Studies on Genetic Aberrations in Acute Myeloid Leukemia

Saman Abbas

Studies on genetic aberrations in acute myeloid leukemia

ISBN: 978-94-6169-095-1

Cover design: Saman Abbas / Image: US National Institutes of Health
Layout: Egied Simons
Printing: Optima Grafische Communicatie, Rotterdam

The work described in this thesis was performed at the Department of Hematology of Erasmus University Medical Center, Rotterdam, The Netherlands. This work was funded by the Dutch Cancer Society
Printing of this thesis was financially supported by Koningin Wilhelmina Fonds, Erasmus University Rotterdam and Skyline Diagnostics.

Studies on Genetic Aberrations in Acute Myeloid Leukemia

Studies aan genetische aberraties in acute myeloïde leukemie

Proefschrift

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam

op gezag van de rector magnificus prof.dr. H.G. Schmidt
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 29 juni 2011 om 11.30 uur

door
Saman Abbas
geboren te Bagdad-Irak

PROMOTIECOMMISSIE

Promotor	Prof.dr. B. Löwenberg
Co-promotor	Dr. P. J. M. Valk
Overige leden:	Prof.dr. L. H. J. Looijenga
	Prof.dr. G. J. Ossenkoppele
	Prof.dr. R. Pieters

Aan mijn ouders

STUDIES ON GENETIC ABERRATIONS IN ACUTE MYELOID LEUKEMIA

CONTENTS

CHAPTER 1

General introduction

NORMAL HEMATOPOIESIS

Hematopoiesis is the formation of blood cellular components. In mammalian embryonic development, the yolk sac and its vasculature are the source of the first blood cells called hematopoietic stem cells (HSCs) from which all blood cells originate [1]. HSCs are produced by the aorta-gonad-mesonephros region, yolk sac and placenta from where they migrate to the fetal liver, where they expand. Hereafter, HSCs transfer to the bone marrow from where they establish the definitive adult hematopoiesis and reside throughout adulthood [1-2].

HSCs are responsible for foundation of the adult blood differentiation hierarchy and provide continuous hematopoietic cell production. The major characteristic of HSCs is their self-renewal capacity, i.e., they proliferate to give rise to all different types of blood cells, but the pool of stem cells does not become depleted [3]. HSCs are pluripotent. They generate more committed progenitor cells or other stem cells, i.e., common myeloid and lymphoid progenitor cells (CMPs and CLPs respectively) which differentiate and give rise to the progeny belonging to these two lineages of blood cells [4-5]. While the lymphoid progenitor cells generate B- and T cells as well as natural killer cells, the myeloid progenitor cells produce the other leukocytes, i.e. granulocytes and monocytes/macrophages, as well as red blood cells (erythrocytes) and platelets (Figure 1).

The life span of mature blood cells is relatively short and cell production process is continuous, therefore it demands tight regulation by hematopoietic growth factors. The hematopoietic growth factors are a family of cytokines that interact with specific receptors on hematopoietic cells. These molecules like stem cell factor (SCF) or KIT ligand (KIT-L), granulocyte-macrophage CSF (GM-CSF), granulocyte CSF (G-CSF), and macrophage CSF (M-CSF), regulate the functional activation of the specific cells with which they interact. Hematopoietic growth factors are required for the survival, proliferation, and differentiation of hematopoietic progenitors [6-7].

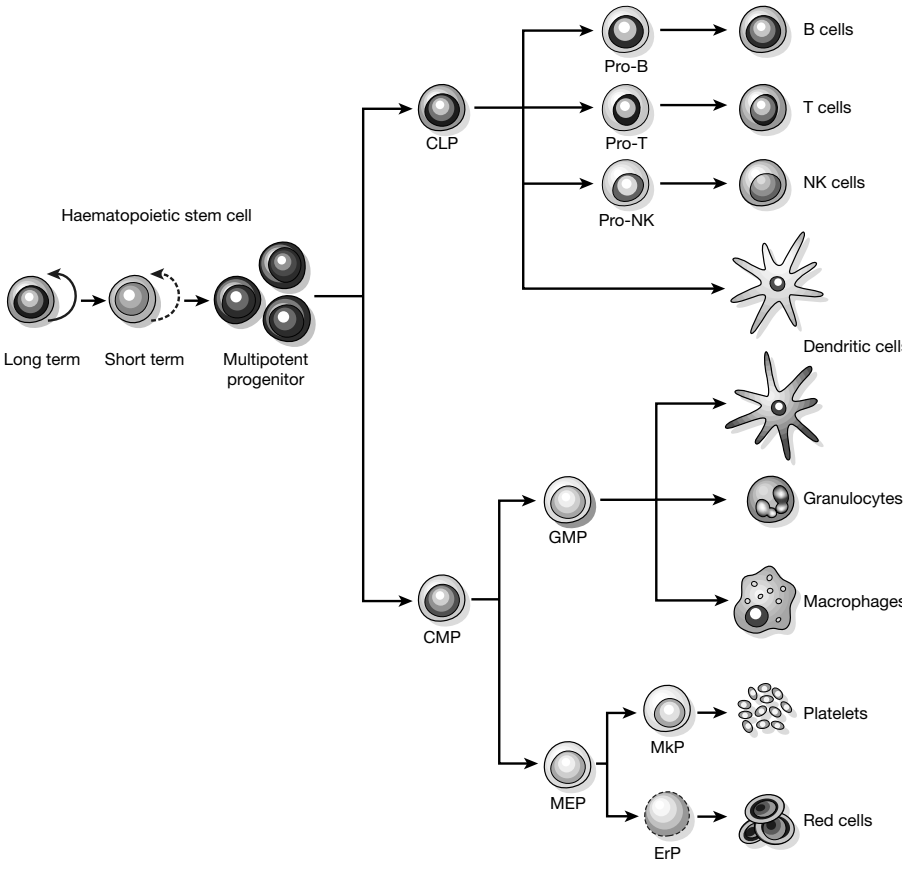


Figure 1. Schematic representation of hematopoiesis. Scheme is adapted from Reya et al [3]. The hematopoietic stem cells (HSCs) can be subdivided into long-term self-renewing HSCs, short term HSCs and multipotent progenitors. They give rise to common lymphoid progenitors (CLPs; the precursor of all lymphoid cells i.e., T-cells, B-cells, natural killer cells (NK cells), and dendritic cells) and common myeloid progenitors (CMPs; the precursor of all myeloid cells). CMPs give rise to granulocyte macrophage precursors (GMPs) and megakaryocyte erythrocyte precursors (MEPs). Subsequently, GMPs give rise to dendritic cells, granulocytes, and macrophages. MEPs give rise to megakaryocyte precursors (MkPs) and erythrocyte precursors (ErP), which further differentiate to platelets and red blood cells, respectively.

ACUTE MYELOID LEUKEMIA (AML)

Acute myeloid leukemia (AML) is a heterogeneous clonal disorder of hematopoietic progenitor cells and is characterized by accumulation of immature myeloid cells, which are impaired in their ability to differentiate towards granulocytes and monocytes that lost the ability to respond to normal regulators of proliferation [8-9]. Due to its heterogeneity, AML represents a group of disorders with variable underlying abnormalities and clinical behavior including response to therapy.

Incidence

AML is among the most common malignant disorders in adults and it accounts approximately 25% of all leukemias in the Western world [10]. Worldwide, the incidence of AML is the highest in the U.S, Australia, and western Europe [11]. The prevelance of AML is 3.8 cases per 100,000 individuals and increases with age, i.e., 17.9 AML cases per 100,000 adults aged over 65 years [8]. The median age for patients at diagnosis for AML is about 70 years, and three men are affected for every two women [8]. In children, AML compromises about 20% of the acute leukemias. The age specific incidence is in contrast with adults and peaks with 11 cases per 1000,000 at the age of 2 years, and decreases to 7 cases per 1000,000 throughout childhood and adolescence [12].

Diagnosis and classification

AML is a heterogeneous disease with variable responses to treatment; therefore the classification of the disease is essential. AML diagnosis is a multidisciplinary process in which morphology, cytology, immunophenotyping, cytogenetics and currently molecular diagnostics are combined. The primary diagnostic approach in AML relies on the morphological identification of leukemic blast cells in bone marrow and blood [9]. For the identification of lineage involvement, cytochemistry and/or immunophenotyping are used. These methods serve as the basis for the traditional French-American-British (FAB) classification system [13]. However, this system has recently been updated by a novel classification model established by the World Health Organization (WHO) which incorporates besides morphology, also cytogenetics and molecular data [14-15]

For the diagnosis of AML, a marrow blast count of over 20% is required, which should be of myeloid origin as confirmed by immunophenotyping and/or cytochemistry [16]. In addition, patients with clonal, recurring cytogenetic abnormalities such as t(15;17)(q22;q12), inv(16)(p13q22) or t(16;16)(p13;q22), and t(8;21)(q22;q22) should be considered to have AML regardless of the blast percentage [16].

It has been suggested that the WHO classification cannot realistically always be used for AML because genetic information is not always available in a timely manner [17]. Furthermore, although the recurring genetic abnormalities are often associated with distinctive morpho-

logical findings, identification of the genetic defect provides a more objective, reproducible means of identifying a specific lesion. For example, the detection of the *CBFβ/MYH11* (inv(16) or t(16;16)) by molecular and/or cytogenetic techniques is reported to correlate with the morphological diagnosis of M4Eo FAB-class in 30% to 100% of cases [18-20].

Therapy and Prognosis

AML therapy consists of two phases. The first phase (induction phase) is applied to achieve complete remission, while the second phase (post-induction) aims to prevent relapse of AML. In the induction therapy, cytarabine has been the backbone of treatment for several decades combined with an anthracyclin, usually daunorubicin [21]. The addition of a third drug has been tested with mixed results for either etoposide or thioguanine [22-23]. There is no convincing evidence of benefit for a third drug [21]. For post-induction therapy, either chemotherapy or bone marrow transplantation (allogeneic or autologous) are applied when patients are at younger age. For patients with older age (>60 years old), treatment options are more limited due to the higher treatment-related toxicity [8-9].

The treatment response and survival of AML patients is dependent on prognostic factors at the time of the presentation of the disease [24]. Beside patient related characteristics that predict the response to treatment like age and disease-related factors, the pre-treatment karyotype of the leukemic cells has an important prognostic impact [8], and provides a mean to divide AML into distinct subgroups with different prognosis (Table 1). However, the limited resolution of cytogenetic analysis leaves smaller structural and numerical abnormalities in the DNA undetected. Thus, the application of cytogenetics only leaves many AML cases unclassified, particularly those with "normal karyotypes", i.e., without detectable cytogenetic abnormalities. AMLs with normal karyotype represent 45% of all cases. A number of prognostic markers, such as mutations and altered expression of particular genes, have recently been identified and now available to risk-stratify AML patients (Table 2).

Table 1. Recurrent cytogenetic abnormalities in adult AML

Cytogenetic abnormality	Frequency in %#	Genes involved	Prognostic significance#
(None) normal cytogenetics	45	Unknown	Intermediate
Complex karyotype (more than 3 abnormatilities)	11	Unknown	Unfavorable
+8	9	Unknown	Intermediate
t(15;17)(q22;q21)	8	*PML-RARA*	Favorable
7/7q-	8	Unknown	Unfavorable
5/5q-	7	Unknown	Unfavorable
t(8;21)(q22;q22)	6	*AML1-ETO*	Favorable
inv(16)(p13q22)/ t(16;16) (p13;q22)	5	*CBFβ-MYH11*	Favorable
Y	4	Unknown	Intermediate
abn(12p)	3	Unknown	Intermediate
plus 21	3	Unknown	Intermediate
abn(17p)	2	Unknown	Intermediate
del(9q)	2	Unknown	Intermediate
inv(3)(q21q26)/ t(3;3)(q21;q26)	2	*EVI1*	Unfavorable
del(11q)	1	Unknown	Intermediate
t/inv(11q23)	1	*MLL*	Favorable
t(9;22)(q34;q11)	1	*BCR-ABL*	Favorable/Unfavorable
t(6;9)(p23;q34)	1	*DEK-CAN*	Unfavorable

Based on Mrozek et al [25]

Genetics in the biology and prognosis of AML

AML is a heterogeneous disease which is illustrated by the variety of cytogenetic abnormalities and molecular mutations present in the leukemic blasts (Table 2). These aberrations are related to different pathogenic mechanisms affecting cell proliferation, differentiation, self-renewal, apoptosis and DNA repair [26-28].

In the past decades, there has been an increasing interest in elucidating the role of molecular genetic abnormalities in AML initiation and development [29-32]. These molecular insights add to the understanding of the pathology of the disease and they might lead towards the development of new targeted drugs [33]. Illustrative of this approach is AML with t(15;17) that

generates the *PML-RARA* fusion gene. Optimized treatment for patients diagnosed with AML and t(15;17), involving all-trans retinoic acid (ATRA), increased the complete remission rate up to 90% to 95% [34-35]. Furthermore, these genetic abnormalities can be used as new markers to refine risk-stratification of AML. Identification of new markers is pivotal particularly for AMLs lacking prognostically informative karyotypes.

It is proposed that leukemogenesis is a multistep process that requires more than one genetic aberration [26-28]. This has been shown in animal models in which a single mutation is not sufficient for leukemogenesis [26,36-38]. In a simplified model, the molecular abnormalities that contribute to AML development can be divided into two groups, one that provides a proliferation advantage to the leukemic cells and another that impairs normal differentiation [39]. In recent years, this simplified model is supported by the fact that in AML cases, these two types of molecular abnormalities occur together and aberrations belonging to the same group are often mutually exclusive [8,29,31-32,40].

Some mutations that provide proliferation advantage are involved in signal transduction pathways causing their aberrant activation. Examples of these mutations are the mutations leading to the activation of the receptor tyrosine kinases *FLT3*, *c-KIT* or RAS. All these mutations affect the RAS signaling pathway (Table 2) [29,31,41].

The chromosomal translocations that lead to impaired differentiation in hematopoietic cells result in the formation of fusion proteins. Such translocations are inv(16)/t(16;16), t(8;21) and t(15;17) that generate the fusion proteins *CBFβ-MYH11, AML1-ETO*, and *PML-RARA* respectively (Table 1).

Table 2: Recurrent molecular abnormalities in AML

Genes	Frequency in %	Association with cytogenetics	Prognostic significance	Reference
Mutation				
Nucleophosmin (*NPM1*)	25-35	Normal	Favorable in absence of *FLT3*-ITD	42-43
DNA methyl transferase 3A (*DNMT3A*)	22	Normal	Unfavorable	44
Fms-related tyrosine kinase 3 (*FLT3*), internal tandem duplication (ITD)	28-33	Normal/t(15;17)/t(6;9)	Unfavorable	45-47
The tet oncogene family member 2 (*TET2*)	12-20	t(3;4)(q26;q24)	Unfavorable	48-50
NAPD-dependent isocitrate dehydrogenase 1/2 (*IDH1/2)*	10-16	Normal	Unfavorable	51-55
Neuroblastoma RAS viral (v-ras) oncogene homolog (*NRAS*)	10-15	Inv(16)	-	56-58
Additional sex comb-like 1 (*ASXL1*)	17-23	Trisomy 8	Unfavorable	57,59
Wilms tumor 1 (*WT1*)	10	Normal	Unfavorable	60-62
Tumor protein p53 (*TP53*)	<10	Complex karyotype	Unfavorable	63-64
Runt-related transcription factor 1 (*RUNX1; AML1*)	6-11	Normal/trisomy21/others	-	65-66
Myeloid/lymphoid or mixed lineage leukemia (*MLL*) partial tandem duplication (PTD)	5-11 of normal karyotypes	Trisomy 11/Normal	Unfavorable	67-68
Fms-related tyrosine kinase 3 (FLT3), tyrosine kinase domain (TKD)	5-10	Normal/inv(16)	?	47,69
CCAAT/enhancer binding protein alpha (*CEBPA*)	5-10	Normal	Favorable	70-71
v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (*KIT*)	28	Inv(16)/t(8;21) (CBF leukemia)/Trisomy 4	Unfavorable in CBF AMLs	40,72
v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolg (*KRAS*)	5	Abn3q	-	56
Protein tyrosine phosphtase non-receptor type 11 (*PTPN11*; *SHP2*)	3	-	-	73-74
Janus kinase 2 (*JAK2*)	2	t(8;21)/Inv(16)	-	75
Casitas B-cell lymphoma (*CBL*)	0,7-1,1	CBF leukemia/11q deletion	-	76-77
Overexpression				
Brain and acute leukemia gene, cytoplasmic (*BAALC*)	~50 of normal karyotypes	Trisomy 8/Normal	Unfavorable	78-79
Meningioma (disrupted in balanced translocation) 1 (*MN1*)	50	Normal	Unfavorable in normal karyotypes	80
v-ets Erythroblastosis virus E26 oncogene homolg (*ERG*)	~25 of normal karyotypes	Complex/ normal, chromosome 21	Unfavorable	81-82
Ecotropic viral integration site 1 (*EVI1*)	~6	Abn3q/ 11q23	Unfvaorable	83-84

In recent years several novel molecular markers have been associated with prognostic classification of AML. For instance, the fms-like tyrosine kinase-3 gene (*FLT3*), a hematopoietic growth factor receptor, is mutated as a result of an internal tandem duplication (ITD) (25-30% of the cases of AML) or by a base pair substitution in the tyrosine kinase domain ((TKD D835) (5-7%) of the cases of AML), both resulting in ligand-independent constitutive activation of the FLT3 receptor. *FLT3* ITDs in particular have been associated with a poor treatment outcome [45,47].

Insertion mutations in the nucleophosmin (*NPM1*) gene, which result in aberrant cytoplasmic localization of the protein, are seen in approximately 35% of all AML cases. These mutations currently represent the most common somatic gene mutation in AML and result in aberrant cytoplasmic localization of NPM1 [42-43] (Table 2). NPM1 is involved in cell proliferation and regulation of tumor suppressors TP53/p53 and ARF [85]. *NPM1* mutations are significantly associated with age, i.e., more frequent in older patients, high white blood cell counts, normal karyotypes, and *FLT3* ITD mutations. *NPM1* mutations predicted for better overall survival, event-free survival, and response to induction therapy in particular in the absence of coexisting *FLT3* ITD mutations [42-43]. In multivariable analysis, *NPM1* mutations express independent favorable prognostic value [42-43,86-88].

Proliferative mutations in the hematopoietic receptor *c-KIT* are present in a variety of diseases, such as systemic mastocytosis, gastrointestinal stromal tumors and AML [89]. In AML, mutations in *c-KIT* are detected in the extracellular domain, i.e., insertion/deletion mutations in exon 8 and intracellular, i.e., tyrosine kinase domain (TKD D816), juxtamembrane domain, and ITD mutations. KIT mutations are exclusively found in Core Binding Factor (CBF) leukemias, i.e., AML with inv(16) and t(8;21) (Table 2) [72]. Patients with AML t(8;21) with c-KIT TKD(D816) mutation have a significantly higher incidence of relapse and a lower overall survival than AML with t(8;21) and wild-type c-*KIT* [90-92]. Poorer responses to treatment were also demonstrated in pediatric AML with t(8;21) and c-KIT TKD(D816) [93].

Mutations in the *CBL* gene is an example of gene mutations that are not directly involved in signal transduction. Mutations in *CBL* will result in increased proliferation or exerting an oncogenic activity which consequently contribute to leukemogenesis [77]. CBL exerts E3-ligase activity [94]. It also has a role in the termination of signaling of receptor tyrosine kinases like *FLT3* and *c-KIT* which are commonly mutated in AML and then give proliferative advantage to leukemic cells. The mutated CBL proteins alter the amplitude and duration of tyrosine kinases signaling events (Table 2) [77,95-96].

Transcription factors regulate gene expression by activating target genes by DNA interaction; thereby they play an important role in activating particular cell types and differentiation stages in the hematopoiesis. Aberrant functioning of transcription factors may result in impaired differentiation as observed in leukemia [97-99]. An example of an aberrant transcription factor implicated in AML is the CCAAT/enhancer binding protein alpha (*CEBPA*) (Table 2). CEBPA is a transcription factor essential for granulocytic differentiation [71]. AML patients

with biallelic mutations in *CEBPA* (approximately 7% of the cases) have been associated with more favorable outcome in the group of patients with standard risk AML [70,100-101]. The favorable value of *CEBPA* gene mutations is maintained after adjustment for cytogenetic risk, *FLT3*-ITD and *CEBPA* expression levels in multivariate analysis [100]. In contrast, low *CEBPA* mRNA expression in AML with intermediate-risk karyotypes may be associated with poor prognosis [100].

Some genes are found to be mutated in AML at comparatively low frequency. Among those are classical P53 mutations and mutations in the GTPases N-RAS and K-RAS (Table 2). RAS mutations do not show any association with clinical outcome in AML [56]. P53 mutations are indicative for poor response to therapy in AML; however, patients with P53 mutations often also carry unfavorable cytogenetic abnormalities, e.g. those involving chromosomes 5 and 7 [102-103].

Recently, a new class of gene mutations has been discovered i.e., mutations in *IDH1* and *IDH2* genes (Table 2) that seem to contribute to leukemogenesis [54,104-105]. The mutations in the *IDH* genes were first identified in gliomas and patients with such tumors had a better outcome than those with wild-type IDH genes [106]. *IDH1* and *IDH2* are members of the β-decarboxylating dehydrogenase in the tricarboxylic acid (TCA) cycle and upon mutation in certain arginine residues, they acquire the ability to convert α-ketoglutarate to 2-hydroxyglutarate, which is a potential oncometabolite and is suggested to be involved in the mutant IDH transforming ability [52-53,106-107]. Studies investigating the prevelance of IDH mutations in AML showed frequencies of 5% to 33% [51-52,54-55,105,108-111] (Table 2). Some of these studies did not show any significant effect of these mutations on survival of AML patients [52,105,108-109,111], but others did [51,54-55,110]. Interestingly, a single nucleotide polymorphism (SNP) located in the same exon of *IDH1* mutational hotspot (IDH1 R132) has a negative impact on prognosis in AML patients [52].

The *EVI1* (ecotropic virus integration site 1) was first identified as the integration site of the ecotropic retrovirus leading to myeloid leukemia in murine model systems [112-113]. In humans, *EVI1* is located on chromosome 3q26, and rearrangements on chromosome 3q26 often activate EVI1 expression in myeloid malignancies, including AML, CML, and MDS [114]. *EVI1* is a transcription factor and contains DNA-binding activity [115]. In addition to its DNA-binding activity, *EVI1* has the potential to recruit diverse proteins, such as SMAD3 and CtBP, thus generating regulatory complexes for transcriptional regulation [114-115]. *EVI1* appears essential for proliferation and maintenance of hematopoietic stem cells (HSC) [116-117]. The role of *EVI1* in HSC regulation would suggest that *EVI1* might participate in the generation of leukemia stem cells (LSCs) [114]. Interestingly, high expression of the transcription factor EVI1 is also apparent in approximately 8% of AML without 3q26 abnormalities [83]. In these cases high *EVI1* expression, like in AML with 3q26 abnormalities, independently predicts for poor prognosis (Table 2) [83-84,118].

High expression of the tumor suppressor Wilms tumor 1 (WT1) has also been proposed as a marker for inferior outcome for patients with AML. However, the definitive prognostic value of this marker remains to be settled [119-123]. *WT1* has been suggested to act as a tumor suppressor in childhood malignancies of the kidney and as a transcription factor with regulating activity on a number of growth and differentiation factors [119,124]. It has been shown that the expression of *WT1* significantly correlates with *BCL2* expression [125] , which has also been proposed as unfavorable prognostic marker in AML, when aberrantly expressed [126]. In addition to aberrant expression of WT1, mutations are also found in 10% to 13% of cytogenetically normal AML [61-62,127]. In some studies, *WT1* mutations have been associated with inferior outcome [61-62], whereas in a larger study, *WT1* mutations did not affect outcome [127]. Furthermore, a recent study showed that *WT1* (SNP) located in the proximity of the *WT1* mutational hotspot was found to be associated with reduced treatment outcome [128].

BAALC, as a single gene, has been proposed as a clinical outcome predictor in AML [78]. High *BAALC* expression appears an independent risk factor for overall survival, event free survival and disease free survival. Patients with high *BAALC* expression showed a significantly worse overall survival (Table 2) [78] which was subsequently confirmed in a larger cohort of AML patients with normal cytogenetics [129]. The *BAALC* protein has been implicated to play a role in leukemia and normal expression of *BAALC* is restricted to hematopoietic progenitors [130].

Epigenetics of AML

Epigenetic control of gene expression has been suggested to play a pivotal role in determining the biological behavior of cells [131]. One such epigenetic mechanism is DNA cytosine methylation, which can alter gene expression by creating new binding sites for methylation-dependent repressor proteins [132-133] or by disrupting the ability of transcription factors to bind to their targets sequences [134-135]. In normal development, the proper distribution of DNA methylation plays a critical role in tissue differentiation and homeostasis [136-137]. Disruption of normal DNA methylation distribution is a hallmark of cancer and can play critical roles in initiation, progression, and maintenance of the malignant phenotype [131]. For example, the most prevalent epigenetic alteration in AML is the transcriptional silencing of *p15INK4b* tumor suppressor gene by DNA hypermethylation [138-141], reported in 40% to 80% of AML, and MDS [138].

Histone modifications play an important role in the gene expression patterns of tumor suppressor genes including *p15INK4b* [142]. Polycomb genes like enhancer of zeste-2 (*EZH2*) and leukemia viral BMI-1 oncogene (*BMI1*) are involved in chromatin and histone modifications. Involvement of polycomb gene repression has been found at tumor suppressor genes in many tumor cell types and embryonic stem cells [143-145]. This suggests that polycomb modifications of tumor suppressor genes are common regulatory mechanisms which contribute to cancer. Furthermore, overexpression of *BMI1* or *EZH2* has been observed in cancer, including AML, and may lead to abnormal repression of genes regulated by polycomb complexes [146]. Muta-

tions in polycomb genes can be a mechanism for the disruption of polycomb genes-directed histone modification exerted on target genes particularly tumor suppressor genes. Recently, mutations have detected in the additional sex comb-like 1 (*ASXL1*) gene. *ASXL1* is an enhancer of trithorax and polycomb gene, and is required to maintain activation and silencing of homeotic loci [147]. A recent study has described *ASXL1* as the most frequent mutated gene in advanced MDS [148], which was confirmed by another report [59]. Thus, disruption of pathways that control the distribution of histone modifications may represent a mechanism by which the affected genes are aberrantly repressed in AML.

Genome-wide approaches to study AML

Gene expression profiling (GEP)

Molecular genetics and epigenetics have provided valuable insights into questions related to classification, progression, and pathogenesis of AML in the past years. The start of the human genome project marks an increasing emphasis in the field of genome-wide analysis [149-151]. Several new gene based approaches enable researchers to investigate changes in normal and malignant cells on genome-wide basis, particularly those involving DNA microarray which has been implicated in the studies on leukemogenesis in an early stage [152-154].

Gene expression profiling (GEP) is the measurement of the abundance of mRNA transcripts of thousands of genes using microarrays. It has been widely and successfully applied in cancer in general, including AML. In pioneer study, using GEP, it was shown that acute lymphoblastic leukemias (ALLs) could be distinguished from AML [155]. Furthermore, ALLs possessing a rearranged *MLL* gene were shown to have a highly uniform and distinct gene expression profiles that clearly distinguishes them from conventional ALL or AML and warrants designation as distinct disease termed mixed-lineage leukemia [156]. This distinction stimulated the development of therapies based on molecular targets [156].

In recent years, GEP has been successfully introduced in clinical AML research. By GEP, cytogenetically as well molecularly well-defined AML subclasses were predicted [157-162] and novel subtypes of AML were discovered [157-159,163]. Attempts to predict response to therapy as well as outcome based on GEP have also been successful [158,164]. More challenging than predicting known classes, is the discovery of novel subtypes of AML using GEP followed by unsupervised cluster analysis, i.e., grouping of samples based on similarities in expression profiling. A study used unsupervised hierarchical cluster analysis to define two novel molecular subclasses of AML, with predominantly normal karyotypes and significant differences in survival times [158]. The outcome predictor predicted overall survival and appeared a strong independent prognostic factor [158]. In another study of a representative cohort of 285 AML patients, sixteen clusters of AML patients were revealed by GEP [157]. In this analysis, novel clusters were characterized by high frequencies of certain molecular lesions or mutations, but also included patients without these molecular lesions. However, due to the heterogeneous

nature of AML, the clustering resulted in fairly small groups of AML, preventing statistically significant survival analysis [157]. Furthermore, smaller subtypes within AML have been distinguished using e.g. *CEBPA* mutations and *EVI1* overexpression, suggesting that a specific combination of acquired molecular lesions result in a distinct molecular signature [157,165-167]. These studies are illustrative of the applicability and potential of gene expression profiling in the discovery of novel AML subtypes.

Genome-wide genotyping and comparative genomic hybridization using microarrays (array-CGH)

Microarray approaches to directly interrogate genetic alterations in a genome-wide fashion are of great interest. Several different microarray platforms are available for the analysis of genetic alterations in cancer that vary in resolution, technical performance, and the ability to detect DNA deletions and gains as well as copy neutral loss of heterozygosity (CN-LOH) [168-171]. CN-LOH also referred to as acquired uniparental disomy (UPD), may reflect reduplication of a chromosomal region harbouring a mutated or silenced tumor suppressor gene [172].

The earliest platforms were bacterial artificial chromosome (BAC) arrays, in which large (up to 200 kb) DNA probes were spotted onto arrays which were subsequently hybridized with test and reference DNA. BAC arrays have been extensively used in studies on cancer and AML [171,173-186]. These have been highly informative but due to the large probe size in BACs have limited ability to identify focal copy number abnormalities (CNA), which are a hallmark in leukemia [170,187-188]. Oligonucleotide arrays are now more commonly used, and comprise up to millions of short nucleotide probes. These oligonucleotide arrays are either used in a comparative genomic hybridization (array-CGH) design, in which test and reference DNA are hybridized to a single array, or single channel array in which either a test or reference array are hybridized alone. The latter encompass SNP arrays that genotype up to hundreds of thousands of SNPs across the genome. This design facilitates genome-wide association studies and the detection of both germline and somatic DNA copy number alterations. However, SNP-arrays are also being used to examine associations of inherited genetic variants, including copy number polymorphisms, SNP genotypes, with tumor susceptibility, treatment response and outcome [189-190].

BAC arrays have been used in studies of leukemia to identify genetic lesions such as deletions of *ETV6* in ALL [175] and gains of chromosome 19 in acute megakaryocytic leukemia [191]. However, the most detailed insights into genetic alterations in acute leukemias have been obtained from higher resolution SNP array studies of diagnostic leukemia samples. An example is the first SNP array study in ALL that utilized a 10000 marker SNP array and demonstrated the ability of this approach to detect focal alterations, most commonly LOH involving *CDKN2A/B* [192].

SNP arrays were initially designed as genotyping tool. Therefore, the distribution of markers across the genome is not even, and many genes exhibit suboptimal coverage, particularly

with lower resolution arrays [193]. Array resolution is thus critical determinant of the ability of microarray analysis to identify all genetic alterations in leukemia. Moreover, detailed bioinformatic analyses is required in order to appropriately normalize raw array data, correct the aneuploidy, and generate sensitive and accurate calls of DNA copy number alterations.

Studies applying SNP array demonstrated the ability of this approach to detect alterations in genes involved in leukemic transformation. When integrated with GEP, as is discussed in chapter 2 and 3 of this thesis, SNP array can be a more powerful tool to elucidate the effect of genetic aberrations on the expression of the affected genes and provide critical insights into the biology of the disease.

Comparative profiling of cytosine methylation

Cytosine methylation is a major component of epigenetic regulation of gene expression and is linked to cancer development [194-196]. Many techniques have been used to test cytosine methylation at multiple loci, but they were not suitable for comparing methylation levels at multiple loci within a genome [194]. In contrast, analogous intragenomic profiling has been successfully developed for studying chromatin organization using chromatin immunoprecipitation with genomic arrays (ChIP-on-chip) [197]. The establishment of a platform for intragenomic profiling that integrates different genomic studies of epigenetic regulation is required.

First assays focusing specifically on CpG island sequences such as Differential Methylation Hybridization (DMH) [198] and Restriction Landmark Genomic Scanning (RLGS) [199-200] revealed valuable information on overall CpG island methylation frequencies and identified tumor suppressors affected by aberrant methylation events. On the other hand, information on DNA methylation changes in regions outside of CpG islands are missed, thus assays that specifically identify hyper- or hypo-methylated sequences in cancer genomes have been designed [201]. For example, Methylated CpG island Amplification (MCA) was successfully used to identify methylated sequences in particular in colon cancer [202].

Meanwhile, a variety of enzymatic approaches have become available to distinguish between methylated and nonmethylated states at CpG sites, which are applied in Microarray based Integrated Analysis of Methylation by Isoschizomers (MIAMI) [203]. Also, the use of HpaII revealed that most of the genome remains high molecular weight following digestion by HpaII [204]. Furthermore, it was subsequently recognized that 55%-70% of HpaII sites in animal genomes are methylated at the central cytosine, which is part of a CpG dinucleotide [205-206]. The minority of the genomic DNA that cuts to a size of hundreds of base pairs was defined as HpaII Tiny Fragments (HTFs). HTF enrichment by ligation-mediated PCR (HELP) [207] is a technique that uses HpaII to cut genomic DNA. It differs from many other assays based on the same approach by using the methylated-insensitive isoschizomer MspI as a control. Consequently, HpaII/MspI strategy allows intragenomic pattern compression and quantification of the degree of methylation between two cell samples.

A recent study applying genome-wide methylation profiling using HELP assay predicted that integration of different genome-wide epigenetic regulatory marks along with gene expression levels provide greater power in capturing biological differences between leukemia subtypes [208]. Measuring gene expression, cytosine methylation and histone H3 lysine 9 (H3K9) acetylation [208] using high-density oligonucleotide microarrays in primary AML and ALL specimens showed that DNA methylation and H3K9 acetylation distinguished these leukemias, but also that an integrative analysis combining the information from each platform revealed hundreds of additional differentially expressed genes that were missed by gene expression arrays alone [208].

Another study using HELP assay showed clustering of AML patients by methylation data into 16 groups. Five of these groups defined new AML subtypes that shared no other known feature. In addition, DNA methylation profiles segregated patients with CEBPA aberrations from other subtypes of leukemia, defined four epigenetically distinct forms of AML with NPM1 mutations, and showed that established AML1-ETO, CBFβ-MYH11, and PML-RARA leukemia entities are associated with specific methylation profiles [131].

HELP assay also revealed that AMLs carrying one recently identified class of mutated i.e., *IDH1* and *IDH2* genes which involved in citrate metabolism [52-53,106,108] display global DNA hypermethylation and a specific hypermethylation signature [209]. Furthermore, the expression of the oncometabolite 2HG [210] by IDH mutated AMLs induced global DNA hypermethylation [209]. This study showed that oncogenic alterations in core cellular metabolic pathways can lead to leukemic transformation by dysregulation of the epigenetic machinery in hematopoietic cells.

The studies applying HELP assay show that this technique may allow insights into the metabolic process of cytosine methylation which is the major mediator of epigenetic regulation. This had not been possible with single-locus studies or intragenomic comparison.

Next generation sequencing (NGS)

Next generation sequencing (NGS) is referred to as a new technology arising after the automated Sanger method which had been first described in 1977 [211]. This technology can generate hundreds of millions of sequences of short DNA fragments in a single run [212-213]. NGS offers many advantages over microarray-based assays. First, NGS provides higher base pair resolution, with the exception of tilling arrays which still require a large number of probes to reach a high resolution. Second, NGS has fewer artefacts, such as noise in the form of cross-hybridization, caused by the hybridization step on microarrays. Third, the genome coverage is not limited in NGS by the repertoire of probe sequences fixed on the array. This is particularly important for the analysis of repetitive regions of the genome, which are typically masked out on microarrays. Finally, NGS has a larger dynamic range and provides high-coverage thereby increasing the confidence of the resultant data [212-213].

