Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

Qualitative and quantitative analysis of molecular markers as a tool for classification, risk assessment and monitoring of acute myeloid leukemia

Simone Silvia Weber

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Prof. Dr. W. Windisch
Prüfende der Dissertation:	1. apl. Prof. Dr. M.W. Pfaffl
	2. apl. Prof. Dr. S. Schnittger
	3. Prof. Dr. H.R. Fries

Die Dissertation wurde am 21.12.2016 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 17.05.2017 angenommen.

Table of content

Abbreviations

AK	Non-complex aberrant karyotype
AK-AML	Non-complex aberrant karyotype AML
AML	Acute myeloid leukemia
ATRA	All-trans retinoic acid
BAALC	Brain and acute leukemia gene
BM	Bone marrow
cDNA	Complementary DNA
Chr	Chromosome
CI	Confidence interval
CK	Complex aberrant karyotype
CK-AML	Complex aberrant karyotype AML
CMP	Common myeloid progenitor
CN	Cytogenetically normal
Cq	Quantification cycles
CR	Complete remission
DFS	Disease-free survival
DOM	Difference of the mean
DS	Down syndrome
DS-AMKL	Down syndrome acute megakaryoblastic leukemia
EFS	Event-free survival
ERG	E26 transforming sequence-related gene
ETP	Early T-cell precursor
ETS	E26 transforming sequence
FAB	French-American-British
FISH	Fluorescence in situ hybridization
FLT3-ITD≥0.5	*FLT3*-ITD to *FLT3* wildtype ratios ≥0.5
GMP	Granulocyte/macrophage progenitor
GOI	Gene of interest
HR	Hazard ratio
HSC	Hematopoietic stem cell
HSCs	Hematopoietic stem cells
ITD	Internal tandem duplication
LT	Long-term
MDS	Myelodysplastic syndrome

MEP	Megakaryocytic/erythroid progenitors
MLP	Multilymphoid progenitors
MPN	Myeloproliferative disease
MPP	Multipotent progenitor
MRD	Minimal residual disease
mRNA	Messenger RNA
mut	Mutation
n.s.	Not significant
$NADP^+/H$	Nicotinamide adenine dinucleotide phosphate
NK	Natural killer
NOS	Not otherwise specified
OR	Odds ratio
OS	Overall survival
PB	Peripheral blood
PCR	Polymerase chain reaction
qPCR	Quantitative real-time polymerase chain reaction
RIN	RNA integrity number
s-AML	Secondary AML
SEM	Standard error of the mean
ST	Short-term
t-AML	Therapy-related AML
TKD	Tyrosine kinase domains
WHO	World Health Organization

Zusammenfassung

Die Akute Myeloische Leukämie (AML) repräsentiert eine klinisch sehr heterogene Erkrankung. In den letzten Jahren ermöglichte die Entwicklung neuer Technologien die Identifizierung von genetischen Veränderungen, darunter Genmutationen und veränderte Genexpressionen, welche zu einem besseren Verständnis der molekularen Pathogenese, einer detaillierteren Klassifizierung und einer verbesserten Risikostratifikation führten. Dennoch stellt die AML weiterhin eine sehr aggressive Erkrankung mit hohen Rezidivraten dar. Speziell bei einer Subgruppe der AML, der AML mit normalem Karyotyp, fehlen weiterhin molekulare Marker, welche die Grundlage für eine verbesserte Risikostratifikation bei Diagnose, aber auch im Verlauf der Erkrankung bilden könnten. Aus diesem Grund sollte in der hier vorliegenden Arbeit die prognostische und biologische Bedeutung der Genexpressionmarker *BAALC* und *ERG* analysiert werden. Beide Marker wurden schon in früheren Studien analysiert, jedoch mit unterschiedlichen Ergebnissen.

In unseren Studien konnten verschieden Assoziationen der *BAALC* und *ERG* Expression zu klinischen und molekularen Parametern dargestellt werden. So war eine hohe Expression beider Gene mit jüngerem Alter und den prognostisch bedeutsamen molekularen Markern *NPM1* und *FLT3* assoziiert. Weiterhin konnten spezifische Assoziationen einer hohen *BAALC* Expression mit Veränderungen in verschiedenen Signalwegen und den kürzlich identifizierten Klassifikationsmarkern *IDH2*R172- und *RUNX1*-Mutationen gezeigt werden.

Bezüglich der Risikostratifikation, konnte beiden Genen eine prognostische Bedeutung zugewiesen werden, wobei der Einfluss der *ERG* Expression auf das Überleben nicht unabhängig von bereits bekannten Markern war. Im Gegensatz dazu war eine aberrante *BAALC* Expression mit einer schlechten Prognose assoziiert, auch unter Berücksichtigung weiterer prognostisch relevanter Marker. Neben der prognostischen Wertigkeit von *BAALC* bei Diagnose konnte gezeigt werden, dass die *BAALC* Expression den klinischen Verlauf wiederspiegelt und dass eine erhöhte *BAALC* Expression an definierten Therapiezeitpunkten mit einem schlechteren Überleben assoziiert ist. Dies lässt vermuten, dass durch die Analyse der *BAALC* Expression, sowohl bei Diagnose als auch im Therapieverlauf, bereits bestehende Risikostratifikationen noch verfeinert werden können.

Neben der prognostischen Bedeutung, wurden auch pathogene Mechanismen der AML untersucht. So konnte gezeigt werden, dass ein Zugewinn von Chromosom 21 mit einer zunehmenden *ERG* Expression einhergeht. Diese Ergebnisse lassen eine aktive Rolle von ERG in der Leukämogenese der AML mit Zugewinn von Chromosom 21 vermuten.

Abstract

Acute myeloid leukemia (AML) is a heterogeneous disease with respect to the clinical picture and therapeutic outcome. Continuous improvements in genomics technology have enabled the identification of several genetic alterations, including gene mutations and deregulated gene expression, which have contributed to a better understanding of the molecular pathogenesis, refined classification and improved prognostication. However, AML still represents an aggressive disease with poor long-term survival rates, demonstrating that we are just at the beginning of unravelling the enormous molecular complexity of AML. Especially in cytogenetically normal AML (CN-AML), there is a tremendous need to specify molecular alterations which enable improved prognostication at diagnosis and over the course of the disease. Furthermore, it becomes increasingly important to consider the clonal complexity of AML. In this context, the prognostic and biological role of *BAALC* and *ERG* gene expression has been evaluated in previous studies with varying results, either focusing on these markers alone or analyzing in the context of concomitant molecular alterations.

In CN-AML, several specific associations were identified in the present study. High *BAALC* and *ERG* expression correlated with younger ages and with the clinically adverse-risk group defined by the mutational status of *NPM1* and *FLT3*. Furthermore, *BAALC* expression was shown to strongly correlate with mutations in transcription factors, in genes that induce proliferation, and with the previously identified markers *IDH2*R172 and *RUNX1* mutations. In terms of *ERG*, no such associations were revealed.

Regarding the prognostic value of *BAALC* and *ERG*, aberrantly high expression levels of both genes were shown to have an impact on survival. Although for *ERG*, this impact was found to be surrogate only for independent markers. *BAALC* strongly impacted on outcome, conveying an independent and additive effect in the context of other adverse prognostic alterations. This suggests that delineation of genetically defined subtypes could improve prognostication in CN-AML. Furthermore, it was evaluated that *BAALC* expression represents a stable marker during follow-up, reflecting the course of the disease. Moreover, high *BAALC* expression at certain clinically defined follow-up time points was strongly correlated to adverse clinical outcome, predisposing *BAALC* assessment as an applicable strategy to detect residual disease.

Finally, it was shown that a gain of chromosome 21 resulted in increased expression of the *ERG* gene, suggesting that altered *ERG* expression represents a causative mechanism contributing to the pathogenesis of AML with a gain of chromosome 21.

These results indicate that a comprehensive screening of molecular alterations provides new biological insight and broadens the information available for risk assessment in AML.

1 Introduction

1.1 AML

1.1.1 Background, epidemiology and etiology

Acute myeloid leukemia (AML) is a heterogeneous malignancy in which genetic alterations lead to inhibition of differentiation and enhanced proliferation of myeloid precursor cells (termed blasts) resulting in the accumulation of these blasts at various stages of maturation (Figure 1). AML is the most common type of acute leukemia in adults. The number of new cases between 2008 and 2012 in the U.S. population was 4.0 per 100,000 per year.[1] Although AML can occur at all ages is it most common in older patients, with a median age at diagnosis of 67 years. Thus, the number of new cases is expected to increase as the population ages. In fact, the rates of new AML cases have been rising on average by 2.2% each year over the last 10 years.[1]

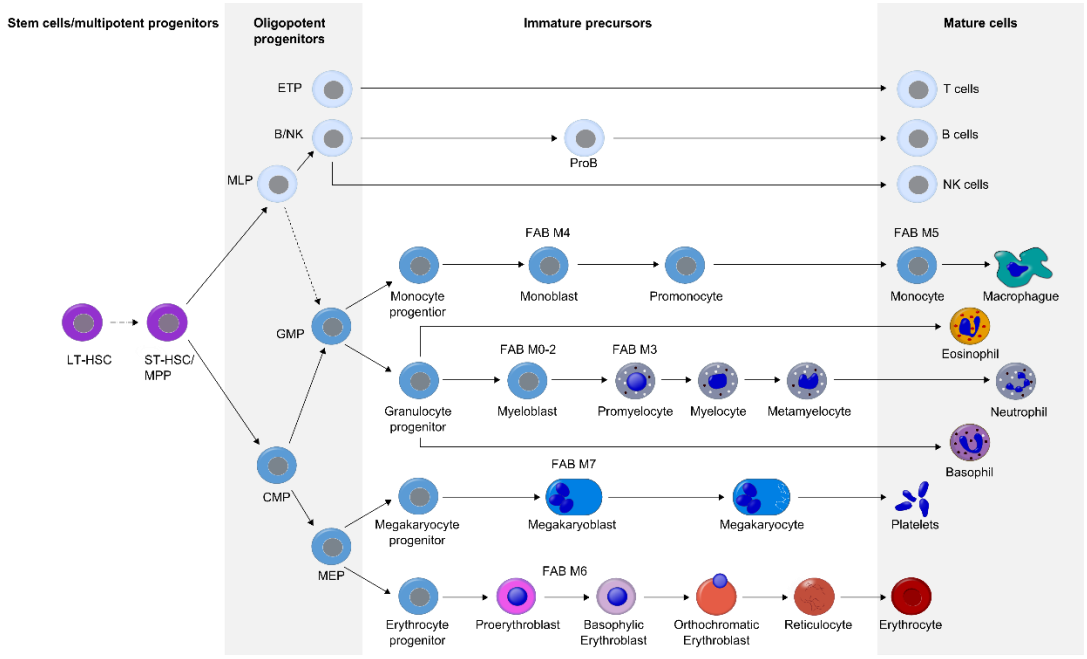


Figure 1: Schematic representation of the current hematopoietic model, also depicting the relation of the different AML subtypes (M0 to M7, cytomophologically defined based on maturational stage accoding to FAB classification[2]) to the different progenies of the myeloid lineages. Modified by Doulatov et al.[3]; edited by Elke Roos; Abbreviations: HSC, Hematopoietic stem cell; LT, Long-term; ST, Short-term; MPP, Multipotent progenitor; ETP, Early T-cell precursor; NK, Natural killer; MLP,

Multilymphoid progenitors; GMP, Granulocyte/macrophage progenitor; CMP, Common myeloid progenitor; MEP, Megakaryocytic/erythroid progenitors; FAB, French-American-British.

The etiology of AML remains largely unknown. However, there are some factors that increase the risk of AML. These include exposure to ionizing radiation, chemical benzene and chemotherapy-related drugs that damage DNA as well as having a history of a blood disorder such as myelodysplastic syndrome (MDS) or myeloproliferative disease (MPN). Further, inherited conditions such as Faconi anemia, Down syndrome (DS) as well as predisposing germline mutation, e.g. in *RUNX1* or *CEBPA*, can increase the risk for development of AML.[4;5] For a diagnosis of AML, a blast count (including myeloblasts, monoblasts, and/or megakaryoblasts) of 20% or more is required according to the WHO classification, except for AML with t(15;17), t(8;21), inv(16) or t(16;16) and some cases of erythroleukemia, where the diagnosis of AML can also be made when the blast percentage is less than 20%.[6] AML, which evolves without a prior history of chemo/radiotherapy or myeloid stem cell disorders is termed "*de novo* AML". On the other hand AML which arises from a pre-existing MDS or other myeloid stem cell disorders such as MPN is defined by the term "secondary AML" (s-AML). AML arising after chemotherapy applied for another previous malignancy is described by "therapy-related AML" (t-AML).

1.1.2 Classification

Two staging systems have been commonly used for the classification of AML. The original classification scheme, proposed by the French-American-British (FAB) Cooperative Group in 1976,[2] is solely based on morphologic and cytochemical features and divides AML into eight subtypes (M0 to M7) reflecting the different stages of maturity of the AML blast cells. (Figure 1; Table 1, lower part).

The modern and now commonly used World Health Organization (WHO) classification system introduced in 2001 and updated in 2008 considered new scientific and clinical findings as it incorporated cytogenetic and molecular genetic characteristics in addition to morphologic and cytochemical criteria for the classification of AML. Table 1 gives an overview of the WHO classification scheme of 2008, where two molecular alteration have been incorporated as provisional entities.[6] For the classification of the category "AML, not otherwise specified (NOS)" criteria similar to original FAB classification system have been maintained. Just lately, a revision of the WHO classification system has been released. Changes proposed to the section of AML also comprise the adjustment of "AML with *NPM1* mutation" and "AML with

CEBPA mutation" from provisional to definite entities as well as the introduction of "AML with *RUNX1* mutation" as a provisional entity.[7]

Table 1: WHO classification of AML[6]

Categories	
Acute myeloid leukemia with recurrent genetic abnormalities	
AML with t(8;21)(q22;q22); *RUNX1-RUNX1T1*	
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*	
APL with t(15;17)(q22;q12); *PML-RARA*	
AML with t(9;11)(p22;q23); *MLLT3-MLL*	
AML with t(6;9)(p23;q34); *DEK-NUP214*	
AML with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); *RPN1-EVI1*	
AML (megakaryoblastic) with t(1;22)(p13;q13); *RBM15-MKL1*	
Provisional entity: AML with mutated *NPM1*	
Provisional entity: AML with mutated *CEBPA*	
Acute myeloid leukemia with myelodysplasia-related changes	
Therapy-related myeloid neoplasms	
Acute myeloid leukemia, not otherwise specified (NOS)	**FAB**
Acute myeloid leukemia with minimal differentiation	M0
Acute myeloid leukemia without maturation	M1
Acute myeloid leukemia with maturation	M2
Acute myelomonocytic leukemia	M4
Acute monoblastic/monocytic leukemia	M5
Acute erythroid leukemia	M6
Acute megakaryoblastic leukemia	M7
Acute basophilic leukemia	
Acute panmyelosis with myelofibrosis (syn.: acute myelofibrosis; acute myelosclerosis)	
For a diagnosis of AML, a blast count of ≥20% is required, except for AML with the recurrent genetic abnormalities t(15;17), t(8;21), inv(16) or t(16;16) and some cases of erythroleukemia.	

1.1.3 Genetics of CN-AML

AML represents a heterogeneous group of leukemia not only clinically but also at the molecular level, reflecting various underlying genetic abnormalities. Non-random clonal chromosome aberrations, such as translocations, inversions, deletions, and numerical abnormalities (e.g. trisomies and monosomies), are detectable in approximately 55% of adults with AML.[8] These chromosomal changes have not only contributed to the classification of the

disease, but have also been recognized as the strongest prognostic factors for survival, and moreover, provide a basis for treatment decisions (Table 2).[9-11]

Particularly in terms of allogeneic stem-cell transplantation it has been shown, that patients with adverse risk diseases (Table 2) would profit from stem-cell transplantation, while those with favorable risk disease would not.

Table 2: Current stratification of molecular genetic and cytogenetic alterations, according to ELN recommendations; revised by Döhner et al.[11]

Risk Profile	Subsets
Favorable	t(8;21)(q22;q22); *RUNX1-RUNX1T1* inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11* Mutated *NPM1* without *FLT3*-ITD (normal karyotype) Biallelic mutated *CEBPA* (normal karyotype)
Intermediate-I†	Mutated *NPM1* and *FLT3*-ITD (normal karyotype) Wildtype *NPM1* and *FLT3*-ITD (normal karyotype) Wildtype *NPM1* without *FLT3*-ITD (normal karyotype)
Intermediate-II	t(9;11)(p22;q23); *MLLT3-MLL* (*KMT2A)* Cytogenetic abnormalities not classified as favorable or adverse‡
Adverse	inv(3)(q21q26.2) or t(3;3)(q21;q26.2); *GATA2–MECOM* t(6;9)(p23;q34); *DEK-NUP214* t(v;11)(v;q23); *MLL* (*KMT2A)* rearranged −5 or del(5q); −7; abnl(17p); complex karyotype§

† This category includes all cases of AML with a normal karyotype except for those included in the favorable subgroup; most of these cases are associated with a poor prognosis, but they should be reported separately because of the potential different response to treatment.
‡ Adequate numbers of most abnormalities have not been studied to draw firm conclusions regarding their prognostic significance.
§ A complex karyotype is defined as three or more chromosomal abnormalities in the absence of one of the World Health Organization–designated recurring translocations or inversions — t(8;21), inv(16) or t(16;16), t(9;11), t(v;11) (v;q23), t(6;9), and inv(3)/t(3;3). About two thirds of patients with AML with a complex karyotype have a mutation of *TP53*, a deletion of *TP53*, or both. *TP53* alterations in AML rarely occur outside a complex karyotype.

However, a large subset of AML, which illustrates enormous heterogeneity with respect to clinical picture and therapeutic outcome, presents with a cytogenetically normal (CN) karyotype - meaning that this subset of AML lacks informative cytogenetic alterations which could allocate the prognostic heterogeneity and moreover which could function as a guide for treatment decisions.[8] But, in recent years, a vast number of gene mutations as well as deregulated gene expression have been identified, which allowed to partly decipher the genetic diversity and differential prognosis of CN-AML. By now, the most useful markers

implicated in risk stratification are mutations in *NPM1*, *FLT3* internal tandem duplication (*FLT3*-ITD), and biallelic *CEBPA* mutations (Table 2).[11] However, the list of newly identified genetic alterations adding more and more prognostic and biological information is continuously growing. Therefore, Table 3 provides an overview of the most commonly mutated genes in CN-AML, also outlining the potential role in hematopoiesis, leukemogenesis, and the clinical significance, if known.

The characterization of these gene mutations has also provided insights into the biology of AML and several lines of evidence have come up showing that different genetic changes cooperate in leukemogenesis. Also from a clinical perspective, data emerge that pretreatment genetic signatures rather than single alterations will become an important tool for risk stratification in therapeutic decision-making processes.[12] For instance, it has been shown that CN-AML patients with mutated *NPM1* in the absence of *FLT3*-ITD (low-risk) have a comparatively better outcome than those with mutated *NPM1* and coexisting *FLT3*-ITD and would therefore no longer benefit from allogeneic stem cell transplantation.[10;13]

Analogically to gene mutations, deregulated expression of genes involved in cell proliferation, survival and differentiation e.g. *BAALC*,[14;15] *ERG*,[16] *MN1*,[16] *WT1*[17] and *EVI1*,[18] have been proposed as prognostic markers especially in CN-AML. More detailed information on *BAALC* and *ERG* gene is given below, since both genes represent the subject of this study.

Table 3: Overview of molecular alterations in CN-AML

Mutated Gene; Localization	**Frequency in CN-AML**	**Type of alteration**	**Impact on hematopoiesis and leukemogenesis; Clinical significance**
NPM1 **5q35**	46-53%	Frameshift mutation; net insertion of 4 nucleotides results in generation of a new C-terminus, with: (i) additional nuclear export signal motif (ii) loss of tryptophan residues	● NPM1 wildtype constantly shuttling between nucleolus and cytoplasm[19] ● various functions: regulation of cell growth and proliferation (e.g. via interaction with ARF/TP53), control of centrosome duplication and mitosis, involved in ribosome biogenesis (nuclear export)[20] ● role in hematopoiesis remains elusive: mutant protein localize aberrantly in the cytoplasm, thereby perturbing some NPM1 functions (e.g. stabilization of ARF), while activating (e.g. shuttle) or gaining (e.g. cytoplasmic interactions) others[21;22] ● significantly associated with better outcome in the absence of *FLT3*-ITD[23-26]
FLT3**-ITD** **13q12**	28-35%	Internal tandem duplications (ITD) within the cytoplasmic juxtamembrane domain; lead to constitutive activation of the tyrosine kinase domains (TKD)	● *FLT3* encodes a class III receptor tyrosine kinase, which activates PI3K/protein kinase B and mitogen activated protein kinase pathways ● *FLT3*-ITD and *FLT3*-TKD lead to constitutive activation of downstream signaling pathways ● FLT3-ITD, but not FLT3-TKD or FLT3 wildtype, induces aberrant signaling including strong activation of signal transducer and activator of transcription 5 and repression of CEBPA and PU1.[27] ● *FLT3*-ITD: associated with unfavorable outcome, particularly in patients with a high mutant(ITD) to wildtype ratio or absence of the wildtype allele[28-30] ● *FLT3*-TKD, in contrast to *FLT3*-ITD, seems not associated with unfavorable outcome[29;31]
FLT3**-TKD** **13q12**	6-14%	Missense mutations in the activation loop, lead to constitutive activation of the TKD	

ASXL1 **20q11**	5-16%	Mainly frameshift and stop mutations in exon 12; predicted to lead to loss of the C-terminal plant homeodomain finger	● member of the enhancer of trithorax and polycomb family, that enlists proteins required for maintenance of activation and silencing of gene expression by modifying chromatin configuration ● ASXL1 can interact with retinoic acid receptor in the presence of retinoic acid and enhance the transcription of some genes while repressing that of others; dependent on the cell context[32] ● *ASXL1* mutations predicted to lead to loss of the C-terminal plant homeodomain finger, which binds specific epigenetic marks on histone tail, recruiting various other factors[33] ● associated with higher age and s-AML[34-36] ● associated with inferior outcome[34-36]
CEBPA **19q13.1**	9-13% (of this ~60% biallelic mutated)	Two main hot spots: N-terminal frameshift mutations and C-terminal in-frame insertions/deletions	● belongs to the basic leucin zipper family of transcription factors ● CEBPA mediates the transition between CMPs and GMPs[37] ● CEBPA essential for long-term HSC function (control of HSC numbers and functions)[38] ● N-terminal frameshift mutations lead to a premature stop of translation of the p42 CEBPA protein, while conserving short p30 isoform ● C-terminal in-frame insertions/deletions disrupt binding to DNA or dimerization[39;40] ● double-mutated (also termed biallelic) *CEBPA* preferentially combine an N-terminal mutation on one allele (sustaining the expression of p30 only) with a C-terminal mutation on the other allele (deficient in dimerization/DNA binding)[41] ● double-mutated (biallelic) *CEBPA* associated with favorable outcome[7;41-43] ● mutations in *CEBPA* have been associated with familial AML[7;42]
DNMT3A **2p23.3**	27-35%	Mainly missense mutations; with ~60% resulting in substitution of Arg882 in the catalytic domain	● *de novo* DNA methyltransferase; catalyzing CpG methylation ● HSCs lacking *DNMT3A* have a competitive growth advantage[44] ● mutations also found in non-leukemic HSCs and in healthy individuals; related to age-related increase of clonal hematopoiesis[44-46] ● murine R878 (equivalent to human R882) showed to abrogate catalytic activity and reduced DNA binding[47] and is suggested to have a dominant negative effect in embryonic stem cells[48] ● clinical significance remains elusive: some report significant associations to inferior survival others only moderate effects on outcome, depending on genetic groups[49-52]

IDH1 **2q33.3** ***IDH2*** **15q26.1**	*IDH1* and *IDH2*R140 ~15% *IDH2*R172 ~1-4%	Missense mutations at 3 specific arginine residues	● IDH1/2 are homodimeric $NADP^+$-dependent enzymes that catalyze oxidative decarboxylation of isocitrate to produce the α-ketoglutarate, NADP, and CO2 required for the Krebs cycle ● mutated enzymes acquire a neomorphic activity that converts α-ketoglutarate to d-2-hydroxyglutarate in a reaction consuming NADPH[53] ● *IDH1*R132, *IDH2*R140, *IDH2*R172 characterize clinical distinct subsets in CN-AML ● clinical significance of *IDH* mutations remains elusive, mainly associated with adverse outcome, though only in *NPM1* mutated/*FLT3*-ITD negative CN-AML for *IDH1* and *IDH2*R140[54-56]
***MLL(KMT2A)*-PTD** **11q23**	7-10%	Partial tandem duplication (PTD) of select exons; producing an in-frame, elongated protein. Mutant protein contains a duplicated N-terminus including AT-hooks and CXXC-domain	● MLL maintains normal function of HSC as a positive regulator of gene expression e.g. regulation HOX gene expression in hematopoietic systems[57] ● $Mll^{PTD/wildtype}$ hematopoietic stem and progenitor cells exhibited a proliferative advantage and reduced myeloid differentiation[58] ● clinical significance remains controversial, some report significant associations to risk of relapse or to shorter complete remission duration,[25;59;60] some reveal no impact[61;62]
NRAS **1p13.2**	9-13%	Missense mutations exclusively located at codons 12, 13 and 61, resulting in loss of intrinsic GTPase activity and constitutive activation of the RAS protein	● mutant NRAS increases HSC proliferation and self-renewal capability and myeloid differentiation bias[63] ● no significant impact on outcome[25;64]

RUNX1 **21q22**	6-26%	Missense, nonsense and frameshift mutations; mostly resulting in loss or destruction of the transactivation domain and the Runt homology domain in some cases	● transcription factor: regulates the expression of hematopoiesis specific genes, such as growth factors (GM-CSF, MPO, IL3), surface receptors (TCRA, TCRB, M-CSF receptor, FLT3), signaling molecules (CDKN1A, BLK, BCL2), and transcription activators (STAT3, MYB)[65] ● RUNX1 essential for the maturation of T-, B-lymphocytes and megakaryocytes/platelets; negatively regulates proliferation of HSCs and myeloid progenitors[66;67] ● *RUNX1* mutations found in pedigrees of a rare inherited human disease familial platelet disorder with propensity to develop AML[7] ● associated with higher age and s-AML[68] ● associated with inferior outcome[68-71]
TET2 **4q24**	6-23%	Missense, nonsense and frameshift mutations; abrogating TET2 enzymatic function	● TET enzymes catalyze the Fe(II)- and α-ketoglutarate-dependent hydroxylation of 5-methylcytosine residues in DNA to form 5-hydroxymethylcytosine, which is involved in epigenetic regulation of gene expression.[72] ● short hairpin RNA-mediated depletion of TET2 in HSCs resulted in skewing toward myeloid differentiation[73] ● Loss of Tet2 resulted in expansion of the HSC compartment in a cell-intrinsic manner and enhanced HSC self-renewal[74] ● associated with higher age and s-AML[75-78] ● mutations also found in non-leukemic HSCs and in healthy individuals; related to age-related increase of clonal hematopoiesis[44-46] ● clinical significance remains controversial: some report significant associations to inferior survival; others report no impact or only an effect in specific genetic groups (e.g. favorable-risk group (Table 2) or in consumption with IDH)[77-80]
WT1 ***11p13***	6-11%	Mostly frameshift mutations in exon 7 and substitutions in exon 9	● transcription factor: either enhance or repress the expression of specific targets, highly context specific ● loss of *WT1* was associated with decreased growth of leukemic cells and rapid induction of apoptosis[81] ● enforced *WT1* expression inhibits differentiation[81] ● elevated expression can be found in high proportion of AML, used for detection of residual disease[82] ● associated with unfavorable outcome, particularly in younger patients[83-86]

1.1.3.1 The *BAALC* gene

The brain and acute leukemia cytoplasmic (*BAALC*) gene expression has originally been described in neuroectoderm-derived tissues and was later, in a search for leukemia associated genes, also found in undifferentiated hematopoietic cells.[87] The *BAALC* gene localizes on human chromosome band 8q22.3. It consists of eight exons, of which exon 2 comprises an alternative termination codon. Up to now, at least 8 differentially spliced transcripts have been described in AML, with the transcript variant 1-8 and 1-6-8 being the most abundant ones. [87] The DNA sequence as well as the expression pattern of *BAALC* is highly conserved among mammals, whereas lower vertebrates lack comparable orthologs. Translation of the different splice variants result in six different protein isoforms, which show a cytoplasmic localization in morphologically immature CD34+ cells. These isoforms only share the N-terminal domain of the first 53 amino acid residues, while the remaining sequence is diverse.[87] The protein sequence of BAALC exhibits no homology to any known protein and does not contain any recognizable motifs or functional domains. A fact, through which the clarification of the functional role of BAALC in biological processes is additionally impeded. In normal hematopoiesis the expression of *BAALC* is restricted to CD34+ hematopoietic progenitor cells including uncommitted CD34+/CD38- cells as well as more committed CD34+ lymphoid, erythroid and myeloid progenitor cells. In vitro studies revealed that *BAALC* expression is down regulated during hematopoietic cell differentiation.[88] Based on this restriction of *BAALC* expression to undifferentiated progenitors in normal hematopoiesis as well as on the association of high *BAALC* expression to the expression of stem cell markers and more immature subtypes of AML it has been suggested that *BAALC* functions as a marker of early hematopoietic cells.[87;89] However, since aberrant expression is also found in other tumors like glioblastoma[87] and malignant melanoma[90] a comprehensive role of BAALC in general pathways has been suggested. In a first attempt to unravel the function of BAALC in the hematopoietic system it has been shown that constitutive activation BAALC did not promote proliferation or survival of hematopoietic cells, but did contribute to myeloid differentiation block in Hoxa9-immortalized cell lines.[91] A more recent study provided the molecular basis for a dual function of BAALC in regulating cell differentiation as well as proliferation. In this study Morita et al.[92] identified two proteins with substantial roles in the ERK pathway, namely MEKK1 and KLF4, as potential interacting partners of BAALC. Subsequent in vitro assays on leukemic cells implied that BAALC on one hand induced cell cycle progression by sustained ERK activation, and on the other hand blocks ERK-mediated differentiation.

Besides the functional role of BAALC in hematopoiesis and leukemogenesis the mechanisms regulating *BAALC* expression in normal and abnormal hematopoiesis are also under investigation. In search for putative activation sides in the upstream regulatory machinery of *BAALC*, different transcription start sides with potential binding signatures for the transcription factors ARID3A, MFZ1, RUNX1 and the SP1/NFkB transactivation complex have been identified.[93;94] Further, Franzoni et al.[94] hypothesized from a cell line based model, that the *BAALC* promoter exhibits a bivalent mark of active and inactive post-translational-histone modifications, suggesting that *BAALC* represents a so called paused gene which can be specifically activated or repressed during differentiation.

Regarding the clinical significance of *BAALC*, collective data suggest *BAALC* as an important risk factor in CN-AML, where *BAALC* gene expression as a single marker is associated with treatment resistance and poor outcome.[15;89;95-99] Further, an association of *BAALC* expression to other molecular prognostic markers, and to a distinct gene-expression signature in CN-AML has been shown. For instance, high *BAALC* expression has been demonstrated to correlate with the mutational status of *FLT3*-ITD, *CEBPA*, *MLL*-PTD and with an unmuted *NPM1* as well as with elevated expression levels of the ETS-related gene *ERG*, the multi-drug resistance gene 1 and the stem cell markers CD133, CD34 and KIT. [14-16;95] Besides its association to several prognostic markers, high *BAALC* expression was independently associated with lower complete remission (CR) rates,[15;95] shorter disease-free survival (DFS)[95] and shorter overall survival (OS),[14;15;95] in some studies, while other studies could not confirm this independent prognostic effect of *BAALC* expression on survival.[16;100]

1.1.3.2 The *ERG* gene

The E26 transforming sequence (ETS)-related gene (*ERG*) is located on human chromosome band 21q22 and encodes a member of the ETS family of transcription factors. Like all members of this family, ERG possesses an evolutionary-conserved ETS domain of about 85 amino acid residues that mediate binding to purine-rich DNA sequences with a central GGAA/T core consensus and additional flanking nucleotides. This ETS DNA-binding motive is located in the C-terminal region, while the N-terminal region shows a so called pointed domain, a helix-loop-helix structure, which confers protein-protein interactions. Further two transcriptional activation domains have been described on either terminal sites. At least five isoforms of the ERG protein are generated by alternative splicing and translation initiation sites, of which isoform 3 represents the major isoform of normal and malignant hematopoietic cells.[101;102] Expression of *ERG* has been shown in several cell lines, including endothelial cells as well as in normal premature (CD34+) and malignant hematopoietic cells, but not in mature lymphoid or

myeloid cells.[102] Regarding the role of ERG in normal hematopoiesis it has been shown that homozygous mice harboring missense mutation in *erg* die at the embryonic stage as they fail to sustain definitive hematopoiesis.[103;104] Mice heterozygous for the same mutation represented with functionally impaired LT-HSCs, significantly lower numbers of committed hematopoietic progenitors and lower platelet numbers.[103] Thus, these data suggest that ERG function is essential to the production and maintenance of the hematopoietic stem cell pool.

Addressing the leukemic potential of *ERG*, ectopic expression of *ERG* was shown to induce megakaryocytic differentiation in the erythroleukemia cell line K562 as well as megakaryoblastic leukemia in mice.[102;105] Further, forced expression of *ERG* has been shown to promote the development and also the maintenance of leukemia in a mouse adult bone marrow (BM) transplantation model[106] and in-vitro assays revealed a growth-promoting effect of *ERG* on various human leukemia cells of erythroid, myeloid, megakaryocytic, T-, and B-cell lineages.[107] In addition, it has been suggested that ERG, like all ETS family members act in concert with other cellular proteins not only under physiological conditions, but also in the pathogenesis of leukemia. For instance, cooperating effects of elevated *ERG* expression and specific mutations in the onset of Down syndrome acute megakaryoblastic leukemia (DS-AMKL) have been described.[106-108] Besides, population-based studies show that the incidence of leukemia is 10 to 20 fold higher in individuals with DS compared with the overall population, with a particularly striking increase of AMKL, suggesting a general dosage-sensitive effect of genes located on chromosome 21 within leukemogenesis.[109] In an attempt to identify the respective genes on chromosome 21 a common Down syndrome critical region on chromosome 21q22 has been defined, which narrowed the number of candidate genes potentially involved in DS-associated leukemogenesis.[110;111] Amongst these genes *ERG* represented one of the most extensively studied, but the results remain controversial.[112;113]

Besides the potential role of an increase in DNA content to elevated *ERG* expression the exact mechanisms regulating *ERG* expression in normal and abnormal hematopoiesis are still under investigation. The human *ERG* locus has at least two recognized promoters and a specific enhancer +85kb downstream of the translation start site.[107] Recently, it has been shown that *ERG* expression is regulated by complexes of HSC transcription factors, namely SCL, LYL1, LMO2, GATA2 or GATA3, RUNX1, FLI1 and ERG itself, binding its promoters and +85 stem cell enhancer. Especially for T-ALL it has been shown, that abnormal *ERG* expression is probably caused by aberrant activation of the +85 enhancer. Even though these findings provide precious insights into transcriptional regulation of *ERG* they still do not explain the mechanisms leading to altered promoter and enhancer activity and therefore to abnormal *ERG* expression.

Finally, it should be mentioned that beside dysregulated expression of *ERG*, the involvement of *ERG* in chromosomal translocations has been implicated in rare cases of leukemia and more commonly in various types of other neoplasms. Therefore, the *ERG* locus can be rearranged and fused with *FUS*/*TLS* in AML[114], with *EWS* in Ewing sarcoma[115] and with *TMPRSS2*[116;117] in prostate cancer. Especially in prostate cancer the fusion has been shown to result in *ERG* overexpression and has been associated with poor outcome.[117]

Regarding the clinical significance of deregulated expression of *ERG*, the impact on clinical outcome remains controversial. Especially in CN-AML, high *ERG* expression has been allocated to lower CR rates, shorter DFS, event-free survival (EFS) and OS in some studies,[16;95;118] while other studies only reported an adverse effect of high *ERG* expression on the achievement of CR and on EFS.[119] Besides the prognostic significance it has been published recently that cell based *ERG* overexpression results in resistance to kinase inhibitors such as sorafenib and dovitinib.[120] This finding is of special therapeutic interest since sorafenib currently is being tested in different clinical AML trials.

1.1.4 Therapy of AML

The prognosis of AML is determined by AML-associated factors such as cytogenetic and molecular genetic aberrations as well as patient-related characteristics such as age, comorbidity and the performance status in general. While the latter predict treatment-related mortality the former factors provide a prediction of resistance to therapy.[11]

The treatment of AML is usually performed within randomized clinical trials and typically involves induction therapy, consolidation therapy and in some cases maintenance.

The goal of induction therapy is to bring about a CR, which means to reduce the amount of all leukemic cells to levels below morphological sensitivity (BM blasts < 5%, absolute neutrophil count > 1.0 x 10^9/L and platelet count > 100 x 10^9 /L).[10] Standard induction therapy of younger adults (aged 18-60) usually consists of one or two cycles of two cytotoxic drugs: cytarabine (also called Ara-C) given as continuous infusion for seven days and an anthracycline (daunorubicin, idarubicin or mitoxantrone) given intravenously for three days, therefore termed the “7+3” regimen. With the 7+3 induction scheme complete response rates of 60-85% can be achieved in younger adults and until recently no other induction regime has been shown to be superior.[11] For older patients (ages >60 years) with favorable-risk and intermediate-risk cytogenetics and no coexisting conditions, the standard induction is the same as in younger adults, while dose reduction, investigational therapy or best supportive care may be considered in some of these patients, who are not eligible to receive standard

intensive chemotherapy.[11] Complete response rates in older patients are 40-60% after induction chemotherapy.

