Molecular Characterization and Residual Disease Monitoring of Acute Myeloid Leukemia

Adil S. A. Al Hinai

Molecular Characterization and Residual Disease Monitoring of Acute Myeloid Leukemia

ADIL AL HINAI

ISBN: 978-94-93197-90-9
Layout: Egied Simons
Cover: Nadia Ayu Lestari
Printing: Off Page
The work described in this thesis was performed at the Department of Hematology at the Erasmus Medical Center, Rotterdam, the Netherlands. This work and the printing of this thesis has been financially supported by Ministry of Health, Sultanate of Oman.

Molecular Characterization and Residual Disease Monitoring of Acute Myeloid Leukemia

Moleculaire Karakterisering en Monitoring van Restziekte van Acute Myeloïde Leukemie

Thesis
Proefschrift

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. L.A. Bredenoord

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Wednesday, 24 November 2021 at 13:00 hrs

by

ADIL S. A. AL HINAI

Born in Muscat, Sultanate of Oman

Erasmus University Rotterdam

Erasmus

DOCTORAL COMMITTEE

Promotor: prof.dr. H.R. Delwel

Other members: prof.dr. I. P. Touw
prof.dr. W.N.M. Dinjens
prof.dr. J. Cloos

Co-promotor: dr. P.J.M. Valk

اللهم علمنا ما ينفعنا
وانفعنا بما علمتنا
وزدنا علما

CONTENTS

Chapter 1

General introduction

HEMATOPOIESIS

The blood system has many different lineages of mature blood cells that perform various and specialized tasks: red blood cells or erythrocytes transport carbon dioxide and provide oxygen and hemoglobin throughout the body. White blood cells, also known as leukocytes, recognized in the blood system as morphologically different subtypes, defend against pathogens and foreign bodies. Megakaryocytes generate platelets that confer the process of blood clotting (Figure 1). Mature blood cells have a short life span and require continuous regeneration to maintain a constant supply of cells to ensure their functions' sustainability. The process of forming all blood cellular components to replenish the blood system is known as hematopoiesis. In adulthood, this process occurs mainly in the bone marrow (BM) and all blood cell lineages arise from hematopoietic stem cells (HSCs).[1-3]

In adult mammals, HSCs reside at the apex of the hematopoietic hierarchy of many progenitor cell stages with progressively restricted lineage potential that gives rise to all blood cells. The HSCs are known to reside in a specific microenvironment called "niches" that orchestrate HSC function.[4] The HSCs have a distinctive ability to self-perpetuate through a process known as self-renewal, in addition to being able to give rise to many mature cell types through differentiation. The HSCs choice between self-renewal and differentiation is a strictly controlled process to ensure that the HSC pool is not depleted and to enable the generation of differentiated cells. The HSC compartment consists of cells with increasingly declined self-renewal aptitude with the maintenance of multi-lineage potential.[2, 5] The HSCs are divided into two populations based on their self-renewal capacity: long-term (LT) and short-term (ST) HSCs. The LT-HSC is at the peak of the hematopoietic hierarchy, is rare and is predominantly dormant. Besides, LT-HSCs have the lifelong proliferative ability and contribute to long-lasting multilineage re-formation. In contrast, the ST-HSCs have limited self-renewing capacity.[5-9]

At the start of HSCs differentiation, LT-HSCs lose their self-renewal ability, differentiating first into ST-HSCs and subsequently to multipotent progenitors (MPPs). MPPs have no evident self-renewal propensity but retain the capacity for multilineage differentiation.[9-11] A significant degree of homeostatic control of mature cells is medicated at the level of multipotent, oligo-potent, and lineage-restricted progenitor cells due to the enormous proliferative and development capacity of these more committed progenitors. The MPPs, in turn, give rise to oligo-potent progenitors, which have more restricted developmental potential. This denotes the branching stage in the hematopoietic hierarchy with the common lymphoid progenitors (CLP), giving rise to mature lymphoid progenitor cells, which include B-cells, T-cells, dendritic and natural killer (NK) cells.[12] MPPs also give rise to common myeloid progenitors (CMP), which can differentiate and generate either megakaryocyte/erythrocyte progenitors (MEPs) or granulocyte/macrophage progenitors (GMPs).[13, 14] MEPs give rise to either erythrocytes or megakaryocytes, which in turn generate platelets. GMPs

differentiate into monocytes/macrophages and granulocytes (neutrophil, eosinophil, and basophil).

The process of hematopoiesis is meticulously controlled to meet a variety of human biological demands throughout life, including normal homeostasis, severe blood loss, or infection. The regulation of HSCs fate is mediated through a wide range of critical factors, including cytokines, growth factors, transcription factors, chromatin modifiers, and cell cycle regulators, and also, both intrinsic and extrinsic signals play a role in the regulation of HSCs. Deregulation of normal HSC fate decision underlines hematological disorder.[2, 15] The abrogation of normal hematopoiesis (or HSC function) via either cell-intrinsic genetic alterations (gene mutations, deletions, amplifications or translocations, and epigenetic changes) and/or cell-extrinsic factors as a result of a defect in BM microenvironment; both can lead to hematological malignancies including leukemia, myelodysplastic syndrome (MDS) and myeloproliferative neoplasms (MPNs).[4]

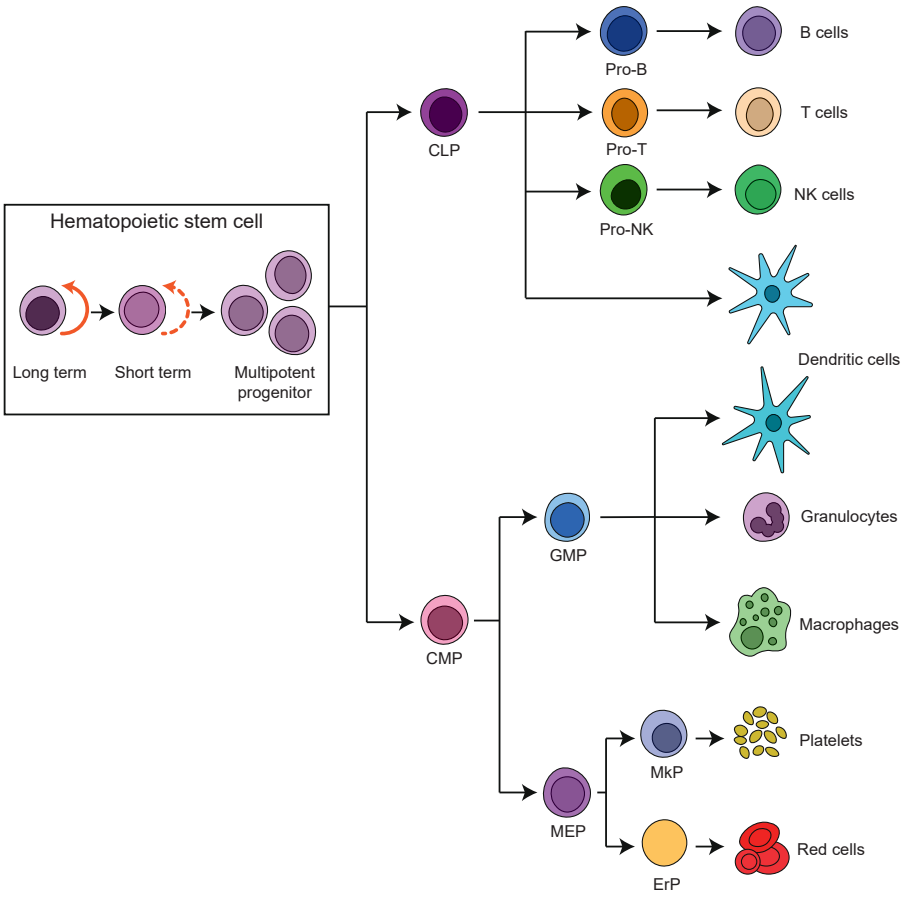


Figure 1: Hematopoiesis. The hematopoietic stem cell (HSC) compartment is divided into long-term self-renewing HSCs (LT-HSC), short-term HSCs (ST-HSC), and multipotent progenitors (MPP). They give rise to common lymphoid progenitors (CLP) and common myeloid progenitors (CMP). The CLP can give rise to Pro-B, Pro-T, and Pro-NK, and upon maturation, they give rise to B-cells, T-cells, and NK-cells, respectively. Besides, CLP can give rise to dendritic cells. The CMP provides granulocyte-macrophage precursors (GMP) and macrophage erythrocyte precursors (MEP). Successively, GMPs give rise to dendritic cells, granulocytes, and macrophages. The MEPs give rise to megakaryocyte precursors (MkP) and erythrocyte precursors (ErP), which mature to platelets and erythrocytes, respectively. Adapted from Reya *et al*. [5]

Acute Myeloid Leukemia

Acute Myeloid Leukemia (AML) is a heterogeneous clonal disorder of hematopoietic progenitor cells characterized by uncontrolled proliferation of immature cells ("myeloblast') and a block in myeloid differentiation that accumulate in BM and peripheral blood (PB). AML can occur as primary or *de novo*, which means patients have no known previous clinical history of preceding hematological malignancy such as MDS or MPN and were previously not exposed to potential leukemogenic agent or treatment.[16] In other cases, AML develops from other malignant hematopoietic conditions such as MDS or MPN (secondary or s-AML), or arises following treatment with genotoxic therapy for unrelated malignancies (therapy-related or t-AML); yet, most AMLs appear as a *de novo*.[17, 18]

Epidemiology and Etiology of AML

AML is a rare disorder, yet an aggressive malignant neoplasm responsible for a large number of cancer-related mortalities.[19] In adult, AML is one of the most common myeloid neoplasms with an incidence of 3.8 cases per 100,000 individuals. The prevalence increases to 17.9 cases per 100,000 persons aged 65 years and older, with a median age of about 70 years at presentation. Also, three men are affected for every two women.[17] In children, AML occurs in 15 – 20% of all acute leukemias. The incidence in infants is 1.5 cases per 100,000 individuals and decreases with age, i.e., 0.4 per 100,000 individuals children aged 5 to 9 years old; subsequently, the prevalence gradually increases into adulthood and accounts for approximately 25% of all leukemias in adults. AML has been associated with several risk factors (Table 1).

The underlying causes of AML remain elusive, but AML generally occurs as a *de novo* malignancy. In elderly patients, AML is often preceded by a pre-leukemic disease state, including MDS and MPN. Individuals suffering from genetic disorders such as Fanconi anemia, or Bloom syndrome, have a high risk for developing AML as a secondary disease. Recent studies have shown that families with a germline mutation involving one of several genes, for instance, *TP53*, *RUNX1*, *GATA2,* and *CEBPA*, have a familial predisposition risk of developing AML.[20] Although this hereditary myeloid malignancy is usually presenting in childhood, it can also be seen in adults generally less than 40 years of age and occasionally in older adults.[21] Moreover, AML may develop as a secondary malignancy in patients who have previously been exposed to radiotherapy or chemotherapy-containing alkylating agents or epipodophyllotoxins.[20]

Table 1: Selected risk factors associated with Acute Myeloid Leukemia

Genetic disorders	Down Syndrome
	Klinefelter syndrome
	Patau syndrome
	Ataxia telangiectasia
	Schwachman syndrome
	Kostman syndrome
	Neurofibromatosis
	Fanconi anemia
	Li-Fraumeni syndrome
Physical and chemical exposures	Benzene
	Drugs such as pipobroman
	Pesticides
	Cigarette smoking
	Embalming fluids
	Herbicides
Radiation exposure	Non-therapeutic, therapeutic radiation
Chemotherapy	Alkylating agents
	Topoisomerase-II inhibitors
	Anthracyclines
	Taxanes

Adapted from Deschler and Lubbert [19]

Diagnosis, Classification, and Risk Stratification

AML is characterized by the accumulation of blasts in the BM and PB, resulting from an increase in the proliferation of undifferentiated myeloid progenitors. The classical and primary diagnosis of AML is based on the morphological identification of leukemic blast in PB and BM. Morphologic, cytochemical, or immunophenotypic features are used to determine the lineage of the neoplastic cells and to assess their maturation. AML is a very heterogeneous disease. At the morphological level, this heterogeneity is exhibited by variability in the degree of commitment and differentiation of the cell lineage. This variability led AML to be categorized into distinct morphologic subgroups. The most commonly used method to classify AML was the French–American–British (FAB), which is based on cytomorphology and cytochemistry.[22] However; this system has been updated by a more comprehensive classification model developed by the world health organization (WHO), which incorporates besides cytomorphology, also cytogenetics and molecular genetics (Table 2).[23-25]

The inclusion of cytogenetics resulted in the identification of chromosomal rearrangements in a relatively large fraction of AML patients, which correlated with specific FAB subtypes. Several of cytogenetic lesions are used to identify patients with distinct

clinical outcomes. Cytogenetic and molecular genetic testing has become an indispensable part of the routine diagnostic workup of a patient with AML. This multidisciplinary approach, including cytomorphology, immunophenotyping, cytogenetics, and molecular testing do not only make it possible to modify therapy based on the sensitivity of biologically defined AML subtype but also provides unique markers that can be used to monitor patient response to treatment.[22]

Diagnosis of AML is confirmed when the marrow or peripheral blood contains at least 20% blast cells of myeloid origin determined immunophenotypically or by cytochemical analyses.[18] The minimum criterion of 20% blast is not a mandate to confirm that the patient has AML or blast transformation; a therapeutic decision must be based on the clinical situation after all information is considered. For instance, cases of core-binding factor (CBF) AML [associated with inv(16)/t(16;16)(p13.1q22), or t(8;21)(q22;q22)], *NPM1* mutant AML, or acute promyelocytic leukemia (APL) [associated with t(15;17)(q22;q21)] are also exempted from the ≥20% blast cut-off; they should be considered to have AML regardless of the blast percentage. The 20% blast criterion is arbitrarily used to distinguish AML from MDS.

Cytogenetic analysis is essential at presentation for establishing the baseline karyotype, judging response to therapy, and detecting genetic evolution at follow-up. In some cases, fluorescence *in situ* hybridization is required to detect rearrangement of well-recognized cytogenetic aberrations or submicroscopic abnormalities not detected by routine karyotyping. Furthermore, gene mutations are ever more being recognized as important diagnostic and prognostic markers in AML; examples of these are mutations in *NPM1*, *FLT3*, *CEBPA*, and *RUNX1*.[21, 23, 26]

Table 2: WHO classification of myeloid neoplasms

Types	Genetic abnormalities
AML 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 *PML-RARA*
	AML with t(9;11)(p21.3;q23.3); *MLLT3-KMT2A*
	AML with t(6;9)(p23;q34.1); *DEK-NUP214*
	AML with inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2*, *MECOM*
	AML (megakaryoblastic) with t(1;22)(p13.3;q13.3); *RBM15-MKL1*
	AML with *BCR-ABL1* (provisional entity)
	AML with mutated *NPM1*
	AML with biallelic mutations of *CEBPA*
	AML with mutated *RUNX1* (provisional entity)
AML with myelodysplasia-related changes	
Therapy-related myeloid neoplasms	
AML, NOS	AML with minimal differentiation
	AML without maturation
	AML with maturation
	Acute myelomonocytic leukemia
	Acute monoblastic/monocytic leukemia
	Acute erythroid leukemia
	Pure erythroid leukemia
	Acute megakaryoblastic leukemia
	Acute basophilic leukemia
	Acute panmyelosis with myelofibrosis
Myeloid sarcoma	
Myeloid proliferations related to Down syndrome	Transient abnormal myelopoiesis
	AML associated with Down syndrome

Abbreviations: AML, acute myeloid leukemia; APL, acute promyelocytic leukemia; NOS, not otherwise specified; WHO, World Health Organization. Adapted from Arber *et al*.[24].

AML is classified based on prognostic risk groups (favorable, intermediate, and adverse) (Table 3). The risk stratification is purely based on cytogenetic abnormalities and gene mutations. This risk stratification system is a recommendation made by an international expert panel on behalf of the European LeukemiaNet (ELN). The ELN guideline is the most widely used for diagnosis, management, and assessing patient risk. The first proposed risk criteria for AML risk stratification was in 2010, and recently the ELN published a revised version of the recommendation (ELN2017). The revision was due to recent discoveries within the genomic landscape of the disease, in the development of genetic testing techniques, detection of minimal residual disease (MRD), and the emergence of novel anti-leukemic therapies. The revised version of the ELN divides AML into three risk categories rather than four risk groups (favorable, intermediate I, intermediate II, and adverse). The reason for this change is that intermediate I and II were separated based on the genetic characteristic, but prognostically these two categories were indistinguishable in older patients, who represent the majority of AML. There were several other modifications to the system, such as biallelic mutated *CEBPA,* that are now considered as a favorable risk in AML. *FLT3*-ITD is deemed to be unfavorable only when the allelic ratio (AR) of mutated to normal alleles is more than 0.5. *NPM1* mutation and *FLT3*-ITD with low AR (<0.5) are regarded as a favorable risk group similar to patients with an *NPM1* mutation. Besides, the adverse risk group has been expanded and includes additional molecular markers and a cytogenetic aberration (i.e., mutated *RUNX1*, mutated *ASXL1,* and the *BCR-ABL1* fusion gene as a provisional AML entity).[27]

Table 3: 2017 ELN risk stratification by genetics

Risk category	Genetic abnormality
Favorable	t(8;21)(q22;q22.1); *RUNX1-RUNX1T1*
	inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*
	Mutated *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low}
	Biallelic mutated *CEBPA*
Intermediate	Mutated *NPM1* and *FLT3*-ITD^{high}
	Wild-type *NPM1* without *FLT3*-ITD or with *FLT3*-ITD^{low} (without adverse-risk genetic lesions)
	t(9;11)(p21.3;q23.3); *MLLT3-KMT2A*
	Cytogenetic abnormalities not classified as favorable or adverse
Adverse	t(6;9)(p23;q34.1); *DEK-NUP214*
	t(v;11q23.3); *KMT2A* rearranged
	t(9;22)(q34.1;q11.2); *BCR-ABL1*
	inv(3)(q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2*, *MECOM*(*EVI1*)
	-5 or del(5q); -7; -17/abn(17p)
	Complex karyotype, monosomal karyotype
	Wild-type *NPM1* and *FLT3*-ITD^{high}
	Mutated *RUNX1*
	Mutated *ASXL1*
	Mutated *TP53*

Abbreviation: ITD: internal tandem duplication. Low: depicts *FLT3*-ITD with an allelic ratio of <0.5, while high represent *FLT3*-ITD with an allelic ratio of >0.5. Adapted from Dohner *et al.*[27]

Genetic landscape of AML

AML is considered a heterogeneous disease due to the variation in morphology, immunophenotype, cytogenetics, molecular aberrations, and response to treatment. The cytogenetic and molecular aberrations are associated with various mechanisms affecting cell proliferation, differentiation, self-renewal, apoptosis, and DNA repair.[23, 28-30] Chromosomal karyotype examination can be considered the first "whole-genome scan" in AML and continues to be an indispensable component of the diagnostic workup for all patients. Early in the 1970s, Rowley and others identified and molecularly characterized recurrent chromosomal rearrangements, including t(8;21), inv(16), and t(15;17).[31] Followed by discovering numerous number of chromosomal rearrangements and copy number variations that also have prognostic value (e.g., -5q/-5, -7, inv(3)/t(3;3), and 11q23 rearrangements).[32]

Cytogenetic abnormalities are found in around 30% - 50% of AML patients at the time of diagnosis and remain the strongest predictor of survival.[27, 33, 34] The remaining 50% of AML patients present with normal cytogenetic. Historically these patients are classified as an intermediate-risk group; however, they are found to exhibit very variable clinical outcomes, which suggests the contribution of other molecular events in the pathogenesis of this AML subcategory. Patients with abnormal cytogenetic at diagnosis regularly present with chromosomal losses and/or balanced reciprocal translocations involving a fusion of a transcription factor essential in maintaining normal hematopoiesis such as CBF or retinoic acid receptor alpha (RARA). Based on the chromosomal structural aberration at diagnosis, patients are classified into three risk categories, favorable, intermediate, and adverse following 2017ELN, as mentioned earlier. Patients with t(15;17) have an excellent prognosis, and this is due to the fact that patients treated with targeted therapy involving all-*trans* retinoic acid (ATRA) and arsenic trioxide (ATO) have up to 90 to 95% achieve complete remission.[35, 36] Patients with CBF alterations have a relatively favorable prognosis. On the other hand, patients with inv(3)(q21q26/t(3;3)(q21;q26), -7 or a complex cytogenetic (i.e., at least three chromosomal aberrations) have an adverse clinical outcome (Table 3). Complex cytogenetics are often associated with the presence of poor-risk cytogenetic aberrations such as -5, del(7q), -17/-17p, -18, or -20. The incidence of the later is higher in patients with secondary AML, and rises with the AML patient's age.[37]

CBF leukemia (i.e., t(8;21) which is associated with the RUNX1-RUNX1T1 fusion protein or inv(16)/t(16;16) with the CBFB-MYH11 fusion protein) represent approximately. 10% of adult *de novo* AML cases. Both translocations disrupt the function of the CBF protein complex leading to a block in myeloid differentiation and, ultimately, leukemia.[27, 38] The chromosomal rearrangement disrupts genes encoding components of the heterodimeric transcription factor complex, including RUNX1 (CBFα) and CBFβ, which plays an essential role in hematopoiesis.[39] Alterations in the receptor tyrosine kinase genes *KIT* and *FLT3* and the proto-oncogenes *NRAS* and *KRAS* have been detected in up to 80% of CBF leukemia patients, suggesting that they function as cooperating factors in CBF leukemogenesis promoting proliferation.[40-42]

Recurrent cytogenetic rearrangements of 11q23 involving *KMT2A* (formerly known as *MLL*) gene are found in approximately 4% of adult AML and 3 to 7% of adults with acute lymphoid leukemia[43] and at a much higher incidence in infants[44, 45]. Till today, 11q23 have been found to have over 60 different translocation partners.[46] *KMT2A* encodes a histone methyltransferase that is essential for regulating gene expression during embryonic development and hematopoiesis. Translocations of the *KMT2A* gene generate chimeric KMT2A fusion proteins that bind to DNA and positively regulate gene transcription. Such activities cause aberrant expression of downstream *KMT2A* targets, including the *HOX* gene, thus resulting in leukemic transformation.[46, 47] Recent Findings showed that the prognosis of 11q23/KMT2A AML is heterogeneous based on the 11q23 fusion partner.[48, 49] Although rearrangements involving *KMT2A* occur more frequently in *de novo* leukemia, a subset of AML patients with *KMT2A* rearrangements is defined by the history of treatment with topoisomerase inhibitors.[50, 51] These cases are frequently associated with rearrangements involving chromosome 11q23 fused to chromosome 4, -9, or -19 and tend to occur with a latency phase of approximately three years post-drug exposure.[52, 53] Besides chromosomal translocations, partial tandem duplication (PTD) of the *KMT2A* gene which is characterized by a duplication of the 5' part of the gene spanning exon 3 to 9, exon 3 to 10, or exon 3 to 11.[54] These mutations detected in 3 to 11% of adult AML patients. *KMT2A*-PTDs very frequently co-occur with trisomy of chromosome 11[54-57], and found in approximately 10% of CN-AML. *KMT2A*-PTDs are found to be associated with unfavorable clinical outcome in AML[56-58].

The advent of next-generation sequencing (NGS) has facilitated new perceptions of myeloid malignancies' molecular origin. Like most human sporadic cancers, AML is a complex, and dynamic disease, characterized by several somatically acquired mutations, concomitant competing clones, and disease evolution over time. Mutated genes in AML can be classified into one of the nine functional categories: transcription factor fusions, the *NPM1* gene, tumor suppressor genes, DNA methylation-related genes, signaling genes, chromatin-modifying genes, myeloid transcription factor genes, cohesin complex genes, and splicing factor genes (Table 4).[59]

Genes that encode transcription factors are one of the most often mutated genes in AML. These genes are typically implicated in chromosomal translocations that lead to an abnormal activation in the hematopoietic compartment, causing impairment of maturation and differentiation of myeloid precursor cells (e.g., CBF translocations and *RARA*- and *KMT2A* rearrangements).[37] In addition to large structural variants, recurrent mutations in genes encoding transcriptions have been implicated in AML pathogenesis, such as *CEBPA* and *RUNX1*. *CEBPA* mutations are detected in about 10% of cytogenetic normal (CN) AML, although these can also be detected in patients with an abnormal karyotype (e.g., del9q). *CEBPA* mutations are never found to coexist with balanced chromosomal rearrangements. Mutations are mainly found to cluster in both the amino- (N-) and carboxy- (C-)terminal

regions; the former results in the expression of the truncated isoform of CEBPA (p30) and loss of the full-length protein (p42).[60] Mutations in the carboxy-terminal affect sites involved in mediating dimerization and DNA binding. Both alleles are mutated in over half of the patients with mutant *CEBPA* gene, combining an upstream mutation in one allele with a downstream mutation in the other allele. Atypically, *CEBPA* mutations are inherited in a subset of patients through germline involvement, with the formation of AML associated with the acquisition of additional *CEBPA* mutations, typically in the other *CEBPA* allele.[61] Studies in murine models have shown that loss of p42 expression (simulating biallelic N-terminal *CEBPA* mutations) or compound heterozygous mutations involving both amino- and carboxy-terminal regions impact hematopoiesis and lead to AML.[62, 63] Early studies revealed that the presence of *CEBPA* mutation predicts a relatively favorable outcome, but subsequent studies showed that the effect was limited to the subgroup with the biallelic mutation. Such AMLs are associated with mutated *GATA2* and usually lack *FLT3*-ITD.[64-68]

RUNX1 is a relatively common somatic mutation in cytogenetic standard risk and CN-AML.[69] Similar to *CEBPA* mutations, germline mutations in the *RUNX1* gene have also been described, which result in familial platelet disorder and predisposition to AML. In adult AML, *RUNX1* mutations are scattered throughout the gene and can also be biallelic. They are usually mutually exclusive of balanced chromosomal rearrangements and mutations involving *NPM1* and *CEBPA* but reported to be associated with the M0 FAB subtype, trisomy of chromosome 13, *KMT2A*-PTD, and *ASXL1* mutations.[70, 71] About 10% of MDS cases also have mutated *RUNX1*[72], indicating that a fraction of AML patients with the secondary disease may evolve from MDS. *RUNX1* mutation has been shown consistently by large cohort studies of young adults with AML to be an independent predictor for a poor outcome.[73-75]

The ectopic virus integration site 1 (*EVI1*) is a transcription located on chromosome 3q26 and plays a vital role in HSC fate.[76, 77] The expression of *EVI1* is affected by recurrent chromosomal rearrangements in AML, MDS[78, 79], and CML[80]. *EVI1* has been shown to be aberrantly expressed in virtually all adult AML patients with inv(3)/t(3;3).[78, 81-83] Thus, abnormal *EVI1* expression has been associated with this subcategory of AML that accounts for 2 to 2.5% of all AMLs.[78, 81, 82] Conversely, overexpression of *EVI1* can be found in other types of AML with 3q rearrangements other than the classical inv(3)/t(3;3) (e.g., t(2;3)(p12;22q26), t(3;7)(q26;q21), t(3;6)(q26;q25) and t(3;17)(q26;q22).[78, 79, 81-88] AML cases with high *EVI1* expression but lack detectable cytogenetic alterations in 3q, may have an enigmatic translocation involving 3q26.[82, 89, 90] Studies on murine models revealed that aberrant expression of EVI1 alone may not be sufficient to cause leukemogenesis[91, 92], indicating additional genetic alterations are required in order to develop leukemia. Mutations in RAS, *RUNX1*, *SF3B1,* and genes encoding epigenetic modifiers recurrently co-occur with the inv(3)/t(3;3) chromosomal aberration.[93, 94] Overexpression of *EVI1* is significantly associated with monosomy 7 and 11q23 rearrangements in non-3q translocations AML.[81, 82, 95] Application of functional genomics and genome engineering on AML samples with

inv(3)/t(3;3) demonstrated that 3q rearrangements play a role in the relocation of a distal *GATA2* enhancer to ectopically activate *EVI1* and simultaneously confer functional haploinsufficiency on *GATA2*.[96, 97] This illustrated that the inv(3) relocates a single enhancer causing a malfunction of two unrelated distal genes in this type of AML. The aberrant expression of *EVI1* by other 3q rearrangements remains elusive. AML patients with high *EVI1* expression, like in AML with 3q26 rearrangements, independently predicts for a worse outcome.[81, 82, 95]

Four base pair insertion mutations in the *NPM1* gene denote the most prevalent AML defining molecular lesion identified to date, detected in a third of cases, including more than 50% of those with CN-AML. More than 40 different *NPM1* mutations have been reported, where types A, B, and D together represent approximately 90% of the cases. Mutations in *NPM1* involve the C-terminal segment of the protein and result in loss of tryptophan residues and formation of a nuclear export signal resulting in delocalization of nucleophosmin from the nucleoli to the cytoplasm. Experiment on murine model has shown that mutant *NPM1* (termed NPM1c in murine) can promote the self-renewal capacity of the hematopoietic progenitors, concomitant with expanded myelopoiesis resulting in the development of AML[98], with *FLT3*-ITD significantly decreasing the latency and enhancing leukemic phenotype penetration[99]. Young adults with AML harboring *NPM1* mutations without *FLT3*-ITD (or low *FLT3*-ITD burden) have a favorable prognosis[100]; whereas, concurrence with (high) *FLT3*-ITD (burden) and/or mutated DNA methyltransferase 3A (*DNMT3A*) have been associated with increased risk of relapse and inferior outcome[101].

The discovery of aberrant protein tyrosine kinase signaling pathways in several human cancer resulted in a significant diagnostic and therapeutic advances in recent years.[102] A typical example is FMS-like tyrosine kinase 3 (*FLT3*), a class three tyrosine kinase receptor that is expressed on the surface of normal hematopoietic lineages and blast cells from many cases of AML.[103] It is constitutively activated in several cases of AML because of mutation(s) in the *FLT3* gene.[104-106] Mutations in the *FLT3* gene are one of the most frequently harbored mutations in AML. These are detected in a third of AML patients and *FLT3* mutations are divided into two types according to the region affected in the gene and the context in which it occurs. The first type involves an in-frame duplication within the juxtamembrane region (*FLT3*-ITD) and found in approximately 25% - 30% of the cases. The second type is point mutations in the tyrosine kinase domain (*FLT3*-TKD) and found in 5 to 7% of the cases. Several studies demonstrated *FLT3*-ITD as an independent predictor for an elevated relapse rate and inferior overall survival in cases with a high mutant allele burden. [107] Whereas, patients with *FLT3*-TKD mutation have been found to perform well.[108] It is interesting how these two classes of mutations on the same gene are associated with such markedly different outcomes. This may reflect variation in the arrays of accompanying cooperative mutations, as well as variation in downstream signaling pathways of *FLT3* affecting disease biology.[34]

Following studies have recognized several other genes encoding signaling pathway

elements, such as *RAS*, *KIT*, *CBL*, *NF1,* and *PTPN11,* as frequently mutated targets in AML. While RAS mutations are considered prognostically neutral[109], multiple studies have demonstrated that the presence of *KIT* mutations and CBF leukemia in a patient predicts for indication of a worse response[110]. The fact that mutations in the signaling protein frequently co-occur with chromosomal translocations that affect hematopoietic transcription factors (for example *RUNX1-RUNX1T1*, *CBFB-MYH11,* and *PML-RARA*) instigated the proposal of a two-hit concept of AML, involving the cooperation of mutations bestowing a proliferative advantage (named class I mutations) and those anticipated to cause a block in myeloid differentiation (named class II mutations).[111] However, based on studies on the AML genome, it became apparent that the state is more complex because about 40% of AML patients have no mutations in genes encoding classical signaling pathway components.[59]

TP53 is a tumor suppressor gene that has a vital role in the regulation of the cell cycle in response to cellular stress. *TP53* mutations are more common in secondary AML (approximately 20 to 25%) than in *de novo* AML (about 2 to 9%)[112-115] and are frequently found in association with complex karyotype but seldom with *CEBPA*, *NPM1*, *FLT3*-ITD, and *RUNX1* mutations.[113] Moreover, mutations in *TP53* are associated with therapy-related AML, chemoresistance, high relapse rate, and poor response.[34, 116, 117] Generally speaking, AML patients with *TP53* mutations have inferior outcome regardless of other predictive factors for instance complex karyotype.[112, 113]

Several recurrent mutations in epigenetic regulators genes have been recently proposed to be associated with clinical outcomes in AML, including *DNMT3A*, *TET2*, *ASXL1,* and *EZH2*.[34, 59, 118, 119] Recurrent mutations in the *DNMT3A* gene occur in approximately 14 to 18%, including 20 to 35% with CN and associated with inferior outcome.[120-126] However, in a study of the German-Austrian AML study group where they investigated 1494 samples of 181 *DNMT3A* mutatant AMLs revealed no association with risk of relapse and OS.[127] Mutations in *DNMT3A*, *NPM1*, *FLT3,* and *IDH1/2* are frequently found to coexist[120, 124], thus inferring cooperativity in AML pathogenesis. Numerous various loss-of-function mutations have been identified in all *DNMT3A* exons, most commonly a single base substitution hotspot mutation encoding arginine at codon 882 (R882). *DNMT3A* is a methyltransferase that transforms cytosine to 5-methylcytosine and produces *de novo* DNA methylation. Abnormal methylation patterns have been associated with tumorigenesis and tumor evolution.[128] The biological implications of *DNMT3A* mutations in leukemia are not clearly understood. It is postulated that mutations cause a reduction in methyltransferase activity of *DNMT3A* by a dominant-negative mechanism.[120, 129] Recent research *in vivo* showed that Dnmt3a-null hematopoietic stem cells have an elevated self-renewal propensity and forfeit their differentiation ability, which was accompanied by hypomethylation of many genes involved in leukemogenesis.[130] In the absence of Dnmt3a, however, pronounced focal hypermethylation has been observed and it thus remains ambiguous which biological function of Dnmt3a contributes to the Dnmt3a-knockout phenotype. Intriguingly, the knockout of *Dnmt3a* is insufficient to cause leukemia in mice.[130]

Mutations in the *TET2* gene have been identified in several myeloid malignancies and occur in 10 to 20% of AML.[131-134] *TET2* mutations are very diverse and scattered across the whole coding sequence.[131-133] Detection of mutations in both *TET2* alleles and loss of heterozygosity in various myeloid malignancies imply a role of *TET2* as a tumor suppressor gene.[135-137] *In vitro* and *in vivo* researches revealed a role for *TET2* in stem and progenitor cells self-renewal and myeloid differentiation[134, 138, 139], but the mechanisms and the downstream impacts of *TET2* remain elusive. The TET2 protein is an enzyme that is involved in the conversion of 5-methylcytosine that causes demethylation of DNA[140] and this enzymatic function has been demonstrated to be impaired when the *TET2* gene is mutated[138]. The prognostic value of *TET2* remains controversial, some studies found an adverse outcome in certain AML subgroups[131, 132], while other studies found no prognostic value[141].

The *ASXL1* gene encodes a chromatin-binding protein involved in the regulation of chromatin remodeling and histone methylation[142, 143], and reported to be mutated in a broad variety of myeloid malignancies, including MDS, MPN, chronic myelomonocytic leukemia (CMML) and AML[144]. In AML, *ASXL1* mutations are found in around 3 to 5%[112, 122, 145] with higher prevalence in patients with intermediate-risk AML including CN-AML (i.e., 11 to 17%). Approximately 15 to 25% of MDS patients harbor mutations in the *ASXL1* gene,[146, 147] *ASXL1* mutations are rarely seen in children (nearly 1%), however, their incidence increases with age especially in patients 60 years or older[115, 147, 148], and the high frequency is also seen in patients with secondary AML (i.e., 30%)[144, 149]. Mutations in the *ASXL1* gene are almost exclusively found in exon 12 and virtually all are heterozygous.[144, 150, 151] They are mainly frameshift and stop mutations that are predicted to result in truncation of the carboxy-terminal plant homeodomain (PHD) finger at the protein level[152], suggesting these mutations are loss-of-function mutations[153]. *ASXL1* mutations frequently coexist with *EZH2*[154], *IDH1/2*, *RUNX1,* and *TET2*[155, 156]; conversely, they are mutually exclusive with *NPM1* mutations[151, 157]. Mutations in *ASXL1* are associated with worse outcome.[113, 114, 151, 155, 158, 159]

IDH mutations of the gene encoding isocitrate dehydrogenase are found in 8 to 16% of cases that harbor *IDH1* with CN-AML[160], and in 12 to 15% of cases with *IDH2*. These mutations are mutually exclusive, however, in rare instances, both genes are mutated in AML.[160-162] Both *IDH1* and *IDH2* convert isocitrate to α-ketoglutarate; conversely, the mutated proteins possess a gain-of-function that results in abnormal upturn of oncometabolite 2-hydroxyglutarate (2-HG).[162, 163] The function of IDH1/2 overlaps with TET2 and consequently these two mutation sets are also mutually exclusive in AML.[131, 132, 134] The epigenetic signature and global DNA hypermethylation of patients with IDH1/2 and TET2 are alike.[134] In addition, it was revealed that the oncometabolite 2-HG which is produced as a consequence of mutation in IDH1/2 hinders TET2 function.[134] Initial studies revealed that *IDH* mutations have an adverse response, however, recent large cohort studies showed that outcomes differ depending on the different sites of the *IDH* mutations and the additional co-occurring mutations.[145, 164] Various studies demonstrated that *IDH1* and *IDH2*^{R172} mutation

may predict a poorer clinical prognosis particularly in CN-AML, while, *IDH2*^{R140} found with *NPM1* mutation may be associated with favorable outcome in AML.[164, 165]

Mutations in splicing factor genes are found in 10% of patients with AML.[166, 167] Spliceosome genes encode elements of the splicing machinery e.g., SF3B1, SRSF2, U2AF1 and ZRSR2, which are involved in pre-messenger RNA (mRNA) processing, before protein translation, and are frequently mutated in MDS as well.[168] They can induce aberrant splicing, impacting the cell's transcriptome and proteome. Complying evidence demonstrates that spliceosome mutations are connected with older age, less proliferative disease, and poor response rate to standard treatment, and worse survival.[166] Splicing factor gene mutations are found to be mutually exclusive and detected in about 50% of MDS cases, where they are regarded to be the initiating mutation and segregate with a distinct subcategory of disease.[72] Primarily, *SF3B1* mutations are detected with the presence of ring sideroblasts,[167] and *SRSF2* mutations are observed in more severe disease and CMML[168, 169]. Recent study on a well-characterized cohort of AML patients revealed that mutation of *SRSF2*, *SF3B1*, *U2AF1,* or *ZRSR2* were among a panel of mutated genes that could be labeled as pathognomonic of secondary AML, developing on a background of MDS.[166]

Application of whole-genome sequencing (WGS) and whole-exome sequencing (WES) have contributed to the identification of recurrent mutations in AML in genes encoding members of the cohesin complex, including SMC1A, SMC3, RAD21, and STAG1/2.[59, 170] These proteins have important roles in DNA repair and looping.[170-172] Mutations in cohesin complex are detected in about 6% of *de novo* AML with no impact on OS. Additionally, they are found in 10 to 15% of MDS and in 20% of secondary AML patients and frequently associated with mutations comprising *RUNX1*, *BCOR,* and *ASXL1* but mutually exclusive with *NPM1* mutations.[171, 172]

Table 4: Prevalence, association, and prognosis of mutations detected in young adults with AML

Category/mutant gene	%	Associated mutations/chromosomal abnormalities	Prognostic implications
DNA methylation			
DNMT3A	20	*NPM1*, *FLT3*-ITD, NK	Adverse
DNA demethylation			
TET2	8	*ASXL1*, NK	Poorer in IR-AML
IDH1	7	*NPM1*, NK	Poorer in *FLT3*-ITD-neg AML
IDH2-R140	7	*NPM1*, NK	Favorable
IDH2-R172	2	NK	Adverse
WT1	9	*FLT3*-ITD	Poorer in NK-AML
Activated signaling			
FLT3-ITD	27	*NPM1*, *DNMT3A*, NK, t(15;17)/*PML-RARA*, t(6;9)/*DEK-NUP214*, t(5;11)/*NUP98-NSD1*	Poorer in IR-AML
FLT3-TKD	7	inv(16)/CBFB-MYH11, t(15;17)/PML-RARA	Variable according to study
NRAS	11	inv(16)/*CBFB-MYH11*, t(3;5)/*NPM1-MLF1*, 11q23/*MLL-X*	NS
KRAS	5	inv(16)/*CBFB-MYH11*, 11q23/*MLL-X*	NS
PTPN11	5	*NPM1*	ND
*NF1**	4	MK, -17/17q	ND
KIT	4	t(8;21)/*RUNX1/RUNX1T1*, inv(16)/*CBFB-MYH11*	Poorer outcome in CBF AML
CBL	1		TBC
Myeloid transcription factors			
RUNX1	5	*MLL*-PTD, *ASXL1, IDH2,* NK, +13	Adverse
bi*CEBPA*	4	*GATA2*, NK	Favorable
Tumor Suppressor /multifactorial			
*TP53**	8	Complex, MK, -5/-5q, -7/-7q, -17/17p	Adverse
NPM1	33	*DNMT3A*, *IDH1*, *IDH2*-R140, *FLT3*-ITD, *PTPN11*, cohesin, NK	Favorable in the absence of *FLT3*-ITD and mutant *DNMT3A*
Chromatin regulation			
ASXL1	5	*RUNX1*, *IDH2*-R140, t(8;21)/*RUNX1-RUNX1T1*, +8	Poorer in IR-AML
MLL-PTD	5	+11, NK, *RUNX1*, *FLT3*	Adverse
PHF6	3	*RUNX1*	TBC
ASXL2	2	t(8;21)/*RUNX1-RUNX1T1*	ND
BCOR	1	NK, *DNMT3A*	TBC
EZH2	1		ND
Spliceosome			
SRSF2	2	+13, *ASXL1, RUNX1, IDH1/2*	ND

Category/mutant gene	%	Associated mutations/chromosomal abnormalities	Prognostic implications
SF3B1	3	*RUNX1*, inv(3)/*GATA2-EVI1*	
U2AF1	2		
ZRSR2	<1		
Cohesin			
RAD21	6-9	*NPM1*	NS
SMC1A			
SMC3			
STAG1			
STAG2			

Mutations within the same functional category are negatively associated and positive associations are listed.
bi: biallelic; IR: intermediate risk; ITD: internal tandem duplication; MK: monosomal karyotype; ND: not determined; NK: normal karyotype; NS: not significant; PTD: partial tandem duplication; TBC: to be confirmed; TKD: tyrosine kinase domain
* Includes mutations and gene deletions.
(Adapted from Grimwade *et al*. [34])

Secondary AML

As stated earlier, s-AML alludes to AML that has emerged from the perspective of antecedent myeloid malignancy (e.g., MDS or MPN). The distinction between secondary AML from *de novo* AML presented in 1997 WHO classification of hematopoietic and lymphoid neoplasms, establishing a classification system for AML that comprises the following subcategorize: AML with recurrent genetic abnormalities, AML with multilineage dysplasia, therapy-related AML and MDS and AML not otherwise categorized.[173] The WHO 2016 classification defines AML with myelodysplastic-related alterations as a diagnosis of AML with a prior six-month history of MDS or MDS/MPN or demonstrated at least 50% dysplasia in two or more myeloid lineages[23, 24, 26]; however, cases harboring an *NPM1* mutation or biallelic *CEBPA* mutation are exempted[174, 175]. In the absence of strong morphological evidence of dysplasia, specific MDS-related cytogenetic aberrations can denote MDS.[23, 24] Recent study by Lindsley and colleagues suggests that patients with secondary-type AML with inferior responses to cytotoxic therapy and outcome, may be identified by detection of mutations involving the spliceosome machinery and chromatin remodeling.[166] However, this has not been integrated into the diagnostic criteria of the WHO.[166]

s-AML occurs more frequently in elderly individuals than in *de novo* AML, which has a low prevalence during life.[176] It has been found by large population-based studies that s-AML accounts for approximately 18 to 20% of all AML cases.[177, 178] Although the risk of progressing to AML varies greatly according to the specific myeloid disease, all myeloid malignancies are associated with an elevated risk for AML. In MDS, patients with excess blasts have a forecasted 25% and 35% risk for AML at 1 and 2 years, respectively. Whereas, the low-risk MDS have 5% and 10% transformation probability over comparable time periods. In MPNs, Primary Myelofibrosis is more likely to transform (6 to 21% at 5 years and about 20% at 10 years) compared to Polycythemia Vera (2% at 10 years and 8% at 20 years) and Essential Thrombocythemia (about 2% at 15 years); indicating that MPNs have variable leukemic transformations risk.[179-183] The MDS/MPN overlap syndromes may pose a significant risk for leukemic transformation, with a 20% and a 40% risk ascribed to chronic myelomonocytic leukemia and an atypical CML, respectively.[184, 185] Leukemic transformation can be displayed in bone marrow failure syndromes, such as Aplastic anemia and Fanconi anemia.[186-190]

The biology of s-AML is distinct from *de novo* AML with many key differentiating attributes such as it is equated with multilineage dysplasia, it frequently harbors complex karyotype with the regular loss of genetic material, it emerges in a clonal hematopoiesis context and it develops as a byproduct of the progressive genetic lesion. In contrast, *de novo* AML is not associated with multilineage dysplasia, but frequently features recurrent balanced translocations and inversions.[176] While *de novo* disease is believed to be an outcome of an inciting genomic event that results in expansion of the leukemic clone, AML originating from MDS or MPN typically progresses in a stepwise manner with multiple hits accumulating with time. Consequently, patients with s-AML tend to have more mutated genes than patients with

de novo AML. Mutations in *SRSF2*, *SF3B1*, *U2AF1*, *ZRSR2*, *ASXL1*, *EZH2*, *BCOR,* and *STAG2* are considered very specific for s-AML. However, there are additional mutated genes that have been implicated in the leukemogenic process such as mutations in signaling pathway (i.e., *FLT3*, *RAS*, *CBL*, *PTPN11*), transcription factors (i.e., *RUNX1*, *WT1*), epigenetic regulators (i.e., *IDH1/2*, *TET2*) and *TP53*. The founding clone is eventually outcompeted by an advantageous subclone, nevertheless evidence of the founder clone consistently persists.[166, 191-197]

Therapy-related AML

Therapy-related AML (t-AML) associated with previous exposure to leukemogenic agents can be classified into two groups with distinctive cytogenetic and morphological features. The first group is associated with exposure to alkylating agents (e.g., cyclophosphamide) therapy or actinotherapy. It is characterized by long latency between exposure to a leukemogenic agent and the development of an overt AML. The progression to AML can take from 5 to 10 years from the start of the exposure. These patients often present with MDS features (e.g., dysplastic changes similar to MDS but with a higher grade of dysplasia). Cytogenetic findings include deletion of the long arm of chromosome 5 or 7 (-5q and -7q, respectively), monosomy of chromosome 5 or 7 (-5 and -7, respectively), or complex karyotype. *TP53* gene is recurrently mutated in this group of AML as well. The second group is a distinct group of therapy-induced AML which comprises patients who were treated with topoisomerase 2 inhibitors (e.g., etoposide, mitoxantrone). In contrast to the first group, patients have relatively short latency (2 to 3 years), do not present with MDS characteristics, and are often associated with balanced chromosomal translocations involving 11q23 (*KMT2A*) or 21q22 (*RUNX1*). Topoisomerase 2 inhibitors have been additionally implicated in more sporadic forms of t-AML, including CBF t-AML and therapy-related APL. The former exhibits rearrangements at CBF genes *RUNX1* at 21q22 and *CBFB* at 16q22 and, similarly to CBF *de novo* AML, respond well to intensive chemotherapy regimens. The latter presents balanced chromosomal translocation between chromosomes 15 and 17 (i.e., t(15;17)) and responds favorably to all-trans retinoic acid-based therapies. Therefore, the spectrum of cytogenetic aberrations in t-AML is similar to that of *de novo* AML; however, the incidence of abnormalities associated with unfavorable and intermediate-risk is higher in t-AML. [22, 198-201] Host propensity plays a crucial role in t-AML; it has been found that many patients with t-AML have germline mutations in cancer predisposition genes.[202-204] Inter-individual heterogeneity in drug metabolism has been implicated.[200, 205] Similar to s-AML, clonal hematopoiesis may play an essential role, particularly in the group of t-AML in which cytotoxic treatment imposes a selective pressure on a minor chemoresistant preleukemic clone, inducing clonal expansion. To that end, studies have shown that mutant clones that harbor mutant *TP53* can be detected in hematopoietic cells years' prior to cytotoxic therapy and eventually pursued by *TP53* mutant t-MDS/t-AML.[206, 207] Thus, the manifestation of clonal hematopoiesis prior to a cytotoxic therapy is much more prevalent in patients who ultimately develop therapy-related myeloid neoplasms.[208, 209]

Treatment

The main objective of AML treatment is to induce remission and prevent relapse. Remission is defined morphologically by the presence of less than 5% blast in BM accompanied by the recovery of PB counts. Treatment of AML consists of two phases: the induction phase to achieve complete remission (CR) and followed by post-induction phase (consolidation therapy), which aims to retain long-lasting CR and prevent relapse. In the induction phase, daunorubicin and cytarabine have been the backbone of AML treatment for more than three decades. With induction therapy, CR can be achieved in 60 to 80 % of patients who are 60 years of age or younger and in about 40 to 60 % of older patients.[22, 210]

To maintain patients in CR, further intensive treatment is required to prevent relapse. Young patients have two options of consolidation therapy: chemotherapy or stem cell transplantation (SCT). SCT is subdivided into two types: allogeneic-SCT ((allo-SCT); which is usually from an HLA of either matched-related or unrelated donor), autologous-SCT (auto-SCT). The choice of post-induction therapy is based on patient AML risk group, response to therapy, and availability of a donor. Treatment choices for older patients are limited because of the high rate of treatment related-toxicity.[18, 22] However, being older should not be the reason for not treat with intensive therapy; evaluating risk factors such as disease-related prognostic factors and comorbidities against treatment intensity is essential.[211] In general, unfit AML patients, regardless of age, will not tolerate intensive chemotherapy; therefore, these patients are limited to supportive care, low-intensity treatment, or clinical trials with investigational drugs.[27]

Clonal Evolution

The advent of high-throughput sequencing has revolutionized and facilitated the study of malignancy and clonal evolution. The sequencing of hundreds of leukemia samples allowed the identification of the recurrent mutations associated with each subset of the disease and the tracking of these mutations over time, and the inference of clonal evolution patterns.[17, 59, 212-215] In AML, NGS studies have also indirectly documented the stepwise acquisition of mutation by demonstrating the difference in the relative proportion of concurring mutations within the bulk tumor at diagnosis. The achievement of greater sequencing depth has enabled these differences to be quantified by variant allele fraction (VAF), where the proportion of reads containing the mutated allele is compared to that of the wild-type allele reads, with the relative proportions capable of inferring clonal architecture. [34] By their VAF, mutations can be classified into clonal (present in the major clone; dominant clone) or subclonal (present in a minor clone), indicating how many cells in a sample carry a specific variant.[216] Such analysis has illustrated the appearance of new clones harboring novel mutations at different times during leukemia evolution.[34]

Several studies have described the genetic and epigenetic evolution of AML from diagnosis to relapse showing that the molecular profile of AML changes during the

disease.[214, 215, 217, 218] It is believed that AML arises from a single HSPC, which with time, acquires somatic mutations that cause a block of differentiation and provide stem cell-like properties of limitless self-renewal capacity.[219, 220] Prior to overt leukemia, a mutation in epigenetic modifiers such as DNMT3A, TET2 and ASXL1 commonly acquired that may provide a growth advantage, nevertheless; they are insufficient to initiate leukemia and are thus widely referred as pre-leukemic.[221-223] Leukemia-initiating mutations are frequently found in genes associated with signaling activation, such as *FLT3* and the gene nucleophosmin (*NPM1*).[59, 213] Individual mutation adds up to the genetic complexity, and ultimately, increased clonal heterogeneity is associated with poor outcome of AML.[224, 225]

During the disease course, each AML population may follow distinct patterns of clonal evolution. The manifestation and the profusion of mutations at various time points portray a picture of dynamic changes. The stepwise acquisition of single mutations is termed linear evolution; whereas, the elimination of the dominant clone and outgrowth of a subclone is known as branching evolution (Figure 2). Cell populations are likely to pursue a linear evolution, gradually increasing in their fitness, when there is no change in the evolutionary pressure. Branching evolution may follow nonetheless linear evolution, or vice versa, mainly when there have been profound alterations in the evolutionary pressure (e.g., at treatment induction or change of therapy). During treatment, the AML clone may evolve by either acquiring additional mutations that mediate resistance to therapy (Figure 2A) or by losing mutations that are, for instance, associated with treatment sensitivity (Figure 2B).

In summary, at relapse, the AML cell population might have evolved from clonal or subclonal cell populations that exist at baseline, convoyed by a prospective gain of additional mutations. Pursuing either linear or branching evolution, cell populations undergo gradual clonal evolution to best adapt to their environment. Since most AML patients relapse following initial response to chemotherapy, few AML cell populations in many patients find a niche to evade chemotherapy and eventually raise out again.[216]

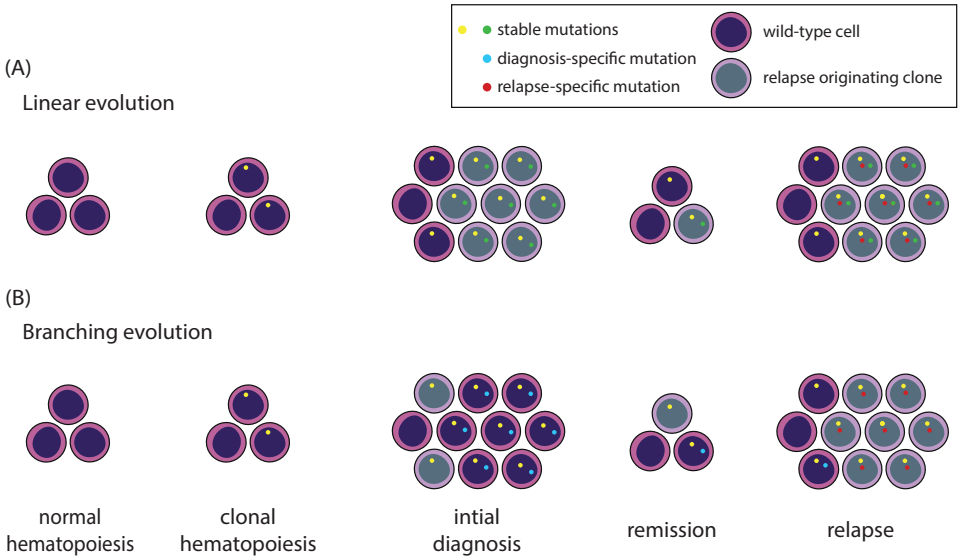


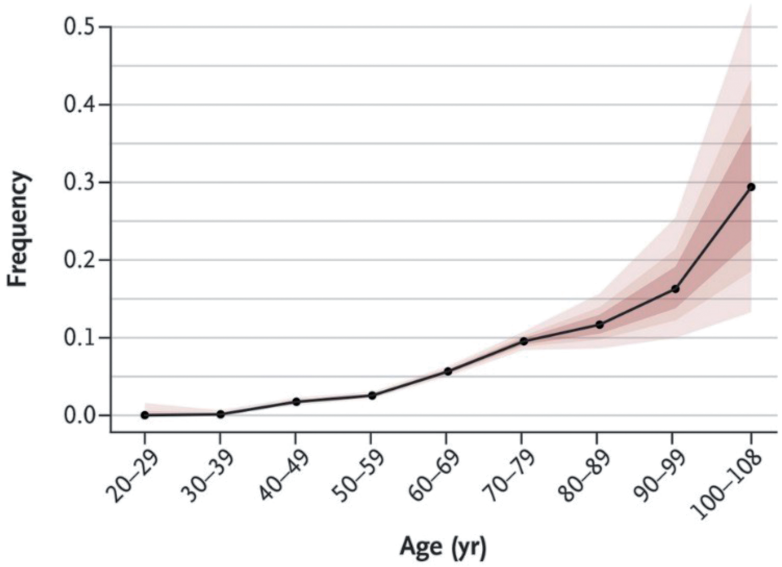
Figure 2: Models of clonal evolution. The schematic diagram illustrates the models of the clonal evolution of AML over time. (A) Linear clonal evolution depicts the sequential gaining of mutations. The relapse initiating cells are a fraction of the major clone at diagnosis. (B) Branching clonal evolution illustrates the elimination of the major clone and is followed by the growth of a secondary clone. Relapse caused by a subclone at diagnosis. The cells without a marker(s) (small colored circle in the nucleus of cells) indicate wild-type cells without somatic mutations. The different small colored circles in the nucleus of cells indicate somatic mutations; colored based on their stability: green and yellow: stable mutations; blue: diagnosis-specific mutation; red: relapse-specific mutation. Cells of different colors with and without somatic mutations represent different clones, and the number of colored cells represents the size of the clone. Cells that originate from the relapse clone are colored differently. (Adapted from Vosberg and Greif [216])

Clonal Hematopoiesis

The manifestation of initiating mutations that result in clonal expansion and the thus premalignant state has long been suspected of preceding the development of most malignancies. Recently, exome sequencing of PB samples from approximately 30,000 healthy individuals without known hematological neoplasms showed recurrent somatic mutations in myeloid malignancy-associated genes in up to 10% of individuals above the age of 65 and more than 20% of individuals over the age of 90 years. This phenomenon is known as clonal hematopoiesis of indeterminate potential (CHIP).[226] CHIP is a term used to describe the presence of a somatic mutation in the white blood cells of apparently healthy individuals with no diagnostic features of hematological malignancies. The somatic mutations occur in leukemia-associated driver genes causing clonal expansion of a genetically identical clone of marrow and blood cells. The majority of patients with CHIP are healthy individuals (Figure 3).[227] Consequently, it is also known as age-related clonal hematopoiesis (ARCH). Individuals with CHIP have a slightly high risk of developing myeloid malignancy. Besides, they have a relatively elevated risk of developing atherosclerosis and cardiovascular disorder. However, a subset of individuals with CHIP does not develop malignancy or cardiovascular disease.[227]

The most frequent recurrent somatic mutations observed in individuals with CHIP occurs in proteins that play a role in DNA methylation or its regulation primarily in *DNMT3A*, *TET2,* and *ASXL1* (Figure 4).[222, 223] These mutations are commonly seen in patients with MDS and AML.[226] Mutation in these proteins accounts for more than 90% of ARCH suggesting that these mutations provide a selective advantage to hematopoietic stem and progenitor cells (HSPCs).[228] Additionally, these mutated HSPCs are still capable of differentiating into granulocytes, monocytes, and lymphocytes.[229, 230] It remains unclear why these mutations are associated with enhanced HSPCs fitness.[228]

Individuals with CHIP have a yearly risk of about 0.5 to 1% for developing myeloid malignancy compared with age-matched control, indicating that the presence of somatic mutations in these genes is not sufficient to cause malignant transformation. Therefore, these mutations can be regarded as initiating events (or pre-leukemia state) that prime cells towards progression to malignancy. Furthermore, this indicates that additional mutations are required to develop overt hematological malignancy.[223]



No. with Mutation	0	1	50	138	282	219	37	14	5
Total	240	855	2894	5441	5002	2300	317	86	17

Figure 3: Prevalence of Somatic Mutations, according to age. Colored bands, in increasingly lighter shades, represent the 50th, 75th, and 95th percentiles. (Adapted from Jaiswal and colleagues[223])

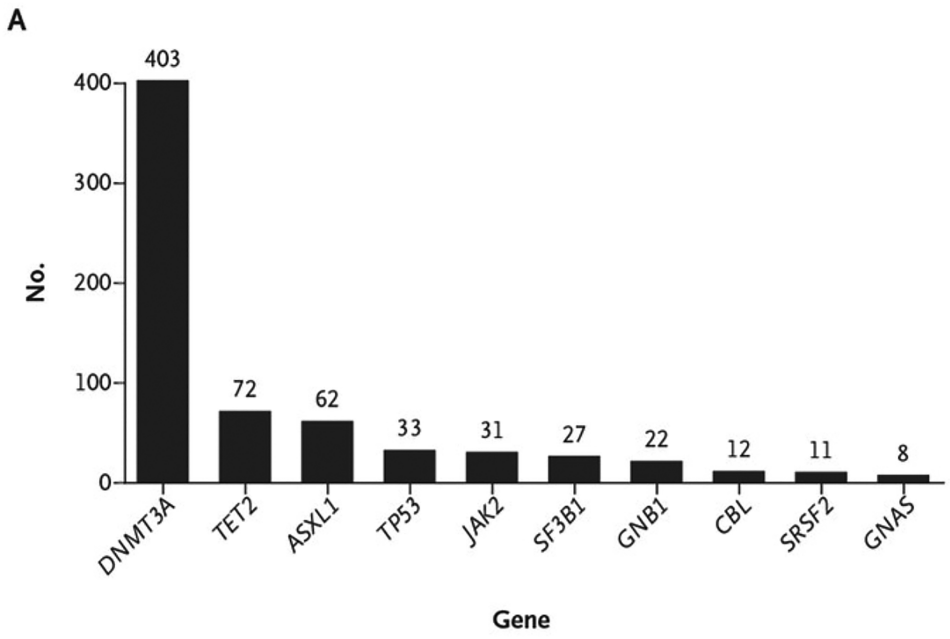


Figure 4: Candidate somatic variant. The bar chart shows the 10 most recurrently mutated genes associated with hematologic cancers. (Adapted from Jaiswal and colleagues[223])

Minimal/Measurable Residual Disease Detection in AML

Despite profound advancement in the therapy of AML, the outcome is still dismal only 30 - 40% of adult patients are curable.[231] Although the majority of patients achieve CR after initial therapy, the relapse rate remains high. [211] For sixty years, early treatment response was judged by the cytomorphological appearance of the BM by means of blast level predicts relapse risk. The detection of five percent blasts or more after achieving morphological CR is associated with a high risk. A substantial fraction of the patients in morphological remission will relapse within five years and die of the disease.[232] The current method to assess patient prognosis classify the patient into three categories and relies on a variety of factors at diagnosis but mainly on cytogenetic and molecular genetic abnormalities.[27, 231] However, these factors cannot predict patient outcome perfectly. These facts have paved the path for the search for more accurate tools to predict which patient will relapse based on the detection of submicroscopic numbers of the leukemic blast that may have the potential to cause the recurrence of leukemia.

In the last two decades, there was a considerable development in tools that can detect submicroscopic remnants of the leukemic blast, including multicolor flow cytometry (MFC), polymerase chain reaction (PCR) techniques, and more recently, next-generation sequencing (NGS). These tools enable the detection of measurable (previously known as minimal) residual disease (MRD) to as low as one leukemic cell in a million non-leukemic cells, which is way past the resolution of the microscope.[232]

MFC

MFC is a rapid method that focuses on the immunophenotype of the leukemic to profile the lineage of the cells. It can be used to determine MRD in the majority of AML patients irrespective of the immunophenotypic, cytogenetic, and molecular genetic aberrations. Thus, it nowadays is the most widely used technique. MRD detection using MFC relies on finding inherent immunophenotypic features of the leukemic cells at the time of diagnosis. This can be done using two different methods. The first MFC approach depends on determining the leukemia-associated immunophenotype (LAIP) at diagnosis and is later used for tracking residual disease during follow-up. This method is applicable to 90 - 95% of the patients. The second approach relies on identifying “different from normal” phenotype, which can be used on the remaining patients or when the diagnostic phenotype is inaccessible. The latter approach is essential because phenotypic abnormalities such as gains or losses of antigens due to disease evolution as compared to normal progenitors could be systematically investigated.

The drawbacks of using MFC for MRD detection are: 1) low sensitivity compared to quantitative PCR (qPCR) (Figure 5); 2) phenotype change is quite often (for at least one antigen), which makes them difficult to detect/track leading to false-negative results; 3) requires a considerable degree of expertise and experience; analysis and data interpretation

have some subjective elements and hence, likely biases (operator-dependent) making it challenging to harmonize the assays across laboratories.[232, 233]

RQ-PCR

The most broadly exploited molecular MRD technique is qPCR. It is used for the detection of chromosomal fusion genes or mutation-specific sequences in nucleic acid derived from PB or BM samples. In general, this methodology is well-established and very sensitive; however, due to the molecular heterogeneity of AML, the application of PCR-based MRD assays is limited to a few molecular subsets of AML patients. This MRD approach is most applicable in AML cases harboring recurrent chromosomal rearrangements such as t(15;17)(q22;q21), t(8;21)(q22;q22.1) or inv(16)(p13.1q22) or patients with a mutation in *NPM1*. These genetic aberrations may be tracked longitudinally using qPCR.[231] However, altogether, these cases represent around 40-50% of all AML cases less than 60 years of age.[233, 234]

NGS

Since many patients, especially elderly patients, do not present with any of the established molecular MRD markers, techniques other than qPCR are needed.[231, 235] NGS is an emerging new technique and offers several advantages over other MRD detection techniques.[231] NGS can be used to detect patient-specific mutation covering the entire genes in virtually all AML patients[236], and due to its ability to do massive parallel sequencing, several samples can be multiplexed simultaneously in a single experiment while testing multiple regions of interest in the genome which makes it more attractive tool for high throughput setting. Thus, NGS is an attractive technology for advancing MRD assessment.[231] It has been shown recently that molecular MRD detection by NGS is applicable to the vast majority of all newly diagnosed AML patients.[237-240] However, NGS-based MRD detection in AML has to overcome several drawbacks before it can be reliably used in routine clinical practice. The technical difficulties for NGS-based MRD detection are owed to the low sensitivity of NGS (Figure 5). Currently, NGS can be reliably used to detect mutations of ≥0.1%. Additionally, the technology has a high intrinsic error rate (0.1 to 1%) and thus may cloud the detection of minor subclones and discriminating a true mutation from PCR and sequencing artifacts. Altogether, these issues hinder the clinical application of NGS MRD tracking during therapy.[231]

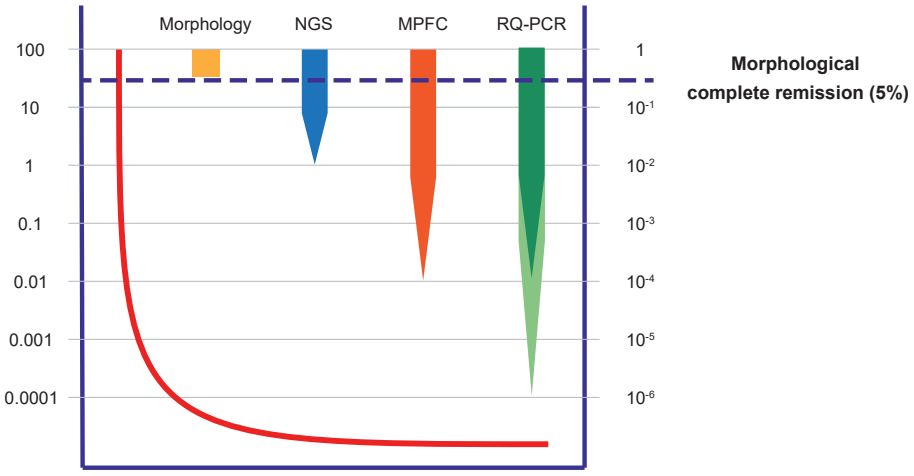


Figure 5: Detection thresholds of various MRD approaches compared to the classical clinical complete remission. An expert hematopathologist, at best, can detect a 2-log reduction in disease burden (1% limit of detection) using microscopic examination of morphology or NGS. Whereas, MPFC and real-time quantitative PCR (RQ-PCR) for overexpressed genes (for instance, *WT1*) are two log more sensitive than microscopic examination (enable detection of 1 abnormal cell in a population of 10,000 cells). RQ-PCR for fusion transcripts (such as CBF AMLs) and hotspot/specific mutations (e.g., *NPM1* mutations) by RQ-PCR can detect residual disease with about 10,000-fold more sensitive than the classical CR criteria. Axis scales are approximate and merely for illustrative purposes. Adapted from Hourigan and Karp[241].

Scope of the thesis

AML is a heterogeneous clonal disease of the hematopoietic progenitor cells that are characterized by a variety of cytogenetic and molecular genetic aberrations that play a role in the pathogenesis of the disease and clinical outcome upon treatment.

The work on this thesis is divided into two parts, all centered around in-depth molecular analyses of AML. The first part is focused on molecular characterization of AML subsets, the role of mutations in AML risk stratification and MRD detection by NGS. The second part provides an insight into a new germline syndrome leading to AML development.

Minimal/measurable residual disease (MRD) detection in AML is mostly done on a handful of AML with specific markers (i.e., *NPM1*, *RUNX1-RUNX1T1*, *CBFB-MYH11,* and *PML-RARA*) by quantitative real-time PCR. However, the majority of AML carry patient-specific mutation or combination of mutations in genes associated with AML pathogenesis that can be used for MRD detection. NGS is a powerful technique to detect mutations in multiple genes simultaneously, which makes it an ideal method to detect mutations in virtually all AML patients. Therefore, in **Chapter 2**, we demonstrated the use of NGS for MRD monitoring to predict impending relapse in 482 AML patients. We investigated the clinical outcome of identifying molecular MRD by NGS in AML patients at complete remission. We found that 51.4% of the cohort had persistence of mutation during CR and many of the persisting mutations were in genes associated with clonal hematopoiesis (i.e., *DNMT3A*, *TET2,* and *ASXL1*; abbreviated *DTA* mutations). Strikingly, the persistence of *DTA* mutations in CR were not associated with increased risk of relapse. However, we found that detection of non-*DTA* mutations in CR to be significantly associated with an elevated risk of relapse and overall survival, which was confirmed by multivariate analysis.

Mutant *TP53* AML is a specific subtype of AML with a dismal outcome. While *TP53* mutant AML generally receives an allogeneic-HSC transplant in first complete remission, relapse rates remain high. In-depth molecular characterization of the clonal architecture of mutant *TP53* AML may shed more light on the biology of mutant *TP53* AML, its response to treatment and the role of MRD detection in mutant *TP53* AML. In **Chapter 3**, we studied a large cohort of AML cases (i.e., 2200 patients) with the aim to study the interactions between the molecular and clinical characteristics of *TP53* mutant patients by meticulously dissecting all clonal features, including *TP53* mutation status, variant allele frequency, concurrent mutations, cytogenetics and molecular MRD. In this study we demonstrated distinct clonal features of *TP53* mutant AML and identified that genomic instability, which is characterized by complex cytogenetic, as surrogate for disease progression and serves as an independent predictor for patients' outcome. Moreover, our study revealed that genomically stable mutant *TP53* patients benefit from an allogeneic-HSC transplant and provides a rationale to identify new treatment approaches for residual disease detection and suitable treatments for genomically instable mutant *TP53* AML.

In **Chapter 4**, mutant *PPM1D* AML has been shown to be associated with clonal hematopoiesis in patients with previous history of malignancy. To determine the contribution of mutant *PPM1D* to AML relapse and the potential of using it as an MRD marker, we studied a large cohort of AML patients (n= 685) and determined the prevalence of *PPM1D* mutations in AML. Subsequently, we studied the kinetics of *PPM1D* mutant clone(s) from diagnosis to CR/relapse and our data indicate that *PPM1D* mutations are rare in AML and do not cause relapse in AML.

Sequential sampling of AML patients for molecular MRD is mostly available for selected AML patients (e.g., AML patients with mutant *NPM1*) and at limited sampling time points during the disease course. Consequently, this represents a hurdle to study the kinetic and the dynamic of AML relapse. However, MGG-stained BM slides are routinely and frequently done for all AML patients and archived in contemporary laboratories. In **Chapter 5**, we explored the feasibility of using DNA derived from archived May-Grünwald Giemsa stained BM (MGG-stained BM) for NGS-based mutation in 18 AML patients. This study revealed that mutations detection in DNA derived from MGG-stained BM slides is feasible and mutation burden is highly comparable to DNA derived from Ficoll purified mononuclear cells. We also demonstrated that MGG-stained BM slides can be used to study mutations kinetics and MRD detection. Thus, this makes MGG-stained BM slides a valuable source for NGS-based mutation analyses.

AML patients with *KMT2A*-PTD mutation are known to have an adverse outcome; furthermore, previous mouse studies showed that Kmt2a-ptd is insufficient to cause AML, suggesting that additional mutations are required to develop leukemia. To shed some light on the biology and clinical outcome of this subset of AML patients in the context of coocuring mutation, in **Chapter 6**, we investigated the prevalence of this AML subset by screening 1998 AML patients and found that this mutation is found in about 5.5% of all AML cases. In line with previous studies, we showed that the presence of *KMT2A*-PTD mutations was significantly associated with a concurrent trisomy of chromosome 11 as compared to the *KMT2A* wild-type of a well-characterized AML cohort (n=561) [hereafter referred to as reference cohort/*KMT2A* wild-type AML]. Next, we determined the molecular and gene expression profile of *KMT2A*-PTD AML and compared it to the *KMT2A* wild-type cohort and to translocation involving 11q23. Our study revealed that *KMT2A*-PTD AML has a specific gene expression signature and concomitant *DNMT3A* and *NRAS* mutations are associated with an adverse outcome.

The second part of this thesis provides some insight into the etiology of clonal hematopoiesis and successive development of AML. DNA methylation on cytosine residues is a significant mutagenic stimulus, as 5-methylcytosine (5mC) is vulnerable to spontaneous deamination to thymine. Studies in mice models revealed that inactivation of methyl-binding domain 4 (*Mbd4*) plays a vital role in the repair of methylation damage. We revealed 3 early-onset AML cases with germline loss of MBD4 and with a common set of driver mutations.

The results of this study are described in **Chapter 7** of this thesis. Our study demonstrated that AMLs with germline MBD4-deficient have high mutational burden than any other AMLs (about 33-fold over typical AML), with >95% being C>T in the context of a CpG dinucleotide. This distinct mutational signature is similar to what has been reported in sporadic cancers with acquired biallelic mutations in *MBD4* and *Mbd4* knockout mice. Our study confirmed that individuals with germline biallelic loss of MBD4 sustain high levels of damage from 5mC deamination throughout their lifetime and experience clonal expansions decades earlier, which leads to the development of AML. Finally, the findings reported in this work are summarized and discussed in **Chapter 8**.

REFERENCES

1. Jagannathan-Bogdan M, Zon LI. Hematopoiesis. *Development* 2013 Jun 15; **140**(12)**:** 2463-2467.
2. Rieger MA, Schroeder T. Hematopoiesis. *Csh Perspect Biol* 2012 Dec; **4**(12).
3. Ng AP, Alexander WS. Haematopoietic stem cells: past, present and future. *Cell Death Discovery* 2017; **3**.
4. Asada N. Regulation of Malignant Hematopoiesis by Bone Marrow Microenvironment. *Frontiers in Oncology* 2018 Apr 23; **8**.
5. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. *Nature* 2001 Nov 1; **414**(6859)**:** 105-111.
6. Kondo M. Lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. *Immunological Reviews* 2010 Nov; **238:** 37-46.
7. Kiel MJ, Yilmaz OH, Iwashita T, Yilmaz OH, Terhorst C, Morrison SJ. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* 2005 Jul 1; **121**(7)**:** 1109-1121.
8. Yang L, Bryder D, Adolfsson J, Nygren J, Mansson R, Sigvardsson M*, et al.* Identification of Lin(-)Sca1(+) kit(+)CD34(+)Flt3- short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. *Blood* 2005 Apr 01; **105**(7)**:** 2717-2723.
9. Christensen JL, Weissman IL. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc Natl Acad Sci U S A* 2001 Dec 04; **98**(25)**:** 14541-14546.
10. Morrison SJ, Wandycz AM, Hemmati HD, Wright DE, Weissman IL. Identification of a lineage of multipotent hematopoietic progenitors. *Development* 1997 May; **124**(10)**:** 1929-1939.
11. Morrison SJ, Weissman IL. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* 1994 Nov; **1**(8)**:** 661-673.
12. Kondo M, Weissman IL, Akashi K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* 1997 Nov 28; **91**(5)**:** 661-672.
13. Bryder D, Rossi DJ, Weissman IL. Hematopoietic stem cells - The paradigmatic tissue-specific stem cell. *American Journal of Pathology* 2006 Aug; **169**(2)**:** 338-346.
14. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* 2000 Mar 9; **404**(6774)**:** 193-197.
15. Blank U, Karlsson G, Karlsson S. Signaling pathways governing stem-cell fate. *Blood* 2008 Jan 15; **111**(2)**:** 492-503.
16. Cheson BD, Bennett JM, Kopecky KJ, Buchner T, Willman CL, Estey EH*, et al.* Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia. *J Clin Oncol* 2003 Dec 15; **21**(24)**:** 4642-4649.
17. Grove CS, Vassiliou GS. Acute myeloid leukaemia: a paradigm for the clonal evolution of cancer? *Disease Models & Mechanisms* 2014; **7**(8)**:** 941-951.
18. Estey E, Dohner H. Acute myeloid leukaemia. *Lancet* 2006 Nov 25; **368**(9550)**:** 1894-1907.
19. Deschler B, Lubbert M. Acute myeloid leukemia: Epidemiology and etiology. *Cancer* 2006 Nov 1; **107**(9)**:** 2099-2107.
20. de Rooij JDE, Zwaan CM, van den Heuvel-Eibrink M. Pediatric AML: From Biology to Clinical Management. *Journal of Clinical Medicine* 2015 Jan; **4**(1)**:** 127-149.
21. Estey EH. Acute myeloid leukemia: 2019 update on risk-stratification and management. *Am J Hematol* 2018 Oct; **93**(10)**:** 1267-1291.

22. Lowenberg B, Downing JR, Burnett A. Medical progress - Acute myeloid leukemia. *New England Journal of Medicine* 1999 Sep 30; **341**(14)**:** 1051-1062.
23. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A*, et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 2009 Jul 30; **114**(5)**:** 937-951.
24. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM*, et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 2016 May 19; **127**(20)**:** 2391-2405.
25. Swerdlow SH CE, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, Revised 4th ed edn. IARC: Lyon, France, 2017, 586pp.
26. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood* 2002 Oct 1; **100**(7)**:** 2292-2302.
27. Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T*, et al.* Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017 Jan 26; **129**(4)**:** 424-447.
28. Schwaller J, Frantsve J, Aster J, Williams IR, Tomasson MH, Ross TS*, et al.* Transformation of hematopoietic cell lines to growth-factor independence and induction of a fatal myelo- and lymphoproliferative disease in mice by retrovirally transduced TEL/JAK2 fusion genes. *EMBO J* 1998 Sep 15; **17**(18)**:** 5321-5333.
29. Wilbanks AM, Mahajan S, Frank DA, Druker BJ, Gilliland DG, Carroll M. TEL/PDGFbetaR fusion protein activates STAT1 and STAT5: a common mechanism for transformation by tyrosine kinase fusion proteins. *Exp Hematol* 2000 May; **28**(5)**:** 584-593.
30. Kelly LM, Liu Q, Kutok JL, Williams IR, Boulton CL, Gilliland DG. FLT3 internal tandem duplication mutations associated with human acute myeloid leukemias induce myeloproliferative disease in a murine bone marrow transplant model. *Blood* 2002 Jan 1; **99**(1)**:** 310-318.
31. Rowley JD. Chromosomes in leukemia and beyond: from irrelevant to central players. *Annu Rev Genomics Hum Genet* 2009; **10:** 1-18.
32. Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH*, 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 Jul 22; **116**(3)**:** 354-365.
33. De Kouchkovsky I, Abdul-Hay M. Acute myeloid leukemia: a comprehensive review and 2016 update. *Blood Cancer Journal* 2016 Jul; **6**.
34. Grimwade D, Ivey A, Huntly BJP. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. *Blood* 2016 Jan 7; **127**(1)**:** 29-41.
35. Tallman MS, Andersen JW, Schiffer CA, Appelbaum FR, Feusner JH, Woods WG*, et al.* All-trans retinoic acid in acute promyelocytic leukemia: long-term outcome and prognostic factor analysis from the North American Intergroup protocol. *Blood* 2002 Dec 15; **100**(13)**:** 4298-4302.
36. Asou N, Kishimoto Y, Kiyoi H, Okada M, Kawai Y, Tsuzuki M*, et al.* A randomized study with or without intensified maintenance chemotherapy in patients with acute promyelocytic leukemia who have become negative for PML-RARα transcript after consolidation therapy: The Japan Adult Leukemia Study Group (JALSG) APL97 study. *Blood* 2007; **110**(1)**:** 59-66.
37. Post SM, Kantarjian H, Quintás-Cardama A. 28 - Biology of Adult Myelocytic Leukemia and Myelodysplasia. In: Mendelsohn J, Gray JW, Howley PM, Israel MA, Thompson CB (eds). *The Molecular Basis of Cancer (Fourth Edition)*. Content Repository Only!: Philadelphia, 2015, pp 421-432.
38. Speck NA, Gilliland DG. Core-binding factors in haematopoiesis and leukaemia. *Nat Rev Cancer* 2002 Jul; **2**(7)**:** 502-513.

39. Goyama S, Mulloy JC. Molecular pathogenesis of core binding factor leukemia: current knowledge and future prospects. *International Journal of Hematology* 2011 Aug; **94**(2)**:** 126-133.
40. Opatz S, Polzer H, Herold T, Konstandin NP, Ksienzyk B, Zellmeier E*, et al.* Exome sequencing identifies recurring FLT3 N676K mutations in core-binding factor leukemia. *Blood* 2013 Sep 5; **122**(10)**:** 1761-1769.
41. Haferlach C, Dicker F, Kohlmann A, Schindela S, Weiss T, Kern W*, et al.* AML with CBFB-MYH11 rearrangement demonstrate RAS pathway alterations in 92% of all cases including a high frequency of NF1 deletions. *Leukemia* 2010 May; **24**(5)**:** 1065-1069.
42. Paschka P, Dohner K. Core-binding factor acute myeloid leukemia: can we improve on HiDAC consolidation? *Hematol-Am Soc Hemat* 2013 Dec**:** 209-219.
43. Cox MC, Panetta P, Lo-Coco F, Del Poeta G, Venditti A, Maurillo L*, et al.* Chromosomal aberration of the 11q23 locus in acute leukemia and frequency of MLL gene translocation: results in 378 adult patients. *Am J Clin Pathol* 2004 Aug; **122**(2)**:** 298-306.
44. Pui CH, Behm FG, Downing JR, Hancock MI, Shurtleff SA, Ribeiro RC*, et al.* 11q23/Mll Rearrangement Confers a Poor-Prognosis in Infants with Acute Lymphoblastic-Leukemia. *Journal of Clinical Oncology* 1994 May; **12**(5)**:** 909-915.
45. Satake N, Maseki N, Nishiyama M, Kobayashi H, Sakurai M, Inaba H*, et al.* Chromosome abnormalities and MLL rearrangements in acute myeloid leukemia of infants. *Leukemia* 1999 Jul; **13**(7)**:** 1013-1017.
46. Muntean AG, Hess JL. The pathogenesis of mixed-lineage leukemia. *Annu Rev Pathol* 2012; **7:** 283-301.
47. Krivtsov AV, Armstrong SA. MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer* 2007 Nov; **7**(11)**:** 823-833.
48. Balgobind BV, Zwaan CM, Pieters R, Van den Heuvel-Eibrink MM. The heterogeneity of pediatric MLL-rearranged acute myeloid leukemia. *Leukemia* 2011 Aug; **25**(8)**:** 1239-1248.
49. Tamai H, Inokuchi K. 11q23/MLL acute leukemia : update of clinical aspects. *J Clin Exp Hematop* 2010; **50**(2)**:** 91-98.
50. Bloomfield CD, Archer KJ, Mrozek K, Lillington DM, Kaneko Y, Head DR*, et al.* 11q23 balanced chromosome aberrations in treatment-related myelodysplastic syndromes and acute leukemia: report from an international workshop. *Genes Chromosomes Cancer* 2002 Apr; **33**(4)**:** 362-378.
51. Felix CA, Kolaris CP, Osheroff N. Topoisomerase II and the etiology of chromosomal translocations. *DNA Repair (Amst)* 2006 Sep 8; **5**(9-10)**:** 1093-1108.
52. Grimwade D, Walker H, Harrison G, Oliver F, Chatters S, Harrison CJ*, et al.* The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial. *Blood* 2001 Sep 1; **98**(5)**:** 1312-1320.
53. Schoch C, Schnittger S, Klaus M, Kern W, Hiddemann W, Haferlach T. AML with 11q23/MLL abnormalities as defined by the WHO classification: incidence, partner chromosomes, FAB subtype, age distribution, and prognostic impact in an unselected series of 1897 cytogenetically analyzed AML cases. *Blood* 2003 Oct 1; **102**(7)**:** 2395-2402.
54. Steudel C, Wermke M, Schaich M, Schakel U, Illmer T, Ehninger G*, et al.* Comparative analysis of MLL partial tandem duplication and FLT3 internal tandem duplication mutations in 956 adult patients with acute myeloid leukemia. *Genes Chromosomes Cancer* 2003 Jul; **37**(3)**:** 237-251.
55. Schnittger S, Kinkelin U, Schoch C, Heinecke A, Haase D, Haferlach T*, et al.* Screening for MLL tandem duplication in 387 unselected patients with AML identify a prognostically unfavorable subset of AML. *Leukemia* 2000 May; **14**(5)**:** 796-804.
56. Caligiuri MA, Schichman SA, Strout MP, Mrozek K, Baer MR, Frankel SR*, et al.* Molecular Rearrangement of the All-1 Gene in Acute Myeloid-Leukemia without Cytogenetic Evidence of 11q23 Chromosomal Translocations. *Cancer Research* 1994 Jan 15; **54**(2)**:** 370-373.

57. Schichman SA, Caligiuri MA, Strout MP, Carter SL, Gu YS, Canaani E*, et al.* All-1 Tandem Duplication in Acute Myeloid-Leukemia with a Normal Karyotype Involves Homologous Recombination between Alu Elements. *Cancer Research* 1994 Aug 15; **54**(16)**:** 4277-4280.
58. Dohner K, Tobis K, Ulrich R, Frohling S, Benner A, Schlenk RF*, 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 Aug 1; **20**(15)**:** 3254-3261.
59. Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson AG*, et al.* Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *New England Journal of Medicine* 2013 May 30; **368**(22)**:** 2059-2074.
60. Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S*, et al.* Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. *Nat Genet* 2001 Mar; **27**(3)**:** 263-270.
61. Smith ML, Hills RK, Grimwade D. Independent prognostic variables in acute myeloid leukaemia. *Blood Rev* 2011 Jan; **25**(1)**:** 39-51.
62. Kirstetter P, Schuster MB, Bereshchenko O, Moore S, Dvinge H, Kurz E*, et al.* Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. *Cancer Cell* 2008 Apr; **13**(4)**:** 299-310.
63. Bereshchenko O, Mancini E, Moore S, Bilbao D, Mansson R, Luc S*, et al.* Hematopoietic stem cell expansion precedes the generation of committed myeloid leukemia-initiating cells in C/EBPalpha mutant AML. *Cancer Cell* 2009 Nov 6; **16**(5)**:** 390-400.
64. Wouters BJ, Erpelinck-Verschueren CAJ, Lowenberg B, Valk PJM, Delwel R. Double, but Not Single, CEBPA mutations Define a Subgroup of Acute Myeloid Leukemia with Favorable Outcome and a Distinct Gene Expression Profile. *Blood* 2008 Nov 16; **112**(11)**:** 58-58.
65. Renneville A, Boissel N, Gachard N, Naguib D, Bastard C, de Botton S*, et al.* The favorable impact of CEBPA mutations in patients with acute myeloid leukemia is only observed in the absence of associated cytogenetic abnormalities and FLT3 internal duplication. *Blood* 2009 May 21; **113**(21)**:** 5090-5093.
66. Green CL, Koo KK, Hills RK, Burnett AK, Linch DC, Gale RE. Prognostic Significance of CEBPA Mutations in a Large Cohort of Younger Adult Patients With Acute Myeloid Leukemia: Impact of Double CEBPA Mutations and the Interaction With FLT3 and NPM1 Mutations. *Journal of Clinical Oncology* 2010 Jun 1; **28**(16)**:** 2739-2747.
67. Taskesen E, Bullinger L, Corbacioglu A, Sanders MA, Erpelinck CAJ, Wouters BJ*, 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 Feb 24; **117**(8)**:** 2469-2475.
68. Greif PA, Dufour A, Konstandin NP, Ksienzyk B, Zellmeier E, Tizazu B*, et al.* GATA2 zinc finger 1 mutations associated with biallelic CEBPA mutations define a unique genetic entity of acute myeloid leukemia. *Blood* 2012 Jul 12; **120**(2)**:** 395-403.
69. Lam K, Zhang DE. RUNX1 and RUNX1-ETO: roles in hematopoiesis and leukemogenesis. *Front Biosci (Landmark Ed)* 2012 Jan 1; **17:** 1120-1139.
70. Mendler JH, Maharry K, Radmacher MD, Mrozek K, Becker H, Metzeler KH*, 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 Sep 1; **30**(25)**:** 3109-3118.
71. You E, Cho YU, Jang S, Seo EJ, Lee JH, Lee JH*, et al.* Frequency and Clinicopathologic Features of RUNX1 Mutations in Patients With Acute Myeloid Leukemia Not Otherwise Specified. *Am J Clin Pathol* 2017 Jul 1; **148**(1)**:** 64-72.