The main limitations of NGS are its current cost and availability. The overall cost of NGS including machine depreciation and reagent expense is much higher than the expenses of microarray analysis. With the improvements of sequencing chemistry and institutional support for the procurement of sequencing platforms, NGS will become a main modality of genome-wide profiling experiments [212].

Significant challenges of NGS technologies are data management and analysis. NGS generates an immense amount of data that can reach terabytes per machine run in raw image files. This makes data storage a challenge even in centers with considerable expertise in the management of genomic data [214]. Genomic alignment and assembly require bioinformatics skills. New and improved algorithms are needed to identify genome enrichment [212]. Taken together, NGS offers higher resolution and well defined data than array-based experiments for profiling the whole-genome [213-214].

The application of NGS has allowed cancer genomics to move from focused approaches such as single-gene sequencing and array analysis to comprehensive genome-wide approaches. NGS can be applied to cancer samples in various ways. These vary by the type of input material (for example, DNA, RNA or chromatin), the proportion of the genome targeted (the whole genome, transcriptome or a subset of genes) and the genomic variation studied (structural change, point mutation, gene expression or chromosomal conformation) [215].

The first NGS study was reported in 2008, a description of the nucleotide sequence of DNA from an AML compared with DNA from normal skin of the same patient [216]. The results obtained from this study revealed eight genes with acquired mutations that had not previously been implicated in AML, such as the protein phosphatise *PTPRT* and the peptide/drug transporter *SLC15A1*, and also two genes that had previously been described i.e., *NPM1* and *FLT3* [216]. This NGS-based study demonstrated the need for unbiased whole-genome approaches to discover all mutations associated with cancer pathogenesis. Since then, several complete sequences of cancer genomes together with matched normal tissue genomes have been reported [44,104,217-220].

A following study using NGS in AML have revealed recurrent mutations in the isocitrate dehydrogenase gene (*IDH1*) [104]. Subsequent work has confirmed and extended this finding, showing that mutations in *IDH1* and related gene *IDH2* are highly recurrent in patients with an intermediate-risk cytogenetic profile (20 to 30% frequency) and are associated with a poor prognosis in selected subgroups of patients [54,110,221]. Very recently, a new study using whole-genome sequence approach in *de novo* AML with an intermediate-risk cytogenetic profile has revealed different mutations in the DNA methyltransferase gene *DNMT3A* at high frequency (22.1% of the patients) [44]. *DNMT3A* mutations may also be independently associated with a poor treatment outcome [44].

It is likely that NGS methods will continue to lead to the comprehensive discovery of a majority of the alterations in the cancer genome. It is also probable that this will be followed by the application of comprehensive sequencing approaches to cancer diagnostics.

Computational analysis will become a central part of these discovery and diagnostic efforts. Information databases that connect genomic findings with clinical parameters to assess the relevance of the genome alterations are also required to realize the potential of cancer genomics [215]. These developments will ultimately potentiate accurate genome-based diagnosis to be applied for an increasingly greater number of patients with cancer.

SCOPE OF THIS THESIS

AML is a heterogeneous disease characterized by a variety of cytogenetic abnormalities and molecular mutations that are assumed to have a role in the pathobiology of the disease and that may have diagnostic and therapeutic value in the clinical practice.

The work in this thesis is divided into two parts. The first part (Chapter 2 and 3) deals with the application of DNA microarray technology (SNP arrays) for the identification of novel abnormalities in AML.

Chapter 2 describes the development of a software tool SNPExpress which allow the concurrent interpretation of genotype, polymorphic copy number variations (CNVs), LOH regions, and copy number variations in a combinatorial and efficient way with gene expression data using various array platforms.

Chapter 3 is concerned with the application of combinational approach of genome-wide genotyping and gene expression profiling (GEP). The molecular data analysis performed by SNPExpress revealed *BCL11B* as a novel recurrent abnormality in AML.

The second part of this thesis represents studies dealing with recently discovered genetic mutations in AML. The impact of these gene mutations on prognosis in large cohorts of patients with AML was investigated.

Chapter 4 presents investigations regarding the distribution of mutations in the *CBL* gene in primary AML patients and the association of these mutations with specific genetically defined AML subtypes.

In chapter 5, we investigated with a high-throughput mutation screening approach the involvement of RNA surveillance mechanism (Nonsense mediated RNA decay) (NMD) in AMLs carrying mutations in the *WT1* gene. These mutations introduce premature termination codon (PTC).

Finally, in chapter 6, by applying high-throughput mutation screening involving denaturing high performance liquid chromatography (dHPLC) the prevalence and the prognostic value of isocitrate dehydrogenase 1 (*IDH1*) and isocitrate dehydrogenase 2 (*IDH2*) mutations in primary AML cohort were examined.

Chapter 7 provides a summary of the findings reported in the thesis and a general discussion of the results and their relationship to future studies.

REFRENCES

1 Dzierzak, E. & Speck, N. A. Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nat Immunol* **9**, 129-136, (2008).

2 oore, M. A. & Metcalf, D. Ontogeny of the haemopoietic system: yolk sac origin of in vivo and in vitro colony forming cells in the developing mouse embryo. *Br J Haematol* **18**, 279-296 (1970).

3 Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105-111, (2001).

4 Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193-197, (2000).

5 Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661-672, (1997).

6 Clark, S. C. & Kamen, R. The human hematopoietic colony-stimulating factors. *Science* **236**, 1229-1237 (1987).

7 Nicola, N. A. Hemopoietic cell growth factors and their receptors. *Annu Rev Biochem* **58**, 45-77, (1989).

8 Estey, E. & Dohner, H. Acute myeloid leukaemia. *Lancet* **368**, 1894-1907, (2006).

9 Lowenberg, B., Downing, J. R. & Burnett, A. Acute myeloid leukemia. *N Engl J Med* **341**, 1051-1062, (1999).

10 Greenlee, R. T., Hill-Harmon, M. B., Murray, T. & Thun, M. Cancer statistics, 2001. *CA Cancer J Clin* **51**, 15-36 (2001).

11 Jemal, A., Thomas, A., Murray, T. & Thun, M. Cancer statistics, 2002. *CA Cancer J Clin* **52**, 23-47 (2002).

12 Deschler, B. & Lubbert, M. Acute myeloid leukemia: epidemiology and etiology. *Cancer* **107**, 2099-2107, (2006).

13 Bennett, J. M. *et al.* Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol* **33**, 451-458 (1976).

14 Vardiman, J. W. *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937-951, (2009).

15 Swerdlow, S. H., Harris, N. L, Jaffe, E. S, Pileri, S. A, Stein, H, et al., editors. WHO classification of tumours of haematopoietic and lymphoid tissues. *IARC: Lyon.* (2008).

16 Vardiman, J. W., Harris, N. L. & Brunning, R. D. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood* **100**, 2292-2302, (2002).

17 Arber, D. A. Realistic pathologic classification of acute myeloid leukemias. *Am J Clin Pathol* **115**, 552-560, (2001).

18 Shurtleff, S. A. *et al.* Heterogeneity in CBF beta/MYH11 fusion messages encoded by the inv(16)(p13q22) and the t(16;16)(p13;q22) in acute myelogenous leukemia. *Blood* **85**, 3695-3703 (1995).

19 Mrozek, K. *et al.* Comparison of cytogenetic and molecular genetic detection of t(8;21) and inv(16) in a prospective series of adults with de novo acute myeloid leukemia: a Cancer and Leukemia Group B Study. *J Clin Oncol* **19**, 2482-2492 (2001).

20 Costello, R. *et al.* Detection of CBFbeta/MYH11 fusion transcripts in acute myeloid leukemia: heterogeneity of cytological and molecular characteristics. *Leukemia* **11**, 644-650 (1997).

21 Burnett, A., Wetzler, M. & Lowenberg, B. Therapeutic Advances in Acute Myeloid Leukemia. *J Clin Oncol*, (2011).

22 Bishop, J. F. *et al.* Etoposide in acute nonlymphocytic leukemia. Australian Leukemia Study Group. *Blood* **75**, 27-32 (1990).

23 Hann, I. M. *et al.* Randomized comparison of DAT versus ADE as induction chemotherapy in children and younger adults with acute myeloid leukemia. Results of the Medical Research Council's 10th AML trial (MRC AML10). Adult and Childhood Leukaemia Working Parties of the Medical Research Council. *Blood* **89**, 2311-2318 (1997).

24 Lowenberg, B., Griffin, J. D. & Tallman, M. S. Acute myeloid leukemia and acute promyelocytic leukemia. *Hematology Am Soc Hematol Educ Program*, 82-101 (2003).

25 Mrozek, K., Heerema, N. A. & Bloomfield, C. D. Cytogenetics in acute leukemia. *Blood Rev* **18**, 115-136, (2004).

26 Kelly, L. M. *et al.* FLT3 internal tandem duplication mutations associated with human acute myeloid leukemias induce myeloproliferative disease in a murine bone marrow transplant model. *Blood* **99**, 310-318 (2002).

27 Wilbanks, A. M. *et al.* TEL/PDGFbetaR fusion protein activates STAT1 and STAT5: a common mechanism for transformation by tyrosine kinase fusion proteins. *Exp Hematol* **28**, 584-593, (2000).

28 Schwaller, J. *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* **17**, 5321-5333, (1998).

29 Renneville, A. *et al.* Cooperating gene mutations in acute myeloid leukemia: a review of the literature. *Leukemia* **22**, 915-931, (2008).

30 Mrozek, K., Marcucci, G., Paschka, P., Whitman, S. P. & Bloomfield, C. D. 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* **109**, 431-448, (2007).

31 Dohner, K. & Dohner, H. Molecular characterization of acute myeloid leukemia. *Haematologica* **93**, 976-982, (2008).

32 Frohling, S., Scholl, C., Gilliland, D. G. & Levine, R. L. Genetics of myeloid malignancies: pathogenetic and clinical implications. *J Clin Oncol* **23**, 6285-6295, (2005).

33 Tallman, M. S., Gilliland, D. G. & Rowe, J. M. Drug therapy for acute myeloid leukemia. *Blood* **106**, 1154-1163, (2005).

34 Asou, N. *et al.* A randomized study with or without intensified maintenance chemotherapy in patients with acute promyelocytic leukemia who have become negative for PML-RARalpha transcript after consolidation therapy: the Japan Adult Leukemia Study Group (JALSG) APL97 study. *Blood* **110**, 59-66, (2007).

35 Tallman, M. S. *et al.* All-trans retinoic acid in acute promyelocytic leukemia: long-term outcome and prognostic factor analysis from the North American Intergroup protocol. *Blood* **100**, 4298-4302, (2002).

36 Castilla, L. H. *et al.* The fusion gene Cbfb-MYH11 blocks myeloid differentiation and predisposes mice to acute myelomonocytic leukaemia. *Nat Genet* **23**, 144-146, (1999).

37 Yuan, Y. *et al.* AML1-ETO expression is directly involved in the development of acute myeloid leukemia in the presence of additional mutations. *Proc Natl Acad Sci U S A* **98**, 10398-10403, (2001).

38 Chan, I. T. *et al.* Conditional expression of oncogenic K-ras from its endogenous promoter induces a myeloproliferative disease. *J Clin Invest* **113**, 528-538, (2004).

39 Dash, A. & Gilliland, D. G. Molecular genetics of acute myeloid leukaemia. *Best Pract Res Clin Haematol* **14**, 49-64, (2001).

40 Care, R. S. *et al.* Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. *Br J Haematol* **121**, 775-777, (2003).

41 Van Etten, R. A. Aberrant cytokine signaling in leukemia. *Oncogene* **26**, 6738-6749, (2007).

42 Dohner, K. *et al.* Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. *Blood* **106**, 3740-3746, (2005).

43 Verhaak, R. G. *et al.* Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. *Blood* **106**, 3747-3754, (2005).

44 Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424-2433, (2010).

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

46 Nakao, M. *et al.* Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia* **10**, 1911-1918 (1996).

47 Yanada, M., Matsuo, K., Suzuki, T., Kiyoi, H. & Naoe, T. Prognostic significance of FLT3 internal tandem duplication and tyrosine kinase domain mutations for acute myeloid leukemia: a meta-analysis. *Leukemia* **19**, 1345-1349, (2005).

48 Abdel-Wahab, O. *et al.* Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114**, 144-147, (2009).

49 Langemeijer, S. M. *et al.* Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet* **41**, 838-842, (2009).

50 Tefferi, A. *et al.* Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML. *Leukemia* **23**, 1343-1345, (2009).

51 Boissel, 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. *J Clin Oncol* **28**, 3717-3723, (2010).

52 Wagner, K. *et al.* Impact of IDH1 R132 mutations and an IDH1 single nucleotide polymorphism in cytogenetically normal acute myeloid leukemia: SNP rs11554137 is an adverse prognostic factor. *J Clin Oncol* **28**, 2356-2364, (2010).

53 Dang, L., Jin, S. & Su, S. M. IDH mutations in glioma and acute myeloid leukemia. *Trends Mol Med* **16**, 387-397, (2010).

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

55 Green, C. L. *et al.* The prognostic significance of IDH1 mutations in younger adult patients with acute myeloid leukemia is dependent on FLT3/ITD status. *Blood* **116**, 2779-2782, (2010).

56 Bowen, D. T. *et al.* RAS mutation in acute myeloid leukemia is associated with distinct cytogenetic subgroups but does not influence outcome in patients younger than 60 years. *Blood* **106**, 2113-2119, (2005).

57 Carbuccia, N. *et al.* Mutual exclusion of ASXL1 and NPM1 mutations in a series of acute myeloid leukemias. *Leukemia* **24**, 469-473, (2010).

58 Valk, P. J. *et al.* Second hit mutations in the RTK/RAS signaling pathway in acute myeloid leukemia with inv(16). *Haematologica* **89**, 106 (2004).

59 Rocquain, J. *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* **10**, 401, (2010).

60 King-Underwood, L. & Pritchard-Jones, K. Wilms' tumor (WT1) gene mutations occur mainly in acute myeloid leukemia and may confer drug resistance. *Blood* **91**, 2961-2968 (1998).

61 Paschka, P. *et al.* Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. *J Clin Oncol* **26**, 4595-4602, (2008).

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

63 Fenaux, P. *et al.* P53 gene mutations in acute myeloid leukemia with 17p monosomy. *Blood* **78**, 1652-1657 (1991).

64 Slingerland, J. M., Minden, M. D. & Benchimol, S. Mutation of the p53 gene in human acute myelogenous leukemia. *Blood* **77**, 1500-1507 (1991).

65 Osato, M. *et al.* Biallelic and heterozygous point mutations in the runt domain of the AML1/PEBP2alphaB gene associated with myeloblastic leukemias. *Blood* **93**, 1817-1824 (1999).

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

67 Caligiuri, M. A. *et al.* The partial tandem duplication of ALL1 in acute myeloid leukemia with normal cytogenetics or trisomy 11 is restricted to one chromosome. *Proc Natl Acad Sci U S A* **94**, 3899-3902 (1997).

68 Dohner, K. *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* **20**, 3254-3261 (2002).

69 Mead, A. J. *et al.* 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* **110**, 1262-1270, (2007).

70 Frohling, S. *et al.* CEBPA mutations in younger adults with acute myeloid leukemia and normal cytogenetics: prognostic relevance and analysis of cooperating mutations. *J Clin Oncol* **22**, 624-633, (2004).

71 Pabst, T. *et al.* Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. *Nat Genet* **27**, 263-270, (2001).

72 Beghini, A. *et al.* C-kit mutations in core binding factor leukemias. *Blood* **95**, 726-727 (2000).

73 Johan, M. F. *et al.* Mutations in PTPN11 are uncommon in adult myelodysplastic syndromes and acute myeloid leukaemia. *Br J Haematol* **124**, 843-844, (2004).

74 Nomdedeu, J. *et al.* Low frequency of exon 3 PTPN11 mutations in adult de novo acute myeloid leukemia. Analysis of a consecutive series of 173 patients. *Haematologica* **90**, 412-413 (2005).

75 Levine, R. L. *et al.* The JAK2V617F activating mutation occurs in chronic myelomonocytic leukemia and acute myeloid leukemia, but not in acute lymphoblastic leukemia or chronic lymphocytic leukemia. *Blood* **106**, 3377-3379, (2005).

76 Reindl, C. *et al.* CBL exon 8/9 mutants activate the FLT3 pathway and cluster in core binding factor/11q deletion acute myeloid leukemia/myelodysplastic syndrome subtypes. *Clin Cancer Res* **15**, 2238-2247, (2009).

77 Sargin, B. *et al.* Flt3-dependent transformation by inactivating c-Cbl mutations in AML. *Blood* **110**, 1004-1012, (2007).

78 Baldus, C. D. *et al.* BAALC expression predicts clinical outcome of de novo acute myeloid leukemia patients with normal cytogenetics: a Cancer and Leukemia Group B Study. *Blood* **102**, 1613-1618, (2003).

79 Tanner, S. M. *et al.* BAALC, the human member of a novel mammalian neuroectoderm gene lineage, is implicated in hematopoiesis and acute leukemia. *Proc Natl Acad Sci U S A* **98**, 13901-13906, (2001).

80 Heuser, M. *et al.* High meningioma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. *Blood* **108**, 3898-3905, (2006).

81 Baldus, C. D. *et al.* Acute myeloid leukemia with complex karyotypes and abnormal chromosome 21: Amplification discloses overexpression of APP, ETS2, and ERG genes. *Proc Natl Acad Sci U S A* **101**, 3915-3920, (2004).

82 Mrozek, K., Heinonen, K., Theil, K. S. & Bloomfield, C. D. Spectral karyotyping in patients with acute myeloid leukemia and a complex karyotype shows hidden aberrations, including recurrent overrepresentation of 21q, 11q, and 22q. *Genes Chromosomes Cancer* **34**, 137-153, (2002).

83 Barjesteh van Waalwijk van Doorn-Khosrovani, S. *et al.* High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients. *Blood* **101**, 837-845, d (2003).

84 Lugthart, S. *et al.* High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression and chromosome 3q26 abnormalities underestimated. *Blood* **111**, 4329-4337, (2008).

85 Tago, K., Chiocca, S. & Sherr, C. J. Sumoylation induced by the Arf tumor suppressor: a p53-independent function. *Proc Natl Acad Sci U S A* **102**, 7689-7694, (2005).

86 Falini, B. *et al.* Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med* **352**, 254-266, (2005).

87 Schnittger, S. *et al.* Nucleophosmin gene mutations are predictors of favorable prognosis in acute myelogenous leukemia with a normal karyotype. *Blood* **106**, 3733-3739, (2005).

88 Thiede, C. *et al.* Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML). *Blood* **107**, 4011-4020, (2006).

89 Lennartsson, J., Jelacic, T., Linnekin, D. & Shivakrupa, R. Normal and oncogenic forms of the receptor tyrosine kinase kit. *Stem Cells* **23**, 16-43, (2005).

90 Schnittger, S. *et al.* KIT-D816 mutations in AML1-ETO-positive AML are associated with impaired event-free and overall survival. *Blood* **107**, 1791-1799, (2006).

91 Nanri, T. *et al.* Mutations in the receptor tyrosine kinase pathway are associated with clinical outcome in patients with acute myeloblastic leukemia harboring t(8;21)(q22;q22). *Leukemia* **19**, 1361-1366, (2005).

92 Cairoli, R. *et al.* Prognostic impact of c-KIT mutations in core binding factor leukemias: an Italian retrospective study. *Blood* **107**, 3463-3468, (2006).

93 Shimada, A. *et al.* KIT mutations, and not FLT3 internal tandem duplication, are strongly associated with a poor prognosis in pediatric acute myeloid leukemia with t(8;21): a study of the Japanese Childhood AML Cooperative Study Group. *Blood* **107**, 1806-1809, (2006).

94 Zeng, S., Xu, Z., Lipkowitz, S. & Longley, J. B. Regulation of stem cell factor receptor signaling by Cbl family proteins (Cbl-b/c-Cbl). *Blood* **105**, 226-232, (2005).

95 Andoniou, C. E. *et al.* The Cbl proto-oncogene product negatively regulates the Src-family tyrosine kinase Fyn by enhancing its degradation. *Mol Cell Biol* **20**, 851-867 (2000).

96 Andoniou, C. E., Thien, C. B. & Langdon, W. Y. Tumour induction by activated abl involves tyrosine phosphorylation of the product of the cbl oncogene. *EMBO J* **13**, 4515-4523 (1994).

97 Rosenbauer, F. & Tenen, D. G. Transcription factors in myeloid development: balancing differentiation with transformation. *Nat Rev Immunol* **7**, 105-117, (2007).

98 Tenen, D. G. Disruption of differentiation in human cancer: AML shows the way. *Nat Rev Cancer* **3**, 89-101, (2003).

99 Tenen, D. G., Hromas, R., Licht, J. D. & Zhang, D. E. Transcription factors, normal myeloid development, and leukemia. *Blood* **90**, 489-519 (1997).

100 Barjesteh van Waalwijk van Doorn-Khosrovani, S. *et al.* Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML. *Hematol J* **4**, 31-40, (2003).

101 Preudhomme, C. *et al.* Favorable prognostic significance of CEBPA mutations in patients with de novo acute myeloid leukemia: a study from the Acute Leukemia French Association (ALFA). *Blood* **100**, 2717-2723, (2002).

102 Nakano, Y. *et al.* Prognostic value of p53 gene mutations and the product expression in de novo acute myeloid leukemia. *Eur J Haematol* **65**, 23-31 (2000).

103 Stirewalt, D. L. *et al.* FLT3, RAS, and TP53 mutations in elderly patients with acute myeloid leukemia. *Blood* **97**, 3589-3595 (2001).

104 Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* **361**, 1058-1066, (2009).

105 Thol, F. *et al.* Prognostic impact of IDH2 mutations in cytogenetically normal acute myeloid leukemia. *Blood* **116**, 614-616, (2010).

106 Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **360**, 765-773, (2009).

107 Chou, W. C. *et al.* Distinct clinical and biological features of de novo acute myeloid leukemia with additional sex comb-like 1 (ASXL1) mutations. *Blood*, (2010).

108 Patel, K. P. *et al.* Acute myeloid leukemia with IDH1 or IDH2 mutation: frequency and clinicopathologic features. *Am J Clin Pathol* **135**, 35-45, (2011).

109 Ho, P. A. *et al.* Molecular alterations of the IDH1 gene in AML: a Children's Oncology Group and Southwest Oncology Group study. *Leukemia* **24**, 909-913, (2010).

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

111 Chou, W. C. *et al.* Distinct clinical and biologic characteristics in adult acute myeloid leukemia bearing the isocitrate dehydrogenase 1 mutation. *Blood* **115**, 2749-2754, (2010).

112 Morishita, K. *et al.* Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. *Cell* **54**, 831-840, (1988).

113 Mucenski, M. L. *et al.* Identification of a common ecotropic viral integration site, Evi-1, in the DNA of AKXD murine myeloid tumors. *Mol Cell Biol* **8**, 301-308, (1988).

114 Goyama, S. & Kurokawa, M. Evi-1 as a critical regulator of leukemic cells. *Int J Hematol* **91**, 753-757, (2010).

115 Nucifora, G., Laricchia-Robbio, L. & Senyuk, V. EVI1 and hematopoietic disorders: history and perspectives. *Gene* **368**, 1-11, (2006).

116 Goyama, S. *et al.* Evi-1 is a critical regulator for hematopoietic stem cells and transformed leukemic cells. *Cell Stem Cell* **3**, 207-220, (2008).

117 Yuasa, H. *et al.* Oncogenic transcription factor Evi1 regulates hematopoietic stem cell proliferation through GATA-2 expression. *EMBO J* **24**, 1976-1987, (2005).

118 Groschel, S. *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* **28**, 2101-2107, (2010).

119 Bergmann, L. *et al.* High levels of Wilms' tumor gene (wt1) mRNA in acute myeloid leukemias are associated with a worse long-term outcome. *Blood* **90**, 1217-1225 (1997).

120 Schmid, D. *et al.* Prognostic significance of WT1 gene expression at diagnosis in adult de novo acute myeloid leukemia. *Leukemia* **11**, 639-643 (1997).

121 Garg, M., Moore, H., Tobal, K. & Liu Yin, J. A. Prognostic significance of quantitative analysis of WT1 gene transcripts by competitive reverse transcription polymerase chain reaction in acute leukaemia. *Br J Haematol* **123**, 49-59, (2003).

122 Barragan, E. *et al.* Prognostic implications of Wilms' tumor gene (WT1) expression in patients with de novo acute myeloid leukemia. *Haematologica* **89**, 926-933 (2004).

123 Yanada, M. *et al.* Multiplex real-time RT-PCR for prospective evaluation of WT1 and fusion gene transcripts in newly diagnosed de novo acute myeloid leukemia. *Leuk Lymphoma* **45**, 1803-1808, (2004).

124 Bergmann, L., Maurer, U. & Weidmann, E. Wilms tumor gene expression in acute myeloid leukemias. *Leuk Lymphoma* **25**, 435-443 (1997).

125 Karakas, T. *et al.* The coexpression of the apoptosis-related genes bcl-2 and wt1 in predicting survival in adult acute myeloid leukemia. *Leukemia* **16**, 846-854, (2002).

126 Karakas, T. *et al.* High expression of bcl-2 mRNA as a determinant of poor prognosis in acute myeloid leukemia. *Ann Oncol* **9**, 159-165 (1998).

127 Gaidzik, V. I. *et al.* Prognostic impact of WT1 mutations in cytogenetically normal acute myeloid leukemia: a study of the German-Austrian AML Study Group. *Blood* **113**, 4505-4511, (2009).

128 Damm, F. *et al.* Single nucleotide polymorphism in the mutational hotspot of WT1 predicts a favorable outcome in patients with cytogenetically normal acute myeloid leukemia. *J Clin Oncol* **28**, 578-585, (2010).

129 Baldus, C. D. *et al.* BAALC expression and FLT3 internal tandem duplication mutations in acute myeloid leukemia patients with normal cytogenetics: prognostic implications. *J Clin Oncol* **24**, 790-797, (2006).

130 Verhaak, R. G. & Valk, P. J. Genes predictive of outcome and novel molecular classification schemes in adult acute myeloid leukemia. *Cancer Treat Res* **145**, 67-83, (2010).

131 Figueroa, M. E. *et al.* DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, 13-27, (2010).

132 Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**, 187-191, (1998).

133 Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386-389, (1998).

134 Kanduri, C. *et al.* Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr Biol* **10**, 853-856, (2000).

135 Watt, F. & Molloy, P. L. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev* **2**, 1136-1143 (1988).

136 Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926, (1992).

137 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257, (1999).

138 Drexler, H. G. Review of alterations of the cyclin-dependent kinase inhibitor INK4 family genes p15, p16, p18 and p19 in human leukemia-lymphoma cells. *Leukemia* **12**, 845-859 (1998).

139 Herman, J. G. *et al.* Distinct patterns of inactivation of p15INK4B and p16INK4A characterize the major types of hematological malignancies. *Cancer Res* **57**, 837-841 (1997).

140 Sherr, C. J. & Roberts, J. M. CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev* **13**, 1501-1512 (1999).

141 Paul, T. A., Bies, J., Small, D. & Wolff, L. Signatures of polycomb repression and reduced H3K4 trimethylation are associated with p15INK4b DNA methylation in AML. *Blood* **115**, 3098-3108, (2010).

142 Gil, J. & Peters, G. Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. *Nat Rev Mol Cell Biol* **7**, 667-677, (2006).

143 Bracken, A. P. *et al.* The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes Dev* **21**, 525-530, (2007).

144 Ezhkova, E. *et al.* Ezh2 orchestrates gene expression for the stepwise differentiation of tissue-specific stem cells. *Cell* **136**, 1122-1135, d (2009).

145 Kia, S. K., Gorski, M. M., Giannakopoulos, S. & Verrijzer, C. P. SWI/SNF mediates polycomb eviction and epigenetic reprogramming of the INK4b-ARF-INK4a locus. *Mol Cell Biol* **28**, 3457-3464, (2008).

146 Simon, J. A. & Lange, C. A. Roles of the EZH2 histone methyltransferase in cancer epigenetics. *Mutat Res* **647**, 21-29, (2008).

147 Fisher, C. L., Randazzo, F., Humphries, R. K. & Brock, H. W. Characterization of Asxl1, a murine homolog of Additional sex combs, and analysis of the Asx-like gene family. *Gene* **369**, 109-118, (2006).

148 Boultwood, J. *et al.* Frequent mutation of the polycomb-associated gene ASXL1 in the myelodysplastic syndromes and in acute myeloid leukemia. *Leukemia* **24**, 1062-1065, (2010).

149 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921, (2001).

150 Venter, J. C., Adams, M. D., Martin-Gallardo, A., McCombie, W. R. & Fields, C. Genome sequence analysis: scientific objectives and practical strategies. *Trends Biotechnol* **10**, 8-11 (1992).

151 Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351, (2001).

152 Brown, P. O. & Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nat Genet* **21**, 33-37, (1999).

153 Ebert, B. L. & Golub, T. R. Genomic approaches to hematologic malignancies. *Blood* **104**, 923-932, (2004).

154 Hoheisel, J. D. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* **7**, 200-210, (2006).

155 Golub, T. R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-537, (1999).

156 Armstrong, S. A. *et al.* MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* **30**, 41-47, (2002).

157 Valk, P. J. *et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* **350**, 1617-1628, (2004).

158 Bullinger, L. *et al.* Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* **350**, 1605-1616, (2004).

159 Debernardi, S. *et al.* Genome-wide analysis of acute myeloid leukemia with normal karyotype reveals a unique pattern of homeobox gene expression distinct from those with translocation-mediated fusion events. *Genes Chromosomes Cancer* **37**, 149-158, (2003).

160 Kohlmann, A. *et al.* Molecular characterization of acute leukemias by use of microarray technology. *Genes Chromosomes Cancer* **37**, 396-405, (2003).

161 Schoch, C. *et al.* Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. *Proc Natl Acad Sci U S A* **99**, 10008-10013, (2002).

162 Virtaneva, K. *et al.* Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc Natl Acad Sci U S A* **98**, 1124-1129, (2001).

163 Vey, N. *et al.* Identification of new classes among acute myelogenous leukaemias with normal karyotype using gene expression profiling. *Oncogene* **23**, 9381-9391, (2004).

164 Heuser, M. *et al.* Gene-expression profiles and their association with drug resistance in adult acute myeloid leukemia. *Haematologica* **90**, 1484-1492 (2005).

165 Wouters, B. J. *et al.* Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood* **110**, 3706-3714, (2007).

166 Wouters, B. J., Lowenberg, B. & Delwel, R. A decade of genome-wide gene expression profiling in acute myeloid leukemia: flashback and prospects. *Blood* **113**, 291-298, (2009).

167 Wouters, B. J. *et al.* Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood* **113**, 3088-3091, (2009).

168 Davies, J. J., Wilson, I. M. & Lam, W. L. Array CGH technologies and their applications to cancer genomes. *Chromosome Res* **13**, 237-248, (2005).

169 Maciejewski, J. P. & Mufti, G. J. Whole genome scanning as a cytogenetic tool in hematologic malignancies. *Blood* **112**, 965-974, d (2008).

170 Mullighan, C. G. *et al.* Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* **446**, 758-764, (2007).

171 Strefford, J. C. & Parker, H. Array-based comparative genomic hybridization as a tool for analyzing the leukemia genome. *Methods Mol Biol* **538**, 151-177, (2009).

172 Tuna, M., Knuutila, S. & Mills, G. B. Uniparental disomy in cancer. *Trends Mol Med* **15**, 120-128, (2009).

173 Davidsson, J. *et al.* Tiling resolution array comparative genomic hybridization, expression and methylation analyses of dup(1q) in Burkitt lymphomas and pediatric high hyperdiploid acute lymphoblastic leukemias reveal clustered near-centromeric breakpoints and overexpression of genes in 1q22-32.3. *Hum Mol Genet* **16**, 2215-2225, (2007).

174 Huhta, T. *et al.* Comparative genomic hybridization and conventional cytogenetic analyses in childhood acute myeloid leukemia. *Leuk Lymphoma* **35**, 311-315 (1999).

175 Jalali, G. R. *et al.* Disruption of ETV6 in intron 2 results in upregulatory and insertional events in childhood acute lymphoblastic leukaemia. *Leukemia* **22**, 114-123, (2008).

176 Kuchinskaya, E. *et al.* Tiling-resolution array-CGH reveals the pattern of DNA copy number alterations in acute lymphoblastic leukemia with 21q amplification: the result of telomere dysfunction and breakage/fusion/breakage cycles? *Leukemia* **21**, 1327-1330, (2007).

177 Larramendy, M. L. *et al.* Comparative genomic hybridization reveals differences in DNA copy number changes between sporadic gastric carcinomas and gastric carcinomas from patients with hereditary nonpolyposis colorectal cancer. *Cancer Genet Cytogenet* **106**, 62-65, (1998).

178 Larramendy, M. L. *et al.* DNA copy number changes in childhood acute lymphoblastic leukemia. *Haematologica* **83**, 890-895 (1998).

179 Larramendy, M. L. *et al.* Comparative genomic hybridization in childhood acute lymphoblastic leukemia. *Leukemia* **12**, 1638-1644 (1998).

180 Lundin, C. *et al.* Tiling resolution array CGH of dic(7;9)(p11 approximately 13;p11 approximately 13) in B-cell precursor acute lymphoblastic leukemia reveals clustered breakpoints at 7p11.2 approximately 12.1 and 9p13.1. *Cytogenet Genome Res* **118**, 13-18, (2007).

181 Pinkel, D. & Albertson, D. G. Comparative genomic hybridization. *Annu Rev Genomics Hum Genet* **6**, 331-354, (2005).

182 Pinkel, D. & Albertson, D. G. Array comparative genomic hybridization and its applications in cancer. *Nat Genet* **37 Suppl**, S11-17, (2005).

183 Rabin, K. R. *et al.* Clinical utility of array comparative genomic hybridization for detection of chromosomal abnormalities in pediatric acute lymphoblastic leukemia. *Pediatr Blood Cancer* **51**, 171-177, (2008).

184 Schoumans, J. *et al.* Characterisation of dic(9;20)(p11-13;q11) in childhood B-cell precursor acute lymphoblastic leukaemia by tiling resolution array-based comparative genomic hybridisation reveals clustered breakpoints at 9p13.2 and 20q11.2. *Br J Haematol* **135**, 492-499, (2006).

185 Steinemann, D. *et al.* Copy number alterations in childhood acute lymphoblastic leukemia and their association with minimal residual disease. *Genes Chromosomes Cancer* **47**, 471-480, (2008).

186 van Vlierberghe, P. *et al.* A new recurrent 9q34 duplication in pediatric T-cell acute lymphoblastic leukemia. *Leukemia* **20**, 1245-1253, (2006).

187 Hehir-Kwa, J. Y. *et al.* Genome-wide copy number profiling on high-density bacterial artificial chromosomes, single-nucleotide polymorphisms, and oligonucleotide microarrays: a platform comparison based on statistical power analysis. *DNA Res* **14**, 1-11, (2007).

188 Lo, K. C. *et al.* Comprehensive analysis of loss of heterozygosity events in glioblastoma using the 100K SNP mapping arrays and comparison with copy number abnormalities defined by BAC array comparative genomic hybridization. *Genes Chromosomes Cancer* **47**, 221-237, (2008).

189 French, D. *et al.* Acquired variation outweighs inherited variation in whole genome analysis of methotrexate polyglutamate accumulation in leukemia. *Blood* **113**, 4512-4520, (2009).