Standard post remission therapy (termed consolidation therapy) aims at the full eradication of the leukemic cell compartment and includes conventional chemotherapy as well as hematopoietic cell transplantation. The decision for hematopoietic cell transplantation depends on the leukemic genetic risk profile (Table 2), the risk of treatment-related death, and specific patient transplantation associated factors (e.g. age, comorbidity). Presently, a predicted risk of relapse of more than 35% is widely considered to warrant hematopoietic cell transplantation during the first remission.[121] Patients with favorable risk genetics (Table 2) and no coexisting conditions, should receive 2-4 cycles of intermediate-dose cytarabine. For patients 16 to 60-65 years with intermediate- or adverse-risk genetics, persistent disease or other high-risk clinical features such as s-AML or t-AML, the aggressive treatment of allogeneic hematopoietic-cell transplantation should be considered. Though the decision for allogeneic hematopoietic-cell transplantation in intermediate-risk patients seems not as clear as most of these patients represent with normal cytogenetics and molecular lesions might provide further prognostic information. For patients 60-65 to 75 years with unfavorable genetic risk allogeneic hematopoietic-cell transplantation can be considered, if they are physically able to undergo this procedure, while for those above 75 years investigational therapy should be considered.[11]

An exception of these recommendations has to be made for the M3 subtype of AML, termed acute promyelocytic leukemia (APL) with a translocation between chromosomes 15 and 17 according to the WHO, where treatment includes the non-chemotherapy drug all-trans retinoic acid (ATRA). ATRA, a derivative of retinoic acid, targets the *RARA* and reverses the differentiation block of promyelocytic blasts induced by the *PML-RARA* chimeric gene, thereby representing the first molecular target-based cancer therapy in AML.

Currently, new compounds targeting different AML specific cellular processes or the mutant proteins directly are under investigation. These compounds include tyrosine kinase inhibitors targeting FLT3 (e.g. midostaurin, quizartinib), inhibitors targeting the mutant metabolic enzymes IDH1 (AG-120) and IDH2 (AG-221), demethylating agents (e.g. azacytidine, decitabine) as well as antibodies, which specifically target antigens on leukemic cells, such as gemtuzumab ozogamicin, a humanized anti-CD33 monoclonal antibody conjugated with the cytotoxic agent calicheamicin. Besides the promising results provided by some of these new compounds in combination with conventional cytotoxic therapies none of them can be used as single agents to cure the disease.[11]

1.1.5 Monitoring of AML during therapy

As described in section "1.1.4 Therapy of AML" antileukemic therapy in AML aims at the eradication of all leukemic cells to achieve cure from the disease. Achieving a morphologically leukemia-free state (termed complete remission) is the aim of the induction chemotherapy and it has consistently been shown that failure of early blast clearance by remission induction therapy is a major independent prognostic factor for prognosis.[122] Despite the relatively high rates of 40-80% of AML patients achieving complete remission, the majority of these patients experience relapse within three years after diagnosis.[10;123-125] The prognosis of patients in relapse is very poor with a five year survival probability of 4 to 46% of the patients (aged 16-60 years), whereby the duration of remission before relapse represents a major prognostic factor as an early relapse (duration of CR <6 month) contributes to a more adverse prognosis.[10] The source of these relapses has been shown to stem from persistent leukemia cells in the majority of cases, existing at levels below morphological sensitivity.[126;127] By now, several studies have shown that it is possible to detect this so called minimal residual disease (MRD) of AML before hematologic manifestation by multiparameter flow cytometry identifying leukemia associated aberrant phenotypes and moreover by quantitative real-time polymerase chain reaction (qPCR) based methods, digital PCR or (ultra) deep sequencing detecting residual leukemia specific targets. Due to the high sensitivity of qPCR-based methods, molecular markers have become key targets for the assessment of treatment response in individual patients to detect early relapse and to allow direct therapeutic intervention.[10;128;129] These targets include fusion genes (e.g. *PML-RARA*,[130;131] *CBFB-MYH11*,[131;132] *RUNX1-RUNX1T1 (*formerly *AML1-ETO*),[131;133;134]) and gene mutations, of which mutations in *NPM1*[129;135] and *MLL*-PTD[136;137] represent the best validated targets amongst the gene mutations. Further, genes highly expressed in some types of leukemia have also been described to represent a suitable MRD target, with the Wilms tumor (*WT1*) gene representing one of the best evaluated target in CN-AML.[17] The prognostic feasibility of the sensitive detection of leukemia specific targets has been validated just recently in a large clinical trial, where the persistence of *NPM1* mutated transcripts after the induction chemotherapy not only provided information of residual disease and predicted risk of relapse, but also was the only independent prognostic factor for death in multivariate analysis. Moreover, this revealed that among patients with a high-risk genotype (*FLT3*-ITD, mutated *DNMT3A*, or both), negative results on qPCR after the second chemotherapy cycle distinguished a group of patients (79%) with a relatively favorable outcome. This result has strong implications for clinical/therapeutic decision making, as this group would have been originally suggested for stem-cell transplantation (Table 2).

1.2 Aim of the dissertation

The general aim of this doctoral thesis was a molecular characterization of AML to delineate the complex inter-relationships between genetic alterations and the expression of selected genes, in order to get further insight into the disease biology and to improve the risk assessment and clinical decision making.

For this purpose, different approaches were applied. First, a large diagnostic cohort of CN-AML was comprehensively investigated for molecular genetic alterations, essentially focusing on *BAALC* and *ERG* gene expression, as the prognostic value of these genes has been investigated in previous studies with varying results. Putative associations of altered *BAALC* and *ERG* gene expression with recently described molecular alterations and with functional biological pathways needed to be revealed, in order to delineate or define new subgroups of CN-AML.

Furthermore, in an attempt to improve prognostication in CN-AML, the impact of deregulated *BAALC* and *ERG* expression on the clinical outcome was examined - thereby focusing on the prognostic impact of these alterations alone, and moreover in the context of concomitant molecular alterations.

The second approach was to address the applicability of *BAALC* assessment during the course of the disease, as the prognostic information of residual disease becomes increasingly important in clinical decision-making. Therefore, *BAALC* expression was analyzed in a considerable number of CN-AML patients at distinct follow-up time points in order to assess the stability and the prognostic value of *BAALC* gene expression during therapy.

In a third approach, further insights into the functional involvement of altered *ERG* gene expression in the pathogenesis of AML needed to be gained. To concern this approach, *ERG* gene expression was analyzed in a group of patients with AML and an acquired gain of chromosome 21, thereby investigating a possible relationship between a gain of *ERG* DNA copy numbers and *ERG* gene expression levels. Besides this, concomitant molecular alterations were analyzed to reveal putative interacting partners of *ERG* in this subtype of AML.

2 Materials and methods

2.1 Sample preparation and nucleic acid isolation

Anticoagulated (heparin, EDTA, citrate) BM and peripheral blood (PB) samples were referred to the MLL Munich Leukemia Laboratory for diagnostic or follow-up assessment. Mononuclear cells were separated by Ficoll density gradient at 1204 g for 20 min using Ficoll (PAA Laboratories GmbH, Pasching, Austria) or Pancoll (PAN Biotech GmbH, Aidenbach, Germany) with a density of 1.077 g/ml. Cell pellets of 2.5–5x10^6 cells, depending on yield, were resuspended in PBS or water (only for DNA) and stored at -80°C if not directly processed. Either mRNA or total RNA was extracted with the MagnaPureLC mRNA Kit I (Roche Applied Science, Mannheim, Germany) or with the MagNA Pure 96 Cellular RNA Large Volume Kit (Roche Applied Science) according to the manufacturer's protocol. Genomic DNA was isolated with QIAamp DNA Mini kit or with QIAsymphony DSP DNA Midi Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. The cDNA synthesis was performed from mRNA or total RNA (from an equivalent of 2.5–5x10^6 cells) using 300 U Superscript II (Life Technologies, Darmstadt, Germany) and random hexamer primers (Roche Applied Science) in a 50 µl reaction. The cDNA and genomic DNA was stored at -20°C and -4°C, respectively.

2.2 *BAALC* and *ERG* assessment

Quantitative assessments of *BAALC* and *ERG* gene expression, *ERG* genomic DNA as well as *ABL1* expression (reference gene) were performed by the use of the Applied Biosystems 7500 Fast Real Time PCR System (Life Technologies). Each sample was analyzed at least in duplicate. *BAALC* and *ABL1* expression was determined as described before (Appendix I).[96] *ERG* expression and *ERG* DNA copy numbers were assessed as outlined in Weber et al.[138] (Appendix II). To calculate *BAALC*, *ERG* and *ABL1* copy numbers, standard curves for the respective assays were generated in every run by 10-fold dilution series of five different plasmid concentrations. *BAALC* and *ERG* gene expression was normalized against the expression of the reference gene *ABL1*, calculated as % GOI/*ABL1*. *ABL1* represents the most commonly used reference gene in hematology. The use of *ABL1* has been proposed by a collaborative initiation within the Europe Against Cancer program, including 26 laboratories in 10 member countries, where 14 potential reference genes have been evaluated on normal and diagnostic leukemic samples.[139] These results of the EAC trial was confirmed by our group, analyzing different reference genes in AML with fusion genes.[140] The use of *ABL1* as reference gene for the assessment of *BAALC* and *ERG* gene expression was further validated for this work in a core-set of patients. The

result on this validation is given in the Appendix V. *ERG* DNA copy numbers were calculated using the comparative Ct method.[141] For quantification of DNA copy numbers, the albumin gene was used for normalization. To calibrate the comparative DNA copy numbers, a mixture of DNA of ten healthy individuals was used. Thus, a relative quantity of 1 corresponds to 2 DNA copies of the *ERG* gene.

2.3 Brief overview of methods used to detect molecular alterations, other than *BAALC* and *ERG* expression

Screening of *FLT3*-TKD,[31] *IDH*,[142] *NRAS*,[64;143] and *NPM1*[23] gene mutations was performed using a melting curve based LightCycler assay (Roche Diagnostics, Mannheim, Germany) with hybridization probes. After 40 cycles of amplification the melting-curve analysis was started at 28°C/40°C and continued to 70°C/95°C with slope of 0.2°C per second and continuous detection with channel F2/F1. LightCycler data were analyzed using LightCycler software (Roche Diagnostics). All cases that revealed an aberrant melting curve as compared to wildtype control were subjected to nucleotide sequence analyses (see below).

To quantify the mutation ratio of *NPM1* qPCR was performed by the use of either the LightCycler 1.5 System (Roche Diagnostics) with the application of hybridization probes as the detection format or by digital PCR using a hydrolysis probe-based assay on 12.765 Digital Arrays (Fluidigm, San Francisco, CA, USA), that separates each PCR reaction into 765 reaction chambers. Digital PCR data analysis was done by EP1 Data Collection v.3.1.2 software (Fluidigm). Positive signals were counted with DIGITAL PCR ANALYSIS v.3.1.3 software (Fluidigm) and transformed to a value of estimated targets. Expression levels were calculated as % *NPM1*mut/*ABL1*. Comprehensive information on these methods are provided in the respective publications.[135;144]

Screening of *MLL*-PTD was performed by qPCR using the LightCycler System (Roche Diagnostics) applying hybridization probes as the detection format. The expression levels are given as % *MLL*-PTD/*ABL1*. Detailed information on primer, probes and the reaction conditions are given in Weisser et al.[136]

Determination of the length of the *FLT3*-ITD and quantification of the *FLT3*-ITD mutation/FLT3 wildtype ratio were realized by genescan analysis, where the forward primer was labeled with 6-FAM. PCR products were analyzed using 3130, 3130*xl* and 3730*xl* Genetic Analysers (ABI, Darmstadt, Germany). The *FLT3*-ITD was quantified as the ratio of the mutation compared to the wildtype allele. Detailed information is given elsewhere.[145;146]

The remaining molecular alterations in *ASXL1, CEBPA*, *DNMT3A*, *RUNX1*, *TET2*, *TP53* and *WT1* were analyzed by either Sanger sequencing or by next-generation sequencing. Sanger sequencing was performed using BigDye chemistry (Applied Biosystems, Weiterstadt, Germany) and the 3130, 3130xl and 3730xl Genetic Analyzers (ABI).

Regarding next generation sequencing, different library preparation and sequencing approaches have been performed in our laboratory. In short:

Library preparation ThunderStorm (Targeted enrichment by droplet-based multiplex PCR)

Sheared genomic DNA (Covaris S220 instrument; Covaris, Woburn, MA) was combined with a PCR reaction mix excluding the oligonucleotide primer molecules. This reaction mastermix and the primer library were loaded separately into the ThunderStorm instrument (RainDance Technologies, Billerica, MA, USA) and picoliter-size droplets containing one primer library per droplet were generated. After amplification, emulsion PCR droplets were broken, purified and quantified. Barcode indices and suitable MiSeq adaptor sequences were added using a second-round PCR step. Equal amounts of each purified, tagged library were then pooled for cluster generation and sequencing. PhiX control libraries (final concentration 1.5%) (Illumina, San Diego, CA, USA) were added to the pooled library for quality control. Detailed information is given in Delic et al.[147]

Library preparation Assay-on-Demand for 454 Pyrosequencing

Target specific regions of interest were amplified using the FastStart High Fidelity PCR System and GC-RICH PCR System kits (Roche Applied Science, Penzberg, Germany) in preconfigured 96-well primer plates containing primer pairs as well as sample specific distinct Multiplexing Identifier. Each PCR product was individually purified using Agencourt AMPure XP beads (Beckman Coulter, Krefeld, Germany) and quantified using the Quant-iT PicoGreen dsDNA kit (Invitrogen, Carlsbad, CA, USA). Detailed information is given in Kohlmann et al.[148]

Library preparation Fluidigm AccessArray

Amplicon generation on the Fluidgm AccessArray System (Fluidigm) reduces the time required for enrichment of targeted sequences by combining amplicon generation with library preparation. The Access Array System workflow consists of three major phases: 1) designing and validating target-specific primers for targeted re-sequencing, 2) running an Access Array Integrated Fluidic Circuit (IFC), and 3) qualifying and quantifying harvested PCR products for sequencing. Input requirements are DNA 50 - 80ng. Detailed information is given elsewhere.[149]

454 Pyrosequencing

The multiplexed amplicon pool was processed according to the manufacturer's recommendations using GS FLX Titanium small volume emPCR Kit (Lib-A) (Roche Applied Science, Penzberg, Germany). Following the emPCR amplification, clonally amplified beads were enriched for 454 next-generation sequencing. Forward (A) and reverse (B) beads were processed separately until combining them at the step of loading the respective PicoTiterPlate (PTP) lanes.

All data were generated using the GS FLX Sequencer Instrument software version 2.3. Image processing and amplicon pipeline analysis was performed using default settings of the GS RunBrowser software version 2.3 (Roche Applied Science). Sequence alignment and variant detection was performed using the GS Amplicon Variant Analyzer software version 2.3 (Roche Applied Science). The results were further processed and visualized in R/Bioconductor using the package R453Plus1Toolbox (version 1.0.1) and the Sequence Pilot software version 4.0.0 (JSI medical systems, Kippenheim, Germany). For the detection of variants, filters were set to display sequence variances occurring in more than 2% of bidirectional reads per amplicon in at least one patient. Amplicons were considered as dropout and excluded from analysis if the coverage at any analyzed position in any of the two paired-end sequences (minimal coverage) was <80 reads.

MiSeq Sequencing

Amplified targets were sequenced on the MiSeq sequencing-by-synthesis benchtop sequencer according to the manufacturer's protocol (Illumina). The sequencing runs were set for 'Resequencing' workflow on the Illumina MiSeq Reporter v2.2.29 software (Illumina) and 500-cycle MiSeq v2 reagent cartridges (Illumina) were used to sequence libraries with paired-end, dual-indexing 251 cycles per read (2 x 251). The quality of the sequence run was monitored by the Sequencing Analysis Viewer (Illumina). Data was automatically demultiplexed by the MiSeq Reporter and corresponding zipped FASTQ files were generated for each barcode index.

Variant interpretation

Validity of the somatic mutations was checked against the publicly accessible COSMIC database,[150] ClinVar (www.ncbi.nlm.nih.gov/clinvar/) and ExAC database (exac.broadinstitute.org/). Functional interpretation was performed using SIFT 1.03 (sift.jcvi.org), PolyPhen 2.0 (Adzhubei et al, Nat Methods, 2010). Additionally, TP53 variants were verified using the IARC repository (Petitjean A et al, Hum Mutat, 2007). Single-nucleotide

polymorphisms according to NCBI dbSNP database (ncbi.nlm.nih.gov/snp), synonymous variants and alterations within introns with the exception of splice-site mutations at position +/- 1 or 2 were not scored.

2.4 Patient samples

All patient samples included in the respective studies were referred to our laboratory for diagnostic or follow-up assessment of AML between September 2005 and July 2015. All patients gave their written informed consent for scientific evaluations. The study was approved by the Internal Review Board and adhered to the tenets of the Declaration of Helsinki. Addressing different scientific aspects following samples were analyzed:

1) *BAALC* and *ERG* gene expression in CN-AML (Appendix I; Appendix IV) [96;151]

To asses *BAALC* gene expression in CN-AML, a total of 326 younger (<65 years, range: 18-65) CN-AML patients were investigated. Of these, 325 patients have also been investigated for *ERG* gene expression. All these patients had, to the best of our knowledge, *de novo* AML without any preceding malignancy or MDS.

2) *BAALC* gene expression for detection of MRD (Appendix I; Appendix III) [96;152]

To address *BAALC* gene expression as a target for MRD monitoring, in total 632 diagnostic and follow-up samples of 142 *de novo* CN-AML patients with high *BAALC* expression at initial diagnosis for whom at least one follow-up sample was available were investigated. Of these, 57 diagnostic and 301 follow-up samples had been published in both studies (Appendix I; Appendix III).[96;152] The median number of follow-up samples per patient was 2 (range: 1-21) with a median follow-up time of 63 months (range: 1-101 months). If multiple samples per patient were available within the time interval of three to six month after induction chemotherapy, only the first evaluation was included in the respective calculations. Further, 18 samples of 9 *de novo* CN-AML patients with low *BAALC* expression at initial diagnosis, for whom a follow-up sample showing complete molecular remission defined by *NPM1* mutation status was available, were analyzed.

2) *ERG* gene expression in patients with AML and acquired gain of chromosome 21 (Appendix II) [138]

In total 479 AML cases, including 25 secondary AML (s-AML) following MDS (n=16) or myeloproliferative neoplasms (n=9) and 22 therapy-related AML (t-AML), were analyzed. The total cohort was subdivided based on karyotype and fluorescence in situ hybridization (FISH) data into 3 subgroups according to the *ERG* copy number state.

Cohort A (further referred to as “Cases with amplification of *ERG*”) included 33 cases with structural aberrations involving the long arm of chromosome 21 and amplification of *ERG* (defined as more than 6 copies) as demonstrated by interphase FISH with probes encompassing the *ERG* gene. Of these, 32 cases showed a complex karyotype (CK) characterized by more than or equal to four chromosomal aberrations.

Cohort B included 95 cases with 3 to 6 *ERG* copies. Of the 95 cases, 41 had a non-complex aberrant karyotype with gain of one (n=40) or two (n=1) chromosomes 21 (further referred as to AK-AML with gain of chr21). The remaining 54 cases had a complex aberrant karyotype with gain of one (n=43), two (n=9), or three (n=2) chromosomes 21. Seven of the latter 54 cases had additional structural aberration involving chromosome arm 21q. *ERG* copy number state of these seven cases were validated by interphase FISH, all of them demonstrating less than or equal to 6 copies. This group is further referred to as to CK-AML with gain of chr21/chr21q.

For comparison, two control cohorts were analyzed. One cohort included the 325 patients with CN-AML, which had been characterized for *BAALC* and *ERG* gene expression.[96] The other cohort included 26 AML patients with a CK but without cytogenetically visible gain of chromosome 21.

2.5 Statistical analysis

The definition of OS was the time from diagnosis to death or last follow-up. EFS was defined as the time from diagnosis to treatment failure, relapse, death, or last follow-up. Survival curves were calculated for OS and EFS according to Kaplan-Meier and compared using the two-sided log rank test. Cox regression analysis was performed for OS and EFS with different parameters as covariates. Median follow-up was calculated taking the respective last observations in surviving cases into account and censoring non-surviving cases at the time of death. Results were considered significant at $p<0.05$ for univariate analyses and at $p<0.1$ for multivariate analyses. Parameters which were significant in univariate analyses were included into multivariate analyses. Dichotomous variables were compared between different groups using the Fisher’s exact test and continuous variables by Student’s t-test. Correlation coefficient was specified as Spearman’s rank correlation. All reported p-values are two-sided. No adjustments for multiple comparisons were performed. SPSS software version 19.0.0 (IBM corporation, Armonk, NY) was used for statistical analysis.

3 Results and discussion

3.1 *BAALC* and *ERG* gene expression in CN-AML

3.1.1 *BAALC* and *ERG* in CN-AML: range of expression and cut-off definition

In diagnostic CN-AML samples, the normalized levels of % *BAALC*/*ABL1* and % *ERG*/*ABL1* varied within a wide range of at least 4 logs (range *BAALC*/*ABL1*: 0.1% to 8019.9%; range *ERG*/*ABL1*: 0.1%-1007.5%; Appendix I; Appendix IV).[96;151] Even though, the expression levels of *ERG* seems more evenly distributed, these results indicate that the genes are differentially expressed in some CN-AML samples (Figure 2).

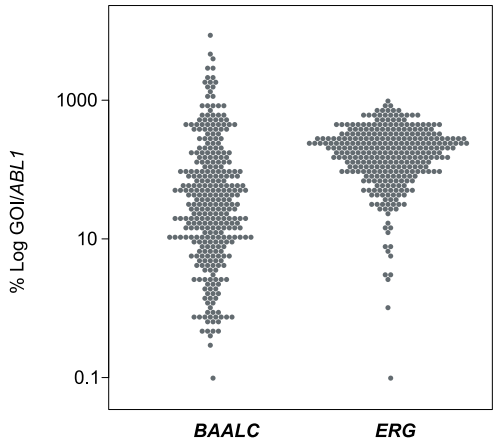


Figure 2: Quantitative analysis of *BAALC* and *ERG* gene expression in CN-AML. Grey circles indicate single cases. The y-axis depicts the % gene of interest (GOI)/*ABL1* on a logarithmic scale.

The translation of gene expression into clinical decision making requires the definition of a cut-off level. Using a survival based method, the level of 33.1% *BAALC*/*ABL1*, accounting for median *BAALC* expression of the total CN-AML cohort, was found to separate high-risk from low-risk patients (Figure 3, Appendix I).[96] This threshold was further validated in Weber et al.[152] where a high proportion (87%, 13/15) of patients with *BAALC* expression levels above 33.1% *BAALC*/*ABL1* during therapy experienced relapse (Appendix III). These results partly reflect the literature, where despite differences in the quantification methods, mean expression levels were used to separate low from high *BAALC* expressers.[14-16;95] However, different cut-off levels have also been suggested by other groups using distinctive approaches to define them.[97-99] Using the survival based approach to define a threshold for *ERG* expression the results were less persuasive (Figure 3). Addressing these result as well as the fact that dichotomization on survival based methods represents a data driven classification,

BAALC and *ERG* expression levels were additionally analyzed as continuous variables, if feasible.

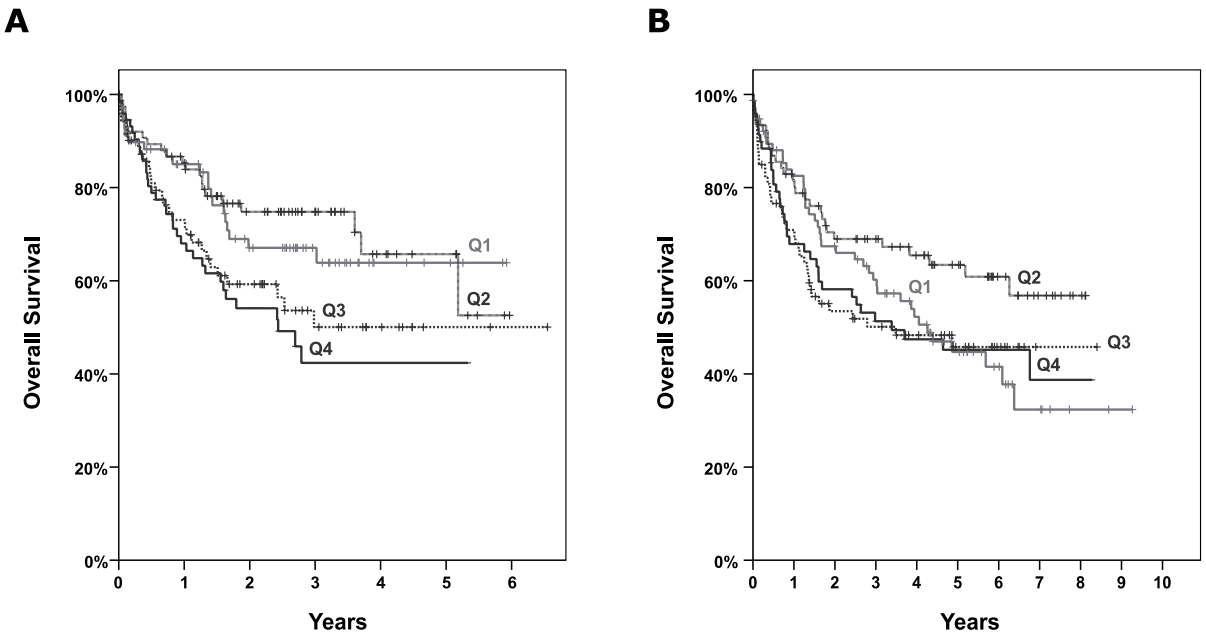


Figure 3: Patients analyzed for *BAALC* and *ERG* expression were initially divided into quartiles (Q1-Q4) according to the levels of expression. (A) Kaplan-Meier curve for *BAALC* expression. Median was chosen for the *BAALC* expression cut-off because the subgroup of Q3 and Q4 showed a clinically distinct outcome with an inferior OS compared with the remaining quartiles (Q1 and Q2). Patients were defined as low *BAALC* with expression levels in Q1 and Q2 and as high *BAALC* with expression levels in Q3 and Q4. (B) Kaplan-Meier curve for *ERG* expression.

3.1.2 Correlation of diagnostic *BAALC* and *ERG* gene expression to clinical and molecular parameters

In terms of patients characteristics at the time of diagnosis, for both, *BAALC* and *ERG*, a negative correlation of gene expression levels to age was revealed, meaning that patients with high *BAALC* and/or *ERG* expression tended to be younger than the respective patients with low *BAALC* and/or *ERG* expression (Appendix I; Appendix IV).[96;151]

With regard to molecular alterations, substantial associations of altered *BAALC* and *ERG* expression to specific molecular aberrations were observed. Figure 4 provides an overview of the analyzed associations between the expression of *BAALC* and *ERG* and the remaining molecular parameters, which have been reported by our group (Appendix I; Appendix IV).[96;151] Analysis of *BAALC* as a categorical variable revealed a strong correlation of high *BAALC* expression to the following molecular alteration: *FLT3*-ITD, especially when *FLT3*-ITD to *FLT3* wildtype ratios were high (≥0.5; further termed *FLT3*-ITD≥0.5), *MLL*-PTD, *IDH2*R172 and to

mutations in *RUNX1*, *CEBPA* and *WT1*. Further, a negative correlation of high *BAALC* expression to mutations in *NPM1* was observed. These correlations were confirmed in terms of *NPM1*, *RUNX1* and WT1 when analyzing *BAALC* expression as a continuous variable. Analyzing *ERG* as a continuous variable, higher *ERG* expression levels were significantly associated with high *BAALC*, *FLT3*-ITD≥0.5 and mutations in *WT1* as well as with the absence of mutations in *IDH1* and *NPM1*. These results corroborate the literature as high *BAALC* expression has been demonstrated to correlate with the mutational status of *FLT3*-ITD, *CEBPA*, *MLL*-PTD, *RUNX1* and *WT1* as well as with *NPM1* wildtype.[14-16;84;95;153] Also for *ERG*, these results were consistent with the literature in terms of *BAALC* and *FLT3*-ITD, though *ERG* has been analyzed as a categorical parameter in these previous studies.[16;95;119]

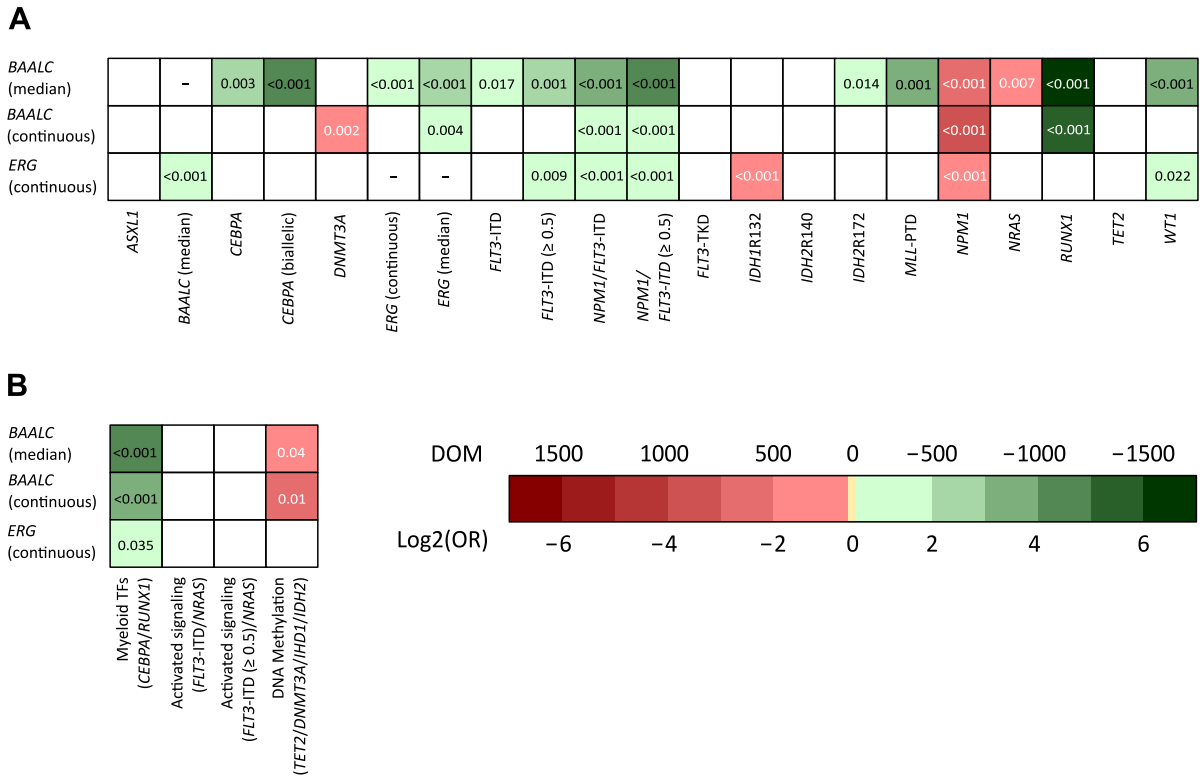


Figure 4: Schematic representation of associations between *BAALC* and *ERG* gene expression and the remaining molecular parameters as reported in Weber et al., 2014[96] and Weber et al., 2016[151] (Appendix I; Appendix IV); edited by Dennis Haupt. Statistical significance was assessed using the Fisher's exact test for categorical variables and the Student's t-test for continuous variables. The difference of the mean (DOM) as well as the binary logarithm of the odds ratio (Log2(OR)) of the association is color coded, and the p-value is given in each field. Green colors indicate a positive association (i.e., 2 characteristics that frequently occurred together in the same patient, or for the association to *BAALC* and *ERG* as continuous variable, a higher value in those carrying the respective molecular alteration). Red colors indicate a negative association. White color indicates no significant association between the tested variables. (A) Pairwise associations between *BAALC* and *ERG* and other

molecular alterations, including the molecular risk group of *NPM1* wildtype or *FLT3*-ITD≥0.5. (B) Pairwise associations between *BAALC* and *ERG* and the functional biological categories.

Despite some shared associations individual correlations of altered *BAALC* and *ERG* expression were observed, presuming that both genes are associated to distinct functional biological categories or, addressing *FLT3*-ITD and *NPM1,* to clinically defined subgroups. Regarding the molecular risk group of *NPM1* wildtype or *FLT3*-ITD≥0.5, both mean *BAALC* and mean *ERG* expression levels were significantly higher as compared to the low-risk group (Figure 4A). Grouping the molecular alterations into functional biological categories, *BAALC* expression revealed distinct associations. For instance, higher *BAALC* expression levels were significantly related to a mutated status in the myeloid transcription factor group (*CEBPA* and *RUNX1*), while substantially lower *BAALC* expression levels were observed in patients harboring mutations in genes involved in DNA methylation, including *DNMT3A*, *TET2*, *IDH1* and *IDH2* (Figure 4B). On the other hand, the association of *ERG* expression to functional biological pathways appeared less pronounced, as only slightly higher *ERG* expression levels were found in patients harboring mutations in one of the myeloid transcription factors, *CEBPA* and *RUNX1* (Figure 4B). Interestingly, aside from the strong correlation to *FLT3*-ITD neither *ERG* expression levels nor *BAALC* expression levels revealed a significant correlation to the activated signaling/proliferation group (Figure 4B; Appendix I; Appendix IV).[96;151]

Taken together, *ERG* expression seems to be associated to the clinically defined subgroup and to age-associated single alterations, but not to specific biological pathways. Contrary to high *BAALC* expression, which seems to gather with certain subgroups of CN-AML, as strong correlations with mutations in transcription factors and genes that induce proliferation, but no or a negative correlation with mutations in epigenetic modifiers were found. An exception represents the epigenetic modifier *IDH2*R172, which is strongly correlated with high *BAALC* expression. This association is of special interest as *IDH2*R172 forms a clinically and molecularly distinct subset in CN-AML as compared to *IDH1*R132 and *IDH2*R140.[54;154] Unlike *IDH1*R132 and *IDH2*R140, which show several co-mutations (mainly with *NPM1*, *FLT3*, and *RAS*), *IDH2*R172 are nearly mutually exclusive of other common molecular alterations in CN-AML. Thus, high *BAALC* expression could represent a co-operating event synergizing in the onset of leukemia further defining the subgroup of *IDH2*R172 mutated AML. Another association with suspected biological consequences is the strong correlation of high *BAALC* expression to mutations in the transcription factor *RUNX1* (Appendix I).[96] Previously, a search for putative activations sides identified binding signatures for RUNX1 in the upstream regulatory machinery of *BAALC* and subsequent in vitro assays revealed an direct effect of overexpression or knockdown of RUNX1 on the expression of *BAALC* in leukemia derived cell

lines.[93;94] However, so far, it remains elusive, if mutated *RUNX1* could also account for elevated *BAALC* expression, since at least some mutations in *RUNX1* have been reported to lead to a loss of protein function by disruption of its DNA binding ability and by now, no correlation of high *BAALC* expression to a distinct *RUNX1* mutation was described (Appendix I).[96;155] Functional studies, which investigate the suspected biological synergisms between *BAALC* expression and *RUNX1* or *IDH2*R172 could provide precious insights in the biology and leukemogenesis of AML.

3.1.3 Prognostic value of *BAALC* and *ERG* expression as single marker

Survival analyses revealed an adverse prognostic impact of high *BAALC* on OS and EFS either as categorical variable (Figure 5) and moreover also as continuous variable (OS: HR 1.27, p=0.009; EFS: HR 1.32, p<0.001; Appendix IV).[151] These results corroborate the literature.[15;89;95;97-99]

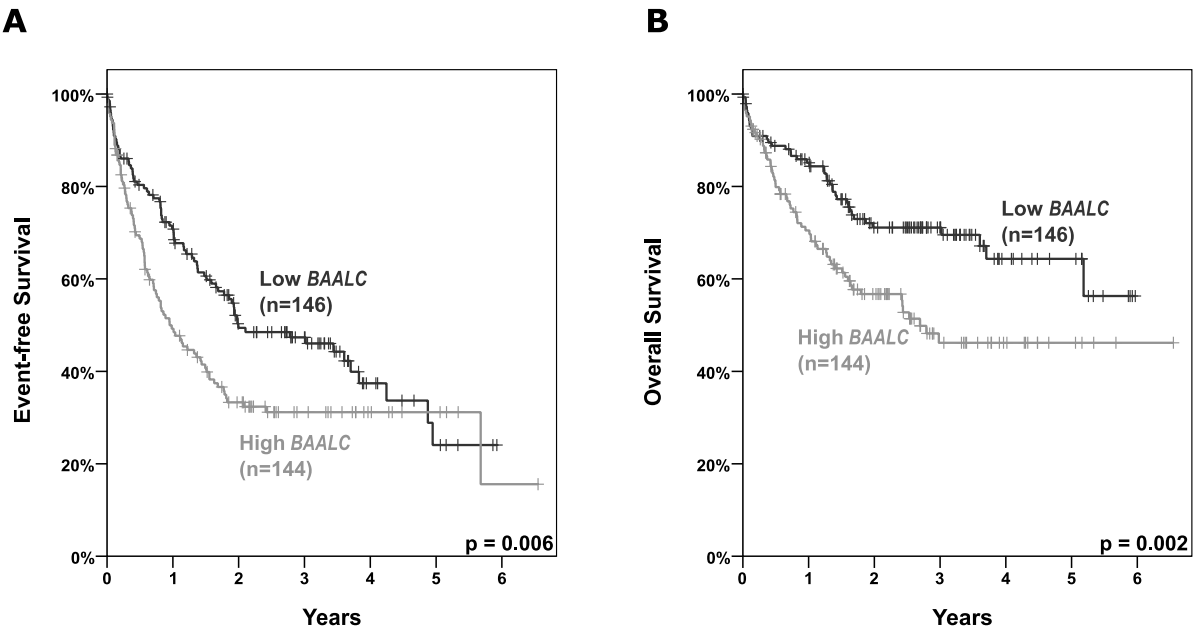


Figure 5: Outcome of 290 intensively treated CN-AML patients aged younger than 65 years with respect to *BAALC* expression as published in Weber et al.[96] (Appendix I). The median expression level was used to dichotomize the total patient cohort into low (black) and high (gray) *BAALC* expressers. (A) EFS at 3 years: 31.2% vs. 47.4%, p=0.006. (B) OS at 3 years: 46.2% vs. 71.1%, p=0.002.