72. Cazzola M, Della Porta MG, Malcovati L. The genetic basis of myelodysplasia and its clinical relevance. *Blood* 2013 Dec 12; **122**(25)**:** 4021-4034.

73. Tang JL, Hou HA, Chen CY, Liu CY, Chou WC, Tseng MH*, 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 Dec 17; **114**(26)**:** 5352-5361.

74. Schnittger S, Dicker F, Kern W, Wendland N, Sundermann J, Alpermann T*, et al.* RUNX1 mutations are frequent in de novo AML with noncomplex karyotype and confer an unfavorable prognosis. *Blood* 2011 Feb 24; **117**(8)**:** 2348-2357.

75. Gaidzik VI, Bullinger L, Schlenk RF, Zimmermann AS, Rock J, Paschka P*, 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 Apr 1; **29**(10)**:** 1364-1372.

76. Morishita K, Parker DS, Mucenski ML, Jenkins NA, Copeland NG, Ihle JN. Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. *Cell* 1988 Sep 9; **54**(6)**:** 831-840.

77. Mucenski ML, Taylor BA, Ihle JN, Hartley JW, Morse HC, 3rd, Jenkins NA*, et al.* Identification of a common ecotropic viral integration site, Evi-1, in the DNA of AKXD murine myeloid tumors. *Mol Cell Biol* 1988 Jan; **8**(1)**:** 301-308.

78. Lugthart S, Groschel S, Beverloo HB, Kayser S, Valk PJ, van Zelderen-Bhola SL*, et al.* Clinical, molecular, and prognostic significance of WHO type inv(3)(q21q26.2)/t(3;3)(q21;q26.2) and various other 3q abnormalities in acute myeloid leukemia. *J Clin Oncol* 2010 Aug 20; **28**(24)**:** 3890-3898.

79. De Braekeleer M, Le Bris MJ, De Braekeleer E, Basinko A, Morel F, Douet-Guilbert N. 3q26/EVI1 rearrangements in myeloid hemopathies: a cytogenetic review. *Future Oncol* 2015; **11**(11)**:** 1675-1686.

80. Johansson B, Fioretos T, Mitelman F. Cytogenetic and molecular genetic evolution of chronic myeloid leukemia. *Acta Haematol* 2002; **107**(2)**:** 76-94.

81. Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, van Putten WL, Valk PJ, van der Poel-van de Luytgaarde S, Hack R*, et al.* High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients. *Blood* 2003 Feb 1; **101**(3)**:** 837-845.

82. Lugthart S, van Drunen E, van Norden Y, van Hoven A, Erpelinck CA, Valk PJ*, et al.* High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression and chromosome 3q26 abnormalities underestimated. *Blood* 2008 Apr 15; **111**(8)**:** 4329-4337.

83. Langabeer SE, Rogers JR, Harrison G, Wheatley K, Walker H, Bain BJ*, et al.* EVI1 expression in acute myeloid leukaemia. *Br J Haematol* 2001 Jan; **112**(1)**:** 208-211.

84. Balgobind BV, Lugthart S, Hollink IH, Arentsen-Peters ST, van Wering ER, de Graaf SS*, et al.* EVI1 overexpression in distinct subtypes of pediatric acute myeloid leukemia. *Leukemia* 2010 May; **24**(5)**:** 942-949.

85. Groschel S, Schlenk RF, Engelmann J, Rockova V, Teleanu V, Kuhn MW*, et al.* Deregulated expression of EVI1 defines a poor prognostic subset of MLL-rearranged acute myeloid leukemias: a study of the German-Austrian Acute Myeloid Leukemia Study Group and the Dutch-Belgian-Swiss HOVON/SAKK Cooperative Group. *J Clin Oncol* 2013 Jan 1; **31**(1)**:** 95-103.

86. Ho PA, Alonzo TA, Gerbing RB, Pollard JA, Hirsch B, Raimondi SC*, et al.* High EVI1 expression is associated with MLL rearrangements and predicts decreased survival in paediatric acute myeloid leukaemia: a report from the children's oncology group. *Brit J Haematol* 2013 Sep; **162**(5)**:** 670-677.

87. Matsuo H, Kajihara M, Tomizawa D, Watanabe T, Saito AM, Fujimoto J*, et al.* EVI1 overexpression is a poor prognostic factor in pediatric patients with mixed lineage leukemia-AF9 rearranged acute myeloid leukemia. *Haematologica* 2014 Nov; **99**(11).

88. Jo A, Mitani S, Shiba N, Hayashi Y, Hara Y, Takahashi H*, et al.* High expression of EVI1 and MEL1 is a compelling poor prognostic marker of pediatric AML. *Leukemia* 2015 May; **29**(5)**:** 1076-1083.

89. Haferlach C, Bacher U, Grossmann V, Schindela S, Zenger M, Kohlmann A*, et al.* Three novel cytogenetically cryptic EVI1 rearrangements associated with increased EVI1 expression and poor prognosis identified in 27 acute myeloid leukemia cases. *Genes Chromosomes Cancer* 2012 Dec; **51**(12)**:** 1079-1085.
90. Volkert S, Schnittger S, Zenger M, Kern W, Haferlach T, Haferlach C. Amplification of EVI1 on cytogenetically cryptic double minutes as new mechanism for increased expression of EVI1. *Cancer Genet-Ny* 2014 Mar; **207**(3)**:** 103-108.
91. Buonamici S, Li D, Chi Y, Zhao R, Wang X, Brace L*, et al.* EVI1 induces myelodysplastic syndrome in mice. *J Clin Invest* 2004 Sep; **114**(5)**:** 713-719.
92. Cuenco GM, Ren RB. Both AML1 and EVI1 oncogenic components are required for the cooperation of AML1/MDS1/EVI1 with BCR/ABL in the induction of acute myelogenous leukemia in mice. *Oncogene* 2004 Jan 15; **23**(2)**:** 569-579.
93. Groschel S, Sanders MA, Hoogenboezem R, Zeilemaker A, Havermans M, Erpelinck C*, et al.* Mutational spectrum of myeloid malignancies with inv(3)/t(3;3) reveals a predominant involvement of RAS/RTK signaling pathways. *Blood* 2015 Jan 1; **125**(1)**:** 133-139.
94. Lavallee VP, Gendron P, Lemieux S, D'Angelo G, Hebert J, Sauvageau G. EVI1-rearranged acute myeloid leukemias are characterized by distinct molecular alterations. *Blood* 2015 Jan 1; **125**(1)**:** 140-143.
95. Groschel S, Lugthart S, Schlenk RF, Valk PJ, Eiwen K, Goudswaard C*, et al.* High EVI1 expression predicts outcome in younger adult patients with acute myeloid leukemia and is associated with distinct cytogenetic abnormalities. *J Clin Oncol* 2010 Apr 20; **28**(12)**:** 2101-2107.
96. Groschel S, Sanders MA, Hoogenboezem R, de Wit E, Bouwman BAM, Erpelinck C*, et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* 2014 Apr 10; **157**(2)**:** 369-381.
97. Yamazaki H, Suzuki M, Otsuki A, Shimizu R, Bresnick EH, Engel JD*, et al.* A remote GATA2 hematopoietic enhancer drives leukemogenesis in inv(3)(q21;q26) by activating EVI1 expression. *Cancer Cell* 2014 Apr 14; **25**(4)**:** 415-427.
98. Vassiliou GS, Cooper JL, Rad R, Li J, Rice S, Uren A*, et al.* Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice. *Nat Genet* 2011 May; **43**(5)**:** 470-475.
99. Mupo A, Celani L, Dovey O, Cooper JL, Grove C, Rad R*, et al.* A powerful molecular synergy between mutant Nucleophosmin and Flt3-ITD drives acute myeloid leukemia in mice. *Leukemia* 2013 Sep; **27**(9)**:** 1917-1920.
100. Kuhnl A, Grimwade D. Molecular markers in acute myeloid leukaemia. *Int J Hematol* 2012 Aug; **96**(2)**:** 153-163.
101. Gale RE, Lamb K, Allen C, El-Sharkawi D, Stowe C, Jenkinson S*, et al.* Simpson's Paradox and the Impact of Different DNMT3A Mutations on Outcome in Younger Adults With Acute Myeloid Leukemia. *J Clin Oncol* 2015 Jun 20; **33**(18)**:** 2072-2083.
102. Krause DS, Van Etten RA. Tyrosine kinases as targets for cancer therapy. *New England Journal of Medicine* 2005 Jul 14; **353**(2)**:** 172-187.
103. Rosnet O, Buhring HJ, Marchetto S, Rappold I, Lavagna C, Sainty D*, et al.* Human FLT3/FLK2 receptor tyrosine kinase is expressed at the surface of normal and malignant hematopoietic cells. *Leukemia* 1996 Feb; **10**(2)**:** 238-248.
104. Stirewalt DL, Radich JP. The role of FLT3 in haematopoietic malignancies. *Nature Reviews Cancer* 2003 Sep; **3**(9)**:** 650-U651.
105. Kottaridis PD, Gale RE, Linch DC. FLT3 mutations and leukaemia. *Brit J Haematol* 2003 Aug; **122**(4)**:** 523-538.
106. Levis M, Small D. FLT3: ITDoes matter in leukemia. *Leukemia* 2003 Sep; **17**(9)**:** 1738-1752.

107. Levis M. FLT3 mutations in acute myeloid leukemia: what is the best approach in 2013? *Hematology* 2013; **2013:** 220-226.

108. Mead AJ, Linch DC, Hills RK, Wheatley K, Burnett AK, Gale RE. FLT3 tyrosine kinase domain mutations are biologically distinct from and have a significantly more favorable prognosis than FLT3 internal tandem duplications in patients with acute myeloid leukemia. *Blood* 2007 Aug 15; **110**(4)**:** 1262-1270.

109. Mrozek K, Marcucci G, Paschka P, Whitman SP, Bloomfield CD. Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification? *Blood* 2007 Jan 15; **109**(2)**:** 431-448.

110. Mrozek K, Marcucci G, Paschka P, Bloomfield CD. Advances in molecular genetics and treatment of core-binding factor acute myeloid leukemia. *Curr Opin Oncol* 2008 Nov; **20**(6)**:** 711-718.

111. Kelly LM, Gilliland DG. Genetics of myeloid leukemias. *Annu Rev Genomics Hum Genet* 2002; **3:** 179-198.

112. Kihara R, Nagata Y, Kiyoi H, Kato T, Yamamoto E, Suzuki K*, et al.* Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients. *Leukemia* 2014 Aug; **28**(8)**:** 1586-1595.

113. Grossmann V, Schnittger S, Kohlmann A, Eder C, Roller A, Dicker F*, et al.* A novel hierarchical prognostic model of AML solely based on molecular mutations. *Blood* 2012 Oct 11; **120**(15)**:** 2963-2972.

114. Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun ZX, Racevskis J*, et al.* Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia. *New England Journal of Medicine* 2012 Mar 22; **366**(12)**:** 1079-1089.

115. Liang DC, Liu HC, Yang CP, Jaing TH, Hung IJ, Yeh TC*, et al.* Cooperating gene mutations in childhood acute myeloid leukemia with special reference on mutations of ASXL1, TET2, IDH1, IDH2, and DNMT3A. *Blood* 2013 Apr 11; **121**(15)**:** 2988-2995.

116. Gaidzik VI, Schlenk RF, Paschka P, Stolzle A, Spath D, Kuendgen A*, et al.* Clinical impact of DNMT3A mutations in younger adult patients with acute myeloid leukemia: results of the AML Study Group (AMLSG). *Blood* 2013 Jun 6; **121**(23)**:** 4769-4777.

117. Devillier R, Mansat-De Mas V, Gelsi-Boyer V, Demur C, Murati A, Corre J*, et al.* Role of ASXL1 and TP53 mutations in the molecular classification and prognosis of acute myeloid leukemias with myelodysplasia-related changes. *Oncotarget* 2015 Apr 10; **6**(10)**:** 8388-8396.

118. Sanders MA, Valk PJM. The evolving molecular genetic landscape in acute myeloid leukaemia. *Current Opinion in Hematology* 2013 Mar; **20**(2)**:** 79-85.

119. Wouters BJ, Delwel R. Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia. *Blood* 2016 Jan 7; **127**(1)**:** 42-52.

120. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE*, et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* 2010 Dec 16; **363**(25)**:** 2424-2433.

121. Thol F, Damm F, Ludeking A, Winschel C, Wagner K, Morgan M*, et al.* Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. *J Clin Oncol* 2011 Jul 20; **29**(21)**:** 2889-2896.

122. Shen Y, Zhu YM, Fan X, Shi JY, Wang QR, Yan XJ*, et al.* Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. *Blood* 2011 Nov 17; **118**(20)**:** 5593-5603.

123. Hou HA, Kuo YY, Liu CY, Chou WC, Lee MC, Chen CY*, et al.* DNMT3A mutations in acute myeloid leukemia: stability during disease evolution and clinical implications. *Blood* 2012 Jan 12; **119**(2)**:** 559-568.

124. Renneville A, Boissel N, Nibourel O, Berthon C, Helevaut N, Gardin C*, 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 Jun; **26**(6)**:** 1247-1254.

125. Marcucci G, Metzeler KH, Schwind S, Becker H, Maharry K, Mrozek K*, 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 Mar 1; **30**(7)**:** 742-750.

126. Markova J, Michkova P, Burckova K, Brezinova J, Michalova K, Dohnalova A*, et al.* Prognostic impact of DNMT3A mutations in patients with intermediate cytogenetic risk profile acute myeloid leukemia. *Eur J Haematol* 2012 Feb; **88**(2)**:** 128-135.

127. Gaidzik VI, Weber D, Paschka P, Kaumanns A, Krieger S, Corbacioglu A*, et al.* DNMT3A mutant transcript levels persist in remission and do not predict outcome in patients with acute myeloid leukemia. *Leukemia* 2018 01; **32**(1)**:** 30-37.

128. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 2007 Apr; **8**(4)**:** 286-298.

129. Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y*, et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* 2011 Mar 13; **43**(4)**:** 309-315.

130. Challen GA, Sun DQ, Jeong M, Luo M, Jelinek J, Berg JS*, et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature Genetics* 2012 Jan; **44**(1)**:** 23-U43.

131. Metzeler KH, Maharry K, Radmacher MD, Mrozek K, Margeson D, Becker H*, et al.* TET2 mutations improve the new European LeukemiaNet risk classification of acute myeloid leukemia: a Cancer and Leukemia Group B study. *J Clin Oncol* 2011 Apr 1; **29**(10)**:** 1373-1381.

132. Chou WC, Chou SC, Liu CY, Chen CY, Hou HA, Kuo YY*, et al.* TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. *Blood* 2011 Oct 06; **118**(14)**:** 3803-3810.

133. Abdel-Wahab O, Mullally A, Hedvat C, Garcia-Manero G, Patel J, Wadleigh M*, et al.* Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* 2009 Jul 2; **114**(1)**:** 144-147.

134. Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A*, et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 2010 Dec 14; **18**(6)**:** 553-567.

135. Delhommeau F, Dupont S, Della Valle V, James C, Trannoy S, Masse A*, et al.* Mutation in TET2 in myeloid cancers. *N Engl J Med* 2009 May 28; **360**(22)**:** 2289-2301.

136. Langemeijer SM, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M*, et al.* Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet* 2009 Jul; **41**(7)**:** 838-842.

137. Jankowska AM, Szpurka H, Tiu RV, Makishima H, Afable M, Huh J*, et al.* Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. *Blood* 2009 Jun 18; **113**(25)**:** 6403-6410.

138. Ko M, Huang Y, Jankowska AM, Pape UJ, Tahiliani M, Bandukwala HS*, et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 2010 Dec 9; **468**(7325)**:** 839-843.

139. Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C*, et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 2011 Jul 12; **20**(1)**:** 11-24.

140. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA*, et al.* Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. *Science* 2011 Sep 2; **333**(6047)**:** 1300-1303.

141. Gaidzik VI, Paschka P, Spath D, Habdank M, Kohne CH, Germing U*, 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 Apr 20; **30**(12)**:** 1350-1357.

142. Fisher CL, Randazzo F, Humphries RK, Brock HW. Characterization of Asxl1, a murine homolog of Additional sex combs, and analysis of the Asx-like gene family. *Gene* 2006 Mar 15; **369:** 109-118.

143. Lee SW, Cho YS, Na JM, Park UH, Kang M, Kim EJ*, et al.* ASXL1 Represses Retinoic Acid Receptor-mediated Transcription through Associating with HP1 and LSD1. *Journal of Biological Chemistry* 2010 Jan 1; **285**(1)**:** 18-29.

144. Gelsi-Boyer V, Brecqueville M, Devillier R, Murati A, Mozziconacci MJ, Birnbaum D. Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. *J Hematol Oncol* 2012 Mar 21; **5:** 12.

145. Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J*, et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med* 2012 Mar 22; **366**(12)**:** 1079-1089.

146. Tian XP, Xu Y, Yin J, Tian H, Chen SN, Wu DP*, et al.* TET2 gene mutation is unfavorable prognostic factor in cytogenetically normal acute myeloid leukemia patients with NPM1(+) and FLT3-ITD- mutations. *International Journal of Hematology* 2014 Jul; **100**(1)**:** 96-104.

147. Schnittger S, Eder C, Jeromin S, Alpermann T, Fasan A, Grossmann V*, et al.* ASXL1 exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome. *Leukemia* 2013 Jan; **27**(1)**:** 82-91.

148. El-Sharkawi D, Ali A, Evans CM, Hills RK, Burnett AK, Linch DC*, et al.* ASXL1 mutations are infrequent in young patients with primary acute myeloid leukemia and their detection has a limited role in therapeutic risk stratification. *Leuk Lymphoma* 2014 Jun; **55**(6)**:** 1326-1331.

149. Fernandez-Mercado M, Yip BH, Pellagatti A, Davies C, Larrayoz MJ, Kondo T*, et al.* Mutation patterns of 16 genes in primary and secondary acute myeloid leukemia (AML) with normal cytogenetics. *PLoS One* 2012; **7**(8)**:** e42334.

150. Carbuccia N, Murati A, Trouplin V, Brecqueville M, Adelaide J, Rey J*, et al.* Mutations of ASXL1 gene in myeloproliferative neoplasms. *Leukemia* 2009 Nov; **23**(11)**:** 2183-2186.

151. Pratcorona M, Abbas S, Sanders MA, Koenders JE, Kavelaars FG, Erpelinck-Verschueren CA*, et al.* Acquired mutations in ASXL1 in acute myeloid leukemia: prevalence and prognostic value. *Haematologica* 2012 Mar; **97**(3)**:** 388-392.

152. Vainchenker W, Delhommeau F, Constantinescu SN, Bernard OA. New mutations and pathogenesis of myeloproliferative neoplasms. *Blood* 2011 Aug 18; **118**(7)**:** 1723-1735.

153. Abdel-Wahab O, Adli M, LaFave LM, Gao J, Hricik T, Shih AH*, et al.* ASXL1 Mutations Promote Myeloid Transformation through Loss of PRC2-Mediated Gene Repression. *Cancer Cell* 2012 Aug 14; **22**(2)**:** 180-193.

154. Abdel-Wahab O, Pardanani A, Patel J, Wadleigh M, Lasho T, Heguy A*, et al.* Concomitant analysis of EZH2 and ASXL1 mutations in myelofibrosis, chronic myelomonocytic leukemia and blast-phase myeloproliferative neoplasms. *Leukemia* 2011 Jul; **25**(7)**:** 1200-1202.

155. Bejar R, Stevenson K, Abdel-Wahab O, Galili N, Nilsson B, Garcia-Manero G*, et al.* Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med* 2011 Jun 30; **364**(26)**:** 2496-2506.

156. Rocquain J, Carbuccia N, Trouplin V, Raynaud S, Murati A, Nezri M*, et al.* Combined mutations of ASXL1, CBL, FLT3, IDH1, IDH2, JAK2, KRAS, NPM1, NRAS, RUNX1, TET2 and WT1 genes in myelodysplastic syndromes and acute myeloid leukemias. *BMC Cancer* 2010 Aug 2; **10:** 401.

157. Carbuccia N, Trouplin V, Gelsi-Boyer V, Murati A, Rocquain J, Adelaide J*, et al.* Mutual exclusion of ASXL1 and NPM1 mutations in a series of acute myeloid leukemias. *Leukemia* 2010 Feb; **24**(2)**:** 469-473.

158. Metzeler KH, Becker H, Maharry K, Radmacher MD, Kohlschmidt J, Mrozek 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 Dec 22; **118**(26)**:** 6920-6929.

159. Paschka P, Schlenk RF, Herzig J, Aulitzky T, Gaidzik VI, Bullinger L*, et al.* ASXL1 Mutations Predict for Resistance to Chemotherapy and Inferior Outcome in Younger Adult Patients with Acute Myeloid Leukemia (AML): A Study of the German-Austrian AMLSG. *Blood* 2011 Nov 18; **118**(21)**:** 189-189.

160. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K*, et al.* Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome. *New England Journal of Medicine* 2009 Sep 10; **361**(11)**:** 1058-1066.

161. Chou WC, Hou HA, Chen CY, Tang JL, Yao M, Tsay W*, et al.* Distinct clinical and biologic characteristics in adult acute myeloid leukemia bearing the isocitrate dehydrogenase 1 mutation. *Blood* 2010 Apr 8; **115**(14)**:** 2749-2754.

162. Ward PS, Patel J, Wise DR, Abdel-Wahab O, Bennett BD, Coller HA*, et al.* The Common Feature of Leukemia-Associated IDH1 and IDH2 Mutations Is a Neomorphic Enzyme Activity Converting alpha-Ketoglutarate to 2-Hydroxyglutarate. *Cancer Cell* 2010 Mar 16; **17**(3)**:** 225-234.

163. Gross S, Cairns RA, Minden MD, Driggers EM, Bittinger MA, Jang HG*, et al.* Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. *J Exp Med* 2010 Feb 15; **207**(2)**:** 339-344.

164. Marcucci G, Maharry K, Wu YZ, Radmacher MD, Mrozek K, Margeson D*, 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. *Journal of Clinical Oncology* 2010 May 10; **28**(14)**:** 2348-2355.

165. Boissel N, Nibourel O, Renneville A, Gardin C, Reman O, Contentin N*, et al.* Prognostic Impact of Isocitrate Dehydrogenase Enzyme Isoforms 1 and 2 Mutations in Acute Myeloid Leukemia: A Study by the Acute Leukemia French Association Group. *Journal of Clinical Oncology* 2010 Aug 10; **28**(23)**:** 3717-3723.

166. Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL*, et al.* Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood* 2015 Feb 26; **125**(9)**:** 1367-1376.

167. Papaemmanuil E, Cazzola M, Boultwood J, Malcovati L, Vyas P, Bowen D*, et al.* Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med* 2011 Oct 13; **365**(15)**:** 1384-1395.

168. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R*, et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 2011 Sep 11; **478**(7367)**:** 64-69.

169. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P*, et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 2013 Nov 21; **122**(22)**:** 3616-3627; quiz 3699.

170. Leeke B, Marsman J, O'Sullivan JM, Horsfield JA. Cohesin mutations in myeloid malignancies: underlying mechanisms. *Exp Hematol Oncol* 2014; **3:** 13.

171. Thol F, Bollin R, Gehlhaar M, Walter C, Dugas M, Suchanek KJ*, et al.* Mutations in the cohesin complex in acute myeloid leukemia: clinical and prognostic implications. *Blood* 2014 Feb 6; **123**(6)**:** 914-920.

172. Thota S, Viny AD, Makishima H, Spitzer B, Radivoyevitch T, Przychodzen B*, et al.* Genetic alterations of the cohesin complex genes in myeloid malignancies. *Blood* 2014 Sep 11; **124**(11)**:** 1790-1798.

173. Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J*, et al.* World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997. *J Clin Oncol* 1999 Dec; **17**(12)**:** 3835-3849.

174. Falini B, Macijewski K, Weiss T, Bacher U, Schnittger S, Kern W*, et al.* Multilineage dysplasia has no impact on biologic, clinicopathologic, and prognostic features of AML with mutated nucleophosmin (NPM1). *Blood* 2010 May 6; **115**(18)**:** 3776-3786.

175. Bacher U, Schnittger S, Macijewski K, Grossmann V, Kohlmann A, Alpermann T*, et al.* Multilineage dysplasia does not influence prognosis in CEBPA-mutated AML, supporting the WHO proposal to classify these patients as a unique entity. *Blood* 2012 May 17; **119**(20)**:** 4719-4722.

176. Head DR. Revised classification of acute myeloid leukemia. *Leukemia* 1996 Nov; **10**(11)**:** 1826-1831.

177. Granfeldt Ostgard LS, Medeiros BC, Sengelov H, Norgaard M, Andersen MK, Dufva IH*, et al.* Epidemiology and Clinical Significance of Secondary and Therapy-Related Acute Myeloid Leukemia: A National Population-Based Cohort Study. *J Clin Oncol* 2015 Nov 1; **33**(31)**:** 3641-3649.

178. Hulegardh E, Nilsson C, Lazarevic V, Garelius H, Antunovic P, Rangert Derolf A*, et al.* Characterization and prognostic features of secondary acute myeloid leukemia in a population-based setting: a report from the Swedish Acute Leukemia Registry. *Am J Hematol* 2015 Mar; **90**(3)**:** 208-214.

179. Tefferi A, Lasho TL, Jimma T, Finke CM, Gangat N, Vaidya R*, et al.* One thousand patients with primary myelofibrosis: the mayo clinic experience. *Mayo Clin Proc* 2012 Jan; **87**(1)**:** 25-33.

180. Tefferi A, Rumi E, Finazzi G, Gisslinger H, Vannucchi AM, Rodeghiero F*, et al.* Survival and prognosis among 1545 patients with contemporary polycythemia vera: an international study. *Leukemia* 2013 Sep; **27**(9)**:** 1874-1881.

181. Gangat N, Caramazza D, Vaidya R, George G, Begna K, Schwager S*, et al.* DIPSS Plus: A Refined Dynamic International Prognostic Scoring System for Primary Myelofibrosis That Incorporates Prognostic Information From Karyotype, Platelet Count, and Transfusion Status. *Journal of Clinical Oncology* 2011 Feb 1; **29**(4)**:** 392-397.

182. Passamonti F, Rumi E, Pungolino E, Malabarba L, Bertazzoni P, Valentini M*, et al.* Life expectancy and prognostic factors for survival in patients with polycythemia vera and essential thrombocythemia. *Am J Med* 2004 Nov 15; **117**(10)**:** 755-761.

183. Barbui T, Thiele J, Passamonti F, Rumi E, Boveri E, Ruggeri M*, et al.* Survival and Disease Progression in Essential Thrombocythemia Are Significantly Influenced by Accurate Morphologic Diagnosis: An International Study. *Journal of Clinical Oncology* 2011 Aug 10; **29**(23)**:** 3179-3184.

184. Padron E, Garcia-Manero G, Patnaik MM, Itzykson R, Lasho T, Nazha A*, et al.* An international data set for CMML validates prognostic scoring systems and demonstrates a need for novel prognostication strategies. *Blood Cancer Journal* 2015 Jul; **5**.

185. Wang SA, Hasserjian RP, Fox PS, Rogers HJ, Geyer JT, Chabot-Richards D*, et al.* Atypical chronic myeloid leukemia is clinically distinct from unclassifiable myelodysplastic/myeloproliferative neoplasms. *Blood* 2014 Apr 24; **123**(17)**:** 2645-2651.

186. Auerbach AD, Weiner MA, Warburton D, Yeboa K, Lu L, Broxmeyer HE. Acute myeloid leukemia as the first hematologic manifestation of Fanconi anemia. *Am J Hematol* 1982 May; **12**(3)**:** 289-300.

187. Velez-Ruelas MA, Martinez-Jaramillo G, Arana-Trejo RM, Mayani H. Hematopoietic changes during progression from Fanconi anemia into acute myeloid leukemia: Case report and brief review of the literature. *Hematology* 2006 Oct-Dec; **11**(5-6)**:** 331-334.

188. Kojima S, Ohara A, Tsuchida M, Kudoh T, Hanada R, Okimoto Y*, et al.* Risk factors for evolution of acquired aplastic anemia into myelodysplastic syndrome and acute myeloid leukemia after immunosuppressive therapy in children. *Blood* 2002 Aug 1; **100**(3)**:** 786-790.

189. Bakhshi S, Gupta A, Kumar L. Acute myeloid leukemia with severe aplastic anemia following immunosuppressive therapy. *Indian J Pediatr* 2006 Nov; **73**(11)**:** 1033-1035.

190. Yoshimi A, Strahm B, Baumann I, Furlan I, Schwarz S, Teigler-Schlegel A*, et al.* Hematopoietic stem cell transplantation in children and young adults with secondary myelodysplastic syndrome and acute myelogenous leukemia after aplastic anemia. *Biol Blood Marrow Transplant* 2014 Mar; **20**(3)**:** 425-429.

191. Rampal R, Ahn J, Abdel-Wahab O, Nahas M, Wang K, Lipson D*, et al.* Genomic and functional analysis of leukemic transformation of myeloproliferative neoplasms. *Proc Natl Acad Sci U S A* 2014 Dec 16; **111**(50)**:** E5401-5410.

192. Rampal R, Mascarenhas J. Pathogenesis and management of acute myeloid leukemia that has evolved from a myeloproliferative neoplasm. *Current Opinion in Hematology* 2014 Mar; **21**(2)**:** 65-71.

193. Takahashi K, Jabbour E, Wang X, Luthra R, Bueso-Ramos C, Patel K*, et al.* Dynamic acquisition of FLT3 or RAS alterations drive a subset of patients with lower risk MDS to secondary AML. *Leukemia* 2013 Oct; **27**(10)**:** 2081-2083.

194. Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K*, et al.* Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med* 2012 Mar 22; **366**(12)**:** 1090-1098.

195. Flach J, Dicker F, Schnittger S, Schindela S, Kohlmann A, Haferlach T*, et al.* An accumulation of cytogenetic and molecular genetic events characterizes the progression from MDS to secondary AML: an analysis of 38 paired samples analyzed by cytogenetics, molecular mutation analysis and SNP microarray profiling. *Leukemia* 2011 Apr; **25**(4)**:** 713-718.

196. Welch JS, Petti AA, Miller CA, Fronick CC, O'Laughlin M, Fulton RS*, et al.* TP53 and Decitabine in Acute Myeloid Leukemia and Myelodysplastic Syndromes. *N Engl J Med* 2016 Nov 24; **375**(21)**:** 2023-2036.

197. Kuykendall A, Duployez N, Boissel N, Lancet JE, Welch JS. Acute Myeloid Leukemia: The Good, the Bad, and the Ugly. *Am Soc Clin Oncol Educ Book* 2018 May 23; **38:** 555-573.

198. Larson RA. Etiology and management of therapy-related myeloid leukemia. *Hematology Am Soc Hematol Educ Program* 2007**:** 453-459.

199. Leone G, Pagano L, Ben-Yehuda D, Voso MT. Therapy-related leukemia and myelodysplasia: susceptibility and incidence. *Haematologica* 2007 Oct; **92**(10)**:** 1389-1398.

200. Bhatia S. Therapy-related myelodysplasia and acute myeloid leukemia. *Semin Oncol* 2013 Dec; **40**(6)**:** 666-675.

201. Pedersen-Bjergaard J, Andersen MK, Christiansen DH, Nerlov C. Genetic pathways in therapy-related myelodysplasia and acute myeloid leukemia. *Blood* 2002 Mar 15; **99**(6)**:** 1909-1912.

202. Churpek JE, Marquez R, Neistadt B, Claussen K, Lee MK, Churpek MM*, et al.* Inherited mutations in cancer susceptibility genes are common among survivors of breast cancer who develop therapy-related leukemia. *Cancer* 2016 Jan 15; **122**(2)**:** 304-311.

203. Schulz E, Valentin A, Ulz P, Beham-Schmid C, Lind K, Rupp V*, et al.* Germline mutations in the DNA damage response genes BRCA1, BRCA2, BARD1 and TP53 in patients with therapy related myeloid neoplasms. *Journal of Medical Genetics* 2012 Jul; **49**(7)**:** 422-428.

204. Berwick M, Vineis P. Markers of DNA repair and susceptibility to cancer in humans: An epidemiologic review. *Jnci-J Natl Cancer I* 2000 Jun 7; **92**(11)**:** 874-897.

205. Evans WE, McLeod HL. Drug therapy - Pharmacogenomics - Drug disposition, drug targets, and side effects. *New England Journal of Medicine* 2003 Feb 6; **348**(6)**:** 538-549.

206. Schulz E, Kashofer K, Heitzer E, Mhatre KN, Speicher MR, Hoefler G*, et al.* Preexisting TP53 mutation in therapy-related acute myeloid leukemia. *Annals of Hematology* 2015 Mar; **94**(3)**:** 527-529.

207. Wong TN, Ramsingh G, Young AL, Miller CA, Touma W, Welch JS*, et al.* Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* 2015 Feb 26; **518**(7540)**:** 552-555.

208. Takahashi K, Wang F, Kantarjian H, Doss D, Khanna K, Thompson E*, et al.* Preleukaemic clonal haemopoiesis and risk of therapy-related myeloid neoplasms: a case-control study. *Lancet Oncol* 2017 Jan; **18**(1)**:** 100-111.

209. Gillis NK, Ball M, Zhang Q, Ma ZJ, Zhao YL, Yoder SJ*, et al.* Clonal haemopoiesis and therapy-related myeloid malignancies in elderly patients: a proof-of-concept, case-control study. *Lancet Oncology* 2017 Jan; **18**(1)**:** 112-121.

210. Hackl H, Astanina K, Wieser R. Molecular and genetic alterations associated with therapy resistance and relapse of acute myeloid leukemia. *J Hematol Oncol* 2017 Feb 20; **10**(1)**:** 51.

211. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. *New England Journal of Medicine* 2015 Sep 17; **373**(12)**:** 1136-1152.

212. Corces-Zimmerman MR, Majeti R. Pre-leukemic evolution of hematopoietic stem cells: the importance of early mutations in leukemogenesis. *Leukemia* 2014; **28:** 2276.

213. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC*, et al.* The Origin and Evolution of Mutations in Acute Myeloid Leukemia. *Cell* 2012 Jul 20; **150**(2)**:** 264-278.

214. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS*, et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 2012 Jan 26; **481**(7382)**:** 506-510.

215. Sun QY, Ding LW, Tan KT, Chien W, Mayakonda A, Lin DC*, et al.* Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD). *Leukemia* 2017 Jan; **31**(1)**:** 1-10.

216. Vosberg S, Greif PA. Clonal evolution of acute myeloid leukemia from diagnosis to relapse. *Genes Chromosomes Cancer* 2019 Dec; **58**(12)**:** 839-849.

217. Cocciardi S, Dolnik A, Kapp-Schwoerer S, Rücker FG, Lux S, Blätte TJ*, et al.* Clonal evolution patterns in acute myeloid leukemia with NPM1 mutation. *Nature Communications* 2019 2019/05/02; **10**(1)**:** 2031.

218. Christen F, Hoyer K, Yoshida K, Hou H-A, Waldhueter N, Heuser M*, et al.* Genomic landscape and clonal evolution of acute myeloid leukemia with t(8;21): an international study on 331 patients. *Blood* 2019; **133**(10)**:** 1140-1151.

219. Dick JE, Lapidot T. Biology of normal and acute myeloid leukemia stem cells. *Int J Hematol* 2005 Dec; **82**(5)**:** 389-396.

220. Short NJ, Rytting ME, Cortes JE. Acute myeloid leukaemia. *Lancet* 2018 Aug 18; **392**(10147)**:** 593-606.

221. 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. *P Natl Acad Sci USA* 2014 Feb 18; **111**(7)**:** 2548-2553.

222. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF*, et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014 Dec 25; **371**(26)**:** 2477-2487.

223. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG*, et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine* 2014; **371**(26)**:** 2488-2498.

224. Bochtler T, Stolzel F, Heilig CE, Kunz C, Mohr B, Jauch A*, et al.* Clonal Heterogeneity As Detected by Metaphase Karyotyping Is an Indicator of Poor Prognosis in Acute Myeloid Leukemia. *Journal of Clinical Oncology* 2013 Nov 1; **31**(31)**:** 3898-+.

225. Medeiros BC, Othus M, Fang M, Appelbaum FR, Erba HP. Cytogenetic heterogeneity negatively impacts outcomes in patients with acute myeloid leukemia. *Haematologica* 2015 Mar; **100**(3)**:** 331-335.

226. Sperling AS, Gibson CJ, Ebert BL. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nature Reviews Cancer* 2017 Jan; **17**(1)**:** 5-19.

227. Valent P, Kern W, Hoermann G, Milosevic Feenstra JD, Sotlar K, Pfeilstoecker M*, et al.* Clonal Hematopoiesis with Oncogenic Potential (CHOP): Separation from CHIP and Roads to AML. *International Journal of Molecular Sciences* 2019 Feb 1; **20**(3).

228. Shlush LI. Age-related clonal hematopoiesis. *Blood* 2018 Feb 1; **131**(5)**:** 496-504.

229. Jan M, Snyder TM, Corces-Zimmerman MR, Vyas P, Weissman IL, Quake SR*, et al.* Clonal Evolution of Preleukemic Hematopoietic Stem Cells Precedes Human Acute Myeloid Leukemia. *Science Translational Medicine* 2012 Aug 29; **4**(149).

230. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V*, et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* 2014 Feb 20; **506**(7488)**:** 328-+.

231. Roloff GW, Lai C, Hourigan CS, Dillon LW. Technical Advances in the Measurement of Residual Disease in Acute Myeloid Leukemia. *J Clin Med* 2017 Sep 19; **6**(9).

232. Ehinger M, Pettersson L. Measurable residual disease testing for personalized treatment of acute myeloid leukemia. *Apmis*; **0**(0).

233. Hourigan CS, Gale RP, Gormley NJ, Ossenkoppele GJ, Walter RB. Measurable residual disease testing in acute myeloid leukaemia. *Leukemia* 2017 Jul; **31**(7)**:** 1482-1490.

234. Ossenkoppele G, Schuurhuis GJ. MRD in AML: does it already guide therapy decision-making? *Hematol-Am Soc Hemat* 2016 Dec**:** 356-365.

235. Buccisano F, Dillon R, Freeman SD, Venditti A. Role of Minimal (Measurable) Residual Disease Assessment in Older Patients with Acute Myeloid Leukemia. *Cancers (Basel)* 2018 Jun 26; **10**(7).

236. Levine RL, Valk PJM. Next generation sequencing in diagnosis and MRD assessment of AML. *Haematologica* 2019**:** haematol.2018.205955.

237. Klco JM, Miller CA, Griffith M, Petti A, Spencer DH, Ketkar-Kulkarni S*, et al.* Association Between Mutation Clearance After Induction Therapy and Outcomes in Acute Myeloid Leukemia. *Jama-J Am Med Assoc* 2015

Aug 25; **314**(8)**:** 811-822.

238. Jongen-Lavrencic M, Grob T, Hanekamp D, Kavelaars FG, Al Hinai A, Zeilemaker A*, et al.* Molecular Minimal Residual Disease in Acute Myeloid Leukemia. *N Engl J Med* 2018 Mar 29; **378**(13)**:** 1189-1199.
239. Morita K, Kantarjian HM, Wang F, Yan YQ, Bueso-Ramos C, Sasaki K*, et al.* Clearance of Somatic Mutations at Remission and the Risk of Relapse in Acute Myeloid Leukemia. *Journal of Clinical Oncology* 2018 Jun 20; **36**(18)**:** 1788-+.
240. Thol F, Gabdoulline R, Liebich A, Klement P, Schiller J, Kandziora C*, et al.* Measurable residual disease monitoring by NGS before allogeneic hematopoietic cell transplantation in AML. *Blood* 2018 Oct 18; **132**(16)**:** 1703-1713.
241. Hourigan CS, Karp JE. Minimal residual disease in acute myeloid leukaemia. *Nat Rev Clin Oncol*. 2013 Aug;10(8):460-71.

Chapter 2

Molecular Minimal Residual Disease in Acute Myeloid Leukemia

M. Jongen-Lavrencic*[1], T. Grob[1]*, D. Hanekamp*[2], F.G. Kavelaars[1], A. al Hinai[1], A. Zeilemaker[1], C.A.J. Erpelinck-Verschueren[1], P.L. Gradowska, R. Meijer[3], J. Cloos[2], B.J. Biemond[4], C. Graux[5], M. van Marwijk Kooy[6], M.G. Manz[7], T. Pabst[8], J.R. Passweg[9], V. Havelange[10], G.J. Ossenkoppele[2], M.A. Sanders[1], G.J. Schuurhuis[2]#, B. Löwenberg#, and P.J.M. Valk[1]#

[1] Department of Hematology, Erasmus University Medical Center, ErasmusMC Cancer Institute, Rotterdam, The Netherlands
[2] Department of Hematology, VU University Medical Center, Amsterdam, The Netherlands
[3] HOVON Data Center, Department of Hematology, Erasmus University Medical Center, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
[4] Department of Hematology, Academic Medical Center, Amsterdam, The Netherlands
[5] UCL Namur (Godinne), Yvoir, Belgium
[6] Isala Hospital, Zwolle, The Netherlands
[7] Hematology, University and University Hospital Zurich, Zurich, Switzerland
[8] University Hospital, Bern, Switzerland
[9] Division of hematology, University Hospital Basel, Basel, Switzerland
[10] Department of Hematology, Cliniques Universitaires Saint-Luc, Brussels, Belgium

* Shared first authors
# Shared last authors

N Engl J Med. 2018 Mar 29;378(13):1189-1199

ABSTRACT

BACKGROUND

Patients with acute myeloid leukemia (AML) often reach complete remission, but relapse rates remain high. Next-generation sequencing enables the detection of molecular minimal residual disease in virtually every patient, but its clinical value for the prediction of relapse has yet to be established.

METHODS

We conducted a study involving patients 18 to 65 years of age who had newly diagnosed AML. Targeted next-generation sequencing was carried out at diagnosis and after induction therapy (during complete remission). End points were 4-year rates of relapse, relapse-free survival, and overall survival.

RESULTS

At least one mutation was detected in 430 out of 482 patients (89.2%). Mutations persisted in 51.4% of those patients during complete remission and were present at various allele frequencies (range, 0.02 to 47%). The detection of persistent *DTA* mutations (i.e., mutations in *DNMT3A*, *TET2*, and *ASXL1*), which are often present in persons with age-related clonal hematopoiesis, was not correlated with an increased relapse rate. After the exclusion of persistent *DTA* mutations, the detection of molecular minimal residual disease was associated with a significantly higher relapse rate than no detection (55.4% vs. 31.9%; hazard ratio, 2.14; P<0.001), as well as with lower rates of relapse-free survival (36.6% vs. 58.1%; hazard ratio for relapse or death, 1.92; P<0.001) and overall survival (41.9% vs. 66.1%; hazard ratio for death, 2.06; P<0.001). Multivariate analysis confirmed that the persistence of non-*DTA* mutations during complete remission conferred significant independent prognostic value with respect to the rates of relapse (hazard ratio, 1.89; P<0.001), relapse-free survival (hazard ratio for relapse or death, 1.64; P = 0.001), and overall survival (hazard ratio for death, 1.64; P = 0.003). A comparison of sequencing with flow cytometry for the detection of residual disease showed that sequencing had significant additive prognostic value.

CONCLUSIONS

Among patients with AML, the detection of molecular minimal residual disease during complete remission had significant independent prognostic value with respect to relapse and survival rates, but the detection of persistent mutations that are associated with clonal hematopoiesis did not have such prognostic value within a 4-year time frame. (Funded by the Queen Wilhelmina Fund Foundation of the Dutch Cancer Society and others.)

Acute myeloid leukemia (AML) is a heterogeneous group of clonal hematopoietic stem-cell disorders with a variable response to therapy.[1-3] Although the majority of patients with newly diagnosed AML have morphologic complete remission after they are treated with intensive induction chemotherapy, relapse rates remain high.2 Decisions about the choice of post-remission therapy in patients with AML currently depend on the identification of a selected set of genetic markers at diagnosis and the detection of residual disease with multiparameter flow cytometry.[2, 4] Quantitative molecular evaluation during complete remission could further improve prognostication of outcomes in patients with AML.

The potential of the detection of molecular minimal residual disease after treatment to predict disease relapse in patients with AML has been explored, but assessment of molecular minimal residual disease is not widely established in clinical practice. Previous studies have dealt with only a few leukemia-specific genetic aberrations.[5-11] Next-generation sequencing enables comprehensive, simultaneous detection of somatic mutations that are often patient-specific, both at diagnosis and during treatment.[5, 12] Initial studies showed the complex dynamics of residual mutations after induction therapy and the possible association between the persistence of certain somatic mutations and risk of relapse.[12, 13]

In determining whether molecular monitoring may be applicable in patients with AML, the phenomenon of age-related clonal hematopoiesis (also known as clonal hematopoiesis of indeterminate potential),[14-17] a condition characterized by the recurrence of gene mutations (allele frequency, >2%) in healthy persons with no evidence of hematologic disease, has added an extra layer of complexity. Persons with age-related clonal hematopoiesis have a slightly increased risk of developing hematologic cancers over time.[14, 15, 18] Mutations in the epigenetic regulators *DNMT3A*, *TET2*, and *ASXL1* (i.e., *DTA* mutations) are most common in persons with age-related clonal hematopoiesis.[14-19] Residual leukemia-specific mutations that are present in the bone marrow during complete remission may represent either residual leukemic cells or age-related clonal hematopoiesis.[14, 15, 17] Whether post-treatment persistence of genetic mutations associated with age-related clonal hematopoiesis in the bone marrow from patients with AML has an effect on the disease course remains unclear.

We evaluated a large cohort of patients with AML to investigate whether targeted molecular monitoring with next-generation sequencing could add clinical value for predicting the recurrence of leukemia.

METHODS

Study Design

The study was designed by the first two and the last two authors, who wrote the manuscript with input from the other authors. The authors vouch for the completeness and accuracy of the data and analysis. No one who is not an author contributed to the manuscript. There was no commercial support for the study.

Patients and Cell Samples

Between 2001 and 2013, we obtained samples of bone marrow or peripheral blood from 482 patients, between the ages of 18 and 65, who had a confirmed diagnosis of previously untreated AML (428 patients) or had refractory anemia with excess of blasts, with a score on the Revised International Prognostic Scoring System of more than 4.5, indicating a high or very high risk of relapse (54 patients). To be included in the study, patients had to be in either complete remission or complete remission with incomplete hematologic recovery (defined according to the European Leukemia Net recommendation; hereafter collectively referred to as complete remission), with less than 5% blast cells in the bone marrow,[2,4] after receiving two cycles of induction chemotherapy (Fig. S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). Among patients in whom at least one mutation was detected at diagnosis, samples were obtained during a defined period of remission, between 21 days and 4 months after the start of the second treatment cycle.

Patients were treated according to the clinical protocol of either the Dutch–Belgian Cooperative Trial Group for Hematology–Oncology (HOVON)[20] or the Swiss Group for Clinical Cancer Research (SAKK). The treatment protocols and patient eligibility criteria have been described previously.[21,22] All the patients provided written informed consent. Details about the patients and cell samples are provided in the Supplementary Appendix.

Targeted Next-Generation Sequencing and Multiparameter Flow Cytometry

To detect the mutations in 54 genes that are often present in patients with hematologic cancers, we used targeted next-generation sequencing with the Illumina TruSight Myeloid Sequencing Panel (Illumina), following the manufacturer's protocol. Detection of residual disease with multiparameter flow cytometry was performed as described previously.[23] Details about these detection methods and data interpretation are provided in the Supplementary Appendix.

Statistical Analysis

The 430 patients in whom at least one mutation was detected at diagnosis were randomly assigned to either a training cohort (283 patients) or a validation cohort (147 patients); the two cohorts had similar clinical, cytogenetic, and molecular characteristics (Table 1,

and Fig. S1 and Table S1 in the Supplementary Appendix). The primary end point was the 4-year cumulative incidence of relapse (defined according to the European Leukemia Net recommendation[4]), and the secondary end points were the 4-year rates of overall survival and relapse-free survival. Within each cohort, the difference in the incidence of relapse between patients in whom residual disease was detected and those in whom residual disease was not detected was evaluated with the use of the method of Gray and the Fine and Gray model for competing risks. The log-rank test and the Cox proportional-hazards model were used for survival analyses. A two-sided P value of 0.05 or less was considered to indicate statistical significance. Details about the statistical analyses are provided in the Supplementary Appendix.

Table 1. clinical, cytogenetic, and molecular characteristics of the 430 Patients.*

Characteristic	Value
Age at diagnosis — yr	
Median	51
Range	18–66
Sex — no. (%)	
Male	216 (50)
Female	214 (50)
White-cell count per microliter at diagnosis — no. (%)	
≤100,000	387 (90)
>100,000	43 (10)
2017 European Leukemia Network risk classification at diagnosis — no. (%)	
Favorable	204 (47)
Intermediate	113 (26)
Adverse	113 (26)
No. of chemotherapy cycles to attain complete remission — no. (%)	
1	360 (84)
2	70 (16)
Consolidation therapy — no. (%)	
None	46 (11)
Chemotherapy	117 (27)
Autologous hematopoietic stem-cell transplantation	78 (18)
Allogeneic hematopoietic stem-cell transplantation	189 (44)
Cytogenetic analysis at diagnosis — no. (%)†	
t(8;21)	27 (6)
inv(16)	24 (6)
Complex karyotype	38 (9)
Monosomal karyotype	30 (7)
Mutation at diagnosis — no. (%)	
ASXL1	31 (7)
CEBPA double mutation	19 (4)
DNMT3A	141 (33)
FLT3	
Tyrosine kinase domain	53 (12)
Internal tandem duplication, low ratio	40 (9)
Internal tandem duplication, high ratio	51 (12)
NPM1	168 (39)
RUNX1	50 (12)
TET2	48 (11)

* The percentages may not sum to 100 because of rounding.
† Karyotyping failed in 13 patients.

RESULTS

Detection of Mutations at Diagnosis

We performed targeted next-generation sequencing to detect gene mutations at diagnosis in samples obtained from 482 patients with AML (Fig. S1 in the Supplementary Appendix). We detected an average of 2.9 mutations per patient; at least 1 single mutation, which could potentially serve as a marker of residual disease, was present in 430 (89.2%) of the patients. Mutations in *NPM1*, *DNMT3A*, *FLT3*, and *NRAS* were among the most common detectable mutations at diagnosis (Table 1 and Fig. 1A, and Table S1 in the Supplementary Appendix).

Detection of Mutations during Complete Remission

We then performed targeted next-generation sequencing to detect persistent mutations after induction therapy in samples of bone marrow obtained from 430 patients who were in complete remission. Persistent mutations were detected in 51.4% of the patients (Fig. 1A, and Fig. S2A in the Supplementary Appendix). The rate at which mutations persisted was highly variable across genes. *DTA* mutations were most common, persisting at rates of 78.7% for *DNMT3A*, 54.2% for *TET2*, and 51.6% for *ASXL1* (Fig. 1A). In contrast, the majority of mutations in genes related to the RAS pathway were cleared after induction therapy, with mutations in *NRAS*, *PTPN11*, *KIT*, and *KRAS* persisting at rates of 4.2%, 7.0%, 13.5%, and 12.5%, respectively.

Of note, the allele frequencies of the mutations that persisted during complete remission ranged from 0.02 to 47% (Fig. 1B). This finding suggests that residual mutation-bearing cells could constitute a minor population of the cells or perhaps even a majority of the cells. An allele frequency of 50% is consistent with the presence of a heterozygous mutation in all cells. Thus, although the patients were in morphologic complete remission, which would typically imply that heterozygous mutations are present at allele frequencies lower than 2.5% (the equivalent of <5% blast cells in the bone marrow), the samples that were obtained during remission often contained mutations with much higher allele frequencies (Fig. 1B).

Mutations that persisted after induction therapy at allele frequencies higher than 2.5% were often *DTA* mutations (Fig. 1, and Fig. S2 and S3 in the Supplementary Appendix). In contrast, mutations in *IDH1*, *IDH2*, *STAG2*, *TP53*, and other genes only occasionally persisted after induction therapy at allele frequencies higher than 2.5%, and thus the allele frequencies of these mutations were typically consistent with the state of morphologic complete remission (<5% blast cells in the bone marrow).

Because *DTA* mutations have been established as the most common gene mutations in persons with age-related clonal hematopoiesis,[14-19] the persistent *DTA* mutations might have represented non-leukemic clones that repopulated the bone marrow after induction therapy. Among patients who had both *DTA* mutations and non-*DTA* mutations at diagnosis, non-*DTA* mutations were generally cleared after induction chemotherapy, whereas *DTA*

mutations often remained detectable during complete remission and were the only persistent mutations in 90 of 133 (67.7%) of those patients (Fig. S2 in the Supplementary Appendix). These observations are consistent with the notion that residual cells bearing *DTA* mutations after induction therapy represent non-leukemic clones rather than persistent malignant disease.

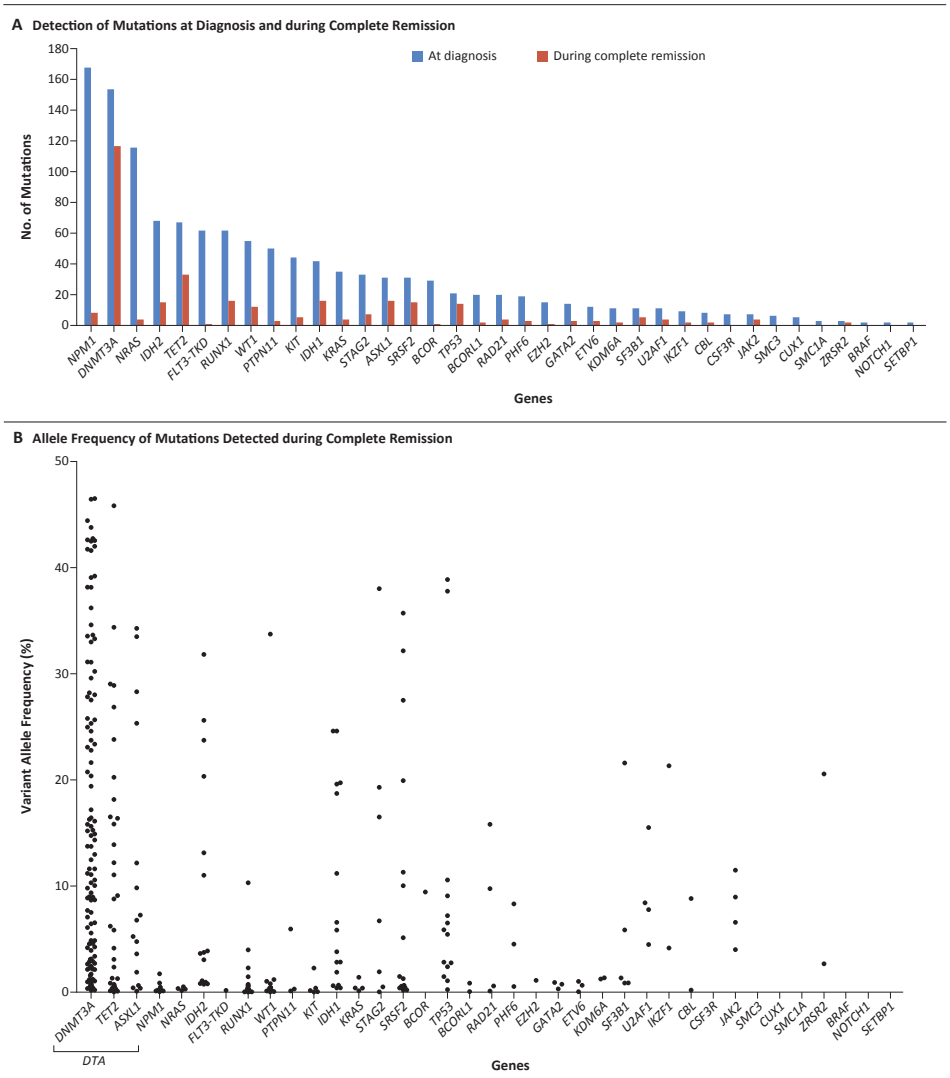


Figure 1. Detection of Mutations at Diagnosis and during Complete Remission and Allele Frequency of Mutations Detected during Complete Remission.

Panel A shows the number of mutations in each leukemia-associated gene, both at diagnosis of acute myeloid leukemia and during complete remission, in 430 patients. Panel B shows the allele frequency of each mutation in each gene during complete remission in 430 patients. In male patients, the variant allele frequencies for *PHF6*, *KDM6A*, *ZRSR2*, *BCOR*, *BCORL1*, and *STAG2* (on the X chromosome) were divided by 2.

Relapse and Survival End Points

In the training cohort (283 patients), we found that the detection of any persistent mutation during complete remission was associated with an increased risk of relapse (4-year relapse rate, 48.2% with detection vs. 32.4% with no detection; P = 0.03) (Fig. S4A in the Supplementary Appendix). We then imposed various thresholds for allele frequency to determine whether the prognostic value of the persistent mutations would improve after the exclusion of mutations with a high allele frequency, which could indicate a state of clonal hematopoiesis. The correlation of persistent mutations with an increased relapse risk appeared to be independent of allele frequency. A correlation with relapse risk generally remained present when we excluded persistent mutations with allele frequencies at or above the following thresholds: 30% (P = 0.09), 20% (P = 0.11), 10% (P = 0.01), 5% (P = 0.04), 2.5% (P = 0.007), and 1% (P = 0.07) (Fig. S4 in the Supplementary Appendix). The exclusion of persistent mutations with certain allele frequencies had no clear effect on the relationship between persistent mutations and an increased relapse risk, thus precluding the identification of a threshold for allele frequency that could be used to distinguish populations at higher or lower risk for relapse. As we mentioned previously, the patients with persistent mutations at high allele frequencies were enriched for *DTA* mutations (Fig. 1B).

We next determined whether persistent *DTA* mutations, which are associated with age-related clonal hematopoiesis, might be correlated with an increased relapse risk. We observed that the detection of persistent *DTA* mutations was not significantly associated with a higher 4-year relapse rate than no detection (P = 0.29). The absence of a correlation was independent of allele frequency. No significant correlation of persistent *DTA* mutations with an increased relapse risk was apparent when we excluded persistent *DTA* mutations with allele frequencies at or above the following thresholds: 30% (P = 0.91), 20% (P = 0.66), 10% (P = 0.89), 5% (P = 0.82), 2.5% (P = 0.53), and 1% (P = 0.92) (Fig. S5 in the Supplementary Appendix). In contrast, among patients who had persistent *DTA* mutations during complete remission, coexisting persistent non-*DTA* mutations had high prognostic value with respect to relapse (4-year relapse rate, 66.7% with detection vs. 39.4% with no detection; P = 0.002) (Fig. 2A). Thus, in patients with persistent *DTA* mutations, the presence of residual disease that specifically included coexisting non-*DTA* mutations represented a predictor of impending relapse.

We next assessed whether persistent non-*DTA* mutations might be correlated with an increased relapse risk. The detection of persistent non-*DTA* mutations at any allele frequency was strongly associated with an increased relapse risk (4-year relapse rate, 55.7% with detection vs. 34.6% with no detection; P = 0.001) (Fig. 2B), as well as with reduced relapse-free survival (4-year rate of relapse-free survival, 36.6% with detection vs. 56.7% with no detection; P = 0.006) and reduced overall survival (4-year rate of overall survival, 43.7% with detection vs. 65.3% with no detection; P = 0.01) (Fig. 2C, and Fig. S6 in the Supplementary Appendix).

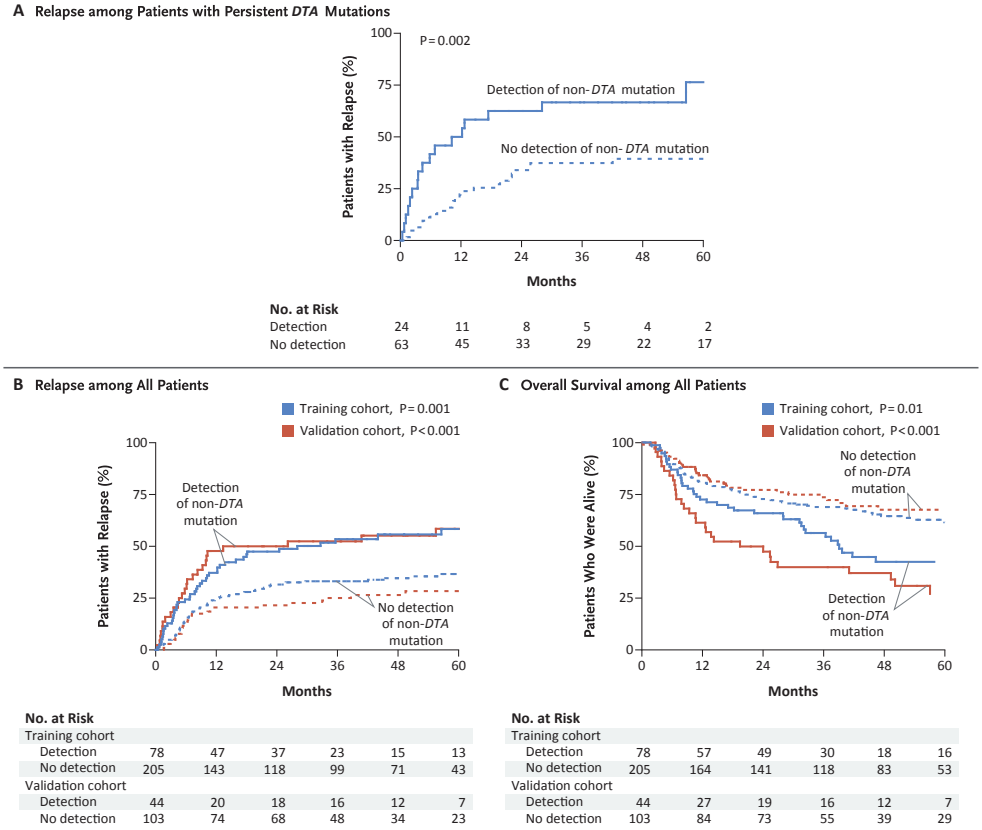


Figure 2. Rates of Relapse and Overall Survival.

Shown is the cumulative incidence of relapse among patients in the training cohort with persistent *DTA* mutations, according to the detection of coexisting persistent non-*DTA* mutations during complete remission (Panel A), and among all patients in the training and validation cohorts, according to the detection of any persistent non-*DTA* mutations during complete remission (Panel B). Panel C shows the rate of overall survival among all patients in the training and validation cohorts, according to the detection of any persistent non-*DTA* mutations during complete remission. *DTA* mutations are mutations in *DNMT3A*, *TET2*, and *ASXL1*.

To assess the reproducibility of these results, we evaluated the effect of sequencing-based detection of persistent non-*DTA* mutations during complete remission on the rates of relapse, relapse-free survival, and overall survival in the validation cohort (147 patients). The rates with detection versus no detection were as follows: 4-year relapse rate, 55.1% versus 26.5% (P<0.001); 4-year rate of relapse-free survival, 35.6% versus 60.6% (P<0.001); and 4-year rate of overall survival, 37.1% versus 67.6% (P<0.001) (Fig. 2B and 2C, and Fig. S6 in the Supplementary Appendix). The results in the validation cohort confirmed the significant findings in the training cohort.

In the combined training and validation cohorts (a total of 430 patients), persistent non-*DTA* mutations were detected during complete remission in 28.4% of the patients. Detection of these mutations was associated with a significantly higher 4-year relapse rate than no detection (55.4% vs. 31.9%; hazard ratio, 2.14; 95% confidence interval [CI], 1.57 to 2.91; P<0.001), as well as with lower 4-year rates of relapse-free survival (36.6% vs. 58.1%; hazard ratio for relapse or death, 1.92; 95% CI, 1.46 to 2.54; P<0.001) and overall survival (41.9% vs. 66.1%; hazard ratio for death, 2.06; 95% CI, 1.52 to 2.79; P<0.001) (Fig. S6 in the Supplementary Appendix).

Multivariate and Sensitivity Analyses

We performed multivariate analyses that accounted for the major established relevant prognostic factors, including age, white-cell count, 2017 European Leukemia Network risk classification, and the number of cycles of induction chemotherapy needed to attain complete remission. Sequencing-based detection of non-*DTA* mutations maintained significant independent prognostic value with respect to the rates of relapse (hazard ratio, 1.89; 95% CI, 1.34 to 2.65; P<0.001), relapse-free survival (hazard ratio for relapse or death, 1.64; 95% CI, 1.22 to 2.20; P = 0.001), and overall survival (hazard ratio for death, 1.64; 95% CI, 1.18 to 2.27; P = 0.003) (Table 2). No significant interactions were apparent between the detection of residual disease and the other prognostic factors in the multivariate model, type of consolidation therapy, or disease entity (AML vs. refractory anemia with excess of blasts) (data not shown).

In sensitivity analyses involving correction for variation in the time at which bone marrow specimens were obtained for sequencing analysis (within the remission period of 21 days to 4 months after the second treatment cycle), the prognostic value of sequencing-based detection of non-*DTA* mutations with respect to the rates of relapse, relapse-free survival, and overall survival remained unaffected (Table S2 in the Supplementary Appendix). In addition, an analysis that included post remission treatment with allogeneic stem cell transplantation as a time-dependent variable conferred no effect on the prognostic value of the detection of residual disease (Table S3 in the Supplementary Appendix).

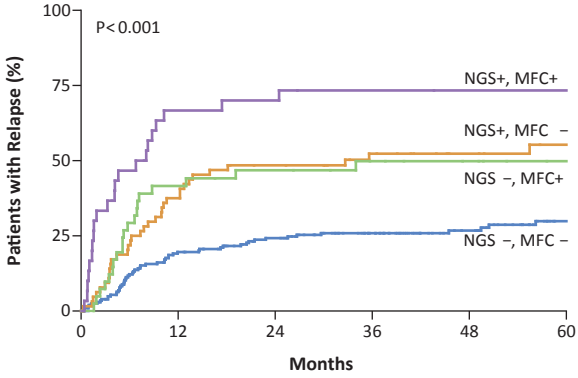
Detection of Residual Disease with Multiparameter Flow Cytometry

Multiparameter flow cytometry is an increasingly used method for predicting relapse in patients with AML who are in complete remission.[7, 24] We compared next-generation sequencing for the detection of persistent non-*DTA* mutations with flow cytometry for the detection of residual disease in a representative subgroup of 340 patients, from whom sufficient samples were obtained for both analyses. Concordant results (either detection or no detection on both assays) were found in 69.1% of the patients (30 patients with detection and 205 with no detection), whereas persistent non-*DTA* mutations were detected only on sequencing in 64 patients and only on flow cytometry in 41 patients. The 4-year relapse rate was 73.3% among patients in whom both assays were positive, 52.3% among those who had residual disease on sequencing but not on flow cytometry, 49.8% among those who had residual disease on flow cytometry but not on sequencing, and 26.7% among those in whom both assays were negative (Fig. 3). In a multivariate analysis that combined the results of sequencing and flow cytometry, the combined use of the two assays for the detection of residual disease conferred independent prognostic value with respect to the rates of relapse (P<0.001), relapse-free (P<0.001), and overall survival (P = 0.003) (Table S4 in the Supplementary Appendix).

Table 2. Multivariate Analysis of Prognostic Factors for Relapse, Relapse-free Survival, and Overall Survival.

Table 2. Multivariate Analysis of Prognostic Factors for Relapse, Relapse-free Survival, and Overall Survival.

Prognostic Factor	Relapse		Relapse-free Survival		Overall Survival	
	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value
Molecular minimal residual disease: detection vs. no detection	1.89 (1.34–2.65)	<0.001	1.64 (1.22–2.20)	0.001	1.64 (1.18–2.27)	0.003
Age: per year	1.01 (0.99–1.03)	0.21	1.02 (1.00–1.03)	0.009	1.03 (1.01–1.04)	0.001
White-cell count per microliter at diagnosis: >100,000 vs. ≤100,000	2.16 (1.31–3.56)	0.003	2.03 (1.34–3.08)	0.001	2.02 (1.27–3.21)	0.003
2017 European Leukemia Network risk classification						
Intermediate vs. favorable	1.67 (1.12–2.49)	0.01	2.01 (1.42–2.83)	<0.001	2.53 (1.72–3.72)	<0.001
Adverse vs. favorable	1.83 (1.26–2.66)	0.002	2.21 (1.58–3.10)	<0.001	2.67 (1.83–3.92)	<0.001
Number of chemotherapy cycles to attain complete remission: 2 cycles vs. 1 cycle	2.17 (1.50–3.15)	<0.001	2.43 (1.74–3.39)	<0.001	2.96 (2.09–4.21)	<0.001



No. at Risk						
NGS+, MFC+	30	8	7	5	4	4
NGS −, MFC+	41	22	18	14	11	7
NGS+, MFC −	64	39	30	22	15	11
NGS −, MFC −	205	153	130	101	69	42

Figure 3. Rate of Relapse According to Results of Next-Generation Sequencing and Multiparameter Flow Cytometry.

Shown is the cumulative incidence of relapse, according to the presence of positive (+) or negative (−) results for the detection of persistent non-*DTA* mutations during complete remission on next-generation sequencing (NGS) and on multiparameter flow cytometry (MFC).

DISCUSSION

In addition to the presence of genetic abnormalities at diagnosis, the continued presence of particular gene mutations during or after treatment carries prognostic information for certain genetically defined AML subtypes.[5-11] This applies, for example, to AML associated with a mutation in *NPM1*, for which the detection of a residual mutation in *NPM1* transcripts during complete remission is indicative of an increased probability of relapse.[8, 9] However, this example is only representative of a single-gene approach. We report the results of a systematic study that involved a large number of patients with AML, in which we used a genome-wide approach to evaluate the persistence of multiple gene mutations for the effect on treatment outcomes. Patients were treated with intensive chemotherapy regimens and attained morphologic complete remission, with a median follow-up exceeding 3 years.

Of note, age-related clonal hematopoiesis,[14-17] which is characterized by recurrent somatic mutations in leukemia-associated genes in persons with no apparent hematologic disease, adds a challenge in the detection of residual disease. Our study showed that the persistence of mutations that are most commonly associated with age-related clonal hematopoiesis (i.e., *DTA* mutations [mutations in *DNMT3A*, *TET2*, and *ASXL1*]) during complete remission did not contribute to a measurably increased risk of relapse within a follow-up period of 4 years in adults with AML who were younger than 65 years of age. This appeared to be true for mutations that were present at various allele frequencies, which suggests that the clone size in age-related clonal hematopoiesis yields no prognostic value with respect to the end points defined in this study.

The cells bearing *DTA* mutations appeared to persist and possess a selective clonal advantage over normal stem cells when they repopulated the bone marrow after induction therapy. This finding is consistent with the competitive clonal advantage of hematopoietic stem cells with deficiencies and mutations in *DNMT3A* and *TET2*, an advantage that has been reported previously.[25-27] The proliferative advantage of hematopoietic stem cells with *DTA* mutations and their capacity to withstand chemotherapy because of inherent resistance may explain why persistent premalignant *DTA* mutations were not correlated with an increased probability of relapse and thereby did not constitute a reliable molecular biomarker for the assessment of relapse risk.

It is possible that gene mutations other than *DTA* mutations also partially reflect clonal hematopoiesis. However, at this time, we cannot rigorously verify the possibility that gene mutations associated with age-related clonal hematopoiesis also reside as sub-fractions among the other gene abnormalities in leukemia cells. In addition, mutations in *TP53*, *IDH1*, and *IDH2*, along with genes related to the RAS pathway and spliceosome genes, have been shown to have distinct biologic features in the context of AML pathogenesis.[28-31] Therefore, in this study, we collectively considered non-*DTA* mutations to be abnormalities that are unrelated to clonal hematopoiesis.

Our study had a median follow-up of almost 40 months. Among patients with AML who have complete remission, most relapses generally occur within the first 4 years. We found that the continued persistence of *DTA* mutations was not associated with an increased relapse risk, and thus these residual cells may not need to be eliminated to prevent relapse. However, the limited follow-up of 40 months does not rule out the possibility that persistent *DTA* mutations represent an increased risk of relapse at a later time point.

Although sequencing-based detection enables assessment for residual disease in virtually all patients with AML, it is imperfect in two ways. First, not all patients with residual mutation-bearing cells have a relapse. Second, some patients with no measurable residual disease have a relapse. It is conceivable that relapse estimation can be improved with the development of technological variations of sequencing-based approaches that have greater sensitivity or a broader scope (e.g., those with molecular barcoding, exome sequencing, or whole-genome sequencing) or with the identification of additional molecular and phenotypic markers so that quantitative minor clones or subclones associated with the leukemia are captured by the assay. In this respect, it is of particular interest that the use of multiparameter flow cytometry[7, 24] which identifies patients with AML who have an increased risk of relapse according to an entirely different approach that is based on a residual leukemia-associated immunophenotype[23, 32] can increase the yield of identification of residual leukemia during complete remission.

In this study, gene sequencing and multiparameter flow cytometry each had independent and additive prognostic value with respect to rates of relapse and survival in patients with AML. The detection of residual leukemia with both methods is associated with an excessively high probability of relapse (approximately 75%), and the absence of detection of residual disease with both methods is correlated with a relatively low probability of relapse (approximately 25%). Thus, the combined use of sequencing and flow cytometry during complete remission warrants further development and evaluation in clinical practice.

In conclusion, targeted sequencing-based detection of molecular minimal residual disease during complete remission was associated with an increased risk of relapse or death in patients with AML. However, over a 4-year follow-up period, the risk of relapse or death was not influenced by the persistence of genetic lesions that are associated with age-related clonal hematopoiesis.

Supported by grants from the Queen Wilhelmina Fund Foundation of the Dutch Cancer Society (EMCR 2015-7525) and the Netherlands Organization for Health Research and Development (846002002). Dr. Sanders is supported by a Rubicon fellowship from the Netherlands Organization for Scientific Research (019.153LW.038), and Mr. al Hinai is a recipient of a Ph.D. scholarship from the Ministry of Health, Oman.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank all the participating centers of the Dutch–Belgian Cooperative Trial Group for Hematology–Oncology (HOVON) and Swiss Group for Clinical Cancer Research (SAKK), where the clinical trials that formed the basis for this study were conducted; H. Berna Beverloo for performing cytogenetic analyses, Eric Bindels for performing next-

generation sequencing, Remco Hoogenboezem for assisting with bioinformatics, and Egied Simons for assisting with preparation of the figures (Erasmus University Medical Center); and Vincent J. van der Velden (Erasmus University Medical Center), Jennichjen Slomp (Medisch Spectrum Twente, Medlon), and Frank Preijers (Radboud University Medical Center) for performing flow cytometry.

APPENDIX

The authors' full names and academic degrees are as follows: Mojca Jongen-Lavrencic, M.D., Ph.D., Tim Grob, M.D., Diana Hanekamp, M.Sc., François G. Kavelaars, Adil al Hinai, M.Sc., Annelieke Zeilemaker, Claudia A.J. Erpelinck-Verschueren, Patrycja L. Gradowska, Ph.D., Rosa Meijer, Ph.D., Jacqueline Cloos, Ph.D., Bart J. Biemond, M.D., Ph.D., Carlos Graux, M.D., Ph.D., Marinus van Marwijk Kooy, M.D., Markus G. Manz, M.D., Thomas Pabst, M.D., Ph.D., Jakob R. Passweg, M.D., Ph.D., Violaine Havelange, M.D., Ph.D., Gert J. Ossenkoppele, M.D., Ph.D., Mathijs A. Sanders, Ph.D., Gerrit J. Schuurhuis, Ph.D., Bob Löwenberg, M.D., Ph.D., and Peter J.M. Valk, Ph.D.

The authors' affiliations are as follows: the Department of Hematology (M.J.-L., T.G., F.G.K., A.H., A.Z., C.A.J.E.-V., M.A.S., B.L.,P.J.M.V.) and HOVON Data Center, Department of Hematology (P.L.G., R.M.), Erasmus University Medical Center, Erasmus MC Cancer Institute, Rotterdam, the Department of Hematology, VU University Medical Center (D.H., J.C., G.J.O., G.J.S.), and the Department of Hematology, Academic Medical Center (B.J.B.), Amsterdam, and Isala Hospital, Zwolle (M.M.K.) — all in the Netherlands; UCL Namur (Godinne), Yvoir (C.G.), and the Department of Hematology, Cliniques Universitaires Saint-Luc, Brussels (V.H.) — both in Belgium; and the Department of Hematology, University Hospital Zurich, Zurich (M.G.M.), University Hospital, Bern (T.P.), and the Division of Hematology, University Hospital Basel, Basel (J.R.P.) — all in Switzerland.

REFERENCES

1. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ*, et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013 May 30; 368(22): 2059-2074.
2. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. *N Engl J Med* 2015 Sep 17; **373**(12)**:** 1136-1152.
3. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND*, et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 2016 Jun 9; **374**(23)**:** 2209-2221.
4. Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T*, et al.* Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017 Jan 26; **129**(4)**:** 424-447.
5. Grimwade D, Freeman SD. Defining minimal residual disease in acute myeloid leukemia: which platforms are ready for "prime time"? *Blood* 2014 Nov 27; **124**(23)**:** 3345-3355.
6. Kayser S, Walter RB, Stock W, Schlenk RF. Minimal residual disease in acute myeloid leukemia--current status and future perspectives. *Curr Hematol Malig Rep* 2015 Jun; **10**(2)**:** 132-144.
7. Hourigan CS, Gale RP, Gormley NJ, Ossenkoppele GJ, Walter RB. Measurable residual disease testing in acute myeloid leukaemia. *Leukemia* 2017 Jul; **31**(7)**:** 1482-1490.
8. Ivey A, Hills RK, Simpson MA, Jovanovic JV, Gilkes A, Grech A*, et al.* Assessment of Minimal Residual Disease in Standard-Risk AML. *N Engl J Med* 2016 Feb 4; **374**(5)**:** 422-433.
9. Kronke J, Schlenk RF, Jensen KO, Tschurtz F, Corbacioglu A, Gaidzik VI*, 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 Jul 1; **29**(19)**:** 2709-2716.
10. Gaidzik VI, Weber D, Paschka P, Kaumanns A, Krieger S, Corbacioglu A*, et al.* DNMT3A mutant transcript levels persist in remission and do not predict outcome in patients with acute myeloid leukemia. *Leukemia* 2018 Jan; **32**(1)**:** 30-37.
11. Bhatnagar B, Eisfeld AK, Nicolet D, Mrozek K, Blachly JS, Orwick S*, et al.* Persistence of DNMT3A R882 mutations during remission does not adversely affect outcomes of patients with acute myeloid leukaemia. *Br J Haematol* 2016 Oct; **175**(2)**:** 226-236.
12. Pastore F, Levine RL. Next-Generation Sequencing and Detection of Minimal Residual Disease in Acute Myeloid Leukemia: Ready for Clinical Practice? *JAMA* 2015 Aug 25; **314**(8)**:** 778-780.
13. Klco JM, Miller CA, Griffith M, Petti A, Spencer DH, Ketkar-Kulkarni S*, et al.* Association Between Mutation Clearance After Induction Therapy and Outcomes in Acute Myeloid Leukemia. *Jama-J Am Med Assoc* 2015 Aug 25; **314**(8)**:** 811-822.
14. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF*, et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine* 2014 Dec 25; **371**(26)**:** 2477-2487.
15. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG*, et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine* 2014 Dec 25; **371**(26)**:** 2488-2498.
16. Zink F, Stacey SN, Norddahl GL, Frigge ML, Magnusson OT, Jonsdottir I*, et al.* Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* 2017 Aug 10; **130**(6)**:** 742-752.
17. Abelson S, Ng SWK, Wiessbrod O, Zuzarte P, Heisler L, Sundaravadanam Y*, et al.* Progression to AML Is Predictable and Distinct from Age Related Clonal Hematopoiesis. *Blood* 2017 Dec 7; **130**.
18. Shlush LI. Age-related clonal hematopoiesis. *Blood* 2018 Feb 1; **131**(5)**:** 496-504.
19. Jan M, Ebert BL, Jaiswal S. Clonal hematopoiesis. *Semin Hematol* 2017 Jan; **54**(1)**:** 43-50.

20. Dutch–Belgian Cooperative Trial Group for Hematology–Oncology. Homepage (http://www .hovon .nl)
21. Pabst T, Vellenga E, van Putten W, Schouten HC, Graux C, Vekemans MC*, et al.* Favorable effect of priming with granulocyte colony-stimulating factor in remission induction of acute myeloid leukemia restricted to dose escalation of cytarabine. *Blood* 2012 Jun 7; **119**(23)**:** 5367-5373.
22. Lowenberg B, Pabst T, Maertens J, van Norden Y, Biemond BJ, Schouten HC*, et al.* Therapeutic value of clofarabine in younger and middle-aged (18-65 years) adults with newly diagnosed AML. *Blood* 2017 Mar 23; **129**(12)**:** 1636-1645.
23. Terwijn M, van Putten WL, Kelder A, van der Velden VH, Brooimans RA, Pabst T*, et al.* High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. *J Clin Oncol* 2013 Nov 1; **31**(31)**:** 3889-3897.
24. Schuurhuis GJ, Heuser M, Freeman S, Bene MC, Buccisano F, Cloos J*, et al.* Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood* 2018 Mar 22; **131**(12)**:** 1275-1291.
25. Challen GA, Sun D, Jeong M, Luo M, Jelinek J, Berg JS*, et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet* 2011 Dec 04; **44**(1)**:** 23-31.
26. Moran-Crusio K, Reavie L, Shih A, Abdel-Wahab O, Ndiaye-Lobry D, Lobry C*, et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 2011 Jul 12; **20**(1)**:** 11-24.
27. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V*, et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* 2014 Feb 20; **506**(7488)**:** 328-333.
28. McKerrell T, Park N, Moreno T, Grove CS, Ponstingl H, Stephens J*, et al.* Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep* 2015 Mar 3; **10**(8)**:** 1239-1245.
29. Wong TN, Ramsingh G, Young AL, Miller CA, Touma W, Welch JS*, et al.* Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. *Nature* 2015 Feb 26; **518**(7540)**:** 552-555.
30. Lindsley RC, Saber W, Mar BG, Redd R, Wang T, Haagenson MD*, et al.* Prognostic Mutations in Myelodysplastic Syndrome after Stem-Cell Transplantation. *N Engl J Med* 2017 Feb 9; **376**(6)**:** 536-547.
31. Desai P, Mencia-Trinchant N, Savenkov O, Simon MS, Cheang G, Lee S*, et al.* Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nature Medicine* 2018 2018/07/01; **24**(7)**:** 1015-1023.
32. Ravandi F, Jorgensen J, Borthakur G, Jabbour E, Kadia T, Pierce S*, et al.* Persistence of minimal residual disease assessed by multiparameter flow cytometry is highly prognostic in younger patients with acute myeloid leukemia. *Cancer* 2017 Feb 1; **123**(3)**:** 426-435.

SUPPLEMENTARY APPENDIX

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Jongen-Lavrencic M, Grob T, Hanekamp D, et al. Molecular minimal residual disease in acute myeloid leukemia. N Engl J Med 2018;378:1189-99. DOI: 10.1056/NEJMoa1716863

SUPPLEMENTARY METHODS

PATIENTS AND CELL SAMPLES

Bone marrow aspirations or peripheral blood samples at diagnosis were taken after informed consent. Follow-up bone marrow samples of 430 out of 482 AML or RAEB patients in CR with mutations at diagnosis (Figure S1) were taken at least 21 days after the start of the second induction cycle. If additional samples were available, the most recent sample prior to start of consolidation therapy was selected. In case no consolidation therapy was given, the last sample that was available within a four month interval from start of the second induction cycle was selected. The probabilities of relapse in AML patients with or without available samples did not differ (p=0.281). The median follow-up of the 430 AML cases was 39.7 months and the residual disease status was not available to the clinical investigator and did not influence the choice of consolidation therapy. Blasts and mononuclear cells at diagnosis were purified by Ficoll-Hypaque (Nygaard, Oslo, Norway) density gradient centrifugation and cryopreserved. Of all 430 AML cases reaching CR, white blood cells were isolated after induction treatment in 385 cases and mononuclear cells were subsequently purified in 45 cases. After thawing, cells were lysed in RLT solution with the addition of DTT (Qiagen, Venlo, The Netherlands).

DNA AND RNA ISOLATION

High quality DNA was extracted using the QIAsymphony (Qiagen, Venlo, The Netherlands). DNA concentration was measured by Qubit Fluorometric Quantitation (Thermo Fisher Scientific, Wilmington, DE). RNA was isolated with RNA-Bee following the protocols of the manufacturer (Bio-Connect BV, Huissen, The Netherlands). *CBFB-MYH11*, *RUNX1-RUNX1T1*, *FLT3* internal tandem duplication (ITD) and *CEBPA* mutations were determined as described previously.[1, 2]

TARGETED NGS

The NGS libraries were paired-end sequenced (2x221bp) on an Illumina HiSeq 2500 System (Illumina, San Diego, CA) in Rapid Run mode. Since CBF fusion transcripts, *CEBPA* mutations and *FLT3* ITDs cannot be reliably assessed with NGS on DNA, these molecular aberrations were excluded from the analyses.

NGS DATA ANALYSIS

The vast majority of amplicon target regions were completely paired-end sequenced. Overlap-based error-correction was utilized to attenuate any form of strand-specific error biases. Error-corrected paired-end reads aligned to the human genome version 19 (hg19) with BBMAP[3] followed by quality control to determine cases with insufficient number of reads for adequate variant calling. Single nucleotide variants (SNVs) and insertions-deletions

(indels) at diagnosis were determined by MuTect[4], Samtools[5], GATK[6], Varscan[7], Indelocator[8] and Pindel[6]. Variant allele frequencies (VAF) of mutations detected at diagnosis were calculated as the ratio between the number of mutant and total reads. The persistence of mutations at follow-up, previously detected at diagnosis, requires the detection of mutations at exceptionally low VAFs. A follow-up background error model was determined by calculating the VAF for each potential SNV within the set of target genes across all follow-up samples.

The detection of variants at low detection level is primarily reserved for highly discriminative insertion or deletion mutations sequenced at sufficient depth in both the follow-up samples of interest and the control set of remission samples. The strength of a site-specific error model is that it models the unique site-and-variant specific noise profile based on a large set of remission samples from patients who did not carry that specific mutation at diagnosis and thereby enables the assessment whether the variant remains persistent, defined as a statistical outlier, in the follow-up sample from the patient of interest. Since more complex insertion or deletion mutations are particularly distinct and the odds of detecting such variants as a consequence of sequencing or alignment errors is exceptionally low, such variants can be detected at higher sensitivity. The detection sensitivity of other mutations is variable and highly dependent on the average coverage for that specific locus for all samples, the observed error variance of the site-specific variant in the control set (a high variance results in decreased detection sensitivity) and the number of control sample available. The unique combination of patient-specific mutations, the application of a site-specific error model and strict detection criteria minimizes the odds that variants are erroneously called to persist.

Quantile normalization of the calculated VAFs was performed per flow cell to mitigate the effect of qualitative differences amongst samples. All SNVs detected across the diagnostic samples were compiled and the background VAF distribution was determined for each individual SNV from follow-up samples lacking this SNV in the matched diagnostic sample. For the remaining follow-up samples the persistence of the SNV was considered confirmed when the VAF was an outlier compared to the background VAF distribution according to the Thompson-Tau test. A one-sided p-value <0.01 was considered statistically significant. Indels were processed and compared similarly, except for quantile normalization as there are infinitely many possible indel-configurations per locus.

MULTI-PARAMETER FLOW CYTOMETRY

Residual disease detection by MFC was performed as described previously[9]. The residual disease percentage was defined as the number of leukemia-associated immuno phenotype (LAIP) cells within the total white blood cell compartment. The threshold between residual and no residual disease based on flow cytometry was established and validated on 0.1%.[9] Multi-parameter flow cytometry was carried out in a subset of 340 of the 430 AML cases

analyzed by NGS, which did not differ significantly from the total series of cases regarding clinical, cytogenetic and molecular characteristics (data not shown).