190 Yang, J. J. *et al.* Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. *JAMA* **301**, 393-403, (2009).

191 Alvarez, S., MacGrogan, D., Calasanz, M. J., Nimer, S. D. & Jhanwar, S. C. Frequent gain of chromosome 19 in megakaryoblastic leukemias detected by comparative genomic hybridization. *Genes Chromosomes Cancer* **32**, 285-293, (2001).

192 Irving, J. A. *et al.* Loss of heterozygosity in childhood acute lymphoblastic leukemia detected by genome-wide microarray single nucleotide polymorphism analysis. *Cancer Res* **65**, 3053-3058, (2005).

193 Xavier, R. J. & Rioux, J. D. Genome-wide association studies: a new window into immune-mediated diseases. *Nat Rev Immunol* **8**, 631-643, (2008).

194 Khulan, B. *et al.* Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res* **16**, 1046-1055, (2006).

195 Baylin, S. B. & Herman, J. G. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet* **16**, 168-174, (2000).

196 Kass, S. U., Pruss, D. & Wolffe, A. P. How does DNA methylation repress transcription? *Trends Genet* **13**, 444-449, (1997).

197 Kim, T. H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876-880, (2005).

198 Huang, T. H., Perry, M. R. & Laux, D. E. Methylation profiling of CpG islands in human breast cancer cells. *Hum Mol Genet* **8**, 459-470, (1999).

199 Hayashizaki, Y. *et al.* Restriction landmark genomic scanning method and its various applications. *Electrophoresis* **14**, 251-258 (1993).

200 Smiraglia, D. J. & Plass, C. The study of aberrant methylation in cancer via restriction landmark genomic scanning. *Oncogene* **21**, 5414-5426, (2002).

201 Claus, R., Plass, C., Armstrong, S. A. & Bullinger, L. DNA methylation profiling in acute myeloid leukemia: from recent technological advances to biological and clinical insights. *Future Oncol* **6**, 1415-1431, (2010).

202 Toyota, M. *et al.* Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res* **59**, 2307-2312 (1999).

203 Hatada, I. *et al.* Genome-wide profiling of promoter methylation in human. *Oncogene* **25**, 3059-3064, (2006).

204 Singer, J., Roberts-Ems, J. & Riggs, A. D. Methylation of mouse liver DNA studied by means of the restriction enzymes msp I and hpa II. *Science* **203**, 1019-1021 (1979).

205 Bestor, T. H., Hellewell, S. B. & Ingram, V. M. Differentiation of two mouse cell lines is associated with hypomethylation of their genomes. *Mol Cell Biol* **4**, 1800-1806 (1984).

206 Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res* **8**, 1499-1504 (1980).

207 Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet* **39**, 61-69, (2007).

208 Figueroa, M. E. *et al.* An integrative genomic and epigenomic approach for the study of transcriptional regulation. *PLoS One* **3**, e1882, (2008).

209 Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell*, (2010).

210 Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744, (2009).

211 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**, 5463-5467 (1977).

212 Huang, Y. W., Huang, T. H. & Wang, L. S. Profiling DNA methylomes from microarray to genome-scale sequencing. *Technol Cancer Res Treat* **9**, 139-147, (2010).

213 Metzker, M. L. Sequencing technologies - the next generation. *Nat Rev Genet* **11**, 31-46, (2010).

214 Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* **10**, 669-680, (2009).

215 Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* **11**, 685-696, (2010).

216 Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66-72, (2008).

217 Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465**, 473-477, (2010).

218 Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191-196, (2010).

219 Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184-190, (2010).

220 Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809-813, (2009).

221 Ward, P. S. *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* **17**, 225-234, (2010).

.

CHAPTER 2

SNPExpress: integrated visualization of genome-wide genotypes, copy numbers and gene expression levels

Mathijs A. Sanders[1], Roel GW. Verhaak[1,2], Wendy MC. Geertsma-Kleinekoort[1], Saman Abbas[1], Sebastiaan Horsman[3], Peter J van der Spek[3], Bob Löwenberg[1], Peter JM Valk[1]

[1] Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands
[2] Department of Medical Oncology, Dana Farber Cancer Institute, Harvard Medical School, Boston, USA; The Broad institute of M.I.T. and Harvard, Cambridge, USA
[3] Department of Bioinformatics, Erasmus University Medical Center, Rotterdam, The Netherlands

BMC Genomics 2008; 9:41

ABSTRACT

Background: Accurate analyses of comprehensive genome-wide SNP genotyping and gene expression data sets is challenging for many researchers. In fact, obtaining an integrated view of both large scale SNP genotyping and gene expression is currently complicated since only a limited number of appropriate software tools are available.

Results: We present SNPExpress, a software tool to accurately analyze Affymetrix and Illumina SNP genotype calls, copy numbers, polymorphic copy number variations (CNVs) and Affymetrix gene expression in a combinatorial and efficient way. In addition, SNPExpress allows concurrent interpretation of these items with Hidden-Markov Model (HMM) inferred Loss-of-Heterozygosity (LOH)- and copy number regions.

Conclusion: The combined analyses with the easily accessible software tool SNPExpress will not only facilitate the recognition of recurrent genetic lesions, but also the identification of critical pathogenic genes.

BACKGROUND

High-density genome-wide views of biological samples, using high-throughput DNA mapping and mRNA gene expression microarrays facilitate the identification of recurrent molecular lesions. Both types of microarrays, which are being produced by different manufacturers, e.g., Nimblegen, Agilent, Sequenom, Applied Biosystems, Illumina and Affymetrix, typically contain large numbers of small oligonucleotides that interrogate the genome. Currently available DNA arrays contain over 500.000 probe sets, while the gene expression arrays target over 20.000 genes. Efficient analysis of these large datasets remains a challenge for many researchers.

The Affymetrix and Illumina DNA mapping platforms have been designed to specifically target sequences containing single nucleotide polymorphisms (SNPs). SNPs are currently estimated to be present at a frequency of 1 out of 300 nucleotides [1]. By including different probe sets to detect the possible SNP variants, genome-wide genotyping is feasible. In fact, these types of arrays have been developed for genome-wide association studies; however, these platforms can easily be applied to determine copy numbers of these chromosomal markers, similar to array comparative genomic hybridization (CGH). Because of the high number of SNPs, sample DNA can be examined with an inter-marker distance of 6 to 12 kb, and (micro) deletions and/or amplifications are detectable. By comparing disease samples to normal germ line DNA, a detailed overview of acquired gains and losses of the genome is obtained. In fact, although our knowledge is still developing, it has recently become apparent that that copy number variation (CNV) accounts for a substantial amount of genetic variation in the human genome [2]. The high-resolution scanning technologies enable the analyses of CNV and associated phenotypes [2].

The power of DNA mapping has been shown extensively in cancer research. Chromosomal gains and losses as well as regions of loss-of-heterozygosity (LOH) have been shown in, for instance, leukemia 3-4, lung cancer 5-7 and colon cancer 8. Recognition of recurrent lesions will ultimately result in the identification of pathogenic genes. For instance, SNP array analysis of a set of cancer cell lines has lead to the identification of the microphthalmia-associated transcription factor MITF as a melanoma oncogene 9.

On the Illumina platform genotypes are determined using hybridization of genomic DNA to BeadChips followed by an enzymatic discrimination step. On the Affymetrix platform, genotype calls and copy numbers are determined by a probe set consisting of mismatch and perfect match probes. In analogy with the expression probe set, the genotype and copy number of an individual SNP is dependent on the balance of genotype calls in the associated probe set. Several methods for genotype calling [10-13] and assessment of copy number [14-15] have been developed. Advanced analysis methods of DNA mapping array data have focused on the identification of regions of LOH, or gains and losses [16-19].

A particular SNP genotype or a numerical change in chromosome copy number can have profound effects on gene expression. A possible relation to tumor development was shown in breast cancer, where a 17q23 amplification was related to increased expression of genes at that locus [20] and in acute myeloid leukemia (AML), where amplification of 8p24 was associated with increased expression of genes such as *MYC* [21]. In fact, SNPs as well as CNVs have recently been shown to have consistent effects, often in *cis*, on gene expression [22-23]. The integrated analysis of gene expression and SNP array data is a prerequisite to recognize these effects. To our knowledge, only one software package is able to visualize chromosome copy number and gene expression levels [19]. Here, we present a package, SNPExpress, which allows concurrent interpretation of genotype, HMM inferred LOH regions, copy number, CNVs, HMM inferred copy number and gene expression data. Due to the simple format of the input data, our package is not restricted to specific methods to determine genotype, copy numbers or expression level. Little knowledge of software is necessary to use SNPExpress, making the tool accessible for a wide audience.

IMPLEMENTATION

SNPExpress, written in JAVA (version 1.5), uses tab-delimited files as input and is currently available for use with Affymetrix DNA mapping arrays (10K 2.0, 100K set and 500K set), Illumina HumanHap550 Genotyping BeadChip and Affymetrix GeneChips (HG-U95Av2, HG-U133A and B, HG-U133 plus 2.0). A file containing a matrix with each column representing the genotypes of one array and rows starting with Illumina or Affymetrix SNP IDs is mandatory. The genotype should be formatted as homozygous 'AA' or 'BB', heterozygous 'AB', or, 'noCall' (Affymetrix)/ 'NC' (Illumina). Similar matrix files containing copy numbers or gene expression values are optional. Copy numbers should be centered around 2, where 2 represents the normal copy number of the autosomes and 1 for the male X chromosome. The maximum displayed copy number is 4, in case the copy number is above 4 this is indicated by the greyblue background. Copy number-, genotype- and gene expression files required for SNPExpress can be generated through tools such as Affymetrix BRLMM [10], GCOS/CNAT 4.0 [24], or dChipSNP [19] with additional formatting in Microsoft Excel. In case of Illumina data, SNP Express includes the non-synonymous SNPs and the MHC region; however, mitochondrial SNPs and Y-chromosome SNPs are not visualized. All files can be optionally uploaded as tab- or comma-delimited .txt files or binary files. These binary files can be created from .txt files by the menu item 'convert data source'.

SNPExpress maps both the SNP IDs (Illumina and Affymetrix) and the expression probe set IDs (Affymetrix) to the genome through internal alignment tables, using annotation provided by the manufacturer [25-26] and [27]. Annotation was generated using NCBI build 36.1.

Regions showing LOH are calculated through a hidden Markov Model, which has been described previously [16]. The probability values for heterogeneous calls required for the HMM have been generated through sets of genotypes of normal samples. For the 100K and 500K

array, 90 samples and 270 samples, respectively, of different ethnical background from the HapMap project are available through the NCBI GEO website (and provided by the manufacturer) [28-29]. For the 10K array normal matched blood samples available through the GEO public repository have been processed [30]. Since reference normal Illumina genotype datasets are currently not publicly available, LOH regions using this platform are not supported in this version of SNPExpress.

SNPExpress includes the option to visualize the results of a novel analytical method that infers the copy number of each SNP based on a HMM model, which is implemented in dChipSNP [19,31]. Also, all CNVs [2], currently cataloged in the Database of Genome Variants [32], can be visualized.

Example expression, copy number, genotype and HMM copy number example files of two AML patients can be downloaded from [33].

RESULTS

Genotypes and copy numbers are displayed as sequential blocks of which color indicates genotype, horizontal coordinate indicates position on the chromosome and vertical coordinate indicates copy number (Figure 1). The colored genotype blocks are drawn sequential in chromosome-wide view and proportional to chromosomal location when zoomed into a region of interest. Gene expression levels are visualized as vertical bar at the chromosomal position of the gene-specific probe set. The height of the bar is proportional to the gene expression value. The default value is 500 and expression higher than 500 is capped at 500, however, these values are user-definable. In the event that multiple probe sets span the same region in the chromosome-wide view the vertical gene expression bars are red and proportional to the highest expression value. Zooming into the location of interest discloses the individual probe sets. Links of SNP IDs to public databases are available by holding the ctrl-key and clicking on a SNP ID.

Distinct background colors are used to accentuate genomic changes. Individual copy numbers are indicated as gain (pink background) or loss (green background) when their value exceeds a user-defined value. The default deviation threshold is 0.5. LOH is highlighted at diploid level by a bold magenta line (Figure 1). All colors can be adapted to the users' preferences.

From the menu, the user is able to choose to visualize either one chromosome of multiple samples or the complete genome of one sample. Detailed information, containing information such as SNP ID, associated gene symbol, probe set ID, cytoband and expression value, is shown on a mouse-over display. Furthermore, a gene of interest is directly visualized through a search function, and its associated SNPs are indicated with an orange background color. The options to display known CNVs (purple background) or the HMM copy number results (thin magenta line) are included (Figure 1C). Finally, relevant data of a particular minimal

deleted of amplified region can be exported (i.e. Sample, Probe_set_id, Chromosome, Location (bp), Cytoband, Associated gene, Genotype, Copy number and Inferred LOH of the selected region) and high-resolution images of the visualization can be saved in the Portable Network Graphic (PNG) format.

To illustrate the power of SNPExpress, DNA mapping array profiles of tumor samples of a series of 48 patients with AML were generated using Affymetrix 250K *Nsp*I DNA mapping arrays. Ficoll separation of the mononuclear cells from AML typically yields >80% pure population of leukemic blast cells. High molecular weight DNA was isolated from these malignant cells and the Affymetrix mapping arrays were used according to the protocol of the manufacturer. Genotypes were calculated using BRLMM and copy numbers were assessed using dChipSNP. Biotin-labeled cRNA of the same AML samples was hybridized on Affymetrix HG-U133 plus 2.0 GeneChips, as described elsewhere [34]. The resulting dataset was imported in SNPExpress for analyses. Large chromosomal regions showing loss or gains of genetic material are known to be apparent in leukemic blasts of AML patients. Well-known examples of chromosomal lesions in AML are monosomies of chromosome 5 and 7, which have been associated with a poor prognosis [35]. Using SNPExpress, monosomies of chromosome 7 were evidently demonstrated in AML samples, previously shown by cytogenetics to have lesions involving chromosome 7 (Figure 1). SNPExpress also correctly predicted the presence of LOH as a result of the absence of one chromosome 7. In fact, 17 out of 21 numerical cytogenetic aberrations, i.e., whole chromosomes and interstitial deletions, in 48 AML samples analyzed, were recognized by using SNPExpress. Four numerical abnormalities, present in less than 30% of the AML cells, were missed. Chromosomal gains, losses as well as uniparental disomy (UPD) may also have other important consequences, such as affecting expression of (imprinted) genes. Combinatorial visualization of genotype, copy number and gene expression is a prerequisite to recognize these aberrations. For example, the majority of genes show located on chromosome 7 show an overall decrease in expression in AML samples with a monosomy 7 (Figure 1).

Large regions of homozygosity are present in approximately 20% of primary AML cases as a result of segmental UPD [4,36]. These regions of UPD seemed to be non-random and may be used to unmask pre-existing recessive mutations in leukemia genes, such as *CEBPA*, *WT*1, *FLT3* and *RUNX*1 [4,37]. SNPExpress adequately identified regions of UPD involving e.g. chromosome 11p (Figure 2), in two patients with a normal karyotype. UPD involving chromosome 11 is associated with homozygous mutations in *WT1* [37]. Interestingly, in 13 out of 48 AML patients (27%) large regions of segmental UPD continuing to the telomere were recognized using SNPExpress.

These examples demonstrate the power of SNPExpress. To our knowledge, no tool is currently available that allows concurrent interpretation of genotype, HMM inferred LOH regions, copy number, CNVs, HMM inferred copy number and gene expression data. Moreover, no specialized knowledge is necessary to work with SNPExpress.

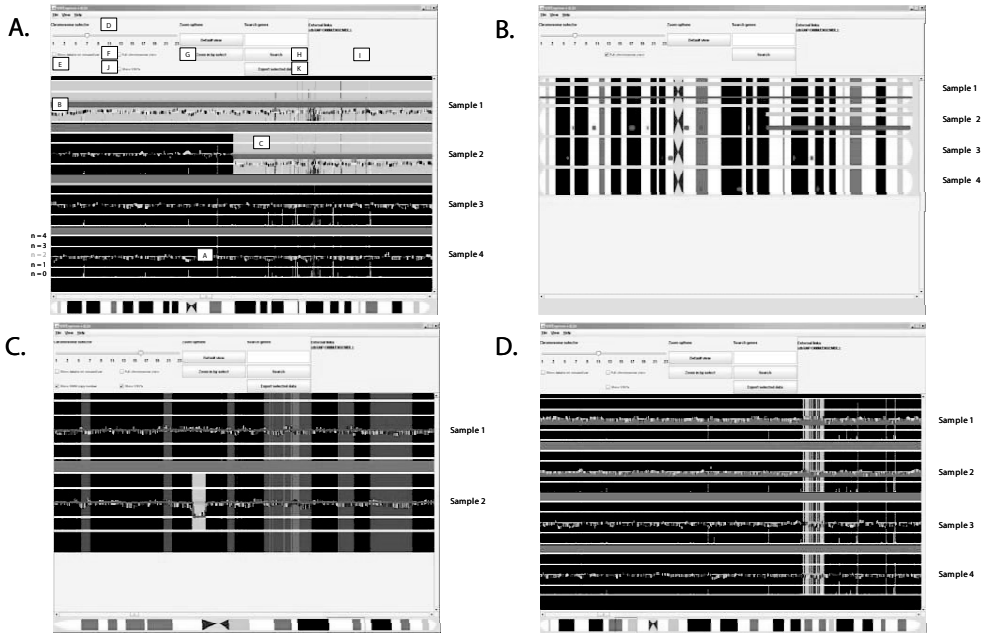


Figure 1

SNPExpress Screenshot. A. DNA mapping array data from the Affymetrix 250K *Nsp*I DNA mapping array was used to sequentially align the genotypes and copy numbers of chromosome 7 of four AML samples. The copy numbers (n=0, 1, 2, 3, 4) are shown for each individual patient by horizontal lines. Copy number n=2 is depicted by a green line (A). The SNP genotypes are sequentially aligned along the chromosome (AA: red; BB: yellow; AB: blue, noCall: white). LOH is indicated by a thick magenta horizontal bar (B), gains (default n>2.5) by a pink (Figure 1C) and losses (default n<1.5) by a turquoise background (C). Gene expression levels are visualized as vertical white bar at the chromosomal position of the gene-specific probe set. In the event that multiple probe sets span the same region in the chromosome-wide view the vertical gene expression bars are red and proportional to the highest expression value. The two upper samples clearly display a decreased copy number as was previously shown by cytogenetics, i.e., a complete monosomy (sample 1) or a deletion of the q-arm of chromosome 7 (sample 2). The overall expression of the majority of genes in the displayed region is decreased in the samples with chromosome 7 abnormalities. The chromosome selector (D; where 23 is the X chromosome), the mouse-over function showing info of each SNP or probe set (E), full chromosome view (F), zoom function (G) gene search function (H), the links to external databases (I), display CNVs (J) and export selected data (K) options are indicated. **B.** Full chromosome view of samples from 1A. **C.** CNV (purple background) and copy number of each SNP based on a HMM model (HMM copy number, magenta line) of the two AML patients from examples [33]. In the event that multiple CNVs span the same region in the chromosome-wide view the background is violet, whereas single CNV are indicated with a rosy brown background. **D.** UPD of chromosome 11 demonstrated using SNPExpress. Example of large scale UPD on chromosome 11 in the upper two AML patients with a normal karyotype in comparison to two other AML samples. The overall copy number is two and large regions of LOH are indicated by the thick magenta line across the chromosome. After using the search function, SNPs associated with *WT1* are depicted with an orange background.

DISCUSSION

Since genome-wide DNA mapping array and mRNA expression studies become more cost effective, the number of samples profiled on these platforms will increase. Specialized user-friendly tools for efficient visualization, such as SNPExpress, will therefore be indispensable. In fact, the initial version of SNPExpress has already been successfully applied in showing segmental uniparental disomy as a recurrent mechanism for homozygous *CEBPA* mutations in acute myeloid leukemia [38].

Other tools for visualizing and processing SNP array data, such as SNPScan [39], SIGMA [40], Array Fusion [41], Partek Genomics Suite [42] and GenePattern [43] have been developed. Most of these tools incorporate visualization options for displaying LOH (GenePattern, Partek Genomics Suite, SNPScan) and copy number (all but ArrayFusion), whereas SNPScan and ArrayFusion have output functionality that facilitates linking SNP data to the UCSC genome browser [39,41]. Some are linked to a private database, which restricts pre-processing of the array data, but gives the advantage of data storage [40]. GenePattern and the Partek Genomics Suite provide normalization and data smoothing functionality. These two packages and SNPScan have also incorporated options for combined analysis of paired samples, i.e., tumor and normal. Like SNPExpress, SNPscan, GenePattern, and the Partek Genomics Suite can detect regions of LOH, amplification and deletion. None of these tools describe the ability to process Illumina BeadArray files. Where SNPExpress may lack the opportunity to directly process raw data files (such as Affymetrix CEL-files), it adds integrated visualization of expression (Affymetrix) and DNA copy number and genotype (Affymetrix and Illumina) data. Moreover, we believe that this is provided in a user-friendly way that does not require specialist computer knowledge.

SNPExpress has some limitations. A full-length chromosome view depicting gains, losses and the regions showing LOH is feasible using SNPExpress. However, the large datasets generated by the 500K mapping array platform makes it impossible to visualize the sequentially aligned SNPs of the full-length chromosomes on one screen. Selecting the most informative SNPs, i.e., representative for particular haplotypes may solve this issue. Such algorithms are currently in development. Furthermore, the current implementation of the HMM could also be improved by implementing a HMM that takes into account the effects of linkage disequilibrium, i.e., LD-HMM [20]. The number of samples to be visualized concurrently is limited by the memory available to the application.

CONCLUSIONS

The power of SNPExpress, as with previously developed tools [44], is its high accessibility and powerful visualization, which facilitates the identification of biologically and clinically relevant entities. We have shown that recurrent biologically relevant entities, such as chromosomal gains or losses and LOH in AML, are accurately identified with SNPExpress. Hence, SNPExpress will be beneficial to genome-wide studies by providing an integrated view of data from DNA mapping and mRNA expression arrays in an easily accessible and accurate way.

AVAILABILITY AND REQUIREMENTS

Project name: SNPExpress
Project homepage: http://www.erasmusmc.nl/hematologie/SNPExpress
(Including downloadable genotype-, copy number-, expression- and HMM copy number example files of two AML patients genotyped with Affymetrix 250K *Nsp*I DNA mapping array and gene expression profiled with Affymetrix U133Plus2.0 GeneChips)

Operating system: Platform independent

Programming language: JAVA

Other requirements: JAVA 1.5 or higher.

License: The tool is available free of charge. Source code is available upon request.

Any restrictions to use by non-academics: None

LIST OF ABBREVIATIONS

AML Acute Myeloid Leukemia
PNG Portable Network Graphics
BRLMM Bayesian robust linear model with Mahalanobis distance classifier
SNP Single nucleotide polymorphism
HMM Hidden Markov Model
CNV Copy Number Variation
LOH Loss-of-heterozygosity

AUTHORS' CONTRIBUTIONS

MAS wrote and designed the software; RGWV designed the software, performed the analysis and wrote the manuscript; WGK performed experiments; SA gave intellectual contributions; SH contributed code; PJS gave intellectual contributions; BL gave intellectual contributions; PJMV designed the study, gave intellectual contributions and wrote the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

The research described was supported by grants from the Erasmus University Medical Center (Revolving Fund) and the Dutch Cancer Society "Koningin Wilhelmina Fonds". We are indebted to Andy Hall for providing Affymetrix 10K DNA mapping array data at the initial set up of SNPExpress.

REFERENCES

1 A haplotype map of the human genome. *Nature* **437**, 1299-1320 (2005).

2 Freeman, J. L. *et al.* Copy number variation: new insights in genome diversity. *Genome Res* **16**, 949-961 (2006).

3 Irving, J. A. *et al.* Loss of heterozygosity in childhood acute lymphoblastic leukemia detected by genome-wide microarray single nucleotide polymorphism analysis. *Cancer Res* **65**, 3053-3058 (2005).

4 Raghavan, M. *et al.* Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res* **65**, 375-378 (2005).

5 Janne, P. A. *et al.* High-resolution single-nucleotide polymorphism array and clustering analysis of loss of heterozygosity in human lung cancer cell lines. *Oncogene* **23**, 2716-2726 (2004).

6 Lindblad-Toh, K. *et al.* Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. *Nat Biotechnol* **18**, 1001-1005 (2000).

7 Zhao, X. *et al.* Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res* **65**, 5561-5570 (2005).

8 Nakao, K. *et al.* High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis* **25**, 1345-1357 (2004).

9 Garraway, L. A. *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117-122 (2005).

10 Affymetrix: BRLMM: an Improved Genotype Calling Method for the GeneChip® Human Mapping 500 K Array Set Santa Clara, CA 2006:1-18 [http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf].

11 Di, X. *et al.* Dynamic model based algorithms for screening and genotyping over 100 K SNPs on oligonucleotide microarrays. *Bioinformatics* **21**, 1958-1963 (2005).

12 Lamy, P., Andersen, C. L., Wikman, F. P. & Wiuf, C. Genotyping and annotation of Affymetrix SNP arrays. *Nucleic Acids Res* **34**, e100 (2006).

13 Rabbee, N. & Speed, T. P. A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* **22**, 7-12 (2006).

14 Huang, J. *et al.* Whole genome DNA copy number changes identified by high density oligonucleotide arrays. *Hum Genomics* **1**, 287-299 (2004).

15 Nannya, Y. *et al.* A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* **65**, 6071-6079 (2005).

16 Beroukhim, R. *et al.* Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. *PLoS Comput Biol* **2**, e41 (2006).

17 Huang, C. C., Taylor, J. M., Beer, D. G. & Kardia, S. L. Hidden Markov model for defining genomic changes in lung cancer using gene expression data. *OMICS* **10**, 276-288 (2006).

18 LaFramboise, T. *et al.* Allele-specific amplification in cancer revealed by SNP array analysis. *PLoS Comput Biol* **1**, e65 (2005).

19 Lin, M. *et al.* dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**, 1233-1240 (2004).

20 Monni, O. *et al.* Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc Natl Acad Sci U S A* **98** (2001).

21 Rucker, F. G. *et al.* Disclosure of candidate genes in acute myeloid leukemia with complex karyotypes using microarray-based molecular characterization. *J Clin Oncol* **24**, 3887-3894 (2006).

22 Stranger, B. E. *et al.* Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* **315**, 848-853 (2007).

23 Stranger, B. E. *et al.* Population genomics of human gene expression. *Nat Genet* **39**, 1217-1224 (2007).

24 Affymetrix GCOS/Affymetrix CNAT4.0 [http://www.affyme trix.com/support/technical/product_updates/gcos_download.affx]. http://www.affymetrix.com/Auth/products/software/download/cnat_terms.affx?p=1.2.1

25 Illumina Annotation HapHumanHap550 Genotyping Bead-Chip [http://www.illumina.com/pages.ilmn?ID=154].

26 Affymetrix Mapping Array Annotation [http://www.affyme trix.com/support/technical/byproduct.affx?cat=dnaarrays].

27 Expression array probe set alignments [http://www.affyme trix.com/Auth/analysis/downloads/psl/HG-U133_Plus_2.link.psl.zip].

28 Reference dataset Affymetrix 500 K Mapping Array [http://www.affymetrix.com/support/technical/sample_data/hapmap_trio_data.affx].

29 Reference dataset Affymetrix 500 K Mapping Array [http://www.affymetrix.com/support/technical/sample_data/500k_data.affx].

30 Reference dataset Affymetrix 10 K Mapping Array [http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2959].

31 Zhao, X. *et al.* An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* **64**, 3060-3071 (2004).

32 Database of Genome Variants [http://projects.tcag.ca/variation].

33 Homepage SNPExpress [http://www.erasmusmc.nl/hematologie/SNPExpress].

34 Valk, P. J. *et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* **350**, 1617-1628 (2004).

35 Mrozek, K., Heerema, N. A. & Bloomfield, C. D. Cytogenetics in acute leukemia. *Blood Rev* **18**, 115-136 (2004).

36 Gorletta, T. A. *et al.* Frequent loss of heterozygosity without loss of genetic material in acute myeloid leukemia with a normal karyotype. *Genes Chromosomes Cancer* **44**, 334-337 (2005).

37 Fitzgibbon, J. *et al.* Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. *Cancer Res* **65**, 9152-9154 (2005).

38 Wouters, B. J. *et al.* Segmental uniparental disomy as a recurrent mechanism for homozygous CEBPA mutations in acute myeloid leukemia. *Leukemia* **21**, 2382-2384 (2007).

39 Ting, J. C., Ye, Y., Thomas, G. H., Ruczinski, I. & Pevsner, J. Analysis and visualization of chromosomal abnormalities in SNP data with SNPscan. *BMC Bioinformatics* **7**, 25 (2006).

40 Chari, R. *et al.* SIGMA: a system for integrative genomic microarray analysis of cancer genomes. *BMC Genomics* **7**, 324 (2006).

41 Yang, T. P., Chang, T. Y., Lin, C. H., Hsu, M. T. & Wang, H. W. ArrayFusion: a web application for multi-dimensional analysis of CGH, SNP and microarray data. *Bioinformatics* **22**, 2697-2698 (2006).

42 Partek Discovery Suite http://www.partek.com/partekds.

43 Reich, M. *et al.* GenePattern 2.0. *Nat Genet* **38**, 500-501 (2006).

44 Verhaak, R. G. *et al.* HeatMapper: powerful combined visualization of gene expression profile correlations, genotypes, phenotypes and sample characteristics. *BMC Bioinformatics* **7**, 337 (2006).

CHAPTER 3

Integrated genome-wide genotyping and gene expression profiling reveals *BCL11B* as a novel oncogene in acute myeloid leukemia

Saman Abbas[1], Annelieke Zeilemaker[1], Wendy M.C. Geertsma-Kleinekoort[1], Zabiollah G.Abbas[1], Souad Mahamoud[1], Isabel W.T. Chu[1], Mathijs Sanders[1], Justine K. Peeters[1], Ellen van Drunen[2], Janneke van Galen[2], H. Berna Beverloo[2], Bob Löwenberg[1] and Peter J.M. Valk[1]

[1] Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands (P.J.M.V., C.A.J.E., S.A., B.W., R.D., B.L.);
[2] Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands (H.B.B.);

Manuscript submitted

ABSTRACT

Acute myeloid leukemia (AML) is a neoplasm characterized by recurrent molecular aberrations traditionally demonstrated by cytogenetic analyses. To reveal cryptic abnormalities in normal karyotype AML, we have used high density genome-wide genotyping and gene expression profiling. By genome-wide genotyping we disclosed a recurrent focal amplification on chromosome 14q23.The *BCL11B, CCNK, C14orf17* and *SETD3* genes are located within the amplified region on 14q32. By gene expression profiling we showed that the two AML samples have consistent high mRNA expression of the *BCL11B* gene, whereas the expression of the other genes was unchanged. FISH analyses confirmed the focal amplifications in the two index AML cases as well as in two additional cases, which were selected based on high *BCL11B* mRNA expression. All newly identified 14q32 aberrations were fused to different partner chromosomes and resulted in increased expression of full-length BCL11B protein. Immuno-phenotyping of the *BCL11B* rearranged AMLs revealed expression of both myeloid and T-cell markers. These biphenotypic acute leukemias all carried *FLT3* internal tandem duplications, a characteristic marker for AML. In AML, *BCL11B* mRNA expression appeared to be strongly associated with expression of other T-cell specific genes. 32D(GCSF-R) cells ectopically expressing *BCL11b* showed decreased proliferation rate and less maturation. In conclusion, by an integrated approach involving high-throughput genome-wide genotyping and gene expression profiling we identified BCL11B as a novel recurrent oncogene in AML.

INTRODUCTION

Acute myeloid leukemia (AML) is a heterogeneous clonal neoplasm characterized by accumulated genetic aberrations, which result in enhanced proliferation, block in differentiation and increased survival, of the leukemic blast cells and variable response to therapy [1-2].

In the past decades a number of recurrent genetic abnormalities have been identified in AML, such as the chromosomal aberrations t(8;21) and inv(16) [1-2]. These recurrent molecular lesions result in the expression of fusion proteins of which the leukemic potential has been demonstrated *in vitro* and *in vivo* models [3-5]. In addition, acquired mutations in disease genes such as *FLT3, NPM1,* and *CEBPA* have recently been demonstrated in AML [1-2]. These have been shown to be involved in leukemogenesis as well [6-8]. The acquired molecular aberrations may carry prognostic value and have been incorporated in the routine molecular analyses of AML [1-2].

Nowadays, various genome-wide approaches, such as gene expression profiling (GEP), genome-wide genotyping and next generation sequencing, enable detailed analyses of hematologic malignancies to identify novel pathogenic genes [9-12]. Examples of aberrations in myeloid proliferative malignancies revealed with these novel technologies are gene mutations in *IDH1, TET2*, *ASXL1*, and *EZH2* [13 14 15 16 17-19].

Besides the balanced translocations, large chromosomal regions showing loss or gains of genetic material are apparent in the leukemic blast of AML patients, e.g., those involving chromosome 5 and 7 [1-2]. In the past two decades, attempts to identify the tumor suppressor genes located on these chromosomes have failed. By genome-wide SNP genotyping it has become possible to simultaneously genotype hundred thousands of single nucleotide polymorphisms (SNPs) in a single assay. In addition, SNP platforms can also be conveniently used to determine chromosomal copy numbers, similarly to array comparative genomic hybridization (CGH) with for instance bacterial artificial chromosomes (BACs). Genomic DNA can be examined with an inter-marker distance of several hundreds of base pairs, which makes it feasible to detect (micro) deletions and/or amplifications that are missed with conventional cytogenetics. The application of high-throughput SNP genotyping has been elegantly demonstrated to be powerful for the identification of disease genes, such as for ALL [11,20-21]. Another major advantage of SNP arrays is the fact that allele losses are directly recognizable as loss-of-heterozygosity (LOH). Moreover, SNP arrays revealed that approximately 20% of AMLs exhibit large non-random regions of homozygosity without changes in copy number as a result of segmental uniparental disomy (UPD), often indicating mutations in genes. These areas of UPD have been associated with mutations in *CEBPA, WT1, FLT3* and *RUNX1* [22-23].

In addition, deletion, amplification and UPD may have other important consequences, such as on gene expression. Juxtaposition of regulatory sequences may result in increased or decreased expression of affected genes. Genome-wide analyses to detect copy number

changes and LOH in the context of gene expression may in fact pinpoint towards these pathogenic genes. We recently developed SNPExpress, an easily accessible software tool to accurately analyze SNP genotype calls, copy number and gene expression in an efficient combinatorial way [24].

In this study, we identified *BCL11B* as a novel oncogene in AML through an integrated approach of genome-wide genotyping and GEP. *BCL11B* is a Kruppel family zinc finger family gene located at 14q32, associated with transcriptional corepressor complexes in mammalian cells and a pivotal regulator of differentiation and survival of haematopoietic cells [25] [26]. We demonstrate that *BCL11B* is involved in a number of cryptic 14q32 translocations in AML, in which *BCL11B* and T-cell associated genes expression levels are increased concomitantly. Overexpression of BCL11B in a murine myeloid cell line model inhibits proliferation.

MATERIALS AND METHODS

Patients samples

After informed consent, bone marrow aspirates or peripheral blood samples of a representative cohort of AML patients were collected. Eligible patients had a diagnosis of primary AML, confirmed by cytological examination of blood and bone marrow. All patients were treated according to the HOVON (Dutch-Belgian Hematology-Oncology Co-operative group) protocols (http://www.hovon.nl). Blasts and mononuclear cells were purified by Ficoll-Hypaque (Nygaard, Oslo, Norway) centrifugation and cryopreserved. The AML samples contained 80-100 percent blast cells after thawing, regardless of the blast count at diagnosis.