ERG expression levels as a continuous parameter did neither affect OS nor EFS. However, when dichotomizing *ERG* expression at distinct cut-off levels (25th percentile, median, 75th percentile) a significant correlation to shorter EFS and a trend towards inferior OS was observed for *ERG* expression levels above the median (Figure 6; Appendix IV).[151] This is in line

with the literature, as one study analyzing *ERG* expression as a continuous parameter did not reveal an impact of *ERG* expression levels on survival,[156] while other studies where *ERG* expression has been dichotomized at certain cut-off levels (median or 75th percentile) showed an association of high *ERG* expression with inferior outcome.[16;95;118;119]

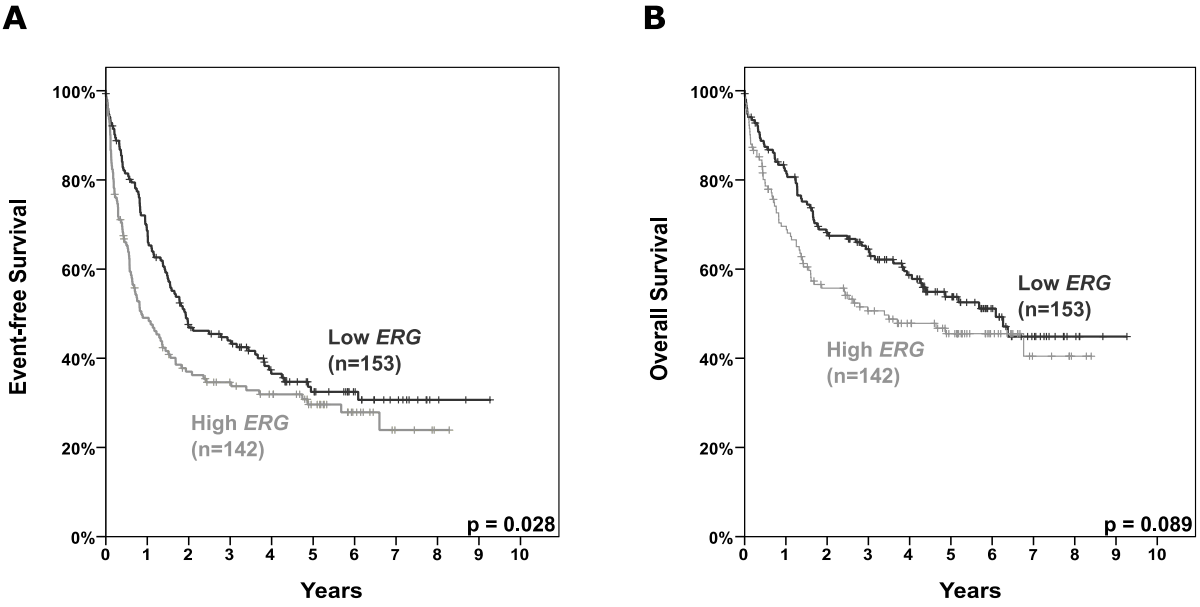


Figure 6: Outcome of 295 intensively treated CN-AML patients aged younger than 65 years with respect to *ERG* expression. The median expression level was used to dichotomize the total patient cohort into low (black) and high (gray) *ERG* expressers. (A) EFS at 3 years: Low *ERG*: 44.0% vs. high *ERG*: 34.6%, p=0.028, (B) OS at 3 years: Low *ERG*: 64.5% vs. high *ERG*: 50.6%, p=0.089 (Appendix IV).[151]

3.1.4 Prognostic value of *BAALC* and *ERG* in the context of other molecular alterations

3.1.4.1 Prognostic value of *BAALC* and *ERG* within the ELN genetic low- and intermediate-risk group defined by mutational status of *NPM1* and *FLT3*-ITD

As *BAALC* and *ERG* expression have been shown to strongly correlate with *NPM1* wildtype and *FLT3*-ITD,[96;151] the prognostic value in the respective low- and intermediate-risk groups was assessed. Regarding *ERG* expression dichotomized at the median, no differences in EFS and OS was observed neither for the low-risk nor for the intermediate-risk group (Appendix IV).[151] This contrasts the study of Metzeler et al.[16], where *ERG* expression dichotomized at the 75th percentile provided additional prognostic information in the intermediate-risk group (*NPM1* wildtype and/or *FLT3*-ITD). This discrepancy most probably results from differences in the analyzed cohorts, as in the study of our group younger AML patients (age 18-65) were investigated, while Metzeler et al.[16] analyzed patients aged 17-83 years. Besides differences in treatment protocols of younger and older AML patients, which itself could account for

varied outcome several studies have shown that the mutational profile as well as its prognostic information differs between younger and older AML patients.[34;36;68;86]

Addressing the impact of *BAALC* expression on the intermediate-risk group, *BAALC* expression (dichotomized at the median) provided additional prognostic information on EFS and OS in the intermediate-risk group of *NPM1* wildtype or *FLT3*-ITD (Figure 7; Appendix IV).[151] An impact of *BAALC* expression on the intermediate-risk group has previously been suggested in a preliminary cohort of 29 AML patients with intermediate-risk chromosomal/molecular abnormalities.[157] Based on these findings a refined algorithm for risk stratification in CN-AML can be proposed, as patients with *NPM1* wildtype or *FLT3*-ITD and high *BAALC* expression rather reflect OS of the ELN intermediate II-risk group (Table 2), while the respective low *BAALC* expressers resemble outcome of the favorable-risk group.[158] Therefore future clinical trials should further evaluate if these patients can benefit from reclassification.

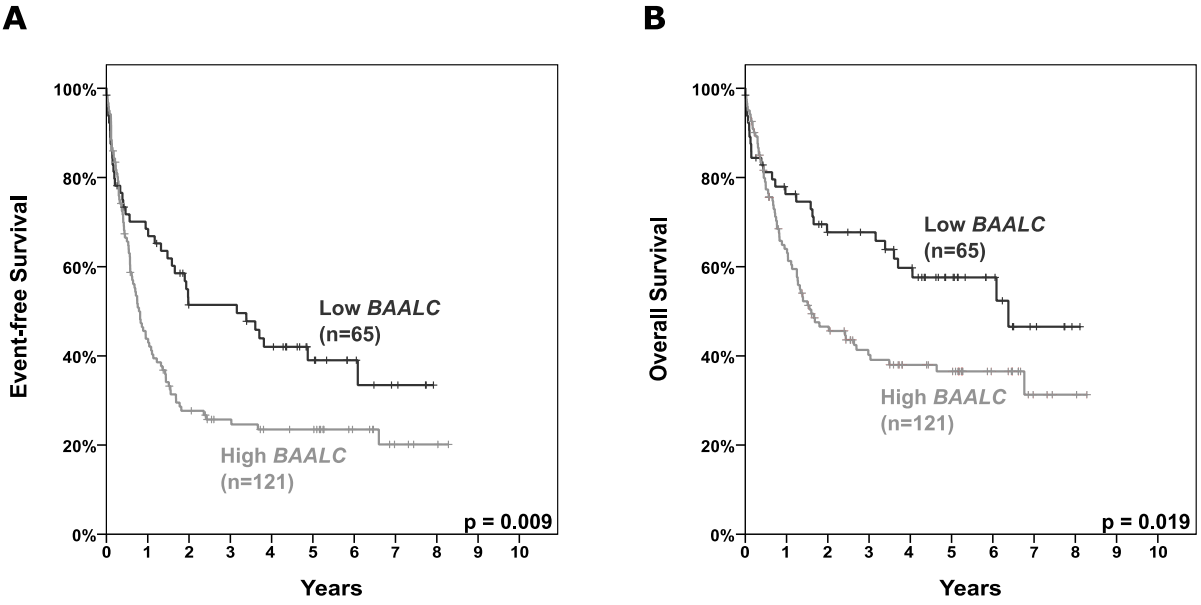


Figure 7: Outcome in the intermediate-risk group of patients with *NPM1* wildtype or *FLT3*-ITD with respect to *BAALC* expression (n=186). The median expression level was used to dichotomize the total patient cohort into low (black) and high (gray) *BAALC* expressers. (A) EFS at 3 years: Low *BAALC*: 51.4% vs. high *BAALC*: 25.7%, p=0.009. (B) OS at 3 years: Low *BAALC*: 67.7% vs. high *BAALC*: 40.2%, p=0.019 (Appendix IV).[151]

3.1.4.2 Multivariate analysis to identify independent prognostic factors in CN-AML

Facing the number of correlations observed between various molecular risk markers, Cox regression analysis was performed to identify those factors that independently predicted prognosis in CN-AML.

In multivariate analysis, high *BAALC* expression, but not *ERG* expression, revealed an independent prognostic impact on EFS and OS. Additional independent factors, besides high *BAALC* expression, were *FLT3*-ITD≥0.5, *MLL*-PTD and *WT1* mutations for EFS (Table 4) as well as *ASXL1* mutations, *FLT3*-ITD≥0.5, *MLL*-PTD and *WT1* mutations for OS (Table 5; Appendix IV).[151]

Table 4: Cox regression analyses for event-free survival in CN-AML

	Univariate			Multivariate		
	HR	***p*-value**	**95% CI**	**HR**	***p*-value**	**95% CI**
Age	1.26*	<0.001	1.12-1.40	1.39*	<0.001	1.24-1.55
***ASXL1*mut**	1.86	0.046	1.01-3.43	-	n.s.	-
Log *BAALC* expression	1.32	<0.001	1.13-1.53			
High *BAALC* (median)	1.68	<0.001	1.27-2.24	1.44	0.024	1.05-1.97
***CEBPA*biallelic**	-	n.s.	-			
***DNMT3A*mut**	1.28	0.083	0.97-1.71			
Log *ERG* expression	-	n.s.	-			
High *ERG* (median)	1.34	0.030	1.03-1.82	-	n.s.	-
High *ERG* (75th percentile)	-	n.s.	-			
***FLT3*-ITD**	-	n.s.	-			
***FLT3*-ITD(≥0.5)**	1.69	0.002	1.22-2.34	1.57	0.012	1.11-2.23
***NPM1*wildtype or *FLT3*-ITD**	1.41	0.021	1.05-1.90			
***NPM1*wildtype or FLT3-ITD(≥0.5)**	1.60	0.001	1.20-2.13			
***FLT3*-TKD**	-	n.s.	-			
***IDH1*R132mut**	-	n.s.	-			
***IDH2*R140mut**	-	n.s.	-			
***IDH2*R172mut**	-	n.s.	-			
***MLL*-PTD**	1.70	0.043	1.02-2.84	1.67	0.057	0.99-2.84
***NPM1*mut**	0.77	0.078	0.58-1.03			
***NRAS*mut**	-	n.s.	-			
***RUNX1*mut**	-	n.s.	-			
***TET2*mut**	-	n.s.	-			
***WT1*mut**	2.18	0.000	1.41-3.38	2.47	<0.001	1.54-3.98

*Per 10 years of increase. Abbreviations: ITD, internal tandem duplication; TKD, tyrosine kinase domain; PTD, partial tandem duplication; mut, mutation; HR, Hazard ratio; CI, Confidence interval; n.s., not significant.

Table 5: Cox regression analyses for overall survival in CN-AML

	Univariate			Multivariate		
	HR	***p*-value**	**95% CI**	**HR**	***p*-value**	**95% CI**
Age	1.38*	<0.001	1.21-1.56	1.53*	<0.001	1.34-1.73
***ASXL1*mut**	2.39	0.012	1.21-4.72	2.47	0.012	1.22-4.98
Log *BAALC* expression	1.27	0.009	1.06-1.52			
High *BAALC* (median)	1.59	0.007	1.14-2.22	1.36	0.099	0.95-1.95
***CEBPA*biallelic**	-	n.s.	-			
***DNMT3A*mut**	-	n.s.	-			
Log *ERG* expression	-	n.s.	-			
High *ERG* (median)	1.33	0.090	0.96-1.85			
High *ERG* (75th percentile)	-	n.s.	-			
***FLT3*-ITD**	1.65	0.003	1.18-2.30			
***FLT3*-ITD(≥0.5)**	2.15	<0.001	1.50-3.08	2.28	<0.001	1.55-3.36
***NPM1*wildtype or *FLT3*-ITD**	1.79	0.002	1.25-2.56			
***NPM1*wildtype or FLT3-ITD(≥0.5)**	1.79	0.001	1.28-2.52			
***FLT3*-TKD**	-	n.s.	-			
***IDH1*R132mut**	-	n.s.	-			
***IDH2*R140mut**	-	n.s.	-			
***IDH2*R172mut**	-	n.s.	-			
***MLL*-PTD**	2.46	0.001	1.46-4.15	2.53	0.001	1.47-4.34
***NPM1*mut**	-	n.s.	-			
***NRAS*mut**	-	n.s.	-			
***RUNX1*mut**	-	n.s.	-			
***TET2*mut**	-	n.s.	-			
***WT1*mut**	1.95	0.010	1.18-3.25	2.57	0.001	1.46-4.52

*Per 10 years of increase. Abbreviations: ITD, internal tandem duplication; TKD, tyrosine kinase domain; PTD, partial tandem duplication; mut, mutation; HR, Hazard ratio; CI, Confidence interval; n.s., not significant.

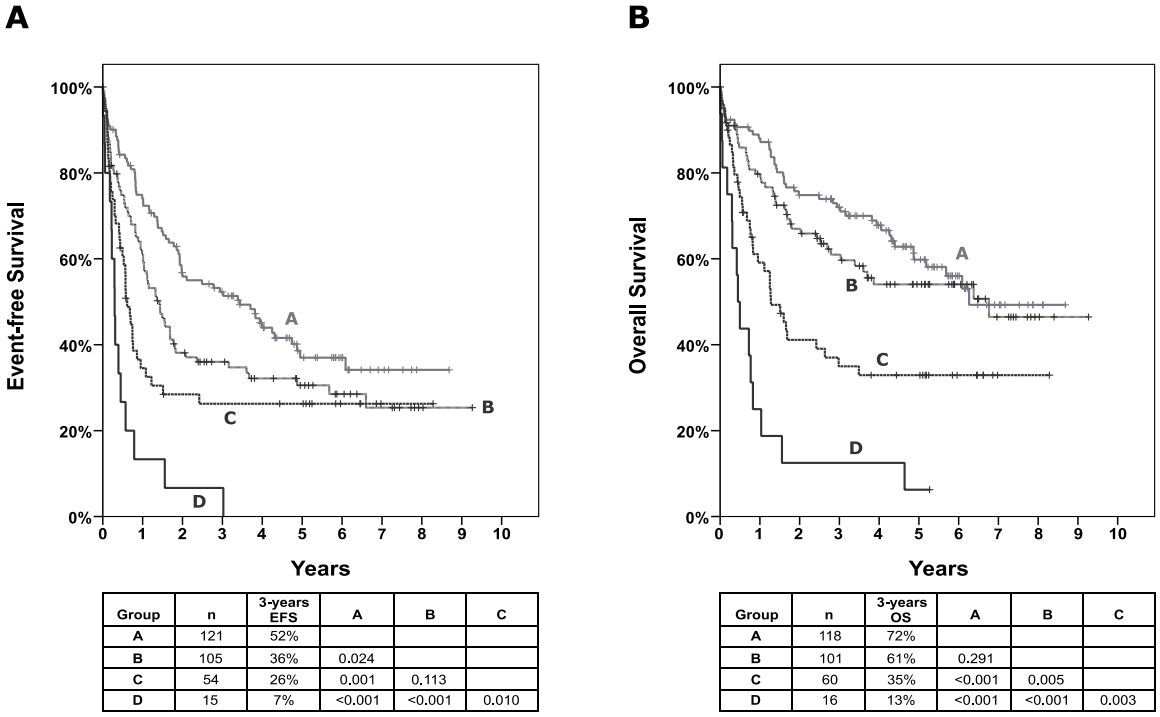
To clarify whether these adverse prognostic markers exhibit an additive effect on survival, the number of independent adverse prognostic parameters for each patient were determined. This resulted in the formation of four subgroups according to the number of the adverse prognostic factors. Included were high *BAALC*, *FLT3*-ITD≥0.5, *MLL*-PTD and *WT1*mut for EFS as well as *ASXL1*mut, high *BAALC*, *FLT3*-ITD≥0.5, *MLL*-PTD and *WT1*mut for OS; with group A: no adverse marker, group B: 1 adverse marker, group C: 2 adverse markers, group D: 3 or 4 adverse markers. None of the patients harbored concomitant alterations in all 5 adverse prognostic factors (Appendix IV).[151]

Kaplan-Meier analysis revealed that the sole accumulation of these adverse prognostic markers stepwise worsened the prognosis (Figure 8; Appendix IV).[151]

Therefore, despite the frequent co-occurrence of some of these adverse markers, they correlated independently and additively with survival (Appendix IV).[151] Similar results have

been shown just previously, were the number of driver mutations, including fusion genes, aneuploidies, gene mutations and small insertions or deletions, correlated with overall survival.[154] These results indicate that, beside the genetic context which defines functional biological and moreover prognostic categories, the sole accumulation of some molecular aberrations influence clinical outcome. This, in part, reflects the situation of cytogenetic aberrations, where specific alterations as well as the accumulation of ≥3 cytogenetic alterations can be allocated to clinical outcome.[11] Therefore, prospective studies are needed to clarify to what extent the described markers influence each other or a given genetic context and which of the markers are beneficial compared to others in terms of prognostic allocation in AML.

Conclusively, these results suggest that a comprehensive screening of molecular genetic alterations provide additional information for risk assessment in CN-AML.



Group	n	3-years EFS	A	B	C
A	121	52%			
B	105	36%	0.024		
C	54	26%	0.001	0.113	
D	15	7%	<0.001	<0.001	0.010

Group	n	3-years OS	A	B	C
A	118	72%			
B	101	61%	0.291		
C	60	35%	<0.001	0.005	
D	16	13%	<0.001	<0.001	0.003

Figure 8: Outcome of 295 intensively treated CN-AML (younger 65 years) at 3 years in the four subgroups allocated according to the number of adverse prognostic markers: group A (no adverse marker), group B (1 adverse marker), group C (2 adverse markers), group D (≥3 adverse markers) (Appendix IV).[151]

3.2 Analysis of *BAALC* gene expression for detection of MRD

Detection of residual leukemia cells during intensified therapies using qPCR-based systems has been found to provide a high prognostic value in AML.[10;128;129] Besides our studies (Appendix I; Appendix III),[96;152] only one prior study has addressed the molecular analysis of *BAALC* expression as a marker to detect MRD.[98] This study indicated the applicability of *BAALC* as an MRD target in a cohort of 34 AML and 11 ALL patients.

3.2.1 Fundamental validation of the feasibility of *BAALC* expression for MRD detection

Since there was no fundamental work establishing the utility of *BAALC* expression for monitoring MRD in CN-AML, these analyses were performed in a first step.

The analysis of serial follow-up PB and BM samples taken from nine patients, in whom diagnostic samples exhibited low *BAALC* gene expression levels, revealed no significant difference of the normalized % *BAALC*/*ABL1* levels during treatment (mean± standard error of the mean (SEM) % *BAALC*/*ABL1* at diagnosis vs. mean±SEM % *BAALC*/*ABL1* at 1st complete molecular remission: 6.2±2.2% vs. 13.8±3.0%, p=0.082). These levels remained below the clinically defined threshold of 33.1% *BAALC*/*ABL1*, indicating that *BAALC* gene expression is not in general modulated by the treatment regimen or within hematopoietic recovery (Appendix I).[96]

In order to confirm the stability of *BAALC* gene expression during follow-up, parallel assessment of diagnostic and relapse samples of 35 CN-AML patients was performed. Despite a high proportion of clonal evolution found in 49% of paired diagnostic and relapse samples as determined by either karyotype instability or the instability of concomitant molecular alterations, a stable high *BAALC* expression was revealed in 31 of these 35 patients at relapse. Mean *BAALC* expression levels did not differ in paired samples between diagnosis and relapse in these cases (mean±SEM % *BAALC*/*ABL1* at diagnosis vs. relapse: 602±209% vs. 312±64%, p=0.157; Appendix III).[152] Therefore, irrespective of the high level of clonal evolution detected in AML shown by us as well as in several other studies,[159-161] *BAALC* expression showed stable expression in 89% of the patients analyzed. This is in good consistency with the literature, where the stability of the commonly used MRD marker has been described to range between 78%-93%.[135;137;162-164]

To further evaluate the utility of *BAALC* expression to monitor therapy response, *BAALC* expression levels of diagnostic and follow-up samples were correlated to either the mutational status or the expression levels of accompanying, well-known MRD markers: *FLT3*-ITD, *MLL*-PTD, *NPM1*mut and *RUNX1*mut. A strong correlation of % *RUNX1*mut and % *MLL*-PTD/*ABL1*

with % *BAALC*/*ABL1* levels (n=42, r=0.824, p<0.001; n=64, r=0.750, p<0.001, respectively) was found. For % *NPM1*mut/*ABL1* and *FLT3*-ITD expression with % *BAALC*/*ABL1* levels (n=257, r=0.500, p<0.001 and n=181; r=0.568, p<0.001, respectively) the correlation was less consistent (Appendix III).[152] The conflicting results on *NPM1*mut and *FLT3*-ITD can be explained by the relatively low level *BAALC* expression found in *NPM1*mut patients (Appendix I),[96] which restricts *BAALC* expression assessment to one log range, as well as on the semi-quantitative approach used to detect *FLT3*-ITD, which does not precisely detect very high *FLT3*-ITD expression levels.

More detailed information on these results are given in the respective publications (Appendix I; Appendix III).[96;152]

3.2.2 Prognostic value of *BAALC* expression levels during therapy

To evaluate the prognostic impact of *BAALC* expression levels during therapy, survival analyses were performed at two clinically defined time points: after the second cycle of induction chemotherapy (before start of consolidation chemotherapy) as well as between 3 and 6 months after completion of induction therapy in 46 and 33 patients, respectively. It is of note that patients with early relapse or persistent disease were excluded from this analysis, as these patients are already detected by the routine cytomorphologic assessment and would therefore not profit from more sensitive assessments.

First, Kaplan-Meier analysis was performed using the previously defined threshold of 33.1% *BAALC*/*ABL1* to separate low from high *BAALC* expression. At both time points analyzed, high *BAALC* expression was found to be associated with shorter EFS. The estimated 1-year EFS for high vs. low *BAALC* expressers was 47% vs. 70% (p=0.026) after second induction and 40% vs. 82% (p=0.021) within 3 to 6 months after completion of induction therapy (Figure 9; Appendix III).[152]

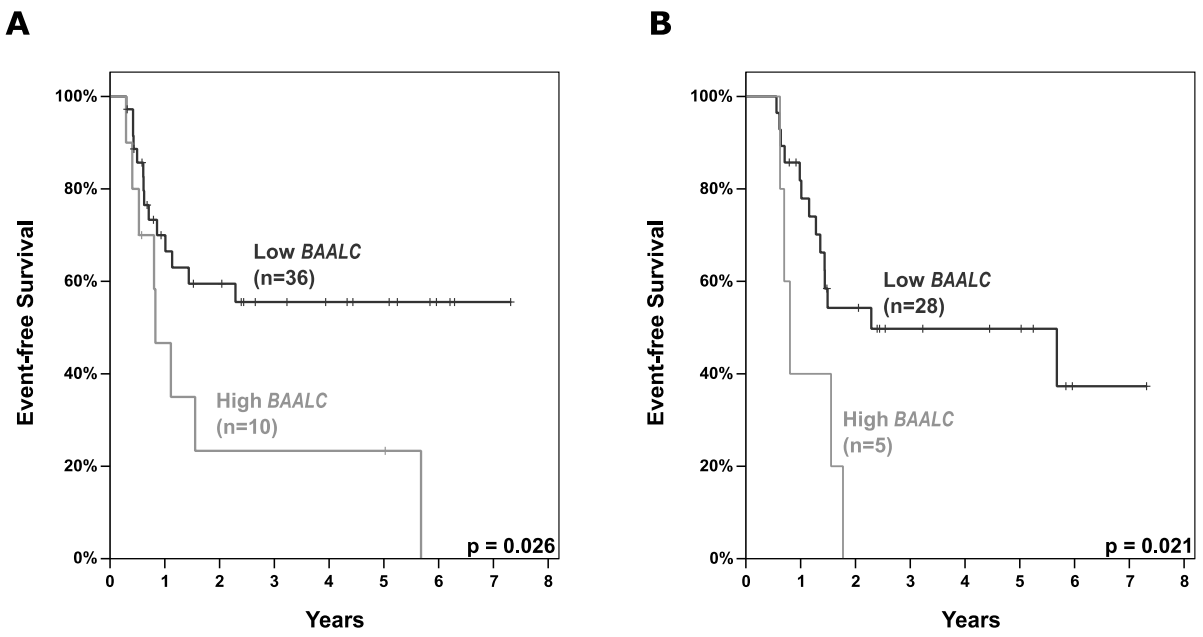


Figure 9: EFS in patients during treatment according to *BAALC* expression as published in Weber et al. [152] (Appendix III). The threshold of 33.1% *BAALC*/*ABL1* was used to dichotomize patients into low (black) and high (grey) *BAALC* expressers. (A) *BAALC* expression levels were assessed in 46 patients after second cycle of induction chemotherapy. EFS at 1 year: Low *BAALC*: 70% vs. high *BAALC*: 47%, p=0.026. (B) *BAALC* expression levels were assessed in 33 patients in whom samples within 3 to 6 months after completion of induction therapy were available. EFS at 1 year: Low *BAALC*: 82% vs. high *BAALC*: 40%, p=0.021.

Subsequently, Cox regression analysis was performed considering *BAALC* expression as continuous variable. After the second cycle of induction chemotherapy, an increase of *BAALC*/*ABL1* levels revealed a significant negative impact on EFS and OS (Table 6). Besides *BAALC*, the prognostic value of age, BM blast count, WBC count and *NPM1*mut transcript level (n=18) were analyzed. Interestingly, *NPM1*mut transcript level was the only factor besides *BAALC* expression level revealing a trend towards inferior OS. In multivariate analysis adjusting for *BAALC* and *NPM1*mut, *NPM1*mut retained its prognostic value on OS (Table 6; Appendix III).[152]

Addressing the prognostic value of continues *BAALC* expression within 3 to 6 months after completion of induction chemotherapy, an increase of 10% *BAALC*/*ABL1* levels as well as higher *NPM1*mut transcript level were again significantly associated with shorter EFS (Table 6). For OS none of the molecular parameters revealed a significant impact, while for higher age a significant association towards inferior OS was found. In multivariate analysis on EFS adjusting for *BAALC* and *NPM1*mut, *NPM1*mut transcript level retained its prognostic value on EFS (Table 6; Appendix III).[152]

Taken together, this data depicts a strong correlation of high *BAALC* expression levels at certain clinically defined time points to adverse clinical outcome as published in *Weber et al.*[152] (Appendix III).

Table 6: Cox regression analyses for event-free and overall survival during therapy

	Event-free survival				**Overall survival**			
	univariate		**multivariate**		**univariate**		**multivariate**	
Variable	HR	p	HR	p	HR	p	HR	p
after 2nd cycle of induction chemotherapy (before consolidation)								
***BAALC* expression (n=46)**	1.14[b]	0.002	-	-	1.12[b]	0.045	1.14	0.289
***NPM1*mut (n=18)**	1.88	0.129	-	-	2.64	0.057	2.82	0.049
within 3 to 6 months after completion of induction therapy								
Age (n=33)	-	n.s.	-	-	1.63[a]	0.030	-	-
***BAALC* expression (n=33)**	1.11[b]	0.004	1.21[b]	0.434	-	n.s.	-	-
***NPM1*mut (n=14)**	1.26	0.046	1.28	0.039	-	n.s.	-	-

[a]Per 10 years of increase; [b]10% *BAALC*/*ABL1* increase; Abbreviations: HR, hazard ratio; mut, mutation.

Focusing more closely on the patients who experience relapse, in fact, eight of ten patients revealing high *BAALC* expression levels after completion of induction therapy experienced relapse with a median of 7 months from sampling to relapse. This was confirmed when analyzing *BAALC* expression levels within 3 to 6 months after completion of induction therapy where all five patients with high *BAALC* expression levels relapsed within a median of 3 months from sampling. Interestingly, five of the 13 patients with high risk of relapse according to elevated *BAALC* expression at either analysis time point, revealed an MRD negative status according to the accompanying molecular alterations detected at diagnosis in *CEBPA*mut or *FLT3*-ITD at sensitivities of 1% to 5%.

However, in 24 relapsed cases *BAALC* expression assessment failed to predict relapse. The analysis of accompanying molecular alterations revealed that only *NPM1*mut (n=9) and *MLL*-PTD (n=1), but not *CEBPA*mut (n=4), *FLT3*-ITD (n=6) and *RUNX1*mut (n=3) provided additional information by disclosing an MRD positive status (Appendix III).[152]

These results on one hand confirm the literature, since it has been shown that quantitation of *NPM1*mut transcripts represents a highly sensitive method of MRD determination,[135;163] which retains its prognostic information independently of other risk factors.[129] On the other hand, these results suggest that the sensitivity of *BAALC* expression during follow-up is,

besides its expression in PB and BM of healthy individuals, at least comparable to that of *FLT3*-ITD, *RUNX1*mut and *CEBPA*mut detection assays, when analyzed by gene scan or sequencing approaches (Appendix I; Appendix III).[96;152]

Further, addressing the high clonal heterogeneity of AML shown by our group[152;162] (Appendix III), but also by several other studies,[44;45;160;165] it can be postulated that the analysis of several alterations characterizing one specific AML could provide a more accurate assessment of residual leukemic burden.

Conclusively, these results predispose *BAALC*-based residual disease monitoring during therapy as an applicable strategy to detect patients at high risk of relapse, especially in *NPM1* wildtype cases.

3.3 *ERG* gene expression in patients with AML and acquired gain of chromosome 21

Addressing the recently described effects of *ERG* overexpression in the onset of DS-AMKL and other leukemia[107;108] and the high incidence of leukemia in individuals with DS,[109] we reasoned that altered *ERG* expression might be the causative mechanism contributing to the pathogenesis of AML with gain of chromosome 21. To address this hypotheses different AML subtypes with amplification or gain of chromosome 21 were analyzed. All of the data presented in this section has already been published in Weber *et al.*[138] (Appendix II).

3.3.1 Characterization of patients with cytogenetically visible structural aberrations of chromosome 21

First, the characterization of 40 patients with cytogenetically visible structural aberrations of chromosome 21 by FISH encompassing the *ERG* and *RUNX1* loci revealed amplification (>6 copies/cell) of *ERG* in 33 (83%) patients and a gain (3-6 copies) of *ERG* in the remaining 7 patients. Array CGH was used to determine the exact length of the respective amplified chromosome region in the 33 cases with *ERG* amplification. Though the pattern of amplification on chromosome arm 21q was heterogeneous (range of length: 501 kbp to 16336 kbp; range of location: 31,793,799 bp and 48,129,895 bp from the 21pter) it disclosed a minimal amplified region at the position 39.6 to 40.0 Mbp from *pter* that harbors *ERG* as the only gene (Figure 10; Appendix II).[138]

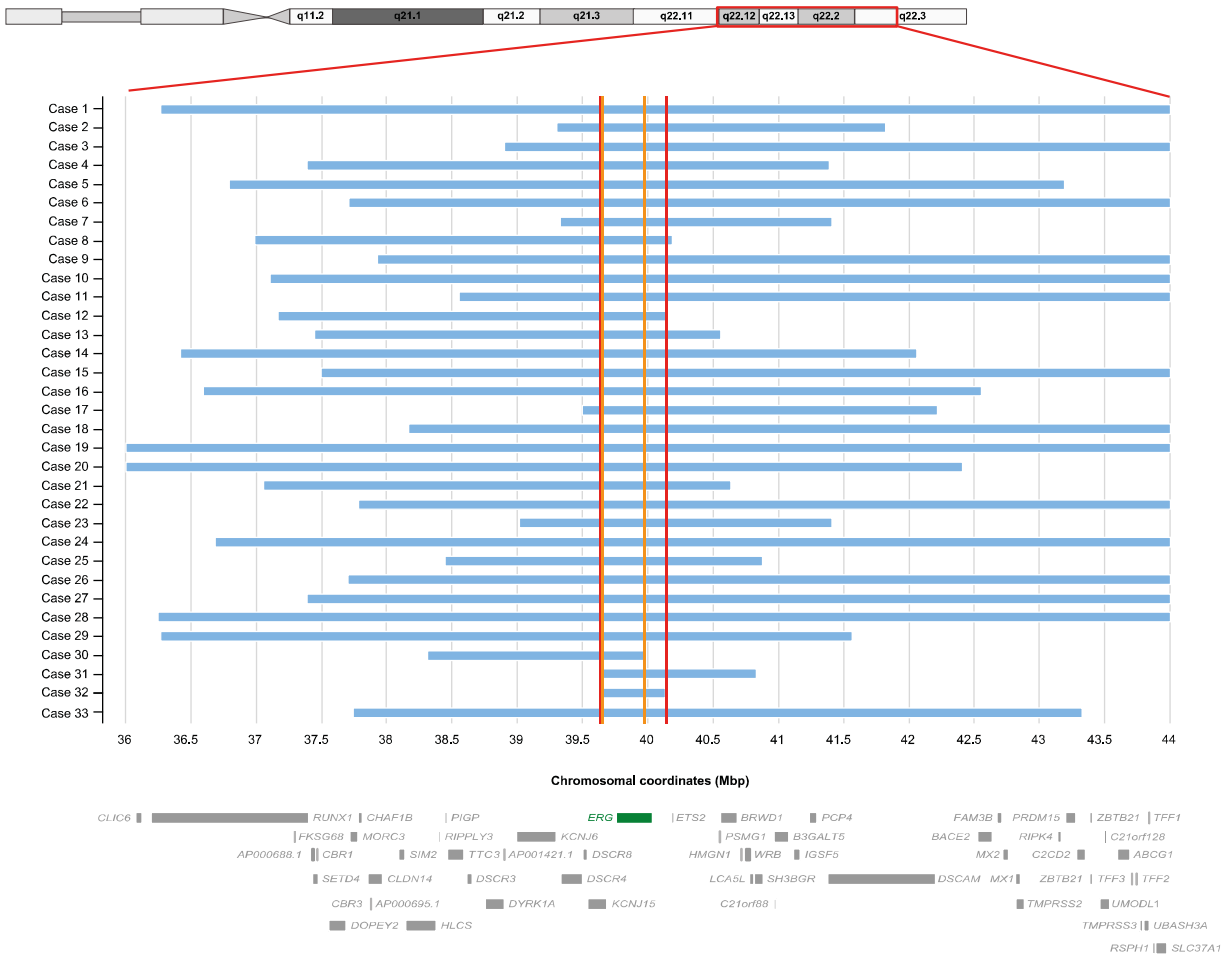


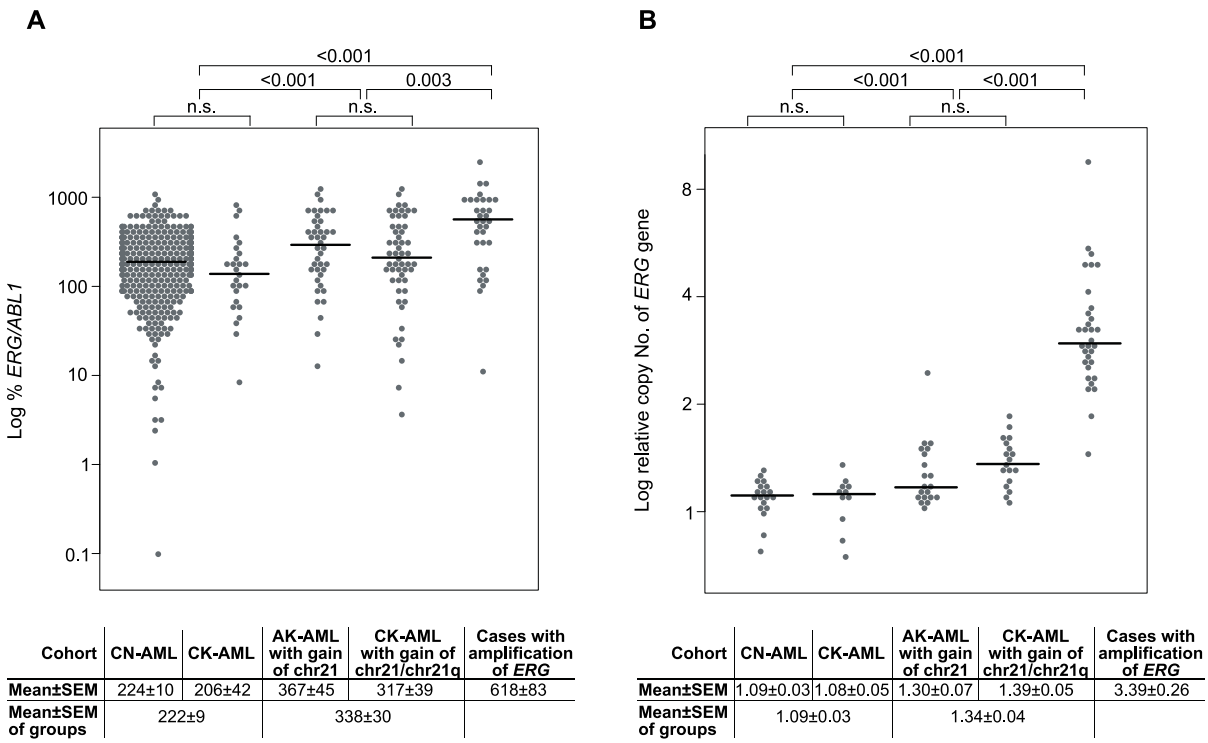
Figure 10: Diagrammatic representation of amplifications on chromosome arm 21q as assessed by array CGH analysis. Array CGH data showing segments of chromosome arm 21q with amplifications (blue horizontal bars) in 33 cases with AML and cytogenetically visible structural aberrations of chromosome 21q and *ERG* amplification. Orange lines represent the minimally amplified region including all 33 cases. Red lines indicate the extended minimally amplified region when excluding case 30. Grey boxes represent the locations of genes within the amplified regions. The *ERG* gene (labeled in green) was the only gene covered by the minimally amplified region. These results have been published in Weber at al.[138] (Appendix II).