STATISTICAL ANALYSES

The complete cohort of 430 AML patients was randomly split using Stata into a training (n=283) and validation cohort (n=147) (Table 1, Figure S1 and Table S1). Each patient received a pseudorandom number from a uniform distribution from 0 to 1. The random numbers generated were shuffled by sorting, allowing for random allocation of patients to the training or validation set. Differences in clinical, cytogenetic and molecular characteristics of the training and validation cohorts or NGS and flow cohorts were tested using the Fisher's exact test for categorical variables and Mann-Whitney U test for continuous variables. Clinical, cytogenetic and molecular characteristics of the training and validation cohorts were not significantly different (Table S1). The primary endpoint of the study was the cumulative incidence of relapse (CIR). Competing-risks regression analysis was performed for relapse with adjustment for non-relapse mortality according to the method of Gray and the Fine & Gray model[10]. The secondary endpoints were relapse free (RFS) and overall survival (OS) which were analyzed using the log-rank test and the Cox proportional hazards model. Relapse and survival time was calculated from the sampling date until the date of the event of interest or censoring. RFS was defined from date of sampling to death, relapse or censoring, whichever came first. All statistical tests were two-sided and p-values <0.05 were considered statistically significant. The proportional hazards assumption was tested by interaction with time and the interactions were evaluated in a standard way. The effect of allogeneic stem cell transplantation on CIR and OS was investigated in both multivariable models as a time-dependent covariat[11]. All p-values are two sided and p-values <0.05 were considered statistically significant. Statistical analyses were performed with Stata Statistical Software, Release 14.1 (Stata, College Station, TX).

SUPPLEMENTARY FIGURES

Figure S1: Consort diagram molecular residual disease study
Abbreviations: **HO**, HOVON-SAKK, Dutch-Belgian Hemato-Oncology Cooperative Group and the Swiss Group for Clinical Cancer Research; **CR**, Complete morphological Remission; **FU2**, Follow-up after induction cycle II; **NGS**, Next Generation Sequencing.

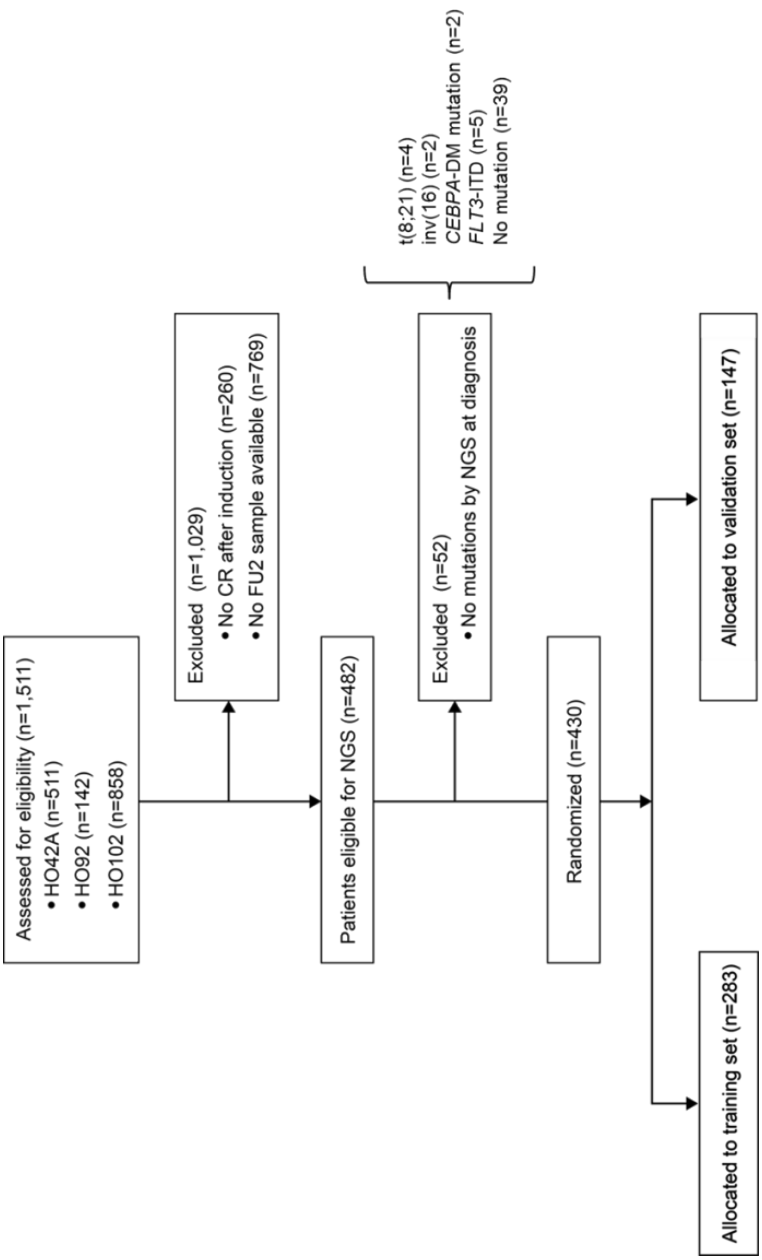


Figure S2: Mutation status at diagnosis and after induction cycle II with various VAF cut-offs.
Overview of the mutation status at diagnosis (present: blue) and after induction cycle II (present: red). Each column represents an individual patient. The upper panel indicates the ELN 2017 risk category (green: favorable risk, orange: intermediate risk, red: adverse risk). The total number of mutations present at diagnosis and the number remaining after induction cycle II are indicated on the right. The total number of residual mutations after induction treatment are summarised in the bottom figure (green: number of residual mutations present after induction treatment; orange: number of residual mutations exclusively in *DNMT3A*, *TET2* and/or *ASXL1*). Figure 2A-E represent the mutation status with distinct VAF cut offs: no cut off (2A), <10% (2B), <5% (2C), <2.5% (2D) and <1.0% (2E). Abbreviations: ELN, European LeukemiaNet. VAF, Variant Allele Frequency.

Figure S2A: **Mutation status at diagnosis and after induction cycle II without VAF cutoff.**

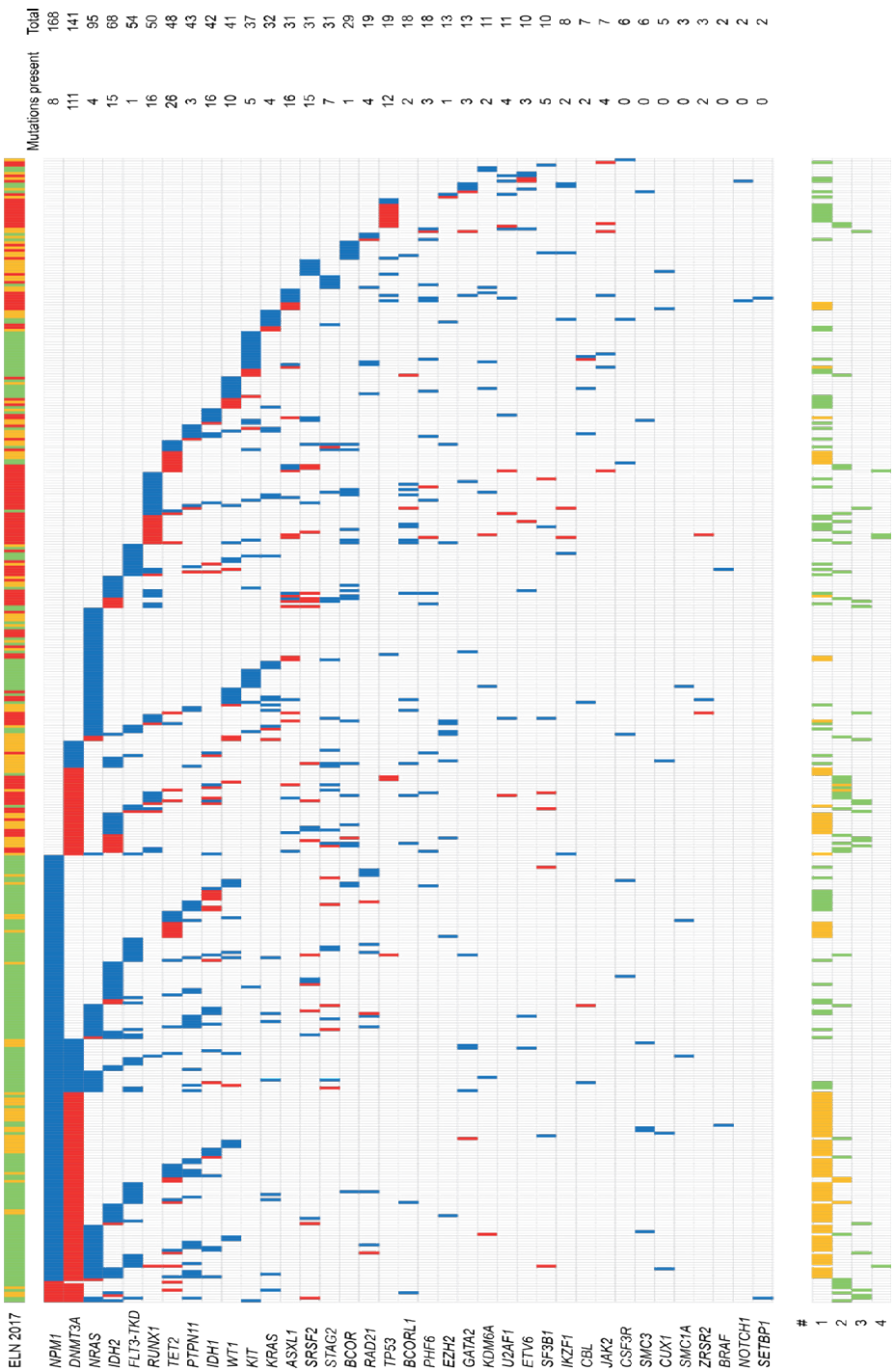


Figure S2B: **Mutation status at diagnosis and after induction cycle II VAF cut-off ≤10%.**

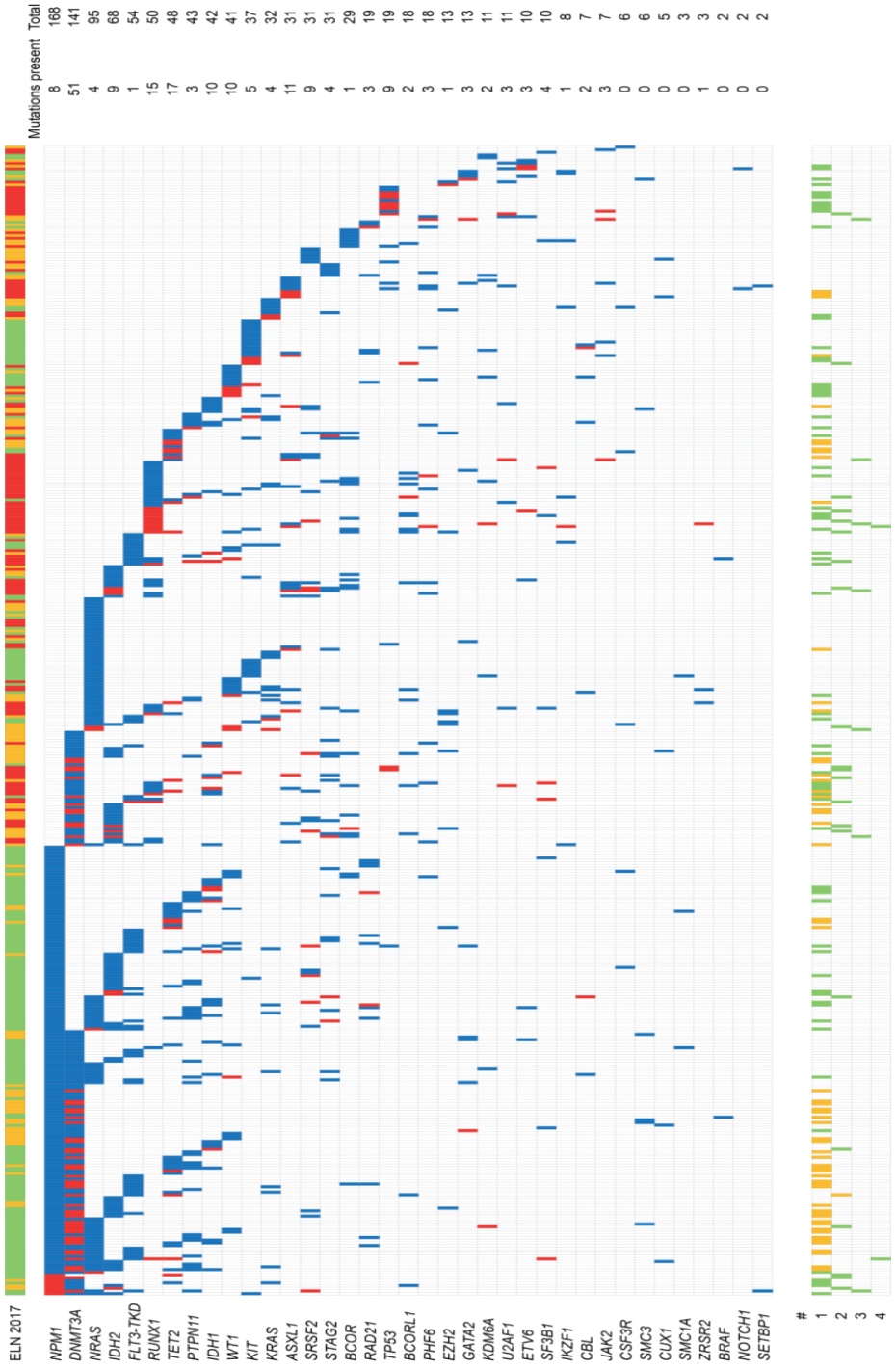


Figure S2C: **Mutation status at diagnosis and after induction cycle II VAF cut-off ≤5%.**

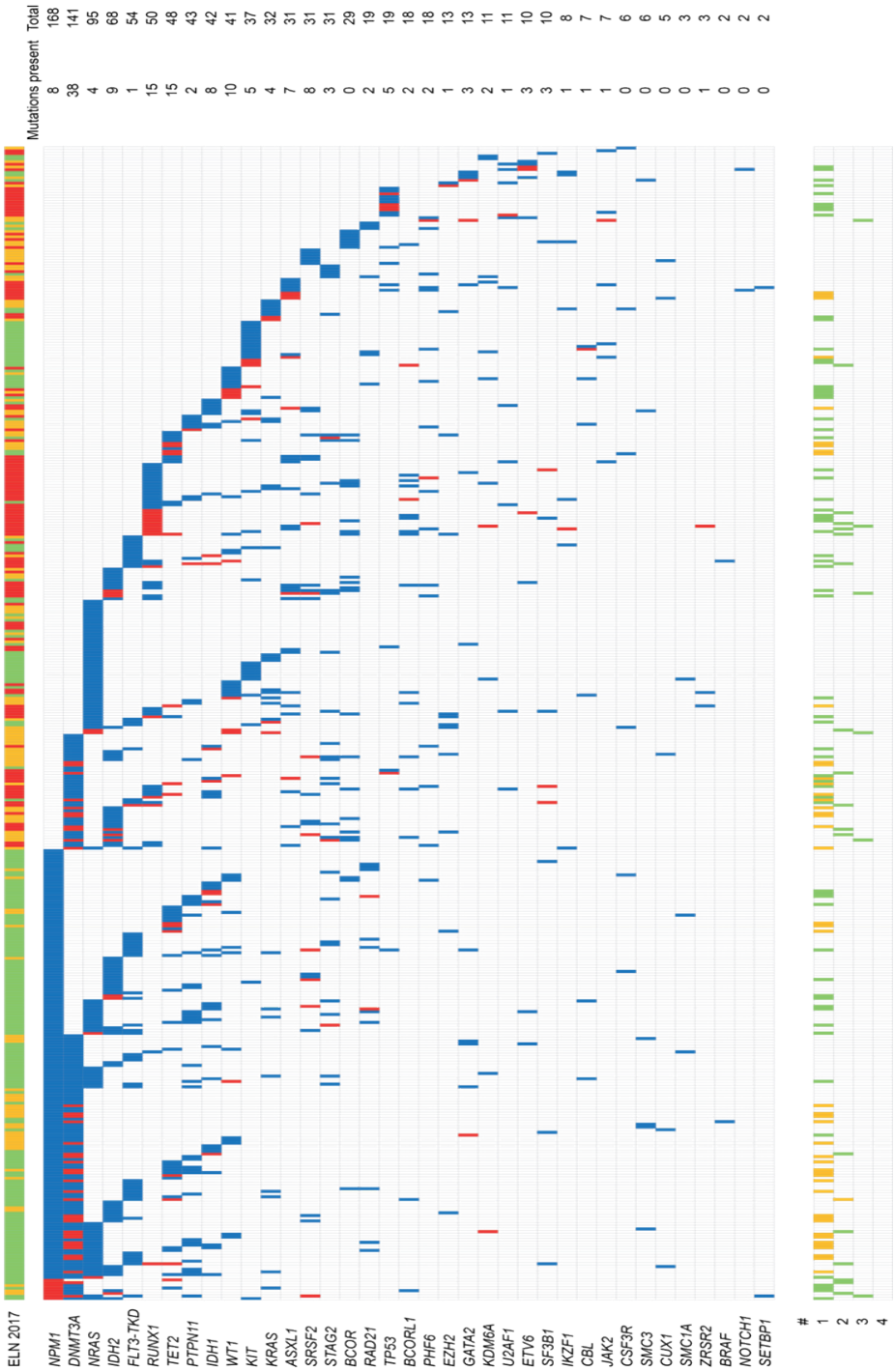


Figure S2D: **Mutation status at diagnosis and after induction cycle II VAF cut-off ≤2.5%.**

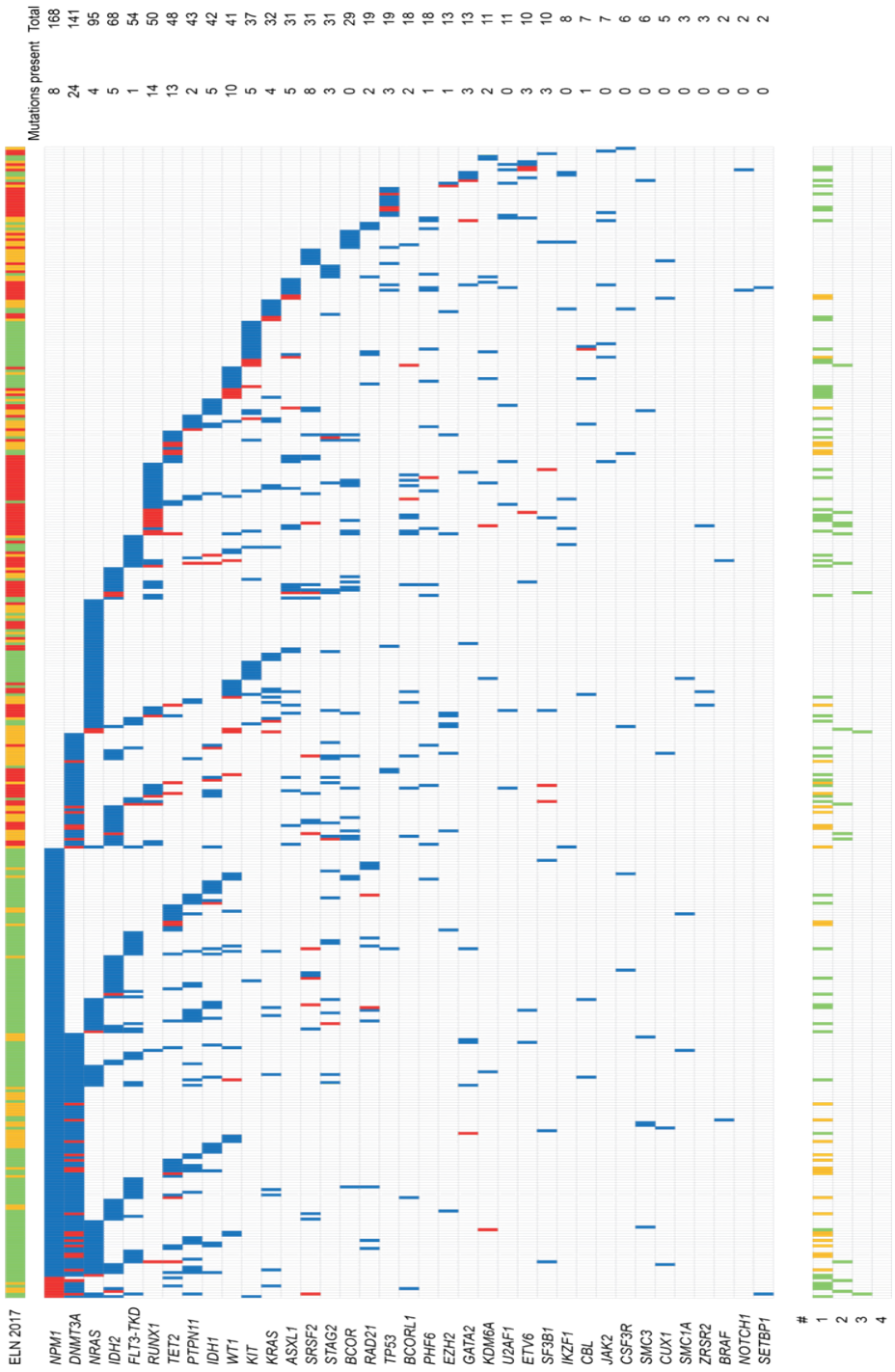


Figure S2E: **Mutation status at diagnosis and after induction cycle II VAF cut-off ≤1%.**

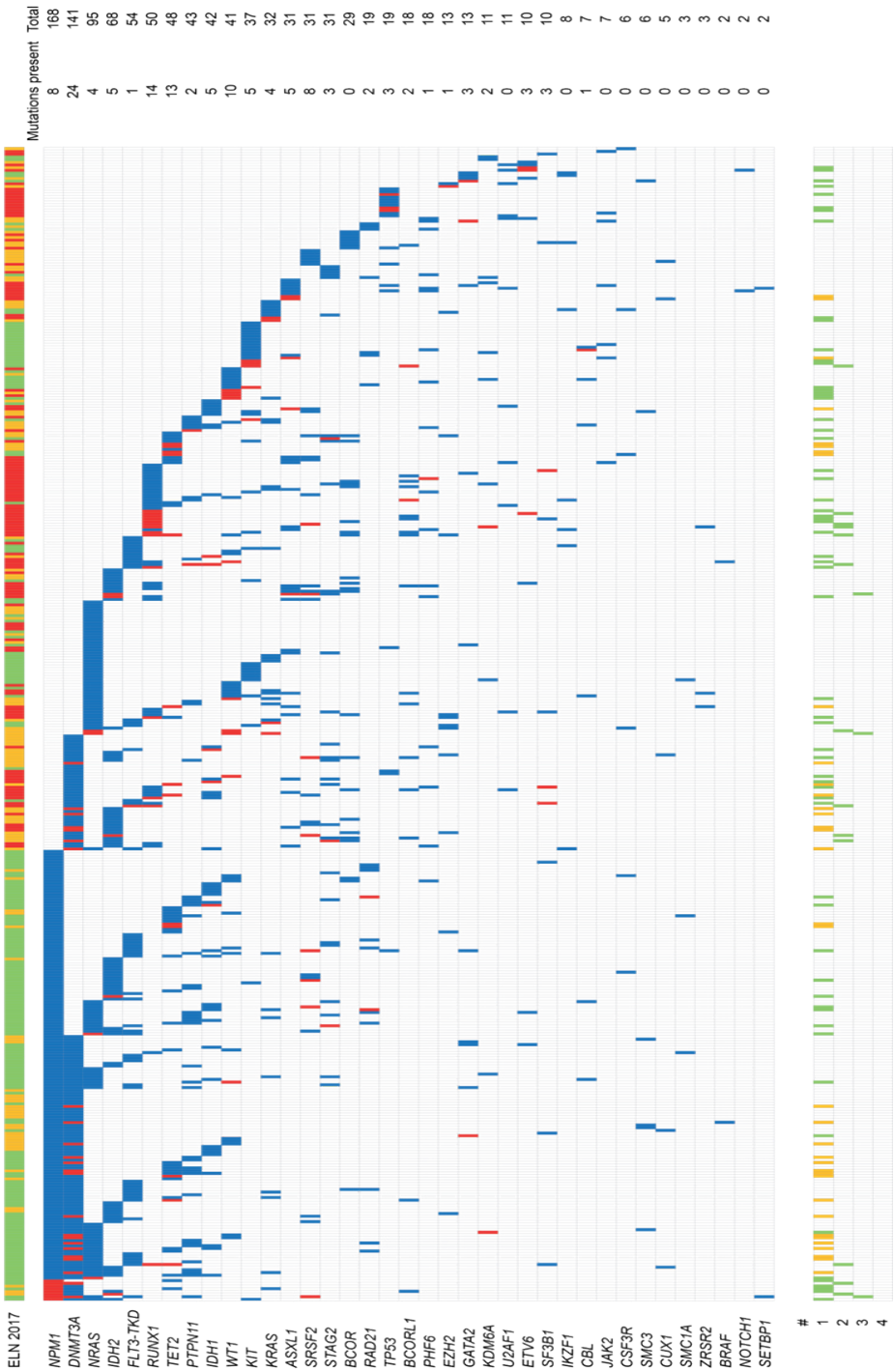


Figure S3: **Total number of mutations at diagnosis and after induction cycle II with various VAF cut-offs.** (blue: at diagnosis, red: without cut off, green: cut off 10%, purple: cut off 5%, black: cut off 2.5%, orange: cut off 1% and light blue: cut off 0.1%).

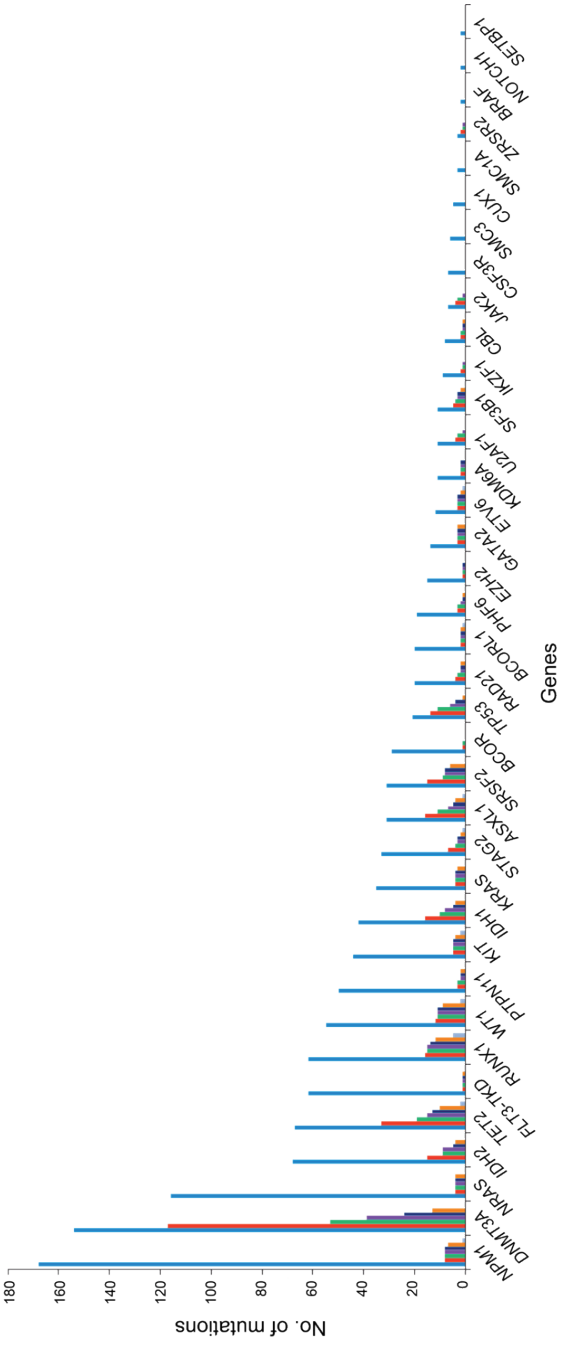


Figure S4: **Cumulative incidence of relapse by any detectable mutation in CR with various VAF cut offs in AML patients allocated to the training set.** (no cut off (A), ≤30% (B), ≤20% (C), ≤10% (D) ≤5% (E), ≤2.5% (F) and ≤1% (G); solid line: mutations detectable in CR (N+); dashed line: mutations not detectable in CR (N-).

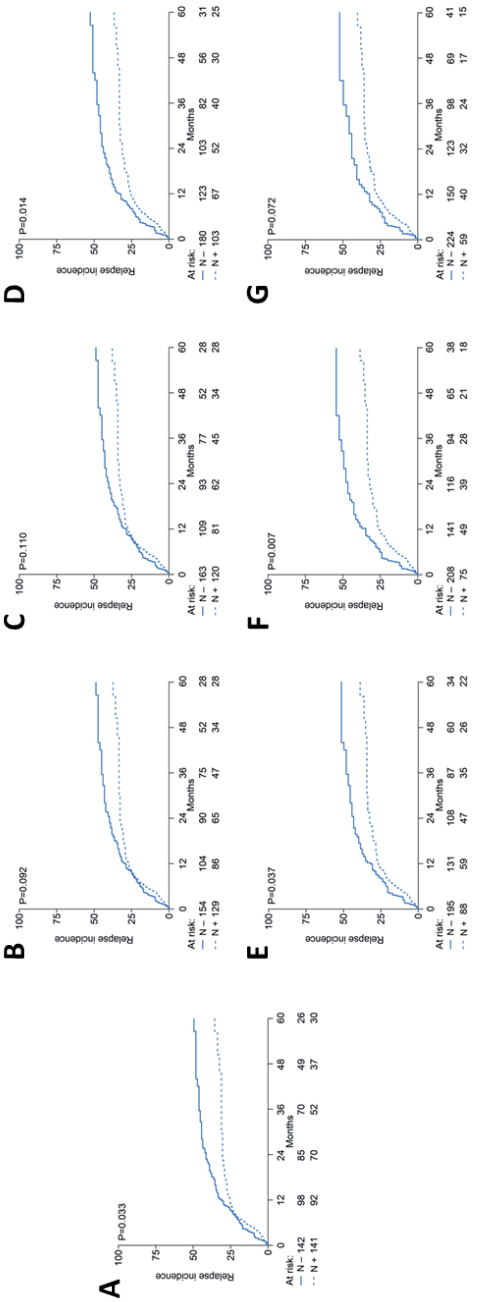
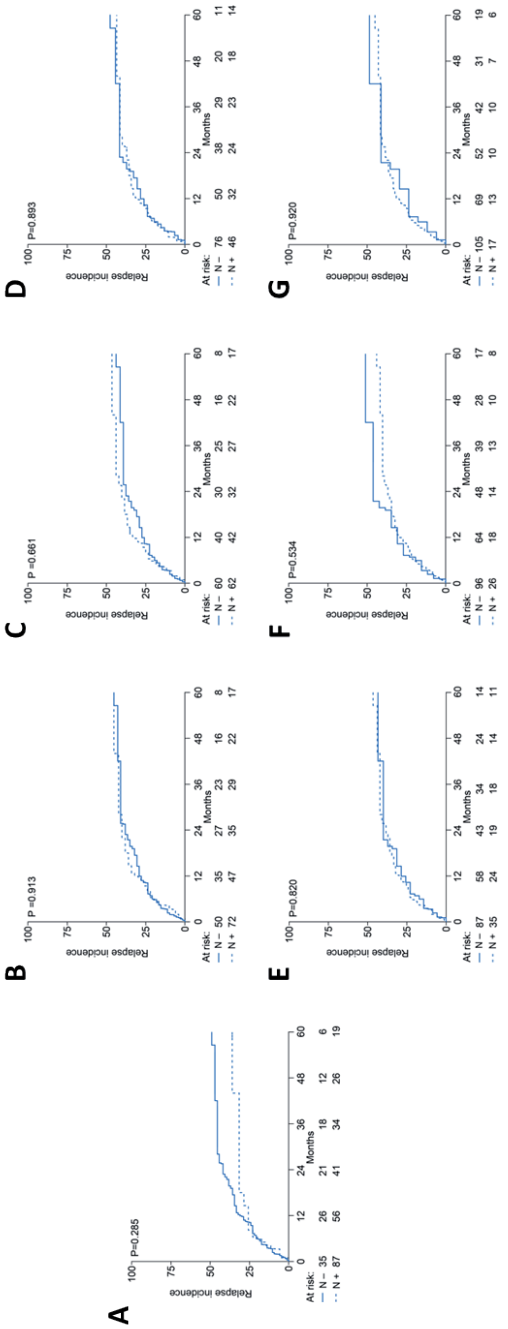
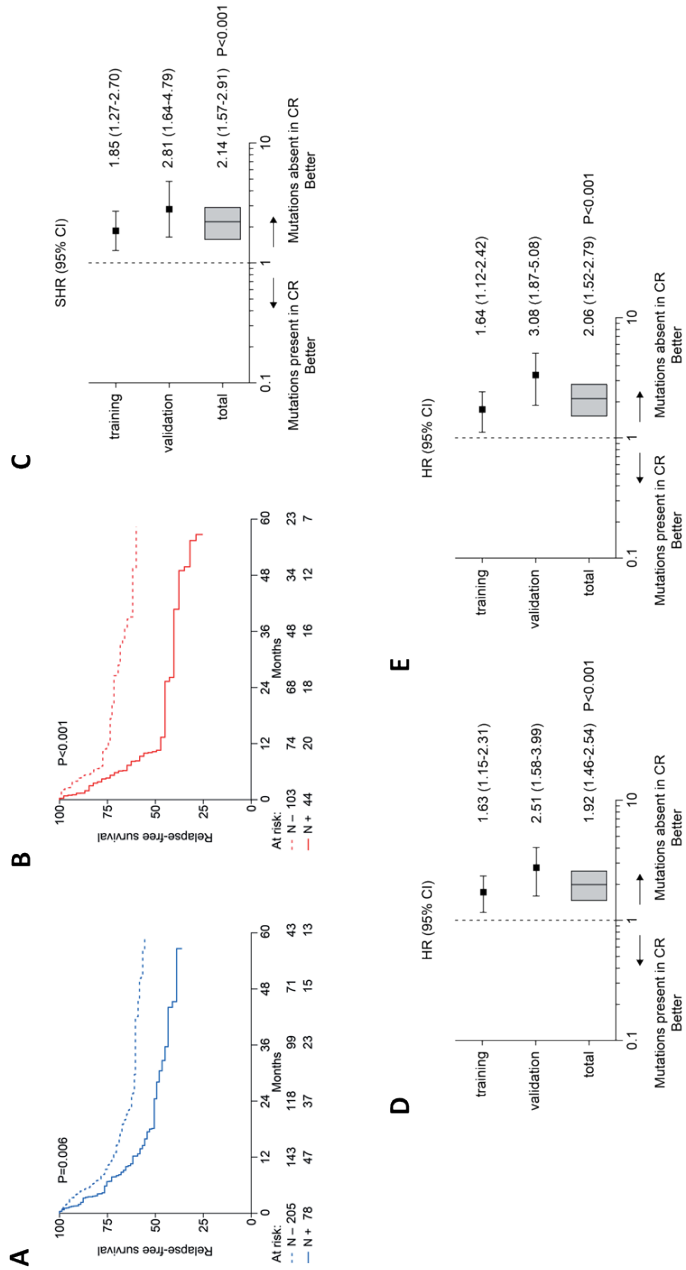


Figure S5: **Cumulative incidence of relapse by detectable *DNMT3A*, *TET2* and/or *ASXL1* mutations in CR with VAF cut offs in *DNMT3A*, *TET2* and/or *ASXL1* mutant AML patients allocated to the training set.** No cut off (A), ≤30% (B), ≤20% (C), ≤10% (D), ≤5% (E), ≤2.5% (F) and ≤1.0% (G); solid line: *DTA* mutations detectable (N+); dashed line: *DTA* mutations not detectable (N-).



Supplementary Figure S6: **Relapse free survival analyses of non-*DTA* persisting mutations in CR in AML patients allocated to the training AML cohort and the validation cohort.** Cumulative incidence of relapse curves of (A) relapse free survival training cohort and (B) relapse free survival validation cohort. Solid line: non-*DTA* mutations detectable (N+); dashed line: non-*DTA* mutations not detectable (N-). Forest plots of (C) relapse, (D) relapse free survival and (E) overall survival, including training, validation and summary measures. In Figures S6A-E all *DTA* persisting mutations are excluded at any VAF.



SUPPLEMENTARY TABLES

Table S1A. Clinical patient characteristics by training and validation cohort.

	Training cohort (N=283)		Validation cohort (N=147)		Total (N=430)		P-value
Age							0.572
Median	51		50		51		
Range	18-66		18-65		18-66		
Sex							0.310
M	137	48%	79	54%	216	50%	
F	146	52%	68	46%	214	50%	
Disease entity							0.862
AML	253	89%	129	88%	382	89%	
RAEB	26	9%	14	9%	40	9%	
unknown	4	2%	1	3%	8	2%	
WBC at diagnosis ($x10^9/l$)							0.735
<=100	256	90%	131	89%	387	90%	
>100	27	10%	16	11%	43	10%	
ELN2017 risk classifiction							0.468
Favorable	140	49%	64	44%	204	48%	
Intermediate	73	26%	40	27%	113	26%	
Adverse	70	25%	43	29%	113	26%	
Number of cycles to CR							0.170
One cycle	242	86%	118	80%	360	84%	
Two cycles	41	14%	29	20%	70	16%	
Consolidation therapy							0.976
None	31	11%	15	10%	46	11%	
Chemotherapy	78	28%	39	26%	117	27%	
Autologous HSCT	52	18%	26	18%	78	18%	
Allogeneic HSCT	122	43%	67	46%	189	44%	
Trial							0.845
HO102	167	59%	87	59%	254	59%	
HO42A	86	30%	47	32%	133	31%	
HO92	30	11%	13	9%	43	10%	

Table S1B. Cytogenetics at diagnosis by training and validation cohort.*

	Training cohort (N=273)		Validation cohort (N=144)		Total (N=417)		P-value
t(8;21)							0.210
Neg	252	92%	138	96%	390	94%	
Pos	51	8%	6	4%	27	6%	
inv(16)							0.270
Neg	260	95%	133	92%	393	94%	
Pos	13	5%	11	8%	24	6%	
t(9;11)							1.000
Neg	269	99%	142	99%	411	99%	
Pos	4	1%	2	1%	6	1%	
t(6;9)							0.697
Neg	269	99%	141	98%	410	98%	
Pos	4	1%	3	2%	7	2%	
t(v;11)							0.188
Neg	271	99%	140	97%	411	99%	
Pos	2	1%	4	3%	6	1%	
abn3q26							1.000
Neg	271	99%	143	99%	414	99%	
Pos	2	1%	1	1%	3	1%	
Monosomy 5							1.000
Neg	268	98%	141	98%	409	98%	
Pos	5	2%	3	2%	8	2%	
del5q							0.119
Neg	268	98%	137	95%	405	97%	
Pos	5	2%	7	5%	12	3%	
Monosomy 7							0.372
Neg	260	95%	134	93%	394	94%	
Pos	13	5%	10	7%	23	6%	
Monosomy 17							0.524
Neg	267	98%	139	97%	406	97%	
Pos	6	2%	5	3%	11	3%	
abn17p							0.720
Neg	267	98%	142	99%	409	98%	
Pos	6	2%	2	1%	8	2%	
Complex karyotype							0.106
Neg	253	93%	126	88%	379	91%	
Pos	20	7%	18	13%	38	9%	
Monosomal karyotype							0.321
Neg	256	94%	131	91%	387	93%	
Pos	17	6%	13	9%	30	7%	

*Cytogenetics are determined by conventional karyotyping and/or FISH. Of 13 patients cytogenetic data is missing

Table S1C. Mutation status at diagnosis by training and validation cohort.

	Training cohort (N=283)		Validation cohort (N=147)		Total (N=430)		P-value
ASXL1							0.237
Neg	266	94%	133	90%	399	93%	
Pos	17	6%	14	10%	31	7%	
BCOR							0.687
Neg	265	94%	136	93%	401	93%	
Pos	18	6%	11	7%	29	7%	
BCORL1							0.447
Neg	273	96%	139	95%	412	96%	
Pos	10	4%	8	5%	18	4%	
BRAF							0.549
Neg	281	99%	147	100%	428	100%	
Pos	2	1%	0	0%	2	0%	
CEBPA-DM							0.808
Neg	271	96%	140	95%	411	96%	
Pos	12	4%	7	5%	19	4%	
CBL							0.431
Neg	277	98%	146	99%	423	98%	
Pos	6	2%	1	1%	7	2%	
CSF3R							0.669
Neg	278	98%	146	99%	424	99%	
Pos	5	2%	1	1%	6	1%	
CUX1							0.344
Neg	281	99%	144	98%	425	99%	
Pos	2	1%	3	2%	5	1%	
DNMT3A							0.516
Neg	194	69%	96	65%	290	67%	
Pos	89	31%	51	35%	140	33%	
ETV6							0.724
Neg	276	98%	145	99%	421	98%	
Pos	7	2%	2	1%	9	2%	
EZH2							0.556
Neg	273	96%	144	98%	417	97%	
Pos	10	4%	3	2%	13	3%	
FLT3-ITD							0.254
Neg	220	78%	119	81%	339	79%	
Low ratio	31	11%	9	6%	40	9%	
High ratio	32	11%	19	13%	51	12%	
FLT3-TKD							1.000
Neg	248	88%	129	88%	377	88%	
Pos	35	12%	18	12%	53	12%	
GATA2							0.234
Neg	272	96%	145	99%	417	97%	
Pos	11	4%	2	1%	13	3%	
IDH1							0.495
Neg	253	89%	135	92%	388	90%	
Pos	30	11%	12	8%	42	10%	
IDH2							0.164
Neg	233	82%	129	88%	362	84%	
Pos	50	18%	18	12%	68	16%	
IKZF1							1.000
Neg	278	98%	144	98%	422	98%	
Pos	5	2%	3	2%	8	2%	
JAK2							1.000
Neg	278	98%	145	99%	423	98%	
Pos	5	2%	2	1%	7	2%	
KDM6A							0.756
Neg	275	97%	144	98%	419	97%	
Pos	8	3%	3	2%	11	3%	
KIT							0.592
Neg	257	91%	136	93%	393	91%	
Pos	26	9%	11	7%	37	9%	

Table S1C. Mutation status at diagnosis by training and validation cohort (continued).

	Training cohort (N=283)		Validation cohort (N=147)		Total (N=430)		P-value
KRAS							0.249
Neg	265	94%	133	90%	398	93%	
Pos	18	6%	14	10%	32	7%	
NOTCH1							0.549
Neg	281	99%	147	100%	428	100%	
Pos	2	1%	0	0%	2	0%	
NPM1							0.835
Neg	171	60%	91	62%	262	61%	
Pos	112	40%	56	38%	168	39%	
NRAS							0.393
Neg	224	79%	111	76%	335	78%	
Pos	59	21%	36	24%	95	22%	
PHF6							0.009
Neg	266	94%	146	99%	412	96%	
Pos	17	6%	1	1%	18	4%	
PTPN11							0.735
Neg	256	90%	131	89%	387	90%	
Pos	27	10%	16	11%	43	10%	
RAD21							0.224
Neg	273	96%	138	94%	411	96%	
Pos	10	4%	9	6%	19	4%	
RUNX1							0.153
Neg	255	90%	125	85%	380	88%	
Pos	28	10%	22	15%	50	12%	
SETBP1							0.549
Neg	281	99%	147	100%	428	100%	
Pos	2	1%	0	0%	2	0%	
SF3B1							1.000
Neg	276	98%	144	98%	420	98%	
Pos	7	2%	3	2%	10	2%	
SMC1A							0.270
Neg	282	100%	145	99%	427	99%	
Pos	1	0%	2	1%	3	1%	
SMC3							0.416
Neg	280	99%	144	98%	424	99%	
Pos	3	1%	3	2%	6	1%	
SRSF2							1.000
Neg	262	93%	137	93%	399	93%	
Pos	21	7%	10	7%	31	7%	
STAG2							0.429
Neg	261	92%	139	95%	400	93%	
Pos	22	8%	8	5%	30	7%	
TET2							0.254
Neg	256	90%	127	86%	383	89%	
Pos	27	10%	20	14%	47	11%	
TP53							1.000
Neg	270	95%	141	96%	411	96%	
Pos	13	5%	6	4%	19	4%	
U2AF1							1.000
Neg	276	98%	143	97%	419	97%	
Pos	7	2%	4	3%	11	3%	
WT1							0.863
Neg	256	90%	134	91%	390	91%	
Pos	27	10%	13	9%	40	9%	
ZRSR2							0.039
Neg	283	100%	144	98%	427	99%	
Pos	0	0%	3	2%	3	1	

Table S2. Multivariable analysis with correction for sampling time

	Relapse incidence			Relapse-free survival			Survival		
	SHR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value
Molecular residual disease (Present vs. Absent)	1.92	1.37-2.70	<0.001	1.66	1.24-2.23	0.001	1.66	1.20-2.30	0.002
Age per year	1.01	1.00-1.03	0.114	1.02	1.01-1.03	0.004	1.03	1.01-1.04	<0.001
WBC at diagnosis (x10^9/l) (>100 vs. ≤100)	2.22	1.37-3.59	0.001	2.08	1.37-3.16	0.001	2.04	1.28-3.24	0.003
ELN2017 risk classification									
Intermediate vs. Favorable	1.72	1.15-2.55	0.008	2.04	1.45-2.87	<0.001	2.58	1.75-3.80	<0.001
Adverse vs. Favorable	1.81	1.25-2.63	0.002	2.21	1.57-3.09	<0.001	2.67	1.82-3.91	<0.001
Number of cycles to CR (2 cycles vs. 1 cycle)	2.10	1.44-3.06	<0.001	2.41	1.73-3.36	<0.001	2.97	2.09-4.22	<0.001
Sampling time in days	0.99	0.98-1.00	0.021	0.99	0.99-1.00	0.121	0.99	0.99-1.00	0.094

Table S3. Multivariable analysis with time-dependent correction for allogeneic HSCT

	Relapse incidence			Relapse-free survival			Survival		
	SHR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value
Molecular residual disease (Present vs. Absent)	1.88	1.34-2.64	<0.001	1.64	1.22-2.21	0.001	1.64	1.18-2.27	0.003
Age per year	1.00	0.99-1.02	0,556	1.02	1.00-1.03	0.024	1.03	1.01-1.04	0.001
WBC at diagnosis (x10^9/l) (>100 vs. ≤100)	2.20	1.34-3.63	0,002	2.04	1.35-3.10	0.001	2.02	1.27-3.20	0.003
ELN2017 risk classification									
Intermediate vs. Favorable	2.31	1.52-3.51	0.001	2.28	1.58-3.28	<0.001	2.56	1.70-3.84	<0.001
Adverse vs. Favorable	2.44	1.64-3.62	0.001	2.51	1.75-3.59	<0.001	2.71	1.81-4.04	<0.001
Number of cycles to CR (2 cycles vs. 1 cycle)	2.07	1.42-3.03	<0.001	2.39	1.72-3.33	<0.001	2.95	2.08-4.20	<0.001
Post-remission therapy (Allo vs. no Allo)	0.45	0.31-0.65	<0.001	0.73	0.54-0.99	0.043	0.97	0.70-1.34	0.848

Table S4. Multivariable analysis

	Relapse incidence			Relapse-free survival			Survival		
	SHR	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**	**HR**	**95% CI**	**p-value**
NGS vs. Flow									
NGS pos/Flow neg vs. double neg	2.23	1.43-3.49	<0.001	1.79	1.21-2.67	0.004	1.82	1.19-2.80	0.006
NGS neg/Flow pos vs. double neg	2.23	1.35-3.68	0.002	1.79	1.11-2.88	0.017	1.63	0.96-2.77	0.069
NGS pos/Flow pos vs. double neg	3.92	2.31-6.67	<0.001	3.22	1.99-5.22	<0.001	2.38	1.41-4.02	0.001
Age per year	1.01	0.99-1.03	0.173	1.02	1.00-1.03	0.013	1.03	1.01-1.04	0.002
WBC at diagnosis (x10^9/l) (>100 vs. ≤100)	2.62	1.66-4.13	<0.001	2.23	1.42-3.51	0.001	1.94	1.17-3.21	0.010
ELN2017 risk classification									
Intermediate vs. Favorable	1.61	1.04-2.49	0.034	1.91	1.29-2.82	0.001	2.42	1.57-3.73	<0.001
Adverse vs. Favorable	1.81	1.19-2.76	0.005	2.07	1.41-3.03	<0.001	2.52	1.64-3.87	<0.001
Number of cycles to CR (2 cycles vs. 1 cycle)	2.29	1.52-3.45	<0.001	2.43	1.66-3.55	<0.001	3.19	2.14-4.76	<0.001

REFERENCES

1. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM*, et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 2004 Apr 15; **350**(16)**:** 1617-1628.
2. Taskesen E, Bullinger L, Corbacioglu A, Sanders MA, Erpelinck CAJ, Wouters BJ*, 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 Feb 24; **117**(8)**:** 2469-2475.
3. BBMap short-read aligner, and other bioinformatics tools. 2016. at http://sourceforge.net/projects/bbmap/.
4. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C*, et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013 Mar; **31**(3)**:** 213-219.
5. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N*, et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009 Aug 15; **25**(16)**:** 2078-2079.
6. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C*, et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011 May; **43**(5)**:** 491-498.
7. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L*, et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 2012 Mar; **22**(3)**:** 568-576.
8. Indelocator. at http://www.broadinstitute.org/cancer/cga/indelocator.
9. Terwijn M, van Putten WL, Kelder A, van der Velden VH, Brooimans RA, Pabst T*, et al.* High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. *J Clin Oncol* 2013 Nov 1; **31**(31)**:** 3889-3897.
10. Fine JP, and Robert J. Gray. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* 1999; **94**(446)**:** 496-509.
11. BD. MN. Evaluation of response-time data involving transient states: an illustration using heart-transplant data. *J Am Stat Assoc* 1974; **69:** 81-86.

Chapter 3

CHROMOSOMAL INSTABILITY DETERMINES OUTCOME IN ACUTE MYELOID LEUKEMIA WITH MUTATED *TP53*

Tim Grob, MD[1*], Adil S. A. Al Hinai[1,2*], Mathijs A. Sanders, PhD[1], François G. Kavelaars[1], Melissa Rijken[1], Patrycja L. Gradowska, PhD[1], Bart J. Biemond, MD, PhD[3], Dimitri A. Breems, MD, PhD[4], Johan Maertens, MD, PhD[5], Marinus van Marwijk Kooy, MD, PhD[6], Thomas Pabst, MD, PhD[7], Okke de Weerdt, MD, PhD[8], Gert J. Ossenkoppele, MD, PhD[3], Arjan A. van de Loosdrecht, MD, PhD[3], Gerwin A. Huls, MD, PhD[9], Jan J. Cornelissen, MD, PhD[1], H. Berna Beverloo, PhD[10], Bob Löwenberg, MD, PhD[1], Mojca Jongen-Lavrencic, MD, PhD[1], Peter J.M. Valk, PhD[1]

[1] Department of Hematology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, the Netherlands

[2] National Genetic Center, Royal Hospital, Ministry of Health, Muscat, Sultanate of Oman

[3] Department of Hematology, Amsterdam University Medical Center, Amsterdam, the Netherlands

[4] Department of Hematology, Ziekenhuis Netwerk Antwerpen, Antwerp, Belgium

[5] Department of Hematology, University Hospital Gasthuisberg, Leuven, Belgium

[6] Department of Hematology, Isala Hospital, Zwolle, the Netherlands

[7] Department of Oncology, University Hospital, Inselspital, Bern, Switzerland

[8] Department of Hematology, Sint Antonius Hospital, Nieuwegein, the Netherlands

[9] Department of Hematology, University Medical Center Groningen, Groningen, the Netherlands

[10] Department of Clinical Genetics, University Medical Center Rotterdam, Rotterdam, the Netherlands

*These authors contributed equally to this work

Submitted

ABSTRACT

BACKGROUND

TP53 mutations in acute myeloid leukemia (AML) confer very poor outcome. However, substantial heterogeneity within the mutant *TP53* AML subgroup on a clinical and molecular level precludes the exact assessment on prognostic impact.

METHODS

We performed next-generation sequencing (NGS) on 2,200 AML specimens and assessed the molecular characteristics of mutant *TP53* AML in detail, including the *TP53* mutant allelic status (mono or bi-allelic), clone size, concurrent mutations, cytogenetics and molecular minimal residual disease. The primary end point was overall survival.

RESULTS

TP53 mutations were detected in 230 (10.5%) AML patients and associated with poor outcome. In 174 (76%) patients bi-allelic *TP53* mutations were found. Complex karyotype (CK) was observed in 185 (84%) patients. The majority of bi-allelic mutant *TP53* patients were marked by CK (97%), which coincided with higher mutant *TP53* variant allele frequencies (VAF). *TP53* mutant allelic status, VAF and concurrent mutations were not significantly associated with survival. Persistence of mutant *TP53* in CR (73% of cases) was not correlated with increased relapse or inferior survival rates. CK in mutant *TP53* AML, was significantly associated with worse overall survival (OS) (2-year OS; 9.2% vs. 34.3%, hazard ratio for death, 1.87; P=0.003), irrespective of *TP53* mutant allelic status and VAF. Among mutant *TP53* AML patients who received an allogeneic hematopoietic stem cell transplant, CK was significantly associated with a higher relapse rate and worse survival (2-year OS; 16.7% vs. 63.6%, hazard ratio for death 2.79; P=0.016). Multivariate analysis established CK as an independent prognostic factor for mutant *TP53* AML.

CONCLUSIONS

CK appeared independently and strongly associated with a higher relapse rate and worse survival in mutant *TP53* AML, irrespective of the type of consolidation therapy.

INTRODUCTION

Mutations in *TP53* are present in approximately 10% of acute myeloid leukemia (AML) patients and represent a unique leukemia subtype with very poor outcome.[1] *TP53* is located on chromosome 17p13 and is essential for cell cycle control and the DNA damage response.[2] Although the exact mechanism of leukemogenesis of mutant *TP53* AML remains unknown, it has been shown that *TP53* mutations drive a dominant negative effect and typically occur in founding clones that expand upon cytotoxic stress.[3,4] In AML, mutant *TP53* is strongly associated with enhanced chromosomal instability as illustrated by the accumulation of numerous cytogenetic aberrations.[5,6] Regarding the poor outcome, mutant *TP53* AML is assigned to the adverse risk category of the 2017 European LeukemiaNet (ELN) risk classification and is recommended to receive intensive consolidation treatments.[7] Although many mutant *TP53* AML patients achieve complete remission (CR) and generally receive an allogeneic hematopoietic stem cell transplant (HSCT), relapse rates remain considerably high.

Clinical and molecular heterogeneity in mutant *TP53* has been explored in myeloid malignancies, but it is unclear which clinical and molecular characteristics are associated with outcome and whether they could be incorporated into clinical practice. The presence of wild type *TP53* is critical for maintaining chromosomal stability and appears to be associated with a more favorable outcome in patients with myelodysplastic syndrome (MDS).[8] Other initial studies into AML suggest that the mutant *TP53* variant allele frequency (VAF) and concurrent cytogenetic aberrations carry prognostic value.[5,6,9,10]

Here, we present an in-depth characterization of mutant *TP53* patients in relation to survival in a large cohort of newly diagnosed AML who were treated with intensive chemotherapy by using demographic, therapeutic and follow-up data in the context of molecular aberrations. We performed next-generation sequencing (NGS) to assess the molecular characteristics of mutant *TP53* AML in detail, including the *TP53* mutant allelic status (mono or bi-allelic), clone size, concurrent mutations, cytogenetics and molecular minimal residual disease (MRD).

METHODS

PATIENTS AND SAMPLES

In total 2,200 AML patients were assessed for eligibility and treated in the Haemato-Oncology Foundation for Adults in the Netherlands and Swiss Group for Clinical Cancer Research (HOVON-SAKK) clinical trials between 2001 and 2017 (Fig. S1). All patients received standard induction chemotherapy and were consolidated according to the HOVON-SAKK study protocols. Details of treatment protocols were described previously (www.hovon.nl).[11-16] All trial participants have provided written informed consent in accordance with the Declaration of Helsinki. DNA was isolated from diagnostic bone marrow samples of 2,200 AML patients and 537 CR samples (Supplementary Methods). In 33 AML patients carrying *TP53* variants with VAF above 40%, DNA from saliva was available to verify the germline status.

CYTOGENETICS

Cytogenetic analysis was carried out at the local reference centers using standard protocols. Cytogenetic data was entered into the database of the HOVON Data Center. This data, including karyotypes and FISH, was centrally peer-reviewed by clinical genetics laboratory specialists. The clonal structural and numerical chromosomal abnormalities were reported in accordance with the International System for Human Cytogenetic Nomenclature and the European LeukemiaNet 2017 recommendations. Complex karyotype is defined by three or more unrelated chromosome abnormalities in the absence of one of the WHO-designated recurring translocations or inversions, that is, t(8;21), inv(16) or t(16;16), t(9;11), t(v;11)(v;q23.3), t(6;9), inv(3) or t(3;3), AML with BCR-ABL1.[7]

TARGETED NEXT-GENERATION SEQUENCING AND *TP53* DEEP SEQUENCING

The TruSight Myeloid Sequencing panel (Illumina, San Diego, CA, USA) was used to detect the presence of driver mutations at diagnosis. Details were described previously.[17] In this study only AML cases carrying pathogenic *TP53* mutations were included. Pathogenic *TP53* variants were defined by occurrence in the COSMIC and IARC TP53 database as well as by analyses *in silico* with programs, such as Polyphen-2, SIFT, FATHMM, MetaSVM, MetaLR, CADD, DANN and ClinVar. The limit of detection was VAF 1% at diagnosis. To detect *TP53* mutations in CR we used Illumina-based deep sequencing, according to the manufacturer's protocol (Illumina, San Diego, CA) (Supplementary Methods). The maximum limit of detection in the follow up samples was VAF 0.001% (variable dependent on *TP53* mutation type). Of note, patients with *TP53* germline mutations were excluded from MRD analyses (n=2).

ALLOCATION OF PATIENTS BASED ON *TP53* MUTANT STATUS

Mutant *TP53* AML patients were considered bi-allelic when (I) two or more *TP53* gene variants were detected, irrespective of the VAF, (II) when at least one *TP53* gene variant co-occurred with a cytogenetic aberration involving chromosome 17p (e.g., abnormality of 17p or monosomy 17), or (III) when *TP53* mutations were detected with a VAF above 55% (Fig. S2). We applied this VAF threshold to ensure homozygosity.

STATISTICAL ANALYSIS

Associations between variables were tested by the Fisher's exact test for categorical variables and by the Mann-Whitney U test for continuous variables. The primary endpoint of the study was overall survival (OS), defined as death from any cause. Survival time was calculated from the start of induction chemotherapy until the event of interest or censoring. Of note, the survival time in the analysis evaluating allogeneic HSCT started at the date of transplant. For the analysis evaluating molecular minimal residual disease and allogeneic HSCT, the endpoints were OS and cumulative incidence of relapse (CIR). Relapse incidence was defined as relapse with correction for competing non-relapse mortality. To compare the survival distributions, we used the log-rank test and the cox proportional hazards model. CIR was evaluated by the method of Gray and the Fine and Gray model for competing risks. The proportional hazards assumption was tested by interaction with time. All p-values were two sided and p-values below 0.05 were considered statistically significant. Statistical analyses were executed with Stata Statistical Software, Release 16.0 (College Station, Texas, USA).

RESULTS

MOLECULAR CHARACTERISTICS OF MUTANT *TP53* AML

We performed NGS to investigate the molecular characteristics of mutant *TP53* AML, including the *TP53* mutant allelic status, clone size and concurrent mutations. We detected 283 *TP53* mutations in 230 out of 2,200 (10.5%) AML patients. Two or more *TP53* mutations were found in 49 cases (Fig. S2). In total 206 missense, 16 nonsense , 38 indels and 23 splice-site mutations were detected (Fig. S3). Nearly all missense mutations occurred in the TP53 DNA binding domain (Fig. S3). Of note, in 112 mutant *TP53* AML patients concurrent chromosomal aberrations involving *TP53* (e.g., abnormality 17p or loss of chromosome 17) were detected. In total 56 (24.3%) patients were considered mono-allelic *TP53* mutant AML patients and 174 (75.7%) were bi-allelic *TP53* mutants (Fig. S2). The mutant *TP53* clone size was normally distributed with a median VAF of 47% (Fig. S4A). In contrast to AML in general, concurrent mutations were detected in only 113 (49%) mutant *TP53* AML patients. The most frequent concurrent mutations were detected in *DNMT3A*, *TET2*, *ASXL1*, *RUNX1* and *SRSF2* (Fig. 1 and Table 1).

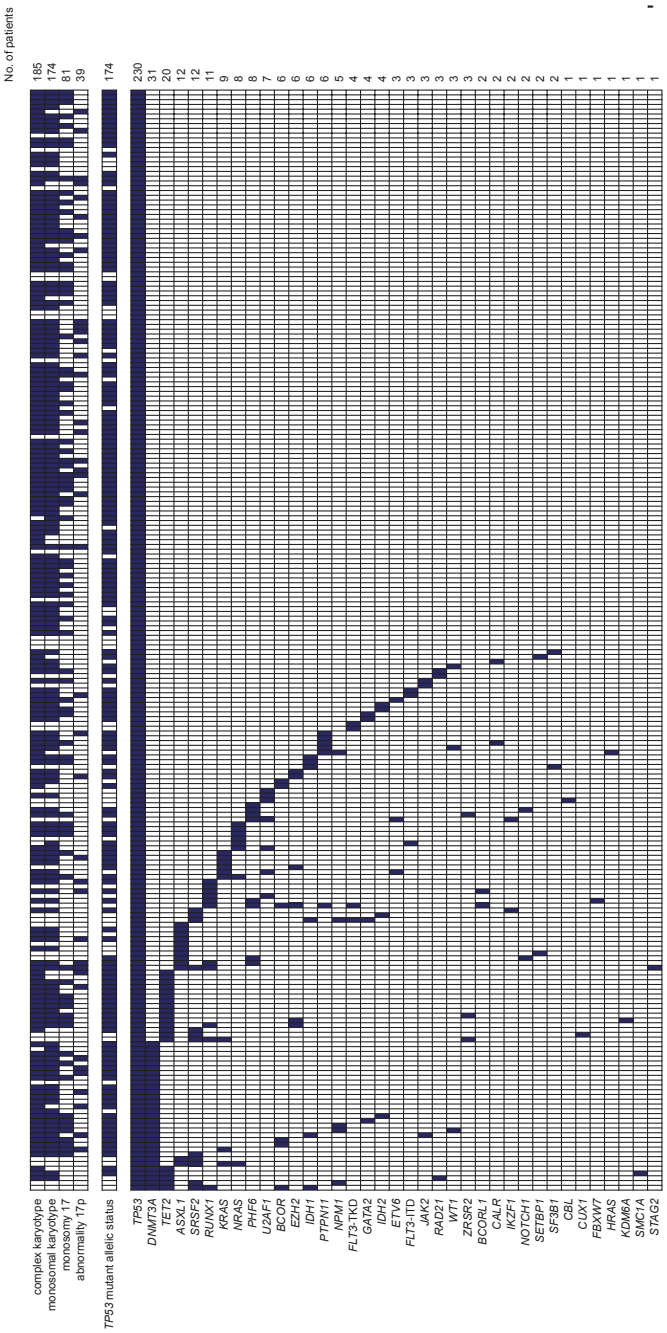
Table 1: Patient characteristics of AML with mutated *TP53* (n=230).

Table 1. Patient characteristics of AML with mutated *TP53*	
Age - yr	
Median	62
Range	18-80
Sex - no. (%)	
M	136 (59)
F	94 (41)
WBC at diagnosis - no. (%) *	
≤100	227 (99)
>100	2 (1)
AML type - no. (%)	
De novo	166 (72)
Secondary	44 (19)
Therapy-related	20 (9)
Last treatment before first CR - no. (%)	
Refractory	80 (35)
Cycle I	119 (52)
Cycle II	31 (13)
Consolidation therapy - no. (%)	
No allogeneic HSCT	170 (74)
Allogeneic HSCT	60 (26)
Cytogenetics - no. (%) †	
monosomy 5	62 (28)
deletion 5q	104 (47)
monosomy 7	72 (33)
monosomy 17	81 (37)
abnormality 17p	39 (18)
complex karyotype	185 (84)
monosomal karyotype	174 (79)
Mutation at diagnosis - no. (%)	
any concurrent	113 (49)
DNMT3A	31 (13)
TET2	20 (9)
ASXL1	12 (5)
RUNX1	11 (5)
SRSF2	12 (5)
***TP53* mutant allelic status - no. (%)**	
mono-allelic	56 (24)
bi-allelic	174 (76)

*Numbers may not sum to 230 because of missing values.

†Cytogenetics failed in 10 patients.

Figure 1: Overview of cytogenetic aberrations and concurrent mutations in mutant TP53 AML (n=230). Each column represents an individual patient and the presence of the aberration is indicated in blue. The upper panel shows the cytogenetic aberrations and the lower panel the concurrent mutations. A bi-allelic TP53 mutant status is also indicated in blue. In case of failed cytogenetics the cytogenetic aberrations were considered negative.

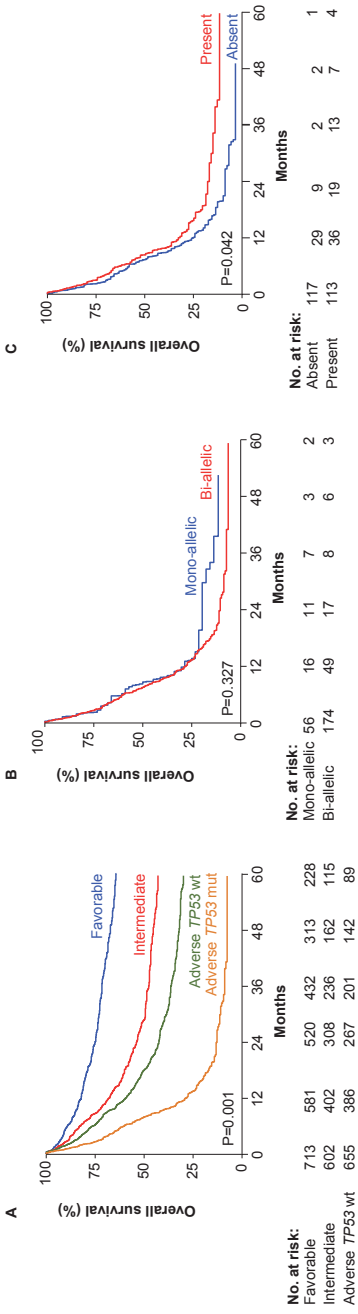


SURVIVAL BY MOLECULAR CHARACTERISTICS IN MUTANT *TP53* AML

We compared outcome of mutant *TP53* AML patients with the established ELN 2017 subgroups. Mutant *TP53* strongly associates with reduced survival in the context of the ELN 2017 risk classification (2-year OS, 12.8% *TP53* mutant vs. 42.5% *TP53* wild type; P<0.001) (Fig. 2A).

Next, we performed survival analysis to evaluate a variety of molecular characteristics and cytogenetic aberrations in relationship to outcome in mutant *TP53* AML. Mono-allelic mutant *TP53* lacked survival difference over its bi-allelic counterpart (P=0.327) (Fig. 2B), nor did the *TP53* mutation burden or aberrations involving chromosome 17 confer better outcome (Fig. S5). Clone size, realized by taking decreasing *TP53* mutation VAF thresholds, was investigated for impact on outcome. None of the *TP53* VAF thresholds appears to significantly associate with survival: VAF 50% (P=0.990), VAF 40% (P=0.257), VAF 30% (P=0.064), VAF 20% (P=0.189), VAF 10% (P=0.161), and VAF 5% (P=0.226) (Fig. S6). Concurrent mutation conferred limited but detectable survival benefit (Fig. 2C), while the presence of specific concurrent mutations provided no further survival advantage (Fig. S7). Hence, in this study none of the molecular characteristics of mutant *TP53* AML evidently relates with outcome.

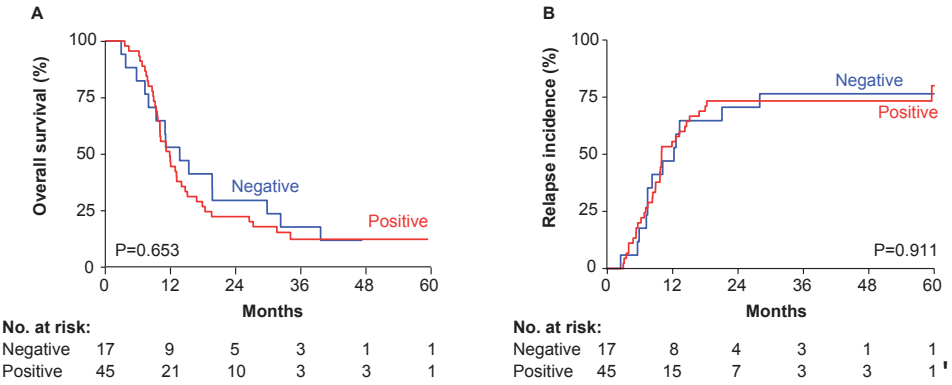
Figure 2: Overall survival of AML patients by the ELN 2017 risk classification (n=2,200). Patient in the adverse risk category are segregated by TP53 wild-type and TP53 mutant (A). Overall survival of mutant TP53 AML by TP53 mutant allelic status (mono-allelic versus bi-allelic) (B) and by the presence or absence of concurrent mutations (C).



MOLECULAR MINIMAL RESIDUAL DISEASE IN MUTANT *TP53* AML

Detection of molecular MRD is an important prognostic marker in AML.[17-19] We performed deep sequencing by targeted NGS on complete morphological remission bone marrow samples from 62 mutant *TP53* AML patients to assess molecular MRD. Mutant *TP53* is often the only suitable marker for molecular MRD detection since the prevalence of concurrent mutations is low and the majority of concurrently mutated genes associates with antecedent clonal hematopoiesis (*DNMT3A*, *TET2* and *ASXL1*). In total, 45 out of 62 AML patients had persistent *TP53* mutations in CR, for which the status did not associate with OS (P=0.653) or CIR (P=0.911) (Fig. 3).

Figure 3: Overall survival (A) and relapse incidence (B) of patients with *TP53* mutations in complete remission. Two patients with *TP53* germline mutations were excluded from the analysis.



CHROMOSOMAL INSTABILITY IN MUTANT *TP53* AML

We next determined whether the molecular characteristics of mutant *TP53* AML were related to chromosomal instability. Cytogenetic data was available for 220 out of 230 mutant *TP53* AML patients. Complex karyotypes, indicating chromosomal instability, were observed in 185 (84%) mutant *TP53* AML patients (Table 1). No association between complex karyotype and standard clinical patient characteristics were found (Table S1). However, major associations were present between complex karyotype, the *TP53* mutant allelic status and the *TP53* VAF (Table 2 and Fig. S4B). Remarkably, complex karyotypes were detected in the majority of bi-allelic *TP53* mutant AML patients (97%). In contrast, complex karyotype was observed in only 20 out of 50 (40%) mono-allelic mutant *TP53* AML patients (Table 2). With the noted exception of a single case, all AML patients with mutant *TP53* and chromosome 17p aberrations were marked by complex karyotype (Table 2). Of note, the VAF of mutant *TP53* was significantly higher in patients with complex karyotypes irrespective of chromosome 17p aberrations, suggesting that in certain AML cases focal 17p abnormalities may have been missed by karyotyping or FISH (Table 2 and Fig. S4B). Concurrent mutations were enriched in AML marked by complex karyotype, yet the prevailing mutated genes (*DNMT3A* and *TET2*) were not significantly associated with complex karyotype (Fig. 1 and Table 2).

Table 2: Molecular characteristics of mutant *TP53* AML by chromosomal instability (n=220).

Table 2. Molecular characteristics of mutant *TP53* AML by chromosomal instability			
	Complex karyotype (n=185)	**Non-complex karyotype (n=35)**	
***TP53* mutant allelic status - no. (%)**			p<0.001
mono-allelic	20 (11)	30 (86)	
bi-allelic	165 (89)	5 (14)	
***TP53* clone size - VAF (%)**			p<0.001
Median	53	11	
Range	2-97	1-97	
Cytogenetics - no. (%)			
monosomy 5	62 (34)	0 (0)	p<0.001
deletion 5q	101 (55)	3 (9)	p<0.001
monosomy 7	68 (37)	4 (11)	p=0.003
monosomy 17	80 (43)	1 (3)	p<0.001
abnormality 17p	39 (21)	0 (0)	p=0.001
monosomal karyotype	171 (92)	3 (9)	p<0.001
Mutation at diagnosis - no. (%)			
any concurrent	84 (45)	26 (74)	p=0.003
DNMT3A	25 (14)	6 (17)	p=0.597
TET2	16 (9)	4 (11)	p=0.534
ASXL1	6 (3)	5 (14)	p=0.017
RUNX1	6 (3)	5 (14)	p=0.017
SRSF2	4 (2)	7 (20)	p<0.001

SURVIVAL BY CHROMOSOMAL INSTABILITY IN MUTANT *TP53* AML

Detectable molecular characteristics and presence of molecular MRD lack clinical value in *TP53* mutant AML. Complex karyotype, manifesting as chromosomal instability, is another hallmark of *TP53* mutant AML that potentially associates with disease progression and treatment response. Complex karyotype strongly associates with reduced outcome in *TP53* mutant AML (2-year OS, 9% complex karyotype vs. 34% non-complex karyotype; P=0.002) (Fig. 4A). Within mono-allelic *TP53* mutant AML, complex karyotype precipitated reduced outcome compared to its non-complex counterpart (P=0.001) (Fig. 4B). Altogether, this indicates that the degree of chromosomal instability has profound prognostic value in mutant *TP53* mutant AML, irrespective of *TP53* mutant allelic status and *TP53* VAF.

CHROMOSOMAL INSTABILITY IN MUTANT *TP53* AND ALLO HSCT

The prognostic value of complex karyotype in the context of allogeneic HSCT was assessed in 59 *TP53* mutant AML patients. Despite receiving allogeneic HSCT, *TP53* mutant AML cases marked by complex karyotype had higher relapse incidence rates (2-year CIR, 69% complex karyotype vs. 36% non-complex karyotype; P=0.056) and worse outcome (2-year OS, 17% complex karyotype vs. 64% non-complex karyotype; P=0.020) (Fig. 5).

Figure 4: Overall survival of mutant *TP53* AML patients by chromosomal instability (A) and of mutant *TP53* AML patients stratified by the *TP53* mutant allelic status and chromosomal instability (B).

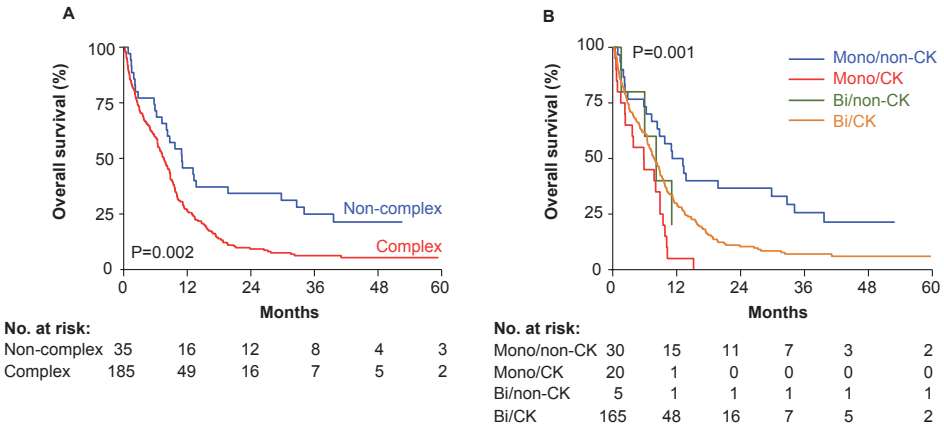
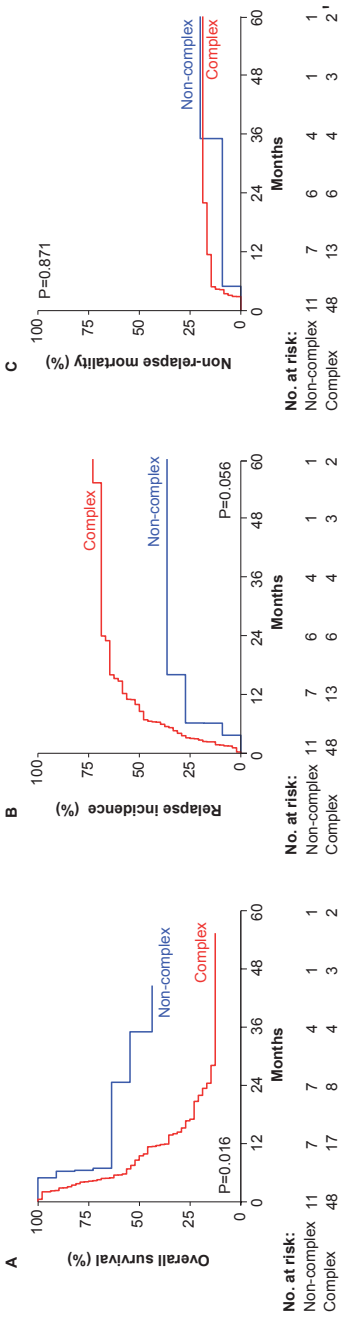


Figure 5: Overall survival (A), relapse incidence (B), and non-relapse mortality (C) of mutant *TP53* AML by chromosomal instability in patients that received allogeneic hematopoietic stem cell transplant. The survival time starts at the date of transplant.



MULTIVARIATE AND SENSITIVITY ANALYSIS

Complex karyotype was subsequently assessed as an independent outcome predictor in multivariate analysis, including standard prognostic markers such as age, AML type (e.g. *de novo*, secondary or therapy-related AML) and the numbers of cycles needed to achieve CR. Complex karyotype appeared independently and strongly associated with outcome (hazard ratio for death, 1.96; 95% CI, 1.29 to 2.99; P=0.002) (Table 3). Sensitivity analysis, performed to identify potential treatment modification within trial protocols, yielded no significant interactions. Similar results were obtained when elderly AML patients were excluded (data not shown).

Table 3: Multivariate analysis of chromosomal instability and other prognostic factors for overall survival (n=220).

Table 3. Multivariate analysis of prognostic factors for overall survival.

	OS		
	Hazard ratio	**95% CI**	**p-value**
Complex karyotype (Complex vs. non-complex)	1.96	1.29-2.99	0.002
Age per year	1.02	1.01-1.04	0.009
AML type			
Secondary AML vs. *de novo* AML	0.85	0.59-1.24	0.403
Therapy-related AML vs. *de novo* AML	0.56	0.34-0.93	0.024
Number of cycles to attain complete remission			
One cycle vs. Refractory	0.44	0.32-0.60	<0.001
Two cycles vs. Refractory	0.30	0.18-0.48	<0.001

DISCUSSION

Mounting evidence revealed that mutant *TP53* AML is a distinct risk group.[6] However, the substantial heterogeneity within the mutant *TP53* AML subgroup on a clinical and molecular level precludes the exact assessment on prognostic impact. Here, we report the detailed molecular characterization of leukemic clones in *TP53* mutant AML.

Although concurrent mutations are present in mutant *TP53* AML, mutant *TP53* itself appeared to be isolated in half of the patients. Of note, the majority of concurrent mutations are known contributors of age-related clonal hematopoiesis.[20,21] Since *TP53* mutations are more frequent in elderly, this observation could therefore at least partially explain why concurrent mutations were not associated with survival in mutant *TP53* AML. In addition, the concurrent mutations may exist in separate age-related clonal hematopoiesis clones. Regardless of the presence of concurrent mutations, mutant *TP53* AML evolves to become chromosomal instable indicating that mutant *TP53* AML is a distinct entity irrespective of other mutations.

While molecular MRD has prognostic value for predicting impending relapse in AML[17-19], we did not observe such association in mutant *TP53* AML. Despite the fact that we used deep sequencing, revealing MRD in the majority of cases, molecular MRD detection in mutant *TP53* AML did not have prognostic value. It is conceivable that all mutant *TP53* AML patients achieving CR have MRD, sometimes at levels that are undetectable with current NGS approaches. In fact, the high relapse rates in AML patients without detectable mutant *TP53* MRD in CR illustrates the critical role of mutant *TP53* in chemotherapeutic response and implies that small refractory clones are present below our NGS detection limit.[3] Therefore, the detection of residual cytogenetic aberrations in mutant *TP53* AML patients, as indicators of chromosomal instability, may be more relevant than detection of molecular MRD.

Our study revealed that chromosomal instability, reflected by complex karyotypes, is an independent marker for outcome in mutant *TP53* AML. In contrast to previous studies,[8-10] we did not observe prognostic value of the *TP53* mutant allelic status or clone size. However, important interactions between these molecular characteristics and chromosomal instability appeared evident. AML patients with complex karyotypes often have higher *TP53* mutation VAF, regardless of cytogenetic abnormalities involving chromosome 17. The high VAF in these patients suggests mutant *TP53* homozygosity. Thus, it is possible that bi-allelic *TP53* mutant patients were erroneously misclassified as mono-allelic *TP53* mutants because of small *TP53* gene deletions missed by conventional cytogenetics, or by epigenetic changes that result in loss of transcription from the remaining wild type *TP53* allele. These findings illustrate the limitations of allocating patients by the *TP53* mutant allelic status based on cytogenetics and NGS. In addition, complex karyotype clearly modifies the prognostic value of the *TP53* mutant allelic status. The great majority of bi-allelic *TP53* AML mutant patients are marked by complex karyotypes and rarely by non-complex karyotypes. However, mono-allelic

mutant *TP53* patients with complex karyotypes have decreased overall survival compared to their chromosomal stable counterparts. From a clinical and molecular perspective, we therefore consider chromosomal instability, as defined by complex karyotype, as the most important prognostic factor in mutant *TP53* AML, irrespective of the *TP53* mutant allelic status and clone size.

In conclusion, in depth clinical and molecular characterization determined that complex karyotype independently associates with higher relapse rates and worse overall survival in mutant *TP53* AML, irrespective of the type of consolidation therapy. These patients therefore qualify for novel experimental therapies.

ACKNOWLEDGEMENTS

We thank all the participating centers of the Dutch–Belgian Cooperative Trial Group for Hematology–Oncology (HOVON) and Swiss Group for Clinical Cancer Research (SAKK), where the clinical trials that formed the basis for this study were conducted; Eric Bindels for performing next-generation sequencing, Remco Hoogenboezem for assisting with bioinformatics, and Egied Simons for assisting with preparation of the figures (Erasmus University Medical Center).

REFERENCES

1. Cancer Genome Atlas Research N, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med 2013;368:2059-74.
2. Kastenhuber ER, Lowe SW. Putting p53 in Context. Cell 2017;170:1062-78.
3. Wong TN, Ramsingh G, Young AL, et al. Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia. Nature 2015;518:552-5.
4. Boettcher S, Miller PG, Sharma R, et al. A dominant-negative effect drives selection of TP53 missense mutations in myeloid malignancies. Science 2019;365:599-604.
5. Rücker FG, Schlenk RF, Bullinger L, et al. TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome. Blood 2012;119:2114-21.
6. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med 2016;374:2209-21.
7. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017;129:424-47.
8. Bernard E, Nannya Y, Hasserjian RP, et al. Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. Nat Med 2020;26:1549-56.
9. Prochazka KT, Pregartner G, Rücker FG, et al. Clinical implications of subclonal TP53 mutations in acute myeloid leukemia. Haematologica 2019;104:516-23.
10. Short NJ, Montalban-Bravo G, Hwang H, et al. Prognostic and therapeutic impacts of mutant TP53 variant allelic frequency in newly diagnosed acute myeloid leukemia. Blood Adv 2020;4:5681-9.
11. Pabst T, Vellenga E, van Putten W, et al. Favorable effect of priming with granulocyte colony-stimulating factor in remission induction of acute myeloid leukemia restricted to dose escalation of cytarabine. Blood 2012;119:5367-73.
12. Löwenberg B, Pabst T, Maertens J, et al. Therapeutic value of clofarabine in younger and middle-aged (18-65 years) adults with newly diagnosed AML. Blood 2017;129:1636-45.
13. Ossenkoppele GJ, Breems DA, Stuessi G, et al. Lenalidomide added to standard intensive treatment for older patients with AML and high-risk MDS. Leukemia 2020;34:1751-9.
14. Janssen J, Löwenberg B, Manz M, et al. Inferior Outcome of Addition of the Aminopeptidase Inhibitor Tosedostat to Standard Intensive Treatment for Elderly Patients with AML and High Risk MDS. Cancers (Basel) 2021;13.
15. Löwenberg B, Pabst T, Maertens J, et al. Addition of lenalidomide to intensive treatment in younger and middle-aged adults with newly diagnosed AML: the HOVON-SAKK-132 trial. Blood Adv 2021;5:1110-21.
16. Dutch–Belgian Cooperative Trial Group for Hematology–Oncology. Home page (http://www.hovon.nl).
17. Jongen-Lavrencic M, Grob T, Hanekamp D, et al. Molecular Minimal Residual Disease in Acute Myeloid Leukemia. N Engl J Med 2018;378:1189-99.
18. Morita K, Kantarjian HM, Wang F, et al. Clearance of Somatic Mutations at Remission and the Risk of Relapse in Acute Myeloid Leukemia. J Clin Oncol 2018;36:1788-97.
19. Thol F, Gabdoulline R, Liebich A, et al. Measurable residual disease monitoring by NGS before allogeneic hematopoietic cell transplantation in AML. Blood 2018;132:1703-13.
20. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med 2014;371:2477-87.
21. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med 2014;371:2488-98.

SUPPLEMENTARY METHODS

PATIENT SAMPLES AND DNA ISOLATION

Follow up samples were taken at least 21 days after the start of the last induction cycle. If additional CR samples were available, the sample preceding consolidation therapy was chosen. When patients did not receive any consolidation therapy, the last sample within four months after the start of the last induction cycle was selected.

The majority of AML samples were purified by Ficoll-Hypaque centrifugation (Nygaard, Oslo, Norway), cryopreserved and subsequently lysed in RLT buffer (Qiagen, Venlo, the Netherlands). High quality DNA was extracted using the QIASymphony DSP DNA Mini Kit according to the manufacturer's instructions (Qiagen, Venlo, the Netherlands). DNA concentration was measured by Qubit Fluorometric Quantitation (Thermo Fisher Scientific, Wilmington, DE).

TARGETED NGS AND DATA ANALYSIS

The TruSight Myeloid Sequencing Panel libraries were paired-end sequenced (2x221bp) on an Illumina HiSeq 2500 System (Illumina, San Diego, CA) in Rapid Run mode or on a Illumina MiSeq System (Illumina, San Diego, CA). The vast majority of amplicon target regions were completely paired-end sequenced. Overlap-based error-correction was utilized to attenuate any form of strand-specific error biases. Error-corrected paired-end reads aligned to the human genome version 19 (hg19) with BBMAP3 followed by quality control to determine cases with insufficient number of reads for adequate variant calling. Single nucleotide variants (SNVs) and insertions-deletions (indels) at diagnosis were determined by MuTect4, Samtools5, GATK6, Varscan7, Indelocator8 and Pindel6. Variant allele frequencies (VAF) of mutations detected at diagnosis were calculated as the ratio between the number of mutant and total reads. Pathogenic *TP53* variants were defined by occurrence in the COSMIC and IARC TP53 database as well as by analyses in silico with programs, such as Polyphen-2, SIFT, FATHMM, MetaSVM, MetaLR, CADD, DANN and ClinVar.

TP53 DEEP SEQUENCING

For deep sequencing template specific primers for *TP53* mutation NGS analysis were designed by using the Ion AmpliSeq Designer software (ThermoFisher Scientific, Bleiswijk). The *TP53* specific primers (see below) were adapted for Illumina-based sequencing by adding an Illumina forward (5'-TCGTCGGCAGCGTCAGATGTGTATAAGAG ACAG-3'-[locus-specific sequence] or reverse (5'-GTCTCGTGGGCTCGGAGATGTGTATAAG AGACAG-3')-[locus-specific sequence] overhang adapter sequence to the *TP53* specific primers.