Genome-wide genotyping and gene expression profiling

Genome-wide genotyping data sets of 48 patients with various subtypes of AML were generated using Affymetrix 500K *Nsp*I/*StyI* DNA mapping arrays and 89 patients with cytogenetically normal AML using Affymetrix 250K *Nsp*I or *StyI* DNA Mapping arrays. High-molecular-weight DNA was isolated using the standard high salt procedures and the Affymetrix mapping arrays were used according to the protocol of the manufacturer (Affymetrix, Santa Clara, CA). In brief, 250 ng of genomic DNA was digested with *Nsp*I or *Sty*I (New England Biolabs, Beverly, MA, USA) and ligated to an *Nsp*I or *Sty*I adapter (Affymetrix) using T4 DNA ligase (New England Biolabs, Ipswich, MA, USA). Samples were then amplified by PCR using TITANIUMTaq polymerase (Clontech, Mountain View, CA, USA). PCR products were pooled and purified using the Clontech purification kit and subjected to fragmentation using DNaseI (Affymetrix). The DNA fragments were subsequently biotin-labeled with terminal deoxynucleotidyl transferase (Affymetrix), hybridized on the array in a GeneChip® Hybridization Oven 640 (Affymetrix), and followed by washing and staining in a GeneChip® Fluidics Station 450 (Affymetrix). Data was obtained using the GeneChip Scan-

ner 3000 7G (Affymetrix). Genotypes were calculated using BRLMM (Affymetrix: BRLMM: an Improved Genotype Calling Method for the GeneChip® Human Mapping 500K Array Set. In. Santa Clara, CA; 2006: 1-18.(http://www.affymetrix.com/support/technical/whitepapers/brlmm_whitepaper.pdf)) and copy numbers were assessed using dChipSNP [27]. The copy numbers of all AML samples were calculated using diploid references, i.e., 15 normal karyotype AML samples that were included in the series of 137 AML samples analyzed.

Gene expression profiles of the same AMLs were generated using Affymetrix HG-U133 plus 2.0, as described elsewhere [9]. These data sets are accessible at the NCBI's Gene Expression Omnibus through GEO Series accession number GSE6891.

The genome-wide genotyping and gene expression profiling data sets were examined using SNPExpress [24]. SNPExpress is an easily accessible software tool to accurately analyze Affymetrix and Illumina SNP genotype calls, copy numbers, polymorphic copy number variations (CNVs) and Affymetrix gene expression in a combinatorial and efficient way (freely available at http://www.erasmusmc.nl/hematologie/SNPExpress), SNPExpress allows concurrent interpretation of these items with Hidden-Markov Model (HMM) inferred Loss-of-Heterozygosity (LOH)- and copy number regions.

Fluorescence In Situ Hybridization (FISH)

Freshly prepared slides from stored fixed cytogenetic suspensions with methanol/ acetic acid (3:1) (at -20 C) were used to carry out dual color fluorescence *in situ* hybridization (FISH) with BAC clones RP11-431B1, RP11-876E22, RP11-830F3, RP11-782I5, RP11-450C22, RP11-57E12, RP11-1069L3 and RP11-242A7 covering the BCL11B region and regions up- and downstream (BACPAC resources, Oakland, USA). Clone isolation and labeling were performed using biotin-16-dUTP and digoxigenin-11-dUTP (Roche Diagnostics Belgium, Vilvoorde, Belgium) according to the manufacturer's protocol. The FISH analysis was performed as previously described [28]. Slides were embedded in DABCO/Vectashield containing 40, 6-diamidino-2-phenylindole (DAPI) as counterstain. Fluorescent signals were visualized with an epi-fluorescence microscope (Zeiss, Axio-Imager Z1, Zeiss, Sliedrecht, the Netherlands) using Metasystems Ikaros software (Metasystems, Altlussheim, Germany).

Both interphase nuclei and metaphase analysis was carried out to determine the presence or absence of the fluorescent signal and to determine it's localization.

Western blot analyses

For Western blot analyses, 100 micrograms of whole cell extracts (lysis buffer: 20 mM Tris pH=8, 137 mM NaCl, 50 mM NaF, 10 mM EDTA, 1% NP40, 10% Glycerol) were heated prior to electrophoresis for 5 minutes at 95°C. Samples were separated on by 10% SDS-PAGE and transferred to Protran BA83 Nitrocellulose membranes (Whatman, Dessel, Germany). Membranes were then blocked in block-buffer (10% BSA (0.6%), 0.5M EDTA (1mM), TBS/0.05% Tween) and immunoblotted with affinity-purified rabbit polyclonal anti-BCL11b

antibody (Novus Biologicals, Littleton, USA). Immune complexes were detected by binding anti-mouse IgG conjugated to horseradish peroxidase (DAKO, Heverlee, Belgium) followed by the enhanced chemiluminescence assay (Amersham Bioscience, Piscataway, NJ) according to the manufacturer's recommendations. GAPDH was stained with primary affinity-purified rabbit polyclonal antibody (α-GAPDH FL-335) (Santa Cruz Biotechnology, California, USA).

DNA constructs and generation of BCL11B expressing 32D/GCSFR cells

A murine Bcl11b cDNA (kindly donated by Dorina Avram, Albany Medical Center, Albany, NY) was subcloned into pLXSN expression vector under control of 5' long terminal repeat (LTR) of the Moloney murine sarcoma virus (MoMSV) (Clontech, Mountainview, USA). Vector constructs were confirmed by nucleotide sequencing and retrovirally transfected into 32D cells that stably express human granulocyte colony-stimulating factor receptor (GCSF-R) [29] using Fugene transfection reagent (Roche, Indianapolis, USA). Cells were cultures in RPMI 1640 medium with 10% fetal calf serum (FCS), penicillin and streptomycin and supplemented with interleukin-3 (IL3, 25ng/ml) and selected with neomycin. To study proliferation and differentiation, 32D cells, transduced with murine BCL11b, were plated in IL3 (25ng/ml) or GCSF (25 ng/ml), counted and assessed for granulocytic differentiation by morphological criteria. Morphological analysis was performed by microscopy on May-Grünwald-Giemsa-stained cytospins (Shandon Holland, Amsterdam, The Netherlands) using a Zeiss Axioskop microscope with 63x plan-apochromat objective. Pictures were taken with a Leica DC 500 camera.

RESULTS

Genome-wide genotyping of cytogenetically abnormal and normal AML cases

In total, DNA mapping array profiles of leukemia samples of 137 patients with AML were generated. Initially a subgroup of 48 AML samples was selected based on previous GEP studies, i.e., 21 AML cases form GEP clusters #4 and #15 (100% *CEBPA* mutant or *CEBPA* silenced [30]), 13 AML cases from GEP cluster #9 (100% inv(16)) and 15 AML cases form GEP cluster #10 (adverse prognosis) [9]. In addition, DNA mapping array profiles, i.e., Affymetrix 250K *Nsp*I or *Sty*I DNA mapping array, of 89 AML cases with normal karyotypes were generated.

With the Affymetrix 500K *Nsp*I/StyI DNA Mapping arrays of the cytogenetically abnormal AML samples, all known numerical cytogenetic aberrations, i.e., whole chromosome and interstitial deletions and amplifications that had been identified with cytogenetic banding analysis, were recognized in the 48 cytogenetically abnormal AML samples, as long as the abnormalities were present in over 30% of the AML cells [24]. Also, in approximately 25% of all cases large regions of segmental uniparental disomy were detected, often involving whole chromosome arms [2]

In addition to the known cytogenetic aberrations, relatively low numbers of small interstitial deletions and amplifications were detected in the 137 AML cases. However, some of these were indicative for the presence of cryptic translocations, such as cryptic t(5;11). t(9;22) and t(4;11), which are known to encode chimeric fusion protein essential for leukemognesis. All fusion transcripts involved in these translocations, i.e., *NUP98-NSD1* (NUP98-NSD1 fusion characterizes a novel poor prognostic group in acute myeloid leukemia with a distinct HOX gene expression pattern, NUP98-NSD1 (Hollink et al. (submitted)), *BCR-ABL* and *MLL-AF4*, were confirmed by RT-PCR. Thus, although relatively small numbers of aberrations were found, most being not recurrent, they may reliably tag relevant leukemic lesions.

Integrated analysis of genome-wide genotypes and gene expression profiles

By an integrated approach using the genome-wide genotyping data and previously determined gene expression profiles [9] of the primary AML samples we searched for genes aberrantly expressed as a result of numerical changes in the AML genome. Using SNPExpress [24], we identified two AML cases with relatively small interstitial amplifications in the 14q32.2 region (#2301 amplification: 482 Kb, 3 copies) and (#7073 amplification: 460 Kb, 3 copies) (Figure 1A, 1B). The amplified region encompassed *BCL11B, CCNK, C14orf177, and SETD3.* Interestingly, *BCL11B* mRNA was highly expressed in the two AML cases with numerical changes, whereas expressions of *C14orf177*, *CCNK* and *SETD3* were not affected as compared to other AML cases (Figure 1A, 1B). This could indicate that as a result of a genomic rearrangement, *BCL11B* has become overexpressed in these AML cases. The small interstitial amplifications in AML #2301 and #7073 may pinpoint towards cryptic translocations as was seen in the AML cases carrying cryptic translocations t(5;11), t(9;22) and t(4;11).

Figure 1

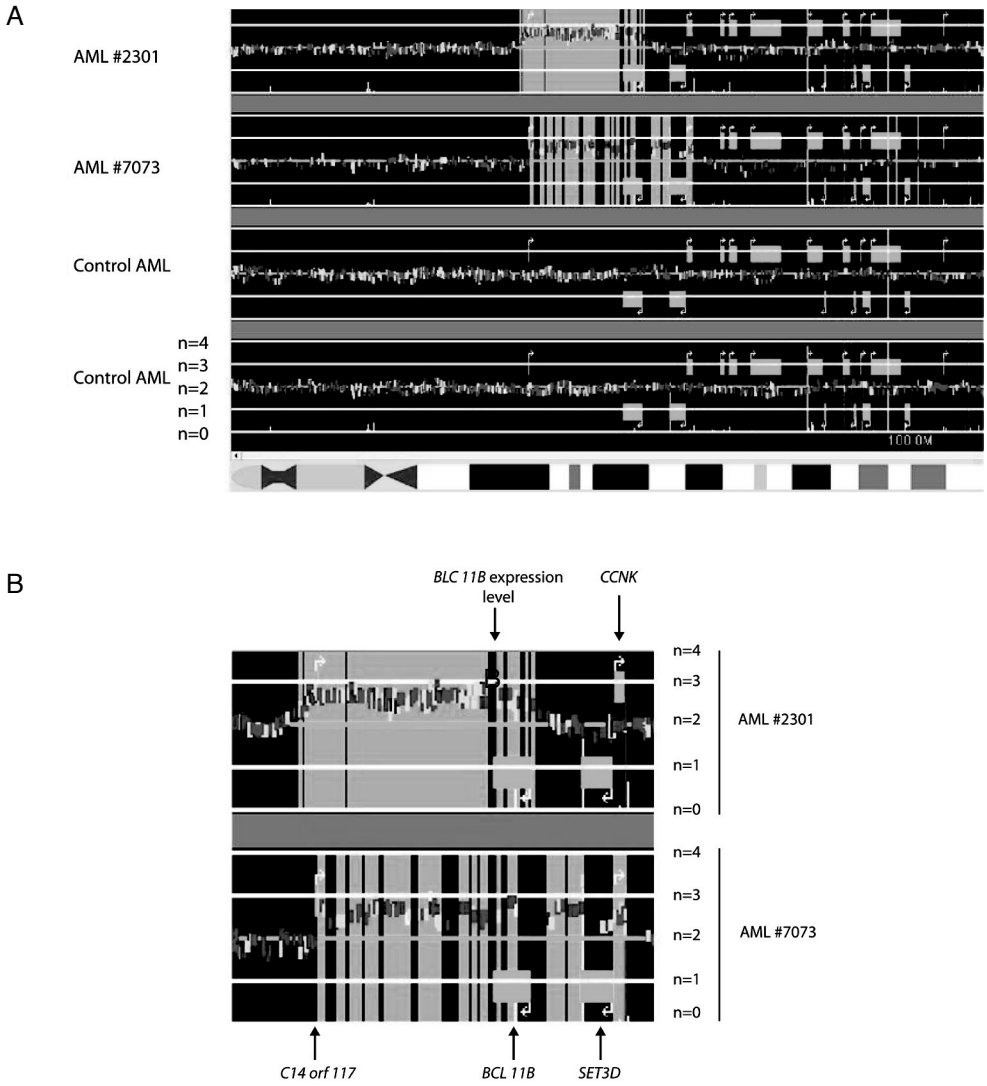


Figure 1: Identification of interstitial amplifications on 14q32.2 using SNPExpress. (A) Sequential alignment of the genotypes with copy numbers from the Affymetrix DNA mapping array of chromosome 14q32.2 of four AML samples [24]. The copy numbers are shown for each individual patient by horizontal lines (n=0, 1, 2, 3, 4). The SNP genotypes are sequentially aligned along the chromosome (AA: red; BB: yellow; AB: blue, No call: white). Gains (default n>2.5) are depicted as pink background. Gene expression levels are visualized as vertical white bars at the chromosomal position of the gene-specific probe set. In the event that multiple probe sets span the same region in the chromosome-wide view the vertical gene expression bars are red and proportional to the highest expression value. The green boxes represent exons of the encoded genes, and the arrows indicate the orientation. In AML #2301 and AML #7073 clear amplifications are visible, whereas these aberrations are absent in the two control AMLs. (B) Snapshot of SNPExpress showing the amplified region in AML case #2301 and #7073 from Figure 1A, showing the genes located within the amplified regions. *C14orf177* and *BCL11B* are amplified in both AML cases, whereas SETD3 and CCNK only in AML #7073. BCL11B expression is increased in AML #2301 and #7073 as indicated by the red bar (multiple probe sets).

FISH reveals translocations in AML #2301 and #7073 involving *BCL11B*

To confirm the amplifications in the *BCL11B* locus in the two AML cases, we performed FISH analysis with a probe covering the *BCL11B* gene (RP11-431B1) and a probe flanking this locus (RP11-242A7) (Figure 2A). On metaphase spreads of both AML cases an additional *BCL11B* allele was apparent (Figure 2B). This is in line with the expected copy number change for the *BCL11B* locus (n=3) as shown in SNPExpress (Figure 1A and B). In fact, through verification using chromosomal paints we showed that *BCL11B* was translocated to chromosome 6 in AML case #2301 and chromosome 8 in AML #7073 (data not shown).

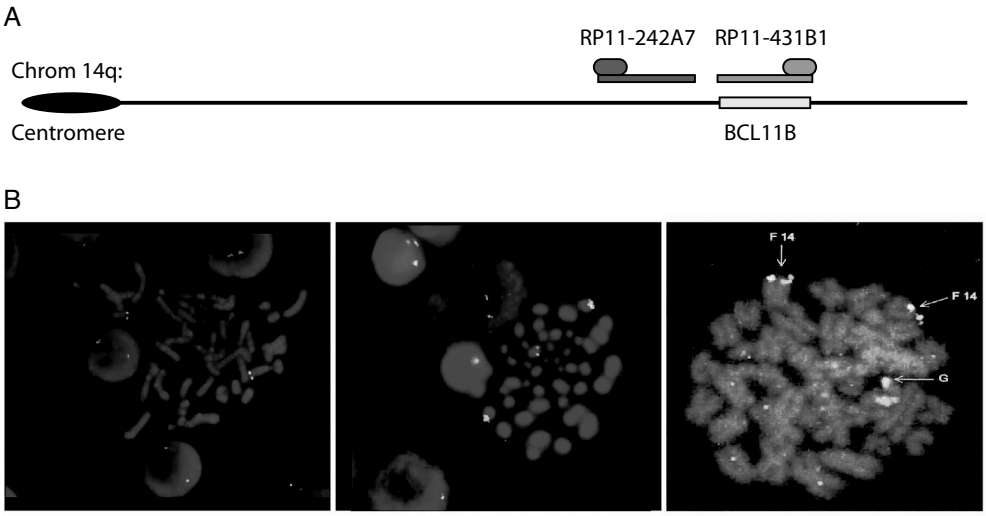


Figure 2: FISH analysis of AML cases #2301 and #7073 using probes specific for *BCL11B* and flanking *BCL11B*. (A) Schematic representation of the FITC-labeled BAC probe (RP11-431B1) covering the *BCL11B* locus and Texas Red-labeled BAC probe (RP11-242A7) covering the region adjacent to *BCL11B.* (B) Microscope images of FISH analysis performed on metaphases chromosomes of AML cases #2301 and #7073 showing additional green signal (RP11-431B1) indicative for an extra copy of the *BCL11B* locus.

AML cases #2301 and #7073 express full-length BCL11B

The translocations involving *BCL11B* could result in increased expression of either full-length BCL11B or a fusion protein involving BCL11B. Next, we examined the expression profiles obtained with Affymetrix Human Exon 1.0 ST Array for AML case #2301. This analysis showed that in AML #2301 all four exons of *BCL11b* were highly expressed at similar high levels (data not shown). The fact that exon 1 of *BCL11B,* containing the ATG start codon was expressed, suggested that full-length BCL11B is expressed instead of a fusion protein in which parts of BCL11B are involved.

Protein lysates of AML case #2301 were available for Western blot analyses. With Western blot analyses of whole, cytoplasmic, and nuclear cell lysates of #2301 we assessed both the size and localization of the BCL11B protein. The immunodetection with BCL11B antibodies

confirmed the expression of full length BCL11B protein (Figure 3). Moreover, BCL11B expression appears to be restricted to the nucleus. Of note, full-length BCL11B was also highly expressed in AML case #2238, an AML without any known aberration involving *BCL11B*..

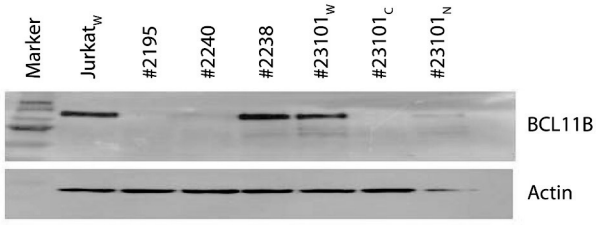


Figure 3: Western blot analysis of BCL11B in AML case #2301. Western blot analysis with a specific BCL11B antibody demonstrates high expression of full length BCL11B in AML case #2301 in the nuclear compartment (upper panel). Whole cell lysates from Jurkat, an acute T-cell leukemia cell line, and AML #2238 show high BCL11B expression. AML cases #2195 and #2240 with low *BCL11B* mRNA expression were used as negative controls (#2301$_w$:whole cell lysate; #2301$_c$: cytoplasmic lysate; #2301$_n$:nuclear lysate). β-actin was used as loading control (lower panel).

FISH analyses of selected AML cases with high *BCL11B* mRNA expression reveals additional cases with *BCL11B* aberrations

The act that FISH analysis of AML cases showing high expressing of *BCL11B* mRNA revealed translocations involving *BCL11B*, raised the possibility that other AML cases with aberrantly high *BCL11B* expression would harbour translocations involving *BCL11B* as well. Gene expression profiling of 461 AML cases [31] showed variable expression of *BCL11B* mRNA in a subset of AML cases, including case #2301 and #7073, but did not show a common gene expression profiling signature (Figure 4A). We selected 40 cases with increased *BCL11B* mRNA expression in AML, i.e., 2.5 times above mean expression of *BCL11B* in AML, and performed FISH analysis of the *BCL11B* chromosomal region.

FISH analyses revealed two additional AML cases with a *BCL11B* translocation (AML #6366 and #6451) (Figure 4B). With specific chromosomal paints, we showed that in AML case #6451 the *BCL11B* locus was translocated to chromosome 7. A similar analysis could not be carried out on AML #6366 due to the lack of material.

Subsequently, we examined seven additional AML cases with known 14q32 cytogenetic abnormalities for *BCL11B* translocations by FISH analysis, but abnormalities affecting the *BCL11B* locus were not detected (data not shown).

Table 1: Molecular analyses of the AML cases with *BCL11B* translocations. The mutation status for *FLT-3, NPM1, N-RAS, K-RAS, CEBPA* and *c-KIT* was determined as described previously 9.

	AML#2301	AML#6451	AML#6366	AML#7073	
FLT3-ITD	Pos	Pos	Pos	Pos	
FLT3-TKD	Neg	Pos	Neg	Neg	
NPM1	Neg	Neg	Neg	Neg	
N-RAS	Neg	Neg	Neg	Neg	
K-RAS	Neg	Neg	Neg	Neg	
CEBP-α	Neg	neg	neg	Neg	
Karyotype	46,XY[21]/?46,XY,inc[9(q46,XY[21]/?46,XY,inc[9]	46,XX,del(7)(q21q35)[5]/46,idem,add(13)(q3?4)[17]/46,idem,add(9)(q3?4)[2]/46,XX[15]	53,XX,+4,+8,+10,+13,+14,+15,+20[4]/46,XX[35]	46,XY[20]	
Immunophenotype	Biphenotypic leukemia	AML: 80% myeloid blasts	AML: 68% myeloid blasts	AML: 80% myeloid blasts	
	CD45(+), HLA-DR-, CD34+, TdT+, MPO partial+, CD1-, CD2+, CD3+, CD4-, CD5-,	CD45(+), HLA-DR+, CD34+, TdT partial+ , MPO partial+, CD11c-, CD13+, CD15-, CD15s partial, CD33-, CD65s-, CD117+, CD133+, CD2+, CD7 partial+ (negative for B and (remaining) T cell markers) CD7+, CD8-, CD11c-, CD13+, CD15-, CD33(+), CD117+,	CD45(+), HLA-DR+, CD34partial+, TdT-, MPO partial+, CD11c partial+, CD13partial+,	CD15partial+, CD33+, CD36partial+, CD56-, CD65s-, CD117partial+, CD133+, CD4partial+,	CD7(+)
cKIT-exon8	Neg	Neg	Neg	Neg	
cKIT-exon17	Neg	Neg	Neg	Neg	
FAB	M1	M1	M2	M4	
Gender	M	F	F	M	
WHO	1 WHO	2 WHO	0 WHO	1 WHO	

Immunophenotyping and molecular analyses of AML cases carrying *BCL11B* aberrations

Immunophenotyping on the AML cases enrolled in the gene expression profiling showed that the AML cases with *BCL11B* translocations expressed besides myeloid markers also lymphoid markers such as CD2, CD3, and CD7 (Table 1). These AML cases appeared to have a biphenotypic signature, i.e., these cases expressed early myeloid as well as T-cell associated markers.

Well-known recurrent molecular abnormalities were determined in the cases carrying *BCL11B* aberrations. These analyses demonstrated that, with no exception, these AML cases carried mutations in the *FLT3* gene mutation, i.e., FLT3 internal tandem duplication (ITD) in all cases and one case with a concurrent *FLT3* mutation in the tyrosine kinase domain (TKD) (Table 1). We did not identify mutations in K-*RAS*, N-*RAS*, c-*KIT* or *CEBPA* in the AML cases carrying a *BCL11B* abnormality.

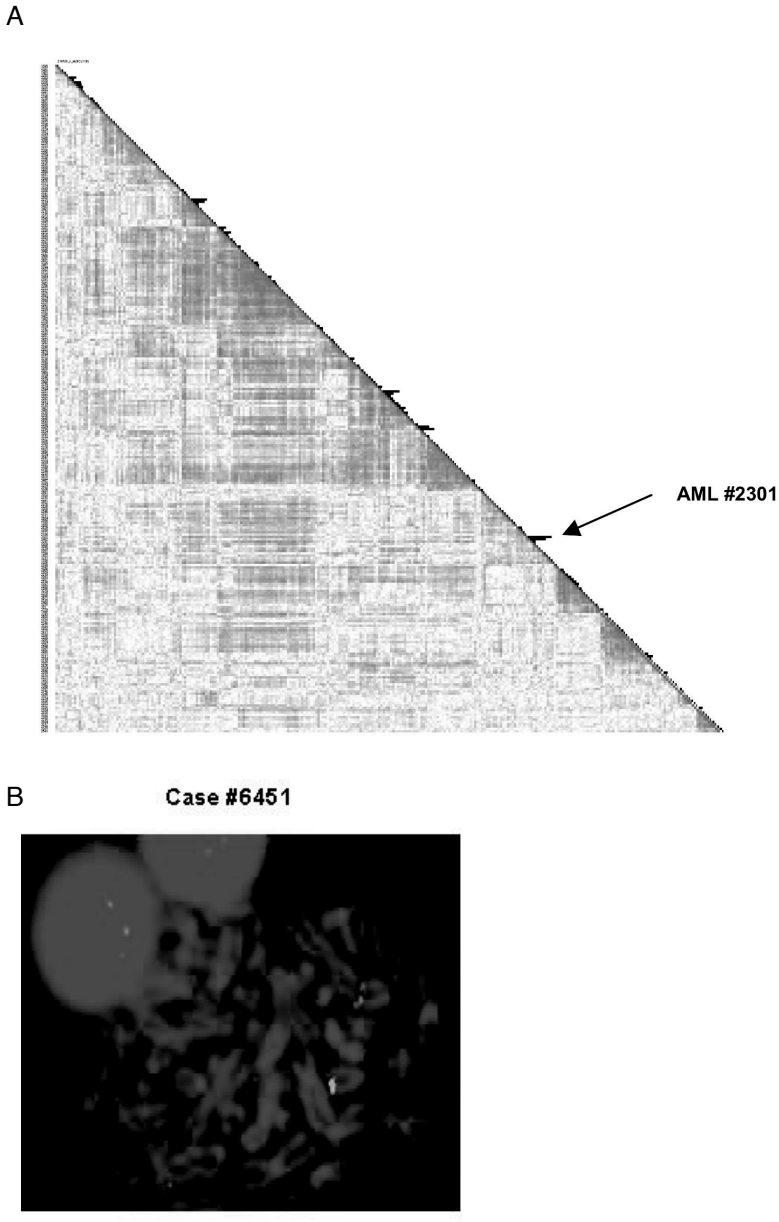


Figure 4: Correlation view based on gene expression profiling of 285 AML cases [9] (A) Correlation view of 285 AML cases showing the gene expression correlation based on 2856 probe sets [9]. The black bars adjacent to the correlation view indicate expression of *BCL11B*, where the size of the bars is proportional to the levels of *BCL11B* expression in the individual AML samples. *BCL11B* expression in case #2301 is indicated by an arrow. Forty cases with increased *BCL11B* expression (2.5 times above average) were selected for FISH analyses. (B) FISH analysis performed on metaphase spreads of AML cases #6451 showing disassociation of the probe RP11-242A7 (red) and RP11-431B1 (green) indicating translocation of *BCL11B*.

BCL11B is aberrantly expressed in AML and associated with T-cell gene expression signature

To investigate whether other AML cases with elevated *BCL11B* mRNA expression show full-length BCL11B expression, we carried out Western blot analyses on a number of AML cases with high *BCL11B* mRNA expression. All analyzed samples with high *BCL11B* mRNA showed full length BCL11B protein expression at variable levels (Figure 5).

To examine which genes are co-expressed with *BCL11B* in AML, we performed a Pearson correlation analyses using Affymetrix gene expression profiling data of 461 AML cases [31]. *BCL11B* co-regulated probe sets were calculated across all AML patients. The top50 *BCL11B* correlating probe sets are highly associated with T-cells and T-cell development (Table 2). In fact, the majority of BCL11B associated genes are T-cell specific genes, such as *CD3, TRBV19, IL32, LCK, TCF7* and *CD2*, among many others (Table 2)

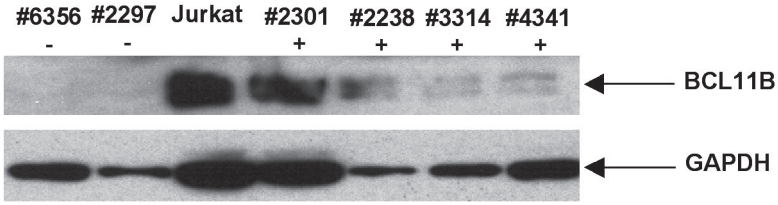


Figure 5: Western blot analyses for BCL11B of 4 primary AML cases. Immuno-detection of the BCL11B protein in AML cases with elevated levels of *BCL11B* mRNA (+) and cases with undetectable levels of *BCL11B* mRNA (-) (upper panel; Jurkat cell lysate as positive control). GAPDH was used as loading control (lower panel).

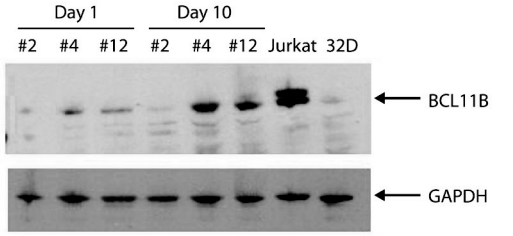
Increased *Bcl11b* expression results in decreased proliferation of the myeloid cell line 32D(GCSF-R)

To investigate the effect of *Bcl11b* expression on proliferation and differentiation, immortalized myeloblast-like murine bone marrow cells stably expressing human GCSF-R (32D(GCSF-R)) were transfected with pLXSN expression vector containing full length murine *Bcl11b* cDNA. Three 32D(GCSF-R) clones expressing *Bcl11b* were selected and incubated for ten days in the presence of interleukin-3 (IL3) or granulocyte stimulating factor (GCSF).

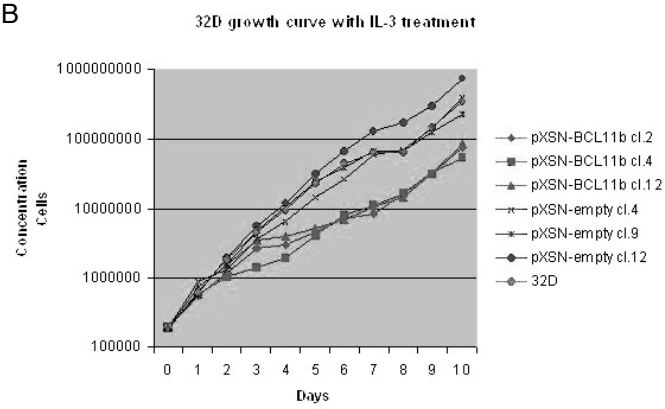
Western blot analyses showed that Bcl11b was expressed at all time points (Figure 6A). Bcl11b expressing 32D(GCSF-R) clones showed a consistent decreased proliferation rate when cultured in the presence of IL3 in comparison to 32D(GCSF-R) clones containing a control empty vector (Figure 6B).

We also evaluated the granulocytic differentiation abilities of the same 32D(GCSF-R) clones in the presence of GCSF. Morphological analyses of cytospins did not show consistent maturation defects in the Bcl11b expressing 32D(GCSF-R) clones. However, less maturation towards granulocytes in 32D(GCSF-R) cells expressing Bcl11b and more undifferentiated blast cells compared to cells with empty vector was present in some 32D(GCSF-R) Bcl11b expressing clones (Figure 6C). This effect was most apparent in 32D(GCSF-R) Bcl11b clones #4 and #12, the 32D(GCSF-R) clones with the highest expression levels of Bcl11b (Figure 6A).

A


Day 1
Day 10
#2 #4 #12 #2 #4 #12 Jurkat 32D
BCL11B
GAPDH

B


32D growth curve with IL-3 treatment
Concentration Cells
1000000000
100000000
10000000
1000000
100000
0 1 2 3 4 5 6 7 8 9 10
Days
pXSN-BCL11b cl.2
pXSN-BCL11b cl.4
pXSN-BCL11b cl.12
pXSN-empty cl.4
pXSN-empty cl.9
pXSN-empty cl.12
32D

C

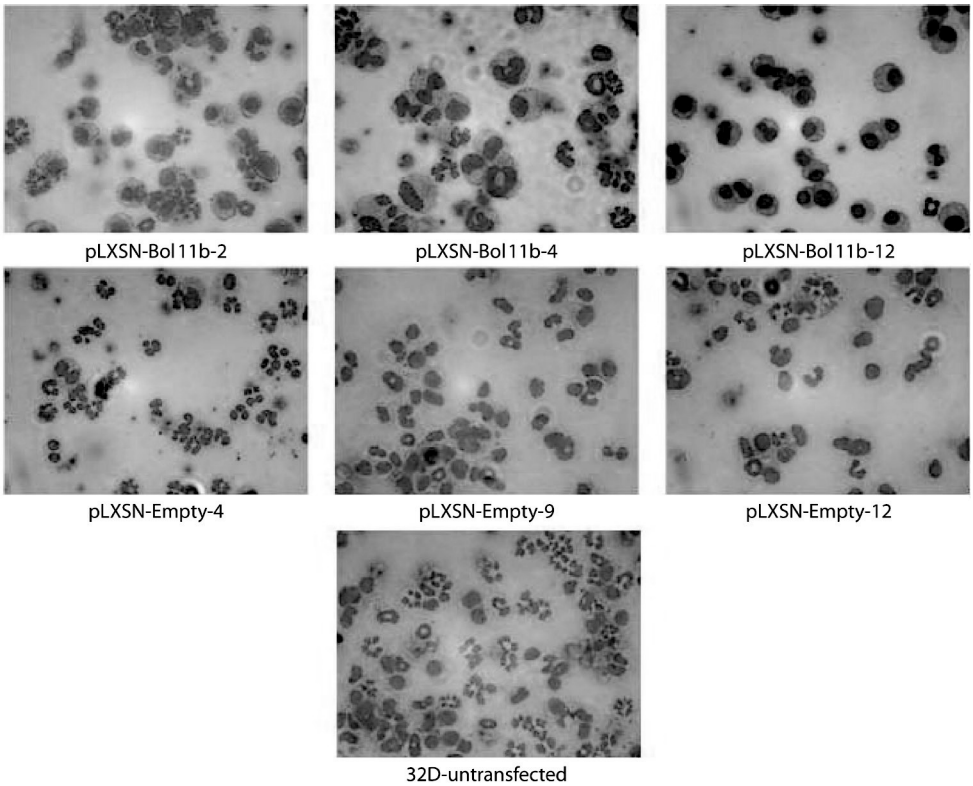

pLXSN-Bol11b-2
pLXSN-Bol11b-4
pLXSN-Bol11b-12
pLXSN-Empty-4
pLXSN-Empty-9
pLXSN-Empty-12
32D-untransfected

Figure 6: Effects of BCL11B overexpression in murine 32D(GCSF-R) cells. (A) Western blot analyses for Bcl11b in 32D(GCSFR) cells. 32D(GCSFR) clones overexpressing *BCL11B* are indicated by #2, #4, and #12 (IL3 1 and 10 days). Lysates obtained from theses clones were immunostained for BCL11b at day 1 and day 10 (Jurkat cells: positive control; 32D: 32D(GCSF-R) cells). GAPDH was used as loading control (lower panel). (B) Growth curve of 32D(GCSFR) cells with and without (empty) *BCL11B* expression incubated with IL3. 32D(GCSFR) cells were counted every 24 hours for ten days (32D: 32D(GCSF-R) cells). (C) May-Grünwald-Giemsa-stained cytospins of 32D(GCSF-R) cells with (upper panel) and without (lower panel) Bcl11b expression incubated with GCSF for 7 days. Granulocytic differentiation is monitored by the presence of cells with segmented nuclei.

