINB9, YPEL2, CCSER1, CYCS, SMARCC1, FRAT2, TNRC18, PDE7B, ZNF518B, SH3PXD2B, HNRNPF, SETD8, SOX12, BRD3, KIF5B, PTRF, CEP97, RNF152, DENND4A, CDCA4, PDE3A, MLXIP, ZNF507, GLUD2, HAS2, RELA, MACC1, ANO5, TMED9, CSNK1G1, CBX2, SRPR, KLHL28, NRIP1, ZNF318, FAM83H, MAATS1, MRFAP1, SLC26A11, JMJD1C, RNF150, ZNF160, THOP1, DENND5B, NFXL1, ZNF518A, SLC19A1, ZNF282, PPID, NBEA, PPM1D, HECTD4, IMP3, PRKRA, JAGN1, VCPIP1, TMEM70, ULK1, LUZP1, BCOR, SDC2, C5orf24, AKAP5, DTWD2, PITPNB, CCBE1, ZBTB21, THAP2, UBXN2A, KLHL15, JUN, TYMS, PRR14L, CD248, PURG, ZBTB41, CTXN1, NCKAP5, XPOT, PTPRM, BUB1, PTTG1IP, MED14, YOD1, C12orf76, SMAD2, GPHN, BRI3BP, SESTD1, DDX3X, IFI30, COL15A1, ANKRD13B, COL5A2, PURA, PCGF3, MAFB, MORF4L1, ZKSCAN4, GMFB, DDI2, ZFP91, PRPF40A, OSTC, RNF39, NRAS, VPS13A, LDOC1L, ZBTB10, HELZ, TET3, STK38L, PRMT6, ANKRD28, SOWAHC, MAFG, FITM2, BCL9L, DNM3, RAB12, KPNA4, TCF4, CTNND1, ZNF292, SPRY4, DHX16, CHM, SLC35B4, LPAR1, HTT, HN1, DCTN1, SRSF10, SCAMP5, ZNF682, ZNF155, LTN1, NOL4L, SIAH1, TPK1, CDPF1, ZNF512B, FNBP1, CCNI2, AFAP1, LAMA2, GIGYF2, PCDHA9, FAN1, COL4A1, GJB3, RUNDC1, C22orf29, ZNF652, CDC42SE1, HIST2H2BF, LAMP1, SLC39A10, R3HDM4, CHAMP1, ATG9A, BAG6, NFKBIL1, TOPORS, TTC30B, ITPRIPL2, TXNRD1, PPP1R10, SND1, S1PR3, WDHD1, RPS4X, ARID2, MICB, ZNF770, SPTAN1, NHLRC2, STYX, COL4A5, ZNF431, RFX6, BACE1, VDAC1, TMEM178B, C7orf73, ERCC6, MMP24, XK, RBM12, TMEM91, CCPG1, BFAR, ATXN7L3B, ISY1, RP11-407N17.3, Mmp24, U2SURP, FGG, HIST1H4A, NUMB, DCAF12, TXNDC16, ACACA, TRIL, DDOST, FGB, WNT4, PRKAB2, IGFBP1, PDHX, WSB2, FGA, ATP5J2, STK19, PSMA2, ZNF45, CEP68, MMP15, PPT1, TIMM44, NUDT3, OIP5, BCL2, MKL2, SOCS7, HUWE1, HIST1H4E, Col6a2, EIF6, AKT3, XKR7, Mmp15, GAN, Ctsk	

Table 7-7: List of experimentally validated targets of miR-141-3p from DIANA Tarbase

Target Gene	miRNA
FOXC1, RB1CC1, MTF2, TNKS2, RAB5A, NR6A1, PREPL, CCNG1, CBX5, PDZD8, TNPO1, RC3H2, REST, TMTC3, SEC24A, BRWD1, UBE2K, SNAP23, ATG2B, SFXN1, B4GALT6, RNF149, CCDC117, SCD, IMPAD1, QKI, NUP155, KCTD12, GATAD2B, AGPAT5, ANKRD52, POU4F1, JOSD1, UBE2Q2, ZBTB10, SRSF11, LDHA, ZNF503, ZNF606, DHX15, MRGBP, ZNF281, NCOA7, U2SURP, KIAA1598, SCRN1, BSG, ABCA13, SRCAP, EEF1D, FDPS,	miR-141-3p

ST3GAL6, KNSTRN, MAN1C1, RBM33, KCNIP2, NFIB, FUS, MAP7, NECAP1, HADHB, CTTN, FGFR1, FBXL17, DICER1, LENG8, PPFIA1, CCDC138, RP11-139J15.7, LMBR1, ZNF121, RBM15B, MAP7D1, BAZ2B, KNTC1, MTMR3, NUPL1, GRHL1, LATS1, GOPC, DDX1, RRM2, TNRC6A, AKAP12, RP11-159G9.5, DYNLL2, ARNTL2, UBE3C, LFNG, DDIT4, HIST1H2BD, UTP20, ZNF442, ATAD2, CMBL, DCTN5, MAST4, GARS, WDR87, HSPA1A, CDK13, DIP2B, EIF2AK2, FNIP2, SOX11, CYP26B1, CHD9, IPO5, ZBTB18, HIPK1, RAP2C, FAM114A2, RABGAP1, RNF11, AK2, DYRK2, OTUD3, C5orf24, MECP2, NCEH1, VAMP3, WDR37, RANBP9, APPBP2, KIF1C, C1GALT1, LY75, LYRM7, MBTD1, MTMR9, HNRNPA0, ADRB1, DSG2, KHNYN, EXPH5, PHTF2, DHX33, SKAP2, CCNE1, MLLT4, CEP57L1, ZHX3, ZNF697, GTPBP3, JKAMP, C6orf89, DNAH12, STRAP, METTL25, SLC39A9, EPN1, RNF4, LSM14B, AP2B1, HNRNPC, SLAIN1, MBTPS2, SLC7A2, ANLN, FGFR3, SLC25A3, RBBP6, KIAA0100, WNK1, TOX4, SFSWAP, TNRC6C, MGAT4A, DP2(hsa), MON2, SMO, SYNE2, MATR3, FUBP1, DENR, ITM2C, ASPM, ZNF805, ERGIC3, TRAF3IP2, SMC4, cyclinD1(hsa), API5, CALM2, VCL, NUP160, ELMO2, DEPDC1B, STK38, ETV3, ABCC1, ZMYND11, CCNL2, KIF1B, GCLC, FBXL5, PRPF4B, PHLPP2, POLR2B, SEMA6A, SAR1A, BRMS1L, TTC28, ULK2, CCNT2, FBXW11, CDC25B, MAPK6, ZNF416, RAB10, PDS5B, TFAP2C, SEC23A, PGR, PSEN1, SEH1L, TRNT1, SIX4, EIF5, ALG9, CRYBG3, ABL1, MEF2A, ZC3H14, SLMO2, ADNP, RBMS2, PTGS2, AGO1, PPP2R5C, CXCL2, SRBD1, GPR137B, CMTM6, VPS35, KLHL20, BTAF1, SESN1, FAT1, CRKL, TDRD1, CYLD, ATP8B1, EXOC5, MAP4K4, EXOC1, ECHDC1, FAP, TRAM1, HSP90AB1, FOSL2, CHD8, KLF6, PICALM, ATG16L1, SRRT, NEDD4, PCM1, ANKRD10, USP33, ITGA6, OSBPL8, STARD7, THUMPD1, CDC42, KDM5A, TCF20, NFKBIA, TMEM131, IKBKAP, TFRC, MSH2, MMP9, CSTF1, ATP1B3, RAE1, HMGXB4, MMP2, CNIH1, MAVS, HECTD1, ZXDC, UNG, RASSF2, HNRNPH3, MAP3K1, PABPC4, RCOR1, PAPOLA, ITCH, HSP90AA1, BAX, DIDO1, STK4, NDST1, DUSP3, E2F3, CXCL12, GLCCI1, GLS, NUP98, KLF5, C3orf14, HMGCS1, INO80D, NR3C1, FBXW7, PGK1, RHEB, PIKFYVE, LPIN2, GABPB1, STRN, TCERG1, SLC25A36, ZFAND5, PVRL1, LIFR, GATA3, HPS5, RNF7, RAB3D, KLF3, XRN1, H2AFV, MED13, GPR126, MET, PTP4A1, OSBP, VAPA, COPZ1, CPSF6, BIRC6, SERINC1, USP31, LUC7L3, TUSC3, SUB1, DHX40, RAB11A, MAPK14, ARRDC3, GGA2, CCSER2, GPI, UBL4A, CSNK2A1, CCND1, LAPTM4B, SPTBN1, DDX6, FNDC3A, TSPAN14, TAF1B, PLEKHA8, FOXP1, STK3, PTGES3, CPEB3, PRLR, CSF3, MRPS31, SETX, SGK3, VEGFA, ENPP5, PRKAR1A, SLC16A6, ASCC3, SLC11A2, GALNT3, TMC5, PLEKHB2, FAM208B, KIAA1432, FAF2, GTF3C2, SLC7A5, CAV2, SH3D19, TRPS1, MTPAP, VAC14, GAPDH, ORC2, GOSR1, CBL, ITGB8, ATP6V0E1, GALNT7, ERBB2IP, RTN4, MYH10, KMT2A, FOXA1, GNA13, HOXC13, CCND2, DEK, MNX1, ANKRD17, PANK3, SSR1, IRF6, LPGAT1, TMX4, S100PBP, ASH1L, CAMSAP2, MED13L, CCNT1, CANX, SUMO1, FAM199X, SFPQ, SNX6, RPL23, TARDBP, ILF3, TTPAL, PHF13, RFX1, DESI2, RAB22A, HOXB5, PRKAA1, NAA50, HNRNPH2, ACSL3, CENPF, FAM63B, FBXL16, H3F3B, AHNAK, EPAS1, SMUG1, JUND, KTN1, AKAP1, IVD, DKC1, ALG2, CRYZ, PPP6C, USP22, RNF170, PMEPA1, EIF2AK4, MORF4L2, DCAF4, RBM38, NUP153, ZNF557, FBXW2, CALU, NEK6, AJUBA, RAB14, PARD6B, YPEL5, KIF17, TRIM25, ERRFI1, RRBP1, TGIF2, USP9X, CDKN1A, FOXO3, RIN2, MPRIP, SYNE1, NOL11, PANK2, RAB32, ATL2, PAIP2, MXD4, SLC19A2, TOP2A, KLHDC10, PDHA1, NPC2, SH3BP4, AMD1, VPS13B, FYTTD1, NCOA3, 42985, SRM, CDK2, TMEM189-UBE2V1, ABL2, PTPRG, RANBP6, GATA6, EPHA2, 42801, EDEM1, NCOA2, SLC20A1, RPRD1A, TRA2B, IREB2, RAPGEF5, IER3IP1, G3BP2, COG3, HNRNPD, ZCCHC14, ABHD10, SPOPL, G3BP1, PPHLN1, MCL1, OGT, IFNAR1, NAA30, APH1B, RNF44, UBXN4, RAC1, KSR1, SBNO1, LMO4, TMOD3, IGF1R, RNF144B, EHF, SRSF1, RERE, ARRDC4, VAV3, AMIGO2, TPBG, THBS1, IQSEC1, SNX27, BRD4, SETD1B, GREB1L, LPP, DCAF7, SETDB1, CCNA2, WTAP, USP53, SERF2, DENND4C, FST, NONO, HMG20A, FNBP1L, RNF185, ANP32E, USP6NL, EPT1, CYP1B1, ZMYM4, KDELR2, SEC11A, CLOCK, TMEM60, NOB1, RHOB, PHIP, DUSP5, FGD4, IL6ST, TNS3, SCAF11, ACP2, GLCE, PIK3R1, DCTN3, TBC1D4, TMEM66, GIGYF1, LBR, ZEB1, ZFHX3, ATP1B1, DSTYK, ARNT, ACVR1B, SETD7, HNRNPA1, IRAK1BP1, RNF38, GNA12, EPRS, NFKBIZ, FAM168B, ZNF385D, FAM160B1, DLC1, OSGIN2, HSPA13, KIAA1430, OTUD4, NFIA, 42986, H2AFZ, ZNF148, ARPC5, UBXN7, CELF1, HEATR2, LMTK2, C16orf87, ARID5B, AGPAT6, UBE2Z, RMND5A, RHOBTB3, VMA21, IL6R, C3orf17, ANKRD50, RAD21, CMPK1, SMG1, TAOK1, GNAQ, LY6K, RASA2, DSCR3, RNF20, EOGT, FMN2, RPRD2, SLC30A6, RAPGEF6, SNIP1, NPY1R, LRRC58, C2CD2, WDR82, REL, ATP5G3, SMIM14, TMED4, SLC25A46, PFKM, UHMK1,	