In the second PCR, by means of these adapter sequences, sample-specific dual indices for sample identification and Illumina sequencing adapters are attached by using unique

dual index primers from the IDT for Illumina Nextera DNA UD Index Kit (Illumina, San Diego, CA, USA). Two multiplex reactions were carried out (see below). The *TP53* loci of interest were amplified by multiplex PCR on 100ng genomic DNA using the Roche FastStart High Fidelity PCR System (Roche) containing 1× Buffer with 1.8mM MgCl2, 0.2mM dNTP, ~0.4µM each primer, 0.1U FastStart Taq DNA polymerase. Amplification was performed using the following thermocycling conditions: 95°C for 5 minutes, 25 cycles of 30 seconds at 95°C, 30 seconds at 60°C, and 30 seconds at 72°C and a final extension for 7 minutes at 72°C. Amplicons from the first step PCR were purified using the Agencourt AMPure XP bead purification kit (Beckman Coulter, Fullerton, CA, USA).

The second step PCR was performed with primers from the IDT for Illumina Nextera DNA UD Index Kit (Illumina, San Diego, CA) using the KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, MA, USA) with the following thermocycling conditions: 95°C for 3 minutes, 10 cycles of 20 seconds at 95°C, 30 seconds at 55°C, and 30 seconds at 72°C and a final extension for 5 minute at 72°C.

The library pool was purified with Agencourt AMPure XP beads and normalized for Illumina-based sequencing, according to the manufacturer's protocol (Illumina, San Diego, CA). The NGS libraries were paired-end sequenced (2×221-bp) on the MiSeq System (Illumina, San Diego, CA) according to manufacturer's recommendation (Illumina, San Diego, CA). NGS data analysis to detect variants at a low detection level is performed as described before (Jongen-Lavrencic et al., 2018). The maximum limit of detection in the follow up samples was VAF 0.001% (dependent on type of *TP53* mutation).

Primer name	Sequence (5' -3')
TP53_AMP1_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGCTTCTGACGCACACCTATTG
TP53_AMP1_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCATGTTCAAGACAGAAGGGCCTGA
TP53_AMP2_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAATCCTATGGCTTTCCAACCTAG
TP53_AMP2_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCAGAGCTGAATGAGGCCTTGGAACT
TP53_AMP3_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTcaggctaggctaagctatgatgttccttaga
TP53_AMP3_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCTTGTTAAAGAGAGCATGAAAATGGTTCT
TP53_AMP4_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAAACGGCATTTTGAGTGTTAGACTGGA
TP53_AMP4_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCTTTCCTTGCCTCTTTCCTAGCACTG
TP53_AMP5_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCATAACTGCACCCTTGGTCTCCTC
TP53_AMP5_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCAACAGCTTTGAGGTGCGTGTTTG
TP53_AMP6_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAAATCGGTAAGAGGTGGGCCCAG
TP53_AMP6_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCCACTACAACTACATGTGTAACAGTTCC
TP53_AMP7_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTttcaactgtgcaatagttaaacccat
TP53_AMP7_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCTGAGGTCTGGTTTGCAACTGGG
TP53_AMP8_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCATCCAAATACTCCACACGCAAATTTC
TP53_AMP8_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCGCTGCTCAGATAGCGATG
TP53_AMP9_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTGCTGTGACTGCTTGTAGATGGC
TP53_AMP9_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCCTGACTTTCAACTCTGTCTCCTTCCTC
TP53_AMP10_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGAAGCCAAAGGGTGAAGAGGAATCCC
TP53_AMP10_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCAAGTCTGTGACTTGCACGGTCAGTTG
TP53_AMP11_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAGCTGCCCTGGTAGGTTTTC
TP53_AMP11_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCACTGAAGACCCAGGTCCAGAT
TP53_AMP12_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTccccagcccAACCCTTGTCCTT
TP53_AMP12_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCAGCCCCCTAGCAGAGACCTG
TP53_AMP13_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGGAACAAGAAGTGGAGAATGTCAG
TP53_AMP13_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCATGTGATGTCATCTCTCCTCCCTGCTT
TP53_AMP14_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTTCCCAGCCTGGGCATCCTT
TP53_AMP14_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCaaccatcttttaactcaggtactgtGT
TP53_AMP15_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTatcgtaagtcaagtagcatctgTATCA
TP53_AMP15_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCGCTCCTGGTTGTAGCTAACTAACTTC
TP53_AMP16_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGAGGAGCTGGTGTTGTTGGG
TP53_AMP16_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCTAAGCGAGGTAAGCAAGCAGGAC
TP53_AMP17_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTTTCTTGCGGAGATTCTCTTCCTCT
TP53_AMP17_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCGGACAGGTAGGACCTGATTTCCTT
TP53_AMP18_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTCTTCCAGTGTGATGATGGTGAG
TP53_AMP18_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCTTGCCACAGGTCTCCCCAAGG
TP53_AMP19_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCACTGACAACCACCCTTAACCC
TP53_AMP19_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCTCTGGCCCCTCCTCAGCATCT
TP53_AMP20_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCCAGCCCCAGCTGCTCAC
TP53_AMP20_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCTTGCCAACTGGCCAAGACCT
TP53_AMP21_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAAACATCTTGTTGAGGGCAGGG
TP53_AMP21_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCTGAGGTGTAGACGCCAACTCTC
TP53_AMP22_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGCATTGAAGTCTCATGGAAGCCA
TP53_AMP22_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCTCCTGGCCCCTGTCATCTTCTGT
TP53_AMP23_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGGAGCAGCCTCTGGCATTCTG
TP53_AMP23_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCTCTGACTGCTCTTTTCACCCATC
TP53_AMP24_Rd1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGCCCTTCCAATGGATCCACTC
TP53_AMP24_Rd2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCAGGGTTGGAAGTGTCTCATGCTGGA

Primer mix 1	Primerstock (pmol/µl)	Pool primerstock (µl)
TP53_AMP1_Rd1	200	5
TP53_AMP1_Rd2	200	5
TP53_AMP2_Rd1	200	5
TP53_AMP2_Rd2	200	5
TP53_AMP3_Rd1	200	5
TP53_AMP3_Rd2	200	5
TP53_AMP4_Rd1	200	5
TP53_AMP4_Rd2	200	5
TP53_AMP5_Rd1	200	5
TP53_AMP5_Rd2	200	5
TP53_AMP6_Rd1	200	5
TP53_AMP6_Rd2	200	5
TP53_AMP7_Rd1	200	5
TP53_AMP7_Rd2	200	5
TP53_AMP8_Rd1	200	5
TP53_AMP8_Rd2	200	5
TP53_AMP9_Rd1	200	5
TP53_AMP9_Rd2	200	5
TP53_AMP10_Rd1	200	5
TP53_AMP10_Rd2	200	5
TP53_AMP11_Rd1	200	5
TP53_AMP11_Rd2	200	5
TP53_AMP12_Rd1	200	5
TP53_AMP12_Rd2	200	5

Primer mix 2	Primerstock (pmol/µl)	Pool primerstock (µl)
TP53_AMP13_Rd1	200	5
TP53_AMP13_Rd2	200	5
TP53_AMP14_Rd1	200	5
TP53_AMP14_Rd2	200	5
TP53_AMP15_Rd1	200	20
TP53_AMP15_Rd2	200	20
TP53_AMP16_Rd1	200	5
TP53_AMP16_Rd2	200	5
TP53_AMP17_Rd1	200	5
TP53_AMP17_Rd2	200	5
TP53_AMP18_Rd1	200	2.5
TP53_AMP18_Rd2	200	2.5
TP53_AMP19_Rd1	200	5
TP53_AMP19_Rd2	200	5
TP53_AMP20_Rd1	200	5
TP53_AMP20_Rd2	200	5
TP53_AMP21_Rd1	200	2.5
TP53_AMP21_Rd2	200	2.5
TP53_AMP22_Rd1	200	5
TP53_AMP22_Rd2	200	5
TP53_AMP23_Rd1	200	5
TP53_AMP23_Rd2	200	5
TP53_AMP24_Rd1	200	5
TP53_AMP24_Rd2	200	5

Table S1: Patient characteristics of mutant *TP53* AML by chromosomal instability (n=220).

Table S1. Patient characteristics of mutant *TP53* AML by chromosomal instability			
	Complex karyotype (n=185)	**Non-complex karyotype (n=35)**	
Age - yr			p=0.786
Median	62	61	
Range	18-78	28-77	
Sex - no. (%)			p=0.851
M	111 (60)	20 (57)	
F	74 (40)	15 (43)	
AML type - no. (%)			p=0.400
De novo	136 (73)	24 (68)	
Secondary	31 (17)	9 (26)	
Therapy related	18 (10)	2 (6)	
Last treatment before first CR - no. (%)			p=0.582
Refractory	63 (34)	9 (26)	
Cycle I	97 (52)	20 (57)	
Cycle II	25 (14)	6 (7)	
Consolidation therapy - no. (%)			p=0.535
No allogeneic HSCT	137 (74)	24 (69)	
Allogeneic HSCT	48 (26)	11 (31)	

Figure S1: Consort diagram of mutant *TP53* study. Abbreviations: HO, HOVON-SAKK, Dutch-Belgian Hemato-Oncology Cooperative Group and the Swiss Group for Clinical Cancer Research.

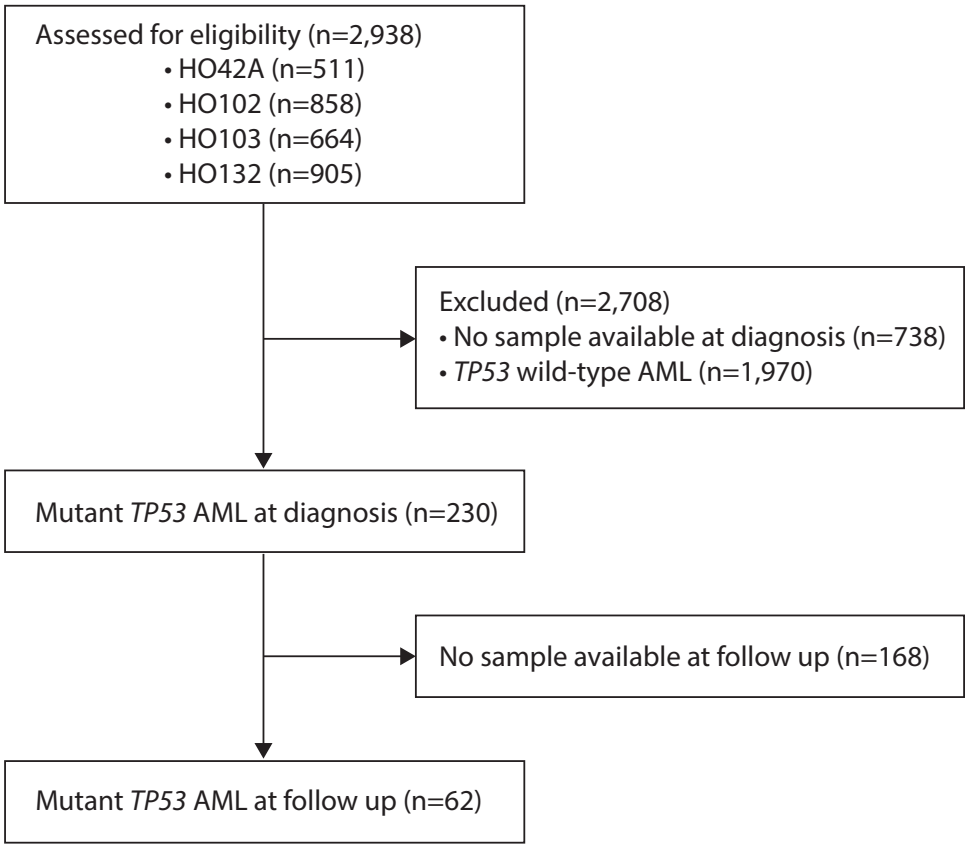


Figure S2: Consort diagram of the mutant *TP53* allelic status of 230 AML patients. AML patients with multipele *TP53* mutations or with mutant *TP53* combined with concurrent chromosome 17 aberrations (e.g., monosomy 17 of abnormality 17p) were assigned as bi-allelic *TP53* mutant. In *TP53* AML without chromosome 17 aberrations, patients with a mutant *TP53* variant allele frequency above 55%, were also considered bi-allelic *TP53* mutant. In 10 patients cytogenetic data was missing. Of note, 4 patients with missing cytogenetics were considered bi-allelic *TP53* mutant based on the presence of multiple *TP53* mutations or a variant allele frequency above 55%.

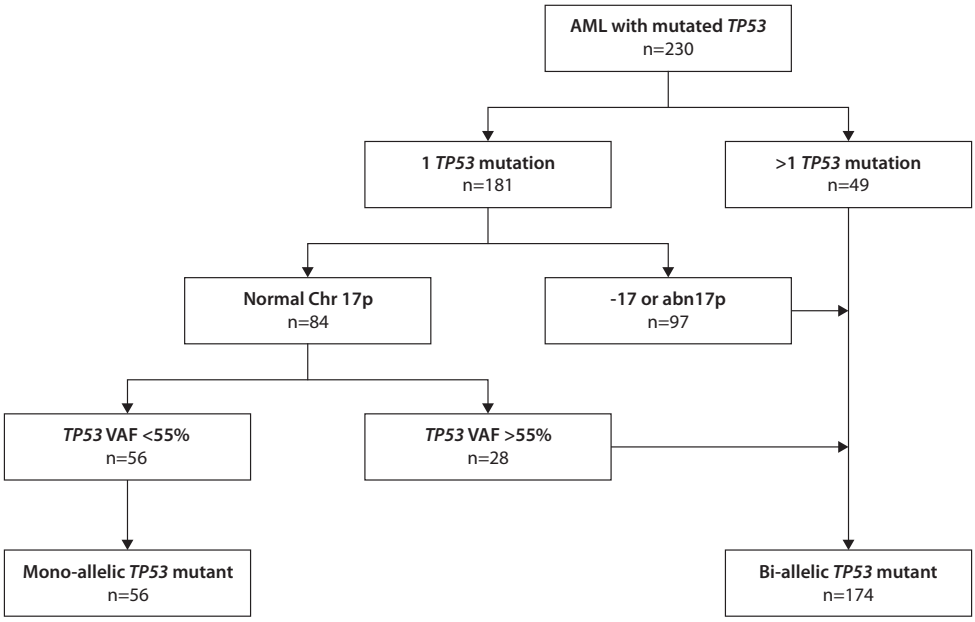


Figure S3: Lollipop plot illustrating the distribution of 228 *TP53* gene variants in 230 mutant *TP53* AML patients. The variants shown are depicted by using the NM_000546 transcript reference sequence. Functional protein domains are indicated in purple (TAD: Trans-activation domain), orange (DNA binding domain) and yellow (oligomerization domain). The upper panel shows the distribution of missense mutations (green) and the lower panel the nonsense and indel mutations. The Y-axis represent the number of detected gene variants. Splice-site mutations (n=23) are not reported and 32 gene variants are out of scope of the reference sequence.

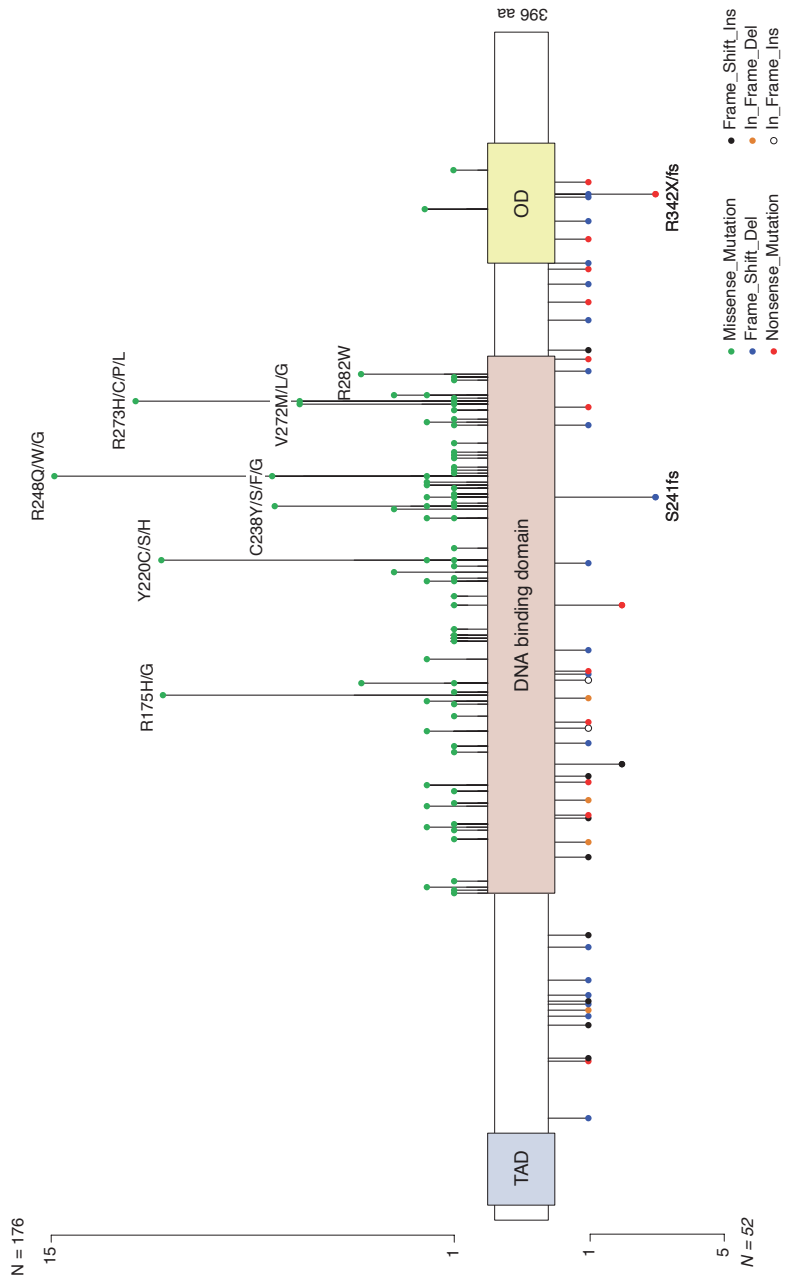


Figure S4: Variant allele frequency of *TP53* mutations (A) and *TP53* variant allele frequency in relation to chromsome 17 aberrations and chromosomal instability (B). *TP53* mutant AML patients without an abnormality involving chromosome 17 (norm17) or chromosomal instability (blue); *TP53* mutant AML patients with norm17 and chromosomal instability (orange) and *TP53* mutant AML patients with an abnormality involving chromosome 17 (abn17) and chromosomal instability (gray).

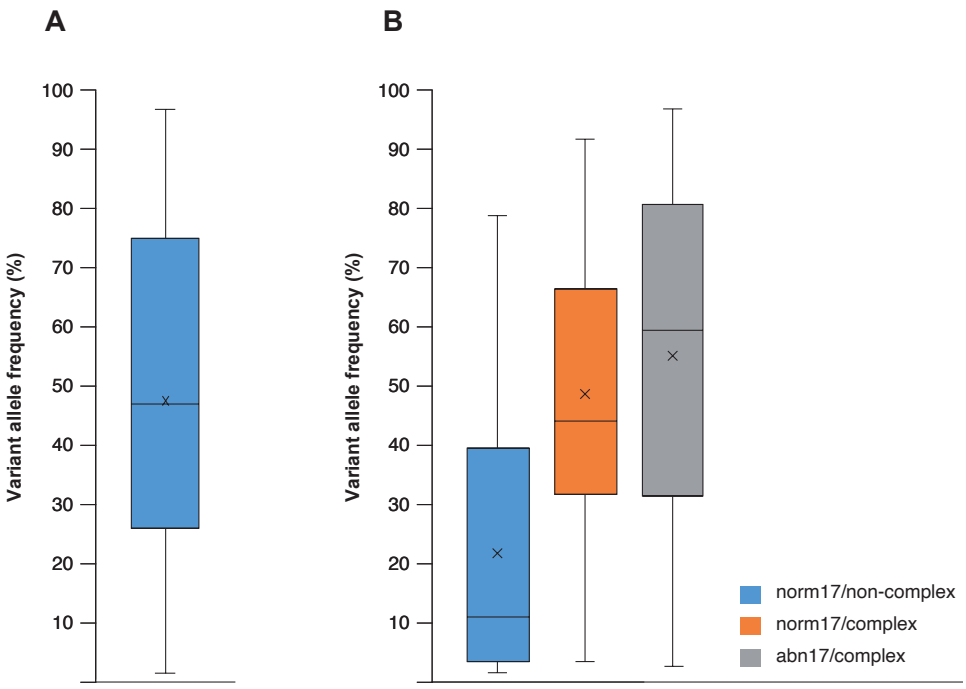


Figure S5: Overall survival of mutant *TP53* AML by the number of *TP53* mutations (A) and concurrent chromosome 17 aberrations (B,C).

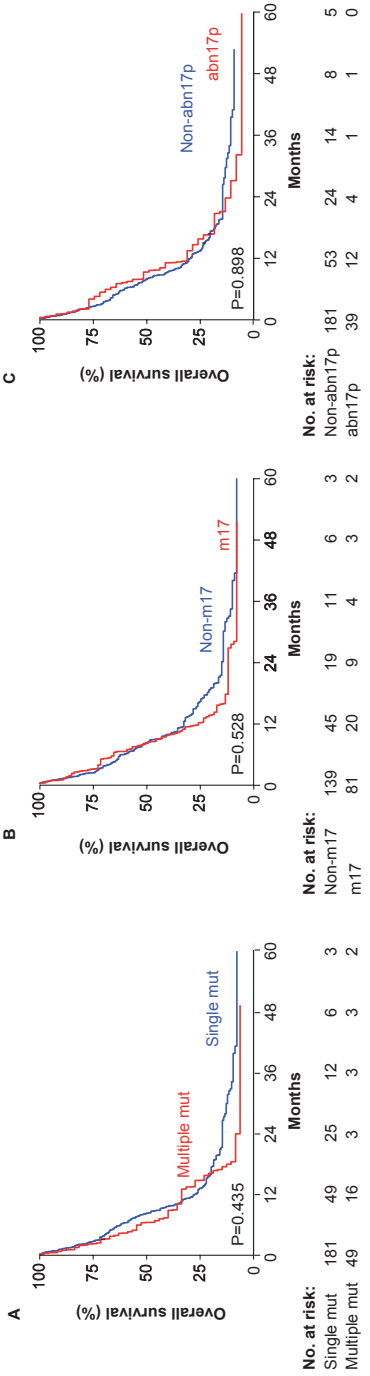


Figure S6: Overall survival by mutant *TP53* variant allele frequency cut off.

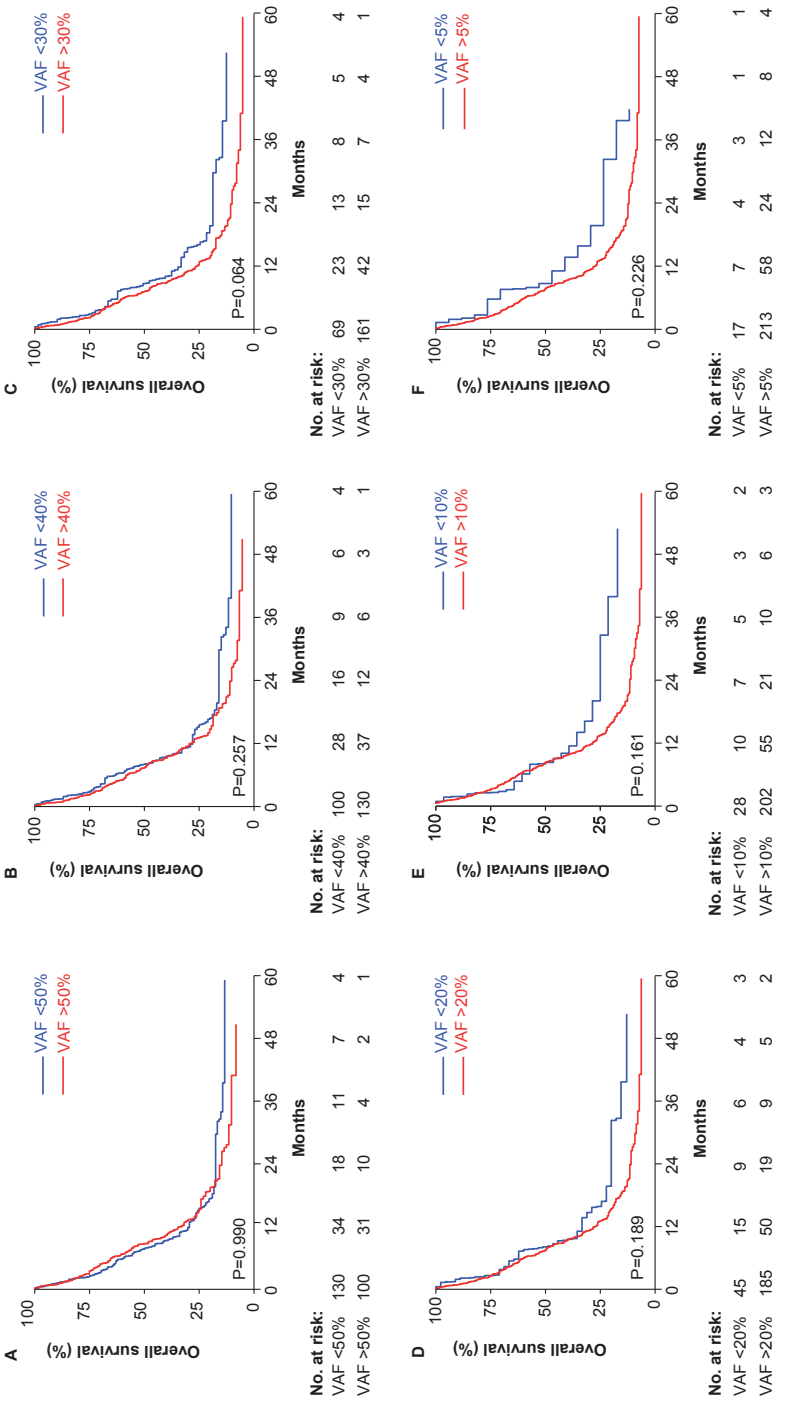
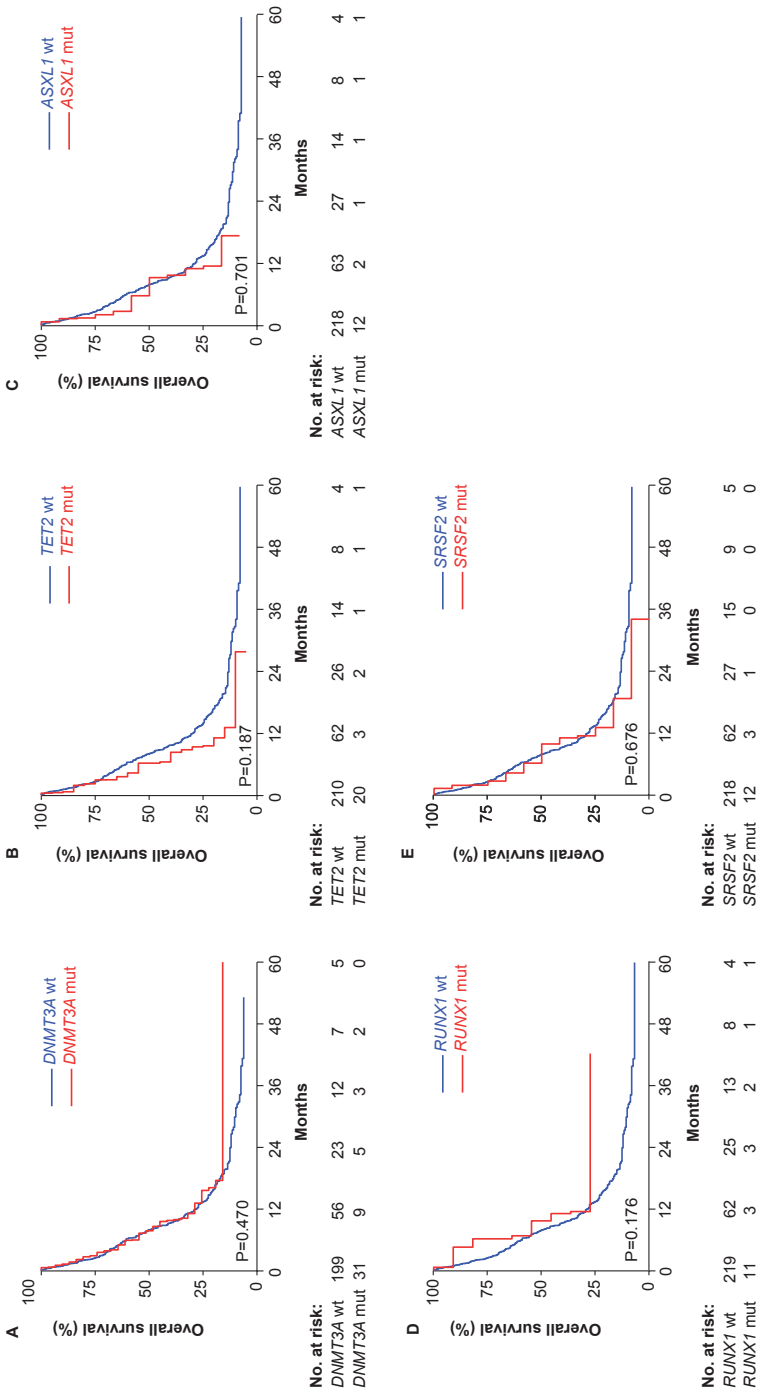


Figure S7: Overall survival of mutant *TP53* AML by individual concurrent mutations.



REFERENCES

1. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C*, et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013 Mar; **31**(3)**:** 213-219.
2. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N*, et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009 Aug 15; **25**(16)**:** 2078-2079.
3. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C*, et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011 May; **43**(5)**:** 491-498.
4. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L*, et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 2012 Mar; **22**(3)**:** 568-576.
5. Indelocator. at http://www.broadinstitute.org/cancer/cga/indelocator.

Chapter 4

PPM1D mutations appear in complete remission after exposure to chemotherapy without predicting emerging AML relapse

Adil S. A. Al Hinai[1,2], Tim Grob[1], Melissa Rijken[1], François G. Kavelaars[1], Annelieke Zeilemaker[1], Claudia A. J. Erpelinck-Verschueren[1], Mathijs A. Sanders[1], Bob Löwenberg[1], Mojca Jongen-Lavrencic[1], Peter J. M. Valk[1]

[1] Department of Hematology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, Rotterdam, the Netherlands

[2] National Genetic Center, Royal Hospital, Ministry of Health, Muscat, Sultanate of Oman

Leukemia. 2021 Sep;35(9):2693-2697

TO THE EDITOR:

Acute myeloid leukemia (AML) is a heterogeneous clonal disorder characterized by the uncontrolled clonal expansion of undifferentiated myeloblasts.[1, 2] Hematopoietic clonal expansion is often the result of age-related acquisition of somatic driver mutations in hematopoietic stem or progenitor cells (HSPCs). In these conditions HSPCs acquire distinct sets of mutations[3], conferring fitness advantage and clonal expansion, ultimately leading to a genetically heterogeneous group of diseases.[4-6] Increased clonal expansion has been observed to occur subsequent to particular biologic conditions and cellular stressors, such as aplastic anemia[7] and cytotoxic therapy[8, 9], resulting in outgrowth of cells carrying specific subsets of mutations, i.e., in *PIGA*, *BCOR*, and *BCORL1* in AA[7] and in *TP53*, *CHEK2*, and *PPM1D*[8, 9] after chemo- and radiotherapy. This observation suggests that cellular stressors affect clonal evolution toward the development of secondary malignancy or relapse. Recently, Wong *et al*. demonstrated that stress from cytotoxic therapy promotes the expansion of clones with mutations in DNA damage response (DDR) genes, such as protein phosphatase Mg2+/Mn2+ 1D (*PPM1D*).[10]

The *PPM1D* gene encodes a protein phosphatase that is upregulated in a p53-dependent manner in response to DNA damage. PPM1D regulates the DDR pathway through dephosphorylation of TP53 and other tumor suppressor target proteins.[11] PPM1D functions as an inhibitor or homeostatic regulator of the DDR by facilitating the return of cells into steady state following DNA damage.[12, 13] Mutations in the *PPM1D* gene have been found to occur in various malignancies, such as breast and ovarian cancers.[14] *PPM1D* mutations are also reported to be present in blood of healthy individual with no history of hematological malignancy; indicating that this gene drives clonal hematopoiesis.[4] In myeloid neoplasms, *PPM1D* mutations are more frequent in patients with therapy related than in de novo disorders.[15] Remarkably, virtually all PPM1D mutations are truncating variants that predominantly occur in exon 6 of *PPM1D* and are considered to be gain-of-function mutations that impair TP53 function.[15, 16] Recently, Kahn *et al*. showed that protein truncating *PPM1D* mutations result in a chemo resistance phenotype leading to expansion of *PPM1D* mutant cells under selective pressure, as a consequence of an altered DDR response.[16] In vitro and in vivo expansion of *PPM1D* mutant clones in the presence of chemotherapy was reversible using PPM1D inhibitors.[16]

In this study we determined the frequency of *PPM1D* mutations in a cohort of 685 newly diagnosed AML and cases of refractory anemia with excess of blasts (RAEB) (Supplementary Tables S1–S5). All cases were enrolled in the AML clinical trials of the Haemato-Oncology Foundation for Adults in the Netherlands and Swiss Group for Clinical Cancer Research (HOVON/SAKK). Mononuclear cells were isolated from all bone marrow and peripheral blood samples at diagnosis. The complete exons 5 and 6 of the *PPM1D* gene were examined using amplicon-based next-generation sequencing (NGS; see Supplementary Data for

details). *PPM1D* mutations were present in only 4 out of the 685 newly diagnosed AML cases (0.6%; Supplementary Table S6). In three, relatively young AML patients virtually all mononuclear cells contained the *PPM1D* mutation (VAF ~45%), whereas in the remaining AML case the *PPM1D* mutation was present in only a minor clone (VAF 1%). Two out of four cases received prior. All PPM1D mutations present at diagnosis were located in exon 6 and resulted in a premature stop of *PPM1D* translation as a consequence of a frame shift (n = 1) or a premature stop codon (n = 3) (Supplementary Table S6).

We subsequently addressed the question whether chemoresistant mutant *PPM1D* AML cells due to their potential selective advantage would become more prevalent after high dose chemotherapy and possibly contribute to disease relapse. Genomic DNA after two cycles of induction treatment was available for 448 out of 685 AML patients who achieved a complete morphological remission (CR). *PPM1D* mutations were detected by NGS in 14 AML patients in CR (Supplementary Table S6). Of the four AML patients carrying a *PPM1D* mutation at diagnosis, one patient retained the same *PPM1D* mutation in CR (#14333; VAF at diagnosis: 49% and in CR: 15%), one patient lost the *PPM1D* mutation (#264; VAF at diagnosis 43%), and the remaining two patients were not evaluable, i.e., no CR (#47642) or follow-up material was unavailable (#29274).

In 13 out of 448 AML patients (2.9%) without detectable *PPM1D* mutations at diagnosis, mutations became apparent at time of CR after induction cycle 2 (Supplementary Table S6). All *PPM1D* mutations were located in exon 6 and introduced frame shifts (n = 6) or premature stop codons (n = 11) (Supplementary Table S6). This phenomenon is analogous to mutant *TP53* cells that can be present at very low frequencies in the bone marrow prior to chemotherapy and grow out after due to a competitive advantage.[10] In total 18 *PPM1D* mutations were detected in CR, with VAFs ranging from 1 to 18%. The majority of the mutant *PPM1D* clones were small, 11 out of 18 *PPM1D* mutations had VAFs between 1 and 3%, indicative for clone sizes of up to 6% of the total mononuclear cells. Remarkably, 3 out of the 13 patients bone marrow carried 2 or even 3 *PPM1D* mutations in CR, whereas these were all absent at diagnosis. Most mutations were unique, however, PPM1D p.(Arg458*) (3×), p.(Cys478*) (2× in CR and 1× at diagnosis (#264)), and p.(Arg552*) (2×) were more recurrently noted (Supplementary Table S6), even in this small series of AML cases. Of note, these mutations all represent C-to-T and C-to-A variants, the type of mutations most frequently seen in clonal hematopoiesis of indeterminate potential (CHIP).[5] The C>T transition mutations in *PPM1D* may indeed be related to aging, whereas the C>A mutations may be the result of errors during transcription-coupled DNA repair.[17] Remarkably, *PPM1D* mutations in CR were significantly more frequently present in elderly (8 out of 125 (6.4%; >60 years) versus 6 out of 323 (1.9%; <60 years) (p = 0.028), which may suggest that these variants are associated with clonal hematopoiesis rather than disease. In fact, *PPM1D* mutations is among the most frequent mutated gene in CHIP after *DNMT3A*, *TET2*, and *ASXL1*.[5] In addition, in our study *PPM1D* mutations were found in CR across various subtypes

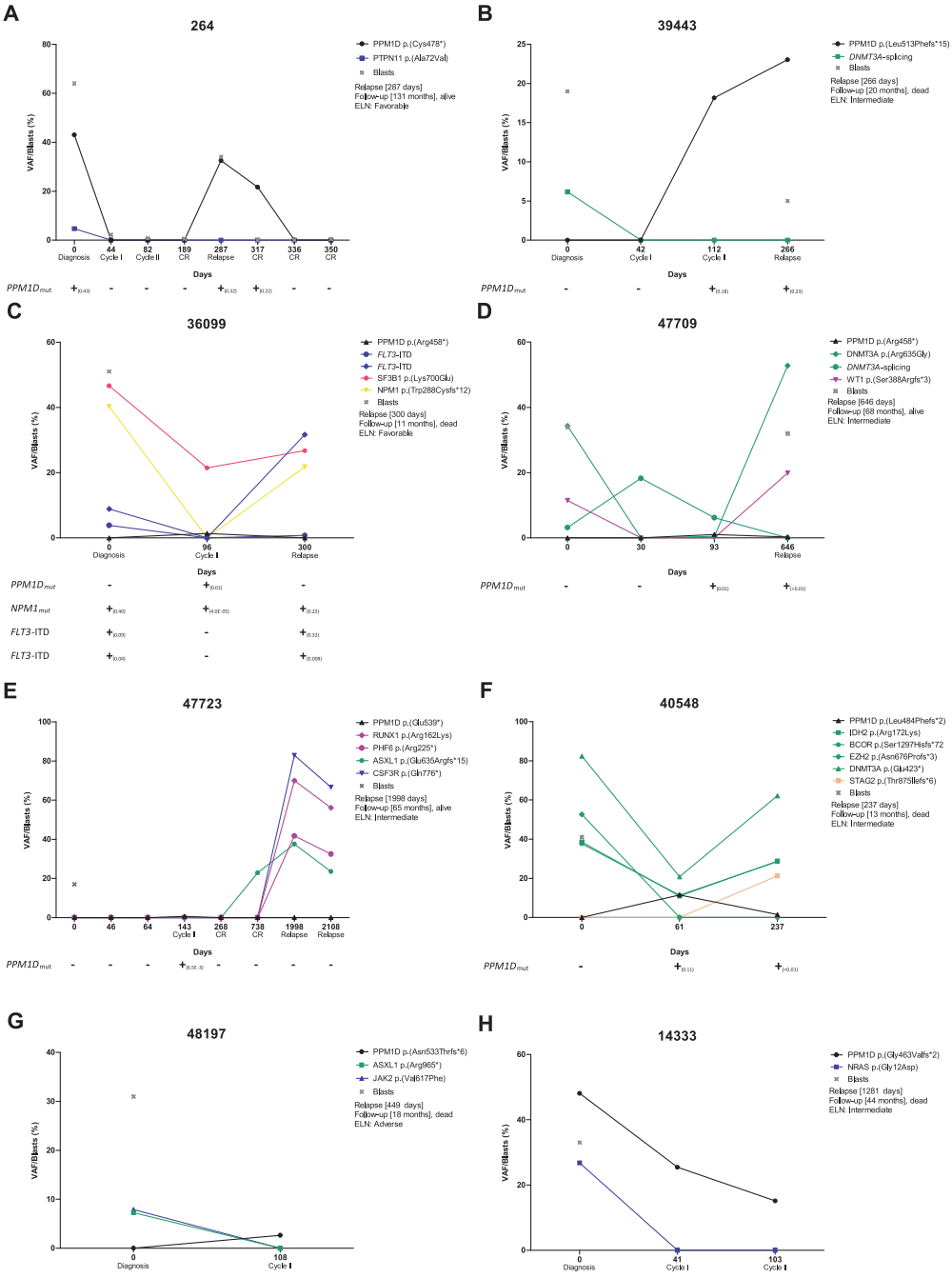


Figure 1: Evolution of *PPM1D* and co-occurring mutations during treatment in AML patients with relapse. Somatic mutations present at AML diagnosis and during the course of disease in all relapsed cases (A–H; n = 8) [VAF variant allele frequency (%), Blasts (%), persistence of **PPM1D**, *NPM1*, and *FLT3*-ITD (including VAF) are indicated below the relevant figures, CR complete remission].

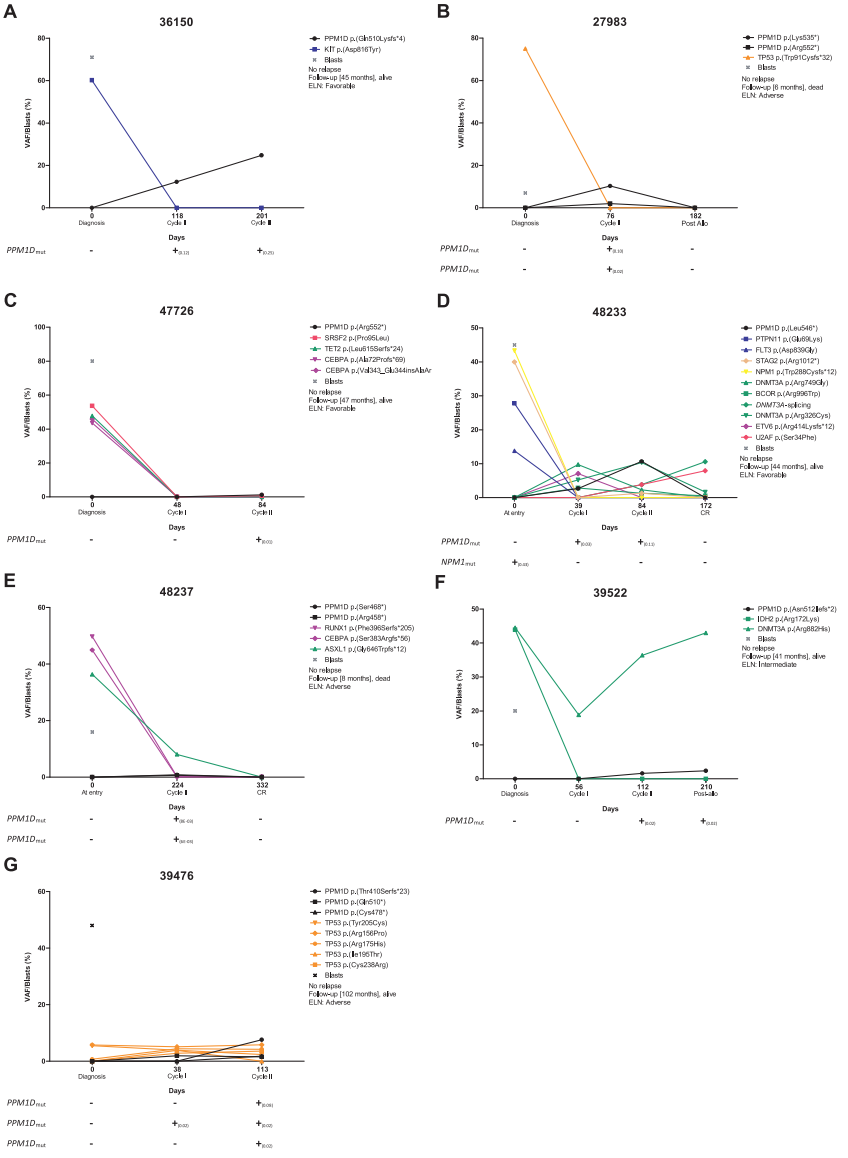


Figure 2: Evolution of PPM1D and co-occuring mutations during treatment in AML patients without relapse. Somatic mutations present at AML diagnosis and during the course of disease in all relapsed cases (A–G; n = 8) [VAF variant allele frequency (%), Blasts (%), persistence of PPM1D, NPM1, and FLT3-ITD (including VAF) are indicated below the relevant figures, CR complete remission].

of AML carrying a wide variety of (cyto)genetic aberrations, e.g., core binding factor translocations, abnormalities involving chromosome 7, and/or various mutations, indicating that the *PPM1D* aberration did not co-segregate with any specific set of mutations when expanding after high dose chemotherapy (Supplementary Table S6). PPM1D mutations are observed in patients with solid tumors following exposure to chemotherapy.[8-10] Seven out of the 17 AML cases with *PPM1D* mutations at diagnosis or in CR had leukemias secondary to antecedent disease (Supplementary Table S6). In five of these cases, who received radio- and/or chemotherapy, the treatment possibly preferentially selected for mutations in genes involved in the DDR pathway, like *PPM1D* .[15] *PPM1D* mutations were significantly more frequently present in AML cases who received previous therapy at diagnosis ($p = 0.027$) and in CR ($p = 0.008$).

We next addressed the question if *PPM1D* mutations appearing in CR would contribute to AML relapse. In addition to *PPM1D* mutations, we monitored all mutations present at AML diagnosis in all follow-up samples (Fig. 1 (relapsed cases (n = 8)) and Fig. 2 (non-relapsed cases (n = 7))). The AML patient who lost the diagnostic *PPM1D* mutation in CR (#264) relapsed after 287 days. In relapse the p.(Asn533Thrfs*6) mutation (VAF 32%/blasts 64% (Fig. 1A)) was present. Of the 13 AML cases with *PPM1D* mutations only present in CR, 7 experienced disease relapse (Fig. 1B–H). Genomic DNA for 5 out of these 7 relapses was available at time of relapse. Only in a single relapse (#39443) the *PPM1D* mutation was present at a high level (VAF 23%) (Fig. 1B), however, the blast count at relapse was 5%, indicating the *PPM1D* mutation most probably represented clonal hematopoiesis rather than frank leukemia. In all other AML cases the low VAF of the *PPM1D* mutations indicates the improbability of their participation in disease relapse. This observation is further augmented by their poor co-segregation with other relapse-specific driver mutations (Fig. 1C–F). Although the remaining AML cases showed stable or even rising levels of mutant *PPM1D* during the course of disease, these patients never experienced a relapse of the disease (Fig. 2). Remarkably, one of these AML cases (#39476), previously treated for primary cerebral lymphoma, carried multiple *TP53* as well as *PPM1D* mutant clones during the disease course (Fig. 2H).

In conclusion, *PPM1D* mutations are rare in de novo AML and RAEB in contrast to therapy-related AML[15] but present at higher frequency after high dose induction chemotherapy at generally low VAFs. Importantly, the presence of *PPM1D* mutations in CR do not seem to herald AML relapse.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

REFERENCES

1. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ*, et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013 May 30; **368**(22)**:** 2059-2074.
2. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. *N Engl J Med* 2015 Sep 17; **373**(12)**:** 1136-1152.
3. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC*, et al.* The Origin and Evolution of Mutations in Acute Myeloid Leukemia. *Cell* 2012 Jul 20; **150**(2)**:** 264-278.
4. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF*, et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014 Dec 25; **371**(26)**:** 2477-2487.
5. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG*, et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine* 2014 Dec 25; **371**(26)**:** 2488-2498.
6. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC*, et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med* 2014 Dec; **20**(12)**:** 1472-1478.
7. Yoshizato T, Dumitriu B, Hosokawa K, Makishima H, Yoshida K, Townsley D*, et al.* Somatic Mutations and Clonal Hematopoiesis in Aplastic Anemia. *N Engl J Med* 2015 Jul 2; **373**(1)**:** 35-47.
8. Coombs CC, Zehir A, Devlin SM, Kishtagari A, Syed A, Jonsson P*, et al.* Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell* 2017 Sep 7; **21**(3)**:** 374-382 e374.
9. Gibson CJ, Lindsley RC, Tchekmedyian V, Mar BG, Shi J, Jaiswal S*, et al.* Clonal Hematopoiesis Associated With Adverse Outcomes After Autologous Stem-Cell Transplantation for Lymphoma. *J Clin Oncol* 2017 May 10; **35**(14)**:** 1598-1605.
10. Wong TN, Miller CA, Jotte MRM, Bagegni N, Baty JD, Schmidt AP*, et al.* Cellular stressors contribute to the expansion of hematopoietic clones of varying leukemic potential. *Nat Commun* 2018 Jan 31; **9**(1)**:** 455.
11. Chuman Y, Kurihashi W, Mizukami Y, Nashimoto T, Yagi H, Sakaguchi K. PPM1D430, a novel alternative splicing variant of the human PPM1D, can dephosphorylate p53 and exhibits specific tissue expression. *J Biochem* 2009 Jan; **145**(1)**:** 1-12.
12. Lu XB, Nguyen TA, Moon SH, Darlington Y, Sommer M, Donehower LA. The type 2C phosphatase Wip1: An oncogenic regulator of tumor suppressor and DNA damage response pathways. *Cancer Metast Rev* 2008 Jun; **27**(2)**:** 123-135.
13. Fiscella M, Zhang HL, Fan SJ, Sakaguchi K, Shen SF, Mercer WE*, et al.* Wip1, a novel human protein phosphatase that is induced in response to ionizing radiation in a p53-dependent manner. *P Natl Acad Sci USA* 1997 Jun 10; **94**(12)**:** 6048-6053.
14. Ruark E, Snape K, Humburg P, Loveday C, Bajrami I, Brough R*, et al.* Mosaic PPM1D mutations are associated with predisposition to breast and ovarian cancer. *Nature* 2013 Jan 17; **493**(7432)**:** 406-410.
15. Hsu JI, Dayaram T, Tovy A, De Braekeleer E, Jeong M, Wang F*, et al.* PPM1D Mutations Drive Clonal Hematopoiesis in Response to Cytotoxic Chemotherapy. *Cell Stem Cell* 2018 Nov 1; **23**(5)**:** 700-713 e706.
16. Kahn JD, Miller PG, Silver AJ, Sellar RS, Bhatt S, Gibson C*, et al.* PPM1D-truncating mutations confer resistance to chemotherapy and sensitivity to PPM1D inhibition in hematopoietic cells. *Blood* 2018 Sep 13; **132**(11)**:** 1095-1105.
17. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV*, et al.* Signatures of mutational processes in human cancer. *Nature* 2013 Aug 22; **500**(7463)**:** 415-421.

SUPPLEMENTARY DATA

Supplementary Materials and Methods

Patients

We selected diagnostic samples of 685 *de novo* AML or refractory anemia with excess of blasts (RAEB) cases enrolled in the Haemato-Oncology Foundation for Adults in the Netherlands and Swiss Group for Clinical Cancer Research (HOVON/SAKK) clinical trials. Follow up material after induction cycle 2 of 448 out of the 685 AML cases was included. Follow-up bone marrow samples were taken at least 21 days after the start of the second induction cycle. If additional samples were available, the most recent sample prior to start of consolidation therapy was selected. In case no consolidation therapy was given, the last sample that was available within a four months interval from start of the second induction cycle was selected. Detailed clinical characteristics of the selected AML cases at diagnosis and follow up are described in Supplementary Table S1 and Supplementary Table S6. The study was conducted according to the declaration of Helsinki.

Cells and DNA isolation

Blasts and mononuclear cells at diagnosis were purified by Ficoll-Hypaque (Nygaard, Oslo, Norway) density gradient centrifugation and cryopreserved. Of the majority of all AML/RAEB cases reaching CR, white blood cells were isolated after induction treatment, whereas of a minority of cases mononuclear cells were subsequently purified. After thawing, cells were lysed in RLT solution with the addition of DTT (Qiagen, Venlo, The Netherlands). High quality DNA was extracted using the QIASymphony DSP DNA Mini Kit according to the manufacturer's instructions (Qiagen, Venlo, the Netherlands). DNA concentration was measured by Qubit Fluorometric Quantitation (Thermo Fisher Scientific, Wilmington, DE).

Next-Generation Sequencing (Custom *PPM1D* Panel)

Primer design

Primers for *PPM1D* mutation NGS analysis were designed by using the Ion AmpliSeq Designer software (ThermoFisher Scientific, Bleiswijk).

The *PPM1D* specific primers (Supplementary Table S2) were adapted for Illumina-based sequencing by adding Illumina forward (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3') and reverse (5'-TCGCGAGTTAATGCAACGATCGTCGAAATTCGC-3') overhang adaptor sequences to the gene specific primers. In the second PCR, by means of these overhang adaptor sequences, sample-specific dual indices for sample identification and Illumina sequencing adaptors are attached by using primers from the Illumina TruSeq Custom Amplicon Index Kit (Illumina, San Diego, CA, USA).

PPM1D NGS library preparation

The *PPM1D* gene was amplified by multiplex PCR on 50ng genomic DNA using the Roche FastStart High fidelity PCR System (Roche) containing 1× Buffer with 1.8mM MgCl2, 0.2mM dNTP, 0.4µM each primer, 0.1U FastStart Taq DNA polymerase. Amplification was performed using the following PCR condition: 95°C for 5 minutes, 25 cycles of 30 seconds at 95°C, 30 seconds at 60°C, and 30 seconds at 72°C and a final extension for 7 minutes at 72°C.

Amplicons from the first step PCR were purified using the Agencourt AMPure XP bead purification kit (Beckman Coulter, Fullerton, CA, USA). The second step PCR was performed with primers from the Illumina TruSeq Custom Amplicon Index Kit ((Illumina, San Diego, CA) using the KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, MA, USA) with the following thermocycling condition: 95°C for 5 minutes, 10 cycles of 20 seconds at 98°C, 30 seconds at 66°C, and 30 seconds at 72°C and a final extension for 1 minute at 72°C. The library pool was purified with Agencourt AMPure XP beads and normalized for Illumina-based sequencing, according to the manufacturer's protocol (Illumina, San Diego, CA).

The NGS libraries were paired-end sequenced (2×221-bp) on the MiSeq according to manufacturer's recommendation (Illumina, San Diego, CA). The mean sequencing depth for *PPM1D* variants at diagnosis was 20383 reads (range 1034-71653) and at CR 7497 (range 930-18446). The limit of detection was VAF 1.5%.

Next-Generation Sequencing (Custom *CEBPA* Panel)

Primer design

Primers for *CEBPA* mutation NGS analysis were designed by using the Ion AmpliSeq Designer software (ThermoFisher Scientific, Bleiswijk).

The *CEBPA* specific primers (Supplementary Table S3) were adapted for Illumina-based sequencing by adding Illumina forward (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3') and reverse (5'-TCGCGAGTTAATGCAACGATCGTCGAAATTCGC-3') overhang adaptor sequences to the gene specific primers. In the second PCR, by means of these overhang adaptor sequences, sample-specific dual indices for sample identification and Illumina sequencing adaptors are attached by using primers from the Illumina TruSeq Custom Amplicon Index Kit (Illumina, San Diego, CA, USA).

CEBPA NGS library preparation

The *CEBPA* gene was amplified by 4 PCR reactions on 50ng genomic DNA using the Roche FastStart High fidelity PCR System (Roche) containing 1× Buffer with 1.8mM MgCl2, 0.2mM dNTP, 0.4µM each primer, 0.1U FastStart Taq DNA polymerase. Amplification was performed using the following PCR condition: amplicons A/B/C2: 94°C for 5 minutes, 25 cycles of 1 minute at 94°C, 1 minute at 56°C, and 1 minute at 72°C and a final extension for 10 minutes at 72°C and amplicon C1: 94°C for 5 minutes, 27 cycles of 1 minute at 94°C, 1 minute at 62°C, and 1 minute at 72°C and a final extension for 10 minutes at 72°C.

Amplicons from the first step PCR were purified using the Agencourt AMPure XP bead purification kit (Beckman Coulter, Fullerton, CA, USA). The second step PCR was performed with primers from the Illumina TruSeq Custom Amplicon Index Kit ((Illumina, San Diego, CA) using the KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, MA, USA) with the following thermocycling condition: 95°C for 5 minutes, 10 cycles of 20 second at 98°C, 30 seconds at 66°C, and 30 seconds at 72°C and a final extension for 1 minutes at 72°C. The library pool was purified with Agencourt AMPure XP beads and normalized for Illumina-based sequencing, according to the manufacturer's protocol (Illumina, San Diego, CA).

The NGS libraries were paired-end sequenced (2×221-bp) on the MiSeq according to manufacturer's recommendation (Illumina, San Diego, CA).

Next-Generation Sequencing (Custom *FLT3*)

Primer design

Primers for *FLT3* mutation NGS analysis were designed by using the Ion AmpliSeq Designer software (ThermoFisher Scientific, Bleiswijk).

The *FLT3* specific primers (Supplementary Table S4) were adapted for Illumina-based sequencing by adding Illumina forward (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3') and reverse (5'-TCGCGAGTTAATGCAACGATCGTCGAAATTCGC-3') overhang adaptor sequences to the gene specific primers. In the second PCR, by means of these overhang adaptor sequences, sample-specific dual indices for sample identification and Illumina sequencing adaptors are attached by using primers from the Illumina TruSeq Custom Amplicon Index Kit (Illumina, San Diego, CA, USA).

FLT3 NGS library preparation

The *FLT3* exon 14 was amplified by PCR on 100ng genomic DNA using the Roche FastStart High fidelity PCR System (Roche) containing 1× Buffer with 1.8mM MgCl2, 0.2mM dNTP, 0.4μM each primer, 0.1U FastStart Taq DNA polymerase. Amplification was performed using the following PCR condition: 95°C for 5 minutes, 25 cycles of 30 seconds at 95°C, 30 seconds at 60°C, and 30 seconds at 72°C and a final extension for 7 minutes at 72°C.

Amplicons from the first step PCR were purified using the Agencourt AMPure XP bead purification kit (Beckman Coulter, Fullerton, CA, USA). The second step PCR was performed with primers from the Illumina TruSeq Custom Amplicon Index Kit ((Illumina, San Diego, CA) using the KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, MA, USA) with the following thermocycling condition: 95°C for 5 minutes, 10 cycles of 20 second at 98°C, 30 seconds at 66°C, and 30 seconds at 72°C and a final extension for 1 minutes at 72°C. The library pool was purified with Agencourt AMPure XP beads and normalized for Illumina-based sequencing, according to the manufacturer's protocol (Illumina, San Diego, CA).

The NGS libraries were paired-end sequenced (2×221-bp) on the MiSeq according to manufacturer's recommendation (Illumina, San Diego, CA).

Next-Generation Sequencing (Custom *NPM1)*

Primer design

Primers for *NPM1* mutation NGS analysis were designed by using the Ion AmpliSeq Designer software (ThermoFisher Scientific, Bleiswijk).

The *NPM1* specific primers (Supplementary Table S5) were adapted for Illumina-based sequencing by adding Illumina forward (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3') and reverse (5'-TCGCGAGTTAATGCAACGATCGTCGAAATTCGC-3') overhang adaptor sequences to the gene specific primers. In the second PCR, by means of these overhang adaptor sequences, sample-specific dual indices for sample identification and Illumina sequencing adaptors are attached by using primers from the Illumina TruSeq Custom Amplicon Index Kit (Illumina, San Diego, CA, USA).

NPM1 NGS library preparation

The *NPM1* exon 12 was amplified by PCR on 100ng genomic DNA using the Roche FastStart High fidelity PCR System (Roche) containing 1× Buffer with 1.8mM MgCl2, 0.2mM dNTP, 0.4μM each primer, 0.1U FastStart Taq DNA polymerase. Amplification was performed using the following PCR condition: 95°C for 5 minutes, 25 cycles of 30 seconds at 95°C, 30 seconds at 60°C, and 30 seconds at 72°C and a final extension for 7 minutes at 72°C.

Amplicons from the first step PCR were purified using the Agencourt AMPure XP bead purification kit (Beckman Coulter, Fullerton, CA, USA). The second step PCR was performed with primers from the Illumina TruSeq Custom Amplicon Index Kit ((Illumina, San Diego, CA) using the KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, MA, USA) with the following thermocycling condition: 95°C for 5 minutes, 10 cycles of 20 second at 98°C, 30 seconds at 66°C, and 30 seconds at 72°C and a final extension for 1 minutes at 72°C. The library pool was purified with Agencourt AMPure XP beads and normalized for Illumina-based sequencing, according to the manufacturer's protocol (Illumina, San Diego, CA).

The NGS libraries were paired-end sequenced (2×221-bp) on the MiSeq according to manufacturer's recommendation (Illumina, San Diego, CA).

Next-Generation Sequencing (TruSight Myeloid Sequencing Panel)

The TruSight Myeloid Sequencing Panel (Illumina, San Diego, CA) was used to generate NGS libraries from 50 ng of DNA according to the TruSight Myeloid Sequencing Reference Guide (Illumina, San Diego, CA, USA). The generated libraries were pooled and paired-end sequenced (2×221-bp) on the Illumina MiSeq according to manufacturer's recommendation (Illumina, San Diego, CA, USA).

DNA sequencing data, alignment, and variant detection

We used our in-house data analysis pipeline for variant calling as previously described.[1] Briefly, overlap-based error-correction was utilized to attenuate any form of strand-specific

error biases. Error-corrected paired-end reads aligned to the human genome version 19 (hg19) with BBMAP3 followed by quality control to determine cases with insufficient number of reads for adequate variant calling. Single nucleotide variants (SNVs) and insertions-deletions (indels) at diagnosis were determined by MuTect[2], Samtools[3], GATK[4], Varscan[5], Indelocator[6] and Pindel[7]. Single nucleotide variants (SNVs) and insertions/deletions (indels) were called for each sample by combining the results produced by MuTect (SNVs), SAMtools (SNVs & indels), GATK (SNV & indels), VarScan (SNVs & indels) and Indelocator (indels). In brief, the union of all variants called across the different variant callers was taken and only variants with a variant allele frequency (VAF) greater than 0.015 (1.5%) were considered candidates. Differences between variant callers in call sets were not considered and a call from a single variant caller was sufficient.

The variant allele frequency (VAF) of both single nucleotide variants (SNVs) and insertions/deletions (indels) were determined by AnnotateBAMStatistics (https://github.com/MathijsSanders/AnnotateBAMStatistics) for all variant caller output. VAFs estimated by the different callers were not considered. AnnotateBAMStatistics is a light-weight multi-threaded application that is paired-end aware. Mutations irrespective of type must be present in both mates when overlapping at the genomic position of interest. In case of discrepancy, the base with the highest base call quality score is selected for SNVs. VAFs of mutations detected at diagnosis were calculated as the ratio between the number of mutant and total reads.

Supplementary Table S1: Detailed clinical characteristics of the selected AML cases at diagnosis (n=685) and at follow up (n=448; after cycle 2).

		n (%)	n (%)
Sex	M	373 (54)	238 (53)
	F	312 (46)	210 (47)
Age (yrs)	median	59	52
Trial	HO42A	123 (18)	120 (27)
	HO92	34 (5)	33 (7)
	HO102	256 (37)	222 (50)
	HO103	272 (40)	73 (16)
AML		633 (92)	419 (94)
tAML		48 (7)	28 (6)
FAB	M0	42 (6)	29 (6)
	M1	127 (19)	89 (20)
	M2	183 (26)	120 (27)
	M3	1 (<1)	0
	M4	77 (12)	63 (14)
	M5	101 (16)	68 (15)
	M6	19 (3)	11 (2)
	M7	1 (<1)	1 (<1)
	RAEB	42 (6)	23 (5)
	RAEB-t	42 (6)	25 (6)
WBC	≤ 20	419 (61)	262 (58)
	20 - 100	201 (29)	140 (31)
	>100	64 (9)	46 (10)
Last treatment before CR	no CR	80 (12)	
	cycle 1	488 (71)	372 (83)
	cycle 2	105 (15)	76 (17)
	later	12 (2)	
Cytogenetics	t(8;21)	30 (4)	25 (6)
	inv(16)	34 (5)	32 (7)
	normal	346 (51)	232 (52)
	other	175 (25)	114 (26)
	monosomy	52 (8)	27 (6)

Supplementary Table S2: Custom *PPM1D* Panel Primer Sequences

Primer name (1st Stage PCR)	Pool name	Primer sequence (including Rd1 or Rd2) for 1st Stage PCR
17-1537 PPM1D_AMP5_Rd1	Pool1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCACAGCGAACACCAAATATT
17-1538 PPM1D_AMP5_Rd2	Pool1	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCACTTCTGGAGAGATGCAGATTACT
17-1539 PPM1D_AMP6_Rd1	Pool1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTATGCATAGATTTGTTGAGTTCTGGGAT
17-1540 PPM1D_AMP6_Rd2	Pool1	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCAATCATGTATCCTTAAAGTCAGGGCTTTAG
17-1541 PPM1D_AMP7_Rd1	Pool1	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAAGGACATTAGAAGAGTCCAATTCTGG
17-1542 PPM1D_AMP7_Rd2	Pool1	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCTAAGTTTGGAAAAACCTCATTTCCCAGAT
17-1551 PPM1D_AMP12_Rd1	Pool2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCGTATGCTCCGAGCAGATAAC
17-1552 PPM1D_AMP12_Rd2	Pool2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCGCTAACCAAAGAACTGGTGTCTATTAAAAC
17-1553 PPM1D_AMP13_Rd1	Pool2	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTCAAAAGATCCAGAACCACTTGA
17-1554 PPM1D_AMP13_Rd2	Pool2	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCACTTAAGCCATTTCGTCTATGCTTCT

Supplementary Table S3: Custom *CEBPA* Panel Primer Sequences

Primer name (1st Stage PCR)	Pool name	Primer sequence (including Rd1 or Rd2) for 1st Stage PCR
CEBPA_A_NGS_Rd1	-	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGCCATGCCGGGAGAACTCT
CEBPA_A_NGS_Rd2	-	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCTTCTCCTGCTGCCGGCTGT
CEBPA_B_NGS_Rd1	-	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCGCCTTCAACGACGAGTT
CEBPA_B_NGS_Rd2	-	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCTTGGCTTCATCCTCCTCGC
CEBPA_C1_NGS_Rd1	-	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGGCCGCTGGTGATCAAG
CEBPA_C1_NGS_Rd2	-	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCACTTCTTGGCCTTGCCCGCG
CEBPA_C2_NGS_Rd1	-	ACACTCTTTCCCTACACGACGCTCTTCCGATCTCCTCCGCGCGAGTGGCGGCA
CEBPA_C2_NGS_Rd2	-	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCCCAGGGCGGTCCCACAGC

Supplementary Table S4: Custom *FLT3* Primer Sequences

Primer name (1st Stage PCR)	Pool name	Primer sequence (including Rd1 or Rd2) for 1st Stage PCR
55-HP_FLT3_ex14_Rd1	-	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGCAAAGACAAATGGTGAGTACGT
56-HP_FLT3_ex14_Rd2	-	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCCTCTATCTGCAGAACTGCCTATTCC

Supplementary Table S5: Custom *NPM1* Primer Sequences

Primer name (1st Stage PCR)	Pool name	Primer sequence (including Rd1 or Rd2) for 1st Stage PCR
47-HP_NPM1_indel-Rd1	-	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTTAACTCTCTGGTGGTAGAATGAAAAATAGA
48-HP_NPM1_indel-Rd2	-	TCGCGAGTTAATGCAACGATCGTCGAAATTCGCGATATCAACTGTTACAGAAATGAAATAAGACG

Genome	Chr	Amplicon Start	Amplicon Stop	Insert Start	Insert Stop	Amplicon length (bp)
hg19	chr17	58733863	58734079	58733885	58734054	217
hg19	chr17	58733863	58734079	58733885	58734054	217
hg19	chr17	58740291	58734079	58740318	58740530	270
hg19	chr17	58740291	58734079	58740318	58740530	270
hg19	chr17	58740699	58740947	58740727	58740918	249
hg19	chr17	58740699	58740947	58740727	58740918	249
hg19	chr17	58734022	58734260	58734044	58734230	239
hg19	chr17	58734022	58734260	58734044	58734230	239
hg19	chr17	58740495	58740763	58740520	58740737	269
hg19	chr17	58740495	58740763	58740520	58740737	269

Genome	Chr	Amplicon Start	Amplicon Stop	Insert Start	Insert Stop	Amplicon length (bp)
hg19	chr19	33793051	33793349	33793071	33793329	299
hg19	chr19	33793051	33793349	33793071	33793329	299
hg19	chr19	33792808	33793110	33792828	33793090	303
hg19	chr19	33792808	33793110	33792828	33793090	303
hg19	chr19	33792492	33792855	33792510	33792830	364
hg19	chr19	33792492	33792855	33792510	33792830	364
hg19	chr19	33792217	33792535	33792237	33792516	319
hg19	chr19	33792217	33792535	33792237	33792516	319

Genome	Chr	Amplicon Start	Amplicon Stop	Insert Start	Insert Stop	Amplicon length (bp)
hg19	chr13	28608130	28608404	28608154	28608379	275
hg19	chr13	28608130	28608404	28608154	28608379	275

Genome	Chr	Amplicon Start	Amplicon Stop	Insert Start	Insert Stop	Amplicon length (bp)
hg19	chr5	170837411	170837635	170837443	170837603	225
hg19	chr5	170837411	170837635	170837443	170837603	225

Supplementary Table S6: Characteristics of *PPM1D* mutant AML patients (diag: diagnosis; fu: follow-up; RAEB: refractory anemia with excess of blasts; WBC: 109/l; platelets: 109/l; bone marrow blasts: %; 2017ELN: risk stratification according to 2017ELN; Tx: (type of) transplant; MDS: myelodysplastic syndrome)

Sample	Sex	Age at diagnosis	*PPM1D* mutation	VAF *PPM1D* diagnosis	NM_003620.3 (*PPM1D*): cDNA	NM_003620.3 (*PPM1D*): protein	FAB diagn	WBC diagn	Platelets diagn	Blast diagn	Cytogenetics at diagnosis
29274	M	53	diagn	48	c.1451T>G	p.(Leu484*)	M1	228	7	86	45,XY,-7[10]
47642	M	68	diagn	1	c.1349T>G	p.(Leu450*)	RAEB	4	88	17	
264	M	45	diagn-lost fu	43	c.1434C>A	p.(Cys478*)	M4	10	125	64	46,XY,add(7)(q22)[4]/46,XY[21] FISH: abberation involving MYH11 and CBFB
14333	F	44	diagn-fu	49	c.1388del	p.(Gly463Valfs*2)	M4	79	83	33	45,XX,der(7;12)(q10;q10)[19]/46,XX,der(16)t(11;16)(q13;q22)[1]
36150	M	59	fu	-	c.1528del	p.(Gln510Lysfs*4)	M2	19	20	71	46,XX,t(8;21)(q22;q22)[2]/47,sl,+4[8]
36099	M	63	fu	-	c.1372C>T	p.(Arg458*)	AML	39	86	51	46,XX,del(9)(q21q37)[2]/46,XX[29]
27983	M	60	fu	-	c.1603A>T c.1654C>T	p.(Lys535*) p.(Arg552*)	RAEB	2	36	7	44~46,X,del(X)(q2?3),del(5)(q14q33),add(12)(p1?0),?16,-17,del(20)(q11q13),-21[12],+mar1[11],+mar2[3][cp16]
47709	F	67	fu	-	c.1372C>T	p.(Arg458*)	unknown	2	105	34	46,XX[20]
47723	M	68	fu	-	c.1615G>T	p.(Glu539*)	RAEB	2	141	17	46,XY[20]
47726	M	77	fu	-	c.1654C>T	p.(Arg552*)	M2	46	27	80	46,XY[20]
48233	M	71	fu	-	c.1636del	p.(Leu546*)	M1	2	21	45	46,XY[20]
48237	M	70	fu	-	c.1372C>T c.1403C>G	p.(Arg458*) p.(Ser468*)	RAEB	6	15	16	46,XY[20]
48197	M	73	fu	-	c.1598del	p.(Asn533Thrfs*6)	unknown	31	31	31	46,XY[20]
39443	M	53	fu	-	c.1538dup	p.(Leu513Phefs*15)	RAEB	3	138	19	46,XY,del(7)(q22q37?)[1]/46,XY[24]
39476	M	36	fu	-	c.1228_1229del c.1528C>T c.1434C>A	p.(Thr410Serfs*23) p.(Gln510*) p.(Cys478*)	M5	1	60	48	42,XY,-2,-4,-6,-12,-13,-15,-16,-17,-18,-20,-21,6mar[cp10]
39522	M	51	fu	-	c.1535del	p.(Asn512Ilefs*2)	M2	2	93	20	46,XY[20]
40548	F	61	fu	-	c.1450_1451dup	p.(Leu484Phefs*2)	M1	2	52	41	46,XX[20]

Mutation status at diagnosis	Relapse	FAB at relapse	WBC count at relapse	Blast at relapse	Cytogenetics at relapse	Mutation status at relapse	Prior chemo	Prior disease
RUNX1, TET2, EZH2, PHF6	No						Yes	Crohn's disease
IDH2, SRSF2, TP53	No						No	
PPM1D, PTPN11	Yes	M2	2	34	46, XY, add(7)(q22)[11]/46, XY[10]	PPM1D	No	
PPM1D, NRAS	Yes	M4	2	30	45,XX,dic(7;12)(p11.2;p11.2)[2]/ 45,sl,del(9)(q34)[8J		Yes	MDS
KIT	No						Yes	Breast carcinoma
SF3B1, NPM1, FLT3-ITD (2x)	Yes	not done	51	3	not done	SF3B1, NPM1, FLT3-ITD (2x)	No	Endometrial cancer
TP53	No						No	
WT1, DNMT3A (2x)	Yes	M2	1	32	46, XX	DNMT3A, WT1	Yes	Breast carcinoma
	Yes	not done	4	49	not done	CSF3R, RUNX1, PHF6, ASXL1, abn involving KMT2A	No	
SRSF2, TET2, CEBPA (2x)	No						No	
FLT3, NPM1, STAG2, PTPN11	No						No	Rectosigmoid carcinoma
ASXL1, RUNX1, CEBPA	No						No	
ASXL1, JAK2	Yes	unknown	1	60	46, XY		No	
DNMT3A	Yes	RAEB	23	5	ND	PPM1D	No	
TP53 (3x)	No						Yes	Cerebral lymphoma
IDH1, DNMT3A	No						No	
IDH2, DNMT3A, EZH2, BCOR	Yes	unknown	unknown	unknown	unknown	IDH2, DNMT3A, BCOR, STAG2	No	

REFERENCES

1. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM*, et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 2004 Apr 15; **350**(16)**:** 1617-1628.
2. BBMap short-read aligner, and other bioinformatics tools. 2016. at http://sourceforge.net/projects/bbmap/.
3. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C*, et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013 Mar; **31**(3)**:** 213-219.
4. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N*, et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009 Aug 15; **25**(16)**:** 2078-2079.
5. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C*, et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011 May; **43**(5)**:** 491-498.
6. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L*, et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 2012 Mar; **22**(3)**:** 568-576.
7. Indelocator. at http://www.broadinstitute.org/cancer/cga/indelocator.

Chapter 5

Archived bone marrow smears are an excellent source for NGS-based mutation detection in acute myeloid leukemia

Adil S. A. Al Hinai[1,2], Tim Grob[2], François G. Kavelaars[2], Melissa Rijken[2], Annelieke Zeilemaker[2], Claudia A. J. Erpelinck-Verschueren[2], Kirsten J. Gussinklo[2], Mathijs A. Sanders[2], Kirsten van Lom[2], Peter J. M. Valk[2]

[1] National Genetic Center, Ministry of Health, Muscat, Sultanate of Oman
[2] Erasmus University Medical Center, Rotterdam, the Netherlands

Leukemia. 2020 Aug;34(8):2220-2224

TO THE EDITOR:

Acute myeloid leukemia (AML) is a heterogeneous disease as illustrated by the enormous diversity in numbers, types and often patient-specific molecular alterations. The current AML prognostic paradigm classifying patient into three 2017ELN risk groups relies solely on cytogenetic and molecular findings at baseline.[1] Besides the established clinically relevant aberrations, numerous acquired mutations have been revealed in AML.[2, 3] The clinical relevance of many of these somatic mutations is currently unknown because sufficient materials of AML patients carrying these less frequent mutations are lacking. Similarly, studies on the evolution of AML between initial diagnosis and relapse and minimal residual disease (MRD) detection have been hampered by inaccurate sampling during the course of the disease. Here we explored the feasibility of using archived May-Grünwald Giemsa stained bone marrow (MGG-stained BM) slides for next-generation sequencing (NGS)-based mutation detection. MGG-stained BM slides, taken at all clinically relevant time points, are widely available in contemporary biobanks. This study shows that archived MGG-stained BM slides are an excellent source for retrospective molecular analyses of AML, which could certainly be valuable for similar analyses of other hematologic malignancies.

We fist demonstrated that DNA could be more efficiently isolated from MGG-stained BM slides, in contrast to archived unstained slides (Supplementary Data). Next, we examined whether targeted NGS using the Illumina TruSight Myeloid sequencing panel was feasible on DNA isolated from the MGG-stained BM slides (Supplementary Figs. S1 and S2 and Supplementary Tables S1–S4). We compared the mutation profiles of DNA samples derived from MGG-stained BM slides and matched DNA extracted from Ficoll-purified BM (FPBM) of 18 paired diagnosis-relapse AML cases. In the majority of cases the mutation profiles i.e., mutation present versus absent, of DNA isolated from the MGG-stained BM slides were identical to FPBM DNA at diagnosis and relapse (Fig. 1A and Supplementary Fig. S3). The majority of MGG-stained slides included were collected after 1997. Interestingly, however, we were able to detected driver mutations in DNA isolated from MGG-stained slides prepared in 1975 (44-year old, Supplementary Data).

We did notice variations in variant allele frequencies (VAFs) in few AML samples. A consistently lower VAF in MGG-stained BM DNA as compared to FPBM DNA could be the result of a variety of cells present on MGG-stained BM slides, such as unmutated stromal or differentiated cells. Ficoll purification enriches for mononuclear cells, which would result in increased percentages of blast cells and higher VAFs in FPBM DNA. In few instances VAFs were higher in MGG-stained BM DNA. These differences were mostly seen in genes known to be associated with clonal hematopoiesis, i.e., *DNMT3A* and *TET2*.[4, 5] In these cases more differentiated cells carrying these mutations may have been lost during Ficoll purification resulting in decreased VAFs in the FPBM samples.

In 5 out of 18 AML pairs we were able to compare the mutation profiles of MGG-stained slides and FPBM DNA at time of complete remission (CR). Targeted NGS analysis on DNA from both the MGG-stained slides and matched FPBM confirmed the feasibility of detecting mutations from stained slides not only at diagnosis and relapse, but also at time of CR (Fig. 1B). The only exception was the SRSF2 (P95H) mutation (VAF: 3%), which was absent in the MGG-stained slide DNA. In contrast, other mutations at VAF of about 1% were detectable in CR in both FPBM and the MGG-stained slide DNA. Importantly, a comparison of the VAFs of 68 mutations in 41 AML and CR samples revealed that the VAFs of the mutations present in FPBM were detectable in MGG-stained BM slides with highly similar VAFs (R^2: 0.87; Fig. 1C), indicating that MGG-stained BM slide DNA is an excellent and reliable source to measure mutation burden in AML.

To demonstrate the feasibility of PCR-based NGS approaches in addition to hybridization-based, we also assessed the feasibility for mutation detection in DNA derived from 4 MGG-stained BM slides with a custom NGS 18 gene panel (Supplementary Table S5). The VAFs detected with the custom NGS hotspot panel were highly similar to those found with the TruSight Myeloid sequencing panel, even at VAFs of 1–2% (Supplementary Fig. S4).

Successively, we addressed the question if MGG-stained BM slides could be utilized to study mutation kinetics in AML during the course of disease. Therefore, we performed NGS on 191 DNA samples derived from MGG-stained BM slides of the 18 AML patients from diagnosis to relapse. On average we sequenced 11 slides per patient (range: 6–16 slides). At diagnosis, we detected 1–6 mutations per patient (mean: 3.1) and at first relapse we detected 0–5 mutations (mean: 3.1).

The power of NGS-based MGG-stained BM DNA mutation detection to retrospectively analyze mutation kinetics in AML is illustrated by four representative cases (Fig. 2 and Supplementary Fig. S5). In the 18 AML patients, we observed four types of clonal evolution patterns as reported previously[6, 7]: the major clone at AML diagnosis (1) is undetectable during treatment but reappears as a relapse clones (n=3) (UPN:25413,25415,25423) (Fig. 2A); (2) gained additional mutations at relapse (gain, n=8) (UPN:25399,25407,25417, 25427,25431,25435,25437,25439) (Fig. 2B); (3) lost mutations at relapse (loss, n=5) (UPN:25411,25421,25425,25429, 25441); and (4) lost mutation(s) and gained new mutation(s) at relapse (gain and loss, n=2) (UPN:25433,25459) (Fig. 2C).

UPN25423 carried at baseline mutations in DNMT3A (R882H), NPM1 (W288fs), and *FLT3*-ITD that were cleared after induction therapy (day 26, Fig. 2A). The patient achieved CR based on morphology and relapsed after 245 days carrying the same set of mutations. The patient reached a second CR but relapsed 209 days post first relapse. In this case, the kinetics of the *NPM1* and bi-allelic *FLT3* mutations nicely follow the BM blast percentages. The mutant *DNMT3A* clone, however, persists and is not entirely eradicated by chemotherapy at various time points, indicating a state of clonal hematopoiesis.

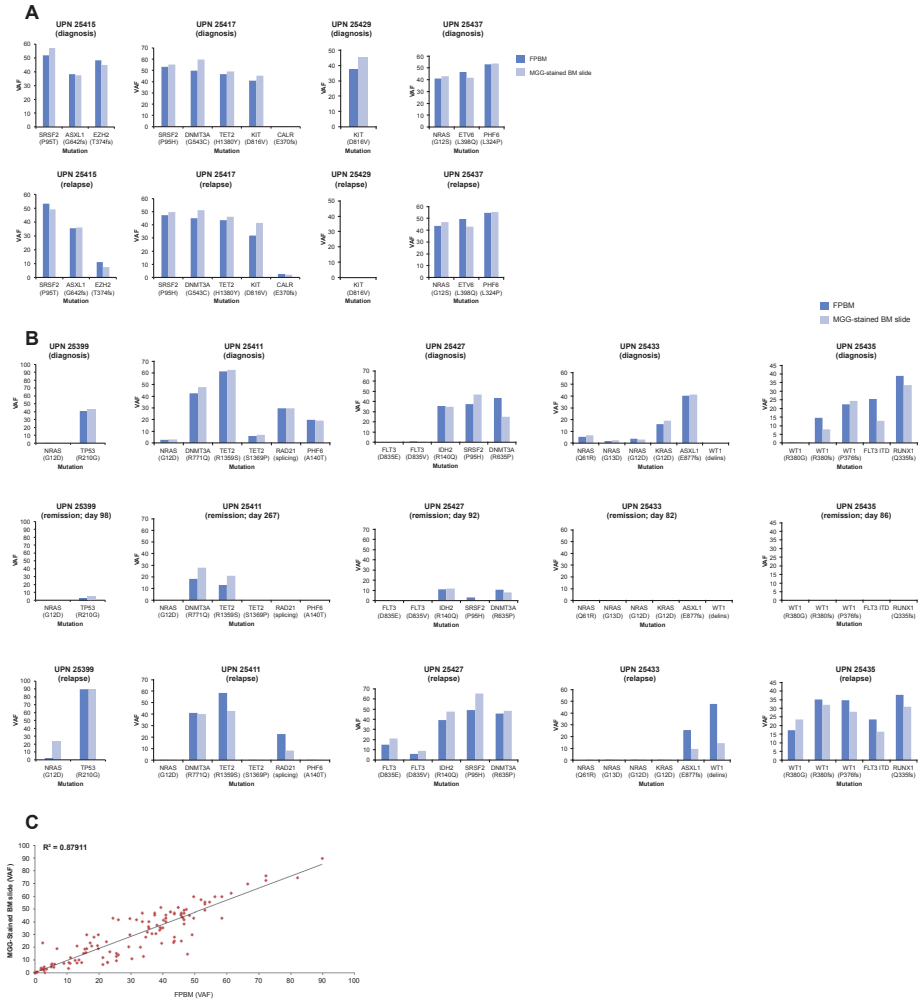


Fig. 1 Mutation detection by targeted NGS on DNA isolated from MGG-stained BM slides and FPBM samples. **a** Mutation profiles of 4 representative AML diagnosis-relapse pairs. **b** Mutation profiles of 5 representative AML diagnosis—CR—relapse trios. **c** Scatter plot showing the correlation between VAFs of mutations detected in FPBM DNA and MGG-stained BM slides. R2 R squared, VAF variant allele frequency.`

UPN25427 had mutations in SRSF2(P95H), IDH2 (R140Q), and DNMT3A (R635P) with approximately similar VAFs. After a single cycle as well as post-induction therapy, the patient achieved a morphologic CR. However, the three mutations present at baseline were not cleared by the two cycles of chemotherapy and remained persistent until relapse. Mutant SRSF2 (P95H) seemed to be cleared post-induction therapy and reappeared before relapse. However, when investigating the FPBM DNA it appeared that the mutant *SRSF2* was still persistent at low level but missed in DNA derived from MGG-stained slide as a result of inadequate coverage (Fig. 1B). At day 334, the patient relapsed, with high blast counts, the same founding mutations and novel mutations in the *FLT3* gene (FLT3-TKD). One of these acquired *FLT3*-TKD mutations (D835V) was eradicated by treatment; however, the second *FLT3*-TKD mutation (D835E) remained stable throughout the relapse phase (Fig. 2b).

UPN25459 is an AML patient with baseline mutations in DNMT3A (R882H), NPM1 (W288fs) and NRAS (G13D) who relapsed twice (Fig. 2C). The NPM1 (W288fs) and NRAS (G13D) mutations were cleared after induction therapy. Interestingly, post induction the DNMT3A (R882H) clone was dramatically reduced but not eradicated (VAF: 2%) and subsequently repopulated the complete BM. The DNMT3A (R882H) mutation remained present during treatment at high VAF and was eradicated only after stem cell transplantation (allogeneic reduced intensity conditioning (RIC) with matched unrelated donor (day 561)). In first relapse, the patient harbored the NPM1 (W288fs) mutation, lost the NRAS (G13D) mutation but gained mutations in IDH1 (R132H) and in FLT3 (D835Y), which were both absent at presentation. The patient received again cycle(s) of chemotherapy that cleared all mutations except for the DNMT3A (R882H) mutant clone. At day 1471 from diagnosis, the patient faced a second very late relapse. At second AML relapse, the patient retained some mutations that were present at diagnosis (NPM1 (W288fs) and DNMT3A (R882H)) and first relapse (IDH1 (R132H)). However, the NRAS (G13D) and FLT3 (D835Y) mutations were lost, suggesting that the patient relapsed on the same founding clone, which acquired additional mutations that were missed with the gene panel used. Interestingly, after transplantation the patient acquired a CR but harbored a new DNMT3A (E30A) that was not previously present in this AML patient. In fact, since the patient had a donor chimerism of >93% after transplant (day 561–855) the BM was repopulated with donor cells carrying a DNMT3A (E30A) variant. This *DNMT3A* variant may either represent donor-derived clonal hematopoiesis or a rare *DNMT3A* germline variant.

At diagnosis, UPN25417 harbored mutations in SRSF2 (P95H), DNMT3A (G543C), TET2 (H1380Y), and KIT (D816V) (Fig. 2D). Sixty nine days (day 73) post induction chemotherapy most mutations were cleared except the *DNMT3A* mutation, which survived. Surprisingly, post high-dose chemotherapy, the patient gained two additional mutations in CALR (E370fs) and JAK2 (V617F) that were not detected at diagnosis (day 73; Fig. 2d). Interestingly, the mutation in the *CALR* gene was detected with a VAF of 50%, indicating that this mutation was present in a heterozygous state in virtually all cells of the bone marrow.