Table 2: Top 50 genes correlating with *BCL11B* expression in AML

#	Probe set	Correlation coefficient	Gene	#	Probe set	Correlation coefficient	Gene
1	205456_at	0.83	CD3E	26	217838_s_at	0.63	EVL
2	206804_at	0.82	CD3G	27	210426_x_at	0.62	RORA
3	210915_x_at	0.81	TRBV19	28	219442_at	0.62	MGC3020
4	213193_x_at	0.80	TRBV19	29	202761_s_at	0.61	SYNE2
5	203828_s_at	0.80	IL32	30	210972_x_at	0.61	TRA@
6	204891_s_at	0.77	LCK	31	203413_at	0.61	NELL2
7	211796_s_at	0.77	TRBV21-1	32	210607_at	0.61	FLT3LG
8	204890_s_at	0.75	LCK	33	206337_at	0.60	CCR7
9	210031_at	0.71	CD3Z	34	209671_x_at	0.60	TRA@
10	206545_at	0.70	CD28	35	206485_at	0.60	CD5
11	212400_at	0.70	C9orf132	36	214470_at	0.58	KLRB1
12	205254_x_at	0.69	TCF7	37	202524_s_at	0.58	SPOCK2
13	205831_at	0.69	CD2	38	213539_at	0.58	CD3D
14	210439_at	0.69	ICOS	39	210948_s_at	0.58	LEF1
15	205798_at	0.68	IL7R	40	214032_at	0.58	ZAP70
16	211893_x_at	0.67	CD6	41	221558_s_at	0.58	LEF1
17	205255_x_at	0.67	TCF7	42	211902_x_at	0.58	TRA@
18	213958_at	0.66	CD6	43	220485_s_at	0.57	SIRPB2
19	211339_s_at	0.66	ITK	44	206966_s_at	0.57	KLF12
20	208602_x_at	0.66	CD6	45	209670_at	0.56	TRAC
21	210479_s_at	0.64	RORA	46	209604_s_at	0.56	GATA3
22	205590_at	0.64	RASGRP1	47	207651_at	0.56	GPR171
23	211900_x_at	0.64	CD6	48	213906_at	0.56	MYBL1
24	204777_s_at	0.64	MAL	49	205259_at	0.56	NR3C2
25	220418_at	0.64	UBASH3A	50	204638_at	0.55	ACP5

DISCUSSION

Simultaneous analyses of genome-wide genotyping and copy number data with gene expression profiling enables the identification of pathogenic genes that as a result of genomic imbalances are aberrantly expressed. By an integrated approach involving genome-wide genotyping and GEP, we identified *BCL11B* as a novel oncogene in AML. Interstitial amplification of 14q32 was initially revealed in two AML cases (#2301 and #7073) by genome-wide genotyping and SNP array, showing relatively small amplifications including *BCL11B,* subsequently confirmed by FISH. Although different genes were affected by the numerical abnormalities, these 14q23 aberrations resulted in unique *BCL11B* mRNA and full-length BCL11B protein overexpression. By FISH in a selection of 40 AML cases with high *BCL11B* mRNA expression, we identified two additional AML cases baring *BCL11B* translocations (AML #6451 and AML #6366).

The BCL11 family has two members BCL11A and BCL11B. Bcl11a was identified as a common retroviral insertion site (Evi9) in murine myeloid leukemias [32-33] and is required for normal B-cell development [34]. BCL11B is encoded by a 4 exon gene located on 14q32, encoding a Kruppel family zinc finger transcription factor. Bcl11b was initially identified as a tumor suppressor gene in T cells [35]. BCL11B is a key regulator of differentiation and survival of thymocytes [25]. *BCL11b* may function as a tumor suppressor gene in T-cell neoplasia, i.e., mice carrying biallelic inactivation of *Bcl11b* developed thymic lymphomas, indicating that loss-of-function mutations in *Bcl11b* contribute to mouse lymohomagenesis and possibly to human cancer development [36].

BCL11B was first associated with hematological malignancies due to its recurrent involvement with the homeobox transcription factor *TLX3* in a relatively high percentage of pediatric and adult T-cell acute lymphoblastic leukemia (T-ALL), carrying the cryptic t(5;14)(q35;q32) [37-39]. Less frequent, T-ALL samples with an inv(14)(q11.2q32.31) carry an in-frame transcript of *BCL11B* and the T-cell receptor gene segment *TRDV1.* These ALL cases do not express wild type *BCL11B* transcripts, suggesting that *BCL11B* disruption may contribute to T-cell malignancies in humans [40]

A number of myeloid, mixed-lineage, and nonlymphocytic leukemias with 14q32 abnormalities have been reported, however, in these instances the affected genes were not identified [41-45]. The first evidence of *BCL11B* involvement in 14q32 translocations in AML was reported by Bezrookove et al. [42]. They reported one case of t(6;14)(q25~q26;q32) in an adult with AML and used bacterial artificial chromosomes to demonstrate the involvement of *BCL11B* in this AML case [42]. Due to lack of patient material, the investigators could neither establish the deregulation of *BCL11B* nor the identification of the partner genomic locus [42]. Of note, the breakpoint in this AML case appeared to be located upstream of the *BCL11B* gene. This is similar to the AML cases described here and suggests that the breakpoints in AML are clustered upstream, whereas in ALL they are downstream of *BCL11B*.

Specific chromosomal paints demonstrated that different partner chromosomes were involved in the AML cases with a *BCL11B* translocation. This may indicate that different regulatory regions on the partner chromosome are capable to activate the *BCL11B* oncogene in AML. Interestingly, BCL11B protein appeared to be expressed in additional primary AML cases that did not carry *BCL11B* translocations. In these AML cases other mutations may be present or *BCL11B* may be activated by other means. The identification of the exact chromosomal regions involved in the translocations and the transcriptional elements responsible for *BCL11B* overexpression, in AML with and without *BCL11B* rearrangements, needs further investigations.

The 14q32 region, including *BCL11B*, has been subject for translocation in T-ALL and acute mixed lineage leukemia [29,41-45]. In fact, the involvement of 14q32 translocations and BCL11B in AML has been debated [46]. However, the immunophenotyping and molecular analyses of the AML samples with *BCL11B* translocations described here show that these leukemias have biphenotypic immunophenotype, but also all carry a common AML-associated *FLT3* ITD mutation. Thus these leukemias share a characteristic genetic feature with AML.

BCL11B is expressed in T-lymphocytes and T-cell leukemias and is a pivotal regulator of number of genes related to T cell proliferation and differentiation such as *IL2*, *NF-kappaB* and *p21* [25-26,47-49]. It was shown recently that the expression of *BCL11B* in T-cell lines resulted in markedly increased apoptosis resistance following treatment with radiomimetic drugs accompanied by a cell cycle delay caused by accumulation of cells at G1 [50]. We examined the consequences of *Bcl11b* overexpression on proliferation and differentiation in a mouse myeloid 32D(GCSF-R) cell line model. 32D(GCSF-R) cells, expressing full length murine Bcl11b cDNA, showed a consistent decreased proliferation rate compared to cells expressing the empty vector or to the parental untransfected cells. Upon stimulation with GCSF, 32D(GCSF-R) cells overexpressing Bcl11b showed less maturation towards granulocytes compared to cells expressing empty vector, giving supporting evidence that Bcl11b is partially blocking or delaying differentiation in 32D(GCSF-R) cells.

In conclusion, we show that *BCL11B* is involved in 14q32 translocations with different putative chromosomal partners in well-characterized AML cases using high-throughput genome-wide genotyping, cytogenetics and GEP. In these translocations, the full length BCL11B transcript is highly expressed and associated with the expression of T-cell markers. We speculate that due to the translocations, *BCL11B* expression is influenced by active transcriptional elements resulting in the high expression of *BCL11B* and consequently T-cell associated genes. Murine cell line overexpressing BCL11B show decreased proliferation and partial delayed differentiation which provides evidence that *BCL11B* may have suppressive and disruptive effects on cell proliferation and differentiation.

ACKNOWLEDGEMENTS

This work was supported by grants from the Dutch Cancer Society (Koningin Wilhelmina Fonds) and performed within the framework of CTMM, the Center for Translational Molecular Medicine (Leukemia BioCHIP project (grant 030-102)). We are indebted to the participants of the HOVON clinical trails, our colleagues from the stem cell transplantation and molecular diagnostics laboratories who provided, collected and analyzed AML cell samples, respectively.

AUTHORSHIP

SA: Performed research, analyzed data and wrote manuscript; AZ: Performed research; WMCG: Performed research; ZGA: Performed research; SM: Performed research; IWTC: Performed research; MS: Analyzed data; JKP: Analyzed data; HBB: Performed research and wrote manuscript; BL: Designed research and wrote manuscript; PJMV: Designed and performed research, analyzed data and wrote manuscript.

REFERENCES

1 Burnett, A., Wetzler, M. & Lowenberg, B. Therapeutic advances in acute myeloid leukemia. *J Clin Oncol* **29**, 487-494, (2011).

2 Marcucci, G., Haferlach, T. & Dohner, H. Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications. *J Clin Oncol* **29**, 475-486, (2011).

3 Kuo, Y. H. *et al.* Runx2 induces acute myeloid leukemia in cooperation with Cbfbeta-SMMHC in mice. *Blood* **113**, 3323-3332, (2009).

4 Wang, Y. Y. *et al.* C-KIT mutation cooperates with full-length AML1-ETO to induce acute myeloid leukemia in mice. *Proc Natl Acad Sci U S A* **108**, 2450-2455, (2011).

5 Frank, R. C., Sun, X., Berguido, F. J., Jakubowiak, A. & Nimer, S. D. The t(8;21) fusion protein, AML1/ETO, transforms NIH3T3 cells and activates AP-1. *Oncogene* **18**, 1701-1710, (1999).

6 Nakao, M. *et al.* Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia* **10**, 1911-1918 (1996).

7 Dohner, K. *et al.* Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. *Blood* **106**, 3740-3746, (2005).

8 Pabst, T. *et al.* Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. *Nat Genet* **27**, 263-270, (2001).

9 Valk, P. J. *et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* **350**, 1617-1628, (2004).

10 Debernardi, S. *et al.* Genome-wide analysis of acute myeloid leukemia with normal karyotype reveals a unique pattern of homeobox gene expression distinct from those with translocation-mediated fusion events. *Genes Chromosomes Cancer* **37**, 149-158, (2003).

11 Huhta, T. *et al.* Comparative genomic hybridization and conventional cytogenetic analyses in childhood acute myeloid leukemia. *Leuk Lymphoma* **35**, 311-315 (1999).

12 Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66-72, (2008).

13 Abdel-Wahab, O. *et al.* Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114**, 144-147, (2009).

14 Delhommeau, F. *et al.* Mutation in TET2 in myeloid cancers. *N Engl J Med* **360**, 2289-2301, (2009).

15 Gelsi-Boyer, V. *et al.* Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br J Haematol* **145**, 788-800, (2009).

16 Carbuccia, N. *et al.* Mutual exclusion of ASXL1 and NPM1 mutations in a series of acute myeloid leukemias. *Leukemia* **24**, 469-473, (2010).

17 Eckerle, S. *et al.* Gene expression profiling of isolated tumour cells from anaplastic large cell lymphomas: insights into its cellular origin, pathogenesis and relation to Hodgkin lymphoma. *Leukemia* **23**, 2129-2138, (2009).

18 Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* **361**, 1058-1066, (2009).

19 Abbas, S. *et al.* Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value. *Blood* **116**, 2122-2126, (2010).

20 Yang, J. J. *et al.* Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. *JAMA* **301**, 393-403, (2009).

21 Alvarez, S., MacGrogan, D., Calasanz, M. J., Nimer, S. D. & Jhanwar, S. C. Frequent gain of chromosome 19 in megakaryoblastic leukemias detected by comparative genomic hybridization. *Genes Chromosomes Cancer* **32**, 285-293, (2001).

22 Raghavan, M. *et al.* Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. *Cancer Res* **65**, 375-378, (2005).

23 Gorletta, T. A. *et al.* Frequent loss of heterozygosity without loss of genetic material in acute myeloid leukemia with a normal karyotype. *Genes Chromosomes Cancer* **44**, 334-337, (2005).

24 Sanders, M. A. *et al.* SNPExpress: integrated visualization of genome-wide genotypes, copy numbers and gene expression levels. *BMC Genomics* **9**, 41 (2008).

25 Wakabayashi, Y. *et al.* Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. *Nat Immunol* **4**, 533-539, (2003).

26 Cismasiu, V. B. *et al.* BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter. *Oncogene* **24**, 6753-6764, (2005).

27 Lin, M. *et al.* dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**, 1233-1240 (2004).

28 Dong, F. *et al.* Distinct cytoplasmic regions of the human granulocyte colony-stimulating factor receptor involved in induction of proliferation and maturation. *Mol Cell Biol* **13**, 7774-7781 (1993).

29 Wouters, B. J. *et al.* Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood* **110**, 3706-3714 (2007).

30 Verhaak, R. G. *et al.* Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* **94**, 131-134, (2009).

31 Akagi, K., Suzuki, T., Stephens, R. M., Jenkins, N. A. & Copeland, N. G. RTCGD: retroviral tagged cancer gene database. *Nucleic Acids Res* **32**, D523-527, (2004).

32 Nakamura, T. *et al.* Evi9 encodes a novel zinc finger protein that physically interacts with BCL6, a known human B-cell proto-oncogene product. *Mol Cell Biol* **20**, 3178-3186 (2000).

33 Liu, P. *et al.* Bcl11a is essential for normal lymphoid development. *Nat Immunol* **4**, 525-532, (2003).

34 Liu, P., Li, P. & Burke, S. Critical roles of Bcl11b in T-cell development and maintenance of T-cell identity. *Immunol Rev* **238**, 138-149, (2010).

35 Wakabayashi, Y. *et al.* Homozygous deletions and point mutations of the Rit1/Bcl11b gene in gamma-ray induced mouse thymic lymphomas. *Biochem Biophys Res Commun* **301**, 598-603, (2003).

36 Bernard, O. A. *et al.* A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. *Leukemia* **15**, 1495-1504 (2001).

37 Berger, R. *et al.* t(5;14)/HOX11L2-positive T-cell acute lymphoblastic leukemia. A collaborative study of the Groupe Francais de Cytogenetique Hematologique (GFCH). *Leukemia* **17**, 1851-1857, (2003).

38 Cave, H. *et al.* Clinical significance of HOX11L2 expression linked to t(5;14)(q35;q32), of HOX11 expression, and of SIL-TAL fusion in childhood T-cell malignancies: results of EORTC studies 58881 and 58951. *Blood* **103**, 442-450, (2004).

39 Cave, H. *et al.* Clinical significance of HOX11L2 expression linked to t(5;14)(q35;q32), of HOX11 expression, and of SIL-TAL fusion in childhood T-cell malignancies: results of EORTC studies 58881 and 58951. *Blood* **103**, 442-450, (2004).

40 Przybylski, G. K. *et al.* Disruption of the BCL11B gene through inv(14)(q11.2q32.31) results in the expression of BCL11B-TRDC fusion transcripts and is associated with the absence of wild-type BCL11B transcripts in T-ALL. *Leukemia* **19**, 201-208, (2005).

41 Batanian, J. R., Dunphy, C. H., Gale, G. & Havlioglu, N. Is t(6;14) a non-random translocation in childhood acute mixed lineage leukemia? *Cancer Genet Cytogenet* **90**, 29-32, (1996).

42 Bezrookove, V. *et al.* A novel t(6;14)(q25-q27;q32) in acute myelocytic leukemia involves the BCL11B gene. *Cancer Genet Cytogenet* **149**, 72-76 (2004).

43 Georgy, M., Yonescu, R., Griffin, C. A. & Batista, D. A. Acute mixed lineage leukemia and a t(6;14)(q25;q32) in two adults. *Cancer Genet Cytogenet* **185**, 28-31, (2008).

44 Hayashi, Y. *et al.* 14q32 translocations are associated with mixed-lineage expression in childhood acute leukemia. *Blood* **76**, 150-156 (1990).

45 Raimondi, S. C. *et al.* Cytogenetics of childhood acute nonlymphocytic leukemia. *Cancer Genet Cytogenet* **40**, 13-27, (1989).

46 MacLeod, R. A., Nagel, S. & Drexler, H. G. BCL11B rearrangements probably target T-cell neoplasia rather than acute myelocytic leukemia. *Cancer Genet Cytogenet* **153**, 88-89, (2004).

47 Cherrier, T. *et al.* p21(WAF1) gene promoter is epigenetically silenced by CTIP2 and SUV39H1. *Oncogene* **28**, 3380-3389, (2009).

48 Cismasiu, V. B. *et al.* BCL11B enhances TCR/CD28-triggered NF-kappaB activation through up-regulation of Cot kinase gene expression in T-lymphocytes. *Biochem J* **417**, 457-466, (2009).

49 Cismasiu, V. B. *et al.* BCL11B participates in the activation of IL2 gene expression in CD4+ T lymphocytes. *Blood* **108**, 2695-2702, (2006).

50 Grabarczyk, P. *et al.* Increased expression of bcl11b leads to chemoresistance accompanied by G1 accumulation. *PLoS One* **5**, (2010).

CHAPTER 4

Exon8 splice site mutations in the gene encoding the E3-ligase CBL are associated with core binding factor acute myeloid leukemias (AML)

Saman Abbas[1], Gabriëlle Rotmans[1], Bob Löwenberg[1], Peter J.M. Valk[1]

[1] Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands

Haematologica. 2008 Oct; 93(10):1595-7.

Acute myeloid leukemia (AML) is a heterogeneous disease with diverse genetic abnormalities and variable response to treatment. In the last decade the diverse genetic abnormalities have refined risk-stratification of AML [1]. Recently, mutations in the Casitas B-cell lymphoma gene encoding the E3-ligase CBL [2] were identified in *de novo* AML [3-4]. In one study, a single case with an inactivating point mutation in exon9 of the *CBL* gene was identified in a cohort of 150 *de novo* AML cases [3]. In a second study, exon8 missense mutations were demonstrated in 3 out of 12 randomly selected AML cases [4]. In an additional AML case, a DNA insertion/deletion mutation in intron7 of in the *CBL* gene resulted in the expression of a *CBL* splice variant, i.e., a *CBL* mRNA lacking exon8 [4].

All published *CBL* mutations are located within the conserved linker region (LR) and ring finger (RF) of the CBL protein [2]. In fact, the mutant *CBL* splice variant without exon8 results in an in-frame deletion, which encodes a CBL protein lacking part of the LR, including two essential tyrosine residues, and almost the entire RF, which is critical for E3 activity [2]. This suggests that mutant CBL may act as a dominant negative protein by inhibiting proper down-regulation of critical activated tyrosine kinases, such as KIT and FLT3 in AML [5].

It is still not clear how frequently mutations in the *CBL* gene occur in newly diagnosed AML. In a diverse population of primary AML (n=319, Table 1) we assessed the frequency of *CBL* mutations, i.e., point mutations in exon8 [4] and exon9 [3], and mutations affecting proper splicing of *CBL* exon8 [4].

Patients had a diagnosis of primary AML, confirmed by cytological examination of blood and bone marrow and were treated according to the HOVON protocols (http://www.hovon.nl). After obtaining patients' informed consent, bone marrow aspirates or peripheral blood samples were taken at diagnosis (n=319). CBL exon8 mRNA splice variants, as well as point mutations in exon8, were determined by cDNA amplification using the primer set CBLe6F 5'-AAACCTCTCTTCCAAGCACTG-3' and CBLe9R 5'-TCCCTCTAGGATCAAACGGA-3' or CBL-exon7-FOR 5'-GTGAACCAACTCCCCAAGAC-3' and CBL-exon9-REV 5'-GGACAGCCCTGACCTTCTG-3'. Mutations in genomic DNA were determined by amplification using CBL-intron7-FOR 5'-GGACCCAGACTAGATGCTTTC-3' and CBL-exon8-REV 5'-GTGCACATGAGGTGTCCACAG-3' (mutations 5' of exon8) or CBL-exon7-FOR and CBL-exon9-REV (mutations 3' of exon8). (0.25mM dNTP, 15 pmol primers, 2mM $MgCl_2$, Taq polymerase and 1xbuffer (Invitrogen Life Technologies, Breda, The Netherlands); 1 cycle at 94˚C for 5 mins., 35 cycles at 94˚C for 1 min., 1' 60˚C for 1 min., 72˚C for 1 min., and 1 cycle at 72˚C for 7 mins.). Samples showing aberrant patterns were sequenced by using forward and reverse primers on the ABI PRISM3100 genetic analyzer (Applied Biosytems, Foster City, CA).

Table 1: Clinical and molecular characteristics of the 319 patients* with *de novo* acute myeloid leukemia.

	#	Percent
Gender		
Male	154	48
Female	165	52
FAB		
M0	11	3
M1	62	19
M2	69	22
M3	19	6
M4	52	16
M5	79	25
M6	6	2
not determined	21	7
Cytogenetic abnormalities**		
t(15;17)	19	6
t(8;21)	21	7
inv(16)/t(16;16)	22	7
+8	23	7
-5/-5(q)	13	4
-7/-7(q)	22	7
3q	10	3
t(6;9)	3	1
t(9;22)	3	1
t(11q23)	14	4
complex (>3 abn.)	18	6
other	77	24
normal	127	40
not determined	14	4
Molecular abnormalities		
FLT3 ITD	90	28
FLT3 TKD	33	10
N-*RAS*	26	8
K-*RAS*	0	0
CEBPA	17	5
KIT (exon8 and D816 mutations)	14	4

*Median age 45 (range 15 - 75), median bone marrow blast count 65 percent (range 0 (for acute promyelocytic leukemia) - 99), median white blood cell (WBC) count 32 (x10^9/l) (range 0.3 - 349) and median platelet count 51 (x10^9/l) (range 3-998).

**All patients with a specific abnormality were considered, irrespective of the presence of additional abnormalities.

All AML cases were screened by RT-PCR for *CBL* exon8 splice variants, whereas a randomly selected subset of 183 out of the 319 AML cases was examined by sequence analysis for the presence of exon8 [4] or exon9 [3] point mutations. We did not find any point mutation in exon8 or exon9 of *CBL* in the subset of 183 AML cases. However, out of the 319 AML cases we did identify two AML cases expressing a *CBL* mRNA splice variant (#2274 and #6717, Figure 1A), which lacked exon8. This aberrant *CBL* mRNA is similar to the splice variant previously identified in the leukemic cell line MOLM13 (Figure 1A) [4]. *CBL* transcripts lacking exon8 were not present in 5 normal bone marrow samples and 3 fluorescence activated cell sorted CD34-positive progenitor cell samples (data not shown).

Interestingly, the two primary AML cases showing aberrant splicing carried an inversion of chromosome 16 (inv(16)) suggesting that *CBL* mutations might be associated with core-binding factor (CBF) leukemias, i.e., AML and inv(16) or t(8;21). In a selected screen of 39 inv(16) AML and 40 t(8;21) AML we did not detect point mutations in *CBL* exon8 or exon9. However, the *CBL* exon8 splice variant was present in 2 additional cases with a t(8;21) and one additional case with an inv(16) (Figure 1A). The *CBL* exon8 splice variant was absent in 4 independent remission samples of AML patient #2549 (data not shown).

By nucleotide sequencing of the flanking sequences of *CBL* exon8 in all AML cases expressing the *CBL* mRNA splice variant, we identified various insertion/deletion mutations in the *CBL* gene (Figure 1B). All insertion/deletion mutations affect the splice acceptor or donor sites of exon8 of *CBL*. In fact, the G to C point mutation in case #7056 is located within the splice acceptor site of *CBL* exon8, resulting in two additional aberrant in-frame *CBL* transcripts (Figure 1A).

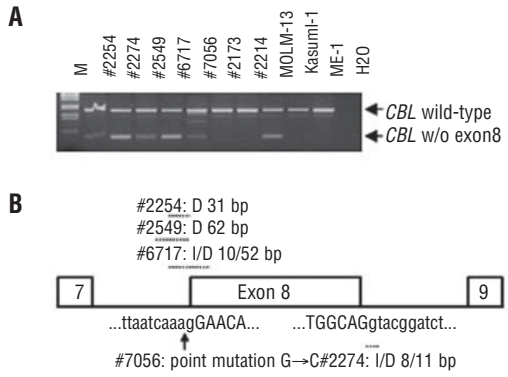


Figure 1.
CBL mRNA splice variant mutations in core-binding factor AML. **(A)** RT-PCR with primers CBL-exon7-FOR and CBL-exon9-REV of primary AML cases showing aberrant splice variants (#2254 (inv(16)), #2549 (t(8;21)), #2274 (inv(16)), #6717 (inv(16)), #7056 (t(8;21)), wild-type *CBL* (#2173 and #2214) and cell lines MOLM13 (*FLT3* internal tandem duplication (ITD)), Kasumi-1 (t(8;21)) and ME-1 (inv(16)). The presence of these *CBL* exon8 splice mutants in the primary AML were confirmed using various primer sets (*data not shown*). AML #7056 consistently expresses two additional CBL mRNA splice variants. Of note, both these *CBL* splice variants are in-frame ((exon7) ACC CAG ATG GGC TCC (exon8) and (exon7) ACC CAG GAT GTA AAG (exon8)) **(B)** Schematic representation of exon 7-to-9 of the *CBL* gene to indicate the insertion (I)/deletion (D) mutations in the *CBL* gene of the 5 CBF AML cases (#2254, #2549, #2274, #6717, #7056).

These results indicate that there is a preferential association between *CBL* exon8 splice variant mutations and CBF leukemias. This raises the possibility of a functional association between impaired CBL function and the CBF-related fusion proteins CBFB-MYH11 and AML1-ETO. Activating point mutations in receptor tyrosine kinase KIT are strongly associated with CBF leukemias [6-8]. In fact, we recently screened 500 cases of AML for exon8/exon17 *KIT* mutations and mutations were present in 25 AML cases (5%). Of the KIT mutant cases 88% carried a proven CBF mutation, such as t(8;21), inv(16) or t(3;21) (*data not shown*). CBL proteins mediate ubiquitination and degradation of KIT upon stimulation with stem cell factor [9]. Thus, the expressed mutant CBL protein, which is still able to bind KIT by its N-terminal tyrosine kinase binding domain, but impaired in downstream ubiquitination [2], may act as a dominant negative protein. Expression of this dominant negative protein could result in impaired routing of KIT and sustained activation, similar to KIT activating mutations in CBF leukemias.

No other AML-specific mutations, such as those affecting *FLT3*, *NRAS*, *KRAS*, *CEBPA* and *NPM1*, were present in the *CBL* mutant AML cases. Interestingly, however, CBL mutant t(8;21) AML case #2549 also carried a KIT D816 mutation. In this AML impaired function of the CBL protein would potentially result in prolonged constitutive activation of KIT.

Our results demonstrate that *CBL* mutations are rare in AML. However, the strong association of these mutations with CBF leukemias suggests that there may be a cooperative activity of mutant CBL with the CBF-related fusion proteins CBFB-MYH11 and AML1-ETO in CBF leukemogenesis, most likely by impaired ubiquitination of KIT.

REFERENCES

1 Mrozek, K., Marcucci, G., Paschka, P., Whitman, S. P. & Bloomfield, C. D. 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 109, 431-448 (2007).

2 Ryan, P. E., Davies, G. C., Nau, M. M. & Lipkowitz, S. Regulating the regulator: negative regulation of Cbl ubiquitin ligases. Trends Biochem Sci 31, 79-88 (2006).

3 Sargin, B. et al. Flt3-dependent transformation by inactivating c-Cbl mutations in AML. Blood 110, 1004-1012 (2007).

4 Caligiuri, M. A. et al. Novel c-CBL and CBL-b ubiquitin ligase mutations in human acute myeloid leukemia. Blood 110, 1022-1024 (2007).

5 Renneville, A. et al. Cooperating gene mutations in acute myeloid leukemia: a review of the literature. Leukemia 22, 915-931 (2008).

6 Valk, P. J. et al. Second hit mutations in the RTK/RAS signaling pathway in acute myeloid leukemia with inv(16). Haematologica 89, 106 (2004).

7 Care, R. S. et al. Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. Br J Haematol 121, 775-777 (2003).

8 Beghini, A. et al. C-kit mutations in core binding factor leukemias. Blood 95, 726-727 (2000).

9 Zeng, S., Xu, Z., Lipkowitz, S. & Longley, J. B. Regulation of stem cell factor receptor signaling by Cbl family proteins (Cbl-b/c-Cbl). Blood 105, 226-232 (2005).

CHAPTER 5

Mutant Wilms' tumor 1 *(WT1)* mRNA with premature termination codons in acute myeloid leukemia (AML) is sensitive to nonsense-mediated RNA decay (NMD)

Saman Abbas[1], Claudia A.J. Erpelinck-Verschueren[1], Chantal Goudswaard[1], Bob Löwenberg[1] and Peter J.M. Valk[1]

[1] Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands

Leukemia **24**, 660-663, (2010)

The Wilms' tumor 1 (*WT1*) gene on chromosome 11p13 encodes a Krupple-like zinc finger transcription factor that may act as tumor suppressor or oncogene in leukemogenesis [1]. Mutations in *WT1* were initially identified in 10-15% of AML and 20% of biphenotypic leukemia [2]. Recently, the frequency of 10-15% *WT1* mutations in adult AML with normal karyotypes was confirmed using relatively large cohorts of AML patients [3-6]. In some studies *WT1* mutations were implicated as an independent adverse prognostic marker for patients with normal karyotype AML (<60 years) [2-5], however, in other *WT1* mutations did not have impact on outcome [6]. Mutations in *WT1* in AML often result in frame shifts introducing premature termination codons (PTC) [2-5], but it is not known whether *WT1* mutant transcripts are degraded by the nonsense-mediated RNA decay (NMD) surveillance system [7].

Here, we analyzed a cohort of 218 AML cases for *WT1* mutations by RT-PCR for mutations in exons 6 to 9 (WT1-FOR3 5'-GAGAGCGATAACCACACAAC-3' (exon6) and WT1-REV4 5'-CTGTATGAGTCCTGGTGTGGG-3' (exon9) (bp1214-1619 (NM_024426)) followed by denaturing high performance liquid chromatography (dHPLC, sensitivity 10-20% (mutation dosage)) and sequencing. In six AML cases missense and in-frame insertion mutations were detectable in *WT1* (no. 1-6, Table1). These *WT1* missense and in-frame mutants are expressed, because these transcripts were readily detectable by RT-PCR. Analyses of a diagnostic and relapse sample of AML case no. 2270 (no. 7, Table 1), carrying a double mutation, showed that the out-of-frame *WT1* mutations were not present on a single allele and both present in the relapse. Thus, the insertions of 11bp and 1bp do not result in the expression of a WT1 protein with an internal frame-shift, but encode truncated WT1 proteins from two alleles.

Following this RT-PCR strategy; we detected relatively small numbers of *WT1* mutations (2.7%) [2-5]. In contrast, out of a diverse AML cohort of 351 AML samples, we were able to reliably detect 21 insertion/deletion mutations in exon 7 or 9 of *WT1* (no. 8-26, Table1) using genomic DNA PCR [3,5] followed by dHPLC (sensitivity 10-20% (mutation dosage)) and sequencing. The overall frequency of *WT1* mutations in AML patients with a normal karyotype was 6.3 and 7.7% in AML patients with normal karyotypes younger than 60 years. *WT1* mutations were present in patients carrying a favorable karyotype, that is, t(8;21) (n=1), inv(16) (n=2) and t(15;17) (n=2) (Table1). This may suggest that mutations in *WT1* cooperate with the fusion proteins AML1/ETO, CBFB/MYH11 and PML/RAR, respectively. *WT1* mutations were more often present in AML with a *FLT3* internal tandem duplication or *NPM1* mutation (data not shown); however, none of these were statistically significant (Fisher's exact test). No association was found between *WT1* mutations and other known acquired mutations in AML, that is, N-*RAS*-, K-*RAS*- or *CEBPA* mutations. By subcloning the *WT1* PCR fragments of the five additional AML double mutants (Table 1), we have shown that these were monoallelic. The overall *WT1* mutation dosage was below 55%, thus, there is no direct indication that acquired uniparental disomy is involved in any of the AML cases. It is interesting to note that *WT1* mutations were present in two out of six cases of a new subgroup of biphenotypic leukaemias, showing hypermethylation and silencing of *CEBPA* as well as *NOTCH1* mutations that were recently identified by gene expression profiling of AML [8].

Table 1 ***WT1*** **mutations in 26 primary AML cases**

No.	*Case*	*RT-PCR*	*PCR*	*Exon*	*Mutation*
1	2226	D	NA	6	T→C (nt 1244)
2	2682	D	NA	8	G→T (nt 1497)
3	2243	D	NA	8	CTCTTT (nt 1498)
4	2497	D	NA	9	G→T (nt 1581)
5	2263	D	NA	9	C→T (nt 1589)
6	2773	D	D	9	GGG (ins) (nt 1567)
7	2270[a]	D	D	7	CTCTTGTACGG (ins) (nt 1336)/
					T (ins) (nt 1393)
8	2299	ND	D	9	GT (ins) (nt 1538)
9	2247	ND	D	7	TC (ins) (nt 1334)
10	2668	ND	D	7	AAGGGAAATT (ins) (nt 1301)
11	2647	ND	D	7	CGGC (ins) (nt 1341)
12	2545	ND	D	7	ACCG (ins) (nt 1334)
13	2170	ND	D	7	ACTCTTG (ins) (nt 1335)
14	2177	ND	D	7	TCGG (ins) (nt 1341)
15	2280[a]	ND	D	7	AGA (ins)/ C (del) (nt 1301)
16	7125	ND	D	7	GACCTGTGAG (ins) (nt 1341)
17	7169	ND	D	7	CGACGTG (del) (nt 1301)
18	4340	ND	D	7	GTCTTGTACGG (ins) (nt 1337)
19	7145	ND	D	7	G (del) (nt 1305)
20	2305[a]	ND	D	7	CC (ins)/ G (del) (nt 1305)
21	3314[a]	ND	D	7	CCCCATGCC (ins)/ G (del) (nt 1305)
22	2275	ND	D	7	CTCTTGT (ins) (nt 1326)
23	2186[a]	ND	D	7	CAGTAGG (ins)/ del TTGTAC (del) (nt 1539)
24	1551	ND	D	7	GA (ins) (nt 1304)
25	5365	D	D	9	CG (del) (nt 1567)
26	5284[a]	D	D	7	A (ins) TC (del) (nt 1326)

Abbreviations: D, mutation detectable; ND, mutation not detectable; NA, material not available; ins, insertion; del, deletion; PTC, premature termination codon; NN, normal karyotype. Mutations are indicated according to Acc. No. NM_024426 WT1 mRNA, [a]Double *WT1* mutant.

Table 1 ***WT1*** **mutations in 26 primary AML cases**

Abnormality	*Protein*	*Karyotype*
missense	Cys350Arg	47,XY,+11[43]/46,XY[5]
missense	Arg434Leu	46,XX,t(2;9;11)(p13;p22;q23)[25]
6 bp ins	434 ins LeuPhe	45,X,-Y,t(8;21)(q22;q22)[19]
missense	Arg462Leu	46,XY,del(1)(p33orp32p36)[10]/46,XY[2]
missense	His465Tyr	46,XX,t(15;17)(q22;q21)[30]/46,XX[2]
3 bp ins	457 ins Gly	46,XY[37]
11bp ins	nt 1539PTC	46,XY,i(8)(q10)[4]/46,XY,i(8)(q10),+ace[2]/
1bp ins	nt 1394PTC	47,XY,+i(8)(q10)[3]/47,XY,+8[3]/46,XY[24]
2 bp ins	nt 1539PTC	47,XY,+21[5]/46,XY [65]
2 bp ins	nt 1539PTC	46,XX,der(2)inv(2)(p2?3p1?5)
		del(2) (p11p1?5)[17]
10 bp ins	nt 1345PTC	47-49,XX,+del(8)(q24?),i(17)(q10),+2mar[cp10]/
		46,XX[2]
4 bp ins	nt 1345PTC	46-47,XY,del(16)(q22)[3],+22[5][cp6)/46,XY[8]
4 bp ins	nt 1345PTC	NN
7 bp ins	nt 1345PTC	46,XY,del(12)(q11,q21),t(15,17)(q22;q12)[25]
4 bp ins	nt 1345PTC	NN
3 bp ins/ 2 bp del	nt 1345PTC	46,XX[37]
10 bp ins	nt 1345PTC	46,XX[21]
7 bp del	nt 1318PTC	failure
11 bp ins	nt 1539PTC	46,XY[20]
1 bp del	nt 1318PTC	failure
2 bp ins/ 1 bp del	nt 1345PTC	46,XY[42]
9 bp ins/ 1 bp del	nt 1318PTC	53,XY,+der(1)add(1)(q?12)del(1)(p22p31),
		der(3)t(3;8)(p26;q21),+8,+11,+13,+13,+14[16]/ 46,XY[5]
7 bp ins	nt 1345PTC	46,XX[29]
7 bp ins/ 6 bp del	nt 1345PTC	45,XY,-7,t(7;8)(q21;p11)[25]
2 bp ins	nt 1318PTC	46,XX[33]
2 bp del	nt 1621PTC	46,X,der(X)(Xpter->Xq1?2::16q22--16qter),
		der(16)(Xqter->Xq1?2::16p13->16q22::16p13-16pter)[17]/
		46XX[2]
1 bp ins/ 2 bp del	nt 1539PTC	46,XY [21]

WT1 mutations have recently been implicated as poor prognostic factor in AML 3-5. We investigated whether *WT* mutations carry prognostic value in our patient cohort. Univariate survival analyses of AML patients younger than 60 years by the method of Kaplan and Meier did not reveal any significant difference in overall (OS) and event-free survival (EFS) between patients with or without *WT1* mutations (data not shown).

