Functional Impact of Preleukemic Mutations on the Human Hematopoietic Stem and Progenitor Cell Compartment

Inaugural-Dissertation

to obtain the academic degree
Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy
of Freie Universität Berlin

by

Friederike Christen

2021

I hereby declare that I am the sole author and composer of my thesis and that no other sources or aids, other than those listed, have been used.

The following work was carried out in the research group of Frederik Damm between June 2016 and December 2020 at the Molecular Cancer Research Centre and Department of Hematology, Oncology and Cancer Immunology at Charité Universitätsmedizin Berlin.

1st Reviewer: Prof. Dr. med. Frederik Damm
Department of Hematology, Oncology and Cancer Immunology
Charité Universitätsmedizin Berlin

2nd Reviewer: Prof. Dr. rer. nat. Peter Robin Hiesinger
Institute of Biology, Neurobiology
Freie Universität Berlin

Date of Defense: December 15th, 2021

Abstract

Dissecting functional consequences of gene mutations contributes to understanding disease development and identifying potential treatment strategies. Mutations in epigenetic regulators (especially in *DNMT3A*, *TET2*, and *ASXL1*, termed DTA mutations) have been identified in acute myeloid leukemia patients, as well as individuals without history of hematologic malignancies. Mutations in these genes are hypothesized to confer a clone a competitive advantage, leading to preferential expansion of this specific clone. DTA mutations are proposed to be preleukemic events that are acquired before leukemia onset and are thought to elevate the risk for acquisition of additional mutations that in turn drive malignant transformation. Clones harboring these mutations are able to survive chemotherapy, potentially increasing the risk for disease recurrence. In this thesis, I investigated the impact of DTA mutations in the human hematopoietic stem and progenitor cell compartment, as well as the role of preleukemic events in patients with AML t(8;21).

Using CRISPR/Cas9, site-specific mutations were introduced in $CD34^+$ progenitor cells freshly isolated from umbilical cord blood samples. By delivering Cas9 nuclease together with an *in vitro* transcribed sgRNA as ribonucleoprotein, I was able to achieve editing efficiencies ranging from 40% for the sgRNA targeting exon 13 of *ASXL1* to over 90% for the sgRNA targeting exon 6 of *TET2*. A known hotspot mutation in *DNMT3A* R882H was introduced with an efficiency of up to 50% in the cell bulk. After successful genetic modification, cells were subjected to *in vitro* assays to assess phenotype, self-renewal properties, differentiation capacities, and long-term culture-initiating potential. Clonal composition and expansion were assessed via deep sequencing and subsequent analysis of introduced specific insertions and deletions. To investigate the role and importance of preleukemic mutations in AML leukemogenesis, a comprehensive genetic in-depth characterization of AML t(8;21) was performed. To this aim, DNA from the time of diagnosis was analyzed via next generation sequencing with a targeted gene panel including exons and coding regions of 66 genes recurrently mutated in hematologic malignancies. From 56 patients, samples from clinical remission were analyzed for the previously identified mutations by targeted re-sequencing of amplicons spanning the target region.

In this thesis, I was able to demonstrate that mutations in *DNMT3A*, *TET2*, and *ASXL1* influence the behavior of hematopoietic stem and progenitor cells (HSPCs) in a gene-specific fashion. Mutations in *ASXL1* had only little effect on HSPCs with no observed changes in expression of differentiation markers after culture for three weeks and no enhanced proliferation in long-term

culture. The only hint that *ASXL1* mutations have an impact on the fitness of HSPCs was increased colony forming ability after long-term culture, meaning that although progenitor cells did not proliferate during the long-term culture period, they were able to sustain and generate colonies later. *DNMT3A* mutations led to a longer retention of CD34 marker expression in differentiation culture experiments, but not enhanced serial replating capacity. *TET2*-mutated cells, in contrast, showed delayed myeloid marker expression in short-term culture and increased self-renewal of committed progenitor cells. Both, *DNMT3A* and *TET2* mutations led to clonal expansion of distinct cell clones during long-term culture with a competitive advantage over wild type cells. The clonal composition of *TET2*- and *DNMT3A*-mutated samples differed, with *TET2*^{mut} samples being more diverse. Collectively my data indicate that *DNMT3A* mutations preferentially influence primitive HSC clones, whereas *TET2* mutations enhance the fitness of committed progenitor cell clones. *ASXL1* mutations might need additional events to drive clonal expansion. Analysis of patient samples revealed that mutations in *DNMT3A* and *TET2* are early events in AML t(8;21) and potentially acquired before the onset of the disease. *ASXL1* mutations are neither early nor late events in these patients but are potentially cooperating events in leukemia transformation.

In this thesis, I was able to demonstrate that DTA mutations enhance the fitness of distinct human hematopoietic stem or progenitor cell clones, in terms of self-renewal and proliferative capacity, leading to clonal expansion. The results obtained here highlight the importance of monitoring specific preleukemic events to prevent disease recurrence as well as development of severe hematologic conditions.

Zusammenfassung

Um die Entwicklung von Krankheiten aufzudecken und etwaige zielgerichtete Behandlungsstrategien zu entwickeln, ist es essenziell den funktionellen Einfluss bestimmter Genmutationen zu untersuchen. Mutationen in epigenetischen Regulatoren wie *DNMT3A, TET2* und *ASXL1* (auch DTA-Mutationen genannt) wurden sowohl in Patienten mit akuter myeloischer Leukämie als auch in Patienten ohne hämatologische Erkrankung gefunden. Diese Mutationen führen zu einem Vorteil bestimmter Zellklone, die daraufhin expandieren. DTA-Mutationen werden auch als prä-leukämische Mutationen bezeichnet, die vor dem Ausbruch der Leukämie bereits in Stamm- und Vorläuferzellen vorkommen und ein erhöhtes Risiko für das Auftreten weiterer Mutationen bedingen. Zusätzliche Mutationen in einem prä-leukämischen Klon können wiederum zur malignen Transformation des Zellklons führen. Des Weiteren können Zellen mit prä-leukämischen Mutationen resistent gegen Chemotherapie sein und somit das Risiko eines Rezidivs erhöhen. In dieser Arbeit habe ich den Einfluss von Mutationen in *DNMT3A, TET2* und *ASXL1* auf humane hämatopoetische Stamm- und Vorläuferzellen untersucht, sowie die Rolle von prä-leukämischen Mutationen in Patienten mit AML t(8;21) erforscht.

Mit Hilfe von CRISPR/Cas9 wurden spezifische Mutationen in CD34-positiven Vorläuferzellen eingeführt, die frisch aus Nabelschnurblut isoliert wurden. Die Cas9-Nuklease wurde zusammen mit *in vitro*-transkribierter sgRNA in Form von Ribonukleoproteinen in die Zellen gebracht. In Exon 13 von *ASXL1* wurden Insertionen und Deletionen in mindestens 40% der Zellen generiert und in Exon 6 des Gens *TET2* konnten in mehr als 90% der Zellen Insertionen und Deletionen entdeckt werden. Die spezifische Mutation *DNMT3A* R882H wurde in bis zu 50% der transfizierten Zellen eingeführt. Die erfolgreich mutierten Zellen wurden anschließend in *in vitro*-Experimenten untersucht in Hinsicht auf Veränderungen im Phänotyp, der Fähigkeit zur Selbsterneuerung, Differenzierungsfähigkeit und der Fähigkeit sich in Langzeitkulturen zu vermehren. Die klonale Zusammensetzung, sowie die Expansion bestimmter Zellklone wurden mit Hilfe von Deep Sequencing und anschließender Analyse der generierten Insertionen und Deletionen untersucht. Patienten mit AML t(8;21) wurden auf prä-leukämische Mutationen untersucht, indem DNA vom Diagnosezeitpunkt via Next-Generation Sequenzierung analysiert wurde. Kodierende Regionen von 66 Genen, die wiederholt in malignen hämatologischen Erkrankungen auftreten, wurden in die Analyse eingeschlossen. Von 56 Patienten wurden DNA-Proben vom Zeitpunkt der klinischen Remission erneut auf die bereits identifizierten Mutationen untersucht.

In dieser Arbeit konnte ich zeigen, dass Mutationen in *DNMT3A*, *TET2* und *ASXL1* unterschiedliche Effekte auf hämatopoetische Stamm- und Vorläuferzellen haben. Mutationen in *ASXL1* hatten den mildesten Effekt. Weder Unterschiede in der Expression von Differenzierungsmarkern noch erhöhte Replatierungsfähigkeit in Vergleich zu Wildtypzellen wurden beobachtet. Des Weiteren zeigten die Zellen kein vermehrtes Wachstum in der Langzeitkultur. Sie waren jedoch in der Lage Kolonien nach der Langzeitkultur zu bilden, was bedeuten könnte, dass diese Zellen trotz des unveränderten Proliferationsverhaltens einen Vorteil gegenüber Wildtypzellen haben und somit in der Lage sind die Langzeitkultur zu überdauern. Zellen mit *DNMT3A* Mutationen exprimierten den Vorläuferzellmarker CD34 länger als Kontrollzellen. Mutationen in *TET2* führten zu einer verspäteten Expression von myeloischen Differenzierungsmarkern und einem erhöhten Selbsterneuerungspotential von Vorläuferzellen. Sowohl in Zellen mit *DNMT3A*- als auch mit *TET2*-Mutationen konnte ich die Expansion bestimmter mutierter Zellklone nachweisen. Die Proben mit eingeführten Mutationen in *TET2* hatten jedoch eine größere Anzahl verschiedener Insertionen und Deletionen als Proben mit *DNMT3A*-Veränderungen. Die Daten dieser Arbeit zeigen, dass Mutationen in *DNMT3A* dazu führen, dass primitive Zellklone einen evolutionären Vorteil erhalten, wohingegen reifere Vorläuferzellen von Mutationen in *TET2* profitieren. Mutationen in *ASXL1* führen möglicherwiese nur in Anwesenheit weiterer genetischer oder zellulärer Veränderungen zu klonaler Expansion. In Patienten mit AML t(8;21) treten Mutationen in *DNMT3A* und *TET2* wahrscheinlich früher auf als andere Mutationen und kommen möglicherweise in Vorläuferzellen vor, bevor sich die eigentliche Erkrankung entwickelt. Mutationen in *ASXL1* scheinen weder frühe noch späte Ereignisse zu sein, sondern treten irgendwann dazwischen auf. Sie kooperieren möglicherweise mit anderen Mutationen, um Zellen zu transformieren.

In dieser Arbeit konnte ich zeigen, dass DTA-Mutationen die Fitness von bestimmten Stamm- und Vorläuferzellen im menschlichen Stammzellkompartiment erhöhen und so zu einer Expansion der mutierten Zellklone führen. Die Ergebnisse dieser Arbeit bestätigen, dass die Untersuchung und Beobachtung prä-leukämischer Klone dazu beitragen könnte, ein Rezidiv oder sogar die Entstehung schwerer hämatologische Erkrankungen zu verhindern.

Table of Contents

Abbreviations

°C	degree Celsius
µg	microgram
µl	microliter
µm	micrometer
µM	micromole/liter
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5hmU	5-hydroxymehtyluracil
5mC	5-methylcytosine
aa	amino acid
ADD domain	ATRX-Dnmt3-Dnmt3L domain
ADP	adenosine diphosphate
AID	activation induced cytidine deaminase
ALL	acute lymphocytic leukemia
AML	acute myeloid leukemia
APOBEC	apolipoprotein B editing complex
APS	ammonium persulfate
Arg/R	arginine
ASXL	additional sex combs-like
Asx	additional sex combs
ATP	adenosine triphosphate
BAP1	BRCA1-associated protein 1
BRCA1	Breast cancer susceptibility protein type 1
BCA	bicinchoninic assay
BER	base excision repair
BFU-E	burst-forming unit erythroid
BM	bone marrow
BMI1	B lymphoma Mo-MLV insertion region 1 homolog
bp	base pair
BSA	bovine serum albumin
Cas9	CRISPR associated nuclease 9
CB	cord blood
PHD	plant homeodomain
CBF	core-binding factor
CD	cluster of differentiation
cDNA	complementary DNA
CFC	colony-forming cell

CFU	colony-forming unit
CFU-E	colony-forming unit erythroid
CFU-GEMM	colony-forming unit granulocyte, erythrocyte, macrophage, megakaryocyte
CFU-GM	colony-forming unit granulocyte, macrophage
CHIP	clonal hematopoiesis of indeterminate potential
CI	confidence interval
CLL	chronic lymphocytic leukemia
CLP	common lymphoid progenitor
CML	chronic myeloid leukemia
CMP	common myeloid progenitor
CMV	cytomegalovirus
CN	cytogenetically normal
CRISPR	clustered regularly interspaced short palindromic repeat
ddPCR	digital droplet polymerase chain reaction
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DNMT	DNA methyltransferase
dNTP	deoxynucleotide triphosphate
dsDNA	double stranded DNA
DTA	*DNMT3A*, *TET2*, *ASXL1*
DTT	dithiothreitol
dUTP	deoxyuridine triphosphate
e.g.	exempli gratia
EDTA	ethylenediaminetetraacetic acid
EPO	erythropoietin
FACS	fluorescence activated cell sorting
FBS	fetal bovine serum
FcR	Fc receptor
FDR	false discovery rate
FLT3-L	FMS-like tyrosine kinase 3 ligand
g	gravitational constant
G-CSF	granulocyte colony stimulating factor
GFP	green fluorescent protein
GM-CSF	granulocyte macrophage colony stimulating factor
GMP	granulocyte-monocyte progenitor
Gy	grey
h	hour
HCl	hydrochloric acid
HDAC	histone deacetylase

HDR	homology directed repair
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIF-1	hypoxia inducible factor 1
His/H	histidine
HLTM	human long-term culture medium
HMT	histone methyltransferase
HRP	horseradish peroxidase
HSC	hematopoietic stem cell
HSPC	hematopoietic stem or progenitor cell
i.e.	id est
ICE	inference of CRISPR edits
IDE	integrated development environment
IL-3	interleukin 3
IL-6	interleukin 6
IMDM	Iscove's modified Dulbecco's medium
Indel	insertion and deletion
IR	ionizing radiation
ITD	internal tandem duplication
IT-HSC	intermediate-term HSC
IVT	*in vitro* transcription
kb	kilo base
KO	knockout
LB	Luria broth
LMPP	lymphoid-primed multipotential progenitor
LOH	loss of heterozygosity
LT	long-term
LTC-IC	long-term culture-initiating cell
M	mole/liter
MACS	magnetic cell separation
MDS	myelodysplastic syndrome
MEP	megakaryocytic-erythroid progenitor
mg	milligram
min	minute
ml	milliliter
MLP	myeloid-lymphoid progenitor
mM	millimole/liter
M-MLV	Moloney Murine Leukemia Virus
MNC	mononuclear cell
MPN	myeloproliferative neoplasm
MPP	multipotent progenitor
mRNA	messenger ribonucleic acid

ms	millisecond
mut	mutated/mutation
ng	nanogram
NHEJ	non-homologous end joining
NK cell	natural killer cell
nM	nanomole/liter
nm	nanometer
No.	Number
NOS	not otherwise specified
OPN	osteopontin
P/S	penicillin/streptomycin
PAGE	polyacrylamide gel electrophoresis
PAM	protospacer adjacent motif
PB	peripheral blood
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PHD	plant homeodomain
PI	propidium iodide
PI3K	phosphatidylinositol 3-kinase
pmol	picomole/liter
PRC 1/2	polycomb repressive complex 1/2
Pro/P	proline
PWWP domain	Proline-Tryptophan-Tryptophan-Proline domain
qRT-PCR	quantitative real-time polymerase chain reaction
RCLB	red cell lysis buffer
RNA	ribonucleic acid
RNase	ribonuclease
RNP	ribonucleoprotein
ROS	reactive oxygen species
RPMI	Roswell Park Memorial Institute medium
RTK	receptor tyrosine kinase
RUNX1-3	runt-related transcription factor 1-3
s	second
SCF	stem cell factor
SDS	sodium dodecyl sulfate
sgRNA	single guide RNA
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
ssDNA	single stranded DNA
ST	short-term
T7E1	T7 endonuclease I
TBS-T	tris-buffered saline with Tween 20

TE	tris/EDTA
TEMED	N,N,N’,N’-tetramethylenediamine
TET	ten-eleven translocation
TGF-β	transforming growth factor β
TIDE	tracking of indels by decomposition
TKD	tyrosine kinase domain
TPO	thrombopoietin
tRFP	tag red fluorescent protein
Tris	tris(hydroxyethyl)aminomethane
Trp/W	tryptophan
U	Units
UV	ultraviolet
VAF	variant allele frequency
WT	wild type

Note: gene symbols are not included in the list of abbreviations because of the internationally standardized nomenclature

List of Figures

List of Tables

1 Introduction

1.1 Hematopoiesis

1.1.1 Hematopoietic System

Hematopoiesis describes the process of continuous production of mature blood cells, which is the main function of the hematopoietic system. Mature blood cells have a short lifespan and have to be generated continuously from rare hematopoietic stem cells (HSCs). The bone marrow, which is localized in the medulla of axial and long bones, is the major blood producing organ in adult mammals and harbors precursors and multipotent stem cells at various differentiation and developmental states. It is comprised of a vascular compartment, the stroma, and the hematopoietic cell compartment. This microenvironment regulates life cycle, maturation, and migration of the cells. Quiescent hematopoietic cells are dependent on stromal cells to maintain self-renewal and multipotency. Stromal cells influence signaling in the cell pools to facilitate proliferation and differentiation via interaction on various levels (i.e., direct cell-cell interaction or indirect via growth factors and cytokines).

Multipotent stem cells in the bone marrow can give rise to HSCs, mesenchymal stem cells, and endothelial precursor cells. These cells have long-term self-renewal capacity to maintain the stem cell pool [1]. By asymmetric division multipotent stem cells generate committed progenitors. With further differentiation of multipotent into oligopotent and unipotent stem cells, the cells lose their self-renewal potential and become lineage committed [2]. A small quiescent HSC pool resides in the bone marrow and is characterized as non- or slowly dividing long-term hematopoietic stem cells (LT-HSC) [3, 4]. Short-term HSCs (ST-HSC) maintain cycling and are generated by asymmetric cell division from the quiescent pool. In the classical model found in most textbooks, hematopoietic differentiation is described as a hierarchical, tree-like system (Figure 1.1) that begins with the differentiation of HSCs into multipotent progenitors (MPP), which in turn differentiate into common lymphoid progenitors (CLP) or common myeloid progenitors (CMP). CMPs divide into megakaryocytic-erythroid progenitors (MEP), which in turn differentiate into erythrocytes or megakaryocytes and platelets, and granulocyte-monocyte progenitors (GMP), which differentiate into monocytes, macrophages, granulocytes, and mast cells. CLPs give rise to committed lymphoid precursors (e.g., pro B cells, pro T cells), which eventually differentiate to B cells, T cells, and natural killer cells (NK cells).

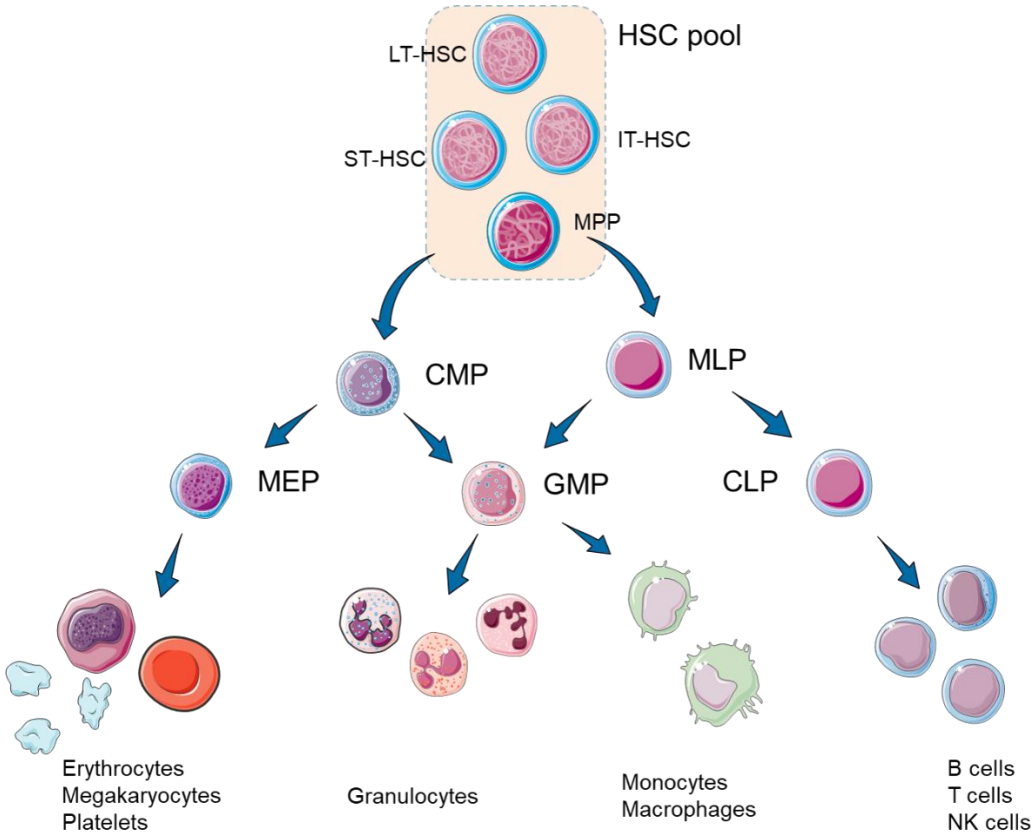


Figure 1.1 Model of human hematopoietic differentiation. The simplified differentiation tree from HSCs to differentiated blood cells begins with an HSC pool comprising different types of cells with heterogeneous self-renewal potential, which can produce all lineages. These differentiate into more committed progenitors, which in turn produce differentiated blood cells. HSC: hematopoietic stem cell, LT: long-term, IT: intermediate-term, ST: short-term, MPP: multipotent progenitor, CMP: common myeloid progenitor, MLP: myeloid-lymphoid progenitor, MEP: megakaryocyte-erythrocyte progenitor, GMP: granulocyte-monocyte progenitor, CLP: common lymphoid progenitor, NK cells: natural killer cells. The graphic was generated with Smart Servier Medical Art templates [5].

Due to improved methodology, new types of hematopoietic cell populations have been identified and the classical differentiation tree revised. Most studies focused on single aspects of the hematopoietic differentiation process, which makes it difficult to paint a comprehensive picture of the hematopoietic development. For example, a fraction of HSCs have an extended repopulating capacity, but not unlimited, and are placed somewhere between ST-HSCs and LT-HSCs (intermediate-term hematopoietic stem cells, IT-HSC) [6]. MPP are a highly heterogeneous population with cells being more similar to ST-HSCs or cells that preferentially generate lymphoid cells but retain myeloid output and are therefore defined as myeloid-lymphoid progenitors (MLP, equivalent to murine lymphoid-primed multipotential progenitors, LMPP [7]), contradicting to some extent the previously described early fate decision between lymphoid and myeloid lineages [8]. Although often described as a step-wise process, hematopoietic differentiation is nowadays believed to be a continuous process and classical boundaries between cell populations are lacking [9, 10]. It has also been shown that the

megakaryocytic lineage can directly rise from HSCs and bypass MPPs [11]. The model is further complicated by a profound heterogeneity within the HSC compartment, in which cells differ in their self-renewal ability and differentiation outputs. These cells possess highly variable chromatin architecture and show lineage-specific priming. Taken together, hematopoiesis is a highly regulated process in which multipotent as well as lineage-restricted hematopoietic stem and progenitor cells (HSPC) are continuously generating terminally differentiated blood cells. With emerging experimental techniques and progress on single cell level, more advances are made, and the hematopoietic differentiation process can be refined. Nevertheless, to dissect altered and malignant hematopoiesis it is essential to understand the normal processes around hematopoietic stem cell development.

1.1.2 Hematopoietic Stem and Progenitor Cells

HSCs are characterized as quiescent and divide rarely (estimated every 175-350 days [12]). The most dormant HSCs have the highest repopulation capacity in mouse studies [13]. Nevertheless, these cells can transiently proliferate, especially in response to stress when the blood system has to be replenished rapidly [13]. Interestingly, although the most dormant HSCs (i.e., LT-HSCs) have the highest repopulation capacity, these cells are not responsible for blood cell production in healthy individuals but rather function as an emergency pool in case of injury [14]. They are less susceptible to cytotoxic agents, e.g., ultraviolet light, ionizing radiation, or chemicals because these agents mainly target cells in M or S phase of the cell cycle. Stem cells have tightly controlled DNA damage response and repair mechanisms [15]. Since HSCs are long-lived, they can accumulate damage over time. Infrequent divisions are a mechanism to protect LT-HSCs from DNA or protein damage. A tight regulation of the stem cell state and stem cell fate is therefore important.

The local microenvironment plays an important role in maintenance of stem cell function and differentiation. The cell state is regulated by extrinsic and intrinsic factors. Extrinsic factors are mainly dependent on components of the stem cell niche, for example cytokines, growth factors, and chemokines that are produced by the surrounding cells [16], as well as oxygen tension and nutrient concentrations [17, 18]. These external signals are translated to modulate intrinsic signaling pathways, transcription factors, and epigenetic marks. It has been shown as early as 1975 that especially quiescent HSCs are located close to the surface of the bone in the endosteal region and move towards the vascular region during differentiation and maturation [19]. It has been confirmed that the localization of cells in specific niches is a dynamic process with cells moving from the quiescence-maintaining endosteal zone to the differentiation-promoting

vascular zone [20, 21]. In the endosteal niche cellular components secrete cytokines and molecules that regulate the HSC state, e.g., osteopontin (OPN) that maintains HSC quiescence [22] and membrane-bound stem cell factor (SCF). Membrane-bound SCF binds and activates the receptor c-Kit, a tyrosine kinase receptor that is expressed in high levels in LT-HSC and in turn leads to activation of stem cell specific gene expression [23]. Additionally, adhesion of HSPCs to stromal cells is mediated by membrane-bound SCF [24, 25]. Although the direct involvement of osteoblasts in the regulation of HSC quiescence has been debated [26], a number of studies has shown that the number of osteoblasts directly correlates with the number of LT-HSCs [20, 27]. The contrasting findings of different studies might result from heterogeneity within the HSC compartment and resulting definition of HSCs. A feasible hypothesis is that especially quiescent HSC reside in the endosteal region in close proximity to osteoblasts, whereas cycling HSCs, as well as MPPs, are found in the perivascular or subendosteal region (reviewed in [28]). Differentiation into myeloid and megakaryocyte progenitors is supported by bone marrow endothelial cells [21]. These are located in the vascular zone, into which HSCs migrate upon release from membrane-bound SCF [29]. This short excursion in HSC regulation by niche components is by far not complete but should demonstrate the various interactions that are necessary for homeostasis in the hematopoietic system. The stem cell state is further regulated by activating and deactivating various extrinsic and intrinsic signaling factors and pathways, like Hedgehog, Notch, transforming growth factor β (TGF-β), as well as cell cycle and proliferation regulation via p53 and phosphatidylinositol 3-kinase (PI3K) pathway. These pathways are regulated by cues from the microenvironment mentioned before resulting in distinct epigenetic profiles that are fluent with the changing needs within the environment. The importance of epigenetic regulation within the hematopoietic system will be elucidated in another chapter.

Oxygen tension is another cue for HSC regulation. HSCs exist in a low oxygen environment, which is maintained in the stem cell niche. It has been reported that HSCs have a hypoxic profile with low oxidative phosphorylation. The cells are therefore dependent on anaerobic glycolysis to produce ATP. Metabolism in a low oxygen environment is regulated by hypoxia inducible factor 1 (HIF-1). HIF-1 transcriptionally activates glycolytic genes and therefore mediates the metabolic switch from oxidative to glycolytic metabolism [30]. Due to the limitation of mitochondrial respiration, endogenous levels of reactive oxygen species (ROS) are low in HSCs. These cells are highly sensitive to ROS and lose stem cell properties. It is therefore suggested, that the distinct metabolism of self-renewing HSCs is a key feature to retain the quiescent state [31]. When cells move from a quiescent to an actively proliferating state (i.e.,

becoming ST-HSC and MPP), more energy in form of ATP is needed. Interestingly, it has been found that these actively proliferating cells are not only proliferating to differentiate but also to produce more progenitors with the same properties (e.g., self-renewal). These cells are the main source of unperturbed hematopoiesis [32]. To accomplish this, cells need to switch to oxidative phosphorylation. Therefore, metabolism seems tightly engaged in the regulation of HSC fate and differentiation. Mitochondrial respiration is a cue for HSCs to enter the cell cycle and compromise self-renewal capacity. This switch is reversible, depending on external signals [33]. The hematopoietic system can react to changes in the environment and produce cells based on the current needs and balance between HSC self-renewal and generation of differentiated progeny. Deregulation of these processes might tip the scale and lead to uncontrolled self-renewal or skewing in mature blood cell production.

1.1.3 Altered and Clonal Hematopoiesis

Alterations in the tightly controlled processes of hematopoietic stem cell maintenance might ultimately lead to the onset of hematologic malignancies. Especially in the aging system cells are exposed to extrinsic as well as intrinsic stress factors that might cause damage and disrupt homeostasis. To avoid accumulation of DNA damage, HSCs divide infrequently and limit their self-renewal and long-term repopulating potential with the number of divisions. This effect might cause the functional decline in the aging stem cell compartment [34]. HSCs from elderly individuals showed decreased repopulation capacities, indicating impaired stem cell functions [35]. It has been demonstrated that immunophenotypically characterized HSCs are increased in elderly individuals but have lower engraftment potential. The former might be a compensation mechanism for decreased HSC functionality [36]. Additionally, the aged hematopoietic system exhibits a myeloid differentiation bias and an increase in myeloid progenitors compared to lymphoid progenitors [37]. It is not clear whether alterations in HSC differentiation or changes in myeloid and lymphoid progenitor cell survival and proliferation are the main cause of this skewing. It is possible that the composition of the heterogeneous stem cell pool is affected in the aging system [38]. As reported in the previous chapter, different niche components might regulate HSC functions differently, balancing self-renewal and lineage output. In the aging system niche homeostasis might be disrupted, favoring distinct lineage output or expansion of the HSC pool and supporting the already intrinsic cellular changes [39]. Interestingly, older hematopoietic progenitor cells are localized more distant from the endosteum of the bone than younger cells, together with impaired adhesion to the niche cells [40]. Molecularly, genes regulating lymphoid differentiation were found to be downregulated in aged HSCs from mice,

whereas the expression of myeloid specific genes was enhanced, underlining the finding of myeloid skewing in aged progenitors [41].

Over the lifespan of an individual, cells are prone to accumulate DNA damage. In HSCs accumulation of DNA damage mainly results from physiological stress (e.g., viral infection or blood loss) that forces the cells to enter the cell cycle. The resulting replicative stress is primarily a consequence of elevated ROS in the cell. When HSCs are forced to enter the cell cycle, mitochondrial respiration leads to accumulation of ROS. Repopulation studies with human cells in mice revealed that intracellular upregulation of ROS led to accumulation of DNA damage accompanied by loss of self-renewal capacity [35]. HSCs that were forced to enter the cell cycle exhibited a reduced repopulating activity and increased myeloid lineage bias [42]. These observations are in line with the hypothesis that hematopoietic stem cells limit their replicative potential and loose stem cell function with increasing number of divisions, as described earlier in this chapter. An increased level of DNA breaks was observed in HSPCs from older mice, with the highest frequency of DNA damage in more primitive cells [43]. Similarly, an increased exposure to oxidative stress leads to accumulation of mitochondrial DNA damage, which affects mitochondrial function. This process drives premature aging of stem cells, independent of physiological aging mechanisms [44]. In contrast, progenitor cells have generally a higher turn-over rate than quiescent HSCs, therefore exhibit a higher baseline level of DNA damage, and have more efficient repair mechanisms. As a result, progenitors do not show a decreased functionality in response to physiological stress or increased age [42].

A consequence of these altered stem cell functions might be clonal skewing in hematopoiesis. It has been reported that especially in older mice a few dominant clones with myeloid differentiation bias are selectively expanded. In these studies the cell clones exhibited a higher self-renewal capacity than unbiased (or “balanced”) HSPCs [38, 45]. In humans a similar phenomenon has been described. The earliest reports of expansion of distinct clones was demonstrated in elderly women with skewed X-inactivation patterns in blood cells [46]. Somatic mutations have been identified in blood cells of older women without hematologic malignancies [47], healthy HSPCs [48], or non-leukemic cells from patients suffering from acute myeloid leukemia (AML) [49]. HSPCs affected by these somatic mutations might gain an advantage, leading to preferential expansion of this clone, an effect that is mainly seen in the aging hematopoietic system. Large sequencing studies have identified mutations in multiple genes in blood cells of elderly individuals suffering from non-hematologic diseases [50, 51] or in blood cells of healthy elderly individuals [51, 52]. Interestingly, the identified mutations were found in the same set of genes (i.e., *DNMT3A*, *TET2*, *ASXL1*, *JAK2*, *PPM1D*, and *SF3B1*) in

different cohorts of elderly individuals without a hematologic malignancy. In 2015, the term clonal hematopoiesis of indeterminate potential (CHIP) was introduced to describe individuals that do not have a hematologic disorder, but carry a detectable somatic mutation in cells of the blood or bone marrow in hematologic malignancy-associated genes [53]. Our group demonstrated that CHIP-associated mutations are found in the stem cell compartment in very primitive progenitors and led to preferential expansion of myeloid cells [54]. Random mutations occur regularly in HSPCs with 0.07-0.86 mutations/year, estimated by the replication fidelity of HSPCs [48]. The incidence of CHIP-associated mutations increases with age [48, 51, 52], suggesting that the already described changes in the aging hematopoietic system make cells more susceptible to mutation acquisition, for example by increased cell divisions. The selective advantage conferred by somatic mutations might be more pronounced in elderly individuals, leading to increased expansion and therefore detection of the clone in the peripheral blood. In the aging hematopoietic system, the incidence of somatic mutations is increased, leading to a higher risk of developing hematologic malignancies when cooperating mutations arise.

1.2 The Onset of Hematologic Malignancies

1.2.1 Preleukemia and Leukemia Stem Cells

Dysregulation of stem cell functions might be a cause for onset of hematologic malignancies. It has often been demonstrated that cancers arise in a clonal fashion, with certain mutations leading to an evolutionary advantage of the cell ([55], reviewed in [56]). It has been reasoned that in the hematopoietic system the heterogeneous HSC compartment might be a target of tumorigenic events since HSCs already have a high self-renewal potential and might therefore need fewer transforming events. With their self-renewal capacity, HSCs also have the advantage of a longer life span and therefore higher chance to accumulate more somatic mutations [57]. Chromosomal abnormalities have been found in phenotypically normal primitive HSCs in patients in long-term remission, indicating that these events are present early on in the stem cell compartment and can expand with the clone when more mutations are acquired [58, 59]. In sequencing studies, HSC populations that carried only a subset of mutations of the leukemic cell population were identified in AML patients [60]. These preleukemic HSCs have been shown to regenerate all blood lineages and have an advantage over “normal” HSCs leading to clonal expansion [49]. Preleukemic cells are able to survive chemotherapy and persist during remission. Many studies reported genes in epigenetic regulators (i.e., DNA methylation and chromatin remodeling) as being involved in the preleukemic state [61]. Based on these observations, leukemogenesis is described as a stepwise process with an initiating event occurring in HSPCs that might already carry hundreds of random background mutations. An initiating event confers advantage to the cell, leading to an expansion of this cell clone carrying background mutations. Over time more mutations are acquired, including driver mutations that lead to the transformation of this clone. In AML cases with unknown drivers, mutations in *DNMT3A* and *IDH1* have been identified as initiating events and these mutations might cooperate in the initiation process [48]. In the same study, several mutations were found in AML with known drivers as wells as in cases with unknown drivers, indicating that these mutations cooperate with the initiating events in transforming the clone (e.g., internal tandem duplication of *FLT3*, *FLT3*-ITD). Similar observations have been made in other sequencing studies: mutations in epigenetic regulators (e.g., *DNMT3A*, *IDH2*) occur before mutations in *FLT3*, confirming the stepwise process of leukemogenesis [49]. Preleukemic cells have been described as functionally normal but are more likely to transform into leukemia. The identification of preleukemic events has substantially added to the understanding of leukemogenesis, which is essential to identify possible treatment regimens.

1.2.2 Acute Myeloid Leukemia

Leukemia arises from HSCs with acquired or inherited genetic abnormalities. The Robert-Koch Institut reported just over 6,000 leukemia cases in women and 7,900 leukemia cases in men in Germany in the year 2016, leading to an age standardized incidence rate of 8.6 in women and 13.5 in men per 100,000 inhabitants [62]. The median age at diagnosis was 74 for women and 71 for men, with 4% of patients being under 15. In children and young adults, the incidence of leukemia decreases with increasing age. Over the age of 30 the relative risk of leukemia onset increases with higher incidence in men (Figure 1.2).

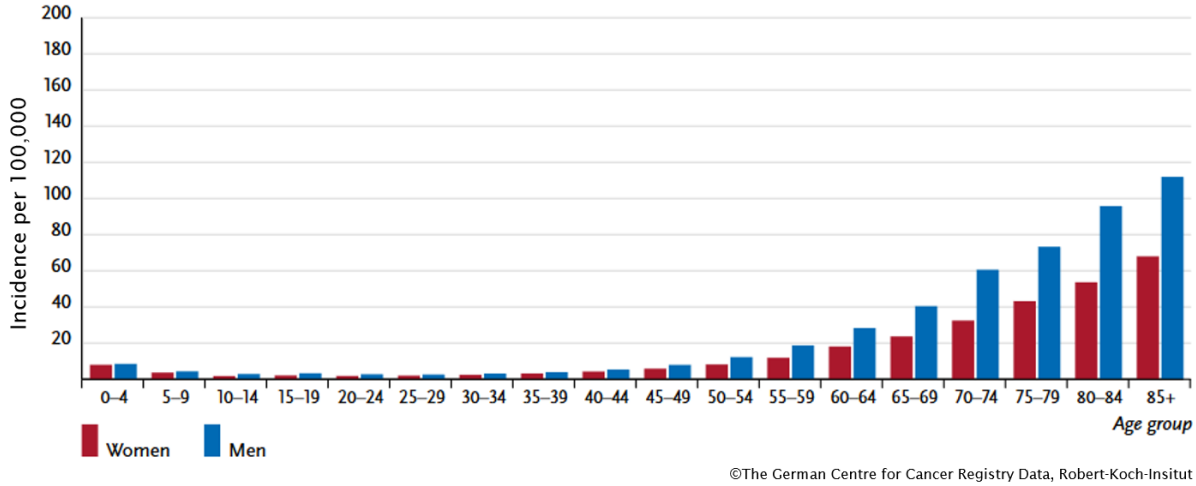


Figure 1.2 Age-specific leukemia incidence rate per sex. Incidence for 2015-2016 in Germany are shown. The graph was modified from The German Centre for Cancer Registry Data at the Robert-Koch-Institut [62].

Especially acute lymphocytic leukemia (ALL) has a higher incidence in children and decreases with age, whereas AML incidence drastically increases in elderly people [63]. In men 39% of newly diagnosed leukemia was chronic lymphocytic leukemia (CLL), 22% AML, 9% chronic myeloid leukemia (CML), and 6% ALL. In women the numbers were comparable, with slightly more AML and less CLL cases. The 5-year survival rate was around 57% for all leukemia subtypes in men and women. The highest 5-year survival rate had female CML patients with 86% and male CLL patients with 83%. The lowest 5-year survival rate was found in AML patients with 21% for both sexes in the overall population [62]. In AML patients between 65-74 years of age the 5-year survival rate is just above 16% and in patients older than 75 years below 5% [63].

AML is characterized by impaired hematopoiesis and bone marrow failure as a result of clonal, abnormally differentiated, and proliferative hematopoietic cells, especially myeloid cells [64]. Most patients show symptoms of bone marrow failure, like anemia and thrombocytopenia, but also fatigue, anorexia, and weight loss [65]. AML diagnosis is based on the identification of blasts (≥20%) in the bone marrow or peripheral blood by morphologic evaluation as well as flow cytometry of surface markers. Additional molecular and cytogenetic analysis is used to determine the subtype of AML, which is essential for evaluation of prognosis and treatment regimen [66]. Based on the world health organization classification from 2016 of AML and related neoplasms, AML is classified into AML with recurrent genetic abnormalities with many more subgroups based on translocations or cytogenetic aberrations, AML with myelodysplasia-related features, therapy-related myeloid neoplasms, AML not otherwise specified (AML, NOS), as well as myeloid sarcoma, and myeloid proliferations related to down syndrome [67].

Prognostic factors that explain variations in treatment outcome are patient-specific demographic as well as clinical parameters, like age, general health, performance status, and comorbidities. In older patients the outcome is generally poorer, partly due to age-specific mutations that increase resistance to chemotherapy [68]. More importantly, genetic abnormalities have been identified to predict disease outcome [69, 70]. AML patients are often mutated in *FLT3* (either an internal tandem duplication, ITD, or tyrosine kinase domain mutation, TKD, 28% of patients), *KRAS* and *NRAS* (12%), *TP53* (8%), or *KIT* (4%). Mutations in these genes are usually associated with active proliferation. As described above, these mutations often occur in combination with other initiating mutations that are associated with impaired hematopoiesis and differentiation (for example in *NPM1*, *CEBPA*, but also epigenetic regulators) [71]. AML patients are grouped into three risk groups (favorable, intermediate, and adverse) to predict treatment outcome based on cytogenetic and molecular alterations (Table 1.1). Poor prognosis is also marked by mutations in *RUNX1* or *ASXL1*, but only if they do not co-occur with favorable prognostic markers. Both mutations are associated with older age [69].

Table 1.1 Classification of AML based on cytogenetic and molecular abnormalities. Prognostic risk groups are listed as reviewed by De Kourchkovsky and Abdul-Hay 2016 [65] and Döhner et al. 2017 [68] CN-AML: cytogenetically normal AML. *without *FLT*3-ITD or *FLT3*-ITD^{low} (allelic ratio <0.5)

Prognostic-risk group	Cytogenetic profile and molecular abnormalities
Favorable	t(8;21)(q22;q22.1) without *KIT*^{mut}
	inv(16)(p13.1q22)
	t(15;17)(q22;q12)
	CN-AML: *NPM1*^{mut} without *FLT3*-ITD *
	CN-AML: biallelic *CEBPA*^{mut}
Intermediate	t(8;21)(q22;q22) with *KIT*^{mut}
	CN-AML: *NPM1*^{mut} with *FLT3*-ITD^{high}
	CN-AML: *NPM1*^{WT} without *FLT3*-ITD *
	t(9;11)(p21.3;q23.3)
	Cytogenetic abnormalities (not included in other groups)
Adverse	*TP53*^{mut}, regardless of cytogenetic profile
	CN with *FLT3*-ITDhigh
	CN with *DNMT3A*^{mut}
	CN with *KMT2A*-PTD
	inv(3)(q21q26.2)
	t(6;9)(p23;q34.1)
	11q abnormalities other than t(9;11)
	−5 or del(5q)
	−7
	Complex karyotype

The first phase of therapy is generally induction therapy with the purpose to achieve morphologic complete remission with reduction of blasts to an undetectable level. Intensive induction chemotherapy is based on treatment with anthracycline and cytarabine [66]. Complete remission is achieved in 60-85% of younger adults and 40-60% of patients older than 60 years [64]. Consolidation therapy serves the purpose to prevent early relapse by controlling minimal residual disease. Either consolidation therapy is managed by chemotherapy (intermediate dose cytarabine) or allogeneic hematopoietic stem cell transplantation. The latter is particularly recommended in patients who are unlikely to have extended complete remission [68]. Changes in treatment regimens are tested continuously but will not be elucidated further in this short overview. Targeted therapy has been shown effective for distinct AML subgroups. For example the multitarget kinase inhibitor midostaurin has been approved for *FLT3*-mutated AML in combination with regular induction and consolidation treatment regimens [72] with improvement of event-free and overall survival [73]. *FLT3*-mutated AML has a poor prognosis

and is associated with high relapse rates [74, 75]. Characterization of AML patient groups with distinct features will lead to a better understanding of patient treatment outcomes and identify possible targets for future personalized therapies.

1.2.3 Core Binding Factor Acute Myeloid Leukemia

Core binding factor acute myeloid leukemia is a common name for a subgroup of AML with cytogenetic aberrations, namely t(8;21)(q22;q22) and inv(16)(p13q22)/t(16;16)(p13;q22). The translocations lead to the formation of fusion genes *RUNX1-RUNX1T1* [76] and *CBFB-MYH11* [77], respectively. The core binding factor is comprised of 3 α subunits (termed Runt related transcription factor 1-3, RUNX1-3), which binds a distinct DNA motif, and one β subunit (core-binding factor β, CBFβ), which stabilizes the interaction with DNA [78]. In normal hematopoiesis RUNX1 and CBFβ are involved in regulation of early and definitive hematopoiesis [79]. The fusion of the Runt domain of *RUNX1* (on chromosome 21) and *RUNXT1* (formerly known as *ETO*, on chromosome 8, Figure 1.3) result in a protein that acts as dominant negative inhibitor of RUNX1- and CBF-mediated transcription leading to impaired hematopoietic development and differentiation [80].

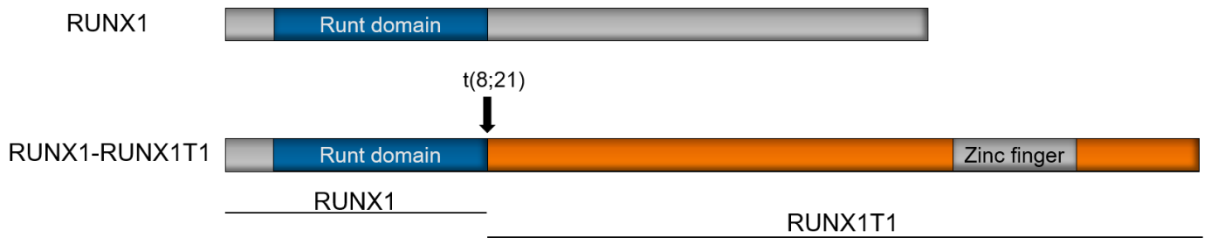


Figure 1.3 Domains of RUNX1 and the fusion product RUNX1-RUNX1T1. The upper schematic represents the wild type RUNX1 protein. The lower schematic depicts the fusion product RUNX1-RUNX1T1 with the translocation site (indicated by a black arrow). Blue: Runt1 DNA-binding domain. Orange: RUNX1T1 protein with its Zinc finger domain (dark grey).

This deregulation is not sufficient to induce leukemia but might serve as an initiating event that primes cells for leukemic transformation. RUNX1-RUNX1T1 has been found in patients before the clinical onset of AML, as well as in patients in clinical remission [59], supporting this hypothesis. Accompanying mutations are therefore the actual drivers of leukemia. Patients with AML t(8;21) are grouped into the favorable risk group with low primary drug resistance and high remission rates (~95%) as well as 5-year survival (over 50%) [81]. Nevertheless, the relapse rates remain high and survival after relapse is low [82].

1.3 Epigenetic Regulation

1.3.1 Epigenetic Regulation in the Hematopoietic System

Epigenetic regulation is crucial in developmental as well as adult hematopoiesis. Since many genes involved in hematologic malignancies or preleukemic conditions are epigenetic regulators, this chapter will elucidate the importance of epigenetic regulation in the hematopoietic system in more detail. As described in the beginning, environmental cues are translated into intrinsic signaling networks, which in turn control transcription factors and the epigenetic state in the cell. Expression of genes is regulated via chromatin accessibility and it has been demonstrated that DNA methylation factors and chromatin modifiers are essential for unperturbed hematopoiesis.