U2AF1, STPG2, HNRNPDL, PLCL2, SPATA2, PPP1R21, SCOC, WDFY3, UBN2, COA7, CDC25A, SORBS2, BAP1, FOXO1, AGL, MBNL1, TRAPPC10, SUN1, MKI67, JAZF1, EI24, PPP4R1, UBAP1, COG5, SON, CGGBP1, ADPGK, SOGA1, ASAP1, DCBLD1, RICTOR, DCAF13, NAA15, MTMR12, RNF111, PTPRD, TCF7L1, ADAR, EXTL2, SAMD8, SHC1, ENDOD1, LPCAT1, ZNF12, CLASP2, SLX4IP, PLAG1, BRD3, ZBTB34, REEP3, DCP2, YWHAG, CHD2, PRKCE, PPM1E, B3GNT2, BBS10, PTEN, KBTBD6, MSL2, RAB8B, YOD1, EPM2AIP1, AKIRIN1, TMEM192, HSPA4, SNRPD1, FAM220A, CTNNB1, GRB2, KLHL15, TBL1XR1, SPRYD3, RNF139, PDIK1L, ZNF318, TMEM70, ARL5B, GNG12, EIF3F, TMEM154, ZNF621, HOXC10, ZNF217, CDK12, VCPIP1, OR1L3, HECTD4, RAB3B, PAFAH1B2, FEM1B, KMT2D, RNF213, XPO6, NFRKB, MTMR10, ORMDL3, ZNF791, HNRNPF, KLHDC2, TRMT112, PCGF5, PWWP2A, ASXL1, ANO6, ATMIN, ZNF592, TMEM41B, ADAM9, ZNF518A, SLC19A1, BEND3, ZNF668, SMCR8, ZNF609, KIF5B, RPL36AL, FAM102A, CLK3, TAF7, RBPJ, DOLK, IRS1, CFL2, CCNE2, FBXO34, LEMD3, CTXN1, CDC40, DPAGT1, KLF11, SP3, CASC4, BCL2, OAZ2, FTH1, RPSA, CRK, KBTBD2, BTRC, ADRBK1, CDH2, ZEB2, LRRN1, TMEM170B, ZNF248, MACC1, BICD2, FAM46C, TCF24, PPT2, IBA57, ANKFY1, LHFP, ZSCAN30, SLC6A9, TMEM57, TMED9, ACADSB, PPP2R2A, PPP3R1, CD2AP, NAP1L4, CUX1, NHLRC2, ZDHHC17, ANAPC7, ERO1L, ENTPD4, TRIM33, BRI3BP, VKORC1L1, PURA, TTC30A, SMIM13, NCR3LG1, NME1, METTL9, EXT1, TOP1, DYNC1H1, TTC3, CDCA2, MVB12B, ATL3, TTC37, KIAA1147, OCLN, HNRNPAB, STK38L, HTT, LMLN, ZNF891, ESRRG, PTCH1, PHF2, ZNF793, TAF15, ZNF736, NAP1L1, SLC6A14, SIAH1, PSG5, LRCH3, ZNF512B, YTHDF3, ZNF397, DMWD, KPNA2, FMN1, NF2, PHB2, TDRKH, METTL7A, CACNA1H, ZCCHC3, ADH5, ABCF1, RBM14, HEXA, SPTSSB, SLC35F6, ZNF335, CDC42SE1, BCOR, NOP9, USP7, SMYD3, EMP2, C16orf72, TXNRD1, PPARA, SF3B3, PRR14L, SIPA1L1, SPIN4, ZKSCAN8, ASB13, TUBB, SRGAP1, GAN, LGR4, TET3, ZBTB44	

Table 7-8: List of experimentally validated targets of miR-195-5p from DIANA Tarbase