3.3.2 Determination of *ERG* gene expression and *ERG* DNA copy number

To assess *ERG* gene expression levels in the 33 patients showing amplifications of *ERG* quantitative real-time reverse transcriptase PCR was performed. Further, *ERG* expression was analyzed in 95 AML cases with gain of chromosome 21. Of these cases 41 had a non-complex aberrant karyotype (AK-AML with gain of chr21) and the remaining 54 cases had a complex aberrant karyotype (CK-AML with gain of chr21/chr21q). In addition, 325 cases with CN-AML and in 26 AML cases with a complex karyotype but without cytogenetically visible gain of chromosome 21 (CK-AML) were analyzed. Detailed information on the cohorts is given in the

material and methods section ("2.4 Patient samples") and in the respective publication (Appendix II).[138]

Although the general expression levels overlapped between the different genetic subgroups, significantly higher mean *ERG* expression levels were found in the 33 patients with amplification of *ERG* (mean±SEM: 618±83% *ERG*/*ABL1*; range: 11-2324% *ERG*/*ABL1*) as well as in 95 patients with gain of chromosome 21 (mean±SEM: 338±30% *ERG*/*ABL1*; range: 3-1214% *ERG*/*ABL1*) compared to the respective controls (mean±SEM: 222±9% *ERG*/*ABL1*; range: 0.1-1007.5% *ERG*/*ABL1*; P<0.001; Figure 11). Furthermore, mean *ERG* expression was also significantly higher in 33 patients with *ERG* amplification compared with the patients with gain of chromosome 21/chromosome 21q (p=0.003; Figure 11), suggesting a marked correlation between *ERG* genomic DNA copy number and *ERG* gene expression. Indeed, the increment of *ERG* gene expression was significantly correlated to the copy number of *ERG* DNA as assessed by qPCR. Pearson product-moment correlation revealed a linear correlation of mean *ERG* expression and mean *ERG* DNA values of the different genetic subgroups (r=0.956, P=0.011; Appendix II).[138] Regarding *ERG* gene expression, this data corroborates the literature since amplification and associated elevated expression of *ERG* has been already described in AML patients with complex karyotypes and abnormal chromosome 21.[166] However, in contrast to the study of Baldus et al.[166] a significant correlation between *ERG* genomic DNA copy numbers and gene expression of *ERG* was observed. This discrepancy might result from different numbers of patients analyzed as well as from different methods used, as in the previous study BAC array CGH and oligonucleotide expression array analyses were used to compare the DNA and mRNA amount of six patients, respectively.[166]



Cohort	CN-AML	CK-AML	AK-AML with gain of chr21	CK-AML with gain of chr21/chr21q	Cases with amplification of *ERG*
Mean±SEM	224±10	206±42	367±45	317±39	618±83
Mean±SEM of groups	222±9		338±30		

Cohort	CN-AML	CK-AML	AK-AML with gain of chr21	CK-AML with gain of chr21/chr21q	Cases with amplification of *ERG*
Mean±SEM	1.09±0.03	1.08±0.05	1.30±0.07	1.39±0.05	3.39±0.26
Mean±SEM of groups	1.09±0.03		1.34±0.04		

Figure 11: Quantitative analysis showing *ERG* gene expression and *ERG* DNA values of the different subgroups (Appendix II).[138] (A) *ERG* gene expression in patients with CN-AML, complex karyotype AML (CK-AML), non-complex aberrant karyotype AML (AK-AML) with gain of chromosome 21 (chr21), and CK-AML with gain of chr21 as well as cases with amplification of *ERG*. Grey dots indicate single cases; black lines indicate mean expression. The y-axis depicts the % *ERG*/*ABL1* on a logarithmic scale; the x-axis depicts the different genetic subgroups. (B) *ERG* copy numbers in patients with CN-AML, CK-AML, AK-AML with gain of chr21, and CK-AML with gain of chr21 as well as cases with amplification of *ERG*. Grey dots indicate single cases; black lines indicate mean DNA values. The y-axis depicts the relative copy number of the *ERG* gene on a logarithmic scale. As a mixture of DNA of ten healthy individuals was used to calibrate the comparative DNA copy numbers a relative quantity of 1 corresponds to 2 DNA copies of the ERG gene. The x-axis depicts the different genetic subgroups.

Besides the linear correlation of the mean *ERG* RNA and mean DNA values, a significant but not linear correlation of *ERG* gene expression and *ERG* DNA values was observed, when correlating the values for individual patient data (r=0.405, P<0.001), which implies that *ERG* expression is additionally influenced (Appendix II).[138] This hypothesis is supported by the fact, that increased *ERG* expression has been observed in some CN-AML cases, in which cryptic gain or amplification of 21q22 has not been described.[16;119;167] In addition, as expression levels did not increase in a linear fashion with an increase in the amount of *ERG* gene amplification, *ERG* expression could also be affected by some kind of negative feedback regulation. This was supported by the fact that *ERG* expression is regulated by complexes of HSC transcription

factors, including ERG itself, binding its promoters and the +85 stem cell enhancer, as recently shown.[156]

3.3.3 Correlation of *ERG* expression with molecular alterations in AML with gain of chromosome 21

Mutations in *RUNX1* are frequently found in AML with trisomy 21.[168] Therefore, a possible relationship between high *ERG* expression and *RUNX1* mutations was investigated. In agreement with the literature,[168] mutations in *RUNX1* were found in 25% (10 of 40) of AML patients with a non-complex aberrant karyotype and gain of one or two chromosome 21 (Appendix II).[138] In these 40 patients, higher mean *ERG* expression levels were observed in cases with *RUNX1* mutation compared with those with *RUNX1* wildtype (mean±SEM: 559±128 vs. 304±40, p=0.084; Figure 12), pointing to a cooperative role of *RUNX1* inactivation and elevated *ERG* expression in the onset of leukemia (Appendix II).[138]

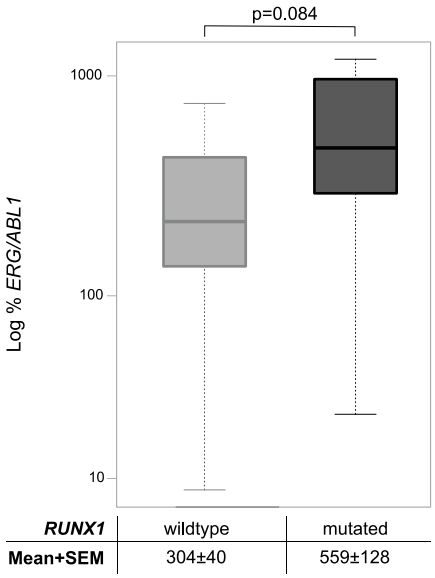


Figure 12: Box plot of *ERG* gene expression in patients with non-complex aberrant karyotype and gain of chromosome 21. Depicted are *ERG* expression levels on a logarithmic scale in *RUNX1* wildtype (n=30) and mutated (n=10) cases. Mean % *ERG*/*ABL1* expression levels were compared using Student's t-test (Appendix II).[138]

However, this correlation seems exclusive for the subtype of AML with gain of chromosome 21 as no association of altered *ERG* expression and mutations in *RUNX1* was observed in CN-AML (Appendix IV).[151] Cooperating effects of elevated *ERG* expression with specific mutations have been suggested before. In mouse models of T-ALL it has been shown that constitutive overexpression of *ERG* in hematopoietic cells drives the development of T-ALL and cooperates with activating *NOTCH1* mutations in leukemogenesis.[106;107] Mutations of *GATA1* are frequent in childhood DS-AMKL, but rarely detected in other DS and non-DS leukemias.[169;170] Studies focusing on the relationship between *GATA1* mutations and trisomy 21 in DS-AMKL revealed a specific synergy between loss of full-length *GATA1* and overexpression of the ETS family members *ERG* and *ETS2* in leukemogenesis.[108;171] Based on these findings it is conceivable that

inactivation of the *RUNX1* gene in combination with altered *ERG* expression could be another cooperative event in leukemogenesis of AML with gain of chromosome 21, which should be further explored.

4 Conclusions

BAALC and *ERG* gene expression levels varied within a wide range in CN-AML. They were allocated to clinical outcome, even though the results on *ERG* expression were less persuasive than those of *BAALC* gene expression. Regarding the inter-relationships between altered *BAALC* and *ERG* gene expression, patients´ characteristics and concomitant genetic alterations, several specific associations were identified. High expression levels of both genes correlated with a younger age and both genes were specifically associated with the clinically adverse-risk group of *NPM1* wildtype and *FLT3*-ITD. Beside these shared associations, *BAALC* expression was shown to gather with certain subgroups of CN-AML. For instance, strong correlations with mutations in transcription factors and genes that induce proliferation, as well as no or a negative correlation with mutations in epigenetic modifiers could be described. Moreover, *BAALC* was found to correlate most specifically with single molecular alterations, namely *IDH2*R172 and mutations in *RUNX1*, both of which represent clinical and molecular distinct classifiers in CN-AML. In terms of *ERG*, no such associations could be revealed in CN-AML until now.

Regarding the prognostic value, both altered *BAALC* and *ERG* expression, were shown to impact on survival. Although for *ERG* gene expression this impact was shown to depend on other molecular alterations. *BAALC* expression was strongly associated with clinical outcome even in the context of other adverse prognostic alterations. Therefore, *BAALC* expression has been shown to provide additional prognostic information on survival in the intermediate-risk group of *NPM1* wildtype or *FLT3*-ITD, suggesting that the delineation of genetically defined subtypes could improve prognostication in CN-AML. Opposed to the genetic categories, defined by functional biological characteristics or prognostic parameters, it was shown that the sole accumulation of some molecular aberrations, including *BAALC* gene expression and alterations in *ASXL1*, *FLT3*, *MLL* and *WT1* influence clinical outcome in an additive manner. These results suggest that a comprehensive screening of molecular genetic alterations can provide additional information for risk assessment in CN-AML.

Based on the strong correlation of *BAALC* expression to clinical outcome at diagnosis, the role of *BAALC* as a marker for detection of MRD was analyzed. *BAALC* was validated to be a stable marker, the expression course of which correlated substantially with the course of the disease and furthermore, with the mutational load or expression levels of well-known and already used MRD markers. Moreover, a strong prognostic value of *BAALC* assessment during therapy was observed, as high *BAALC* expression at certain clinically defined follow-up time points was shown to strongly correlate to adverse clinical outcome. These results predispose *BAALC*-

based residual disease monitoring during therapy as an applicable strategy for detecting patients at high risk of relapse, especially in *NPM1* wildtype cases.

Finally, it was shown that a gain of chromosome 21 resulted in an increased expression of the *ERG* gene, suggesting that an altered *ERG* expression represents a causative mechanism, contributing to the pathogenesis in AML with a gain of chromosome 21. However, the correlation of chromosomal gain and elevated *ERG* gene expression was not linear, which strengthens the assumption that *ERG* expression is influenced by additional parameters or alterations. Though the cooperating event could not be entirely defined - neither in CN-AML nor in AML with a gain of chromosome 21 - a strong correlation of *ERG* expression and mutations in *RUNX1* was observed in AML with a gain of chromosome 21. This correlation seemed to be subtype specific as, besides an appropriately high number of *RUNX1* mutated patients in CN-AML, no such association could be revealed.

5 Perspectives

Within the presented studies it was possible to identify several specific clusters of genetic alterations, which in some cases could even be allocated to clinical outcome. However, these results need to be validated in independent and extended patient cohorts. It should be strongly suggested to integrate this validation into clinical trials, as these would not only be applicable for data validation, but could also investigate several additional approaches. For instance, it has previously been shown that *ERG* expression results in resistance to kinase inhibitors such as sorafenib, a drug currently tested for treatment of AML. Furthermore, it has been suggested that *BAALC* mediated treatment resistance might be reversed by the use of MEK inhibitors. Clinical trials would be most appropriate to simultaneously address these previously described effects and their consequences in large and uniformly treated patient cohorts. Moreover, as clinical trials conform to stringent sampling intervals, these studies would also provide the conditions to study the relapse kinetics of high *BAALC* expressing leukemic clones. This would allow optimal sampling intervals for high *BAALC* expressing AML subtype to be revealed and therefore most probably even strengthen the information which can be gained from *BAALC* assessment during therapy.

In addition, it has to be pointed out that the results presented here, though representing a comprehensive analysis and providing precious prognostic and biologic information, still represent a snapshot of the enormous heterogeneity of AML. Nowadays it is clear that the outcomes of AML patients depend on a complex interplay of factors, including specific combinations of genetic lesions, patient conditions, the therapeutic approaches used and the magnitude of treatment response. After the great improvements in genomics technology, there is now a tremendous need for developing or adapting analysis tools, which allow the translation of the massively accumulating genetic information into clinically useful and valuable information. For instance, adjusted multivariate projection methodologies could provide the potential to generate discriminative models, favoring the delineation of new subtypes in AML, as multivariate projection methodologies were originally invented to uncover relevant information from huge data sets. Subsequently, these AML subtypes could be studied further to evaluate clinically applicable risk stratification algorithms and individual therapies. However, addressing the dilemma that subclassification by definition accompanies declining sample size, it would be desirable that these studies and subsequent integration of genetic risk factors into clinically applicable algorithms occurs through collaborations of multiple study groups, in order to increase the sample size and therefore the study validity.

6 References

1. U.S.National Institutes of Health, National Cancer Institute. SEER: Surveillance Epidemiology and End Results. http://seer cancer gov/. 2012.

2. Bennett JM, Catovsky D, Daniel MT, et al. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. Br J Haematol. 1976;33(4):451-458.

3. Doulatov S, Notta F, Laurenti E, Dick JE. Hematopoiesis: a human perspective. Cell Stem Cell. 2012;10(2):120-136.

4. West AH, Godley LA, Churpek JE. Familial myelodysplastic syndrome/acute leukemia syndromes: a review and utility for translational investigations. Ann N Y Acad Sci. 2014;1310:111-118.

5. Churpek JE, Pyrtel K, Kanchi KL, et al. Genomic analysis of germ line and somatic variants in familial myelodysplasia/acute myeloid leukemia. Blood. 2015;126(22):2484-2490.

6. Arber DA, Brunning RD, Le Beau MM, Falini B, Vardiman J, Porwit A et al. Acute myeloid leukemia with recurrent genetic abnormalities. In: Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H et al. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, vol. 4th. Lyon: International Agency for Research on Cancer (IARC), 2008:110-23.

7. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia. Blood. 2016;127(20):2391-2405.

8. Mrozek K, Heerema NA, Bloomfield CD. Cytogenetics in acute leukemia. Blood Rev. 2004;18(2):115-136.

9. Grimwade D, Hills RK, Moorman AV, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood. 2010;116(3):354-365.

10. Dohner H, Estey EH, Amadori S, 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(3):453-474.

11. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. N Engl J Med. 2015;373(12):1136-1152.

12. Patel JP, Gonen M, Figueroa ME, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med. 2012;366(12):1079-1089.

13. Koreth J, Schlenk R, Kopecky KJ, et al. Allogeneic stem cell transplantation for acute myeloid leukemia in first complete remission: systematic review and meta-analysis of prospective clinical trials. JAMA. 2009;301(22):2349-2361.

14. Baldus CD, Thiede C, Soucek S, et al. BAALC expression and FLT3 internal tandem duplication mutations in acute myeloid leukemia patients with normal cytogenetics: prognostic implications. J Clin Oncol. 2006;24(5):790-797.

15. Langer C, Radmacher MD, Ruppert AS, et al. High BAALC expression associates with other molecular prognostic markers, poor outcome, and a distinct gene-expression signature in cytogenetically normal patients younger than 60 years with acute myeloid leukemia: a Cancer and Leukemia Group B (CALGB) study. Blood. 2008;111(11):5371-5379.

16. Metzeler KH, Dufour A, Benthaus T, et al. ERG expression is an independent prognostic factor and allows refined risk stratification in cytogenetically normal acute myeloid leukemia: a comprehensive analysis of ERG, MN1, and BAALC transcript levels using oligonucleotide microarrays. J Clin Oncol. 2009;27(30):5031-5038.

17. Cilloni D, Renneville A, Hermitte F, et al. Real-time quantitative polymerase chain reaction detection of minimal residual disease by standardized WT1 assay to enhance risk stratification in acute myeloid leukemia: a European LeukemiaNet study. J Clin Oncol. 2009;27(31):5195-5201.

18. Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, van Putten WL, et al. High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients. Blood. 2003;101(3):837-845.

19. Borer RA, Lehner CF, Eppenberger HM, Nigg EA. Major nucleolar proteins shuttle between nucleus and cytoplasm. Cell. 1989;56(3):379-390.

20. Grisendi S, Mecucci C, Falini B, Pandolfi PP. Nucleophosmin and cancer. Nat Rev Cancer. 2006;6(7):493-505.

21. Falini B, Mecucci C, Tiacci E, et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. N Engl J Med. 2005;352(3):254-266.

22. Falini B, Nicoletti I, Martelli MF, Mecucci C. Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. Blood. 2007;109(3):874-885.

23. Schnittger S, Schoch C, Kern W, et al. Nucleophosmin gene mutations are predictors of favorable prognosis in acute myelogenous leukemia with a normal karyotype. Blood. 2005;106(12):3733-3739.

24. Thiede C, Koch S, Creutzig E, et al. Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML). Blood. 2006;107(10):4011-4020.

25. Schlenk RF, Dohner K, Krauter J, et al. Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia. N Engl J Med. 2008;358(18):1909-1918.

26. Becker H, Marcucci G, Maharry K, et al. Favorable prognostic impact of NPM1 mutations in older patients with cytogenetically normal de novo acute myeloid leukemia and associated gene- and microRNA-expression signatures: a Cancer and Leukemia Group B study. J Clin Oncol. 2010;28(4):596-604.

27. Choudhary C, Schwable J, Brandts C, et al. AML-associated Flt3 kinase domain mutations show signal transduction differences compared with Flt3 ITD mutations. Blood. 2005;106(1):265-273.

28. Whitman SP, Archer KJ, Feng L, et al. Absence of the wild-type allele predicts poor prognosis in adult de novo acute myeloid leukemia with normal cytogenetics and the internal tandem duplication of FLT3: a cancer and leukemia group B study. Cancer Res. 2001;61(19):7233-7239.

29. Frohling S, Schlenk RF, Breitruck J, et al. Prognostic significance of activating FLT3 mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML Study Group Ulm. Blood. 2002;100(13):4372-4380.

30. Schnittger S, Bacher U, Haferlach C, et al. Diversity of the juxtamembrane and TKD1 mutations (Exons 13-15) in the FLT3 gene with regards to mutant load, sequence, length, localization, and correlation with biological data. Genes Chromosomes Cancer. 2012;51(10):910-924.

31. Bacher U, Haferlach C, Kern W, Haferlach T, Schnittger S. Prognostic relevance of FLT3-TKD mutations in AML: the combination matters--an analysis of 3082 patients. Blood. 2008;111(5):2527-2537.

32. Cho YS, Kim EJ, Park UH, Sin HS, Um SJ. Additional sex comb-like 1 (ASXL1), in cooperation with SRC-1, acts as a ligand-dependent coactivator for retinoic acid receptor. J Biol Chem. 2006;281(26):17588-17598.

33. Vainchenker W, Delhommeau F, Constantinescu SN, Bernard OA. New mutations and pathogenesis of myeloproliferative neoplasms. Blood. 2011;118(7):1723-1735.

34. Metzeler KH, Becker H, Maharry 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(26):6920-6929.

35. Paschka P, Schlenk RF, Gaidzik VI, et al. ASXL1 mutations in younger adult patients with acute myeloid leukemia: a study by the German-Austrian Acute Myeloid Leukemia Study Group. Haematologica. 2015;100(3):324-330.

36. Schnittger S, Eder C, Jeromin S, et al. ASXL1 exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome. Leukemia. 2013;27(1):82-91.

37. Zhang P, Iwasaki-Arai J, Iwasaki H, et al. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. Immunity. 2004;21(6):853-863.

38. Ye M, Zhang H, Amabile G, et al. C/EBPa controls acquisition and maintenance of adult haematopoietic stem cell quiescence. Nat Cell Biol. 2013;15(4):385-394.

39. Pabst T, Mueller BU, Zhang P, et al. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. Nat Genet. 2001;27(3):263-270.

40. Nerlov C. C/EBPalpha mutations in acute myeloid leukaemias. Nat Rev Cancer. 2004;4(5):394-400.

41. Wouters BJ, Lowenberg B, Erpelinck-Verschueren CA, et al. Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. Blood. 2009;113(13):3088-3091.

42. Taskesen E, Bullinger L, Corbacioglu A, et al. Prognostic impact, concurrent genetic mutations, and gene expression features of AML with CEBPA mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for CEBPA double mutant AML as a distinctive disease entity. Blood. 2011;117(8):2469-2475.

43. Dufour A, Schneider F, Hoster E, et al. Monoallelic CEBPA mutations in normal karyotype acute myeloid leukemia: independent favorable prognostic factor within NPM1 mutated patients. Ann Hematol. 2012;91(7):1051-1063.

44. Shlush LI, Zandi S, Mitchell A, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia 2. Nature. 2014;506(7488):328-333.

45. Corces-Zimmerman MR, Hong WJ, Weissman IL, Medeiros BC, Majeti R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. Proc Natl Acad Sci U S A. 2014;111(7):2548-2553.

46. Genovese G, Kahler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014;371(26):2477-2487.

47. Gowher H, Loutchanwoot P, Vorobjeva O, et al. Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. J Mol Biol. 2006;357(3):928-941.

48. Kim SJ, Zhao H, Hardikar S, et al. A DNMT3A mutation common in AML exhibits dominant-negative effects in murine ES cells. Blood. 2013;122(25):4086-4089.

49. Gaidzik VI, Schlenk RF, Paschka P, et al. Clinical impact of DNMT3A mutations in younger adult patients with acute myeloid leukemia: results of the AML Study Group (AMLSG). Blood. 2013;121(23):4769-4777.

50. Marcucci G, Metzeler KH, Schwind S, et al. Age-Related Prognostic Impact of Different Types of DNMT3A Mutations in Adults With Primary Cytogenetically Normal Acute Myeloid Leukemia. J Clin Oncol. 2012;30(7):742-750.

51. Renneville A, Boissel N, Nibourel O, et al. Prognostic significance of DNA methyltransferase 3A mutations in cytogenetically normal acute myeloid leukemia: a study by the Acute Leukemia French Association. Leukemia. 2012;26(6):1247-1254.

52. Thol F, Damm F, Ludeking A, et al. Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. J Clin Oncol. 2011;29(21):2889-2896.

53. Dang L, White DW, Gross S, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature. 2009;462(7274):739-744.

54. Marcucci G, Maharry K, Wu YZ, et al. IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. J Clin Oncol. 2010;28(14):2348-2355.

55. Thol F, Damm F, Wagner K, et al. Prognostic impact of IDH2 mutations in cytogenetically normal acute myeloid leukemia. Blood. 2010;116(4):614-616.

56. Paschka P, Schlenk RF, Gaidzik VI, et al. IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. J Clin Oncol. 2010;28(22):3636-3643.

57. Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer. 2007;7(11):823-833.

58. Zhang Y, Yan X, Sashida G, et al. Stress hematopoiesis reveals abnormal control of self-renewal, lineage bias, and myeloid differentiation in Mll partial tandem duplication (Mll-PTD) hematopoietic stem/progenitor cells. Blood. 2012;120(5):1118-1129.

59. Schnittger S, Kinkelin U, Schoch C, et al. Screening for MLL tandem duplication in 387 unselected patients with AML identify a prognostically unfavorable subset of AML. Leukemia. 2000;14(5):796-804.

60. Dohner K, Tobis K, Ulrich R, et al. Prognostic significance of partial tandem duplications of the MLL gene in adult patients 16 to 60 years old with acute myeloid leukemia and normal cytogenetics: a study of the Acute Myeloid Leukemia Study Group Ulm. J Clin Oncol. 2002;20(15):3254-3261.

61. Whitman SP, Ruppert AS, Marcucci G, et al. Long-term disease-free survivors with cytogenetically normal acute myeloid leukemia and MLL partial tandem duplication: a Cancer and Leukemia Group B study. Blood. 2007;109(12):5164-5167.

62. Whitman SP, Caligiuri MA, Maharry K, et al. The MLL partial tandem duplication in adults aged 60 years and older with de novo cytogenetically normal acute myeloid leukemia. Leukemia. 2012;26(7):1713-1717.

63. Wang J, Kong G, Liu Y, et al. Nras(G12D/+) promotes leukemogenesis by aberrantly regulating hematopoietic stem cell functions. Blood. 2013;121(26):5203-5207.

64. Bacher U, Haferlach T, Schoch C, Kern W, Schnittger S. Implications of NRAS mutations in AML: a study of 2502 patients. Blood. 2006;107(10):3847-3853.

65. Michaud J, Scott HS, Escher R. AML1 interconnected pathways of leukemogenesis. Cancer Invest. 2003;21(1):105-136.

66. Tsuzuki S, Hong D, Gupta R, et al. Isoform-specific potentiation of stem and progenitor cell engraftment by AML1/RUNX1. PLoS Med. 2007;4(5):e172.

67. Ichikawa M, Asai T, Saito T, et al. AML-1 is required for megakaryocytic maturation and lymphocytic differentiation, but not for maintenance of hematopoietic stem cells in adult hematopoiesis. Nat Med. 2004;10(3):299-304.

68. Mendler JH, Maharry K, Radmacher MD, et al. RUNX1 Mutations Are Associated With Poor Outcome in Younger and Older Patients With Cytogenetically Normal Acute Myeloid Leukemia and With Distinct Gene and MicroRNA Expression Signatures. J Clin Oncol. 2012;30(25):3109-3118.

69. Tang JL, Hou HA, Chen CY, et al. AML1/RUNX1 mutations in 470 adult patients with de novo acute myeloid leukemia: prognostic implication and interaction with other gene alterations. Blood. 2009;114(26):5352-5361.

70. Schnittger S, Dicker F, Kern W, et al. RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis. Blood. 2011;117(8):2348-2357.

71. Gaidzik VI, Bullinger L, Schlenk RF, et al. RUNX1 Mutations in Acute Myeloid Leukemia: Results From a Comprehensive Genetic and Clinical Analysis From the AML Study Group. J Clin Oncol. 2011;29(10):1364-1372.

72. Ito S, Shen L, Dai Q, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300-1303.

73. Ko M, Huang Y, Jankowska AM, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature. 2010;468(7325):839-843.

74. Moran-Crusio K, Reavie L, Shih A, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell. 2011;20(1):11-24.

75. Saint-Martin C, Leroy G, Delhommeau F, et al. Analysis of the ten-eleven translocation 2 (TET2) gene in familial myeloproliferative neoplasms. Blood. 2009;114(8):1628-1632.

76. Abdel-Wahab O, Manshouri T, Patel J, et al. Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. Cancer Res. 2010;70(2):447-452.

77. Metzeler KH, Maharry K, Radmacher MD, et al. TET2 Mutations Improve the New European LeukemiaNet Risk Classification of Acute Myeloid Leukemia: A Cancer and Leukemia Group B Study. J Clin Oncol. 2011;29(10):1373-1381.

78. Weissmann S, Alpermann T, Grossmann V, et al. Landscape of TET2 mutations in acute myeloid leukemia. Leukemia. 2012;26(5):934-942.

79. Gaidzik VI, Paschka P, Spath D, et al. TET2 Mutations in Acute Myeloid Leukemia (AML): Results From a Comprehensive Genetic and Clinical Analysis of the AML Study Group. J Clin Oncol. 2012;30(12):1350-1357.

80. Damm F, Markus B, Thol F, et al. TET2 mutations in cytogenetically normal acute myeloid leukemia: clinical implications and evolutionary patterns. Genes Chromosomes Cancer. 2014;53(10):824-832.

81. Yang L, Han Y, Suarez SF, Minden MD. A tumor suppressor and oncogene: the WT1 story. Leukemia. 2007;21(5):868-876.

82. Cilloni D, Gottardi E, De Micheli D, et al. Quantitative assessment of WT1 expression by real time quantitative PCR may be a useful tool for monitoring minimal residual disease in acute leukemia patients. Leukemia. 2002;16(10):2115-2121.

83. Virappane P, Gale R, Hills R, et al. Mutation of the Wilms' tumor 1 gene is a poor prognostic factor associated with chemotherapy resistance in normal karyotype acute myeloid leukemia: the United Kingdom Medical Research Council Adult Leukaemia Working Party. J Clin Oncol. 2008;26(33):5429-5435.

84. Paschka P, Marcucci G, Ruppert AS, et al. Wilms Tumor 1 Gene Mutations Independently Predict Poor Outcome in Adults With Cytogenetically Normal Acute Myeloid Leukemia: A Cancer and Leukemia Group B Study. J Clin Oncol. 2008;26(28):4595-4602.

85. Renneville A, Boissel N, Zurawski V, et al. Wilms tumor 1 gene mutations are associated with a higher risk of recurrence in young adults with acute myeloid leukemia: a study from the Acute Leukemia French Association. Cancer. 2009;115(16):3719-3727.

86. Krauth MT, Alpermann T, Bacher U, et al. WT1 mutations are secondary events in AML, show varying frequencies and impact on prognosis between genetic subgroups. Leukemia. 2015;29(3):660-667.

87. Tanner SM, Austin JL, Leone G, et al. BAALC, the human member of a novel mammalian neuroectoderm gene lineage, is implicated in hematopoiesis and acute leukemia. Proc Natl Acad Sci U S A. 2001;98(24):13901-13906.

88. Baldus CD, Tanner SM, Kusewitt DF, et al. BAALC, a novel marker of human hematopoietic progenitor cells. Exp Hematol. 2003;31(11):1051-1056.

89. Baldus CD, Tanner SM, Ruppert AS, et al. BAALC expression predicts clinical outcome of de novo acute myeloid leukemia patients with normal cytogenetics: a Cancer and Leukemia Group B Study. Blood. 2003;102(5):1613-1618.

90. Schrama D, Keller G, Houben R, et al. BRAFV600E mutations in malignant melanoma are associated with increased expressions of BAALC. J Carcinog. 2008;7:1.

91. Heuser M, Berg T, Kuchenbauer F, et al. Functional role of BAALC in leukemogenesis. Leukemia. 2012;26(3):532-536.

92. Morita K, Masamoto Y, Kataoka K, et al. BAALC potentiates oncogenic ERK pathway through interactions with MEKK1 and KLF4. Leukemia. 2015;29(11):2248-2256.

93. Eisfeld AK, Schwind S, Patel R, et al. Intronic miR-3151 within BAALC drives leukemogenesis by deregulating the TP53 pathway. Sci Signal. 2014;7(321):ra36.

94. Franzoni A, Passon N, Fabbro D, et al. Histone post-translational modifications associated to BAALC expression in leukemic cells. Biochem Biophys Res Commun. 2012;417(2):721-725.

95. Schwind S, Marcucci G, Maharry K, et al. BAALC and ERG expression levels are associated with outcome and distinct gene and microRNA expression profiles in older patients with de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. Blood. 2010;116(25):5660-5669.

96. Weber S, Alpermann T, Dicker F, et al. BAALC expression: a suitable marker for prognostic risk stratification and detection of residual disease in cytogenetically normal acute myeloid leukemia. Blood Cancer J. 2014;4:e173.

97. Bienz M, Ludwig M, Mueller BU, et al. Risk assessment in patients with acute myeloid leukemia and a normal karyotype. Clin Cancer Res. 2005;11(4):1416-1424.

98. Najima Y, Ohashi K, Kawamura M, et al. Molecular monitoring of BAALC expression in patients with CD34-positive acute leukemia. Int J Hematol. 2010;91(4):636-645.

99. Diaz-Beya M, Brunet S, Nomdedeu J, et al. The expression level of BAALC-associated microRNA miR-3151 is an independent prognostic factor in younger patients with cytogenetic intermediate-risk acute myeloid leukemia. Blood Cancer J. 2015;5:e352.

100. Haferlach C, Kern W, Schindela S, et al. Gene expression of BAALC, CDKN1B, ERG, and MN1 adds independent prognostic information to cytogenetics and molecular

mutations in adult acute myeloid leukemia. Genes Chromosomes Cancer. 2012;51(3):257-265.

101. Duterque-Coquillaud M, Niel C, Plaza S, Stehelin D. New human erg isoforms generated by alternative splicing are transcriptional activators. Oncogene. 1993;8(7):1865-1873.

102. Rainis L, Toki T, Pimanda JE, et al. The proto-oncogene ERG in megakaryoblastic leukemias. Cancer Res. 2005;65(17):7596-7602.

103. Loughran SJ, Kruse EA, Hacking DF, et al. The transcription factor Erg is essential for definitive hematopoiesis and the function of adult hematopoietic stem cells. Nat Immunol. 2008;9(7):810-819.

104. Taoudi S, Bee T, Hilton A, et al. ERG dependence distinguishes developmental control of hematopoietic stem cell maintenance from hematopoietic specification. Genes Dev. 2011;25(3):251-262.

105. Salek-Ardakani S, Smooha G, de Boer J, et al. ERG is a megakaryocytic oncogene. Cancer Res. 2009;69(11):4665-4673.

106. Tsuzuki S, Taguchi O, Seto M. Promotion and maintenance of leukemia by ERG. Blood. 2011;117(14):3858-3868.

107. Thoms JA, Birger Y, Foster S, et al. ERG promotes T-acute lymphoblastic leukemia and is transcriptionally regulated in leukemic cells by a stem cell enhancer. Blood. 2011;117(26):7079-7089.

108. Stankiewicz MJ, Crispino JD. ETS2 and ERG promote megakaryopoiesis and synergize with alterations in GATA-1 to immortalize hematopoietic progenitor cells. Blood. 2009;113(14):3337-3347.

109. Khan I, Malinge S, Crispino J. Myeloid leukemia in Down syndrome. Crit Rev Oncog. 2011;16(1-2):25-36.

110. Ronan A, Fagan K, Christie L, et al. Familial 4.3 Mb duplication of 21q22 sheds new light on the Down syndrome critical region. J Med Genet. 2007;44(7):448-451.

111. Delabar JM, Theophile D, Rahmani Z, et al. Molecular mapping of twenty-four features of Down syndrome on chromosome 21. Eur J Hum Genet. 1993;1(2):114-124.

112. Malinge S, Izraeli S, Crispino JD. Insights into the manifestations, outcomes, and mechanisms of leukemogenesis in Down syndrome. Blood. 2009;113(12):2619-2628.

113. Canzonetta C, Hoischen A, Giarin E, et al. Amplified segment in the 'Down syndrome critical region' on HSA21 shared between Down syndrome and euploid AML-M0 excludes RUNX1, ERG and ETS2. Br J Haematol. 2012;157(2):197-200.

114. Ichikawa H, Shimizu K, Hayashi Y, Ohki M. An RNA-binding protein gene, TLS/FUS, is fused to ERG in human myeloid leukemia with t(16;21) chromosomal translocation. Cancer Res. 1994;54(11):2865-2868.

115. Giovannini M, Biegel JA, Serra M, et al. EWS-erg and EWS-Fli1 fusion transcripts in Ewing's sarcoma and primitive neuroectodermal tumors with variant translocations. J Clin Invest. 1994;94(2):489-496.

116. Soller MJ, Isaksson M, Elfving P, et al. Confirmation of the high frequency of the TMPRSS2/ERG fusion gene in prostate cancer. Genes Chromosomes Cancer. 2006;45(7):717-719.

117. Petrovics G, Liu A, Shaheduzzaman S, et al. Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome. Oncogene. 2005;24(23):3847-3852.

118. Marcucci G, Baldus CD, Ruppert AS, et al. Overexpression of the ETS-related gene, ERG, predicts a worse outcome in acute myeloid leukemia with normal karyotype: a Cancer and Leukemia Group B study. J Clin Oncol. 2005;23(36):9234-9242.

119. Marcucci G, Maharry K, Whitman SP, et al. High expression levels of the ETS-related gene, ERG, predict adverse outcome and improve molecular risk-based classification of cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B Study. J Clin Oncol. 2007;25(22):3337-3343.

120. Bock J, Mochmann LH, Schlee C, et al. ERG transcriptional networks in primary acute leukemia cells implicate a role for ERG in deregulated kinase signaling. PLoS ONE. 2013;8(1):e52872.

121. Cornelissen JJ, Gratwohl A, Schlenk RF, et al. The European LeukemiaNet AML Working Party consensus statement on allogeneic HSCT for patients with AML in remission: an integrated-risk adapted approach. Nat Rev Clin Oncol. 2012;9(10):579-590.

122. Kern W, Haferlach T, Schoch C, et al. Early blast clearance by remission induction therapy is a major independent prognostic factor for both achievement of complete remission and long-term outcome in acute myeloid leukemia: data from the German AML Cooperative Group (AMLCG) 1992 Trial. Blood. 2003;101(1):64-70.

123. Lowenberg B, van Putten W, Theobald M, et al. Effect of priming with granulocyte colony-stimulating factor on the outcome of chemotherapy for acute myeloid leukemia. N Engl J Med. 2003;349(8):743-752.

124. Büchner T, Hiddemann W, Wormann B, et al. Double induction strategy for acute myeloid leukemia: the effect of high-dose cytarabine with mitoxantrone instead of standard-dose cytarabine with daunorubicin and 6-thioguanine: a randomized trial by the German AML Cooperative Group. Blood. 1999;93(12):4116-4124.

125. Burnett AK, Hills RK, Milligan D, et al. Identification of patients with acute myeloblastic leukemia who benefit from the addition of gemtuzumab ozogamicin: results of the MRC AML15 trial. J Clin Oncol. 2011;29(4):369-377.

126. Hokland P, Ommen HB. Towards individualized follow-up in adult acute myeloid leukemia in remission. Blood. 2011;117(9):2577-2584.

127. Ommen HB, Schnittger S, Jovanovic JV, et al. Strikingly different molecular relapse kinetics in NPM1c, PML-RARA, RUNX1-RUNX1T1, and CBFB-MYH11 acute myeloid leukemias. Blood. 2010;115(2):198-205.

128. Freeman SD, Jovanovic JV, Grimwade D. Development of minimal residual disease-directed therapy in acute myeloid leukemia. Semin Oncol. 2008;35(4):388-400.

129. Ivey A, Hills RK, Simpson MA, et al. Assessment of Minimal Residual Disease in Standard-Risk AML. N Engl J Med. 2016;374(5):422-433.

130. Sanz MA, Grimwade D, Tallman MS, et al. Management of acute promyelocytic leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. Blood. 2009;113(9):1875-1891.

131. Schnittger S, Weisser M, Schoch C, et al. New score predicting for prognosis in PML-RARA+, AML1-ETO+, or CBFBMYH11+ acute myeloid leukemia based on quantification of fusion transcripts. Blood. 2003;102(8):2746-2755.

132. Corbacioglu A, Scholl C, Schlenk RF, et al. Prognostic impact of minimal residual disease inCBFB-MYH11-positive acute myeloid leukemia. J Clin Oncol. 2010;28(23):3724-3729.

133. Viehmann S, Teigler-Schlegel A, Bruch J, et al. Monitoring of minimal residual disease (MRD) by real-time quantitative reverse transcription PCR (RQ-RT-PCR) in childhood acute myeloid leukemia with AML1/ETO rearrangement. Leukemia. 2003;17(6):1130-1136.

134. Weisser M, Haferlach C, Hiddemann W, Schnittger S. The quality of molecular response to chemotherapy is predictive for the outcome of AML1-ETO-positive AML and is independent of pretreatment risk factors. Leukemia. 2007;21(6):1177-1182.