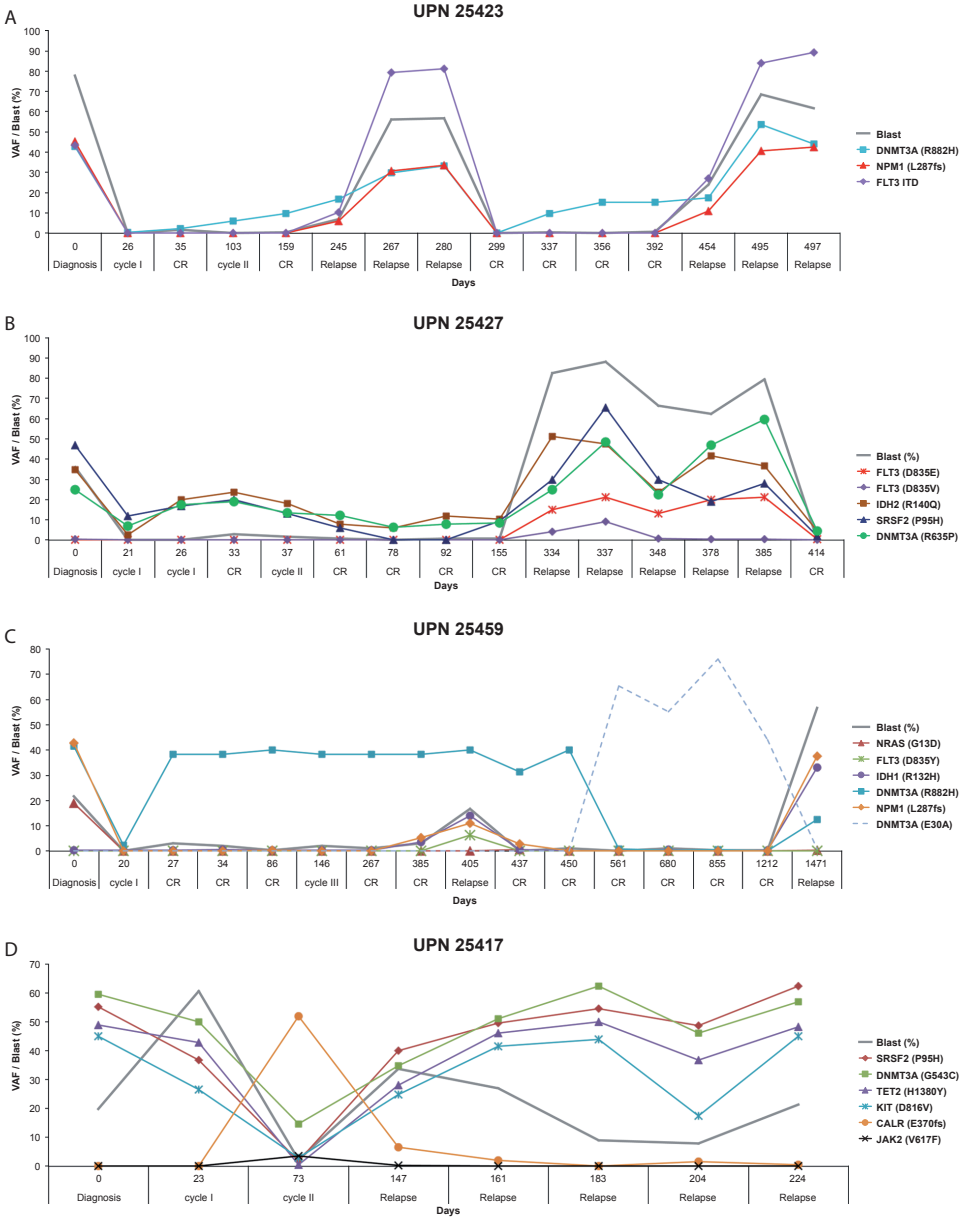


Fig. 2 Longitudinal mutation detection by targeted NGS on DNA derived from MGG-stained BM slides. a–d Sequential mutation analysis from diagnosis until the last available evaluation sampling time point (days). Each line represents the VAF of an acquired mutation in the indicated gene. The thick gray line depicts the BM blast percentage.

However, the mutation burden in the *JAK2* gene at this sampling point was low (VAF: 4%). The co-occurrence of CALR (E370fs) and JAK2 (V617F) mutations in this constellation has been reported in patients diagnosed with myeloproliferative neoplasms.[8, 9] On day 147, the patient relapsed with the same mutations detected at diagnosis and also carried a minor CALR mutant clone without a detectable JAK2 mutation. Although the blast percentages drop over time, the mutation burden in *SRSF2*, *DNMT3A* and *TET2* remain high (VAF≈50%). This indicates that even in the presence of this combination of mutations cells are able to differentiate in a state of clonal hematopoiesis and that sequential mutational analyses during treatment enable clear discrimination between mutations representing true residual disease and clonal hematopoiesis.[5]

In this study, we demonstrate that reliable mutation detection by using NGS on DNA derived from archived AML MGG-stained BM slides taken during the course of disease is feasible. Although we have performed NGS analyses on DNA of archived MGG-stained BM slides of AML patients, we are convinced that the retrospective molecular monitoring using BM sides will be beneficial for other hematologic malignancies as well. It would be of interest to determine if DNA of archived MGG-stained BM slides would also be suitable for NGS-based copy number or structural variant analyses to further classify AML cases according to 2017ELN.

Acknowledgements

We thank all the participating centers of the Dutch–Belgian Cooperative Trial Group for Hematology–Oncology (HOVON) and Swiss Group for Clinical Cancer Research (SAKK), where the clinical trials that formed the basis for this study were conducted; H. Berna Beverloo for performing cytogenetic analyses; Eric Bindels for performing next-generation sequencing, Remco Hoogenboezem for assisting with bioinformatics, and Egied Simons for assisting with preparation of the figures.

Author contributions

ASAAH: performed research, analyzed data and wrote the paper; TG, FGK, MR, AZ, CAJ, KJG, MAS, and KvL: performed research, and analyzed data; and PJMV: designed and performed research analyzed data and wrote the paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest. Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T*, et al.* Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017 Jan 26; **129**(4)**:** 424-447.
2. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ*, et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013 May 30; **368**(22)**:** 2059-2074.
3. Grimwade D, Ivey A, Huntly BJ. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. *Blood* 2016 Jan 7; **127**(1)**:** 29-41.
4. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E*, et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med* 2017 Jul 13; **377**(2)**:** 111-121.
5. Shlush LI. Age-related clonal hematopoiesis. *Blood* 2018 Feb 1; **131**(5)**:** 496-504.
6. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS*, et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 2012 Jan 11; **481**(7382)**:** 506-510.
7. Yilmaz M, Wang F, Loghavi S, Bueso-Ramos C, Gumbs C, Little L*, et al.* Late relapse in acute myeloid leukemia (AML): clonal evolution or therapy-related leukemia? *Blood Cancer J* 2019 Jan 16; **9**(2)**:** 7.
8. Xu N, Ding L, Yin C, Zhou X, Li L, Li Y*, et al.* A report on the co-occurrence of JAK2V617F and CALR mutations in myeloproliferative neoplasm patients. *Annals of Hematology* 2015 2015/05/01; **94**(5)**:** 865-867.
9. Rashid M, Ahmed RZ, Ahmed S, Nadeem M, Ahmed N, Shamsi TS. Coexisting JAK2V617F and CALR Exon 9 Mutation in Essential Thrombocythemia. *Indian J Hematol Blood Transfus* 2016 Jun; **32**(Suppl 1)**:** 112-116.

SUPPLEMENTARY MATERIALS AND METHODS

Patients

We selected samples of 18 de novo AML pairs (diagnosis - relapse) enrolled in the Haemato-Oncology Foundation for Adults in the Netherlands and Swiss Group for Clinical Cancer Research (HOVON/SAKK) clinical trials, and the study was conducted according to the declaration of Helsinki. Detailed clinical characteristics of the selected patients are described in Supplementary Table S1.

The collected samples were archived MGG-stained BM slides and matched FPBM; n=191 and n=41, respectively. The majority of archived MGG-stained BM slides were prepared between 1998 and 2014 at different clinical evaluation points. Two MGG-stained BM prepared in 1975 were analysed. In the comparison between the MGG-stained BM slide and FPBM DNA we examined AML at diagnosis (n=18), complete remission (CR) (n=5) and relapse (n=18).

DNA isolation from MGG-stained BM slides

Before storage, the MGG bone marrow smears were covered using a coverslip and an embedding agent. BM slides were first dipped in xylene in separate glass jars to remove the coverslip. Once the slide was detached from the coverslip, the slide was rinsed with running tap water and then air-dried in a fume hood. Next, RLT buffer (QIAGEN, Venlo, the Netherlands) was dripped over the entire smear (approximately 250 µL). Using the P1000 pipette, the smear was washed off by pipetting the RLT buffer up and down gently several times. Then the entire volume was transferred to a 1.5 ml microfuge tube and incubated at 56°C for 1 hour, vortexed and briefly centrifuged at high speed. Subsequently, the volume was toped-up to 1 ml with PBS. Next, DNA was extracted using QIASymphony DSP DNA Mini Kit according to the manufacturer's instructions (Qiagen, Venlo, the Netherlands). Subsequently, the DNA was concentrated using Amicon Ultra-0.5 mL Centrifugal Filters 30K as described by the manufacturer (Millipore, Amsterdam, the Netherlands). The DNA concentration was measured with Qubit using the broad-range kit as per manufacturer's recommendations (Thermofisher Scientific, Bleiswijk, the Netherlands).

DNA isolation from ficoll-purified mononuclear bone marrow cells (FPBM)

Genomic DNA derived from FPBM was extracted using QIASymphony DSP DNA Mini Kit according to the manufacturer's instructions (Qiagen, Venlo, the Netherlands).

Screening for FLT3-ITD by fragment length analysis

FLT3 internal tandem duplications (ITD) were determined with fragment length analysis as previously described.[1]

Next-Generation Sequencing (TruSight Myeloid Sequencing Panel)

The TruSight Myeloid Sequencing Panel (Illumina, San Diego, CA) was used to generate NGS libraries from 50 ng of DNA or and a fixed volume of 8.5µL of DNA derived from archived MGG-stained BM slide and prepared according to the TruSight Myeloid Sequencing Reference Guide (Illumina, San Diego, CA, USA). The generated libraries were pooled and paired-end sequenced (2×221-bp) on the Illumina HiSeq 2500 in Rapid Mode according to manufacturer's recommendation (Illumina, San Diego, CA, USA). Bioinformatics analysis was performed as previously described.[2]

DNA sequencing data, alignment, and variant detection

We used our in-house data analysis pipeline for variant calling as previously described.[2] Moreover, a site-specific error model that models the unique site-and-variant specific noise profile from a large set of complete remission samples (n=480) from patients who did not carry the variant of interest was used for reliable variant detection at low levels.[2] Variants with a p-value < 0.001 was considered detected.

Next-Generation Sequencing (Custom Hotspot Panel)

Primer design

Primers for hotspot mutation NGS analysis were designed by using the Ion AmpliSeq Designer software (ThermoFisher Scientific, Bleiswijk).

Target specific primers were adapted for Illumina-based sequencing by adding Illumina forward (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3') and reverse (5'-TCGCGAGTTAATGCAACGATCGTCGAAATTCGC-3') overhang adaptor sequences to the respective primers. The second PCR, by means of these overhang adaptor sequences, attaches sample-specific dual indices for sample identification and Illumina sequencing adaptors using primers from the Illumina TruSeq Custom Amplicon Index Kit (Illumina, San Diego, CA, USA).

Multiplex PCR-based custom NGS library preparation

The multiplex PCR assay to be analyzed by NGS was processed as follows: First, the hotspot gene targets were amplified by multiplex PCR using the Roche FastStart High fidelity PCR System (Roche) containing 1× Buffer with 1.8mM $MgCl_{2,}$ 0.2mM dNTP, 0.4µM each primer, 0.1U FastStart Taq DNA polymerase and 50ng and 4µL DNA input of FPBM and MGG-stained BM, respectively. Amplification was performed using the following PCR condition: 95°C for 5 minutes, 25 cycles of 30 second at 95°C, 30 seconds at 60°C, and 30 seconds at 72°C and a final extension for 7 minutes at 72°C.

Amplicons from the first step PCR were purified using the Agencourt AMPure XP bead purification kit (Beckman Coulter, Fullerton, CA, USA). The second step PCR was performed with primers from the Illumina TruSeq Custom Amplicon Index Kit ((Illumina, San Diego, CA)

using the KAPA HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, MA, USA) using the following thermocycling condition: 95°C for 5 minutes, 10 cycles of 20 second at 98°C, 30 seconds at 66°C, and 30 seconds at 72°C and a final extension for 1 minutes at 72°C. The library pool was purified with Agencourt AMPure XP beads and normalized for Illumina-based sequencing, according to the manufacturer's protocol (Illumina, San Diego, CA).

The NGS libraries were paired-end sequenced (2×151-bp) on the MiSeq according to manufacturer's recommendation (Illumina, San Diego, CA). Bioinformatics analysis was performed as previously described.[2]

Custom Hotspot Panel Primer Sequences

Target	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
ASXL1_G643-G645	**CAGACATTAAAGCCCGTGCTC**	**TCTGCCACCTCCCTCATCGG**
BCORL1_rs4830173	**AAACTTGGGCACAAGTCAGAAGA**	**GTCAGATGCGGGAAGCTTGA**
BRAF_V600	**TCAGTGGAAAAATAGCCTCAATTCTTACC**	**CTTCATGAAGACCTCACAGTAAAAATAGGT**
CSF3R_T618	**TACCCTCCAAACAGCCATCTCT**	**CAGTCTGTATCACATCCACCTCAT**
DNMT3A_R882	**TCCCTTACACACACGCAAAATACT**	**TGGTTTCCCAGTCCACTATACTGA**
ETNK1_H243-N244	**TTTTATGCGGTTTTGTTTTAAACAGGCTA**	**TCCTGTGGGAATGAGAGAGAAATACTT**
FLT3_D835	**ATAACGACACAACACAAAATAGCCGT**	**CCACGGGAAAGTGGTGAAGATATG**
IDH1_R132	**AGAATAAAACACATACAAGTTGGAAATTTCTGG**	**CTTGTGAGTGGATGGGTAAAACCTAT**
IDH2_R140	**GACTAGGCGTGGGATGTTTTTG**	**GTGGGACCACTATTATCTCTGTCCT**
IDH2_R172	**GGCCTTGTACTGCAGAGACAA**	**CTGGACCAAGCCCATCACCAT**
JAK2_V617	**TTCCTTAGTCTTTCTTTGAAGCAGCA**	**AGATGCTCTGAGAAAGGCATTAGAAAG**
KIT_D816	**TATTCACAGAGACTTGGCAGCCAGAA**	**GCAGAGAATGGGTACTCACGTTTC**
KIT_L416_R420	**AGCACTCTGACATATGGCCATTT**	**CCTGGACAAAAATACCAATCTATTGTGG**
KRAS_G12-G13	**TACCTCTATTGTTGGATCATATTCGTCCA**	**TATTATAAGGCCTGCTGAAAATGACTGAAT**
KRAS_Q61	**CATGTACTGGTCCCTCATTGCA**	**GTAATAATCCAGACTGTGTTTCTCCCTT**
MPL_W515	**GCCGAAGTCTGACCCTTTTTGT**	**GGTACCTGTAGTGTGCAGGAAA**
MYD88_L265	**CTTGCAGGTGCCCATCAGA**	**GGCGAGTCCAGAACCAAGATTT**
NRAS_G12-G13	**GTGGGATCATATTCATCTACAAAGTGGT**	**GATTACTGGTTTCCAACAGGTTCTTG**
NRAS_Q61	**GAGGTTAATATCCGCAAATGACTTGC**	**AAACCTGTTTGTTGGACATACTGGA**
SETBP1_H860-K884	**CTCCCTAAAGGAAATCACGCTGT**	**AGTACCTCCTTCGGGATTCTGA**
SF3B1_K700	**ACCCTTCCATAAAGGCTTTAACACA**	**TGTTTGGTTTTGTAGGTCTTGTGGA**
U2AF1_Q157-R156	**GAACTGTGCTCAGTCACGTCA**	**TGACTTGAATAACCGTTGGTTTAATGGA**
U2AF1_S34	**AAAAAGGCAAACAAACCTGGCTA**	CCCAGCAAAATAATCAGCTCTCATTTTC

SUPPLEMENTARY RESULTS

DNA isolation from archived MGG-stained and unstained BM slides and FPBM

We initially tested the DNA isolation protocol on matched MGG-stained and unstained BM morphology slides, which were both archived over the years. We isolated genomic DNA from the MGG-stained BM slides (n=16) and unstained BM slides (n=16). However, the success rate of obtaining DNA from the MGG-stained BM slides was higher compared to the unstained BM slides (16 out of 16 (100%) versus 11 out of 16 (69%)). Furthermore, the DNA yield from the unstained slides was much lower as compared to the MGG-stained slides (range: 0 – 85.6 ng/µl; mean: 40 ng/µl versus range: 2.94 – 340 ng/µl; mean: 127.6 ng/µl) (Supplementary Table S2). The reason for this difference in DNA yields might be due to the fact that the DNA may be better conserved on the MGG-stained morphology slides following MGG staining and the coverslip protection of the BM smears.

We subsequently isolated DNA from 191 MGG-stained BM slides and 41 FPBM samples in total. The mean concentration of DNA derived from MGG-stained BM and DNA extracted from FPBM [with a mean cell count of 35.7 million cells/ml (range: 18.9 – 80 million cells)] was 1µg (range: 40.8ng – 14.3µg) and 14.5µg (range: 3.9µg – 37.4µg), respectively. The DNA concentrations of all individual DNA samples derived from MGG-stained BM slides are depicted in Figure S1. The indicated DNA concentration of the MGG-stained BM DNA is after concentrating the DNA using Amicon Ultra Centrifugal Filters. The average size of the DNA fragments of DNA derived from MGG-stained BM slides is approximately 8-10 Kb (Supplementary Figure S2). The OD260/280 values of 11 representative DNAs derived from MGG-stained BM slides are depicted in Supplementary Table S3. We did not experience any differences in success rate regarding DNA isolation, yield or NGS between older or younger MGG-stained BM slides. Overall the *FLT3* ITD ratios of all FPBM and MGG-stained slides of the *FLT3* ITD-mutant cases were comparable between FPBM and MGG-stained slides (Supplementary Table S4).

DNA isolation from very old archived MGG-stained BM slides

We collected 2 slides (both prepared in 1975) from our MGG-stained slide collections. The coverslips of these very old MGG-stained BM slides did not loosen in xylene. Most probably these coverslips were attached using other chemicals. We have tried some other types of dissolvent to loosen the coverslips and were successful with toluene. We next performed NGS-based panel sequencing on the DNA derived from the 44 years-old MGG-stained BM slides and were able to detect known driver mutations [slide 1: NPM1 (VAF 0.35) and slide 2: NPM1 W288fs (VAF 0.61), IDH1 R132H (VAF 0.05), IDH2 R140Q (VAF 0.38) and KRAS G12S (VAF 0.08). Unfortunately, there were no FPBM samples available for comparison.

SUPPLEMENTARY FIGURES AND TABLES

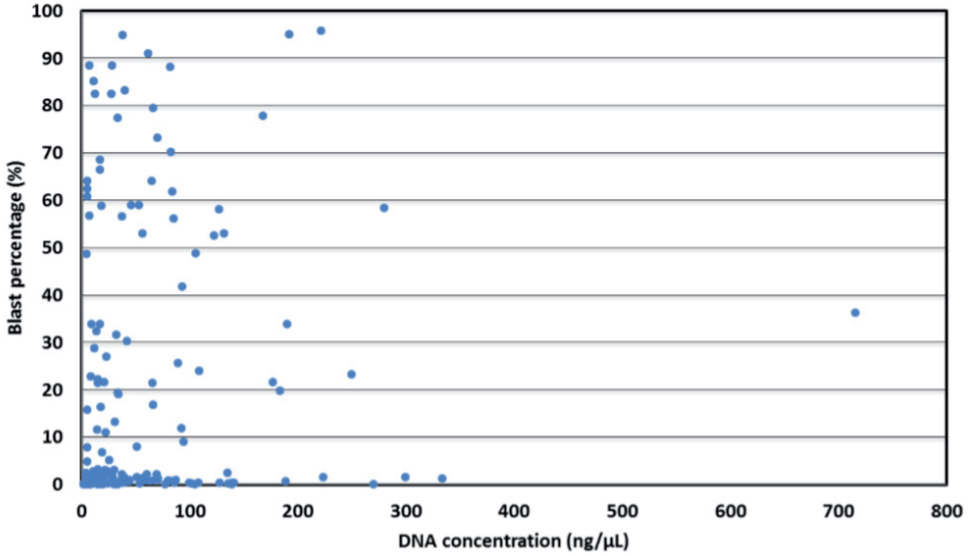


Figure S1: Correlation between DNA concentration (ng/µL) and blast percentage in BM at diagnosis (%).

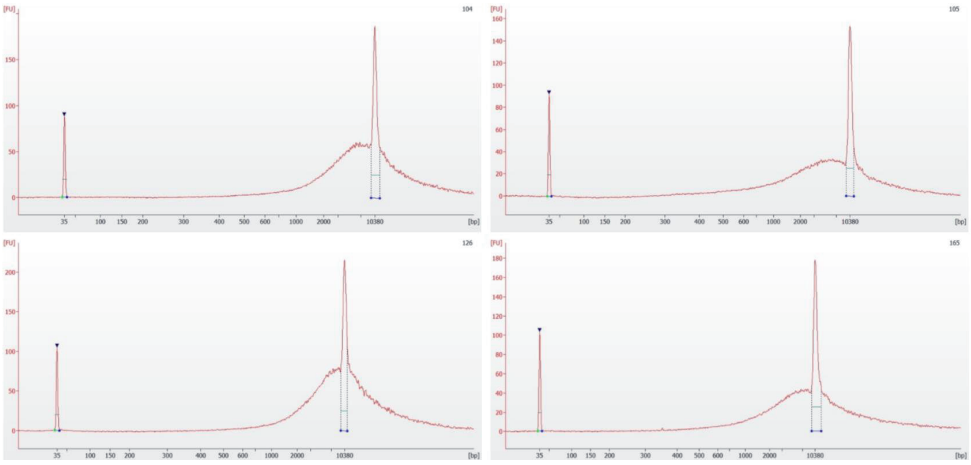


Figure S2: Four representative BioAnalyzer Agilent High Sensitivity DNA Assay measurements of DNA derived from MGG-stained BM slides (104: 16.8 ng/µL and OD260/280 2.08; 105: 19.6 ng/µL and OD260/280 1.78; 126: 14.4 ng/µL and OD260/280 1.92; 165: 42.9 ng/µL and OD260/280 1.90).

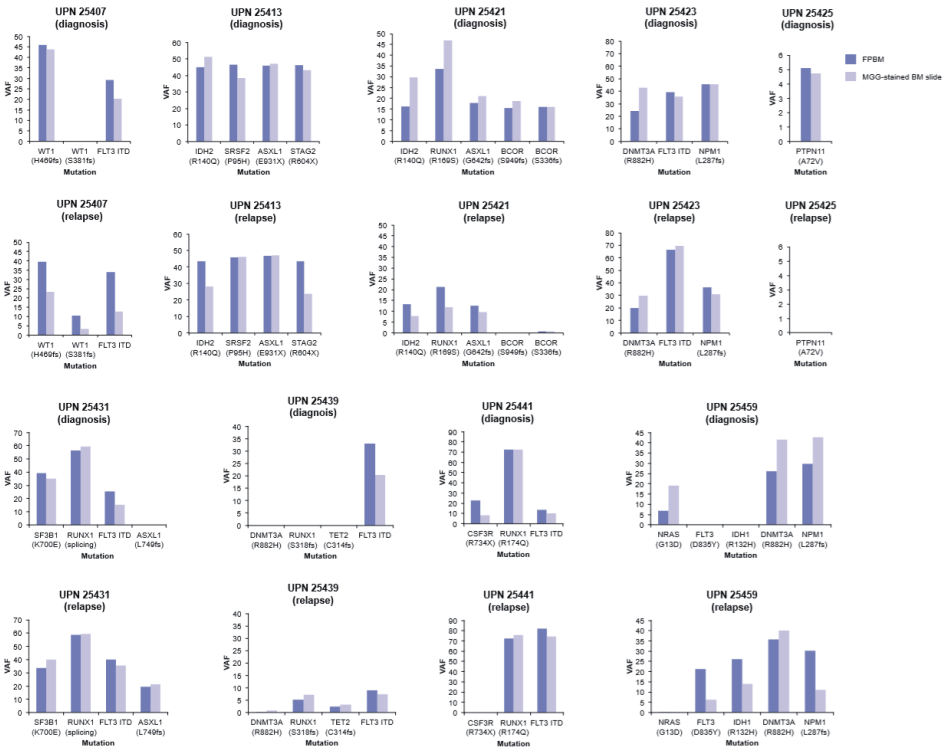
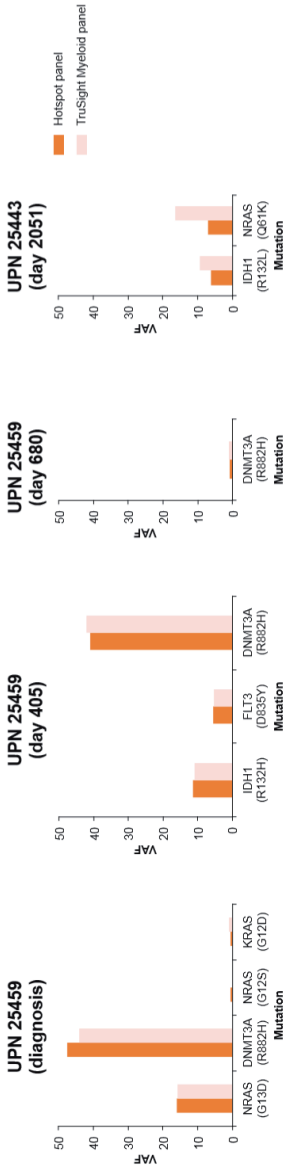
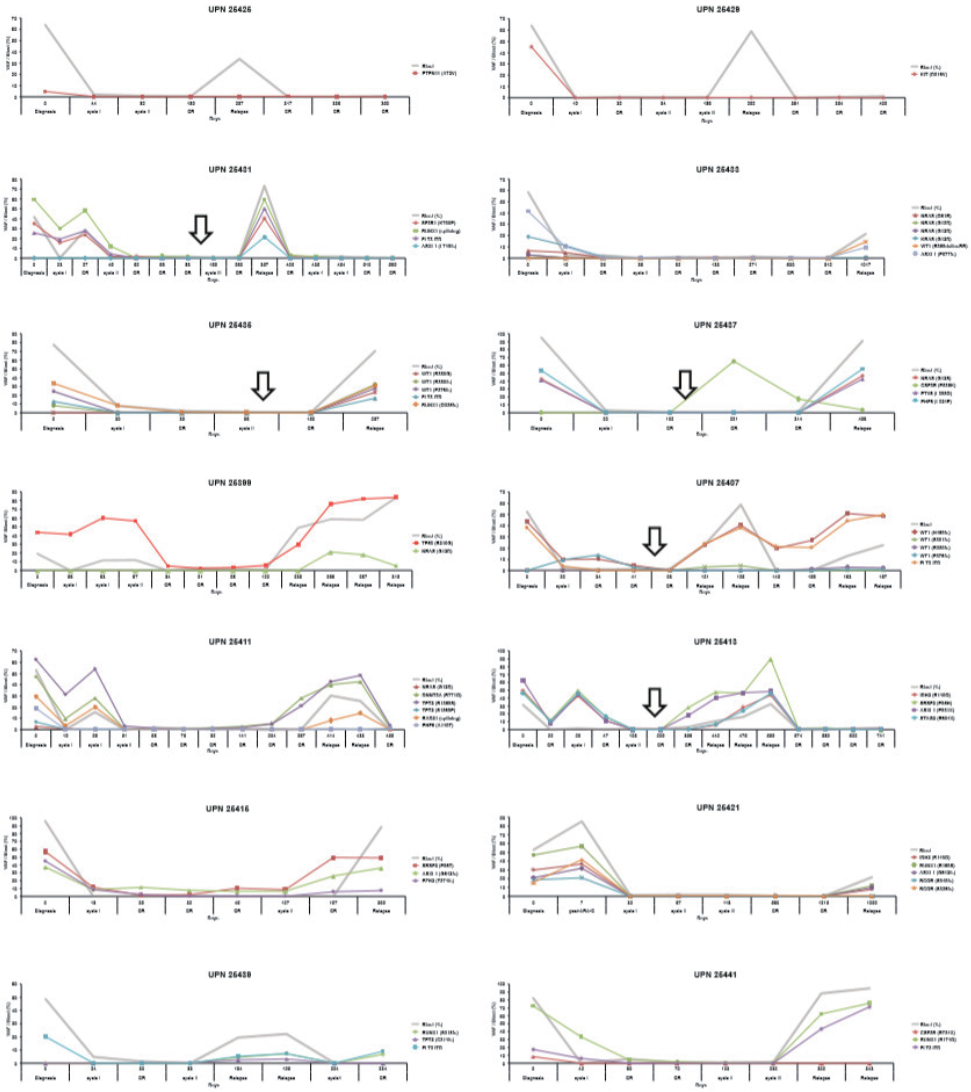


Figure S3: Comparison of variant allele frequencies (VAF) of mutations detected in DNA derived from Ficoll-purified BM and MGG-stained BM slide. The bar charts illustrating the VAFs of mutations detected at diagnosis and relapsein four acute myeloid leukemia patients.



Supplementary Figure S4: MGG-stained BM slide mutation comparison between laboratory developed hotspot panel and TruSight Myeloid Panel. Comparison of variant allele frequencies (VAF) of mutations detected inDNA derived from MGG-stained BM slide using hotspot panel and TruSight Myeloid Panel in DNA derived from MGG-stained BM slide. The bar charts illustrate the VAFs of mutations detected using two NGS panels at different sampling day.



Supplementary Figure S5: Longitudinal mutation detection by targeted NGS on DNA derived from MGG-stained BM slides. Sequential mutation analysis from diagnosis until the last available evaluation sampling time point (in days). Each line represents the VAF of an acquired mutation in the indicated gene. The thick transparent gray line depicts the BM blast count. The black arrow indicates the time of stem cell transplant.

Supplementary Table S1: Patients clinical characteristics

Supplementary Table S1: Patient Clinical Characteristics

Nr.	UPN	FAB at diagnosis	Sex	Age at diagnosis	Date of diagnosis	Cytogenetics at diagnosis	Mutation status known at diagnosis	WBC count at diagnosis	Blast percent at diagnosis (clinically determined)	Maintenance treatment post diagnosis	FAB at relapse	Date of relapse diagnosis	WBC count at relapse	Blast percent at relapse	Cytogenetics at relapse	Mutation status at relapse
1	25399	RAEB-t	M	65	11/26/2007	56~58,XY,+X,der(5)del(5)(p1?2p1?4)del(5)(q1?3q15),+8,+10,+11,+11,+13,+19,+21,+22,+mar1,+mar2[8]/56~58,idem,add(2)(q3?6)[16]/46,XY[30]	TP53	2.4	19	no SCT		8/18/2008	14.1	59	53-58,XY,+X[5], +1[5], add(2)(q3?6)[6],der(5)del(5)(p1?2p1?4)del(5)(q1?3q15)[3],+8, +10[5],+11,+11[6],12?[4],+13-17[4],+19[5],+21[4],+22[6],1-6mar,inc[cp7]	TP53, NRAS
2	25407	M2	M	32	6/19/2007	46,XY [21]	WT1 (2), FLT3-ITD	42	53	AlloHLA	M2	10/22/2007	8.4	59	46,XY,del(3)(q?13.3q2?4) [7]/46,XY [12]	WT1 (2), FLT3-ITD
3	25411	M2	M	37	1/6/1999	46, XY, del (11) (q14 q25) [3]/46, XY [60]	DNMT3A,PHF6,TET2 (2),CEBPA (double mutant), NRAS,RAD21	35	53	Auto	M2	3/14/2000	3	26	46,XY	DNMT3A,TET2,CEBPA (double mutant), RAD21
4	25413		M	63	3/12/2009	47,XY,+8[1]/46,XY[34]	IDH2,ASXL1,SRSF2,STAG2	4.9	32	Auto		9/22/2010	0.7	32	46,XY	IDH2,ASXL1,SRSF2,STAG2
5	25415		M	71	5/17/2010	49,XY,+13,+13,+19 [21]	SRSF2,EZH2		96					88	49,XY,+13,+13,+19 [2], 49,idem,-7,+mar1 [20]	SRSF2
6	25417	RAEB-t	M	72	2/24/2003	46,XY,t(3;21)(q26;q22),del(12)(p12p13)[20]	TET2,KIT,SRSF2,DNMT3A	54	20	no SCT	M0	8/4/2003	3.1	27	46,XY,t(3;21)(q26q22),del(12)(p12p13)	TET2,KIT,SRSF2,DNMT3A,CALR
7	25421	M0	M	62	7/18/2000	46, XY [68]	RUNX1,IDH2,ASXL1,BCOR (2)	1.1	53	no SCT	M6	1/16/2004	1.6	22	46,XY	RUNX1, IDH2, ASXL1, BCOR
8	25423	M5b	F	30	9/1/2004	46,XX[21]	NPM1,FLT3-ITD,DNMT3A	174	78	no SCT		6/8/2005	6.2	57	46,XX	NPM1,FLT3-ITD,DNMT3A
9	25425	M4eo	M	45	3/8/2006	46,XY,add(7)(q22)[4]/46,XY[21]	INV(16), PTPN11	9.7	64	AlloHLA	M4eo	12/20/2006	2.4	34	46,XY,add(7)(q22)	INV(16)
10	25427	M4	M	43	10/17/1998	47, XY, +11 [14]/ 46, XY [13]	DNMT3A,IDH2,SRSF2, FLT3	20	36	AlloMUD	M4	9/9/1999	12.2	88	47,XY,+11	DNMT3A,IDH2,SRSF2, FLT3 (2)
11	25429	M4eo	M	41	9/15/2006	46,XY,inv(16)(p13q22)[21]	INV(16),KIT	114	64	AlloHLA	M4eo	7/4/2007	5.7	59	46,XY,inv(16)(p13q22)	INV(16)
12	25431	M4	M	37	11/20/2001	47,XY,+13[11]/46,XY[28]	RUNX1,SF3B1,FLT3-ITD	23	42	Auto	M4	11/12/2002	38	73	47,XY,+13	RUNX1,SF3B1,ASXL1,FLT3-ITD
13	25433	M2	F	43	6/25/2008	46,XX,del(11)(q11q14),inv(16)(p13q22)[16]	INV(16),ASXL1, NRAS	20	58	no SCT	M2	4/20/2011	1.5	39	46,XX,del(11)(q11q14),inv(16)(p13q22)	INV(16),ASXL1
14	25435	M5A	F	49	3/3/2008	46,XX[20]	FLT3-ITD,RUNX1,WT1(2)	8.2	77	Auto	M5A	12/16/2008	2.1	70	46,XX,t(3;14)(p1?2;q32)	FLT3-ITD,RUNX1,WT1(3)
15	25437	M1	M	29	9/25/2003	47,XY,del(5)(q2?1),add(15)(q2?5),+19 [14] /46,XY [11]	NRAS,ETV6,PHF6	218	95	AlloMUD		2/2/2005	10.6	85	47,XY,del(5)(q2?1),add(15)(q2?5),+19	NRAS,ETV6,PHF6, CSF3R
16	25439	M4	M	42	12/27/2005	46,XY[20]	FLT3-ITD	138	49	AlloHLA	M4	7/10/2006	3.1	22	46,XY	FLT3-ITD, RUNX1, TET2
17	25441	M1	M	53	2/12/2007	46,XY[23]	RUNX1,FLT3-ITD, CSF3R	3.4	83	no SCT	M1	1/22/2008	147	95	ND	RUNX1,FLT3-ITD
18	25459		F	62	10/12/2010	46,XX [20]	NPM1,NRAS,DNMT3A	0	22	Allo RIC MUD		11/29/2011	2.3	17	46,XX,t(2;17;6)(q2?4;q1?2;q2?3)?del(2)(q24q31)	NPM1,FLT3-TKD,IDH1, DNMT3A

Supplementary Table S2: Comparison of DNA concentration between DNA derived from MGG-stained and unstained BM slides

Nr.	Sample ID	Sampling date	Blast (%)	MGG-stained BM slide gDNA concentration (ng/µL)	Unstained BM slide gDNA concentration (ng/µL)
1	1	9/16/1998	84	340	85.6
2	2	10/5/1998	51	3.98	too low
3	3	10/12/1998	32	14.5	8.4
4	4	11/2/1998	3	2.94	too low
5	5	11/9/1998	0.8	136	64.4
6	6	11/18/1998	0.4	115	77.8
7	7	12/1/1998	0.8	79.6	30.6
8	8	1/18/1999	0.6	9.46	16.4
9	9	3/3/1999	1.2	81.8	19.9
10	10	5/26/1999	1.8	258	too low
11	11	9/1/1999	0.4	108	too low
12	12	1/24/2000	1.6	330	25.8
13	13	1/3/2001	16.2	185	75
14	14	1/15/2001	7.6	268	24.2
15	15	2/8/2001	0.8	45.2	too low
16	16	2/21/2001	2.4	63.4	12

Supplementary Table S3: OD260/280 of 11 representative MGG-stained BM samples

Nr.	Sample ID	NanoDrop [ng/µl]	A260/280
1	104	16.8	2.15
2	105	19.6	2.36
3	126	14.4	2.47
4	134	8	2.21
5	150	179.48	1.89
6	165	42.9	2.04
7	175	739.8	1.92
8	215	165.98	1.95
9	228	21.73	2.63
10	265	10.5	2.86
11	300	44.48	2.11

Supplementary Table S4: FLT3-ITD ratios FPBM and MGG-stained BM samples.

Sample ID	Disease time-point	Material	FLT3ITD size	Allelic ratio (ELN2017)
25407	Diagnosis	FPBM	72	0.90
		MGG-stained	72	0.62
	Relpase	FPBM	72	1.16
		MGG-stained	72	0.32
25423	Diagnosis	FPBM	24	0.85
		MGG-stained	24	0.77
	Relpase	FPBM	24	2.73
		MGG-stained	24	3.85
25431	Diagnosis	FPBM	48	0.34
		MGG-stained	48	0.18
	Relpase	FPBM	48	0.67
		MGG-stained	48	0.55
25435	Diagnosis	FPBM	60	0.34
		MGG-stained	60	0.14
	Relpase	FPBM	60	0.30
		MGG-stained	60	0.20
25439	Diagnosis	FPBM	141	0.50
		MGG-stained	141	0.25
	Relpase	FPBM	141	0.09
		MGG-stained	141	0.05
25441	Diagnosis	FPBM	24	0.21
		MGG-stained	24	0.21
	Relpase	FPBM	24	6.13
		MGG-stained	24	0.76

Supplementary Table S5: List of genes in Custom Hotspot Panel

Gene	Amino Acid
ASXL1	G643-G645
BRAF	V600
CSF3R	T618
DNMT3A	R882
ETNK1	H243-N244
FLT3 TKD	D835
IDH1	R 132
IDH2	R140 and R172
JAK2	V617
KIT	L416-R420, D816
KRAS	G12, G13, Q61
MPL	W515
MYD88	L265
NRAS	G12, G13, Q61
SETBP1	H860-K884
SF3B1	K700
U2AF1	S34, Q157-R156

REFERENCES

1. Versluis J, In 't Hout FE, Devillier R, van Putten WL, Manz MG, Vekemans MC*, et al.* Comparative value of post-remission treatment in cytogenetically normal AML subclassified by NPM1 and FLT3-ITD allelic ratio. *Leukemia* 2017 Jan; **31**(1)**:** 26-33.
2. Jongen-Lavrencic M, Grob T, Hanekamp D, Kavelaars FG, Al Hinai A, Zeilemaker A*, et al.* Molecular Minimal Residual Disease in Acute Myeloid Leukemia. *N Engl J Med* 2018 Mar 29; **378**(13)**:** 1189-1199.

Chapter 6

The Landscape of *KMT2A*-PTD AML: Concurrent Mutations, Gene Expression Signatures, and Clinical Outcome

Adil S.A. Al Hinai[1,2], Marta Pratcorona[3], Tim Grob2, François G. Kavelaars[2], Elena Bussaglia[3], Mathijs A. Sanders[2], Josep Nomdedeu[3], Peter J.M. Valk[2]

[1] National Genetic Center, Ministry of Health, Muscat, Sultanate of Oman
[2] Erasmus University Medical Center, Rotterdam, the Netherlands
[3] Hospital de la Santa Creu i Sant Pau, Barcelona, Spain.

Hemasphere. 2019 Mar 5;3(2):e181

TO THE EDITOR:

Acute myeloid leukemia (AML) patients with partial tandem duplications (PTDs) in the Mixed Lineage Leukemia (*MLL*) officially known as the Lysine (K)-specific Methyltransferase 2A (*KMT2A*) gene, generally have adverse outcomes. Previous mouse studies have shown that Kmt2a-ptd is insufficient to cause AML, indicating additional mutations are required for leukemogenesis. Herein, we evaluated the mutational landscape, gene expression signatures and prognosis of *KMT2A*-PTD adult AML in comparison to a well-characterized adult AML cohort without *KMT2A*-PTD. Our study demonstrates that *KMT2A*-PTD AML has a distinct gene expression signature and that concomitant DNMT3A and *NRAS* mutations were associated with adverse clinical outcome in this subset of AML.

AML patients with *KMT2A*-PTD is characterized by an internal duplication spanning exon 3 to 9, exon 3 to 10, or exon 3 to 11 (Fig. S1A).[1] *KMT2A*-PTDs occur in 3.2 to 11% of adult de novo AML and are more frequently present in AML with normal cytogenetics and AML with trisomy of chromosome 11 as a sole cytogenetic aberration.[2] The presence of a KMT2A-PTD has been shown to associate with adverse outcome in AML.[2, 3]

The partial-tandem duplications within KMT2A result in in-frame additions of extra N-terminal amino acids and maintain functional proteins, which contribute to leukemogenesis.[4] Under normal physiological conditions, the *KMT2A* gene encodes a SET domain-containing protein, which mediates methylation of histone 3 lysine 4 (H3K4).[5] The effect of the PTD on normal KMT2A function and its role in leukemogenesis are currently unknown. 11q23-rearrangements involving *KMT2A* gene (3–4% of adult AML) result in abrogation of KMT2A transactivation and histone methyltransferase function.[6] *Kmt2a*-PTD alone appeared insufficient to cause AML.[7] These findings support the notion that additional genetic hits are required for the development of *KMT2A*-PTD leukemia.[7, 8]

We aimed to evaluate the mutational landscape, gene expression signatures and prognosis of *KMT2A*-PTD adult AML in comparison to a well-characterized adult AML cohort without *KMT2A*-PTD (hereafter referred as reference cohort) treated according to the international multicenter HOVON-SAKK AML clinical trials (www.hovon.nl).

cDNA from 1998 AML patients was screened for *KMT2A*-PTD mutations using RT-PCR and confirmed by Sanger sequencing (Fig. S1B and Supplementary method). *KMT2A*-PTDs were present in 5.5% (109 out of 1998) of all AML cases. The *KMT2A*-PTD was examined in the context of a number of clinical parameters (summarized in Table S1, and detailed in Table S4). The median age of AML patients with or without a *KMT2A*-PTD was 56 and 51 years, respectively (p=0.0016). The majority of *KMT2A*-PTD AML patients had a normal karyotype (57.3%). However, this did not significantly differ from AML patients without a *KMT2A*-PTD (48.8%) (p=0.158). In line with previous studies, the presence of *KMT2A*-PTDs was significantly associated with a concurrent trisomy 11 as compared to the *KMT2A* wild-type AML reference cohort (7.3% vs. 1.1%; respectively, p=0.002). Genomic DNA (gDNA)

was available for 85 out of the 109 *KMT2A*-PTD AML cases. We performed next-generation sequencing (NGS) on the 85 *KMT2A*-PTD AML cases to determine the concurrent driver mutations, with as reference the *KMT2A* wild-type AML cohort (n=561). All AML cases were sequenced using the Illumina TruSight Myeloid Sequencing Panel (Table S6) on the HiSeq 2500 in Rapid mode following the manufacturer's recommendations (Illumina, San Diego, CA). *FLT3* internal tandem duplications (ITD) were determined with fragment length analyses as previously described.[9] The number of mutations detected in the *KMT2A*-PTD AML cohort ranged from 0 to 6 mutations, with an average of 2.7 mutations per *KMT2A*-PTD AML patient, whereas, the AML reference cohort carried 0 to 9 mutations with an average of 3.0 mutations per AML patient. The average number of mutations was not significantly different between both cohorts (p=0.1).

The most frequently mutated genes in *KMT2A*-PTD AML were *DNMT3A* (38.8%), *FLT3*-ITD (34.9%), *IDH2* (21.2%), *RUNX1* (20%), *IDH1* (18.8%), *TET2* (16.5%), *WT1* (14.1%), and *SRSF2* (12.9%) (Fig. 1). In contrast to what has been reported, we found three cases with a concurrent mutation in the *NPM1* gene (Fig. 1).[10, 11]

We next examined whether the concurrently mutated genes were significantly associated with *KMT2A*-PTD AML compared to the *KMT2A* wild-type AML reference cohort. *KMT2A*-PTD concurrent mutations were present in proteins involved in epigenetic regulation, signaling, transcription, splicing, chromosome segregation, and tumor suppression (Fig. 1A). None of these mutational categories was significantly associated with *KMT2A*-PTD AML. However, a number of mutated genes were significantly more frequent in *KMT2A*-PTD AML compared to other types of AML, i.e., *FLT3*-ITD (34.9% vs. 23%; p=0.028) as well as mutations in *IDH1* (18.8% vs. 9.1%; p=0.012), *U2AF1* (9.4% vs. 3.2%; p=0.014) and *IDH2* (21.2% vs. 12.1%; p=0.027) [Fig. 1B and Fig. S2]. Similar associations with *IDH2* and *U2AF1* mutations were demonstrated by Papaemmanuil *et al.*[12]; however, *DNMT3A*, *RUNX1*, and *STAG2* mutations were not significantly associated with *KMT2A*-PTD AML in our cohort. Sun *et al.* demonstrated that *KMT2A*-PTD AML carried more frequently *FLT3*, *DNMT3A*, *RUNX1*, *IDH1* and *IDH2* mutations,[10] whereas Kao *et al.* showed this correlation for *FLT3*, *U2AF1*, *RUNX1*, *STAG2*, *PTPN11*, *WT1* and *EZH2* mutations.[11] However, in the latter 2 studies results were not compared to an internal *KMT2A* wild-type AML cohort, which could potentially result in misinterpretation as a result of selection biases regarding these positive associations. In our AML cohort, we did see trends for associations, for example, mutations in *DNMT3A*, *RUNX1*, *PTPN11*, and *WT1*, but these did not reach statistical significance. In contrast, mutations in *NPM1* (3.5% vs. 33.9%; p<0.001), *TP53* (3.5% vs. 10.3%; p=0.046) and *NRAS* (5.9% vs. 22.8%; p<0.001) were significantly less frequent in KMT2A-PTD AML in our cohort (Fig. 1 and Fig. S2). Mutual exclusivity between NPM1 mutations and KMT2APTDs have been shown before in AML.[10-12]

We next investigated which genes were differentially expressed between *KMT2A*-PTD AML and all other AMLs, in particular, AML with t(11q23), using our previously published gene expression profile (GEP) dataset (n=513 AML).[13] Interestingly, multiple homeobox-related gene family members were consistently overexpressed in *KMT2A*-PTD AML. The top-35 differentially expressed genes included HOX- and TALE-related genes, such as *HOXB5*, *HOXB6*, *HOXB7*, *HOXB8*, *HOXB9*, and *NKX2.3*, whereas *KMT2A* itself appeared to be the most consistently overexpressed gene (Table S2). Since mutations in *NPM1* are also associated with dysregulation of *HOX* gene expression, we next used an association model to see if those differentially expressed HOX genes were also dysregulated in *NPM1* mutant AML.[14, 15] In this association model, which takes a number of relevant clinically and genetically defined subsets of AML into account, *HOX*-related genes, such as *HOXA7*, *HOXA9*, and *HOXA5*, seemed to be more significantly differentially expressed in mutant *NPM1* AML (Table S2). In contrast, these specific gene expression changes were absent when a similar analysis was performed in AML with t(11q23) involving *KMT2A* (Table S2). Thus, *KMT2A*-PTD may induce overexpression of *HOX*-related genes in different ways than t(11q23)-related fusion proteins, suggesting that the *KMT2A*-PTD induces leukemogenesis by mechanisms distinct from t(11q23) abnormalities involving *KMT2A*.

The prognostic impact of *KMT2A*-PTD on overall survival (OS) and event-free survival (EFS) of this AML cohort appeared to be not significantly different to wild-type KMT2A AML (p= 0.44) (Figure S3). This is in contrast to what has been shown before but in line with more recent publications.[2, 3] In normal karyotype AML, this association was also absent (p=0.7, data not shown). We next addressed the question whether the concurrent mutations might carry prognostic value within *KMT2A*-PTD AML. Interestingly, in *KMT2A*-PTD AML, coexisting *DNMT3A* mutations were significantly associated with inferior overall survival (HR: 2.06; 95%CI: 1.19–3.58; p= 0.010) (Fig. 2A), as was suggested before.[11] Moreover, *KMT2A*-PTD AML patients that harbor *NRAS* mutations also have an inferior outcome (HR: 6.54; 95%CI: 2.45–17.49; p<0.001) (Fig. 2B). RAS-related mutations such as *FLT3*-TKD mutations recently were shown to confer an inferior outcome to patients with *KMT2A*-PTD AML. Survival analysis of *DNMT3A* and NRAS mutations in *KMT2A* wild-type AML patients revealed that mutations in *DNMT3A* were not associated with treatment outcome (p=0.99) (Fig. S4), whereas mutations in *NRAS* only showed a borderline association with positive outcome (p= 0.044) (Fig. S5). In multivariable analysis, including white blood cell count (WBC) and cytogenetics, *DNMT3A* and *NRAS* mutations remained significantly associated with adverse outcome (Table S3). Thus, although *KMT2A*-PTD did not associate with outcome in AML in general, specific mutational subtypes within *KMT2A*-PTD AML appear to carry poor prognostic value.

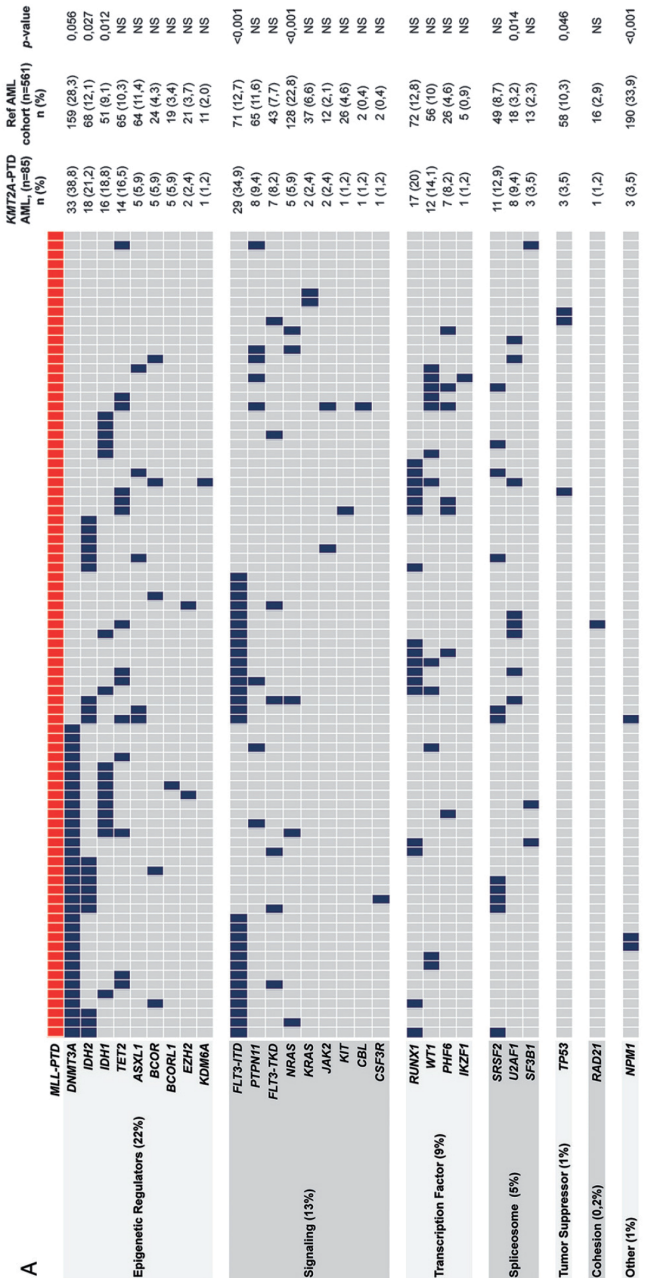
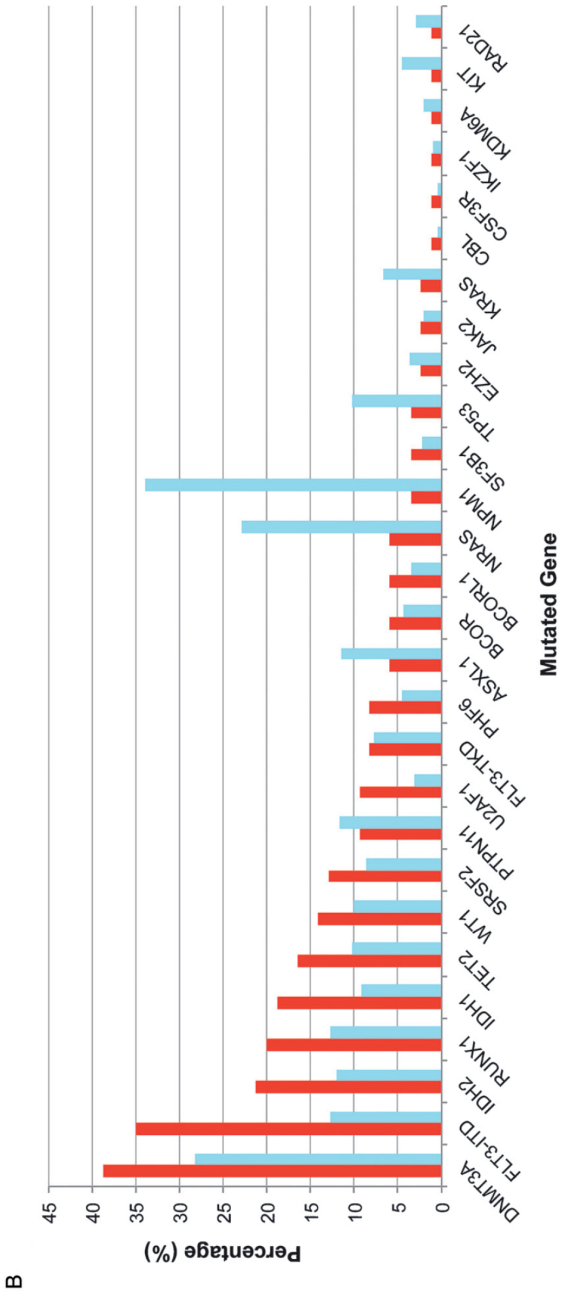


Figure 1. Mutational landscape of *KMT2A*-PTD AML.
KMT2A-PTD mutation landscape. Each column represents an individual AML patient. The gene mutations are categorized according to gene function and family. The number and the percentage of the gene mutations in *KMT2A*-PTD AML and the AML reference cohort and p-value are indicated. B) The frequency of concurrent mutations of *KMT2A*-PTD AML (red) and the reference cohort (blue).


B
Percentage (%)
0
5
10
15
20
25
30
35
40
45
Mutated Gene
DNMT3A
FLT3-ITD
IDH2
RUNX1
IDH1
TET2
WT1
SRSF2
PTPN11
U2AF1
FLT3-TKD
PHF6
ASXL1
BCOR
BCORL1
NRAS
NPM1
SF3B1
TP53
EZH2
JAK2
KRAS
CBL
CSF3R
IKZF1
KDM6A
KIT
RAD21

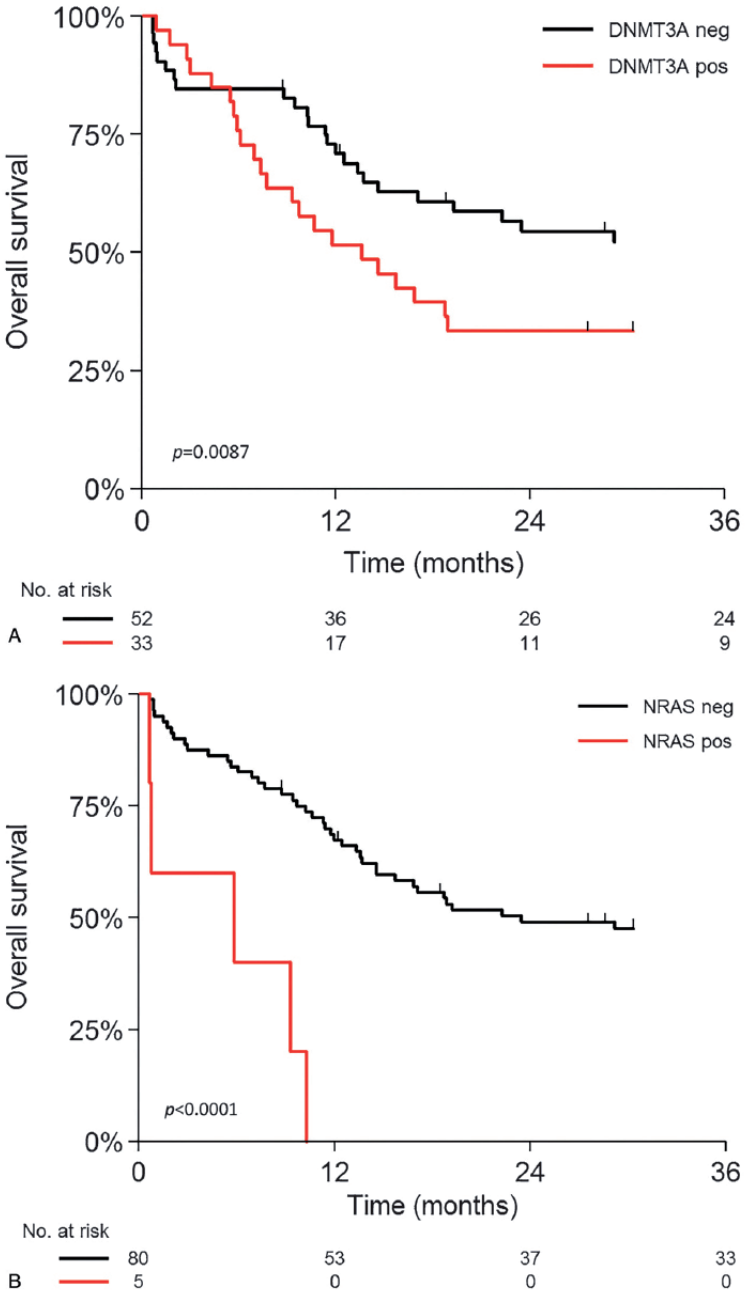


Figure 2. Overall survival analysis *KMT2A*-PTD AML and concurrent mutations.
KMT2A-PTD with and without *DNMT3A* mutations (P=0.0087). B) *KMT2A*-PTD with and without *NRAS* mutations (P<0.0001).

To validate our findings in *KMT2A*-PTD AML, we investigated an independent KMT2A-PTD AML cohort of patients included in the treatment protocols of the CETLAM cooperative group. The validation cohort contained 27 *KMT2A*-PTD AML patients with a median age of 56 years (Table S5), who were sequenced using Qiagen Human Myeloid Neoplasms GeneRead DNAseq Targeted Panel V2 (Table S6) (Qiagen, Hilden, Germany) on the MiSeq (Illumina, San Diego, CA) as per manufacturer instructions. The most frequently mutated genes in *KMT2A*-PTD AML validation cohort were similar to our 85 AML patients cohort (DNMT3A (48%), *FLT3*-ITD (37%), *IDH2* (22.2%), *WT1* (14.8%), *IDH1* (14.8%) and *RUNX1* (11.1%)) (Fig. S2). In the validation cohort, we confirmed that *KMT2A*-PTD AML patients with concurrent mutant *DNMT3A* have inferior outcome (13 out of 27; p=0.0017; Fig. S6). Unfortunately, the low numbers in the validation cohort precluded survival analyses of coexisting NRAS mutations, which should be confirmed in a larger *KMT2A*-PTD AML cohort.

In summary, we revealed within the molecular landscape of *KMT2A*-PTD AML, which carry specific *HOX* gene expression signatures, that concurrent *DNMT3A* mutations and *NRAS* mutations are associated with an adverse outcome.

Acknowledgments

We thank all the participating centers of the Dutch–Belgian Cooperative Trial Group for Hematology–Oncology (HOVON) and Swiss Group for Clinical Cancer Research (SAKK), where the clinical trials that formed the basis for this study were conducted; H. Berna Beverloo for performing cytogenetic analyses, Eric Bindels for performing next-generation sequencing, Remco Hoogenboezem for assisting with bioinformatics.

REFERENCES

1. Steudel C, Wermke M, Schaich M, Schakel U, Illmer T, Ehninger G*, et al.* Comparative analysis of MLL partial tandem duplication and FLT3 internal tandem duplication mutations in 956 adult patients with acute myeloid leukemia. *Genes Chromosomes Cancer* 2003 Jul; **37**(3)**:** 237-251.
2. Basecke J, Whelan JT, Griesinger F, Bertrand FE. The MLL partial tandem duplication in acute myeloid leukaemia. *Br J Haematol* 2006 Nov; **135**(4)**:** 438-449.
3. Shimada A, Taki T, Tabuchi K, Taketani T, Hanada R, Tawa A*, et al.* Tandem duplications of MLL and FLT3 are correlated with poor prognoses in pediatric acute myeloid leukemia: a study of the Japanese childhood AML Cooperative Study Group. *Pediatr Blood Cancer* 2008 Feb; **50**(2)**:** 264-269.
4. Whitman SP, Liu S, Vukosavljevic T, Rush LJ, Yu L, Liu C*, et al.* The MLL partial tandem duplication: evidence for recessive gain-of-function in acute myeloid leukemia identifies a novel patient subgroup for molecular-targeted therapy. *Blood* 2005 Jul 1; **106**(1)**:** 345-352.
5. Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD*, et al.* MLL targets SET domain methyltransferase activity to Hox gene promoters. *Mol Cell* 2002 Nov; **10**(5)**:** 1107-1117.
6. Visani G, Piccaluga PP, Martinelli G, Rossi M, Malagola M, Baccarani M. Sustained molecular remission in advanced acute promyelocytic leukemia with combined pulsed retinoic acid and arsenic trioxide. Clinical evidence of synergistic effect and real-time quantification of minimal residual disease. *Haematologica* 2003 Apr; **88**(4)**:** ELT15.
7. Zhang Y, Yan X, Sashida G, Zhao X, Rao Y, Goyama S*, 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 Aug 2; **120**(5)**:** 1118-1129.
8. Gilliland DG. Hematologic malignancies. *Curr Opin Hematol* 2001 Jul; **8**(4)**:** 189-191.
9. Versluis J, in 't Hout FEM, Devillier R, van Putten WLJ, Manz MG, Vekemans MC*, et al.* Comparative value of post-remission treatment in cytogenetically normal AML subclassified by NPM1 and FLT3-ITD allelic ratio. *Leukemia* 2017 2017/01/01; **31**(1)**:** 26-33.
10. Sun QY, Ding LW, Tan KT, Chien W, Mayakonda A, Lin DC*, et al.* Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD). *Leukemia* 2017 Jan; **31**(1)**:** 1-10.
11. Kao HW, Liang DC, Kuo MC, Wu JH, Dunn P, Wang PN*, et al.* High frequency of additional gene mutations in acute myeloid leukemia with MLL partial tandem duplication: DNMT3A mutation is associated with poor prognosis. *Oncotarget* 2015 Oct 20; **6**(32)**:** 33217-33225.
12. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND*, et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 2016 Jun 9; **374**(23)**:** 2209-2221.
13. Verhaak RG, Wouters BJ, Erpelinck CA, Abbas S, Beverloo HB, Lugthart S*, et al.* Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* 2009 Jan; **94**(1)**:** 131-134.
14. Mullighan CG, Kennedy A, Zhou X, Radtke I, Phillips LA, Shurtleff SA*, et al.* Pediatric acute myeloid leukemia with NPM1 mutations is characterized by a gene expression profile with dysregulated HOX gene expression distinct from MLL-rearranged leukemias. *Leukemia* 2007 Sep; **21**(9)**:** 2000-2009.
15. Alcalay M, Tiacci E, Bergomas R, Bigerna B, Venturini E, Minardi SP*, et al.* Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. *Blood* 2005 Aug 1; **106**(3)**:** 899-902.

SUPPLEMENTARY INFORMATION

MATERIALS AND METHOD:

Patient material, RNA isolation cDNA synthesis, and DNA isolation

Bone marrow aspirates or peripheral blood samples of cohorts of patients with AML were collected after a written informed consent in accordance with the Declaration of Helsinki. These procedures were performed as previously described [1]. DNA was isolated as per the standard molecular techniques.

Screening for *KMT2A*-PTD by conventional PCR and Sanger sequencing

The primer set used for the detection of partial tandem duplication (PTD) of the *KMT2A* gene are flanking exon 8 and exon 4 (Supplementary figure S1). The primer set used to detect the common type PTD of the *KMT2A* gene (ex9/ex3, ex10/ex3, and ex11/ex3) were adopted from Caligiuri *et al*. 1998 [2]. The sequence of the forward primer KMT2A-ex8: 5'-GGAAGTCAAGCAAGCAGGTC and reverse primer KMT2A-ex4: 5'-AGGAGAGAGTTTACCTGCTC-3' were used at 200nM for PCR in a total reaction volume of 50µL using 3µL cDNA, 1×PCR buffer, 1.5mM $MgCl_2$, 200nM dNTPs (Thermo Fisher, Walthan, MA) and 200nM of each primer. The thermocycling condition was 94°C for 3min, followed by 35 cycles of 94°C for 15sec, 63°C for 1min, 72°C for 1min, ending with a final extension for 10min at 72°C. The PCR product was then visualized on a 2% agarose gel. Serial dilution of the EOL-1 cell line was taking along each PCR as a positive control. The PCR products of the positive cases were purified using MultiScreen plates (Merck Millipore, Amsterdam, the Netherlands) and Sanger sequenced on the ABI PRISM 3130*xl* genetic analyzer (Thermo Fisher, Waltham, MA).

Next-Generation Sequencing

Genomic DNA of 50ng used for generating libraries using multiplexed oligonucleotide probes as per the protocol described by the manufacturer of TruSight Myeloid Sequencing Panel (Illumina, San Diego, CA). Subsequently, sample-specific indexes were ligated to each library. The libraries were then pooled, and paired-end sequenced (2× 221-bp) on the Illumina HiSeq 2500 in Rapid Mode according to manufacturer's recommendation (Illumina, San Diego, CA). Bioinformatics analysis was performed as previously described [3].

EVI1 Expression assay

EVI1 expression was assessed by RQ-PCR assay as previously described [4].

Gene Expression Profiling

Samples were analyzed using Affymetrix Human Genome U133Plus2.0 GeneChips (Affymetrix, Santa Clara, CA, USA) as previously described [5].

Statistical analysis

The Fisher's exact test and the Mann-Whitney U test were used for categorical and continuous variables, respectively. Survival time was calculated from the date of diagnosis until the date of death from any cause or censoring. Event-free survival was defined as the date of diagnosis to relapse or death from any cause or censoring, whichever came first. Survival estimates were graphically represented with the Kaplan-Meier method. The log-rank test was used to compare the survival distributions of the groups and the Cox proportional hazards model for multivariable analysis. P-values are two-sided and p-values <0.05 were considered statistically significant. Statistical analysis was performed in STATA statistical software, Release 15.1 (Stata, College Station, TX, USA).

SUPPLEMENTARY TABLES

Supplementary Table S1. Clinical, cytogenetic and molecular characteristics study cohort

	n=85	
Age		
Median	55	
Range	22-77	
Sex		
M	47	55%
F	38	45%
WBC at diagnosis (x10^9/l)		
<=100	81	95%
>100	4	5%
Platelet count at diagnosis (x10^9/l)		
Median	74	
Range	4-574	
FAB		
M0	9	11%
M1	21	25%
M2	27	32%
M4	6	7%
M5	9	11%
M6	3	3%
RAEB	7	8%
unknown	2	3%
Cytogenetics*		
CN-X-Y	50	61%
Aberrant	32	39%
trisomy 11	6	7%
Monosomal karyotype	2	2%
Consolidation therapy		
None	31	37%
Chemotherapy	8	9%
Autologous HSCT	8	9%
Allogeneic HSCT	38	45%

*karyotyping failed in 3 cases

Supplementary Table S2: Top 40 genes differentially expressed in (A) KMT2A-PTD versus wild type AML, (B) mutant NPM1 versus wild type AML and (C) 11q23-rearranged versus wild type AML (wt: wild type; FC: fold change in gene expression, red: KMT2A; green: HOX genes)

A

Probe set	Gene name	FC PTD vs wt AML	p-value
212078_s_at	KMT2A	6.100988902	2.33394E-14
212079_s_at	KMT2A	19.73953864	1.02495E-12
1553808_a_at	NKX2-3	1.022567441	6.88611E-11
205366_s_at	HOXB6	1.547327782	8.11932E-10
228904_at	HOXB3	0.917893534	1.60868E-09
205600_x_at	HOXB5	3.307453996	6.20886E-09
204778_x_at	HOXB7	2.643319605	3.10232E-08
1565436_s_at	KMT2A	3.47970953	2.27485E-07
205453_at	HOXB2	0.603641169	2.90262E-07
222862_s_at	AK5	5.286210652	8.16665E-07
231354_at	LOC780529	10.48315094	1.34598E-06
1570098_at	NA	3.100523414	1.55844E-06
205601_s_at	HOXB5	7.336914322	1.91627E-06
204779_s_at	HOXB7	11.48637668	3.08961E-06
232825_s_at	DSEL	46.69080825	3.30862E-06
229667_s_at	HOXB8	8.762352672	5.62887E-06
228527_s_at	SLC25A37	2.98898028	6.75693E-06
235753_at	HOXA7	-0.030254168	7.83851E-06
230896_at	BEND4	3.185653194	8.02864E-06
242426_at	NRG4	0.217367227	9.59947E-06
206578_at	NKX2-5	4.309394807	1.19302E-05
219308_s_at	AK5	2.414967749	1.31367E-05
226928_x_at	SLC25A37	2.765591507	1.86969E-05
201069_at	MMP2	0.064885427	2.31784E-05
241764_at	NA	1.751489388	2.39565E-05
236892_s_at	HOXB-AS3	0.121697452	3.03957E-05
229309_at	ADRB1	1.107685005	3.26071E-05
213844_at	HOXA5	0.621531944	3.44483E-05
206847_s_at	HOXA7	-0.069635652	3.54635E-05
231767_at	HOXB4	0.479556104	3.55637E-05
203911_at	RAP1GAP	2.982770355	4.21619E-05
231376_at	UPP2	55.77251546	4.92584E-05
204030_s_at	NA	0.3523649	5.40626E-05

B

Probe set	Gene name	FC NPM1 mutant vs wt AML	p-value
236892_s_at	HOXB-AS3	1.741690173	8.51082E-62
228904_at	HOXB3	1.618634337	1.76939E-45
235521_at	HOXA3	2.024427044	3.10526E-45
206847_s_at	HOXA7	0.779008038	6.00413E-42
213844_at	HOXA5	2.007972201	4.41751E-40
205366_s_at	HOXB6	1.954353483	3.72671E-36
235753_at	HOXA7	0.528136238	6.1178E-35
230743_at	HOXB-AS3	0.768834889	4.81135E-34
1559477_s_at	MEIS1	1.563442939	1.16102E-33
205453_at	HOXB2	1.1135191	7.12794E-33
205600_x_at	HOXB5	3.931908042	3.0609E-32
231767_at	HOXB4	1.386425631	6.97303E-32
1553808_a_at	NKX2-3	1.028058532	1.33673E-31
204082_at	PBX3	12.08113723	1.50273E-29
242172_at	NA	1.268095354	3.91119E-29
204069_at	MEIS1	0.928222702	7.33551E-29
208414_s_at	HOXB3	1.790543223	1.26787E-28
206289_at	HOXA4	2.198685019	1.35004E-27
227210_at	SFMBT2	-1.987223222	1.6518E-26
239791_at	HOXB-AS3	2.611836463	2.65466E-26
213147_at	HOXA10	0.885267448	2.22908E-25
213110_s_at	COL4A5	4.010142033	1.33022E-24
209905_at	NA	0.920253485	1.59738E-24
232979_at	MIR10A	6.321346838	1.65533E-24
209439_s_at	PHKA2	1.649850734	3.67841E-24
241464_s_at	NA	2.567473398	7.26378E-24
213150_at	HOXA10	1.418745474	8.39501E-24
201664_at	SMC4	3.304772268	2.11363E-23
209582_s_at	CD200	-43.54878255	4.24983E-23
209583_s_at	CD200	-125.2486339	9.97692E-23
200602_at	APP	-10.07742298	1.10584E-22
209543_s_at	CD34	-0.70726961	1.3334E-22
214651_s_at	NA	1.201273068	2.20951E-22

C

Probe set	Gene Name	FC 11q23 vs wt AML	p-value
242828_at	FIGN	9.732778006	5.20E-22
234074_at	NA	10.67890221	3.97E-21
1559265_at	SKIDA1	2.140504756	5.32E-21
1559266_s_at	SKIDA1	7.601250705	7.11E-20
239710_at	FIGN	7.216195867	2.99E-19
1569041_at	NA	5.039665697	5.06E-19
1552665_at	JMJD1C-AS1	11.06755497	7.51E-19
1561002_at	NA	4.057752603	3.17E-18
222367_at	NA	4.509213034	7.38E-18
203733_at	DEXI	3.360873999	1.57E-17
225752_at	NIPA1	8.450241635	5.24E-17
222956_at	FIGN	44.06358522	1.56E-15
1557480_a_at	PPP1R27	149.0877593	1.66E-14
212173_at	AK2	2.459632659	1.06E-13
204082_at	PBX3	13.63136753	1.09E-13
225905_s_at	ST3GAL3	5.302916839	1.18E-13
213908_at	NA	4.343178816	1.95E-13
1555702_a_at	ST3GAL3	10.76849761	2.11E-13
1562342_at	NA	6.332638541	2.40E-13
220666_at	NA	11.76308026	2.43E-13
1557261_at	NA	5.1764378	2.87E-13
1554867_a_at	PRR16	4.666250891	7.55E-13
217520_x_at	NA	2.769242463	7.96E-13
239503_at	CASC10	2.910166597	1.29E-12
228412_at	LOC643072	0.327627528	1.36E-12
216456_at	NA	9.136858157	2.46E-12
215811_at	NA	52.18046212	2.65E-12
220014_at	PRR16	1.922466027	3.29E-12
1560758_at	NA	82.90691424	3.68E-12
1569577_x_at	RP11-462G12.1	3.316025048	5.02E-12
1567101_at	NA	2.136336295	5.62E-12
219738_s_at	PCDH9	54.88666491	5.90E-12
220595_at	PDZRN4	3.500590793	6.22E-12

Supplementary Table S3. Multivariable analysis

	Overall survival		
	HR	95% CI	p-value
Age per year	1.02	0.99-1.05	0.171
WBC at diagnosis (>100 vs. ≤100)	3.15	1.07-9.28	0.038
Cytogenetics (Abnormal vs. Normal)	1.91	1.11-3.29	0.020
***DNMT3A* mutation** (Positive vs. Negative)	1.93	1.09-3.41	0.024
***NRAS* mutation** (Positive vs. Negative)	5.13	1.89-13.97	0.001

Supplementary Table S4: Clinical information of the KMT2A-PTD AML patients analyzed by targeted DNA NGS using the Illumina TruSight Myeloid Sequencing Panel (main cohort)

UPID	Gender	Age	WBC	Platelate	FAB Classification	Cytogenetic (Karyotype)	Mutated Gene	Mutation	VAF	Remarks
1	M	51	4,7	152	M2	46,XY	RUNX1	P350fs	0,16	weak KMT2A-PTD
							SF3B1	K700E	0,45	
							DNMT3A	R882H	0,42	
2	M	50	60	83	M5	46,XY[53]	TET2	C1271fs	0,48	
							DNMT3A	L547H	0,5	
							FLT3 ITD			
							FLT3 TKD	D835Y	0,28	
3	M	43	20	63	M4	47, XY, +11 [14]/ 46, XY [13]	IDH2	R140Q	0,51	
							DNMT3A	R635P	0,47	
							SRSF2	P95H	0,51	
4	F	33	23	34	M1	46, XX,del(7) (q22q35) [27]/46, XX [4]	IDH2	R172K	0,47	
							DNMT3A	S393fs	0,89	
							BCOR	L1612fs	0,41	
5	F	47	0,8	64	M1	46, XX, der(2) inv(2) (p2? 3p1? 5) del(2) (p11 p1? 5) [17]	TET2	S1284F	0,41	
							WT1	R380fs	0,41	
6	M	40	67,3	66	M0	46,XY [21]	EZH2	D730fs	0,49	
							FLT3 ITD			
							FLT3 TKD	D835Y	0,37	
7	F	25	3,8	129	M1	46,XX [30]	RUNX1	D93delinsYPD	0,34	
							WT1	R369fs	0,03	
							U2AF1	R83H	0,34	
							BCOR	P1587fs	0,38	
							KDM6A	V1112_ V1113delinsALGLGPX	0,29	
8	F	57	7,5	111	M1	46,XX,8?p,del(9)(q12q34) [11]/46,XX[24]	IDH1	R132C	0,43	

UPID	Gender	Age	WBC	Platelate	FAB Classification	Cytogenetic (Karyotype)	Mutated Gene	Mutation	VAF	Remarks
							DNMT3A	R882H	0,44	
9	M	50	11	66	RAEB-t	46,XY,der(8)(p?)[26]	DNMT3A	G266R	0,48	
10	F	47	11,8	160	M4	46,XX[32]	NPM1	L287fs	0,45	
							DNMT3A	R882H	0,48	
							FLT3 ITD			
11	M	60	33	110	M2	46,XY?del (5)(q14q15)[9]> 46, XY[15]	TET2	Y867H	0,42	
							U2AF1	S34F	0,52	
							RAD21	E301fs	0,16	
							FLT3 ITD			
12	F	58	0,9	157	M0	46, XX [20]	ASXL1	G642fs	0,14	
							WT1	S381X	0,08	
13	M	56	2,6	62	M6	46,XY[22]	IDH2	R140Q	0,36	
14	F	60	2,2	107	M0	47,XX,+8[16]/46,XX[17]	RUNX1	M347fs	0,12	
							IDH2	R140Q	0,24	
							DNMT3A	D768fs	0,24	
							SRSF2	P95R	0,02	
							FLT3 ITD			
15	M	50	10,1	113	M2	46,XY[20]	TP53	R174X	0,02	weak KMT2A-PTD
							RUNX1	G135C	0,16	
							TET2	Q947X	0,03	
16	F	55	12,4	77	M0	46,XX[24]	TET2	L1721W	0,48	weak KMT2A-PTD
							JAK2	V617F	0,88	
							ASXL1	G642fs	0,06	
							PTPN11	Q79R	0,02	
							PHF6	S120fs	0,32	
							CBL	C396R	0,02	

UPID	Gender	Age	WBC	Platelate	FAB Classification	Cytogenetic (Karyotype)	Mutated Gene	Mutation	VAF	Remarks
17	M	54	56,5	55	M1	47,X,inv(Y)(p11.2q11.2),+11[10]	IDH1	R132L	0,46	
							FLT3 TKD	D835Y	0,26	
18	M	57	2,5	139	M2	46,XY[20]	IDH1	R132C	0,42	
							DNMT3A	R882H	0,47	
19	F	56	2,1	173	M2	46,XX[23]	TET2	S1369fs	0,37	
							DNMT3A	R882H	0,43	
20	M	57	4,1	51	M2	47,XY,+4[18]/46,XY[4]	KIT	D812V	0,45	
							RUNX1	G367fs	0,43	
							TET2	Q947X	0,01	
							PHF6	T300fs	0,6	
21	M	46	67,6	53	M5	46,XY,ins(10;11)(p12;q23q13)[10]	*			weak KMT2A-PTD
22	M	55	183,9	49	M4	46,XY,t(6;11)(q27;q23)[12]	EVI1 overexpression			weak KMT2A-PTD
							NRAS	G12D	0,04	
							PTPN11	S502L; G503A	0,42	
23	F	40	1,3	160	M1	46,XX[20]	IDH2	R172K	0,34	
24	F	49	8,2	311	M5	46,XX[20]	RUNX1	Q335fs	0,37	weak KMT2A-PTD
							WT1	P376fs	0,18	
							FLT3 ITD		36	
25	M	57	32,1	53	M2	46,XY[20]	IDH2	R140Q	0,45	weak KMT2A-PTD
							DNMT3A	W698R	0,43	
							SRSF2	P95H	0,5	
							CSF3R	T618I	0,1	
26	F	48	46	182	M4	trisomy 11,inc[5]/40~46,XX,+11,inc[cp5]	TET2	Q947X	0,05	weak KMT2A-PTD
							IDH1	R132C	0,5	
							DNMT3A	R882H	0,5	

UPID	Gender	Age	WBC	Platelate	FAB Classification	Cytogenetic (Karyotype)	Mutated Gene	Mutation	VAF	Remarks
							NRAS	Q61H	0,05	
27	F	48	7,4	111	M1	47,XX,+8[4]/46,XX[7]	RUNX1	Q335X	0,52	weak KMT2A-PTD
							IDH2	R172K	0,44	
28	M	52	64	66	M5	47,XY,+11[20]	IDH2	R140Q	0,5	
							NRAS	Q61K	0,2	
							U2AF1	S34F	0,49	
							FLT3 TKD	D835V	0,06	
							FLT3 ITD			
29	M	71	7,4	152	M1	46,XY[23]	IDH1	R132G	0,35	
							DNMT3A	N553fs	0,71	
							FLT3 ITD			
30	F	77	74	102	M5	46,XX[36]	ASXL1	G642fs	0,07	
							NPM1	L287fs	0,37	
							TET2	E887X	0,49	
							IDH2	R140W	0,05	
							SRSF2	P95L	0,46	
							FLT3 ITD			
31	F	65	2,6	361	RAEB-t	45,XX, -7 [19]	IDH1	R132C	0,25	
							DNMT3A	R320X	0,46	
							EZH2	R288L	0,12	
32	M	70	26	65	M4	46,XY,del(12)(p12)[14]/46,idem,add(21)(p1?2)[5]/46,XY[2]	ASXL1	H630fs	0,37	
							RUNX1	S141P	0,33	
							SRSF2	P95L	0,49	
33	M	63	4,9	136	M2	47,XY,+8[1]/46,XY[34]	ASXL1	E931X	0,46	
							IDH2	R140Q	0,48	

UPID	Gender	Age	WBC	Platelate	FAB Classification	Cytogenetic (Karyotype)	Mutated Gene	Mutation	VAF	Remarks
							SRSF2	P95H	0,35	
34	M	47	68,8	40	M2	46,XY[20]	IDH1	R132H	0,46	
							U2AF1	R83H	0,48	
							FLT3 ITD			
35	M	36	5,9	17	RAEB-t	47,XY,+21[6]/46,XY,der(6) t(6;11)(p23;q13)[2]/46,XY[14]	U2AF1	S34Y	0,39	weak KMT2A-PTD
							FLT3 ITD			
36	M	56	1	61	M2	46,XX[20]	WT1	R370fs	0,05	
							SRSF2	P95H	0,42	
							PHF6	L269delinsLS	0,34	
37	F	52	1,7	89	M1	46,XX[20]	FLT3 ITD			
							BCOR	N1425S	0,11	
38	M	65	1,8	80	M2	46,XY[25]	IDH1	R132H	0,42	
							SRSF2	P96fs	0,02	
39	M	56	1,7	150	M1	46,XY[20]	RUNX1	P125fs	0,91	
							TET2	S460F	0,49	
							PHF6	R274X	0,89	
40	F	48	28,6	45	M5	46,XX[9]	RUNX1	S318fs	0,4	
							DNMT3A	R882H	0,42	
							FLT3 TKD	D835Y	0,43	
41	M	49	86,6	43	M2	N/A	DNMT3A	R882H	0,47	
							FLT3 ITD			
42	F	56	12,9	43	M1	46,XX[31]	IDH2	R140Q	0,48	
							DNMT3A	H821fs	0,38	
							FLT3 ITD			
43	F	49	1,7	70	M2	46,XX[20]	RUNX1	R201fs	0,16	
							PHF6	R129X	0,24	
							FLT3 ITD			

UPID	Gender	Age	WBC	Platelate	FAB Classification	Cytogenetic (Karyotype)	Mutated Gene	Mutation	VAF	Remarks
44	M	64	13,6	14	AML-not M3	46,XY,add(14)(p10)[19]/	U2AF1	S34Y	0,47	
45	M	54	2,1	121	M2	46,XY[20]	*			
46	M	64	15,1	161	M2	46,XY[20]	TET2	Y867H	0,51	
							DNMT3A	R882H	0,43	
							FLT3 ITD			
47	M	60	2,1	241	M2	N/A	*			weak KMT2A-PTD
48	F	50	1,6	134	RAEB-t	46,XX[20]	IDH2	R172K	0,51	
49	F	59	14,9	129	M0	45,XX,-7[9]/	EVI1 overexpression			
							IDH1	R132C	0,46	
							DNMT3A	P627fs	0,84	
							PTPN11	D61N	0,41	
50	M	47	2,7	71	M1	N/A	IDH1	R132C	0,48	
							WT1	H465N	0,08	
51	F	39	0	0	M5	46,XX,t(9;11)(p22;q23)[7]/	EVI1 overexpression			
							WT1	S381X	0,02	
							PTPN11	D61N	0,02	
							IKZF1	P32fs	0,14	
52	M	55	17,4	350	RAEB		TET2	1106_1108del	0,48	
						46,XY[20]	SF3B1	K666T	0,49	
							PTPN11	A72T	0,02	
53	M	55	5,8	255	M6	46,XY[20]	IDH1	R132H	0,48	
54	M	63	10,5	42	M6		ASXL1	G642fs	0,07	
							IDH2	R140Q	0,51	
						46,XY[20]	SRSF2	P95R	0,49	
							FLT3 ITD			
55	F	57	11,1	93	AML-not M3	46,XX,add(7)(?q32)[12]/	*			weak KMT2A-PTD
56	M	56	1,8	45	M1	46,XY[20]	DNMT3A	R635W	0,29	

UPID	Gender	Age	WBC	Platelate	FAB Classification	Cytogenetic (Karyotype)	Mutated Gene	Mutation	VAF	Remarks
57	F	39	2,2	63	M2	46,XX[28]	EVI1 overexpression			weak KMT2A-PTD
							KRAS	G12V	0,22	
58	M	44	1,7	190	RAEB	47,XY,add(7)(q22),+11[4]/	FLT3 ITD			
59	F	43	8,4	32	RAEB-t	46,XX[20]	RUNX1	A329fs	0,35	
							TET2	P1367A	0,38	
							PTPN11	D61V	0,24	
							FLT3 ITD			
60	M	40	31,9	13	M5	46,XY,t(6;11)(q27;q23)[17]/	EVI1 overexpression			weak KMT2A-PTD
							KRAS	G12D	0,03	
61	M	44	0,6	19	M2	46,XY[22]	IDH2	R140Q	0,41	
							DNMT3A	R676W	0,41	
							SRSF2	P95R	0,37	
62	M	60	34,6	26	M4	47,XY,+11[12]/	IDH2	R140Q	0,46	
							DNMT3A	R882H	0,44	
							SRSF2	P95H	0,49	
							FLT3 TKD	D835A	0,36	
63	F	60	72,5	152	M2	46,XX[40]	IDH2	R140Q	0,51	
							DNMT3A	R882H	0,51	
							NRAS	G12D	0,04	
							FLT3 ITD			
64	M	38	156,7	62	M2	46,XY[20]	DNMT3A	R882C	0,49	
							WT1	R370fs	0,49	
							FLT3 ITD			
65	F	53	11,9	29	M2	45,XX,del(5)(q13q33),-18[3]/	DNMT3A	W330X	0,13	
							WT1	R462Q	0,47	
							PTPN11			