Target Gene	miRNA
CCNE1, WEE1, FBXW7, IPO7, E2F3, SKI, AGO1, SON, CDC27, CDC42SE2, NUFIP2, SLC25A22, ZNF367, TMEM245, ANKRD52, NRBP1, PRKAB2, EIF3A, EIF5B, HSPA1B, SIK1, SUCLA2, FAM199X, PPRC1, TMEM8A, KPNB1, AKAP10, CAMSAP2, NEMF, OTUD3, ILF3, HIAT1, OGT, DYRK2, YWHAG, TBC1D13, B4GALNT3, ISG20L2, TAF7, PSMC2, TMEM127, SEC23A, CUX1, CDC14A, ATP2A2, C1orf85, KTN1, DCP1A, VIM, ALDOA, ANKS6, DHRS13, PHC3, DNAJC5, MARCKS, NFE2L1, SNRPB, VEGFA, RBM25, RBM4, CSDE1, RHBDD3, FASTK, LY6E, ATP2B1, FAM73A, MGA, MDM2, ZBTB47, PIAS1, RBM14, SUGT1, FAT1, ATG9A, TNRC6A, PRKAR1A, MAT2A, DYNLL2, CLNS1A, ASCC3, TMCC1, SFPQ, APLP2, ASPM, DUS1L, LRRC59, PRRC2C, NAB1, GIGYF1, APP, PDIA6, ZFHX3, FZD6, HSP90AA1, TAOK2, SMURF2, PLAGL2, TM7SF2, FAM189B, PDCD4, PRDM4, SLIT2, CACUL1, ARL2, CHAC1, CCND2, CBX4, RBMS1, YRDC, USP14, PAFAH1B1, DLL1, SESN1, BCL2L2, ATP13A3, PSAT1, MIB1, SPRTN, RIF1, ARHGDIA, SMIM13, NSD1, ENTPD4, MLK4, SESN3, TMEM189-UBE2V1, ZHX1, STK35, B4GALT1, TBL1XR1, CCNT1, RPS14, CCNG2, CCDC47, NUP160, GFPT1, C2orf69, CDC7, TMEM181, TM9SF2, PHF12, EDC3, CDS1, HBP1, CCDC59, SEC61A2, PLA2G12A, UST, SDC4, AGPAT2, UTP15, TCP1, SYMPK, CLTC, RPS25, TNPO2, RPL19, EPN1, BCL2L12, DDOST, UHMK1, PRRC2B, CHD4, PXYLP1, MICA, NAT8L, TMUB1, SLC29A2, SCARF2, FBXL19, KIAA0100, PILRB, SALL1, TNRC6B, EHMT1, AP000721.4, GATAD2A, DDX1, DLD, LY6G5B, TMEM121, MIDN, SRRM2, SACS, NPEPPS, XPO6, LBR, PTOV1, ZNF224, INPP5E, MCL1, PPP1CC, TMEM161A, HIST1H1C, SMAD2, SCAMP3, MSI2, ACTB, ARPP19, SPEN, CBX6, BTG2, TAOK1, FBXO3, TTPAL, IPO9, 42799, ZNF264, PPP1R14C, CALM1, WDR82, ANKRD28, CBX5, SP1, NUCKS1, NDST1, TLN1, LAMTOR1, PLXNA1, ZFR, MIA3, SOGA1, LARP1, PLEKHG4B, TRIM44, RPL27A, FITM1, DCBLD1, KIAA1328, SLC12A7, LRRC47, ANKH, KIAA1467, NKX3-1, NOTCH3, RAI1, SPIRE2, ENGASE, INPP4A, SURF4, SLC4A2, DYRK1B, TRAP1, TNK2, ISYNA1, NME4, PGLYRP2, LTB4R2, VGLL4, HSPBAP1, HNRNPDL, MBIP, ARFGAP3, PIGO, TPH2, PMPCA, ACTG1, DICER1, LENG8, STK11, PML, HINT1, POLR3H, FRMPD3, TNKS1BP1, TGIF1,	miR-195-5p

COPS7A, OTUB1, PCM1, SLC39A1, ROGDI, GALNT2, CKAP4, THTPA, TCF20, IRS4, CSNK1A1, SRGAP3, PFN2, AEBP2, RASA1, PIGS, PPP1R11, WBP1, POLR2L, TNKS2, SQLE, NUP98, GAN, GTF3C4, FTCD, ZNFX1, ELAVL2, FRRS1L, CLASP2, EPB41L2, SIPA1L3, SLC9A2, SLC20A2, RECK, KIAA1432, SOBP, KIF5B, EZH1, IGF1R, TBPL1, ACSL3, CCND1, SKIL, KBTBD2, CELF1, HDGF, GBF1, SEC22C, CANX, ALAD, ZBTB33, IPO8, GET4, PVRL3, ADIPOR1, ATXN1L, UBE3C, VMA21, PTPN12, STRIP1, RABL6, RBM12B, SLC6A8, MAFG, TIMM13, SLFN11, TTYH3, DDB1, ZNF580, PLXNB2, STK33, HNRNPU, EEF1D, BAI2, AGRN, CIRBP, RGL4, DHRS2, EMD, OGFR, CAMTA2, PTBP3, UCKL1, NLGN2, SLC25A3, KLC1, NME3, C19orf48, CALU, ABCF1, TMBIM6, CENPN, KMT2C, XPO1, LSM14B, TMEM134, CDKN1A, CALR, USP42, RARS, FAM118A, EIF4G1, TMED10, PPP2R1A, RBM19, ADAM23, REPIN1, HNRNPUL2, POLR2D, MLLT4, FAM46A, HECTD1, SORBS3, H2AFV, AZIN1, CERCAM, UNG, AMOT, ACBD5, RFK, U2AF2, CCT4, TECTA, HIST1H4C, ACLY, PTTG1IP, GUSB, POLR1C, FAM120A, ZBTB34, KDSR, ZBTB39, MTMR3, N4BP1, BTAF1, SPRED1, CHD6, PPAPDC2, HSDL2, ARL5B, PWWP2B, MYO5A, PHLDA3, TMEM199, TSN, DCTN5, TBC1D9, IRF2BP1, GATAD2B, BSDC1, SEC61A1, STAT3, GPR157, MRPS25, NUP210, DCP2, CPXM2, RP6-24A23.6, TCFL5, CYP51A1, BRWD1, FAM160B1, TNFAIP1, SCML1, AGPAT3, RRAGD, KPNA5, MLEC, CRAMP1L, PPP2CA, C3orf62, CAMTA1, GPC4, ARHGEF11, CYLD, CPEB2, LRRC41, GDPD5, ABCA13, CLCA1, MARK1, MBD3, SCAMP5, VDAC3, TNFRSF12A, POLR2A, C3orf67, CEP170, TMEM258, ANLN, ZC3HC1, KIF5C, NONO, PCDH9, DNAH14, UBAP1, OTUD7B, CNOT1, CKB, JOSD2, TUSC2, ADCY6, KCTD20, SRRM1, TOP1, MYL12A, PIDD1, ZIC5, UBR3, BAG6, NAE1, NPPC, PNPLA4, EIF4A2, MAPRE3, ACTR3, SYAP1, CCDC127, ZSWIM6, ATP6V1G1, TAB2, ANKRD23, NTN1, DOCK6, COL4A5, IFNGR2, KIAA1522, GPR137B, TENC1, WIPI2, SLC2A3, CDCA4, AKAP11, KIF1B, GNAI3, HIPK2, WAPAL, ASB1, ANKIB1, ABCF2, MAPK9, PI4K2B, MAP3K9, EIF2AK2, TBC1D20, ADSS, HGF, UHRF1BP1, HSPA5, RC3H2, CREBBP, QSER1, DHX33, MTMR1, KIF2A, RRM2B, TTC17, RAI14, FECH, ARF3, VMP1, ELOVL5, CUL1, ALDH18A1, LAMC2, VEZT, JKAMP, PIGV, KARS, ZC3H11A, SPPL2B, NATD1, PSD, DDX21, ALS2, AP2B1, RNH1, ELMO2, TRAM1, FUCA2, GLI2, NTMT1, REV3L, CFLAR, SPAST, MAP4, MDH1, CD44, TNPO3, RANBP9, TMEM206, FAM214A, LTBP1, MON2, EIF4G3, MATR3, NDC1, PUM2, LYPLA2, PHKA1, TNFRSF10B, CTPS2, TBC1D25, RAD51C, GDE1, NCAPD2, CDH1, UBA6, LAPTM4A, CHPF2, cyclinD1(hsa), HIVEP2, TFPI, DCBLD2, NSUN2, PKM, LSG1, PNPLA6, COASY, ZMYND11, HEATR6, TMEM59, VTA1, MAST4, TMEM159, YIPF2, ACSL4, PHLPP2, DESI1, PTPN4, CCNT2, PTPN3, RSBN1, BCAP29, MNT, RAB10, BZW1, PAG1, KHNYN, PPP2R5C, FNDC3B, VTI1B, PIAS2, RBFOX2, ARCN1, IL2RB, SLC44A1, TNPO1, ZNF532, PDS5B, PRKACA, CRKL, TCF3, ZFAND6, DNAJA1, BIK, ERMP1, SUPT16H, TRAF4, SPTLC2, OGFOD1, CAD, MAVS, CDV3, C12orf5, RBL1, LMAN1, ST6GALNAC2, GOLGA3, TM9SF3, SCFD1, COMT, TRAF5, BAZ2A, SRCAP, RAB18, ECHDC1, ACOX3, HSP90AB1, FOSL2, SMEK1, EPS15, DNM1L, PREP, KIFAP3, RPS6KA5, FRYL, TOX4, ZW10, IFT74, ABCB1, HUWE1, TNRC6C, NIN, NFATC3, MKNK2, LMF2, NSF, BAMBI, UBE2D1, EP300, PIGB, CLASP1, CDC42, ZMPSTE24, BLVRB, NLRP1, THOC5, TXLNA, RBM7, CNIH1, HADHA, ICAM1, RTCB, MTMR2, PABPC4, RCOR1, RBM22, BIRC5, GPATCH2, RDH11, TMED2, UBE2T, KAT6A, ATXN7L3, PSMD5, SMAD7, CDC37L1, PVRL1, WSB1, KCNN4, CAND1, BFAR, KATNAL1, CDK6, MMD, COL12A1, UBFD1, GALNT7, USP31, PLA2G15, TRIM35, CLCN3, MDN1, CPSF6, MON1B, CDS2, SMC3, TSC22D1, TWISTNB, MAPK14, ARMC1, RBL2, STK38, ABHD17B, ADNP2, CA2, COQ5, UBE2R2, CDIPT, NKAP, PDHX, PPP2CB, DMXL2, GATA3, KCTD10, GOSR1, SH3D19, KLF5, MMP9, SLC1A5, DNM1, TBC1D17, CDK2AP1, TRIM37, RNF43, DDX6, RAPGEF1, ERICH1, FNDC3A, ARHGEF18, CCNC, METTL4, KIAA1549L, IKBKB, FRK, LARP4B, ANKRD12, CPEB3, XIAP, EIF4G2, SLC39A14, SETX, FNBP4, AIP, MAN1A1, CA9, PLEKHA1, SEC23IP, PSMD7, PGRMC1, ABCC3, DHX15, DDX5, TLE4, SNX16, C11orf58, SLC11A2, RABAC1, DNAJA3, MAPK8, MGP, PNN, LFNG, CPT1A, KRT18, RHOV, SLC7A5, CHPT1, KIAA0391, DSE, TRPS1, PIM2, SLC38A1, ERLIN1, ASH1L, ACVR2A, LY9, ELL2, CCND3, STXBP3, PARD6B, PRKAR2A, CPEB4, SENP5, BVES, ODC1, RAD23B, HELLS, SLC35F5, KPNA1, KLF7, VAMP7, IRF2BPL, ARFGEF2, MRPS2, AIM1, QSOX1, PLEKHA3, PLEKHB2, HEATR1, ATP6V0B, RAB14, POF1B, CNOT6, CHPF, IRF1, ST3GAL5, FLVCR2, PDS5A, KIF14, CHMP3, TTK, PIK3CA, NUDCD1, HSPA9, FAM98A, MYB, ANXA11, TBP, MFAP2, UBE3A, DNAJC16, KLHDC3, ACO1, TFG, RREB1, QKI, ZBTB45, SF3B1, DIEXF, RBBP6, ABCC10, NUP153, SLC25A12,	