Chromatin is organized in nucleosomes, consisting of DNA and histone octamers. Posttranslational modifications at N-terminal tails of histones lead to conformation changes of DNA and changing accessibility of regulatory and genetic regions [83]. Histone modification generally involves acetylation, methylation, phosphorylation, ADP-ribosylation, sumoylation, and ubiquitination. While acetylation of lysine residues is associated with active transcription and deacetylation with gene repression, histone methylation is more complex depending on the residue and the number of methyl groups transferred. Histone modification is tightly linked with the methylation status of the DNA. In mammalian cells, DNA methyltransferases catalyze the addition of a methyl group to the fifth carbon of the cytidine ring, primarily within CpG dinucleotides. CpG dinucleotides are specifically enriched (compared to their distribution over the rest of the genome) in regulatory regions of genes, for example promotor regions. These enriched segments are called CpG islands [84]. DNA methyltransferase (DNMT) 1 is essential for maintaining methylation marks after cell division and targets hemimethylated DNA. Unmethylated DNA is targeted by DNMT3A and DNMT3B which are *de novo* DNA methyltransferases [85, 86]. CpG islands in regulatory regions are mostly unmethylated. During development, hematopoiesis, and hematopoietic differentiation, the methylation status of regulatory regions is modified. Methylated DNA is in turn a target for proteins with methyl-CpG-binding domains. These proteins are then responsible for the recruitment of histone methyltransferases (HMT) and histone deacetylases (HDAC). By applying specific marks at amino acid residues of histones, the chromatin gets a more compact conformation, and gene expression is repressed. Conversely, unmethylated DNA is associated with an open chromatin conformation and expression of genes. Active DNA demethylation plays an important role in the switch from hematopoietic stem cell maintenance to proliferation and differentiation, as

well as direct lineage specification. Ten-eleven translocation (TET) family of proteins have been identified to catalyze the demethylation of 5-methylcytosine (5mC) [87, 88]. Generally, epigenetic regulation is important for balancing self-renewal and differentiation. DNA methylation prevents premature differentiation. Members of polycomb repressive complexes, like BMI1 (B lymphoma Mo-MLV insertion region 1 homolog), have been shown to be essential for early hematopoietic development as well as for maintenance of HSC self-renewal [89]. The maintenance methyltransferase DNMT1 prevents early myeloid differentiation, which has been demonstrated in mice with *Dnmt1* knockout cells showing upregulation in myeloid specific genes and increased myeloid cell output. In normal hematopoiesis, restriction of myeloid differentiation seems to be essential to permit lymphoid lineage output [90]. DNA methylation is dynamic, with increased methylation in promotor regions in stem cell specific genes (e.g., *HOXA9, HOXB5*) and decreasing methylation in lineage specific genes in more differentiated progenitor populations [91].

Mutations in epigenetic regulators have been found in various hematologic malignancies (summarized in [92]), as well as in individuals without history of blood cancers (see chapter 1.1.3). It seems therefore highly likely (and has often been hypothesized) that mutations in these regulators pay a role in the onset of leukemia. In the next chapter, the role of some genes most effected in hematologic malignancies as well as CHIP will be reviewed in more detail.

1.3.2 Mutations in Epigenetic Regulators and Aberrant Hematopoiesis

1.3.2.1 DNMT3A

DNMT3A is one of the two *de novo* methyltransferases identified in mammals [86]. Mutations in *DNMT3A* are often found in hematopoietic cells and in association with hematologic malignancies [93, 94], with the implication of being early events [49]. Encoded on chromosome 2p23, *DNMT3A* consists of 23 exons that are transcribed in different splice forms. *DNMT3A1* is a full length transcript that has an alternative transcription start site, *DNMT3A2* is an N-terminal truncated isoform and expressed in embryonic stem cells [95]. The catalytic domain with methyltransferase activity is located at the C-terminus, is dependent on binding to S-adenosyl methionine, and transfers a methyl group at the C5 position of cytidine. The regulatory regions consist of the DNA-binding PWWP-domain and the ADD domain, which is responsible for interaction with histone tails (Figure 1.4). In homeostasis, DNMT3A acts as an oligomer of either heterodimeric DNMT3A-DNMT3L or DNMT3A-homodimers. The different domains of DNMT3A have been demonstrated to interact with various binding

partners on protein, histone, and DNA level, to modulate its catalytic activity and influence DNA methylation patterns [96].

With the help of knockout studies, it has been demonstrated that *Dnmt3a*$^{-/-}$ mice do not survive longer than one month [86]. Conditional deletion of *Dnmt3a* has shown, that mouse HSCs shift towards self-renewal and hinder differentiation, with preferential expansion of *Dnmt3a* knockout HSCs over wild type HSCs. *Dnmt3a*-null HSCs have the same phenotype as wild type HSCs and are able to generate all lineages, but the overall output of differentiated cells is compromised [97]. Loss of *Dnmt3a* in mouse HSCs leads to loss of DNA methylation, especially at genes regulating self-renewal, like *Hoxa9* and *Meis1*. In more differentiated *Dnmt3a*-null progenitors, these self-renewal and multipotency genes are not properly repressed, leading to disturbed differentiation [97, 98]. *Dnmt3a* knockout (KO) HSCs are immortalized, meaning they can be transplanted into mice and generate more HSCs, but are not capable of producing differentiated progeny [99].

DNMT3A has first been found mutated in AML patients by targeted sequencing of tumor samples in around 20% of patients investigated [100, 101]. The majority of *DNMT3A* mutations are found in the catalytic domain affecting amino acid Arginine 822 (R882, Figure 1.4).

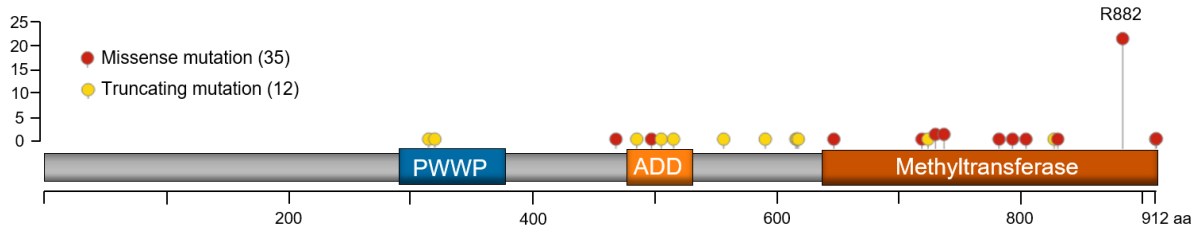


Figure 1.4 Mutations in *DNMT3A*. Lollipop plot depicting the mutations in *DNMT3A* found in the TCGA AML cohort [71]. The data were accessed via cbioportal [102, 103]. PWWP: Proline-Tryptophan-Tryptophan-Proline domain, ADD: ATRX-Dnmt3-Dnmt3L domain, aa: amino acid.

This specific mutation impairs the ability to form homodimers and decreases catalytic activity [104]. It has been hypothesized that disruption of *DNMT3A* function results in diverse effects depending on the differentiation state of the cell [105]. *DNMT3A* mutations have been found in preleukemic HSCs in patients with AML at diagnosis, in remission, and relapse [49]. Preleukemic cells carrying *DNMT3A* mutations survive induction chemotherapy, are able to reconstitute the bone marrow after therapy, and generate clonal mature populations [61]. In individuals with CHIP, *DNMT3A* is most commonly affected and mutations are found in various cells of the hematopoietic differentiation tree [54]. CHIP does not per se lead to the onset of leukemia [51, 52], but increases the risk of hematologic malignancies, indicating that clones carrying mutations for example in *DNMT3A* are more likely to transform if given enough time.

1.3.2.2 TET2

The function of ten-eleven translocation (TET) enzymes was only discovered after it has been found mutated in myeloid malignancies [106]. All three family members, TET1, TET2, and TET3, actively convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), and are therefore DNA oxidases [87]. The process is Fe(II) and α-ketoglutarate dependent. TET proteins can catalyze the oxidation of 5mC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) [107]. The formation of intermediate products such as 5hmC can lead to active and passive DNA demethylation. The latter follows replication when DNMT1 is not able to recognize 5hmC and does not maintain the methylation mark at this specific cytosine residue [108]. Active DNA demethylation is realized via DNA repair pathways. First, 5hmC is converted to 5-hydroxymethyluracil (5hmU) by activation induced cytidine deaminase (AID)/apolipoprotein B editing complex (APOBEC), then leading to base excision repair (BER) pathway dependent conversion to unmethylated cytosine. Similarly, 5caC is converted into unmethylated cytosine via BER. The enzymatic function of TET2 is localized between amino acids (aa) 1129 and 1936 (of total 2002 aa) within cysteine rich regions and a double-stranded β helix [109].

TET2 mutations have been identified in individuals with CHIP, although not as frequently as *DNMT3A* mutations [50–52], as well as myelodysplastic syndrome (MDS)/myeloproliferative neoplasms (MPN) [110, 111] and AML in 10-20% of the patients [112–114]. The mutation status of *TET2* can function as a prognostic marker, leading to poorer overall survival in intermediate-risk AML [70, 114]. In myeloid cancers, mutations in *TET2* have been connected with impaired 5mC hydroxylation [115]. The role of *TET2* mutations and especially functional knockouts has been investigated intensively in mouse studies. In multipotent progenitors as well as myeloid progenitors isolated from mice, *Tet2* expression is high and silencing of *Tet2* led to an increase in myeloid cell production [115]. Loss of *Tet2* led to a disruption of hematopoiesis with an upregulation of HSC self-renewal and expansion of the stem and progenitor cell compartment [116–118]. Another study demonstrated that silencing of *Aid* led to a similar effect, connecting impaired DNA demethylation with *Tet2*-knockdown phenotype [119]. Nevertheless, it seems that *TET2* mutations alone are not drivers of hematologic malignancies and cooperating mutations have often been detected [69]. These observations indicate that alterations in *TET2* might play a similar role as *DNMT3A* mutations, increasing the risk of leukemic transformation in the stem cell compartment.

1.3.2.3 ASXL1

Additional sex combs-like (ASXL) 1 is, in contrast to TET2 and DNMT3A, not involved in DNA methylation and demethylation but functions as chromatin remodeler by interacting with polycomb repressive complexes (PRC). The mammalian regulators are paralogs of *Drosophila melanogaster* Additional sex combs (Asx), which regulates *Hox* gene expression [120]. The human *ASXL1* gene is located on chromosome 20q11 and encodes a protein consisting of three functional domains. The ASXN domain forms the N-terminus and is essential for DNA-binding of the protein. An ASX homology (ASXH) domain, also called deubiquitinase adaptor, is located in the N-terminal region and binds deubiquitinase BRCA1 (breast cancer susceptibility type 1 protein)-associated protein 1 (BAP1). A plant homeodomain (PHD) finger forms the C-terminus and is essential for recognition of histone modifications [121].

ASXL1 has diverse roles in histone modification and chromatin remodeling. Disruption of *Asxl1* in mice led to impaired lymphopoiesis as well as myelopoiesis, but is not sufficient to induce leukemia [122]. Other groups reported that heterozygous loss of *Asxl1* led to development of MDS/MPN-like conditions [123]. Deletion of *Asxl1* in hematopoietic cells resulted in a loss of the repressive trimethylation of lysine 27 on histone 3, for example in the vicinity of the *HoxA* gene cluster. *HoxA* gene expression is in turn upregulated in *Asxl1* mutated cells, disrupting myelopoiesis [124]. A C-terminal truncating mutation of *Asxl1* is associated with clonal advantage of the cell, promoting leukemic transformation [125]. *ASXL1* is often mutated in MDS [126, 127], MPN [128], and AML [129, 130] and has also been found in individuals with CHIP [51, 52]. In AML patients, *ASXL1* is more often mutated in older patients [131, 132], as well as in patients with t(8;21) [132]. The majority of *ASXL1* mutations are nonsense or frameshift mutations in the last exon (exon 12 in hg19, exon 13 in hg38). The C-terminal truncation of *ASXL1* has been investigated intensively in recent years. Some groups demonstrated that mutant *Asxl1* interfered with the catalytic activity of PRC2 [133], whereas another group found that truncated *Asxl1* led to development of myeloid malignancies in mice [134]. Several groups demonstrated that knock-in of truncated *Asxl1* in mice did not lead to transformation but disrupts hematopoiesis by interfering with chromatin regulation of PRC 1/2 [135–137].

2 Aim of the Thesis

In this thesis, the goal is to investigate the impact of mutations in *DNMT3A, TET2,* and *ASXL1* on the hematopoietic stem and progenitor cell compartment. In mouse models, it has been demonstrated that knockouts of either of the three genes (also termed DTA mutations), disrupt normal hematopoiesis, lead to expansion of the stem cell pool, and favor skewed differentiation. Studies on human cells have been carried out in xenograft mouse models or as retrospective studies of cell expansion in elderly individuals or cancer patients.

In healthy, young hematopoietic stem and progenitor cells from umbilical cord blood, genetic modifications will be introduced using CRISPR/Cas9 and the cells will be monitored in various *in vitro* culture systems regarding altered phenotype, self-renewal properties, differentiation capacities, and long-term culture initiating potential. The advantage of cord blood cells is the low rate of DNA damage due to the young age and therefore few numbers of replications. The cells have not been exposed to various environmental factors, making it possible to study the unperturbed effects of the mutations themselves. To dissect if the mutations lead to clonal expansion in *in vitro* settings, I will evaluate the introduced mutations and clonal composition of the samples. In this thesis, the hypothesis that DTA mutations have a direct impact on the human stem and progenitor cell compartment and lead to preferential expansion of the investigated cells will be tested.

In a second, independent part of the project, I will investigate AML patients with t(8;21) to dissect the mutational composition and disease progression retrospectively. Samples from the time point of diagnosis will be screened for mutations using a targeted gene panel. For some patients, samples from clinical remission will be re-sequenced to elucidate whether distinct mutations are present after induction therapy. Here, I will investigate whether distinct clones are able to survive chemotherapy and can be drivers of disease progression. Both parts will elucidate the role of DTA mutations in the onset and development of diseases of hematopoietic stem and progenitor cells, including myeloid malignancies in general and AML especially. By characterizing the functional impact of these gene targets, disease monitoring can be improved, and distinct therapy options evaluated.

3 Material and Methods

3.1 Material

3.1.1 Equipment

Name	Provider
2100 Bioanalyzer Instrument	Agilent
4200 TapeStation System	Agilent
Amaxa™ Nucleofector™ II	Lonza
AlphaImager EC	AlphaInnotech
BD FACSAria™ II SORP	BD Biosciences
BD FACSCanto™ II Cell Analyzer	BD Biosciences
Biological Safety Cabinet Class II	Nuaire
Bioruptor Plus	Diagenode
Cell culture incubator	Binder
Centrifuge	various models
Heating block/shaker	various models
ImageQuant™ LAS 4000	GE Healthcare
Incucyte S3	Sartorius
KS 4000 i control (incubator shaker)	IKA systems
LUNA™ Automated Cell Counter	Logos Biosystems
Microbiological incubator Heraeus function line	Thermo Scientific
Microscope	various models
MiSeq sequencer	Illumina®
MJ Mini PCR cycler	Bio-Rad
NanoDrop 1000 Spectrophotometer	Thermo Scientific
Neon™ Transfection System	Invitrogen™
NextSeq 550 sequencer	Illumina®
Pipette controller	various models
BioShake XP	QInstruments
Quantus™ Fluorometer	Promega
QX200 droplet generator	Bio-Rad
QX200 droplet reader	Bio-Rad
S3e sorter	Bio-Rad
StepOnePlus™ Real-Time PCR System	Applied Biosystems™
Vacusafe vacuum pump	Integra
Water bath	various models

3.1.2 Consumables

Name	Supplier (Catalog No.)
8 Well PCR tube strips plus strips of 8 domed caps	4ti-0780
96 Well Non-Skirted PCR Plate	4ti-0750
Abgene™ 96 Well 0.8mL Polypropylene Deepwell Storage Plate	ThermoFisher Scientific (AB0859)
Biosphere® Filter Tip 10 µl	Sarstedt (70.1130.210)
Biosphere® Filter Tip 100 µl	Sarstedt (70.760.212)
Biosphere® Filter Tip 1000 µl	Sarstedt (70.762.211)
Biosphere® Filter Tip 20 µl	Sarstedt (70.760.213)
Biosphere® Filter Tip 200 µl	Sarstedt (70.760.211)
Combitips advanced® 0.1 ml	Eppendorf (0030089405)
Combitips advanced® 0.2 ml	Eppendorf (0030089413)
Combitips advanced® 0.5 ml	Eppendorf (0030089421)
Combitips advanced® Biopur® 0.2 ml	Eppendorf (0030089626)
Combitips advanced® Biopur® 1.0 ml	Eppendorf (0030089642)
Combitips advanced® Biopur® 5.0 ml	Eppendorf (0030089677)
Cord blood collection bag	Macopharma (MSC1200PU)
CryoPure Tube 1.6ml white	Sarstedt (72.380)
Disposable needles Sterican® long bevel facet, 70 mm, 0.90 mm	Carl Roth (C722.1)
Domed 8 well cap strips	4ti-0752
Falcon® 1 mL Serological Pipet	Corning (357521)
Falcon® 10 mL Serological Pipet	Corning (357551)
Falcon® 100 µm Cell Strainer	Corning (352360)
Falcon® 12-well Clear Flat Bottom TC-treated Multiwell Cell Culture Plate	Corning (353043)
Falcon® 15 mL High Clarity PP Centrifuge Tube	Corning (352096)
Falcon® 2 mL Serological Pipet	Corning (357507)
Falcon® 24-well Clear Flat Bottom TC-treated Multiwell Cell Culture Plate	Corning (353047)
Falcon® 25 mL Serological Pipet	Corning (357525)
Falcon® 35 mm TC-treated Easy-Grip Style Cell Culture Dish	Corning (353001)
Falcon® 40 µm Cell Strainer	Corning (352340)
Falcon® 5 mL Round Bottom Polystyrene Test Tube	Corning (352052)
Falcon® 5 mL Round Bottom Polystyrene Test Tube, with Cell Strainer Snap Cap	Corning (352235)
Falcon® 5 mL Serological Pipet	Corning (357543)
Falcon® 50 mL High Clarity PP Centrifuge Tube	Corning (352070)

Falcon® 50 mL Serological Pipet	Corning (357550)
Falcon® 6-well Clear Flat Bottom TC-treated Multiwell Cell Culture Plate	Corning (353046)
Falcon® 96-well Clear Flat Bottom TC-treated Culture Microplate	Corning (353072)
Falcon® 96-well Clear Round Bottom TC-treated Cell Culture Microplate	Corning (353077)
Folded filters ROTILABO® Type: 113P	Carl Roth (CA10.1)
Hypodermic needle 16G 1,60x40mm	VWR (613-2032)
Inoculation spreader	Sarstedt (86.1569.005)
LUNA™ Cell Counting Slides	Logos Biosystems (L12001)
MACS LS Columns	Miltenyi Biotec (130-042-401)
MACS MS Columns	Miltenyi Biotec (130-042-201)
Micro tube 2.0ml	Sarstedt (72.691)
MicroAmp™ Fast Optical 96-Well Reaction Plate with Barcode	Applied Biosystems (4346906)
MicroAmp™ Optical Adhesive Film	Applied Biosystems (4311971)
Neoprene gloves	VWR (112-0027)
Nitrocellulose membrane Amersham™ Protran® Premium	GE Healthcare (10600003)
Nunc EasYFlask 25cm^2	ThermoFisher Scientific (156340)
Nunc EasYFlask 75cm^2	ThermoFisher Scientific (156499)
Nunc™ 96-Well Polypropylene DeepWell™ Storage Plates	ThermoFisher Scientific (260251)
Nylon membrane Amersham Hybond-N+	GE Healthcare (RPN203B)
Petri dish 92x16mm	Sarstedt (82.1472)
Reagent Reservoirs, 25 ml	Carl Roth (EKT7.1)
SafeSeal tube 1.5ml	Sarstedt (72.706)
Spritzenfilter ROTILABO® PVDF, 0,22 µm	Carl Roth (P666.1)
TapeStation 4200 Loading Tips	Agilent (5067-5599)
Tube 13ml, 100x16mm, PP	Sarstedt (62.515.006)
Vasco® Nitril blue glove S	B. Braun (9209817)
Vasco® Nitril white glove S	B. Braun (9208410)

3.1.3 Chemicals and Reagents

Reagent/Chemical	Supplier (Cat. No.)
2-Mercaptoethanol	AppliChem (A1108)
2-Propanol	Carl Roth (6752)
4',6-Diamidino-2-Phenylindole, Dilactate (DAPI)	BioLegend (422801)
Albumin Fraktion V	Carl Roth (8076)
Amersham ECL Prime Western Blotting Detection Reagent	GE Healthcare (RPN2232)
Ammonium acetate	Fluka (9688)
Ammonium peroxydisulfate (APS)	Carl Roth (9592)
Ampicillin sodium salt	Carl Roth (HP62)
AMPure XP	Beckman Coulter (A63881)
BD FACSFlow™ Sheath Fluid	BD Bioscience (342003)
BD FACS™ Clean Solution	BD Bioscience (340345)
BD FACS™ Rinse Solution	BD Bioscience (340346)
BD FACS™ Shutdown Solution	BD Bioscience (334224)
Biocoll separating solution	Merck (L6115)
Bovine serum albumin, heat shock fraction	Sigma Aldrich (A7906)
Collagen coating solution	Sigma Aldrich (125-50)
cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail	Roche (4693159001)
ddPCR™ Droplet Generator Oil for Probes	Bio-Rad (186-3005)
ddPCR™ Supermix for Probes (No dUTP)	Bio-Rad (186-3025)
Dimethyl sulfoxide (DMSO)	Sigma Aldrich (41639)
DMEM	Gibco™ (11960044)
dNTP Set (100mM)	Invitrogen™ (10297018)
Ethanol Rotipuran 99.8% p.a.	Carl Roth (9065.2)
Ethylene diamine tetraacetic acid (EDTA)	Carl Roth (8040.3)
Fetal Bovine Serum	Sigma Aldrich (F7524)
GelRed Nucleic Acid Gel Stain	VWR (730-2957)
GeneRuler 100 bp Plus DNA Ladder	Thermo Scientific™ (SM0322)
Glycin Pufferan 99% p.A.	Carl Roth (3908.3)
Glycerol Ultra Pure	Invitrogen™ (15514-011)
Hydrocortisone	Sigma-Aldrich (H0888)
Hygromycin B (50mg/ml)	Gibco™ (10687010)
IMDM	Gibco™ (21980065)
LB-Agar (Luria/Miller)	Carl Roth (X969.1)
LB-Medium (Luria/Miller)	Carl Roth (X968.1)
MEM α, nucleosides	Gibco™ (12571063)

Methanol	Carl Roth (CP43.3)
MethoCult™ H4434 Classic	STEMCELL Technologies (04434)
MethoCult™ H4435 Enriched	STEMCELL Technologies (04435)
MyeloCult™ H5100	STEMCELL Technologies (05150)
N,N,N′,N′-Tetramethylethylenediamine (TEMED)	Carl Roth (2367.3)
Nuclease-Free Water	Invitrogen™ (AM9937)
PBS, pH 7.2	Gibco™ (20012068)
Penicillin-Streptomycin (10,000 U/mL)	Gibco™ (15140122)
PhosSTOP	Sigma-Aldrich (4906845001)
Ponceau S	Carl Roth (5938.1)
Potassium acetate	Sigma-Aldrich (P1190)
Propidium iodide	Sigma-Aldrich (P4864)
Protein Marker VI	AppliChem (A8889)
Proteinase K	Fisher Scientific (BP1700-100)
Random Primers	Invitrogen™ (48190011)
Recombinant human Flt3-Ligand	PeproTech (300-19)
Recombinant human IL-3	PeproTech (200-03)
Recombinant human IL-6	PeproTech (200-06)
Recombinant human SCF	PeproTech (300-07)
Recombinant human TPO	PeproTech (300-18)
RNase A Solution	Promega (A7973)
Rotiphorese® 50x TAE Puffer	Carl Roth (CL86.1)
RPMI 1640 Medium	Gibco™ (21875034)
Sodium chloride	Carl Roth (3957.1)
Sodium hydroxide	Carl Roth (6771.2)
StemSpan™ SFEM II	STEMCELL Technologies (09655)
TRIS	Carl Roth (5429.2)
Triton® X 100	Carl Roth (3051.2)
Trypan Blue Solution	VWR (K940)
Trypsin-EDTA (0.05%), phenol red	Gibco™ (25300054)
Trypsin-EDTA (0.25%), phenol red	Gibco™ (25200056)
Trypsin-EDTA (0.5%)	Gibco™ (15400054)
UltraPure™ Agarose	Invitrogen™ (16500500)

3.1.4 Kits and Enzymes

Kit	Supplier (Cat. No.)
Agilent DNA 1000 Kit	Agilent (5067-1504)
Agilent High Sensitivity DNA Kit	Agilent (5067-4626)
AllPrep DNA/RNA Mini Kit	Qiagen (80204)
BsmBI	New England BioLabs® (R0580)
Cell Line Nucleofector™ Kit V	Lonza (VACA-1003)
ddPCR™ KRAS G12/G13 Screening Kit	Bio-Rad (863506)
DNA Clean & Concentrator-5	Zymo Research (D4013)
HotStarTaq DNA Polymerase	Qiagen (203203)
KAPA HiFi HotStart ReadyMix	Roche (KK2601)
Luna® Universal qPCR Master Mix	New England BioLabs® (M3003)
Methylated DNA Standard Kit	Active Motif ® (55008)
MiSeq® Reagent Kit v2 300 cycles	Illumina (MS-102-2002)
M-MLV Reverse Transcriptase	Invitrogen™ (28025021)
NEB Next Multiplex Oligos For Illumina (96)	New England BioLabs® (E7600S)
NEB Next Ultra DNA Library Prep Kit for Illumina	New England BioLabs® (E7370L)
Neon™ Transfection 10 µl-Kit	Invitrogen™ (MPK1096)
NextSeq® 500/550 Mid Output Kit v2 (300 cycles)	Illumina (FC-404-2003)
NucleoBond Xtra Midi Plus Kit	Macherey Nagel (740412.50)
NucleoSpin Tissue XS Kit	Macherey Nagel (740901.50)
Pierce™ BCA™ Protein-Assay	Thermo Scientific™ (23227)
PrimePCR™ ddPCR™ Mutation Assay: *DNMT3A* p.R882H, Human	Bio-Rad (dHsaCP2000089)
PrimePCR™ ddPCR™ Mutation Assay: *DNMT3A* WT for p.R882H, Human	Bio-Rad (dHsaCP2000090)
PrimePCR™ ddPCR™ Mutation Assay: *JAK2* p.V617F, Human	Bio-Rad (dHsaCP2000061)
PrimePCR™ ddPCR™ Mutation Assay: *JAK2* WT for p.V617F, Human	Bio-Rad (dHsaCP2000062)
QIAmp DNA Mini Kit	Qiagen (51306)
QIAquick Gel Extraction Kit	Qiagen (28115)
QIAquick PCR Purification Kit	Qiagen (28104)
QuantiFluor® dsDNA System	Promega (E2670)
QuantiFluor® RNA System	Promega (E3310)
RNA Clean & Concentrator-5	Zymo Research (R1013)
RNase-free DNase Set	Qiagen (79254)
RNeasy Mini Kit	Qiagen (74104)
T7 endonuclease I	New England BioLabs® (M0302L)
TapeStation D1000 ScreenTape System	Agilent (5067-5582/5583/5586)
TapeStation High Sensitivity D1000 ScreenTape System	Agilent (5067-5584/5585/5587)
TruSeq Custom Amplicon Kit (96 samples)	Illumina (FC-130-1001)

3.1.5 Plasmids

Plasmid	Supplier (No.)
pCas9_GFP	gift from Kiran Musunuru (Addgene plasmid # 44719) [138]
pLKO5.sgRNA.EFS.tRFP	gift from Benjamin Ebert (Addgene plasmid # 57823) [139]

3.1.6 Oligonucleotides

Table 3.1 PCR and sequencing primers. Primers were designed using Benchling [140] or Primer3plus [141]. The M13-tag was added to the sequence manually.

Name	Sequence
*DNMT3A*_ex14_M13_F	tgtaaaacgacggccagtGCTGCTCTTTGGTTCTGTCC
*DNMT3A*_ex14_M13_R	caggaaacagctatgaccAGGTGTGCTACCTGGAATGG
*DNMT3A*_ex6_M13_F	tgtaaaacgacggccagtTGTGGGAAGGAGAGGAAGTG
*DNMT3A*_ex6_M13_R	caggaaacagctatgaccGCTGAAGGAGCAGATGAACC
*DNMT3A*_ex23_M13_F	tgtaaaacgacggccagtGCACATGGTTGGTGATCTGA
*DNMT3A*_ex23_M13_R	caggaaacagctatgaccGGAAGGGAGCTTGGTTTTGT
*DNMT3A*_ex4_F	tgtaaaacgacggccagtAAACAGGCTAAGCCCACTGA
*DNMT3A*_ex4_R	caggaaacagctatgaccAGCCAAGTCCCTGACTCTCA
*ASXL1*_ex13_F	tgtaaaacgacggccagtGGTCAGATCACCCAGTCAGTT
*ASXL1*_ex13_R	caggaaacagctatgaccAGCCCATCTGTGAGTCCAAC
*ASXL1*_ex7_M13_F	tgtaaaacgacggccagtTTTGTGGCTCTGCAGTTGAC
*ASXL1*_ex7_M13_R	caggaaacagctatgaccCCAGCAATACTGGGATGCTT
*ASXL1*_ex8_M13_F	tgtaaaacgacggccagtGGTTCTAGTGCTGGGCTCTG
*ASXL1*_ex8_M13_R	caggaaacagctatgaccAAAATAGAGGGCCACCCAAG
*TET2*_ex6_F	tgtaaaacgacggccagtGCTGCAAGTGACCCTTGTTT
*TET2*_ex6_R	caggaaacagctatgaccCTTATCGCATGACTGCCAAA
*DNMT3A*_ex23_T7E1_F	GCACATGGTTGGTGATCTGA
*DNMT3A*_ex23_T7E1_R	GGAAGGGAGCTTGGTTTTGT
*TET2*_ex6_T7E1_F	AAGAGGTCCATTCTAGTGCC
*TET2*_ex6_T7E1_R	GCTGATCATTTAGCAGCAGT
pLKO_U6_SEQ_forw	TTTGCTGTACTTTCTATAGTG

Table 3.2 quantitative real-time PCR primers. Primers were designed using PrimerQuest tool from Integrated DNA technologies.

Name	Sequence
*ASXL1*_qPCR_F	GGAGAAGGATGAAGGACAAACA
*ASXL1*_qPCR_R	GTCATTGGAGCATCCGAGTAG
*DNMT3A*_qPCR_F	GCCCATTCGATCTGGTGATT
*DNMT3A*_qPCR_R	GGCGGTAGAACTCAAAGAAGAG
*TET2*_qPCR_F	AGGTTTGGACAGAAGGGTAAAG
*TET2*_qPCR_R	CGAACCACCCACTTAGCAATA
*GAPDH*_qPCR_F	GAAGGTGAAGGTCGGAGTCA
*GAPDH*_qPCR_R	AATGAAGGGGTCATTGAT

Table 3.3 Oligonucleotides for CRISPR/Cas9 sgRNA-vector cloning or RNP sgRNA *in vitro* transcription. Single guide RNAs were designed using CCTop [142], CRISPRscan [143], and Benchling [140]. Additional bases were added based on the prospective cloning sites for vector cloning or for *in vitro* transcription of the sgRNA [144]. RNP: ribonucleoprotein

Name	Delivery	Ordered Oligonucleotide
*DNMT3A*_T1_F	Vector	CACC-gCGTCTCCAACATGAGCCGCT
*DNMT3A*_T1_R	Vector	AAAC-AGCGGCTCATGTTGGAGACGC
*DNMT3A*_T3_F	Vector	CACC-gCTCGCCAAGCGGCTCATGT
*DNMT3A*_T3_R	Vector	AAAC-ACATGAGCCGCTTGGCGAGC
*TET2*_sg_F	Vector	CACC-ACGGCACGCTCACCAATCGC
*TET2*_sg_R	Vector	AAAC-GCGATTGGTGAGCGTGCCGT
*DNMT3A*_ex23_T1	RNP	taatacgactcactataGGTCTCCAACATGAGCCGCTgttttagagctagaaatagc
*DNMT3A*_ex6_sg	RNP	taatacgactcactataGGCTCGGCATGGGCCGCTGAgttttagagctagaaatagc
*DNMT3A*_ex4_sg	RNP	taatacgactcactataGGAAAGCGGTGACACGCCAAgttttagagctagaaatagc
*DNMT3A*_ex14_sg	RNP	taatacgactcactataGGGTACCAGTACGACGACGAgttttagagctagaaatagc
*ASXL1*_ex8_sg	RNP	taatacgactcactataGGCCACGCCGATGGCGAGAGgttttagagctagaaatagc
*ASXL1*_ex7_sg	RNP	taatacgactcactataGGCCACGTGGAATCTGCATCgttttagagctagaaatagc
*ASXL1*_ex13_sg	RNP	taatacgactcactataGGCACCACTGCCATCGGAGGgttttagagctagaaatagc
*TET2*_ex6_sg	RNP	taatacgactcactataGGCGGCACGCTCACCAATCGCgttttagagctagaaatagc

Table 3.4 Oligonucleotides for homology directed repair.

Name	Sequence
*DNMT3A*_HR	CTGGCCAGCACTCACCCTGCCCTCTCTGCCTTTTCTCCCCCAGGGTATTTGGTT TCCCAGTCCACTATACTGACGTCTCCAACATGAGCCACTTGGCGAGGCAGAG ACTGCTGGGCCGGTCATGGAGCGTGCCAGTCATCCGCCACCTCTTCGCTCCGC TGAAGGAGTATTTTGCGTGTG
*DNMT3A*_anti-sense	CACACGCAAAATACTCCTTCAGCGGAGCGAAGAGGTGGCGGATGACTGGCAC GCTCCATGACCGGCCCAGCAGTCTCTGCCTCGCCAAGTGGCTCATGTTGGAGA CGTCAGTATAGTGGACTGGGAAACCAAATACCCTGGAAGAGAAAAGGCAGA GAGGGCAGGGTGAGTGCTGGCCAG
*ASXL1*_HR	ACCCTCGCAGACATTAAAGCCCGTGCTCTGCAGGTCCGAGGGGCGAGAGGTC ACCACTGCCATAGAGAGGCGGCCACCACTGCCATCGGATGGCCCGGGTGGAG GTGGCGGCGGGGCCACCGATGAGGGAGGTGGCAGAGGCAGCAGCAGTGGTG ATGGTGGTGAGGCCTGTGGCCACCC
*DNMT3A*_T1_ PAM65-35	CTGCCTTTTCTCCCCCAGGGTATTTGGTTTCCCAGTCCACTATACTGACGTCTC CAACATGAGCCACCTTGCGAGGCAGAGACTGCTGGGCCGGTCATGGA
*DNMT3A*_T1_ nonPAM65-35	GAGCGAAGAGGTGGCGGATGACTGGCACGCTCCATGACCGGCCCAGCAGTCT CTGCCTCGCAAGGTGGCTCATGTTGGAGACGTCAGTATAGTGGACTGGG
*ASXL1*ex13_ PAM65-35	ACTGCCATAGAGAGGCGGCCACCACTGCCATCGGATGGCCCGGGTGGAGGTG GCGGCGGGGCCACCGATGAGGGAGGTGGCAGAGGCAGCAGCAGTGGTG
*ASXL1*ex13_ nonPAM 65-35	CATCGGTGGCCCCGCCGCCACCTCCACCCGGGCCATCCGATGGCAGTGGTGG CCGCCTCTCTATGGCAGTGGTGACCTCTCGCCCCTCGGACCTGCAGAG

3.1.7 Antibodies

Table 3.5 Flow cytometry antibodies.

Antibody	Dilution	Clone	Supplier (No)
Anti-human CD14 APC	1:20	61D3	eBioscience (17-0149)
Anti-human CD14 Brilliant Violet 510	1:100	M5E2	BioLegend (301841)
Anti-human CD19 PE-Cy7	1:100	HIB19	BD Biosciences (560728)
Anti-human CD235a FITC	1:100	HI264	BioLegend (349104)
Anti-human CD3 APC	1:100	HIT3a	BioLegend (300311)
Anti-human CD34 APC	1:10	581	BD Biosciences (555824)
Anti-human CD34 FITC	1:100	561	BioLegend (343604)
Anti-human CD34 PE	1:50	581	BD Biosciences (555822)
Anti-human CD38 AlexaFluor 700	1:20	HIT2	BD Biosciences (560676)
Anti-human CD38 PE-Cy7	1:50	HIT2	BioLegend (303516)
Anti-human CD41 FITC	1:100	HIP8	BioLegend (303704)
Anti-human CD56 PE-Cy7	1:100	5.1H11	BioLegend (362510)
Anti-human CD66b Pacific Blue	1:60	G10F5	BioLegend (305112)
Anti-human CD90 PerCP-Cy5.5	1:20	5E10	BioLegend (328118)
Anti-human FcεRIα	1:20	AER-37	BioLegend (334610)

Table 3.6 Protein or DNA immunoblotting antibodies.

Antibody	Dilution	Supplier (No)
5-Hydroxymethylcytosine (5hmC) antibody (pAb)	1:10,000	Active Motif (39770)
5-Methylcytosine (5mC) (D3S2Z) Rabbit mAb	1:500	Cell Signaling Technologies (28692)
ASXL1 Antibody - middle region	1:1,000	Aviva (OAAB07377)
DNMT3A Antibody (C-12)	1:200	Santa Cruz (sc-365769)
TET2 Antibody (S-13)	1:200	Santa Cruz (sc-136926)
β-Actin Antibody (C-4)	1:1,000	Santa Cruz (sc-47778)
Swine Anti-Rabbit Immunoglobulins - HRP	1:1000	Agilent Dako
Goat Anti-Mouse Immunoglobulins - HRP	1:2000	Agilent Dako

3.1.8 Buffers and Media

Table 3.7 Buffer recipes.

Buffer	Components
10x Red Cell Lysis Buffer	1.55 M NH_4Cl 100 mM $KHCO_3$ 1 mM EDTA, pH 7.4
Alkaline Lysis Buffer I	50 mM glucose 25 mM Tris-Cl (pH 8.0) 10 mM EDTA (pH8.0)
Alkaline Lysis Buffer II	0.2 N NaOH 1% SDS
Alkaline Lysis Buffer III	60 ml 5 M potassium acetate 11.5 ml glacial acetic acid 28.5 ml H_2O
10x PBS	1.37 M NaCl 27 mM KCl 100 mM Na_2HPO_4 18 mM KH_2PO_4
10x TBS	1.54 M NaCl 1.3 M Tris
10x Transfer Buffer	2 M Glycin 250 mM Tris
10x Electrophoresis Buffer	2 M Glycin 250 mM Tris 35 mM SDS
Ponceau Staining Solution	2% Ponceau S 30% Trichloroacetic acid 30% Sulfosalicylic acid
Stacking Gel Buffer	0.5 M Tris-HCl, pH 6.8
Resolving Gel Buffer	2 M Tris-HCl, pH 8.0

Table 3.8 Cell culture media recipes.

Medium	Components
StemSpan Expansion Medium	StemSpan SFEM II L-Glutamine (2 mM) P/S (1 U) SCF (100 ng/µl) FLT3-L (100 ng/µl) TPO (100 ng/µl)
Human Long-Term Culture Medium (HLTM)	MyeloCult H5100 Hydrocortisone (1 µM)
MyeloCult Expansion Medium	MyeloCult H5100 L-Glutamine (2 mM) P/S (1 U) SCF (100 ng/µl) FLT3-L (100 ng/µl) TPO (20 ng/µl)
StemSpan Liquid Culture Medium	StemSpan SFEM II L-Glutamine (2 mM) P/S (1 U) SCF (50 ng/µl) FLT3-L (50 ng/µl) TPO (20 ng/µl) IL-6 (20 ng/µl) IL-3 (20 ng/µl)
K-562 Culture Medium	RPMI-1640 FBS (10%) L-Glutamine (2 mM) P/S (1 U)
M2-10B4 Culture Medium	RPMI-1640 FBS (10%) L-Glutamine (2 mM) P/S (1 U)
Sl/Sl Culture Medium	DMEM FBS (10%) L-Glutamine (2 mM) P/S (1 U)

3.1.9 Software

Software	Use	Provider
AlphaImager	Gel imaging	AlphaInnotech
BD FACSDiva™ software	Flow cytometry	BD Biosciences
FlowJo v10	Analysis of flow cytometry	BD Biosciences
GraphPad Prism 6	Statistical testing	GraphPad Software
ImageJ	Blot quantification	-
ImageQuant® LAS 4000 v 1.3	Western and Dot blot detection	GE Healthcare
Integrative genomics viewer v2.3.97	Visualization of genomic data	Broad Institute, UC
Mutation surveyor v.5.0.0	Mutation analysis	Softgenetics®
Nanodrop-1000 v3.7.1	Nucleic acid quantification	Thermo Scientific
IncuCyte® Software (v2019B)	Live cell analysis	Sartorius
QuantaSoft™ v1.7.4	ddPCR analysis	Bio-Rad
R v.3.4.0-4.0.2.	Statistics and graphics	-
RStudio v0.99.903-1.2.5042	IDE for R	-
StepOnePlus™ Software v2.3	qRT-PCR	Applied Biosystems™

*The table does not include web-based applications that were used for analysis as indicated in the methods section (i.e., ICE, TIDE, BaseSpace sequence Hub)

3.1.10 Cell Lines

Cell Line	Species	Type	Modification	Provider
K-562	*Homo sapiens*	CML, lymphoblast		ATCC (CRL-1972)
M2-10B4 (IL-3, G-CSF)	*Mus musculus*	Bone marrow fibroblasts	Produce human cytokines IL-3 and G-CSF [145]	StemCell Technologies (00301)
Sl/Sl (IL-3, SCF)	*Mus musculus*	Bone marrow fibroblasts	Produce human cytokines IL-3 and SCF [145]	StemCell Technologies (00302)

3.1.11 Primary Material and Patient Samples

3.1.11.1 Umbilical Cord Blood Collection

Cord blood of 11 donors was obtained in cooperation with the obstetrics clinic at Charité Universitätsmedizin Berlin between March 2019 and May 2020. The anonymous sampling of cord blood was approved by the Charité's Ethics Committee (process number EA2/234/18). The collection of umbilical cord blood was performed *ex utero* by an obstetrician or trained midwife with an umbilical cord blood collection bag from Macopharma. The umbilical cord blood sample was handled anonymously to respect the privacy of mother and child.

3.1.11.2 Patient Samples

We received DNA extracted from peripheral blood or bone marrow samples from 331 patients (age between 15 and 84 years) with t(8;21) AML from collaborating institutions in Germany (n=65), Netherlands (n=50), Taiwan (n=80), and United Kingdom (n=136). Written consent was obtained in accordance with the declaration of Helsinki and ethical approval was obtained from the local ethics committees of the cooperating institute.

3.2 Methods

3.2.1 Cell Biological Methods

3.2.1.1 Culture and Maintenance of Cell Lines

Cell lines were cultivated in a humid atmosphere at 37 °C and 5% CO_2 in a cell culture incubator. All cell lines were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS), 100 U/ml Penicillin/Streptomycin (P/S), and 2 mM L-glutamine (Glut). The suspension cell line K-562 was passaged twice a week by transferring the whole content of the cell culture flask into a 15 ml centrifuge tube and centrifuging at 400x g for 5 min. The supernatant was removed and the cell pellet resuspended in fresh culture medium. The required fraction of the cell suspension was added to a fresh cell culture flask and medium was filled to 15 ml. Adherent cells were passaged once or twice a week, depending on the confluence. After removing the cell culture medium from the flask, the cell surface was carefully rinsed with phosphate buffered saline (PBS) and 0.05% Trypsin/EDTA was added onto the adherent cell layer. After incubation at 37 °C, the cells were resuspended in fresh cell culture medium and a fraction of the cell suspension was transferred into a new cell culture flask. Cell lines were passaged for up to 3 months.

Upon the first passage, a fraction of the culture cells was cryopreserved by resuspending the cells in cryo-preservation medium containing 50% culture medium, 40% FBS, and 10% DMSO. The cell suspension was transferred into cryo-preservation tubes and stored in a freezing container at -80 °C for 24 h. For long-term storage the cells were removed from the freezing container and transferred into liquid nitrogen. Cells were thawed at 37 °C in a water bath, carefully resuspended in pre-warmed culture medium, and centrifuged at 400x g for 5 min to remove the remaining cryo-preservation medium containing DMSO.

Cell lines were regularly checked for mycoplasma contamination via polymerase chain reaction (PCR) and the identity of the cell lines was confirmed (Human Cell Line Authentication Services, Eurofins Genomics).

3.2.1.2 Preparation and Irradiation of Feeder Cell Lines

The feeder cell lines Sl/Sl and M2-10B4 were obtained as a kind gift from colleagues (Prof. Dr. Michael Heuser, Hannover Medical School, Germany). Both cell lines are engineered to produce the human cytokines IL-3 (interleukin-3) and SCF or IL-3 and G-CSF (granulocyte-colony stimulating factor), respectively [145, 146]. Every third passage the cells were selected

by adding a defined concentration of selection antibiotics to the culture medium, as listed in Table 3.9.

Table 3.9 Final concentration of selection antibiotics for feeder cell lines.

Cell line	Geneticin (G418) [mg/ml]	Hygromycin B [mg/ml]
M2-10B4 (IL-3, G-CSF)	0.4	0.06
Sl/Sl (IL-3, SCF)	0.8	0.125

Feeder cell lines were maintained as described above. For co-culture experiments, feeder cells were mixed at a 1:1 ratio and seeded in collagen coated cell culture plates. Cells were seeded at a density of approximately $2.5x10^4$ cells/cm^2. After letting the cells adhere, they were irradiated with 80 Gy, using a γ irradiation source, to inhibit proliferation. The irradiated feeder cells were incubated for at least 24 h before adding primary test cell material.

3.2.1.3 Isolation of Mononuclear Cells from Umbilical Cord Blood by Density Centrifugation

Cord Blood (CB) collection and handling procedures are described in section 3.1.11.1. The freshly collected cord blood was filtered through a 40 µm cell strainer, transferred into 50 ml reaction tubes and diluted with 1 volume of PBS supplemented with 1 mM EDTA (PBS/EDTA). To separate mononuclear cells, density centrifugation was performed. For this purpose, 15 ml Biocoll density centrifugation separation solution was transferred into a 50 ml reaction tube and overlaid with 30 ml of the diluted blood. After centrifugation at 760x g for 30 min, mononuclear cells (MNC) were collected from the interphase and transferred into a fresh 50 ml reaction tube. The cells were washed by filling up the tube with PBS/EDTA and centrifuging at 300x g for 15 min. The supernatant was removed and the cell pellet was resuspended in 10-20 ml red cell lysis buffer (RCLB). Erythrocyte lysis was performed for 30 min at 4 °C. After incubation, the cells were washed twice by adding PBS/EDTA and centrifuging at 400x g for 7 min. Cells were quantified by trypan blue exclusion between the two washing steps.

3.2.1.4 Culture of Human Primary $CD34^+$ Cells from Umbilical Cord Blood

Primary mononuclear cells from umbilical cord blood were enriched by magnetic separation using MACS (magnetic cell separation) CD34 MicroBead Kit UltraPure. Up to 10^8 cells were resuspended in 300 µl PBS supplemented with 0.5% BSA and 2 mM EDTA (MACS buffer). Afterwards 100 µl FcR Blocking Reagent and 100 µl CD34 MicroBeads UltraPure were added to the cell suspension. After 30 min incubation at 4 °C, the cells were washed by adding 5 ml MACS buffer and centrifuged at 300x g for 10 min. The supernatant was removed and up to

10^8 cells were resuspended in 500 µl MACS buffer. Before applying the cell suspension, the separation column was placed onto a magnetic stand and rinsed with 500 µl MACS buffer. The cell suspension was then transferred onto the separation column and the flow-through containing unlabeled cells was collected. The column was washed three times with 500 µl of buffer. To elute the enriched cell population, the column was removed from the magnetic stand and flushed with 1 ml buffer. Purified cells were immediately quantified by trypan blue exclusion. Isolated $CD34^+$ cells were resuspended in StemSpan SFEM II medium supplemented with 100 ng/ml human SCF, 100 ng/ml human FLT3-L (FMS-like tyrosine kinase 3 ligand), and 100 ng/ml human thrombopoietin (TPO), at a concentration of approximately $2.5x10^5$ cells/ml at 37 °C, 5% CO_2 for 24 to 48 h.