UPID	Gender	Age	WBC	Platelate	FAB Classification	Cytogenetic (Karyotype)	Mutated Gene	Mutation	VAF	Remarks
66	M	61	1,8	132	M2	46,XY[22]	CEBPA (dm)			
							NRAS	G12C	0,02	
							PHF6	G226fs	0,28	
67	F	51	1,2	85	M1	46,XX[20]	FLT3 ITD			
68	F	63	2,5	68	M2	47,XX,+8[12]/	IDH2	R132C	0,44	
							DNMT3A	W698X	0,3	
							BCORL1	W702X	0,11	
69	F	59	3,6	93	M1	46,XX[21]	IDH2	R172K	0,45	
							DNMT3A	R882C	0,44	
70	M	58	20,6	53	AML-not M3	46,XY[20]	RUNX1	G 2 1 7 _ S218delinsAAETRX	0,21	
							TET2	W1198X	0,4	
							U2AF1	S34Y	0,43	
							FLT3 ITD			
71	F	59	4,7	306	M2	46,XX[20]	IDH1	R132C	0,42	
							SF3B1	K700E	0,4	
							DNMT3A	p.567_568del	0,45	
72	M	56	2,4	61	M2	46,XY[20]	PTPN11	A72T	0,05	
							U2AF1	S34F	0,41	
							BCOR	C1491fs	0,8	
73	F	62	1,4	160	M0	46,XX[7]	RUNX1	R169G	0,23	
74	F	22	73,3	22	M1	46,XX,t(6;11)(q27;q23)[9]/	*			
75	M	55	43,8	357	M2	46,XY[20]	NPM1	L287fs	0,37	weak KMT2A-PTD
							DNMT3A	R882H	0,47	
							FLT3 ITD			
76	M	54	27,2	7	M0	48,XY,+4,+10,-20,+?22,inc[cp18]	DNMT3A	R803S	0,43	weak KMT2A-PTD
							PHF6	C20fs	0,85	

UPID	Gender	Age	WBC	Platelate	FAB Classification	Cytogenetic (Karyotype)	Mutated Gene	Mutation	VAF	Remarks
77	F	61	174,7	70	M1	46,XX[15]	RUNX1	D451fs	0,49	
							IDH1	R132H	0,53	
							WT1	S381X	0,03	
							FLT3 ITD			
78	F	65	69,1	83	M5	46,XX[25]	DNMT3A	R366G	0,39	
							WT1	P376fs	0,26	
							FLT3 ITD			
79	F	27	63,3	54	M2	46,XX,t(1;11)(q43-44;q23)[20]/	RUNX1	R166L	0,44	
							DNMT3A	R882H	0,46	
							FLT3 ITD			
							BCOR	L1612fs	0,46	
80	M	60	3	121	M1	46,XY,add(16)(q1?1)[15]/	TP53	R234C	0,38	
81	M	55	81,8	4	M1	46,XY[20]	RUNX1	S318fs	0,46	
							FLT3 ITD			
82	F	45	0,4	21	M0	46,XX,+1,der(1;7)(q10;q10)[3]/	IDH2	R172K	0,48	weak KMT2A-PTD
							JAK2	V617F	0,56	
83	M	59	21,7	574	M1	46,XY,del(9)(q13q33)[18]	CEBPA (dm)			
							DNMT3A	R882H	0,42	
							FLT3 ITD			
84	F	51	52,9	48	M0	46,XX[20]	TP53	R136H	0,1	weak KMT2A-PTD
							FLT3 TKD	D835Y	0,2	
85	M	60	19	38	M1	46,XY[21]	IDH1	R132H	0,49	

NA: Not available
* No mutation detected

Supplementary Table S5: Clinical information of the KMT2A-PTD AML patients analyzed by targeted DNA NGS using the QiagenSequencing Panel (validation cohort)

UPN	Gender	Age	WBC	Platelate	Cytogenetic (Karyotype)	Mutated Gene	Amino Acid Change	VAF
40072	M	46	3,5	11	46,XX[20]	IDH2	R140Q	44,31
40311	F	61	119,2	109	NR	DNMT3A	R882H	45,59
						NPM1	W288CfsTer12	41,67
						FLT3-ITD		
41030	F	61	45,5	187	46,XX[20]	DNMT3A	R882H	38,09
						IDH2	R140Q	8,4
40711	F	37	15,1	51	46,XX,add(3)(p25),del(7)(p13)[20]	NRAS	G12D	40,29
52306	M	58	42,2	85	46,XY	NRAS	G12D	14,66
						U2AF1	S34F	43,61
						FLT3-ITD		
60323	M	59	15	54	46,XY[20]	TET2	P29R	48,82
61627	M	62	2	194	46,XY[20]	DNMT3A	P904L	42,03
						IDH1	R132C	26,27
61581	M	59	12,6	180	47,XY,+11,i(11)(q10)[18]/46,XY[2]	DNMT3A	P904L	
						IDH2	R140Q	43,58
61718	M	52	62,4	65	46,XY[20]	DNMT3A	R882S	46,73
						WT1	R380TfsTer5	77,19
						FLT3-ITD		
61784	M	66	48	53	47,XY,+11	IDH2	R140Q	47,27
						FLT3-ITD		
61786	M	45	68,6	87	46,XY[20]	DNMT3A	W893S	41,54
						IDH1	R132C	44,68
						NRAS	G13D	45,26
61945	M	42	6,3	40	46,XY	RUNX1	Q390Lfs*211	45,6
71000	M	56	1,26	100	46,XY,inv(9)(p11q13)	WT1	S381fs*5	37,6

UPN	Gender	Age	WBC	Platelate	Cytogenetic (Karyotype)	Mutated Gene	Amino Acid Change	VAF
72218	M	50	90,9	68	46,XY[4]	BCOR	R1400W	100
						RUNX1	V164EfsTer10	48,83
120243	M	54	117	63	46,XY[20]	DNMT3A	R882H	8,42
120285	F	49	3,6	87	46,XX[20]	EZH2	R690H	97,09
						IDH1	R132C	50,62
						KIT	D816V	40,07
120916	M	48	48,86	58	46,XY	DNMT3A	S714C	43,06
						SF1	G497V	46,33
						TET2	Y867H	54,13
						TET2	H1778R	46,69
						FLT3-ITD		
121867	F	21	4	200	46,XX[30]	U2AF1	R156H	9,09
130975	M	68	176	90	46,XY[20]	FLT3	L262F	47,94
						FLT3-ITD		
140761	M	55	1,4	47	46,XY[20]	SETBP1	D682RfsTer2	20,99
						WT1	E384IfsTer2	6,1
141304	M	51	1,58	63	46,XY[20]	DNMT3A	R882H	43,32
						IDH2	R140Q	45,54
						JAK3	T8M	46,54
141465	M	61	42,48	38	46,XY[20]	DNMT3A	Y735C	42,87
141968	F	66	136	309	47,XX +8 [6], 46,XX [14]	DNMT3A	P904L	48,7
						FLT3-ITD		
151315	F	65	2,5	114	46,XX[30]	BCOR	T1250LfsTer44	21,75
						CBL	V363_T364insS	5,7
						DNMT3A	D768Y	45,44
						FLT3	D839G	5,52
						IDH1	R132C	34,55
						KMT2A	D877PfsTer8	5,79

UPN	Gender	Age	WBC	Platelate	Cytogenetic (Karyotype)	Mutated Gene	Amino Acid Change	VAF
						RAD21	L451R	48,71
						RUNX1	T300SfsTer301	31,39
						SF3B1	R625C	31,69
						FLT3-ITD		
161029	M	64			46,XY[20]	DNMT3A	R882H	47,69
						WT1	A211PfsTer76	6,97
						FLT3-ITD		
161084	M	60	102	88	46,XY[20]	EZH2	**R690H**	**96,6**
						ZSRS2	**R186X**	**98,06**
161099	**F**	**49**	**2,4**	**103**	**46,XX[30]**	**IDH2**	**R172K**	**37,59**

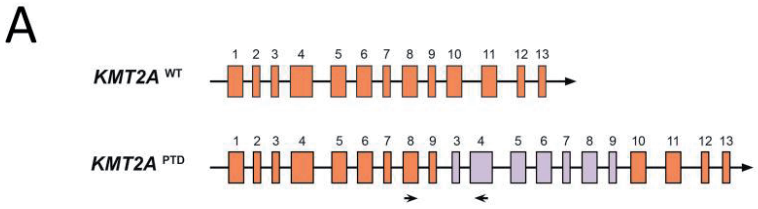
Supplementary Table S6: List of genes present in the kits used for targeted DNA sequencing of the main cohort and validation cohort; TruSight Myeloid Sequencing Panel, and Human Myeloid Neoplasms GeneRead DNAseq Targeted Panel V2, respectively

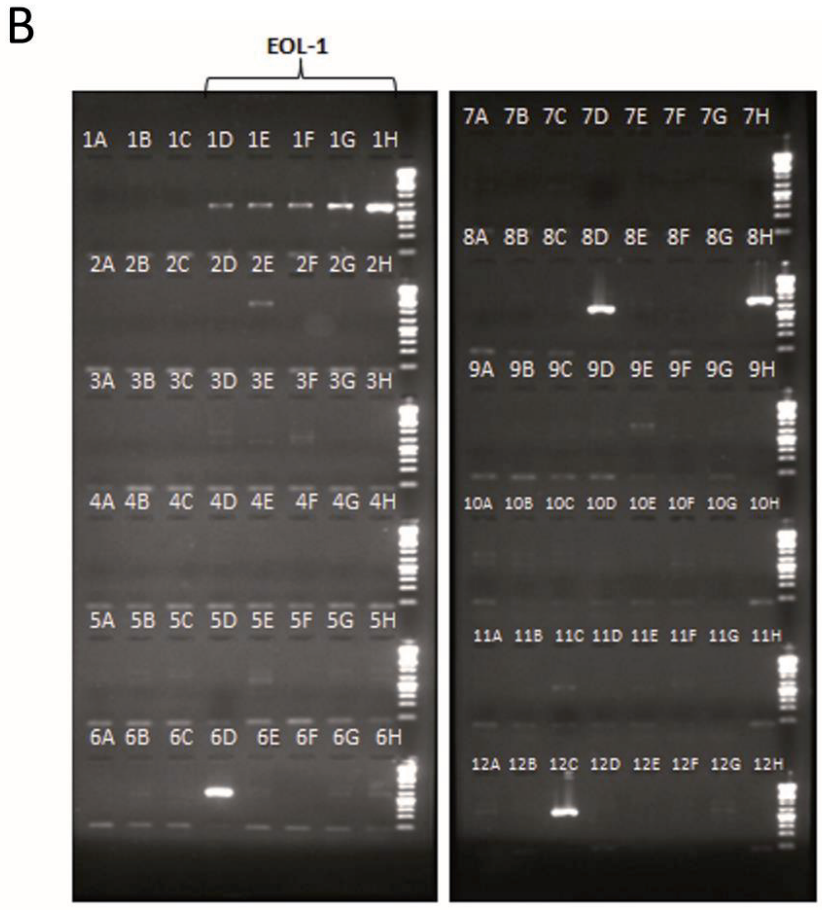
Illumina	**Qiagen**
TruSight™ Myeloid Sequencing Panel	Human Myeloid Neoplasms GeneRead DNAseq Targeted Panel V2
15 genetargets, full gene (exon only) + 39 genetargets (hotspots only)	50 genetargets,full gene (exon only)
568 amplicons (~250bp amplicon), 141kb target region	2536 amplicons (~150bp amplicon) 236kb target region
ASXL1	ASXL1
FLT3	FLT3
KIT	KIT
NPM1	NPM1
RUNX1	RUNX1
TP53	TP53
CEBPA	CEBPA
ABL1	ABL1
ATRX	ATRX
BCOR	BCOR
BCORL1	BCORL1
CBL	CBL
CBLB	CBLB
DNMT3A	DNMT3A
ETV6/TEL	ETV6
EZH2	EZH2
GATA1	GATA1
GNAS	GNAS
IDH1	IDH1
IDH2	IDH2
IKZF1	IKZF1
JAK2	JAK2
JAK3	JAK3
KRAS	KRAS
MPL	MPL
NRAS	NRAS
PHF6	PHF6
PTPN11	PTPN11
RAD21	RAD21
SETBP1	SETBP1

Illumina	Qiagen
SF3B1	SF3B1
SMC3	SMC3
STAG2	STAG2
TET2	TET2
U2AF1	U2AF1
WT1	WT1
ZRSR2	ZRSR2
KMT2A (MLL)	KMT2A (MLL)
BRAF	DAXX
CALR	EED
CBLC	JAK1
CKDN2A	KAT6A
CSF3R	NF1
CUX1	PRPF40B
FBXW7	RB1
GATA2	SF1
HRAS	SF3A1
KDM6A	SH2B3
MYD88	SMC1A
NOTCH1	SUZ12
PDGFRA	U2AF2
PTEN	
SMC1A	
SRSF2	

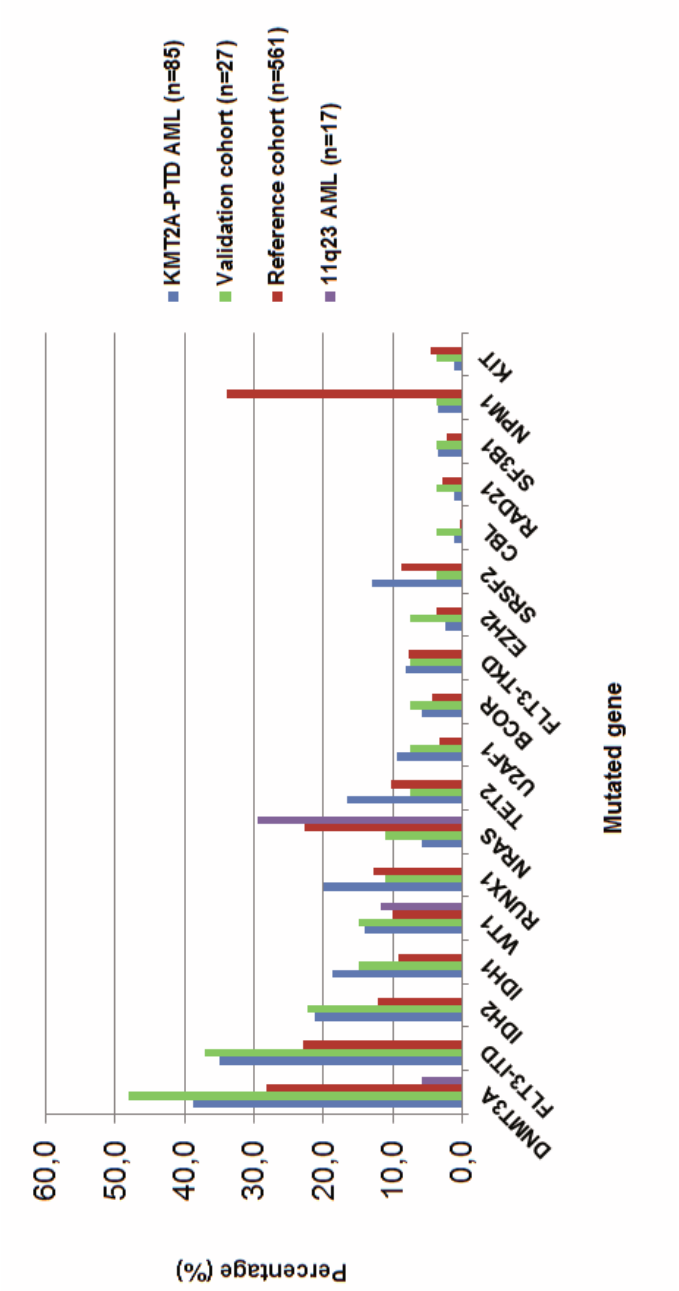
SUPPLEMENTARY FIGURES

Supplementary Figure S1: Schematic diagram of the DNA structure of a *KMT2A* wild-type (WT) and partial tandem duplication of the *KMT2A* gene. A) The boxes represent an exon, and the number above each box represents exon number. The *KMT2A* gene is composed of 36 exons, but for illustration, only the first 13 exons are displayed in the above diagram. The black small arrows indicate the location of the primers used for the analysis (adapted from Caligiuri *et al*., 1996). The grey boxes illustrate the exons that are duplicated in an e3-e9 *KMT2A*-PTD in the *KMT2A* gene. The PTD can also involve other exons of the *KMT2A* gene. B) Representative RT-PCR analyses KMT2A-PTD. The EOL1 cell line carries *KMT2A-PTD* and served as positive control (1H). The assay limit of detection was determined by diluting the EOL1 cell line (1:10 dilutions (1H-1D (100% - 0.01%)) in HL-60, a cell line without *KMT2A*-PTD.

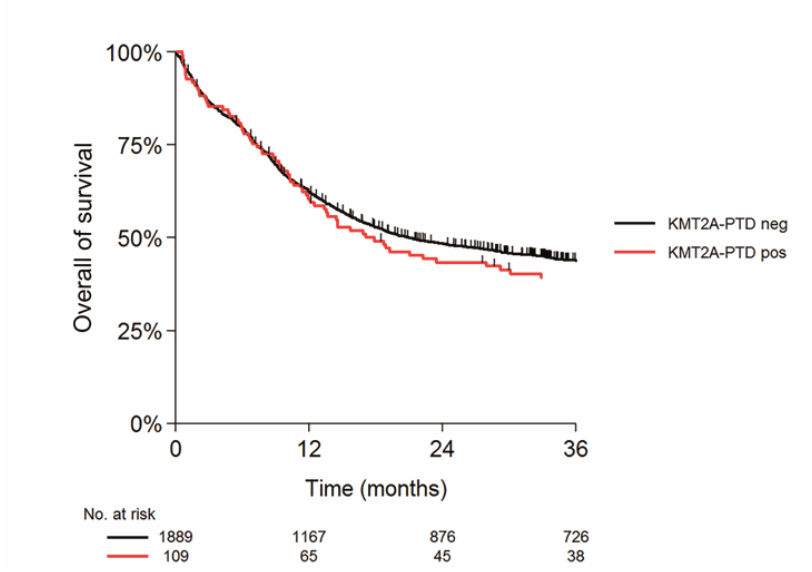




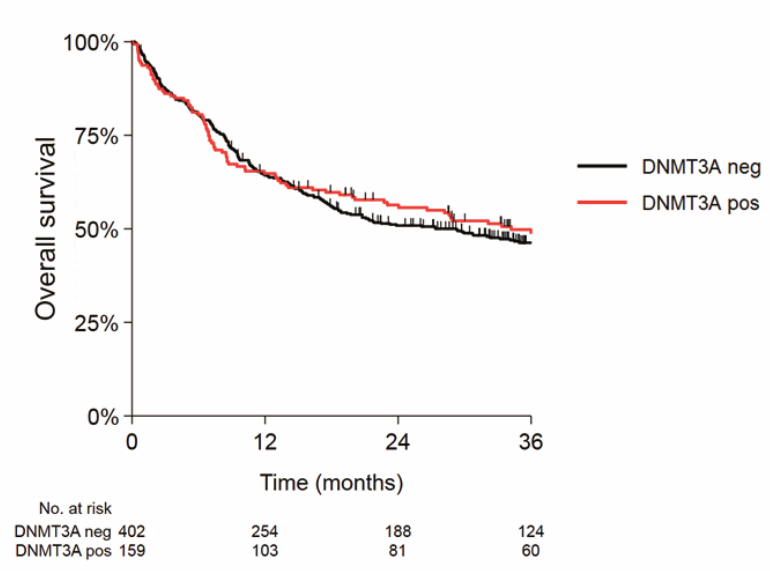
Supplementary Figure S2: *KMT2A*-PTD AML mutation frequency in study (*KMT2A*-PTD AML), validation cohort (*KMT2A*-PTD AML), AML reference cohort and t(11q23) AML. The frequencies of concurrent mutations in *KMT2A*-PTD AML study cohort (blue), *KMT2A*-PTD AML validation cohort (green), reference cohort (red) and t(11q23) (purple).



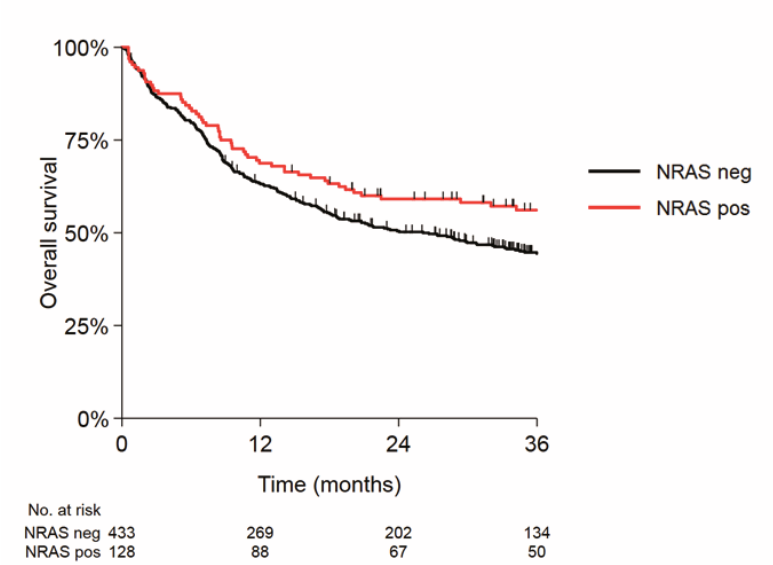
Supplementary Figure S3: Survival analysis *KMT2A*-PTD AML vs. *KMT2A*-WT AMLs. A) Kaplan-Meier graph OS of *KMT2A*-PTD AMLs (red) vs. *KMT2A* wild-type AML (black).



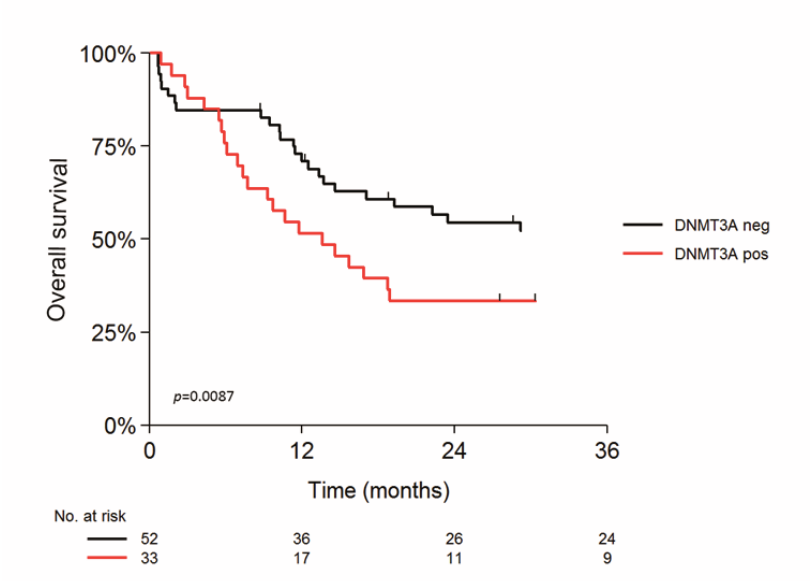
Supplementary Figure S4: Survival analysis of *KMT2A* wild-type cohort and concurrent mutations. *KMT2A* wild-type with and without *DNMT3A* mutations (p=0.99).



Supplementary Figure S5: Survival analysis of *KMT2A* wild-type cohort and concurrent mutations. *KMT2A* wild-type with and without *NRAS* mutations (p=0.044).



Supplementary Figure S6: Survival analysis of *KMT2A*-PTD validation cohort and concurrent mutations. *KMT2A*-PTD with and without *DNMT3A* mutations (p=0.0017).



REFERENCES

1. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM*, et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 2004 Apr 15; **350**(16)**:** 1617-1628.
2. Caligiuri MA, Strout MP, Lawrence D, Arthur DC, Baer MR, Yu F*, et al.* Rearrangement of ALL1 (MLL) in acute myeloid leukemia with normal cytogenetics. *Cancer Res* 1998 Jan 01; **58**(1)**:** 55-59.
3. Jongen-Lavrencic M, Grob T, Hanekamp D, Kavelaars FG, Al Hinai A, Zeilemaker A*, et al.* Molecular Minimal Residual Disease in Acute Myeloid Leukemia. *N Engl J Med* 2018 Mar 29; **378**(13)**:** 1189-1199.
4. Groschel S, Lugthart S, Schlenk RF, Valk PJ, Eiwen K, Goudswaard C*, et al.* High EVI1 expression predicts outcome in younger adult patients with acute myeloid leukemia and is associated with distinct cytogenetic abnormalities. *J Clin Oncol* 2010 Apr 20; **28**(12)**:** 2101-2107.
5. Verhaak RG, Wouters BJ, Erpelinck CA, Abbas S, Beverloo HB, Lugthart S*, et al.* Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* 2009 Jan; **94**(1)**:** 131-134.

Chapter 7

MBD4 guards against methylation damage and germ line deficiency predisposes to clonal hematopoiesis and early-onset AML

Mathijs A. Sanders,[1,*] Edward Chew,[2-5,*] Christoffer Flensburg,[2,4] Annelieke Zeilemaker,[1] Sarah E. Miller,[2] Adil S. al Hinai,[1,6] Ashish Bajel,[3,5] Bram Luiken,[1] Melissa Rijken,[1] TamaraMclennan,[7] Remco M. Hoogenboezem,[1] François G. Kavelaars,[1] Stefan Fröhling,[8-10] Marnie E. Blewitt,[4,7] Eric M. Bindels,[1] Warren S. Alexander,[2,4] Bob Löwenberg,[1] Andrew W. Roberts,[2-5] Peter J. M. Valk,[1,†] and Ian J. Majewski[2,4,†]

[1] Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands;
[2] Division of Cancer and Haematology, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia;
[3] Clinical Hematology, Peter MacCallum Cancer Center, Royal Melbourne Hospital, Parkville, VIC, Australia;
[4] Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Parkville, VIC, Australia;
[5] Victorian Comprehensive Cancer Centre, Parkville, VIC,Australia;
[6] National Genetic Center, Royal Hospital, Ministry of Health, Muscat, Sultanate of Oman;
[7] Division of Molecular Medicine, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia;
[8] Division of Translational Oncology, National Center for Tumor Diseases Heidelberg and German Cancer Research Center, Heidelberg, Germany;
[9] German Cancer Consortium, Heidelberg, Germany; and
[10] Section for Personalized Oncology, Heidelberg University Hospital, Heidelberg, Germany

Blood. 2018;132(14):1526-1534

ABSTRACT

The tendency of 5-methylcytosine (5mC) to undergo spontaneous deamination has had a major role in shaping the human genome, and this methylation damage remains the primary source of somatic mutations that accumulate with age. How 5mC deamination contributes to cancer risk in different tissues remains unclear. Genomic profiling of 3 early-onset acute myeloid leukemias (AMLs) identified germ line loss of MBD4 as an initiator of 5mC dependent hypermutation. MBD4-deficient AMLs display a 33-fold higher mutation burden than AML generally, with >95% being C>T in the context of a CG dinucleotide. This distinctive signature was also observed in sporadic cancers that acquired biallelic mutations in MBD4 and in Mbd4 knockout mice. Sequential sampling of germ line cases demonstrated repeated expansion of blood cell progenitors with pathogenic mutations in *DNMT3A*, a key driver gene for both clonal hematopoiesis and AML. Our findings reveal genetic and epigenetic factors that shape the mutagenic influence of 5mC. Within blood cells, this links methylation damage to the driver landscape of clonal hematopoiesis and reveals a conserved path to leukemia. Germ line MBD4 deficiency enhances cancer susceptibility and predisposes to AML.

KEY POINTS

- The DNA glycosylase MBD4 acts as a safeguard against damage from 5mC deamination.
- Germ line MBD4 deficiency stimulates clonal hematopoiesis and guides the development of leukemia via recurrent mutations in *DNMT3A*.

INTRODUCTION

Cells are exposed to a variety of stresses that damage DNA. Most damage arises from endogenous sources, including exposure to reactive molecules and replication errors.[1] Although the vast majority of these events are repaired, some are propagated and introduce mutations. This decay in genomic integrity has major implications for our health, particularly for modulating cancer incidence as we age. Fanconi anemia provides an illustration of this within the hematopoietic system. The specific DNA repair defects that underpin this family of diseases set the stage for a high risk of development of myelodysplasia and acute myeloid leukemia (AML) at an early age.[2]

DNA methylation on cytosine residues provides a major mutagenic stimulus, as 5-methylcytosine (5mC) has a tendency to undergo spontaneous deamination to hymine.[3] Therefore, it is not surprising that CG>TG mutations are a prominent feature of age-related DNA damage, as detected in human cancers,[4] normal stem cells,[5] and de novo mutations passed through the germ line.[6] This form of damage is so ubiquitous that it has been proposed as a molecular clock to track aging.4 CG>TG mutations make an important contribution to the somatic mutation landscape of cancer,[7] and it is important to delineate how the repair pathways that restrict methylation damage modify cancer risk.

Methylation damage is repaired by the base excision repair (BER) pathway. After deamination of 5mC, removal of the mispaired thymine is accomplished by 1 of 2 DNA glycosylases, methyl binding domain 4 (MBD4)[8] or thymine DNA glycosylase (TDG).[9] Inactivation of Mbd4 in mice confirmed a functional role in repair of methylation damage,[10, 11] but whether it protects against cancer remains unclear. In this report, we characterize familial cases with germ line inactivation of MBD4 and demonstrate its crucial role in safeguarding against methylation damage and vulnerability to the development of AML and some solid cancers.

METHODS

Patient characteristics and sample collection

Patients provided informed consent in accordance with the Declaration of Helsinki for participation in research and for collection of samples over the course of their treatment. This research project was approved by our respective human research ethics committees (HRECs) (Erasmus Medical Center [EMC] Medical Review Ethics Committee project MEC 2015-155, Walter and Eliza Hall Institute of Medical Research [WEHI] HREC project 13/01, Melbourne Health HREC project 2012.274). EMC-AML-1, WEHIAML-1, and WEHI-AML-2 were diagnosed with AML and treated with combination chemotherapy as per the protocols at their respective institutions.

EMC-AML-1 was 33 years old when diagnosed with acute monocytic leukemia (AML, World Health Organization [WHO] International Classification of Diseases [ICD] 9891/3). The AML had trisomy 11 on karyotyping and was negative for *NPM1*, *FLT3*, and *CEBPA* mutations. His medical history included colonic polyps requiring a hemicolectomy 2 years prior to his AML diagnosis. His AML was refractory to induction chemotherapy (standard dose cytarabine and daunorubicin). Repeat induction with intermediate dose cytarabine resulted in complete morphologic and cytogenetic remission. He then had an autologous hematopoietic stem cell transplant (HSCT) with BU-CY conditioning (busulfan and cyclophosphamide). He relapsed 2 years and 3 months post-autologous HSCT. The AML at relapse had a normal karyotype and was negative for *NPM1*, *FLT3*, and *CEBPA* mutations. Salvage induction chemotherapy (high-dose cytarabine, mitoxantrone, and etoposide) resulted in complete morphologic remission. This was followed by an allogeneic HSCT from a matched unrelated donor with myeloablative and total body irradiation conditioning. He achieved complete morphologic remission with full donor chimerism. He developed extensive graft-versus-host disease with secondary graft failure responsive to steroids and Epstein-Barr virus reactivation requiring rituximab. He died 2 years post-allogeneic HSCT with relapsed AML.

WEHI-AML-1 was 31 years old when diagnosed with AML withmyelodysplasia-related changes (myelodysplastic syndrome–associated cytogenetic abnormality, monosomy 7,WHOICD9895/3). The AML was negative for *NPM1*, *FLT3*, and *CEBPA* mutations. She had induction chemotherapy (high-dose cytarabine, idarubicin, and etoposide) and achieved complete morphologic and cytogenetic remission. This was followed by 2 cycles of consolidation chemotherapy (standard-dose cytarabine, idarubicin, and etoposide). Early morphologic relapse was detected on bone marrow examination prior to allogeneic HSCT from her female sibling (WEHI-AML-2) with BU-CY conditioning. Bone marrow examination 5 weeks postallogeneic HSCT showed complete morphologic and cytogenetic remission, as well as full donor chimerism. Relapsed AML (of WEHI-AML-1 origin) occurred 11 weeks post-allogeneic HSCT. Salvage therapy with FLAG chemotherapy regimen (fludarabine, cytarabine, and filgrastim) proved unsuccessful. WEHIAML-1 died of relapsed AML ,12 months after diagnosis.

WEHI-AML-2 was 30 years old when she donated peripheral blood stem cells to WEHI-AML-1. Her medical history included iron deficiency anemia secondary to menorrhagia and bleeding from descending colon and rectal polyps. Her full blood count was normal at the time of stem cell donation. Her routine full blood count 4 years later, at 34 years old, showed pancytopenia. A diagnosis of AML with myelodysplasia-related changes (myelodysplastic syndrome–associated cytogenetic abnormality, monosomy 7, WHO ICD 9895/3) was made on bone marrow examination. The AML was negative for *NPM1*, *FLT3*, and *CEBPA* mutations. She had induction chemotherapy (high-dose cytarabine and idarubicin) and achieved complete morphologic and cytogenetic remission. This was followed by 1 cycle of consolidation chemotherapy (standard-dose cytarabine, idarubicin, and etoposide). She then had an allogeneic HSCT using 2 partially HLA-matched umbilical cord blood units following FLU-CY-TBI conditioning (fludarabine; cyclophosphamide, and total body irradiation). She developed grade 1 graft-versus-host disease of the gut. She remains in complete morphologic and cytogenetic remission.

Samples from bone marrow and peripheral blood were collected over the course of their treatment (supplemental Table 1; available on the Blood Web site). WEHI-AML-2 was the donor for an allogeneic HSCT for WEHI-AML-1 and had peripheral blood taken for chimerism analysis at time of donation that was available for analysis.

Whole exome sequencing and whole genome sequencing

Whole exome sequencing on EMC-AML-1 was performed as previously described.[12] For WEHI-AML-1 and WEHI-AML-2, 50 to 100 ng of DNA and the TruSeq Nano DNA Sample Preparation Kit (Illumina) were used to generate indexed DNA libraries. Whole genome sequencing was performed on a HiSeq X Ten (Illumina). Exome capture was performed with the Human All Exon v5_UTR Capture Library and the SureSelectXT2 Target Enrichment System (Agilent Technologies) before sequencing on a HiSeq2500 (Illumina). Alignment and variant calling are detailed in the supplemental Methods.

Assessment of MBD4 status and proportion of CG>TG mutations in TCGA

To assess the frequency of CG>TG mutations in The Cancer Genome Atlas (TCGA) samples, somatic single nucleotide variant (SNV) calls available through the National Cancer Institute Genomic Data Commons were filtered to restrict the analysis to variants with a variant allele frequency >20%, with minimum 20 reads coverage and that were recognized by at least 3 out of the 4 callers: SomaticSniper, VarScan2, MuTect2, and MuSE. This approach correlated well with results from our own analysis pipeline. Candidate germ line loss-of-function variants impacting *MBD4* were sourced from Genomic Data Commons (September 2016) and analysis restricted to variants with a variant allele frequency >10%, found with

a population frequency <1% in ExAC (non-TCGA cohort).[13] The variant allele frequency and local copy number around *MBD4* were assessed in the matched cancer sample to designate cases as either monoallelic or biallelic inactivation.

Reduced representation bisulfite sequencing (RRBS)

For WEHI-AML-1 and WEHI-AML-2, RRBS libraries were made from 75 to 100 ng of DNA using the Ovation RRBS Methyl-Seq System (NuGEN) with bisulfite conversion using the Epitect kit (Qiagen). The libraries were sequenced on a HiSeq2500. Enhanced RRBS data from EMC-AML-1 were available through the Database of Genotypes and Phenotypes (dbGaP) (phs001027), and RRBS data from a glioblastoma, GBM1063T, were available from Gene Expression Omnibus (GSE70175).[14] RRBS sequencing reads were trimmed to remove adapters and low-quality sequence with Trim_Galore. Diversity adaptors were removed with a NuGEN python script (trimRRBSdiversityAdaptCustomers.py). Alignment to hg19 was performed with Bismark 0.13.0, and methylation status was assessed using bismark_methylation_extractor, ignoring 5 bases at the 5' end of each read.[15]

Whole genome sequencing of Mbd4 wild-type and knockout mice

Mbd4 knockout mice (JAX stock #004989) were obtained from Jackson Laboratory.[11] The mice were backcrossed an additional generation to C57BL/6, prior to intercrossing. All animal studies were approved by the WEHI Animal Ethics Committee (Project 2014.010). Mouse bone marrow cells were collected in Dulbecco modified Eagle medium (Thermo Fisher Scientific) containing 10% HyClone bovine calf serum, iron supplemented (Thermo Fisher Scientific). Ten thousand cells were cultured in 1 mL Dulbecco modified Eagle medium with 20% bovine calf serum, 0.3% agar (BD), 100 ng/mL murine stem cell factor, 10 ng/mL murine interleukin-3 (IL-3), and 2 IU erythropoietin.[16] Cultures were incubated for 11 days at 37°C in a fully humidified atmosphere with 10% CO2. Individual colonies were isolated, and DNA was extracted using QIAamp DNA Micro Kit (Qiagen). DNA from individual colonies was amplified using the TruePrime WGA Kit (SYGNIS), and the amplified DNA was purified using QIAamp DNA Mini Kit (Qiagen). Mouse bone marrow DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen). DNA was measured using the Agilent 2200 Tapestation Genomic DNA ScreenTape Assay (Agilent Technologies). Whole genome sequencing was performed on a NovaSeq (Illumina). DNA sequencing data were aligned to the mouse genome (mm10) using bwa-mem. Alignment, variant calling, and calculation of relative mutation rate were performed using the same approach outlined for the human sequencing data. Welch's t test was used to compare between the groups of samples (n = 3 per group).

Genomic profiling of single-cell–derived colonies (SCDCs) from EMC-AML-1

EMC-AML-1's autologous stem cell transplant and diagnosis peripheral blood samples were used to obtain single hematopoietic progenitor cell colonies. Briefly, cells were thawed and

sequentially diluted in Iscove modified Dulbecco medium (Thermo Fisher) supplemented with 5% human serum albumin (Thermo Fisher) and 20 U/mL heparin (ie, initially 1:1; after 10 minutes, 1:10; and after 20 minutes, 1:20). The cell suspension was centrifuged at 4°C, and cells were resuspended in cold phosphate-buffered saline. Cells were plated at different densities (0.04 to 23105 cells per mL) in MethoCult GF H84434 Methylcellulose medium with cytokines (Stemcell Technologies) for 14 days at 37°C and 5% CO_2. DNA was isolated from individual colonies using the QiaAmp DNA Micro Kit (Qiagen) and quantified using Qubit DNA HS assay kit (Life Technologies). The Illumina TruSight Myeloid Sequencing Panel (Illumina) was applied to detect mutations in genes frequently mutated in myeloid malignancy.

MBD4 glycosylase activity assays

MBD4 glycosylase activity assays were performed as previously described with the following modifications.[17] The glycosylase activity of MBD4 protein (0.5 μM) on double-stranded FAM-labeled 32bp-oligonucleotides (0.1μM) was assessed and monitored by denaturing gel electrophoresis. The resulting FAM labeled single-stranded DNA was visualized using the 473-nm laser (Blue LD Laser) and 530DF20 emission filter on a Typhoon FLA9500 (GE Healthcare).

The following 32-bp oligonucleotides were obtained from Integrated DNA Technologies: (FAM)-5'-TCGGATGTTGTGGGTCAGXGCATGATAGTGTA-3' (where X = C or T); 5'-TACACTATCATGCGCTGACCCACAACATCCGA-3'. The double-stranded FAM-labeled matched and mismatched oligonucleotides were prepared by hybridization whereby 100 μM of oligodinucleotides were mixed in 50 μL annealing buffer containing 10 mM tris (hydroxymethyl) aminomethane HCl, 1 mM EDTA, and 50 mM NaCl (pH 8.0), then incubated at 95°C for 2 minutes, followed by a steady temperature reduction over 45 minutes to 25°C. The double-stranded duplexes were cooled and stored at 4°C.

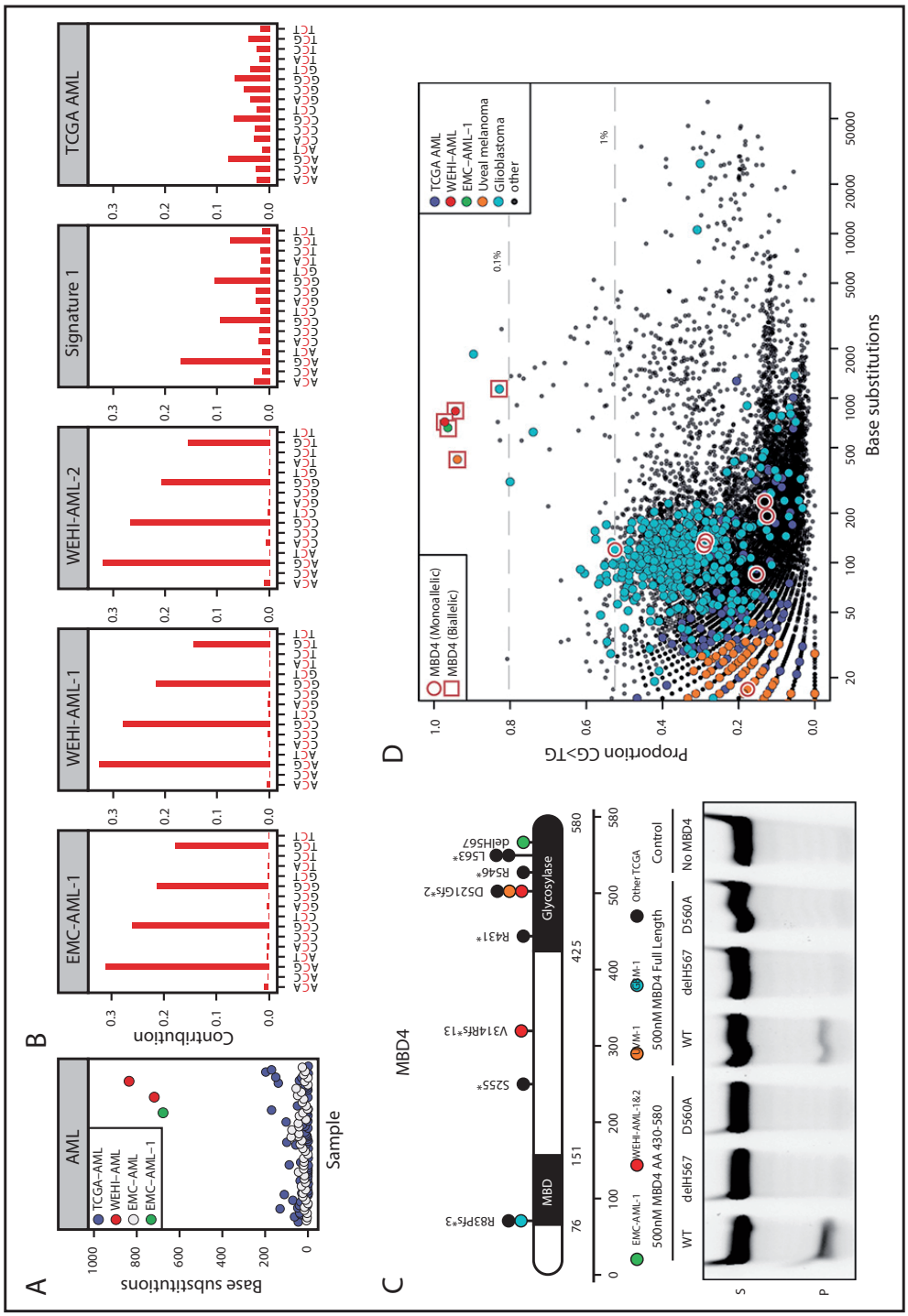
RESULTS

Germ line loss of MBD4 predisposes to AML with a novel mutational signature

We identified 3 patients with AML, including 2 siblings, that were distinctive because of their high mutational burden (~33-fold above what is typical for AML) and unique mutational signature, where .95% of mutations were CG>TG (Figure 1A-B; supplemental Figure 1A). This signature differs from the distribution of C>T mutations generally observed in AML and is more refined than the mutational signature ascribed to aging,4 suggesting a near complete dependence on 5mC deamination. Although CG>TG mutations are an integral feature of age-related DNA damage and AML is most commonly a disease of older age (median age of onset is >70 years), all 3 patients were younger than 35 years at diagnosis.

Sequencing germ line DNA from the 3 cases identified loss-of-function variants in the gene encoding the DNA glycosylase MBD4, which plays a key role in initiating repair after 5mC deamination[8] (Figure 1C; supplemental Table 2). Case EMC-AML-1 carried a homozygous deletion of Histidine 567 (H567) in the glycosylase domain of MBD4. An in vitro glycosylase assay confirmed that loss of H567 results in a catalytically inactive MBD4 protein (Figure 1C). The siblings (WEHI-AML-1,WEHI-AML-2) were compound heterozygotes with a frameshift in exon 3 and a variant that disrupts the splice acceptor of exon 7 of MBD4 (Figure 1C; supplemental Figure 2A). Analysis of the MBD4 messenger RNA allowed for phasing of the variants to distinct alleles and confirmed aberrant splicing that excludes exon 7 and disrupts the glycosylase domain (supplemental Figure 2B). MBD4 has not previously been associated with hematological malignancy, but somatic mutations, predominantly frameshifts, have been detected in sporadic colon cancers with mismatch repair deficiency.[18, 19] Two patients (EMCAML-1, WEHI-AML-2) also had colorectal polyps, a common manifestation of DNA repair defects, including those associated with loss of BER components MUTYH[20] and NTHL1.[21]

Figure 1. MBD4-deficient cancers exhibit a distinctive mutational signature. (A) Mutation burden in AML, presented as number of base substitutions per exome. Data sourced from dbGaP; cases are ordered on patient identifier (EMC: phs00102712 and TCGA: phs00017824). (B) Trimer context of C.T mutations in 3 MBD4-deficient AML cases. The center of origin is reflected in the sample label. For comparison, we show signature 1, the established signature associated with 5mC deamination, and all C.T mutations present in TCGA-AML. (C) Schematic representation of MBD4, highlighting germ line loss-of-function variants detected in the AML cases and cases within TCGA (at top). A glycosylase assay was performed to assess the activity of recombinant MBD4 (either AA430-580 or full length), wild-type (WT), delH567, or the catalytically inactive mutant D560A. Substrate (S) and product (P). Consistent results were obtained in 5 experiments for MBD4 AA430-580 and 3 experiments for full length. (D) The proportion of CG.TG mutations observed is set out against the total number of base substitutions detected for all TCGA samples. Samples with germ line MBD4 loss-of-function variants were designated either as heterozygous (monoallelic) or completely inactivated (biallelic) based on the genotype of the cancer (includes somatic mutations). Gray lines mark the top 1% and 0.1% of cases with the highest proportion of CG.TG mutations. A select set of tumor types are highlighted.


A
AML
TCGA–AML
WEHI–AML
EMC–AML
EMC–AML–1
Base substitutions
Sample
B
EMC-AML-1
WEHI-AML-1
WEHI-AML-2
Signature 1
TCGA AML
Contribution
C
MBD4
MBD
Glycosylase
R83Pfs*3
S255*
V314Rfs*13
R431*
D521Gfs*2
R546*
L563*
delH567
EMC-AML-1
WEHI-AML-18.2
AML-1
Other TCGA
500nM MBD4 AA 430-580
500nM MBD4 Full Length
Control
No MBD4
WT
delH567
D560A
S
P
D
MBD4 (Monoallelic)
MBD4 (Biallelic)
TCGA AML
WEHI-AML
EMC–AML–1
Uveal melanoma
Glioblastoma
other
0.1%
1%
Proportion CG>TG
Base substitutions

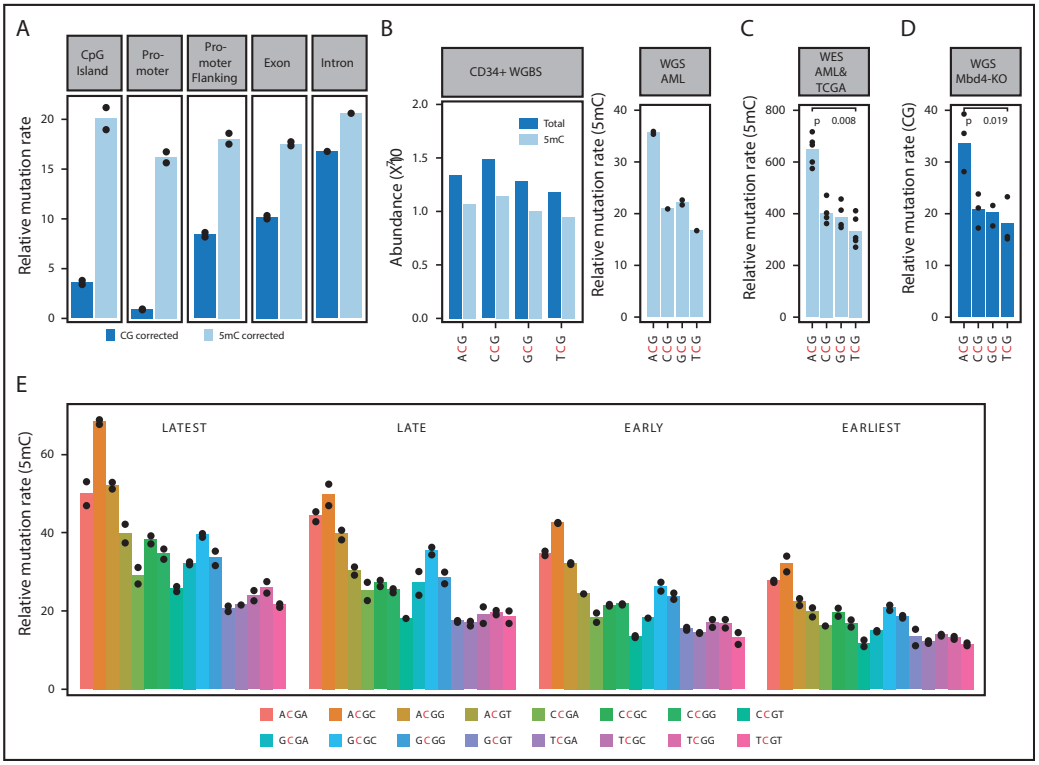


Figure 2. Damage introduced by 5mC deamination is influenced by genetic and epigenetic features. (A) Observed relative mutation rates (RMRs) at different genomic features in whole genome sequencing from WEHI-AML-1 and WEHI-AML-2, calculated per Mb of CG dinucleotides (CG corrected), or corrected for methylation status in normal CD341 cells (5mC corrected). (B) Abundance and methylation status for NCG trimers from whole genome bisulfite sequencing derived from normal CD341 cells.[37] An RMR value was calculated for WEHI-AML-1 and WEHI-AML-2 for each NCG trimer, accounting for differences in abundance and 5mC status in normal CD341 cells and scaled to account for total mutation load (see supplemental Methods). Individual values are plotted (n 5 2), and bars show the mean. (C) RMR values were calculated from exome data for the 5 MBD4-deficient cancers. There was a significant enrichment of mutations in the ACG context compared with TCG (P 5 .0079, Mann-Whitney U test). (D) RMR values were calculated from whole genome sequencing data generated from Mbd4 knockout (Mbd4-KO) murine blood cell progenitors at 4 months of age. Values from individual colonies are plotted (n53), and the bar shows the mean. There was a significant enrichment of mutations in the ACG context compared with TCG (P5.019, Welch's t test). (E) RMR values were calculated for NCGN tetramers in WEHI-AML-1 and WEHI-AML-2, then separated by replication timing (n 5 2).

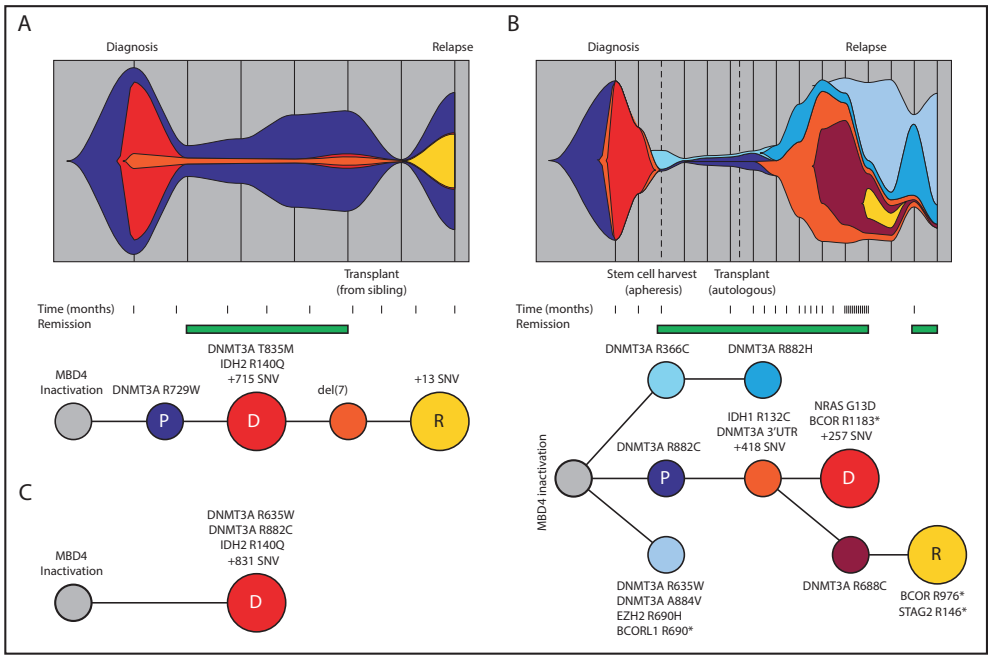


Figure 3. Germ line MBD4-deficient patients share a common path to AML. Clonal evolution and phylogenetic tree diagram highlighting the acquisition of key driver mutations and clonal dynamics in WEHI-AML-1 (A) and in EMC-AML-1 (B). (C) The phylogenetic tree diagram for key driver mutations in WEHI-AML-2. Variant allele frequencies were derived from whole exome sequencing data or deep sequencing for all cases. For EMC-AML-1 single-cell genotyping was used to resolve the clonal relationships. Clones are represented by different colors, and the vertical lines in the top panels indicate sampling points. The premalignant clone (P, in dark blue) and the AML clones evident at diagnosis (D, in red) and relapse (R, in yellow) are designated. Both WEHI-AML-1 and EMC-AML-1 experienced clonal hematopoiesis during remission. The transplant for WEHIAML-1 was provided by WEHI-AML-2, which occurred 4 years prior to her own diagnosis of AML.

MBD4 deficiency drives a common path of clonal evolution to AML

The 3 cases of AML with germ line MBD4 deficiency exhibited common molecular features, including biallelic *DNMT3A* mutations and *IDH1* or *IDH2* hot spot mutations, all of which were CG>TG (Figure 3A-C). This is a relatively rare path to AML, affecting <3% of patients in TCGA-AML[24]; therefore, it is highly unlikely that these 3 individuals share this pattern of driver mutations by chance. Analysis of sequential bone marrow biopsies taken during treatment and single-cell genotyping allowed us to refine the order of somatic mutation acquisition in 2 cases (EMC-AML-1, WEHI-AML-1), with *DNMT3A* mutations preceding IDH mutations (Figure 3A-B; supplemental Figure 7). *DNMT3A* mutations present in the AML at diagnosis were also detected in non-leukemic bone marrow populations in both cases, indicating that these mutations are among the first acquired. Mutations in *DNMT3A* are known to alter the self-renewal capacity of hematopoietic stem cells (HSCs)[25] and are associated with age-related clonal hematopoiesis (ARCH), also known as clonal hematopoiesis of indeterminate potential.[26-29] For both cases, a marked expansion of clones carrying *DNMT3A* mutations occurred in the remission phase following treatment (Figure 3A-B). EMC-AML-1 experienced multiple clonal outgrowths, with 9 distinct *DNMT3A* mutations, and repeated selection of clones with biallelic *DNMT3A* mutations, which appears to be a key step in the development of leukemia in these patients. Broader testing of other AMLs with biallelic *DNMT3A* mutations demonstrated that 24 out of 30 (80%) have coincident mutations in *IDH1* or *IDH2*, suggesting cooperation between these mutations that may explain this conserved path to leukemia.

MBD4 deficiency stimulates clonal hematopoiesis through inactivation of DNMT3A

To determine the influence of this mutational process on the composition of MBD4-deficient bone marrow, we genotyped additional single cells, or SCDCs, isolated from EMC-AML-1 at multiple points during treatment. As expected, the leukemic clones were dominant at the time of diagnosis and relapse, but genotyping individual cells revealed that they continue to acquire CG>TG mutations (supplemental Figure 8). When HSCs collected at remission were examined, we found that 20 of 30 (67%) SCDCs carried mono- or biallelic CG>TG mutations in *DNMT3A* that were mostly distinct (Figure 4A). A further 2 (7%) SCDCs carried CG>TG mutations in *TP53* (Figure 4B). Deep variant calling across all EMC-AML-1 samples uncovered additional CG.TG mutations in ARCH-associated genes: 28 in *DNMT3A*, 10 in *TP53*, 5 in *ASXL1*, and 7 in *TET2* (Figure 4A-D). When these findings are extrapolated to the entire bone marrow compartment, it suggests a rich diversity of clones carrying mutations in ARCH-associated genes, predominantly in *DNMT3A*. Three observations support the notion that the mutations in *DNMT3A* are functionally important: first, their repeated expansion in the blood indicates a fitness advantage; second, there is clear enrichment of nonsynonymous mutations (assessed with dNdScv,[30] q = 4.63e-05, Benjamini-Hochberg corrected); and third, the majority of mutations (65%) have been observed in ARCH[26, 28, 31] (Figure 4A). Taken together, these results emphasize the importance of 5mC damage as a source of mutations that drive clonal expansion in the blood, representing a key contributor to ARCH.

Inactivation of MBD4 is associated with a methylation damage signature across different types of cancer

We mined large cancer databases to explore the link between MBD4 deficiency and the distinctive CG>TG signature. Analysis of TCGA, comprising 10 683 cancers (including 200 AMLs), identified 9 cases that carried germline loss-of-function variants in MBD4 (Figure 1C; supplemental Figure 1A-B and supplemental Table 2). In 2 of these cases, a uveal melanoma (TCGA-UVM-1) and a glioblastoma multiforme (TCGA-GBM-1), splice site mutations were accompanied by loss of the wild-type MBD4 allele (supplemental Figure 3A). Analysis of RNA sequencing from both tumors confirmed aberrant splicing of MBD4, predicted to result in protein truncation and loss of function (supplemental Figure 3B). Both cases exhibited an elevated mutation rate and strong enrichment for CG.TG mutations, similar to the MBD4-deficient AMLs (Figure 1D; supplemental Figure 1A). This signature was also observed in a glioma cell line, SW1783, that carries a homozygous truncating variant in MBD4 at Leucine 563 (supplemental Figure 1A). Cancers that retained a wild-type allele did not display a prominent CG>TG signature (Figure 1D; supplemental Figure 1A). These results suggest both alleles of *MBD4* must be inactivated to inhibit its repair activity, which is consistent with other BER-associated cancer syndromes.[20, 21]

Genetic and epigenetic features that impact methylation damage

Whole-genome sequencing and methylation profiling were performed to refine the mutational signature associated with MBD4 deficiency in AML. Overall, >15 000 substitution mutations were identified in each AML genome, of which >90% were CG>TG (supplemental Figure 1B). Insertions and deletions were uncommon, suggesting the mismatch repair pathway remains intact. The mutation rate was linked to 5mC abundance. Sparsely methylated regions, such as promoters and CG islands, were rarely mutated (Figure 2A). Correcting for 5mC abundance measured in normal CD34+ cells revealed a consistent mutation rate across different genomic features (Figure 2A). Direct assessment of the methylation status in MBD4-deficient cancers, or matched control tissue, confirmed that mutations occurred at methylated CG sites (supplemental Figure 4).

We next assessed the influence of genetic and epigenetic features on the mutation rate.[22] When we examined the local sequence context, we observed that the proportion of mutations was higher in the context of the ACG triplet and lower in the context of TCG, with CCG and GCG being intermediate. The preference for ACG remained after correction for trimer abundance and methylation status (Figure 2B) and was found to be significant in the exome data from 5 MBD4-deficient cancers (P = .007937, Mann-Whitney U test) (Figure 2C). The same mutational signature, including the preference for the ACG trimer, was recapitulated in blood cell progenitors isolated from Mbd4 knockout mice, both at 4 months of age (Figure 2D) and in animals aged for over a year, which had a higher mutation burden (supplemental Figure 5). The ACA trimer was the most commonly mutated site

outside of a CG context in the cancers, and this matches the most common site of non-CG methylation.[23] Extending the analysis of sequence context to include 1 base on either side of the CG identified higher mutation rates in the context of a 3' cytosine (NCGC). The relative mutation rate was not influenced by the transcriptional strand (supplemental Figure 6A) but was higher in late replicating regions (Figure 2E) and at lowly expressed genes (supplemental Figure 6B). The differences between tetramers and enrichment in late replicating regions were also evident in rare germ line CG>TG single nucleotide polymorphisms from the gnomAD database[13] (supplemental Figure 6C). Collectively, these results suggest that although 5mC is the dominant factor contributing to the mutation rate, the local sequence context, replication timing, and expression status also contribute.

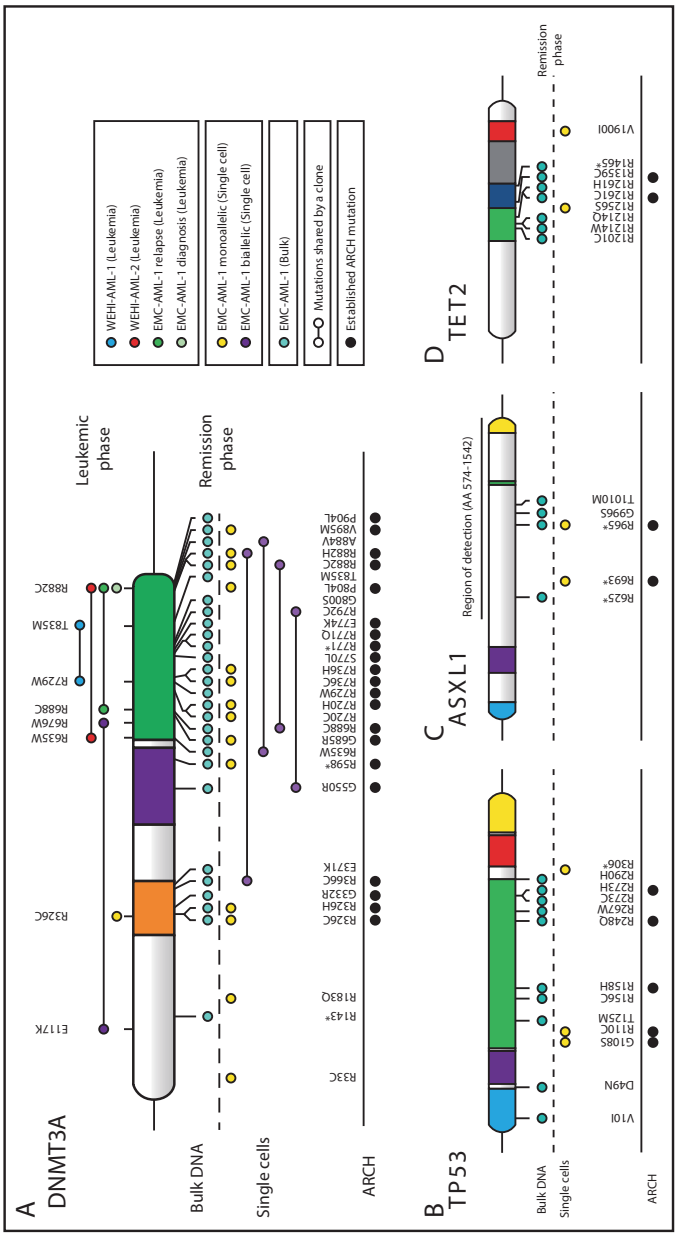


Figure 4. Recurrent C>T mutations in genes implicated in age-related clonal hematopoiesis (ARCH). (A) *DNMT3A* mutations were detected in MBD4-deficient patients at time of disease (leukemic phase) or remission (remission phase). EMC-AML-1 had additional *DNMT3A* mutations that were detected through sequencing of bulk DNA, SCDCs obtained from diagnostic bone marrow, and SCDCs from autologous stem cells collected during complete remission. The majority of the *DNMT3A* mutations had been detected in healthy individuals with ARCH. Additional point mutations were identified in remission material from EMC-AML-1, in *TP53* (B), *ASXL1* (C), and *TET2* (D). A more detailed phylogenetic tree is provided in supplemental Figure 8.

DISCUSSION

Here we describe a new genetic predisposition to cancer, in which germ line MBD4 deficiency is associated with the development of early-onset AML, through the acquisition of pathogenic mutations in driver genes, most particularly *DNMT3A*. Although additional investigation is required to determine the frequency with which MBD4 deficiency contributes to familial cancer predisposition and to refine the disease spectrum and penetrance, our results highlight a crucial role for MBD4 in safeguarding against the damage wrought by 5mC deamination. Concomitantly, 2 other groups have also identified the link betweenMBD4 inactivation and methylation damage, through identification of sporadic solid cancers with a combination of germ line and somatic mutations (Rodrigues et al[32] and Jan Korbel, manuscript submitted November 2017). Our study, in addition, reveals the impact of constitutive inactivation of MBD4 on the development of early-onset AML and reveals that blood cell progenitors are particularly sensitive to methylation damage.

As noted earlier, methylation damage accumulates as part of normal aging.[4, 5]Our current understanding of how methylation damage manifests largely depends on mutational profiles garnered from large collections of human cancers,[4] but distilling a clear signature has been complicated by the diverse DNA damage processes and repair defects present in those cancers. MBD4-deficient cancers, particularly cases with constitutive loss, provide a unique opportunity to refine the mutational signature for methylation damage, and we have identified genetic and epigenetic factors that shape its influence. This distinctive damage signature was recapitulated in blood cells from *Mbd4* knockout mice, indicating that the DNA repair pathway guarding against methylation damage is broadly conserved. The ubiquitous nature of methylation damage means even small fluctuations in mutation rate are relevant if we wish to understand its influence on genomic integrity. Our results demonstrate a profound link between methylation damage and the development of hematological malignancy, which is reshaping our understanding of how 5mC contributes to cancer risk over a lifetime.

One manifestation of methylation damage is clonal hematopoiesis, a phenomenon typically observed in people >70 years of age.[26-29] The influence of methylation damage is reflected in the prevalence of C.T mutations in clonal hematopoiesis, which has been noted previously.[28, 33] Individuals with biallelic loss of *MBD4* in the germ line confirm this link; they sustain high levels of damage from 5mC deamination throughout their lifetime and experience clonal expansions decades earlier, which eventually progress to AML. Repeated sampling and single-cell genotyping of blood cell progenitors revealed a rich diversity of mutations that overlap the driver landscape of clonal hematopoiesis, including mutations in *DNMT3A* particularly, but also in *TP53*, *ASXL1*, and *TET2*. The coexistence of this diverse array of mutant clones, and our ability to monitor their prevalence dynamically, offers new insight into the fitness landscape of clonal hematopoiesis. Future studies will need to explore the

latency and degree of penetrance of mutations associated with clonal hematopoiesis and AML in *Mbd4* knockout mice as they age, in order to fully investigate the human disease pathogenesis we have identified.

There are >40 million 5mC residues in the genome, yet the 3 individuals that lack MBD4 constitutively all developed the same type of cancer, AML, with a common set of driver mutations. A small set of genes have been defined that predispose to AML (reviewed by Godley and Shimamura[34]), including DNA repair genes, such as those in the Fanconi anemia pathway, but to our knowledge none exhibit such a conserved path to malignancy.

Our results indicate this convergence results from the combination of a highly restricted mutational signature, which accesses a select set of driver genes, and the role of DNMT3A, which regulates HSC self-renewal capacity and protects against transformation.[25, 35, 36] This interaction between mutational process, driver landscape, and stem cell biology may explain the tissue-restricted pattern of disease in this and other cancer predisposition syndromes and has broader implications for understanding how the aging process shapes cancer risk.

Acknowledgments

The authors thank S. He, A. Rijneveld, K. van Lom, and K. Gussinklo for providing clinical information; M. Wall for assistance with cytogenetics; N. Sprigg for assistance with sample collection; L. Di Rago for assistance with mouse agar colonies; E. Rombouts for assistance with single-cell sorting; I. Martincorena and F. Abascal for advice on the dNdScv model; S. van Rossum and J. Lebbink for assistance with recombinant protein isolation; the Australasian Leukaemia and Lymphoma Group for access to clinical samples; and S. Wilcox for technical assistance with sequencing. Additional sequencing was performed at the Australian Genome Research Facility (Melbourne, VIC, Australia) and the Kinghorn Centre for Clinical Genomics (Sydney, NSW, Australia). Sean Grimmond, Jason Wong, Oliver Sieber, Alicia Oshlack, and Stephen Nutt provided valuable feedback on the manuscript.

This work was supported by the Australian National Health and Medical Research Council (NHMRC) (program grant 1113577 [W.S.A. and A.W.R.] and project grant 1145912 [I.J.M.]), an Independent Research Institutes Infrastructure Support Scheme Grant (9000220), a Victorian State Government Operational Infrastructure Support Grant, The Netherlands Organisation for Scientific Research (NWO), and the Center for Translational Molecular Medicine (CTMM). M.A.S. is supported by a grant from CTMM (GR03O-102) and a Rubicon fellowship from NOW (019.153LW.038). E.C. is supported by a PhD scholarship from the Leukaemia Foundation of Australia. A.S.a.H. is supported by a PhD scholarship from theMinistry of Health, Sultanate ofOman. M.E.B. is supported by the Bellberry-Viertel fellowship. W.S.A. and A.W.R. are supported by fellowships from NHMRC (1058344 and 1079560, respectively). I.J.M. is supported by the Victorian Cancer Agency. The authors wish to acknowledge the generous philanthropic support of the Felton Bequest, Malcolm Broomhead, and BHP Billiton.

The results are based, in part, on data generated by the TCGA Research Network (http://cancergenome.nih.gov/) and the Epigenetic studies in Acute Myeloid Leukemia (phs001027), which was supported by National Institutes of Health, National Cancer Institute (K08CA169055) (F. E. Garrett-Bakelman), Starr Cancer Consortium I4-A442 (A. M. Melnick, R. Levine, and C. E. Mason), and LLS SCOR 7006-13 (A. M. Melnick). Sequencing data from WEHI-AML-1 andWEHI-AML-2 have been deposited at the EuropeanGenome Phenome Archive (EGA) (EGAS00001002581). The data are available for ethically approved research into hematological malignancy upon completion of a data transfer agreement. Sequencing data from EMC-AML-1 were sourced from the dbGaP under accession phs001027. Sequencing data from the Mbd4 knockout mice is available through the National Center for Biotechnology Information (NCBI) Short Read Archive (SRP126117). The code for reproducing figures is made available through GitHub (https://github.com/MathijsSanders/AML-RoaMeR).

Authorship

Contribution: M.A.S., E.C., A.W.R., P.J.M.V., and I.J.M. conceived and designed research; M.A.S., E.C., C.F., A.Z., S.E.M., A.S.a.H., A.B., B. Luiken, M.R., T.M., R.M.H., F.G.K., A.W.R., P.J.M.V., and I.J.M. developed methodology and performed research; M.A.S., E.C., C.F., A.Z., S.E.M., R.M.H., F.G.K., S.F., M.E.B., E.M.B., W.S.A., A.W.R., P.J.M.V., and I.J.M. analyzed data; and M.A.S., E.C., C.F.,W.S.A., B. Löwenberg, A.W.R., P.J.M.V., and I.J.M. wrote the manuscript or contributed to revision of the manuscript.

Conflict-of-interest disclosure

The authors declare no competing financial interests.

Correspondence

Ian J. Majewski, Cancer and Haematology Division, The Walter and Eliza Hall Institute, 1G Royal Parade, Parkville 3052, VIC, Australia; e-mail: majewski@wehi.edu.au.

Footnotes

Submitted 21 May 2018; accepted 18 July 2018. Prepublished online as Blood First Edition paper, 26 July 2018; DOI 10.1182/blood-2018-05-852566.

*M.A.S. and E.C. are joint first authors.

†P.J.M.V. and I.J.M. are joint senior authors.

The online version of this article contains a data supplement.

There is a Blood Commentary on this article in this issue.

REFERENCES

1. Lindahl T, Wood RD. Quality control by DNA repair. *Science* 1999 Dec 3; **286**(5446)**:** 1897-1905.
2. Ceccaldi R, Sarangi P, D'Andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol* 2016 Jun; **17**(6)**:** 337-349.
3. Duncan BK, Miller JH. Mutagenic deamination of cytosine residues in DNA. *Nature* 1980 Oct 9; **287**(5782)**:** 560-561.
4. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV*, et al.* Signatures of mutational processes in human cancer. *Nature* 2013 Aug 22; **500**(7463)**:** 415-421.
5. Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N*, et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 2016 Oct 13; **538**(7624)**:** 260-264.
6. Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Turki SA*, et al.* Timing, rates and spectra of human germline mutation. *Nat Genet* 2016 Feb; **48**(2)**:** 126-133.
7. Cooper DN, Youssoufian H. The CpG dinucleotide and human genetic disease. *Hum Genet* 1988 Feb; **78**(2)**:** 151-155.
8. Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites (vol 401, pg 301, 1999). *Nature* 2000 Mar 30; **404**(6777)**:** 525-525.
9. Wiebauer K, Jiricny J. In vitro correction of G.T mispairs to G.C pairs in nuclear extracts from human cells. *Nature* 1989 May 18; **339**(6221)**:** 234-236.
10. Millar CB, Guy J, Sansom OJ, Selfridge J, MacDougall E, Hendrich B*, et al.* Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* 2002 Jul 19; **297**(5580)**:** 403-405.
11. Wong E, Yang K, Kuraguchi M, Werling U, Avdievich E, Fan K*, et al.* Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci U S A* 2002 Nov 12; **99**(23)**:** 14937-14942.
12. Li S, Garrett-Bakelman FE, Chung SS, Sanders MA, Hricik T, Rapaport F*, et al.* Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nature Medicine* 2016 Jul; **22**(7)**:** 792-799.
13. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T*, et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016 Aug 18; **536**(7616)**:** 285-291.
14. Lee EJ, Rath P, Liu J, Ryu D, Pei L, Noonepalle SK*, et al.* Identification of Global DNA Methylation Signatures in Glioblastoma-Derived Cancer Stem Cells. *J Genet Genomics* 2015 Jul 20; **42**(7)**:** 355-371.
15. Yin DQ, Ritchie ME, Jabbari JS, Beck T, Blewitt ME, Keniry A. High concordance between Illumina HiSeq2500 and NextSeq500 for reduced representation bisulfite sequencing (RRBS). *Genom Data* 2016 Dec; **10:** 97-100.
16. Alexander WS, Roberts AW, Nicola NA, Li R, Metcalf D. Deficiencies in progenitor cells of multiple hematopoietic lineages and defective megakaryocytopoiesis in mice lacking the thrombopoietic receptor c-Mpl. *Blood* 1996 Mar 15; **87**(6)**:** 2162-2170.
17. Hashimoto H, Liu Y, Upadhyay AK, Chang Y, Howerton SB, Vertino PM*, et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res* 2012 Jun; **40**(11)**:** 4841-4849.
18. Bader S, Walker M, Hendrich B, Bird A, Bird C, Hooper M*, et al.* Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. *Oncogene* 1999 Dec 23; **18**(56)**:** 8044-8047.
19. Riccio A, Aaltonen LA, Godwin AK, Loukola A, Percesepe A, Salovaara R*, et al.* The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. *Nature Genetics* 1999 Nov; **23**(3)**:** 266-268.

20. Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT*, et al.* Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nat Genet* 2002 Feb; **30**(2)**:** 227-232.
21. Weren RD, Ligtenberg MJ, Kets CM, de Voer RM, Verwiel ET, Spruijt L*, et al.* A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. *Nat Genet* 2015 Jun; **47**(6)**:** 668-671.
22. Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM*, et al.* Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell* 2016 Jan 28; **164**(3)**:** 538-549.
23. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J*, et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009 Nov 19; **462**(7271)**:** 315-322.
24. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ*, et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013 May 30; **368**(22)**:** 2059-2074.
25. Challen GA, Sun DQ, Jeong M, Luo M, Jelinek J, Berg JS*, et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature Genetics* 2012 Jan; **44**(1)**:** 23-U43.
26. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF*, et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014 Dec 25; **371**(26)**:** 2477-2487.
27. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC*, et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med* 2014 Dec; **20**(12)**:** 1472-1478.
28. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG*, et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine* 2014; **371**(26)**:** 2488-2498.
29. McKerrell T, Park N, Moreno T, Grove CS, Ponstingl H, Stephens J*, et al.* Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Rep* 2015 Mar 3; **10**(8)**:** 1239-1245.
30. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P*, et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* 2017 Nov 16; **171**(5)**:** 1029-1041 e1021.
31. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E*, et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N Engl J Med* 2017 Jul 13; **377**(2)**:** 111-121.
32. Rodrigues M, Mobuchon L, Houy A, Fievet A, Gardrat S, Barnhill RL*, et al.* Outlier response to anti-PD1 in uveal melanoma reveals germline MBD4 mutations in hypermutated tumors. *Nat Commun* 2018 May 14; **9**(1)**:** 1866.
33. Yoshizato T, Dumitriu B, Hosokawa K, Makishima H, Yoshida K, Townsley D*, et al.* Somatic Mutations and Clonal Hematopoiesis in Aplastic Anemia. *N Eng J Med* 2015 Jul 2; **373**(1)**:** 35-47.
34. Godley LA, Shimamura A. Genetic predisposition to hematologic malignancies: management and surveillance. *Blood* 2017 Jul 27; **130**(4)**:** 424-432.
35. Cole CB, Russler-Germain DA, Ketkar S, Verdoni AM, Smith AM, Bangert CV*, et al.* Haploinsufficiency for DNA methyltransferase 3A predisposes hematopoietic cells to myeloid malignancies. *J Clin Invest* 2017 Oct 2; **127**(10)**:** 3657-3674.
36. Mayle A, Yang L, Rodriguez B, Zhou T, Chang E, Curry CV*, et al.* Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. *Blood* 2015 Jan 22; **125**(4)**:** 629-638.
37. Spencer DH, Russler-Germain DA, Ketkar S, Helton NM, Lamprecht TL, Fulton RS*, et al.* CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression. *Cell* 2017 Feb 23; **168**(5)**:** 801-+.

SUPPLEMENTARY DATA

Methylation damage due to germline MBD4-deficiency predisposes to clonal hematopoiesis and early-onset AML

Authors
Mathijs A. Sanders, Edward Chew, Christoffer Flensburg, Annelieke Zeilemaker, Sarah E. Miller, Adil S. al Hinai, Ashish Bajel, Bram Luiken, Melissa Rijken, Tamara Mclennan, Remco M. Hoogenboezem, François G. Kavelaars, Stefan Fröhling, Marnie E. Blewitt, Eric M. Bindels, Warren S. Alexander, Bob Löwenberg, Andrew W. Roberts, Peter J.M. Valk, Ian J. Majewski.

Corresponding author
Ian J. Majewski PhD
Cancer & Haematology Division
The Walter and Eliza Hall Institute
1G Royal Parade, Parkville, 3052
Australia
Ph: +61393452169
Em: majewski@wehi.edu.au

This file includes
Supplementary Methods
Supplementary References
Supplementary Tables 1-2
Supplementary Figures 1-8
Other Supplementary Materials for this manuscript include the following:
Supplementary Data Sets (Data S1 to S4)

SUPPLEMENTARY METHODS

Validation and phasing of WEHI-AML-1 *MBD4* variants

Primers flanking the *MBD4* variants in exon 3 and exon 7 were designed using Primer3. Non-amplified DNA extracted from remission material from WEHI-AML-1 was used as a template. PCR of exon 7 (MBD4_Ex7-F: 5'-CCTCTTGGTCTCTACGATCTTC-3' and MBD4_Ex7-R: 5'-TCGGTAAGAGTCGTTGCCAT-3') was performed using Platinum *Taq* DNA Polymerase (Thermo Fisher Scientific) (5 minutes at 94°C, 35 cycles of 20 seconds at 94°C, 30 seconds at 60°C and 60 seconds at 72°C, and a final 5 minutes at 72°C). PCR of exon 3 (MBD4_Ex3-F: 5'-GCTGAAAGTGAACCTGTTGC-3' and MBD4_Ex3-R: 5'-TGTGTTCTGAGTCTTTGGCTG-3') was performed using Phusion Hot Start II High-Fidelity DNA Polymerase (30 seconds at 98°C, 32 cycles of 10 seconds at 98°C, 30 seconds at 60°C and 15 seconds at 72°C, and a final 10 minutes at 72°C). Phusion Hot Start II High-Fidelity DNA polymerase was used as its enhanced proof-reading ability was preferable for amplifying the poly-A stretch in exon 3. The PCR products were used directly for Sanger sequencing and analysed on Applied Biosystems 3730 or 3730*xl* capillary sequencers (Thermo Fisher Scientific). Sanger sequencing results were analysed using SeqMan (DNASTAR) and by comparison to the human genome with BLAT (https://genome.ucsc.edu/cgi-bin/hgBlat). PCR products were also cloned using Zero Blunt TOPO PCR Cloning Kit and One Shot Top10 Electrocomp *E. coli* (Thermo Fisher Scientific) prior to Sanger sequencing, to enable resolution of the mixed signal produced by frameshift variants. Plasmid DNA was obtained using PureLink HiPure Plasmid Miniprep Kit (Thermo Fisher Scientific).

Total RNA from WEHI-AML-1 and WEHI-AML-2 at the time of first remission was used to make cDNA with SuperScript III Reverse Transcriptase (Thermo Fisher Scientific). Primers flanking *MBD4* exons 3 to 8 were designed using Primer3 (MBD4_Phasing1-F 5'-GAGACCCTCAGTGTGACCAG-3' with MBD4_Phasing1-R 5'-GCTGGAAAGGTGGTTGGTTG-3', and MBD4_Phasing2-F 5'-GCTGAAAGTGAACCTGTTGC-3' with MBD4_Phasing2-R 5'-GCTGGAAAGGTGGTTGGTTG-3'). PCR was performed using Phusion Hot Start II High-Fidelity DNA Polymerase (30 seconds at 98°C, 30 cycles of 10 seconds at 98°C, 30 seconds at 60°C and 2 minutes at 72°C, and a final 10 minutes at 72°C). The PCR products were cloned, plasmids were purified and Sanger sequenced, as above.

Alignment and variant detection

DNA sequencing data from WEHI-AML-1 and WEHI-AML-2 (exome, genome and small amplicons) were aligned to the human genome (hg19) using bwa-mem v0.7.10-r789[1]. Additional sequencing data were sourced from the NIH-NCI Genomic Data Commons (GDC) Data Portal [phs001027 and phs000178] through an authorised access request. Aligned

sequencing data (BAM files aligned to hg38) was downloaded from GDC, together with variant calls. Preliminary variant detection of WEHI-AML-1, WEHI-AML-2, EMCAML-1, the EMC cohort [phs001027][2] and all TCGA-AML samples [phs000178][3] was performed with samtools version 0.1.19-44428cd, followed by Varscan v2.3.6[4]. These preliminary calls were refined with superFreq 0.9.15, which also provided clonality estimates for copy number and somatic SNV calls (https://github.com/ChristofferFlensburg/superFreq) and clonal relationships were visualized[5]. Somatic mutation calls for SW1783 were available through the Catalogue of Somatic Mutations in Cancer (v81)[6]. The MutationalPatterns R Package was used to assess the trimer base context for somatic mutations[7]. Additional variants in a target gene set in EMC-AML-1 were identified using a site-specific error model detailed below.

RNA sequencing alignment and analysis

For WEHI-AML-1 and WEHI-AML-2, total RNA was extracted using TRIzol (Thermo Fisher Scientific). One microgram of RNA, except for the WEHI-AML-1 diagnosis time point where only 170ng was available, was used to generate RNA libraries using the TruSeq RNA Sample Preparation Kit v2 (Illumina). The fragmentation protocol was modified to generate fragments with a median insert size of 200 bp. The libraries were sequenced on a HiSeq2500 and data aligned to hg19 with Tophat v2.0.12[8]. Aligned sequencing data for TCGA samples (BAM files aligned to hg38) was downloaded from GDC. Aligned RNA-seq data was counted over genes with featureCounts[9]. Reads spanning splice sites were used to calculate the proportion of reads involved in canonical and non-canonical splicing.

Assessment of the relative mutation rate and its association with genetic and epigenetic features

The incidence of CG>TG mutations was assessed in different genomic features, defined by the sequence context (NCG trimer, NCGN tetramer) or functional classification (CG island, promoter, promoter flanking, exon or intron). CG island annotation was obtained at the University of California, Santa Cruz (UCSC) genome browser[10]. Promoter and promoter flanking annotations were obtained from the Ensembl website through biomaRt[11]. Intronic and exonic regions were determined from the Ensembl v75database[12] (GRCh37).

As the mutations occurred almost exclusively in a CG context, the rate of CG>TG mutations per CG was calculated for each genomic feature. Each CG provides two mutational substrates and was therefore counted twice. Whole genome bisulfite (WGBS) data from CD34+ specimens from 5 healthy individuals was used to estimate methylation status for CG sites[13]. Due to the variable coverage in WGBS, the analysis was restricted to CG sites covered in at least 1 of the WGBS CD34+ samples, and a weighted average was used to estimate methylation status based on the coverage in each sample. The restriction to CG sites with coverage in WGBS ensured that a methylation call was available for all sites and omitted difficult to align regions.

A relative mutation rate (RMR) was calculated for each feature, either corrected for CG abundance (RMR–CG) or for 5mC abundance (RMR–5mC), using RoaMeR (https://github.com/MathijsSanders/RoaMeR). RMR is calculated in a way that is analogous to the fragments per kilobase of exon per million reads mapped statistic (FPKM). RMR-CG reflects the number of mutations expected in 1 Mb of CG sites in a particular genomic feature if 1000 somatic mutations were randomly selected from the total mutation load:

$$\widetilde{RMR}_{ij}^{CG} = \frac{\mu_{ij}}{\left(\frac{\alpha_i}{10^6}\right) \cdot \left(\frac{\tau_j}{10^3}\right)}$$

Where μ_{ij} represents the CG>TG mutation count for feature i within sample j, α_i represents the total number of CGs present in feature i, and τ_j is the total CG>TG mutation burden in sample j. The RMR-5mC replaces α_i with the weighted average methylation methylation level for a feature:

$$\widetilde{RMR}_{ij}^{5mC} = \frac{\mu_{ij}}{\left(\frac{\omega_i}{10^6}\right) \cdot \left(\frac{\tau_j}{10^3}\right)}$$

Where ω_i represents the sum of the weighted average methylation level for all CG sites in feature i across the CD34+ WGBS cases:

$$\bar{\beta} = \frac{\sum_{j=1}^{5} cov_j \cdot \beta_j}{\sum_{j=1}^{5} cov_j} \qquad \omega_i = \sum_{CG \in i} \bar{\beta}_{CG}$$

Where cov_j reflects the local coverage and β_j is the methylation level for this CG, both for sample j. For a fully methylated feature the CG-abundance (α_i) is equal to the 5mCG-abundance (ω_i), meaning the RMR-CG and RMR-5mC are equivalent.

RMRs were calculated for each genomic feature, then further stratified based on replication timing, transcriptional strand or expression level.

Replication timing: A conserved DNA replication timing profile was generated from 14 Repli-seq data sets from ENCODE[14] downloaded through the UCSC genome browser[10] (hg19). The data comprises genome-wide wavelet-smoothed values per 1-kb bin for 14 cell lines (BG02ES, BJ, GM06990, GM12801, GM12812, GM12813, GM12878, Hela-S3, HepG2, HUVEC, IMR90, K562, MCF-7 and NHEK). The median value over all cell lines was calculated per 1-kb bin, and then the median of these values was calculated for contiguous 10-kb blocks. Each 10-kb block was distributed among 4 equally-sized domains, designated the latest (< 29), late (≥ 29 & < 47), early (≥ 47 & < 63) and earliest (≥ 63) replication domains.

Transcriptional strand and expression level: Transcriptional strand bias analysis was performed by determining the template and non-template strands per gene as reported in Ensembl v75[12]. Overlapping gene bodies on the same strand were merged while overlapping regions defined by genes located on opposite strands were excluded. The number of CG>TG mutations on the template and non-template strands was counted genome-wide and corrected for the total CG-abundance within gene bodies. Transcriptional bias was assessed by binning genes into categories based on expression. The average FPKM value per gene was calculated from RNA-seq data available from WEHI-AML-1 and WEHI-AML-2. Genes with an average FPKM value ≤ 0.5 were considered to have very low expression (n=7464) and were allocated to the none-gene expression bin. The remaining genes were ordered by average FPKM value and distributed among 4 equally-sized expression bins (n=4509 per bin), designated lowest-, low-, high- and highest- gene expression bins. A high correlation between gene expression level and replication domain was noted.
RMR-5mC calculation for gnomAD: The Genome Aggregation Database[15] (gnomAD) was provided by the Broad Institute and downloaded from Google Cloud Storage (release-170228). Rare CG>TG germline polymorphisms with a minor allele frequency between 0.0001 and 0.001 were considered for further analysis. SNPs were used in place of somatic mutations to calculate an RMR-5mC for all NCGN tetramers, either globally or for each replication domain.