OBSL1, FBXW2, CREB1, PCMT1, AUP1, RPA2, SGK1, PLAU, ITPR2, TMEM214, ABCC5, TGIF2, AKR1A1, SLC1A4, AKT3, PEPD, MOB4, DARS, GALNT3, CD46, NR3C1, STX16, TMTC4, RCAN3, RAB29, HES1, EPC1, PTCH2, PHF3, BYSL, MAPKAP1, GNA13, EFEMP1, PLEKHM2, ATF2, AREL1, SPARC, IGFBP5, RTN4, GOLGA1, UBE4B, RNF138, XPO7, POLR3F, ACTR2, CAPRIN1, PPP2R1B, NAPG, ANKRD17, SLC41A2, AGO4, ZFYVE20, ARL8B, SOX5, LATS1, PPT1, CAB39, UNC13A, ADAM10, CD164, LRRC1, ERN2, ARMCX1, ANAPC13, TNFSF9, DCTN4, HS3ST3B1, MOSPD2, RRAS2, SMYD5, KIAA1468, BTBD3, COG3, GRSF1, TTBK2, EPB41L4A, REEP5, IREB2, EIF2S3, TFCP2, COPB1, ANKEF1, STT3A, SERINC3, FADS2, VTCN1, MYO5C, EIF5A, NCLN, HSD17B4, REPS1, HMGA1, HEATR5A, SCRN1, YAP1, DCAF7, ANKRD36, MBOAT7, EMC4, ARHGAP32, MYO1G, CIDEB, UBAP2, NSRP1, MAP7, FHDC1, VPS36, ZCCHC11, ZBTB46, SMARCA4, ARPC1B, MACF1, PER2, ROMO1, KDELR2, UBR4, KIAA0368, IL6ST, STARD13, ACP2, CDC73, TMEM2, DMTF1, COL5A1, MYC, PUM1, NAV1, MED1, VPS13B, 42801, VHL, KIF23, PRDX5, LMO7, SBF2, TMED7, CDK4, PLAA, BECN1, FGF2, SREK1, MBNL2, CACNA2D1, ABL2, ZNF697, GNA12, NUPL1, PPIG, ABI2, ARRDC4, DOPEY2, DENR, LRPPRC, TMEM87B, GALNT1, CREG1, ASNSD1, FREM2, EPG5, ZNF827, GUCD1, PHIP, ARID5B, TSPAN3, AMMECR1L, TMTC3, ERLIN2, LRIG1, EML4, G3BP1, YTHDF1, HIATL1, BARD1, WNT9A, SNX30, SNRNP200, TMOD3, SECISBP2L, HNRNPLL, MUT, MFAP1, QTRTD1, IGFBP3, SETD1B, CNKSR3, USO1, CCNA2, WTAP, KANSL2, C11orf49, FBN2, PAM, NFIB, SHPRH, AIG1, SLC15A4, RNF217, LRRK1, KDM6A, TUT1, RMND5A, NACC2, BCL2L11, ASXL2, SCARB2, CENPU, EIF2B5, RHBDD1, CCDC102B, KIF21A, ATP8B2, FGD4, PDGFC, SLC20A1, NARF, GLCE, TEX261, TGOLN2, RBM26, DDX46, CBLC, PARP16, AHCTF1, ASAP1, HSPG2, ACP1, TMEM123, DST, ZCCHC2, MSN, MDH2, PTPRF, MIPOL1, TCF12, COPS7B, TACC1, SCAMP2, ZCRB1, WNT3A, IHH, BTRC, DNAJB4, TMEM183A, TMEM55B, OTUD4, ATP5G3, PSKH1, ENAH, AMOTL1, SLC30A8, PALM2-AKAP2, RRAGA, FRS2, C9orf91, ACOX1, RNF149, FOXK1, GNAQ, TP53INP1, FSTL1, FAM160A1, NPTN, TMEM55A, SLC30A7, XKR8, PAFAH2, ZER1, SGPP2, HGSNAT, SQSTM1, RNF168, KIAA1161, PRKCD, TAB3, DVL3, CXCL5, PRRC1, LARP4, UBLCP1, CLMN, RPRD2, RUSC1, HEATR2, HPS3, SLC18A2, U2SURP, FAM126B, ZDHHC12, SUCLG1, PEX19, MGAT4B, SHISA5, SHANK2, GART, BAG5, FYCO1, GYLTL1B, ZFYVE1, PPM1K, RACGAP1, SPATA2, PPP1R21, TPRG1L, UBN2, CDC25A, ARHGAP12, RAB11FIP1, CLDN12, DYRK1A, HK1, ZNF326, ZNF281, C2orf47, POLR3K, FAM208A, VCP, GPBP1L1, ELMSAN1, TATDN2, CGGBP1, CREBRF, POLR2H, PBXIP1, C1orf52, RICTOR, ANXA5, GRAMD3, CXXC1, CNST, VCAM1, SNRK, BACH1, UVSSA, DUSP2, SCAF4, PINK1, CDK5R1, SOCS6, BCL2, ZNF622, HECTD4, PAFAH1B2, ZMAT3, SPRYD3, RPS6KA3, TUBA1A, OSCAR, RNF213, CHD2, VPS37C, SNTB2, SNCG, KMT2D, ATG13, CTDSP2, GRAMD2, CHD9, KLHL26, CCNE2, TMC8, VCPIP1, CLPX, CEP83, ZBTB5, TMEM43, FAM111A, TADA2B, SERPINB9, TOMM20, CTNNB1, ATF7, KCMF1, RFWD3, PRR15L, HSPA4, IRF2, S1PR1, RBM4B, FEM1B, SH3BP5L, DHX36, TMX3, NR2C2, 42980, BDKRB2, ORMDL3, DNAJC24, CHD7, ASXL1, JMJD1C, PPFIBP2, IP6K1, SLC38A9, GNB2, BNC2, TMED3, SPSB4, ZNF146, WIBG, ENDOV, EFNA1, SMCR8, MRPL1, KIAA1551, NBEA, PDP2, HDHD2, FN3K, CNBP, YES1, GPR25, KIAA0232, GOLGB1, BRD3, APBB1, ZBTB21, TAPT1, SEMA4C, UQCRH, RNF34, PBX3, NAA20, RNF187, ALCAM, SETD5, YWHAB, SP3, IRF2BP2, NET1, CD300A, CRK, C12orf76, ADORA2B, CXCL8, SERTAD3, PLAG1, PAPPA, CD2AP, SRPR, FAM110C, MAFK, PURA, NHLRC2, ST8SIA3, CAPZA2, STYX, OOEP, SH3BGRL2, CMTM4, ZNRF3, C16orf72, VPS13A, ZDHHC23, ARMCX2, NDNL2, MME, HTT, XPNPEP3, CTNNBIP1, SESTD1, BRI3BP, GRB2, NCOR2, NPM1, PSAP, SMDT1, TBL3, BET1L, NRIP1, EDARADD, TTC3, ANXA4, LDOC1L, METTL9, DDI2, CCSER1, DYNC1H1, DDX28, GREB1, MFSD5, TTC30B, HEXIM1, SVIP, NOC2L, MYADM, ZNF43, ZNF181, MAGED1, CLDN4, LRBA, ZNF627, ZC3H6, NOL4L, ENTPD6, PPP1R2, HLA-DQA1, WDSUB1, SMTN, RAB40C, SERPINA1, BTBD9, MBP, PDE4DIP, SUMO3, RIC8A, AGPAT1, MKL2, PDGFA, SVIL, TACSTD2, PRKRA, TRRAP, ZBTB14, TLK1, NF1, TSC22D2, FAM179A, TRIM33, KANK2, COX20, CXorf40A, MSL1, ARID2, C14orf119, SPTY2D1, ATOX1, CD151, NDUFAF3, CRIPAK, PRPF40A, IST1, GPAA1, FITM2, C5orf30	