3.2.1.5 Electroporation of K-562 Cell Line and Primary $CD34^+$ Cells

Amaxa Nucleofection

K-562 cells were seeded at $2x10^5$ cells/ml 2 days before electroporation. Before electroporation 12-well plates were filled with 1.5 ml culture medium per well and pre-incubated at 37 °C in the incubator. Per reaction $1x10^6$ cells were centrifuged at 200x g for 10 min. The supplement was added to the Cell Line Nucleofector® solution V according to the manufacturer's instructions. Per reaction 100 µl of Nucleofector® solution was pre-mixed with 2-5 µg plasmid DNA. After centrifugation, the supernatant was removed, the cell pellet resuspended in 100 µl pre-mixed Nucleofector® solution and directly transferred into the electroporation cuvette. K-562 cells were electroporated with the program T-016. After electroporation 500 µl of pre-warmed culture medium was added to the cuvette and the cell suspension transferred to the pre-warmed 12-well plates.

For transfection of primary cells from cord blood, either frozen mononuclear cells were thawed, cultured for up to 7 days, and then subjected to electroporation, or enriched $CD34^+$ cells were expanded for 2 days prior to electroporation. For nucleofection of primary cells approximately $5x10^5$ cells were used per condition and treated as described above. The cells were resuspended in Human $CD34^+$ Cell Nucleofector® Solution and electroporated with program U-008. After electroporation 500 µl of pre-warmed culture medium was added to the cuvette and the cell suspension transferred to the pre-warmed 12-well plates.

Neon transfection

Before electroporation, a 24-well plate was filled with 500 µl antibiotic free culture medium per well and placed at 37 °C. Cells were counted by trypan blue exclusion. For each electroporation reaction $1x10^5$ viable cells were collected. Electroporation was always performed in duplicates for each target. Cells were washed twice with PBS, centrifuging at 400x g for 5 min and removing the supernatant. The cells were resuspended in 10 µl electroporation buffer per reaction. For K-562 electroporation buffer R was used, for primary cells electroporation buffer T. For each electroporation reaction either 1 µg of plasmid DNA or up to 1 µl of Cas9-sgRNA ribonucleoprotein (RNP) was transferred into a reaction tube. To each tube 10 µl of cell suspension was added and mixed. After setting up the NEON device according to the manufacturer's instructions, cells were electroporated using 10 µl NEON tips with the established electroporation conditions listed in Table 3.10. After the electric pulse, cells were directly pipetted into the pre-warmed culture plate.

Table 3.10 NEON electroporation conditions for different cells. CB: cord blood

	K-562	**CB CD34$^+$ cells**
Voltage [V]	1450	1600
Pulse duration [ms]	10	10
No. of pulses	3	3

The target cells were electroporated with Cas9-sgRNA-ribonuleoprotein. For generating knockouts via non-homologous end joining (NHEJ), 3 pmol recombinant Cas9 nuclease (0.5 µg) were mixed with 30 pmol guide RNA (molar ratio 1:10) per replicate and incubated for 30 min at room temperature. For homology directed repair (HDR)-mediated specific modification, Cas9 and sgRNA were mixed in a molar ratio of 1:5 (3 pmol Cas9 and 15 pmol sgRNA) and incubated for 30 min at room temperature. Directly before electroporation 12.5 pmol of the HDR-template was added to the reaction. Cells were electroporated as described above and cultured in the respective culture medium without antibiotics before setting up down-stream experiments.

3.2.1.6 Colony Forming Assay

Colony Forming assays were carried out in Methylcellulose based medium supplemented with cytokines (MethoCult® H4434 classic or MethoCult® H4435 enriched). Transfected cells were resuspended in IMDM supplemented with 2% FBS at a final concentration of $1x10^4$ cells/ml. For duplicate assays 250 µl of the cell suspension was added to 2.5 ml MethoCult® medium and mixed thoroughly. Two 35 mm cell culture dishes were placed into a 100 mm petri dish

and 1 ml cell suspension was dispensed into each 35 mm cell culture dish. To assure optimal humidity, a third 35 mm culture dish was placed into the 100 mm petri dish and filled with 2.5 ml sterile H_2O. The cultures were incubated at 37 °C in a cell culture incubator for 10-14 d. Colonies were enumerated based on their morphology and categorized into burst-forming unit erythroid (BFU-E), colony-forming unit erythroid (CFU-E), colony-forming unit granulocyte, macrophage (CFU-GM), and colony-forming unit granulocyte, erythrocyte, macrophage, megakaryocyte (CFU-GEMM).

To assess serial replating capacity, quantified cultures were resuspended in IMDM/2%FBS, washed two times, and cells were quantified by trypan blue exclusion. Depending on the cell concentration and viability, 1,000-2,500 cells were again plated in MethoCult®, as described above. The process was repeated up to 4 times, or until no colonies were detected.

3.2.1.7 Long-term Culture-Initiating Cell Assay

For long-term culture-initiating cell (LTC-IC) assays test cells were seeded on irradiated feeder cells (see 3.2.1.2). LTC-IC assay was carried out in 24-well format with approximately $2.5x10^4$ M2-10B4:Sl/Sl feeder cells seeded for irradiation. Medium was removed from the wells and 250-1,000 test cells were seeded in 500 µl MyeloCult™ H5100 supplemented with 1 µM hydrocortisone (human long-term culture medium, HLTM). The cells were cultured for up to 9 weeks at 37 °C in a humidified incubator. Cultures were maintained by gently rotating the plate, removing half of the medium, and replacing it with freshly prepared HLTM once a week.

After up to 9 weeks of culture the number of colony-forming cells was enumerated via colony forming assay. All adherent and non-adherent cells were harvested from the well and washed twice with IMDM/2%FBS. Cells were counted by trypan blue exclusion and approximately $2x10^4$ cells were seeded into MethoCult H4435 as described above.

3.2.1.8 Live Cell Proliferation Assay

To determine cell proliferation, $4x10^3$ transfected cells were seeded in 96-well flat-bottom plates in 200 µl Stem Cell liquid medium (n=4 per transfection replicate). Cells were cultured and monitored in the IncuCyte Live-Cell Analysis Systems (Sartorius) for 7 to 14 days. Cell density was measured every 2 hours using HD phase imaging. The first picture was taken 30 min after the cells were put inside the IncuCyte to allow them to settle on the bottom of the well. To reduce evaporation, border wells were filled with 200 µl PBS. Confluence data were analyzed using the IncuCyte software and normalized to correct for seeding densities.

3.2.1.9 Flow Cytometry Analysis of Cell Cycle

For flow cytometry analysis of the cell cycle, cells were harvested in 1.5-ml micro centrifuge tubes and washed once with cold PBS. After centrifugation at 400x g for 5 min, the supernatant was removed carefully, and the pellet was resuspended in 300 µl cold PBS. While mixing gently, 700 µl ice-cold ethanol (100%) was added dropwise to the cell suspension. After fixing for 30 min at 4 °C, cells were either subjected to staining or stored for up to 8 weeks at -20 °C. To visualize mitotic cells, an antibody against phosphorylated histone H3 was used. Cells were centrifuged at 1,000x g for 3 min and the supernatant was removed. After washing the cells once with PBS containing 0.5% Triton-X-100 and 1% BSA, 50 µl of the same solution containing primary antibody targeting phosphorylated histone H3 (1:1600, from goat) was added. The suspension was incubated at room temperature for 1 h in the dark. After incubation, the cells were washed once with PBS/Triton-X-100/BSA, resuspended in 50 µl secondary antibody solution (Alexa-Fluor 488 conjugated anti-goat antibody, 1:500), and incubated for 1 h at room temperature in the dark. The cells were washed once and resuspended in PBS containing propidium iodide (PI) and RNase A. After incubating for 30 min, cells were acquired on a BD FACS Canto II flow cytometer. Analysis and gating of cell populations was performed in FlowJo v10.

3.2.1.10 Flow Cytometry Analysis of Cell Surface Markers

To determine expression of cell surface markers via flow cytometry, cells were harvested in 1.5-ml microcentrifuge tubes and washed twice with PBS containing 0.5% BSA. While centrifuging at 400x g for 5 min, antibody staining solution was prepared. The fluorescence conjugated antibodies were diluted in PBS/0.5% BSA as indicated in Table 3.5. To avoid unspecific binding of antibodies, FcR blocking reagent was added to the antibody mix. After centrifugation, the supernatant was removed as completely as possible without disturbing the cell pellet, and the cells were resuspended in 10 µl antibody solution. Depending on the timing of the experiment, the cells were incubated for 15 min at room temperature or 1 h at 4 °C. After incubation, the cell suspension was centrifuged, the supernatant removed, and the cells washed once in PBS/0.5% BSA. For analysis, the cells were resuspended in at least 200 µl PBS/0.5% BSA and transferred into 5 ml round bottom test tubes. Cells were acquired on a BD FACS Canto II flow cytometer. Analysis and gating of cell populations was performed in FlowJo v10.

3.2.2 Microbiological Methods

3.2.2.1 Reconstitution of Plasmid Stocks from Supplier

Plasmids were obtained from Addgene as stab cultures. With the help of a pipette tip, bacteria were streaked onto a Luria Broth (LB) agar plate containing ampicillin to obtain single colonies. After incubation overnight at 37 °C, single colonies were picked and transferred into LB medium containing ampicillin for overnight culture. To check the sequence of the plasmid, isolated colonies were cultured in 2 ml LB medium overnight. For generating plasmid stocks, 100-200 ml of LB medium with ampicillin were inoculated with isolated colonies.

3.2.2.2 Isolation of Plasmid DNA from E. coli

To obtain plasmid DNA for Sanger sequencing, alkaline lysis with SDS (sodium dodecyl sulfate) was performed (Minipreps). After overnight incubation, 1.5 ml of the 2 ml culture was poured into a 1.5 ml microcentrifuge tube and centrifuged at maximum speed for 1 min. The remaining 500 µl of the culture were stored at 4 °C. After centrifugation, the medium was aspirated, and the pellet resuspended in 100 µl of ice-cold alkaline lysis solution I. Alkaline lysis solution II (200 µl) was added to the suspension and mixed by inverting the tube at least 5 times before storing the tube on ice. Before incubating the solution for 5 min on ice, 200 µl of alkaline lysis solution III was added and the contents mixed by inverting the tube. After incubation, the lysate was centrifuged for 5 min at maximum speed and the supernatant transferred into a fresh microcentrifuge tube. Plasmid DNA was recovered by adding 1 volume of isopropanol. The properly mixed solution was incubated at room temperature for 2 min before centrifugation at maximum speed for 5 min. After removing the supernatant, the pellet was washed in 500 µl 70% ethanol, centrifuged, and dried by standing the tube in an inverted position on a paper towel. The nucleic acids were dissolved in 50 µl TE buffer (10 mM Tris, 1 mM EDTA, pH 7.6).

After confirming the plasmids had the correct sequences, the rest of the bacterial suspension was cultured in 100 ml LB medium with ampicillin for Midiprep. Isolation of plasmid DNA was performed after incubation overnight at 37 °C with the NucleoBond Xtra Midi Plus Kit from Macherey Nagel according to the manufacturer's instruction. Before DNA isolation, 500 µl of the bacteria suspension was mixed with glycerol in a cryopreservation tube and stored at -80 °C. The isolated plasmid DNA was stored at -20 °C.

3.2.2.3 Transformation of E. coli with Plasmid DNA

To amplify plasmid DNA, chemically competent *E. coli* DH5alpha or TOP10 were transformed with plasmid DNA. The bacteria were thawed on ice and carefully flicked to mix the contents

of the tube. Up to 1 ng of plasmid DNA or 10 µl ligation reaction were added to the bacteria suspension and the contents of the tube were mixed by flicking the tube gently. The suspension was incubated on ice for 30 min. Subsequently, a heat-shock was performed by incubating the bacteria suspension for 90 s in a water bath at 42 °C. After cooling the suspension down on ice for 3 min, 200 µl of LB medium was added and the reaction tubes placed in a shaker at 37 °C for 1 h. Finally, 200 µl of the bacteria suspension was plated onto LB agar plates containing ampicillin and incubated overnight at 37 °C. Single colonies were isolated the following day.

3.2.3 Molecular Biological Methods

3.2.3.1 Isolation of Genomic DNA and RNA from Cultured Cells

After the appropriate culture time, cells were harvested, washed once with PBS and pelleted at 400x g for 5 min. The pellets were either stored at -20 °C until further use or directly subjected to DNA or RNA isolation. DNA isolation was carried out using different methods, depending on the downstream application and the number of cells that were collected as detailed below.

DNA isolation of samples containing between $1x10^5$ and $5x10^6$ cells was carried out using Qiagen DNA Mini Kit. The cell pellet was resuspended in 200 µl PBS and 20 µl proteinase K were added to the suspension. Cells were then lysed and DNA extracted according to the manufacturer's instructions. For samples containing less than $1x10^5$ cells, NucleoSpin Tissue XS Kit from Macherey Nagel was used according to the manufacturer's instructions.

For DNA extraction by isopropanol precipitation, the cell pellet was resuspended in 200 µl 10 mM Tris-HCl containing 400 µg/ml proteinase K. The samples were mixed thoroughly and incubated at 56 °C for at least 30 min. After incubation, the samples were spun down and one volume isopropanol was added to the cell lysate. The samples were mixed thoroughly and centrifuged for 5 min at maximum speed (17,000x g) in a tabletop centrifuge. The supernatant was removed and the pellet was washed with 70% ethanol. After centrifuging again for 5 min at maximum speed, the supernatant was discarded and the pellets were dried at room temperature. The dried DNA pellet was dissolved in 50-100 µl TE buffer. For genotyping experiments, a fast extraction protocol was used. Cells were transferred in 96-well PCR reaction wells and centrifuged at 2,500xg for 15 min to remove the supernatant. The cells were washed once in 10 mM Tris-HCl (pH 8.0)/proteinase K and afterwards resuspended in the same buffer. Cells were lysed at 56 °C for 15 min, and proteinase K was deactivated by heating the reaction to 95 °C for 15 min. The cells were centrifuged again and the supernatant was used for PCR and genotyping experiments.

RNA isolation was performed using the RNeasy RNA isolation kit from Qiagen or Qiagen AllPrep DNA/RNA Kit according to the manufacturer's instructions.

DNA and RNA were quantified using a NanoDrop 1000 spectrophotometer or Quantus™ Fluorometer with the respective QuantiFluor® dsDNA or QuantiFluor® RNA System.

3.2.3.2 PCR Amplification of Genomic DNA

Amplification of DNA regions by polymerase chain reaction (PCR) was generally performed using HotStarTaq DNA Polymerase.

Table 3.11 HotStar Taq DNA Polymerase reaction set-up for 25 µl reaction.

Component	Volume
HotStar Taq polymerase	0.125 µl
10x PCR buffer	2.5 µl
dNTP mix (25 mM each)	0.2 µl
Forward primer (10 µM)	1 µl
Reverse primer (10 µM)	1 µl
Genomic DNA	2-100 ng
H_2O	Ad 25 µl

Table 3.12 HotStar Taq DNA Polymerase PCR program.

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	15 min	
Denaturation	94 °C	20 s	30x
Annealing	50-65 °C	20 s	
Extension	72 °C	1 min/kb	
Final extension	72 °C	7 min	
Hold	15 °C	∞	

Amplified PCR products were separated on 1-2% agarose gels in TAE buffer for 30-60 min at 115 V. Prior to loading the agarose gel, PCR products were mixed with 6x loading dye and 10x GelRed. The agarose gels were imaged with UV light on an AlphaImager transilluminator. After confirming the correct amplicon size on the agarose gel, PCR products were purified using AMPure Beads XP or Qiagen PCR purification kit according to the manufacturer's instructions.

3.2.3.3 Digital Droplet PCR

Digital droplet polymerase chain reaction (ddPCR) was performed to detect and validate specific gene mutations. For ddPCR, droplets are formed in a water-oil emulsion that separates the DNA molecules. In each individual droplet a TaqMan probe-based PCR amplification is carried out. Two different fluorophores are used for mutant and wild type probes. By reading the fluorescence in both channels of each droplet individually, the amount of target DNA molecules can then be calculated. The final DNA input for ddPCR was generally 10-20 ng in order to achieve a sensitivity of up to 0.5% variant allele frequency. The reaction was set up according to the manufacturer's instructions for each assay.

Table 3.13 ddPCR reaction set-up for *DNMT3A* R882C and *KRAS* G12/G13 mutation screening.

Component	Volume/amount
2x ddPCR Supermix for Probes (no dUTP)	10 µl
20x multiplex primers/probes (FAM + HEX)	1 µl
Genomic DNA	10-20 ng
H_2O	ad 20 µl

Table 3.14 ddPCR reaction set-up for *DNMT3A* R882H and *JAK2* V617F mutation screening.

Component	Volume/amount
2x ddPCR Supermix for Probes (no dUTP)	10 µl
20x target primer/probe (FAM)	1 µl
20x reference primer/probe (HEX)	1 µl
Genomic DNA	10-20 ng
H_2O	ad 20 µl

Digital droplet PCR experiments were always carried out with one positive control (mutated DNA) and one negative control (wild type DNA) for each target. The reaction set-up was prepared in PCR strips and mixed thoroughly. Subsequently, 20 µl of the reaction mix was loaded into a DG8™ Cartridge followed by 70 µl droplet generation oil. Droplets were generated in a QX200 droplet generator according to the manufacturer's instructions. The droplets were transferred into a clean 96-well plate and the plate was sealed. Following cycling conditions were used to amplify the genomic DNA:

Table 3.15 Cycling conditions for ddPCR assays.

Step	Temperature	Time	Ramp rate	Cycles
Enzyme activation	95 °C	10 min		
Denaturation	94 °C	30 s	2 °C/s	40x
Annealing/extension	55 °C	1 min		
Enzyme deactivation	98 °C	10 min		
Hold	4 °C	∞	1 °C/s	

After thermal cycling, the plate was placed in a QX200 Droplet reader for droplet quantification. Data acquisition and analysis was performed using QuantaSoft™ analysis software. The software calculates the DNA concentration in copies per µl. Variant allele frequency can be calculated by dividing the droplet concentrations of target (FAM channel) and reference (HEX channel).

3.2.3.4 Reverse Transcription of mRNA to cDNA

Synthesis of cDNA (complementary DNA) was performed prior to quantitative real-time polymerase chain reaction (qRT-PCR) by reverse transcription of RNA using M-MLV reverse transcriptase. The reaction was set up according to the manufacturer's instructions with a minimum of 10 ng RNA. RNA input was equalized for all samples in one experimental set-up and brought to 10 µl with RNase free water. To each reaction 200 ng random primer and 10 mM dNTP mix was added and incubated at 65 °C for 5 min. After chilling the samples on ice, 4 µl 5x first stand buffer and 0.1 M dithiothreitol (DTT) was added to each reaction and incubated at 37 °C for 2 min. 1 µl of M-MLV was added to each reaction and the PCR strip was placed in a thermal cycler for cDNA synthesis.

Table 3.16 Thermal cycler program for cDNA synthesis.

Temperature	Time
25 °C	10 min
37 °C	50 min
70 °C	15 min

3.2.3.5 Quantitative Real Time PCR

The previously generated cDNA samples were used for qRT-PCR. Per reaction 5-25 ng cDNA were used.

Table 3.17 Luna Universal reaction set-up for qRT-PCR.

Component	Volume/amount
Luna Universal qPCR Master Mix	10 µl
Forward primer (10 µM)	0.5 µl
Reverse primer (10 µM)	0.5 µl
cDNA	5-25 ng
H_2O	ad 20 µl

Quantitative real time PCR was carried out in an Applied Biosystems StepOnePlus® machine with "fast" ramp setting. For each experiment a melt curve analysis was performed.

Table 3.18 Cycling conditions for qRT-PCR.

Step	Temperature	Time	Cycles
Initial Denaturation	95 °C	60 s	
Denaturation	95 °C	15 s	40x
Extension	60 °C	30 s	
Hold	4 °C	∞	

Data acquisition and analysis was performed via StepOne Software. Relative target expression was calculated via the $\Delta\Delta C_T$-method.

3.2.3.6 CRISPR Vector Cloning

To introduce knockouts or precise genetic modifications in the target genes, the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) method was used. Two components are necessary to perform genetic modification: a single guide RNA (sgRNA) and a CRISPR-associated (Cas) nuclease. Guide RNAs were designed using the algorithm of the online platforms CCTop [142], CRISPRscan [143], and Benchling. SgRNA target sequences were purchased as single stranded DNA-oligonucleotides from Eurofins Genomics with overhangs complementary to *BsmBI* cut sites in the pLKO5-vector backbone. Forward and reverse strand of each guide RNA were hybridized by incubating at 95 °C for 15 min and cooling down to room temperature. The DNA fragments were ligated into the pLKO5.sgRNA.EFS.tRFP vector (Addgene no. 57823, [139]). The correct insertion of the sgRNA was verified by Sanger sequencing with a U6 sequencing primer (Table 3.1).

For *DNMT3A* and *ASXL1* point mutations should be introduced. This was achieved by designing templates for homology directed repair. At the beginning of CRISPR-system establishment, a 200 bp, single-stranded oligonucleotide was used, carrying the desired point mutation in the middle. With refinement of the CRISPR-strategy, I modified the design of the HDR templates to optimize the HDR frequency in the cells.

3.2.3.7 In vitro Transcription of guide RNA for Ribonucleoprotein

For modifying primary cells from umbilical cord blood, recombinant Cas9 nuclease was complexed with *in vitro* transcribed or synthetic guide RNA, generating Cas9-sgRNA ribonucleoprotein (RNPs). Synthetic guide RNAs were ordered from Synthego and diluted to a concentration of 30 pmol/µl.

For *in vitro* transcription (IVT) oligonucleotides were designed containing a T7 promotor sequence, the actual target sequence, and an overlap sequence for the scaffold primer (Figure 3.1).

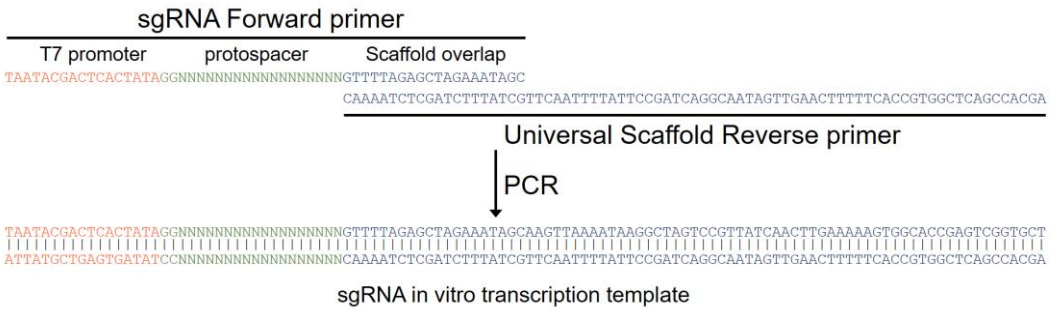


Figure 3.1 Schema of sgRNA oligonucleotide. A PCR is performed using the sgRNA oligonucleotide and a universal scaffold primer to obtain the template for *in vitro* transcription of the sgRNA. The figure was modified from [144].

To obtain the sgRNA DNA template for IVT, an overlapping PCR was performed as described by Brunetti et al. [144].

Table 3.19 Overlapping PCR reaction mix for sgRNA DNA template synthesis.

Component	Volume
2x KAPA HiFi HotStart ReadyMix	10 µl
sgRNA oligonucleotide (10 µM)	2 µl
Universal reverse scaffold primer (10 µM)	2 µl
Nuclease-free H_2O	ad 20 µl

Table 3.20 sgRNA DNA template synthesis cycling conditions.

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	3 min	
Denaturation	98 °C	5 s	6x
Annealing	60 °C	5 s	
Extension	72 °C	10 s	
Final extension	72 °C	1 min	
Hold	4 °C	∞	

The PCR products were purified using Zymo Research DNA Clean & Concentrator-5 kit according to the manufacturer's instructions and eluted with 11.5 µl elution buffer. DNA concentration was determined using a NanoDrop1000 spectrophotometer. The purified PCR products were used as templates for IVT with the HiScribe™ T7 High Yield RNA Synthesis Kit. The components of the kit were mixed as described in Table 3.21 and the reaction was incubated at 37 °C for 16 h.

Table 3.21 *In vitro* transcription reaction mix.

Component	Volume
10x reaction buffer	1 µl
ATP (100 mM)	1 µl
CTP (100 mM)	1 µl
GTP (100 mM)	1 µl
TTP (100 mM)	1 µl
purified PCR product	4 µl
T7 RNA polymerase enzyme mix	1 µl

After incubation, each RNA sample was brought up to a total volume of 50 µl and purified using the Zymo Research RNA Clean & Concentrator-5 kit following the manufacturer's instructions. The RNA was eluted in 35 µl of nuclease free water and quantified using a NanoDrop1000 spectrophotometer.

3.2.3.8 CRISPR/Cas9 Indel Detection via T7 Endonuclease I Digestion

For CRISPR-indel detection, genomic DNA was extracted (3.2.3.1) earliest 24 h after electroporation. The expected target sites were amplified by PCR (3.2.3.2) resulting in DNA fragments of 400-700 bp. PCR products were purified and eluted in 15 µl nuclease-free water. For each PCR product a 20 µl reaction was prepared as shown in Table 3.22 and hybridized in a PCR cycler (Table 3.23).

Table 3.22 Reaction set-up for T7 endonuclease I hybridization.

Component	Volume
purified PCR product	5 µl
NEB buffer 2	2 µl
Nuclease-free water	13 µl

Table 3.23 Hybridization conditions for T7 endonuclease I assay.

Step	Temperature	Time	Ramp rate
Initial denaturation	95 °C	10 min	
Annealing	85 °C	20 s	-2 °C/s
	25 °C	20 s	-0.1 °C/s
Hold	4 °C	∞	

After hybridization, 10 µl of the reaction was transferred into a fresh reaction tube and mixed with 2 U T7 endonuclease I (T7E1). The reaction was incubated at 37 °C for 15 min, stopped by adding 1 µl 0.25 M EDTA, and loaded on a 1.5% agarose gel. T7 endonuclease I cuts mismatched heteroduplex DNA that results from hybridization of wild type and mutated strands. In samples with CRISPR-induced mutations more than one band should be visible in the agarose gel.

3.2.3.9 Detection of 5-Hydroxymethylcytosine and 5-Methylcytosine by Dot Blot

DNA was extracted as described in 3.2.3.1. The concentration was adjusted for all samples, DNA denatured at 99 °C for 5 min, and spotted on a positively charged nylon membrane. All samples were spotted on two separate membranes in at least 3 different concentrations (2-fold dilutions). The membrane was air-dried for 30 min, UV-cross-linked for 3 minutes on a trans-illuminator (350 nm) and blocked overnight at 4 °C in 5% milk powder in tris-buffered saline with 0.5% Tween-20 (TBS-T). After washing the membrane 3 times for 5 min in TBS-T, primary antibody against 5-hydroxymethylcytosine (active motif, 39770, 1:10,000 in 5% BSA in TBS-T) or 5-methylcyotsine (cell signaling technologies, 28692, 1:500 in 5% BSA in TBS-T) was added over night at 4 °C. The blots were washed three times for 5 min in TBS-T and

incubated with the secondary horseradish peroxidase (HRP)-conjugated antibody for 90 min at room temperature. After three final wash steps with TBS-T and incubating the membrane with ECL prime western blotting detection reagent, the chemiluminescence signal was detected using the ImageQuant LAS 4000 system. To stain total DNA that was spotted on the membrane, it was incubated in 0.2% methylene blue in 0.3% sodium acetate for 15 min and rinsed with A. dest. to remove the background signal. Intensities of all signals were determined using ImageJ 1.48v. The dot blot signal was normalized to the methylene blue loading control and calculated as quotient of control and test sample. Specificity of the antibodies was tested using the methylated DNA Standard Kit (active motif, 55008, Supplement: Figure 8.1).

3.2.4 Protein Biochemistry

3.2.4.1 Isolation of Whole Cell Protein

Cells were harvested from culture vessels and centrifuged at 400x g for 5 min at 4 °C and washed once with cold PBS. The supernatant was carefully removed and the cell pellet was resuspended in protein lysis buffer containing protease and phosphatase inhibitors. After incubating on ice, the cells were disrupted by sonification in a Bioruptor (15 cycles, 30 s on/30 s off). To remove cell debris, the samples were centrifuged at maximum speed for 15 min at 4 °C. The supernatant was transferred into a new reaction tube and stored at -20 °C.

3.2.4.2 Pierce BCA Assay to Determine Protein Concentrations

Protein concentrations were determined using the Pierce™ BCA (bicinchoninic assay) Protein Assay Kit (Thermo Fischer) according to the manufacturer's instructions. The assay was carried out in duplicates in microwell plates by mixing 150 µl working solution (reagent A : reagent B, 50:1) with either 10 µl BSA standard, 1 µl of unknown sample diluted in 9 µl lysis buffer, or 10 µl lysis buffer as blank. BSA solutions with concentrations ranging from 125 µg/µl to 2,000 µg/µl were used for creating a standard curve. The reaction was incubated for 30 min at 37 °C and absorbance measured at 562 nm in Tecan infinite plate reader.

3.2.4.3 SDS Polyacrylamide Gel Electrophoresis

To separate proteins by molecular weight, sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed. Polyacrylamide gels were prepared in a Mini-PROTEAN® Tetra Cell Casting Module (Bio-Rad) using glass plates with 1 mm spacer. Separation gels were prepared as described in Table 3.24, poured between the glass plates and covered with 70% ethanol to ensure a smooth interface between the gels. After polymerization of the separation gel, a stacking gel was prepared, laid over the separation gel, and a comb inserted between the glass plates. After complete polymerization, the polyacrylamide gels were stored in wet tissues at 4 °C for up to one week.

Cell lysates were thawed on ice, homogenized thoroughly, and 40 µg of each sample were mixed with 4x laemmli buffer containing 10% β-mercaptoethanol. All samples were denatured for 5 min at 95 °C before loading on the SDS polyacrylamide gel together with a protein standard. Gels were run in the Mini-PROTEAN® Tetra Cell system from Bio-Rad at initially 80 V, increasing the voltage up to 140 V, for 1.5-2 h.

Table 3.24 Gel components for SDS-PAGE.

	separation gel				stacking gel
Components	**8%**	**10%**	**12%**	**16%**	
water	3.3 ml	2.8 ml	2.2 ml	1.1 ml	2.0 ml
Tris-HCl 1.5 M, pH 8.8	2.0 ml				
Tris-HCl 1 M pH 6.8					1.3 ml
Acrylamide (40%)	2.2 ml	2.7 ml	3.3 ml	4.4 ml	700 µl
Glycerin (40%)	4.0 ml				1.0 ml
SDS (20%)	50 µl				25 µl
Ammonium persulfate (APS, 10%)	50 µl				40 µl
N,N,N',N'-tetramethylenediamine (TEMED)	10 µl				10 µl

3.2.4.4 Western Blot and Immunodetection of Proteins

Proteins were transferred from the polyacrylamide gels onto 0.2 µm nitrocellulose membranes by western blotting. Nitrocellulose membranes were activated in distilled water for 5 min and equilibrated in transfer buffer for 10 min. The gels were removed from the electrophoresis chamber and equilibrated in transfer buffer for 15 min to remove remaining SDS. The transfer was carried out with a Mini-PROTEAN® Tetra Cell wet blotting system (Bio-Rad) at 340 mA for 80 min or at 150 mA overnight in a cold room. The protein transfer was checked by staining the membranes for 15 min in Ponceau S. The stain was removed by rinsing the membrane with distilled water. Membranes were then blocked in 5 % BSA in TBS-T for at least 1 h to prevent non-specific antibody binding.

Before immunodetection of proteins, the membrane was washed for 5 min in TBS-T. The primary antibodies were diluted in 5% BSA in TBS-T and added to the membrane. After incubating the primary antibody overnight at 4 °C, the membrane was washed three times for 10 min in TBS-T. The HRP-conjugated secondary antibody was also diluted in 5% BSA in TBS-T, added to the membrane, and incubated for 90 min at room temperature. After washing the membrane again for three times with TBS-T, it was incubated with ECL western blot detection reagent and chemiluminescence detected with an ImageQuant LAS 4000 system. Western Blot band intensities were quantified via ImageJ 1.48v. All bands were normalized to the loading control (housekeeper) and a relative intensity was calculated as quotient of control sample and test sample.

3.2.5 Sequencing Methods

3.2.5.1 Sanger Sequencing

Sanger sequencing was performed by GATC services of Eurofins Genomics GmbH. Purified PCR products, as described in section 3.2.3.2, were quantified and brought to a concentration between 2 and 10 ng/µl. Depending on the available barcode and the number of samples, 5 µl PCR product were mixed with 5 µl sequencing primer (5 µM) in 1.5 ml reaction tubes or 96-well PCR-plates, and directly sent for sequencing. The obtained sequences were either checked manually for mutations in Benchling [140] or Mutation Surveyor v5.0.0.

Raw ab1 files were used to check for introduced CRISPR insertions and deletions via sequence decomposition algorithms with the online tools TIDE (tracking of indels by decomposition, [147]) or ICE (inference of CRISPR edits, [148]).

3.2.5.2 Targeted Panel Sequencing

AML t(8;21) patients were screened using a TruSight Myeloid Custom Sequencing panel (66 genes recurrently mutated in AML, Supplement: Table 8.1) according to the manufacturer's instructions. Paired-end sequencing of the libraries was performed on a MiSeq sequencer using the MiSeq reagent Kit v2 (300 cycles; n=42) or on a NextSeq Sequencer (n=289) using the NextSeq 500/550 Mid Output v2 kit (300 cycles, all from Illumina). Samples were excluded if the coverage was below 300x on MiSeq and 500x on NextSeq.

3.2.5.3 Amplicon-Based Deep Sequencing

For amplicon-based deep sequencing, PCR products of approx. 100-200 bp length, covering the area of the expected variant, were generated from genomic DNA. Amplicons were subsequently pooled for library preparation and indexed using the NEBNext Ultra DNA Library Prep Kit (New England Biolabs) according to the manufacturer's instructions. The libraries were single-end sequenced (300 cycles) for validation of CRISPR indels and paired-end sequenced for mutation validation (2x 150 cycles) on an Illumina MiSeq sequencer using the MiSeq Reagent Kit v2 (300 cycles, Illumina).

3.2.6 Data Processing and Statistical Analyses

3.2.6.1 Bioinformatics Processing of Sequencing Data

Targeted Panel Sequencing

FastQ generation was performed by the MiSeq/NextSeq sequencer automatically. The BaseSpace™ App TruSeq Amplicon version 2.0.0 was used to align the amplicon reads of the sequencing panel to the human genome hg19 and generate variant call files. Only reads aligning to regions covered by the targeted sequencing panel were used. Variant annotation was performed using Illumina Variant Studio 3.0 (BaseSpace™ Annotation Engine version 1.4.2.60). The detected variants were filtered using R version 3.4.3. Filtering criteria were i) a minimum VAF of 5%, ii) a minimum reading depth of 50/200, and iii) a minimum of 20/50 variant supporting reads (for sequencing on MiSeq/NextSeq, respectively). Single nucleotide polymorphisms (SNPs) were excluded when reported in dbSNP and ExAC databases with a population frequency >1% and have not been reported in COSMIC as hematopoietic or somatic mutation. I excluded intronic and synonymous variants, as well as variants >5 bases from the exonic regions. Ten healthy subjects were sequenced and variant calling performed as described above, to detect recurrent sequencing artefacts. Variants called in >60% of all samples and detected in healthy subjects were manually checked in Integrative Genome Viewer (IGV; Broad Institute, version 2.4) for repetitive sequences or strand bias and excluded if necessary. A proportion of the obtained variants (281/729, 39%) was validated by targeted re-sequencing (n=245), ddPCR (n=27), or Sanger sequencing (n=9) to estimate the stringency of the filter criteria and exclude false-positive variants.

Amplicon-Based Deep Sequencing

FastQ generation was performed by the MiSeq sequencer automatically. Reads were aligned to human genome hg19 [149] using BWA-MEM [150]. The reads were checked manually for the presence of mutations at the positions previously identified through targeted panel sequencing. Longer insertions and deletions were checked virtually in IGV.

CRISPRseq Analysis

Reads were aligned with BWA-MEM [150] to hg19 and used as input for the “unknown indel analysis”-pipeline of CRISPRseq as described by Tothova et al. [151]. Aligned reads were filtered for those mapping to the target amplicons of *ASXL1*, *DNMT3A*, and *TET2*. The tool iterates through the reads and labels each read that spans the target region with wild type or mutant, depending on whether an insertion or deletion was detected within 25 bp from the

expected CRISPR/Cas9 cut site. A second iteration step for reads that did not span the amplicon was conducted to check if those reads have an insertion or deletion larger than 25 bp. The size of the indel is estimated by the width of the gap in the read for insertions and of the reference for deletions. A Fisher's exact test was performed to estimate, whether an indel was to be expected at that position. P-values were corrected for multiple testing using Benjamini-Hochberg [152]. Total indel fraction for each sample was calculated based on reads defined as mutated and the total number of reads. Single nucleotide variants introduced via HDR were called with DeepVariant [153] (version 0.9.0) in WGS mode limited to the gene of interest (*DNMT3A*). Mutated reads were barcoded based on whether an insertion or deletion was found, the size of the indel, and the start position. The barcode uniquely identifies each indel.

3.2.6.2 Shannon Entropy Analysis

To determine the diversity of the mutational spectrum of a sample, Shannon entropy analysis was performed. Entropy in information theory is a measure for the level of information in a variable [154]. If the entropy is high, a sample contains a high variety of detected insertions in the analyzed sequencing reads. Conversely, if the entropy is low, the same insertions or deletions are found in a high number of reads, hinting to low diversity. Analysis of entropy was performed using the R package "entropy" [155]. Counts for each indel per sample and target obtained by deep sequencing were used as input for the entropy calculation. All indels that were found more than 10 times were included in the analysis.

3.2.6.3 Bradley Terry Model

Data of 208 patients were used to estimate temporal order of mutations with the help of Bradley-Terry model. Bradley-Terry model is a probability model used for predicting outcomes of pairwise comparisons. It estimates the probability whether the assumption individual A is ranked higher than individual B is true [156]. In the case of the analysis performed here, it estimates the probability that the comparison mutation A has a higher variant allele frequency than mutation B, and did therefore occur earlier, is true. The data used here included all patients with at least 2 mutations in different genes and all genes that were mutated in at least 4 patients. Confidence intervals (CI, 95%) were calculated for all variants taking the total depth into account and correcting for the copy number state. Based on these confidence intervals, mutations were compared in a pairwise fashion, assigning the mutation with the higher VAF as winner. The modeling was performed using the R package "BradleyTerry2" [157].

3.2.6.4 Statistical Analysis of in vitro Data

Data obtained from *in vitro* culture experiments were tested for statistical significance using Wilcoxon rank-sum test (also referred to as Mann-Whitney-test) [158]. Test groups were either compared to the wild type group or against each other. This test was used since the data are not normally distributed or the sample size was too small to test for normal distribution (Shapiro wilk test [159]). If multiple groups were compared within one analysis, the resulting p-values were corrected for multiple testing using Benjamini-Hochberg method [152] or Holm's method [160]. Growth curve analysis was performed using the 'compareGrowthCurves' function of the R package "statmod" [161, 162].

4 Results

4.1 Mutations in Epigenetic Regulators in AML

4.1.1 Mutation Spectrum in AML t(8;21)

This first part of the thesis revolves around AML patients harboring the cytogenetic abnormality t(8;21). The described results are part of an article published in the journal Blood in 2019 [163], the respective paper is attached to this thesis. First, we unraveled the mutation spectrum of 331 patients by targeted panel sequencing. Second, we investigated the clonal composition at the time point of diagnosis and after induction chemotherapy. Our goal was to dissect the role of DTA mutations in this particular AML subtype as well as identifying mutations that are critical for disease progression. By investigating coding regions and hotspots of 66 genes recurrently mutated in AML, we found 729 mutations in 216 patients with a mean of 2.2 mutations per patient (maximum: 11 mutations, supplemental spreadsheet file). Mutations were detected in 49 genes, of which 42 were recurrently mutated. The majority of the patients harbored mutations in Ras/receptor tyrosine kinase (RTK) signaling genes (63.4%, Figure 4.1), for example in *KIT* (113 mutations in 89 patients), *FLT3* (61 mutations in 54 patients), and *NRAS* (14.8% of patients). Mutations in epigenetic regulators of transcription were found in 45% of the patients, predominantly in *ASXL2* (15.7%), but also *ASXL1* (12.4%), and *TET2* (7.9%).

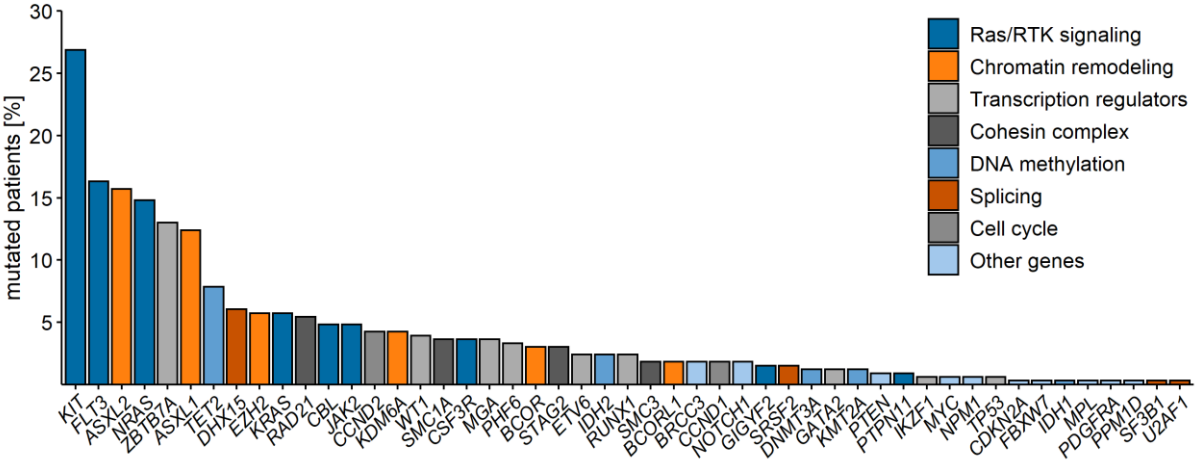


Figure 4.1 Mutation spectrum in AML t(8;21). A total of 331 patients with AML t(8;21) were sequenced with a targeted sequencing panel. All 49 mutated genes are shown. The bars indicate the percentage of patients with at least one mutation in the respective gene. The color indicates the functional category of the gene.

Hotspot mutations were especially found in *KIT*, where the majority of the mutations (86/113) are located in exon 17, mainly at codons D816 and N822. In exon 8 mainly in-frame indels at position 416 and 419 were identified (n=17). *KIT* showed the highest double mutation rate with

21 out of 89 patients harboring at least two mutations in the same gene. At least two mutations in Ras/RTK signaling genes were found in 34% of the patients.

4.1.2 DTA Mutations as Early Events in AML t(8;21)

By comparing variant allele frequencies, we wanted to estimate the sequential order of mutation acquisition. For this analysis, we selected all genes that were mutated in at least four patients. In the complete cohort we observed a median variant allele frequency of 0.28. The highest median VAF had mutations in regulators of DNA methylation, like *DNMT3A* (0.43), *TET2* (0.4), and *IDH2* (0.39, Figure 4.2). Mutations that are found with a high mutation burden in the blood or bone marrow are thought to be present in the majority of the cells, indicating that these mutations occur in stem and/or progenitors that generate clonal mature cell populations. Mutations in tyrosine kinase signaling genes like *KIT* (0.21), *NRAS* (0.16), and *FLT3* (0.11), as well as transcriptional regulators (e.g., *RUNX1*, *GATA2*, *ETV6*) had lower VAFs, indicating that these mutations might be acquired later. It has to be noted that internal tandem duplications in *FLT3* (*FLT*3-ITD) were not included in this analysis, since these mutations cannot be reliably identified by next generation sequencing, but were identified via GeneScan analysis.

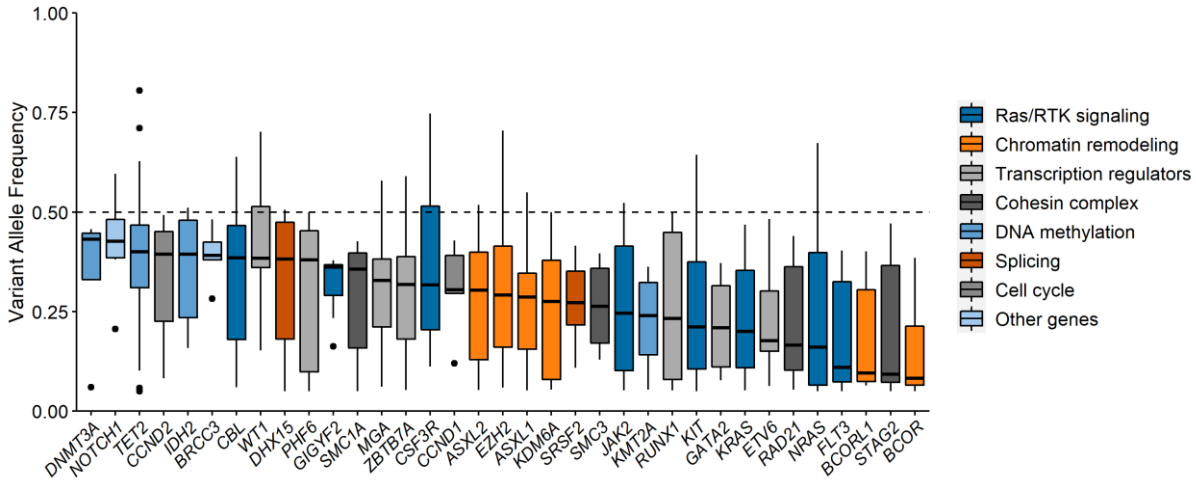


Figure 4.2 Variant allele frequencies for all mutations per gene. Variant allele frequencies are plotted for all gene mutations found in at least 4 patients. The dashed line indicates a variant allele frequency of 0.5. Box plots are color coded according to the functional group. *FLT3* does not include internal tandem duplications (*FLT3*-ITD).

To confirm this observation, I modeled the relative timing of mutation acquisition by pairwise mutation comparison. I used the same cohort as for the VAF analysis, but due to analysis restrictions, excluded patients with less than two mutations. For all mutations 95% confidence intervals (CI) were calculated and the gene with the higher VAF based on the 95% CI assigned as winner and rewarded with one point. This was performed for all pairwise associated genes. The resulting list was used as input for Bradley Terry analysis to model temporal order of mutation acquisition. *MGA* was used as a reference point in this analysis (Figure 4.3). This

model confirmed that mutations in DNA methylation regulators are acquired early (e.g., *TET2*, *KMT2A*, *DNMT3A*) as well as mutations in *CCND2*, *DHX15*, and *ZBTB7A*. The latest events and therefore possible drivers of malignancies are mutations in *JAK2*, *FLT3*, *NRAS*, and *KIT*. Interestingly, mutations in chromatin remodelers are also acquired later in comparison to other events.