These survival analyses were performed on all AML patients, except those with favourable cytogenetics (n = 233) (OS P= 0.09 and EFS P = 0.08) or AML patients with normal karyotypes (n = 134) (OS P = 0.35 and EFS P = 0.25).

All AML cases with genomic *WT1* mutations were examined using RT-PCR. In 17 out of 21 cases *WT1* mutations were detectable by PCR/dHPLC but undetectable by RT-PCR/dHPLC (no. 8-24, Table1). The majority of these out-of-frame mutations were present in exon 7 of *WT1*. Thus, most AML cases do not express PTC-containing *WT1* transcripts or at very low levels. To show the involvement of NMD, we selected three primary AML cases (no. 2275, no. 2305 and no. 3314 (no. 20-22, Table 1)), which carried a *WT1* insertion mutation as was shown by PCR on genomic DNA (*WT1* mutation dosage 50% in all three AML cases based on PCR analysis), but not cDNA (Table1). We cultured these AMLs in the presence and absence of emetine (6 hours, 100 μg/ml). Emetine is known to block translation-mediated NMD [7]. The *WT1* mutant mRNA transcripts are not detectable or present at low levels in AML cells cultured in the absence of emetine (Figure 1A). Interestingly, however, the translation inhibitor prominently stabilized the *WT1* mutant mRNA transcripts. After treatment with emetine the mutant WT1 transcripts are present at similar levels or even higher as the wild-type *WT1* mRNA transcripts (Figure 1A). By RQ-PCR with *WT1* wild-type- and mutant-specific primers we showed a fourfold increase of mutant *WT1* transcripts after emetine treatment relative to wild-type in case no. 2305. In fact, in case no. 3314 *WT1* mutant transcripts were extremely low to undetectable before emetine treatment and highly abundant (>100000-fold increase) after treatment. The specificity of the mutant-specific primers was confirmed using AML cases without *WT1* mutations. These results strongly indicate that *WT1* transcripts with PTCs are sensitive to translation-mediated NMD. Importantly, in a diagnostic setting *WT1* mutations may therefore be missed when mutation detection is based on RT-PCR.

PTCs are introduced as a result of insertion/deletion mutations in 77% of the AML cases. The three PTCs that are introduced in the majority of mutant *WT1* transcripts (nt 1318PTC, nt 1345PTC and nt 1539PTC) are located over 55 bp upstream of the next intron boundary and are therefore prone to NMD [7]. However, PTC1621 that is present in AML patient no. 5365 should in theory not be affected by NMD because this PTC is located only 5 bp from the next intron boundary. In fact, this mutant transcript was indeed detectable by RT-PCR dHPLC. This case is in line with the involvement of NMD in the majority of *WT1* mutant AML patients. In contrast, however, *WT1* mRNA transcripts in AML cases no. 2270 and no. 5284, which contain nt 1394/1539 PTCs and nt 1539 PTC, respectively, were expressed. In these cases NMD may reduce the levels of PTC-containing mRNA suboptimal and truncated WT1 proteins may be expressed. In fact, it has been shown that PTC-containing mRNA is reduced by NMD to 5-25% levels [7]. At levels of >10% WT1 mutants will be detectable using the RT-PCR/dHPLC strateg

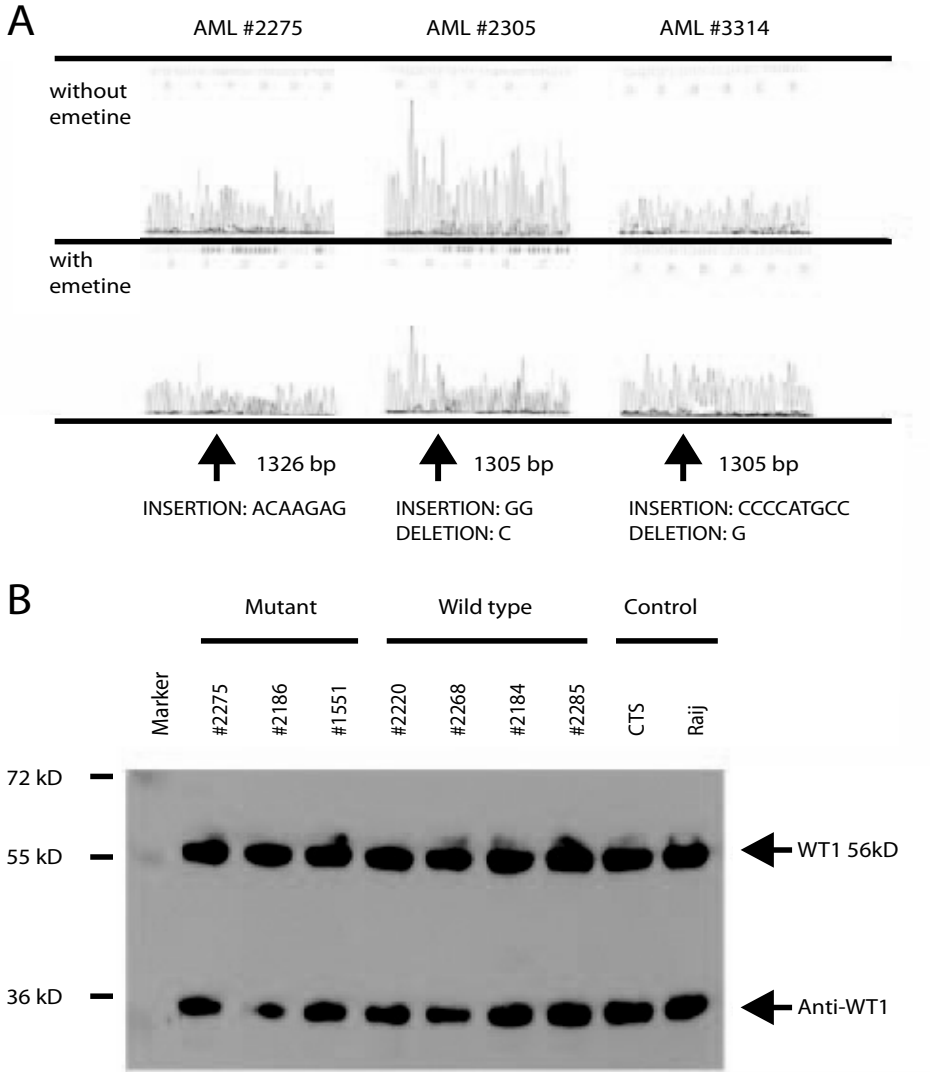


Figure 1 (A) Direct nucleotide sequence analyses of *WT1* RT-PCR products (WT1-FOR3 and WT1-REV4) with primer WT1-REV4 of three cases of AML (no. 2275, no. 2305 and no. 3314) cultured in the absence or presence of the translation inhibitor emetine. **(B)** Immunoprecipitations of *WT1*-mutant (no. 2275, no. 2186, no. 1551) and wild-type (no. 2220, no. 2268, no. 2184, no. 2285) AML cases with an N-terminal antibody against WT1. The cell lines CTS (GA (del) (nt1398) 2 bp del (detectable on gDNA and cDNA)) and Raji were included as mutant and wild-type WT1 controls, respectively.

To investigate the levels of WT1 protein in *WT1* mutant cases, we selected three AML cases with *WT1* insertion/deletion mutations (no. 2275, no. 2186, no. 1551 (no. 22-24, Table 1) and control wild-type *WT1* AML samples and performed immunoprecipitations with an antibody specific for the N terminus of human WT1, which recognizes residues 1-180 (Figure1B). The predicted truncated WT1 proteins were not detectable in the selected *WT1* mutant cases (no. 2275:42 kDa, no. 2186:42 kDa; no. 1551:41 kDa) and CTS cell line (49 kDa). It is interesting to note that mutant *WT1* mRNA is expressed in CTS (data not shown), which would suggest that the truncated WT1 protein is instable in this cell line. In contrast, wild-type WT1 (56 kDa) is expressed in all AML cases irrespective of the presence of *WT1* mutations.

In conclusion, we identified in 218 AML patients a relatively low number of *WT1* mutations by RT-PCR (2.7%), whereas by PCR on 351 AML cases additional *WT1* mutations (6.0%) were detectable. *WT1* mutations were present not only in normal karyotype AML, but also in cytogenetically abnormal AML (including AML with t(8;21), inv(16) and t(15;17)). AML cases with WT1 missense and in-frame insertion mutations express mutant *WT1* mRNA at *WT1* wild-type levels. These transcripts most probably encode proteins with impaired function because the mutations are located within the WT1 zinc fingers. The majority of *WT1* aberrations were insertion/deletion mutations that introduced PTCs. Here, we show for the first time that mutant *WT1* mRNA transcripts that carry PTCs are sensitive to NMD in primary AML. Consequently, these AML cases do not express the predicted truncated WT1 proteins. In these AML cases, mutations in *WT1* may result in haploinsufficiency rather than expression of truncated WT1 proteins with impaired function. Haploinsufficiency and impaired function have been proposed as two possible mechanisms by which mutant WT1 may affect normal hematopoietic development. Surprisingly, in the *WT1* mutant AML cases the wild-type WT1 protein is expressed at relatively high levels. Although, this phenomena has been shown in sporadic Wilms' tumors and some related syndromes, it is currently unclear whether the WT1 protein is pathogenic in these case or whether it is related to the etiology of the cell of origin [1].

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Supported by grants from the Dutch Cancer Society (Koningin Wilhelmina Fonds)

REFERENCES

1 Yang, L., Han, Y., Suarez Saiz, F. & Minden, M. D. A tumor suppressor and oncogene: the WT1 story. *Leukemia* **21**, 868-876 (2007).

2 King-Underwood, L. & Pritchard-Jones, K. Wilms' tumor (WT1) gene mutations occur mainly in acute myeloid leukemia and may confer drug resistance. *Blood* **91**, 2961-2968 (1998).

3 Summers, K. *et al.* Wilms' tumour 1 mutations are associated with FLT3-ITD and failure of standard induction chemotherapy in patients with normal karyotype AML. *Leukemia* **21**, 550-551; author reply 552 (2007).

4 Paschka, P. *et al.* Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. *J Clin Oncol* **26**, 4595-4602 (2008).

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

6 Gaidzik, V. I. *et al.* Prognostic impact of WT1 mutations in cytogenetically normal acute myeloid leukemia: a study of the German-Austrian AML Study Group. *Blood* **113**, 4505-4511 (2009).

7 Kuzmiak, H. A. & Maquat, L. E. Applying nonsense-mediated mRNA decay research to the clinic: progress and challenges. *Trends Mol Med* **12**, 306-316 (2006).

8 Wouters, B. J. *et al.* Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood* **110**, 3706-3714 (2007).

CHAPTER 6

Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia (AML): prevalence and prognostic value

Saman Abbas[1], Sanne Lugthart[1], Francois G. Kavelaars[1], Anita Schelen[1], Jasper Koenders[1], Annelieke Zeilemakers[1], Wim J.L. van Putten[2], Anita W. Rijneveld[1], Bob Löwenberg[1], and Peter J.M. Valk[1]

[1] Department of Hematology, Erasmus University Medical Center, Rotterdam, The Netherlands
[2] Department of Trials and Statistics, Erasmus University Medical Center, Rotterdam, The Netherlands

Blood 2010;116:2122-6

ABSTRACT

Somatic mutations in isocitrate dehydrogenase 1 and 2 (IDH1 and IDH2) were recently demonstrated in acute myeloid leukemia (AML), but their prevalence and prognostic impact remain to be explored in large extensively characterized AML series, and also in various other haematological malignancies. Here, we demonstrate in 893 newly diagnosed cases of AML mutations in the *IDH1* (6%) and *IDH2* (11%) genes. Moreover, we identified *IDH* mutations in two *JAK2* V617F myeloproliferative neoplasias (n=96), a single case of acute lymphoblastic leukemia (n=96) and none in chronic myeloid leukemias (n=81). In AML, *IDH1* and *IDH2* mutations are more common among AML with normal karyotype and *NPM1*^{mutant} genotypes. *IDH1* mutation status is an unfavorable prognostic factor as regards survival in a composite genotypic subset lacking *FLT3*^{ITD} and *NPM1*^{mutant}. Thus, *IDH1* and *IDH2* mutations are common genetic aberrations in AML, and *IDH1* mutations may carry prognostic value in distinct subtypes of AML.

INTRODUCTION

Somatic mutations in the genes encoding the isocitrate dehydrogenases IDH1 and IDH2 were revealed in more than 70% of WHO grade II and III astrocytomas, oligodendrogliomas and glioblastomas [1-3]. Mutations in *IDH1* and *IDH2* were mutually exclusive and affected the arginines on position 132 of IDH1 and position 172 of IDH2 [3]. Patients with malignant gliomas with IDH1 or IDH2 mutations showed a better response to therapy than those with wild-type *IDH* genes [3]. Mutations in these residues of IDH significantly disturb the function of both isocitrate dehydrogenases, as demonstrated by impaired production of nicotinamide adenine dinucleotide phosphate (NADP) [3-4]. In acute myeloid leukemia (AML), mutant IDH enzyme activity results in accumulation of the cancer-associated metabolite 2-hydroxyglutarate [5-6].

Recently, acquired mutations in the gene encoding IDH1 were identified in 8% [7] and 5.5% [8] of newly diagnosed AML cases. *IDH1* mutations were significantly associated with normal karyotype and *NPM1* mutations [7-8]. Overall, the *IDH1* mutation status did not suggest a relationship with overall survival (OS), but the sample sizes were limited in these studies [7-8]. However, a trend for an adverse effect on OS was suggested in normal karyotype AML with *NPM1*$^{\text{wild-type}}$ [7].

The prevalence and prognostic value of *IDH* mutations in AML, as well as other hematologic malignancies, remain to be further established. In this study, we determined the frequencies of both *IDH1* and *IDH2* mutations in cohorts of AML, acute lymphoblastic leukemia (ALL), chronic myeloid leukemia (CML) and *JAK2* V617F myeloproliferative neoplasia (MPN). In a cohort of 893 cases of AML, we investigated their distribution in relationship with cytogenetic and molecular risk categories as well as recurrent gene mutations commonly apparent in AML and we evaluated the impact of *IDH* mutations on treatment outcome.

METHODS

Bone marrow aspirates or peripheral blood samples of cohorts of patients with various hematologic malignancies were collected after written informed consent in accordance with the Declaration of Helsinki. All experiments described were approved by Erasmus University Medical Center Institutional Review Board. AML, ALL and CML patients were treated according to the HOVON (Dutch-Belgian Hematology-Oncology Co-operative group) AML protocols HO04, HO04A, HO29, HO42, HO42A and HO43, ALL protocols HO18, HO37, HO70 and HO71 and CML protocol HO51 (http://www.hovon.nl). The MPN samples were collected and the *JAK2* V617F mutation was determined in our routine molecular diagnostics facility.

IDH1 and *IDH2* mutations in AML, refractory anemia with excess blasts, ALL, CML and JAK2 V617F MPN were determined by cDNA amplifications using FW1-IDH1 cDNA WAVE 5'-CTTCAGAGAAGCCATTATCTG-3' and REV2-IDH1 cDNA WAVE 5'- TCACTTG-GTGTGTAGGTTATC-3' (IDH1 R132), FW1-IDH2 cDNA WAVE 5'- GAACTATCCG-GAACATCCTG-3' and REV2-IDH2 cDNA WAVE 5'- CTTGACACCACTGCCATC-3' (IDH2 R172) or FW-IDH2-Ex4 5'-GTTCAAGCTGAAGAAGATGTG-3' and REV-IDH2-Ex5-6 cDNA WAVE 5'-TGAGATGGACTCGTCGGTG-3' (IDH2 R140). All polymerase chain reaction (PCR) reactions were carried out at an annealing temperature of 60^0C in the presence of 25mM deoxynucleoside triphosphate, 15 pmol primers, 2mM $MgCl_2$, Taq polymerase and 1 times buffer (Invitrogen). Cycling conditions were as follows: 1 cycle 5 minutes at 94^0C, 30 cycles 1 minute at 94^0C, 1 minute at annealing temperature, 1 minute at 72^0C, and 1 cycle 7 minutes at 72^0C. All *IDH1* and *IDH2* reverse-transcribed PCR products were subjected to denaturing high performance liquid chromatography (dHPLC) analyses using a Transgenomics WAVE system. Samples were run at 61.4^0C (IDH1 R132), 57.7^0C (IDH2 R172) or 60.1^0C (IDH2 R140). PCR products showing aberrant dHPLC profiles were purified using the Multiscreen-PCR 96-well system (Millipore) followed by direct sequencing with the appropriate forward and reversed primers using an ABI-PRISM3100 genetic analyzer (Applied Biosystems). PCR products were sequenced with FW-IDH1 cDNA WAVE (IDH1 R132), FW-IDH2 cDNA WAVE (R172 IDH2) or FW-IDH2-Ex4 (R140 IDH2). We validated this strategy using 350 cases of de novo AML that were previously analyzed using PCR on genomic DNA followed by direct sequencing.

Information on the *IDH1* and *IDH2* mutation status of all AML cases is available as Supplementary Table 1 (available on the *Blood* Website; see the Supplemental Materials link at the top of the online article) and of all AML cases that were previously gene expression-profiled at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo, accession no. GSE6891).

The relation between *IDH* mutations and various patient characteristics were determined by the Student's *t* test, equal variances not assumed (continuous variables) and the Fisher exact test (categorical variables).

We distinguished the following cytogenetic risk categories: (1) favorable: t(8;21), inv(16) or t(15;17); (2) unfavorable: inv(3)/t(3;3), t(6;9), 11q23 abnormalities other than t(9;11), -5, 5q-, -7, 7q- or t(9;22) (cytogenetically abnormal [CA] unfavorable); (3) very unfavorable: monosomal karyotypes (MK) [9]; (4) intermediate risk I: cytogenetically normal (CN) and (V) intermediate risk II: the remaining AML cases (CA rest).

OS endpoints were death (failure) and alive at last follow-up (censored), as measured from entry onto trial. Event-free survival (EFS) endpoints were remission induction failure, disease relapse, or death from any cause, measured from entry onto trial. Distribution estimations and survival distributions of OS and EFS were calculated by the Kaplan-Meier method and the log-rank test.

Table 1. Distribution of IDH1 and IDH2 mutations in 893 cases of AML

	IDH1 mutant	*IDH2* mutant	Wild-type	*P*
Mean age at diagnosis, years (range)	50 (20-71)	50 (18-72)	45 (15-77)	.002*
Mean WBC at diagnosis, x 109/L (range)	48 (1-400)	42 (18-72)	46 (0-510)	
Mean platelets at diagnosis, x 109/L (range)	131 (14-494)	104 (11-884)	83 (3-998)	.05*
Mean blasts at diagnosis, percentage (range)	64 (0-96)	65 (15-95)	57 (0-98)	
Total, † no. (%)	55 (6)	97 (11)	743	
Female, no. (%)	30 (7)	52 (12)	384	
Male, no. (%)	25 (6)	45 (10)	395	
FAB, no. (%)				
M0	_		40	
M1	21 (12)	36 (21)	113	
M2	14 (6)	27 (11)	195	
M3	_	_	22	
M4	7 (5)	10 (7)	130	
M5	8 (5)	14 (8)	151	
M6	_	1 (6)	17	
M7	_	_	1	
RAEB	3 (15)	_	17	
RAEB-t‡	1 (2)	1 (2)	43	
Unknown	1 (6)	3 (17)	14	
Karyotype classification,§ no. (%)				
t(8;21)	_	2 (4)	51	
inv(16)	_	_	50	
t(15;17)	_	_	21	
CA unfavorable	2 (2)	6 (7)	74	
MK	_	3 (4)	72	
CN	39 (10)	58 (15)	283	<.001*
CA rest	11 (6)	23 (13)	141	
Unknown	3 (5)	5 (8)	51	
Mutations‖, no. (%)				
FLT3 ITD	15 (7)	19 (9)	177	
FLT3 TKD	9 (10)	12 (13)	69	
NPM1	35 (13)	40 (15)	191	.001*
NPM1^{wt}*FLT3*^{wt}	19 (3)	50 (9)	475	
NPM1^{mut}*FLT3*^{wt}	21 (15)	28 (20)	91	
NPM1^{mut}*FLT3*^{ITD}	14 (11)	12 (10)	100	
NPM1^{wt}*FLT3*^{ITD}	1 (1)	7 (8)	77	
N-*RAS*¶	3 (6)	3 (6)	43	
K-*RAS*¶	1 (17)	_	4	
CEBPA¶	1 (3)	4 (11)	30	

WBC indicates white blood cell count at diagnosis; FAB, French-American-British classification; _, not applicable; RAEB, refractory anemia with excess blasts; and RAEB-t, refractory anemia with excess blasts in transformation.

* IDH^{mut} vs. IDH$^{wild\text{-}type}$.

† Includes 2 AML patients with *IDH1* and *IDH2* mutation.

‡ At the time of diagnosis, these cases were classified as RAEB-t but would now be classified as AML.

§ Karyotypes were centrally reviewed. CA unfavorable: inv(3)/t(3;3), t(6;9), 11q23 abnormalities except t(9;11), -5, 5q-, -7, 7q- or t(9;22); MK: monosomal karyotypes (very unfavorable); CN: normal cytogenetics or –X or –Y as single abnormalities only (intermediate-risk I); CA rest: any other abnormal cytogenetics not included in any of the other categories (intermediate-risk II).

‖ Mutation detection in *FLT3*(ITD or TKD), *NPM1*, N-*RAS*, K-*RAS*, and *CEBPA* was performed as described previously [10-13]

¶ A total of 518 cases were analyzed

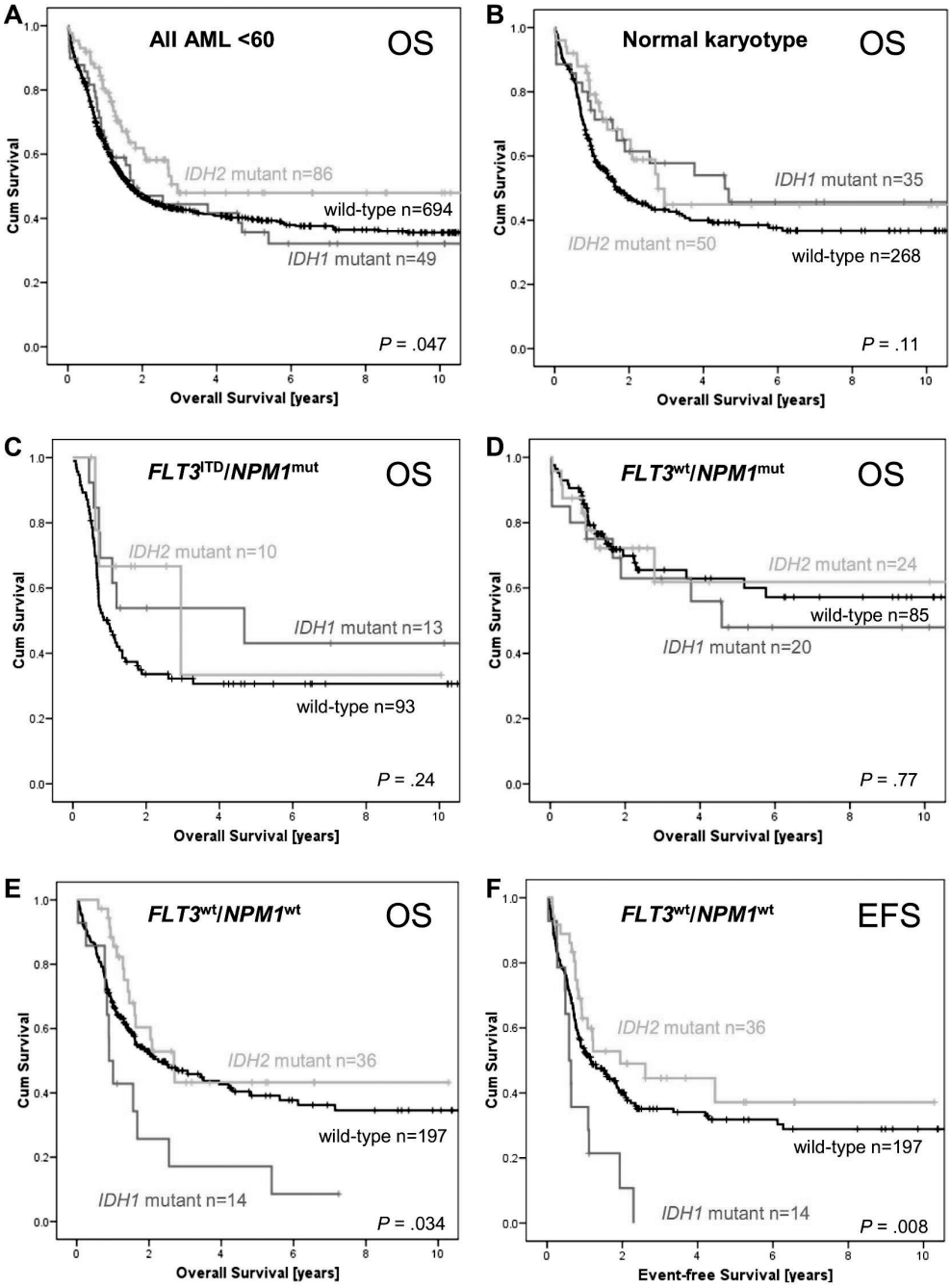


Figure 1. Survival analyses of patients of AML with or without *IDH1* and *IDH2* mutations. (A) Kaplan-Meier estimates of OS for all AML patients. (B) OS for AML patients with normal karyotypes. (C) OS for patients with intermediate-risk AML and *FLT3*^{mutant} and *NPM1*^{mutant}. (D) OS for patients with intermediate-risk AML and *FLT3*$^{wild\text{-}type}$ and *NPM1*^{mutant}. (E) OS for patients with intermediate-risk AML and *FLT3*$^{wild\text{-}type}$ and *NPM1*$^{wild\text{-}type}$. (F) EFS for patients with intermediate-risk AML and *FLT3*$^{wild\text{-}type}$ and *NPM1*$^{wild\text{-}type}$. Survival curves in red represent cases with *IDH1*^{mutant}; those in green, *IDH2*^{mutant}; and those in black, cases with *IDH1*$^{wild\text{-}type}$ and *IDH2*$^{wild\text{-}type}$, respectively. The log-rank *P* value is indicated per Kaplan-Meier analysis.

RESULTS AND DISCUSSION

To determine the frequencies of *IDH1* and *IDH2* mutations in AML, we screened cDNA of 893 newly diagnosed AMLs by reverse-transcribed PCR/dHPLC followed by direct sequencing (Table 1). *IDH1* mutations were identified in 55 AML cases (6%) and *IDH2* mutations in 97 cases (11%). A total of 152 (17%) mutations in either *IDH1* or *IDH2* were apparent in 150 cases. *IDH1* and *IDH2* mutations were mutually exclusive except in 2 cases of AML (nos. 7272 and 10400) with dual mutations in *IDH1* and *IDH2*. The R132H mutation was the most prevalent mutation in IDH1 (n=31, 56%). In addition, various other IDH1 protein mutations were identified (R132C, n=15, 28%; R132G n=6, 11%; R132L n=3, 6%). We identified 74 IDH2 R140Q mutations, [6,14] 22 cases with an IDH2 R172K mutation and a single case with a R172M substitution (no. 7309).

In addition to AML, we investigated the prevalence of *IDH1* and *IDH2* mutations in *JAK2* V617F MPN (n=96), ALL (n=96), including cases with *BCR-ABL* (n=21), *MLL* fusions (*MLL-AF4*, *MLL-AF9* or *MLL-ENL*) (n=6), *SIL-TAL* (n=2), *E2A-PBX* (n=2), and *SET-NUP* (n=1)) and CML in chronic phase (n=81). We identified a mutation in IDH1 (R132C) and IDH2 (R140Q) in 2 independent cases of *JAK2* V617F MPN, indicating that these mutations can be present before leukemic transformation [14]. In addition, we identified an IDH2 R140Q mutation in a single case of ALL. No *IDH* mutations were present in CML.

AML with $IDH1^{mut}$ and $IDH2^{mut}$ are more prevalent at older age and present with significantly higher average platelet counts at diagnosis as compared to AML with $IDH^{wild\text{-}type}$ (Table 1). *IDH1* and *IDH2* mutations were significantly more frequently present among cytogenetically defined intermediate risk AML (P=<0.001 (CN and CA rest)), as well as cytogenetically normal AML (P<0.001, CN; Table 1). In addition, *IDH* mutations appear to be significantly associated with $NPM1^{mutant}$ (P<.001; Table 1). The specificity of the pathogenetic involvement of *IDH* gene mutations in AML is also suggested by the observations that they did not significantly associate with various other recurrent mutations (i.e. $FLT3^{ITD}$ [internal tandem duplication] or $FLT3^{TKD}$ [tyrosine kinase domain], N-*RAS*, K-*RAS* or *CEBPA* gene mutations).

To investigate the prognostic value of *IDH1* mutations, 829 AML patients younger than 60 years were considered for survival analysis. The median follow-up of these patients is 33.2 months. The OS of patients with AML with or without $IDH1^{mutant}$ or $IDH2^{mutant}$ genotypes among the entire series of patients with AML did not differ (P = .05; Figure 1A). OS of IDH^{mutant} patients in the subgroups with intermediate risk cytogenetics (P = .13), normal karyotypes (Figure 1B, P = .11), and intermediate risk cytogenetics with $FLT3^{wild\text{-}type}$ (P = .32), $FLT3^{ITD}$ (P = .09), $NPM1^{wild\text{-}type}$ (P = .06), or $NPM1^{mutant}$ (P = .25) genotypes were not significantly different from those with $IDH^{wild\text{-}type}$. Similar results were obtained in analyses as regards EFS. Of note, IDH^{mutant} patients within the AML subtype $NPM1^{wild\text{-}type}$ were associated with an inferior EFS (P = .02).

Because there is significant overlap in the occurrence of mutations in *NPM1*^{mutant} and *FLT3*^{ITD} we also assessed the value of *IDH* gene mutations in each of the 4 composite variants, but no significant prognostic effect of *IDH* mutations was apparent as regards OS or EFS among *FLT3*^{ITD}/*NPM1*^{mutant} (OS, $P = .24$; EFS, $P = .24$) and *FLT3*$^{wild\text{-}type}$/*NPM1*^{mutant} (OS, $P = .77$; EFS, $P = .75$; Figure 1C and 1D, respectively). Only 8 AML patients with *IDH* mutations were identified among *FLT3*^{ITD}/*NPM1*$^{wild\text{-}type}$, which prevents reliable survival analysis. However, among the *FLT3*$^{wild\text{-}type}$/*NPM1*$^{wild\text{-}type}$ AML subtype, the presence of *IDH1* mutations (n = 14 cases) predicted for both significantly reduced OS (Figure 1E, $P = .032$) and EFS ($P = .005$). These data suggest an only moderate prognostic effect of *IDH1*^{mut} because it is not evident in genetically heterogeneous series of AML, but only in intermediate risk AML in the absence of *NPM1*^{mut} and *FLT3*^{ITD}. Apparently the *NPM1*^{mutant} and *FLT3*^{ITD} markers override the prognostic effect of *IDH1*^{mut}. In this regard we wish to note that the numbers of the 4 composite subgroups, even though this study was performed in a relatively large series of AML, become obviously increasingly small which limits the statistical power of these analyses and prohibits the interesting exploratory analysis for *IDH1* and *IDH2* mutations separately.

Acquired *IDH* gene mutations (i.e. not only in *IDH1* but also *IDH2* [6]) are common abnormalities in AML. The results of the current study demonstrate that the frequency of *IDH2* mutations exceeds those of *IDH1*. Together, *IDH1* and *IDH2* mutations account for a considerable frequency of approximately 17% in adult AML. The presence of *IDH* gene mutations appears to be associated with normal karyotypes and *NPM1* mutations. The observation that *IDH1* mutations appear to correlate with significantly inferior outcome in patients *FLT3*$^{wild\text{-}type}$/*NPM1*$^{wild\text{-}type}$ AML, requires confirmation in future studies.

ACKNOWLEDGMENTS

The authors thank Dr Gert J. Ossenkoppele (Free University Medical Center, Amsterdam, The Netherlands), Dr Jaap Jan Zwaginga (Sanquin, Amsterdam, The Netherlands), Dr Edo Vellenga (University Hospital, Groningen, The Netherlands), Dr Leo F. Verdonck (University Hospital, Utrecht, The Netherlands), Dr Gregor Verhoef (Hospital Gasthuisberg, Leuven, Belgium) and Dr Matthias Theobald (Johannes Gutenberg-University Hospital, Mainz, Germany), who provided AML cell samples; our colleagues from the stem cell transplantation and molecular diagnostics laboratories for storage of the samples and molecular analyses; and H. Berna Beverloo (Erasmus University Medical Center) for cytogenetic analyses.

This work was supported by grants from the Dutch Cancer Society (Koningin Wilhelmina Fonds).

REFERENCES

1 Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807-1812 (2008).

2 Bleeker, F. E. *et al.* IDH1 mutations at residue p.R132 (IDH1(R132)) occur frequently in high-grade gliomas but not in other solid tumors. *Hum Mutat* **30**, 7-11 (2009).

3 Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **360**, 765-773 (2009).

4 Zhao, S. *et al.* Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1alpha. *Science* **324**, 261-265 (2009).

5 Gross, S. *et al.* Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations. *J Exp Med* **207**, 339-344 (2010).

6 Ward, P. S. *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* **17**, 225-234 (2010).

7 Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* **361**, 1058-1066 (2009).

8 Chou, W. C. *et al.* Distinct clinical and biological features of de novo acute myeloid leukemia with additional sex comb-like 1 (ASXL1) mutations. *Blood* (2010).

9 Breems, D. A. *et al.* Monosomal karyotype in acute myeloid leukemia: a better indicator of poor prognosis than a complex karyotype. *J Clin Oncol* **26**, 4791-4797 (2008).

10 Valk, P. J. *et al.* Second hit mutations in the RTK/RAS signaling pathway in acute myeloid leukemia with inv(16). *Haematologica* **89**, 106 (2004).

11 Barjesteh van Waalwijk van Doorn-Khosrovani, S. *et al.* Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML. *Hematol J* **4**, 31-40 (2003).