Variant validation using a small amplicon panel

Primers flanking variants of interest were designed using Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/). Whole genome amplification (WGA) was performed on DNA from WEHI-AML-1 and WEHI-AML-2 using the REPLI-g Mini/Midi Kit (Qiagen). Amplified DNA was purified using Agencourt AMPure XP (Beckman Coulter) before use in PCR, which was performed using Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific) (30 seconds at 98°C, 31 cycles of 10 seconds at 98°C, 20 seconds at 60°C and 15 seconds at 72°C, and a final 5 minutes at 72°C). PCR products were analysed on an agarose gel, pooled and purified using Agencourt AMPure XP. A second PCR was performed using indexed forward and reverse primers that introduce the P5 and P7 sequences (30 seconds at 98°C, 24 cycles of 10 seconds at 98°C, 20 seconds at 60°C and 15 seconds at 72°C, and a final 5 minutes at 72°C). PCR products were analysed on an Agilent 2200 Tapestation D1000 ScreenTape Assay (Agilent Technologies), purified using Agencourt AMPure XP, quantified and diluted before sequencing on a MiSeq (Illumina, San Diego, CA, USA).

Mutation analysis on bone marrow smears and isolated single cells

EMC-AML-1 DNA was isolated from May-Grunwald Giemsa (MGG) stained bone marrow smears. Material was removed from glass slides and dissolved in RLT plus lysis buffer (Qiagen). Genomic DNA was isolated with the QIAsymphony. A Becton Dickson FACS

Aria II was used to sort intact and living cells into 96-well plates at one cell per well. WGA was performed on each cell using the REPLI-g Single Cell DNA Library Kit (Qiagen). Selected samples were analysed using the Illumina TruSight Myeloid Sequencing Panel (Illumina). Sanger sequencing was also performed to assess mutations in *DNMT3A*, with the following PCR primers: R635-F: 5'-CAGGGTGTGTGGGTCTAGGA-3' and R635-R: 5'-AAGCTTCCCCTTTGGGATAA-3'; R688-F: 5'-CAGGGAGATGGCTCCAAGTA-3' and R688-R: 5'-TTTGCCCTTTACCCTCTCAA-3'; R882/A884-F: 5'-AGGAGTTGGTGGGTGTGAGT-3' and R882/A884-R: 5'-TCTCCATCCTCATGTTCTTGG-3'; 3'UTR-F: 5'-TTCTAGAAGCCGCTGTTACCTC-3' and 3'UTR-R: 5'-CCTCATCTAGCCCCCTTTTT-3'. PCR was performed using Taq DNA polymerase (Thermo Fisher Scientific) (4 minutes at 94°C, 35 cycles of 1 minute at 94°C, 1 minute at 60°C and 1 minute at 72°C, and a final 7 minutes at 72°C). PCR products were purified using a Millipore Microscreen™ PCR cleanup plate (Merck) and sequenced on an Applied Biosystems 3130xl Genetic Analyzer (Thermo Fisher Scientific).

Deep variant detection in serial samples from EMC-AML-1

We constructed a site-specific error model that models the unique site-and-variant specific noise profile from a large set of complete remission samples (n=480) from patients who never carried the variant of interest in the leukemia[16]. Variant allele frequencies (VAFs) for each potential SNV within the target gene set was calculated across all serial samples from EMC-AML-1 and the set of remission samples. Quantile normalization of the calculated VAFs was performed to mitigate the effect of qualitative differences amongst samples. The VAFs for variants of interest in the serial samples of EMC-AML-1 was compared to the site-and-variant specific noise profile constructed from the set of remission samples by the modified Thompson-Tau test. A p-value < 0.0001 was considered statistically significant, implying that the variants was present at that time point, and only variants present at 3 or more time points were considered detected. The strict criteria imposed limit the number of erroneously detected variants.

Assessment of positive selection for somatic mutations in EMC-AML-1

Positive selection for individual genes was assessed by an unmodified version of dNdScv[17] for the whole exome sequencing data, while a modified substitution model (12 substitution classes for non-CG mutations: A>C, A>G, A>T, C>A, C>G, C>T, G>A, G>C, G>T, T>A, T>C and T>G + the 4 different CG-contexts: ACG>ATG, CCG>CTG, GCG>GTG and TCG>TTG) within dNdScv was used for the variants detected for the single cell derived colonies (SCDCs) expanded from the autologous hematopoietic stem cell transplant material of EMC-AML-1. Any duplicate variants detected in the SCDCs were omitted from further analysis as it could not be guaranteed that they were independent events.

Site-directed mutagenesis and cloning

Full length *MBD4* cDNA was PCR amplified from a pReceiver-B13 vector (GeneCopoeia) and cloned into pET28-MHL expression vector (GenBank accession EF456735, Addgene) by applying BD-BioScience In-Fusion enzyme-mediated directional recombination between complementary 15 nucleotide DNA sequences at the end of the PCR products, using the following primers for MBD4: sense 5'-TTGTATTTCCAGGGCGGCACGACTGGGCTGGAGAGTCT and anti-sense 5'-TTGTATTTCCAGGGCGGCACGACTGGGCTGGAGAGTCT-3'. The truncated form of MBD4, residues 430-580, in the pET28 backbone was sourced from Addgene.

QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) was used to generate the MBD4 mutants, including the catalytically inactive mutant D560A[18]. The primer sequences were as follows, with the site of deletion or mutation underlined: MBD4 H567 deletion: sense 5'-CCCTGAAGACCACAAATTAAATAAATAT___GACTGGCTTTGGGAA-3' and anti-sense 5'-TTCCCAAAGCCAGTC___ATATTTATTTAATTTGTGGTCTTCAGGG-3'; MBD4 D560A: sense 5'-GAAGCAGGTGCACCCTGAAGCCCACAAATTAAATAAATATCA-3' and anti-sense 5'-TGATATTTATTTAATTTGTGGGCTTCAGGGTGCACCTGCTTC-3'; Successful mutagenesis was confirmed by enzymatic digestion and DNA sequencing.

Expression and purification of recombinant MBD4

The pET28-MHL containing hexahistidine-tagged MBD4 full length and mutants, MBD4 residues 430-580 and mutants were expressed in *Escherichia coli* BL21(DE3) Gold and BL21(DE3) pLysS cells (Thermo Fisher Scientific). A single BL21(DE3) colony containing the pET28 expression vector with the desired insert was cultured for 16 hours at 37°C in Luria Broth (LB) medium supplemented with 50 µg/ml kanamycin. The culture was diluted 25 times in LB medium without selection marker and grown at 37°C until it reached an OD600 of 0.5 to 0.8. Cells were induced by adding 0.5 mM isopropyl b-D-1-thiogalactopyranoside and incubated for 4 hours at 37°C. Cell cultures were centrifuged at 38,000 x g for 30 minutes at 4°C and the cell pellets resuspended in lysis buffer containing 20 mM sodium phosphate, 500 mM NaCl, 10 mM imidazole (pH 7.4), 1 mg/ml lysozyme, 200 µg/ml DNAse, and 1x SIGMA*FAST* Protease Inhibitor, EDTA-Free (Sigma-Aldrich), before incubation on ice for 30 minutes. The cells were sonicated on ice for 6 minutes at amplitude 60%, using 10 second bursts followed by a 20 second reprieve with a Branson Digital Sonifier. The lysate was clarified by centrifugation at 38,000 x g for 30 minutes at 4°C. The hexahistidine-tagged proteins were isolated from the crude lysate using Ni-NTA Superflow (Qiagen) and a 2.5 x 10cm Econo-chromatography column (Bio-Rad). The Ni-NTA resin was washed twice with cold washing buffer containing 20 mM sodium phosphate, 500 mM NaCl and 10 mM imidazole (pH 7.4) to remove non-specific interacting proteins. Hexahistidine-tagged proteins were eluted from the Ni-NTA group on the matrix with cold elution buffer, containing 20 mM sodium phosphate, 500 mM NaCl and 500 mM imidazole (pH 7.4). Directly after elution into buffer containing 500 mM imidazole, the protein suspensions were dialyzed in a Slide-

A-Lyzer G2 Dialysis Cassette (gamma irradiated, 10K MWCO) (Thermo Fisher Scientific), for 2 hours at 4°C against 300x volume dialysis buffer containing 50 mM Tris-HCl and 150 mM NaCl (pH 7.6). The dialysis buffer was refreshed twice, with further incubation at 4°C for 2 hours and 16 hours. Proteins were quantified using Qubit protein assay kit and Qubit 3.0 fluorometer (Thermo Fisher Scientific). Proteins were verified by SDS-PAGE using a NuPage Novex 4-12% Bis-Tris Protein Gel run in a Bis-Tris XCell SureLock™ Mini-Cell system (Thermo Fisher Scientific) with 1x MOPS at 200V for 90 minutes. Blots were incubated in blocking buffer containing 5% BSA, 0.1% Tween and 1xPBS with the appropriate antibodies: α-His H-15 (sc-803, Santa Cruz Biotechnology) and α-MBD4 (ab12187, Abcam). Proteins were visualized using a Li-Cor Odyssey 3.0.

SUPPLEMENTARY TABLES:

Table S1: Clinical sample description

WEHI-AML-1 with AML with myelodysplasia-related changes (MDS-associated cytogenetic abnormality, monosomy 7, WHO ICD 9895/3)

Sample	Time from diagnosis days (months)	Clinical time point	Tissue	Blast % morphology	Pathology summary	Analysis
1	0 (0)	Diagnosis	BM	21	Diagnosis of AML	WGS, WES, RNA-seq, RRBS
2	35 (1)	Post induction	BM	3	In morphologic and cytogenetic remission	WGS, WES, RNA-seq, RRBS
3	69 (2)	Post consolidation cycle 1	BM	2	In morphologic and cytogenetic remission	WES, RNA-seq, RRBS
4	115 (3)	Post consolidation cycle 2	BM	4	In morphologic and cytogenetic remission	WES, RNA-seq
5	148 (4)	Pre allogeneic HSCT	BM	13	Relapsed AML	WES, RNA-seq
6	202 (6)	Post allogeneic HSCT	BM	3.5	In morphologic and cytogenetic remission	WES
7	245 (8)	Relapse	BM	36	Relapsed AML	WES, RNA-seq, RRBS

WEHI-AML-2 with AML with myelodysplasia-related changes (MDS-associated cytogenetic abnormality, monosomy 7, WHO ICD 9895/3)

Sample	Time from diagnosis days (months)	Clinical time point	Tissue	Blast % morphology	Pathology summary	Analysis
1	-1410 (-46)	HSCT donor to WEHI-AML-1	PB	0	Normal full blood count and blood film	WES
2	0	Diagnosis	BM	22	Diagnosis of AML	WGS, WES, RNA-seq, RRBS
3	33 (1)	Post induction	BM	1.5	In morphologic and cytogenetic remission	WGS, WES, RNA-seq, RRBS

EMC-AML-1 with AML, not otherwise specified (acute monocytic leukaemia, WHO ICD 9891/3)

Sample	Time from diagnosis days (months)	Clinical time point	Tissue	Blast % morphology	Pathology summary	Analysis
1	0	Diagnosis	BM	84	Diagnosis of AML	WES, RNA-Seq, Panel, eRRBS
2	27 (1)	Induction cycle 1 (Day 25)	BM	32	Persistent AML	Panel
3	48 (1)	Induction cycle 2 (Day 17)	BM	3	Hypocellular bone marrow	Panel
4	55 (2)	Induction cycle 2 (Day 25)	BM	0.8	In morphologic remission	Panel
5	64 (2)	Complete remission	BM	0.4	In morphologic remission	Panel
6	77 (3)	Pre autologous HSCT	BM	0.8	In morphologic remission	Panel
7	125 (4)	Post autologous HSCT	BM	0.6	In morphologic remission	Panel
8	169 (6)	Complete remission	BM	1.2	In morphologic remission	Panel
9	253 (8)	Complete remission	BM	1.8	In morphologic remission	Panel
10	351 (12)	Complete remission	BM	0.4	In morphologic remission	Panel
11	496 (16)	Complete remission	BM	1.6	In morphologic remission	Panel
12	841 (28)	Relapse timepoint 1	BM	16.2	Relapsed AML	Panel, Single cells
13	853 (28)	Relapse timepoint 2	BM	7.6	Relapsed AML	WES, RNA-Seq, Panel
14	877 (29)	Complete remission	BM	0.8	In morphologic remission	Panel
15	890 (29)	Complete remission	BM	2.4	In morphologic remission	Panel

Description of clinical samples and molecular profiling approaches applied to each. HSCT: Hematopoietic stem cell transplant, BM: Bone marrow, PB: Peripheral blood, AML: Acute myeloid leukemia, MDS: Myelodysplastic syndrome, WGS: Whole genome sequencing, WES: Whole exome sequencing, RNA-seq: RNA sequencing, RRBS: Reduced representation bisulfite sequencing, eRRBS: Enhanced reduced representation bisulfite sequencing, Panel: Illumina TruSight Myeloid Sequencing Panel.

Table S2: Candidate germline loss-of-function variants in *MBD4*.

code	cancer type	chr	pos	ref	var	normal		cancer		tumour content	dbSNP	gene	transcript	annotation	type
						ref	var	ref	var						
EMC-AML-1	Acute Myeloid Leukaemia	3	129150385	ATG	-	1	11	2	12	90%	rs775848563	MBD4	NM_003925	p.His567del	inframe_deletion
SW1783	Glioma (cell line)	3	129150399	A	T						rs200758755	MBD4	NM_003925	p.Leu563Ter	stop_gained
TCGA-GBM-2	Glioblastoma Multiforme	3	129150399	A	T	105	80	127	108	90%	rs200758755	MBD4	NM_003925	p.Leu563Ter	stop_gained
TCGA-OV-2	Ovarian Serous Cystadenocarcinoma	3	129150399	A	T	74	53	48	232	90%	rs200758755	MBD4	NM_003925	p.Leu563Ter	stop_gained
TCGA-UCEC-1	Uterine Corpus Endometrial Carcinoma	3	129151375	G	A	19	26	26	57	90%		MBD4	NM_003925	p.Arg546Ter	stop_gained
TCGA-OV-1	Serous Ovarian Cancer	3	129151450	C	A	70	74	109	215	90%	rs778697654	MBD4	NM_003925	c.1562-1G>T	splice_acceptor
WEHI-AML-1	Acute Myeloid Leukaemia	3	129151450	C	A	26	25	41	32	80%	rs778697654	MBD4	NM_003925	c.1562-1G>T	splice_acceptor
WEHI-AML-2	Acute Myeloid Leukaemia	3	129151450	C	A	42	28	24	38	80%	rs778697654	MBD4	NM_003925	c.1562-1G>T	splice_acceptor
TCGA-UVM-1	Uveal Melanoma	3	129151450	C	A	43	43	1	32	90%	rs778697654	MBD4	NM_003925	c.1562-1G>T	splice_acceptor
TCGA-THYM-1	Thymoma	3	129152813	G	A	14	28	35	31	<10%		MBD4	NM_003925	p.Arg431Ter	stop_gained
WEHI-AML-1	Acute Myeloid Leukaemia	3	129155547	C	CT	84	45	133	64	80%	rs558765093	MBD4	NM_003925	p.Val314ArgfsTer13	frameshift
WEHI-AML-2	Acute Myeloid Leukaemia	3	129155547	C	CT	126	84	117	69	80%	rs558765093	MBD4	NM_003925	p.Val314ArgfsTer13	frameshift
TCGA-STAD-1	Stomach adenocarcinoma	3	129155723	G	C	39	38	98	85	70%	rs749739092	MBD4	NM_003925	p.Ser255Ter	stop_gained
TCGA-GBM-1	Glioblastoma Multiforme	3	129156562	C	T	51	67	15	55	90%	rs552296498	MBD4	NM_003925	c.335+1G>A	splice_donor
TCGA-CHOL-1	Cholangiocarcinoma	3	129156562	C	T	13	15	31	22	80%	rs552296498	MBD4	NM_003925	c.335+1G>A	splice_donor

Code: Sample identifiers, TCGA samples were re-coded. Genomic position in hg19 (chr: Chromosome. pos: Position. ref: reference. var: variant. dbSNP: rs identifier). Coverage values derived from exome data for the cancers and matched normal samples. An estimate of tumour purity was calculated based on copy number alterations and somatic SNV frequency. Annotation: Gene, transcript, annotation and variant type.

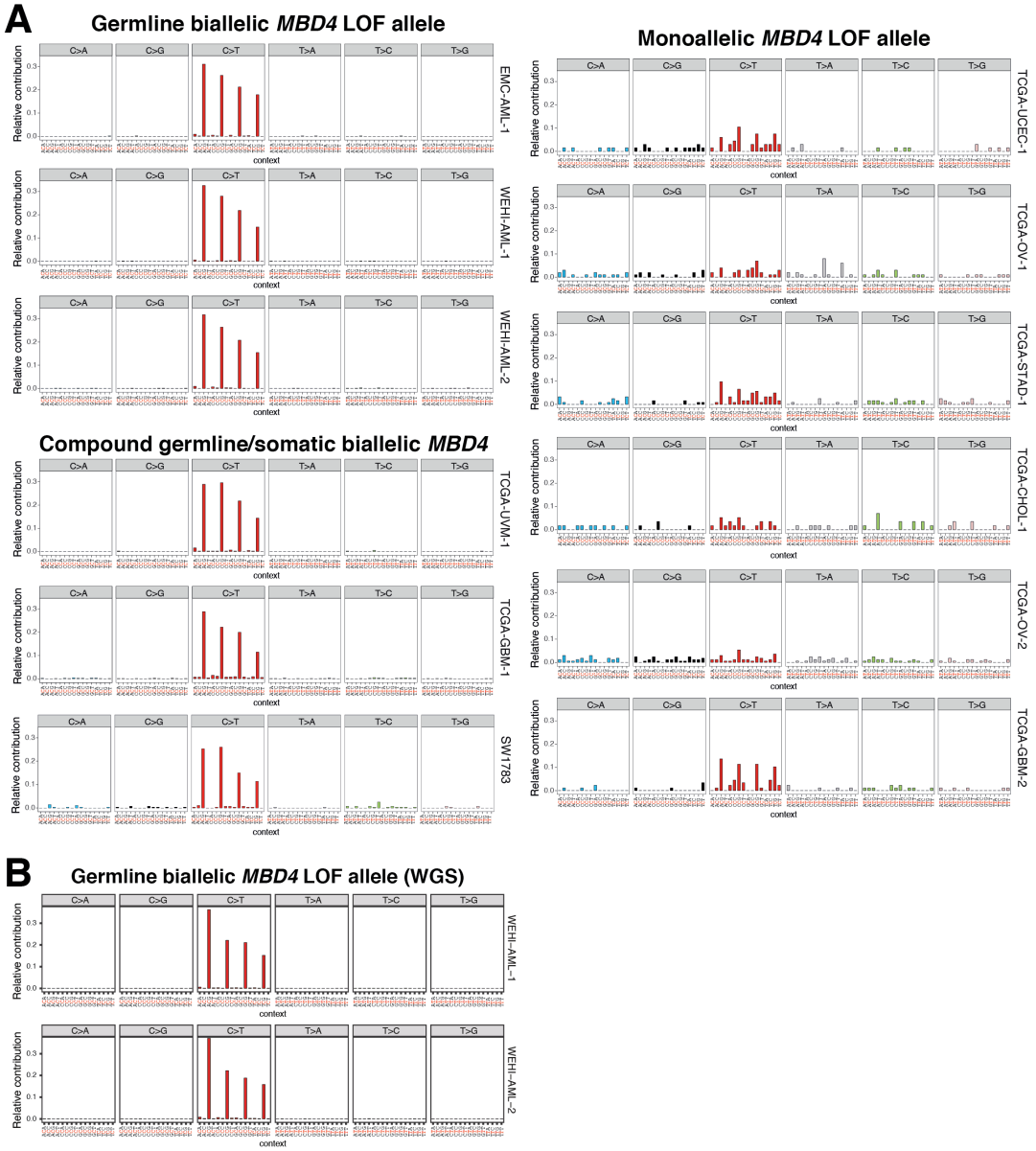


Figure S1. Mutation profiles in cancers with germline loss-of-function variants in *MBD4*. (**A**) Trimer mutation profiles determined from diagnostic AML exome data (EMCAML-1, WEHI-AML-1 and WEHI-AML-2), cancers from TCGA with germline loss-of-function variants and the SW1783 cell line (COSMIC). Samples with biallelic MBD4 inactivation are shown left (separated into germline biallelic and compound germline/somatic), and cases with monoallelic MBD4 inactivation are shown at right. TCGA-THYM-1 was excluded due to low purity. There was a significant enrichment of mutations in the ACG context compared to TCG in MBD4-deficient exomes (p=0.007937, Mann-Whitney U test). (**B**) Trimer context for somatic mutations identified in whole genome sequencing (WGS) from WEHI-AML-1 (n=14,313 mutations) and WEHI-AML-2 (n=15,543 mutations).

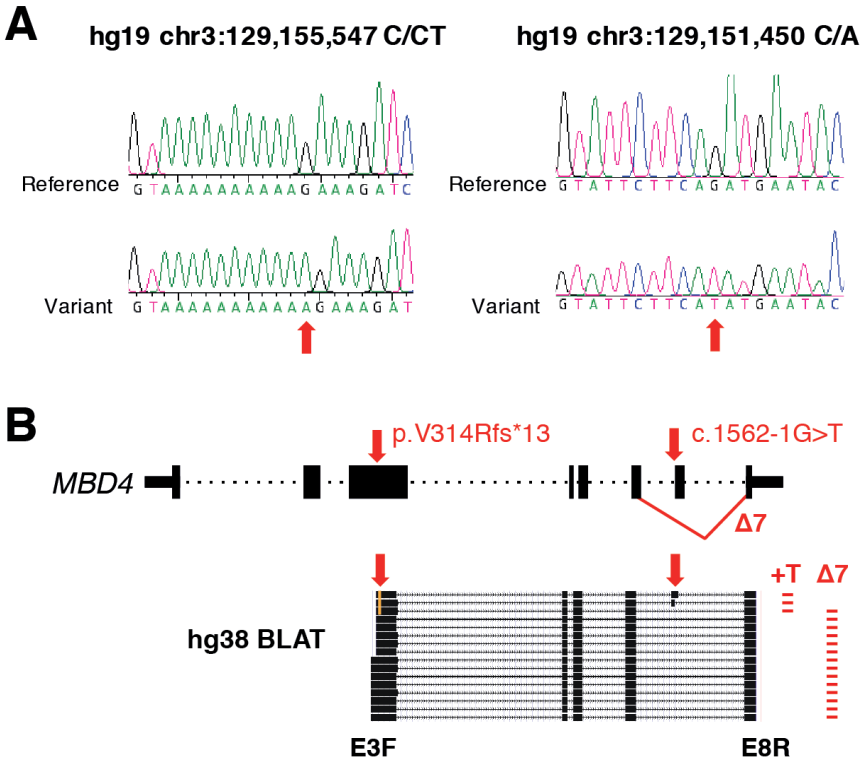


Figure S2. Confirmation and phasing of loss-of-function *MBD4* variants present in the genomes of the WEHI-AML cases. (**A**) Confirmation of the loss-of-function *MBD4* variants present in the genomes of WEHI-AML-1 and WEHI-AML-2. Sanger sequencing traces were generated from cloned PCR products after amplification from DNA (top). The reverse complement of the genome sequence is shown, which represents the open reading frame 5′-3′. (**B**) A schematic of the *MBD4* gene is shown. PCR was performed on cDNA to generate fragments that covered both variant positions, using primers in exon 3 and exon 8. Products cloned into TOPO were sequenced and BLAT results are shown to demonstrate transcript structure. Each product was assessed for the presence of the +T insertion (+T) and aberrant splicing of exon 7 (Δ7). The variants were mutually exclusive in fifteen out of sixteen clones. In one clone the variants were coincident, likely due to instability in the poly-T stretch (during reverse transcription or subsequent PCR amplification).

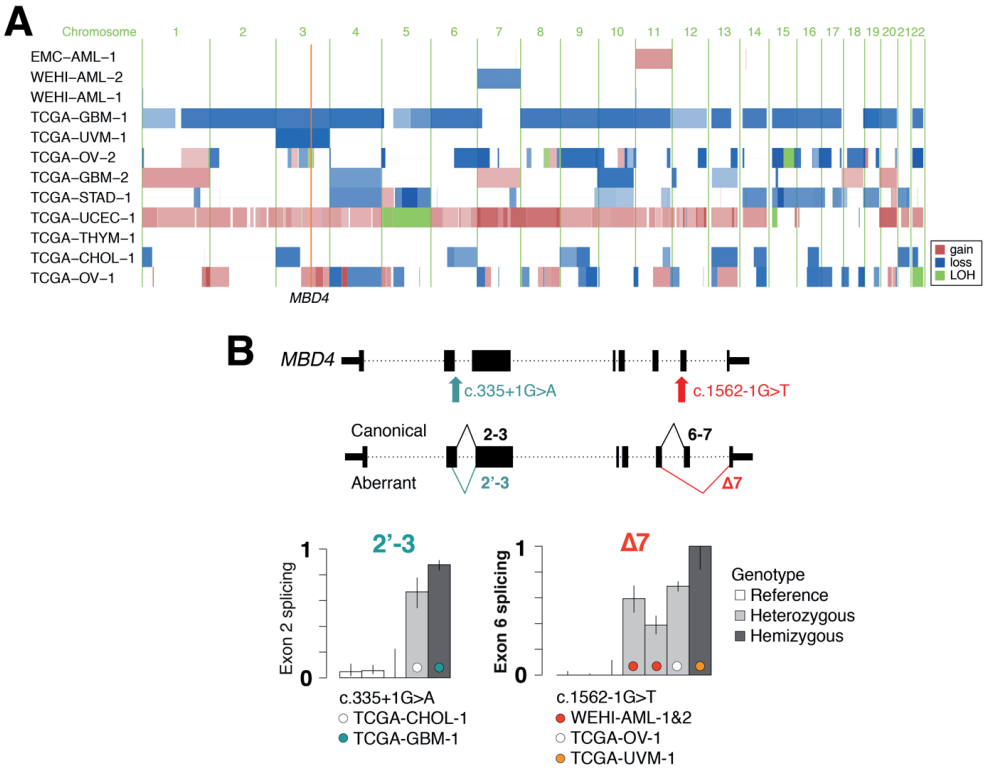


Figure S3. Copy number changes and RNA splicing defects in cancers with germline loss-of-function variants in *MBD4.* (**A**) Copy number profiles were generated from cancer exome data by tracking shifts in allele frequencies at heterozygous germline SNPs and coverage (summarised at the gene level). Sample codes are shown; gain is represented in red, loss in blue, and loss of heterozygosity in green. The position of *MBD4* is highlighted. The intensity of the colour is proportional to the clonality, with darker colours representing higher clonality estimates. (**B**) A schematic of the *MBD4* gene is shown at top together with the position of two candidate loss-of-function variants that impact splice sites. The genotype status is indicated, which is informed based on the copy number calls. Canonical and aberrant splice products are shown below, which were quantified with RNA-seq by counting reads that span splice junctions. Three unrelated control samples were selected to serve as controls for canonical splicing. Each bar represents data from a single sample. The error bars reflect the level of coverage and are set at $1/\sqrt{N}$, where N is the total number of spliced reads.

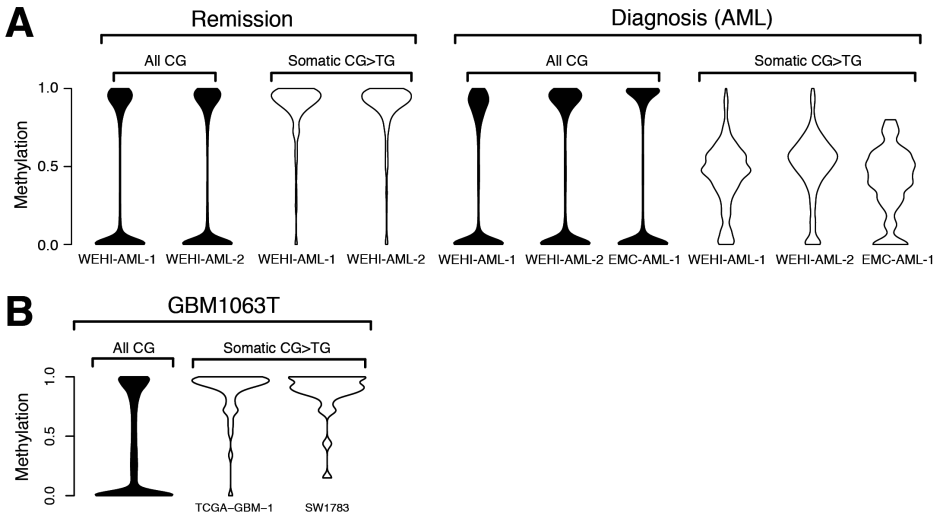


Figure S4. Methylation status at somatic mutation positions in AML and glioblastoma. (**A**) Methylation levels were assessed at CG dinucleotides following bisulfite conversion, either globally (all CG) or at sites of somatic mutations. Normal bone marrow was taken at remission. Assessment of the mutated sites in each AML directly revealed ~50% methylation, indicating the non-mutated CG site on the alternate allele was methylated. (**B**) Methylation status was extracted for CG sites from RRBS data from an unrelated glioblastoma (GBM1063T)[19], either at all CG sites (All CG) or at sites of somatic CG>TG mutations in TCGA-GBM-1 or the cell line SW1783. Sites with mutations were typically fully methylated in the control sample.

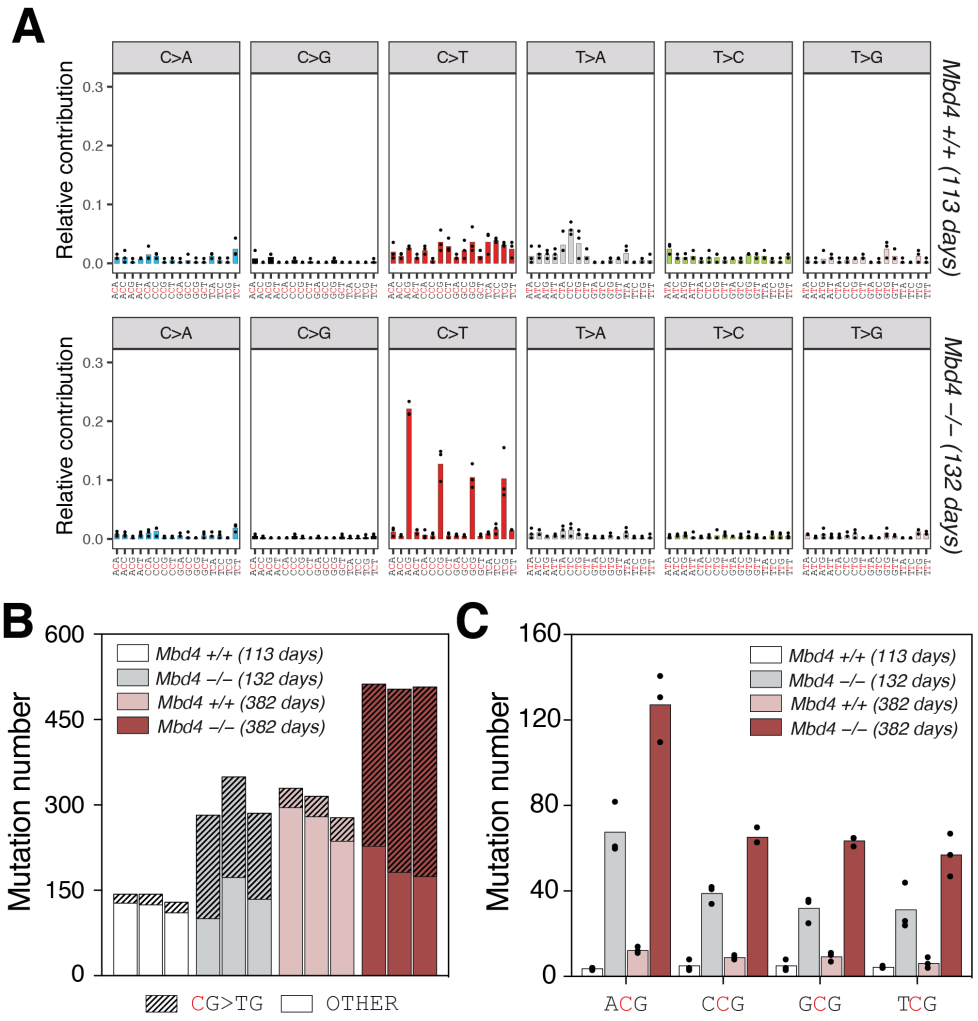


Figure S5. Mutation profiles in myeloid progenitors from *Mbd4* knockout mice.
(**A**) Mutation profiles determined from genomes from myeloid progenitors from *Mbd4-/-* mice[20] and controls (*Mbd4+/+*). The proportion of mutations in each trimer context is shown; individual values are plotted (3 colonies per genotype) and the bar shows the mean. (**B**) The number of base substitutions identified in myeloid colonies from *Mbd4-/-* and *Mbd4+/+* mice at different ages, separated into CG>TG or other (3 colonies per mouse). Analysis was restricted to high quality somatic variants on autosomes. There was a 9.4-fold elevation in CG>TG mutations in *Mbd4-/-* colonies at 4 months (mean 170±16.6 compared to 18±1.7). A similar increase was observed after a year, but with a higher number of CG>TG mutations (8.5-fold increase, mean 313±25.1 compared to 37±3.6). The increase in CG>TG mutations was significant at both timepoints (p=0.0037 and p=0.0023, respectively, Welch's t-test). **(C)** The NCG trimer context for CG>TG mutations identified in myeloid colonies from *Mbd4-/-* and *Mbd4+/+* mice. There were significantly more mutations in the ACG context compared to TCG in the *Mbd4-/-* animals (p=0.019 at 4 months and p=0.0051 for the older mice, Welch's t-test).

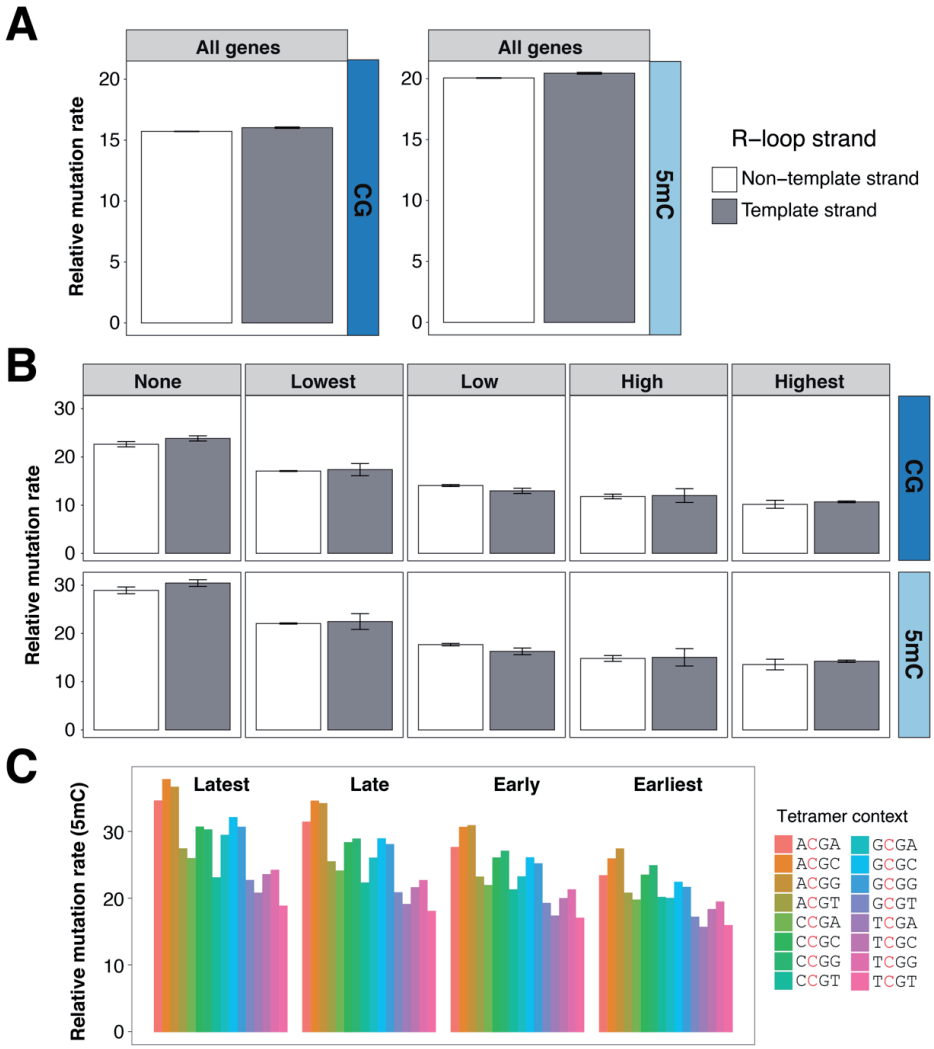


Figure S6. Assessment of genetic and epigenetic factors that impact mutation rate.
(**A**) Mutations from WEHI-AML-1 and WEHI-AML-2 were tested for association with transcriptional strand. The relative mutation rate was calculated per megabase of CG sites (dark blue) or per megabase of 5mCG sites (light blue). The values are corrected to account for total mutation load (see Methods). Individual values are plotted (n=2) and the bar shows the mean. (**B**) Genes were divided into bins based on their expression (see Methods). The relative mutation rate was calculated per bin based on CG or 5mCG abundance. Individual values are plotted (n=2) and the bar shows the mean. There is considerable correlation between replication domain and expression level. (**C**) The NCGN tetramer context was determined for rare germline CG>TG SNPs from the gnomAD database15. The dataset was split based on replication timing, with a quarter of the genome ascribed to each bin. Note that the correction for methylation status uses data derived from normal CD34+ blood cells.

A

Cell	BCOR R976X	BCORL1 R609X	DNMT3A R635W	DNMT3A R688C	DNMT3A R882C	DNMT3A A884V	EZH2 R690H	IDH1 R132C	Clone
SC1	-	-	-/-	+/-	+/-	-/-	-	+	
SC2	-	-			-	-	-	-	
SC3	-	+	+/-	-/-	-/-	+/-	+/-	-/-	
SC4	-	-	-/-	-/-	-/-	-/-	-/-	-/-	
SC5									
SC6	+	-	-/-	+/-	+/-	-/-	-/-	+/-	
SC7									
SC8	-	-	-/-	-/-	-/-	-/-		-/-	
SC9	+	-	-/-	+/-			-		
SC10	-	-	-/-	-/-	G>T	-/-	-	-/-	
SC11	-	-			+	-		-/-	
SC12	+	-	-/-	+/-	+/-	-/-	-/-	+	
SC13	-	-	-/-	-/-	-/-	-/-	-/-	-/-	
SC14	-	-	-	-			-	+	
SC15	-	-	-/-	+/-	+/-	-/-	-/-		
SC16		-			+/-	-/-	-/-	+/-	
SC17		-	-	+			-	+	
SC18					+/-	-	-	+/-	
SC19	+		-/-	+			-/-	+/-	
SC20	+	-	-/-	+/-	+/-	-/-	-/-	+/-	
SC21	+	-	-/-	+/-	+/-	-/-	-/-	+/-	
SC22	-	-	-/-	-/-	-/-	-/-	-/-	-/-	
SC23	-	-	-/-	-/-	-/-	-/-	-/-	-/-	
SC24	+		-	+			-/-	+/-	
SC25	+	-			+/-	-/-	-/-	+/-	
SC26	-	-			-	-	-	-/-	
SC27	-	+			-/-	+/-	+/-	-/-	
SC28	+				+/-	-	-/-	+/-	
SC29	+	-	-/-	+/-	+/-	-/-	-/-	+	
SC30									
SC31	+	-			+	-		+/-	
SC32	+	-	-/-	+/-	+/-	-/-	-/-	+/-	
SC33	-		-	-			-/-		
SC34	+		-/-	+/-	+/-	-/-	-/-	+/-	

B

Cell	DNMT3A R635W	DNMT3A R688C	DNMT3A R882C	DNMT3A A884V	DNMT3A 3' UTR
SC1	-	+/-	+/-	-	+/-
SC2	-	-	-	-	-
SC3	-	-	-	+/-	-
SC4	-	-	-	-	-
SC5					
SC6	-	+	+	-	-
SC7	-	-	-	-	-
SC8	-	-	-	-	-
SC9	-	+/-	-	-	+/-
SC10	-	-	G>T	-	-
SC11	-	-	+	-	+/-
SC12	-	+/-	+/-	-	+/-
SC13	-	-	-	-	-
SC14	-	-	-	-	+
SC15	-	+/-	+/-	-	+/-
SC16	-	-	+/-	-	+/-
SC17	-	+	-	-	+
SC18	-	-	+/-	-	+/-
SC19	-	+	-	-	
SC20	-	+	-	-	+
SC21	-		+/-	-	+/-
SC22	-	-	-	-	-
SC23	-	-	-	-	-
SC24	-	+	-	-	+
SC25	-	-	+	-	+/-
SC26	-	-	-	-	-
SC27	-	-	-	+/-	-
SC28	-	-	+/-	-	+
SC29	-	+/-	+/-	-	+/-
SC30	+		-	-	-
SC31	-		+	-	-
SC32	-	+/-	+/-	-	+/-
SC33	-	-	-	-	-
SC34	-	-	+/-	-	+/-

Allelic dropout
Lack of coverage
+ One mutated allele present
- One unmutated allele present
+/- Two alleles present of which one mutated
-/- Two alleles present of which both unmutated

C

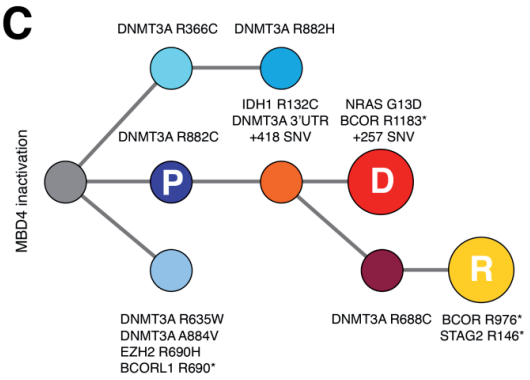


Figure S7 - Single cell genotyping for EMC-AML-1 at relapse. Single cells isolated at relapse (SC1-SC34, Sample #12, 841 days post-diagnosis) were genotyped for key mutations after whole genome amplification. Cells were assessed using the TruSight Myeloid Sequencing Panel (**A**), or by PCR followed by Sanger sequencing (**B**). The genotyping results were used to assign each cell to a clone, with colours matching the phylogenetic tree presented in Figure 3 (repeated here for clarity (**C**)). No clonal assignment was made if the genotyping results were ambiguous due to technical failure (white). Discordant results between the platforms are highlighted using red text and outline (in (**B**)) and generally reflect the lower sensitivity of detection for Sanger sequencing. For SC10 multiple alleles were detected at the site of the DNMT3A R882C mutation (denoted with a G>T symbol).

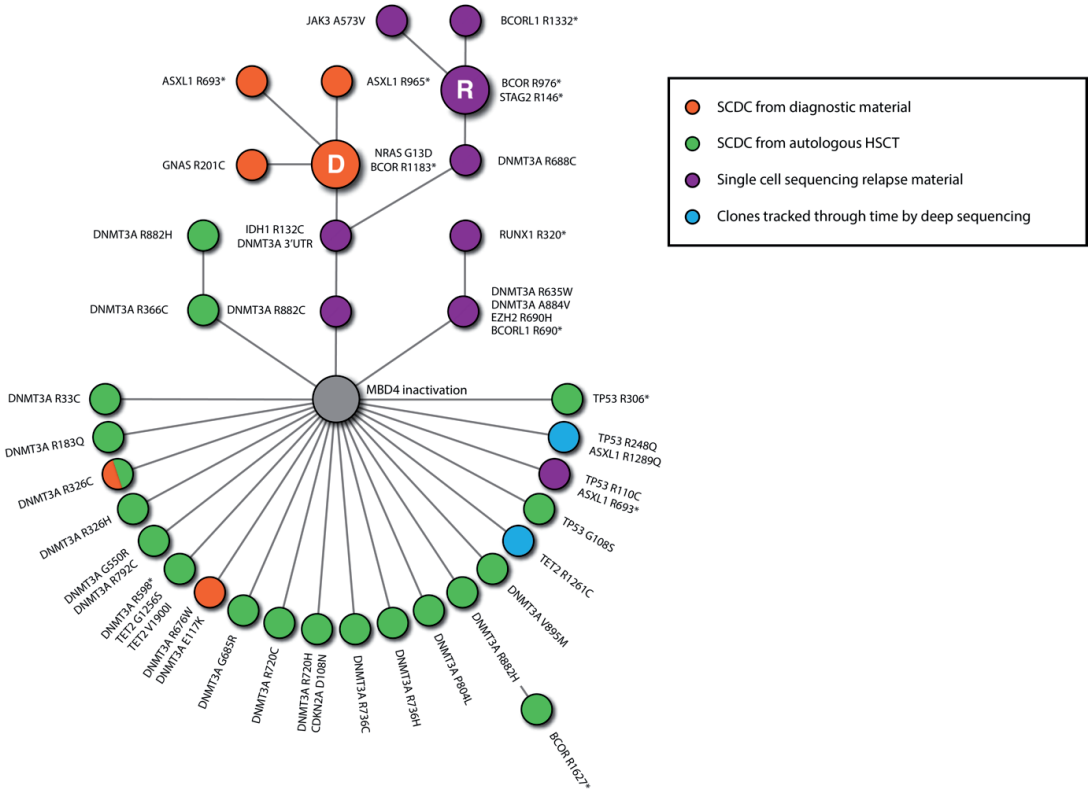


Figure S8. Mutations in ARCH-associated genes detected in SCDCs, single cells or bulk DNA from EMC-AML-1. The phylogenetic tree diagram highlights disparate clones with key driver mutations, which were detected in EMC-AML-1 in SCDCs from diagnostic and autoHSCT material, single cells at relapse or from bulk DNA sequencing.

Supplementary Data Set S1

Somatic mutations detected in MBD4-deficient AML at diagnosis (hg19). VCFv4.0 format. A quality score is provided (SOMATICP), variants with a score >0.5 were used for mutation signature analysis. Each call set presented in a tab of an xlsx file.

WEHI-AML-1.vcf [Diagnosis, Bone Marrow, Exome]
WEHI-AML-1.G.vcf [Diagnosis, Bone Marrow, Genome]
WEHI-AML-2.vcf [Diagnosis, Bone Marrow, Exome]
WEHI-AML-2.G.vcf [Diagnosis, Bone Marrow, Genome]
EMC-AML-1.vcf [Diagnosis, Bone Marrow, Exome]

Supplementary Data Set S2

Somatic mutations calls from *Mbd4* knockout myeloid progenitors (mm10). VCFv4.0 format. Variant calls were restricted to substitutions with a SOMATICP score >0.5. Each call set presented in a tab of an xlsx file.

MBD4_WT_113d_C36.G.vcf
MBD4_WT_113d_C40.G.vcf
MBD4_WT_113d_C45.G.vcf
MBD4_MUT_132d_C16.G.vcf
MBD4_MUT_132d_C22.G.vcf
MBD4_MUT_132d_C25.G.vcf
MBD4_WT_382d_C19.G.vcf
MBD4_WT_382d_C20.G.vcf
MBD4_WT_382d_C21.G.vcf
MBD4_MUT_382d_C13.G.vcf
MBD4_MUT_382d_C3.G.vcf
MBD4_MUT_382d_C7.G.vcf

Supplementary Data Set S3

Average FPKM values from WEHI-AML-1 and WEHI-AML-2. Columns are chromosome, start, end, gene, average FPKM, strand.

SDataset3-expression.bed.xlsx

Supplementary Data Set S4

Calculated relative mutation rates derived from sequencing data of MBD4-deficient cancers and for the *Mbd4* knockout myeloid progenitors from young mice. For the human data, the analysis includes mutations covered in the WGBS controls. For the mouse data the analysis includes CG sites with coverage of ≥ 10x in at least one sample.

SDataset4-RMR.xlsx

SUPPLEMENTARY REFERENCES

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009 Jul 15; **25**(14)**:** 1754-1760.
2. Li S, Garrett-Bakelman FE, Chung SS, Sanders MA, Hricik T, Rapaport F, *et al.* Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nature Medicine* 2016 Jul; **22**(7)**:** 792-799.
3. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013 May 30; **368**(22)**:** 2059-2074.
4. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 2012 Mar; **22**(3)**:** 568-576.
5. Miller CA, McMichael J, Dang HX, Maher CA, Ding L, Ley TJ, *et al.* Visualizing tumor evolution with the fishplot package for R. *BMC Genomics* 2016 Nov 7; **17**(1)**:** 880.
6. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, *et al.* COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Research* 2017 Jan 4; **45**(D1)**:** D777-D783.
7. Blokzijl F, Janssen R, Van Boxtel R, Cuppen E. MutationalPatterns: an integrative R package for studying patterns in base substitution catalogues. *bioRxiv* 2016**:** 071761.
8. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; **25**(9)**:** 1105-1111.
9. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 2013; **30**(7)**:** 923-930.
10. Tyner C, Barber GP, Casper J, Clawson H, Diekhans M, Eisenhart C, *et al.* The UCSC Genome Browser database: 2017 update. *Nucleic Acids Res* 2017 Jan 4; **45**(D1)**:** D626-D634.
11. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 2009; **4**(8)**:** 1184-1191.
12. Aken BL, Achuthan P, Akanni W, Amode MR, Bernsdorff F, Bhai J, *et al.* Ensembl 2017. *Nucleic Acids Res* 2017 Jan 4; **45**(D1)**:** D635-D642.
13. Spencer DH, Russler-Germain DA, Ketkar S, Helton NM, Lamprecht TL, Fulton RS, *et al.* CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression. *Cell* 2017 Feb 23; **168**(5)**:** 801-+.
14. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012 Sep 6; **489**(7414)**:** 57-74.
15. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016 Aug 18; **536**(7616)**:** 285-291.
16. Jongen-Lavrencic M, Grob T, Hanekamp D, Kavelaars FG, Al Hinai A, Zeilemaker A, *et al.* Molecular Minimal Residual Disease in Acute Myeloid Leukemia. *N Engl J Med* 2018 Mar 29; **378**(13)**:** 1189-1199.
17. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, *et al.* Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* 2017 Nov 16; **171**(5)**:** 1029-1041 e1021.
18. Morera S, Grin I, Vigouroux A, Couve S, Henriot V, Saparbaev M, *et al.* Biochemical and structural characterization of the glycosylase domain of MBD4 bound to thymine and 5-hydroxymethyuracil-containing DNA. *Nucleic Acids Res* 2012 Oct; **40**(19)**:** 9917-9926.
19. Lee EJ, Rath P, Liu J, Ryu D, Pei L, Noonepalle SK, *et al.* Identification of Global DNA Methylation Signatures in Glioblastoma-Derived Cancer Stem Cells. *J Genet Genomics* 2015 Jul 20; **42**(7)**:** 355-371.
20. Wong E, Yang K, Kuraguchi M, Werling U, Avdievich E, Fan K, *et al.* Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci U S A* 2002 Nov 12; **99**(23)**:** 14937-14942.

Chapter 8

Summary and General Discussion

SUMMARY

Acute myeloid leukemia (AML) is a heterogeneous group of disorders with variable treatment outcomes attributable to the intrinsic genetic complexity, present at initial diagnosis and/or evolving through disease evolution.[1, 2] These genetic lesions have a profound impact on patients response to treatment and prognosis. The advent of next-generation sequencing has tremendously increased our knowledge and understanding of AML pathogenesis, heterogeneity, and clinical outcome.[1, 3] In addition, it revealed genetic precursor lesions associated with age-related clonal hematopoiesis or clonal hematopoiesis of indeterminate potential (CHIP). [4, 5] These advancements in the field of genomics of cancers, in general, and AML in particular, has led to improvements in the WHO classification of AML in 2016 and updates of the ELN risk stratification of AML in 2017.[6, 7]

In **Chapter 2**, we studied the utility of molecular NGS-based minimal residual disease (MRD) detection in AML. We analyzed a large number of adult patients with AML (i.e., 482 patients) by targeted NGS panel, at diagnosis and complete hematologic remission (CR) achieved after 2 cycles of intensive induction therapy. We used NGS to detect mutations in 54 genes known to be recurrently mutated in myeloid malignancies. The endpoints were 4-year rates of relapse, relapse-free survival, and overall survival (OS). We divided the cohort into a training set (283 patients) and a validation set (147 patients); two sets with similar clinical, cytogenetic, and molecular genetic characteristics. At diagnosis, mutations were present in 430 out of 482 patients (i.e., 89.2%). Persistent mutations were detected in 51.4% of those patients during CR. The frequency at which mutations persisted across genes was highly variable, frequently at high variant allele frequencies (VAF; range: 0.02 up to 47%). Persistence of mutations known to be associated with age-related clonal hematopoiesis (i.e., mutation in *DNMT3A*, *TET2,* and *ASXL1* [abbreviated *DTA* mutations]) were commonly detected at high VAFs in CR. However, these persisting mutations did not have prognostic value in both the training and the validation sets. In contrast, detection of non-*DTA* mutations during CR was significantly associated with an elevated risk of relapse, confirmed in the validation set. Along the same line, we showed that persistence of non-*DTA* mutations was associated with shorter relapse-free survival and overall survival. When we performed multivariate analyses and corrected for the major established relevant prognostic factors (e.g., age, white-cell count, 2017 ELN risk classification, and the number of cycles of induction chemotherapy), we confirmed that persistence of non-*DTA* mutations during complete remission conferred significant independent prognostic value with respect to the rates of relapse, relapse-free survival, and OS. Furthermore, we compared NGS for detection of persistent non-*DTA* mutations with multiparameter flow cytometry for the detection of residual disease according to the leukemia-associated immune phenotype from which sufficient samples were available for both analyses (i.e., 340 patients). Using multivariate analysis, we demonstrated that the combined use of the two assays for the detection of

residual disease conferred the strongest independent prognostic value with respect to the rates of relapse, relapse-free, and overall survival. In this unprecedented study of NGS-based MRD, we demonstrated that persistence of mutations in genes associated with pre-leukemia (i.e., *DTA* mutations) in CR were not predictive for relapse, and we provided evidence that NGS-based MRD is a powerful prognostic tool that is applicable to the vast majority of AML patients.

In Chapter 3, we sequenced a large cohort of AML patients (i.e., 2200 patients) using NGS to comprehensively investigate the mutation and clonal features of AML with mutated *TP53*. Mutated *TP53* was detected in 10.5% (230 out of 2200) of AML patients. The majority of mutant *TP53* AML had bi-allelic *TP53* mutations. Interestingly, concurrent mutations were found in 49% (113 out of 230) of the mutant *TP53* AML patients and the most frequent concurrent mutations were in genes known to be associated with clonal hematopoiesis (i.e., *DNMT3A*, *TET2,* and *ASXL1*). We observed that 84% of mutant *TP53* AML patients had genomic instability; defined by the presence of a complex karyotype. Remarkably, genomic instability was more prevalent in bi-allelic than in mono-allelic *TP53* mutant AML patients, i.e., 97.1% and 40%; respectively. We observed no significant association between survival and various *TP53* VAF thresholds (from 50%, down to 5%). Moreover, we did not observe any survival advantage of mono-allelic over bi-allelic *TP53* mutant AML patients. Since the majority of mutant *TP53* AML concurrent mutations were associated with clonal hematopoiesis, *TP53* was considered the only suitable mutation for molecular MRD detection by deep NGS. Seventy-three percent of mutated *TP53* at diagnosis had persistence *TP53* mutation in CR. Yet, detection of *TP53* mutation in CR was not associated with relapse incidence (p= 0.911) or with OS (p= 0.653). Although the association between *TP53* mutations and the complex karyotype is well known, our study revealed that the presence of genomic instability was strongly associated with reduced OS in *TP53* mutant AML (2-year rate of OS, 34% vs. 9% genomic stable and genomic instable, respectively; p= 0.002). Furthermore, we showed that mono-allelic mutant *TP53* AML who are genomically stable have better OS than mutant *TP53* AML with genomic instability (p= 0.001). This implies that the degree of genomic instability in *TP53* mutant AML, regardless of *TP53* allelic status, has a profound prognostic value. In order to recognize who would benefit from allogeneic HSCT, we evaluated the prognostic value of genomic instability in 59 mutant *TP53* AML patients received allogeneic-HSCT. We showed that genomic instability in mutant *TP53* AML patients was associated with shorter OS (2-year rate of OS, 60% genomically stable vs. 16.7% genomically instable; HR 2.44, 95% CI: 1.03 to 5.79; p=0.02). Multivariate analysis confirmed that genomic instability in mutant *TP53* AML confers an independent prognostic value for OS.

In Chapter 4, we examined the occurrence of *PPM1D* mutations at diagnosis, in CR, and at the time of relapse**.** We demonstrated that *PPM1D* mutations are rare in *de novo* AML and refractory anemia with excess of blasts (RAEB) in contrast to therapy-related AML, present at higher frequency after high dose chemotherapy at generally low VAFs and not

part of the AML relapse. Thus, *PPM1D* mutations should not be considered as a marker for measurable (minimal) residual disease.

In Chapter 5, we investigated the feasibility of using May-Grünwald Giemsa stained bone marrow (MGG-stained BM) slides for NGS-based mutation detection. We demonstrated that DNA could be more efficiently isolated from MGG-stained BM slides, in contrast to archived unstained slides. We showed that in the majority of cases the mutation profiles of DNA derived from the MGG-stained BM slide were identical to the DNA derived from Ficoll-purified BM (FPBM) at diagnosis and relapse. We also revealed the possibility to detect driver mutations in DNA isolated from old MGG-stained slides (i.e., prepared in 1975). We did not only show that mutations were detectable in DNA derived from MGG-stained BM slides, but the variant allele frequencies (VAFs) of the mutations were highly similar to the mutations detected in FPBM-derived DNA (R^2: 0.87). This finding indicates that MGG-stained BM slide DNA is an excellent and reliable source to measure mutation burden in AML. Besides demonstrating the feasibility of mutation detection by hybridization-based NGS approach (Illumina TruSight Myeloid sequencing panel), we showed that mutations can also be detected using PCR-based NGS custom approach. We demonstrated that detection of mutations in DNA derived from MGG-stained slides is feasible not only at diagnosis and relapse but also during treatment. We used MGG-stained slide DNA to study mutation kinetics in 18 AML patients during the course of the disease. The mutation patterns were in line with the treatment and clonal hematopoiesis could be effectively distinguished from residual disease. Thus, archived MGG-stained BM slides are an excellent source for retrospective molecular analyses of AML during the course of disease and treatment. Similar molecular analyses are for certain applicable to other hematological malignancies as well. This study provides a new avenue for studying molecular dynamics and kinetics of mutations during the course of disease of patients when limited samples are available at critical time points. In addition, this approach has the potential to discriminate early mutations, which are often associated with clonal hematopoiesis of indeterminate potential, from oncogenic driver mutations.

In Chapter 6, we evaluated the mutational landscape, gene expression signatures and prognosis of a subset of AML patients harboring *KMT2A*-PTD mutations in comparison to a well-characterized adult AML cohort without *KMT2A*-PTD. We showed that *KMT2A*-PTDs were present in 5.5% of all AML cases. *KMT2A*-PTDs were significantly associated with concurrent trisomy of chromosome 11 in line to previous studies. When we analyzed the mutational landscape of AMLs with and without *KMT2A*-PTDs we found that a number of mutated genes were significantly more prevalent in *KMT2A*-PTD AML (i.e., *FLT3*-ITD, *IDH1*, *U2AF1,* and *IDH2*). In contrast, mutations in *NPM1*, *TP53,* and *NRAS* were significantly less frequent in *KMT2A*-PTD AML. Using our previously published gene expression profile dataset[8], we demonstrated that multiple homeobox-related gene family members were consistently overexpressed in *KMT2A*-PTD AML. Furthermore, using an association model

that takes a large number of clinically relevant and genetically defined subsets of AML into account, we revealed that *KMT2A*-PTDs induce overexpression of a subset of *HOX*-related genes differently as compared to t(11q23)-related KMT2A fusion proteins, indicating that the *KMT2A*-PTD prompts leukemogenesis by a distinct mechanism. Finally, we demonstrated that *KMT2A*-PTD AML with concurrent *DNMT3A* or *NRAS* mutations are associated with adverse clinical outcomes. This detailed study of a subset of AML with *KMT2A*-PTD mutations provides further insights into the biology and the clinical outcome of *KMT2A*-PTD AML and evidence that *KMT2A*-PTD AML is a different entity than other 11q23-rearranged AML.

In **Chapter 7**, we investigated familial cases with germline inactivation of methyl-binding domain 4 (MBD4) which is associated with the development of early-onset AML by acquiring pathogenic mutations in driver genes, most particularly *DNMT3A*. Our results provided insight into the vital role of MBD4 in safeguarding against the damage produced by 5-methylcytosine (5mC) deamination. We revealed the impact of constitutive inactivation of MBD4 on the development of early-onset AML and revealed that MBD4-deficient blood cell progenitors are particularly sensitive to methylation damage. We showed that there is a profound link between methylation damage and the development of molecularly defined hematological malignancy, AML with bi-allelic mutations in *DNMT3A,* and mutant *IDH1* or *IDH2*. Our study revealed that patients with germline biallelic *MBD4* mutations, experience accelerated clonal expansions of cells carrying mutations strongly associated with clonal hematopoiesis, however, decades earlier than healthy individuals. We have revealed an entirely new germline syndrome with a predisposition to myeloid malignancy. Although this syndrome appears to be rare it can potentially give important insights into the etiology of clonal hematopoiesis and subsequent development of AML.

GENERAL DISCUSSION

Next-generation sequencing in AML

Over the last decade, our knowledge about the mutational landscape of AML has increased tremendously, largely due to advances in novel sequencing techniques.[1, 3, 9] Next-generation sequencing (NGS), which is also known as high-throughput sequencing, is the term that is used to refer to several different modern massively parallel sequencing technologies. The main NGS technologies include whole-genome sequencing (WGS), whole-exome sequencing (WES), and targeted gene sequencing. These technologies have enabled sequencing of multiple genes or even whole genomes and exomes rapidly, precisely, and cost-effectively within a short amount of time, which has revolutionized genomic research.[10]

Targeted NGS sequencing is used to sequence selected or clinically relevant genes and hotspot mutations. This method has gained popularity in recent years due to the reduced cost and time of sequencing.[10] In addition to reducing the costs and turnaround time required for the single diagnostic assay, such myeloid panels also enable significantly higher levels of coverage of a given region of interest compared with Sanger sequencing. While Sanger Sequencing has a detection limit of 20% of allele frequency, targeted NGS reaches sensitivities up to at least 1% allelic burden. In contrast to PCR, NGS also allows for single-base resolution.[11, 12] This approach has been shown to be effective for targeted molecular genotyping of AML patients.[1] Many laboratories around the world nowadays are using targeted NGS in routine clinical diagnostics and prognostics of AML. A number of commercially available gene panels focusing on genes frequently mutated in myeloid malignancies have been introduced, e.g., the Illumina TruSight Myeloid panel, the Archer VariantPlex Core Myeloid panel, the Human Myeloid Neoplasms QIASeq DNA Panel and the AmpliSeq for Illumina Myeloid panel among many others. For example, the Hemato-oncology laboratory, unit molecular diagnostics of the department of Hematology at the Erasmus University Medical Center, Rotterdam, as a reference laboratory for the multicenter international AML HOVON-SAKK clinical trials uses the TruSight Myeloid Sequencing Panel from Illumina, to determine all clinically relevant genetic markers. The kit was designed to sequence 54 genes that are known to be frequently mutated in myeloid malignancies. This panel has been extensively validated before applying it in routine clinical diagnosis and is now used by many laboratories worldwide. AML patients are classified locally according to 2017 ELN, only *NPM1, CEBPA, FLT3, RUNX1, ASXL1*, and *TP53* need to be included in a small and cost-effective gene panel. Of note, mutations in *CEBPA*, a notoriously difficult gene to be sequenced due to its GC-content cannot be reliably detected with the TruSight Myeloid panel. Thus, for the identification of bi-allelic *CEBPA* mutant cases, separate custom assays need to be developed. However, some commercially available targeted NGS myeloid panels claim that their chemistry is able to sequence the entire gene with a good coverage.

WGS and WES were elegantly used by The Cancer Genome Atlas (TCGA)[3] to determine

the mutational profile of 200 *de novo* AML patients (50 and 150 patients, respectively), along with analysis of RNA and microRNA expression and DNA methylation. In contrast to most adult cancers such as breast, lung or pancreatic cancer, AML genomes have fewer mutations. Recurrently mutated genes were classified into nine distinct categories based on gene function or pathway involvement (i.e., signaling genes, DNA methylation-associated genes, myeloid transcription factor gene fusions or mutations, chromatin modifier genes, the *NPM1* gene, tumor suppressor genes, spliceosome complex genes and cohesin complex genes) revealing many potentially essential biologic relationships. Several genes were found mutated in 1 or 2 patients, however, clear nonrandom mutational patterns of co-occurrence and mutual exclusivity were revealed. Using the variant allele frequency (VAF), the TCGA characterized the clonal architecture of AML. Mutations with the highest VAF were considered to define the founding clones, and lower VAFs represented subclones. More than 50% of the patients exhibited both the founding clone and at least one subclone.[3, 13, 14] This study provided a comprehensive picture of the AML genome, however, in a relatively small series of AML.

In a recent comprehensive study of 1540 AML patients treated in different intensive chemotherapy trials, Papaemmanuil *et al*.[1] identified 5234 driver mutations across 76 genes or genomic regions. These driver mutations were found in 96% of the patients. Over 80% of the patients harbored 2 or more driver mutations. Combining NGS and cytogenetic data enabled the classification of the cohort into 11 subgroups or classes with distinctive diagnostic features and clinical outcomes.[1] Clearly more than what was shown by the TCGA[3] as a result of the size of the cohort, which enabled better and refined definitions of the AML subtypes. Remarkably, three AML subgroups representing novel entities were revealed: AML with mutations in genes involved in chromatin remodeling, the spliceosome or both (18%); AML with mutations in the *TP53* gene, chromosomal aneuploidies or both (13%); and AML with *IDH2*^{R172} mutations (1%).[1] The study showed that patients with the first two subgroups had the poorest outcomes. AML with mutated *IDH2*^{R172} showed absence of co-occurring class-defining mutations and featured by favorable clinical outcome.[1] Intriguingly, 84% of patients in the chromatin – spliceosome category would initially be categorized as intermediate risk, although these patients showed lower induction chemotherapy response rates and elevated relapse rates comparable to adverse risk outcome rates. Papaemmanuil and colleagues were able to propose a clinically relevant and systematic detailed molecularly driven classification and in addition, provide fertile grounds to improve WHO and ELN classification systems.[15-17]

Recent NGS-based analyses of well-defined AML cohorts have enriched our knowledge about the genetic landscape of AML in general. These analyses also demonstrated that AML is a group of diseases rather than one disease, which should thus be studied as separate entities as we did in **Chapter 5** where we investigated the mutational landscape, gene expression signatures, and prognosis of a subset of AML patients with *KMT2A*-PTD

mutations (app. 5% of AML). We demonstrated that *KMT2A-PTD* AML with concurrent *DNMT3A* or *NRAS* mutations is associated with adverse clinical outcome. We also showed that *KMT2A*-PTD AMLs may cause leukemogenesis in a different fashion than 11q23 AMLs. The improved knowledge of small genetically defined subsets may form the basis of more individualized treatment of the AML patient in the future. The various NGS-based AML genomic profiling studies [1, 3, 9, 18] already resulted in the improvement of the classification, prognostic stratification, treatment and response assessment of AML[10].

NGS-based MRD detection in AML

The genetic landscape of AML is characterized by numerous somatic mutations and each mutation could potentially be used as an MRD marker.[1, 3] This feature makes NGS technologies an appealing approach, not only for detection of somatic mutations at AML diagnosis, but also as a tool for MRD measurements during the course of disease. Several studies demonstrated the ability of NGS for MRD detection, initially focusing on specific single markers. In 2012, Thol *et al*.[19] demonstrated in 80 AML patients the feasibility of MRD detection by NGS. Their study focused on tracking MRD levels in patients with AML carrying *FLT3*-ITD or *NPM1* mutations. They showed that DNA sequencing by NGS is a reliable method for the quantitative assessment of *NPM1* mutation burden during the disease course. This NGS-based MRD approach was found to be concordant with qPCR in 95% of the AML patients with mutated *NPM1*.[19] In a prospective cohort of 814 AML patients, Kohlman and colleagues[20] investigated the potential of amplicon deep sequencing of *RUNX1* mutations as an MRD marker. This approach revealed that patients can be categorized into good or bad responders on the basis of median *RUNX1* mutational burden at CR, following distinct patterns of event-free survival (EFS) and overall survival (OS), respectively.[20]

Successively, in recent years several studies illustrated the potential of NGS-based MRD detection by targeting multiple markers using amplicon-based sequencing (Table 1). Klco and colleagues[21] performed WGS or WES on samples obtained at diagnosis from 71 AML patients treated with standard induction chemotherapy. Subsequently, they performed enhanced deep exon sequencing targeting 264 recurrently mutated genes on paired diagnosis and CR samples from 50 patients. While using a fairly less sensitive method (VAF cut-off of 2.5%), they reported that patients who cleared their somatic mutations at CR had prolonged event-free survival (EFS) and overall survival (OS) with hazard ratio of 6.0 (CI 1.93 – 7.11) and 2.86 (CI 1.39 – 5.88), respectively.[21]

Morita *et al*.[22] used a 295 gene panel to evaluate 131 AML patients who achieved CR of whom 93% had at least one mutation pretreatment. Subsequently, BM specimens taken 30 days post-induction chemotherapy were sequenced and examined using maximum VAF cut-offs of 2.5%, 1.0% and undetectable. Morita and colleagues reported that day 30 persistent mutations with VAFs of <1% were associated with a substantially better OS than that of patients with mutation VAFs >1%. They also showed that patients with undetectable

mutations at day 30 had significantly better EFS in multivariate analysis after adjusting for age, cytogenetic risk, allogeneic stem cell transplantation (allo-SCT), and flow cytometry-based MRD. Furthermore, they demonstrated that removal of solely persisting *DTA* mutations in CR from the MRD analysis produced stronger prognostic associations. This study confirmed our findings in **Chapter 2**, where we reported that the detection of the *DTA* mutations in CR have no prognostic impact.[23]

NGS-based MRD has also been used in transplantation setting of AML patients. Getta *et al*.[24] investigated if NGS can be used for MRD detection by comparing a targeted 28 gene NGS panel to multicolor flow cytometry (MFC) at diagnosis and prior to allogeneic HSCT. MRD positivity by NGS was defined as having a mutation burden of more than 5%. They observed in pre-allo-HSCT the persistence of mutations in *DNMT3A*, *TET2* and *JAK2* at rather high VAFs, similar to our findings. The two MRD detection methodologies, i.e., NGS and MFC, exhibited a concordance of 71% and patients with MRD positivity detected by both techniques were associated with an elevated risk of relapse, compared to either modality alone, similar to our findings (**Chapter 2**). This work demonstrated that NGS is complementary to MFC and provides actionable clinical information before transplantation.[24]

Thol and colleagues[25] evaluated 116 pre-transplant AML patients using a 46 targeted genes panel to identify persisting somatic mutations. They detected at least one trackable mutation in 93% of the patients. They used error-corrected sequencing that is based on unique molecular index (UMI) and applied it on pre-allogeneic transplant BM or peripheral blood (PB) specimens with a sensitivity of <0.02%. The study revealed that 45% of patients were NGS MRD positive with a median VAF of 0.33% (range: 0.016 to 4.91%). Based on competing risk analysis, these NGS MRD positive pre-transplant patients had a higher cumulative incidence of relapse (CIR) (hazard ratio 5.58; $P < 0.001$). Surprisingly, no difference was found between MRD positive and MRD negative patients in OS. In a similar study, Press *et al*.[26] investigated 42 AML patients for MRD pre- and post-allogeneic transplant using 42 gene NGS panel with a sensitivity of 0.5% VAF. The CIR was significantly higher in pre-transplant MRD-positive patients ($P= 0.014$). They found pre-transplant MRD positivity was associated with an elevated risk of relapse (hazard ratio 7.3; $P= 0.05$), reduced progression-free survival ($P= 0.038$), and marginally reduced OS ($P= 0.068$) in multivariate analysis.

Using a 13 genes NGS panel, Hourigan and colleagues[27] examined PB samples from pre-transplant AML patients in morphologic CR who were randomly allocated to either myeloablative (MAC) or reduced-intensity pre-transplant conditioning (RIC). They observed significant differences in relapse (19 vs. 67%; $P < 0.001$) and OS (61 vs. 43%; $P= 0.02$) between patients with MAC or RIC pre-transplant, respectively, for patients with detectable mutations. Compared to patients who underwent MAC, they found that patients who had RIC pre-transplant demonstrated a higher risk of relapse (hazard ratio 6.38; 95% CI 3.37 to 2.10), reduced relapse-free survival (hazard ratio 2.94; 95% CI, 1.84–4.69), and decreased OS (hazard ratio 1.97; 95% CI, 1.17– 3.30). These data indicate that MAC may result in better outcome for MRD-positive pre-transplant AML patients.

Similar results were reported by Kim *et al*.[28] in a cohort of 104 AML patients who underwent allo-SCT. They used an NGS panel of 84 genes on pre-transplant BM and paired CD3+ T-cells. Eighty percent of the patients had a traceable mutation. Post-transplant BM specimens were collected 21 days post-transplantation and sequenced utilizing a computational error correction approach and applied a cut-off of 0.2% VAF. They found that detection of MRD post-transplant was associated with an increased risk of relapse (56.2 vs. 16.0% at 3 years; $P < 0.001$) and decreased OS (36.5 vs. 67.0% at 3 years; $P = 0.006$) compared to patients with undetectable MRD. In multivariate analysis when taking into account 2017 European LeukemiaNet risk groups, these associations remained significant.

While these studies addressing slightly different prognostic time points of MRD assessment (i.e., after 2 cycles of chemotherapy, before and/or after transplantation) with different NGS-based MRD approaches and variable MRD sensitivities/thresholds, most of them showed that detectable NGS-based MRD in CR is associated with worse clinical outcomes irrespective to the time of MRD measurement. The above studies range in cohort size from 30 to 482 patients, markers were measured with variable sequencing depth and sensitivity ranging from 575x to >100.000x and from 2.5 to 0.001%, respectively. The number of markers targeted by the different studies was also highly variable (range from 13 to 295 genes). In comparison to all the aforementioned studies, our study (**Chapter 2**) was carried out using the largest AML study cohort. However, among the aforementioned studies, Hourigan and colleagues[27] performed the most sensitive NGS-based MRD approach in detecting residual disease due to the sequencing depth and error-corrected sequencing (mean MRD coverage >100.000x, and a sensitivity of 0.001% VAF) (Table 1). However, not all studies that used UMI error-corrected sequencing achieved the same sensitivity, which may be caused by different factors, such as the higher number of amplicons in the gene panel, insensitivity of the gene panel due to suboptimal design, the bioinformatics analysis tools used or lower DNA input. In our study, we used position-based error correction, similar to two other studies.[26, 28] However, one of these studies had lower sensitivity/threshold (0.5% VAF) [26], which might indicate that they possibly had high rates of sequencing artifacts. Clearly, there are a number of factors that influence the efficiency of MRD assessment by NGS, which requires harmonization to be able to compare all international AML clinical trials where NGS-based molecular MRD detection is applied.

We showed that the presence of most mutations at the time of CR was associated with an elevated risk of relapse. In contrast, the persistence of *DTA* mutations, which are known to be frequently mutated in aged healthy individuals with clonal hematopoiesis of indeterminate potential (CHIP)[4, 29], did not have prognostic value and this finding was confirmed by two studies, the study of Morita and colleagues[22] and by the study of Hourigan *et al*.[27], were they found that the prognostic significance becomes more pronounced when *mutation associated with clonal hematopoiesis* (i.e., *DNMT3A*, *TET2*, *JAK2*) were excluded. However, the other studies were not able to show this because of too small sample sizes.

Currently, MRD detection by MFC is the golden standard for patients with AML. There are only limited studies with a rigorous comparison between NGS- and multiparameter flow cytometric MRD detection besides ours. Getta *et al*.[24] findings supported our data that NGS-based MRD may complement MFC MRD assay. Collectively, these data from the different studies demonstrated the profound ability of NGS for MRD assessment. However, there are a number of issues that need to be addressed to reliably implement NGS-based MRD detection in routine clinical practice.

Table 1: Summary of NGS-based multi-gene MRD studies in AML.

Author	Year	Key findings	Disease state	Technique	Study size (n)	Mean MRD coverage	Maximum sensitivity (VAF)
Klco	2015	Clearance of disease-specific mutations at 30 days post induction correlated with better EFS and OS	AML, post induction	WES or WGS, and amplicon-based sequencing with paired normal tissue	50	543x (exome) 14,780x (amplicon)	2.5%
Jongen – Lavrencic	2018	Persistence of non-"DTA" mutations correlated with decreased RFS and OS. Combining NGS and flow MRD data produced strong correlations with outcome when methods were concordant and defined an intermediate prognosis group when methods were discordant	AML, post induction	54 Gene tumor-only NGS panel with position-based error correction	482	3500x	0.02%
Morita	2018	Patients with clearance of disease-associated mutations (<1% VAF) at 30 days post induction had better OS and better EFS after multivariate analysis; this correlation was strengthened by the exclusion of "DTA mutations" from the analysis	AML, post induction	295 Gene tumor only NGS panel	131	575x	<1.0%
Thol	2018	Detection of disease-associated mutations post-transplant was associated with higher incidence of relapse, but no difference in OS	AML, pre-transplant	46 Gene custom amplicon panel with UMI-based error correction	116	6,100x	0.02%
Kim	2018	MRD positive patients had a higher incidence of relapse and lower OS	AML, post-transplant	84 Gene NGS panel with paired normal T-cells and position-based error correction	104	1726x	0.02%
Balagopal	2019	Found evidence of MRD in 18/30 (60%) of post-transplant AML patients who showed no evidence of disease by standard engraftment studies	AML, post-transplant	22 Gene NGS panel with UMI-based error correction	30	>10,000x	0.10%
Press	2019	NGS MRD positive pre-transplant patients had higher risk of relapse in multivariate analysis	AML, pre-transplant	42 Gene NGS panel, coverage-based error correction	42	1900x	0.50%
Hourigan	2019	NGS MRD positive pre-transplant patients with reduced-intensity conditioning had increased relapse rates, decreased overall survival, and decreased OS compared to patients who underwent myelo-ablative conditioning	AML, pre-transplant	13 Gene NGS panel with UMI-error correction run in peripheral blood samples.	190	>100,000x	0.001%

Abbreviations: EFS, event free survival; OS, overall survival; NGS, next generation sequencing; WES, whole exome sequencing; WGS, whole genome sequencing; RFS, relapse free survival; MRD, measurable residual disease; UMI, unique molecular identifier.
Adapted from Yoest and colleagues[30]

NGS-based MRD detection in elderly AML

Following to the success in showing a profound relapse prediction by NGS-based MRD detection in AML of the young (<65 years), we sought to study MRD in elderly AML patients (>65 years). In this follow-up study, we used the same NGS panel and determined mutations at diagnosis and in CR similarly to what we did in **Chapter 2**; however, in a cohort of 157 elderly AML patients (manuscript in preparation). As expected the mutational landscape in AML of the elderly was as heterogeneous as of the young adults at diagnosis. Ninety percent of the patients (n=140 patients) had at least one mutation at diagnosis and on average more mutations were present in the elderly compared to the young AML. We found that 38% of the mutation were persistent in 68% of patients in CR which is more than what we found in the young AML patients (i.e., 26% of the mutations persisted in 50% of patients in CR; **Chapter 2**) and the frequency of persistent mutations was again highly variable. Similar to the young AML patients, we saw a high frequency of persistent *DTA* mutations in CR. However, we found other mutations to be more frequently persisting in CR of the elderly compared to young AMLs, including *IDH2*, *RUNX1*, *STAG2*, *BCOR,* and *TP53* (Figure 1, figure 1A in **Chapter 2**). Many mutations in the elderly persisted at high VAF in CR, similar to young AML; however, in the elderly several mutations other than *DTA* mutations persisted at high VAF (Figure 2, figure 1B in **Chapter 2**). Surprisingly, when we looked at the CIR of all mutations present versus absent in CR, we did not see any association between the persistence of mutations in CR and the risk of relapse (Figure 3A). This finding is different than what we revealed within the young AML patients (see Figure S4A in **Chapter 2**). When we only considered non-*DTA* mutations as MRD, again we did not see any association with risk of relapse, suggesting that the response to treatment and/or the biology of AML of the elderly is different than in the young AML patients (Figure 3B). Of note, due to the small size of patients with mutations high VAF in CR other than the *DTA* mutations, we could not analyze the impact in risk of relapse when these mutations would be included in the *DTA* mutation category.

Looking at the landscape of mutations in AML of the elderly at diagnosis it was clear that mutations types and frequencies in the elderly are different as compared to the young AMLs. Several mutations were found to be more frequently mutated in the elderly than in young AMLs, including *RUNX1*, *TET2*, *ASXL1*, *SRSF2*, *ZRSR2*, *U2AF1*, *BCOR*, *EZH2*, and *TP53*. Whereas, other mutations were less prevalent at diagnosis in elderly AML than in the young patients, such as *NPM1*, *WT1*, *KIT,* and *NRAS*. These results directed our focus to the findings of Lindsley *et al*.[31] published in 2015, where they showed in clinically proven *de novo* and s-AML that some mutations appear to be highly specific for *de novo* AML and others are highly specific for s-AML. They revealed that the mutations (i.e., 8 of all mutations) that were highly specific for s-AML involved mutations in the spliceosome, and chromatin-modifying genes. Whereas, core-binding factor fusion genes, 11q23 rearrangements and

mutant *NPM1* were highly specific to *de novo* AML.

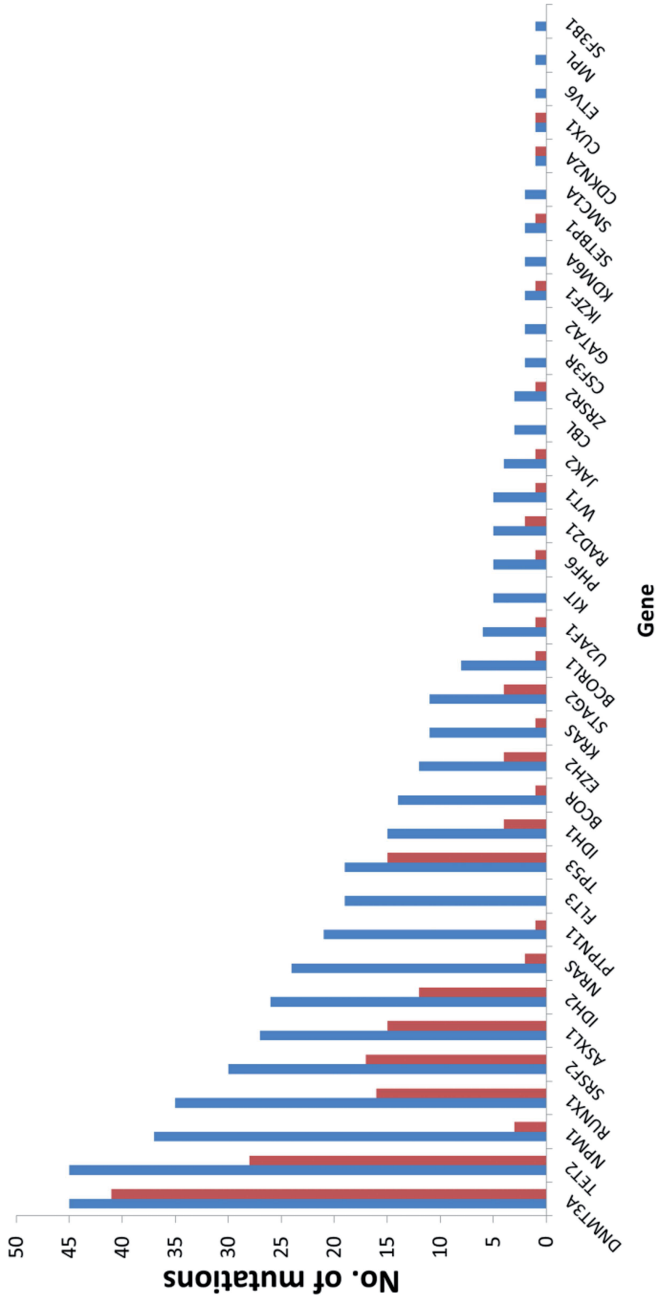


Figure 1: Mutations profile of elderly AML patients at diagnosis and in complete remission. The figure depicts the number of mutations on the y-axis and the gene mutations on the x-axis. The number of mutations found at diagnosis is illustrated by blue bars and the number of mutations detectable in CR is depicted in red bars.

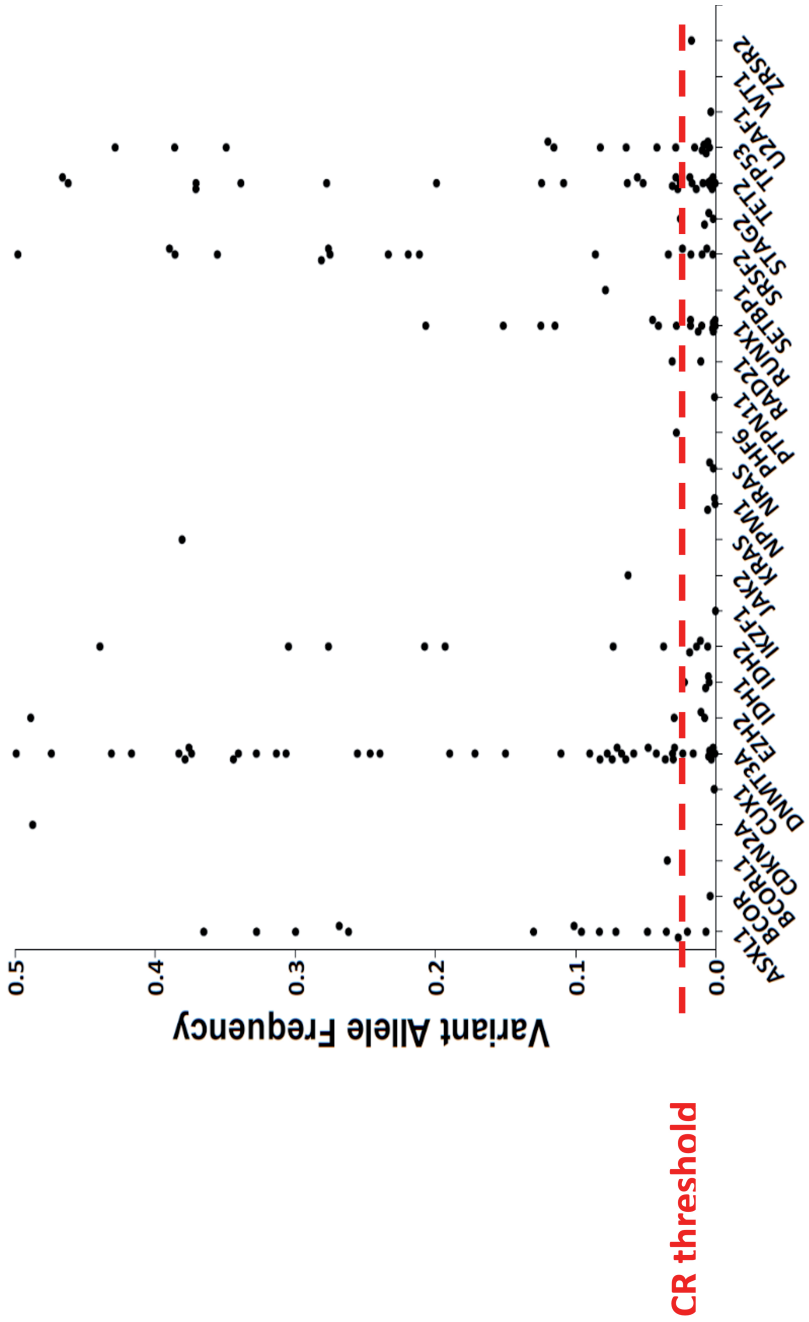


Figure 2: Variant allele frequencies of mutations detected in CR. The graph depicts the VAF of mutations detected in elderly AML patients at CR. Each dot in the graph represents an individual mutation. The y-axis depicts the VAF and on the x-axis the names of the genes.