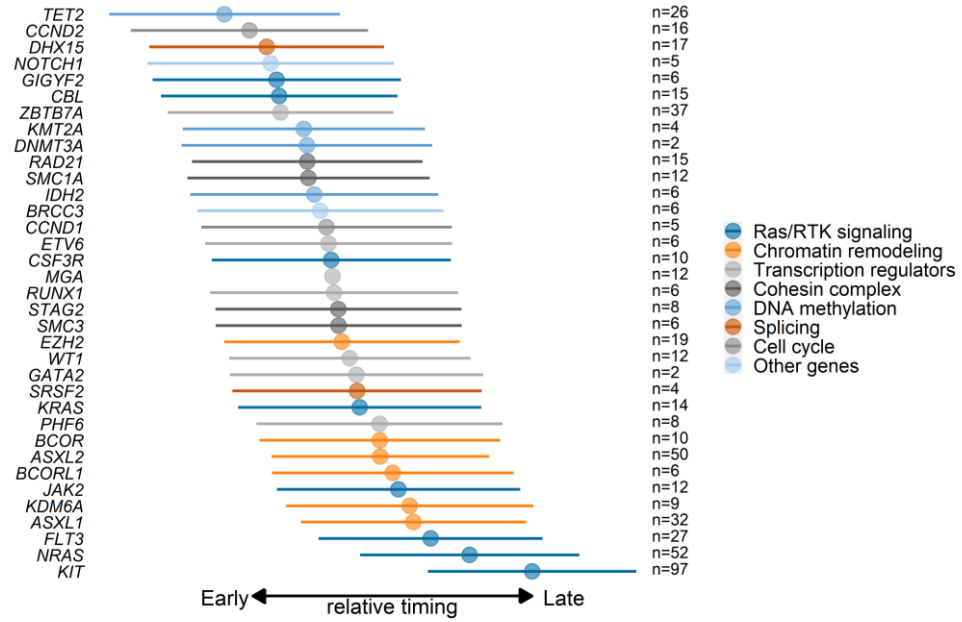


Figure 4.3 Bradley Terry Model of relative timing of mutation acquisition. Relative timing of mutations was modeled by pairwise comparisons. The estimates are calculated in relation *MGA* as the reference point, and standard errors are shown as horizontal bars. Genes are colored by their biological function. The analysis was performed in 208 patients. *FLT3* does not include internal tandem duplications (*FLT3*-ITD). N = number of mutations in the respective gene that were included in the analysis.

Since especially mutations in DNA methylation regulators can be found as early events with a high variant allele frequency in patient samples at diagnosis, they are very likely to represent disease-initiating events. Preleukemic mutations have been implicated in clonal expansion at remission, due to their ability to survive induction chemotherapy, and serve as potential risk factor for relapse. We analyzed data from complete remission samples of 56 patients to determine which clones might be able to survive induction chemotherapy. Simultaneously, we obtained data about the *RUNX1-RUNX1T1* fusion mRNA. The fusion mRNA was found in 26 patients in complete remission and a mutation was found in 12 cases. In 7 patients the fusion mRNA as well as a low level mutation was detected, whereas in 25 patients neither was present in the complete remission samples (Table 4.1). In the remission samples, *RUNX1* and *DNMT3A* variants were found in two patients (out of 8 and 4 patients in the diagnostic cohort, respectively) and variants in *ASXL2*, *CCND1*, *GATA2*, *NOTCH1*, and *NRAS* were each found in one patient. The genes that were frequently mutated in the diagnosis cohort *ZBTB7A* and *KIT* were found in remission samples of 3 and 4 patients, respectively.

Table 4.1 Analysis of 56 complete remission samples. Data for the presence of the fusion mRNA was provided by the respective center. All samples were analyzed by targeted re-sequencing for the presence of the mutations found in the diagnostic sample of the respective patient. Fusion: *RUNX1-RUNX1T1* mRNA.

	Mutation	**no mutation**	
fusion	7	19	26
no fusion	5	25	30
	12	44	56

Relapse data was available in 51 out of the 56 cases and a relapse was diagnosed in 9 of these cases. In five patients with a relapse neither a panel mutation nor the *RUNX1-RUNX1T1* fusion mRNA was detected in complete remission. Three patients had the fusion in the remission sample, and in one patient we found a *DNMT3A* mutation in the remission sample (VAF: 5%). Because of only few patients remaining in this analysis, a further conclusion cannot be drawn. Interestingly, I found in 2 out of 4 patients harboring a *DNMT3A* mutation at diagnosis, the same mutation at remission. In addition, the patient with the more prominent clone developed a relapse. Although more investigations would be needed, this is in line with the previously reported fact that *DNMT3A* mutations persist in remission and might prone the cells to acquire more mutations. Taken together, the results from investigating 331 patients with AML t(8;21), indicate that mutations in *DNMT3A* as well as *TET2* are acquired early, drive expansion of a clone and potentially play a role in initiation of the disease. *ASXL1* mutations are acquired later and might cooperate with other events in disease progression. To dissect the functional role of these gene mutations, I established an *in vitro* model system in human HSPCs.

4.2 Cord Blood Cells as Model System

4.2.1 Cell Culture Feasibility

After identifying main mutations in CHIP being present in early HSCs [54], and identifying mutations in epigenetic regulators as early events in AML, I investigated the functional impact of these gene mutations on hematopoietic stem and progenitor cells.

As a first step in creating a fitting model system, I investigated different sources for the presence of $CD34^+$ HSPCs. As depicted in Figure 4.4, mononuclear cells (MNCs) extracted from peripheral blood (PB), bone marrow (BM), and cord blood (CB) were inspected for their cellular composition as indicated by the presence of distinct cell surface markers. Frequencies of all quantified populations were normalized to 100%, correcting for cell populations that were present in the sample, but not analyzed (i.e., erythrocytes and platelets from incomplete density centrifugation). As expected, I was able to find more than 1% of $CD34^+$ cells among MNCs from bone marrow and almost 1% of $CD34^+$ cells in cord blood MNCs, whereas I detected less than 0.1% $CD34^+$ cells in the peripheral blood. BM and CB MNCs for this analysis were frozen previously and thawed 4 to 20 h before cytometry, whereas the peripheral blood was freshly drawn.

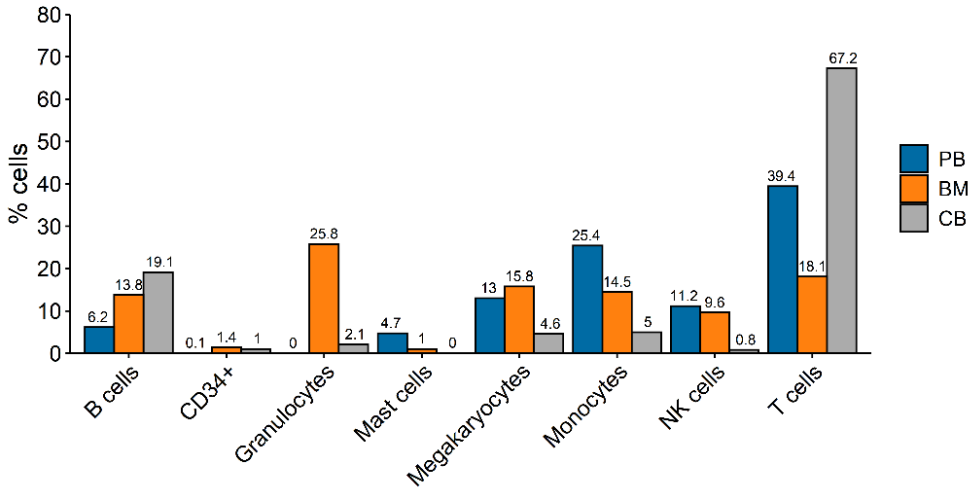


Figure 4.4 Flow cytometry of mononuclear cells from different sources. Freshly isolated or thawed cell populations from peripheral blood (PB), bone marrow (BM), or cord blood (CB) were analyzed for cell surface markers expressed on the main blood cell populations: CD19 B cells, CD34 progenitor cells, CD66b granulocytes, FcεRIα mast cells and basophiles, CD41 megakaryocytes, CD14 monocytes, CD56 natural killer (NK) cells, CD3 T cells. Frequencies were normalized to 100% for all depicted cell populations. The numbers above the bars indicate the frequency of the respective population in %.

These results indicate that cord blood or bone marrow can function as sources for $CD34^+$ cells for our purpose. Among other deciding factors, cord blood can be collected less invasive and often leads to a higher cell output, therefore I selected cord blood cells as model system. The main reason for using umbilical cord blood as cell source was the young biological age and

therefore low mutation burden of the cells, facilitating unperturbed investigation of introduced mutations in the cells.

Next, I addressed the question of the appropriate culture system for CD34$^+$ cells. Since the aim of the project was to investigate the functionality of stem and progenitor cells, I needed a culture system in which the cells retain their stem cell potential as long as possible. I tested four different culture conditions for partially enriched CD34$^+$ cells (35% CD34$^+$ cells, Figure 4.5 A) from frozen CB MNCs. The cells were cultured for 8 days: i) on M2-10B4 feeder cells in MyeloCult expansion medium, ii) on a 1:1 mix of both feeder cell lines in MyeloCult expansion medium, iii) on Sl/Sl feeder cells in MyeloCult expansion medium, or iv) in StemSpan SFEM II expansion medium.

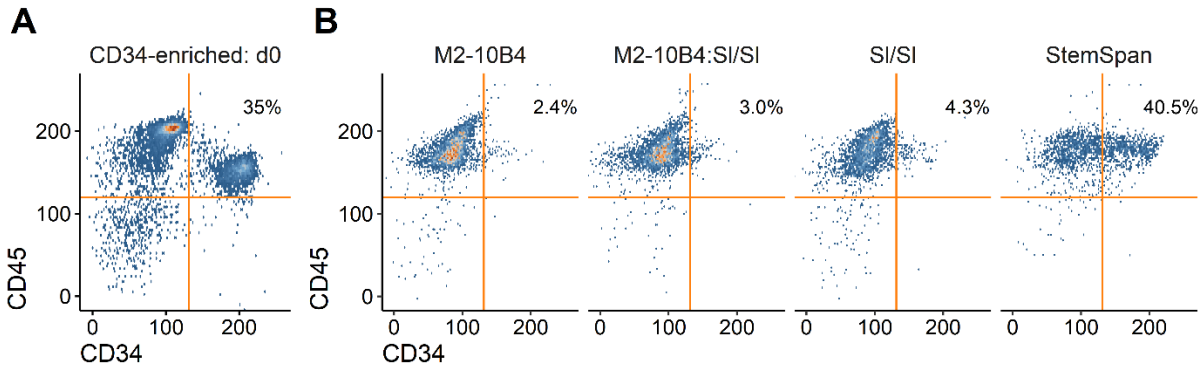


Figure 4.5 Flow cytometry of surface markers CD45 and CD34. A) Flow cytometry of CD34-enriched cells after magnetic separation (MACS). 35% of all cells express CD34 and a low level of CD45. B) Flow cytometry of cells cultured for 8 days in different conditions. Percentages indicate cells that express CD34 and CD45. Different culture conditions are indicated above the plots. M2-10B4: co-culture on M2-10B4 feeder cells in MyeloCult expansion medium, M2-10B4:Sl/Sl: co-culture on a 1:1 mix of M2-10B4 and Sl/Sl feeder cells in MyeloCult expansion medium, Sl/Sl: co-culture on Sl/Sl feeder cells in MyeloCult expansion medium, StemSpan: culture in StemSpan SFEM II expansion medium.I observed an enhanced total cell expansion when cultivating the cells on feeder layers (126-fold and 109-fold expansion in total cell number), whereas the total cell number only increased 4.7-fold without feeder layer (Table 4.2). Focusing on CD34$^+$ cells, the expansion on feeder layers was 8- to 9-fold, so other cell populations expanded more preferentially in these conditions. In StemSpan expansion medium, CD34$^+$ cells expanded 4.5-fold, indicating steady expansion of all cell types. Consistently, the frequency of CD34-expressing cells remained higher in StemSpan expansion medium, whereas in the feeder cell cultures, the level of CD34 expression was decreased (Figure 4.5 B). Therefore, I chose to culture the CD34-enriched cells without feeder cells to retain a high level of CD34-expressing progenitor cells in the first week of culture.

Table 4.2 Cell count after 8-day expansion of CD34-enriched cells. Partially CD34-enriched cells were cultured on various feeder layers (M2-10B2 cells, Sl/Sl cells, or a1:1 mix of both cell lines), or in StemSpan expansion medium for 8 days. Total cell counts were determined via trypan blue exclusion. $CD34^+$ cell frequencies were determined via flow cytometry. Fold expansion was calculated by dividing the total cell count after 8 days by the cell number at seeding (10,000).

	Total cells		**$CD34^+$ cells**	
	count*	fold expansion	count*	fold expansion
M2-10B4	$1.5x10^6$	126	$3.6x10^4$	8.5
M2-10B4:Sl/Sl	$1.3x10^6$	109.4	$3.9x10^4$	9.3
Sl/Sl	$7.5x10^5$	62.5	$3.3x10^4$	7.7
StemSpan	$4.7x10^4$	3.9	$1.9x10^4$	4.5

*cell counts were rounded to increase readability of the table. Fold increase was calculated with not rounded numbers.

To finalize the culture system establishment, I tested various cytokine cocktails as medium supplement for CD34-enriched hematopoietic cells. The purpose of this experiment was to determine which cytokine cocktail supports i) maintenance of stem cell phenotype and ii) differentiation without generating a differentiation bias. All media were prepared with StemSpan SFEM II containing 2 mM L-Glut and 100 U/ml P/S. The different cytokine combinations are listed in Table 4.3.

Table 4.3 Cytokine cocktails for $CD34^+$ cell culture. SCF: stem cell factor, FLT3-L: FMS-like tyrosine kinase 3 ligand, TPO: thrombopoietin, IL-3: interleukin 3, IL-6: interleukin 6, G-CSF: granulocyte-colony stimulating factor, GM-CSF: granulocyte macrophage-colony stimulating factor, EPO: erythropoietin.

	Medium A	**Medium B**	**Medium C**	**Medium D**	**Medium E**	**Medium F**
SCF [ng/ml]	.	100	100	50	100	25
FLT3-L [ng/ml]	.	100	100	50	100	10
TPO [ng/ml]	.	100	.	20	100	10
IL-3 [ng/ml]	.	.	20	10	100	10
IL-6 [ng/ml]	.	.	20	20	100	10
G-CSF [ng/ml]	.	.	.	.	100	20
GM-CSF [ng/ml]	.	.	.	.	100	5
EPO [ng/ml]	.	.	.	.	.	10

$CD34^+$ cells (>90% enriched, Supplement; Figure 8.2, CB002) were seeded in 6 different media at a density of $3x10^5$ cells/ml and 1 volume of medium was added every 3 to 4 days. I investigated cell viability and cell growth over a period of 3 weeks. Every 7 days cells were checked for surface marker expression via flow cytometry. In the medium without cytokines (medium A) the cell viability was very low and no cell growth was observed, indicating that the presence of cytokines is essential for primary $CD34^+$ cells. In all other conditions the cells had a high viability (>60% viable cells) and expanded at similar rates (Figure 4.6 A and B). Differences in cell growth were mainly observed after day 18 when the cells reached the limits of the culture vessels and might have used up all available culture supplements. After 7 days of

culture, the cells lost CD34 and started expressing CD66b (granulocyte marker) and CD14 (monocytes and macrophages), especially when G-CSF and GM-CSF were supplemented (medium E and medium F, Figure 4.6 C). In medium B (SCF, FLT3-L, and TPO), the cells retained the highest CD34 expression level, with a detectable proportion of cells retaining an immature $CD34^{+}CD38^{-}$ phenotype after 3 weeks of culture. In medium C and medium D, the cells began to highly express CD66b and CD14 after day 7 and gradually lost CD34 expression. In medium D, the frequency of CD34-expressing cells remained higher at day 14 and day 21 compared to medium C, but not as high as in medium B. For stem cell expansion before transfection, cytokine cocktail B was selected since I observed the highest retention of CD34 marker expression. For short-term culture experiments, with the purpose of differentiation to various lineages, I used cytokine cocktail D.

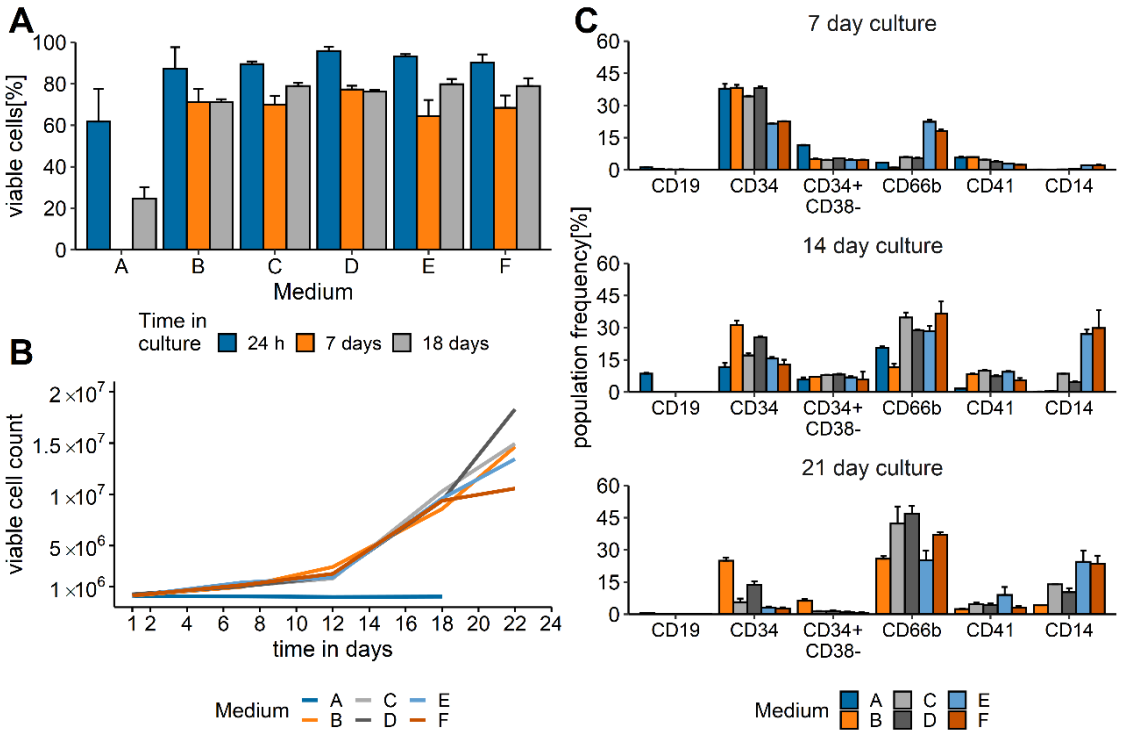


Figure 4.6 $CD34^{+}$ cell culture with different cytokine cocktails. A) Viability of the cells in media with different cytokine cocktails (see Table 4.3) as estimated by trypan blue exclusion at three different time points after seeding (24 h, 7 days, and 18 days). For medium A, 7-day data could not be collected. The bars show mean and standard deviation of two replicates. B) Total cell count in media with different cytokines was obtained by trypan blue exclusion. For medium A, only two time points were counted (24 h and 18 days). For all other time points, cells were counted 24 h, 7, 18, and 22 days after seeding. C) Flow cytometry analysis of cell surface markers for major blood cell populations after 7, 14, and 21 days in culture: $CD19^{+}$ B cells, $CD34^{+}$ and $CD34^{+}CD38^{-}$ progenitor cells, $CD66b^{+}$ granulocytes, $CD41^{+}$ megakaryocytes, and $CD14^{+}$ monocytes. The bars show mean and standard deviation of duplicates.

4.2.2 Target Gene Expression

In order to characterize the chosen model system, I determined the abundance of target gene expression in CD34-enriched cells via quantitative real-time PCR (qRT-PCR). The amount of *ASXL1*, *DNMT3A*, and *TET2* mRNA in CD34-enriched cells was compared with the cell line K-562. All targets were at least 2-fold enriched in $CD34^+$ cells from cord blood after positive selection by magnetic separation and 2-day expansion in expansion medium (Figure 4.7 A, day 0, purity: see Supplement Figure 8.2). After three days in liquid culture the expression of *DNMT3A* mRNA was reduced compared to day 0, whereas the difference for the other targets was less pronounced (Figure 4.7 B).

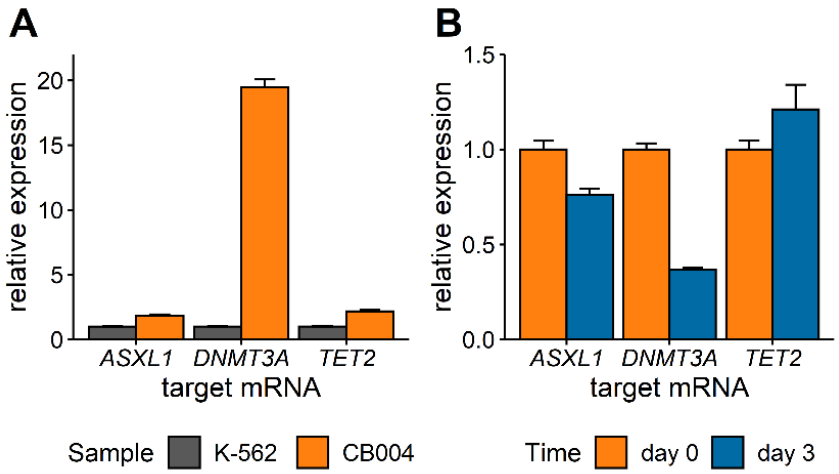


Figure 4.7 Relative expression of *ASXL1*, *DNMT3A*, and *TET2* mRNA in cord blood $CD34^+$ cells. A) Expression of all three targets in $CD34^+$ cells of specimen CB004 at day 0 was compared to the test cell line K-562. B) Relative expression of all three targets at day 3 of liquid culture. Relative expression was calculated with the $\Delta\Delta C_T$-method. The bars show $2^{(-\Delta\Delta CT)}$ (mean+SD) of three technical replicates, to set the reference sample to a value of 1. *GAPDH* was used as endogenous control.

4.3 Modeling Preleukemic Mutations by Genetic Modification

4.3.1 Vector-based Delivery of CRISPR/Cas9

Multiple studies investigated functional knockouts and overexpression models of *DNMT3A* [97, 99, 164–167], *TET2* [116–118, 168–170], and *ASXL1* [123–125, 171], with a preference for murine models. To investigate the functional impact of these mutations in human hematopoietic stem and progenitor cells, I decided to model previously identified mutations in human HSPCs *in vitro*. We were interested in modeling site- and domain-specific mutations, representing the most frequently found mutations in humans, instead of functional knockouts of the target genes. We established a CRISPR-based genetic modification approach that was first tested in the cell line K-562. In the next chapters the process of establishing the optimal delivery and modification approach will be described.

First, site-specific mutations were introduced in the cell line K-562 by delivering a plasmid expressing Cas9-nuclease and green fluorescent protein (GFP) under control of a CMV promotor (pCas9_GFP, Figure 4.8 A) together with a plasmid expressing a single guide RNA and red fluorescent protein (RFP) (pLKO5.sgRNA.EFS.tRFP, Figure 4.8 B) via Amaxa Nucleofection®. The purpose of this experiment was to test the efficiency of the designed guide RNAs.

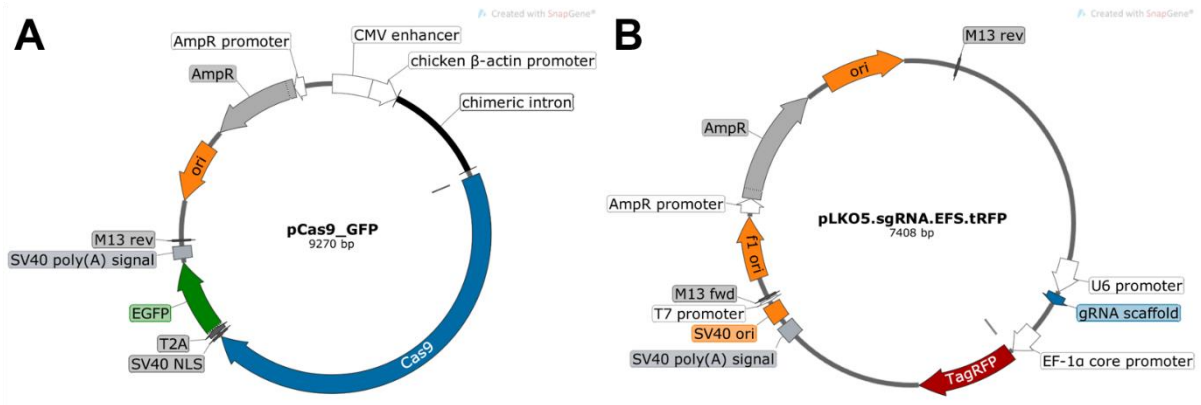


Figure 4.8 Vector maps of pCas9_GFP and pLKO5.sgRNA.EFS.tRFP. A) The vector pCas9_GFP expression vector was used to transiently express Cas9 nuclease together with green fluorescent protein (GFP) in the target cells. B) A single guide RNA target sequence was cloned into the pLKO5.sgRNA.EFS.tRFP expression vector, which already has a gRNA scaffold. This vector expresses red fluorescent protein (tRFP), to trace the vector expression in the target cell. The vector maps were created using SnapGene software (from Insightful Science; available at snapgene.com).

Two different sgRNAs targeting *DNMT3A* exon 23 (T1 and T3) and one sgRNA targeting *TET2* exon 6 were tested. After confirming delivery efficiency (GFP/RFP-positive cells), I extracted DNA at various time points after transfection to investigate knockout efficiency. The cells showed decreased viability in the first three days post transfection (Figure 4.9 A).

Electroporation of the plasmids seemed to interfere with the viability of the cells. Between 3 and 6 days after transfection, the cells regained their capacity and were proliferating again. One day post transfection, 50% to 70% of the cells were GFP/RFP-positive (Figure 4.9 B). I enriched the positive fraction by fluorescence activated cell sorting and tracked the GFP/RFP expression for 9 more days. After 6 days the GFP/RFP-positive cell fraction decreased notably, especially for the enriched cells. After 10 days, less than 2% of the bulk cells showed plasmid expression. The GFP/RFP-positive cell fraction decreased faster after the cell viability increased. K-562 have a relatively short doubling time of around 24 h [172] and loose ectopically expressed plasmids fast. Next, DNA was extracted to investigate whether our CRISPR-knockout strategy with ectopically expressed plasmid was successful for this cell line. First, Cas9 nuclease activity in the target region was investigated via T7 endonuclease I assay (T7E1).

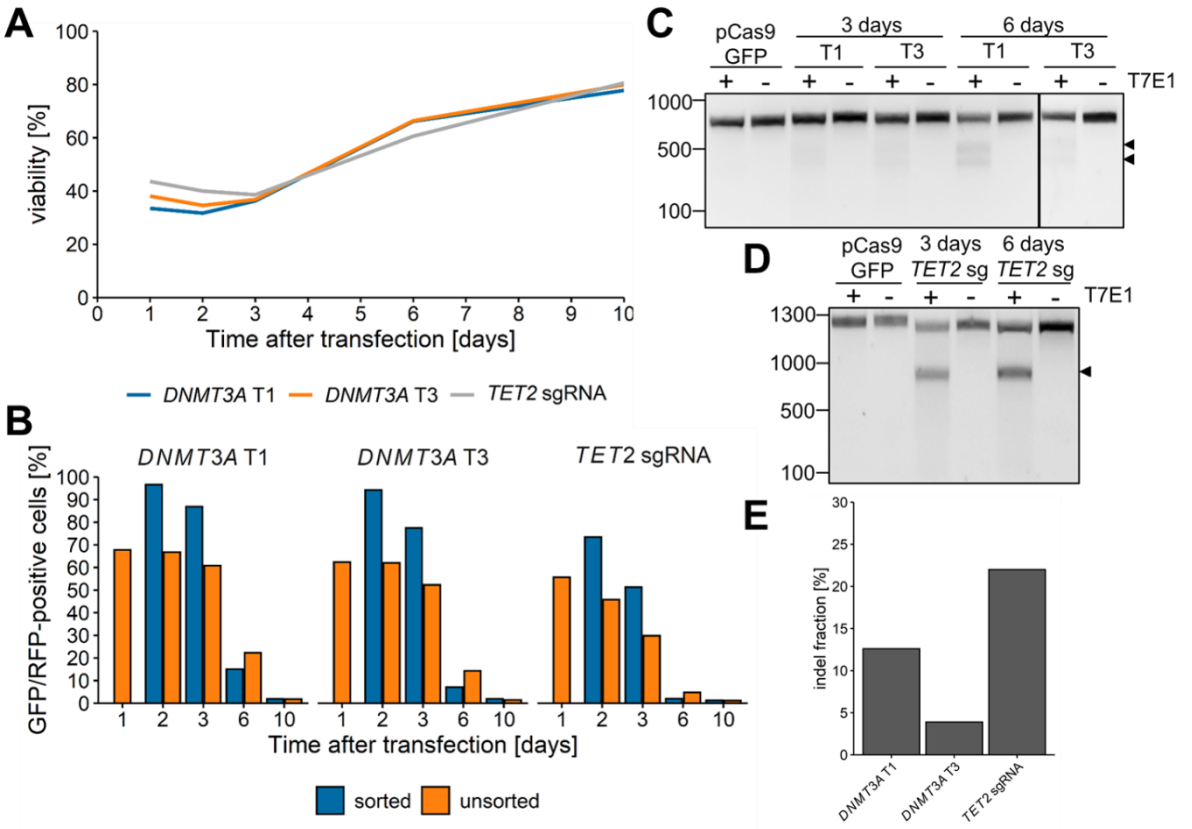


Figure 4.9 CRISPR-mediated knockout in the cell line K-562. A) Viability of K-562 cells after transfection with the vectors pCas9_GFP and pLKO5.sgRNA.EFS.tRFP carrying the sgRNAs targeting *DNMT3A* (*DNMT3A* T1 and *DNMT3A* T3) or *TET2* (*TET2* sgRNA). Viability was determined via flow cytometry. B) Flow cytometry analysis of GFP and RFP after transfection. The analysis was carried out in unsorted cells and cells that were sorted 2 days after transfection. The analysis was carried out once. C + D) T7 endonuclease I assay to determine CRISPR cutting for *DNMT3A* and *TET2* sgRNAs, respectively. The arrows on the right side indicate smaller cleavage bands. E) Indel fraction for each sgRNA as determined by TIDE analysis. The analysis was performed once per target.

In the *DNMT3A*-targeted samples only weak cleavage bands were visible 3 days after transfection, increasing slightly after 6 days (Figure 4.9 C), indicating a small fraction of cells carrying insertions or deletions (indels). The result from the T7E1 indicates that *DNMT3A* T1 targeted the mutation site more efficiently. With *TET2* sgRNA on the other hand, a higher proportion of cells seemed to have acquired indels (Figure 4.9 D). The observations from the T7E1 were confirmed by Sanger sequencing and tracking of indels by decomposition (TIDE [147]). In detail, the calculated indel fraction for cells targeted with *DNMT3A* T1 was 12.6% after 10 d, whereas the indel fraction for cells targeted with *DNMT3A* T3 was 3.9%. The indel fraction in *TET2*-targeted cells was 22% (Figure 4.9 E). All in all, the general targeted knockout strategy produced indels in the cells with different efficiency, depending mainly on the guide RNA.

In the next step, introduction of point mutations with this CRISPR-strategy was tested in the cell line K-562. For this approach, I introduced a 200 bp single stranded homology directed repair-template together with the CRISPR-vector system targeting *DNMT3A* exon 23 via electroporation. One day after transfection the majority of the cells expressed GFP and RFP, confirming a successful delivery of the plasmid. At the same time GFP/RFP-positive single cells were sorted into 96-well plates to investigate the frequency of cells carrying the desired point mutation. Of 48 sorted single cell clones, mutations were confirmed in 9 clones. The mutation frequency in the single cell clones was higher than the bulk data indicated. For the same experiment, I was not able to confirm indels via sequence decomposition after 5 to 9 days of bulk cultures (Supplement: Figure 8.3). This observation might indicate that the wild type cells in the bulk culture outcompeted the mutated cells faster than expected and indels were not detectable in the bulk after 5 days of culture, although a toxic effect of the mutations is a more likely explanation for the observed effect. By sorting single cells 24 h after transfection, mutated cells were captured. One clone carried the desired point mutation in exon 23 (amino acid change: R882H) in one allele (Figure 4.10 A). The VAF was determined via digital droplet PCR (ddPCR) in three replicates and was 0.33±0.02 (Figure 4.10 B). The cell line K-562 has been karyotyped extensively in several studies [173, 174] and has more than two alleles for most chromosomes. In our case the cell line carries three copies of chromosome 2, explaining a mutation burden of 33%. Nevertheless, I was able to introduce the desired point mutation with this genetic modification strategy.

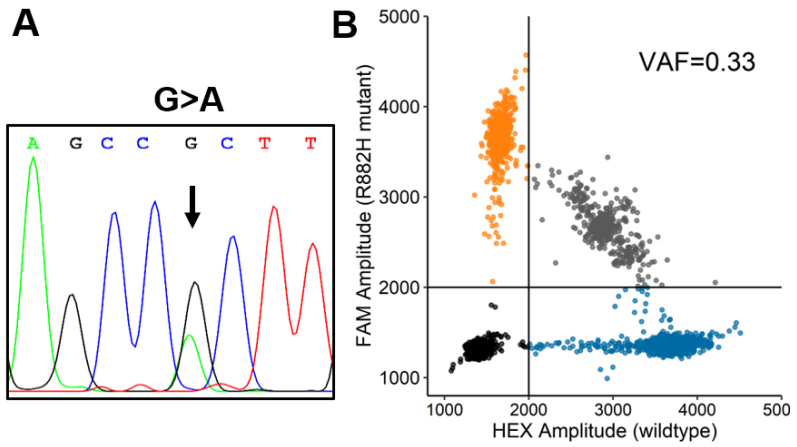


Figure 4.10 CRISPR-mediated point mutation in K-562.A) Sanger sequencing track of the single cell clone carrying the desired point mutation G to A, which translates to the mutation R882H in *DNMT3A*. B) Validation of the mutation via digital droplet PCR. The variant allele frequency (VAF) was 0.33, which results from the karyotype of the cell line K-562. Digital droplet PCR was performed with three replicates.

Next, I investigated whether this approach was applicable to the target cell system. Since primary cells are not easily transfected, I tested various delivery and enrichment strategies. First, the idea was to transfect mononuclear cells with the expression vectors and sort GFP/RFP-positive $CD34^+$ cells from the bulk. For this test, I used frozen cord blood mononuclear cells and although I depleted the sample for lineage positive cells before sorting, $CD34^+$ cell fraction was very low (Supplement: Figure 8.4). To improve this outcome, the cells were first enriched for CD34 by magnetic cells sorting (MACS), expanded for up to one week, and electroporated with the vector constructs. The Amaxa Nucleofector system does not allow modification of the electroporation parameters, therefore the cells were electroporated with the same condition, established for $CD34^+$ cells (U-008), but with varying cell numbers and concentrations to test if downscaling was possible (Supplement: Table 8.2). Different electroporation parameters were tested for Neon transfection system. GFP-expressing cells and viability were tracked after transfection with pCas9_GFP. Every condition was only tested once, due to a limited number of cells available for this experiment. One day after electroporation the maximum viability was around 20% for both Amaxa and Neon transfection systems, decreasing further after 48 h (Figure 4.11 A). GFP expression was higher after transfection with the Amaxa system, using optimized transfection conditions (condition 1), but was below 15% 48 h after transfection (Figure 4.11 B). Since only 10% of the cells survive the transfection procedure, very few cells were left in total.

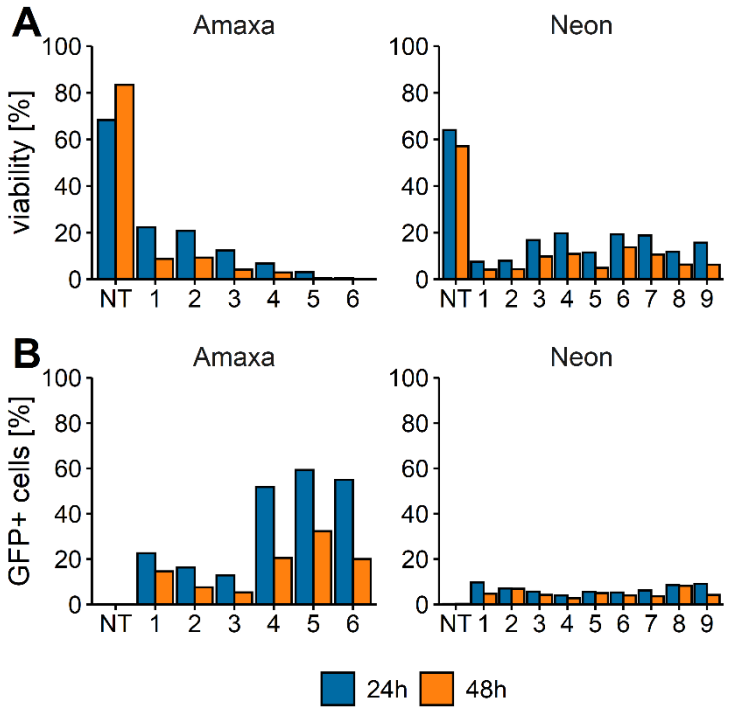


Figure 4.11 Transfection optimization for CD34⁺ cells.A) Viability of transfected cells 24 h and 48 h after transfection with the vector pCas_GFP using either Amaxa nucleofection or Neon transfection system. Viability was determined by flow cytometry. B) Frequency of GFP-positive cells after transfection with the vector pCas9_GFP using either Amaxa nucleofection or Neon transfection system. GFP-positive cells were determined as a fraction of viable cells by flow cytometry. Every condition was tested once.

Taken together, neither electroporation system is optimal for delivering the CRISPR-vector system into primary CD34-enriched cells. The vector system is inefficient for primary cells, leaving very little cell material for downstream experiments. Since working with single cell clones in primary cells is not feasible, I had to find a method to improve viability of the cells after transfection and increase the CRISPR-efficiency so that it would be possible to work in a cell bulk. These preliminary experiments, helped to understand the CRISPR strategy and served as a baseline for improving the experimental set-up.

4.3.2 Modeling Clonal Hematopoiesis by RNP transfection

A second CRISPR-delivery system without expression vectors was tested after determining that using a vector-based system is highly inefficient in primary hematopoietic cells. This system is based on the formation of a Cas9-sgRNA RNP. Commercially available Cas9-nuclease is complexed with an *in vitro*-transcribed or synthetic guide RNA. The RNP-based approach has the advantage, that the protein-RNA complex is already in its final form and can target the DNA. A disadvantage of this system is, that it is not possible to select the positively transfected cells, therefore the editing efficiency has to be very high.

The already established guide RNAs targeting *DNMT3A* and *TET2* (see chapter 4.3.1) and sgRNAs targeting *ASXL1* were *in vitro*-transcribed, as described in the methods section, or ordered as synthetic guide RNAs from Synthego (indicated by *, Figure 4.12), and tested in the cell line K-562. Electroporation was carried out with the Neon transfection system, which has more flexibility and needs less cells for one transfection reaction.

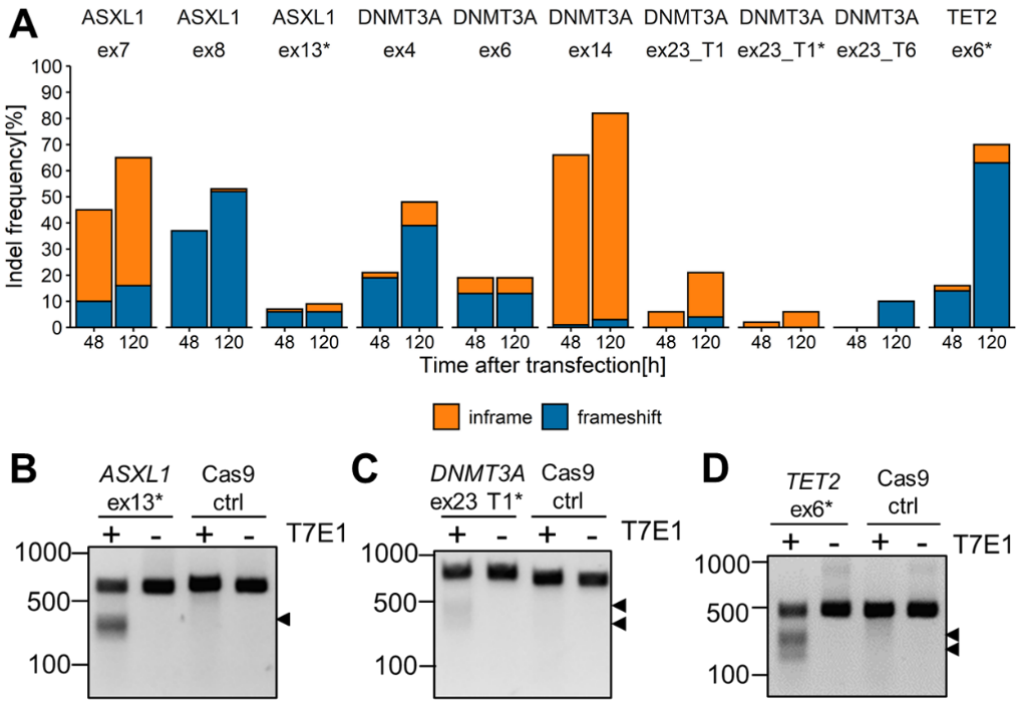


Figure 4.12 CRISPR-editing in K-562 using Cas9-riboneucleoprotein. A) Indel frequency as determined via ICE sequence decomposition for 10 different sgRNAs targeting the three genes *ASXL1*, *DNMT3A*, and *TET2*. Inframe indels are all indels that have a multiple of 3 bp deletion or insertion. B-D) T7 endonuclease I assay of K-562 cells transfected with B) *ASXL1* exon 13 RNP, C) *DNMT3A* exon 23 T1 RNP and D) *TET2* exon 6 RNP or Cas9 protein without sgRNA 5 days after transfection. The arrowheads show cleavage bands. Cas9 ctrl: Cells transfected with Cas9 without guide RNA. T7E1: T7 endoculease I. * = synthetic guide RNA.

After 48 h and 7 days, indel frequency was calculated via Sanger trace sequence decomposition, using an improved online tool (ICE) that is less dependent on the quality of the Sanger sequence [148]. The TIDE algorithm tends to overestimate the indel fraction in low quality Sanger sequencing files. Except for *TET2*, more than one guide RNA was tested, to select the ones with

highest knockout efficiency. As depicted in Figure 4.12 A, two guide RNAs led to high proportions of predicted inframe deletions and were not followed up further (*ASXL1* exon 7 and *DNMT3A* exon 14). The highest predicted frameshift indel frequencies were achieved using *ASXL1* exon 8 sgRNA, *DNMT3A* exon 4 sgRNA, and *TET2* exon 6* sgRNA. Generally, the editing efficiency in the test cell line was higher using the RNP approach than with the vector approach, especially 5 days after transfection. Despite showing not optimal editing outcomes, *ASXL1* exon 13* guide RNA and *DNMT3A* exon 23 T1* guide RNA were used for establishing experiments, since these guides were needed for introducing point mutations. In the above-described experiment, a molar ratio of 1:5 (Cas9:sgRNA) was used for IVT guides and a molar ratio of 1:3 was used for synthetic guides. T7E1 assay confirmed that the three selected sgRNAs were able to introduce mutations in the target genes 5 days after transfection (Figure 4.12 B, C, and D for the three synthetic guide RNAs). The cleavage band in the *ASXL1* targeted sample is relatively pronounced, indicating that the decomposition algorithm might have underestimated the editing efficiency.

In the next experiment, molar ratios from 1:2 to 1:10 of Cas9-protein and *DNMT3A* exon 23 T1* were tested in K-562 to improve the editing efficiency. Editing efficiencies were modeled with sequence decomposition after 48 h and 7 days. Increasing the amount of guide RNA with constant amount of Cas9 protein greatly improved editing in the cell bulk (Figure 4.13).

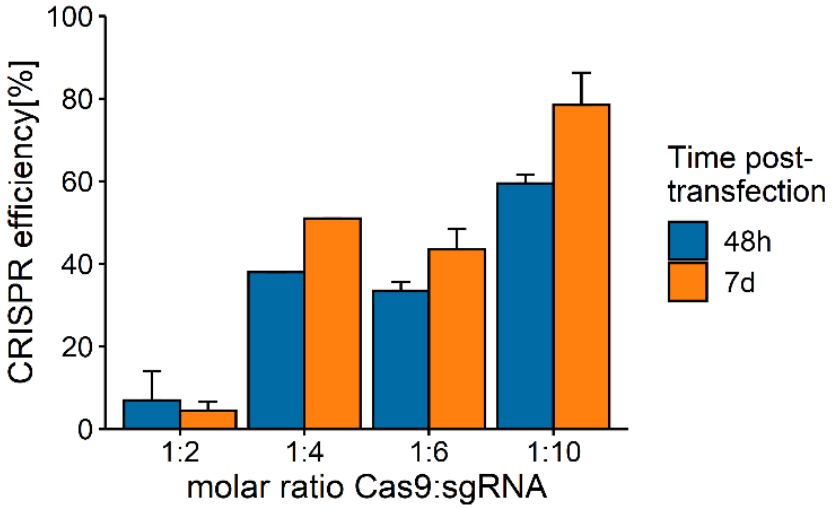


Figure 4.13 Optimization of RNP-mediated CRISPR editing in K-562. CRISPR efficiency in K-562 transfected with *DNMT3A* exon 23 T1* RNPs at different molar ratios of Cas9:sgRNA. ICE analysis was performed 48 h and 7 d after transfection. The editing efficiency includes inframe and frameshift indels. The bars show mean+SD of the calculated CRISPR efficiency for duplicates.

In this setting an editing efficiency (frameshift and inframe indels) of around 80% could be achieved with a 1:10 ratio 7 days after transfection. The estimated knockout score for this condition was 37% (data not shown).

Furthermore, the HDR-mediated editing efficiency was tested in combination with RNP transfection. To this end, two different single-stranded DNA (ssDNA) HDR-templates were tested in two different concentrations (10 pmol and 30 pmol). The templates were designed with asymmetric homology arms, with a 35 bp homology arm 3' of the estimated cut site and 65 bp homology arm 5' of the estimated cut site. One template was designed homologous to the strand on which the protospacer adjacent motif (PAM) was located (PAM strand) or homologous to the complementary strand (non-PAM strand, Figure 4.14).

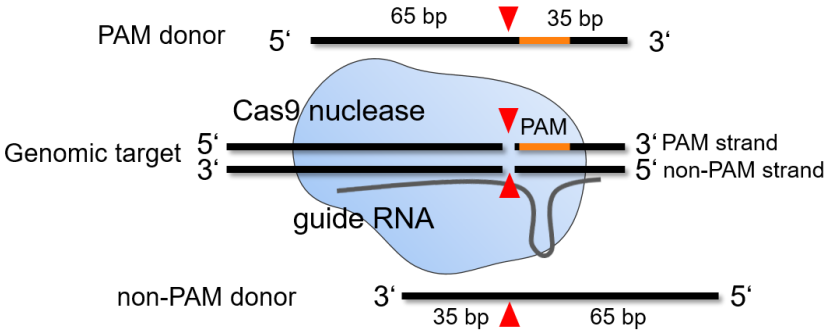


Figure 4.14 Donor design for homology directed repair. Single stranded DNA donors for homology directed repair, to introduce point mutations, were designed homologous to the PAM strand (PAM donor) or the non-PAM strand (non-PAM donor). All templates have asymmetric homology arms (5' of the estimated cut site 65 bp, 3' of the estimated cut site 35 bp). The cut sites are indicated with red arrowheads. PAM: protospacer adjacent motif.

The desired point mutation in exon 23 of *DNMT3A* could be introduced with all donors (Figure 4.15). The lower concentration of each ssDNA donor was more effective in direct comparison. Overall, the HDR template homologous to the PAM strand generated the highest proportion of point mutations (25.5%±7.8%).