Table 7-9: List of experimentally validated targets of miR-486-5p from DIANA Tarbase

Target Gene	miRNA
ST6GALNAC6, ZNF367, GXYLT1, G3BP1, PITPNA, H3F3B, CPSF2, RELT, ARL6IP1, AGTPBP1, FAM13B, BRWD1, DDX3X, MET, MTF2, MDN1, PLCG1, DPP9, FGFR1OP, KLHL20, CLPTM1, BCMO1, CCDC57, VEZT, CLCC1, UROS, VIM, ZNF791, SLC7A6, CERS1, RER1, MSH6, UBE3B, DCUN1D2, ATP6AP1, SLC35A3, KIF5C, FYCO1, RALGPS1, NME3, GDF2, WDR6, TMEM130, GPSM1, MAP4, TMF1, MYO1B, CDC123, RPS21, AP1B1, XPO1, NOP9, LDLR, HIGD2A, MLXIP, SH2D6, ROBO1, ZNF326, HNRNPH1, MRE11A, CELSR3, PUM2, CDK17, DESI1, ZNF451, CASP9, GLCE, LDHA, SRRM2, ZFAND6, MYPOP, ARID1A, GPBP1L1, SPEN, SMAD4, ZNF711, STX16, C6orf99, KLHDC10, COPB2, CRAMP1L, C15orf39, SEPT7, MTMR2, ETF1, MSL2, PHF3, RAB21, TIAM1, NT5DC3, APH1A, HBP1, BCL2, GAN, LMBRD2, ZDHHC20, FEM1B, RCC1, ARGLU1, ACTB, SRSF3, BTAF1, PLAGL2, TANC1, CREBL2, EMP1, PYCR2, ITGB8, TFAM, G3BP2, BIRC6, COX16, KLHL42, SPRTN, NR3C1, SYPL1, CDC23, KMT2A, ZNF800, GLIPR1, FAF2, MAP1B, GID8, TERF2, VPS4B, LRRC16A, MTR, NCOA2, SLAIN2, EEA1, LRRFIP1, BTN3A3, DIS3, TNFRSF10B, ZC3H14, METTL20, C14orf105, MAPK14, SPHK2, BRD4, PCNXL4, SIRT1, CLIP1, USP40, C20orf24, SMC2, GDA, NR4A1, KCNAB2, ITPK1, FAM65A, FBXW2, SYT16, DSP, NCALD, RPS6KB1, ECE1, SLC39A14, KIAA0141, COLGALT1, SETD1A, MKNK2, BTBD7, NCBP1, BNIP3L, KIFC3, ABLIM1, ZNF644, CD4, ARID4B, MRPS27, SERPINB1, RPL28, C6orf62, CHAC1, RSRC2, TTK, SLIRP, PPT1, C2CD5, FAM208B, NAV1, SEC63, PRRC2C, SAMD1, PRCC, ADIPOR2, TAB2, ENSA, CDCA8, NPC1, ZNF275, PHGDH, NUCKS1, ATXN7L3, VASH2, SAR1A, TBCD, MLK4	miR-486-5p

Table 7-10: List of experimentally validated targets of miR-29c-3p from miRTarBase

Target Gene	miRNA
COL3A1, COL1A1, TDG, SPARC, LAMC1, COL15A1, COL4A2, COL4A1, COL1A2, GAPDH, DNMT3B, DNMT3A, CDK6, MCL1, BCL2, FBN1, SRSF10, CD276, MMP15, MMP24, FGG, FGB, FGA, COL7A1, COL21A1, TFAP2C, WNT4, IGFBP1, PPP1R13B, CDC42, MKL2, C12orf76, NUMB, ZNF45, SH3GLB1, IRF2BP2, WSB2, PURG, CDV3, JUN, TIPRL, OIP5, SCAF11, GNB4, R3HDM4, ADO, FAM102B, KCTD15, U2SURP, KLHL28, BFAR, CDC42SE1, PPP1R14C, HMGCR, SOWAHC, TXNDC16, ELMSAN1, IRGQ, PDHX, CEP68, NKIRAS2, VEGFA, ZBTB10, AKT3, SPIN4, KIAA0895, CDC23, DDX21, BCL11A, CCNT2, DICER1, FREM2, TMTC3, AMER1, KIAA1549, DCAF12, ANKRD13C, HDGF, MORF4L1, PLAG1, RAB30, KMT5C, OTULIN, PRKAB2, KDM5B, BACH2, RNF19A, MYCN, BACE1, MMS19, C15orf52, MTHFD2, CTC1, XK, TIMM44, PSMA2, CHIC2, HUWE1, STK19, PPT1, BCKDHA, PPM1D, FAM126B, MMP2, SIRT1, RCC2, CCND2, PTEN, AKT2, CTNND1, PER1, TMEM132A, COL6A2, BTG2, RNF138, CALM3, INSIG1, TET3, VHL, TET2, ABCE1, REST, GOLGA7, MAZ, TGIF2, ZFP91, MXD1, ULBP2, CRYBG1, LAMC2, ITGA6, COL5A2, COL10A1, SERPINH1, LOX, PDGFRB, PHACTR2, TUBB2A, EMP1, SNX24, AMFR, RIOK3, WDR26, DSC2, CD274, TIAM1, NEDD9, MORF4L2, RPL22, TRAM2, SLC2A14, SLC29A2, SLC16A1, SGK1, REL, RAB40C, NAA40, P3H1, KLHDC3, FRK, FOS, DUSP2, CSRNP2, COMMD2, CCNA2, CCDC117, CAND1, BBC3, C1QTNF6, RAET1L, FSCN1, SLC7A5P2, PPY, TMEM237, GLDN, RTL6, GAS2L3, MDM2, DYNLT1, CASP8, HECW1, FJX1, ENPP2, DDX6, C21orf91, PTP4A1, ISG20L2, FEM1B, WWTR1, ZBTB5, LIMS1, SPRTN, CBX6, PIGS, COX7A2L, ENTPD1, RAB11FIP1, TESPA1, YY2, ZNF286A, YAE1D1, EPHX2, MRM3, ZNF850, SEC31A, TRIM72, ZFPM1, PRY2, PRY, FBRS, CTNNBIP1, CNBP, CBX2, ASXL2, OTUD4, FAM71F2, COLEC10, IFRD1, MAPKBP1, FAM193A, ITGB1, CREB5, SP1, ADAM12, KLF4, NASP, TARBP1, RFX7, PHLDB2, FRAT2, LRP6, FZD4, FZD5, CNOT6, AGAP1, C4orf26, DENND6A, DOT1L, EREG, FAM53C, FMNL3, GTDC1, ID3, KDM6B, RHBDD1, TNRC18, EDC3, GLRX3, SLC30A10, SURF2, UBE2Q1, METTL15, BMT2, ZBTB34	miR-29c-3p

Table 7-11: List of experimentally validated targets of miR-141-3p from miRTarBase

Target Gene	miRNA
ZEB2, ZEB1, DLX5, BAP1, KLF5, STK3, TGFB2, SFPQ, CLOCK, BRD3, UBAP1, PTEN, ZFPM2, TRAPPC2B, EIF4E, CTBP2, CDYL, ACVR2B, MAPK14, PPARA, NR0B2, YWHAG, ELAVL4, MAPK9, TFDP2, E2F3, SHC1, VAC14, TCF7L1, ELMO2, RASSF2, KLHL20, RIN2, SEPT7, HOXB5, ERBIN, KLF11, PTPRD, WDR37, STAT4, YAP1, CDC25C, HDGF, KEAP1, TAZ, TIAM1, TM4SF1, CDC25A, PHLPP1, PHLPP2, PPP1R15B, RAB8B, MED13, QKI, ADNP2, ARL5B, TNKS2, SPRYD4, TET3, OGT, PGK1, ZBTB34, HNRNPAB, IPO5, TET1, MAP4K4, HNRNPD, ZMPSTE24, STXBP2, SEPT8, DPY19L1, CHML, NDST4, TRAM1, YRDC, PSMD11, MCL1, CYP1B1, BICD2, PRELID2, USP53, HNRNPF, GNA13, FUT11, EPHA2, TMOD3, MARCH6, KIAA1147, ZNF805, ZMAT3, MALT1, KIAA1549, ELAVL2, ATXN7L1, PIGW, GATA6, EPHA7, C18orf25, SLC35D1, PGAM4, H2AFZ, TRMT112, XPOT, CDV3, KIAA1549L, LHX1, YME1L1, APOOL, ZNF621, ANGPTL7, PRKAA2, MACC1, CCDC18, DNAJC28, ZNF292, RPL12, A1BG, POLR3F, COX6B1, RBM28, LPP, CCDC71L, TROVE2, BICRAL, ERO1A, HSPA4, SCD5, CLDND1, PEX11A, CELF1, STAT5A, S100A6, RB1, KLF12, HIPK2, IGF1R, MALAT1, PAPPA, C16orf58, IRF2BPL, MDM4, PHB2, RAP2C, TNFRSF10B, UQCRFS1	miR-141-3p

Table 7-12: List of experimentally validated targets of miR-195-5p from miRTarBase