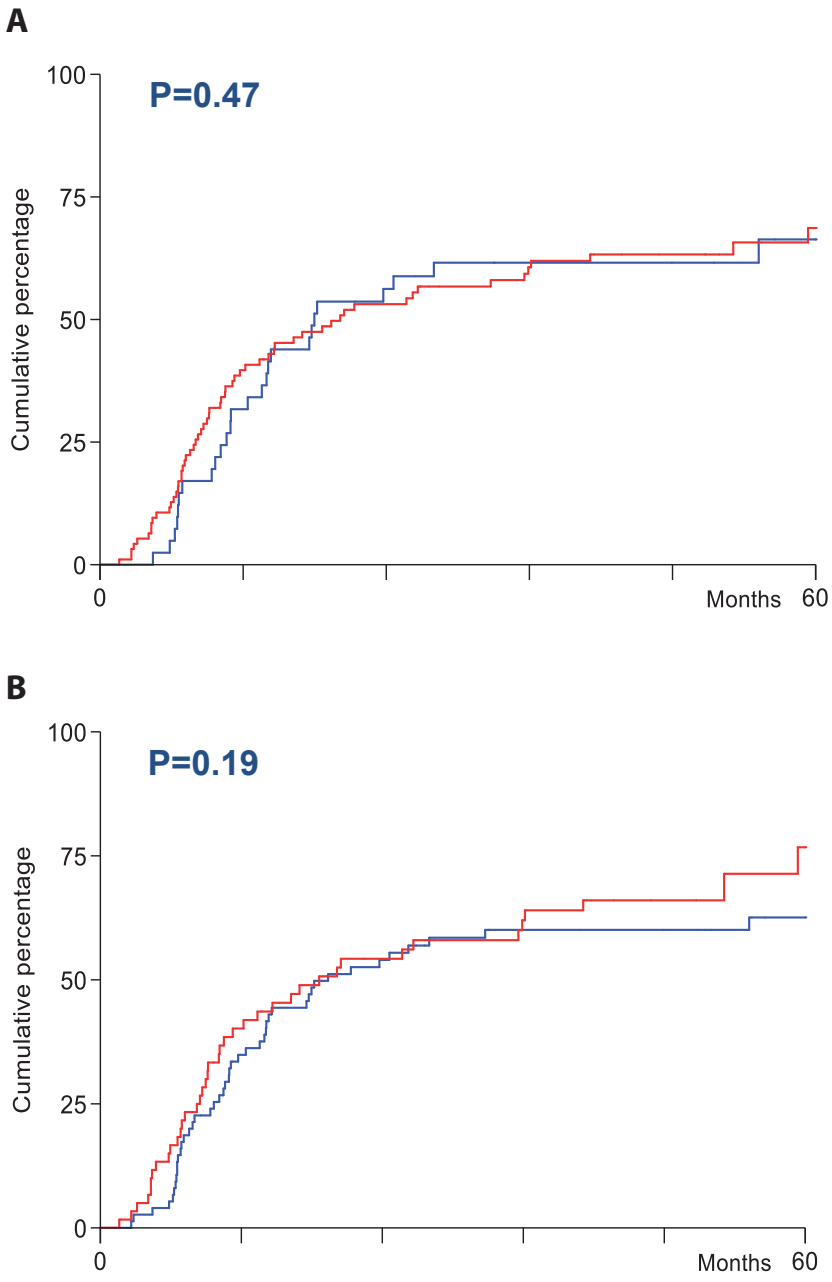


Figure 3: Cumulative incidence of relapse in elderly AML. The figure depicts the CIR in elderly AML patients when all mutations were detected in CR (3A) and when non-*DTA* mutations were considered as MRD (3B).

The s-AML characterized by the 8 mutations defined a group of AML patients from now on termed secondary-type AML. This group of patients is highly similar to the group of patients with mutations in genes encoding chromatin, RNA splicing regulators or both that were defined in a large cohort by Papaemmanuil and colleagues[1] in 2016. They classified AML patients into different subtypes of mutual exclusivity based on patterns of co-existence of mutations and these subsets of patients (i.e., patients with mutations in chromatin and RNA splicing regulators genes) represented 18% of the AML patients who are highly similar to secondary-type AMLs. AML with *TP53* mutations were considered a separate entity[31].

Table 2: Distribution of young and elderly AML patients in CR in the three AML categories (ontogeny)

	n	%
AML <65 years (n= 430)		
de novo AML	295	69%
secondary type AML	116	27%
mutant *TP53* AML	19	4%
AML >65 years (n= 139)		
de novo AML	51	37%
secondary type AML	71	51%
mutant *TP53* AML	17	12%

We next defined our cohort of AML patients in remission molecularly into *de novo* AML, secondary-type AML (patients with mutations in *SRSF2*, *ZRSR2*, *SF3B1*, *ASXL1*, *BCOR*, *EZH2*, *U2AF1*, and *STAG2*), and AML patients with mutant *TP53* based on mutations present at diagnosis as per Lindsley paper[31]. By looking at the prevalence of mutations in the three categories, it is clear that molecularly defined *de novo* AML is more frequent in AML of the young (69% vs. 37%), whereas, molecularly defined secondary-type AML is more frequent in the elderly than in young AML (51% vs. 27%) like mutant *TP53* AML (4% vs. 17%; young and elderly, respectively) (Table 2). When we determined the CIR in patients who achieved complete morphological remission (combined young and elderly), we found that patients with mutant *TP53* have increased risk of relapse independent of MRD status (Figure 4), but strikingly, there was no difference in the risk of relapse between patients with *de novo* and secondary-type AML mutations. This similarity might be due to the fact that all patients with *de novo* and secondary-type AML mutations in this cohort achieved CR. When we looked at MRD using any marker (presence vs. absence) in molecularly defined *de novo* AML at CR (n=346), patients with residual disease have a higher risk of relapse than patients with no residual disease (data not shown, p=0.01). This difference in risk of relapse became more pronounced when the persistence of *DT* mutations only was excluded (Figure 5). Similarly, when we examined MRD using any marker (presence vs. absence) in molecularly defined secondary AML at CR (n=187), there was no difference in risk of relapse (data not shown). Interestingly, however, when considering the secondary-type AML cases with >3 persistent mutations in CR, these appeared to have a higher relapse risk compared to those cases with <3 persistent mutations in CR (Figure 6). These findings suggest that there is a difference in approach when considering NGS-based MRD detection in secondary-type AML compared to *de novo* AML, possibly determined by the etiology of this type of AML. In addition, there was no difference between AML patients with or without detectable mutant *TP53* in CR, suggesting that MRD detection is not beneficial for mutant *TP53* AML patients, as shown in **Chapter 5**. This unpublished study suggests that there is a quantitative and qualitative difference in mutations between age groups at diagnosis and persisting in CR and that NGS-based MRD detection in the young and elderly may be affected by the differences in types of AML, *de novo,* and secondary type AML. Moreover, this study indicates that mutant *TP53* AMLs should be considered as a separate AML entity (**Chapter 3**). Further studies are warranted to decipher the proper way on how to apply NGS-based MRD detection in these types of AML.

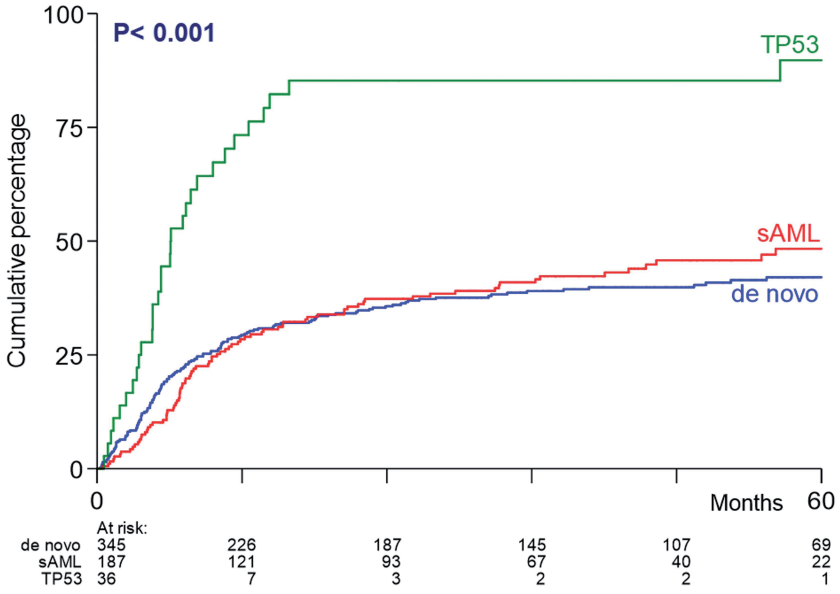


Figure 4: Cumulative incidence of relapse in AML patients. The graph demonstrated the risk of relapse in patients molecularly defined *de novo* AML, secondary-type AML, and *TP53* mutant AMLs.

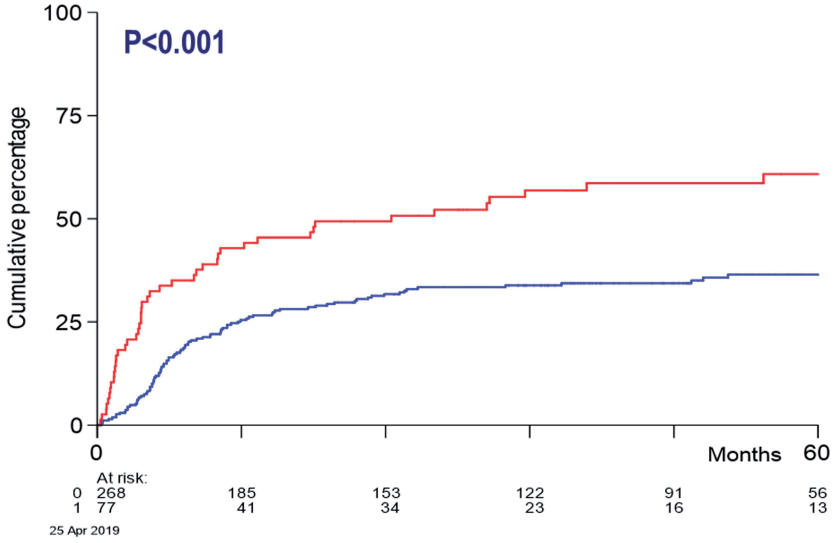


Figure 5: Cumulative incidence of relapse in CR in molecularly defined *de novo* AML when considering persisting mutations excluding *DNMT3A* and *TET2* mutations (non-*DT* mutations). The graph demonstrated the risk of relapse in patients with and without detectable MRD in non-*DT* de novo AML in CR. The red line represents MRD-positive and the blue line MRD-negative.

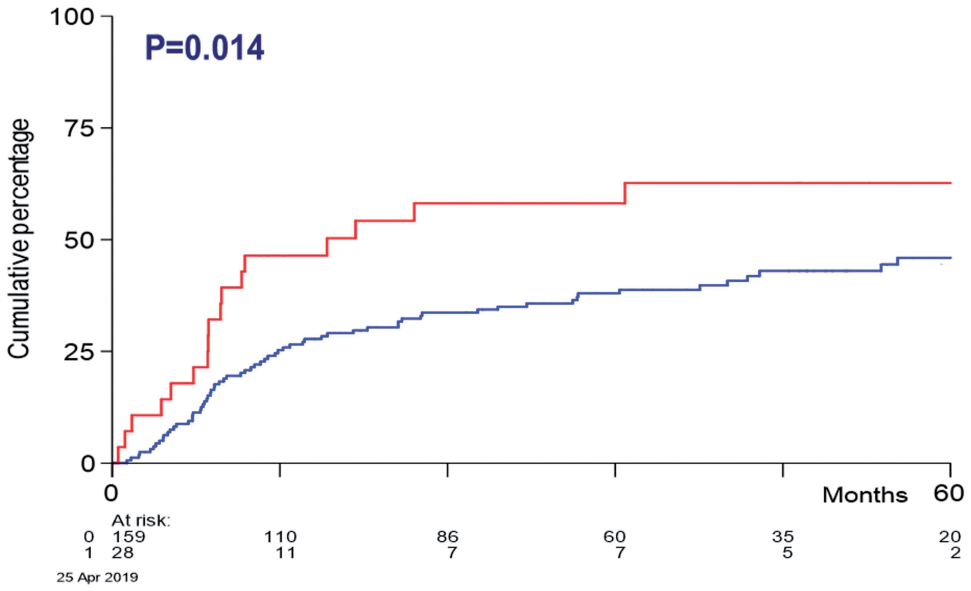


Figure 6: Cumulative incidence of relapse in CR in molecularly defined secondary-type AML. The graph demonstrated the risk of relapse in patients molecularly defined as secondary type AML. The red line represents patients who carry 3 or more mutations in CR and the blue line represents patients that carry less than three mutations in CR.

Limitations of NGS-based MRD detection

The ideal MRD detection method should be able to identify and quantify the smallest subclone(s) of leukemic cells in AML patients accurately during the disease course.[32] NGS has the potential to detect mutations in virtually all AML patients and thus is by far the most ideal molecular technique for MRD detection as demonstrated in **Chapter 2**. However, in order to implement NGS-based MRD detection in AML, the technique has to overcome several hurdles before it can be reliably used in routine clinical practice.

Sensitivity and specificity of NGS-based MRD detection

One of the limitations for NGS-based MRD detection is owed to the low sensitivity of NGS. Currently, NGS can be reliably used to detect mutations of ≥1%, because low-level mutational artifacts are introduced from isolation of DNA up to the actual NGS. Additionally, the NGS technology itself has a high intrinsic error rate (0.1 to 1%) and thus may cloud the detection of minor subclones and the discrimination of true mutation from PCR and sequencing artifacts. The limited sensitivity hinders the clinical application of NGS MRD tracking during therapy.[33] Several approaches have been developed to try to overcome these hurdles. Firstly, this could be done biochemically, for instance, the use of proof-reading polymerases can reduce the rate of PCR artifacts.[34] Secondly, it has been shown that computational approaches can be

used to correct for background noise to discriminate minor clone from artifact. For instance in **Chapter 2 and 3** we used mathematical and computational approaches for which all SNVs detected across the diagnostic samples were compiled and the background VAF distribution was determined for each specific SNV from follow-up samples missing this SNV in the matched diagnostic sample. For the remaining follow-up samples the presence of the SNV was considered confirmed when the VAF was an outlier compared to the background VAF distribution according to the Thompson-Tau test. A one-sided p-value <0.01 was considered statistically significant. Indels were processed and compared similarly, except for quantile normalization as there are infinitely many possible indel-configurations per locus.[23, 34] Thirdly, a method that relies on the incorporation of unique molecular identifiers (UMIs) can improve the accurate detection of mutations at low level by NGS. UMIs have been shown to have the ability to enhance the accuracy of identifying rare and subclonal mutations.[35] UMIs are a string of entirely random degenerate nucleotides that can be utilized to distinguish a true variant from artifacts/errors that were introduced during sample preparation or sequencing. This approach is known as error-corrected sequencing (ECS). This method has been recently applied by several groups, such as Thol[25], Balagopal[36] and Hourigan[27] (Table 1) for MRD detection in PB and BM and demonstrated marked reduction in false positive noise and appear to have better sensitivity over the standard NGS approach.

We sought to compare the accuracy of the application of error-corrected sequencing by using UMIs over the computational and statistical approaches that we used in **Chapters 2 and 3**. We randomly selected 24 AML patient samples at CR with known mutational profiles at diagnosis from our cohort in the study of Jongen–Lavrencic and colleagues[23] (Table 1). For testing the UMI method/ECS, we used VariantPlex Core Myeloid Panel kit (VPCMP) from ArcherDx (37 genes), and for the standard NGS method we used the TruSight Myeloid Sequencing Panel (TSM) from Illumina (54 genes); the same panel that we used in **Chapters 2 and 4**. The libraries were prepared as per the manufacturer's instructions and sequenced on the Illumina NovaSeq and variants we called similar to the computational and bioinformatics approach we applied in **Chapter 2**. Variants in samples sequenced with the ArcherDx kit were determined using the ArcherDx analysis tool. We obtained coverage of 9444 and 3261 reads/amplicon for VPCMP and TSM, respectively. The mutations detected and their VAFs were highly concordant. VAFs called by both error correction methods were extremely similar, when we examined all mutations (R^2= 0.976; Figure 7A) and the same was true when we focused on mutations with VAF <0.1% (R^2= 0.961; Figure 7B). The latter mutations are deletion and insertion mutations which can be detected with higher sensitivity than point mutations. The indel mutations are unique and the probability of detecting such variants as sequencing error is exceptionally low. These results indicate that applying computational error correction is as accurate as using the ECS method; additionally, it indicates that ECS might be superior when analyzing individual samples. In contrast to ECS, the computational approach requires to have control samples in each run

to use for background noise discrimination from true mutation. However, more samples and rigorous analyses are warranted to demonstrate the superiority of the ECS approach in MRD assessment of individual AML cases.

Discrimination of MRD from clonal hematopoiesis

We (**Chapter 2**) and others showed that persistence of mutations in preleukemic genes (i.e., *DTA* mutations), in the absence of non-preleukemic mutations, post-therapy indicate a state of clonal hematopoiesis rather than residual disease or pre-relapse condition.[22, 23, 27] However, there are other genes that were reported to be associated with CHIP, including *TP53*, *JAK2*, *SF3B1*, *SRSF2*, *PPM1D*, *CBL* and *IDH2*[4, 5], which may or may not be predictive for impending relapse. In **Chapter 3**, we showed that mutant *TP53* MRD is not beneficial. In addition, there are many other mutated genes of unknown prognostic significance when detectable in CR due to their low prevalence and so far analyzed within the bulk of non-*DTA* mutations.[23] Since the incidence of most of these genes is low in AML, it precluded the discrimination between residual disease and clonal hematopoiesis. Large AML cohort studies are warranted to distinguish between the two to refine and improve risk prediction.

Multiple time points (as illustrated in **Chapter 5**) enable discrimination of clonal hematopoiesis and true leukemia during CR. Studying the order of mutations also enables the discrimination. Mutations in the founding clone are the ones that may represent clonal hematopoiesis, whereas, the late mutations are the ones we need to aim at, such as *NPM1*, *FLT3*, and *RAS*. When such discrimination between clonal hematopoiesis and true leukemia, designing a small NGS panel that includes only the true leukemia mutated gene will enable deeper sequencing. Our preliminary data in the elderly AMLs show that NGS MRD might be different due to the persistence of more clonal hematopoiesis-related mutations.

Design challenges for accurate NGS-based MRD detection

Another limitation of NGS technologies is the sequencing of the *CEBPA* gene and detection of internal tandem duplication (ITD) in the *FLT3* gene in commercial NGS gene panels in a multiplex setting at diagnosis or for MRD monitoring. *CEBPA* (i.e., bi-allelic *CEBPA*) and *FLT3*-ITD are recognized as important AML prognostic factors in the ELN risk stratification of AML.[7] The *CEBPA* gene is a single exon gene that is generally difficult to amplify using PCR and in turn difficult to sequence as well since the gene is GC-rich. Detection of *FLT3*-ITD mutations by NGS is challenging because the size of the *FLT3*-ITD vary dramatically[37] producing a large range of variable insertions within the Juxtamembrane (JM) domain (range from 3 to 400bp)[38], which makes it challenging to align the mutation to a reference sequence. However, there are nowadays some commercially available NGS gene panels that are able to sequence *CEBPA* gene successfully and there is a company that provides kits and analysis tool to detect mutations in these two genes (i.e., ArcherDx).

Recently a group established a single target custom high-coverage approach to detect *FLT3*-ITD mutations using NGS and developed a novel open-source analysis program called "getITD". In this study, 57 samples from 28 AML patients were sequenced, as well 3 human AML cell lines and 2 healthy volunteers and demonstrated that this method can reliably detect a wide range of ITD lengths (up to 244bp) at high sensitivity (0.0067%) with high accuracy and precision.[39] In addition, they performed longitudinal analysis on 10 out of 28 AML patients and revealed that this method can be reliably used for MRD detection of the vast majority of *FLT3*-ITD mutations accurately and deeply.[39] Nonetheless, more extensive testing of this approach is required to validate it for clinical MRD detection. Conversely, it would be better if the mutations in these two genes could be sequenced in gene panels rather than separately for ease of library preparation and detection of other mutation(s) that might emerge during disease course/CR and cause relapse.

The NGS technology is not yet optimal for MRD detection due to the fact that it has limited sensitivity in detecting mutation <0.1%. To improve the sensitivity and specificity of NGS technology for MRD detection in AML patients, a comprehensive targeted NGS panel with ECS feature is desirable. The design should contain all the driver mutations found in major studies in AML genomics such as the TCGA[3], and Papaemmanuil[1]. The design should exclude genes that are known to be associated with clonal hematopoiesis and shown to be not predictive of relapse. WGS or WES may be applied in the future for mutation detection at diagnosis, subsequently, deep sequencing focusing on those mutations should be performed at follow-ups.

Is there a role for WGS and WES in MRD detection?

WGS enables the detection of the entire range of genomic lesions, including point mutations, indels, copy number variations (CNV), and structural rearrangements such as translocations, inversions, cryptic and complex rearrangements.[40] On the other hand, WES is an approach to selectively sequence the coding region of the genome, accounting for only 1% of the genome.[40] These approaches could be considered ideal since all the possible variants could be detected in a single run; additionally, they could provide valuable information into clonal evolution during treatment on mutated genes beyond the panel to predict for impending relapse. However, currently, there are several limitations preventing WGS and WES to be used in routine analyses of malignancies and MRD surveillance in particular. These limitations for WES and WGS include i) the two technologies have a relatively low sensitivity (i.e., 30-60X and 100-500X, respectively), which may result in missing subclonal mutations at diagnosis (which may be predictive for impending relapse) and thus not optimal for MRD detection; ii) sequencing the entire genome or exome, limits the number of samples to be sequenced in a single run; (iii) consequently it is expensive (approximately $1000 to $2000 and $5000 - $15000 per sample, respectively); iv) the time taken until reporting the result to the clinic is long with the current processing time; v) it requires a high amount of starting

material (approximately 3 µg and 1µg of genomic DNA, respectively).[40] Both WGS and WES technologies are more powerful than targeted NGS; however, WGS and WES are still considered quite expensive for routine clinical diagnosis as well as MRD detection. Moreover, both approaches generate a massive amount of genome data that in turn requires large and probably multiple servers to store the data, but nevertheless, they could be of an additional clinical value.

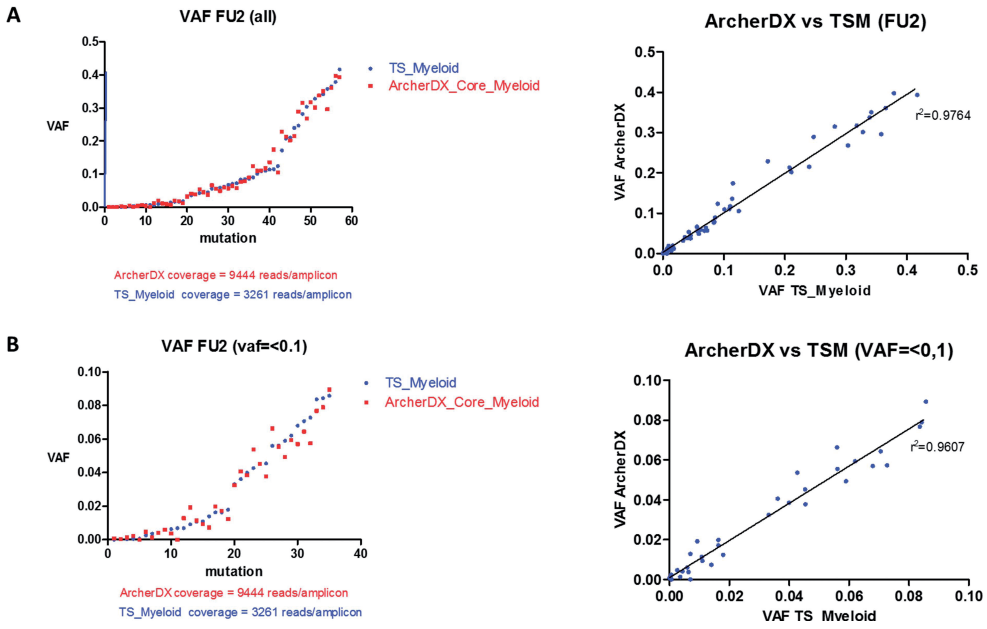


Figure 7: MRD with error-corrected NGS versus site-and-variant-specific error correction model. A) Correlation of variant allele frequency of all mutations detected by TruSight Myeloid Sequencing panel versus VariantPlex Core Myeloid Panel in 24 samples (n= 24). In the left graph, the mutations are highly concordant between the two panels, and in the right graph, there is a high correlation between the two panels. B) Same as in Figure 7A but focusing on mutations with low VAF (<0.1%). VAF: variant allele frequency; r^2: correlation coefficient.

Peripheral blood as an alternative source for NGS-based MRD assessment

PB is an attractive material for both patients and clinicians for NGS-based MRD tracking as an alternative for the more invasive BM aspiration. There are some studies that suggested similar findings using both sources of assay material[41, 42], however, most of the available clinical studies on the prognosis and predictive impact of MRD are based on BM specimens. That is due to the fact that BM likely provides additional sensitivity with detectable MRD levels of about 1-log higher than those of PB.[43] For instance, in CBF AML, an MRD negative PB assay might miss an MRD positive BM in up to 40% of cases during therapy and 10 to 15% of cases during follow-up.[44] However, Ivey *et al*.[45] used qPCR in a large study to monitor mutant *NPM1* levels in 2569 samples obtained from 346 patients with *NPM1* mutated AMLs treated with intensive chemotherapy in both BM and PB after each cycle of chemotherapy.[45] They demonstrated that association with survival was better when PB was used than when BM was used. This finding suggests that the source of MRD assessment is likely dependent on the assay, as well as regimen and time.[46] Due to the lower sensitivity of PB over BM there is a need to improve the sensitivity of NGS technology to detect residual disease by ECS. Besides, the European LeukemiaNet MRD Working Party recently produced a consensus on the frequency of monitoring MRD in PB which will allow early detection of residual disease.[43]

Hourigan and colleagues[27] used UMI ECS combined with ultra-deep sequencing to determine MRD status in frozen whole blood specimens of AML collected during CR before the conditioning regimen. The result they obtained in this study indicates that PB can also be used for NGS-based MRD assessment to predict patient's performance when a sensitive NGS method is used.[27] Our group also did an experiment on a small subset of AML patients in CR (n=30) and looked at NGS-based MRD in BM versus PB. We found that all mutations detected in BM were also detected PB; however, at lower VAF which indicates that the detection of MRD in PB is less sensitive (Figure 8). However, this approach remains to be proven with a large cohort and perhaps with more frequent sampling intervals as recommend by the ELN working party, and eventually, it has to be validated for routine clinical MRD tracking.

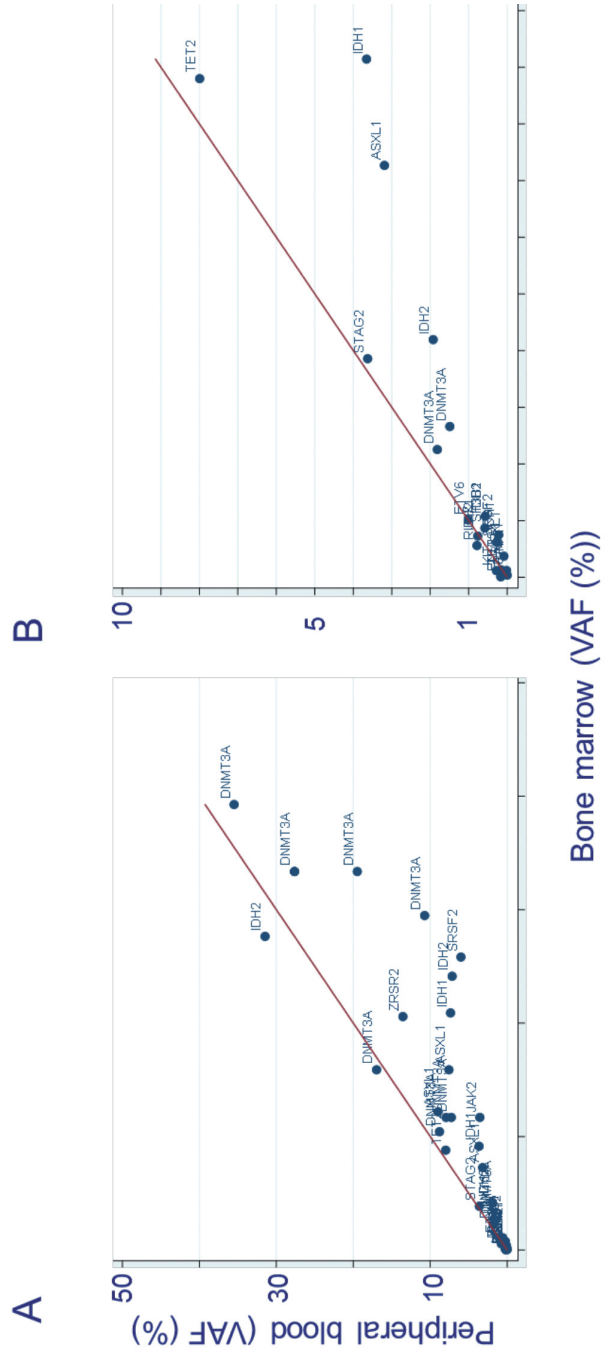


Figure 8: **NGS-based MRD detection in bone marrow versus peripheral blood**. A) Correlation of variant allele frequency of all mutations detected in BM and PB. B) Same as in Figure 8A but looking at mutations with variant allele frequencies.

Harmonization and standardization of NGS-based MRD detection

Currently, NGS based assays are becoming more routinely used in clinical diagnosis setting of AML and since NGS is hypothetically applicable to all AML patients, there is an urgent need to utilize NGS for MRD monitoring, because nowadays the most widely used molecular MRD technique (i.e., Quantitative real-time PCR) is applicable to approximately 40% of AML patients (i.e., CBF AMLs, *NPM1* mutant AML and *PML-RARA*).[43] MRD testing has many clinical values, including but not limited to determining the risk of relapse after induction therapy (as demonstrated in **Chapter 2**); determine prognosis and assignment to maintenance following completion of standard therapy, and the efficacy of the treatment.[43, 47]

A number of studies revealed the prognostic value of NGS MRD after induction therapy[21-23], pre- and post-stem cell transplantation[25]. However, prior to the implementation of MRD measurement by NGS in routine clinical practice, there should be harmonization to enable comparability of results. The harmonization will allow independent laboratories to continue using the platform and the methodologies developed and validated similarly to what has been done for major *BCR-ABL1* fusion gene transcript monitoring.[48] There should be consensus on the following simple and complex issues, but not limited to i) DNA input during follow-up (minimum requirement to detect residual disease with the appropriate sensitivity); ii) sampling time-points, sampling tissue (BM and/or PB); iii) molecular markers for tracking MRD (for instance focused on ELN 2017 recommendation or more), and single markers or combination, iv) sequencing depth and minimum variant reads; tissue/material of choice to distinguish somatic from germline mutations; v) methodologies to discriminate artifact from true mutations (ECS/UMI or computational approach); vi) adequately discriminate MRD from clonal hematopoiesis.

Conclusion and future perspective

Although the majority of AML patients achieve complete morphological remission, unfortunately, many patients will eventually relapse. Thus, suggesting that relying on pretreatment risk stratification only is insufficient to improve patient outcome. Therefore, there is still a great need for adequate prediction of impending relapse in order to adapt the treatment accordingly and improve the outcome of patients at risk of relapse. The work described in this thesis mainly deals with the question on how we can further refine and improve AML risk stratification focusing on MRD monitoring by NGS. Several conclusions can be drawn from this thesis; however, the main conclusions lay around the application of molecular MRD by NGS during CR. We demonstrated that targeted NGS-based identification of molecular MRD during CR was associated with an elevated risk of relapse and mortality over years of follow-up in patients with AML. Moreover, NGS has a profound independent prognostic value for patients with AML and NGS-based MRD is highly concordant with multiparameter flow cytometry for detection of persistent non-*DTA* mutations and provides an additive prognostic value. The new and important finding in **Chapter 2**, was the discovery

that detection of *DTA* mutations in CR have no association with the risk of relapse.[23] This discovery was confirmed by a number of later studies.[22, 27] However, we found that there are other mutated genes that persist in CR, but, in low frequency. This finding suggests that there might be additional mutated genes that are also associated with clonal hematopoiesis. This in turn means that these mutations may not be suitable for MRD monitoring. Unfortunately, due to the low prevalence of these mutations in our study cohort, their distinction was hindered and analyses in a larger cohort are warranted. Hence, increasing the cohort size and the sensitivity of NGS-based MRD will enable the discrimination of true residual disease from clonal hematopoiesis.

The mechanism by which mutations in genes associated with clonal hematopoiesis occur remains elusive and the same is true for late mutations that lead to the development of AML. Therefore, understanding the mechanism by which mutations in genes involved in DNA repair pathway such as MBD4 and TDG and their contribution to the development of mutations is another dimension that warrant investigation. Future plans in this part would include, i) increase patient's number with germline inactivation of MBD4 by identifying patients with colon and/or rectal polyps and screening for mutations in *MBD4* gene; ii) recapitulate what we found in the three AML patients in mice model with inactive MBD4 protein and the introduction of mutations in *DNMT3A*, *IDH* and *TDG* using CRISPR-cas9 technology; iii) measure the activity of MBD4 and TDG under various single and double mutant *DNMT3A* cells/mice/patient to understand which mutations contribute to AML development and or clonal hematopoiesis. Studying syndromes that predispose to AML may enhance our knowledge and understanding AML pathogenesis and might provide more insight on MRD tracking of patients with a predisposition to AML.

Although we successfully monitored MRD by NGS as shown in **Chapters 2 and 5**, the Illumina TruSight Myeloid NGS panel used in these studies was not developed for MRD detection and produces relatively high sequencing error rates. Thus, the Illumina TruSight Myeloid NGS panel is far from optimal when it comes to sensitivity and specificity. There is a need to develop an NGS-based MRD detection panel that has better sensitivity and specificity to enable correct discrimination between residual leukemia and clonal hematopoiesis. In addition, a larger panel is warranted to study mutations that are less frequent in AML in a large AML cohort.

Our future plan in improving NGS-based MRD detection would involve developing a comprehensive NGS AML panel that takes into account all relevant mutations in driver genes that have been reported in large cohort studies, such as the TCGA (n= 200), the 2016 Papaemmanuil *et al*. study (n= 1540) and also the well-annotated HOVON AML patients cohort (n=2600) to enable adequate detection of all relevant mutations in the vast majority of AML patients. The methodology of choice is a PCR-based approach which is already operational in our molecular diagnostic laboratory for the detection of *JAK2* (V617F), *JAK2* exon 12, *MPL* and *CALR* in MPN. This NGS approach has been proven to produce a significantly low

sequence error rate than the Illumina TruSight Myeloid NGS Panel. To enhance the sensitivity and specificity of this approach we would incorporate the ECS approach which will enable to distinguish true mutation from an artifact. This approach will ultimately enable us to detect and monitor MRD more accurately and precisely. Alternatively, using commercially available error-corrected sequencing kits, such as ArcherDx is also a valid option.

We aim also to use this custom NGS panel to investigate the sensitivity of NGS-based MRD detection in peripheral blood as an alternative to the more invasive BM aspiration in a large cohort of AML patients (approx. 400 patients). To further enable the detection of molecular MRD in virtually all AML, we aim to develop an RNASeq/cDNA approach to detect the various fusion gene transcripts such as *RUNX1-RUNX1T1* (5% in AML), *CBFB-MYH11* (5% in AML), *DEK-NUP214* (2% in AML) and rearrangement involving 11q23 (approx. 10% in AML). The result from peripheral blood MRD detection will be compared to the data obtained from bone marrow.

To further enhance MRD detection, we aim to explore the feasibility of NGS-based MRD detection by using single-cell analyses. So far NGS-based MRD detection have been applied on the bulk of AML cells, which lack the resolution and accuracy for characterizing small MRD subclones that may drive the impending relapse. We will apply single-cell gene expression analyses and genotyping to determine the identity of the cell and deconstruct the clonal architecture at AML diagnosis and MRD in CR (i.e., clonal hematopoiesis and residual leukemia). These analyses will reveal the cellular identity and molecular patterns of the cell populations persistent in CR that are responsible for AML relapse. Furthermore, these analyses will reveal whether single-cell MRD detection is superior to bulk AML cells MRD detection.

All the aforementioned future plans will enable and aid in better distinction of residual disease from clonal hematopoiesis as well as a true mutation from sequencing artifact. This will result in a significant improvement in molecular MRD detection specificity and sensitivity by NGS. To improve the clinical outcome of patients at risk of relapse, it is essential to perform longitudinal MRD detection which will definitely advance/refine relapse prediction to the level of an individual patient. For longitudinal MRD detection, using a less invasive material such as blood for longitudinal MRD detection by NGS will be more alluring to both patients and clinicians. Standardization and harmonization of NGS-based MRD detection across laboratories are paramount and will speed up the incorporation of this MRD tool into routine clinical practice.

REFERENCES

1. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND*, et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 2016 Jun 9; **374**(23)**:** 2209-2221.
2. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ*, et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 2013 May 30; **368**(22)**:** 2059-2074.
3. Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson AG*, et al.* Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *New England Journal of Medicine* 2013 May 30; **368**(22)**:** 2059-2074.
4. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF*, et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014 Dec 25; **371**(26)**:** 2477-2487.
5. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG*, et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *New England Journal of Medicine* 2014; **371**(26)**:** 2488-2498.
6. Swerdlow SH CE, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*, Revised 4th ed edn. IARC: Lyon, France, 2017, 586pp.
7. Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T*, et al.* Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 2017 Jan 26; **129**(4)**:** 424-447.
8. Verhaak RG, Wouters BJ, Erpelinck CA, Abbas S, Beverloo HB, Lugthart S*, et al.* Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* 2009 Jan; **94**(1)**:** 131-134.
9. Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J*, et al.* Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med* 2012 Mar 22; **366**(12)**:** 1079-1089.
10. Leisch M, Jansko B, Zaborsky N, Greil R, Pleyer L. Next Generation Sequencing in AML-On the Way to Becoming a New Standard for Treatment Initiation and/or Modulation? *Cancers (Basel)* 2019 Feb 21; **11**(2).
11. Shendure J, Ji H. Next-generation DNA sequencing. *Nature Biotechnology* 2008 2008/10/01; **26**(10)**:** 1135-1145.
12. Schuster SC. Next-generation sequencing transforms today's biology. *Nat Methods* 2008 Jan; **5**(1)**:** 16-18.
13. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC*, et al.* The Origin and Evolution of Mutations in Acute Myeloid Leukemia. *Cell* 2012 Jul 20; **150**(2)**:** 264-278.
14. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS*, et al.* Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 2012 Jan 26; **481**(7382)**:** 506-510.
15. Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK*, et al.* Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 2010 Jan 21; **115**(3)**:** 453-474.
16. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A*, et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 2009 Jul 30; **114**(5)**:** 937-951.
17. Shumilov E, Flach J, Kohlmann A, Banz Y, Bonadies N, Fiedler M*, et al.* Current status and trends in the diagnostics of AML and MDS. *Blood Reviews* 2018 Nov; **32**(6)**:** 508-519.
18. Kihara R, Nagata Y, Kiyoi H, Kato T, Yamamoto E, Suzuki K*, et al.* Comprehensive analysis of genetic alterations and their prognostic impacts in adult acute myeloid leukemia patients. *Leukemia* 2014 Aug; **28**(8)**:** 1586-1595.

19. Thol F, Kolking B, Damm F, Reinhardt K, Klusmann JH, Reinhardt D*, et al.* Next-generation sequencing for minimal residual disease monitoring in acute myeloid leukemia patients with FLT3-ITD or NPM1 mutations. *Genes Chromosomes Cancer* 2012 Jul; **51**(7)**:** 689-695.
20. Kohlmann A, Nadarajah N, Alpermann T, Grossmann V, Schindela S, Dicker F*, 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 Jan; **28**(1)**:** 129-137.
21. Klco JM, Miller CA, Griffith M, Petti A, Spencer DH, Ketkar-Kulkarni S*, et al.* Association Between Mutation Clearance After Induction Therapy and Outcomes in Acute Myeloid Leukemia. *Jama-J Am Med Assoc* 2015 Aug 25; **314**(8)**:** 811-822.
22. Morita K, Kantarjian HM, Wang F, Yan YQ, Bueso-Ramos C, Sasaki K*, et al.* Clearance of Somatic Mutations at Remission and the Risk of Relapse in Acute Myeloid Leukemia. *Journal of Clinical Oncology* 2018 Jun 20; **36**(18)**:** 1788-+.
23. Jongen-Lavrencic M, Grob T, Hanekamp D, Kavelaars FG, Al Hinai A, Zeilemaker A*, et al.* Molecular Minimal Residual Disease in Acute Myeloid Leukemia. *N Engl J Med* 2018 Mar 29; **378**(13)**:** 1189-1199.
24. Getta BM, Devlin SM, Levine RL, Arcila ME, Mohanty AS, Zehir A*, et al.* Multicolor Flow Cytometry and Multigene Next-Generation Sequencing Are Complementary and Highly Predictive for Relapse in Acute Myeloid Leukemia after Allogeneic Transplantation. *Biol Blood Marrow Tr* 2017 Jul; **23**(7)**:** 1064-1071.
25. Thol F, Gabdoulline R, Liebich A, Klement P, Schiller J, Kandziora C*, et al.* Measurable residual disease monitoring by NGS before allogeneic hematopoietic cell transplantation in AML. *Blood* 2018 Oct 18; **132**(16)**:** 1703-1713.
26. Press RD, Eickelberg G, Froman A, Yang F, Stentz A, Flatley EM*, et al.* Next-generation sequencing-defined minimal residual disease before stem cell transplantation predicts acute myeloid leukemia relapse. *American Journal of Hematology* 2019 Aug; **94**(8)**:** 902-912.
27. Hourigan CS, Dillon LW, Gui G, Logan BR, Fei M, Ghannam J*, et al.* Impact of Conditioning Intensity of Allogeneic Transplantation for Acute Myeloid Leukemia With Genomic Evidence of Residual Disease. *J Clin Oncol* 2020 Apr 20; **38**(12)**:** 1273-1283.
28. Kim T, Moon JH, Ahn JS, Kim YK, Lee SS, Ahn SY*, et al.* Next-generation sequencing-based posttransplant monitoring of acute myeloid leukemia identifies patients at high risk of relapse. *Blood* 2018 Oct 11; **132**(15)**:** 1604-1613.
29. Zink F, Stacey SN, Norddahl GL, Frigge ML, Magnusson OT, Jonsdottir I*, et al.* Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* 2017 Aug 10; **130**(6)**:** 742-752.
30. Yoest JM, Shirai CL, Duncavage EJ. Sequencing-Based Measurable Residual Disease Testing in Acute Myeloid Leukemia. *Frontiers in Cell and Developmental Biology* 2020 2020-May-08; **8**(249).
31. Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL*, et al.* Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood* 2015 Feb 26; **125**(9)**:** 1367-1376.
32. Hourigan CS, Gale RP, Gormley NJ, Ossenkoppele GJ, Walter RB. Measurable residual disease testing in acute myeloid leukaemia. *Leukemia* 2017 Jul; **31**(7)**:** 1482-1490.
33. Roloff GW, Lai C, Hourigan CS, Dillon LW. Technical Advances in the Measurement of Residual Disease in Acute Myeloid Leukemia. *J Clin Med* 2017 Sep 19; **6**(9).
34. Levine RL, Valk PJM. Next generation sequencing in diagnosis and MRD assessment of AML. *Haematologica* 2019**:** haematol.2018.205955.
35. Salk JJ, Schmitt MW, Loeb LA. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nature Reviews Genetics* 2018 May; **19**(5)**:** 269-285.

36. Balagopal V, Hantel A, Kadri S, Steinhardt G, Zhen C, Kang W*, et al.* Measurable Residual Disease Monitoring for Patients with Acute Myeloid Leukemia Following Hematopoietic Cell Transplantation Using Error-Corrected Hybrid-Capture Next-Generation Sequencing. *Journal of Molecular Diagnostics* 2019 Nov; **21**(6)**:** 1149-1149.
37. Stirewalt DL, Meshinchi S, Kussick SJ, Sheets KM, Pogosova-Agadjanyan E, Willman CL*, et al.* Novel FLT3 point mutations within exon 14 found in patients with acute myeloid leukaemia. *Brit J Haematol* 2004 Feb; **124**(4)**:** 481-484.
38. Small D. FLT3 mutations: biology and treatment. *Hematology Am Soc Hematol Educ Program* 2006**:** 178-184.
39. Blatte TJ, Schmalbrock LK, Skambraks S, Lux S, Cocciardi S, Dolnik A*, et al.* getITD for FLT3-ITD-based MRD monitoring in AML. *Leukemia* 2019 Oct; **33**(10)**:** 2535-2539.
40. Riva L, Luzi L, Pelicci PG. Genomics of acute myeloid leukemia: the next generation. *Front Oncol* 2012; **2:** 40.
41. Maurillo L, Buccisano F, Spagnoli A, Del Poeta G, Panetta P, Neri B*, et al.* Monitoring of minimal residual disease in adult acute myeloid leukemia using peripheral blood as an alternative source to bone marrow. *Haematol-Hematol J* 2007 May; **92**(5)**:** 605-611.
42. Zeijlemaker W, Kelder A, Oussoren-Brockhoff YJM, Scholten WJ, Snel AN, Veldhuizen D*, et al.* Peripheral blood minimal residual disease may replace bone marrow minimal residual disease as an immunophenotypic biomarker for impending relapse in acute myeloid leukemia. *Leukemia* 2016 Mar; **30**(3)**:** 708-715.
43. Schuurhuis GJ, Heuser M, Freeman S, Bene MC, Buccisano F, Cloos J*, et al.* Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. *Blood* 2018 Mar 22; **131**(12)**:** 1275-1291.
44. Yin JA, O'Brien MA, Hills RK, Daly SB, Wheatley K, Burnett AK. Minimal residual disease monitoring by quantitative RT-PCR in core binding factor AML allows risk stratification and predicts relapse: results of the United Kingdom MRC AML-15 trial. *Blood* 2012 Oct 4; **120**(14)**:** 2826-2835.
45. Ivey A, Hills RK, Simpson MA, Jovanovic JV, Gilkes A, Grech A*, et al.* Assessment of Minimal Residual Disease in Standard-Risk AML. *N Engl J Med* 2016 Feb 4; **374**(5)**:** 422-433.
46. Short NJ, Ravandi F. How close are we to incorporating measurable residual disease into clinical practice for acute myeloid leukemia? *Haematologica* 2019 Jul 31; **104**(8)**:** 1532-1541.
47. Hourigan CS, Karp JE. Minimal residual disease in acute myeloid leukaemia. *Nature Reviews Clinical Oncology* 2013 2013/08/01; **10**(8)**:** 460-471.
48. Hughes T, Deininger M, Hochhaus A, Branford S, Radich J, Kaeda J*, et al.* Monitoring CML patients responding to treatment with tyrosine kinase inhibitors: review and recommendations for harmonizing current methodology for detecting BCR-ABL transcripts and kinase domain mutations and for expressing results. *Blood* 2006 Jul 1; **108**(1)**:** 28-37.

Addendum

A

Addendum

NEDERLANDSE SAMENVATTING

Acute Myeloïde Leukemie (AML) is een heterogene groep hematologische aandoeningen met verschillende behandeluitkomsten, hetgeen toe te wijzen is aan de intrinsieke genetische complexiteit die reeds aanwezig is bij diagnosis en voortkomend uit evolutie tijdens progressie van de ziekte. Deze genetische afwijkingen hebben een sterke invloed op zowel de respons van patiënten op de behandeling en de hierbij behorende prognose. De opkomst van Next Generation Sequencing (NGS) technieken heeft de kennis naar de pathogenese, heterogeniteit, en klinische uitkomst in AML-patiënten sterk verbeterd, en heeft daarnaast bijgedragen aan het identificeren van vroege genetische laesies die geassocieerd zijn met leeftijd-gerelateerde klonale hematopoëse, ook bekend als 'clonal hematopoiesis of indeterminate potential (CHIP)'. Door de ontwikkelingen in het genetisch onderzoek binnen de oncologie, en in het bijzonder binnen studies naar AML, zijn er in 2016 verbeteringen aangebracht in de WHO classificatie van AML, en is in 2017 de European Leukemia Net (ELN) risico stratificatie herzien en bijgewerkt.

In hoofdstuk 2 wordt de toepassing van NGS-methoden voor het detecteren van moleculaire restziekte (minimal/measurable residual disease (MRD)) in AML beschreven. In een groot cohort van volwassen AML-patiënten (n = 482), hebben we m.b.v. een targeted NGS genpanel analyses uitgevoerd op AML-samples afgenomen bij diagnose en tijdens complete hematologische remissie (CR), bereikt na twee cycli intensieve inductie chemotherapie. Hiermee werd de aan- en afwezigheid van mutaties bepaald in 54 genen, waarvan bekend is dat deze gemuteerd kunnen zijn in myeloïde maligniteiten. De eindpunten van de studie waren het optreden van een recidief binnen 4 jaar, recidiefvrije overleving en algehele overleving. Het cohort was opgedeeld in een training- (283 AML patiënten) en validatie set (147 AML patiënten), beiden vergelijkbaar op klinisch, cytogenetisch en moleculair gebied. Tijdens diagnose werden er mutaties gedetecteerd in 430 van de 482 patiënten (89,2%), waarvan 51.4% persisterend in CR. Het persisteren van mutaties was zeer variabel tussen genen, dikwijls bij een hoge variant allel frequentie (VAF), variërend tussen 0,02 en 47%. Mutaties in genen geassocieerd met CHIP (*DNMT3A, TET2*, en *ASXL1;* afgekort *DTA* mutaties) werden vaak gedetecteerd met een hoge VAF in CR, echter de aanwezigheid van deze mutaties in CR hadden geen prognostische waarde in zowel de training als de validatie set. Daartegenover was de detectie van andere mutaties, de non-*DTA* mutaties, in CR significant geassocieerd met een verhoogd risico op een recidief, wat ook bevestigd werd in de validatie set. Hiermee in overeenstemming hebben we tevens laten zien dat het persisteren van non-*DTA* mutaties was geassocieerd met een kortere recidiefvrije- en algehele overleving. Dit hield stand na het uitvoeren van een multivariate analyse, waarbij gecorrigeerd was voor relevante prognostische factoren waaronder leeftijd, witte bloedcellen, ELN risico classificatie 2017, en het aantal cycli inductie chemotherapie. Daarnaast hebben we een vergelijking gemaakt

tussen het detecteren van restziekte door het identificeren van persisterende non-*DTA* mutaties met NGS en leukemisch geassocieerde immunofenotypes met multiparameter flow cytometrie in 340 AML patiënten. Na het uitvoeren van een multivariate analyse, toonden we aan dat de combinatie van NGS en flow cytometrie de sterkste onafhankelijke prognostische waarde had voor het detecteren van restziekte bij alle eindpunten. Samengevat, in deze studie naar MRD-detectie met NGS-technieken, hebben we kunnen aantonen dat persisterende mutaties in genen die geassocieerd zijn met pre-leukemie (*DTA* mutaties) in CR niet voorspellend waren voor het ontwikkelen van een recidief. Daarnaast hebben we laten zien dat MRD-detectie met NGS een sterke prognostische techniek is, die van toepassing kan zijn op een groot deel van de AML-patiënten.

In hoofdstuk 3 worden de mutationele en klonale kenmerken van AML met een gemuteerd *TP53* gen in kaart gebracht d.m.v. het sequensen van een groot cohort AML-patiënten (n = 2200). Een mutatie in *TP53* was hierbij gedetecteerd in 10,5% (230 van de 2200 AML-patiënten), waarbij de meerderheid bestond uit bi-allelische *TP53* mutaties. In 49% (113/230) van deze patiënten werd een co-mutatie gevonden, waarbij de meest frequent voorkomende mutaties in genen waren die geassocieerd zijn met klonale hematopoiese, namelijk in *DNMT3A, TET2,* en *ASXL1*. Daarnaast vertoonde 84% van de AML-patiënten met een *TP53* mutatie genomische instabiliteit in de vorm van een complex karyotype. Opvallend hierbij was dat genomische instabiliteit vaker voorkwam in AML met bi-allelische *TP53* mutaties (97,1%) ten opzichte van AML met mono-allelische mutaties (40%). Betreffende overleving was er geen significant verschil tussen verschillende *TP53* VAF drempelwaardes, getest van 50% tot 5%, en daarnaast was er geen positief effect op algehele overleving voor AML-patiënten met mono-allelische *TP53* mutaties. Aangezien de meerderheid van de co-mutaties voorkomend met *TP53* mutaties waren geassocieerd met klonale hematopoiese, werd voor het vervolg *TP53* beschouwd als de enige geschikte marker voor moleculaire MRD-detectie d.m.v. deep sequencing NGS. Van de *TP53* mutaties gedetecteerd bij diagnose, was 73% persisterend in CR, echter was de detectie hiervan niet geassocieerd met een verhoogde kans op recidief (p = 0,911) of een verlaagde overleving (p = 0,653). Hoewel de associatie tussen *TP53* mutaties en een complex karyotype al bekend is, heeft deze studie aangetoond dat de aanwezigheid van genomische instabiliteit sterk geassocieerd is met een verlaagde overleving in deze patiënten (2 jaars overleving van 34% t.o.v. 9% in AML-patiënten met een complex karytopye; p = 0,002). Daarnaast, toont deze studie aan dat patiënten met een mono-allelische *TP53* mutatie zonder complex karyotype een betere algehele overleving hebben dan patiënten met een complex karyotype (p = 0,001), wat impliceert dat de mate van genomische instabiliteit in *TP53* gemuteerde AML-patiënten van prognostische waarde is, ongeacht de *TP53* allel status. Om te achterhalen welke patiënten baat hebben bij een allogene stamceltransplantatie (HSCT), werd de prognostische waarde bepaald van genomische instabilitieit van 59 AML-patiënten met een *TP53* mutatie die

ook een allogene HSCT hadden ontvangen. Hieruit kwam naar voren dat in deze patiënten genomische instabiliteit was geassocieerd met een kortere overleving (2 jaars overleving ratio, 60% genomisch stabiel vs. 16,7% genomisch instabiel; HR 2,44, 95% CI: 1,03 – 5,79; p = 0,02), bevestigd in een multivariate analyse.

In hoofdstuk 4 wordt de aanwezigheid van *PPM1D* mutaties tijdens diagnose, in CR, en tijdens recidief onderzocht. We tonen aan dat mutaties in *PPM1D* mutaties zelden voorkomen in *de novo* AML en refractoire anemie met een overmaat aan blasten (refractory anemia with excess of blasts; RAEB) in tegenstelling tot therapie-gerelateerde AML. Daarnaast komen *PPM1D* mutaties vaker voor na hoge dosis chemotherapie met een relatief lage VAF, en maken ze geen deel uit van het AML-recidief. In conclusie, aangezien *PPM1D* mutaties niet zijn geassocieerd met een opkomend recidief zijn deze mutaties niet bruikbaar als MRD-marker in AML.

In hoofdstuk 5 wordt de haalbaarheid getest van het gebruik van beenmerg-glaasjes gekleurd met May-Grünwald Giemsa (MGG) voor de detectie van mutaties m.b.v. NGS. We hebben aangetoond dat isolatie van DNA efficiënter verloopt in MGG-gekleurde glaasjes dan bij gearchiveerde ongekleurde glaasjes, doordat in de meeste gevallen de mutatie profielen afkomstig van MGG-gekleurde beenmerg monsters identiek waren aan de mutaties gevonden met DNA afkomstig van Ficoll gezuiverd beenmerg (FPBM) bij zowel diagnose als in recidief. Daarnaast onderzochten we de mogelijkheid om driver mutaties te detecteren in DNA afkomstig van oudere MGG-gekleurde glaasjes, geprepareerd in 1975. We konden hierbij niet alleen aantonen dat de mutaties detecteerbaar waren, maar ook dat de VAFs van deze mutaties zeer vergelijkbaar waren met de VAFs van de mutaties gedetecteerd in DNA afkomstig van FPBM (R2: 0,87). MGG-gekleurde beenmerg glaasjes zijn een goede en betrouwbare bron om mutaties te meten in AML. Naast het aantonen van de mogelijkheid tot het detecteren van mutaties met een NGS-methode gebaseerd op hybridisatie (met het Illumina TruSight Myeloid sequensing panel), konden we ook laten zien dat deze mutaties detecteerbaar waren met een op maat gemaakt NGS genpanel gebaseerd op PCR, bij zowel bij diagnose en recidief, en tijdens behandeling. Tevens werd de kinetiek van de mutaties gedurende behandeling onderzocht m.b.v. MGG-gekleurd beenmerg DNA bij 18 AML patiënten. De mutatie profielen waren in lijn met de behandeling, waarbij er duidelijk onderscheid kon worden gemaakt tussen klonale hematopoiese en leukemie. Hieruit volgt dat gearchiveerde MGG-gekleurde beenmerg glaasjes uitstekend kunnen fungeren als bron voor retrospectieve moleculaire analyses naar AML gedurende het ziekteverloop. Deze benadering zou ook kunnen gebruikt in vergelijkbare studies aan andere hematologische aandoeningen. Deze studie biedt daarom nieuwe mogelijkheden voor het onderzoeken van de moleculaire kinetiek van mutaties vooral in patiënten waar van gelimiteerd testmateriaal beschikbaar is. Daarnaast heeft deze aanpak de potentie om vroege mutaties, vaak

geassocieerd met CHIP te kunnen onderscheiden van de oncogene transformerende driver mutaties.

In hoofdstuk 6 wordt in een subset van AML-patiënten met een *KMT2A-PTD* mutatie gekeken naar het spectrum van mutaties, gen expressie profielen, en prognoses en vergeleken met een cohort AML-patiënten waarbij deze mutatie niet voorkomt. Hierbij konden we aantonen dat *KMT2A-PTD*s aanwezig waren in 5,5% van alle AML gevallen, en dat ze significant geassocieerd waren met een gelijktijdig voorkomende trisomie van chromosoom 11, zoals vermeld in eerdere studies. Gekeken naar de mutatie profielen in AML-patiënten met en zonder *KMT2A-PTD* mutaties, waren er een aantal genen significant vaker gemuteerd in *KMT2A-PTD* AML, waaronder *FLT3-ITD, IDH1, U2AF1*, en *IDH2*. Hiertegenover, werden mutaties in *NPM1*, *TP53*, en *NRAS* significant minder vaak gevonden in deze patiënten. Met ons eerder gepubliceerde genexpressie dataset, hebben we laten zien dat meerdere homeobox-gerelateerde gen families consistent over-expressie vertoonden in patiënten met een *KMT2A-PTD* mutatie. Daaropvolgend, gebruikmakend van een associatie model waarbij een groot aantal relevante klinisch- en genetisch gedefinieerde subsets van AML werden meegenomen, hebben we laten zien dat *KMT2A-PTD*s een over-expressie in een subset van HOX-gerelateerde genen veroorzaakt op een andere manier dan t(11q23)-gerelateerde KMT2A fusie eiwitten. Dit kan erop duiden dat *KMT2A*-PTD leukemogenese aanstuurt via een ander mechanisme. Daarnaast hebben we aangetoond dat *KMT2A*-PTD AML met een bijbehorende *DNMT3A* of *NRAS* mutatie geassocieerd is met een verslechterd klinische uitkomst. Deze gedetailleerde studie aan een subset van AML geeft hierdoor meer inzicht in de onderliggende biologie en klinische uitkomst van *KMT2A-PTD* AMLs en laat zien dat dit type AML zich onderscheid van andere 11q23 gemuteerde AMLs.

In hoofdstuk 7 wordt een studie beschreven naar familiale gevallen met een kiembaan inactivatie van het Methyl-binding Domain 4 (MBD4) gen, geassocieerd met de ontwikkeling van een een specifiek type AML, welke ontstaat als gevolg van pathogene mutaties in driver genen, voornamelijk in *DNMT3A*. Onze resultaten geven meer inzicht in de rol die MBD4 speelt in de bescherming tegen 5-methylcytosine (5mC) geïnduceerde DNA schade. We hebben aangetoond wat de impact is van een constitutieve inactivatie van MBD4 op de ontwikkeling van voorloper AML door een extreme gevoeligheid voor DNA methylatie schade. Daarnaast konden we aantonen dat er een sterke link is tussen methylatie schade en de ontwikkeling van een moleculaire gedefinieerde hematologische maligniteit, AML met bi-allelische *DNMT3A* mutaties, en mutaties in *IDH1* of *IDH2*. Onze studie onthuld ook dat patiënten met een kiembaan bi-allelische MBD4 mutatie een versnelde klonale groei van cellen hebben die sterk geassocieerd is met klonale hematopoiese tientallen jaren eerder dan gezonde personen. We hebben hiermee een compleet nieuw kiembaan syndroom ontdekt met een predispositie voor een myeloide maligniteit. Hoewel dit syndroom zeldzaam lijkt te

zijn kan het in potentie belangrijke inzichten geven in de etiologie van klonale hematopoiese en de daaruit volgende ontwikkeling tot AML.

In hoofdstuk 8, de algemene discussie, worden de resultaten van alle hoofdstukken besproken en vergeleken met de huidige kennis en nieuwe ontwikkelingen betreffende moleculaire analyses aan en minimale restziekte detectie voor patiënten met AML.

LIST OF ABBREVIATIONS

ABL1	c-Abl oncogene1
Allo-HSCT	Allogeneic hematopoietic stem cell transplantation
AML	Acute myeloblastic/myeloid leukemia
AML1	Acute myeloid leukemia 1
APL	Acute promyelocytic leukemia
AR	Allelic ratio
ARCH	Age-Related Clonal Hematopoiesis
ASXL1	Additional sex combs like 1(Drosophila)
ATO	arsenic trioxide
ATRA	all-trans retinoic acid
Auto-HSCT	Autologus hematopoietic stem cell transplantation
BCOR	BCL6 corepressor
BCR	Breakpoint cluster region
BER	Base excision repair
BM	Bone marrow
bp	base pair
C	Cytosine
CALR	Calreticulin
CBF	Core binding factor
CBFB	Core binding factor beta subunit
CBL	Casitas B-lineage Lymphoma
CEBPA	CCAAT/enhancer binding protein alpha
CHEK2	Checkpoint Kinase 2
CHIP	Clonal Hematopoiesis of Indeterminate Potential
CI	Confidence interval
CIR	Cumulative incidence of relapse
CK	Complex Karyotype
CLP	Common lymphoid progenitor
CML	Chronic myeloid leukemia
CMML	Chronic myelomonocytic leukemia
CMP	Common myeloid progenitor
CN	Cytogenetic normal
CNV	copy number variation
CR	Complete remission
DDR	DNA damage response
DEK	DEK Proto-Oncogene
DNA	Deoxyribonucleic acid
DNMT3A	DNA methyltransferase 3A

dNTP	Deoxyribonucleotide triphosphate
DTA	DNMT3A, TET2 and ASXL1
ECS	Error corrected sequencing
EFS	Event-free survival
ELN	European LeukemiaNet
ErP	erythrocyte precursors
ETO	Eight twenty one
EVI1	Ectopic virus integration site 1
EZH2	Enhancer of zeste homolog 2
FAB	French-American-British
Fig	Figure
FISH	Flourescence in situ hybridization
FLT3	fms-related tyrosine kinase 3
FPBM	Ficoll-purified BM
GATA2	GATA binding prtein 2
gDNA	genomic DNA
GMP	Granulocyte monocyte progenitor
GMP	Granulocyte macrophage precursor
HOVON	Dutch-Belgian Hemato-Oncology Cooperative Group
HOVON	Dutch-Belgian Cooperative Trial Group for Hematology-Oncology
HOX	Homeobox
HR	Hazard ratio
HSC	Hematopoietic stem cell
HSCP	Hematopoietic stem cell progenitor
IDH1	Isocitrate dehydrogenase 1
IDH2	Isocitrate dehydrogenase 2
Indel	Insertion or deletion
ITD	Internal tandem duplication
JAK2	Janus kinase 2
KMT2A	lysine (K)-specific methyltransferase 2A
KRAS	V-Ki-ras2 Kirsten rat sarcoma viral oncogen homolog
LAIP	leukemia-associated immunophenotype
LT-HSC	long-term Hematopoietic stem cell
MBD4	methylbinding domain 4
MDS	Myelodysplastic syndrome
MECOM	MDS1 and EVI1 complex locus
MEP	Megakaryocyte erythrocyte precursor
MFC	Multicolor Flow Cytometry
MgCl2	Magnesium chloride

MGG-stained BM	May-Grünwald Giemsa stained bone marrow slides
MKL1	Megakaryoblastic Leukemia 1
MkP	Megakaryocyte precursor
MLF1	Myeloid Leukemia Factor 1
MLL	Mixed lineage leukemia
MLLT3	Mixed-Lineage Leukemia (Trithorax Homolog); Translocated To, 3
MPFC	Multiparameter Flow Cytometry
MPN	Myeloproliferative Neoplasms
MPP	Multipotent progenitor
MRD	Minimal/Measurable Residual Disease
mRNA	Messenger RNA
MYH11	Myosine heavy chain 11
NF1	Neurofibromatosis type 1
NGS	Next-generation sequencing
NK	Normal karyotype
NPM1	Nucleophosmin 1
NRAS	Neuroblastoma RAS viral oncogene homolog
NS	Not significant
NSD1	Nuclear Receptor SET Domain-Containing Protein 1
NUP214	Nucleoporin 214
NUP98	Nucleoporin 98
OS	Overall survival
p	Short arm of a chromosome
P	*P*-value
PB	Peripheral blood
PCR	Polymerase chain reaction
PHD	Plant homeodomain
PML	Promyelocytic leukemia
PPM1D	Protein Phosphatase, Mg2+/Mn2+ Dependent 1D
PTD	Partial tandem duplication
PTPN11	Protein Tyrosine Phosphatase Non-Receptor Type 11
q	Long arm of a chromosome
qPCR	Quantitative PCR
RAEB	Refractory anemia with excess of blast
RARA	Tetionic acid receptor, alpha
RBM15	RNA Binding Motif Protein 15
RFS	Relapse-free survival
RIC	Reduced intensity conditioning
RNA	Ribonucleic acid

RQ-PCR	Real-time quantitative PCR
RQ-PCR	Real-time Quantitative PCR
RRBS	Reduced representation bisulfite sequencing
RTK	Receptor tyrosince kinase
RT-PCR	Reverse transcription PCR
RUNX1	Runt-related transcription factor 1
RUNX1T1	RUNX1 translocation partner 1
SAKK	Swiss Group for Clinical Cancer Research
s-AML	Secondary Acute Myeloid Leukemia
SCDC	Single-cell-derived colonies
SCT	Stem Cell Transplantation
SF3B1	Splicing factor 3B subunit 1
SMC1A	Structural Maintenance Of Chromosomes 1A
SMC3	Structural Maintenance Of Chromosomes 3
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variation
SRSF2	Serine And Arginine Rich Splicing Factor 2
STAG2	Stromal Antigen 2
ST-HSCs	short-term Hematopoietic stem cell
SV	Structural variant
T	Thymidine
t-AML	therapy-related Acute Myeloid Leukemia
TCGA	The cancer genome atlas
TDG	Thymine DNA glycosylase
TET2	Tet methylcytosine dioxygenase 2
TKD	Tyrosine kinase domain
t-MDS	therapy-related Myelodysplastic Syndrome
TP53	Tumor Protein P53
U2AF1	U2 Small Nuclear RNA Auxiliary Factor 1
UMI	Unique molecular identifier
VAF	Variant allele frequency
vs.	Versus
WBC	White blood count
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World health organization
WT1	Wilms' tumor suppressor gene1
ZRSR2	Zinc Finger CCCH-Type, RNA Binding Motif And Serine/Arginine Rich 2
5mC	5-methylcytosine

CURRICULUM VITEA

Adil Salim Abdullah al Hinai was born on the 1st of April 1982 in Muscat, Sultanate of Oman. After obtaining his high secondary school certificate he studied Medical Laboratory Sciences at the Institute of Health Sciences – Ministry of Health, Oman. After completion of his diploma study in 2003, he had a one-year internship where he got exposed to various medical laboratory disciplines/departments, such as Hematology, Biochemistry, Histopathology, Microbiology, Blood Transfusion, and Virology. This was done at different levels of health institutions and specialized healthcare centers in Oman (secondary and tertiary health institutions, Central Public Health Laboratories, and Central Blood Bank). When he completed his internship program he was appointed as a laboratory technologist in the Human Genetic Department at Central Public Health Laboratories.

In 2004, he got a scholarship to obtain a bachelor in science (B.Sc) degree at Cardiff Metropolitan University (previously known as University of Wales, Institute Cardiff) in Cardiff, United Kingdom. He completed his B.Sc (Hons.) Biomedical Science with a graduation project entitled "The association between CD44 and CD47 with red cell membrane integrity and Rh genotype". Subsequently, he did his master of science (M.Sc) in Cancer Cell and Molecular Biology at the University of Leicester in Leicester, United Kingdom. During his M.Sc he investigated "the mechanism of anticancer drug-induced cell death using an RNAi approach" under the supervision of Dr. Raj Patel and graduated with a Merit.

After obtaining his master's degree he returned to Oman where he was given the responsibility to establish molecular diagnostics of hematological malignancies within the Genetics Center, Muscat. He underwent training in several molecular diagnostics laboratories abroad, including the Hannover Medical School, Hannover, Germany; the Charité Campus Virchow Clinic, Berlin, Germany; the London Royal Hospital, London, United Kingdom, and Erasmus University Medical Center, Rotterdam, the Netherlands. In 2014, he got a scholarship from the Ministry of Health, Oman to pursue a doctoral program.

In August 2014, he joined Dr. Peter J.M. Valk's research group in the Hematology Department at the Erasmus University Medical Center, Rotterdam as a Ph.D candidate under the supervison of Dr. Peter J.M. Valk (promotor Prof. Dr. Ruud Delwel).

PHD PORTFOLIO

Name Ph.D student:	Adil S.A. Al Hinai	**Ph.D period:**	Aug 2014 – Nov 2021
Erasmus MC Department:	Hematology	**Promotor:**	prof.dr. H.R. Delwel
Research School:	Molecular Medicine	**Co-promotor:**	dr. P.J.M. Valk

Phd training	**Year**	**ECTS**
General courses		
Workshop on Photoshop and Illustrator CS5 for PhD-students and other researchers.	2014	0.30
Workshop on Microsoft Access 2010: Basic	2015	0.30
Workshop on Microsoft Access 2010: Advanced	2015	0.40
Browsing Genes and Genomes with UCSC (Basic)	2015	0.35
Browsing Genes and Genomes with UCSC (Advance)	2015	0.35
Workshop on InDesign for PhD-students and other researchers	2015	0.15
Workshop on Microsoft Excel 2010: Advanced	2016	0.4
Basic Introduction Course on SPSS	2017	1.00
In-depth Courses & Workshops		
Genomic Resequencing in Medical Diagnostics	2015	1.00
Course on R	2015	1.40
Course on Molecular Diagnostics X	2016	1.00
Survival Analysis Course	2018	0.70
Course and Master Classes on Molecular Aspects of Hematological Disorders (3×)	2016-2019	2.10
Scientific Meetings at Department of Hematology		
Work discussion (weekly)	2014-2019	8.00
Erasmus Hematology Lectures (monthly)	2014-2019	2.00
Ph.D lunch with invited speaker (monthly)	2014-2019	2.50
Journal club (bi-monthly)	2014-2019	7.00
National & International Conferences		
Dutch Hematology congress	2015	0.30
Molecular Medicine Day (5×)	2015 - 2019	3.50
Annual Conference of European Hematology Association (2×)	2017 & 2019	2.00
Annual Conference of American Society of Hematology (2×)	2017 & 2018	2.00
MODHEM meeting (half annual meetings) (5×)	2016-2019	1.5
Cancer Institue Research day	2018	0.70
Presentations		
Departmental floor discussions (oral 8×)	2014 - 2019	4.00
Departmental journal clubs (oral 2×)	2014 - 2019	1.00
Annual Conference of European Hematology Association (Poster, 1×)	2016	1.00
Annual Conference of American Society of Hematology (Poster, 1×)	2017 - 2018	1.00
Total ECTS		**43.05**

LIST OF PUBLICATIONS

1 C.M. Vonk*, **A.S.A. Al Hinai***, D. Hanekamp and P.J.M. Valk, Molecular minimal residual disease detection in acute myeloid leukemia. Cancers, *in press*. *Shared authors

2 Cucchi, D.G.J., Van Alphen, C., Zweegman, S., Van Kuijk, B., Kwidama, Z.J., **Al Hinai, A.**, Henneman, A.A., Knol, J.C., Piersma, S.R., Pham, T.V., Jimenez, C.R., Cloos, J. & Janssen, J. (2021) Phosphoproteomic Characterization of Primary AML Samples and Relevance for Response Toward FLT3-inhibitors. Hemasphere, 5, e606.

3 **Al Hinai, A.S.A.**, Grob, T., Rijken, M., Kavelaars, F.G., Zeilemaker, A., Erpelinck-Verschueren, C.A.J., Sanders, M.A., Lowenberg, B., Jongen-Lavrencic, M. & Valk, P.J.M. (2021) PPM1D mutations appear in complete remission after exposure to chemotherapy without predicting emerging AML relapse. Leukemia, 35, 2693-2697.

4 **Al Hinai, A.S.A.**, Grob, T., Kavelaars, F.G., Rijken, M., Zeilemaker, A., Erpelinck-Verschueren, C.A.J., Gussinklo, K.J., Sanders, M.A., van Lom, K. & Valk, P.J.M. (2020) Archived bone marrow smears are an excellent source for NGS-based mutation detection in acute myeloid leukemia. Leukemia, 34, 2220-2224.

5 **Al Hinai, A.**, Pratcorona, M., Grob, T., Kavelaars, F.G., Bussaglia, E., Sanders, M.A., Nomdedeu, J. & Valk, P.J.M. (2019) The Landscape of KMT2A-PTD AML: Concurrent Mutations, Gene Expression Signatures, and Clinical Outcome. Hemasphere, 3, e181.

6 Sanders, M.A., Chew, E., Flensburg, C., Zeilemaker, A., Miller, S.E., **Al Hinai, A.S.**, Bajel, A., Luiken, B., Rijken, M., McLennan, T., Hoogenboezem, R.M., Kavelaars, F.G., Frohling, S., Blewitt, M.E., Bindels, E.M., Alexander, W.S., Lowenberg, B., Roberts, A.W., Valk, P.J.M. & Majewski, I.J. (2018) MBD4 guards against methylation damage and germ line deficiency predisposes to clonal hematopoiesis and early-onset AML. Blood, 132, 1526-1534.

7 Jongen-Lavrencic, M., Grob, T., Hanekamp, D., Kavelaars, F.G., **Al Hinai, A.**, Zeilemaker, A., Erpelinck-Verschueren, C.A.J., Gradowska, P.L., Meijer, R., Cloos, J., Biemond, B.J., Graux, C., van Marwijk Kooy, M., Manz, M.G., Pabst, T., Passweg, J.R., Havelange, V., Ossenkoppele, G.J., Sanders, M.A., Schuurhuis, G.J., Lowenberg, B. & Valk, P.J.M. (2018) Molecular Minimal Residual Disease in Acute Myeloid Leukemia. N Engl J Med, 378, 1189-1199.

8 **Al Hinai, A.** & Valk, P.J. (2016) Review: Aberrant EVI1 expression in acute myeloid leukaemia. Br J Haematol, 172, 870-878.

9 Pont, M.J., Honders, M.W., Kremer, A.N., van Kooten, C., Out, C., Hiemstra, P.S., de Boer, H.C., Jager, M.J., Schmelzer, E., Vries, R.G., **Al Hinai, A.S.**, Kroes, W.G., Monajemi, R., Goeman, J.J., Bohringer, S., Marijt, W.A., Falkenburg, J.H. & Griffioen, M. (2016) Microarray Gene Expression Analysis to Evaluate Cell Type Specific Expression of Targets Relevant for Immunotherapy of Hematological Malignancies. PLoS One, 11, e0155165.

10 Sanders, M.A., Kavelaars, F.G., Zeilemaker, **A., Al Hinai**, A.S., Abbas, S., Beverloo, H.B., van Lom, K. & Valk, P.J. (2015) RNA sequencing reveals a unique fusion of the lysine (K)-specific methyltransferase 2A and smooth muscle myosin heavy chain 11 in myelodysplastic syndrome and acute myeloid leukemia. Haematologica, 100, e1-3**.**

WORD OF THANKS

Although this is the last section of the thesis, it is the first section most people read so I have to do my best not to forget to acknowledge and appreciate anyone who supported and helped me all the way to complete and close this chapter of my life. The work described in this thesis would not have been possible without the guidance and support of my supervisor, input of all PIs, and the contribution, support, and motivation of colleagues, family, and friends. With great pleasure and honor, I would like to express my sincere gratitude to all of those who made my Ph.D journey less bumpy, enjoyable, and an unforgettable experience.

First and foremost I would like to express my sincere appreciation and gratitude to my supervisor and mentor **Peter Valk**. Dear Peter, I remember when I met you for the first time in March 2010 when I came for training in molecular diagnostics. You and the colleagues at the molecular diagnostic unit (Hematology Department, EMC) made my stay and training enjoyable and a wonderful experience. I gained and learned a lot from you and your molecular diagnostic team during my training that motivated me to come back for a Ph.D at Erasmus MC and to your group specifically. I learned even more from you during my Ph.D. I am very thankful for giving me the opportunity to get trained in your lab and for accepting me as a Ph.D student in your group. Peter, many thanks for offering advice and encouragement with a perfect blend of insight and humor. You and your family made not only me but also my family, in the 5 years we stayed in the Netherlands, feel like home. Peter, you were very supportive, and cooperative throughout my Ph.D path. I enjoyed the brainstorming in the metro whenever we went together for dinner at your place. Whenever I wanted help or advice you were always available and easy to approach. Many thanks for making this journey possible and achievable. Furthermore, many thanks for your patience and for providing continuous support, motivation, and for providing immense knowledge that helped me to achieve this milestone. I would like to extend my heartfelt thanks and appreciation to **Barbara**, **Menno**, and **Lennard** for the warm welcome and the delicious food, and the enjoyable conversation that we had, and the generous and warm hospitality. Peter, on behalf of my country I am profoundly grateful for your help in establishing molecular diagnostics of hematological malignancies. I am proud of, and grateful for, my time working with you. I wish you all the best and I wish you a prosperous life and scientific career.

Besides my direct supervisor, I would like to express my sincere gratitude to my promoter, **Ruud Delwel**. Dear Ruud, many thanks for being my promoter and I highly appreciate your comments during the Friday floor discussions and the joint Delwel and Valk research groups as well as your feedback on my thesis.

Dear prof.dr. **I.P. Touw**, prof.dr. **W.N.M. Dinjens** and prof.dr. **J. Cloos** members of the doctoral assessment/sub-committee, I would like to thank you for your time in critically reading my thesis and for your feedback which helped to improve my thesis. I would like to extend my thanks and appreciation to the complete doctoral committee for making the time for my defense ceremony. I am looking forward to the day of defense.

I would like to express my gratitude to all group leaders (or PIs) in the department including, **Ivo Touw**, **Tom Cupedo**, **Marc Raaijmakers**, **Eric Braakman**, **Frank Leebeek**, **Pieter Sonneveld**, **Mojca Jongen-Lavrencic**, **Rebekka Schneider**, **Emma de Pater, Anita Rijneveld**, **Jan Cornelissen**, **Ruben Bierings, Moniek de Maat, Bas Wouters**, **Mathijs Sanders**, and ex-group leader in the hematology department **Bob Löwenberg** and **Stefan Erkeland** for their comments, feedback, and helpful advice during the discussions. These have enhanced my scientific thinking, research skills and helped to shape my thesis the way it is today.

I am indebted also to my colleague in Valk research group: **François Kavelaars**, **Annelieke Zeilemaker,** and **Melissa Rijken.** Dear Francois, Annelieke, and Melissa, there is a lot to say to you and the time and work we did together but the most important is I would like to express my gratitude and appreciation for your support, time, and efforts with all the experiments. Most of the work would not have been finished on time without your support and cooperation. I enjoyed working and having conversions with you. Hope you can visit me in Oman one day. Dear Annelieke, I'm honored to know you and to have worked with you, it was sad to know that you left the group but I am happy that you are following your passion. I wish you success and best of luck in your path and life. Dear **Christian**, the new member of the Valk research group, we have not met yet but I am sure you will do great in your Ph.D because you have a great supervisor and a great team. I would like to wish you all the best in your Ph.D and future career.

Since most of the work in this thesis involved next-generation sequencing, I would like to thank our bioinformaticians **Mathijs Sanders**, **Remco Hoogenboezem, and Elodie Stoetman** for your inputs and contributions to my projects. Dear Mathijs, thank you very much for your help with statistics and your inputs on the different projects. I hope your research group will grow bigger and you will excel in your research field and I hope we can work together again in the near future. Remco, thanks for your help and support and I hope you can come to cycle in Oman. Dear Elodie, thank you for your contributions to the different projects and your help in analyzing the custom NGS fusion gene panel. When mentioning NGS I can think of the Ph.D student's big brother and mentor, **Eric Bindels**. Dear Eric, I would like to thank you for helping me with doing RNA-sequencing and for your support with sequencing our libraries for the MRD project and other projects as well and for your inputs in the departmental discussion.

Special thanks to the special ladies in molecular diagnostics **Wendy**, **Pauline**, **Isabel**, **Sonja**, **Chantal**, **Lisa**, **Larissa**, **Milena Isis**, **Farzaneh** and to the ex-staff as well, **Antoinette** and **Marloes**. I would like to express my appreciation to Wendy, Pauline, and Isabel for their help and time during my first visit.

I would like to thank one of the most energetic and enthusiastic people in the department whom I worked a lot with during the last two years of my Ph.D, **Tim Grob**. Dear Tim, I enjoyed working and discussing the different projects with you. Thank you very much for your help with statistical analysis and all other contributions. You have a lot of potentials

and I am sure you will have a bright future. I wish you all the best in finishing your thesis, defending your Ph.D, and in your clinical and scientific career.

I would like to express my thanks and appreciation to **Eric Braakman, Kirsten van Lom** and **Kirsten Gussinklo** for their support and contributions to my PhD projects. I would like to extend my thanks and appreciation to **Leenke**, **Annelies**, and **Tessa** for their help and for facilitating logistics and administrative work during my Ph.D journey. Dear **Egied**, I highly appreciate your help, time, and efforts in making the tables and figures for the manuscripts, and many thanks for making the layout of this thesis.

In addition, I would like to express my appreciation to **Claudia, Anke**, **Hans**, and **Eline** for their help, efforts, and contributions to the different projects. I would like to thank **Onno** for his continuous support and help during my time in the department and when I was away. Dear **Andrea**, **Aniko**, **Anita**, **Arie**, **Bella**, **Burak, Cansu**, **Carla**, **Claire**, **Davine**, **Dennis**, **Dorien**, **Ed**, **Eline P**, **Elwin**, **Emanuele**, **Erik**, **Ferdaws**, **Ferry**, **Hans de Looper**, **Helene**, **Inge**, **Iris**, **Jacqueline**, **Jan van Kapel**, **Jasper**, **Jess M**, **Jess P**, **Joyce**, **Joke**, **Judith**, **Julia**, **Julien**, **Kasia**, **Keans**, **Leonie**, **Lianne**, **Marije**, **Mark**, **Mariette**, **Martijn**, **Maurice**, **Michael**, **Monica**, **Natalie**, **Natasja**, **Neom**, **Nils**, **Patricia D**, **Patricia Olofsen**, **Paola**, **Paulette**, **Peter van Geel**, **Pia**, **Ping**, **Roger**, **Rowan**, **Ruth**, **Saman**, **Shiraz**, **Si**, **Stanley** and **Zoltan.** First, it was a great pleasure to know you. Second, I would like to thank you all for your help, directly or indirectly, and for the little conversions that we had in the department and the lab days. I would like to wish all the Ph.D students the best of luck in finishing Ph.D and in writing and defending the thesis.

I would like to extend my thanks and gratitude to our collaborators in Spain (**Marta Pratcorona** and colleagues), Australia (**Ian Majewski** and colleagues), and the Netherlands. It was my great pleasure to work on joint projects and share knowledge.

My stay in the Netherlands has granted me the pleasure to get new friends that are like brothers to me. Dear **Zuhair**, **Ihsane**, **Mustafa**, **Mohamed**, **Yassir**, **Du**, and **Enny**. I enjoyed your friendship and brotherhood. I enjoyed my time with you and I hope we can meet face to face from time to time. I wish you all the best in your life and career. So, Zuhair and Ihsane where to go in our next vacation?

I would like to acknowledge and express my sincere gratitude and appreciation to the **Ministry of Health** and to the **National Genetic Center** for giving me the opportunity to pursue my Ph.D. I would like to extend my appreciation and special thanks to dr. **Anna Rajab** has been instrumental in pushing us towards perusing higher education, for being inspirational to develop our careers and expand our horizons for the benefit of Oman.

Dear **Salma**, thank you very much for your support, motivation, encouragement, help, and guidance during my Ph.D journey. I would like to express my appreciation to my colleagues in molecular diagnostics of hematological malignancies at the National Genetics Center in Muscat (**Nassra**, **Hilal**, **Ahlam,** and **al Nasser**) for trying to reduce the burden on me to focus on finishing my thesis. I also thank **Mashael** for supporting the team whenever we needed her assistance with the workload. I would like to extend my thanks to my colleagues

Kamla, **Fatma**, **Hind**, **Fahad**, and **Ali** for motivating me to finish this thesis.

Lastly, my family deserves endless gratitude for caring, looking after me like angels, filling my life with joy and happiness, and for providing me with unconditional love and unlimited trust: my dear **parents**, my wife **Muzna**, and my two children (**Retaal** and **Salim**). Dear father and mother, thank you for everything you did for me since I was born. Thank you for your continuous support, motivation, encouragement throughout my life and your supplications (Du'aa). I wouldn't have reached this stage without you, your warmth, and your blessings. Immense gratitude goes to my dear wife Muzna. Muzna thank you for your patience and support and for taking care of our children on your own when you had to go back to Oman to continue your study. It wasn't easy even with the help of your parents but you did great and you manage to finish your degree, bravo! My Ph.D was easier with you around me and I apologize to you and to our children for having to stay late and sometimes away from you to finish this work. May almighty Allah bless you all.

I would like to express my sincere thanks and appreciation to my childhood best friend **Sulaiman Khalifa** and to my cousin **Qasim** for their support, encouragement, and motivation during my Ph.D journey. Dear Sulaiman, thanks for giving me a call now and then to check on me and to update me on what is going in your life and what is going in the karate club and on your journey in becoming one of the World Karate Federation judges, and for becoming Oman's Chief referee, congratulations in your achievements! I'm so proud of you and I have to do a lot to get to your level. Dear Qasim thanks for offering to help me whenever we met and for your advices.

The sad part of leaving the Netherlands for me and my wife is leaving friends who were always there to ease our stay, which made us feel welcomed and happy in our second homeland. They shared our happiness and welcomed our guests into their house. We will cherish these memories for the rest of our lives and look forward to the day that they come to Oman.

The best thing that happened to me and my wife during our stay in the Netherlands is the birth of our two children. The Netherlands for both of us isn't the place where I did my Ph.D, it is the place where our family grew and the joys of our life were born. The Netherlands will always have a very special place in our hearts.

Last but not least I'm looking forward to seeing those who said will visit me in Oman, you know who you are!

I would like to finish by apologizing for missing to mention anyone who has contributed, supported, and helped me during my Ph.D journey. Thank you all and I wish you all the best!


NX1
FLT3-ITD
KIT
MLL-PTD
NPMI
JAK2
WT1
SETBP1
U2AF1
RUN
MYD88
CEBPA
SRSF2
MBD4
NRAS
DNMT3A
IDH1
PTPN11
CEBPA
BRAF
EZH2
IDH2
TEN
ASXL1
TP53
SF3B1
SMC1A
KRAS
MPL
PHF6
TP53
CEBPA
CBL
PTEN
STAG2
CSF3R
SMC1A
SMC3
ZRSR2
CALR
TP53
MYD88
CBL
JAK2
FLT3-ITD
TP53
PHF6