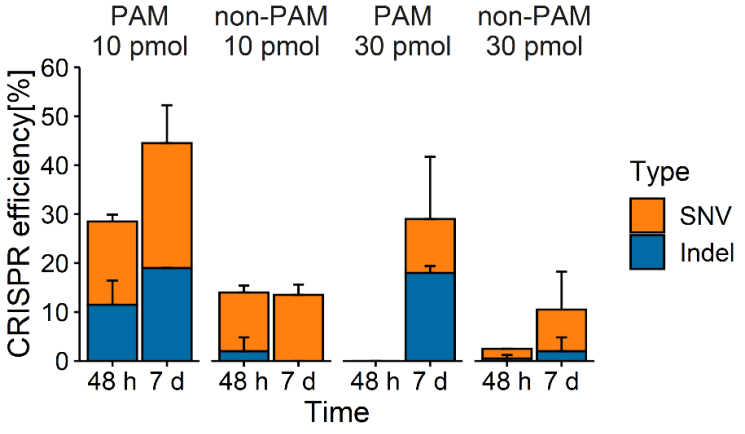


Figure 4.15 CRISPR point mutations via HDR in K-562. CRISPR efficiency in K-562 transfected with *DNMT3A* exon 23 T1* RNPs and a single stranded homology directed repair template at different concentrations. ICE analysis was performed 48 h and 7 d after transfection. The editing efficiency includes inframe and frameshift indels (blue) and point mutations (single nucleotide variant, SNV). The bars show mean+SD of the calculated CRISPR efficiency for duplicates. PAM: HDR template was designed homologous to the strand on which the protospacer adjacent motif (PAM strand) is located. Non-PAM: HDR template is homologous to complementary strand (non-PAM strand).

Finally, to confirm that the RNP-based delivery method also worked for primary material, mononuclear cells were isolated and expanded for 2 days. The cells were enriched for CD34 (Supplement: Figure 8.5) and expanded for four more days before transfection with RNPs and the vector system both targeting *TET2* exon 6. The viability of the cells was around 60% one day after transfection for all conditions (Figure 4.16 A). After 48 h, the cells transfected with vectors exhibited decreased viability, which was even more pronounced after 7 days of culture. The cells transfected with RNPs exhibited a higher viability in general. By sequence decomposition analysis, I was able to confirm a high proportion of indels in $CD34^+$ cells targeted with *TET2* ex6 RNPs (Figure 4.16 B). Ribonucleoprotein-based delivery of Cas9 nuclease and guide RNA was highly efficient as well as less cytotoxic than vector-based delivery in primary $CD34^+$ cells and is therefore the optimal experimental set-up for my purpose.

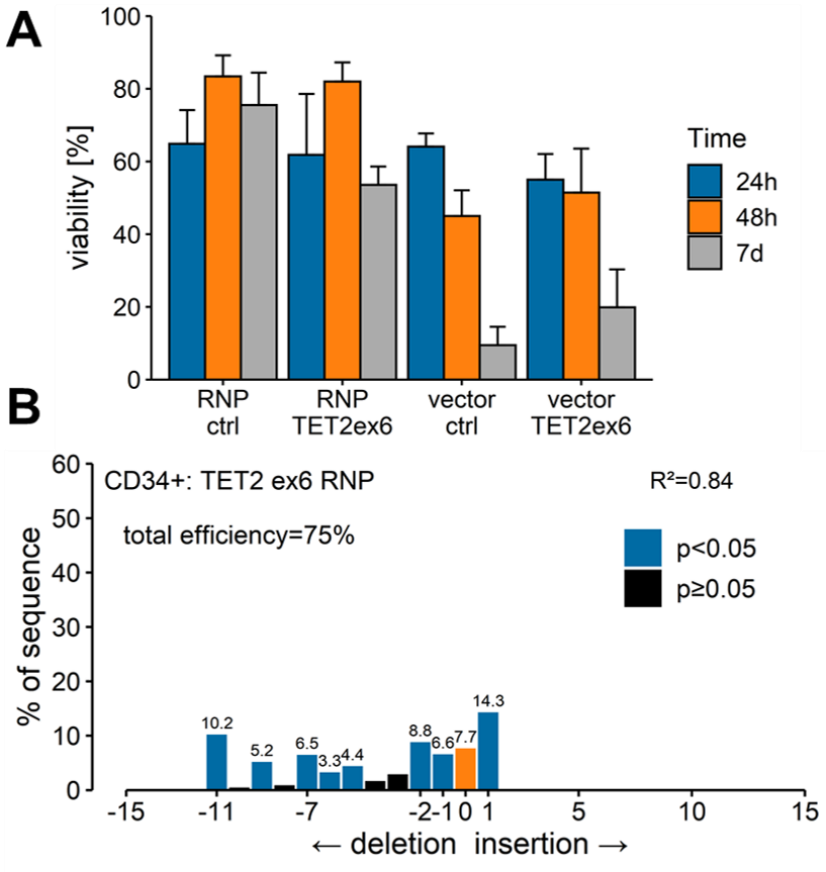


Figure 4.16 Comparison of CRISPR editing with vectors and RNP in $CD34^+$ cells. Viability $CD34^+$ cells after transfection with RNPs or the vector system (pCas9_GFP and pLKO5.sgRNA.EFS.tRFP:TET2ex6 sgRNA) targeting exon 6 of TET2. Viability was determined via trypan blue exclusion (mean+SD). RNP ctrl: Cas9 protein without sgRNA; vector ctrl: pCas9_GFP. B) Indel distribution in $CD34^+$ cells targeted with TET2 exon 6 RNPs. R^2 and p-value calculations are automatically performed by the TIDE algorithm, where R^2 is calculated to assess the goodness of fit and the p-value is based on the estimated abundance of each indel [147].

4.4 Functional Impact of Mutations in Epigenetic Regulators on HSPCs

4.4.1 Characterizing the Knockout

Parts of the following results are currently submitted to be published. The general experimental procedure was performed as follows: MNC were isolated from freshly drawn umbilical cord blood by Biocoll density centrifugation. Subsequently $CD34^+$ cells were collected by magnetic separation and cultured for 2 days in StemSpan expansion medium. After expansion, the cells were transfected with RNPs targeting *ASXL1* exon 13, *TET2* exon 6, and *DNMT3A* exon 23 plus the homology directed repair templates to introduce the mutation R882H in *DNMT3A* or an 8 bp deletion in *ASXL1* (Figure 4.17).

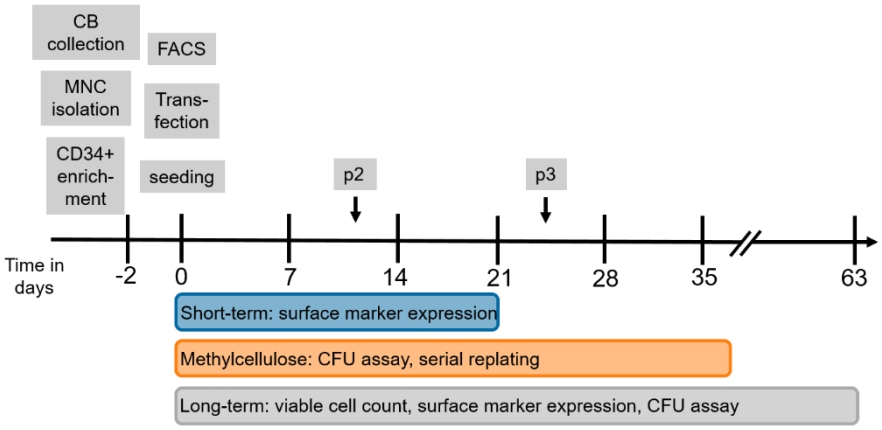


Figure 4.17 Schema of general experimental procedure. $CD34^+$ cells were isolated from cord blood and expanded for 2 days in StemSpan expansion medium. Surface marker expression was checked via flow cytometry to assure $CD34^+$ cell enrichment before transfection. Directly after transfection, cells were seeded for short-term culture in liquid medium, in methylcellulose for CFU assay and replating experiments, and on irradiated feeder cells for long-term culture and long-term culture initiating cell assay (LTC-IC). P2 and p3 indicate the time points of CFU counting and serial replating in methylcellulose.

Introduced mutations were checked via Sanger sequencing and sequence decomposition (ICE analysis) at various time points as described above for all experiments. In Figure 4.18 an overview over the predicted mutations per target are shown. For *DNMT3A*, the indel VAF (knockout, KO) and the VAF of the introduced mutations (single nucleotide variant, SNV) are shown separately. The mean indel frequency was around 40% for $DNMT3A^{mut}$. In a high proportion of samples, SNVs could be introduced with an equally high VAF. For $ASXL1^{mut}$ all mutations were summed up as total 'KO', since the introduced mutation is an 8 bp deletion, and had a mean VAF of around 40%. The high variance in predicted mutations might result from the quality of the Sanger sequencing and the size of the generated insertions and deletions. Indels over 21 bp are less frequently detected by the decomposition algorithm, especially when the signal has a high background noise. Samples targeted with *TET2* sgRNA had the highest indel proportion (mean=62.2%). I could detect mutations in all samples, except for one *ASXL1* targeted sample with a predicted indel frequency of 1%.

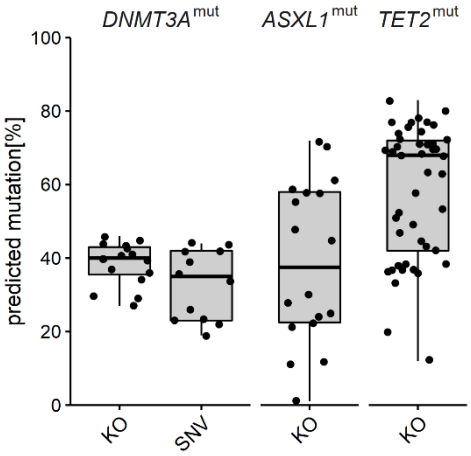


Figure 4.18 Overview of mutation efficiency in cord blood CD34⁺ cells. CD34⁺ enriched cells were transfected with RNPs targeting *DNMT3A* exon 23, *ASXL1* exon 13, and *TET2* exon 6 to introduce indels (KO = knockout). Point mutations (SNV) were introduced by addition of a single stranded DNA template carrying the mutation. Mutation efficiencies were estimated via sequence decomposition of Sanger sequencing tracks of at least 5 biological replicates and 2 transfection replicates at various time points after transfection.

In Figure 4.19 examples of the predicted sequence contributions are shown for *TET2*, *DNMT3A*, and *ASXL1* targeted samples, respectively. The pictures were chosen randomly from all analyzed samples and were trimmed to show only the most frequent indels. Especially *ASXL1*^{mut} samples (Figure 4.19 A) have longer insertions and deletions that could possibly be missed in bad quality Sanger sequencing tracks. Nevertheless, the sequence decomposition allowed to estimate the frequency and the type of indels in the bulk sample.

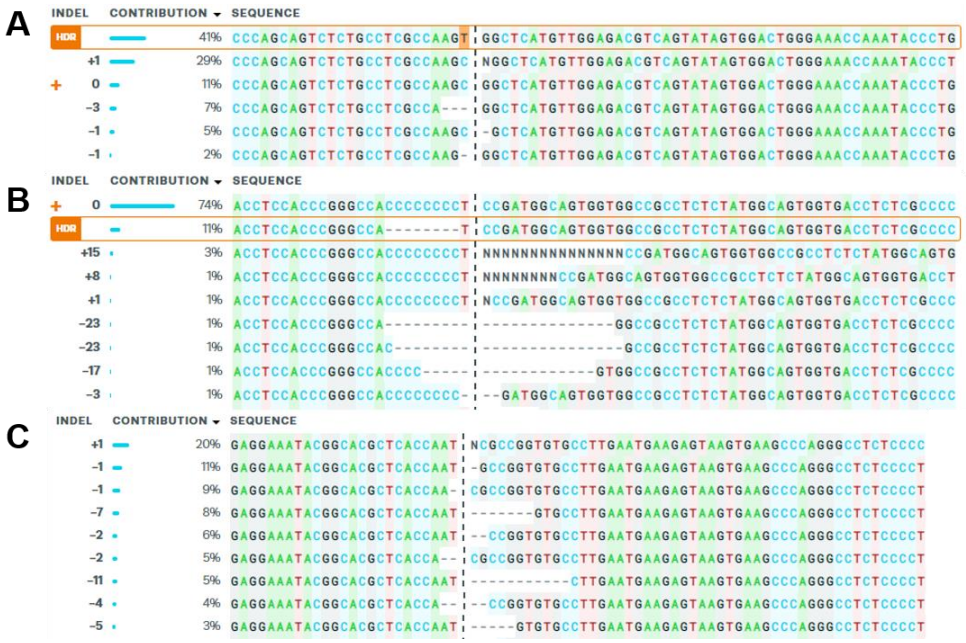


Figure 4.19 Indel contribution as estimated by ICE analysis. CD34⁺ cells were transfected with RNPs targeting A) *DNMT3A* exon 23, B) *ASXL1* exon 13, and C) *TET2* exon 6. Mutations introduced by homology directed repair (HDR) are highlighted in orange. The wild type sequence is indicated by a '+' on the left side. The size of each indel (+ for insertion, - for deletion) and the contribution are also shown on the left side.

In order to confirm the knockout of the target genes, mRNA expression was quantified for two independent biological samples. RNA was extracted 3, 7, or 14 days after transfection. Three days after transfection only for *TET2*^{mut} cells a decrease in target mRNA expression was detectable (Figure 4.20 A), indicating that the genetic knockout did not lead to an immediate reduction of *DNMT3A* and *ASXL1* mRNA expression. This observation could be confirmed in a second biological sample, in which only the *TET2* targeted sample exhibited reduced target mRNA 7 days after transfection (Figure 4.20 B). Nevertheless, after 14 days a reduction in mRNA of all targets was observed.

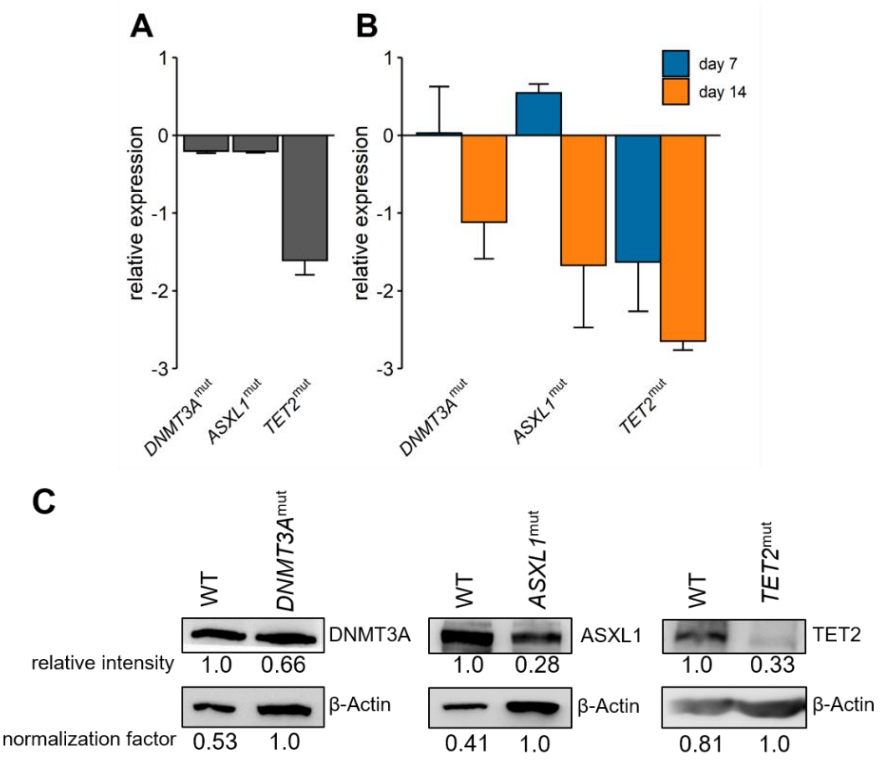


Figure 4.20 Target knockout in CD34^{+} cells. A + B) Quantitative real-time PCR analysis of *ASXL1*, *DNMT3A*, and *TET2* mRNA after CRISPR knockout of the respective targets. Relative expression was calculated via the -$\Delta\Delta C_T$-method, with the sample transfected only with Cas9 protein as reference sample (relative expression = 0). *GAPDH* was used as endogenous control. A) CB004 3 days after transfection. B) CB006 7 and 14 days after transfection. Bars and error bars show mean and SD of three technical replicates. C) Western Blot analysis of sample CB004 6 days (TET2) and 24 days after transfection (ASXL1 and DNMT3A). β-Actin was used as loading control. The normalization factor was calculated relative to the highest signal of β-Actin and used to normalize the signal intensities. WT: Cells transfected with Cas9 protein.

After confirming that the genetic knockout led to a decrease in target mRNA, I was interested in protein expression. Protein lysates were prepared 6 days (*TET2*^{mut}) and 4 weeks (*DNMT3A*^{mut} and *ASXL1*^{mut}) after transfection. TET2 protein was strongly decreased in the mutated cells (Figure 4.20 C, relative intensity vs. ctrl: 0.33). After normalization to the housekeeper β-actin, a decrease in target protein could also be confirmed in *ASXL1* and *DNMT3A* mutated cells (relative intensity vs. ctrl: 0.28 and 0.66, respectively). Western Blots of the target proteins were only performed once for one biological sample to confirm the observations on RNA level.

Since DNMT3A and TET2 play a crucial role in methylation and hydroxymethylation of DNA, a functional knockout might lead to an altered global methylation status of the DNA. The amount of 5-hydroxymethycytosine (5hmC) and 5-methylcytosine (5mC) was quantified via dot blot. Of three independent biological samples DNA was extracted at different time points and spotted onto a nylon membrane. The signals were analyzed using ImageJ software and normalized to the methylene blue loading control for each time point and DNA input. Relative levels compared to the wild type sample for all quantified spots of three biological samples are shown in Figure 4.21 A and B.

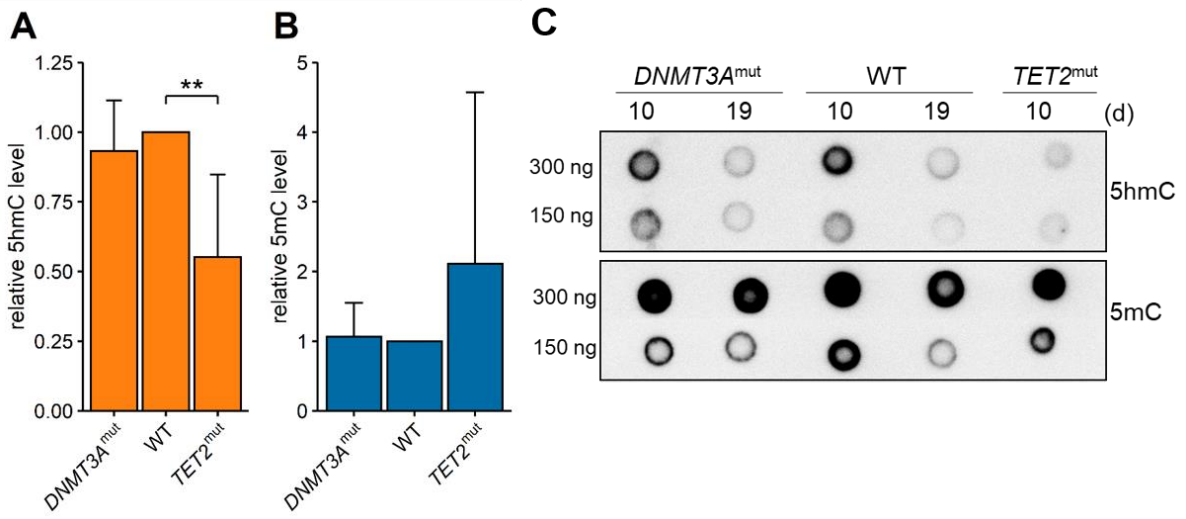


Figure 4.21 Dot Blot analysis of 5-hydroxymethylcytosine and 5-methylcytosine levels. A) Relative 5hmC levels of *DNMT3A*^{mut} and *TET2*^{mut} cells compared to wild type cells (Cas9 protein only). Signal intensities were quantified via ImageJ, normalized to the methylene blue loading control, and calculated relative to the wild type sample for each time point and input amount of three biological samples. ** $p<0.01$, one-sample Wilcoxon signed rank exact test with $\mu=1$. B) Relative 5mC levels of *DNMT3A*^{mut} and *TET2*^{mut} cells compared to wild type cells (Cas9 protein only). The bars show mean+SD of the relative signal intensity (3 biological samples with 3 or 4 technical replicates). C) One representative example of 5hmC and 5mC dot blot. Different time points are indicated above the blot (d = days after transfection), and different input amounts are shown on the left.

As anticipated, I observed a decrease of 5-hmC levels in *TET2*-mutated cells compared to wild type cells ($p<0.01$, one sample Wilcoxon signed rank exact test with $\mu=1$, $n=9$). The 5-methylcytosine levels vary greatly between the different replicates of *TET2*-mutated cells and did not follow a distinct pattern. *DNMT3A* mutations did not appear to have an effect on the 5mC or 5hmC levels in the cells. I observed that the global 5hmC levels decreased over time, whereas the 5mC level was less effected by time (Figure 4.21 C and Supplement: Figure 8.6).

4.4.2 Baseline Cell Characteristics

For all following experiments a total of 10 biological samples was collected (Supplement: Figure 8.2) and the experimental procedures were carried out in the same manner (see Figure 4.17 in chapter 4.4.1), to avoid any confounding variables. To investigate whether the introduction of mutations in the model system led to a functional difference in the cells, I first targeted the cells with all three guide RNAs simultaneously (multi KO). In the second part, I targeted each gene individually and compared the obtained results to the multi KO model to investigate how the aberrations act individually on the cells. Not all experiments were carried out in all biological samples and some data could not be analyzed due to culture contamination issues. After transfection, viable cells were counted via trypan blue exclusion 3, 5, and 7 days later to estimate the effect of the transfection and the introduction of CRISPR mutations on cell viability. The viability of the cells was between 30% at day 3 and up to 80% at days 5 and 7 after transfection and differed depending on the target (Figure 4.22). Especially in the multi KO samples the viability was reduced in the first days after transfection. One week after transfection the cells regenerated and showed a similar high viability of around 70% for all conditions.

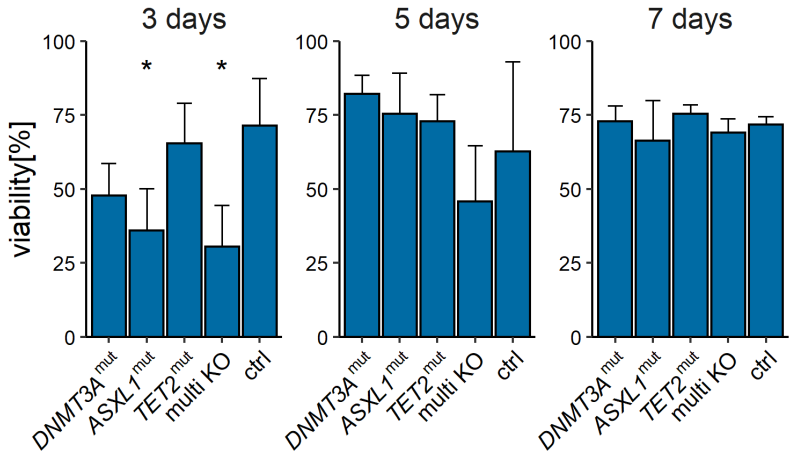


Figure 4.22 Viability of transfected cord blood CD34$^+$ cells. A) Viability of the cells was determined via trypan blue exclusion 3, 5, and 7 days after transfection with RNPs for three biological and two transfection replicates. The bars show mean+SD. Ctrl: cells transfected with Cas9 protein, multi KO: cells transfected with a mix of all three guide RNAs+Cas9 protein. * $p<0.05$ Wilcoxon rank-sum test compared to control group, corrected for multiple testing.

To assess whether the mutations have an impact on the baseline cycling and proliferative capacities of the cells, they were subjected to cell cycle analysis 5 days after transfection. For the bulk cells targeted with all three guide RNAs, I observed a different distribution of cells in the cell cycle phases in a preliminary experiment, with cells accumulating in G2 and S phase rather than G1 phase (Figure 4.23 A). In the cells targeted with the guide RNAs individually, no difference to the control group (WT) could be observed. This might be an effect of increased stress and DNA damage due to the cutting of Cas9 nuclease in multiple genes, leading to an

arrest in G2 phase. It has to be noted that cell cycle analysis of the multi KO cells was only performed with one biological replicate, therefore statistical testing was omitted.

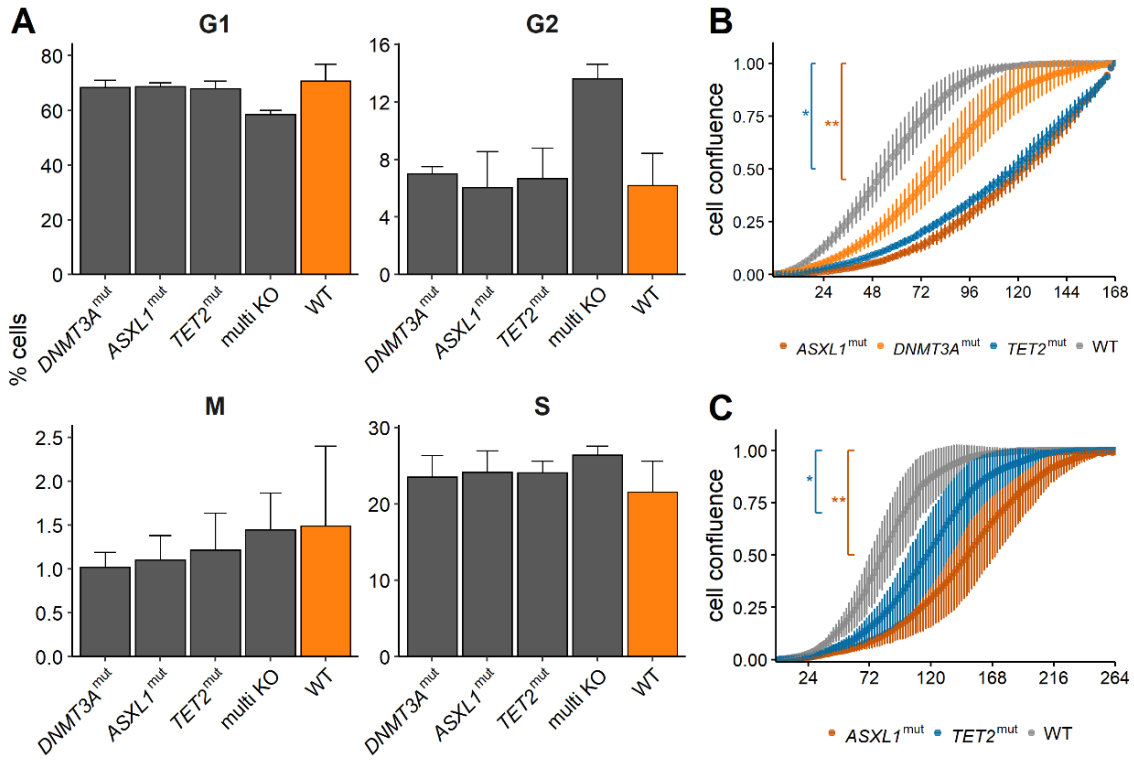


Figure 4.23 Cell cycle and proliferation analysis of transfected cord blood CD34⁺ cells. A) Cell cycle analysis was carried out in three biological and two technical replicates 5 days after transfection. The proportion of G1-, G2-, and S-phase was determined by flow cytometry of propidium iodide treated cells. Cells in mitosis were determined via the flow cytometry analysis of phosphorylated histone H3. Bars show mean and standard deviation. The wild type group is highlighted in orange (WT = transfection with Cas9 protein). No significant differences were detected using Wilcoxon rank-sum test and correcting for multiple testing. B+C) Cell proliferation analysis via cell confluence measurement in the IncuCyte live cell imaging system of two biological replicates. Each experiment was carried out with 4 technical replicates per condition. The points and error bars show mean±SD. *p<0.05, **<0.01 statistical testing was performed using the R package "statmod".

I did not observe any change in mitotic cells, indicating that the mutations did not affect the proliferative capacities of the cell bulk. I could confirm this observation by growth curve estimation via IncuCyte. Two to four days after transfection the cells were seeded in 96-well plates and the confluence was monitored via IncuCyte Live Cell Imaging System (Figure 4.23 B, C). This assay is strongly dependent on the seeding density of the cells and is not optimal when all conditions have to be seeded individually. Wild type cells grew fastest in both proliferation experiments, hinting to a toxic effect of the transfection procedure in the first days of culture.

4.4.3 Differentiation and Phenotypic Changes

To assess whether the introduced mutations have an influence on the differentiation of the cells, expression of differentiation specific cell surface markers was monitored via flow cytometry. Marker expression of all samples was checked before transfection (day 0, Supplement: Figure 8.2) and again every 7 days for three weeks. In general, and as expected, very few to no cells expressing lymphocytic markers could be found in the liquid culture (CD3 or CD19). The frequency of $CD34^+$ cells remained high after 7 days of liquid culture but was gradually lost over time (Figure 4.24). CD66b expression (neutrophil granulocytes) was very low in the beginning of the differentiation experiment but increased after 14 days. CD14 (monocytes and macrophages) expression increased mainly between 14 and 21 days of the experiment, but remained relatively low, compared to CD66b. These general patterns could be observed for all biological samples, whereas the expression of other markers varied more between the different donors (i.e., erythroid marker CD235a).

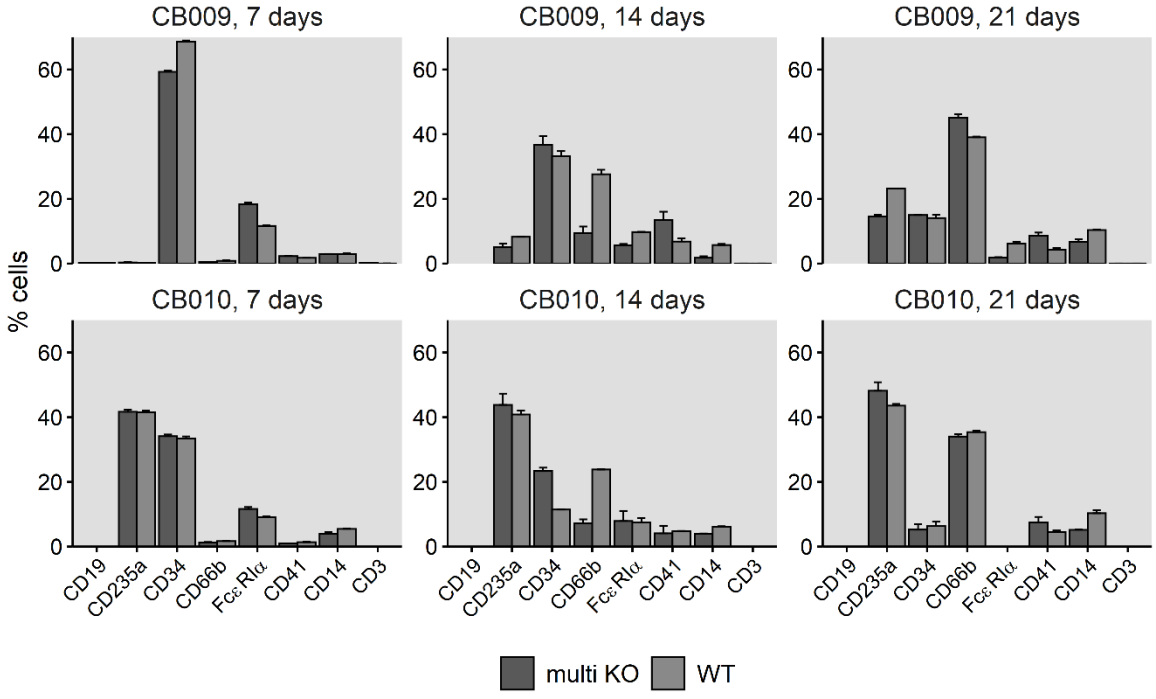


Figure 4.24 Cell surface marker analysis of multi KO $CD34^+$ cells. Transfected $CD34^+$ cells were analyzed for cell surface marker expression via flow cytometry 7, 14, and 21 days after transfection: CD19 B cells, CD235a erythroid cells, CD34 progenitor cells, CD66b granulocytes, FcεRIα mast cells and basophiles, CD41 megakaryocytes, CD14 monocytes, CD3 T cells. Frequencies were normalized to 100% for all depicted cell populations. Two biological samples are shown. Bars and error bars show mean and standard deviation of two transfection replicates. No statistical testing was performed for the individual biological samples with two technical replicates.

After 21 days, the frequency of CD66b expressing multi KO cells reached a level similar to WT cells, indicating that the expression of this neutrophil granulocyte marker is delayed in multi KO cells. A similar pattern was observed for CD14, with fewer multi KO cells expressing CD14

after 14 days as compared to WT cells. A trend that was consistent even after 21 days in liquid culture. Introducing knockouts in bulk cells enriched for hematopoietic progenitors seemed to affect differentiation marker expression *in vitro*. To dissect which mutation led to the observed patterns, I investigated the targets individually.

For *DNMT3A*^{mut} and *AXSL1*^{mut} 4 biological samples were analyzed. I did not observe a different marker expression pattern than in the WT group (Figure 4.25 A, two examples). Although the actual percentage of CD34-expressing cells did not differ between WT and mutated cells for several biological samples, the CD34 marker expression decreased slower in *DNMT3A*^{mut} than in the wild type cells (Figure 4.25 A, CB009). Fold change calculations of CD34 marker expression from day 7 to day 14 revealed that *DNMT3A*^{mut} cells retained more of their CD34 expression over this time period (Figure 4.25 B, Wilcoxon rank-sum test $p<0.05$), but this effect was lost after 21 days.

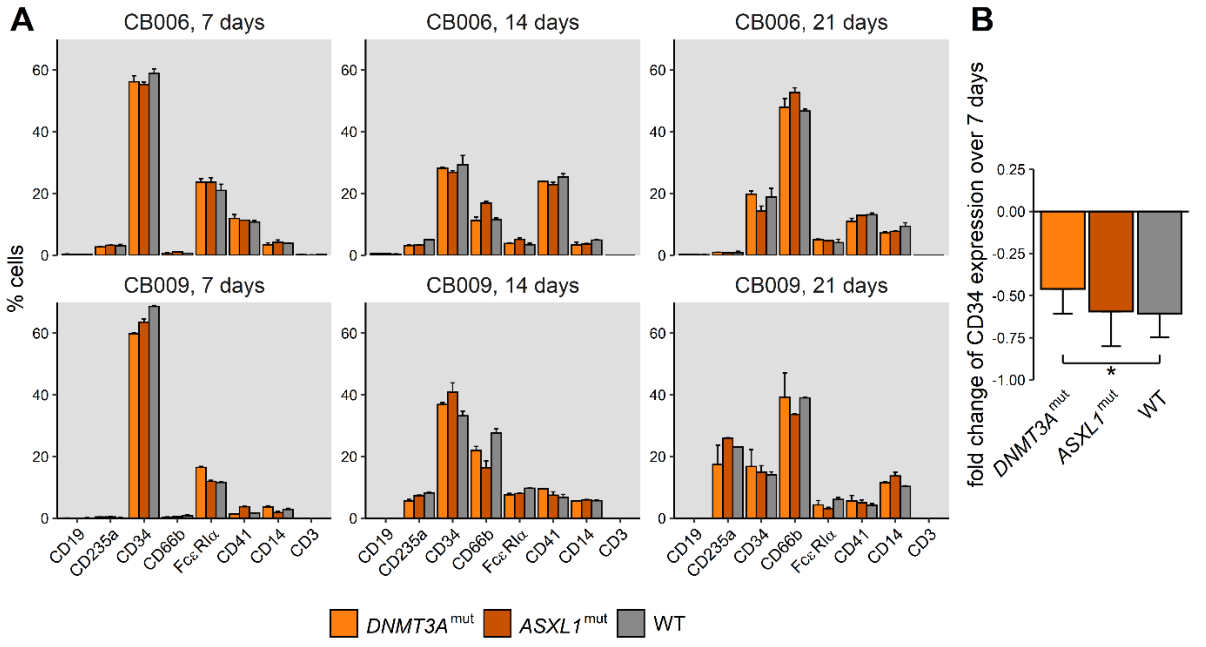


Figure 4.25 Cell surface marker analysis of *ASXL1*^{mut} and *DNMT3A*^{mut} CD34^{+} cells. A) Transfected CD34^{+} cells were analyzed for cell surface marker expression via flow cytometry 7, 14, and 21 days after transfection: CD19 B cells, CD235a erythroid cells, CD34 progenitor cells, CD66b granulocytes, FcεRIα mast cells and basophiles, CD41 megakaryocytes, CD14 monocytes, CD3 T cells. Frequencies were normalized to 100% for all depicted cell populations. Two biological samples are shown. Bars and error bars show mean+SD of two transfection replicates. No statistical testing was performed for the individual biological samples with two technical replicates. B) Loss of CD34 marker expression was calculated as fold change between day 7 and day 14. The bars show mean-SD. * $p<0.05$, Wilcoxon rank-sum test.

Investigating *TET2*^{mut} cells, I detected a similar pattern as for the multi KO cells. First, 7 days after transfection very few cells expressed CD66b (Figure 4.26, left panels). Second, after 14 days in liquid culture, the proportion of CD66b expressing cells was higher in the wild type group than in *TET2*^{mut} cells (Figure 4.26, middle panels), an effect that was lost after 21 days in liquid culture (Figure 4.26, right panels). CD14 expression seemed to differ already after 7 days and remained this way even after 21 days of culture. For the two examples shown in Figure 4.26, no individual statistical testing was performed, so only trends were reported.

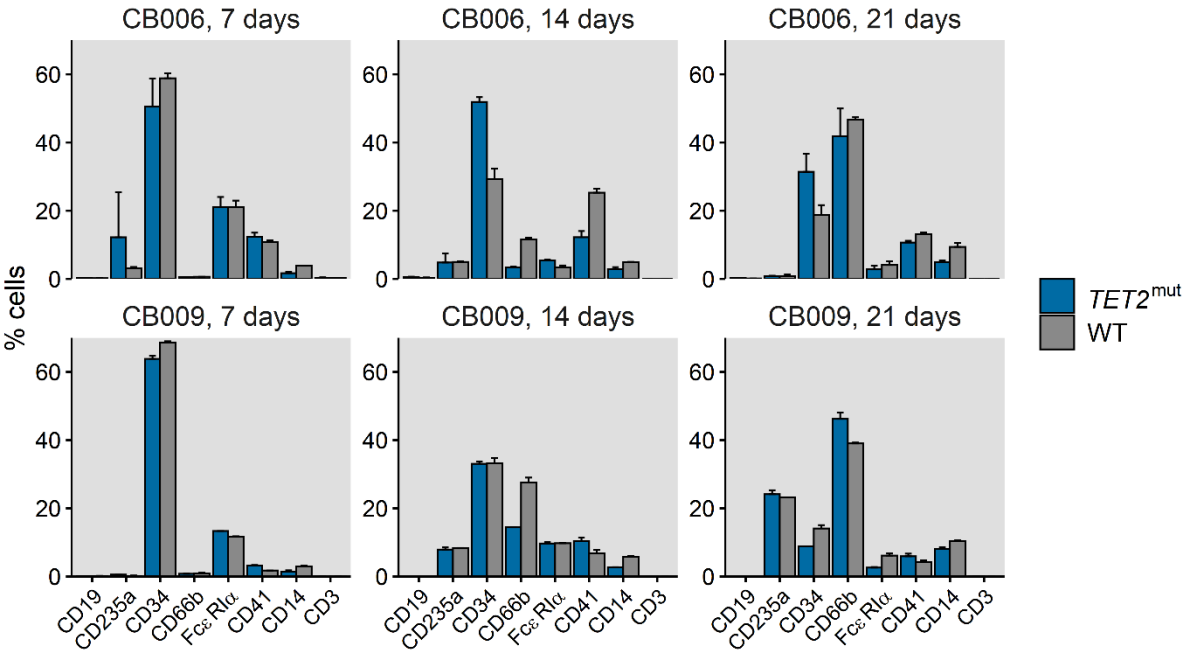


Figure 4.26 Cell surface marker analysis via flow cytometry of *TET2*^{mut} CD34^{+} cells. *TET2*^{mut} CD34^{+} cells were analyzed for cell surface marker expression via flow cytometry 7, 14, and 21 days after transfection: CD19 B cells, CD235a erythroid cells, CD34 progenitor cells, CD66b granulocytes, FcεRIα mast cells and basophiles, CD41 megakaryocytes, CD14 monocytes, CD3 T cells. Frequencies were normalized to 100% for all depicted cell populations. Two biological samples are shown. Bars and error bars show mean+SD of two transfection replicates. No statistical testing was performed for the individual biological samples with two technical replicates.

In total 7 biological replicates were measured for *TET2*^{mut} to confirm this observation. For all replicates CD14 expression was decreased in *TET2*^{mut} cells compared to the WT (Figure 4.27 A, Wilcoxon rank-sum test $p<0.05$ for all time points), with the highest discrepancy at day 14, indicating that CD14 might be expressed delayed and the cells might reach the same level as the control group at a certain time. As described before, CD66b marker expression was generally very low after 7 days in liquid culture and remained lower in *TET2*^{mut} cells at 14 days (Figure 4.27 B). *TET2*^{mut} cells increased CD66b marker expression after 14 days and reached the same level as the control group after 3 weeks of culture. Comparing these results with the multi KO cells, a similar pattern as described for only *TET2*^{mut} cells can be observed (Figure 4.27 A and B). Since no distinct patterns were found in the *DNMT3A*^{mut} or *ASXL1*^{mut} cells, it seemed that the effects observed in the multi KO resulted from the aberrations in *TET2*. CD34

marker expression was retained longer in multi KO cells as determined by fold change calculation (Figure 4.27 C), as observed for the *DNMT3A*^{mut} cells. This means the effects of *TET2* mutations as well *DNMT3A* mutations could be observed in the multi KO sample.

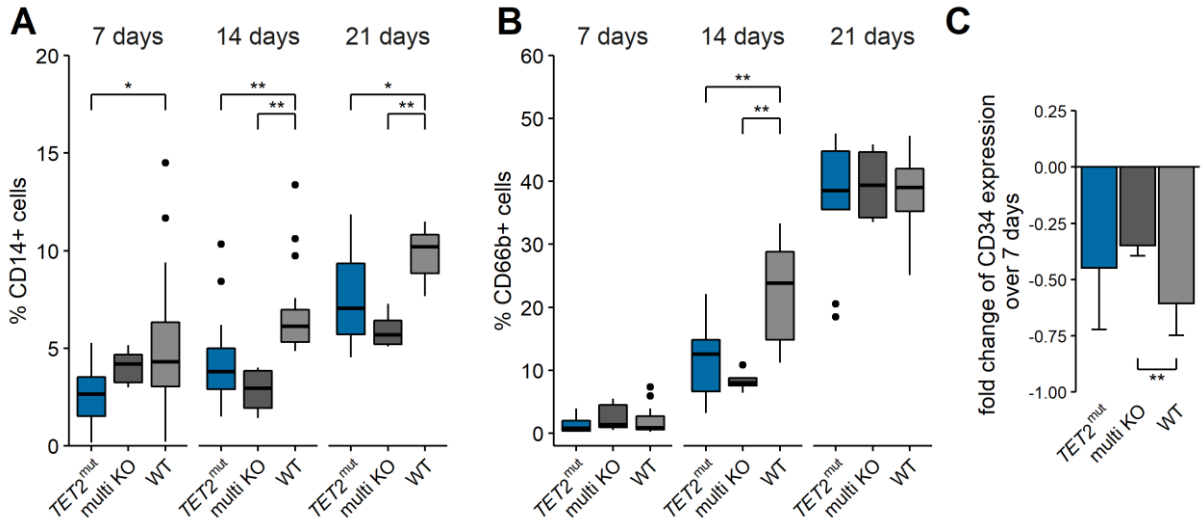


Figure 4.27 CD14 and CD66b analysis via flow cytometry of *TET2*^{mut} and multi KO CD34^{+} cells. A) CD14^{+} cell fraction in *TET2*^{mut}, multi KO, and wild type cells 7, 14, and 21 days after transfection for all analyzed biological and technical replicates (n=14 for *TET2*^{mut} and WT, n=4 for multi KO). * p<0.05, **p<0.01, Wilcoxon rank-sum test corrected for multiple testing. B) CD66b^{+} cell fraction in *TET2*^{mut}, multi KO, and wild type cells 7, 14, and 21 days after transfection for all analyzed biological and technical replicates (n=14 for *TET2*^{mut} and WT, n=4 for multi KO). *p<0.05, **p<0.01, Wilcoxon rank-sum test corrected for multiple testing. C) Loss of CD34 marker expression was calculated as fold change between day 7 and day 14. ** p<0.01, Wilcoxon rank-sum test.

Altogether a delayed differentiation marker expression could be observed in *TET2*^{mut} cells as well as an increased retention of the primitive marker CD34 in *DNMT3A*^{mut} cells, indicating that a disruption in these genes might lead to impaired differentiation and therefore altering the balance of the tightly regulated hematopoietic system.

4.4.4 Self-renewal and Long-term Culture Initiating Potential

To get more details about the differentiation potential of the mutated cells, colony forming assays were performed. Colony forming assays are used to identify committed progenitors (colony forming cells, CFC) in a cell sample by stimulating them to develop and produce mature colonies. It is possible to detect progenitors with multi-lineage differentiation potential or more committed mature progenitors, but not primitive progenitors or HSCs. Colonies that arise from different progenitors can be characterized based on morphology (Figure 4.28). Every quantified colony therefore stands for a distinct progenitor present in the sample [175, 176]. CFU assays were carried out for at least three biological samples per transfection set-up. Per condition, 1,000 cells were seeded in duplicates in methylcellulose-based medium. Colonies were quantified after 10 to 14 days of culture, and classified as colony forming unit granulocyte, macrophage (CFU-GM), colony forming unit granulocyte, erythrocyte, macrophage, megakaryocyte (CFU-GEMM), colony forming unit erythroid (CFU-E), or burst forming unit erythroid (BFU-E). CFU-E and BFU-E both identify erythroid progenitors, with the difference that BFU-E are more immature and produce more cells per cluster. CFU-GM produce granulocytes, macrophages, or both. A sub-classification can be performed (CFU-G and CFU-M) [175], but was not done for all samples and the CFUs are therefore summed up as CFU-GM in the experiments performed here.

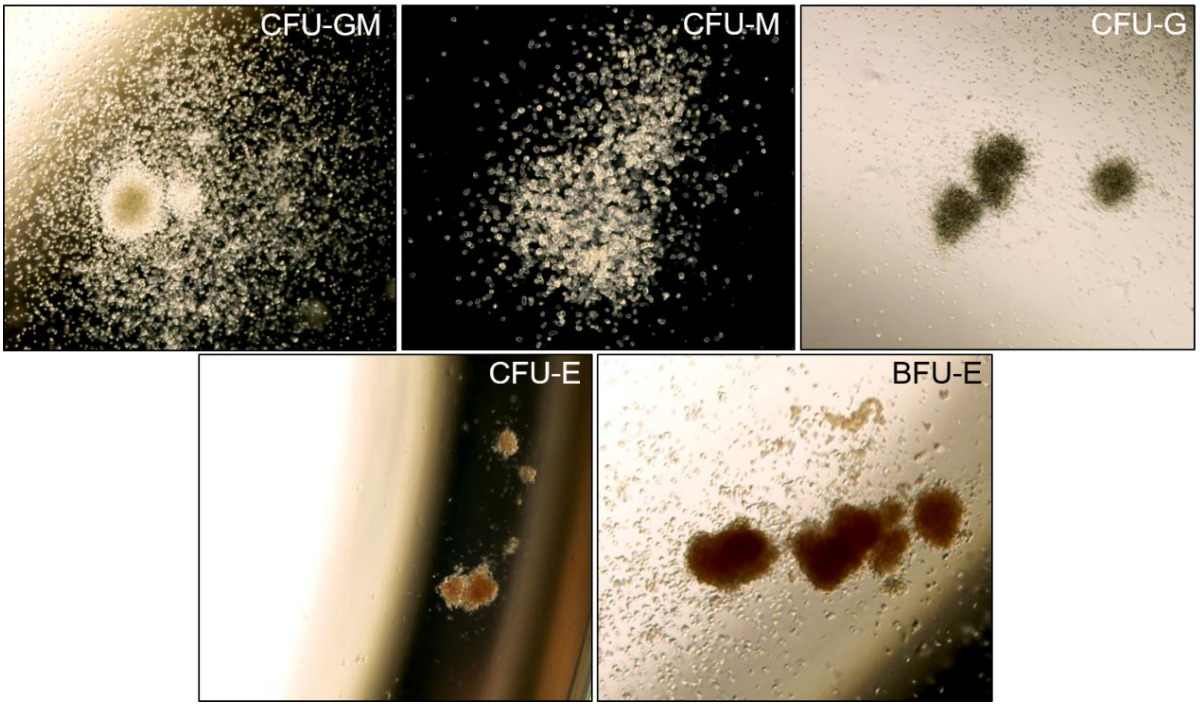


Figure 4.28 Examples of colony forming unit types. CFU-GM: colony forming unit granulocyte, macrophage, CFU-M: colony forming unit macrophage, CFU-G: colony forming unit granulocyte, CFU-E colony forming unit erythroid, BFU-E: burst forming unit erythroid.