Target Gene	miRNA
WEE1, E2F3, CDK6, CCND1, BCL2L11, MECP2, VEGFA, SKI, CCL4, KRT7, BCL2, RAF1, RUNX2, SLC2A3, TBCCD1, CCND3, CDK4, CDC42, CAB39, CHUK, TAB3, MBD1, CCNE1, BCL2L2, JAK2, CAMKV, AGER, LSM11, ABCB7, ZNF280C, SPTBN1, NOLC1, CAND1, COPB1, RPL10, TMC6, TPI1, SH3BGRL2, AGO1, DICER1, FASN, ARL2, BIRC5, WNT7A, MYB, ATG14, SHOC2, PLEKHA1, CEP55, ZDHHC16, STXBP3, C1orf21, AMOTL1, BTG2, ETNK1, CCND2, TARBP2, MTFR1L, USP15, AKAP11, FCF1, ARIH1, PAGR1, SNTB2, VPS4A, C16orf72, GOSR1, MINK1, RPS6KB1, NAPG, GALNT1, PNPLA6, ZCCHC3, SOWAHC, CCNT2, UBR3, ZNRF3, YWHAH, SEPT2, B3GNT2, ACTR2, CDV3, U2SURP, RARB, ITPR1, BHLHE40, PI4K2B, PURA, CANX, PCMT1, PPP1R11, HSPA1B, PIM1, CD2AP, CALU, UBN2, EN2, MAFK, FOXK1, USP42, DMTF1, ZFHX4, RAD23B, ZBTB34, RECK, SLC9A6, ZNF275, OGT, LUZP1, CHAC1, SCAMP4, PAK2, PDCD4, ENTPD1, UBE4A, ARCN1, PTPRJ, TMEM138, RASSF5, MAPKAPK2, BCL7A, RCOR1, SPRED1, SMAD3, SCAMP5, UBE2Q2, DCTN5, PSKH1, ANKRD13B, EZH1, PNPO, SNX11, RNF138, IER2, TNFSF9, ENTPD6, KIF3B, NOL4L, RASSF2, CDS2, ATP5G3, TNRC6B, SELENOI, PEX13, TET3, PRKCD, CPEB2, TADA2B, GRAMD2B, CREBRF, HMGA1, CDKN1A, IRF4, KLHDC10, UBE3C, SUN1, HMBOX1, RNF38, PSAT1, ZBTB33, PAFAH1B2, RPRD2, PNRC2, CDADC1, AGO4, GPR180, PPM1A, SLC39A9, FAM103A1, RBBP6, PAFAH1B1, TAOK1, SNRPB2, HSPE1-MOB4, MOB4, BZW1, NUP50, FGF2, HSPA4L, RBPJ, TBPL1, FZD6, BAG4, ZNF449, YIPF6, DDX3Y, MSL1, ENTPD7, PRRC2C, ELK4, GABARAPL1, PDIK1L, ATXN7L3B, ZNF691, DYNLL2, SRPRB, CDC42SE2, WIPI2, TBRG1, CHEK1, TMEM109, C1orf226, SIRT4, SETD1B, HOXC8, RAB3IP, PPP2R5C, SLCO3A1, MLLT6, CBX2, ZFP28, CHMP4B, GABPA, STRADB, MTMR3, CRIM1, SOCS5, CTDSPL, ITGA2, PIK3R1, PPIL1, CAPZA2, ZNRF2, IFT74, OCRL, CHIC1, CRKL, ALOX12, PHACTR2, CGNL1, KIF21A, RASGEF1B, TUFT1, PDIA6, SNX16, TRIP10, CDC25A, BTRC, INSR, ELN, RET, Bace1, SIDT2, ASCC1, DMRT2, TLL1, LUC7L3, ALDH3B1, EFTUD2, CSNK1E, TPM2, ZNF460, FGFR4, DOCK11, ACTR3B, ZNF367, UBE2V1, UBE2Q1, TSC22D2, TOB2, TMEM189-UBE2V1, TMEM189, TM4SF1, TFAP2A, STK38, SSRP1, SIK1, SH3BP4, SEC24A, RNF168, REL, PISD, PDE4D, NR6A1, NFIC, NAA25, LAMC1, RUBCN, IVNS1ABP, IPPK, HOXA3, HEYL, HDGF, GNB1, GNAL, FURIN, EIF1AX, DYRK3, CPSF7, CARD10, AVL9, AKT3, AGO2, ABL2, ABHD2, ABCC6, PRKAR2A, VSIR, PHYHIP, CASKIN1, ZNRF1, CD180, KIAA0895, ORC4, ODF2L, DNAJA1, L2HGDH, ZNF622, ZMAT3, USP53, SYPL1, SRPRA, SREK1, SMAD7, PRICKLE2, LRIG2, KIF5B, FAM122B, E2F7, DDX3X, CDCA4, CDC37L1, ATG9A, ASGR2, ZNF620, HAUS3, YTHDC1, TMEM245, TMEM100, SRSF1, SESTD1, RIMS3, RCAN3, PPIG, PLRG1, PLAG1, PHKA1, MYO5A, KIF23, HNRNPDL, CDK1, CCNE2, CBX6, AXIN2, CASK, DMPK, ATAD5, AKR1B10, GPATCH8, ARHGDIA, CPEB3, JARID2, CAMSAP1, TRIM35, FLCN, NNT, SBNO1, POM121C, NUFIP2, LAMP2, EFNB2,	miR-195-5p

RPL14, GNAT1, HOXA10, CBX4, PHC3, PDCD1, BAZ2A, APP, ZNF585B, ANAPC13, PRSS21, RALGAPB, GSG1, POLDIP3, AP5Z1, CLSPN, UGT2B4, MAP4K2, CCDC83, DECR1, ZBTB10, YWHAQ, USP48, SALL1, RUNX1T1, RTN4, PNISR, PHLPP2, PAG1, NUCKS1, MKX, MBD4, LRRFIP2, KPNA3, KPNA1, HIGD1A, HCFC2, GRB2, FKBP1A, FBXL20, CRK, CLIP4, CDK17, CACUL1, C11orf24, ASH1L, AMOT, SLC29A1, OSCAR, MTHFR, FAM229B, EPM2AIP1, ZNF267, SSU72, DNAJC10, ZNF704, YRDC, TPM3, TMEM161B, TM7SF3, TAF13, SZRD1, RNF149, RACGAP1, RAB23, PTPRD, PRKAA1, PRDM4, PLPP3, LURAP1L, KANK1, HIST2H2BE, GPR27, EXT1, CYP26B1, CREBL2, CNKSR3, CA8, BTN3A3, ARHGAP12, OSBPL3, KRT33B, TUBB2A, MSANTD4, LANCL1, HNRNPA2B1, KIAA1456, SLC25A12, DLGAP3, THRAP3, SMDT1, RAPH1, CCNT1, ZNF391, CCDC80, ZBTB16, XKR7, WNK3, VAV2, TGFBR3, RASEF, NCKAP1, MAP3K7, KLHL15, GNG12, FZD9, CMTM4, CCDC88C, ARMC12, AHNAK2, ACVR2A, TLK1, UBE2H, TTLL5, RIF1, SERBP1, PHF19, PLPBP, CLEC2D, N4BP1, TRAK1, ADRA2B, ANKMY1, RNF41, GPRC5A, C3orf36, BSPRY, ANKRD36, KLHL40, NOTCH2, EIF2B2, CUL3, DCAF17, RS1, GLP2R, FLOT2, HNRNPA1L2, NEGR1, MCFD2, HNRNPA1, SLC35E2B, ARHGAP32, RAB15, ADORA3, PPIP5K2, SYNRG, ZNF91, ZBTB5, VPS33B, CHMP3, VCL, USP3, USP31, TUBB, TRAM1, TMEM69, PIP4P1, TMEM135, TLE4, TKTL1, TIMM13, TFB1M, TCF3, TBL1XR1, TBC1D20, TBC1D14, TASP1, SUPT16H, STX17, SRPK1, SPTLC1, SMURF1, SLC9A1, SLC7A5, SLC25A29, SLC25A22, RPS6KA3, RPS5, RPL36, RNPS1, RNMT, RNH1, RFWD2, REXO1, RELT, RAP2C, RAB40B, RAB11FIP2, PSMB5, PPP6C, POU2AF1, POLE4, PGD, PEX12, PANK1, TM9SF2, OTUB1, NR2C2, NCOR2, MTMR4, MRPL40, MLXIP, MIB1, MED11, MCM3AP-AS1, ETFRF1, LRRC57, LRPPRC, LITAF, KLC2, TECPR2, KATNAL1, IRAK1BP1, HYOU1, GSK3B, GGA3, GANAB, GABARAP, FRYL, MIGA1, AMER1, EDC3, DSCR3, DPP8, DNAJC9, CYLD, CYB561A3, CUL2, CSDE1, CREG1, CPNE1, RHOV, CDKN2AIPNL, CDC27, CBFA2T3, C6orf106, C2orf42, LRIF1, C15orf39, BSG, B4GALT1, ATP13A3, ASXL1, AP3M1, AFF4, ACOX1, NKD1, CDK8, YAP1, KIAA0100, KDR, AP2B1, AURKAIP1, BCL2L12, C16orf58, CLU, CLUH, DENND6A, DIAPH1, FBXL18, GATAD2A, HSPA8, KMT2D, MAP2K3, PLEKHB2, POLR2E, PPP6R3, RPRD1B, SEC61A1, SNCG, WDR13, C21orf62, CARM1, CD274, JPT2, RFK, TXNIP, VOPP1, ZNF284	