135. Schnittger S, Kern W, Tschulik C, et al. Minimal residual disease levels assessed by NPM1 mutation-specific RQ-PCR provide important prognostic information in AML. Blood. 2009;114(11):2220-2231.

136. Weisser M, Kern W, Schoch C, et al. Risk assessment by monitoring expression levels of partial tandem duplications in the MLL gene in acute myeloid leukemia during therapy. Haematologica. 2005;90(7):881-889.

137. Ommen HB, Hokland P, Haferlach T, et al. Relapse kinetics in acute myeloid leukaemias with MLL translocations or partial tandem duplications within the MLL gene. Br J Haematol. 2014;165(5):618-628.

138. Weber S, Haferlach C, Jeromin S, et al. Gain of chromosome 21 or amplification of chromosome arm 21q is one mechanism for increased ERG expression in acute myeloid leukemia. Genes Chromosomes Cancer. 2016;55(2):148-157.

139. Beillard E, Pallisgaard N, van der Velden VHJ, et al. Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (RQ-PCR) - a Europe against cancer program. Leukemia. 2003;17(12):2474-2486.

140. Weisser M, Haferlach T, Schoch C, Hiddemann W, Schnittger S. The use of housekeeping genes for real-time PCR-based quantification of fusion gene transcripts in acute myeloid leukemia. Leukemia. 2004;18(9):1551-1553.

141. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc. 2008;3(6):1101-1108.

142. Schnittger S, Haferlach C, Ulke M, et al. IDH1 mutations are detected in 6.6% of 1414 AML patients and are associated with intermediate risk karyotype and unfavorable prognosis in adults younger than 60 years and unmutated NPM1 status. Blood. 2010;116(25):5486-5496.

143. Nakao M, Janssen JW, Seriu T, Bartram CR. Rapid and reliable detection of N-ras mutations in acute lymphoblastic leukemia by melting curve analysis using LightCycler technology. Leukemia. 2000;14(2):312-315.

144. Bacher U, Dicker F, Haferlach C, et al. Quantification of rare NPM1 mutation subtypes by digital PCR. Br J Haematol. 2014;167(5):710-714.

145. Schnittger S, Schoch C, Dugas M, 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(1):59-66.

146. Thiede C, Steudel C, Mohr B, et al. Analysis of FLT3-activating mutations in 979 patients with acute myelogenous leukemia: association with FAB subtypes and identification of subgroups with poor prognosis. Blood. 2002;99(12):4326-4335.

147. Delic S, Rose D, Kern W, et al. Application of an NGS-based 28-gene panel in myeloproliferative neoplasms reveals distinct mutation patterns in essential thrombocythaemia, primary myelofibrosis and polycythaemia vera. Br J Haematol. 2016.

148. Kohlmann A, Klein HU, Weissmann S, et al. The Interlaboratory RObustness of Next-generation sequencing (IRON) study: a deep sequencing investigation of TET2, CBL and KRAS mutations by an international consortium involving 10 laboratories. Leukemia. 2011;25(12):1840-1848.

149. Lange V, Bohme I, Hofmann J, et al. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. BMC Genomics. 2014;15(1):63.

150. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015;43(Database issue):D805-D811.

151. Weber S, Haferlach T, Haferlach C, et al. Comprehensive study on ERG gene expression in normal karyotype acute myeloid leukemia: ERG expression is of limited prognostic value, whereas the accumulation of adverse prognostic markers stepwise worsens the prognosis. Blood Cancer J. 2016;6(12):e507.

152. Weber S, Haferlach T, Alpermann T, et al. Feasibility of BAALC gene expression for detection of minimal residual disease and risk stratification in normal karyotype acute myeloid leukaemia. Br J Haematol. 2016;175(5):904-916.

153. Greif PA, Konstandin NP, Metzeler KH, et al. RUNX1 mutations in cytogenetically normal acute myeloid leukemia are associated with a poor prognosis and up-regulation of lymphoid genes. Haematologica. 2012;97(12):1909-1915.

154. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016;374(23):2209-2221.

155. Harada Y, Harada H. Molecular pathways mediating MDS/AML with focus on AML1/RUNX1 point mutations. J Cell Physiol. 2009;220(1):16-20.

156. Diffner E, Beck D, Gudgin EJ, et al. Activity of a heptad of transcription factors is associated with stem cell programs and clinical outcome in acute myeloid leukemia. Blood. 2013;121(12):2289-2300.

157. Santamaria C, Chillon MC, Garcia-Sanz R, et al. BAALC is an important predictor of refractoriness to chemotherapy and poor survival in intermediate-risk acute myeloid leukemia (AML). Ann Hematol. 2010;89(5):453-458.

158. Mrozek K, Marcucci G, Nicolet D, et al. Prognostic Significance of the European LeukemiaNet Standardized System for Reporting Cytogenetic and Molecular Alterations in Adults With Acute Myeloid Leukemia. J Clin Oncol. 2012;30(36):4515-4523.

159. Kern W, Haferlach T, Schnittger S, et al. Karyotype instability between diagnosis and relapse in 117 patients with acute myeloid leukemia: implications for resistance against therapy. Leukemia. 2002;16(10):2084-2091.

160. Ding L, Ley TJ, Larson DE, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481(7382):506-510.

161. Krönke J, Bullinger L, Teleanu V, et al. Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia. Blood. 2013;122(1):100-108.

162. Kohlmann A, Nadarajah N, Alpermann T, et al. Monitoring of residual disease by next-generation deep-sequencing of RUNX1 mutations can identify acute myeloid leukemia patients with resistant disease. Leukemia. 2014;28(1):129-137.

163. Krönke J, Schlenk RF, Jensen KO, et al. Monitoring of minimal residual disease in NPM1-mutated acute myeloid leukemia: a study from the German-Austrian acute myeloid leukemia study group. J Clin Oncol. 2011;29(19):2709-2716.

164. McCormick SR, McCormick MJ, Grutkoski PS, et al. FLT3 mutations at diagnosis and relapse in acute myeloid leukemia: cytogenetic and pathologic correlations, including cuplike blast morphology. Arch Pathol Lab Med. 2010;134(8):1143-1151.

165. Welch JS, Ley TJ, Link DC, et al. The origin and evolution of mutations in acute myeloid leukemia. Cell. 2012;150(2):264-278.

166. Baldus CD, Liyanarachchi S, Mrozek K, et al. Acute myeloid leukemia with complex karyotypes and abnormal chromosome 21: Amplification discloses overexpression of APP, ETS2, and ERG genes. Proc Natl Acad Sci U S A. 2004;101(11):3915-3920.

167. Baldus CD, Burmeister T, Martus P, et al. High expression of the ETS transcription factor ERG predicts adverse outcome in acute T-lymphoblastic leukemia in adults. J Clin Oncol. 2006;24(29):4714-4720.

168. Preudhomme C, Warot-Loze D, Roumier C, et al. High incidence of biallelic point mutations in the Runt domain of the AML1/PEBP2 alpha B gene in Mo acute myeloid leukemia and in myeloid malignancies with acquired trisomy 21. Blood. 2000;96(8):2862-2869.

169. Rainis L, Bercovich D, Strehl S, et al. Mutations in exon 2 of GATA1 are early events in megakaryocytic malignancies associated with trisomy 21. Blood. 2003;102(3):981-986.

170. Wechsler J, Greene M, McDevitt MA, et al. Acquired mutations in GATA1 in the megakaryoblastic leukemia of Down syndrome. Nat Genet. 2002;32(1):148-152.

171. Hollanda LM, Lima CS, Cunha AF, et al. An inherited mutation leading to production of only the short isoform of GATA-1 is associated with impaired erythropoiesis. Nat Genet. 2006;38(7):807-812.

Acknowledgements

I would like to thank my advisor Prof. Dr. Susanne Schnittger for her continuous support with my PhD study, for her encouragement and supervision during the research and writing of this thesis. Thank you for your constant understanding and open attitude and for the scientific inspiration.

I want to sincerely thank Prof. Dr. Michael W. Pfaffl, who gave me the opportunity to do my PhD thesis at the Chair of Physiology, thereby supporting my work. Thank you for your straightforwardness and for always providing help whenever I needed it.

Additionally, my thanks go to Prof. Dr. Wolfgang Kern, Prof. Dr. Dr. Torsten Haferlach and Prof. Dr. Claudia Haferlach not only for giving me the possibility to do my PhD thesis at the Munich Leukemia Laboratory, but also for sharing their truthful and illuminating views on my PhD-related research.

Many thanks go to all my colleagues from MLL, especially to those from the molecular department. I was always able to count on helping hands and a friendly working atmosphere. I would also like to thank Elke Roos and Niroshan Nadarajah for their creative help with the figure design. A special acknowledgement goes to my office colleague, Sabine Jeromin. You have been a true friend and supportive in every way.

I particularly want to thank my family, my relatives and also my partner´s family. Their support has been unconditional throughout the writing of this thesis and my life in general. Thank you for your caring support.

My sincere thanks to my partner, Harald Müller, for supporting me in everything. I can’t thank you enough for encouraging me throughout this experience, for always cheering me up in hard times and cherishing every great moment with me.

Last but not least, special thanks go to my friends, who have always been on my side, sharing all the good and bad times. Your friendship is irreplaceable.

List of scientific communications

Research Papers

Simone Weber, Torsten Haferlach, Claudia Haferlach, Wolfgang Kern

Comprehensive study on *ERG* gene expression in normal karyotype acute myeloid leukemia: *ERG* expression is of limited prognostic value, whereas the accumulation of adverse prognostic markers stepwise worsens the prognosis

Blood Cancer J. 2016 Dec 09; 4:e173. doi:10.1038/bcj.2016.120.

Simone Weber, Torsten Haferlach, Tamara Alpermann, Karolína Perglerová, Susanne Schnittger, Claudia Haferlach, Wolfgang Kern

Feasibility of *BAALC* gene expression for detection of minimal residual disease and risk stratification in normal karyotype acute myeloid leukaemia

British Journal of Haematology. 2016 Dec;175(5):904-916. doi: 10.1111/bjh.14343. Epub 2016 Sep 23.

Torsten Haferlach, Simone Weber, Rabea Konietschke, Niroshan Nadarajah, Anna Stengel, Wolfgang Kern, Claudia Haferlach, Manja Meggendorfer

Robustness of comprehensive DNA and RNA based assays at diagnosis of acute myeloid leukemia using blood and bone marrow stored on filter cards

Leukemia. 2016 Jun 10. doi: 10.1038/leu.2016.156.

Simone Weber, Claudia Haferlach, Sabine Jeromin, Niroshan Nadarajah, Frank Dicker, Louisa Noël, Melanie Zenger, Tamara Alpermann, Wolfgang Kern, Torsten Haferlach, Susanne Schnittger

Gain of chromosome 21 or amplification of chromosome arm 21q is one mechanism for increased *ERG* expression in acute myeloid leukemia

Genes Chromosomes Cancer. 2016 Feb;55(2):148-57. doi: 10.1002/gcc.22321. Epub 2015 Nov 6.

Simone Weber, Tamara Alpermann, Frank Dicker, Sabine Jeromin, Niroshan Nadarajah, Christiane Eder, Annette Fasan, Alexander Kohlmann, Manja Meggendorfer, Claudia Haferlach, Wolfgang Kern, Torsten Haferlach, Susanne Schnittger

***BAALC* expression: a suitable marker for prognostic risk stratification and detection of residual disease in cytogenetically normal acute myeloid leukemia**

Blood Cancer J. 2014 Jan 10; 4:e173. doi: 10.1038/bcj.2013.71.

Poster presentations

Simone Weber, Tamara Alpermann, Karolína Perglerová, Susanne Schnittger, Torsten Haferlach, MD, Claudia Haferlach Wolfgang Kern

***BAALC* Expression Is a Feasible Marker for Risk Stratification and Detection of Minimal Residual Disease in Cytogenetically Normal Acute Myeloid Leukemia**

57th ASH Annual Meeting and Exposition; December 5-8, 2015; San Diego, CA

Simone Weber, Manja Meggendorfer, Niroshan Nadarajah, Karolína Perglerová, Susanne Schnittger

Molecular Characterization of Philadelphia Chromosome Positive Acute Myeloid

Leukemia - New Provisional Entity?

57th ASH Annual Meeting and Exposition; December 5-8, 2015; San Diego, CA

Annette Fasan, Wolfgang Kern, Niroshan Nadarajah, Simone Weber, Sonja Schindela, Nicole Schlenther, Susanne Schnittger, Torsten Haferlach, Claudia Haferlach

Three Steps to the Diagnosis of Adult Ph-like ALL

57th ASH Annual Meeting and Exposition; December 5-8, 2015; San Diego, CA

Andreas Roller, Simone Weber, Alexander Kohlmann, Melanie Zenger, Marita Staller, Wolfgang Kern, Susanne Schnittger, Torsten Haferlach, Claudia Haferlach

Gene amplifications in 84 patients with acute myeloid leukemia and 31 patients with myelodysplastic syndrome investigated by array CGH

55th ASH Annual Meeting and Exposition; December 7-10, 2013; New Orleans, LA

Simone Weber, Claudia Haferlach, Louisa Noel, Tamara Alpermann, Niroshan Nadarahja, Wolfgang Kern, Torsten Haferlach, Susanne Schnittger

***ERG* Overexpression is Highly Associated with *ERG* Gene Amplifications in Patients with Myeloid Malignancies**

55th ASH Annual Meeting and Exposition; December 7-10, 2013; New Orleans, LA

Simone Weber, Tamara Alpermann, Christiane Eder, Frank Dicker, Sabine Jeromin, Alexander Kohlmann, Annette Fasan, Claudia Haferlach, Wolfgang Kern, Torsten Haferlach, Susanne Schnittger

High *BAALC* Expression in Cytogenetically Normal Acute Myeloid Leukemia Strongly Correlates with Adverse Markers Such As *RUNX1*mut, *MLL*-PTD and *FLT3*-ITD and Is Useful for Disease Monitoring

54th ASH Annual Meeting and Exposition; December 8-11, 2012; Atlanta, GA

Scientific talks

Simone Weber

Amplification of chromosome 21 as a mechanism for *ERG* overexpression in patients with myeloid malignancies

18th Congress of the European Hematology Association; June 13–16, 2013; Stockholm, Sweden

Simone Weber

In acute myeloid leukemia high *BAALC* expression levels strongly correlate with adverse outcome but also with established adverse markers such as *RUNX1*mut, *MLL*-PTD and *FLT3*mut/wt ratio>0.5

Gemeinsame Jahrestagung 2012 der DGHO, ÖGHO, SGMO und SGH+SSH; October 19-23, 2012; Stuttgart, Germany

Curriculum vitae

Simone Weber

Date of birth	**June, 12th 1984**
Place of birth	**Freiburg im Breisgau, Germany**
Nationality	**German**
Since 12/2012	Scientific Associate and PhD student Münchner Leukämielabor GmbH and Technical University Munich, Germany
10/2010 – 12/2012	Scientific Associate Münchner Leukämielabor GmbH, Munich, Germany Department of Molecular Genetics
05/2010 – 10/2010	Scientific Associate Institute for molecular Microbiology and Hygiene, Freiburg, Germany Department of Virology
09/2006 – 05/2010	Advanced studies in Biology (Diploma) Albert-Ludwigs-University Freiburg, Germany Degree: Diploma in Biology
09/2004 – 09/2006	Basic studies in Biology Albert-Ludwigs-University Freiburg, Germany Degree: Intermediate diploma
07/1990 – 07/2003	Primary and Secondary School, Freiburg, Germany Graduation: Abitur (A-levels)

Appendix I - VI

Appendix I

Simone Weber et al.

***BAALC* expression: a suitable marker for prognostic risk stratification and detection of residual disease in cytogenetically normal acute myeloid leukemia**

Blood Cancer Journal. 2014 Jan 10; 4:e173. doi: 10.1038/bcj.2013.71.

Appendix II

Simone Weber et al.

Gain of chromosome 21 or amplification of chromosome arm 21q is one mechanism for increased *ERG* expression in acute myeloid leukemia

Genes Chromosomes Cancer. 2016 Feb;55(2):148-57. doi: 10.1002/gcc.22321. Epub 2015 Nov 6.

Appendix III

Simone Weber et al.

Feasibility of *BAALC* gene expression for detection of minimal residual disease and risk stratification in normal karyotype acute myeloid leukaemia

British Journal of Haematology. 2016 Dec;175(5):904-916. doi: 10.1111/bjh.14343. Epub 2016 Sep 23.

Appendix IV

Simone Weber et al.

Comprehensive study on *ERG* gene expression in normal karyotype acute myeloid leukemia: *ERG* expression is of limited prognostic value, whereas the accumulation of adverse prognostic markers stepwise worsens the prognosis

Blood Cancer J. 2016 Dec 09; 4:e173. doi:10.1038/bcj.2016.120.

Appendix V

Validation of *ABL1* for the assessment of *BAALC* and *ERG* gene expression

The use of *ABL1* as a reference gene has been proposed by a collaborative initiation within the Europe Against Cancer (EAC) program, including 26 laboratories in 10 member countries, where 14 potential reference genes have been evaluated on normal and diagnostic leukemic samples.[139] Therefore, *ABL1* represents the most commonly used reference gene in hematology. However, when testing *BAALC* and *ERG* expression the magnitude of quantification cycles (Cq) of *ABL1* represented a quite broad range from 23 to 34 cycles, which is beyond the expected range of common reference genes. The differential Cq values are most probably caused by differences in yield (cell numbers) and variations in RNA and cDNA synthesis. But, other factors such as differences in RNA quality or correlated regulation of *ABL1* and *BAALC* or *ERG* gene expression, could also cause the differential expression of *ABL1*. To exclude these parameters as source of differential expression we performed additional analysis. Since all analysis were performed on primary patient samples we had to deal with limited amount of material. Therefore, we defined a core-set of 10 primary AML samples with sufficient material, reflecting the full magnitude of Cq. In this core-set the RNA quality and stability of *ABL1* was re-evaluated. RNA integrity and concentration was assessed using the RNA 6000 Pico Kit (Agilent Technologies, Inc., Santa Clara, CA). Afterwards, 0.1µg of total RNA and 0.01µg of mRNA of each sample was reverse transcribed and the quantity of *ABL1* was analyzed as outlined above. The newly defined Cq values were compared with the originally measured Cq values.

The estimation of the integrity of RNA resulted in a median RIN value (RNA integrity number) of 8.4 demonstrating intact RNA in these cases (range: 6.3 – 9.2). Quantifying *ABL1*, the magnitude of Cq ranged from 27.3 to 29.3 when analyzing cDNA processed from total RNA and from 24.6 to 27.9 when analyzing cDNA processed from mRNA. Conclusively these results suggest that differential Cq values are most probably caused by differences in yield (cell numbers) and variations in RNA and cDNA synthesis, as the quality of the RNA was acceptable and moreover since the broad magnitude of Cq value narrowed down when analyzing cDNA processed from defined amounts of RNA.

Appendix VI

Table S1: Primer sequences used to analyze the respective genes by 454 Pyrosequencing

Gene-Identifier	Material	Sequence: Forward-Primer(5´-3´)	Sequence: Reverse-Primer(5´-3´)
ASXL1-E13.01	gDNA	GGTCAGATCACCCAGTCAGTT	CCTCACCACCATCACCACT
ASXL1-E13.02	gDNA	ATGAGGGAGGTGGCAGAG	ACTGAGGTTTGGGAGGACAG
ASXL1-E13.03	gDNA	AGGAGATGCCTCCCAACTC	TGTTCTGCAGGCAATCAGTC
ASXL1-E13.04	gDNA	AGAAAGGAACTGGCCAAGC	CAGCTTTCTCCCACTCCTCT
ASXL1-E13.05	gDNA	GAGAACACATACCATCTGTTGAGC	TGGGTTTCTCATCCTTGTCC
ASXL1-E13.06	gDNA	AGTCCACGGATACAGCCTCT	AGCAACTGCATCACAAGTGG
ASXL1-E13.07	gDNA	ATGCCTGGGTCCTCAGTG	GGACTGCCTTGCAATTCTTT
ASXL1-E13.08	gDNA	ACTGGTCTTGCCAGGATTG	GGCCTGTAACATTGCTCTGA
ASXL1-E13.09	gDNA	CACCTCGTTTCTCATCTCCA	GGCATTTGCCTTAAGAGGAC
ASXL1-E13.10	gDNA	CCAAGCACAAACTCCATGTC	CTGGTGGAACTCAGTTGGAG
ASXL1-E13.11	gDNA	CAGCATCAAGCAGGCATTT	CAATGTTTCCCATGGCCATA
CEBPA-E01.01	gDNA	GCCATGCCGGGAGAACT	CCCGGGTAGTCAAAGTCG
CEBPA-E01.02	gDNA	CCTTCAACGACGAGTTCCTG	CGGCTGGTAAGGGAAGAGG
CEBPA-E01.03	gDNA	GAGGAGGATGAAGCCAAGC	CTCGTTGCTGTTCTTGTCCA
CEBPA-E01.04	gDNA	TGGCAGCGCGCTCAAG	CCAGGGCGGTCCCACA
CEBPA-E01.XL	gDNA	ATCGACATCAGCGCCTACAT	GCAGCGTGTCCAGTTCG
DNMT3A-E02	gDNA	GCCTCCAAAGACCACGATAA	TGCTCCATCACCAGAATTGA
DNMT3A-E03	gDNA	ACAGGCCTGGAATGCTACAC	GTGAGGGTGGGATTTGAAGA
DNMT3A-E04	gDNA	AAACAGGCTAAGCCCACTGA	AAGAGGCTGCCCCTGGT
DNMT3A-E05	gDNA	GGTGGGCCATCATATTTAACA	CACTTCCTCTCCTTCCCACA
DNMT3A-E06	gDNA	CCACCCTAATGCCCTAATGTC	GCTGAAGGAGCAGATGAACC
DNMT3A-E07	gDNA	TTCCTGGAGAGGTCAAGGTG	TGGAGAGAGGAGAGCAGGAC
DNMT3A-E08	gDNA	GCCTCGTGACCACTGTGTAA	ACCCACCACAGGCAGAGTAG
DNMT3A-E09	gDNA	CTCCTCTTTGCATCGGGTAA	ACCTGCACTCCAACTTCCAG
DNMT3A-E10	gDNA	TGTGCCACCCTCACTACTCA	TCCCTAAGCATGGCTTTCC
DNMT3A-E11-12	gDNA	GACCTTGGCACCTGCTTTC	CCACACTAGGAGTGCCAGAGTT
DNMT3A-E13	gDNA	GGTCACAGTGCCTCCCTTT	ACCCTGTACATGCCCAGAAG
DNMT3A-E14	gDNA	CACAGGCAGATGAGGTTTCC	CCCAGCTAAGGAGACCACTG
DNMT3A-E15	gDNA	CCCTAGCCATGCTCCAGAC	CCCACAACCAAGGCTCAG
DNMT3A-E16	gDNA	CAGGGTGTGTGGGTCTAGGA	TGCATACGTTTCCACTTCACA
DNMT3A-E17	gDNA	AAAGATAGGACTTGGGCCTACA	CTGCCTCCAGGTGCTGAG
DNMT3A-E18	gDNA	TGGTCCCGTTCTTGTTTAGG	CAAGGAGGAAGCCTATGTGC
DNMT3A-E19	gDNA	GACAGCTATTCCCGATGACC	GCTCCACAATGCAGATGAGA
DNMT3A-E20	gDNA	TGTGTGGCTCCTGAGAGAGA	CATGGCAGAGCAGCTAGTCA
DNMT3A-E21	gDNA	TGGTGGATTTGTGTCTTTGC	CATCCTGCCCTTCCTTCTC
DNMT3A-E22	gDNA	CTGCGAACTCTGCTCACTCA	AGCAAGCACAGCAATCAGAA
DNMT3A-E23	gDNA	CACTCACCCTGCCCTCTCT	AAAGCCCTCCGGTATTTCC
DNMT3A-FE06-08	cDNA	GGGGGACCCCTACTACATCA	CCTGGCCACCAGGAGAAG
DNMT3A-FE07-09	cDNA	GTGCAGCAGCCCACTGAC	GGGCTGCTTGTTGTACGTG
DNMT3A-FE08-11	cDNA	GTTCGGAGACGGCAAATTCT	AGGTTCCACCCACATGTCC
DNMT3A-FE10-14	cDNA	GGAGCCACCAGAAGAAGAGA	CAGCAGATGGTGCAGTAGGA
DNMT3A-FE14-16	cDNA	TTCTGGAGTGTGCGTACCAG	GATGGGCTTCCTCTTCTCAG
DNMT3A-FE15-19	cDNA	GGAATTTGACCCTCCAAAGG	AAAGAAGAGCCGGCCAGT
DNMT3A-FE18-22	cDNA	CAATGACCTCTCCATCGTCA	TGCTGAACTTGGCTATCCTG
DNMT3A-FE21-23	cDNA	CCACTGTGAATGATAAGCTGGA	TTGTGTCGCTACCTCAGTTTG
FLT3-E13	gDNA	TCTGTTTCATCGCTGAGTGAC	TCTTGGAAACTCCCATTTGA
FLT3-E14	gDNA	TCTGCAGAACTGCCTATTCCT	GCAACCTGGATTGAGACTCC
FLT3-E15	gDNA	GTGACCGGCTCCTCAGATAA	TGCATCTTTGTTGCTGTCCT
FLT3-E20	gDNA	GGCACAGCCCAGTAAAGATAAG	CACCACAGTGAGTGCAGTTG
IDH1-E04	gDNA	AAACTTTGCTTCTAATTTTTCTCTTTC	GCAAAATCACATTATTGCCAAC
IDH2-E04	gDNA	TGCAGTGGGACCACTATTATCTC	CACCACTGCCATCTTTTGG
NPM1-E06	gDNA	TGAAGAGGAGGAGGATGTGAAA	CATAGGATGGCTTTTTGTAAATGACT
NPM1-E07	gDNA	TCGATAAACATGGGTGCACAA	GAAAACAGACCATTCCCCACT
NPM1-E11	gDNA	TTTTCCAGGCTATTCAAGAT	ATATTTTTGTCTCCCCACCA
NRAS-E02	gDNA	GGCCGATATTAATCCGGTGT	TCCGACAAGTGAGAGACAGG
NRAS-E03	gDNA	CCCTCCCTGCCCCCTTAC	CAATGTCAAACAACCTAAAACCA
NRAS-E04	gDNA	TGTTGCCCAGGCTAATCTCA	GCAAACTCTTGCACAAATGC
RUNX1-E01	gDNA	GCTGTTTGCAGGGTCCTAAC	GGCCTCCGCCTGTCCTC
RUNX1-E02	gDNA	CATTGCTATTCCTCTGCAACC	GTTTGTTGCCATGAAACGTG
RUNX1-E03	gDNA	AAATTCCGGGAGTGTTGTCA	GAAAGGTTGAACCCAAGGAA
RUNX1-E04	gDNA	TGATCTCTTCCCTCCCTCCT	CAGTTGGTCTGGGAAGGTGT
RUNX1-E05	gDNA	ATTTGAACAAGGGCCACTCA	AATGTTCTGCCAACTCCTTCA
RUNX1-E06.01	gDNA	CTCCGCAACCTCCTACTCAC	CCCACCATGGAGAACTGGTA
RUNX1-E06.02	gDNA	CCCGTTCCAAGCCAGCTC	GCTTGTCGCGAACAGGAG
RUNX1-FE01-03	cDNA	AGCATGGTGGAGGTGCTG	TGATGGCTCTGTGGTAGGTG
RUNX1-FE02-04	cDNA	TGAGCTGAGAAATGCTACCG	AAGGCAGTGGAGTGGTTCAG
RUNX1-FE03-05	cDNA	ACAAACCCACCGCAAGTC	ACAGAAGGAGAGGCAATGGA
RUNX1-FE04-06	cDNA	ATGAGGGTCAGCCCACAC	GTGAAGGCGCCTGGATAGT
TET2-E03.01	gDNA	ATTCAACTAGAGGGCAGCCTTG	ACTGTGCGTTTTATTCCTCCAT
TET2-E03.02	gDNA	GAATACCCTGTATGAAGGGAAGC	CCCACTGCAGTTATGTGTTGAA
TET2-E03.03	gDNA	TGTAGCCCAAGAAAATGCAG	TGGGTGAGTGATCTCACAGG
TET2-E03.04	gDNA	CATCTCACATAAATGCCATTAACA	AGCTTGCAAATTGCTGCTG
TET2-E03.05	gDNA	GAAAATAACATCCAGGGAACCA	CCCTCTATTTTCACTTCCCTTAAA
TET2-E03.06	gDNA	GGAGTTTTAGAAGAACACCACCA	TCGACCCTTCAGAATCTCTTG

TET2-E03.07	gDNA	CCAATTTTTGGTAGCAGTGGA	CCAGCTGTGTTGTTTTCTGG
TET2-E03.08	gDNA	TGACCTCCAAACAATACACTGG	TGAGTTTGAAAATGGCTCAGTC
TET2-E03.09	gDNA	CCCAGTGTTGAAACAGCA	ACTTCCTCCAGTCCCATTTG
TET2-E03.10	gDNA	TGGTGAAAATCAGTATTCAAAATCA	CCCTGTAGAACTGAAGCTTGTTG
TET2-E03.11	gDNA	CTTCTTCACAGGTGCTTTCAAG	ATACAGGCATGTGGCTTGC
TET2-E03.12	gDNA	TTGCCATAGTCAGATGCACAG	CTGAAGAAGTTGTTTGCTGCTCT
TET2-E03.13	gDNA	TTGACTAGACAAACCACTGCTG	TTTATGAGCCTTTACAAATTGCTG
TET2-E04	gDNA	TGGCACATTTTCTAATAGATCAGTC	CTTTGTGTGTGAAGGCTGGA
TET2-E05	gDNA	AAACCGTTCATTTCTCAGGATG	GTAATGTTCTTTTTAACTGGCATGA
TET2-E06	gDNA	TGACCCTTGTTTTGTTTTGG	CGCTGAACTCTCTTCCTTTCA
TET2-E07	gDNA	ATAGACACCTATAATATCAGCTGCAC	CAGTTTGGGAAAAACTTTGATTA
TET2-E08	gDNA	CCATATATTGTGTTTGGGATTCAA	GCAGTGGTTTCAACAATTAAGAG
TET2-E09	gDNA	TGCTCTATTTTGTGTCATTCCATT	CAGTGTGAGAACAGACTCAACAG
TET2-E10.01	gDNA	GGGACCTGTAGTTGAGGCTGT	GGGGCTGACTTTTCCTTTTC
TET2-E10.02	gDNA	GAGTTTGGGAGTGTGGAAGC	GGGGGCAAAACCAAAATAAT
TET2-E11.01	gDNA	GCCTTCATAAAATAATCATCAACA	CTGCAGCTTGAGATGAGGTG
TET2-E11.02	gDNA	CCAATCCAGTTAGTCCTTATCCA	AAAACTCTGGCTATTTCCAAACC
TET2-E11.03	gDNA	CAAGCCAAGACCCTCTGTCT	GCATGAAGAGAGCTGTTGAA
TET2-E11.04	gDNA	GGTGAACATCATTCACCTTCTC	GAATTGACCCATGAGTTGGAG
TET2-E11.05	gDNA	AGACAGCGAGCAGAGCTTTC	AAGTTTCATGTGGCTCAGCA
TET2-E11.06	gDNA	AGCCCGTGAGAAAGAGGAAG	ACTGTGACCTTTCCCCACTG
TET2-FE03-04	cDNA	CACGCCAAGTCGTTATTTGA	GGACCTGCTCCTAGATGGGTA
TET2-FE03-06	cDNA	AACAGCTGCTTCTGTTCTCAATAA	CCCGCACCAAACACAGTAG
TET2-FE03.07	cDNA	CCAATTTTTGGTAGCAGTGGA	CCAGCTGTGTTGTTTTCTGG
TET2-FE05-07	cDNA	TCATCTATACTGGTAAAGAAGGCAAA	ACATGCTCCATGAACAACCA
TET2-FE06-08	cDNA	GTGAGGCTGCAGTGATTGTG	TGGTGCCATAAGAGTGGACA
TET2-FE07-10	cDNA	TGGAGCATGTACTACAATGGATG	GGTTTTCCTCCAAATTCTCG
TET2-FE09-10	cDNA	ATTCTCAGGGGTCACTGCAT	TGTTCTCCAGGGAGGAAAGC
TET2-FE10-11	cDNA	AGACTTGCCGACAAAGGAAA	GGCCGTCTCATGTATGGATT
TP53-E04	gDNA	ACCTGGTCCTCTGACTGCTC	CAGGCATTGAAGTCTCATGG
TP53-E05	gDNA	CACTTGTGCCCTGACTTTCA	CACTCGGATAAGATGCTGAGG
TP53-E06	gDNA	CAGATAGCGATGGTGAGCAG	TTGCACATCTCATGGGGTTA
TP53-E07	gDNA	GCACTGGCCTCATCTTGG	AAGAGGTCCCAAAGCCAGAG
TP53-E08	gDNA	GGACAGGTAGGACCTGATTTC	TCTCCATCCAGTGGTTTCTTC
TP53-E09	gDNA	AAAGGGGAGCCTCACCAC	TGTCTTTGAGGCATCACTGC
TP53-E10	gDNA	GCTGTATAGGTACTTGAAGTGCAG	CTGCCTTTGACCATGAAGG
TP53-E11	gDNA	AGGGAAAAGGGGCACAG	CCCCACAACAAAACACCAGT
WT1-E01.01	gDNA	TCCTCGCCGCGATCCTG	GGCGCAAAGTCCAGCAC
WT1-E01.02	gDNA	CAGCGCTGAACGTCTCCA	GCTCAGGCACTGCTCCTC
WT1-E01.03	gDNA	GCCTCACTCCTTCATCAAAC	GGGTAAGAGCTGCGGTCA
WT1-E02	gDNA	CGCTGACACTGTGCTTCTCT	TACTTGGGATGGAGGGAGAC
WT1-E03	gDNA	CAGGCTCAGGATCTCGTGTC	GTCTCGTGCCTCCAAGACC
WT1-E04	gDNA	TCACATCCCTTCTGCTGTGT	CTTTGAAATGGTTCAAACAGG
WT1-E05	gDNA	ATCTGATTCCAAAGCCCAAG	TCAGTCCTAACTCCTGCATTG
WT1-E06	gDNA	AGAACCCTGCATCTAAAGTGG	AAGGAACTAAAGGGCCGGTA
WT1-E07	gDNA	GCTCACTCTCCCTCAAGACC	AGCAGTGCTTACTTTCCATCC
WT1-E08	gDNA	GGATCATCCTACCCTAACAAGC	AAACTAAACACATGGCTGACTCTC
WT1-E09	gDNA	CAGGCATGGCAGGAAATG	GCCACGCACTATTCCTTCTC
WT1-E10	gDNA	ACTTCACTCGGGCCTTGATA	CACCGGTATCTTGTCTTGGA
ASXL1-E13.01	gDNA	ACACTGACGACATGGTTCTACATCCTAGTTTTGCTTTACAGTCCC	TACGGTAGCAGAGACTTGGTCTGCAGTGGTGACCTCTCGC
ASXL1-E13.02	gDNA	ACACTGACGACATGGTTCTACAACCCTCGCAGACATTAAAGC	TACGGTAGCAGAGACTTGGTCTCGGCAGTAGTTGTGTTCGC
ASXL1-E13.03	gDNA	ACACTGACGACATGGTTCTACAAGTGGTGATGGTGGTGAGG	TACGGTAGCAGAGACTTGGTCTGGGGAGTTGGGAGGCATC
ASXL1-E13.04	gDNA	ACACTGACGACATGGTTCTACATCTCTGAGAAAGGAGGAAAGCT	TACGGTAGCAGAGACTTGGTCTGTGGGTCCTTGCTCCTCAT
ASXL1-E13.05	gDNA	ACACTGACGACATGGTTCTACAGAACTGAATGTGAGTCTGGCA	TACGGTAGCAGAGACTTGGTCTATGAGCCACCAAGCCCTAAT
ASXL1-E13.06	gDNA	ACACTGACGACATGGTTCTACACCAGTTCCACACCTGAATCCT	TACGGTAGCAGAGACTTGGTCTCCAACCTGGGGCTCAACA
ASXL1-E13.07	gDNA	ACACTGACGACATGGTTCTACATCGAATGATGAGGTAGTGAAACAG	TACGGTAGCAGAGACTTGGTCTTTCAGCTTTTCAATGTCCACCT
ASXL1-E13.08	gDNA	ACACTGACGACATGGTTCTACAGACTGTGCCATCTCGAGGAG	TACGGTAGCAGAGACTTGGTCTCATTCACCTTGGACAGTGGG
ASXL1-E13.09	gDNA	ACACTGACGACATGGTTCTACAAGGGATCTTCGGTGGACAAG	TACGGTAGCAGAGACTTGGTCTCATCACAAGTGGGTTAGTGGC
ASXL1-E13.10	gDNA	ACACTGACGACATGGTTCTACAGATTCCCTACTGCTGGCCA	TACGGTAGCAGAGACTTGGTCTCCATCAACCATGCCACTGTT
ASXL1-E13.11	gDNA	ACACTGACGACATGGTTCTACACATGGCTCGCTACGCATG	TACGGTAGCAGAGACTTGGTCTCCATTTCTTCCAGTTTCCTAGAGG
ASXL1-E13.12	gDNA	ACACTGACGACATGGTTCTACAGGCAGTCCCAAGTTTTGACT	TACGGTAGCAGAGACTTGGTCTTCTGCTCTGGACCAAAGGAG
ASXL1-E13.13	gDNA	ACACTGACGACATGGTTCTACAGCCCAGGAGATCTTACTACCT	TACGGTAGCAGAGACTTGGTCTAGACATGGAGTTTGTGCTTGG
ASXL1-E13.14	gDNA	ACACTGACGACATGGTTCTACATTTTGGCTCTGGGAATGTGG	TACGGTAGCAGAGACTTGGTCTATGCCCAGTAGCTTTCCTGT
ASXL1-E13.15	gDNA	ACACTGACGACATGGTTCTACATTTGTTGGCAGCGTCAAGAA	TACGGTAGCAGAGACTTGGTCTAGCTTCCCATAAAATGCCTGC
ASXL1-E13.16	gDNA	ACACTGACGACATGGTTCTACAGGGAAGGGGCTCAGTGAG	TACGGTAGCAGAGACTTGGTCTTCCACCGTGCTGCTGTCA
ASXL1-E13.17	gDNA	ACACTGACGACATGGTTCTACAGCAAACTTTGGTGCGAGCC	TACGGTAGCAGAGACTTGGTCTAGATCAATCATTATCAAAATACACACA
DNMT3A-E02	gDNA	ACACTGACGACATGGTTCTACACCCACTGATCCTCTTCTCTCC	TACGGTAGCAGAGACTTGGTCTCGGCTGTCATCACATAGGG
DNMT3A-E03	gDNA	ACACTGACGACATGGTTCTACAACAGGCCTGGAATGCTACAC	TACGGTAGCAGAGACTTGGTCTGTGAGGGTGGGATTTGAAGA
DNMT3A-E04	gDNA	ACACTGACGACATGGTTCTACAAAACAGGCTAAGCCCACTGA	TACGGTAGCAGAGACTTGGTCTAAGAGGCTGCCCCTGGT
DNMT3A-E05	gDNA	ACACTGACGACATGGTTCTACAGGTGGGCCATCATATTTAACA	TACGGTAGCAGAGACTTGGTCTCACTTCCTCTCCTTCCCACA
DNMT3A-E06	gDNA	ACACTGACGACATGGTTCTACACCACCCTAATGCCCTAATGTC	TACGGTAGCAGAGACTTGGTCTGCTGAAGGAGCAGATGAACC
DNMT3A-E07	gDNA	ACACTGACGACATGGTTCTACATTCCTGGAGAGGTCAAGGTG	TACGGTAGCAGAGACTTGGTCTTGGAGAGAGGAGAGCAGGAC
DNMT3A-E08	gDNA	ACACTGACGACATGGTTCTACAGCCTCGTGACCACTGTGTAA	TACGGTAGCAGAGACTTGGTCTACCCACCACAGGCAGAGTAG
DNMT3A-E09	gDNA	ACACTGACGACATGGTTCTACACTCCTCTTTGCATCGGGTAA	TACGGTAGCAGAGACTTGGTCTACCTGCACTCCAACTTCCAG
DNMT3A-E10	gDNA	ACACTGACGACATGGTTCTACATGTGCCACCCTCACTACTCA	TACGGTAGCAGAGACTTGGTCTTCCCTAAGCATGGCTTTCC
DNMT3A-E11-12	gDNA	ACACTGACGACATGGTTCTACAGACCTTGGCACCTGCTTTC	TACGGTAGCAGAGACTTGGTCTCCACACTAGGAGTGCCAGAGTT
DNMT3A-E13	gDNA	ACACTGACGACATGGTTCTACATTGGTGCTCTGAGATGATGG	TACGGTAGCAGAGACTTGGTCTTGGACACAGTCAGCCAGAAG
DNMT3A-E14	gDNA	ACACTGACGACATGGTTCTACACACAGGCAGATGAGGTTTCC	TACGGTAGCAGAGACTTGGTCTCCCAGCTAAGGAGACCACTG
DNMT3A-E15	gDNA	ACACTGACGACATGGTTCTACACCCTAGCCATGCTCCAGAC	TACGGTAGCAGAGACTTGGTCTCCCACAACCAAGGCTCAG
DNMT3A-E16	gDNA	ACACTGACGACATGGTTCTACACAGGGTGTGTGGGTCTAGGA	TACGGTAGCAGAGACTTGGTCTTGCATACGTTTCCACTTCACA
DNMT3A-E17	gDNA	ACACTGACGACATGGTTCTACAAAAGATAGGACTTGGGCCTACA	TACGGTAGCAGAGACTTGGTCTCTGCCTCCAGGTGCTGAG
DNMT3A-E18	gDNA	ACACTGACGACATGGTTCTACATGGTCCCGTTCTTGTTTAGG	TACGGTAGCAGAGACTTGGTCTCAAGGAGGAAGCCTATGTGC
DNMT3A-E19	gDNA	ACACTGACGACATGGTTCTACAGACAGCTATTCCCGATGACC	TACGGTAGCAGAGACTTGGTCTGCTCCACAATGCAGATGAGA
DNMT3A-E20	gDNA	ACACTGACGACATGGTTCTACATGTGTGGCTCCTGAGAGAGA	TACGGTAGCAGAGACTTGGTCTCATGGCAGAGCAGCTAGTCA
DNMT3A-E21	gDNA	ACACTGACGACATGGTTCTACATGGTGGATTTGTGTCTTTGC	TACGGTAGCAGAGACTTGGTCTCATCCTGCCCTTCCTTCTC
DNMT3A-E22	gDNA	ACACTGACGACATGGTTCTACACTGCGAACTCTGCTCACTCA	TACGGTAGCAGAGACTTGGTCTAGCAAGCACAGCAATCAGAA