12 Care, R. S. *et al.* Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. *Br J Haematol* **121**, 775-777 (2003).

13 Verhaak, R. G. *et al.* Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. *Blood* **106**, 3747-3754 (2005).

14 Green, A. & Beer, P. Somatic mutations of IDH1 and IDH2 in the leukemic transformation of myeloproliferative neoplasms. *N Engl J Med* **362**, 369-370 (2010).

CHAPTER 7

General discussion

GENERAL DISCUSSION

Acute myeloid leukemia (AML) is a heterogeneous disease that results from malignantly transformed hematopoietic progenitor cells. Leukemia develops though series of accumulating genetic changes, resulting in impaired abilities of cellular survival, proliferation and/or differentiation [1-2]. These alterations impact also clinical behavior and the response to therapy.

The work presented in this thesis has been divided into two main sections, each covering several chapters. The results will be discussed here and will be accompanied with directions for future research.

I- Genome-wide molecular analysis in AML

Instead of classical gene based applications, genome-wide approaches, in particular those involving DNA microarray, allow researchers to simultaneously investigate changes in gene expression as well as numerical changes in malignant and normal cells at high resolution. These techniques have been successfully applied to the study of leukemia [3-5].

One of the techniques based on DNA microarray technology is gene expression profiling (GEP). GEP has been shown to be a powerful approach to classify and predict AML subtypes with specific genetic aberrations and different survival outcomes [6-12] (for review see [11]). Genomic technologies have the capacity to address the complexity of molecular networks in transformed cell. Pathways that might be targetable pharmacologically are highlighted, which provide keys to drug development.

Microarray approaches that directly interrogate numerical genetic alterations in a genome-wide fashion are of great interest. SNP-array is a genome-wide approach that enables to genotyping hundreds of thousands of SNPs across the genome and detects DNA deletions and gains as well as CN-LOH [13-15].

In chapter 2, we described the development of a software tool SNPExpress that allows accurate analyses of SNP genotype calls, copy numbers, CNVs combined with the concurrent analysis of gene expression data obtained from Affymetrix and Illumina DNA mapping platforms. The power of SNPExpress was illustrated by generating DNA mapping profiles of series of 48 AML patients, of which the gene expression profiles had previously been determined [8]. SNPExpress correctly confirmed all cytogenetic abnormalities that had been found in these AML patients. In addition, SNPExpress revealed BCR-ABL, MLL-AF4 and NUP98-NSD1 cryptic translocations in AML cases which had been missed by cytogenetic analysis. With SNPExpress one can identify small chromosomal regions showing gains, losses or UPD. These events can result in increased expression of oncogenes or point towards losses of possible tumor suppressor genes, haploinsufficiency or mutated as well as recessive mutations in case of UPD. UPD may also have important consequences as regards gene expression. For instance, UPD might affect expression of (imprinted) genes. The recognition of these aber-

rations by SNPExpress has been possible due to the combinatorial visualization of genotype, copy number and gene expression.

What are the limitations of SNPExpress? The large datasets generated by mapping array platform make it difficult to visualize the sequentially aligned SNPs of the full-length chromosomes on a single screen for genotype analysis. The selection of the most informative SNPs that represent particular haplotypes may solve this issue. However, SNPExpress has been applied in the study of alternative exon-splicing in AML. That will allow distinguishing between different isoforms of a gene which opens the door to detect specific alterations in exon usage that may play a central role in disease mechanism and etiology. SNPExpress has also been used to investigate the expression and aberrations in microRNAs which facilitates the accurate measurement of microRNAs that are involved in gene regulation. Furthermore, SNPExpress can be applied for the analysis of genome-wide methylation data.

The contribution of genome-wide SNP-arrays in AML research is valuable and software tools like SNPExpress which facilitate the identification of biologically and clinical relevant entities are useful for genome-wide studies. The future perspective of genome-wide analyses is the integration of the massive amount of data obtained from different genome-wide approaches and the development of software platforms that allow for an integrated evaluation. The development of SNPExpress has set the first step towards this perspective. SNP-arrays also allow the investigation of genetic aberrations not only in genes known to be involved in leukemia, but it may also reveal novel aberrations in AML (chapter 3).

In chapter 3, an integrated approach using genome-wide SNP-array molecular data analyses by SNPExpress and gene expression profiling revealed *BCL11B* as a novel oncogene in AML. *BCL11B* is involved in several recurrent translocations in ALL [16-18], but we showed that *BCL11B* is involved in a number of cryptic 14q32 chromosomal translocations in primary AML as well. *BCL11B* is a Kruppel zinc finger transcription factor which is pivotal regulator of differentiation and survival of hematopoietic cells and mediates the function of transcriptional corepressor complex NURD [19-20]. The rearranged *BCL11B* locus was confirmed by FISH analyses in several AML cases. The *BCL11B* locus, located on chromosome 14, appeared to be translocated to different chromosomes in each case of AML. All translocations of *BCL11B* resulted in high expression. Apparently, there is no fusion BCL11B protein. What causes the high expression of *BCL11B* in the AML cases with 14q32 translocations currently remains elusive. The most likely explanation would be that *BCL11B* allele involved in the translocation will be regulated by an active promoter or by enhancer elements from the partner chromosome. This type of abnormal gene regulation has been reported in T-ALL translocations [16], where a remote enhancer of *BCL11B* caused the activation of *TLX3* and *NKX2-5* genes in T-ALL with translocation t(5;14)(q35;q32). The identification of these promoters and/or enhancer elements will be subject of future studies to help understand the mechanism leading to increased BCL11B expression and the role of *BCL11B* overexpression in leukemogenesis.

BCL11B overexpression has been observed in squamous cell carcinoma [21]. When T-cell lines overexpress Bcl11b, it triggers resistance to apoptosis and radiomimetic drugs accompanied by cell cycle delay [22]. What is the consequence of high expression of BCL11B in hematopoietic cells? This question was addressed in a murine 32D cell line model. 32D cells overexpressing Bcl11b showed a consistent decrease in proliferation and reduced granulocytic maturation ability. Thus, Bcl11b may partially block or delay cellular differentiation in hematopoietic cells. *BCL11B* is required for normal development of gene regulation in thymocytes and T-cells [19,23]. AMLs with *BCL11B* translocation show a biphenotypic signature with expression of both T-cell and myeloid markers. However, the presence of the FLT3-ITD mutation in these cases of leukemia strongly suggests that these leukemias are of myeloid origin.

What role do these "T-cell genes" play in myeloid cells when expressed? This is a question to be addressed in further studies using functional approaches in animal/cell line models. It is possible that *BCL11B* activate groups of T-cell genes thus causing the biphenotype and/or contribute to leukemogenesis. It would also be of great interest to investigate the presence of possible mutations in the *BCL11B* gene, especially in the promoter region of *BCL11B* that could lead to the overexpression of the gene and, if so, whether it would be a marker for clinical outcome. The argument for addressing this question is that not all AML cases that aberrantly express BCL11B carry chromosomal translocations involving *BCL11B* and thus there might be undisclosed gene mutations.

Recent years have witnessed a great interest in the application of genome-wide approaches to the study of leukemia. These studies have identified multiple novel genetic alterations targeting critical cellular pathways that contribute to leukemogenesis. These include alterations of genes regulating lymphoid development, tumor suppressors, and oncogenes, for example, genes targeted by lesions in ALL, such as *IKZF1* [24-25]. Overall, fewer recurrent alterations have been identified in AML [26] than in ALL so that in AML the necessity is greater to apply other genome-wide approaches such as NGS to identify new lesions. Also, it will be important to perform simultaneous analysis of sequence variation, epigenetic changes, and gene expression in a genome-wide fashion. This is now becoming attainable with the rapid advances in next generation sequencing that may replace microarray analysis of genetic alterations [27-29].

II- Mutation analyses in AML

The identification of new gene mutations will shed light on the biology of AML, but can also be used to define diagnostically and prognostically relevant subgroups of AML. This is useful for in AMLs that currently lack prognostically informative markers. Novel markers may also provide better insights into the biology of the disease. This has for instance previously been well documented for gene mutations in *FLT3*-ITD and *NPM1* in AML with normal cytogenetics [30-34]. In the chapters 4 through 6 of this thesis, we investigated known genetic mutations in AML focusing on the frequency and the prognostic impact of these genes in a large cohort of well-characterized AML patient samples.

Chapter 4 describes the mutation analyses of the *CBL* gene in a cohort of primary AML patients. CBL is an ubiquitin protein ligase (E3) that regulates signal transduction pathways by tyrosine kinases. These tyrosine kinases are crucial for cell function and development [35-37]. Disruption of the normal regulatory mechanisms that control CBL function can lead to pathological conditions and malignant diseases [38-41]. All published *CBL* mutations are located within the conserved linker region (LR) and ring finger (RF) resulting in a CBL protein lacking the entire RF which is critical for CBL E3 activity [42]. The expression of the truncated CBL may act as a dominant negative protein by inhibiting proper downregulation of critical activated tyrosine kinases, such as KIT and FLT3 in AML [43].

Our analyses of an AML cohort for *CBL* mutations in exon 8 and 9 that affect the proper splicing of the gene showed that the frequency of these mutations in AML is low (0.6%). All mutations affect the splice acceptor or donor site of CBL exon 8. Interestingly, all identified CBL gene mutants are apparent in core binding leukemias, i.e., those with inv(16) or t(8;21), indicating that there is preferential association between *CBL* exon 8 splice variant mutations and CBF leukemia. The strong association of these mutations with aberrations affecting *RUNX1* was subsequently confirmed in various studies [44-45].

Why is there such a strong association between these two genetic events? Mutations in receptor tyrosine kinase KIT are also strongly associated with CBF leukemias [46-48]. In fact, we have shown that KIT mutations are present in 88% of CBF leukemias and CBL proteins mediate ubiquitination and degradation of KIT upon stimulation with stem cell factor (SCF) [49]. Therefore, the expressed mutant CBL that lacks E3 activity still binds KIT receptor but is impaired in the downstream ubiquitination. Thus, mutant CBL may act as a dominant negative protein and will lead to sustained KIT activation, similar to KIT activating mutations in CBF leukemias. This hypothesis is supported by the fact that none of *CBL* mutant AML cases carry AML-specific mutations affecting RAS signalling, such as mutations in *FLT3*, *NRAS* or *KRAS* except in one AML case with KIT D816 mutation which would potentially result in prolonged constitutive activation of KIT.

We previously identified 16 groups of AML patients on the basis of their gene expression signature [8]. One of the groups is driven by inv(16) which is associated with KIT mutations. If we evaluate this group in greater detail, we can see that it can be divided into two sub-clusters based on the gene expression signatures. KIT mutations are more prevalent in one subcluster than in the other. This raises general question. What is the role that CBL mutations might play in these sub-groups? Do CBL mutants contribute to the development of AMLs that lack KIT mutation by prolonging constitutive activation of KIT receptor? This might opens a window for therapeutic approach, such as imatinib-based therapy for patients with extracellular domain mutations in c-*KIT*.

In chapter 5, we investigated Wilm's tumor 1 (*WT1*) mutations' frequency and their sensitivity to nonsense-mediated RNA decay (NMD) surveillance mechanisms in series of AML patients. *WT1* is a zinc finger transcription factor, which has been shown to act as tumor

suppressor or oncogene in leukemia [50]. Mutations in WT1 in AML are present at different frequencies and associated with normal karyotype AML [51-54]. In some studies, mutations in *WT1* have been implicated as an adverse prognostic marker for AML [52], but in other studies no impact on survival was observed [55]. The mutations detected in *WT1* often result in frame shifts leading to the introduction of premature termination codons (PTCs). PTCs trigger NMD to degrade the mutant transcripts, but is unknown whether *WT1* mutant transcripts are degraded by NMD [56].

We detected *WT1* mutations in 2.7% of the AMLs cases on cDNA level, but the mutations' prevalence was higher using genomic DNA, i.e. 6%. There was no association between *WT1* mutations and other acquired mutations in AML and survival analysis did not reveal any significant difference in overall (OS) and event-free survival (EFS). The presence of mutations at a higher percentage in the genomic DNA than cDNA that result in premature terminations codons detected in *WT1* gene, triggered our interest to investigate the involvement of NMD in leukemic *WT1* mutations AML.

The involvement of NMD is confirmed when emetine, which is known to block translation-mediated NMD [56], stabilized the *WT1* mutant mRNA transcripts. The PTCs are introduced as a result of insertion/deletion mutations in *WT1* in 77% of AML cases. When a PTC is located over 55 bp from the intron boundary, as demonstrated in our *WT1* mutation analyses, the PTC makes the gene more prone to NMD which leads to the degradation of the mutant transcript. But the mechanism of NMD is not waterproof. In few AML cases carrying PTCs that should trigger NMD, the mutant transcript still expressed. This shows, in line with other studies [56], that NMD may reduce the levels of PTC-containing transcripts rather than entirely eliminate transcripts of truncated WT1 proteins that may still be expressed. However, on the protein level the truncated WT1 protein was not expressed, suggesting the instability of the truncated WT1 protein. On the other hand, in *WT1* mutant AMLs, the wild-type WT1 protein is expressed at relatively high level. Haploinsufficiency of WT1 might take place in AML cases that do not express the truncated WT1 rather than expression of truncated WT1 proteins with impaired function. The wild-type *WT1* is highly expressed in AML and this has been shown previously to be associated with cell death [57-58]. These observations support a role of WT1 maintaining the survival of AML cells [50].

The demonstration of the involvement of NMD in *WT1* mutations in AML opens the door for questions concerning the activation of NMD upon mutations in other genes in AML. As previously mentioned, pre-termination codons have been introduced in 77% of AML cases with insertion/deletion mutations in *WT1* and may be responsible for degradation of the mutant transcripts. To reveal the real frequency of a mutation, analyses using genomic DNA are warranted in mutations containing PTCs as this gives additional insight on mechanisms such as nonsense-mediated RNA decay that lead to possible haploinsufficiency of the involved genes which may affect normal hematopoietic development. This objective can be pursued by applying next- generation sequencing with AML cells treated with translation inhibitors such

as emetine. Next-generation sequencing will be informative in revealing, with high base pair resolution and genome-wide fashion, the transcripts expressed upon the inhibition of NMD which will give better understanding on the scale of NMD involvement in AML.

Chapter 6 presents studies regarding *IDH1/2* gene mutations in AML. These are new class of oncogenic transcripts that have recently been identified in brain tumors [59-61] and also in leukemia [62]. We showed that in AML, *IDH2* mutations were more prevalent (11%) than *IDH1* mutations (6%). Both *IDH1* and *IDH2* gene mutations were mutually exclusive with the exception of two AML cases. The mutations affected mostly specific arginine residues in *IDH1* and *IDH2* as has been shown by others [59,62]. These arginine residues are located in the active site of IDH1 and IDH2 enzymes and participate in isocitrate binding [63]. The studies also confirmed that *IDH1*/2 gene mutations are more frequent in cytogenetically normal AML and AML with *NPM1* mutations. We also identified mutations in *IDH1* and *IDH2* in myeloproliferative neoplasms (MPNs) with *JAK2* V617F mutations. Importantly, mutations in IDH enzymes have been linked to the leukemic transformation of myeloproliferative neoplasms, an event that is associated with poor clinical prognosis [64]. This mutational data in myeloid malignancies and previously in malignant gliomas suggest that *IDH1/2* mutations can occur early in the disease and may drive tumorgenesis [65].

Survival analysis of IDH mutants vs. non mutants showed that in AML subtype $FLT3^{\text{wild-type}}/NPM1^{\text{wild-type}}$, the presence of *IDH1* mutations predicted an inferior outcome (Chapter 6). This shows a moderate prognostic effect of $IDH1^{\text{mutant}}$ since it is only evident in intermediate risk AML without *NPM1* and *FLT3* mutations and not in more heterogeneous series of AML. *NPM1* and *FLT3* mutations may override the prognostic effect of *IDH1* mutations. Studies investigating the prognostic value of IDH mutations in AML showed controversial outcomes. Number of studies showed a significant impact of *IDH* mutations on survival [66-69], while did not show any impact of *IDH* mutations on survival [70-74]. Noteworthy, recent studies across large patient cohorts revealed that the presence of *IDH1* mutations is associated with a worse prognosis in AML patients with mutated *NPM1* without *FLT3*-ITD [68-69], which was not observed in our study. Further studies dealing with the prognostic value of *IDH1* mutations in AML in relation to specific genotypes, e.g., $FLT3^{\text{wild-type}}/NPM1^{\text{wild-type}}$, are warranted.

AML is characterized by recurrent chromosomal translocations and somatic mutations that define biologically distinct disease subtypes. Among these abnormalities are mutations or rearrangements of genes encoding aberrant transcription factors that directly perturb gene expression and disrupt cell differentiation and survival. Others include gain-of-function mutations of kinases involved in transduction of growth and proliferation signals. The revealing of *IDH1* and *IDH2* as a new class of gene mutations that contributes to leukemogenesis, stimulated research regarding the oncogenic mechanism of these mutant enzymes. In general, only one copy of the *IDH1* and *IDH2* genes has been found to be mutated in tumors [59-61,75-77], suggesting that the mutations give enzymatic gain of function rather than

a loss of function. One wild-type allele remains expressed which may be required for the normal cellular metabolism. Normally, IDH1 and IDH2 are $NADP^+$ enzymes that catalyze the conversion of isocitrate to alpha-ketoglutarate (αKG). The mutations affecting certain arginine residues in *IDH1* and *IDH2* disrupt the ability of the enzymes to convert isocitrate to αKG [59,78] and thereby acquiring a neomorphic enzymatic activity by catalyzing the reduction of αKG to 2-hydroxyglutarate (2HG) [79]. In fact, the R140 mutation was identified by monitoring 2HG levels [80]. It has been suggested that the production of 2HG contributes to malignant transformation by impairing the function of enzymes that require αKG as a substrate.

Distinct subtypes of AML can be distinguished not only by gene expression profiling but also based on DNA methylation which is another hallmark of AML [81]. A recent study showed the consequences of the production of 2HG caused by *IDH1* and *IDH2* mutations on progression towards leukemogenesis. Increased 2HG levels interfered with the normal cycle of DNA methylation and demethylation through inhibiting αKG-dependent enzymes such as TET2 [82]. *TET2* is a αKG-dependent enzyme that converts 5-methyl-cytosine (5mC) to 5-hydroxymethylcytosine (5hmC) in DNA [83-84] and its loss of function lead αKG reduction will result in hypermethylation. Furthermore, a recent report showed that *TET2* mutations associated with myeloid malignancies, including AML, impair hydroxylation of 5mC [85]. Thus, mutations in *TET2* or its loss of function induced by *IDH1* and *IDH2* mutations appear to result in hypermethylation.

Unveiling the leukemogenic mechanism orchestrated by *IDH1/2* mutations will make way for targeted therapies that inhibit the neomorphic function of mutant IDH enzymes and the production of 2HG that will promote cellular differentiation and improve AML patient's outcomes. It was not anticipated that this new class of mutations affecting metabolic enzymes normally involved in citrate metabolism would defined a new subtype of AML with distinct oncogenic mechanism. Extended studies may reveal more cellular pathways affected by αKG deregulation. Further studies may serve to specifically address the effects of IDH1/2 mutants on stem- or progenitor cell function and on hematopoietic differentiation.

The genome-wide technology with probably the greatest impact in the field of cancer genomics is next-generation sequencing. It enables the sequencing of targeted, whole genome, and whole transcriptomes [26,62,86-87]. Also, NGS can produce data to characterize gene expression, methylation, histone packaging, and transcription factors [88]. It offers many advantages over microarray-based assays, like higher base pair resolution, less artefacts, and larger dynamic range and high coverage. The development of next-generation sequencing of cancer genomes will also have considerable value in cancer diagnostics. It allows diagnosis from smaller samples, including circulating tumor cells and DNA [89-90]. Computational analysis will become a central part of discoveries and diagnostic efforts. This will lead to the generation of immense amount of data, which will make effective and accessible storage of data a main challenge as well.

Next-generations sequencing will facilitate the detection of molecular markers in AML and will catalyze the molecular characterization of AML [26]. These improvements in the definition of new genetic markers will provide a firm basis for the identification and characterization of novel molecular targets as has been realized by the introduction of *FLT3* inhibitors in the treatment of *FLT3*-mutated AML [91]. It is foreseen that this will allow a further translation of molecular research into clinical practice in AML.

REFRENCES

1 Estey, E. & Dohner, H. Acute myeloid leukaemia. *Lancet* **368**, 1894-1907, (2006).

2 Lowenberg, B., Downing, J. R. & Burnett, A. Acute myeloid leukemia. *N Engl J Med* **341**, 1051-1062, (1999).

3 Ebert, B. L. & Golub, T. R. Genomic approaches to hematologic malignancies. *Blood* **104**, 923-932, (2004).

4 Hoheisel, J. D. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* **7**, 200-210, (2006).

5 Armstrong, S. A. *et al.* MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet* **30**, 41-47, (2002).

6 Bullinger, L. *et al.* Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med* **350**, 1605-1616, (2004).

7 Valk, P. J., Delwel, R. & Lowenberg, B. Gene expression profiling in acute myeloid leukemia. *Curr Opin Hematol* **12**, 76-81, (2005).

8 Valk, P. J. *et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* **350**, 1617-1628, (2004).

9 Verhaak, R. G. *et al.* Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* **94**, 131-134, (2009).

10 Wouters, B. J. *et al.* Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood* **110**, 3706-3714, (2007).

11 Wouters, B. J., Lowenberg, B. & Delwel, R. A decade of genome-wide gene expression profiling in acute myeloid leukemia: flashback and prospects. *Blood* **113**, 291-298, (2009).

12 Wouters, B. J. *et al.* Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood* **113**, 3088-3091, (2009).

13 Irving, J. A. *et al.* Loss of heterozygosity in childhood acute lymphoblastic leukemia detected by genome-wide microarray single nucleotide polymorphism analysis. *Cancer Res* **65**, 3053-3058, (2005).

14 Yang, J. J. *et al.* Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. *JAMA* **301**, 393-403, (2009).

15 Jankowska, A. M. *et al.* Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. *Blood* **113**, 6403-6410, (2009).

16 Nagel, S. *et al.* Activation of TLX3 and NKX2-5 in t(5;14)(q35;q32) T-cell acute lymphoblastic leukemia by remote 3'-BCL11B enhancers and coregulation by PU.1 and HMGA1. *Cancer Res* **67**, 1461-1471, (2007).

17 Bernard, O. A. *et al.* A new recurrent and specific cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. *Leukemia* **15**, 1495-1504 (2001).

18 Nagel, S., Kaufmann, M., Drexler, H. G. & MacLeod, R. A. The cardiac homeobox gene NKX2-5 is deregulated by juxtaposition with BCL11B in pediatric T-ALL cell lines via a novel t(5;14)(q35.1;q32.2). *Cancer Res* **63**, 5329-5334 (2003).

19 Cismasiu, V. B. *et al.* BCL11B functionally associates with the NuRD complex in T lymphocytes to repress targeted promoter. *Oncogene* **24**, 6753-6764, (2005).

20 Wakabayashi, Y. *et al.* Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. *Nat Immunol* **4**, 533-539, (2003).

21 Ganguli-Indra, G. *et al.* CTIP2 expression in human head and neck squamous cell carcinoma is linked to poorly differentiated tumor status. *PLoS One* **4**, e5367, (2009).

22 Grabarczyk, P. *et al.* Increased expression of bcl11b leads to chemoresistance accompanied by G1 accumulation. *PLoS One* **5**, (2010).

23 Albu, D. I. *et al.* BCL11B is required for positive selection and survival of double-positive thymocytes. *J Exp Med* **204**, 3003-3015, (2007).

24 Mullighan, C. G. *et al.* Genomic analysis of the clonal origins of relapsed acute lymphoblastic leukemia. *Science* **322**, 1377-1380, (2008).

25 Mullighan, C. G. *et al.* Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med* **360**, 470-480, (2009).

26 Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66-72, (2008).

27 Mullighan, C. G. Genomic analysis of acute leukemia. *Int J Lab Hematol* **31**, 384-397, (2009).

28 Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* **40**, 722-729, (2008).

29 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628, (2008).

30 Dohner, K. *et al.* Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. *Blood* **106**, 3740-3746, (2005).

31 Verhaak, R. G. *et al.* Mutations in nucleophosmin (NPM1) in acute myeloid leukemia (AML): association with other gene abnormalities and previously established gene expression signatures and their favorable prognostic significance. *Blood* **106**, 3747-3754, (2005).

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

33 Nakao, M. *et al.* Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. *Leukemia* **10**, 1911-1918 (1996).

34 Yanada, M., Matsuo, K., Suzuki, T., Kiyoi, H. & Naoe, T. Prognostic significance of FLT3 internal tandem duplication and tyrosine kinase domain mutations for acute myeloid leukemia: a meta-analysis. *Leukemia* **19**, 1345-1349, (2005).

35 Dikic, I., Szymkiewicz, I. & Soubeyran, P. Cbl signaling networks in the regulation of cell function. *Cell Mol Life Sci* **60**, 1805-1827, (2003).

36 Duan, L., Reddi, A. L., Ghosh, A., Dimri, M. & Band, H. The Cbl family and other ubiquitin ligases: destructive forces in control of antigen receptor signaling. *Immunity* **21**, 7-17, (2004).

37 Thien, C. B. & Langdon, W. Y. Cbl: many adaptations to regulate protein tyrosine kinases. *Nat Rev Mol Cell Biol* **2**, 294-307, (2001).

38 Mancini, A., Koch, A., Wilms, R. & Tamura, T. c-Cbl associates directly with the C-terminal tail of the receptor for the macrophage colony-stimulating factor, c-Fms, and down-modulates this receptor but not the viral oncogene v-Fms. *J Biol Chem* **277**, 14635-14640, (2002).

39 Peschard, P. *et al.* Mutation of the c-Cbl TKB domain binding site on the Met receptor tyrosine kinase converts it into a transforming protein. *Mol Cell* **8**, 995-1004, (2001).

40 Roussel, M. F., Shurtleff, S. A., Downing, J. R. & Sherr, C. J. A point mutation at tyrosine-809 in the human colony-stimulating factor 1 receptor impairs mitogenesis without abrogating tyrosine

kinase activity, association with phosphatidylinositol 3-kinase, or induction of c-fos and junB genes. *Proc Natl Acad Sci U S A* **87**, 6738-6742 (1990).

41 Wu, W. J., Tu, S. & Cerione, R. A. Activated Cdc42 sequesters c-Cbl and prevents EGF receptor degradation. *Cell* **114**, 715-725, (2003).

42 Ryan, P. E., Davies, G. C., Nau, M. M. & Lipkowitz, S. Regulating the regulator: negative regulation of Cbl ubiquitin ligases. *Trends Biochem Sci* **31**, 79-88, (2006).

43 Renneville, A. *et al.* Cooperating gene mutations in acute myeloid leukemia: a review of the literature. *Leukemia* **22**, 915-931, (2008).

44 Reindl, C. *et al.* CBL exon 8/9 mutants activate the FLT3 pathway and cluster in core binding factor/11q deletion acute myeloid leukemia/myelodysplastic syndrome subtypes. *Clin Cancer Res* **15**, 2238-2247, (2009).

45 Haferlach, C. *et al.* AML with CBFB-MYH11 rearrangement demonstrate RAS pathway alterations in 92% of all cases including a high frequency of NF1 deletions. *Leukemia* **24**, 1065-1069, (2010).

46 Valk, P. J. *et al.* Second hit mutations in the RTK/RAS signaling pathway in acute myeloid leukemia with inv(16). *Haematologica* **89**, 106 (2004).

47 Care, R. S. *et al.* Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. *Br J Haematol* **121**, 775-777, (2003).

48 Beghini, A. *et al.* C-kit mutations in core binding factor leukemias. *Blood* **95**, 726-727 (2000).

49 Zeng, S., Xu, Z., Lipkowitz, S. & Longley, J. B. Regulation of stem cell factor receptor signaling by Cbl family proteins (Cbl-b/c-Cbl). *Blood* **105**, 226-232, (2005).

50 Yang, L., Han, Y., Suarez Saiz, F. & Minden, M. D. A tumor suppressor and oncogene: the WT1 story. *Leukemia* **21**, 868-876, (2007).

51 King-Underwood, L. & Pritchard-Jones, K. Wilms' tumor (WT1) gene mutations occur mainly in acute myeloid leukemia and may confer drug resistance. *Blood* **91**, 2961-2968 (1998).

52 Paschka, P. *et al.* Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. *J Clin Oncol* **26**, 4595-4602, (2008).

53 Summers, K. *et al.* Wilms' tumour 1 mutations are associated with FLT3-ITD and failure of standard induction chemotherapy in patients with normal karyotype AML. *Leukemia* **21**, 550-551; author reply 552, (2007).

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

55 Gaidzik, V. I. *et al.* Prognostic impact of WT1 mutations in cytogenetically normal acute myeloid leukemia: a study of the German-Austrian AML Study Group. *Blood* **113**, 4505-4511, (2009).

56 Kuzmiak, H. A. & Maquat, L. E. Applying nonsense-mediated mRNA decay research to the clinic: progress and challenges. *Trends Mol Med* **12**, 306-316, (2006).

57 Yamagami, T. *et al.* Growth inhibition of human leukemic cells by WT1 (Wilms tumor gene) antisense oligodeoxynucleotides: implications for the involvement of WT1 in leukemogenesis. *Blood* **87**, 2878-2884 (1996).

58 Ito, K. *et al.* Antiapoptotic function of 17AA(+)WT1 (Wilms' tumor gene) isoforms on the intrinsic apoptosis pathway. *Oncogene* **25**, 4217-4229, (2006).

59 Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **360**, 765-773, (2009).

60 Bleeker, F. E. *et al.* IDH1 mutations at residue p.R132 (IDH1(R132)) occur frequently in high-grade gliomas but not in other solid tumors. *Hum Mutat* **30**, 7-11, (2009).

61 Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807-1812, (2008).

62 Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* **361**, 1058-1066, (2009).

63 Xu, X. *et al.* Structures of human cytosolic NADP-dependent isocitrate dehydrogenase reveal a novel self-regulatory mechanism of activity. *J Biol Chem* **279**, 33946-33957, (2004).

64 Abdel-Wahab, O. *et al.* Genetic analysis of transforming events that convert chronic myeloproliferative neoplasms to leukemias. *Cancer Res* **70**, 447-452, (2010).

65 Green, A. & Beer, P. Somatic mutations of IDH1 and IDH2 in the leukemic transformation of myeloproliferative neoplasms. *N Engl J Med* **362**, 369-370, (2010).

66 Boissel, 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. *J Clin Oncol* **28**, 3717-3723, (2010).

67 Green, C. L. *et al.* The prognostic significance of IDH1 mutations in younger adult patients with acute myeloid leukemia is dependent on FLT3/ITD status. *Blood* **116**, 2779-2782, (2010).

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

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

70 Wagner, K. *et al.* Impact of IDH1 R132 mutations and an IDH1 single nucleotide polymorphism in cytogenetically normal acute myeloid leukemia: SNP rs11554137 is an adverse prognostic factor. *J Clin Oncol* **28**, 2356-2364, (2010).

71 Patel, K. P. *et al.* Acute myeloid leukemia with IDH1 or IDH2 mutation: frequency and clinicopathologic features. *Am J Clin Pathol* **135**, 35-45, (2011).

72 Ho, P. A. *et al.* Molecular alterations of the IDH1 gene in AML: a Children's Oncology Group and Southwest Oncology Group study. *Leukemia* **24**, 909-913, (2010).

73 Thol, F. *et al.* Prognostic impact of IDH2 mutations in cytogenetically normal acute myeloid leukemia. *Blood* **116**, 614-616, (2010).

74 Chou, W. C. *et al.* Distinct clinical and biologic characteristics in adult acute myeloid leukemia bearing the isocitrate dehydrogenase 1 mutation. *Blood* **115**, 2749-2754, (2010).

75 Watanabe, T., Nobusawa, S., Kleihues, P. & Ohgaki, H. IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas. *Am J Pathol* **174**, 1149-1153, (2009).

76 Hartmann, C. *et al.* Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: a study of 1,010 diffuse gliomas. *Acta Neuropathol* **118**, 469-474, (2009).

77 Balss, J. *et al.* Analysis of the IDH1 codon 132 mutation in brain tumors. *Acta Neuropathol* **116**, 597-602, (2008).

78 Zhao, S. *et al.* Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1alpha. *Science* **324**, 261-265, (2009).

79 Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744, (2009).

80 Ward, P. S. *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* **17**, 225-234, d (2010).

81 Figueroa, M. E. *et al.* DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* **17**, 13-27, (2010).

82 Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell*, (2010).

83 Ito, S. *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133, (2010).

84 Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935, (2009).

85 Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839-843, (2010).

86 Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59, (2008).

87 Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191-196, (2010).

88 Mardis, E. R. & Wilson, R. K. Cancer genome sequencing: a review. *Hum Mol Genet* **18**, R163-168, (2009).

89 Leary, R. J. *et al.* Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med* **2**, 20ra14, (2010).

90 Maheswaran, S. *et al.* Detection of mutations in EGFR in circulating lung-cancer cells. *N Engl J Med* **359**, 366-377, (2008).

91 Tallman, M. S. New agents for the treatment of acute myeloid leukemia. *Best Pract Res Clin Haematol* **19**, 311-320, (2006).

NEDERLANDSE SAMENVATTING

Acute myeloïde leukemie (AML) is een heterogene ziekte waarin een verscheidenheid aan cytogenetische abnormaliteiten en moleculaire veranderingen plaatsvinden in de leukemie cellen. Deze afwijkingen spelen een rol bij de ontwikkeling van leukemia. Ze beïnvloeden de celdeling (proliferatie), uitrijping van de cel (differentiatie), zelf-vernieuwing van de vroeg hematopoietische cellen, apoptosis en de reparatie van DNA. Verscheidene nieuwe gen-gebaseerde technieken zoals gen expressie profielen, DNA microarrays, analyse van cytosine methylering en next generation sequencing, waren toegepast voor het onderzoek van AML. Deze technieken gaven wetenschappers de mogelijkheid om normal en kwaadaardige cellen op een genoom-brede en moleculair niveau te onderzoeken en verzorgde waardevolle inzichten in vragen hadden betrekking op classificatie, vooruitgang en ontwikkeling van AML. In de afgelopen twee decennia is er veel onderzoek gedaan aan de genetische afwijkingen in AML. Daarnaast zijn er in de afgelopen jaren nieuwe mutaties in ziektegenen geïdentificeerd. Naast het feit dat deze afwijkingen een rol spelen bij de pathogene van leukemie, kunnen ze ook een prognostische waarde hebben. Zo is het bekend dat sommige mutaties of chromosomale abnormaliteiten en mutaties geassocieerd zijn met een relatief goede prognose, terwijl bij andere afwijkingen de vooruitzichten minder goed gesteld zijn. Het is daarom voor de behandeling van AML essentieel om de verschillende vormen van deze ziekte goed te kunnen onderscheiden. Door de genetische afwijkingen te kunnen identificeren die geassocieerd zijn met bepaalde (sub)-types van AML betrokken bij initiatie van de ziekte.

Het werk in dit proefschrift is verdeeld in twee delen. Het eerste deel (hoofdstukken 2 en 3) bespreekt de toepassing van de microarray technologie van DNA (SNP arrays) om nieuwe abnormaliteiten in AML te ontdekken. Het tweede deel van dit proefschrift (hoofdstukken 4, 5 en 6) behandelt de studies over onlangs ontdekte genetische veranderingen in AML. Het effect van gen veranderingen op prognose in een grote cohort van AML patiënten werd onderzocht.