Table 7-13: List of experimentally validated targets of miR-486-5p from miRTarBase

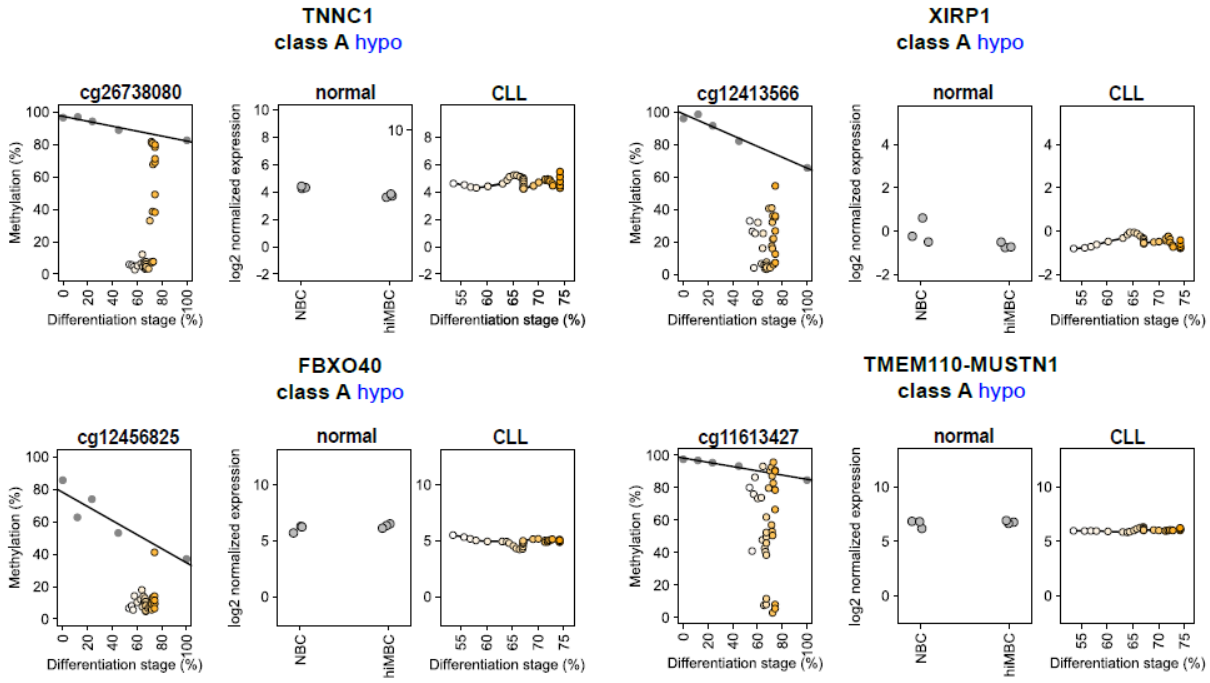
Target Gene	miRNA
CD40, TMED1, ARHGAP5, OLFM4, SERPINE1, ID4, PIM1, ZNRF2, IGF1R, DOCK3, CADM1, H3F3B, PCCA, ZDHHC20, UBASH3B, RCOR3, ZNF460, ZIC5, BTF3L4, CCDC14, CMSS1, TTC8, RBM12B, MBD4, FAM217B, IGSF3, YAE1D1, SEL1L3, FOXP1, TM4SF20, MARCH4, SP4, SEC23IP, BAG2, METTL27, DENND5B, MACROD2, UTP4, LTBP2, ARF6, RBM22, SAPCD2, BASP1, CENPN, CRIPT, DCTN4, SPRTN, EPGN, HPGD, FAM46A, ABCF2, FBN1, CDK4, CIT, SMAD2, SNAI1, NRP2, FOXO1, PTEN, CLDN10, HAT1, PIK3R1, G3BP2, HMGA1, UBE2S, FPR1, ZNF286B	miR-486-5p

7.6 CLL-specific protein-coding genes

Table 7-14: List of correlation coefficients (rho) of CLL-specific protein-coding genes identified using cell-of-origin as a control for aberrant methylation calls

Gene	Promoter	CpG	p-value	rho
MIB2	chr1:1549294-1551294	cg19246761	0.009	-0,44
MOV10	chr1:113215547-113217547	cg03417473	0.016	-0.40
AMICA1	chr11:118095309-118097309	cg06957943	0.013	-0.42
ARNTL2	chr12:27484286-27486286	cg26165146	0.002	-0.51
CLLU1	chr12:92813806-92815806	cg04845867	0.016	0.40
CLLU1	chr12:92813806-92815806	mean	0.008	0.44

C17orf99	chr17:76140933-76142933	cg27606002	0.016	-0.41
C18orf32	chr18:47013144-47015144	cg23756872	0.002	-0.50
PHLDB3	chr19:44008485-44010485	cg17121205	0.015	-0.41
LILRB4	chr19:55172770-55174770	cg24140775	0.007	-0.45
RAD21L1	chr20:1205263-1207263	cg22227354	0.015	0.41
C22orf46	chr22:42085046-42087046	cg14856679	0.003	-0.49
XIRP1	chr3:39233585-39235585	cg12413566	0.0002	-0.59
TNNC1	chr3:52487557-52489557	cg26738080	0.014	-0.41
TMEM110-MUSTN1	chr3:52931097-52933097	cg11613427	0.0001	-0.62
FBXO40	chr3:121310669-121312669	cg12456825	0.007	0.45
EFCAB12	chr3:129146994-129148994	cg03221715	0.013	-0.42
EFCAB12	chr3:129146994-129148994	mean	0.005	-0.47
ARL14	chr3:160393447-160395447	cg11896170	0.002	0.50
C4orf32	chr4:113065052-113067052	cg04831870	0.018	0.40
GABBR1	chr6:29600462-29602462	cg04324598	0.0009	-0.54
GABBR1	chr6:29600462-29602462	mean	0.0005	-0.57
DOK2	chr8:21770705-21772705	mean	0.01	-0.43



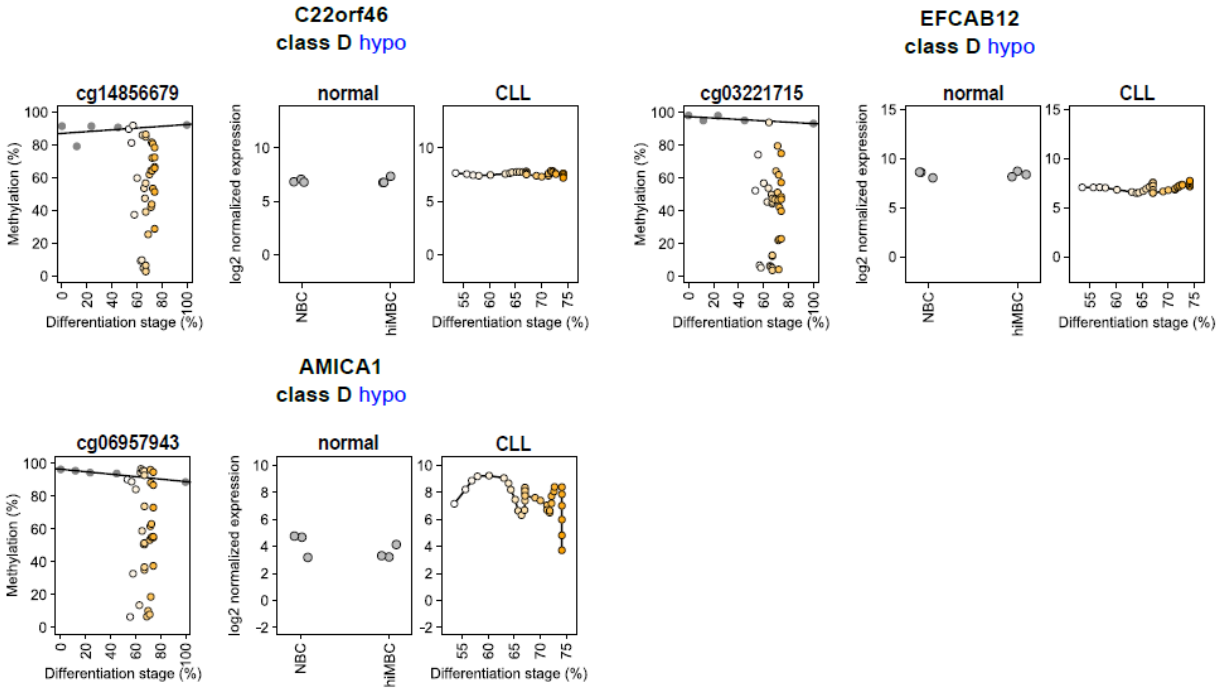


Figure 7-4: Candidate protein-coding genes associated with CLL-specific aberrant DNA methylation. A) Left panel, CLL-specific differentially methylated CpGs identified in gene promoters grouped by subclasses (class A, C, D and E). Epigenetic programming during normal B cell differentiation is represented as a grey line. Average methylation values are represented as dots; normal B cell subpopulations (grey dots); CLL samples (white-orange dots). Y-axis represents methylation levels (%), while X-axis differentiation stage assigned either to normal B cell or CLLs. **Middle panel, expression levels of protein-coding genes in normal B cells** (log2 normalized expression values, NBC; naive B cells, hiMBC; high maturity memory B cells). **Right panel, expression levels of protein-coding genes in CLL** (log2 normalized expression values). The y-axis represents log2 normalized expression values of protein-coding genes, and the x-axis assigns differentiation stage either to normal B cells or to CLLs