DNMT3A-E23	gDNA	ACACTGACGACATGGTTCTACACACTCACCCTGCCCTCTCT	TACGGTAGCAGAGACTTGGTCTAAAGCCCTCCGGTATTTCC
FLT3-E20X	gDNA	ACACTGACGACATGGTTCTACAGCACTCCAGGATAATACACATCA	TACGGTAGCAGAGACTTGGTCTACAGTGAGTGCAGTTGTTTACC
IDH1-E04	gDNA	ACACTGACGACATGGTTCTACAAAACTTTGCTTCTAATTTTTCTCTTTC	TACGGTAGCAGAGACTTGGTCTGCAAAATCACATTATTGCCAAC
IDH2-E04	gDNA	ACACTGACGACATGGTTCTACATGCAGTGGGACCACTATTATCTC	TACGGTAGCAGAGACTTGGTCTCACCACTGCCATCTTTTGG
NPM1-E11	gDNA	ACACTGACGACATGGTTCTACAAGATGTTGAACTATGCAAAGAGACA	TACGGTAGCAGAGACTTGGTCTACCAAGCAAAGGGTGGAGTT
NRAS-E02	gDNA	ACACTGACGACATGGTTCTACATACTGTAGATGTGGCTCGCC	TACGGTAGCAGAGACTTGGTCTTGATCCGACAAGTGAGAGACA
NRAS-E03	gDNA	ACACTGACGACATGGTTCTACACCCTCCCTGCCCCCTTAC	TACGGTAGCAGAGACTTGGTCTAGATTCAGAACACAAAGATCATCCT
TET2-E03.01	gDNA	ACACTGACGACATGGTTCTACAATTCAACTAGAGGGCAGCCTTG	TACGGTAGCAGAGACTTGGTCTACTGTGCGTTTTATTCCTCCAT
TET2-E03.02	gDNA	ACACTGACGACATGGTTCTACAGAATACCCTGTATGAAGGGAAGC	TACGGTAGCAGAGACTTGGTCTCCCACTGCAGTTATGTGTTGAA
TET2-E03.03	gDNA	ACACTGACGACATGGTTCTACATGTAGCCCAAGAAAATGCAG	TACGGTAGCAGAGACTTGGTCTTGGGTGAGTGATCTCACAGG
TET2-E03.04	gDNA	ACACTGACGACATGGTTCTACACATCTCACATAAATGCCATTAACA	TACGGTAGCAGAGACTTGGTCTAGCTTGCAAATTGCTGCTG
TET2-E03.05	gDNA	ACACTGACGACATGGTTCTACAGAAAATAACATCCAGGGAACCA	TACGGTAGCAGAGACTTGGTCTCCCTCTATTTTCACTTCCCTTAAA
TET2-E03.06	gDNA	ACACTGACGACATGGTTCTACAGGAGTTTTAGAAGAACACCACCA	TACGGTAGCAGAGACTTGGTCTTCGACCCTTCAGAATCTCTTG
TET2-E03.07	gDNA	ACACTGACGACATGGTTCTACACCAATTTTTGGTAGCAGTGGA	TACGGTAGCAGAGACTTGGTCTCCAGCTGTGTTGTTTTCTGG
TET2-E03.08	gDNA	ACACTGACGACATGGTTCTACATGACCTCCAAACAATACACTGG	TACGGTAGCAGAGACTTGGTCTTGAGTTTGAAAATGGCTCAGTC
TET2-E03.09	gDNA	ACACTGACGACATGGTTCTACACCCAGTGTTGAAACAGCA	TACGGTAGCAGAGACTTGGTCTACTTCCTCCAGTCCCATTTG
TET2-E03.10	gDNA	ACACTGACGACATGGTTCTACATGGTGAAAATCAGTATTCAAAATCA	TACGGTAGCAGAGACTTGGTCTTTGACCAGACATATCTTGGTTTC
TET2-E03.11	gDNA	ACACTGACGACATGGTTCTACACTTCTTCACAGGTGCTTTCAAG	TACGGTAGCAGAGACTTGGTCTATACAGGCATGTGGCTTGC
TET2-E03.12	gDNA	ACACTGACGACATGGTTCTACATTGCCATAGTCAGATGCACAG	TACGGTAGCAGAGACTTGGTCTCTGAAGAAGTTGTTTGCTGCTCT
TET2-E03.13	gDNA	ACACTGACGACATGGTTCTACATTGACTAGACAAACCACTGCTG	TACGGTAGCAGAGACTTGGTCTTTTATGAGCCTTTACAAATTGCTG
TET2-E04	gDNA	ACACTGACGACATGGTTCTACATGGCACATTTTCTAATAGATCAGTC	TACGGTAGCAGAGACTTGGTCTCTTTGTGTGTGAAGGCTGGA
TET2-E05	gDNA	ACACTGACGACATGGTTCTACAAAACCGTTCATTTCTCAGGATG	TACGGTAGCAGAGACTTGGTCTGTAATGTTCTTTTTAACTGGCATGA
TET2-E06	gDNA	ACACTGACGACATGGTTCTACAGCCCTTATCTGCTGCAAGTG	TACGGTAGCAGAGACTTGGTCTTTGGGCTTTCCTATCAGTGG
TET2-E07	gDNA	ACACTGACGACATGGTTCTACAATAGACACCTATAATATCAGCTGCAC	TACGGTAGCAGAGACTTGGTCTCAGTTTGGGAAAAACTTTGATTA
TET2-E08	gDNA	ACACTGACGACATGGTTCTACACCATATATTGTGTTTGGGATTCAA	TACGGTAGCAGAGACTTGGTCTGCAGTGGTTTCAACAATTAAGAG
TET2-E09	gDNA	ACACTGACGACATGGTTCTACATGCTCTATTTTGTGTCATTCCATT	TACGGTAGCAGAGACTTGGTCTCAGTGTGAGAACAGACTCAACAG
TET2-E10.01	gDNA	ACACTGACGACATGGTTCTACAGGGACCTGTAGTTGAGGCTGT	TACGGTAGCAGAGACTTGGTCTGGGGCTGACTTTTCCTTTTC
TET2-E10.02	gDNA	ACACTGACGACATGGTTCTACAGAGTTTGGGAGTGTGGAAGC	TACGGTAGCAGAGACTTGGTCTGGGGGCAAAACCAAAATAAT
TET2-E11.01	gDNA	ACACTGACGACATGGTTCTACAGCCTTCATAAAATAATCATCAACA	TACGGTAGCAGAGACTTGGTCTCTGCAGCTTGAGATGAGGTG
TET2-E11.02	gDNA	ACACTGACGACATGGTTCTACACCAATCCAGTTAGTCCTTATCCA	TACGGTAGCAGAGACTTGGTCTAAAACTCTGGCTATTTCCAAACC
TET2-E11.03	gDNA	ACACTGACGACATGGTTCTACACAAGCCAAGACCCTCTGTCT	TACGGTAGCAGAGACTTGGTCTGCATGAAGAGAGCTGTTGAA
TET2-E11.04	gDNA	ACACTGACGACATGGTTCTACAGGTGAACATCATTCACCTTCTC	TACGGTAGCAGAGACTTGGTCTGAATTGACCCATGAGTTGGAG
TET2-E11.05	gDNA	ACACTGACGACATGGTTCTACAAGACAGCGAGCAGAGCTTTC	TACGGTAGCAGAGACTTGGTCTAAGTTTCATGTGGCTCAGCA
TET2-E11.06	gDNA	ACACTGACGACATGGTTCTACAAGCCCGTGAGAAAGAGGAAG	TACGGTAGCAGAGACTTGGTCTACTGTGACCTTTCCCCACTG
TP53-E04	gDNA	ACACTGACGACATGGTTCTACAACCTGGTCCTCTGACTGCTC	TACGGTAGCAGAGACTTGGTCTCAGGCATTGAAGTCTCATGG
TP53-E05	gDNA	ACACTGACGACATGGTTCTACACACTTGTGCCCTGACTTTCA	TACGGTAGCAGAGACTTGGTCTCACTCGGATAAGATGCTGAGG
TP53-E06	gDNA	ACACTGACGACATGGTTCTACACAGATAGCGATGGTGAGCAG	TACGGTAGCAGAGACTTGGTCTTTGCACATCTCATGGGGTTA
TP53-E07	gDNA	ACACTGACGACATGGTTCTACAGCACTGGCCTCATCTTGG	TACGGTAGCAGAGACTTGGTCTAAGAGGTCCCAAAGCCAGAG
TP53-E08	gDNA	ACACTGACGACATGGTTCTACAGGACAGGTAGGACCTGATTTC	TACGGTAGCAGAGACTTGGTCTTCTCCATCCAGTGGTTTCTTC
TP53-E09	gDNA	ACACTGACGACATGGTTCTACAAAAGGGGAGCCTCACCAC	TACGGTAGCAGAGACTTGGTCTTGTCTTTGAGGCATCACTGC
TP53-E10	gDNA	ACACTGACGACATGGTTCTACAGCTGTATAGGTACTTGAAGTGCAG	TACGGTAGCAGAGACTTGGTCTCTGCCTTTGACCATGAAGG
TP53-E11	gDNA	ACACTGACGACATGGTTCTACAAGGGAAAAGGGGCACAG	TACGGTAGCAGAGACTTGGTCTCCCCACAACAAAACACCAGT
WT1-E01.01	gDNA	ACACTGACGACATGGTTCTACATCCTCGCCGCGATCCTG	TACGGTAGCAGAGACTTGGTCTGGCGCAAAGTCCAGCAC
WT1-E01.02	gDNA	ACACTGACGACATGGTTCTACACAGCGCTGAACGTCTCCA	TACGGTAGCAGAGACTTGGTCTGCTCAGGCACTGCTCCTC
WT1-E01.03	gDNA	ACACTGACGACATGGTTCTACAGCCTCACTCCTTCATCAAAC	TACGGTAGCAGAGACTTGGTCTGGGTAAGAGCTGCGGTCA
WT1-E02	gDNA	ACACTGACGACATGGTTCTACACGCTGACACTGTGCTTCTCT	TACGGTAGCAGAGACTTGGTCTTACTTGGGATGGAGGGAGAC
WT1-E03	gDNA	ACACTGACGACATGGTTCTACACAGGCTCAGGATCTCGTGTC	TACGGTAGCAGAGACTTGGTCTGTCTCGTGCCTCCAAGACC
WT1-E04	gDNA	ACACTGACGACATGGTTCTACATCACATCCCTTCTGCTGTGT	TACGGTAGCAGAGACTTGGTCTCTTTGAAATGGTTCAAACAGG
WT1-E05	gDNA	ACACTGACGACATGGTTCTACAATCTGATTCCAAAGCCCAAG	TACGGTAGCAGAGACTTGGTCTTCAGTCCTAACTCCTGCATTG
WT1-E06	gDNA	ACACTGACGACATGGTTCTACAAGAACCCTGCATCTAAAGTGG	TACGGTAGCAGAGACTTGGTCTAAGGAACTAAAGGGCCGGTA
WT1-E07	gDNA	ACACTGACGACATGGTTCTACAGCTCACTCTCCCTCAAGACC	TACGGTAGCAGAGACTTGGTCTAGCAGTGCTTACTTTCCATCC
WT1-E08	gDNA	ACACTGACGACATGGTTCTACAGGATCATCCTACCCTAACAAGC	TACGGTAGCAGAGACTTGGTCTAAACTAAACACATGGCTGACTCTC
WT1-E09	gDNA	ACACTGACGACATGGTTCTACACAGGCATGGCAGGAAATG	TACGGTAGCAGAGACTTGGTCTGCCACGCACTATTCCTTCTC
WT1-E10	gDNA	ACACTGACGACATGGTTCTACAACTTCACTCGGGCCTTGATA	TACGGTAGCAGAGACTTGGTCTCACCGGTATCTTGTCTTGGA

Table S2: Primer sequences used to analyze the respective genes by MiSeq Sequencing

Gene-Identifier	Material	Sequence: Forward-Primer(5´-3´)	Sequence: Reverse-Primer(5´-3´)
ASXL1-E13.01	gDNA	ACACTGACGACATGGTTCTACATCCTAGTTTTGCTTTACAGTCCC	TACGGTAGCAGAGACTTGGTCTGCAGTGGTGACCTCTCGC
ASXL1-E13.03	gDNA	ACACTGACGACATGGTTCTACAACCCTCGCAGACATTAAAGC	TACGGTAGCAGAGACTTGGTCTCGGCAGTAGTTGTGTTCGC
ASXL1-E13.04	gDNA	ACACTGACGACATGGTTCTACAAGTGGTGATGGTGGTGAGG	TACGGTAGCAGAGACTTGGTCTGGGGAGTTGGGAGGCATC
ASXL1-E13.05	gDNA	ACACTGACGACATGGTTCTACATCTCTGAGAAAGGAGGAAAGCT	TACGGTAGCAGAGACTTGGTCTGTGGGTCCTTGCTCCTCAT
ASXL1-E13.06	gDNA	ACACTGACGACATGGTTCTACAGAACTGAATGTGAGTCTGGCA	TACGGTAGCAGAGACTTGGTCTATGAGCCACCAAGCCCTAAT
ASXL1-E13.07	gDNA	ACACTGACGACATGGTTCTACACCAGTTCCACACCTGAATCCT	TACGGTAGCAGAGACTTGGTCTCCAACCTGGGGCTCAACA
ASXL1-E13.08	gDNA	ACACTGACGACATGGTTCTACATCGAATGATGAGGTAGTGAAACAG	TACGGTAGCAGAGACTTGGTCTTTCAGCTTTTCAATGTCCACCT
ASXL1-E13.09	gDNA	ACACTGACGACATGGTTCTACAGACTGTGCCATCTCGAGGAG	TACGGTAGCAGAGACTTGGTCTCATTCACCTTGGACAGTGGG
ASXL1-E13.10	gDNA	ACACTGACGACATGGTTCTACAAGGGATCTTCGGTGGACAAG	TACGGTAGCAGAGACTTGGTCTCATCACAAGTGGGTTAGTGGC
ASXL1-E13.11	gDNA	ACACTGACGACATGGTTCTACAGATTCCCTACTGCTGGCCA	TACGGTAGCAGAGACTTGGTCTCCATCAACCATGCCACTGTT
ASXL1-E13.12	gDNA	ACACTGACGACATGGTTCTACACATGGCTCGCTACGCATG	TACGGTAGCAGAGACTTGGTCTCCATTTCTTCCAGTTTCCTAGAGG
ASXL1-E13.13	gDNA	ACACTGACGACATGGTTCTACAGGCAGTCCCAAGTTTTGACT	TACGGTAGCAGAGACTTGGTCTTCTGCTCTGGACCAAAGGAG
ASXL1-E13.14	gDNA	ACACTGACGACATGGTTCTACAGCCCAGGAGATCTTACTACCT	TACGGTAGCAGAGACTTGGTCTAGACATGGAGTTTGTGCTTGG
ASXL1-E13.15	gDNA	ACACTGACGACATGGTTCTACATTTTGGCTCTGGGAATGTGG	TACGGTAGCAGAGACTTGGTCTATGCCCAGTAGCTTTCCTGT
ASXL1-E13.16	gDNA	ACACTGACGACATGGTTCTACATTTGTTGGCAGCGTCAAGAA	TACGGTAGCAGAGACTTGGTCTAGCTTCCCATAAAATGCCTGC
ASXL1-E13.17	gDNA	ACACTGACGACATGGTTCTACAGGGAAGGGGCTCAGTGAG	TACGGTAGCAGAGACTTGGTCTTCCACCGTGCTGCTGTCA
ASXL1-E13.18	gDNA	ACACTGACGACATGGTTCTACAGCAAACTTTGGTGCGAGCC	TACGGTAGCAGAGACTTGGTCTAGATCAATCATTATCAAAATACACACA
CEBPA-E01.01	gDNA	ACACTGACGACATGGTTCTACACCATGCCGGGAGAACTCTA	TACGGTAGCAGAGACTTGGTCTGTCGTTGAAGGCGGCC
CEBPA-E01.02	gDNA	ACACTGACGACATGGTTCTACAATCTGCGAGCACGAGACG	TACGGTAGCAGAGACTTGGTCTCTCCAGCCTGCCGTCC
CEBPA-E01.03	gDNA	ACACTGACGACATGGTTCTACAGGCGACTTTGACTACCCGG	TACGGTAGCAGAGACTTGGTCTCGGCTGGTAAGGGAAGAGG
CEBPA-E01.04	gDNA	ACACTGACGACATGGTTCTACAGCCGCTGGTGATCAAGC	TACGGTAGCAGAGACTTGGTCTGACCGGGCTGCAGGTG
CEBPA-E01.05	gDNA	ACACTGACGACATGGTTCTACAGCAGTTCCAGATCGCGC	TACGGTAGCAGAGACTTGGTCTCGGTACTCGTTGCTGTTCTT
CEBPA-E01.06	gDNA	ACACTGACGACATGGTTCTACACTCAAGGGGCTGGGCG	TACGGTAGCAGAGACTTGGTCTCAGCTGGCGGAAGATGC
CEBPA-E01.07	gDNA	ACACTGACGACATGGTTCTACAGTGGACAAGAACAGCAACGA	TACGGTAGCAGAGACTTGGTCTCGCAGCGTGTCCAGTTC
CEBPA-E01.08	gDNA	ACACTGACGACATGGTTCTACAGCAAGCGGGTGGAACA	TACGGTAGCAGAGACTTGGTCTAGAGGAAGGGAGGGGACAC
DNMT3A-E07.01	gDNA	ACACTGACGACATGGTTCTACATTCCTGGAGAGGTCAAGGTG	TACGGTAGCAGAGACTTGGTCTTTTGGTGGCATTCTTGTCCC

DNMT3A-E07.02	gDNA	ACACTGACGACATGGTTCTACAGGAAGAAAACCAGGGGCCC	TACGGTAGCAGAGACTTGGTCTGAGGAGCTGGCAGTGGAAG
DNMT3A-E08	gDNA	ACACTGACGACATGGTTCTACACTCGTGACCACTGTGTAATGAT	TACGGTAGCAGAGACTTGGTCTGCCCTGGGATCAAGAACCTT
DNMT3A-E09	gDNA	ACACTGACGACATGGTTCTACAGGTGGGTGCTTGCAAGTG	TACGGTAGCAGAGACTTGGTCTACCTGCACTCCAACTTCCAG
DNMT3A-E10	gDNA	ACACTGACGACATGGTTCTACACTGACAACCCCAACCCTG	TACGGTAGCAGAGACTTGGTCTGCCCTGGTGTGGATCTGC
DNMT3A-E11	gDNA	ACACTGACGACATGGTTCTACACTGTAACTGACCTTGGCACCT	TACGGTAGCAGAGACTTGGTCTTCGGATGCAGGCCTCCT
DNMT3A-E12	gDNA	ACACTGACGACATGGTTCTACAAGAAGCCCAAGGTCAAGGAG	TACGGTAGCAGAGACTTGGTCTCTCTGCTACTCTGCCCCATG
DNMT3A-E13	gDNA	ACACTGACGACATGGTTCTACATTTTGTGGCTGGTCTGGTGC	TACGGTAGCAGAGACTTGGTCTAGAAGCGGTGGACACAGTC
DNMT3A-E14	gDNA	ACACTGACGACATGGTTCTACAGAGCCTCCCTTCGTCCTG	TACGGTAGCAGAGACTTGGTCTGGCCCAGCTAAGGAGACC
DNMT3A-E15.01	gDNA	ACACTGACGACATGGTTCTACACCAGGGCTGAGAGTCTCCT	TACGGTAGCAGAGACTTGGTCTGCCCAGCACTCACAAATTCC
DNMT3A-E15.02	gDNA	ACACTGACGACATGGTTCTACAGTGGAGTGTGTGGACCTCTT	TACGGTAGCAGAGACTTGGTCTCCCCACAACCAAGGCTCA
DNMT3A-E16	gDNA	ACACTGACGACATGGTTCTACACTGGGCCTGCATCTGACC	TACGGTAGCAGAGACTTGGTCTGTTTTGCCAGAGTTGCCCAC
DNMT3A-E17	gDNA	ACACTGACGACATGGTTCTACAAGATGGCTCCAAGTAACGGTG	TACGGTAGCAGAGACTTGGTCTGTGCAGGGAGGGGAAGAC
DNMT3A-E18	gDNA	ACACTGACGACATGGTTCTACACTTCCTGTCTGCCTCTGTCC	TACGGTAGCAGAGACTTGGTCTCATCGGGAATAGCTGTCCCA
DNMT3A-E19	gDNA	ACACTGACGACATGGTTCTACAGAGCCACACCACTGTCCTAT	TACGGTAGCAGAGACTTGGTCTCCATTAGTGAGCTGGCCAAAC
DNMT3A-E20	gDNA	ACACTGACGACATGGTTCTACACCTTGGCTCATCTTCAAACCG	TACGGTAGCAGAGACTTGGTCTAGCAGCTAGTCATTCAGCAGA
DNMT3A-E21	gDNA	ACACTGACGACATGGTTCTACAACCCTGTGAACTAGTGGCTG	TACGGTAGCAGAGACTTGGTCTCATCCTGCCCTTCCTTCTCC
DNMT3A-E22	gDNA	ACACTGACGACATGGTTCTACAGCATATTTGGTAGACGCATGACC	TACGGTAGCAGAGACTTGGTCTCACAGCAATCAGAACAGCCAC
DNMT3A-E23.02	gDNA	ACACTGACGACATGGTTCTACACACTCACCCTGCCCTCTC	TACGGTAGCAGAGACTTGGTCTTGTTTAACTTTGTGTCGCTACCT
FLT3-E06	gDNA	ACACTGACGACATGGTTCTACATGCAGCTGTAAAGAAGAAAGTCC	TACGGTAGCAGAGACTTGGTCTGAACCGGTCACTGAAAATGAA
FLT3-E20	gDNA	ACACTGACGACATGGTTCTACAGCACTCCAGGATAATACACATCA	TACGGTAGCAGAGACTTGGTCTACAGTGAGTGCAGTTGTTTACC
FLT3-E24	gDNA	ACACTGACGACATGGTTCTACATACTACAGAATTGCCCATTATTGAA	TACGGTAGCAGAGACTTGGTCTTGACAACCATAGCTGCCTACA
FLT3-I01.02	gDNA	ACACTGACGACATGGTTCTACAACCCTGGGTGACAGAGACAG	TACGGTAGCAGAGACTTGGTCTAGCCTCCCGAGTAACTAGGATT
FLT3-I01.03	gDNA	ACACTGACGACATGGTTCTACAGGAGAGAATGGAGGAAAGGAA	TACGGTAGCAGAGACTTGGTCTCAAGAAGTCTGCCAAATGAATG
FLT3-I01.06	gDNA	ACACTGACGACATGGTTCTACATTTGCCCAAATTGATCTAGTGA	TACGGTAGCAGAGACTTGGTCTGATGCCAATTCCACATTGTCT
FLT3-I01.07	gDNA	ACACTGACGACATGGTTCTACATTCAGTTAGGAGGGTAAGTTCTGG	TACGGTAGCAGAGACTTGGTCTAGGTGGAGATTGCAGTTGGT
FLT3-I01.09	gDNA	ACACTGACGACATGGTTCTACAGGAATGGGGTTGGTAGCAG	TACGGTAGCAGAGACTTGGTCTTACTTGGGAGGCTGAAGTGG
FLT3-I01.11	gDNA	ACACTGACGACATGGTTCTACAAGGCAGCAAGCAAGAGAGAG	TACGGTAGCAGAGACTTGGTCTTTTTCCCCAAATGACTGACAA
FLT3-I02	gDNA	ACACTGACGACATGGTTCTACATTTGCTCCTGTGTCCTTTCA	TACGGTAGCAGAGACTTGGTCTGGCACACCCAACTCTCAGAT
FLT3-I09.01	gDNA	ACACTGACGACATGGTTCTACAGCTGGCCATTTGTAGAAAACC	TACGGTAGCAGAGACTTGGTCTTGGCTTTAAACTGCCTTTCC
FLT3-I09.02	gDNA	ACACTGACGACATGGTTCTACACTTGAGGGCAGGAGTTTGAG	TACGGTAGCAGAGACTTGGTCTCCACCACAGCCAGCTAATTT
FLT3-I09.04	gDNA	ACACTGACGACATGGTTCTACATAGTGCCACTGCACTCCAAC	TACGGTAGCAGAGACTTGGTCTCCAAGGAGGAGACAAAGCAG
FLT3-I11	gDNA	ACACTGACGACATGGTTCTACATGGACCTGGAAGAAGTGTTCA	TACGGTAGCAGAGACTTGGTCTGGATTTCTGCAGCGAGTTCT
FLT3-I15.02	gDNA	ACACTGACGACATGGTTCTACACTGGGTACCAGCAATTCTCC	TACGGTAGCAGAGACTTGGTCTTGGGGAATGAAGGAAAAACA
FLT3-I15.03	gDNA	ACACTGACGACATGGTTCTACATGGAGAGGGAAGTCCAGAGA	TACGGTAGCAGAGACTTGGTCTAACTGCCAAGCAGGCTAGAA
FLT3-I17	gDNA	ACACTGACGACATGGTTCTACATTTTAAGTGCAGGCCTCTGG	TACGGTAGCAGAGACTTGGTCTAGGATAGCCCCTCTGTTCCA
FLT3-I19.02	gDNA	ACACTGACGACATGGTTCTACAACTCCCAACTGGCAGTCAAG	TACGGTAGCAGAGACTTGGTCTTCAAGGGCCTACAGAGGAAA
FLT3-I20	gDNA	ACACTGACGACATGGTTCTACAGGCAGGCACCTAAAATACCA	TACGGTAGCAGAGACTTGGTCTTCCTTGTCCTCCCTGTGAAG
FLT3-I22	gDNA	ACACTGACGACATGGTTCTACAGCATTCCGGTTGATGCTAAC	TACGGTAGCAGAGACTTGGTCTCCAACACAGCAAGACCTCAC
FLT3-I23.01	gDNA	ACACTGACGACATGGTTCTACATCAAAAGGCCAGAGGAAATG	TACGGTAGCAGAGACTTGGTCTGATCATTCCACTGCACTCCA
FLT3-I23.02	gDNA	ACACTGACGACATGGTTCTACAACCTAGGGTTTGTTTCACAATTTC	TACGGTAGCAGAGACTTGGTCTAGGCTGGTCTTGCACTCCT
IDH1-E04.X	gDNA	ACACTGACGACATGGTTCTACATCAAGTTGAAACAAATGTGGAAATCA	TACGGTAGCAGAGACTTGGTCTCATACAAGTTGGAAATTTCTGGGC
IDH2-E04	gDNA	ACACTGACGACATGGTTCTACATTCTGGTTGAAAGATGGCGG	TACGGTAGCAGAGACTTGGTCTCAGGTCAGTGGATCCCCTC
NPM1-E11	gDNA	ACACTGACGACATGGTTCTACATGTCTATGAAGTGTTGTGGTTCC	TACGGTAGCAGAGACTTGGTCTACAGGCATTTTGGACAACACA
NRAS-E02	gDNA	ACACTGACGACATGGTTCTACATACTGTAGATGTGGCTCGCC	TACGGTAGCAGAGACTTGGTCTTGATCCGACAAGTGAGAGACA
NRAS-E03	gDNA	ACACTGACGACATGGTTCTACACCCTCCCTGCCCCCTTAC	TACGGTAGCAGAGACTTGGTCTAGATTCAGAACACAAAGATCATCCT
RUNX1-E01.01	gDNA	ACACTGACGACATGGTTCTACACTGCTTGCTGAAGATCCGC	TACGGTAGCAGAGACTTGGTCTGCCAGCACCTCCACCATG
RUNX1-E01.02	gDNA	ACACTGACGACATGGTTCTACACCCAGGCAAGATGAGCGA	TACGGTAGCAGAGACTTGGTCTCTCCGCCTGTCCTCCCA
RUNX1-E02.01	gDNA	ACACTGACGACATGGTTCTACATCCTCTGCAACCTAAAAAGAAATCA	TACGGTAGCAGAGACTTGGTCTTGCAACCTGGTTCTTCATGG
RUNX1-E02.02	gDNA	ACACTGACGACATGGTTCTACAGGTGTTTAGGTGGTGGCCC	TACGGTAGCAGAGACTTGGTCTATTTTGAAATGTGGGTTTGTTGCC
RUNX1-E03	gDNA	ACACTGACGACATGGTTCTACAACAGATATGTTCAGGCCACCA	TACGGTAGCAGAGACTTGGTCTGACATGGTCCCTGAGTATACCA
RUNX1-E04.01	gDNA	ACACTGACGACATGGTTCTACACTGATCTCTTCCCTCCCTCC	TACGGTAGCAGAGACTTGGTCTTGAGGCTGAGGGTTAAAGGC
RUNX1-E04.02	gDNA	ACACTGACGACATGGTTCTACAGAGCTTGTCCTTTTCCGAGC	TACGGTAGCAGAGACTTGGTCTAGTTGGTCTGGGAAGGTGTG
RUNX1-E05	gDNA	ACACTGACGACATGGTTCTACAAAATCCCACCCCACTTTACATATAA	TACGGTAGCAGAGACTTGGTCTGCTCAGCTGCAAAGAATGTG
RUNX1-E06.01	gDNA	ACACTGACGACATGGTTCTACAGGGCAGAGGGAAGAGCTG	TACGGTAGCAGAGACTTGGTCTGGCAGGTAGGTGTGGTAGC
RUNX1-E06.02	gDNA	ACACTGACGACATGGTTCTACACTGCCCTCCATCTCCGAC	TACGGTAGCAGAGACTTGGTCTCATGGAGAACTGGTAGGAGCC
RUNX1-E06.03	gDNA	ACACTGACGACATGGTTCTACAGTTCCAAGCCAGCTCGC	TACGGTAGCAGAGACTTGGTCTCTCCACACGGCCTCCTC
RUNX1-E06.04	gDNA	ACACTGACGACATGGTTCTACACTGCACCAACGCCTCCA	TACGGTAGCAGAGACTTGGTCTTCGCGAACAGGAGGCC
TET2-E03.01	gDNA	ACACTGACGACATGGTTCTACAATTCAACTAGAGGGCAGCCTTG	TACGGTAGCAGAGACTTGGTCTTCTGGCTTCCCTTCATACAGG
TET2-E03.02	gDNA	ACACTGACGACATGGTTCTACACCCACTGCCTGAGAGAGC	TACGGTAGCAGAGACTTGGTCTCCCGAAGTTACGTCTTTCTCC
TET2-E03.03	gDNA	ACACTGACGACATGGTTCTACATTCTCTCTCTGGGCTCCTTC	TACGGTAGCAGAGACTTGGTCTTCCCCTCCTGCTCATTCAG
TET2-E03.04	gDNA	ACACTGACGACATGGTTCTACATCAACACATAACTGCAGTGGG	TACGGTAGCAGAGACTTGGTCTTGTGAGATGTGGTTTTCTGCA
TET2-E03.05	gDNA	ACACTGACGACATGGTTCTACAGGTGAACTCCTGGAAAAAACACTG	TACGGTAGCAGAGACTTGGTCTGCATTATCAGCATCATCAGCATCA
TET2-E03.06	gDNA	ACACTGACGACATGGTTCTACATGCCTCCAAAGCCAGCTG	TACGGTAGCAGAGACTTGGTCTATACCGTTCAGAGCTGCCAC
TET2-E03.07	gDNA	ACACTGACGACATGGTTCTACAAAGCTAGCGTCTGGTGAAGA	TACGGTAGCAGAGACTTGGTCTACTCCACCATTCAGAGTGCT
TET2-E03.08	gDNA	ACACTGACGACATGGTTCTACACCTCCTCCTCTTCCACAGG	TACGGTAGCAGAGACTTGGTCTTCCCTGCAGTCTGTATGTCA
TET2-E03.09	gDNA	ACACTGACGACATGGTTCTACACCCTTCTCCGATGCTTTCTG	TACGGTAGCAGAGACTTGGTCTCAAGATCTCGTGTTTGCTCCT
TET2-E03.10	gDNA	ACACTGACGACATGGTTCTACACAGTGGAGAGCTACAGGACA	TACGGTAGCAGAGACTTGGTCTTGGAGGTCATTTGATTGGAGAGA
TET2-E03.11	gDNA	ACACTGACGACATGGTTCTACAGCCCCTCGTTTTCACCAAG	TACGGTAGCAGAGACTTGGTCTTGGAACTGGAGATGTTGGTCC
TET2-E03.12	gDNA	ACACTGACGACATGGTTCTACAGTCACAAATGTACCAAGTTGAAATGA	TACGGTAGCAGAGACTTGGTCTAGGTGTGAGTTTGAAAATGGCT
TET2-E03.13	gDNA	ACACTGACGACATGGTTCTACAAAAACTTATGTCCCCAGTGTTGA	TACGGTAGCAGAGACTTGGTCTTGGCCAAAGAATGATCCTTCTC
TET2-E03.14	gDNA	ACACTGACGACATGGTTCTACAAGCAACAGCAGCAAAAATTACAA	TACGGTAGCAGAGACTTGGTCTGCATGCACTTGATTTCATGGTC
TET2-E03.15	gDNA	ACACTGACGACATGGTTCTACACAAATGGGACTGGAGGAAGTAC	TACGGTAGCAGAGACTTGGTCTTGCTCCTGTTCTTGAAAGCAC
TET2-E03.16	gDNA	ACACTGACGACATGGTTCTACAGCAGGAAACAAGACCCAAAAC	TACGGTAGCAGAGACTTGGTCTTGAGTGTGACTTCCTCCCTG
TET2-E03.17	gDNA	ACACTGACGACATGGTTCTACAACAAGCTTCAGTTCTACAGGGA	TACGGTAGCAGAGACTTGGTCTCCTGTGCATCTGACTATGGC
TET2-E03.18	gDNA	ACACTGACGACATGGTTCTACAACAGAAGCAAGAACAGCAGC	TACGGTAGCAGAGACTTGGTCTGACTTGGCGTGAAACTGCTT
TET2-E03.19	gDNA	ACACTGACGACATGGTTCTACAGAATCCACCTGCAAGCTGTG	TACGGTAGCAGAGACTTGGTCTTGTTCTTTTGGTTGGTGTCTTTTC
TET2-E03.20	gDNA	ACACTGACGACATGGTTCTACACTAGACAAACCACTGCTGCA	TACGGTAGCAGAGACTTGGTCTTGCTAATTCTGGATAAACGCCA
TET2-E04	gDNA	ACACTGACGACATGGTTCTACAAATGTGTAGTTGGGGGTTAAGC	TACGGTAGCAGAGACTTGGTCTTGTGTGAAGGCTGGAAAAATTCT
TET2-E05	gDNA	ACACTGACGACATGGTTCTACACCGTTCATTTCTCAGGATGTGG	TACGGTAGCAGAGACTTGGTCTTGCCCAAGATTTAAGACCAAAGG
TET2-E06.01	gDNA	ACACTGACGACATGGTTCTACAGCCCTTATCTGCTGCAAGTG	TACGGTAGCAGAGACTTGGTCTGCCGTATTTCCTCAGCGTCT
TET2-E06.02	gDNA	ACACTGACGACATGGTTCTACAACTGTGTTTGGTGCGGGA	TACGGTAGCAGAGACTTGGTCTGCTGAACTCTCTTCCTTTCAACC
TET2-E07.02	gDNA	ACACTGACGACATGGTTCTACATCAGCTGCACAGCCTATATAATG	TACGGTAGCAGAGACTTGGTCTTAATGAATTCTCTATGAAAAATAAAGC
TET2-E08	gDNA	ACACTGACGACATGGTTCTACAAGGCACCATATATTGTGTTTGGG	TACGGTAGCAGAGACTTGGTCTTAAGTTGTTACAATTGCTGCCA
TET2-E09	gDNA	ACACTGACGACATGGTTCTACAAACTACTTTCGCATTCACACACA	TACGGTAGCAGAGACTTGGTCTCAGCTGCTAAGCTGTCCTCA
TET2-E10.01	gDNA	ACACTGACGACATGGTTCTACACACGTTTTCTTTGGGACCTGTA	TACGGTAGCAGAGACTTGGTCTCTTGACTGGCTCTGCTAACATC
TET2-E10.02	gDNA	ACACTGACGACATGGTTCTACAGTGGATGAGTTTGGGAGTGTG	TACGGTAGCAGAGACTTGGTCTCTGGCTTGCGTTTTCAGTTTG