After the first CFU quantification, the cells were harvested, washed, and resuspended to obtain a single cell suspension. After determination of viable cell count, cells were seeded a second time in methylcellulose and cultured for 10 to 14 days. This procedure was repeated until no CFUs were detected. In the first passage mainly CFU-GM were quantified, but also BFU-E, and in lower numbers CFU-E and CFU-GEMM. Since the overall number of colonies varied widely between the biological replicates (range: 36±20 to 194±83 colonies, mean±standard deviation), I compared the relative distribution of the different progenitors present in the samples (Table 4.4). For all samples, the majority of CFU counted were granulocyte and macrophage committed, with the multi target knockout having the highest proportion of CFU-GM and the wild type cells having the lowest. Simultaneously, in the control group the highest proportion of BFU-E, CFU-E, and multi lineage progenitors could be seen. Nevertheless, the standard deviation was very high for all targets, so no further conclusions can be drawn (no significant differences using Wilcoxon rank-sum test and multiple testing correction).

Table 4.4 Relative frequency of different CFU. Percentage of each CFU type calculated for duplicates of at least three biological samples. All values are mean ± standard deviation. No significant differences were found using Wilcoxon rank-sum test and multiple testing correction.

	CFU-GM	**BFU-E**	**CFU-E**	**CFU-GEMM**
DNMT3A^{mut}	73.8 ± 12%	18.9 ± 8.4%	0.5 ± 1.6%	6.8 ± 8.6%
ASXL1^{mut}	69.3 ± 20.6%	24.7 ± 18.4%	1.3 ± 4.0%	4.7 ± 5.2%
TET2^{mut}	73.6 ± 14.8%	19.2 ± 8.9%	3.8 ± 5.6%	3.4 ± 4.0%
multi KO	78.4 ± 13.8%	12.9 ± 11.5%	1.8 ± 2.9%	6.7 ± 3.0%
WT	60.3 ± 12.8%	25.9 ± 11.1%	5.4 ± 6.7%	8.4 ± 7.6%

With regard to overall colony numbers per 1,000 seeded cells, CFU count was higher in the WT compared to multi KO in two out of three biological replicates in the first passage (Figure 4.29, CB010 and CB011). For all biological samples, more CFUs were counted after the first replating, for CB009 it was even possible to detect CFUs after another round of replating.

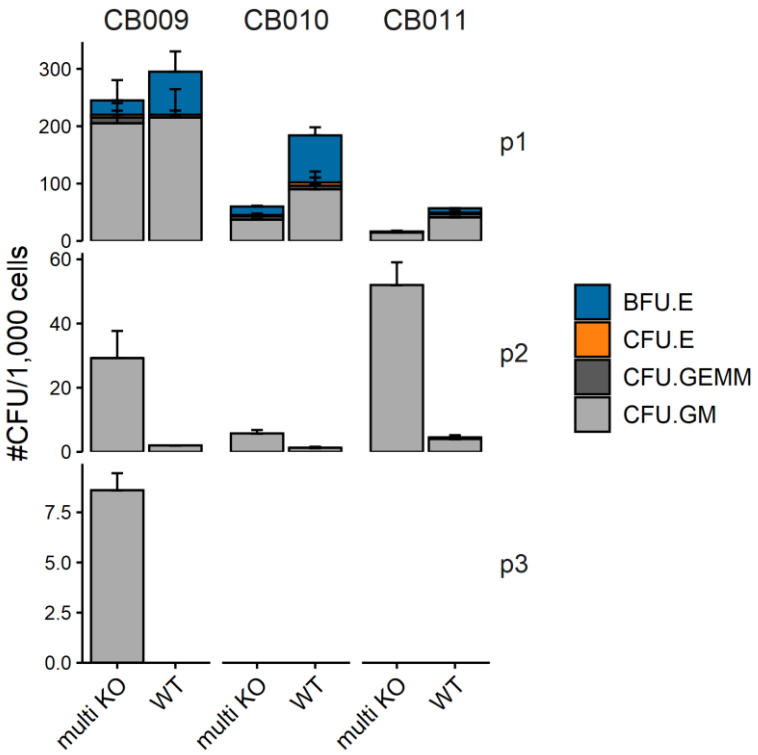


Figure 4.29 Colony forming unit and serial replating assays of multi KO CD34$^+$ cells. Transfected cells were directly seeded in methylcellulose (MethoCult H4435) and cultured for 10 to 14 days. After determining colony forming units with distinct phenotypes, cells were harvested and replated in methylcellulose. Colonies were determined as BFU-E: burst forming unit erythroid, CFU-E: colony forming unit erythroid, CFU-GM: colony forming unit granulocyte, macrophage, CFU-GEMM: colony forming unit granulocyte, erythrocyte, macrophage, megakaryocyte per 1,000 plated cells. Biological replicates are indicated above the bars. The number of the passage is shown on the right (p1: first seeding, p2: second seeding, p3: third seeding). All bars show mean and standard deviation of two technical replicates.

Similar to the multi KO cells, no difference between *AXSL1*^{mut}, *DNMT3A*^{mut}, and the WT group was observed in the first passage, except for one sample in which more colonies were counted in the wild type sample (Figure 4.30, CB009). In one biological replicate, I was able to find colonies in the third passage for the *ASXL1* and *DNMT3A* mutants (sample CB005), but not in the wild type sample. Since the colony forming ability was highly variable between donors, three biological replicates are not enough to find patterns in the behavior of the cells. Unfortunately, several replicates were lost due to contamination issues.

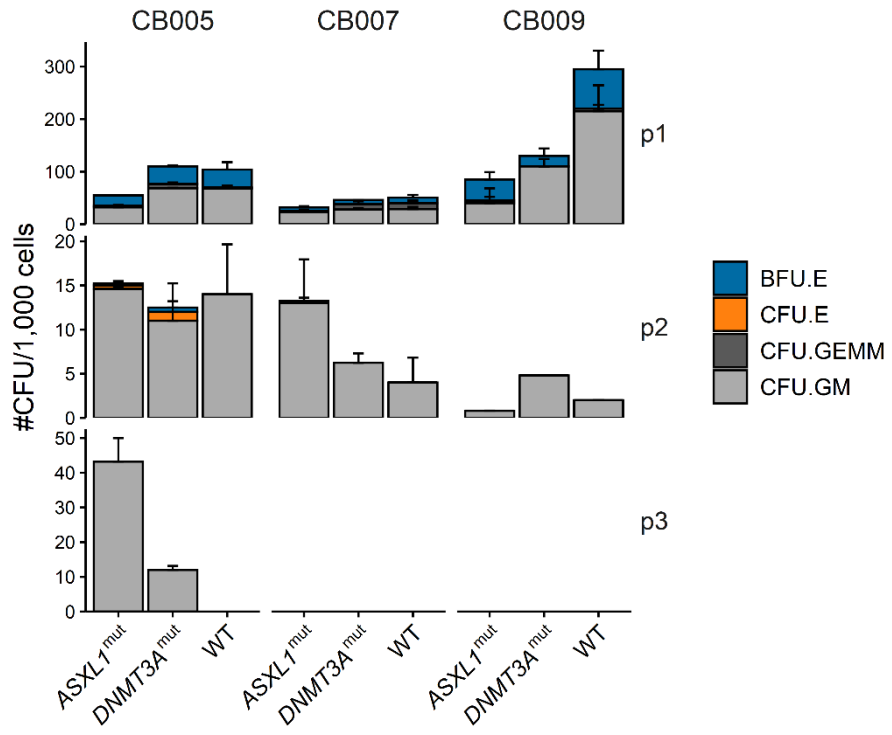


Figure 4.30 Colony forming unit and serial replating assays of *ASXL1*^{mut} and *DNMT3A*^{mut} CD34^{+} cells. Transfected cells were directly seeded in methylcellulose (MethoCult H4435) and cultured for 10 to 14 days. After determining colony forming units with distinct phenotypes, cells were harvested and replated in methylcellulose. Colonies were determined as BFU-E: burst forming unit erythroid, CFU-E: colony forming unit erythroid, CFU-GM: colony forming unit granulocyte, macrophage, CFU-GEMM: colony forming unit granulocyte, erythroid, macrophage, megakaryocyte per 1,000 plated cells. Biological replicates are indicated above the bars. The number of the passage is shown on the right (p1: first seeding, p2: second seeding, p3: third seeding). All bars show mean and standard deviation of two technical replicates.

TET2^{mut} cells were examined in 7 biological replicates (Figure 4.31 A). In the first passage there was no significant difference in CFU numbers between *TET2*^{mut} and WT samples (Figure 4.31 B). In the second passage, significantly more CFU were forund in the *TET2*^{mut} cell bulk than in the WT cells (adjusted p=0.001, Wilcoxon rank-sum test). For some samples colonies emerged in the third passage, indicating that even after replating, progenitors were present in the sample leading to the formation of mature cell colonies.

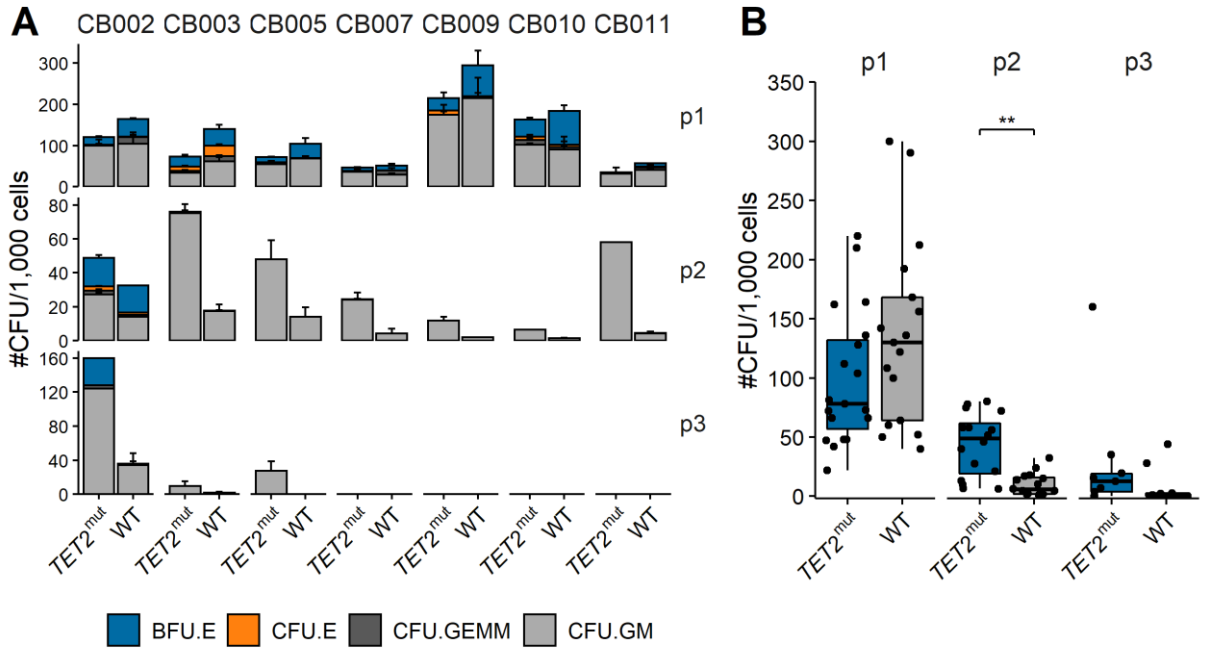


Figure 4.31 Colony forming unit and serial replating assays of *TET2*^{mut} CD34^{+} cells Transfected cells were directly seeded in methylcellulose (MethoCult H4435) and cultured for 10 to 14 days. After determining colony forming units with distinct phenotypes, cells were harvested and replated in methylcellulose. A) Bar plots showing the different biological samples. Colonies were determined as BFU-E: burst forming unit erythroid, CFU-E: colony forming unit erythroid, CFU-GM: colony forming unit granulocyte, macrophage, CFU-GEMM: colony forming unit granulocyte, erythroid, macrophage, megakaryocyte per 1,000 plated cells. Biological replicates are indicated above the bars. The number of the passage is shown on the right (p1: first seeding, p2: second seeding, p3: third seeding). All bars show mean and standard deviation of two technical replicates. B) Boxplot showing the total CFU count of all analyzed replicates for *TET2*^{mut} vs. WT cells for statistical analysis. The passage is indicated above the plot. ** p<0.01, Wilcoxon rank-sum test.

To investigate if primitive progenitors were retained in the samples, long-term culture-initiating cell (LTC-IC) assays were performed. Cells were cultured on a feeder layer that supports attachment of more primitive progenitor cells which in turn generate committed progenitors and sustain myelopoiesis [146]. Transfected cells were plated on irradiated feeder cells producing human cytokines G-CSF, SCF, and IL-3 (M2-10B4 and Sl/Sl mix) at a concentration of 500 to 1,000 cells per 500 µl HLTM. After 4 to 9 weeks the cells were harvested and viable cells were counted via trypan blue exclusion. The total number of viable cells was normalized to the number of plated cells per experiment. Viable cell numbers in *DNMT3A*^{mut} and *ASXL1*^{mut} samples differed strongly between donors, nevertheless I observed an increase in cell numbers compared to the WT sample. In *TET2*^{mut} and multi KO samples the variability was less strong and therefore showed a strong increase in cell numbers compared to the wild type samples (Figure 4.32). This first observation indicated that a number of primitive progenitors and/or committed progenitors supporting continuous production of mature myeloid cells was present in the samples.

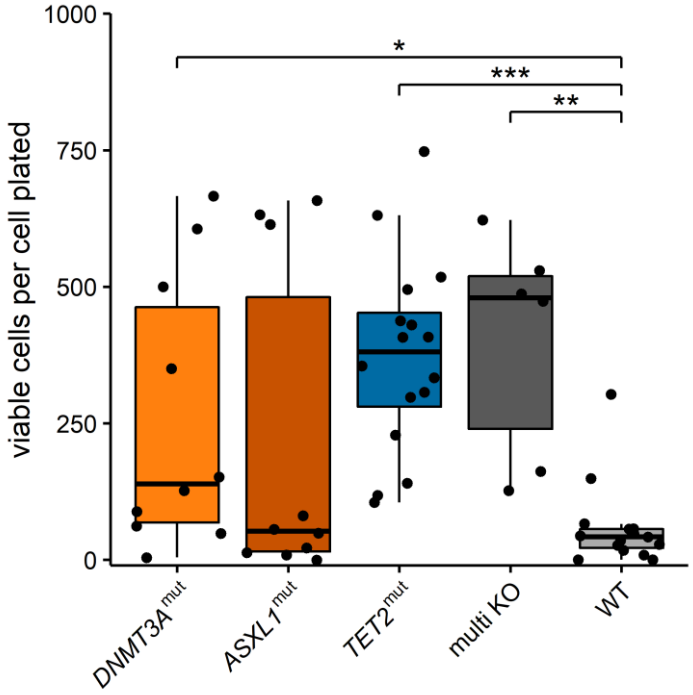


Figure 4.32 Viable cells after long-term culture. Transfected cells were seeded onto irradiated feeder cells and cultured for up to 9 weeks. Viable cells were determined via trypan blue exclusion after harvesting the complete well. The number of viable cells determined was divided by the number of cells plated. The experiment was performed in triplicates of at least two biological samples per condition. Statistical testing was performed using Wilcoxon rank-sum test of each group vs. control group (WT = wild type). P-values were adjusted for multiple testing. *p<0.05, **p<0.01, ***p<0.001

To assess the phenotype of the detected cells after long-term culture, cell surface marker analysis via flow cytometry was performed. Mainly CD14 (monocyte and macrophage) and CD66b (neutrophil granulocyte) marker expressing cells were detected after long-term culture (Figure 4.33 A, B, C, representative examples from two biological samples). In some samples the proportion of basophiles and mast cells (FcεRIα expression) was equally high (data not shown), but very few CD41 (megakaryocyte) positive cells were detected. This observation confirmed that myelopoiesis was maintained over 9 weeks of long-term culture. Interestingly, the frequency of CD34 expressing cells was higher in the mutants than in the wild type sample in the biological replicate depicted in Figure 4.33 A (*DNMT3A*^{mut}: 1.31±0.4%, *ASXL1*^{mut}: 2.71±1.77%, *TET2*^{mut}: 0.52±0.38%, WT: 0.10±0.12%). Only for this biological sample (with 3 technical replicates) the cytometry data had a sufficiently high quality to quantify CD34 expressing cells, so statistical testing was omitted for this analysis.

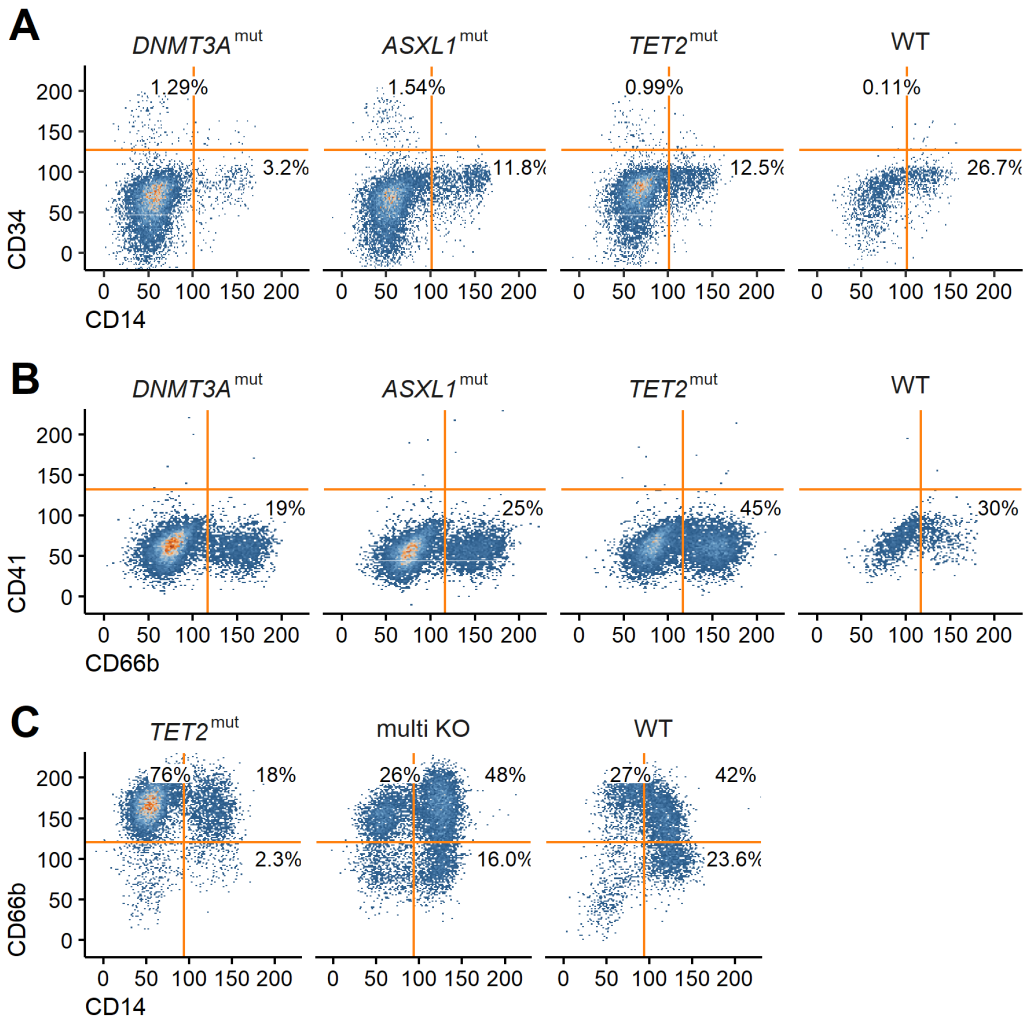


Figure 4.33 Exemplary flow cytometry analysis of surface marker expression after long-term culture. A) Flow cytometry analysis of CD34 and CD14 marker expression in cells from CB007 after long-term culture. B) Flow cytometry analysis of CD41 and CD66b in cells from CB007 after long-term culture. C) Flow cytometry analysis of CD66b and CD14 in cells from CB010 after long-term culture. CD14: monocyte/macrophage, CD34: progenitor cells, CD41: megakaryocytes, CD66b: granulocytes. One representative sample is shown for each condition.

Next, I performed a CFU assay after long-term culture to determine if colony forming cells were still present in the culture. Colony forming cells represent committed progenitors. Of the harvested cells, 20,000 were seeded in methylcellulose. For all mutated conditions LTC-IC derived CFU were found, whereas only in few of the WT samples colonies were detected (each condition vs WT $p<0.05$, Wilcoxon rank-sum test). The standard deviation for CFUs determined was high because of the variance between the biological samples. For visualization, the y-axis was square root transformed (Figure 4.34). In *DNMT3A*^{mut} a high number of LTC-IC derived CFU per 20,000 test cells were found (8.8±10.4), but only in half of the replicates colonies were detected (56%, n=9). *TET2*^{mut} cells did not produce as many colonies per seeded replicate (5.9±12.7), but a higher proportion of replicates were able to generate colonies (71%, n=17). In *ASXL1*^{mut} less LTC-IC derived CFU were detected per replicate (2.7±3.6) and in 56%

(n=9) of the seeded samples CFU were found. In the multiple target knockout samples, colonies were detected in 71% of the replicates (n=7), with a mean CFU number of 19.

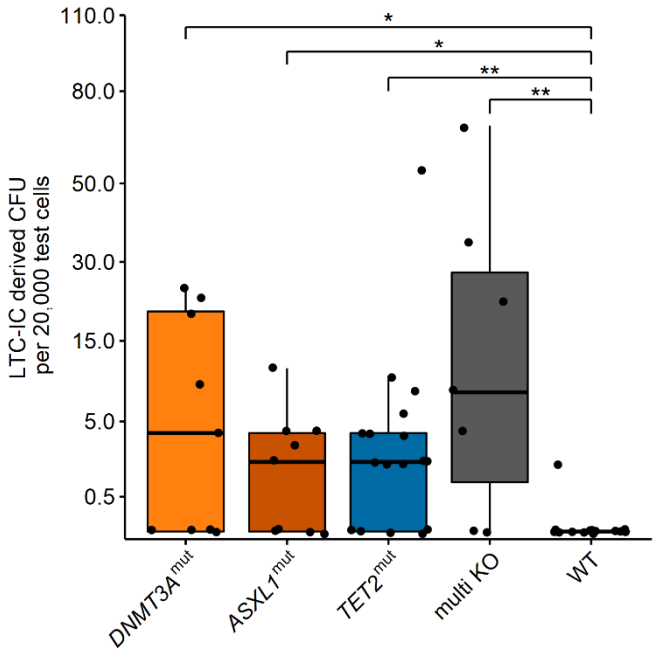


Figure 4.34 Long-term culture-initiating cell derived colony forming units. After long-term culture all cells were harvested, viable cells were determined via trypan blue exclusion, and 20,000 viable cells were seeded. For each condition duplicates of at least two biological replicates were seeded in methylcellulose and colony forming units (CFU) were quantified. The y-axis was square root transformed to make the data points more visible. Statistical testing was performed using Wilcoxon rank-sum test of each group vs. control group (WT = wild type). P-values were corrected for multiple testing. * $p<0.05$, ** $p<0.01$

4.5 Mutational Analysis and Clonal Expansion

4.5.1 Indel Detection via CRISPRseq

To determine the indel composition of the mutated cells, deep sequencing of an amplicon spanning the mutation target site was performed in single mutated bulk cells from various time points. The different culture condition support different cell output, therefore all samples from each condition were analyzed separately to determine differences in indel composition. CRISPRseq mutational analysis was performed, as described by Tothova et al. [151] with minor modifications in the library preparation and bioinformatics pipeline. This analysis allows the detection of large indels that are missed by sequence decomposition algorithms of Sanger sequencing files. The sequencing was performed in two sequencing runs with 140 and 221 individual amplicons for the three target regions. A mean sequencing depth of approx. 18,000 and 30,000 reads was achieved for run 1 and run 2, respectively. Since the sequencing depth was high, all insertions and deletions that were found less than 10 times were omitted from further analysis (list of all detected mutations: supplemental spreadsheet file). Indel sizes ranged from 1 to 86 bp (Figure 4.35 A), with more deletions than insertions (73% deletions). The majority of the deletions were smaller than 10 bp (71% of all deletions) and the most frequently observed insertion was 1 bp (84 % of all insertions).

Generally, more indels were detected per sample via CRISPRseq than with Sanger sequencing and sequence decomposition (Figure 4.35 B). The indel frequency ranged from 21% to 89% for the *ASXL1* exon 13 sgRNA, which was a much higher variation than for the *DNMT3A* or *TET2* targets (47-67% and 75-99%, respectively).

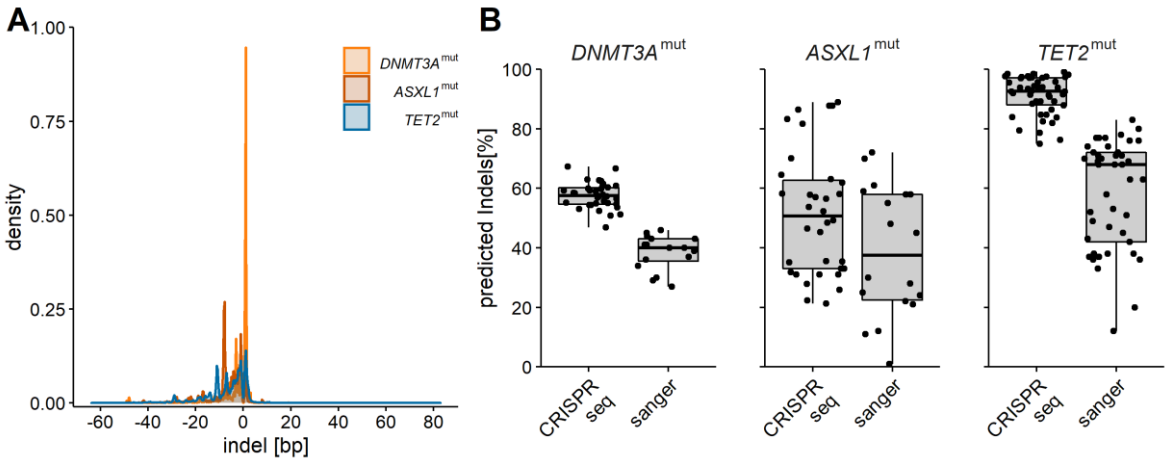


Figure 4.35 Indel prediction via CRISPRseq. A) Density distribution of insertion and deletion size in base pairs (bp) found with the CRISPRseq algorithm in deep sequencing reads of amplicons spanning the targeting region of each sgRNA. B) Predicted indel frequency for all samples from short-term culture experiments analyzed via CRISPRseq or Sanger sequence decomposition. Samples from different biological samples and time points are grouped together based on the target gene.

Between the different time points of short-term culture, I did not observe a difference in predicted indel frequency for all targets, indicating that the wild type cells were not overgrowing the mutant in short-term culture (Figure 4.36 A). Methylcellulose and long-term culture are selective culture conditions, supporting the expansion of distinct cell types and are carried out over longer times. Similar to short-term culture experiments, the indel frequencies remained constant for the majority of the samples over longer culture periods (Figure 4.36 B, Wilcoxon rank-sum test, adjusted for multiple testing). The point mutation in *DNMT3A* was found with a mean VAF of 32.7±11.2% and did not vary between the time points.

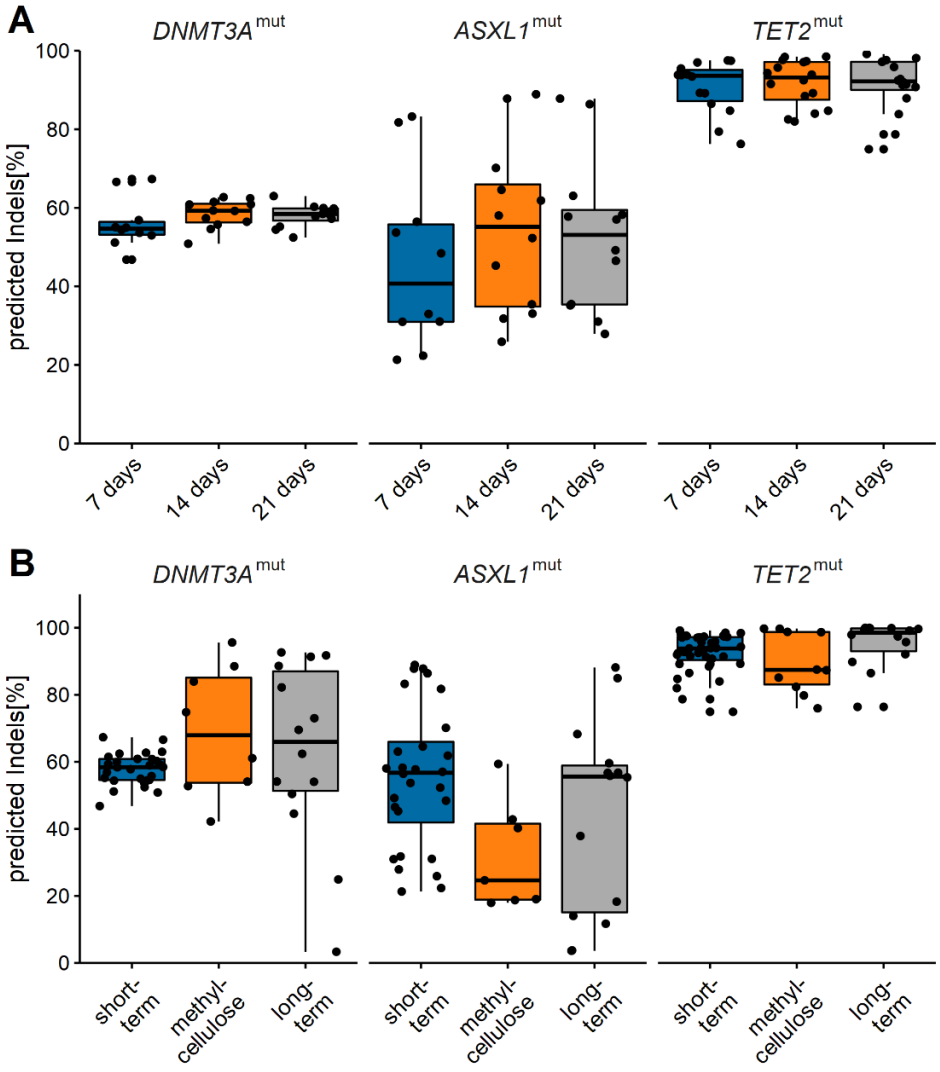


Figure 4.36 Predicted indel frequency for different time points and culture conditions. Indel frequencies were predicted via CRISPRseq. A) Samples from short-term culture experiments were compared: the different time points are shown on the x-axis. B) Indel frequencies were compared between different culture methods. Samples from different biological samples were grouped together based on the target gene and the time point at which the sample was taken. The colors of the box plots highlight the different time points. Statistical testing was performed using Wilcoxon rank-sum test. No significant differences were detected after correction for multiple testing.

4.5.2 Mutational Composition of CRISPR-targeted HSPCs

In the next step, I analyzed the indel distribution of samples targeted with one sgRNA from the various culture conditions. As described before, no difference in indel frequency was observed between the different time points of short-term culture experiments, which was also true for samples from methylcellulose and long-term cultures. The number of indels detected per sample did not differ between time points of the different culture conditions (Figure 4.37 A, B, C). Indel numbers per sample were significantly different between culture conditions and therefore time of overall culture (Figure 4.37 D). Numbers of indels decreased significantly from short-term culture experiments to methylcellulose cultures and was lowest in samples from long-term cultures.

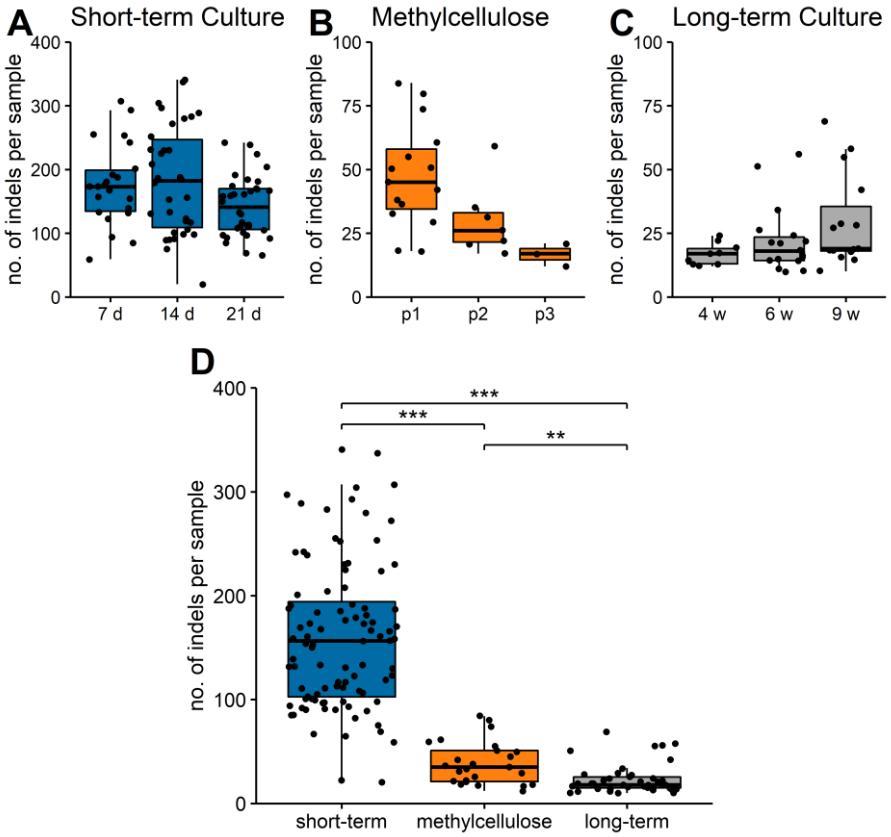


Figure 4.37 Number of indels per sample as determined via CRISPRseq. A) Samples were collected at three different time points of short-term culture: 7, 14, and 21 days after transfection. B) Samples were collected every 10-14 days when the cells were passaged. C) Long-term cultures were seeded in individual wells for every time point and replicate. The different time points are therefore independent. D) Samples are grouped based on culture condition, independent of biological replicate and time of sample collection. Statistical analysis was performed using Wilcoxon rank-sum test. The p-values were adjusted for multiple testing. *p<0.05, **p<0.01, ***p<0.001

The overall indel frequency did not change over time while the number of distinct indels decreased, leading to the hypothesis that a selection and/or expansion of distinct indels, and therefore distinct cell clones, took place over longer culture periods.

To test this hypothesis, Shannon entropy analysis was performed based on the insertions and deletions found in each sample after deep sequencing and CRISPRseq analysis. Entropy is a measure of diversity of a sample. All determined indels were assigned a unique code (e.g., D4:31022439 for a 4 bp deletion at chromosome position 31022439). Shannon entropy was calculated based on the read counts for each insertion or deletion. If in one sample a high number of sequencing reads with the same indel is detected the entropy is low. Conversely, the entropy is high when a high number of different indels is detected. Shannon entropy analysis confirmed what I estimated by investigating the number of indels. In samples from short-term cultures the entropy was highest, decreasing in methylcellulose cultures, and was lowest in samples from long-term cultures (Figure 4.38). These results strongly suggest that an expansion of distinct indels took place during long-term culture. To dissect this in more detail, I analyzed the spectrum of indels that were identified in the samples.

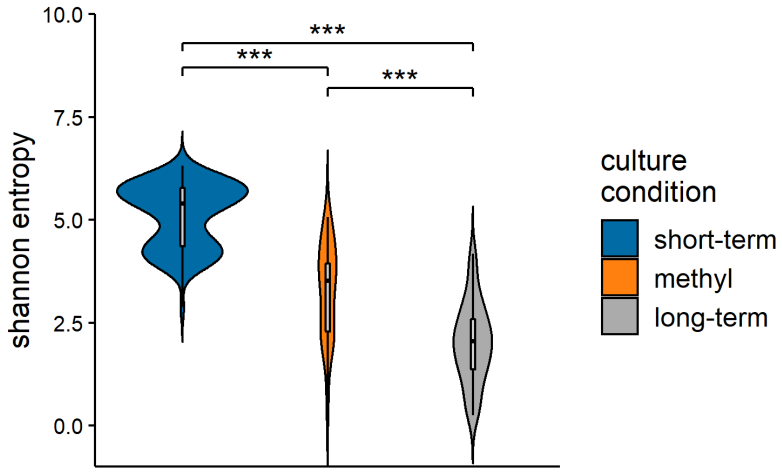


Figure 4.38 Shannon entropy analysis of indel composition. Shannon entropy analysis was performed on the indel count data for each sample. Samples were grouped based on their culture condition, independent of sampling time and biological sample. Median, 25% and 75% quantile, and 95% confidence intervals of the entropy values are indicated by the boxplot within the violin plot. The violin plot shows the density distribution of the values. Statistical analysis was performed using Wilcoxon rank-sum test. The p-values were adjusted for multiple testing. *p<0.05, **p<0.01, ***p<0.001

4.5.3 Preleukemic Mutations Lead to Clonal Expansion of HSPCs

To compare the indel spectrum between different samples and time points, the 15 insertions and deletions (and point mutation for *DNMT3A*) with the highest frequency were investigated in detail. The most frequent indel in the *ASXL1*^{mut} cells from short-term culture was the 8 bp deletion that was introduced via HDR (Figure 4.39, D:8 (31022442)). In both examples, the frequency of this deletion increased in the first 14 days of culture but was outgrown by other cell clones at later time points. The indel composition was the same after culture in methylcellulose, since the cells were seeded from the same transfected original sample. The overall frequency was slightly reduced, indicating that more wild type clones were present in this sample. In contrast, after 6 weeks of long-term culture especially two deletions have expanded in CB006 (4 bp and 8 bp, Figure 4.39 A), whereas a different 8 bp deletion was predominately found after 9 weeks of long-term culture in CB009. In contrast, in an independent sample after 4 weeks of culture, the overall indel frequency was greatly reduced (Figure 4.39 B).

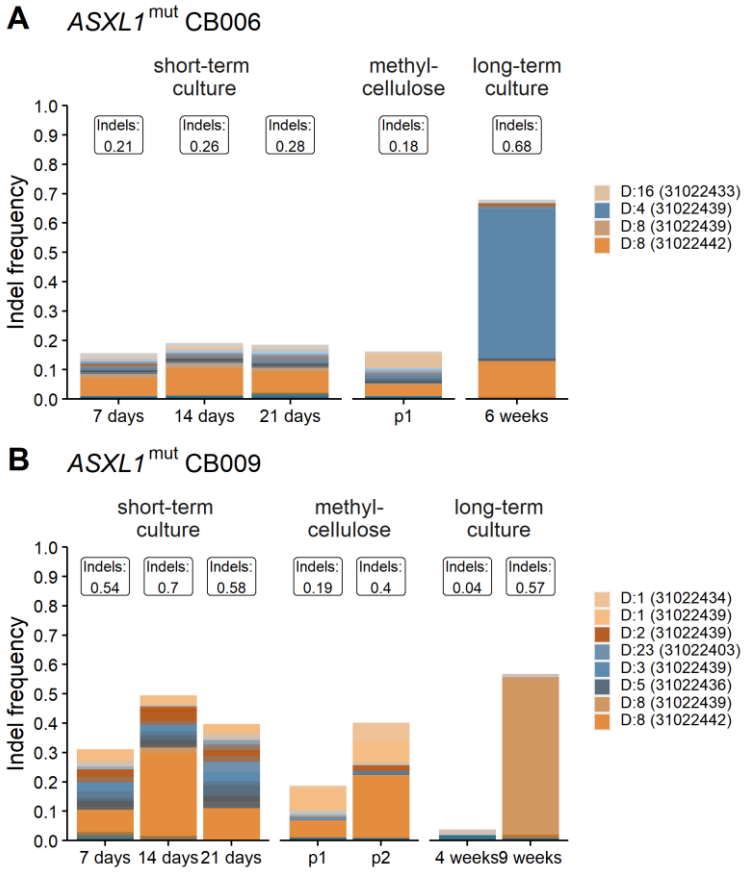


Figure 4.39 Clonal composition of *ASXL1*^{mut} cells. Bulk cells from different culture conditions and time points were analyzed via CRISPRseq. A) One transfection replicate from CB006 and B) one transfection replicate from CB009. The 15 most frequent insertions and deletions are depicted in the stacked bar graph. In the legend only the most prominent indels are depicted (A: VAF>1.2%, B: VAF>3%). The labels above the bars indicate the overall indel frequency in the sample. The indel codes are unique for each indel by integrating information about the type of indel (D/I), the size of the indel, and the start site of the mutation.

I investigated different time points after long-term culture. It has to be noted that the time points are independent of each other, since per sample and time point a complete well has to be analyzed. An expansion of one or two distinct indel types was observed in multiple *ASXL1*^{mut} samples, indicating that cell clones carrying these specific deletions might have a growth advantage in this culture condition. However, only in two samples the overall VAF after long-term culture was higher than in the liquid culture. In 5 out of 14 samples, the overall VAF did not change after long-term culture, whereas in 50% of the *ASXL1*^{mut} samples examined (7 out of 14) wild type cells overgrew the mutant cells (example: Figure 4.39 B, 4 weeks). This is consistent with the observation that the *ASXL1*^{mut} bulk did not generate more viable cells after long-term culture than the wild type group (chapter 4.4.4, Figure 4.32).

Looking at *DNMT3A*^{mut} cells, I could see that approximtely 50-60% of the bulk cells in the liquid culture system carried indels and approximately 40-50% have the introduced point mutation (M: R882H, Figure 4.40). The frequency of the mutations changed only little over the period of 21 days in the liquid culture system. Interestingly, a second major mutation was observed, which is a 1 bp insertion at the CRISPR cut site (I:1 (25457243)). In methylcellulose culture, the point mutation decreased with passaging in both examples shown here, whereas the overall mutation frequency did not change. In CB008, an increase in the 1 bp insertion was observed at the cost of other indels, whereas the mutation frequency was comparable to that of short-term culture samples (Figure 4.40 A). Also, in CB009 an expansion of the 1 bp insertion was observed with an accompanying decrease of the point mutation (Figure 4.40 B). In contrast to the *ASXL1*^{mut} cells, in only one *DNMT3A*^{mut} replicate the overall mutation frequency decreased over the long-term culture period (from 97% to 77%), whereas all other replicates retained a high overall mutation burden (>80%). Opposing dynamics could be observed in the majority of the samples with either the point mutation expanding and indel frequency decreasing (3 out of 13) or the other way round (7 out of 13). In three samples the composition was stable. Moreover, in 11 of the 14 long-term culture samples the same 1 bp insertion was found to contribute majorly to the mutation burden. This observation indicates that a clonal selection took place in long-term culture and that clones carrying these specific mutations had an advantage over wild type and possibly other mutated cells.

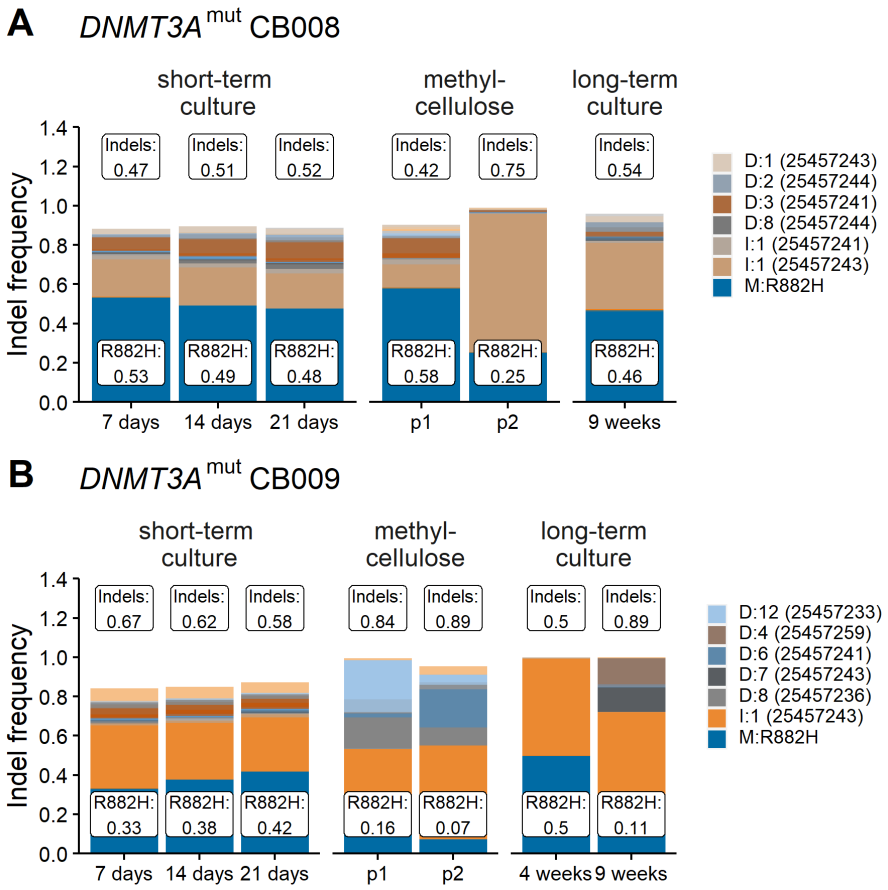


Figure 4.40 Clonal composition of *DNMT3A*^{mut} cells. Bulk cells from different culture conditions and time points were analyzed via CRISPRseq. A) One transfection replicate from CB008 and B) one transfection replicate from CB009. The 15 most frequent insertions and deletions are depicted in the stacked bar graph. In the legend only the most prominent indels are depicted (A: VAF≥2.4%, B: VAF≥10%). The labels above the bars indicate the overall indel frequency in the sample. The labels at the blue portion of the graph show the frequency of the point mutation R882H. The indel codes are unique for each indel by integrating information about the type of indel (D/I), the size of the indel, and the start site of the mutation. M:R882H indicates the point mutation.

In *TET2*^{mut} samples a high number of different insertions and deletions was introduced, therefore the 15 insertions and deletions with the highest frequency made up only approximately 60% of all indels detected in the short-term culture samples (Figure 4.41). The most frequent indels were 1 bp insertions and deletions around the expected Cas9 cut-site (D/I:1 (106164912) or D/I:1 (106164913)), but there was no obvious selection for this type of indel. In long-term culture samples, a decrease in number of indels was observed with an accompanying increase in the specific indel frequency. No selection for a distinct indel type was observed after long-term culture. Nevertheless, clones with mutations predominately produced myeloid cells in the long-term culture, indicating a general fitness advantage of *TET2*^{mut} cells over wild type cells.