Table 7-15: List of correlation coefficients (rho) of protein-coding genes identified previously as epigenetically deregulated in CLL

Gene	Promoter	CpG	p-value	rho
ZAP70	chr2:98328530-98330530	cg25095518	0.45	0.13
ZAP70	chr2:98328530-98330530	cg08859278	0.58	0.1
ZAP70	chr2:98328530-98330530	cg09006159	0.26	0.2
ZAP70	chr2:98328530-98330530	cg15933451	0.94	0.01
ZAP70	chr2:98328530-98330530	cg13853141	0.76	0.05
ZAP70	chr2:98328530-98330530	cg12332902	0.07	0.32
ZAP70	chr2:98328530-98330530	cg21773162	0.88	0.03
ZAP70	chr2:98328530-98330530	mean	0.42	0.14
TWIST2	chr2:239755172-239757172	cg04840356	0.23	-0.21
TWIST2	chr2:239755172-239757172	cg06738242	0.65	0.08
TWIST2	chr2:239755172-239757172	mean	0.73	-0.06
HOXA4	chr7:27169899-27171899	cg23884241	0.39	-0.15
HOXA4	chr7:27169899-27171899	cg16651126	0.5	0.12
HOXA4	chr7:27169899-27171899	cg11015251	0.57	0.1

HOXA4	chr7:27169899-27171899	cg06942814	0.77	-0.05
HOXA4	chr7:27169899-27171899	cg15196806	0.19	0.23
HOXA4	chr7:27169899-27171899	mean	0.84	0.04
DAPK1	chr9:90111255-90113255	cg08719486	0.91	0.02
DAPK1	chr9:90111255-90113255	cg13814950	0.17	0.24
DAPK1	chr9:90111255-90113255	cg22571217	0.74	0.06
DAPK1	chr9:90111255-90113255	cg13932603	0.54	0.11
DAPK1	chr9:90111255-90113255	cg20401521	0.64	0.08
DAPK1	chr9:90111255-90113255	cg08797471	0.82	0.04
DAPK1	chr9:90111255-90113255	mean	0.93	0.02

8. PUBLICATIONS AND POSTER PRESENTATIONS

PUBLICATIONS - PEER-REVIEWED JOURNALS

Wierzbinska JA, Toth R, Rippe K, Mallm JP, Mertens D, Zenz T, Hielscher T, Seifert M, Küppers R, Oakes CC, Byrd J, Lipka DB, Plass C. **Modeling of the epigenome of cell-of-origin identifies cancer-specific DNA methylation patterns in CLL** (in preparation).

Mansouri L, Wierzbinska JA, Plass C, Rosenquist R. **Epigenetic deregulation in chronic lymphocytic leukemia: Clinical and biological impact**. *Seminars in Cancer Biology*, Review, 2018.

Lipka DB, Witte T, Toth R, Yang J, Wiesenfarth M, Nöllke P, Fischer A, Brocks D, Gu Z, Park J, Strahm B, Wlodarski M, Yoshimi A, Claus R, Lübbert M, Busch H, Boerries M, Hartmann M, Schönung M, Kilik U, Langstein J, Wierzbinska JA, Pabst C, Garg S, Catalá A, De Moerloose B, Dworzak M, Hasle H, Locatelli F, Masetti R, Schmugge M, Smith O, Stary J, Ussowicz M, van den Heuvel-Eibrink MM, Assenov Y, Schlesner M, Niemeyer C, Flotho C, Plass C. **RAS-pathway mutation patterns define epigenetic subclasses in juvenile myelomonocytic leukemia**. *Nature Communications*, 2017.

Wilhelm T, Lipka DB, Witte T, Wierzbinska JA, Fluhr S, Helf M, Mücke O, Claus R, Konermann C, Nöllke P, Niemeyer CM, Flotho C, Plass C. **Epigenetic silencing of AKAP12 in juvenile myelomonocytic leukemia.** *Epigenetics*, 2016.

POSTER PRESENTATIONS

Modeling of the epigenome of cell-of-origin identifies cancer-specific DNA methylation patterns in CLL.
Wierzbinska JA, Toth R, Rippe K, Mallm JP, Mertens D, Zenz T, Hielscher T, Seifert M, Küppers R, Oakes CC, Byrd J, Lipka DB, Plass C. *Cambridge Lymphoma Biology International Symposium*, Cambridge, UK, 07/2018.

A novel strategy for identification of CLL-specific methylation events with a focus on epigenetic deregulation of microRNAs in CLL.
Wierzbinska JA, Toth R, Rippe K, Mallm JP, Mertens D, Zenz T, Hielscher T, Seifert M, Küppers R, Oakes CC, Byrd J, Lipka DB, Plass C. *International Cancer Evolution Symposium*, Munich, 03/2017.

Epigenetic regulation of microRNA expression in CLL.
Wierzbinska JA, Toth R, Rippe K, Mallm JP, Mertens D, Zenz T, Hielscher T, Seifert M, Küppers R, Oakes CC, Byrd J, Lipka DB, Plass C. *Ph.D. poster presentation*, DKFZ, Heidelberg, Germany, 11/2016.
Genome-wide mapping of epigenetically regulated promoters and enhancers of microRNAs in CLL.

Wierzbinska JA, Toth R, Rippe K, Mallm JP, Mertens D, Zenz T, Hielscher T, Seifert M, Küppers R, Oakes CC, Byrd J, Lipka DB, Plass C. *Ph.D. retreat,* Weil-der Stadt, Germany, 07/2016.

ORAL PRESENTATIONS AT INTERNATIONAL MEETINGS

The identification of the cell-of-origin and cancer-specific DNA methylation patterns in CLL.
Wierzbinska JA. *Cancer Core Europe Lymphoma Meeting*, Cambridge, UK, 07/2018.

9. ACKNOWLEDGMENTS

I would not have been able to successfully complete my Ph.D. thesis without the strong support of numerous people. This is just a modest way of showing my gratitude. I would like to thank all of you:

- Prof. Dr. Christoph Plass for giving me a chance to join his team, acquire new skills, for providing me with a scientific support and mentorship during this time. I am very grateful for all discussions, meetings, and for all the patience during this Ph.D. journey. Thanks for showing me how to be passionate about science.

- Dr. Daniel Lipka for the scientific support, mentorship, and all the time invested in discussions, meetings, corrections, analyses, bioinformatical teaching, and for all the patience. Also, for mensa-time and the non-scientific chatting

- Dr. Reka Toth for unconditional bioinformatical support and mentorship. Without you, I would not be able to learn computational methods so fast. Thank you so much for all your help, and discussions, during which you patiently answered all my dummy questions☺

- My Thesis Advisory Committee (TAC) members: Dr. Michael Milsom, PD. Dr. Odilia Popanda and Prof. Dr. Stefan Fröhling for the helpful discussions and suggestions.

- Dr. Annika Baude, Dr. Simin Oz, Dr. Pavlo Lutsik for all discussions related to my project.

- The fellow Ph.D. students/Post-Docs who, by now, have become my close friends: Annika Baude, Melanie Weiss, Tania Witte, Simin Oz, Monika Helf, Michael Daskalakis, Reka Toth. The time shared throughout these years has been unique. Thanks for your friendship, laughing, chatting, personal support and the time spent outside the lab doing sports, cooking or just having fun.

- My new office friends Ashish Goyal, Alexander Kühn and Ana Banito. It has been a great pleasure to share an office with you over the last months. Without your personal support during this time, the writing would not have been easy.

- Monika Helf, Marion Bähr, Oliver Mücke for their technical support and help when needed, e.g. cell culture, MassARRAY and all the other methods …

- All the members of our group whom I might have missed in the previous points: Daniela, Aurore, Joschka, Yassen, David: it was really great to work with you and to get to know you. Thank you for your support☺

Acknowledgments

- To my lovely flatmates, Bouchra and Siao-Han: I'm blessed to have you as friends. Thank you for your support, help, and late-night chats☺ Friendship is such a special thing and I'm so glad I get to share it with you ladies. The power of friendship is a lot more than fun times. It is a pity I don't always acknowledge it at the time. Today, I take this chance to say thank you☺ You're amazing!

- At last but not least, I would like to thank some other friends (it's not possible to mention all of you by name: Magda, Marta, Alicia, Max, Korbinian) and family (my parents, brothers) for being there whenever I needed. Your support has been invaluable during these last four years of my life. I am beyond lucky to have you in my life.

- To my parents: Mom and Dad, I have no words to acknowledge the sacrifices you made and the dreams you had to let go, just to give me a shot at achieving mine. Thanks a lot for that! Dziekuje z calego serca!