TET2-E10.03	gDNA	ACACTGACGACATGGTTCTACAGTTAGCAGAGCCAGTCAAGAC	TACGGTAGCAGAGACTTGGTCTTACAAGTTGATGGGGGCAAAAC
TET2-E11.01	gDNA	ACACTGACGACATGGTTCTACATCAACATCAAAGATACCTGTTTCTGT	TACGGTAGCAGAGACTTGGTCTGATAAGGACTAACTGGATTGGGC
TET2-E11.02	gDNA	ACACTGACGACATGGTTCTACATGTCAACTCTTATTCTGCTTCTGG	TACGGTAGCAGAGACTTGGTCTCCACTGATAGGTTTCCATTGCA
TET2-E11.03	gDNA	ACACTGACGACATGGTTCTACATACCCTGGGCTTTTGAATCAGA	TACGGTAGCAGAGACTTGGTCTTCCCTGCATATTTTGGTTTCCA
TET2-E11.04	gDNA	ACACTGACGACATGGTTCTACACTCAGTCTACCACCCATCCAT	TACGGTAGCAGAGACTTGGTCTTGTTCACCATTTTTATAGTCCATGTT
TET2-E11.05	gDNA	ACACTGACGACATGGTTCTACACACTCATGAGATGGATGGCC	TACGGTAGCAGAGACTTGGTCTGCAGTTCTATCATGGTTAAGAGCT
TET2-E11.06	gDNA	ACACTGACGACATGGTTCTACAAATGACATGCTTTCCCACACAG	TACGGTAGCAGAGACTTGGTCTCATGAGTTGGAGCCACGG
TET2-E11.07	gDNA	ACACTGACGACATGGTTCTACAGGTCAGACAGCGAGCAGA	TACGGTAGCAGAGACTTGGTCTTCTTTCTCACGGGCTTTTTCAG
TET2-E11.08	gDNA	ACACTGACGACATGGTTCTACAGCATGCCACAACCCCTTTAA	TACGGTAGCAGAGACTTGGTCTGCAGGTAAGTGGGCTCTGAA
TET2-E11.09	gDNA	ACACTGACGACATGGTTCTACATATGGCCCAGACTATGTGCC	TACGGTAGCAGAGACTTGGTCTGGTTGGTTGTGGTCTTTTCAAG
TP53-E04.01	gDNA	ACACTGACGACATGGTTCTACATCCTCTGACTGCTCTTTTCAC	TACGGTAGCAGAGACTTGGTCTGGTAGGTTTTCTGGGAAGGGA
TP53-E04.02	gDNA	ACACTGACGACATGGTTCTACAGTCCAGATGAAGCTCCCAGA	TACGGTAGCAGAGACTTGGTCTGATACGGCCAGGCATTGAAG
TP53-E05	gDNA	ACACTGACGACATGGTTCTACATGTGCCCTGACTTTCAACTC	TACGGTAGCAGAGACTTGGTCTCCCTGTCGTCTCTCCAGC
TP53-E06	gDNA	ACACTGACGACATGGTTCTACAGGAGAGACGACAGGGCTG	TACGGTAGCAGAGACTTGGTCTGGAGAAAGCCCCCCTACTG
TP53-E07	gDNA	ACACTGACGACATGGTTCTACAGCACTGGCCTCATCTTGGG	TACGGTAGCAGAGACTTGGTCTAGAGGTGGATGGGTAGTAGTATG
TP53-E08	gDNA	ACACTGACGACATGGTTCTACAGGGTGGTTGGGAGTAGATGG	TACGGTAGCAGAGACTTGGTCTCTCCACCGCTTCTTGTCCT
TP53-E09	gDNA	ACACTGACGACATGGTTCTACACAGTTATGCCTCAGATTCACTTTT	TACGGTAGCAGAGACTTGGTCTCCAGGAGCCATTGTCTTTGA
TP53-E10	gDNA	ACACTGACGACATGGTTCTACAACTTGAACCATCTTTTAACTCAGGT	TACGGTAGCAGAGACTTGGTCTTGAAGGCAGGATGAGAATGGA
ASXL1-E13.02	gDNA	CTTACCAGATATGCCCCCG	ACCACCATCACCACTGC
ASXL1-E13.03	gDNA	ACTGCCATCGGAGGGG	TCTCCTAGCTCTGGACATGG
RUNX1-E01.01	gDNA	GTAATAAAGGCCCCTGAACG	ACGCCTCGCTCATCTTG
RUNX1-E01.02	gDNA	TTGTGATGCGTATCCCCG	CAGAGGAAGTTGGGGCTG
RUNX1-E01.03	gDNA	CAAGCTGAGGAGCGGC	CCTGTCCTCCCACCACC
RUNX1-E02.01	gDNA	ACATCCCTGATGTCTGCATT	GTACCTCTTCCACTTCGACC
RUNX1-E02.02	gDNA	ACTGGTGTTTAGGTGGTGG	GTTTGTTGCCATGAAACGTG
RUNX1-E03	gDNA	CCACCAACCTCATTCTGTTTT	GTCCCTGAGTATACCAGCCT
RUNX1-E04.01	gDNA	TCTGATCTCTTCCCTCCCT	GCTGACCCTCATGGCTG
RUNX1-E04.02	gDNA	GAGCTTGTCCTTTTCCGAGC	AGGAGAGGCGGGCAGTG
RUNX1-E05.01	gDNA	AATCCCACCCCACTTTACAT	GAAAGTTCTGCAGAGAGGGT
RUNX1-E05.02	gDNA	TGGTCCTACGATCAGTCCTA	CATGCACCTCTAGTCTCCTG
RUNX1-E06.01	gDNA	CTACTCACTTCCGCTCCG	GCAGGTAGGTGTGGTAGC
RUNX1-E06.02	gDNA	CTTCACCTACTCCCCGAC	CCACCATGGAGAACTGGTAG
RUNX1-E06.03	gDNA	CCCGTTCCAAGCCAGCTC	CTCCACCACGTCGCTCTG
RUNX1-E06.04	gDNA	CCGGCTCCTACCAGTTCTC	CTCCACACGGCCTCCTC
RUNX1-E06.05	gDNA	CAGCCTCCCGAACCAGAG	GTCGCGAACAGGAGGCC
ASXL1-E13.01	gDNA	TCCTAGTTTTGCTTTACAGTCCC	GCAGTGGTGACCTCTCGC
ASXL1-E13.02.A	gDNA	ACTTACCAGATATGCCCCCG	CCTCACCACCATCACCACTG
ASXL1-E13.03	gDNA	ACCCTCGCAGACATTAAAGC	CGGCAGTAGTTGTGTTCGC
ASXL1-E13.04	gDNA	AGTGGTGATGGTGGTGAGG	GGGGAGTTGGGAGGCATC
ASXL1-E13.05	gDNA	TCTCTGAGAAAGGAGGAAAGCT	GTGGGTCCTTGCTCCTCAT
ASXL1-E13.06	gDNA	GAACTGAATGTGAGTCTGGCA	ATGAGCCACCAAGCCCTAAT
ASXL1-E13.07	gDNA	CCAGTTCCACACCTGAATCCT	CCAACCTGGGGCTCAACA
ASXL1-E13.08	gDNA	TCGAATGATGAGGTAGTGAAACAG	TTCAGCTTTTCAATGTCCACCT
ASXL1-E13.09	gDNA	GACTGTGCCATCTCGAGGAG	CATTCACCTTGGACAGTGGG
ASXL1-E13.10	gDNA	AGGGATCTTCGGTGGACAAG	CATCACAAGTGGGTTAGTGGC
ASXL1-E13.11	gDNA	GATTCCCTACTGCTGGCCA	CCATCAACCATGCCACTGTT
ASXL1-E13.12	gDNA	CATGGCTCGCTACGCATG	CCATTTCTTCCAGTTTCCTAGAGG
ASXL1-E13.13	gDNA	GGCAGTCCCAAGTTTTGACT	TCTGCTCTGGACCAAAGGAG
ASXL1-E13.14	gDNA	GCCCAGGAGATCTTACTACCT	AGACATGGAGTTTGTGCTTGG
ASXL1-E13.15	gDNA	TTTTGGCTCTGGGAATGTGG	ATGCCCAGTAGCTTTCCTGT
ASXL1-E13.16	gDNA	TTTGTTGGCAGCGTCAAGAA	AGCTTCCCATAAAATGCCTGC
ASXL1-E13.17	gDNA	GGGAAGGGGCTCAGTGAG	TCCACCGTGCTGCTGTCA
ASXL1-E13.18	gDNA	GCAAACTTTGGTGCGAGCC	AGATCAATCATTATCAAAATACACACA
CEBPA-E01.01	gDNA	CCATGCCGGGAGAACTCTA	GTCGTTGAAGGCGGCC
CEBPA-E01.02	gDNA	GCACGAGACGTCCATCGA	GTCCAGGTAGCCGGCG
CEBPA-E01.03	gDNA	GGCGACTTTGACTACCCGG	CGGCTGGTAAGGGAAGAGG
CEBPA-E01.04	gDNA	GCCGCTGGTGATCAAGC	GACCGGGCTGCAGGTG
CEBPA-E01.05	gDNA	GCAGTTCCAGATCGCGC	CGGTACTCGTTGCTGTTCTT
CEBPA-E01.06	gDNA	CTCAAGGGGCTGGGCG	CAGCTGGCGGAAGATGC
CEBPA-E01.07	gDNA	GTGGACAAGAACAGCAACGA	CGCAGCGTGTCCAGTTC
CEBPA-E01.08	gDNA	GCAAGCGGGTGGAACA	AGAGGAAGGGAGGGGACAC
DNMT3A-E07.01	gDNA	TTCCTGGAGAGGTCAAGGTG	TTTGGTGGCATTCTTGTCCC
DNMT3A-E07.02	gDNA	GGAAGAAAACCAGGGGCCC	GAGGAGCTGGCAGTGGAAG
DNMT3A-E08	gDNA	CTCGTGACCACTGTGTAATGAT	GCCCTGGGATCAAGAACCTT
DNMT3A-E09	gDNA	GGTGGGTGCTTGCAAGTG	ACCTGCACTCCAACTTCCAG
DNMT3A-E10	gDNA	CTGACAACCCCAACCCTG	GCCCTGGTGTGGATCTGC
DNMT3A-E11	gDNA	CTGTAACTGACCTTGGCACCT	TCGGATGCAGGCCTCCT
DNMT3A-E12	gDNA	AGAAGCCCAAGGTCAAGGAG	CTCTGCTACTCTGCCCCATG
DNMT3A-E13	gDNA	TTTTGTGGCTGGTCTGGTGC	AGAAGCGGTGGACACAGTC
DNMT3A-E14	gDNA	GAGCCTCCCTTCGTCCTG	GGCCCAGCTAAGGAGACC
DNMT3A-E15.01	gDNA	CCAGGGCTGAGAGTCTCCT	GCCCAGCACTCACAAATTCC
DNMT3A-E15.02	gDNA	GTGGAGTGTGTGGACCTCTT	CCCCACAACCAAGGCTCA
DNMT3A-E16	gDNA	CTGGGCCTGCATCTGACC	GTTTTGCCAGAGTTGCCCAC
DNMT3A-E17	gDNA	AGATGGCTCCAAGTAACGGTG	GTGCAGGGAGGGGAAGAC
DNMT3A-E18	gDNA	CTTCCTGTCTGCCTCTGTCC	CATCGGGAATAGCTGTCCCA
DNMT3A-E19	gDNA	GAGCCACACCACTGTCCTAT	CCATTAGTGAGCTGGCCAAAC
DNMT3A-E20	gDNA	CCTTGGCTCATCTTCAAACCG	AGCAGCTAGTCATTCAGCAGA
DNMT3A-E21	gDNA	ACCCTGTGAACTAGTGGCTG	CATCCTGCCCTTCCTTCTCC
DNMT3A-E22	gDNA	GCATATTTGGTAGACGCATGACC	CACAGCAATCAGAACAGCCAC
DNMT3A-E23.01.A	gDNA	GGTCCTGCTGTGTGGTTAGA	GTGTCGCTACCTCAGTTTGC
DNMT3A-E23.02	gDNA	CACTCACCCTGCCCTCTC	TGTTTAACTTTGTGTCGCTACCT
FLT3-E16.01	gDNA	TGCCTGGCTTCTCTCATAATT	GTGATTTTCGTGGAAGTGGGT
FLT3-E16.02.A	gDNA	CTGAGCTGAGAAAAATTTTATTTAATGAC	TGTCATCAAGCTACAGTCTTTTTGA
FLT3-E20.01.A	gDNA	GAAGAAAGATTGCACTCCAGGA	GATAACGACACAACACAAAATAGC
FLT3-E20.02	gDNA	GCACTCCAGGATAATACACATCA	ACAGTGAGTGCAGTTGTTTACC
IDH1-E04.01.A	gDNA	GAAGCTATAAAGAAGCATAATGTTGGC	TGCCAACATGACTTACTTGATCC

IDH1-E04.02	gDNA	TCAAGTTGAAACAAATGTGGAAATCA	CATACAAGTTGGAAATTTCTGGGC
IDH2-E04.01	gDNA	TTCTGGTTGAAAGATGGCGG	CAGGTCAGTGGATCCCCTC
IDH2-E04.02.A	gDNA	TGTCCTCACAGAGTTCAAGCT	AGGATGGCTAGGCGAGGAG
NPM1-E11.01.A	gDNA	TGATGTCTATGAAGTGTTGTGGT	GGCAATAGAACCTGGACAACA
NPM1-E11.02	gDNA	TGTCTATGAAGTGTTGTGGTTCC	ACAGGCATTTTGGACAACACA
NPM1-FE10-11	cDNA	TCCCAAAGTGGAAGCC	GGAAAGTTCTCACTCTGC
NRAS-E02.01.A	gDNA	GGAAGGTCACACTAGGGTTTTC	GATCAGGTCAGCGGGCTAC
NRAS-E02.02	gDNA	TACTGTAGATGTGGCTCGCC	TGATCCGACAAGTGAGAGACA
NRAS-E03.01.A	gDNA	AGTTAGATGCTTATTTAACCTTGGCA	AGAGGTTAATATCCGCAAATGACT
NRAS-E03.02	gDNA	CCCTCCCTGCCCCCTTAC	AGATTCAGAACACAAAGATCATCCT
RUNX1-E01.01	gDNA	CTGCTTGCTGAAGATCCGC	GCCAGCACCTCCACCATG
RUNX1-E01.02.A	gDNA	TATCCCCGTAGATGCCAGC	CAGGGTCTTGTTGCAGCG
RUNX1-E01.03	gDNA	CCCAGGCAAGATGAGCGA	CTCCGCCTGTCCTCCCA
RUNX1-E02.01	gDNA	TCCTCTGCAACCTAAAAAGAAATCA	TGCAACCTGGTTCTTCATGG
RUNX1-E02.02	gDNA	GGTGTTTAGGTGGTGGCCC	ATTTTGAAATGTGGGTTTGTTGCC
RUNX1-E03	gDNA	ACAGATATGTTCAGGCCACCA	GACATGGTCCCTGAGTATACCA
RUNX1-E04.01	gDNA	CTGATCTCTTCCCTCCCTCC	TGAGGCTGAGGGTTAAAGGC
RUNX1-E04.02	gDNA	GAGCTTGTCCTTTTCCGAGC	AGTTGGTCTGGGAAGGTGTG
RUNX1-E05	gDNA	AAATCCCACCCCACTTTACATATAA	GCTCAGCTGCAAAGAATGTG
RUNX1-E06.01	gDNA	GGGCAGAGGGAAGAGCTG	GGCAGGTAGGTGTGGTAGC
RUNX1-E06.02	gDNA	CTGCCCTCCATCTCCGAC	CATGGAGAACTGGTAGGAGCC
RUNX1-E06.03	gDNA	GTTCCAAGCCAGCTCGC	CTCCACACGGCCTCCTC
RUNX1-E06.04	gDNA	CTGCACCAACGCCTCCA	TCGCGAACAGGAGGCC
TET2-E03.01	gDNA	ATTCAACTAGAGGGCAGCCTTG	TCTGGCTTCCCTTCATACAGG
TET2-E03.02	gDNA	CCCACTGCCTGAGAGAGC	CCCGAAGTTACGTCTTTCTCC
TET2-E03.03	gDNA	TTCTCTCTCTGGGCTCCTTC	TCCCCTCCTGCTCATTCAG
TET2-E03.04	gDNA	TCAACACATAACTGCAGTGGG	TGTGAGATGTGGTTTTCTGCA
TET2-E03.05	gDNA	GGTGAACTCCTGGAAAAAACACTG	GCATTATCAGCATCATCAGCATCA
TET2-E03.06	gDNA	TGCCTCCAAAGCCAGCTG	ATACCGTTCAGAGCTGCCAC
TET2-E03.07	gDNA	AAGCTAGCGTCTGGTGAAGA	ACTCCACCATTCAGAGTGCT
TET2-E03.08	gDNA	CCTCCTCCTCTTCCACAGG	TCCCTGCAGTCTGTATGTCA
TET2-E03.09	gDNA	CCCTTCTCCGATGCTTTCTG	CAAGATCTCGTGTTTGCTCCT
TET2-E03.10	gDNA	CAGTGGAGAGCTACAGGACA	TGGAGGTCATTTGATTGGAGAGA
TET2-E03.11	gDNA	GCCCCTCGTTTTCACCAAG	TGGAACTGGAGATGTTGGTCC
TET2-E03.12	gDNA	GTCACAAATGTACCAAGTTGAAATGA	AGGTGTGAGTTTGAAAATGGCT
TET2-E03.13	gDNA	AAAACTTATGTCCCCAGTGTTGA	TGGCCAAAGAATGATCCTTCTC
TET2-E03.14	gDNA	AGCAACAGCAGCAAAAATTACAA	GCATGCACTTGATTTCATGGTC
TET2-E03.15	gDNA	CAAATGGGACTGGAGGAAGTAC	TGCTCCTGTTCTTGAAAGCAC
TET2-E03.16	gDNA	GCAGGAAACAAGACCCAAAAC	TGAGTGTGACTTCCTCCCTG
TET2-E03.17	gDNA	ACAAGCTTCAGTTCTACAGGGA	CCTGTGCATCTGACTATGGC
TET2-E03.18	gDNA	ACAGAAGCAAGAACAGCAGC	GACTTGGCGTGAAACTGCTT
TET2-E03.19	gDNA	GAATCCACCTGCAAGCTGTG	TGTTCTTTTGGTTGGTGTCTTTTC
TET2-E03.20	gDNA	CTAGACAAACCACTGCTGCA	TGCTAATTCTGGATAAACGCCA
TET2-E04	gDNA	AATGTGTAGTTGGGGGTTAAGC	TGTGTGAAGGCTGGAAAAATTCT
TET2-E05	gDNA	CCGTTCATTTCTCAGGATGTGG	TGCCCAAGATTTAAGACCAAAGG
TET2-E06.01	gDNA	GCCCTTATCTGCTGCAAGTG	GCCGTATTTCCTCAGCGTCT
TET2-E06.02	gDNA	ACTGTGTTTGGTGCGGGA	GCTGAACTCTCTTCCTTTCAACC
TET2-E07.01.A	gDNA	GACACCTATAATATCAGCTGCACA	ACATCAGGAAGTAAACAAACCTCT
TET2-E07.02	gDNA	TCAGCTGCACAGCCTATATAATG	TAATGAATTCTCTATGAAAAATAAAGC
TET2-E07.03.A	gDNA	TTCAGGAGAACTTGCGCC	AAGATATGTCATATTGTTCACTTCATCT
TET2-E08	gDNA	AGGCACCATATATTGTGTTTGGG	TAAGTTGTTACAATTGCTGCCA
TET2-E09	gDNA	AACTACTTTCGCATTCACACACA	CAGCTGCTAAGCTGTCCTCA
TET2-E10.01	gDNA	CACGTTTTCTTTGGGACCTGTA	CTTGACTGGCTCTGCTAACATC
TET2-E10.02	gDNA	GTGGATGAGTTTGGGAGTGTG	CTGGCTTGCGTTTTCAGTTTG
TET2-E10.03	gDNA	GTTAGCAGAGCCAGTCAAGAC	TACAAGTTGATGGGGGCAAAAC
TET2-E11.01	gDNA	TCAACATCAAAGATACCTGTTTCTGT	GATAAGGACTAACTGGATTGGGC
TET2-E11.02	gDNA	TGTCAACTCTTATTCTGCTTCTGG	CCACTGATAGGTTTCCATTGCA
TET2-E11.03	gDNA	TACCCTGGGCTTTTGAATCAGA	TCCCTGCATATTTTGGTTTCCA
TET2-E11.04	gDNA	CTCAGTCTACCACCCATCCAT	TGTTCACCATTTTTATAGTCCATGTT
TET2-E11.05	gDNA	CACTCATGAGATGGATGGCC	GCAGTTCTATCATGGTTAAGAGCT
TET2-E11.06	gDNA	AATGACATGCTTTCCCACACAG	CATGAGTTGGAGCCACGG
TET2-E11.07	gDNA	GGTCAGACAGCGAGCAGA	TCTTTCTCACGGGCTTTTTCAG
TET2-E11.08	gDNA	GCATGCCACAACCCCTTTAA	GCAGGTAAGTGGGCTCTGAA
TET2-E11.09	gDNA	TATGGCCCAGACTATGTGCC	GGTTGGTTGTGGTCTTTTCAAG
TP53-E04.01	gDNA	TCCTCTGACTGCTCTTTTCAC	GGTAGGTTTTCTGGGAAGGGA
TP53-E04.02	gDNA	GTCCAGATGAAGCTCCCAGA	GATACGGCCAGGCATTGAAG
TP53-E05	gDNA	TGTGCCCTGACTTTCAACTC	CCCTGTCGTCTCTCCAGC
TP53-E06	gDNA	GGAGAGACGACAGGGCTG	GGAGAAAGCCCCCCTACTG
TP53-E07	gDNA	GCACTGGCCTCATCTTGGG	AGAGGTGGATGGGTAGTAGTATG
TP53-E08	gDNA	GGGTGGTTGGGAGTAGATGG	CTCCACCGCTTCTTGTCCT
TP53-E09	gDNA	CAGTTATGCCTCAGATTCACTTTT	CCAGGAGCCATTGTCTTTGA
TP53-E10	gDNA	ACTTGAACCATCTTTTAACTCAGGT	TGAAGGCAGGATGAGAATGGA
TP53-E11	gDNA	AGGGAAAAGGGGCACAGAC	TGCAAGCAAGGGTTCAAAGA
WT1-E07	gDNA	AGCCTCCCTTCCTCTTACTCT	CCTTAGCAGTGTGAGAGCCT
WT1-E09	gDNA	CCAGCTGCCGGAAGTCAG	CTTTTCCAATCCCTCTCATCACA
ASXL1-E13.01	gDNA	CTAGGTCAGATCACCCAGTC	CCTCTCTATGGCAGTGGTG
ASXL1-E13.02	gDNA	CTTACCAGATATGCCCCCG	ACCACCATCACCACTGC
ASXL1-E13.03	gDNA	ACTGCCATCGGAGGGG	TCTCCTAGCTCTGGACATGG
ASXL1-E13.04	gDNA	ACGTCAGATCTACAGCGAAC	ATGGCTGGTCCCCAGT
ASXL1-E13.05	gDNA	GAGGAAAGCTGCCTACTACA	GAAGTGGTGCCAGACTCA
ASXL1-E13.06	gDNA	TGTCCTCCCAAACCTCAGTA	CTTGGCCAGTTCCTTTCTC
ASXL1-E13.07	gDNA	TGGCACCACTTCCTGG	GAGCCACCAAGCCCTAA
ASXL1-E13.08	gDNA	AGAAAGGAACTGGCCAAGC	GGGATGGGTATCCAATGCAA
ASXL1-E13.09	gDNA	AATTAGGGCTTGGTGGCT	TGTGTTCTCTGGATTCTGGTT
ASXL1-E13.10	gDNA	CCTCCTATGAGGGAAAGTGAT	GTCAAATCCCCAGGCAATG
ASXL1-E13.11	gDNA	GGTAGTGAAACAGCCCAAAC	TAACTGCCATTGCTGTCACT

ASXL1-E13.12	gDNA	GAGAAAGCTGCTCCCACC	GACCTTCAAAGTCAGAGGCT
ASXL1-E13.13	gDNA	TGTCAACAGGTGGACATTGA	TGTAACCAGACGCATGTCA
ASXL1-E13.14	gDNA	TGCAGTGACAAAGGGATCTT	ACTGAGGACCCAGGCAT
ASXL1-E13.15	gDNA	CTGGGTGTCTCGAGTATGTG	GTAAGTGGTACTTGTGGGGA
ASXL1-E13.16	gDNA	ACTTGTGATGCAGTTGCTG	TGGCAGAAGAGGCTCC
ASXL1-E13.17	gDNA	GCATGGGATCTTTACATGGT	GAGGGAGTCAAAACTTGGGA
ASXL1-E13.18	gDNA	AAAACAGTGGCATGGTTGAT	ATTTGTCACTGGATGGAGGG
ASXL1-E13.19	gDNA	TCCTGGAGCACCCCAA	TTTCCTGGAGCAGCATTTGA
ASXL1-E13.20	gDNA	AAAGAATTGCAAGGCAGTCC	GAGATGAGAAACGAGGTGTT
ASXL1-E13.21	gDNA	AGAGCAGTTCTCTTCCTTTAGT	CTTCTTCCCTTGGCCTGTAA
ASXL1-E13.22	gDNA	AAATGCTGCTCCAGGAAAGA	CTGGGAAAAACTGGAGGGAT
ASXL1-E13.23	gDNA	TTACAGGCCAAGGGAAGAAG	CCAACAAAGGCATGTGGC
ASXL1-E13.24	gDNA	AAGCACAAACTCCATGTCTG	AATTTCCAGAAGGGCAAGTC
ASXL1-E13.25	gDNA	CTGGGCTCCAAAGCCAC	GAGTTGGGAGGGGAGAGA
ASXL1-E13.26	gDNA	GGACTTGCCCTTCTGGAAAT	CCTTTGGGAAAGGTGGGAG
ASXL1-E13.27	gDNA	CAAGCAGGCATTTTATGGGA	CGCACACTGGAGCGAG
ASXL1-E13.28	gDNA	CTAGCTCTCCCACCTTTCC	CAGAGCTTTGAGGGTCCAAT
ASXL1-E13.29	gDNA	CCACAGTGCATCACTTTCC	AGATCAATCATTATCAAAATACACACA
DNMT3A-E07.01	gDNA	AAACATGGTCCCCTTGAGTG	TAGCCACAGTGGGGGATG
DNMT3A-E07.02	gDNA	TCAGAAGGTGGAGGAGGC	GAGCTGGCAGTGGAAGG
DNMT3A-E08.01	gDNA	CGTGACCACTGTGTAATGAT	CAAACCCCACAACTTACCAC
DNMT3A-E08.02	gDNA	AATGATTTCTGCTCCTTGGG	ACCCCAGGCTACTGCC
DNMT3A-E09	gDNA	GTAAGCCTCGGCAAACAAG	ACCTGCACTCCAACTTCC
DNMT3A-E10.01	gDNA	CTACTCAGAGTCTGGCCTTG	CCATTCAATCATGGGCTTGT
DNMT3A-E10.02	gDNA	AAGCTGTTCCCGGTGTG	CCTCCCTAAGCATGGCTTTC
DNMT3A-E11.01	gDNA	AGTTTCCTGTCAGCCTGTAA	CAGGCCAACTACCTCTTGT
DNMT3A-E11.02	gDNA	ACAAAGAAGTGTACACGGAC	AGCTGGCGTCAGAGGA
DNMT3A-E12	gDNA	CACAAGAGGTAGTTGGCCT	CACACTAGGAGTGCCAGAG
DNMT3A-E13	gDNA	GGTGGTACTCACCCCATC	CCTCAACGGCACCTCTC
DNMT3A-E14.01	gDNA	GGTCATGTCTTCAGGGCTTA	ATGGTGCAGTAGGACTGGTA
DNMT3A-E14.02	gDNA	GTGGTTTCTGACCCTTCCC	GCTAAGGAGACCACTGGAG
DNMT3A-E15.01	gDNA	CAGGGCTGAGAGTCTCCT	CGTAGGTACCCTTGTGCC
DNMT3A-E15.02	gDNA	GTCTCCTTCCAGGTGCTTTT	CCCAGCACTCACAAATTCC
DNMT3A-E15.03	gDNA	CCATTAAGGAAGACCCCTGG	TCCTAGACCCACACACCC
DNMT3A-E16	gDNA	GGCCTGCATCTGACCTG	CATCCTGGGACAAGGCG
DNMT3A-E17.01	gDNA	AGATGGCTCCAAGTAACGG	TACATGATCTTCCCCTGGTG
DNMT3A-E17.02	gDNA	AAGGACTTGGGCATTCAGG	GAACAAAATGAAAGGAGGCAA
DNMT3A-E18	gDNA	GCAGGCCCATCACGTT	GGAAGCCTATGTGCGGAA
DNMT3A-E19.01	gDNA	TATTCCCGATGACCCTGTCT	GCCACCACATTCTCAAAGAG
DNMT3A-E19.02	gDNA	TCTCAGAGGGCACTGGC	AAGCAGCAGTCCAAGGTAG
DNMT3A-E20	gDNA	CCTCCTTGGCTCATCTTCAA	CAGTCCCAGCCCACAG
DNMT3A-E21	gDNA	CCAGGTTTCTGTTGTTACAGTC	CCTTCCTTCTCCCTGCC
DNMT3A-E22.01	gDNA	GTGTTGGCTGGTGAATGAAT	CAGTGCACCATAAGATGTCC
DNMT3A-E22.02	gDNA	CGTGCTTATTCCTCTTTTCTCC	CAGCAAGCACAGCAATCAG
DNMT3A-E23.01	gDNA	CCAGCACTCACCCTGC	TTTGTGTCGCTACCTCAGTT
DNMT3A-E23.02	gDNA	CCCTCTCTGCCTTTTCTCC	TGTTTGTTTAACTTTGTGTCGC
FLT3-E20.01	gDNA	AAGAAAGATTGCACTCCAGG	AAATAGCAGCCTCACATTGC
FLT3-E20.02	gDNA	TGACAGTGTGTTCACAGAGA	AGTGAGTGCAGTTGTTTACC
IDH1-E04.X	gDNA	TGTGGAAATCACCAAATGGC	CATACAAGTTGGAAATTTCTGGG
IDH2-E04.01	gDNA	TTGTTGCTTGGGGTTCAAAT	ACTAGGCGTGGGATGTTTTT
IDH2-E04.02	gDNA	TGTGGAAAAGTCCCAATGGA	CAAGAGGATGGCTAGGCG
NPM1-E11	gDNA	TCTATGAAGTGTTGTGGTTCC	GCCAGATATCAACTGTTACAGAA
NRAS-E02	gDNA	ACCCTGATTACTGGTTTCCA	ATCCGACAAGTGAGAGACAG
NRAS-E03.01	gDNA	CCTCCCTGCCCCCTTAC	TCCGCAAATGACTTGCTATT
NRAS-E03.02	gDNA	AGATGGTGAAACCTGTTTGTT	CTCATTTCCCCATAAAGATTCAGA
TET2-E03.01	gDNA	TAGAGGGCAGCCTTGTGG	AAAGAGTGCCACTTGGTGT
TET2-E03.02	gDNA	TTCCTGATACCATCACCTCC	GCAAACACTTGGAATACCCT
TET2-E03.03	gDNA	ACCAAGTGGCACTCTTTCAA	AGCCTTTTGGTCTTGTTTCA
TET2-E03.04	gDNA	AAGCCAGAATAGTCGTGTGA	CTTGGCTTACCCCGAAGTTA
TET2-E03.05	gDNA	TGGAGGAATAAAACGCACAG	GGTGAAATCTTTAACTGCATTTTC
TET2-E03.06	gDNA	AGAAAGACGTAACTTCGGGG	TTTTCCCCTCCTGCTCATTC
TET2-E03.07	gDNA	GCCCAAGAAAATGCAGTTAAAG	TTCACCATGTGTGTGTTCCA
TET2-E03.08	gDNA	AGGAGGGGAAAAGTGCTAAT	TGAGATGTGGTTTTCTGCAC
TET2-E03.09	gDNA	TGCTACAGTTTCTGCCTCTT	CAGAGTTAGAGGTCTGTGCG
TET2-E03.10	gDNA	AGATTGTGTTTCCATTGCGG	CATCATCAGCATCACAGGC
TET2-E03.11	gDNA	CTGTGAGATCACTCACCCAT	TGGGCATATCTCAAAAACTGATT
TET2-E03.12	gDNA	GCTGATGATGCTGATAATGC	CTTGCAAATTGCTGCTGG
TET2-E03.13	gDNA	TAACATCCAGGGAACCACAA	GAGGGGGAGAAAGAAGCAAT
TET2-E03.14	gDNA	TCAAGCAAAGCTCAGTGTTC	TTGTGTTACTTTGGTTGGGG
TET2-E03.15	gDNA	TTCCTTTTCTGCCACTACCA	GCCTCAGGTTTACCCTCTAT
TET2-E03.16	gDNA	ACCCCAACCAAAGTAACACA	TGGAACAGTCATTGTCCCTG
TET2-E03.17	gDNA	ATAGAGGGTAAACCTGAGGC	TGCTTGAGGTGTTCTGACAT
TET2-E03.18	gDNA	CTACACATGTATGCAGCCCT	GTTGTCCTGTAGCTCTCCAC
TET2-E03.19	gDNA	CATACAGACTGCAGGGACAA	ACAAGATCTCGTGTTTGCTC
TET2-E03.20	gDNA	TGTCAGAACACCTCAAGCAT	TTAGATGGGATTCCGCTTGG
TET2-E03.21	gDNA	ATTCTGAAGGGTCGAGACAA	CCAGTGTATTGTTTGGAGGTC
TET2-E03.22	gDNA	TGGATTGAATTGAAGGCCCC	CAGCTGTGTTGTTTTCTGGG
TET2-E03.23	gDNA	TAATGAGGCATCACTGCCAT	ATGTTGGTCCACTGTACCTT
TET2-E03.24	gDNA	ACCTCCAAACAATACACTGGA	GGTCTGTTTTGGAGAAGTGC
TET2-E03.25	gDNA	ACACCCAGAAAACAACACAG	TCTGCTCTTTGTTGAAAATGAA
TET2-E03.26	gDNA	CCAGTTCCAAAAACCCTCAC	CCTGTTGATTCAAGTGCTGT
TET2-E03.27	gDNA	TACCAAAAGCTCATGTGCAG	GTGTTTGTGCTGCCTGTTTA
TET2-E03.28	gDNA	ACTTATGTCCCCAGTGTTGAA	TTGCTGGTTTTGAGGGAGAT
TET2-E03.29	gDNA	GCTTCAGAGACTGAGCCATT	GTCTGGCCAAAGAATGATCC
TET2-E03.30	gDNA	CACATCTCCCTCAAAACCAG	ACTTCCTCCAGTCCCATTTG
TET2-E03.31	gDNA	TTTGGCCAGACTAAAGTGGA	TCTCTGAAACTAGGTGTGTATTG