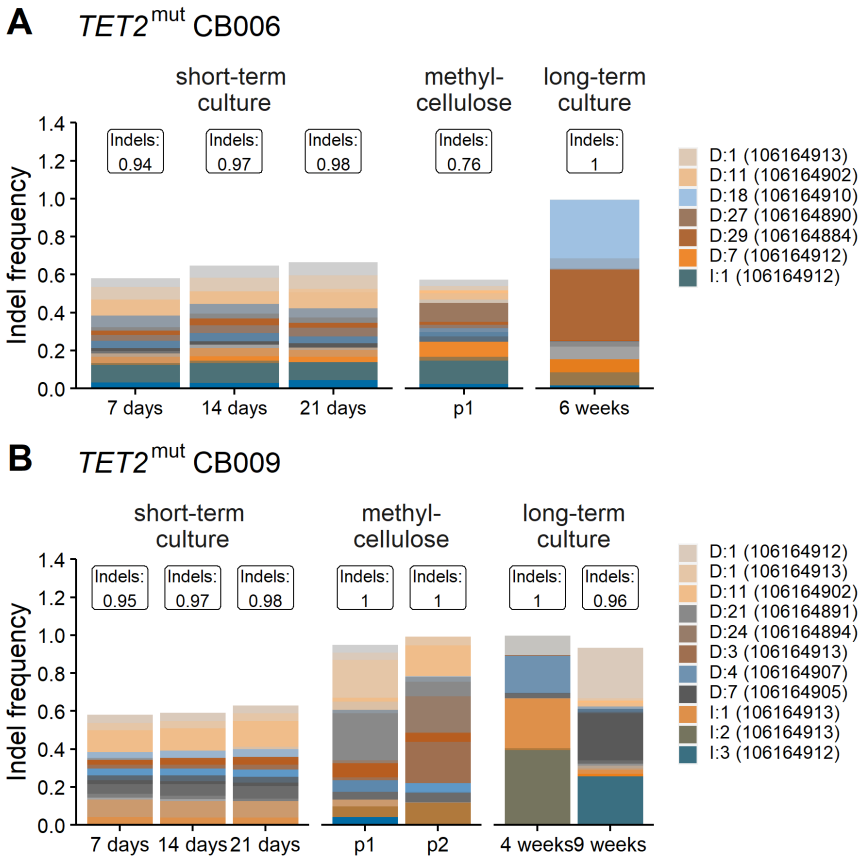


Figure 4.41 Clonal composition of *TET2*^{mut} cells. Bulk cells from different culture conditions and time points were analyzed via CRISPRseq. A) One transfection replicate from CB006 and B) one transfection replicate from CB009. The 15 most frequent insertions and deletions are depicted in the stacked bar graph. In the legend only the most prominent indels are depicted (A: VAF>7%, B: VAF>11%). The labels above the bars indicate the overall indel frequency of the sample. The indel codes are unique for each indel by integrating information about the type of indel (D/I), the size of the indel, and the start site of the mutation.

5 Discussion

Leukemia and especially AML is a disease of hematopoietic stem and progenitor cells, characterized by impaired hematopoiesis [66]. Fine-tuned regulation of hematopoietic stem and progenitor cell fate is a key component of hematopoiesis and mature blood cell production. AML is a condition often associated with increased age. Mutations found in AML have also been identified in elderly individuals without a known hematologic condition in the bone marrow or blood [51, 52] or in various cells of the stem cell compartment, including HSCs [54]. Similar mutations have been identified in non-leukemic cells of AML patients, leading to the assumption that these aberrations define a preleukemic state, leading to increased self-renewal of HSCs and a competitive proliferation advantage. Understanding the molecular processes leading to enhanced fitness of a cell clone is crucial to dissect leukemia onset.

In this thesis, mutations implicated in the preleukemic process were analyzed in detail. The first part of the thesis revolved around patients with AML t(8;21), an established and well defined AML subentitiy, and how different mutations interplay to affect disease pathogenesis and patient outcome. Sequencing of a large patient cohort with a targeted next generation sequencing approach was the method of choice to investigate the mutational landscape and the clonal dynamics of this disease. Mutations in the genes *DNMT3A*, *TET2*, and *ASXL1* have been termed preleukemic mutations in AML and are thought to be disease-initiating events. The main question addressed in this thesis was how mutations identified in elderly individuals with CHIP and in non-leukemic cell populations functionally effect the stem cell compartment. With this thesis, I contributed to the strongly forming base of knowledge about preleukemic mutations, by investigating their functional impact on healthy human stem and progenitor cells. So far, the investigations were limited to mouse models and based their observations on complete or conditional knockouts. Here, I established a model that makes it possible to study direct mutational effects independent of confounding factors like age or environmental stress in young, healthy cells from umbilical cord blood. In the last 5 to 10 years, genome editing techniques have been improved continuously, making it possible to target genomic hotspots. Simultaneously, continuous improvement of next-generation sequencing techniques and bioinformatics pipelines made it possible to overcome some limitations of *in vitro* models.

5.1 Cord Blood Cells as *in vitro* Model System

In this project, I focused on the three most frequently mutated genes in patients with clonal hematopoiesis of indeterminate potential (*DNMT3A, TET2, ASXL1*) [51, 52], which have also been reported as preleukemic events in AML [49, 61].

To introduce site-specific mutations, I chose to work with fresh cord blood cells as a model system. I had to consider the collection process and the ethics involved in using primary material. In both cases, cord blood has an advantage over whole blood or bone marrow as cell source since it can be collected non-invasive after birth. Bone marrow has a high cell density, together with a relatively high frequency of $CD34^+$ progenitor cells (~1.4%, reported 1-3% [177]). In whole blood of healthy individuals, the $CD34^+$ frequency is rather low (~0.1%). Although the frequency of $CD34^+$ cells in cord blood is lower than in the bone marrow (~1%, reported 1% [177]), several advantages can be found. CB cells are younger, have undergone less replications in their lifetime, and are therefore less likely to have acquired somatic mutations. It has been demonstrated that CB cells have a lower mutation burden than blood cells from healthy adults and that the number of mutations found in HSPCs increases as a function of age. The researchers estimated that 0.13 ± 0.02 somatic mutations are acquired per year [48]. Due to the young age, CB cells have not been exposed to external stress factors, making it possible to investigate the effect of the mutations themselves. CB cells have a higher proportion of primitive HSPCs (cell surface markers $CD34^+CD38^-$) [178]. Compared to BM cells, they have a higher expansion as well as proliferation potential [179, 180]. In line with this, a higher replating potential of CB cells was reported [181]. These features make progenitor cells from umbilical cord blood more applicable in *in vitro* culture systems and downstream assays.

Hematopoietic stem and progenitor cell culture is not a new concept and many researchers have focused on improving HSC *in vitro* culture (reviewed: [182]). I focused on finding an appropriate culture system for HSPC expansion and downstream experimental testing. The expansion medium contained the cytokines SCF, TPO, and FLT3-L, which are early acting factors and support the expansion of progenitors [183]. The feeder cells used here are engineered to produce G-CSF, SCF, and IL-3, supporting production of mature cells [184], which in turn outgrew the progenitor population. HSPC expansion and culture was better supported without feeder cells. I then examined different cytokine combinations, analogous to Choi et al. [185], to find the optimal culture condition for HSPC expansion and differentiation. Culture with SCF, FLT3-L, and TPO led to the longest retention of CD34 marker expression and consequently the highest proportion of progenitor cells. This medium composition was

suited for HSPC expansion. By adding IL-3 and IL-6 to the culture, more cells expressed myeloid differentiation markers, making it possible to study differences in primitive as well as mature differentiation marker expression over the course of three weeks. G-CSF and GM-CSF stimulated granulocyte and monocyte marker expression early in the culture (like what was observed in the feeder cell culture) making it difficult to study effects of introduced mutations in the progenitor cell compartment. I therefore used a culture medium without G-CSF and GM-CSF. Similar *in vitro* culture protocols have been published in recent years [151, 186, 187], confirming my decision for liquid culture with supplemented cytokines. It should be kept in mind that the *in vitro* culture system forces HSCs to proliferate and generate progeny due to the lack of niche cell interaction and proper nutrient and oxygen concentrations. The cells potentially behave differently in physiological conditions. Nevertheless, these differences might help understand how genetically altered HSPCs can cope with changing environments in terms of enhanced fitness.

5.2 Modeling Preleukemic Mutations in Cord Blood HSPCs

Investigating functional gene disruption in a model system has vastly contributed to the understanding of cancer development. For many years, researchers relied on knockout or overexpression models to investigate gene functions. Around the start of this project, site-specific gene editing was mainly achieved by using Zinc-finger nucleases (ZFN) [188], transcription-activator like effector nucleases (TALENs) [189], or CRISPR-associated (Cas) 9 nucleases [190]. Disadvantages of the first two systems are complicated design strategies and a need to synthesize new proteins for each target since these techniques rely on protein-DNA interaction. Cas9 nucleases are guided by simple base pairing of single guide RNAs to the DNA target sequence, making the system more flexible. One delivery system for CRISPR/Cas9 is based on the electroporation of Cas9-expression plasmids together with sgRNA-expression plasmids into the target cells. I adapted protocols published before [151, 191], using a plasmid expressing Cas9 together with GFP as a reporter. Mandal et al. introduced mutations in *B2M* and *CCR5* in HEK293T and K-562 cell lines, respectively. They achieved target editing in 20-48% of cells, depending on the sgRNA. In $CD34^+$ cells the mutation frequencies were similar to those in K-562 [191]. Tothova et al. reported a mutation efficiency of up to 53% when targeting *STAG2* in $CD34^+$ cells [151]. Testing the system in K-562, I achieved an editing efficiency of up to 22% for the most efficient target. For *DNMT3A*, I was limited in the location of the target site since I wanted to introduce a point mutation at the defined position. I could therefore not test many different sgRNAs and select the most efficient one. The 22% indel frequency for *TET2* sgRNA are in the range of the efficiencies reported by Mandal et al. by

targeting K-562. Using the same approach in $CD34^+$ cells isolated from frozen CB MNC, I was not able to achieve genetic modification. The viability of the cells was very low after electroporation. Tothova et al. used freshly isolated $CD34^+$ from umbilical cord blood or adult peripheral blood [151]. Mandal et al. used commercially available isolated $CD34^+$ cells, but from G-CSF mobilized peripheral blood [191]. Both groups were able to successfully introduce mutations in the $CD34^+$ cells, meaning that the use of freshly isolated primary material or already isolated commercially available $CD34^+$ was essential.

With editing efficiencies around or below 20%, I would not be able to work in the cell bulk but had to enrich the cells after transfection. A second delivery system was established recently for primary hematopoietic cells, like $CD4^+$ T cells [192] or $CD34^+$ HSPCs [186], around the start time of this project. This system used recombinant Cas9-protein precomplexed with *in vitro* transcribed sgRNA forming a ribonucleoprotein (RNP). RNPs can be electroporated into the target cells with or without a single stranded DNA donor for homology directed repair. A disadvantage of this system is that positive selection is not possible, because only RNPs consisting of Cas9 and the sgRNA are delivered into the cell. The problem of selecting transfected cells was the main reason I tested the vector-based strategy first. Since this delivery method was inefficient, I adopted the RNP approach. Other groups reported very high editing efficiencies in primary cells, making the need for enrichment obsolete (over 60% in HSPCs [186, 193]). With electroporation of Cas9-sgRNA RNPs, I was able to achieve editing efficiencies ranging from 20% to over 80% in the cell bulk, depending on the target gene and sgRNA. By modifying the molar ratio of Cas9 protein and sgRNA, I was able to enhance the mutation frequency to around 80% for *DNMT3A* exon23 T1 sgRNA. With use of the HDR donor, I was also able to overcome the problem of a high proportion of in frame deletions of codon R882 with *DNMT3A* exon23 T1 sgRNA. The efficiency of HDR could be improved greatly by modifying the location and the length of the single stranded DNA template (achieving 25% SNV in K-562 and up to 50% in primary HSPCs). I chose to use asymmetric single stranded DNA donors with 65 bp distal and 35 bp proximal of the expected cut site. I also found that a donor complementary to the PAM strand is most efficient in introducing SNVs, whereas a non-PAM donor was less efficient overall but inhibited indel generation at the target site. Schumann and colleagues reported the use of symmetric single stranded DNA as donors for HDR in primary human T cells with an efficiency of 25% [192] and Gundry et al. reported HDR-mediated gene editing in primary human $CD34^+$ cells with a mean of 22%. They achieved the highest editing efficiency (27%) by using 30 pmol of an asymmetric ssDNA donor [186]. My donor choice was based on the results reported by Richardson et al. [194] and Liang

et al. [195]. Richardson et al. reported that the highest efficiency was achieved by using an asymmetric donor complementary to the non-target strand (PAM strand by my definition) with 36 bp overlap 3' and 91 bp overlap 5' of the expected cut site. They have shown that by varying donor design, the improvement of the editing efficiency is greater than by chemical intervention [194]. Liang et al. reported that the location of the sgRNA in relation to the desired mutation site is important and might influence the outcome of the editing [195]. The results from my establishing experiments are in line with the reported protocols. After thorough evaluation, delivering Cas9-nuclease and sgRNAs as RNPs was the most efficient method for my experiments. This method was superior to vector-based delivery regarding cell viability as well as editing efficiency.

I confirmed the expression of *DNMT3A*, *ASXL1*, and *TET2* mRNA in the selected cell model early after transfection, marking the starting point of the experiments. The confirmation of mRNA expression of the target genes was performed for one biological sample with three technical replicates. I had to limit these preliminary experiments due to very little starting material at day 0. After isolation, I started the expansion culture with 1-1.5x10^6 cells. I could confirm that gene editing of the three target genes led to a downregulation of mRNA after a certain time (*DNMT3A*^{mut} and *ASXL1*^{mut} at 14 days, *TET2*^{mut} after 3 days). The target sites of the qRT-PCR primers are located upstream of the mutation site, therefore I did not expect to see a decrease in mRNA expression of the targets. The high number of indels and the resulting frameshift mutations might lead to degradation of mRNA because of stop-loss or acquisition of premature stop codons [196]. Simultaneously, the turnover rate of *TET2* might be higher, leading to an early loss of mRNA due to degradation processes. Additionally, *DNMT3A* and *ASXL1* mRNAs, in which the mutations are located in the last exons, might escape nonsense-mediated decay, resulting in higher mRNA and protein levels compared to *TET2*. I confirmed the downregulation of all proteins by western blot. Here, I observed that the downregulation of TET2 was more pronounced than of the other targets. It has to be noted that the samples were taken at different time points and the material was sparse, so western blots could not be repeated.

For *TET2*^{mut}, I was able to confirm the technical knockout on a functional level by showing that the overall level of 5-hydroxymethylcytosine (5hmC) was reduced compared to the wild type cells. Although I observed less 5hmC levels in *TET2*^{mut} compared to wild type samples, which is in line with a functional decline of *TET2* on protein level [88], I did not find a change in 5mC levels. Other researchers have found similar results [117, 197]. The overall level of 5mC is much higher throughout the genome than 5hmC [87] and little changes might not be detected

with the methods described here. I detected highly variable 5mC levels in the *TET2*^{mut} bulk analyzed from different samples. This might also be due to a saturation of the antibodies with very high levels being underestimated. I did not observe a change in 5mC levels in *DNMT3A*^{mut} cells. This does not necessarily mean, that I did not generate a functional knockout. First, DNMT3A is mainly involved in *de novo* methylation and not in maintenance of the methylation level and therefore the impact of DNMT3A disruption on the global methylation level might not be detectable. Second, DNMT3B might cope for the loss of DNMT3A and maintain methylation levels. My observation is in line with previously reported results [97]. For *AXSL1*^{mut} the possibilities of confirming a functional knockout are rather slim. The functions of ASXL1 are diverse as interaction partner of PRC and an *ASXL1* knockout/mutation might have a variety of immediate effects. As for *DNMT3A*^{mut}, I had to rely on the genetic, transcriptional, and translational observations. Overall, the results show that introducing knockouts is feasible in primary CB CD34^{+} cells by genetic modification using CRISPR/Cas9.

5.3 Mutations in *TET2* and *DNMT3A* Have Diverse Effects on Differentiation and Self-renewal of HSPCs

Introduction of DTA mutations did not have a negative influence on general proliferation or cycling behavior of investigated cells. I observed a shift in cell cycle phases in the multi KO sample. I cannot rule out that the cells targeted with three sgRNAs simultaneously exhibited increased stress due to induction of DNA damage, indicated by the low viability in these samples. Accumulation of DNA damage leads to G2-arrest of cells to initiate DNA repair mechanisms and avoid replication of damaged DNA [198]. These general observations regarding cell proliferation might mean that the introduced preleukemic mutations have a more subtle effect on HSPC differentiation and self-renewal.

I got a first hint that introduction of DTA mutations led to impaired differentiation of HSPCs by investigating the surface marker expression over the course of three weeks. The frequency of granulocyte and monocyte/macrophage marker expressing cells was lower in multi KO than in wild type cells, meaning that myeloid differentiation marker expression was delayed in mutated cells. Conversely, the cells retained expression of the progenitor cell marker longer than wild type cells. *DNMT3A*^{mut} cells exhibited a retention of CD34 marker expression over a longer time than wild type cells. In contrast, *TET2*^{mut} cells had similar delayed myeloid marker expression as observed in multi KO cells, which was consistent for all biological samples. This means a combination of both effects is achieved in multi KO cells, with *TET2*^{mut} leading to a delayed expression of differentiation markers and *DNMT3A*^{mut} leading to a longer retention of

primitive surface marker expression. It would be interesting to follow up further and investigate whether the presence of mutations in all three genes enhances the observed effects. At the first glance, the effects seemed to be stronger, but to draw a conclusion in this direction more biological samples would be needed. The introduced mutations in *ASXL1* did not have a short-term effect on the cells in *in vitro* culture. It is possible that the introduced mutations in the bulk are not prominent enough and a selection of mutated cell clones would be needed. Another possible explanation is that the effects of *ASXL1* aberrations are not observed immediately since the function of *ASXL1* is more diverse in its interaction with PRC1/2.

A second observation that pointed in the direction of DTA mutations altering the differentiation and self-renewal properties of HSPCs was the increased serial replating potential of *TET2*^{mut} cells. I observed increased CFU numbers after replating in the preliminary tests with multi KO cells and later in the *TET2*^{mut} samples in all replicates. An enhanced replating capacity for *Tet2* knockout models has been described in mice [116, 118, 168]. My results differed in the numbers of passages that could be performed with the mutated cells. All the mentioned studies used either conditional or complete knockout models resulting in a complete loss of *Tet2*, which have a more severe effect on stem cell function. The increased serial replating capacity is hypothesized to reflect increased self-renewal potential of the progenitor cells investigated in CFU assays. For *Dnmt3a* knockout models similar observations have been reported [165]. I could not confirm an increased serial replating potential of *DNMT3A*^{mut} cells in the *in vitro* setting, indicating that mutations in *TET2* and *DNMT3A* have a more diverse functional impact. Self-renewal potential is determined *in vivo* by the capacity of cells to serially transplant from one mouse into another. Recently it has been demonstrated that *Dnmt3a* KO HSCs can be serial transplanted indefinately in mice. The cells loose their differentiation capacity and cannot reconstitute blood cell lineages [99]. *Dnmt3a* KO therefore has a severe impact on hematopoiesis in mice. *DNMT3A* mutations in myeloid conditions are often heterozygous point mutations in specific hotspots (e.g., Arg882), leading to a dominant-negative inhibition of wild type *DNMT3A* [104]. The mutations found in patients and introduced in the cells here are therefore functionally different from a complete knockout. This might be one explanation for the results observed here and described in the literature. Another reason might be that the cells profiting from *DNMT3A* mutations cannot be detected in a CFU assay, mainly supporting outgrowth of committed progenitors. In long-term culture initiating cell (LTC-IC) assays more primitve cells can be detected. In wild type samples, I found very few viable cells and in only one replicate CFC were detected after long-term culture. The seeding concentration was too low to support colony formation in wild type cells. Therefore it is even more remarkable that I

detected a high number of viable cells in the mutants. *ASXL1*^{mut} supported myeloid cell production only in one biological sample, indicating that the introduced mutation in this gene is not enough to confer a proliferative advantage to the cells in this setting. I did nevertheless detect CFUs after long-term culture of *ASXL1*^{mut} cells. The cells seem to have acquired a fitness advantage over wild type cells, but are not able to regenerate the culture. It might be that the fitness advantage is more of an enduring kind than of actual self-renwal or proliferation.

DNMT3A^{mut} and *TET2*^{mut} in contrast led to a strong increase in cell numbers after long-term culture compared to wild type cells. It is possible that these muations confer a proliferative advantage to the mutated cell clone. Even after 4-9 weeks of culture a higher proportion of progenitor cells was present in the mutated samples, indicated by the increased frequency of $CD34^+$ cells. This means that a number of primitive progenitors present in the sample supports continuous production of mature cells. This observation is supported by the detection of CFUs for all mutant conditions after long-term culture. Interestingly, the highest number of colonies was derived in *DNMT3A*^{mut} samples, whereas in *TET2*^{mut} samples, more replicates were able to generate colonies. This observation supports the hypothesis that mutations in these genes have a different impact on the stem and progenitor cell compartment. In *DNMT3A*^{mut} samples, the benefitting cell clones might be more primitive and therefore lead to generation of a high number of colonies. These cells seemed to be rarer and were not captured in all replicates due to low cell numbers when seeding. Supporting this hypothesis is the result that I did not find differences in the expression of mature cell markers but rather of the stem and progenitor cell marker. Neither did I observe an advantage of *DNMT3A*^{mut} cells in the CFU assay, supporting the hypothesis further. In *TET2*^{mut} conditions in contrast, it might be that more committed progenitors have acquired a self-renewal advantage and are continuously generating mature myeloid cells. An increased serial replating capacity in CFU assays also points to an increased self-renewal potential of *TET2*^{mut} progenitors. This observation matches the results discussed in the beginning of this chapter (retention of CD34 marker in *DNMT3A*^{mut} and delayed myeloid marker expression in *TET2*^{mut}). Studies performed by our group and others are in line with the described distinct effects of *DNMT3A* and *TET2* mutations on the hematopoietic stem and progenitor cell compartment. Arends et al. reported that *DNMT3A* mutations could be detected in T cells (with higher VAFs than other mutations), indicating that primitive HSCs or multi-potent progenitors expanded preferentially in these individuals. *TET2* mutations in contrast led to a stronger myeloid commitment and were found with a high allelic burden in mature myeloid cells but also myeloid committed progenitors [54]. Buscarlet and colleagues reported similar results with the hypothesis that *DNMT3A* mutations occur in more primitive HSCs and *TET2*

mutations occur in committed HSPCs [199]. Other than the retrospective studies on sorted cell populations carrying mutations, I transfected a heterogeneous pool of $CD34^+$ HSPCs simultaneously. My reported results strongly suggest that distinct mutations led to a fitness advantage of either primitive HSCs carrying *DNMT3A* mutations or more committed progenitors with *TET2* alterations. Single cell analysis of a multi KO sample would elucidate this hypothesis further by connecting the gene expression signature and the mutation status.

5.4 *DNMT3A*^{mut} and *TET2*^{mut} Promote Clonal Expansion of HSPC

In the last part of this thesis, I investigated whether distinct clones expanded after long-term culture, clarifying if some mutations led to the preferential expansion of a clone even in *in vitro* settings. The number of different indels found in cells from short- and long-term cultures differed significantly while the overall indel frequency was stable, indicating that an expansion of distinct indels and therefore distinct cell clones took place. Shannon entropy analysis confirmed this observation by demonstrating that long-term culture samples are less diverse than early samples. Deep sequencing of the target region made it possible for me to trace individual insertions and deletions. Given that one cell clone carries distinct mutations in the target region, I could identify expanding clones in the bulk sample. Bulk analysis serves as an internal competition assay, where I can identify the clones with growth advantage, if there are any. *ASXL1*^{mut} samples were the only ones in which I observed the wild type overgrowing mutated clones in almost half of the samples. Although I noted the preferential expansion of distinct indels in some samples, the total indel frequency increased in only two samples. As hypothesized before, the introduced *ASXL1* mutations do not confer an advantage to a distinct cell clone in the conditions analyzed here. In line with the hypothesis is that only few samples supported mature cell production. *ASXL1* is often mutated in individuals with CHIP as well as older AML patients [200]. There might be other factors that are needed for *ASXL1* mutations to lead to preferential expansion. Age might be an influencing variable, since the mutation is more often found in older people. Interestingly, it was recently demonstrated that an aging microenvironment influences clonal selection and expansion of HSCs transplanted from young donor mice [201]. Although the researchers only investigated *Dnmt3a*^{R878H} mutations (the murine homologue to R882H mutations), this might also be true for other clonal hematopoiesis-associated mutations. I investigated healthy, young cells, in which *ASXL1* alterations might function as passenger mutations. Exposing the cells to external stress factors (for example irradiation or UV light) might be a way to trigger the expansion of *ASXL1*^{mut} clones. It has recently been reported that mutations in *ASXL1* are strongly associated with current and past smoking [202], supporting the hypothesis that external factors are needed to trigger clonal

expansion of *ASXL1*^{mut} cells. Furthermore, *ASXL1* mutations might cooperate with disruptions of PRC1/2 members (e.g., *EZH2*), promoting transformation, as hypothesized by Abdel-Wahab and colleagues [124].

In my *in vitro* experimental approach, I found that mutations in *TET2* and *DNMT3A* led to a preferential expansion of distinct cell clones in long-term cultures. The observation that in *DNMT3A*^{mut} samples one or two mutations contributed to the majority of the observed mutations supports the hypothesis that primitive cells are preferentially expanding in these samples. Primitive mutated LTC-IC clones generate committed progenitors and later mature cells that all carry the same mutation. In 4 of the 14 investigated replicates, the predominantly expanding mutation was the introduced point mutation R882H. In these cells, DNA methylation at distinct promotors is disrupted by inhibition of tetramer formation, leading to enhanced self-renewal capacities of HSCs [104, 203]. In the majority of the investigated samples, a 1 bp insertion was found as the dominant clone. The resulting frameshift leads to severe changes in the amino acid composition of the N-terminal methyltransferase domain, possibly interfering with the function of DNMT3A to form tetramers, as well as DNA interaction, and enzymatic activity. In *TET2*^{mut} samples, the mutation composition of long-term culture samples appeared to be more diverse. In this case, different mutated committed progenitors with enhanced self-renewal potential might contribute to the mature cell population. This analysis confirms that *DNMT3A*^{mut} as well as *TET2*^{mut} clones gain a fitness advantage over wild type clones and in turn expand and contribute to the majority of blood cell production. Although I could not confirm a proliferative advantage of *ASXL1*^{mut} clones, the cells nevertheless have acquired endurance to generate mature colonies in CFU assays. This is in line with CHIP clones preferentially expanding in elderly individuals as well as the emergence of clones carrying similar mutations in patients in complete remission. Acquisition of DTA mutations might therefore potentially pose a risk for further mutation acquisition in an expanding cell clone, promoting leukemogenesis.

5.5 DTA Mutations in Early Leukemogenesis

Loss of *DNMT3A, TET2*, and *ASXL1* has severe effects in mouse models. For example, a complete knockout of *Dnmt3a* in mice led to accumulation of HSCs [97], expansion of myeloid cells, impaired erythroid maturation [165], and malignant transformation [167]. Similar effects have been described for *Tet2* knockout mice accompanied by premature myeloid development [116, 118, 170, 204]. *Asxl1* knockout led to severe conditions in mice including myelodysplasia [123, 205]. Investigations of truncated *Asxl1* revealed milder effects [135, 136], but cooperating mutations, for example in *Runx1*, induced malignant transformation [125]. A repopulation advantage of *DNMT3A*^{mut} and *TET2*^{mut} has been demonstrated in various xenograft models. It has been established that a knockout of either gene can have severe consequences in mice. A double-knockout showed even more severe features with bone marrow failure and infiltration of non-hematopoietic organs [206]. The group discovered that altered gene expression of double KO mice was similar to *Tet2* KO with the addition of a few differentially expressed genes. In wild type HSPCs, *Tet2* represses specific differentiation associated genes (for example *Ikfz1* for lymphoid development or *Epor* for erythroid differentiation), meaning that *Tet2* loss might lead to enhanced formation of progenitors. *Dnmt3a* loss leads to de-repression of the stem cell program [97]. The impact of introduced mutations here on the stem and progenitor cell compartment in humans and therefore the development of hematologic malignancies is more subtle. I was able to confirm that human HSPCs carrying these mutations show similar features in *in vitro* models and that the mutations led to clonal expansion of HSPCs.

As discussed before, mutations in epigenetic regulators have been identified as preleukemic events. This means that the mutations themselves do not drive leukemic transformation of a cell clone. In patients with AML t(8;21), we identified mutations in epigenetic regulators as second most common mutations (45% of patients with at least one mutation in one of the genes). The most common mutations were found in members of Ras/RTK signaling pathways (more than 60% of the patients). It is highly likely that mutations in effectors of DNA methylation (including *DNMT3A* and *TET2*) are early events, as identified by comparing variant allele frequencies, as well as pairwise mutation precedence. Chromatin remodelers (i.e., *ASXL1*) are in the context of AML t(8;21) acquired later. Mutations in Ras/RTK effectors are rather late events. These observations seem in line with mutations in *DNMT3A* and *TET2* being acquired even before the onset of the disease. It has to be noted that compared to other AML entities *DNMT3A* mutations are less frequent in AML with t(8;21) (4 out of 331 investigated patients). A similar observation has been made by other researchers [100]. The RUNX1-RUNX1T1 fusion protein interacts with de novo DNA methyltransferase DNMT1 and HDAC1 to silence

promotors of RUNX1 target genes [207] and aberrantly represses differentiation of myeloid progenitors [208]. Another group reported that RUNX1-RUNX1T1 promotes transcription of *DNMT3A* and global hypermethylation of promotors [209]. It is possible that hypermethylation is achieved by enhanced transcriptional activation of *DNMT3A* as well as direct interaction with DNMT1 by RUNX1-RUNX1T1. Although *DNMT3A* mutations have been reported to lead to global hypomethylation of DNA, promoters of distinct genomic regions were found to be hypermethylated [97, 101]. Therefore, the effects of *DNMT3A* mutations and the RUNX1-RUNX1T1 might be similar and the events therefore redundant.

It has been demonstrated that mutations in epigenetic regulators persist in remission [210]. In the patient cohort, I found in 12 out of 56 patients persisting mutations in complete remission, including two patients with *DNMT3A* variants. Clones carrying mutations in epigenetic regulators survive induction chemotherapy [61]. It has been proposed that mutations in *DNMT3A* might be responsible for treatment resistance in the relapse. It has also been reported that the variant allele frequency of *DNMT3A* mutations at relapse is at least as high as in the remission or higher [49]. Unfortunately, I did not have samples from the relapse of these patients to investigate this further. A feature of preleukemic cells is that they are able to generate all hematopoietic lineages [49]. The mutations investigated here do not disrupt hematopoiesis but rather alter distinct types of stem or progenitor cells to preferentially expand by limiting differentiation and conferring enhanced self-renewal. In the investigated patients, mutations in epigenetic regulators did not have an impact on the patients' survival or disease outcome. We rather found that aberrations in Ras/RTK signaling genes like *KIT* and *FLT3*, especially when the mutation burden is high, have a negative impact on survival [163]. It is therefore likely that these mutations are the drivers of full-blown leukemia when acquired in a preleukemic clone. A similar observations was made in mouse models with a *Dnmt3a*-null phenotype, in which *c-Kit* is the driver of malignant transformation [211]. In 2018 a group investigated residual mutations after induction chemotherapy in AML patients and detected especially DTA mutations with high variant allele frequencies present in the remission samples. They did not report an elevated relapse risk for patients with residual DTA mutations, but rather with mutations other than those. Those findings indicate that DTA mutations do not drive leukemogenesis but are able to survive induction chemotherapy (in contrast to mutations in other genes) and are, probably due to enhanced clonal expansion, repopulating the bone marrow after therapy. The researchers did not follow-up further whether therapy resistance is observed in relapse clones with DTA mutations. They stated themselves that they did not follow up longer than 40 months and longer follow up times might lead to different results [210]. Although

remission rates are high, enhanced treatment resistance is one major problem in relapsing AML patients. Therefore, researchers have proposed that identifying and targeting preleukemic clones might reduce relapse rates. Effects of *TET2* disruption for example can be corrected by high-dose vitamin C in the mouse model [157]. Hypomethylating agents like 5-azacytidine or decitabine are used for treatment of MDS and were evaluated for the use in AML patients in combination the Bcl-2 inhibitor venetoclax [212]. Using low-dose hypomethylating agents in patients with persisting preleukemic mutations might decrease relapse rates. Similarly, have different inhibitors of RUNX1-RUNX1T1 been evaluated and are promising in preclinical settings [213, 214]. Overall, monitoring and targeting preleukemic mutations might be a solution to prevent the onset of hematologic malignancies, reduce relapse rates, and improve the long-term outcome.

6 Summary and Conclusion

Introducing specific genetic modifications in cord blood HPSCs is an adaptable and very efficient method to investigate functional consequences of gene mutations in the stem cell compartment. I was able to establish a model that allowed testing of clonal hematopoiesis-associated mutations without confounding factors, like the aging microenvironment or a complex mutational landscape. The model can be adapted and used for studying other mutational events, complex mutational compositions, mimicking different stages of myeloid malignancies. Integration of external stress factors or pharmacological intervention can broaden the spectrum of *in vitro* approaches.

The goal of the thesis was to demonstrate the influence of DTA mutations on the hematopoietic stem and progenitor cell compartment. Here, I found that mutations in *DNMT3A*, *TET2*, and *ASXL1* influence stem and progenitor cell behavior differently. Although mutations in *ASXL1* seem to have a minor impact on cell fitness, I could not confirm enhanced clonal expansion in the *in vitro* setting. It might be possible that co-factors such as increased age and exposure to external stress factors, but also other aberrations are cooperating with *ASXL1* to disrupt stem cell functions. *DNMT3A* and *TET2* on the other hand had detectable effects on stem and progenitor cell behavior. Both single mutants showed enhanced long-term culture initiating potential, indicating that the cells acquire self-renewal properties to sustain cell production in long-term culture. I observed clonal expansion of mutant clones in the long-term culture and a competitive advantage over the wild type. *DNMT3A*^{mut} and *TET2*^{mut} seem to influence different cell types preferentially. In *TET2*^{mut} cells, I detected an enhanced serial replating capacity, which I could not confirm in the other mutants, indicating that especially committed progenitors profit from these mutations. The clonal composition of *TET2*^{mut} cell bulks was much more diverse than of *DNMT3A*^{mut} cell bulks. It seems therefore that *DNMT3A* mutations affect more primitive cells. Together, these findings support the hypothesis that *DNMT3A* and *TET2* mutations confer a fitness advantage to hematopoietic stem and progenitor cell clones driving clonal expansion. I did not observe any severe changes in the proliferative capacity or the phenotype of the mutated cells, indicating that DTA mutations are not drivers of leukemia but might prime the cell clone for mutation acquisition.

By investigating AML patients with AML t(8;21), I could dissect temporal mutation acquisition, with mutations in *DNMT3A* and *TET2* appearing early, maybe even in non-leukemic cells. Events that are potentially acquired later influence disease outcome and are therefore more likely leukemic drivers. *ASXL1* mutations seem to play a different role in this

AML entity. Although DTA mutations are not drivers of leukemic transformation, their impact on human stem and progenitor cell function poses a risk factor for subsequent mutation acquisition. Additional investigations in respect to resistance to external stress factors or chemotherapy would contribute to risk assessment of individuals with DTA aberrations.

7 References

1. **Orkin SH, Zon LI**. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell*. 2008; 132: 631–44.
2. **Doulatov S, Notta F, Laurenti E, Dick JE**. Hematopoiesis: A Human Perspective. *Cell stem cell*. 2012; 10: 120–36.
3. **Majeti R, Park CY, Weissman IL**. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell*. 2007; 1: 635–45.
4. **Notta F, Doulatov S, Laurenti E, Poeppl A, Jurisica I, Dick JE**. Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science (New York, N.Y.)*. 2011; 333: 218–21.
5. SMART - Servier Medical Art. https://smart.servier.com/ (accessed September 14, 2020).
6. **Benveniste P, Frelin C, Janmohamed S, Barbara M, Herrington R, Hyam D, *et al.*** Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential. *Cell Stem Cell*. 2010; 6: 48–58.
7. **Goardon N, Marchi E, Atzberger A, Quek L, Schuh A, Soneji S, *et al.*** Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer cell*. 2011; 19: 138–52.
8. **Doulatov S, Notta F, Eppert K, Nguyen LT, Ohashi PS, Dick JE**. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat Immunol*. 2010; 11: 585–93.
9. **Laurenti E, Göttgens B**. From haematopoietic stem cells to complex differentiation landscapes. *Nature*. 2018; 553: 418–26.
10. **Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, *et al.*** Human haematopoietic stem cell lineage commitment is a continuous process. *Nat Cell Biol*. 2017; 19: 271–81.
11. **Notta F, Zandi S, Takayama N, Dobson S, Gan OI, Wilson G, *et al.*** Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science (New York, N.Y.)*. 2016; 351: aab2116.
12. **Catlin SN, Busque L, Gale RE, Guttorp P, Abkowitz JL**. The replication rate of human hematopoietic stem cells in vivo. *Blood*. 2011; 117: 4460–66.
13. **Wilson A, Laurenti E, Oser G, van der Wath RC, Blanco-Bose W, Jaworski M, *et al.*** Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell*. 2008; 135: 1118–29.
14. **Sun J, Ramos A, Chapman B, Johnnidis JB, Le L, Ho Y-J, *et al.*** Clonal dynamics of native haematopoiesis. *Nature*. 2014; 514: 322–27.
15. **Blanpain C, Mohrin M, Sotiropoulou PA, Passegué E**. DNA-Damage Response in Tissue-Specific and Cancer Stem Cells. *Cell stem cell*. 2011; 8: 16–29.
16. **Smith JNP, Calvi LM**. Concise review: Current concepts in bone marrow microenvironmental regulation of hematopoietic stem and progenitor cells. *Stem cells (Dayton, Ohio)*. 2013; 31: 1044–50.
17. **Suda T, Takubo K, Semenza GL**. Metabolic regulation of hematopoietic stem cells in the hypoxic niche. *Cell Stem Cell*. 2011; 9: 298–310.
18. **Wilson A, Trumpp A**. Bone-marrow haematopoietic-stem-cell niches. *Nat Rev Immunol*. 2006; 6: 93–106.

19. **Shackney SE, Ford SS, Wittig AB**. KINETIC-MICROARCHITECTURAL CORRELATIONS IN THE BONE MARROW OF THE MOUSE. *Cell Prolif*. 1975; 8: 505–16.

20. **Zhang J, Niu C, Ye L, Huang H, He X, Tong W-G, *et al.*** Identification of the haematopoietic stem cell niche and control of the niche size. *Nature*. 2003; 425: 836–41.

21. **Kopp H-G, Avecilla ST, Hooper AT, Rafii S**. The Bone Marrow Vascular Niche: Home of HSC Differentiation and Mobilization. *Physiology*. 2005; 20: 349–56.

22. **Nilsson SK, Johnston HM, Whitty GA, Williams B, Webb RJ, Denhardt DT, *et al.*** Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells. *Blood*. 2005; 106: 1232–39.

23. **Lyman SD, Jacobsen SEW**. c-kit Ligand and Flt3 Ligand: Stem/Progenitor Cell Factors With Overlapping Yet Distinct Activities. *Blood*. 1998; 91: 1101–34.

24. **Driessen RL, Johnston HM, Nilsson SK**. Membrane-bound stem cell factor is a key regulator in the initial lodgment of stem cells within the endosteal marrow region. *Experimental hematology*. 2003; 31: 1284–91.

25. **Kimura Y, Ding B, Imai N, Nolan DJ, Butler JM, Rafii S**. c-Kit-Mediated Functional Positioning of Stem Cells to Their Niches Is Essential for Maintenance and Regeneration of Adult Hematopoiesis. *PLOS ONE*. 2011; 6: e26918.

26. **Ding L, Saunders TL, Enikolopov G, Morrison SJ**. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nature*. 2012; 481: 457–62.

27. **Calvi LM, Adams GB, Weibrecht KW, Weber JM, Olson DP, Knight MC, *et al.*** Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature*. 2003; 425: 841–46.

28. **Cordeiro-Spinetti E, Taichman RS, Balduino A**. The Bone Marrow Endosteal Niche: How Far from the Surface? *J. Cell. Biochem.* 2015; 116: 6–11.

29. **Heissig B, Hattori K, Dias S, Friedrich M, Ferris B, Hackett NR, *et al.*** Recruitment of Stem and Progenitor Cells from the Bone Marrow Niche Requires MMP-9 Mediated Release of Kit-Ligand. *Cell*. 2002; 109: 625–37.

30. **Simsek T, Kocabas F, Zheng J, DeBerardinis RJ, Mahmoud AI, Olson EN, *et al.*** The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. *Cell Stem Cell*. 2010; 7: 380–90.

31. **Takubo K, Nagamatsu G, Kobayashi CI, Nakamura-Ishizu A, Kobayashi H, Ikeda E, *et al.*** Regulation of Glycolysis by Pdk Functions as a Metabolic Checkpoint for Cell Cycle Quiescence in Hematopoietic Stem Cells. *Cell stem cell*. 2013; 12: 49–61.

32. **Busch K, Klapproth K, Barile M, Flossdorf M, Holland-Letz T, Schlenner SM, *et al.*** Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature*. 2015; 518: 542–46.

33. **Vannini N, Girotra M, Naveiras O, Nikitin G, Campos V, Giger S, *et al.*** Specification of haematopoietic stem cell fate via modulation of mitochondrial activity. *Nat Commun*. 2016; 7: 298.

34. **Bernitz JM, Kim HS, MacArthur B, Sieburg H, Moore K**. Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell*. 2016; 167: 1296-1309.e10.

35. **Yahata T, Takanashi T, Muguruma Y, Ibrahim AA, Matsuzawa H, Uno T, *et al.*** Accumulation of oxidative DNA damage restricts the self-renewal capacity of human hematopoietic stem cells. *Blood*. 2011; 118: 2941–50.

36. **Geiger H, Haan G** de**, Florian MC**. The ageing haematopoietic stem cell compartment. *Nat Rev Immunol*. 2013; 13: 376–89.

37. **Pang WW, Price EA, Sahoo D, Beerman I, Maloney WJ, Rossi DJ, *et al.*** Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108: 20012–17.

38. **Beerman I, Bhattacharya D, Zandi S, Sigvardsson M, Weissman IL, Bryder D, *et al.*** Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. *PNAS*. 2010; 107: 5465–70.

39. **Wagner W, Horn P, Bork S, Ho AD**. Aging of hematopoietic stem cells is regulated by the stem cell niche. *Experimental gerontology*. 2008; 43: 974–80.

40. **Köhler A, Schmithorst V, Filippi M-D, Ryan MA, Daria D, Gunzer M, *et al.*** Altered cellular dynamics and endosteal location of aged early hematopoietic progenitor cells revealed by time-lapse intravital imaging in long bones. *Blood*. 2009; 114: 290–98.

41. **Rossi DJ, Bryder D, Zahn JM, Ahlenius H, Sonu R, Wagers AJ, *et al.*** Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102: 9194–99.

42. **Walter D, Lier A, Geiselhart A, Thalheimer FB, Huntscha S, Sobotta MC, *et al.*** Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature*. 2015; 520: 549–52.

43. **Beerman I, Seita J, Inlay MA, Weissman IL, Rossi DJ**. Quiescent Hematopoietic Stem Cells Accumulate DNA Damage during Aging that Is Repaired upon Entry into Cell Cycle. *Cell stem cell*. 2014; 15: 37–50.

44. **Norddahl GL, Pronk CJ, Wahlestedt M, Sten G, Nygren JM, Ugale A, *et al.*** Accumulating mitochondrial DNA mutations drive premature hematopoietic aging phenotypes distinct from physiological stem cell aging. *Cell Stem Cell*. 2011; 8: 499–510.

45. **Dykstra B, Olthof S, Schreuder J, Ritsema M, Haan G** de. Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. *Journal of Experimental Medicine*. 2011; 208: 2691–703.

46. **Busque L, Mio R, Mattioli J, Brais E, Blais N, Lalonde Y, *et al.*** Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood*. 1996; 88: 59–65.

47. **Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z, *et al.*** Recurrent somatic *TET2* mutations in normal elderly individuals with clonal hematopoiesis. *ng*. 2012; 44: 1179–81.

48. **Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, *et al.*** The Origin and Evolution of Mutations in Acute Myeloid Leukemia. *Cell*. 2012; 150: 264–78.

49. **Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, *et al.*** Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature*. 2014; 506: 328–33.

50. **Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, *et al.*** Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nature medicine*. 2014; 20: 1472–78.

51. **Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, *et al.*** Age-related clonal hematopoiesis associated with adverse outcomes. *The New England journal of medicine*. 2014; 371: 2488–98.

52. **Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, *et al.*** Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *The New England journal of medicine*. 2014; 371: 2477–87.

53. **Steensma DP**. Clinical consequences of clonal hematopoiesis of indeterminate potential. *Blood advances*. 2018; 2: 3404–10.

54. **Arends CM, Galan-Sousa J, Hoyer K, Chan W, Jäger M, Yoshida K, *et al.*** Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. *Leukemia*. 2018; 32: 1908–19.

55. **Nowell PC**. The clonal evolution of tumor cell populations. *Science (New York, N.Y.)*. 1976; 194: 23–28.

56. **Greaves M, Maley CC**. Clonal evolution in cancer. *Nature*. 2012; 481: 306–13.

57. **Reya T, Morrison SJ, Clarke MF, Weissman IL**. Stem cells, cancer, and cancer stem cells. *Nature*. 2001; 414: 105–11.

58. **George AA, Franklin J, Kerkof K, Shah AJ, Price M, Tsark E, *et al.*** Detection of leukemic cells in the CD34(+)CD38(-) bone marrow progenitor population in children with acute lymphoblastic leukemia. *Blood*. 2001; 97: 3925–30.

59. **Miyamoto T, Weissman IL, Akashi K**. AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97: 7521–26.

60. **Jan M, Snyder TM, Corces-Zimmerman MR, Vyas P, Weissman IL, Quake SR, *et al.*** Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Science translational medicine*. 2012; 4: 149ra118.

61. **Corces-Zimmerman MR, Hong W-J, Weissman IL, Medeiros BC, Majeti R**. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *PNAS*. 2014; 111: 2548–53.

62. Krebs - Leukämien. https://www.krebsdaten.de/Krebs/DE/Content/Krebsarten/Leukaemien/leukaemien_node.html (accessed October 14, 2020).

63. **Kraywinkel K, Spix C**. Epidemiologie akuter Leukämien in Deutschland. *Onkologe*. 2017; 23: 499–503.

64. **Döhner H, Weisdorf DJ, Bloomfield CD**. Acute Myeloid Leukemia. *The New England journal of medicine*. 2015; 373: 1136–52.

65. **Kouchkovsky I** de**, Abdul-Hay M**. 'Acute myeloid leukemia: a comprehensive review and 2016 update'. *Blood cancer journal*. 2016; 6: e441.

66. **Döhner H, Estey EH, Amadori S, Appelbaum FR, Büchner T, Burnett AK, *et al.*** Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood*. 2010; 115: 453–74.

67. **Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, *et al.*** The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016; 127: 2391–405.

68. **Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, *et al.*** Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017; 129: 424–47.

69. **Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, *et al.*** Genomic Classification and Prognosis in Acute Myeloid Leukemia. *The New England journal of medicine*. 2016; 374: 2209–21.

70. **Patel JP, Gönen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, *et al.*** Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *The New England journal of medicine*. 2012; 366: 1079–89.

71. **Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson AG, *et al.*** Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *The New England journal of medicine*. 2013; 368: 2059–74.

72. **Levis M**. Midostaurin approved for FLT3-mutated AML. *Blood*. 2017; 129: 3403–06.