TET2-E03.32	gDNA	CAAATGGGACTGGAGGAAGT	TTGGGTCTTGTTTCCTGCAA
TET2-E03.33	gDNA	TGCAAAATACAGGTTTCTTGTT	TTATATCCCTGTAGAACTGAAGC
TET2-E03.34	gDNA	TGCAGGAAACAAGACCCAAA	TGACCAGACATATCTTGGTTTC
TET2-E03.35	gDNA	ATGTGATCCCAAAGCAAGAT	GGTCTGAGTGTGACTTCCTC
TET2-E03.36	gDNA	AGCTTCAGTTCTACAGGGAT	GTTTGCTGCTGTTCTTGCTT
TET2-E03.37	gDNA	TGACCAGGGAGGAAGTCA	GTTCCACCTTAATTGGCCTG
TET2-E03.38	gDNA	CCCCAGAAGGACACTCAAA	TGCTTAGTTACCTTTTTCCATGT
TET2-E03.39	gDNA	GAAGCAAGAACAGCAGCAAA	TTGCTGCACATTATCACAGC
TET2-E03.40	gDNA	ACAGGCCAATTAAGGTGGAA	TGTGATTTGAGAGTAAGAGCCT
TET2-E03.41	gDNA	AGCTGTGATAATGTGCAGC	AAGTTCTGCAGCAGTGGTTT
TET2-E03.42	gDNA	AAGGCTCTTACTCTCAAATCAC	AGTAATTTGGAAGGTGACTCTATAA
TET2-E03.43	gDNA	CTGCAGAACTTGATAGCCAC	CATCTGCAAGATGGGAAATCA
TET2-E03.44	gDNA	ACAGCTGCTTCTGTTCTCAA	TCCCATATCTGAAGATAAATTTGC
TET2-E04	gDNA	TGTGGATGTAGCCTTTATATTTAGTA	ATAAACGTTAATCTGCCCTGT
TET2-E05	gDNA	TGCCTCTTGAATTCATTTGC	CAAAATGCCCAAGATTTAAGACC
TET2-E06.01	gDNA	TGTTTTGTTTTGGTTGGGGT	TCCGAGTAGAGTTTGTCAGC
TET2-E06.02	gDNA	GAAGCAGCAGTGAAGAGAAG	ACTCTTCATTCAAGGCACAC
TET2-E06.03	gDNA	GGCTGCAGTGATTGTGATTC	TCAACCAAAGATTGGGCTTT
TET2-E07.01	gDNA	CAGCTGCACAGCCTATATAAT	TCATCCCCAAGCAGCTTAAA
TET2-E07.02	gDNA	GGTGCCTCCTTCTCTTTTGG	CACTTCATCTAAGCTAATGAATTCTC
TET2-E08	gDNA	GGAATAATCTAACTGATAGTCTCTTTT	AAGTTGTTACAATTGCTGCC
TET2-E09.01	gDNA	TTTTTAAAGTTCTAAATGGTCTAAATACTA	ATGTGCTGCCATTCTGC
TET2-E09.02	gDNA	TGAACACAGAGCACCAGAG	ACAGCTGCTAAGCTGTCC
TET2-E09.03	gDNA	CTGCATGTTTGGACTTCTGT	CCTCATTTGCCTTCAGCTAT
TET2-E10.01	gDNA	ACACACACGTTTTCTTTGGG	TCAGTACCTGAATGGCACC
TET2-E10.02	gDNA	TGGAGGAAAACCTGAGGATG	CTGCAGCTTTCTTGGCTT
TET2-E10.03	gDNA	GGTACTGAGTTCTTTTCGGC	GCTTGCGTTTTCAGTTTGTT
TET2-E10.04	gDNA	TGCCGACAAAGGAAACTAGA	ATACCACACAACACATTTATCTAC
TET2-E11.01	gDNA	CATCAACATCAAAGATACCTGTTT	GATTGGTGGATCCAGAAGCA
TET2-E11.02	gDNA	ACCCTGTCCACAGAACTTTT	ACTAACTGGATTGGGCCG
TET2-E11.03	gDNA	CCAGCAGCAGCAGAGAC	AGCTTGAGATGAGGTGGAAT
TET2-E11.04	gDNA	CAATCCATACATGAGACGGC	CACTGATAGGTTTCCATTGCAT
TET2-E11.05	gDNA	TCAAGCTGCAGGTTCATATTT	ATACAGATCCATCGGCTGAG
TET2-E11.06	gDNA	CTTACCCTGGGCTTTTGAATC	AAAGTGTATGGATGGGTGGT
TET2-E11.07	gDNA	TCCCCATATCTGGGTTCCTA	ACAACTGCTGAAACCATCTC
TET2-E11.08	gDNA	GTATCCAAGCCAAGACCCT	GAAGTGGCCATCCATCTCA
TET2-E11.09	gDNA	TACACTTTACCAGCCAAGGT	TTGGATTGCTCAGATTGGGT
TET2-E11.10	gDNA	CAGGGAGATGGTTTCAGC	TGATGTTCACCATTTTTATAGTCCA
TET2-E11.11	gDNA	ATGGATGGCCACTTCATGG	CATTAGCTGTGTGGGAAAGC
TET2-E11.12	gDNA	TGGGAGCCACCTCTAGATTA	CATGGTTAAGAGCTGGAAGC
TET2-E11.13	gDNA	GGTGAACATCATTCACCTTCT	GACCATTAGCATCACTTAATTTGT
TET2-E11.14	gDNA	CCCTGCATCTCCAAAACAAG	GACCAGACCTCATCGTTGT
TET2-E11.15	gDNA	AGAACTGCTTGTGTCCAAGG	ATTGACCCATGAGTTGGAGC
TET2-E11.16	gDNA	TGCTAATGGTCAGGAAAAGC	GCTCACGCTTTGCACAC
TET2-E11.17	gDNA	ACAACGATGAGGTCTGGTC	CTGGTAAAAGACGAGGGAGA
TET2-E11.18	gDNA	GATCCTGACATTGGGGGAG	GGGCTTTTTCAGCCATTTTG
TET2-E11.19	gDNA	CAAAGCGTGAGCTGCAT	CTGGGCCATACTTTTCACAC
TET2-E11.20	gDNA	CCCTCGTCTTTTACCAGCAT	GAGACTTGATGAAACGCAGG
TET2-E11.21	gDNA	GAGTGTGAAAAGTATGGCCC	TGACCCGAGTGAAGGCAT
TET2-E11.22	gDNA	CATGAAACTTCAGAGCCCAC	ACCAACAAAAGGGGGTGATA
TET2-E11.23	gDNA	GACCACAGACTCCACAGTAA	CTGGTGAGCTGAGGTTTTTC
TP53-E04.01	gDNA	ACAACGTTCTGGTAAGGACA	GAACCATTGTTCAATATCGTCC
TP53-E04.02	gDNA	CTGACTGCTCTTTTCACCCA	GAAGATGACAGGGGCCAG
TP53-E04.03	gDNA	TGGATGATTTGATGCTGTCC	TAGCTGCCCTGGTAGGTTT
TP53-E04.04	gDNA	CAGCAGCTCCTACACCG	CAAAAGAAATGCAGGGGGAT
TP53-E05.01	gDNA	TTGTGCCCTGACTTTCAACT	CCGTCATGTGCTGTGACT
TP53-E05.02	gDNA	TCTTCCTACAGTACTCCCCT	TGAGCAGCGCTCATGG
TP53-E05.03	gDNA	GCAGCTGTGGGTTGATTC	CCTGGGCAACCAGCC
TP53-E06.01	gDNA	CTGGTTGCCCAGGGTC	CCAGAGACCCCAGTTGC
TP53-E06.02	gDNA	CTCCTCAGCATCTTATCCGA	CAGCAGGAGAAAGCCCC
TP53-E07	gDNA	CTGGCCTCATCTTGGGC	CAGGCCAGTGTGCAGG
TP53-E08	gDNA	GGACAGGTAGGACCTGATTT	TCTCCTCCACCGCTTCT
TP53-E09	gDNA	TTATGCCTCAGATTCACTTTTATC	GCATTTTGAGTGTTAGACTGG
TP53-E10.01	gDNA	GGTACTTGAAGTGCAGTTTCT	CTGGCTCCTTCCCAGC
TP53-E10.02	gDNA	CTTCTCCCCCTCCTCTGTT	AAGGCAGGATGAGAATGGAA
TP53-E11	gDNA	TGATGTCATCTCTCCTCCCT	AGGCTGTCAGTGGGGAA
WT1-E02	gDNA	GCTGACACTGTGCTTCTCT	GAGGAGGATAGCACGGAAG
WT1-E03	gDNA	CTTGGGGCGCACTCG	GTCCCAAGGACCCAGAC
WT1-E04	gDNA	TGTGGTTATGTGTTTCTAACTCTA	ACTGTGGAAAGGCAATGGAA
WT1-E05	gDNA	GGGCTTTTCACTGGATTCTG	GCCTACGCCATTTGCTTTG
WT1-E06	gDNA	CATTTCCAAATGGCGACTGT	GGTAAGTAGGAAGAGGCAGT
WT1-E07.01	gDNA	CTCCCTCAAGACCTACGTG	TGCATCTGTAAGTGGGACAG
WT1-E07.02	gDNA	CCTCCCTTCCTCTTACTCTC	GAACCATGTTTGCCCAAGA
WT1-E08	gDNA	GAGAGGTTGCCTTTAATGAGA	CAGCTGCCAGCAATGAG
WT1-E09	gDNA	CACTGTGCCCACATTGTT	CTCTCATCACAATTTCATTCCAC
WT1-E10.01	gDNA	CAGGGACAGAATGATGGGAA	ATGTTGTGATGGCGGACTAA
WT1-E10.02	gDNA	GTCAACAGGTGAAAAGCCC	GGAGTGGAGAGTCAGACTTG
ASXL1-E13.01	gDNA	GGTCAGATCACCCAGTCAGT	GTCCAACTGTAGCCCTCTGT
ASXL1-E13.02	gDNA	GGACCTGCCTTCTCTGAGA	TCTGGATTCTGGTTTGGGCT
ASXL1-E13.03	gDNA	CCAAGGCTCTCGTTTCTAACAG	CATACTCGAGACACCCAGCT
ASXL1-E13.04	gDNA	ACATGCGTCTGGTTACAAGG	AGCACGGACTTCCTTCTGAT
ASXL1-E13.05	gDNA	ACAAATCCCATTACATCCTCTAGG	CTCGGGGTAATTTCCAGAAGG
ASXL1-E13.06	gDNA	TGCCGAGAACAGGAAAGCTA	TTTTGGGGGAAGGCAAGAGT
ASXL1-FE13.01	cDNA	AAGCCACAGCCCACTAAAGA	GTCCAACTGTAGCCCTCTGT
ASXL1-FE13.02	cDNA	GGACCTGCCTTCTCTGAGA	TCTGGATTCTGGTTTGGGCT
ASXL1-FE13.03	cDNA	CCAAGGCTCTCGTTTCTAACAG	CATACTCGAGACACCCAGCT

ASXL1-FE13.04	cDNA	ACATGCGTCTGGTTACAAGG	AGCACGGACTTCCTTCTGAT
ASXL1-FE13.05	cDNA	TGGATTCCAAAGAGCAGTTCTCTTC	CATGACAAAGGGCATCCCTTCCAA
ASXL1-FE13.06	cDNA	ACAGGAAAGCTACTGGGCATAGTC	CAAGAGTGCTCCTGCCTAAAGAGT
CEBPA-E01.01	gDNA	GCCATGCCGGGAGAACT	CCCGGGTAGTCAAAGTCG
CEBPA-E01.02	gDNA	CCTTCAACGACGAGTTCCTG	CGGCTGGTAAGGGAAGAGG
CEBPA-E01.03	gDNA	GAGGAGGATGAAGCCAAGC	CTCGTTGCTGTTCTTGTCCA
CEBPA-E01.04	gDNA	TGGCAGCGCGCTCAAG	CCAGGGCGGTCCCACA
CEBPA-FE01.01	cDNA	GCCATGCCGGGAGAACT	CCCGGGTAGTCAAAGTCG
CEBPA-FE01.02	cDNA	CCTTCAACGACGAGTTCCTG	CGGCTGGTAAGGGAAGAGG
CEBPA-FE01.03	cDNA	GAGGAGGATGAAGCCAAGC	CTCGTTGCTGTTCTTGTCCA
CEBPA-FE01.04	cDNA	TGGCAGCGCGCTCAAG	CCAGGGCGGTCCCACA
DNMT3A-E07	gDNA	TTCCTGGAGAGGTCAAGGTG	TGGAGAGAGGAGAGCAGGAC
DNMT3A-E08	gDNA	GCCTCGTGACCACTGTGTAA	ACCCACCACAGGCAGAGTAG
DNMT3A-E09	gDNA	CTCCTCTTTGCATCGGGTAA	ACCTGCACTCCAACTTCCAG
DNMT3A-E10	gDNA	TGTGCCACCCTCACTACTCA	TCCCTAAGCATGGCTTTCC
DNMT3A-E11E12	gDNA	GACCTTGGCACCTGCTTTC	CCACACTAGGAGTGCCAGAGTT
DNMT3A-E13	gDNA	GGTCACAGTGCCTCCCTTT	ACCCTGTACATGCCCAGAAG
DNMT3A-E14	gDNA	CACAGGCAGATGAGGTTTCC	CCCAGCTAAGGAGACCACTG
DNMT3A-E15	gDNA	CCCTAGCCATGCTCCAGAC	CCCACAACCAAGGCTCAG
DNMT3A-E16	gDNA	CAGGGTGTGTGGGTCTAGGA	TGCATACGTTTCCACTTCACA
DNMT3A-E17	gDNA	AAAGATAGGACTTGGGCCTACA	CTGCCTCCAGGTGCTGAG
DNMT3A-E18	gDNA	TGGTCCCGTTCTTGTTTAGG	CAAGGAGGAAGCCTATGTGC
DNMT3A-E19	gDNA	GACAGCTATTCCCGATGACC	GCTCCACAATGCAGATGAGA
DNMT3A-E20	gDNA	TGTGTGGCTCCTGAGAGAGA	CATGGCAGAGCAGCTAGTCA
DNMT3A-E21	gDNA	TGGTGGATTTGTGTCTTTGC	CATCCTGCCCTTCCTTCTC
DNMT3A-E22	gDNA	CTGCGAACTCTGCTCACTCA	AGCAAGCACAGCAATCAGAA
DNMT3A-E23	gDNA	CACTCACCCTGCCCTCTCT	AAAGCCCTCCGGTATTTCC
DNMT3A-FE06-08	cDNA	GGGGGACCCCTACTACATCA	CCTGGCCACCAGGAGAAG
DNMT3A-FE07-09	cDNA	GTGGCTACCACGCCTGAG	GACCTCGTAGATGGCTTTGC
DNMT3A-FE08-11	cDNA	GTTCGGAGACGGCAAATTCT	AGGTTCCACCCACATGTCC
DNMT3A-FE10-14	cDNA	GGGGCTTCCAGCCTTCT	CGTCGTCGTCGTACTGGT
DNMT3A-FE13-16	cDNA	CCTCTTCGTTGGAGGAATGT	GGGACAGGTGGGTAAACCTT
DNMT3A-FE15-18	cDNA	GGCACAAGGGTACCTACGG	CCCAATCACCAGATCGAATG
DNMT3A-FE17-20	cDNA	CCAGGGGAAGATCATGTACG	TGCAGCTGACACTTCTTTGG
DNMT3A-FE19-22	cDNA	CCCTTCTTCTGGCTCTTTGA	CTTTCCATTTCAGTGCACCA
DNMT3A-FE21-23	cDNA	CCACTGTGAATGATAAGCTGGA	TTGTGTCGCTACCTCAGTTTG
FLT3-FE16-17	cDNA	CAGCTCTGAAAGAGAGGC	CTTCTTAGATAGTTGAGAAGATCACC
FLT3-FE20-22	cDNA	CCGCCAGGAACGTGCTTG	ATGCCAGGGTAAGGATTCACACC
IDH1-E04	gDNA	AAACTTTGCTTCTAATTTTTCTCTTTC	GCAAAATCACATTATTGCCAAC
IDH1-FE04-05	cDNA	GCTTGTGAGTGGATGGGTAA	TATGTACCAGGTATGTCACCTT
IDH2-E04	gDNA	TGCAGTGGGACCACTATTATCTC	CACCACTGCCATCTTTTGG
IDH2-FE03-05	cDNA	ACTGGCCACCCAGAAGTA	TTGTACACTTCCCACTCC
NPM1-FE10-11	cDNA	TCCCAAAGTGGAAGCC	GGAAAGTTCTCACTCTGC
NRAS-E02	gDNA	GGCCGATATTAATCCGGTGT	TCCGACAAGTGAGAGACAGG
NRAS-E03	gDNA	CCCTCCCTGCCCCCTTAC	CAATGTCAAACAACCTAAAACCA
NRAS-FE02-03	cDNA	GGGAAAAGCGCACTGACAAT	CCTTCGCCTGTCCTCATGTA
NRAS-FE02-03	cDNA	GCTGTGGTCCTAAATCTGTCC	AGGTACATCATCCGAGTCTTTTAC
RUNX1-E01	gDNA	GCTGTTTGCAGGGTCCTAAC	GGCCTCCGCCTGTCCTC
RUNX1-E02	gDNA	CATTGCTATTCCTCTGCAACC	GTTTGTTGCCATGAAACGTG
RUNX1-E03	gDNA	AAATTCCGGGAGTGTTGTCA	GAAAGGTTGAACCCAAGGAA
RUNX1-E04	gDNA	TGATCTCTTCCCTCCCTCCT	CAGTTGGTCTGGGAAGGTGT
RUNX1-E05	gDNA	ATTTGAACAAGGGCCACTCA	AATGTTCTGCCAACTCCTTCA
RUNX1-E06.01	gDNA	CTCCGCAACCTCCTACTCAC	CCCACCATGGAGAACTGGTA
RUNX1-E06.02	gDNA	CCCGTTCCAAGCCAGCTC	GCTTGTCGCGAACAGGAG
RUNX1-FE01-02	cDNA	TGCAGGGTCCTAACTCAATC	CATTGCCAGCCATCACAGTGAC
RUNX1-FE02-04	cDNA	TTTCAAGGTGGTGGCCCTA	CTGAGGGTTAAAGGCAGTGGAGT
RUNX1-FE04-06	cDNA	CCGGGAGCTTGTCCTTTTCC	CGGCAGGTAGGTGTGGTAG
RUNX1-FE06-06	cDNA	CAGGCGCCTTCACCTACTC	TGACCTACAGCGAGATCCTG
RUNX1-FE06-06	cDNA	CAGGCGCCTTCACCTACTC	CTCAGTAGGGCCTCCACACG
RUNX1-FE06.01	cDNA	CAGGCGCCTTCACCTACTC	TGACCTACAGCGAGATCCTG
RUNX1-FE06.01	cDNA	CAGGCGCCTTCACCTACTC	GTCGCTCTGGTTCGGGAG
RUNX1-FE06.02	cDNA	CTCCTACCACCTGTACTACGG	CTGACCTACAGCGAGATCCTG
TET2-E03.01	gDNA	ATTCAACTAGAGGGCAGCCTTG	ACTGTGCGTTTTATTCCTCCAT
TET2-E03.02	gDNA	GAATACCCTGTATGAAGGGAAGC	CCCACTGCAGTTATGTGTTGAA
TET2-E03.03	gDNA	TGTAGCCCAAGAAAATGCAG	TGGGTGAGTGATCTCACAGG
TET2-E03.04	gDNA	CATCTCACATAAATGCCATTAACA	AGCTTGCAAATTGCTGCTG
TET2-E03.05	gDNA	GAAAATAACATCCAGGGAACCA	CCCTCTATTTTCACTTCCCTTAAA
TET2-E03.06	gDNA	GGAGTTTTAGAAGAACACCACCA	TCGACCCTTCAGAATCTCTTG
TET2-E03.07	gDNA	CCAATTTTTGGTAGCAGTGGA	CCAGCTGTGTTGTTTTCTGG
TET2-E03.08	gDNA	TGACCTCCAAACAATACACTGG	TGAGTTTGAAAATGGCTCAGTC
TET2-E03.09	gDNA	CCCAGTGTTGAAACAGCA	ACTTCCTCCAGTCCCATTTG
TET2-E03.10	gDNA	TGGTGAAAATCAGTATTCAAAATCA	CCCTGTAGAACTGAAGCTTGTTG
TET2-E03.11	gDNA	CTTCTTCACAGGTGCTTTCAAG	ATACAGGCATGTGGCTTGC
TET2-E03.12	gDNA	TTGCCATAGTCAGATGCACAG	CTGAAGAAGTTGTTTGCTGCTCT
TET2-E03.13	gDNA	TTGACTAGACAAACCACTGCTG	TTTATGAGCCTTTACAAATTGCTG
TET2-E04	gDNA	TGGCACATTTTCTAATAGATCAGTC	CTTTGTGTGTGAAGGCTGGA
TET2-E05	gDNA	AAACCGTTCATTTCTCAGGATG	GTAATGTTCTTTTTAACTGGCATGA
TET2-E06	gDNA	TGACCCTTGTTTTGTTTTGG	CGCTGAACTCTCTTCCTTTCA
TET2-E07	gDNA	ATAGACACCTATAATATCAGCTGCAC	CAGTTTGGGAAAAACTTTGATTA
TET2-E08	gDNA	CCATATATTGTGTTTGGGATTCAA	GCAGTGGTTTCAACAATTAAGAG
TET2-E09	gDNA	TGCTCTATTTTGTGTCATTCCATT	CAGTGTGAGAACAGACTCAACAG
TET2-E10.01	gDNA	GGGACCTGTAGTTGAGGCTGT	GGGGCTGACTTTTCCTTTTC
TET2-E10.02	gDNA	GAGTTTGGGAGTGTGGAAGC	GGGGGCAAAACCAAAATAAT
TET2-E11.01	gDNA	GCCTTCATAAAATAATCATCAACA	CTGCAGCTTGAGATGAGGTG
TET2-E11.02	gDNA	CCAATCCAGTTAGTCCTTATCCA	AAAACTCTGGCTATTTCCAAACC
TET2-E11.03	gDNA	CAAGCCAAGACCCTCTGTCT	GCATGAAGAGAGCTGTTGAA

TET2-E11.04	gDNA	GGTGAACATCATTCACCTTCTC	GAATTGACCCATGAGTTGGAG
TET2-E11.05	gDNA	AGACAGCGAGCAGAGCTTTC	AAGTTTCATGTGGCTCAGCA
TET2-E11.06	gDNA	AGCCCGTGAGAAAGAGGAAG	ACTGTGACCTTTCCCCACTG
TP53-E04	gDNA	ACCTGGTCCTCTGACTGCTC	CAGGCATTGAAGTCTCATGG
TP53-E05E06	gDNA	CACTTGTGCCCTGACTTTCA	GCCACTGACAACCACCCTTA
TP53-E07	gDNA	CTTGCCACAGGTCTCCCC	AAGAGGTCCCAAAGCCAGAG
TP53-E08E09	gDNA	GACAAGGGTGGTTGGGAGTA	ACAGTCAAGAAGAAAACGGCA
TP53-E10	gDNA	ACTTGAACCATCTTTTAACTCAGGT	TCTGTGCAGGGCTGGGAC
TP53-FE02-06	cDNA	CAGTCAGATCCTAGCGTCGAG	ACACGCAAATTTCCTTCCAC
TP53-FE03-05	cDNA	TCAGACCTATGGAAACTACTTCCTG	GGCAAAACATCTTGTTGAGG
TP53-FE04-06	cDNA	GGCCCCTGTCATCTTCTGT	ACACGCAAATTTCCTTCCAC
TP53-FE05-08	cDNA	TGGCCATCTACAAGCAGTCA	AGCTGTTCCGTCCCAGTAGA
TP53-FE06-11	cDNA	CTCAGCATCTTATCCGAGTGG	TTATGGCGGGAGGTAGACTG
TP53-FE07-10	cDNA	TGGCTCTGACTGTACCACCA	CCTCATTCAGCTCTCGGAAC
TP53-FE09-11	cDNA	CCAGCCAAAGAAGAAACCAC	TTCTGACGCACACCTATTGC
WT1-E07	gDNA	GACCTACGTGAATGTTCACATG	ACAACACCTGGATCAGACCT
WT1-E09	gDNA	TGCAGACATTGCAGGCATGGCAGG	GCACTATTCCTTCTCTCAACTGAG

Table S3: Primer sequences used to analyze the respective genes by Sanger Sequencing

Gene-Identifier	Material	Sequence: Forward-Primer(5´-3´)	Sequence: Reverse-Primer(5´-3´)
ASXL1-E13.01	gDNA	GGTCAGATCACCCAGTCAGT	GTCCAACTGTAGCCCTCTGT
ASXL1-E13.02	gDNA	GGACCTGCCTTCTCTGAGA	TCTGGATTCTGGTTTGGGCT
ASXL1-E13.03	gDNA	CCAAGGCTCTCGTTTCTAACAG	CATACTCGAGACACCCAGCT
ASXL1-E13.04	gDNA	ACATGCGTCTGGTTACAAGG	AGCACGGACTTCCTTCTGAT
ASXL1-E13.05	gDNA	ACAAATCCCATTACATCCTCTAGG	CTCGGGGTAATTTCCAGAAGG
ASXL1-E13.06	gDNA	TGCCGAGAACAGGAAAGCTA	TTTTGGGGGAAGGCAAGAGT
ASXL1-FE13.01	cDNA	AAGCCACAGCCCACTAAAGA	GTCCAACTGTAGCCCTCTGT
ASXL1-FE13.02	cDNA	GGACCTGCCTTCTCTGAGA	TCTGGATTCTGGTTTGGGCT
ASXL1-FE13.03	cDNA	CCAAGGCTCTCGTTTCTAACAG	CATACTCGAGACACCCAGCT
ASXL1-FE13.04	cDNA	ACATGCGTCTGGTTACAAGG	AGCACGGACTTCCTTCTGAT
ASXL1-FE13.05	cDNA	TGGATTCCAAAGAGCAGTTCTCTTC	CATGACAAAGGGCATCCCTTCCAA
ASXL1-FE13.06	cDNA	ACAGGAAAGCTACTGGGCATAGTC	CAAGAGTGCTCCTGCCTAAAGAGT
CEBPA-E01.01	gDNA	GCCATGCCGGGAGAACT	CCCGGGTAGTCAAAGTCG
CEBPA-E01.02	gDNA	CCTTCAACGACGAGTTCCTG	CGGCTGGTAAGGGAAGAGG
CEBPA-E01.03	gDNA	GAGGAGGATGAAGCCAAGC	CTCGTTGCTGTTCTTGTCCA
CEBPA-E01.04	gDNA	TGGCAGCGCGCTCAAG	CCAGGGCGGTCCCACA
CEBPA-FE01.01	cDNA	GCCATGCCGGGAGAACT	CCCGGGTAGTCAAAGTCG
CEBPA-FE01.02	cDNA	CCTTCAACGACGAGTTCCTG	CGGCTGGTAAGGGAAGAGG
CEBPA-FE01.03	cDNA	GAGGAGGATGAAGCCAAGC	CTCGTTGCTGTTCTTGTCCA
CEBPA-FE01.04	cDNA	TGGCAGCGCGCTCAAG	CCAGGGCGGTCCCACA
DNMT3A-E07	gDNA	TTCCTGGAGAGGTCAAGGTG	TGGAGAGAGGAGAGCAGGAC
DNMT3A-E08	gDNA	GCCTCGTGACCACTGTGTAA	ACCCACCACAGGCAGAGTAG
DNMT3A-E09	gDNA	CTCCTCTTTGCATCGGGTAA	ACCTGCACTCCAACTTCCAG
DNMT3A-E10	gDNA	TGTGCCACCCTCACTACTCA	TCCCTAAGCATGGCTTTCC
DNMT3A-E11E12	gDNA	GACCTTGGCACCTGCTTTC	CCACACTAGGAGTGCCAGAGTT
DNMT3A-E13	gDNA	GGTCACAGTGCCTCCCTTT	ACCCTGTACATGCCCAGAAG
DNMT3A-E14	gDNA	CACAGGCAGATGAGGTTTCC	CCCAGCTAAGGAGACCACTG
DNMT3A-E15	gDNA	CCCTAGCCATGCTCCAGAC	CCCACAACCAAGGCTCAG
DNMT3A-E16	gDNA	CAGGGTGTGTGGGTCTAGGA	TGCATACGTTTCCACTTCACA
DNMT3A-E17	gDNA	AAAGATAGGACTTGGGCCTACA	CTGCCTCCAGGTGCTGAG
DNMT3A-E18	gDNA	TGGTCCCGTTCTTGTTTAGG	CAAGGAGGAAGCCTATGTGC
DNMT3A-E19	gDNA	GACAGCTATTCCCGATGACC	GCTCCACAATGCAGATGAGA
DNMT3A-E20	gDNA	TGTGTGGCTCCTGAGAGAGA	CATGGCAGAGCAGCTAGTCA
DNMT3A-E21	gDNA	TGGTGGATTTGTGTCTTTGC	CATCCTGCCCTTCCTTCTC
DNMT3A-E22	gDNA	CTGCGAACTCTGCTCACTCA	AGCAAGCACAGCAATCAGAA
DNMT3A-E23	gDNA	CACTCACCCTGCCCTCTCT	AAAGCCCTCCGGTATTTCC
DNMT3A-FE06-08	cDNA	GGGGGACCCCTACTACATCA	CCTGGCCACCAGGAGAAG
DNMT3A-FE07-09	cDNA	GTGGCTACCACGCCTGAG	GACCTCGTAGATGGCTTTGC
DNMT3A-FE08-11	cDNA	GTTCGGAGACGGCAAATTCT	AGGTTCCACCCACATGTCC
DNMT3A-FE10-14	cDNA	GGGGCTTCCAGCCTTCT	CGTCGTCGTCGTACTGGT
DNMT3A-FE13-16	cDNA	CCTCTTCGTTGGAGGAATGT	GGGACAGGTGGGTAAACCTT
DNMT3A-FE15-18	cDNA	GGCACAAGGGTACCTACGG	CCCAATCACCAGATCGAATG
DNMT3A-FE17-20	cDNA	CCAGGGGAAGATCATGTACG	TGCAGCTGACACTTCTTTGG
DNMT3A-FE19-22	cDNA	CCCTTCTTCTGGCTCTTTGA	CTTTCCATTTCAGTGCACCA
DNMT3A-FE21-23	cDNA	CCACTGTGAATGATAAGCTGGA	TTGTGTCGCTACCTCAGTTTG
FLT3-FE16-17	cDNA	CAGCTCTGAAAGAGAGGC	CTTCTTAGATAGTTGAGAAGATCACC
FLT3-FE20-22	cDNA	CCGCCAGGAACGTGCTTG	ATGCCAGGGTAAGGATTCACACC
IDH1-E04	gDNA	AAACTTTGCTTCTAATTTTTCTCTTTC	GCAAAATCACATTATTGCCAAC
IDH1-FE04-05	cDNA	GCTTGTGAGTGGATGGGTAA	TATGTACCAGGTATGTCACCTT
IDH2-E04	gDNA	TGCAGTGGGACCACTATTATCTC	CACCACTGCCATCTTTTGG
IDH2-FE03-05	cDNA	ACTGGCCACCCAGAAGTA	TTGTACACTTCCCACTCC
NPM1-FE10-11	cDNA	TCCCAAAGTGGAAGCC	GGAAAGTTCTCACTCTGC
NRAS-E02	gDNA	GGCCGATATTAATCCGGTGT	TCCGACAAGTGAGAGACAGG
NRAS-E03	gDNA	CCCTCCCTGCCCCCTTAC	CAATGTCAAACAACCTAAAACCA
NRAS-FE02-03	cDNA	GGGAAAAGCGCACTGACAAT	CCTTCGCCTGTCCTCATGTA
NRAS-FE02-03	cDNA	GCTGTGGTCCTAAATCTGTCC	AGGTACATCATCCGAGTCTTTTAC
RUNX1-E01	gDNA	GCTGTTTGCAGGGTCCTAAC	GGCCTCCGCCTGTCCTC
RUNX1-E02	gDNA	CATTGCTATTCCTCTGCAACC	GTTTGTTGCCATGAAACGTG
RUNX1-E03	gDNA	AAATTCCGGGAGTGTTGTCA	GAAAGGTTGAACCCAAGGAA
RUNX1-E04	gDNA	TGATCTCTTCCCTCCCTCCT	CAGTTGGTCTGGGAAGGTGT
RUNX1-E05	gDNA	ATTTGAACAAGGGCCACTCA	AATGTTCTGCCAACTCCTTCA
RUNX1-E06.01	gDNA	CTCCGCAACCTCCTACTCAC	CCCACCATGGAGAACTGGTA
RUNX1-E06.02	gDNA	CCCGTTCCAAGCCAGCTC	GCTTGTCGCGAACAGGAG

RUNX1-FE01-02	cDNA	TGCAGGGTCCTAACTCAATC	CATTGCCAGCCATCACAGTGAC
RUNX1-FE02-04	cDNA	TTTCAAGGTGGTGGCCCTA	CTGAGGGTTAAAGGCAGTGGAGT
RUNX1-FE04-06	cDNA	CCGGGAGCTTGTCCTTTTCC	CGGCAGGTAGGTGTGGTAG
RUNX1-FE06-06	cDNA	CAGGCGCCTTCACCTACTC	TGACCTACAGCGAGATCCTG
RUNX1-FE06-06	cDNA	CAGGCGCCTTCACCTACTC	CTCAGTAGGGCCTCCACACG
RUNX1-FE06.01	cDNA	CAGGCGCCTTCACCTACTC	TGACCTACAGCGAGATCCTG
RUNX1-FE06.01	cDNA	CAGGCGCCTTCACCTACTC	GTCGCTCTGGTTCGGGAG
RUNX1-FE06.02	cDNA	CTCCTACCACCTGTACTACGG	CTGACCTACAGCGAGATCCTG
TET2-E03.01	gDNA	ATTCAACTAGAGGGCAGCCTTG	ACTGTGCGTTTTATTCCTCCAT
TET2-E03.02	gDNA	GAATACCCTGTATGAAGGGAAGC	CCCACTGCAGTTATGTGTTGAA
TET2-E03.03	gDNA	TGTAGCCCAAGAAAATGCAG	TGGGTGAGTGATCTCACAGG
TET2-E03.04	gDNA	CATCTCACATAAATGCCATTAACA	AGCTTGCAAATTGCTGCTG
TET2-E03.05	gDNA	GAAAATAACATCCAGGGAACCA	CCCTCTATTTTCACTTCCCTTAAA
TET2-E03.06	gDNA	GGAGTTTTAGAAGAACACCACCA	TCGACCCTTCAGAATCTCTTG
TET2-E03.07	gDNA	CCAATTTTTGGTAGCAGTGGA	CCAGCTGTGTTGTTTTCTGG
TET2-E03.08	gDNA	TGACCTCCAAACAATACACTGG	TGAGTTTGAAAATGGCTCAGTC
TET2-E03.09	gDNA	CCCAGTGTTGAAACAGCA	ACTTCCTCCAGTCCCATTTG
TET2-E03.10	gDNA	TGGTGAAAATCAGTATTCAAAATCA	CCCTGTAGAACTGAAGCTTGTTG
TET2-E03.11	gDNA	CTTCTTCACAGGTGCTTTCAAG	ATACAGGCATGTGGCTTGC
TET2-E03.12	gDNA	TTGCCATAGTCAGATGCACAG	CTGAAGAAGTTGTTTGCTGCTCT
TET2-E03.13	gDNA	TTGACTAGACAAACCACTGCTG	TTTATGAGCCTTTACAAATTGCTG
TET2-E04	gDNA	TGGCACATTTTCTAATAGATCAGTC	CTTTGTGTGTGAAGGCTGGA
TET2-E05	gDNA	AAACCGTTCATTTCTCAGGATG	GTAATGTTCTTTTTAACTGGCATGA
TET2-E06	gDNA	TGACCCTTGTTTTGTTTTGG	CGCTGAACTCTCTTCCTTTCA
TET2-E07	gDNA	ATAGACACCTATAATATCAGCTGCAC	CAGTTTGGGAAAAACTTTGATTA
TET2-E08	gDNA	CCATATATTGTGTTTGGGATTCAA	GCAGTGGTTTCAACAATTAAGAG
TET2-E09	gDNA	TGCTCTATTTTGTGTCATTCCATT	CAGTGTGAGAACAGACTCAACAG
TET2-E10.01	gDNA	GGGACCTGTAGTTGAGGCTGT	GGGGCTGACTTTTCCTTTTC
TET2-E10.02	gDNA	GAGTTTGGGAGTGTGGAAGC	GGGGGCAAAACCAAAATAAT
TET2-E11.01	gDNA	GCCTTCATAAAATAATCATCAACA	CTGCAGCTTGAGATGAGGTG
TET2-E11.02	gDNA	CCAATCCAGTTAGTCCTTATCCA	AAAACTCTGGCTATTTCCAAACC
TET2-E11.03	gDNA	CAAGCCAAGACCCTCTGTCT	GCATGAAGAGAGCTGTTGAA
TET2-E11.04	gDNA	GGTGAACATCATTCACCTTCTC	GAATTGACCCATGAGTTGGAG
TET2-E11.05	gDNA	AGACAGCGAGCAGAGCTTTC	AAGTTTCATGTGGCTCAGCA
TET2-E11.06	gDNA	AGCCCGTGAGAAAGAGGAAG	ACTGTGACCTTTCCCCACTG
TP53-E04	gDNA	ACCTGGTCCTCTGACTGCTC	CAGGCATTGAAGTCTCATGG
TP53-E05E06	gDNA	CACTTGTGCCCTGACTTTCA	GCCACTGACAACCACCCTTA
TP53-E07	gDNA	CTTGCCACAGGTCTCCCC	AAGAGGTCCCAAAGCCAGAG
TP53-E08E09	gDNA	GACAAGGGTGGTTGGGAGTA	ACAGTCAAGAAGAAAACGGCA
TP53-E10	gDNA	ACTTGAACCATCTTTTAACTCAGGT	TCTGTGCAGGGCTGGGAC
TP53-FE02-06	cDNA	CAGTCAGATCCTAGCGTCGAG	ACACGCAAATTTCCTTCCAC
TP53-FE03-05	cDNA	TCAGACCTATGGAAACTACTTCCTG	GGCAAAACATCTTGTTGAGG
TP53-FE04-06	cDNA	GGCCCCTGTCATCTTCTGT	ACACGCAAATTTCCTTCCAC
TP53-FE05-08	cDNA	TGGCCATCTACAAGCAGTCA	AGCTGTTCCGTCCCAGTAGA
TP53-FE06-11	cDNA	CTCAGCATCTTATCCGAGTGG	TTATGGCGGGAGGTAGACTG
TP53-FE07-10	cDNA	TGGCTCTGACTGTACCACCA	CCTCATTCAGCTCTCGGAAC
TP53-FE09-11	cDNA	CCAGCCAAAGAAGAAACCAC	TTCTGACGCACACCTATTGC
WT1-E07	gDNA	GACCTACGTGAATGTTCACATG	ACAACACCTGGATCAGACCT
WT1-E09	gDNA	TGCAGACATTGCAGGCATGGCAGG	GCACTATTCCTTCTCTCAACTGAG