73. **Stone RM, Mandrekar SJ, Sanford BL, Laumann K, Geyer S, Bloomfield CD, *et al.*** Midostaurin plus Chemotherapy for Acute Myeloid Leukemia with a FLT3 Mutation. *The New England journal of medicine*. 2017; 377: 454–64.

74. **Schnittger S, Schoch C, Dugas M, Kern W, Staib P, Wuchter C, *et al.*** Analysis of FLT3 length mutations in 1003 patients with acute myeloid leukemia: correlation to cytogenetics, FAB subtype, and prognosis in the AMLCG study and usefulness as a marker for the detection of minimal residual disease. *Blood*. 2002; 100: 59–66.

75. **Kottaridis PD, Gale RE, Frew ME, Harrison G, Langabeer SE, Belton AA, *et al.*** The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood*. 2001; 98: 1752–59.

76. **Erickson P, Gao J, Chang KS, Look T, Whisenant E, Raimondi S, *et al.*** Identification of breakpoints in t(8;21) acute myelogenous leukemia and isolation of a fusion transcript, AML1/ETO, with similarity to Drosophila segmentation gene, runt. *Blood*. 1992; 80: 1825–31.

77. **Liu P, Tarlé SA, Hajra A, Claxton DF, Marlton P, Freedman M, *et al.*** Fusion between transcription factor CBF beta/PEBP2 beta and a myosin heavy chain in acute myeloid leukemia. *Science (New York, N.Y.)*. 1993; 261: 1041–44.

78. **Wang S, Wang Q, Crute BE, Melnikova IN, Keller SR, Speck NA**. Cloning and characterization of subunits of the T-cell receptor and murine leukemia virus enhancer core-binding factor. *Molecular and cellular biology*. 1993; 13: 3324–39.

79. **Wang Q, Stacy T, Binder M, Marin-Padilla M, Sharpe AH, Speck NA**. Disruption of the Cbfa2 gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proceedings of the National Academy of Sciences of the United States of America*. 1996; 93: 3444–49.

80. **Speck NA, Gilliland DG**. Core-binding factors in haematopoiesis and leukaemia. *Nature reviews. Cancer*. 2002; 2: 502–13.

81. **Nishii K, Usui E, Katayama N, Lorenzo F, Nakase K, Kobayashi T, *et al.*** Characteristics of t(8;21) acute myeloid leukemia (AML) with additional chromosomal abnormality: concomitant trisomy 4 may constitute a distinctive subtype of t(8;21) AML. *Leukemia*. 2003; 17: 731–37.

82. **Marcucci G, Mrózek K, Ruppert AS, Maharry K, Kolitz JE, Moore JO, *et al.*** Prognostic factors and outcome of core binding factor acute myeloid leukemia patients with t(8;21) differ from those of patients with inv(16): a Cancer and Leukemia Group B study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2005; 23: 5705–17.

83. **Bannister AJ, Kouzarides T**. Regulation of chromatin by histone modifications. *Cell research*. 2011; 21: 381–95.

84. **Antequera F, Bird A**. Number of CpG islands and genes in human and mouse. *Proceedings of the National Academy of Sciences of the United States of America*. 1993; 90: 11995–99.

85. **Okano M, Xie S, Li E**. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature genetics*. 1998; 19: 219–20.

86. **Okano M, Bell DW, Haber DA, Li E**. DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell*. 1999; 99: 247–57.

87. **Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, *et al.*** Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, N.Y.)*. 2009; 324: 930–35.

88. **Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y**. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*. 2010; 466: 1129.

89. **Park I-k, Qian D, Kiel M, Becker MW, Pihalja M, Weissman IL, *et al.*** Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature*. 2003; 423: 302–05.

90. **Bröske A-M, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M, *et al.*** DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *ng*. 2009; 41: 1207–15.

91. **Bock C, Beerman I, Lien W-H, Smith ZD, Gu H, Boyle P, *et al.*** DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. *Molecular Cell*. 2012; 47: 633–47.

92. **Shih AH, Abdel-Wahab O, Patel JP, Levine RL**. The role of mutations in epigenetic regulators in myeloid malignancies. *nrc*. 2012; 12: 599–612.

93. **Yang L, Rau R, Goodell MA**. DNMT3A in haematological malignancies. *Nature reviews. Cancer*. 2015; 15: 152–65.

94. **Brunetti L, Gundry MC, Goodell MA**. DNMT3A in Leukemia. *Cold Spring Harbor perspectives in medicine*. 2017; 7.

95. **Chen T, Ueda Y, Xie S, Li E**. A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *The Journal of biological chemistry*. 2002; 277: 38746–54.

96. **Bird A**. DNA methylation patterns and epigenetic memory. *Genes & development*. 2002; 16: 6–21.

97. **Challen GA, Sun D, Jeong M, Luo M, Jelinek J, Berg JS, *et al.*** Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature genetics*. 2012; 44: 23–31.

98. **Jeong M, Sun D, Luo M, Huang Y, Challen GA, Rodriguez B, *et al.*** Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat Genet*. 2014; 46: 17–23.

99. **Jeong M, Park HJ, Celik H, Ostrander EL, Reyes JM, Guzman A, *et al.*** Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo. *Cell Reports*. 2018; 23: 1–10.

100. **Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, *et al.*** DNMT3A mutations in acute myeloid leukemia. *The New England journal of medicine*. 2010; 363: 2424–33.

101. **Yan X-J, Xu J, Gu Z-H, Pan C-M, Lu G, Shen Y, *et al.*** Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet*. 2011; 43: 309–15.

102. **Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, *et al.*** The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012; 2: 401–04.

103. **Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, *et al.*** Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science signaling*. 2013; 6: pl1.

104. **Russler-Germain DA, Spencer DH, Young MA, Lamprecht TL, Miller CA, Fulton R, *et al.*** The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by Blocking Its Ability to Form Active Tetramers. *Cancer cell*. 2014; 25: 442–54.

105. **Sato H, Wheat JC, Steidl U, Ito K**. DNMT3A and TET2 in the Pre-Leukemic Phase of Hematopoietic Disorders. *Front. Oncol.* 2016; 6: 187.

106. **Lorsbach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, Downing JR**. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia*. 2003; 17: 637–41.

107. **Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, *et al.*** Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science (New York, N.Y.)*. 2011; 333: 1300–03.

108. **Valinluck V, Sowers LC**. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer research*. 2007; 67: 946–50.

109. **Bowman RL, Levine RL**. TET2 in Normal and Malignant Hematopoiesis. *Cold Spring Harbor perspectives in medicine*. 2017; 7.

110. **Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Massé A, *et al.*** Mutation in TET2 in myeloid cancers. *The New England journal of medicine*. 2009; 360: 2289–301.

111. **Langemeijer SMC, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M, *et al.*** Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet*. 2009; 41: 838–42.

112. **Abdel-Wahab O, Mullally A, Hedvat C, Garcia-Manero G, Patel J, Wadleigh M, *et al.*** Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood*. 2009; 114: 144–47.

113. **Nibourel O, Kosmider O, Cheok M, Boissel N, Renneville A, Philippe N, *et al.*** Incidence and prognostic value of TET2 alterations in de novo acute myeloid leukemia achieving complete remission. *Blood*. 2010; 116: 1132–35.

114. **Metzeler KH, Maharry K, Radmacher MD, Mrózek K, Margeson D, Becker H, *et al.*** TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2011; 29: 1373–81.

115. **Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS, *et al.*** Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant *TET2*. *Nature*. 2010; 468: 839.

116. **Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C, *et al.*** Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer Cell*. 2011; 20: 11–24.

117. **Quivoron C, Couronne L, Della Valle V, Lopez CK, Plo I, Wagner-Ballon O, *et al.*** TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer cell*. 2011; 20: 25–38.

118. **Ko M, Bandukwala HS, an J, Lamperti ED, Thompson EC, Hastie R, *et al.*** Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proceedings of the National Academy of Sciences of the United States of America*. 2011; 108: 14566–71.

119. **Kunimoto H, McKenney AS, Meydan C, Shank K, Nazir A, Rapaport F, *et al.*** Aid is a key regulator of myeloid/erythroid differentiation and DNA methylation in hematopoietic stem/progenitor cells. *Blood*. 2017; 129: 1779–90.

120. **Milne TA, Sinclair DA, Brock HW**. The Additional sex combs gene of Drosophila is required for activation and repression of homeotic loci, and interacts specifically with Polycomb and super sex combs. *Molecular & general genetics : MGG*. 1999; 261: 753–61.

121. **Fisher C**. A human homolog of Additional sex combs, ADDITIONAL SEX COMBS-LIKE 1, maps to chromosome 20q11. *Gene*. 2003; 306: 115–26.

122. **Fisher CL, Pineault N, Brookes C, Helgason CD, Ohta H, Bodner C, *et al.*** Loss-of-function Additional sex combs like 1 mutations disrupt hematopoiesis but do not cause severe myelodysplasia or leukemia. *Blood*. 2010; 115: 38–46.

123. **Wang J, Li Z, He Y, Pan F, Chen S, Rhodes S, *et al.*** Loss of Asxl1 leads to myelodysplastic syndrome-like disease in mice. *Blood*. 2014; 123: 541–53.

124. **Abdel-Wahab O, Adli M, LaFave LM, Gao J, Hricik T, Shih AH, *et al.*** ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. *Cancer cell*. 2012; 22: 180–93.

125. **Nagase R, Inoue D, Pastore A, Fujino T, Hou H-A, Yamasaki N, *et al.*** Expression of mutant Asxl1 perturbs hematopoiesis and promotes susceptibility to leukemic transformation. *Journal of Experimental Medicine*. 2018; 215: 1729–47.

126. **Thol F, Friesen I, Damm F, Yun H, Weissinger EM, Krauter J, *et al.*** Prognostic significance of ASXL1 mutations in patients with myelodysplastic syndromes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2011; 29: 2499–506.

127. **Boultwood J, Perry J, Pellagatti A, Fernandez-Mercado M, Fernandez-Santamaria C, Calasanz MJ, *et al.*** Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. *leu*. 2010; 24: 1062–65.

128. **Carbuccia N, Murati A, Trouplin V, Brecqueville M, Adélaïde J, Rey J, *et al.*** Mutations of ASXL1 gene in myeloproliferative neoplasms. *leu*. 2009; 23: 2183–86.

129. **Schnittger S, Eder C, Jeromin S, Alpermann T, Fasan A, Grossmann V, *et al.*** *ASXL1* exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome. *leu*. 2013; 27: 82–91.

130. **Rocquain J, Carbuccia N, Trouplin V, Raynaud S, Murati A, Nezri M, *et al.*** Combined mutations of ASXL1, CBL, FLT3, IDH1, IDH2, JAK2, KRAS, NPM1, NRAS, RUNX1, TET2 and WT1 genes in myelodysplastic syndromes and acute myeloid leukemias. *BMC cancer*. 2010; 10: 401.

131. **Metzeler KH, Becker H, Maharry K, Radmacher MD, Kohlschmidt J, Mrózek K, *et al.*** ASXL1 mutations identify a high-risk subgroup of older patients with primary cytogenetically normal AML within the ELN Favorable genetic category. *Blood*. 2011; 118: 6920–29.

132. **Krauth M-T, Eder C, Alpermann T, Bacher U, Nadarajah N, Kern W, *et al.*** High number of additional genetic lesions in acute myeloid leukemia with t(8;21)/RUNX1-RUNX1T1: frequency and impact on clinical outcome. *leu*. 2014; 28: 1449–58.

133. **Inoue D, Kitaura J, Togami K, Nishimura K, Enomoto Y, Uchida T, *et al.*** Myelodysplastic syndromes are induced by histone methylation–altering ASXL1 mutations. *The Journal of clinical investigation*. 2013; 123: 4627–40.

134. **Yang H, Kurtenbach S, Guo Y, Lohse I, Durante MA, Li J, *et al.*** Gain of function of ASXL1 truncating protein in the pathogenesis of myeloid malignancies. *Blood*. 2018; 131: 328–41.

135. **Hsu Y-C, Chiu Y-C, Lin C-C, Kuo Y-Y, Hou H-A, Tzeng Y-S, *et al.*** The distinct biological implications of Asxl1 mutation and its roles in leukemogenesis revealed by a knock-in mouse model. *Journal of hematology & oncology*. 2017; 10: 139.

136. **Uni M, Masamoto Y, Sato T, Kamikubo Y, Arai S, Hara E, *et al.*** Modeling ASXL1 mutation revealed impaired hematopoiesis caused by derepression of p16Ink4a through aberrant PRC1-mediated histone modification. *leu*. 2019; 33: 191–204.

137. **Wu X, Bekker-Jensen IH, Christensen J, Rasmussen KD, Sidoli S, Qi Y, *et al.*** Tumor suppressor ASXL1 is essential for the activation of INK4B expression in response to oncogene activity and anti-proliferative signals. *Cell research*. 2015; 25: 1205–18.

138. **Ding Q, Regan SN, Xia Y, Oostrom LA, Cowan CA, Musunuru K**. Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell*. 2013; 12: 393–94.

139. **Heckl D, Kowalczyk MS, Yudovich D, Belizaire R, Puram RV, McConkey ME, *et al.*** Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nature biotechnology*. 2014; 32: 941–46.

140. Cloud-Based Informatics Platform for Life Sciences R&D | Benchling: Biology Software. https://benchling.com (accessed November 24, 2020).

141. **Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, *et al.*** Primer3--new capabilities and interfaces. *Nucleic Acids Research*. 2012; 40: e115.

142. **Stemmer M, Thumberger T, Del Sol Keyer M, Wittbrodt J, Mateo JL**. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PloS one*. 2015; 10: e0124633.

143. **Moreno-Mateos MA, Vejnar CE, Beaudoin J-D, Fernandez JP, Mis EK, Khokha MK, *et al.*** CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting *in vivo*. *nmeth*. 2015; 12: 982–88.

144. **Brunetti L, Gundry MC, Kitano A, Nakada D, Goodell MA**. Highly Efficient Gene Disruption of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. *Journal of visualized experiments : JoVE*. 2018.

145. **Hogge DE, Lansdorp PM, Reid D, Gerhard B, Eaves CJ**. Enhanced detection, maintenance, and differentiation of primitive human hematopoietic cells in cultures containing murine fibroblasts engineered to produce human steel factor, interleukin-3, and granulocyte colony-stimulating factor. *Blood*. 1996; 88: 3765–73.

146. **Sutherland HJ, Eaves CJ, Lansdorp PM, Thacker JD, de Hogge**. Differential regulation of primitive human hematopoietic cells in long- term cultures maintained on genetically engineered murine stromal cells. *Blood*. 1991; 78: 666–72.

147. **Brinkman EK, Chen T, Amendola M, van Steensel B**. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Research*. 2014; 42: e168.

148. **Hsiau T, Conant D, Rossi N, Maures T, Waite K, Yang J, *et al.*** Inference of CRISPR Edits from Sanger Trace Data; 2018.

149. GRCh37 - hg19 - Genome - Assembly - NCBI. https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/ (accessed December 6, 2020).

150. **Li H**. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM; 2013.

151. **Tothova Z, Krill-Burger JM, Popova KD, Landers CC, Sievers QL, Yudovich D, *et al.*** Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia. *Cell stem cell*. 2017; 21: 547-555.e8.

152. **Benjamini Y, Hochberg Y**. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995; 57: 289–300.

153. **Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, *et al.*** A universal SNP and small-indel variant caller using deep neural networks. *Nature biotechnology*. 2018; 36: 983–87.

154. **Shannon CE**. A Mathematical Theory of Communication. *Bell System Technical Journal*. 1948; 27: 379–423.

155. **Hausser J, Strimmer K**. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. 2009.

156. **Bradley RA, Terry ME**. Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*. 1952; 39: 324.

157. **Turner H, Firth D**. Bradley-Terry Models in R : The BradleyTerry2 Package. *J. Stat. Soft.* 2012; 48.

158. **Mann HB, Whitney DR**. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann. Math. Statist.* 1947; 18: 50–60.

159. **SHAPIRO SS, WILK MB**. An analysis of variance test for normality (complete samples). *Biometrika*. 1965; 52: 591–611.

160. **Holm S**. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*. 1979; 6: 65–70.

161. **Göknur Giner, Gordon K. Smyth**. statmod: Probability Calculations for the Inverse Gaussian Distribution. *The R Journal*. 2016; 8: 339–51.

162. **Baldwin T, Sakthianandeswaren A, Curtis JM, Kumar B, Smyth GK, Foote SJ, *et al.*** Wound healing response is a major contributor to the severity of cutaneous leishmaniasis in the ear model of infection. *Parasite Immunology*. 2007; 29: 501–13.

163. **Christen F, Hoyer K, Yoshida K, Hou H-A, Waldhueter N, Heuser M, *et al.*** Genomic landscape and clonal evolution of acute myeloid leukemia with t(8;21): an international study on 331 patients. *Blood*. 2019; 133: 1140–51.

164. **Dai Y-J, Wang Y-Y, Huang J-Y, Xia L, Shi X-D, Xu J, *et al.*** Conditional knockin of Dnmt3a R878H initiates acute myeloid leukemia with mTOR pathway involvement. *Proceedings of the National Academy of Sciences of the United States of America*. 2017; 114: 5237–42.

165. **Guryanova OA, Lieu YK, Garrett-Bakelman FE, Spitzer B, Glass JL, Shank K, *et al.*** *Dnmt3a* regulates myeloproliferation and liver-specific expansion of hematopoietic stem and progenitor cells. *leu*. 2016; 30: 1133–42.

166. **Koya J, Kataoka K, Sato T, Bando M, Kato Y, Tsuruta-Kishino T, *et al.*** DNMT3A R882 mutants interact with polycomb proteins to block haematopoietic stem and leukaemic cell differentiation. *Nature Communications*. 2016; 7: 10924.

167. **Mayle A, Yang L, Rodriguez B, Zhou T, Chang E, Curry CV, *et al.*** Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. *Blood*. 2015; 125: 629–38.

168. **Cimmino L, Dolgalev I, Wang Y, Yoshimi A, Martin GH, Wang J, *et al.*** Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. *Cell*. 2017; 170: 1079-1095.e20.

169. **Ito K, Lee J, Chrysanthou S, Zhao Y, Josephs K, Sato H, *et al.*** Non-catalytic Roles of Tet2 Are Essential to Regulate Hematopoietic Stem and Progenitor Cell Homeostasis. *Cell Reports*. 2019; 28: 2480-2490.e4.

170. **Pan F, Wingo TS, Zhao Z, Gao R, Makishima H, Qu G, *et al.*** Tet2 loss leads to hypermutagenicity in haematopoietic stem/progenitor cells. *Nature Communications*. 2017; 8: 15102.

171. **Li Z, Zhang P, Yan A, Guo Z, Ban Y, Li J, *et al.*** ASXL1 interacts with the cohesin complex to maintain chromatid separation and gene expression for normal hematopoiesis. *Science advances*. 2017; 3: e1601602.

172. **Rutherford T, Clegg JB, Higgs DR, Jones RW, Thompson J, Weatherall DJ**. Embryonic erythroid differentiation in the human leukemic cell line K562. *Proceedings of the National Academy of Sciences of the United States of America*. 1981; 78: 348–52.

173. **Selden JR, Emanuel BS, Wang E, Cannizzaro L, Palumbo A, Erikson J, *et al.*** Amplified C lambda and c-abl genes are on the same marker chromosome in K562 leukemia cells. *Proceedings of the National Academy of Sciences of the United States of America*. 1983; 80: 7289–92.

174. **Naumann S, Reutzel D, Speicher M, Decker H-J**. Complete karyotype characterization of the K562 cell line by combined application of G-banding, multiplex-fluorescence in situ hybridization, fluorescence in situ hybridization, and comparative genomic hybridization. *Leukemia Research*. 2001; 25: 313–22.

175. **Kronstein-Wiedemann R, Tonn T**. Colony Formation: An Assay of Hematopoietic Progenitor Cells. *Methods in molecular biology (Clifton, N.J.)*. 2019; 2017: 29–40.

176. **Miller CL, Lai B**. Human and mouse hematopoietic colony-forming cell assays. *Methods in molecular biology (Clifton, N.J.)*. 2005; 290: 71–89.

177. **Mayani H, Lansdorp PM**. Biology of human umbilical cord blood-derived hematopoietic stem/progenitor cells. *Stem cells (Dayton, Ohio)*. 1998; 16: 153–65.

178. **Cardoso AA, Li ML, Batard P, Hatzfeld A, Brown EL, Levesque JP, *et al.*** Release from quiescence of CD34+ CD38- human umbilical cord blood cells reveals their potentiality to engraft adults. *Proceedings of the National Academy of Sciences of the United States of America*. 1993; 90: 8707–11.

179. **Lansdorp PM, Dragowska W, Mayani H**. Ontogeny-related changes in proliferative potential of human hematopoietic cells. *The Journal of experimental medicine*. 1993; 178: 787–91.

180. **Hao QL, Shah AJ, Thiemann FT, Smogorzewska EM, Crooks GM**. A functional comparison of CD34 + CD38- cells in cord blood and bone marrow. *Blood*. 1995; 86: 3745–53.

181. **Carow CE, Hangoc G, Broxmeyer HE**. Human multipotential progenitor cells (CFU-GEMM) have extensive replating capacity for secondary CFU-GEMM: an effect enhanced by cord blood plasma. *Blood*. 1993; 81: 942–49.

182. **Koestenbauer S, Zisch A, Dohr G, Zech NH**. Protocols for hematopoietic stem cell expansion from umbilical cord blood. *Cell transplantation*. 2009; 18: 1059–68.

183. **Piacibello W, Sanavio F, Garetto L, Severino A, Bergandi D, Ferrario J, *et al.*** Extensive Amplification and Self-Renewal of Human Primitive Hematopoietic Stem Cells From Cord Blood. *Blood*. 1997; 89: 2644–53.

184. **Angeli S** de**, Baiguera S, Del Pup L, Pavan E, Gajo GB, Di Liddo R, *et al.*** Middle-term expansion of hematopoietic cord blood cells with new human stromal cell line feeder-layers and different cytokine cocktails. *International journal of molecular medicine*. 2009; 24: 837–45.

185. **Choi Y-S, Noh S-E, Lim S-M, Kim D-I**. Optimization of ex vivo hematopoietic stem cell expansion in intermittent dynamic cultures. *Biotechnology letters*. 2010; 32: 1969–75.

186. **Gundry MC, Brunetti L, Lin A, Mayle AE, Kitano A, Wagner D, *et al.*** Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. *Cell Reports*. 2016; 17: 1453–61.

187. **Genovese P, Schiroli G, Escobar G, Di Tomaso T, Firrito C, Calabria A, *et al.*** Targeted genome editing in human repopulating haematopoietic stem cells. *Nature*. 2014; 510: 235–40.

188. **Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD**. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics*. 2010; 11: 636–46.

189. **Joung JK, Sander JD**. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol*. 2013; 14: 49–55.

190. **Sander JD, Joung JK**. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology*. 2014; 32: 347–55.

191. **Mandal PK, Ferreira LMR, Collins R, Meissner TB, Boutwell CL, Friesen M, *et al.*** Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. *Cell stem cell*. 2014; 15: 643–52.

192. **Schumann K, Lin S, Boyer E, Simeonov DR, Subramaniam M, Gate RE, *et al.*** Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proceedings of the National Academy of Sciences of the United States of America*. 2015; 112: 10437–42.

193. **DeWitt MA, Magis W, Bray NL, Wang T, Berman JR, Urbinati F, *et al.*** Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Science translational medicine*. 2016; 8: 360ra134.

194. **Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE**. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology*. 2016; 34: 339–44.

195. **Liang X, Potter J, Kumar S, Ravinder N, Chesnut JD**. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *Journal of biotechnology*. 2017; 241: 136–46.

196. **Garneau NL, Wilusz J, Wilusz CJ**. The highways and byways of mRNA decay. *Nature reviews. Molecular cell biology*. 2007; 8: 113–26.

197. **Agathocleous M, Meacham CE, Burgess RJ, Piskounova E, Zhao Z, Crane GM, *et al.*** Ascorbate regulates haematopoietic stem cell function and leukaemogenesis. *Nature*. 2017; 549: 476–81.

198. **Cuddihy AR, O'Connell MJ**. Cell-cycle responses to DNA damage in G2. In: International Review of Cytology: Elsevier; 2003. p. 99–140.

199. **Buscarlet M, Provost S, Zada YF, Bourgoin V, Mollica L, Dubé M-P, *et al.*** Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A. *Blood*. 2018; 132: 277–80.

200. **Asada S, Fujino T, Goyama S, Kitamura T**. The role of ASXL1 in hematopoiesis and myeloid malignancies. *Cellular and Molecular Life Sciences*. 2019; 76: 2511–23.

201. **SanMiguel JM, Loberg M, Heuer S, Stearns T, Young K, Trowbridge J**. Cell-Extrinsic Stressors from the Aging Bone Marrow (BM) Microenvironment Promote Dnmt3a-Mutant Clonal Hematopoiesis. *Blood*. 2019; 134: 5.

202. **Dawoud AAZ, Tapper WJ, Cross NCP**. Clonal myelopoiesis in the UK Biobank cohort: ASXL1 mutations are strongly associated with smoking. *leu*. 2020; 34: 2660–72.

203. **Anteneh H, Fang J, Song J**. Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. *Nat Commun*. 2020; 11: 2294.

204. **Rasmussen KD, Jia G, Johansen JV, Pedersen MT, Rapin N, Bagger FO, *et al.*** Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes & development*. 2015; 29: 910–22.

205. **Abdel-Wahab O, Gao J, Adli M, Dey A, Trimarchi T, Chung YR, *et al.*** Deletion of Asxl1 results in myelodysplasia and severe developmental defects in vivo. *The Journal of experimental medicine*. 2013; 210: 2641–59.

206. **Zhang X, Su J, Jeong M, Ko M, Huang Y, Park HJ, *et al.*** DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells. *Nature genetics*. 2016; 48: 1014–23.

207. **Liu S, Shen T, Huynh L, Klisovic MI, Rush LJ, Ford JL, *et al.*** Interplay of RUNX1/MTG8 and DNA methyltransferase 1 in acute myeloid leukemia. *Cancer research*. 2005; 65: 1277–84.

208. **Fazi F, Zardo G, Gelmetti V, Travaglini L, Ciolfi A, Di Croce L, *et al.*** Heterochromatic gene repression of the retinoic acid pathway in acute myeloid leukemia. *Blood*. 2007; 109: 4432–40.

209. **Gao XN, Yan F, Lin J, Gao L, Lu XL, Wei SC, *et al.*** AML1/ETO cooperates with HIF1α to promote leukemogenesis through DNMT3a transactivation. *leu*. 2015; 29: 1730–40.

210. **Jongen-Lavrencic M, Grob T, Hanekamp D, Kavelaars FG, Al Hinai A, Zeilemaker A, *et al.*** Molecular Minimal Residual Disease in Acute Myeloid Leukemia. *The New England journal of medicine*. 2018; 378: 1189–99.

211. **Celik H, Mallaney C, Kothari A, Ostrander EL, Eultgen E, Martens A, *et al.*** Enforced differentiation of Dnmt3a-null bone marrow leads to failure with c-Kit mutations driving leukemic transformation. *Blood*. 2015; 125: 619–28.

212. **Mei M, Aldoss I, Marcucci G, Pullarkat V**. Hypomethylating agents in combination with venetoclax for acute myeloid leukemia: Update on clinical trial data and practical considerations for use. *American journal of hematology*. 2019; 94: 358–62.

213. **Oo ZM, Illendula A, Grembecka J, Schmidt C, Zhou Y, Esain V, *et al.*** A tool compound targeting the core binding factor Runt domain to disrupt binding to CBFβ in leukemic cells. *Leukemia & lymphoma*. 2018; 59: 2188–200.

214. **Schanda J, Lee C-W, Wohlan K, Müller-Kuller U, Kunkel H, Coco IQ-L, *et al.*** Suppression of RUNX1/ETO oncogenic activity by a small molecule inhibitor of tetramerization. *Haematologica*. 2017; 102: e170-e174.

8 Appendix

8.1 Supplemental Data

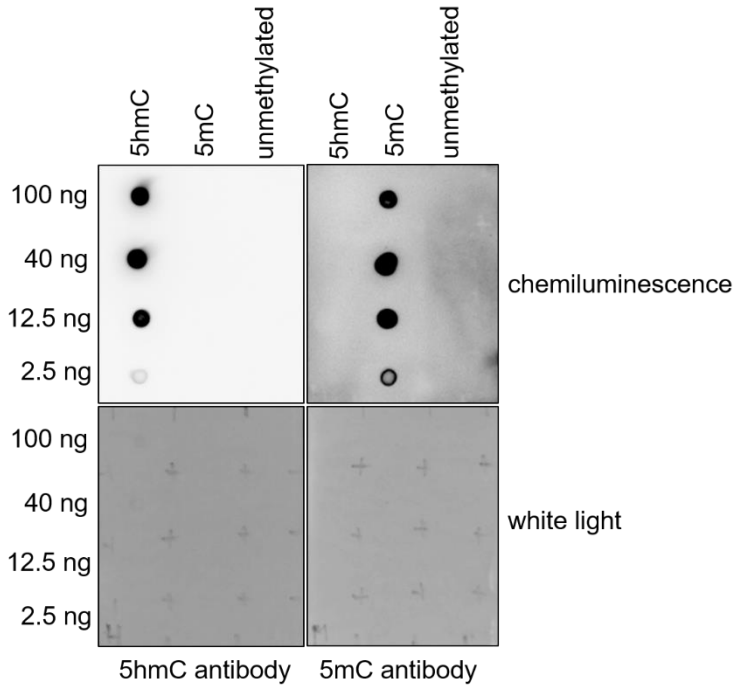


Figure 8.1 Specificity test for 5hmC and 5mC antibodies. PCR products containing 5-hydroxymethylated (5hmC), 5-methylated, and unmethylated cytosine (Methylated DNA Standard Kit, Active Motif) were spotted on two nylon membranes. The upper panels show the chemiluminescence signal of the HRP-conjugated antibodies. The lower panels show the white light exposure with the position marks on the membrane.

Table 8.1 List of genes and target regions included in the customized version of the TruSightMyeloid panel.

Gene	Target region (exon)	Gene	Target region (exon)
ABL1	4–6	*KDM6A*	Full
ASXL1	12	*KIT*	2, 8–10, 13, 17
ASXL2	12, 13 (exon 13 CDS only)	*KLHDC8A*	Full
BCOR	Full	*KMT2A*	1, 3, 5-8, 27
BCORL1	Full	*KRAS*	2, 3
BRAF	15	*MGA*	3-8
BRCC3	Full	*MPL*	4, 10, 12
CALR	9	*MYB*	Full
CBL	8	*MYC*	1-3
CBLB	9, 10	*MYD88*	3–5
CBLC	9, 10	*NOTCH1*	26-28, 34
CCND1	4, 5	*NPM1*	11
CCND2	4, 5	*NRAS*	2, 3
CDKN2A	Full	*PDGFRA*	12, 14, 18
CSF3R	16, 17	*PHF6*	Full
CUX1	Full	*PPM1D*	5, 6 (exon 6 CDS only)
DHX15	3, 4, 5	*PTEN*	5, 7
DNMT3A	Full	*PTPN11*	3, 13
ETV6/TEL	Full	*RAD21*	Full
EZH2	Full	*RUNX1*	Full
FBXW7	9, 10, 11	*SETBP1*	4 (partial)
FLT3	14, 15, 20	*SF3B1*	14–16
GATA1	2	*SMC1A*	2, 11, 16, 17
GATA2	2–6	*SMC3*	10, 13, 19, 23, 25, 28
GIGYF2	22	*SRSF2*	1
GNAS	8	*STAG2*	full
GNB1	2, 3, 4, 5	*STAT3*	20, 21
HRAS	2, 3	*TET2*	3–11
IDH1	4	*TP53*	2–11
IDH2	4	*U2AF1*	2, 6
IKZF1	Full	*WT1*	7, 9
JAK2	12, 13, 14	*ZRSR2*	Full
JAK3	13	*ZBTB7A*	Full

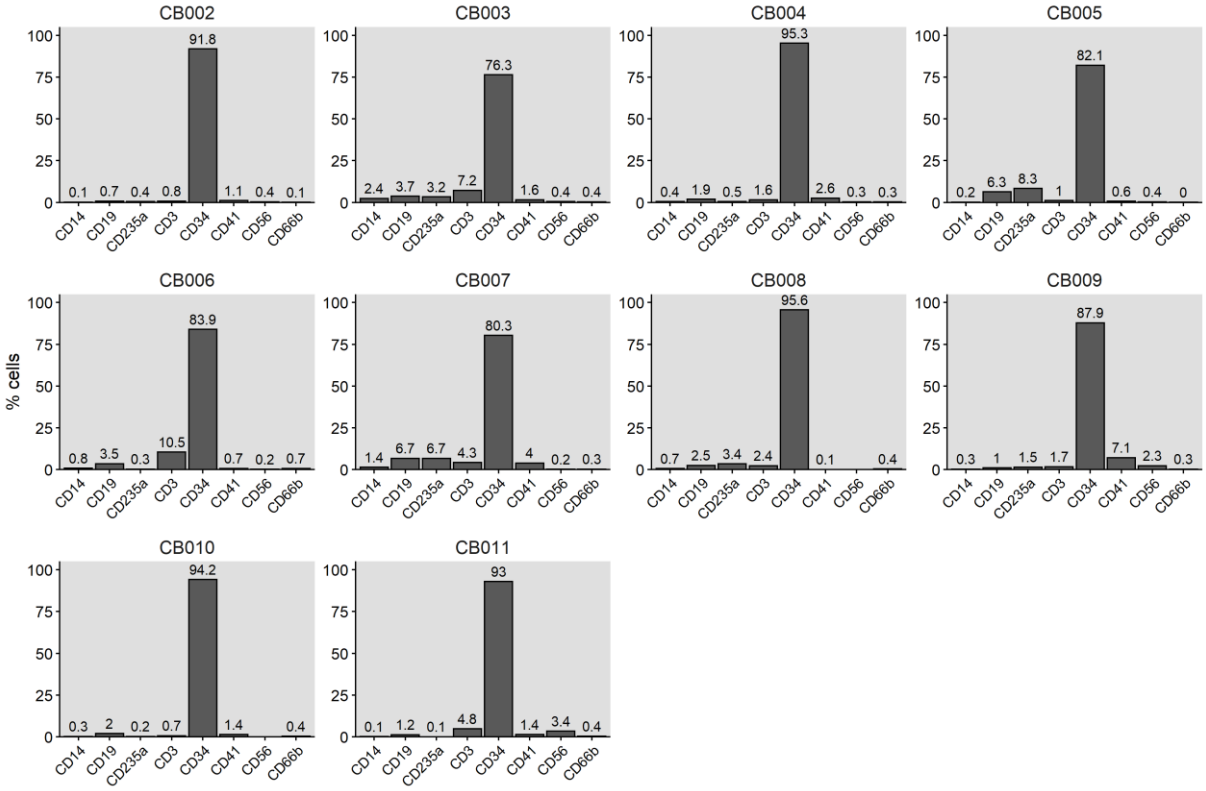
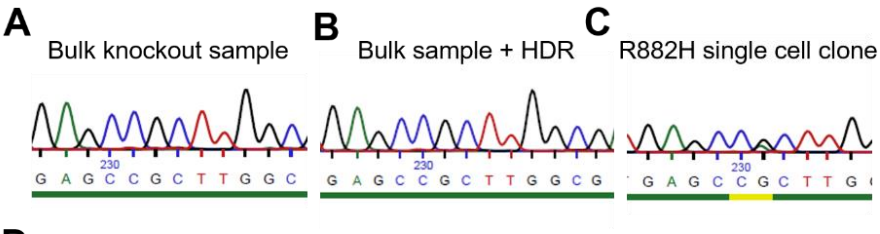


Figure 8.2 Flow cytometry analysis of CD34⁺ enriched cells. Mononuclear cells were isolated from freshly collected umbilical cord blood samples and further enriched for CD34⁺ cells by magnetic separation (MACS). After 2 days of culture in expansion medium, the cells were analyzed via flow cytometry before transfection. CD14 monocytes, CD19 B cells, CD235a erythrocytes, CD3 T cells, CD34 progenitor cells, CD41 megakaryocytes, CD56 NK cells, CD66b granulocytes. The numbers above the bars indicate the frequency of the respective population in %.



D

	Number of clones	Efficiency
sequenced	48	
Carrying mutations	9	18.75%
R882 deletion	3	6.25%
R882H SNV	1	2%

Figure 8.3 Sanger sequencing of sorted K-562 single cell clones. A – C) Sanger sequencing tracks of bulk sorted samples and single cell clone. D) Statistics of mutations found via Sanger sequencing of single cell clones.

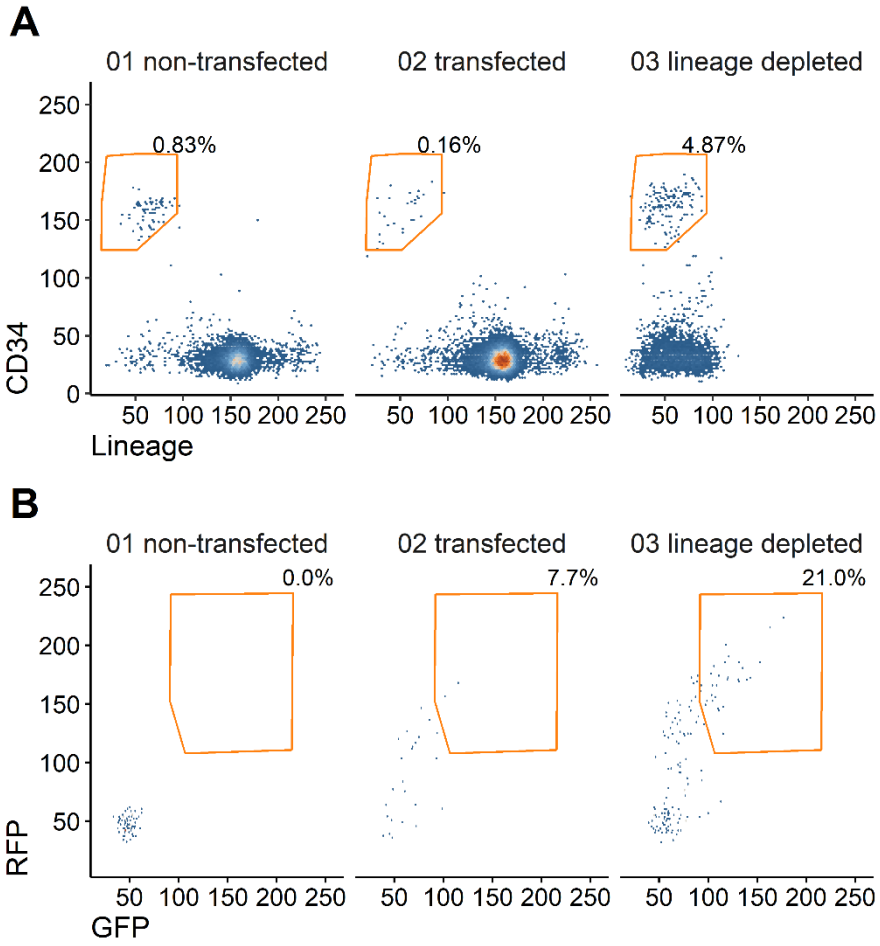


Figure 8.4 Flow cytometry analysis of transfected cord blood mononuclear cells. Based on forward and sideward scatter, living cells were selected. Of those single cells were gated. A) The cells were stained with CD34 and an antibody cocktail against lineage markers. B) Of the $Lin^{-}CD34^{+}$ fraction the $GFP^{+}RFP^{+}$ cells were gated.

Table 8.2 Electroporation conditions tested with Amaxa Nucleofector™ and NEON® transfection system.

Sample	Amaxa Nucleofector			Neon		
	cells/µl	buffer volume[µl]	program	voltage [V]	duration [ms]	pulses
1	$1x10^4$	100	U-008	1700	20	1
2	$5x10^3$	100	U-008	1400	30	1
3	$2.5x10^3$	100	U-008	1200	40	1
4	$5x10^3$	50	U-008	1200	20	2
5	$2x10^3$	50	U-008	1400	20	2
6	$1x10^3$	50	U-008	1150	30	2
7				1400	10	3
8				1500	10	3
9				1600	10	3

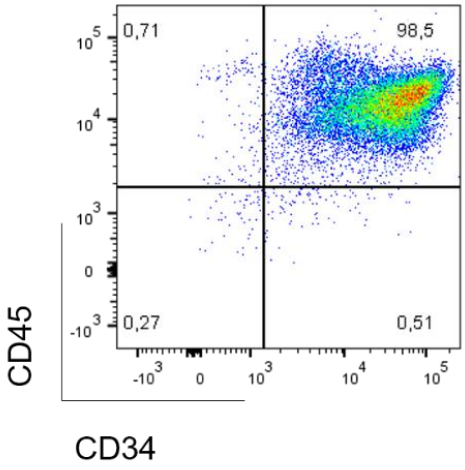


Figure 8.5 Flow cytometry analysis of expanded CD34⁺ cells after MACS enrichment. Antibodies against the cell surface protein CD45 (nuclear cells) and CD34 (progenitor cells) were used.

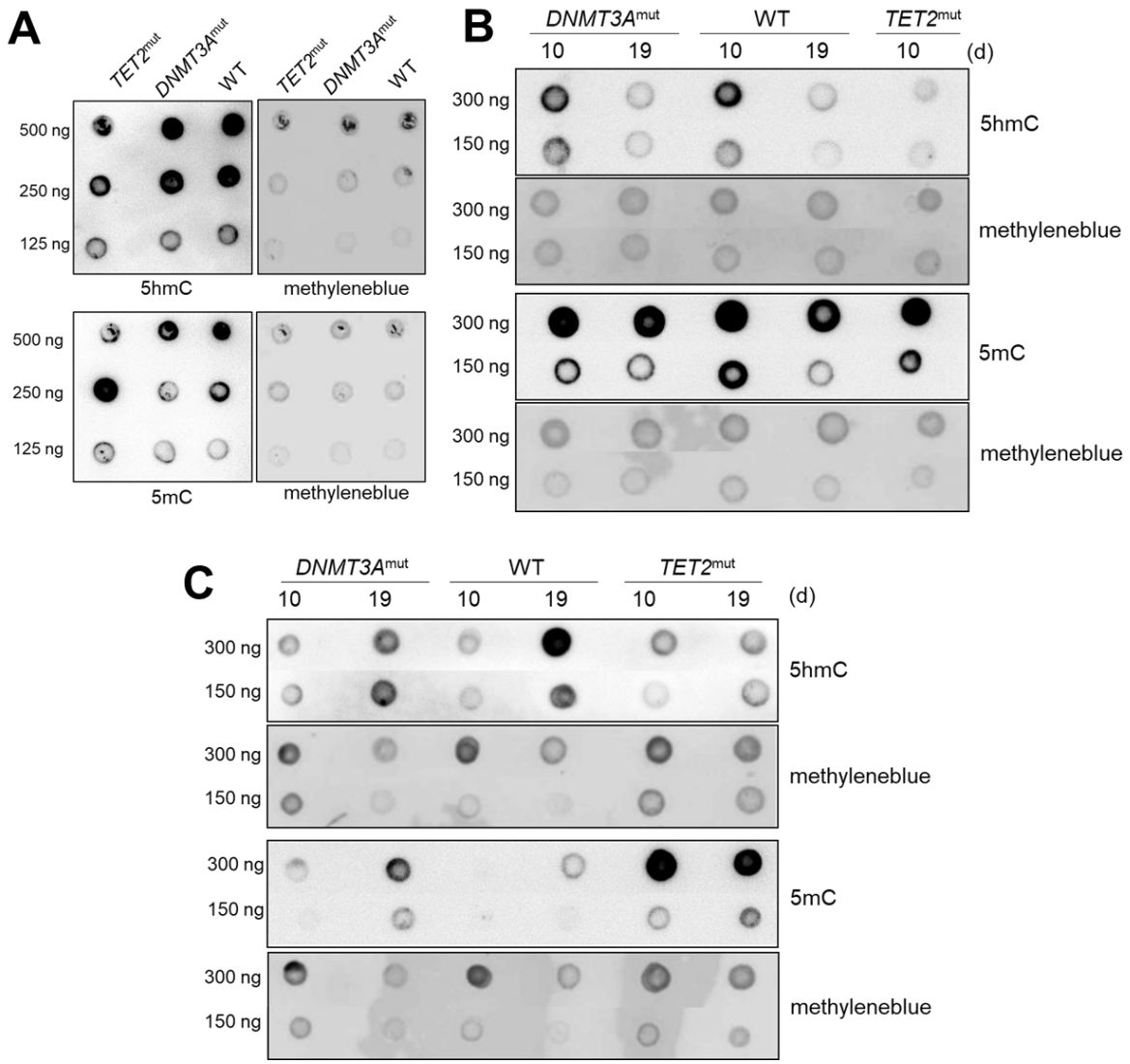


Figure 8.6 Dot Blot analysis of 5-hydroxymethylcytosine and 5-methylcytosine levels. A) DNA from sample CB004 was spotted after 19 days of culture. Left panels: chemiluminescence analysis after incubation with 5hmC and 5mC antibodies respectively. Right panels: Methylene blue staining of the membranes after chemiluminescence detection. B) DNA from CB005 was used (two different time points). C) DNA from CB007 was used (two different time points)

8.2 List of Publications

8.2.1 Articles

Christen F, Hablesreiter R, Hoyer K, Hennch C, Maluck-Böttcher A, Segler A, Madadi A, Frick M, Bullinger L, Briest F, Damm F. Modeling clonal hematopoiesis in umbilical cord blood cells by CRISPR/Cas9. **2021 Apr** *currently under revision in Leukemia*

Hoyer K, Hablesreiter R, Inoue Y, Yoshida K, Briest F, **Christen F**, […] Ogawa S, Sinn M, Damm F. A genetically defined signature of responsiveness to erlotinib 1 in early-stage pancreatic cancer patients: results from the CONKO-005 trial. EBioMedicine **2021 Apr.** DOI: 10.1016/j.ebiom.2021.103327. PMID: 33862582

Arends CM, Weiss M, **Christen F**, […], Schreiber A , Damm F. Clonal hematopoiesis in patients with anti-neutrophil cytoplasmic antibody-associated vasculitis. Haematologica. **2020 Jun**;105(6):e264-e267.

Mylonas E, Yoshida K, Frick M, Hoyer K, **Christen F**, […], Damm F. Single-cell analysis based dissection of clonality in myelofibrosis. Nat Commun. **2020 Jan** 7;11(1):73.

Christen F, Hoyer K, Yoshida K, Hou HA, […], Ogawa S, Damm F. Genomic landscape and clonal evolution of acute myeloid leukemia with t(8;21): an international study on 331 patients. Blood. **2019 Mar** 7;133(10):1140-1151.

Frick M, Chan W, Arends CM, Hablesreiter R, Halik A, Heuser M, Michonneau D, Blau O, Hoyer K, **Christen F**, […], Ogawa S., Damm F. Role of Donor Clonal Hematopoiesis in Allogeneic Hematopoietic Stem-Cell Transplantation. J Clin Oncol. **2019 Feb** 10;37(5):375-385. doi: 10.1200/JCO.2018.79.2184. Epub 2018 Nov 7. PMID: 30403573.

Arends CM, Galan-Sousa J, Hoyer K, Chan W, Jäger M, Yoshida K, Seemann R, Noerenberg D, Waldhueter N, Fleischer-Notter H, **Christen F**, […], Frick M, Damm F. Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. Leukemia. **2018 Sep**;32(9):1908-1919.

8.2.2 Posters and Presentations

Christen F. Investigating functional consequences of clonal hematopoiesis on stem cell properties by *in vitro* modeling. BSIO Future Visions in Cancer Research Symposium 2019, Berlin, Germany. Presentation.

Christen F, Hoyer K, Yoshida K, Hou HA, […], Ogawa S, Damm F. Genomic landscape and clonal evolution of acute myeloid leukemia with t(8;21): an international study on 331 patients. Frontiers in Cancer Conference 2018, Heidelberg, Germany. Poster

8.3 Acknowledgements

The acknowledgement is not included in the online version for data protection reasons.

8.4 Curriculum Vitae

The curriculum vitae is not included in the online version for data protection reasons.