Hoofdstuk 2 beschrijft de ontwikkeling van een softwarehulpmiddel, genaamd SNPExpress, die de gezamenlijke interpretatie van genotype, de veelvormige kopie nummer variatie (in het Engels "copy number variation", afgekort als CNV), verlies van heterozygositeit (in het Engels "loss of heterozygosity", afgekort LOH) gecombineerd met de gegevens van gen expressie profielen (GEP) op diverse platforms mogelijk maakt. Hoofdstuk 3 bespreekt de toepassing van combinatiebenadering van genoom-brede genotypering en GEP data om nieuwe afwijkingen in AML te identificeren. De moleculaire analyse die door SNPExpress werd uitgevoerd openbaarde *BCL11B* als nieuwe oncogen in AML. Translocaties, waarin *BCL11B* is betrokken, waren aller eerst geïdentificeerd in T-cel lymfomen. Aberraties in *BCL11B* op chromosoom 14 leiden tot overexpressie van *BCL11B* in AML patiënten. Als *BCL11B* tot overexpressie komt, reguleert dit in co-expressie van T-cel gerelateerde genen maar ook myeloide markers. Wij toonden aan dat *Bcl11b* overexpressie leidt tot verminderde

proliferatie en blokkeert differentiatie van hematopoëtische cellen in een muis beenmerg cellijn aan.

Hoofdstuk 4, beschrijft het onderzoek aan de distributie van mutaties in het *CBL* gen in een primaire AML cohort en de associatie van deze mutaties met specifieke genetische subtypes van AML. Het *CBL* gen codeert voor een ubiquitine ligase die de signaal transductie via tyrosine kinases receptoren zoals, FLT3 en KIT, reguleert. Onze mutatie analyse toont aan dat CBL mutaties plaats vinden in de sequenties die coderen voor zogenaamde de linker en de zink finger domeinen, die belangrijk zijn voor de binding van CBL aan de tyrosine kinases receptoren. Bovendien bleken CBL mutatie sterk zijn geassocieerd met AML subtypes inv(16) en t(8;21), of zogeheten core binding factor leukemie (CBF). 88% van de CBF gevallen toont mutaties in KIT.

Hoofdstuk 5, beschrijft het onderzoek aan de frequentie en de gevolgen van mutaties in het transcriptie factor Wilms Tumor (*WT1*) gen. Hierin, tonen we voor het eerst aan dat een belangrijk RNA toezicht en het regulatie mechanisme (in het Engels "nonsense-mediated RNA decay", afgekort NMD) betrokken is bij de degradatie van het mutant mRNA transcripten van *WT*. Wanneer een insertie/deletie mutatie plaats vindt op een bepaalde positie in *WT1*, 55 base paren en meer van intron/exon grens, leidt dit tot de vorming van stop codons (in het Engels (premature termination codon PTC). De vorming van PTCs activeert NMD en leidt tot de afbraak van het mutant transcript en dient als een bescherming mechanisme voor de cel tegen de expressie van afwijkende mutant eiwitten.

Tot slot, hoofdstuk 6 bediscussieert de frequentie in een AML cohort de mutaties in isocitrate dehyrogenases 1/2 (afgekort *IDH1* en *IDH2*). *IDH1* en *IDH2* behoren tot een nieuwe klasse oncogenen die eerder werden beschreven in hersentumoren. IDH enzymen zijn betrokken bij productie van nicotinamide adenine dinucleotide phosphate-oxidase (afgekort NADPH), hetgeen een beschermend effect tegen cellulaire stress. In onze AML cohort detecteerden we *IDH1/2* mutaties in 17% van de gevallen. *IDH1* mutaties treffen namelijk het arginine R132 en in *IDH2* de arginines R140 en R172. We identificeerden tevens *IDH1/2* mutaties in andere myeloproliferatieve aandoeningen, namelijk myeloproliferatieve neoplasma (afgekort MPN) met *JAK2* V617F mutatie. De overlevingsanalyse, waarbij AML patiënten met *IDH1* mutatie vergeleken worden met AML patiënten zonder *IDH1* mutatie, toont aan dat AML subtype $FLT3^{\text{wild-type}}/NPM1^{\text{wild-type}}$ met IDH1 mutaties slechtere overlevingskansen hebben dan AML patiënten zonder *IDH1* mutatie. De ontdekking van nieuwe moleculaire markers en afwijkingen in AML door genoom-brede benaderingen zal onze kennis over AML biologie verbeteren. Bovendien, het identificeren en karakteriseren van nieuwe moleculaire markers zal nieuwe mogelijkheden bieden voor medicijnen ontwikkeling en het vertalen van moleculair onderzoek naar de kliniek zal bevorderen.

ABBEREVIATIONS

ALL	Acute Lymphoblastic Leukemia
AML	Acute Myeloid Leukemia
ASXL1	Additional Sex combs-Like 1
BAC	Bacterial Artificial Chromosome
BCL11B	B-Cell Lymphoma/leukemia 11B
BMI1	Polycomb ring finger oncogene
CBL	Casitas B-lineage Lymphoma
CEBPA	CCAAT/Enhancer-Binding Protein Alpha
CGH	Comparative Genomic Hybridization
CLP	Common Lymphoid Progenitor
CML	Chronic Myeloid Leukemia
CMP	Common Myeloid Progenitor
CNA	Copy Number Abnormalities
CN-LOH	Copy Neutral Loss of Heterozygosity
CNV	Copy Number Variations
dHPLC	High Performance Liquid Chromatography
DNMT3A	DNA Methyl Transferase 3A
EFS	Event Free Survival
ETP	Enhancer of Trithorax and Polycomb
EZH2	Enhancer of Zeste-2
FAB	French-American-British
FISH	Fluorescent In Situ Hybridization
FLT3-ITD	FMS-Like Tyrosine Kinase Internal Tandem Repeat
FLT3-TKD	FLT3-Tyrosine Kinase Domain
G-CSF	Granulocyte Colony-Stimulating Factor
GEP	Gene Expression Profiling
GM-CSF	Granulocyte-Macrophage Colony-Stimulating Factor
HELP	HpaII Enriched Ligated mediate PCR assay
HMM	Hidden Markov Model
HOVON	Stichting Hemato-Oncologie voor volwassenen Nederland
HSC	Hematopoietic Stem Cell
IDH1	Isocitrate Dehydrogenase 1
IDH2	Isocitrate Dehydrogenase 2
JAK2	Janus Kinase 2
KDa	Kilo Dalton
K-RAS	V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LOH	Loss of Heterozygosity

LTR	Long Terminal Repeat
M-CSF	Macrophage Colony-Stimulating Factor
MK	Monosomal Karyotype
MN1	Meningioma 1 gene
MoMSV	Moloney Murine Sarcoma Virus
MPN	Myeloproliferative Neoplasia
NADP	Nicotinamide Adenine Dinucleotide Phosphate
NGS	Next Generation Sequencing
NMD	Nonsense Mediated RNA Decay
NPM1	Nuceophsomin 1 gene
N-RAS	Neuroblastoma RAS viral oncogene homolog
NURD complex	Nucleosome remodelling and histone deacetylase complex
OS	Overall Survival
PCR	Polymerase chain reaction
PTC	Premature Termination Codon
Q-PCR	Quantative Polymerase Chain Reaction
SCF	Stem Cell Factor
SNP	Single Nucleotide Polymorphism
TCA	Tricarboxylic acid
UPD	Uniparental Disomy
WBC	White Blood Cells
WHO	World Health Organization
WT1	Wilm's Tumor 1

DANKWOORD

Graag wil ik van de gelegenheid gebruiken maken om alle mensen te bedanken die ieder op hun eigen manier hebben bijgedragen aan de totstandkoming van dit proefschrift. Ik hoop dat ik niemand te kort doe in dit dankwoord en als wel, dan bied ik op voorhand mijn excuses aan.

De meeste dank ben ik verschuldigd aan mijn promotor Bob Löwenberg en co-promotor Peter Valk. Bob Löwenberg, beste Bob, ik wil je bedanken voor de gelegenheid om te promoveren binnen jouw afdeling en voor je steun, betrokkenheid, visie en samenwerking in alle projecten. Het is een waar genoegen om op afdeling hematologie onder jouw leiding te hebben gewerkt. Mijn co-promotor, Peter Valk. Beste Peter, Ik ben je veel dank verschuldigd voor jou enorme bijdrage en steun. Ik had me geen betere begeleider kunnen wensen. Het was een voorrecht om ruim vijf jaar bij je te mogen werken. Bedankt voor je gezelligheid, je enthousiasme en ik kijk ernaar uit om met je te blijven samenwerken.

Graag wil ik de leden van de kleine commissie, professor Leendert Looijenga, professor Rob Pieters en professor Gert Ossenkoppele bedanken voor hun bijdrage en het beoordelen van het manuscript.

Beste Ruud Delwel, ik wil je graag danken voor jouw scherpe inzichten, waardevolle persoonlijke gesprekken en je steun vooral tijdens mijn zoektocht naar een postdoc positie. Van Ivo Touw heb ik veel geleerd. Zijn enthousiasme en kritische noten bij werkbesprekingen waren motiverend om mijn onderzoek te verbeteren. Marieke van Linderen wil ik bedanken voor haar steun en behulpzaamheid vooral tijdens het plannen en uitvoeren van experimenten.

Mijn twee paranimfen, Annelieke Zeilmaker en Claudia Erpelinck, bedankt dat jullie samen met mij deze dag tot een succes willen maken. Annelieke, mijn onderzoek partner, ik wil jouw speciaal bedanken voor je bijdragen aan het onderzoek. Zonder jou zou ik veel experimenten niet kunnen afronden en artikelen niet kunnen publiceren. Beste Claudia, als je er niet was, kon ik veel patiënten cDNA/gDNA/RNA niet vinden. Ik kon altijd op je rekenen en bij jou terecht voor een gezellig gesprek.

To my friends and colleagues; Marta Pratcorona, Ana Ribeiro and Lucila Solerno. Dear Marta, it was a pleasure to know you and be experiments partners. You have been a very active member in our group and accomplished a lot with your research in the nine months you stayed in our lab. I am glad to be friends with you. I wish you good luck with your carrier and I am sure that we will continue hearing from each other and you will be the main reason for me to visit the lovely city of Barcelona. Dear Ana I had a lot of fun going to lunch together and chat very often. I wish you and your family all the best in life and let us keep in touch and I promise I try my best to be in Brazil during the World Cup. Lucila, I knew you as a very kind and caring person. I wish you a life full of joy and happiness with your newborn baby and the rest of your family.

De secretariële ondersteuning op de afdeling hematologie was altijd behulpzaam en aardig; Ans Mannens, Monique Mes en Leenke de Jong-de Visser bedankt voor alles. Egied Simons, bedankt voor jouw hulp met het uitwerken van dit proefschrift. Je hebt een prachtig beroep. Jan van Kapel, bedankt voor jouw hulp en leuke tijd. Tomasia, veel succes en ik hoop dat we elkaar weer ontmoeten.

The bioinformatics team Roel Verhaak, Matijs Sanders en Justine Peters. Your input and contribution to our hematology research is enormous. With these tools and methods we conducted nice studies and opened new fields to explore in the hematology department. I want to thank you all for your input and nice times in the department and wish you good luck and success.

De studenten die hun steentje hebben bijgedragen aan het onderzoek: Gabriëlle Rotmans en Zabiollah Gholam. Dank voor jullie inzet. De HOVON cohort managers; Francois Kavelaars en Jasper Koenders, bedankt voor jullie inzet en het bevoorraden van patiënten DNA voor de mutatie studies.

De dames van de moleculaire diagnostiek; Isabel Chu, Wendy Geertsma-Kleinekoort, Chantal Goudswaard, Pauline Hogenbirk-Hupkes, Antoinette Hoven van-Beijen, Sonja van der Poel en Marloes van der Werf, ook dames en heren van BMT; Eric Braakman, Lianne Broeders, Hannie Busking-van der Lelij, Pia van Eersel, Peter van Geel, Hans Hoogerbrugge, Arie Prins en Anita Rijken-Schelen, bedankt voor jullie hulp en de leuke tijd in het lab.

De AIOs op de 13e verdieping; Su Ming, Sophie, Annemarie, Karishma, Elnaz, Renee, Simone, Rasti, Andrzej, Kerim, Erdogan, Kasia, Davine, Patricia, allen veel succes met het afronden van jullie promotie en de toekomst plannen.

Nog niet genoemd (oud-)collega's, Goran Rudez, Mojca Jongen-Lavrencic, Stefan Erkeland, Menno Dijkstra, Mahban Irandoust, Marijke Valkhof, Onno Roovers, Tanja Jevdjovic, Arturo Sanz Sanz, Paulette Strien, Yvonne de Knegt, Martijn Schoester, Roya Sarwari, Eric Bindels, Jurgen Haanstra, Bas Wouters, Sanne Lugthart en Marije Havermans, jullie wil ik ook bedanken voor jullie hulp en leuke tijden.

Pap, Mam en mijn broertjes Araz en Salar. Bedankt dat jullie mij altijd ondersteund hebben in mijn keuzes. Bedankt ook voor jullie onvoorwaardelijke steun, liefde en vertrouwen in mij. Zonder jullie was dit nooit gelukt. Jullie compenseerden voor alle dipjes en stress die promoveren soms opleverden. Voor jullie ben ik eeuwig dankbaar.

CURRICULUM VITAE

Saman Abbas werd op 23 Julie 1974 geboren in Bagdad-Irak. Hij sinds 1997 woonachtig in Nederland. In 1999 begon hij aan de studie biologie en medisch laboratorium onderzoek aan Saxion Hogeschool IJselland in Deventer en behaalde zijn diploma in 2003. Direct na, stroomde hij in naar master opleiding in oncologie aan de Vrije Universiteit van Amsterdam. In de laatste fase van zijn master opleiding, deed hij zes maanden onderzoek naar het effect van histone deacetylases inhibitors in prostaat kanker cellen in het laboratorium van Dr. Manel Esteller aan de Spaanse Nationale Kanker Onderzoek Centrum in Madrid (Spanje). Na het behalen van zijn master diploma, begon hij in december 2005 als promovendus in de onderzoekgroep van Dr. Peter Valk op de afdeling hematologie van het Erasmus MC (promotor Prof.dr. Bob Löwenberg). Aldaar vond het onderzoek beschreven in dit proefschrift plaats.

PUBLICATONS

1 Abou El Hassan, M. A., van der Meulen-Muileman, I., **Abbas, S.** & Kruyt, F. A. Conditionally replicating adenoviruses kill tumor cells via a basic apoptotic machinery-independent mechanism that resembles necrosis-like programmed cell death. *J Virol* **78**, 12243-12251, (2004).

2 Smeets, S. J., Braakhuis, B. J., **Abbas, S.**, Snijders, P. J., Ylstra, B., van de Wiel, M. A., Meijer, G. A., Leemans, C. R., Brakenhoff, R. H. Genome-wide DNA copy number alterations in head and neck squamous cell carcinomas with or without oncogene-expressing human papillomavirus. *Oncogene* **25**, 2558-2564, (2006).

3 Sanders, M. A., Verhaak, R. G., Geertsma-Kleinekoort, W. M., **Abbas, S.**, Horsman, S., van der Spek, P. J., Lowenberg, B. Valk, P. J. SNPExpress: integrated visualization of genome-wide genotypes, copy numbers and gene expression levels. *BMC Genomics* **9**, 41, (2008).

4 Verhaak, R. G., Verhaak, R. G., Wouters, B. J., Erpelinck, C. A., **Abbas, S.**, Beverloo, H. B., Lugthart, S. Lowenberg, B., Delwel, R., Valk, P. J. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* **94**, 131-134, (2009).

5 **Abbas, S**., Rotmans, G., Lowenberg, B. & Valk, P. J. Exon 8 splice site mutations in the gene encoding the E3-ligase CBL are associated with core binding factor acute myeloid leukemias. *Haematologica* **93**, 1595-1597, (2008).

6 **Abbas, S.**, Erpelinck-Verschueren, C. A., Goudswaard, C. S., Lowenberg, B. & Valk, P. J. Mutant Wilms' tumor 1 (WT1) mRNA with premature termination codons in acute myeloid leukemia (AML) is sensitive to nonsense-mediated RNA decay (NMD). *Leukemia* **24**, 660-663, (2010).

7 **Abbas, S.**, Lugthart, S., Kavelaars, F. G., Koenders, J. E., Zeilemaker, A., van Putten, W. J., Rijneveld, A. W., Lowenberg, B., Valk, P. J. Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value. *Blood* **116**, 2122-2126, (2010).

8 Veronika Rockova, **Saman Abbas,** Bas J. Wouters, Claudia A.J. Erpelinck, H. Berna Beverloo, Ruud Delwel, Wim L.J. van Putten, Bob Löwenberg and Peter J.M.Valk. Risk-stratification of intermediate-risk acute myeloid leukemia: integrative analysis of a multitude of gene mutation and gene expression markers. *Blood* (2011), in press.

Chapter 1, Figure 1

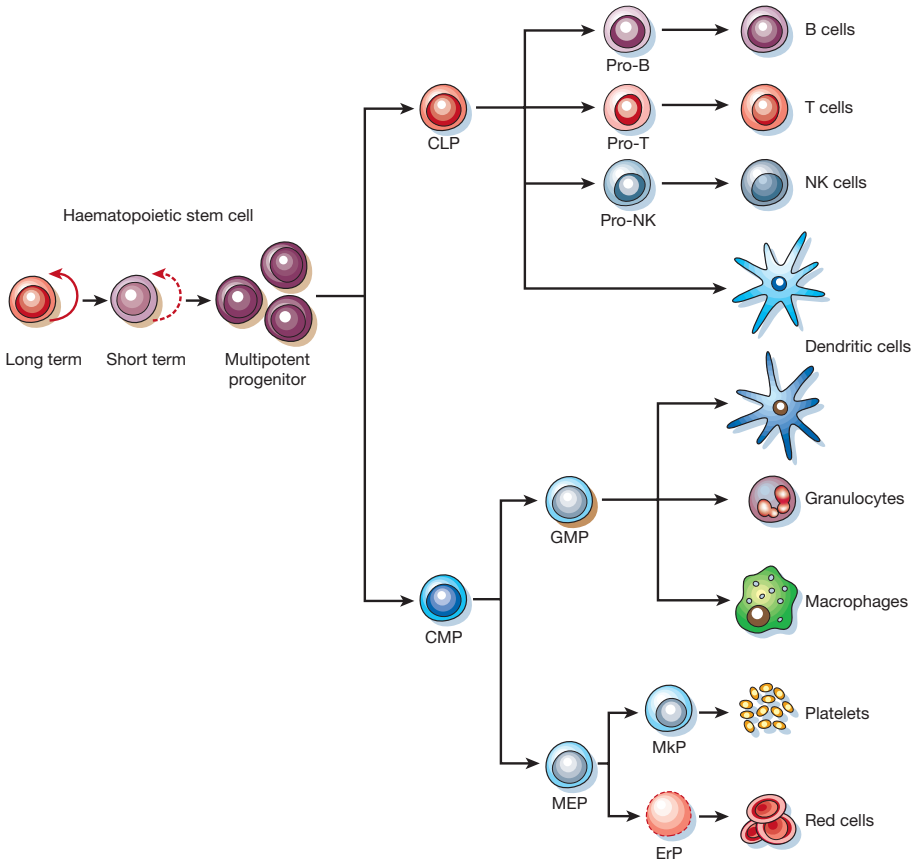


Figure 1. Schematic representation of hematopoiesis. Scheme is adapted from Reya et al [3]. The hematopoietic stem cells (HSCs) can be subdivided into long-term self-renewing HSCs, short term HSCs and multipotent progenitors. They give rise to common lymphoid progenitors (CLPs; the precursor of all lymphoid cells i.e., T-cells, B-cells, natural killer cells (NK cells), and dendritic cells) and common myeloid progenitors (CMPs; the precursor of all myeloid cells). CMPs give rise to granulocyte macrophage precursors (GMPs) and megakaryocyte erythrocyte precursors (MEPs). Subsequently, GMPs give rise to dendritic cells, granulocytes, and macrophages. MEPs give rise to megakaryocyte precursors (MkPs) and erythrocyte precursors (ErP), which further differentiate to platelets and red blood cells, respectively.

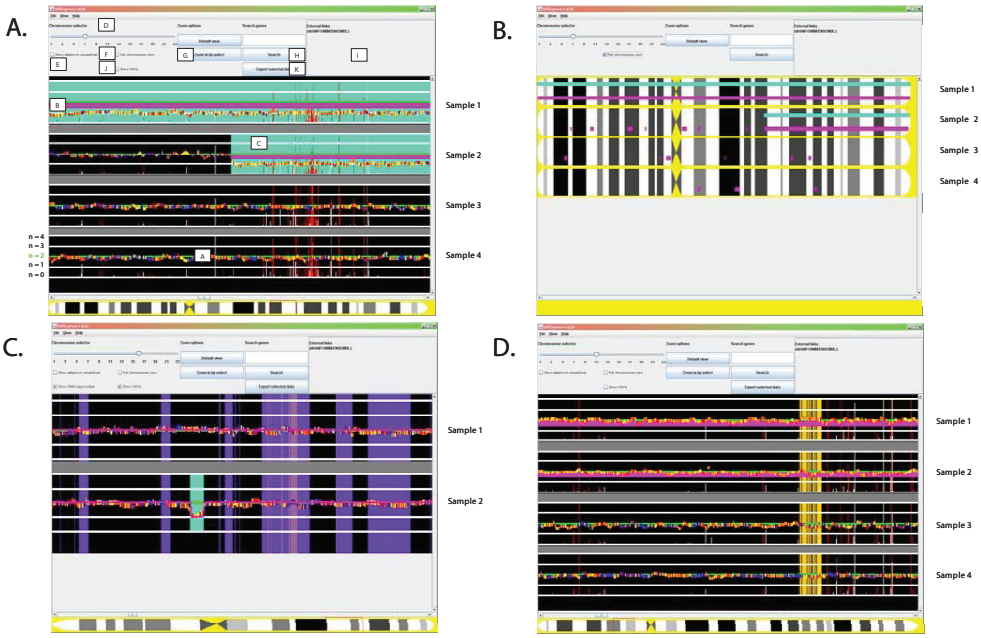


Figure 1

SNPExpress Screenshot. A. DNA mapping array data from the Affymetrix 250K *Nsp*I DNA mapping array was used to sequentially align the genotypes and copy numbers of chromosome 7 of four AML samples. The copy numbers (n=0, 1, 2, 3, 4) are shown for each individual patient by horizontal lines. Copy number n=2 is depicted by a green line (A). The SNP genotypes are sequentially aligned along the chromosome (AA: red; BB: yellow; AB: blue, noCall: white). LOH is indicated by a thick magenta horizontal bar (B), gains (default n>2.5) by a pink (Figure 1C) and losses (default n<1.5) by a turquoise background (C). Gene expression levels are visualized as vertical white bar at the chromosomal position of the gene-specific probe set. In the event that multiple probe sets span the same region in the chromosome-wide view the vertical gene expression bars are red and proportional to the highest expression value. The two upper samples clearly display a decreased copy number as was previously shown by cytogenetics, i.e., a complete monosomy (sample 1) or a deletion of the q-arm of chromosome 7 (sample 2). The overall expression of the majority of genes in the displayed region is decreased in the samples with chromosome 7 abnormalities. The chromosome selector (D; where 23 is the X chromosome), the mouse-over function showing info of each SNP or probe set (E), full chromosome view (F), zoom function (G) gene search function (H), the links to external databases (I), display CNVs (J) and export selected data (K) options are indicated. **B.** Full chromosome view of samples from 1A. **C.** CNV (purple background) and copy number of each SNP based on a HMM model (HMM copy number, magenta line) of the two AML patients from examples [33]. In the event that multiple CNVs span the same region in the chromosome-wide view the background is violet, whereas single CNV are indicated with a rosy brown background. **D.** UPD of chromosome 11 demonstrated using SNPExpress. Example of large scale UPD on chromosome 11 in the upper two AML patients with a normal karyotype in comparison to two other AML samples. The overall copy number is two and large regions of LOH are indicated by the thick magenta line across the chromosome. After using the search function, SNPs associated with *WT1* are depicted with an orange background.

Chapter 3
Figure 1

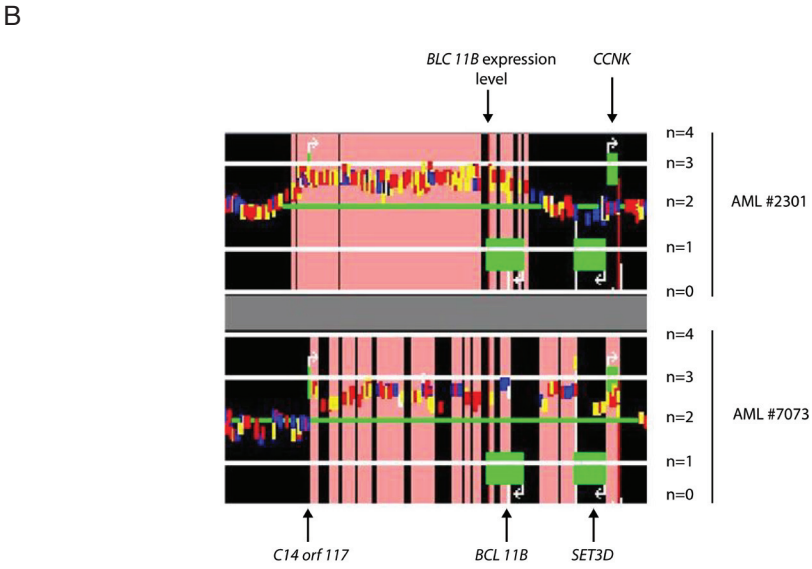


Figure 1: Identification of interstitial amplifications on 14q32.2 using SNPExpress. (A) Sequential alignment of the genotypes with copy numbers from the Affymetrix DNA mapping array of chromosome 14q32.2 of four AML samples [24]. The copy numbers are shown for each individual patient by horizontal lines (n=0, 1, 2, 3, 4). The SNP genotypes are sequentially aligned along the chromosome (AA: red; BB: yellow; AB: blue, No call: white). Gains (default n>2.5) are depicted as pink background. Gene expression levels are visualized as vertical white bars at the chromosomal position of the gene-specific probe set. In the event that multiple probe sets span the same region in the chromosome-wide view the vertical gene expression bars are red and proportional to the highest expression value. The green boxes represent exons of the encoded genes, and the arrows indicate the orientation. In AML #2301 and AML #7073 clear amplifications are visible, whereas these aberrations are absent in the two control AMLs. (B) Snapshot of SNPExpress showing the amplified region in AML case #2301 and #7073 from Figure 1A, showing the genes located within the amplified regions. *C14orf177* and *BCL11B* are amplified in both AML cases, whereas SETD3 and CCNK only in AML #7073. BCL11B expression is increased in AML #2301 and #7073 as indicated by the red bar (multiple probe sets).

Chapter 3 Figure 2

A

B

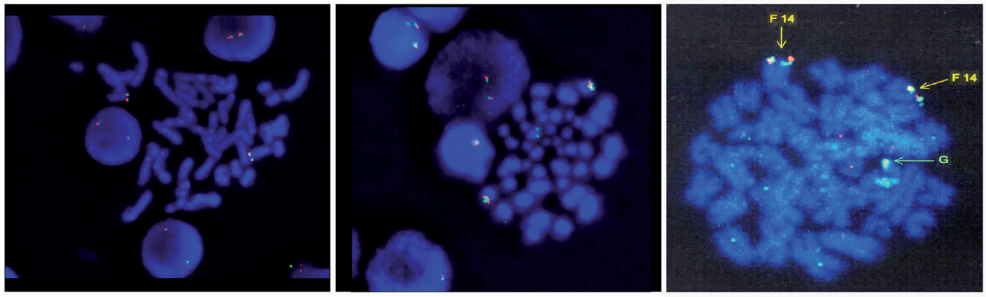
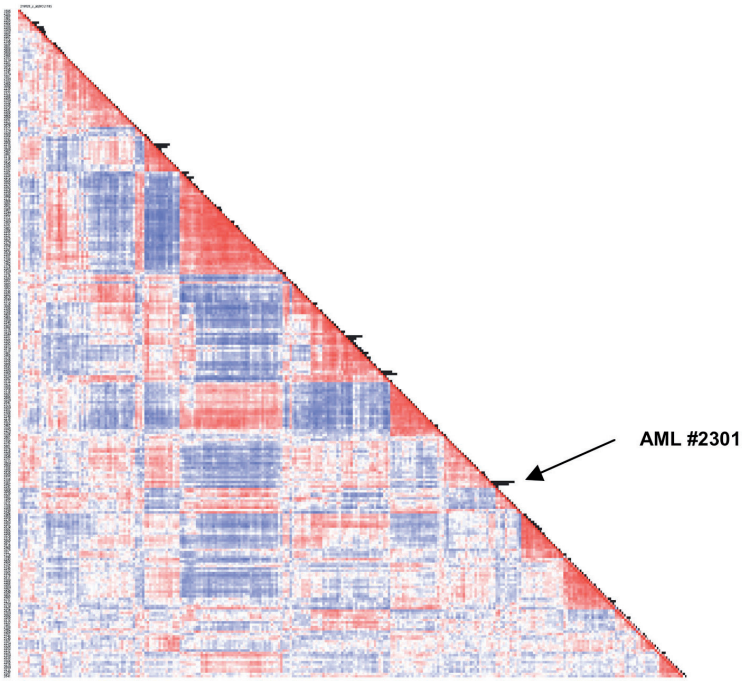


Figure 2: FISH analysis of AML cases #2301 and #7073 using probes specific for *BCL11B* and flanking *BCL11B*. (A) Schematic representation of the FITC-labeled BAC probe (RP11 431B1) covering the *BCL11B* locus and Texas Red-labeled BAC probe (RP11 242A7) covering the region adjacent to *BCL11B*. (B) Microscope images of FISH analysis performed on metaphases chromosomes of AML cases #2301 and #7073 showing additional green signal (RP11431B1) indicative for an extra copy of the *BCL11B* locus.

A



B

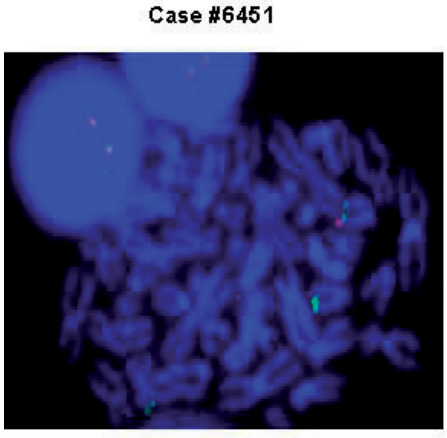
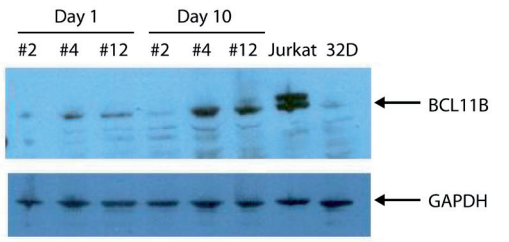


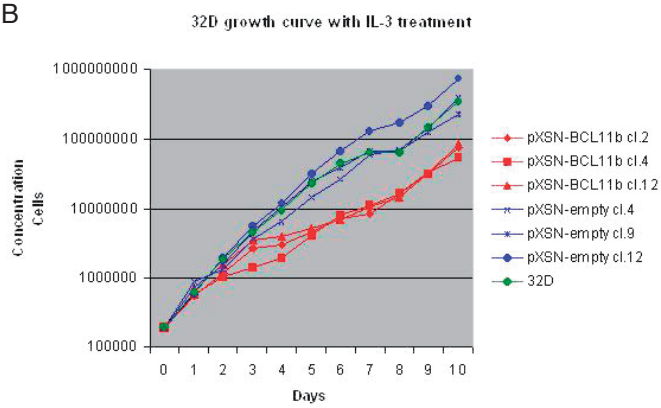
Figure 4: Correlation view based on gene expression profiling of 285 AML cases [9] (A) Correlation view of 285 AML cases showing the gene expression correlation based on 2856 probe sets [9]. The black bars adjacent to the correlation view indicate expression of *BCL11B*, where the size of the bars is proportional to the levels of *BCL11B* expression in the individual AML samples. *BCL11B* expression in case #2301 is indicated by an arrow. Forty cases with increased *BCL11B* expression (2.5 times above average) were selected for FISH analyses. (B) FISH analysis performed on metaphase spreads of AML cases #6451 (and #6366) showing disassociation of the probe RP11242A7 (red) and RP11431B1 (green) indicating translocation of *BCL11B*.

Chapter 3 Figure 6

A



B



C

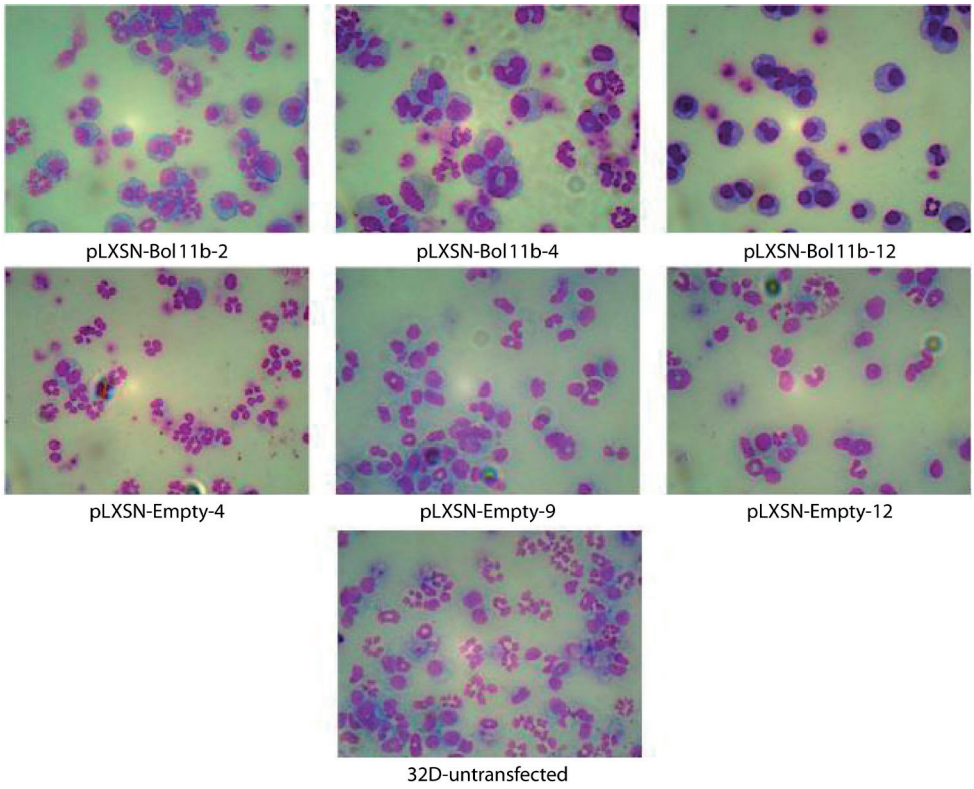


Figure 6: Effects of BCL11B overexpression in murine 32D(GCSF-R) cells. (A) Western blot analyses for Bcl11b in 32D(GCSFR) cells. 32D(GCSFR) clones overexpressing *BCL11B* are indicated by #2, #4, and #12 (IL3 1 and 10 days). Lysates obtained from theses clones were immunostained for BCL11b at day 1 and day 10 (Jurkat cells: positive control; 32D: 32D(GCSF-R) cells). GAPDH was used as loading control (lower panel). (B) Growth curve of 32D(GCSFR) cells with and without (empty) *BCL11B* expression incubated with IL3. 32D(GCSFR) cells were counted every 24 hours for ten days (32D: 32D(GCSF-R) cells). (C) May-Grünwald-Giemsa-stained cytospins of 32D(GCSF-R) cells with (upper panel) and without (lower panel) Bcl11b expression incubated with GCSF for 7 days. Granulocytic differentiation is monitored by the presence of cells with segmented nuclei.

Chapter 5 Figure 1

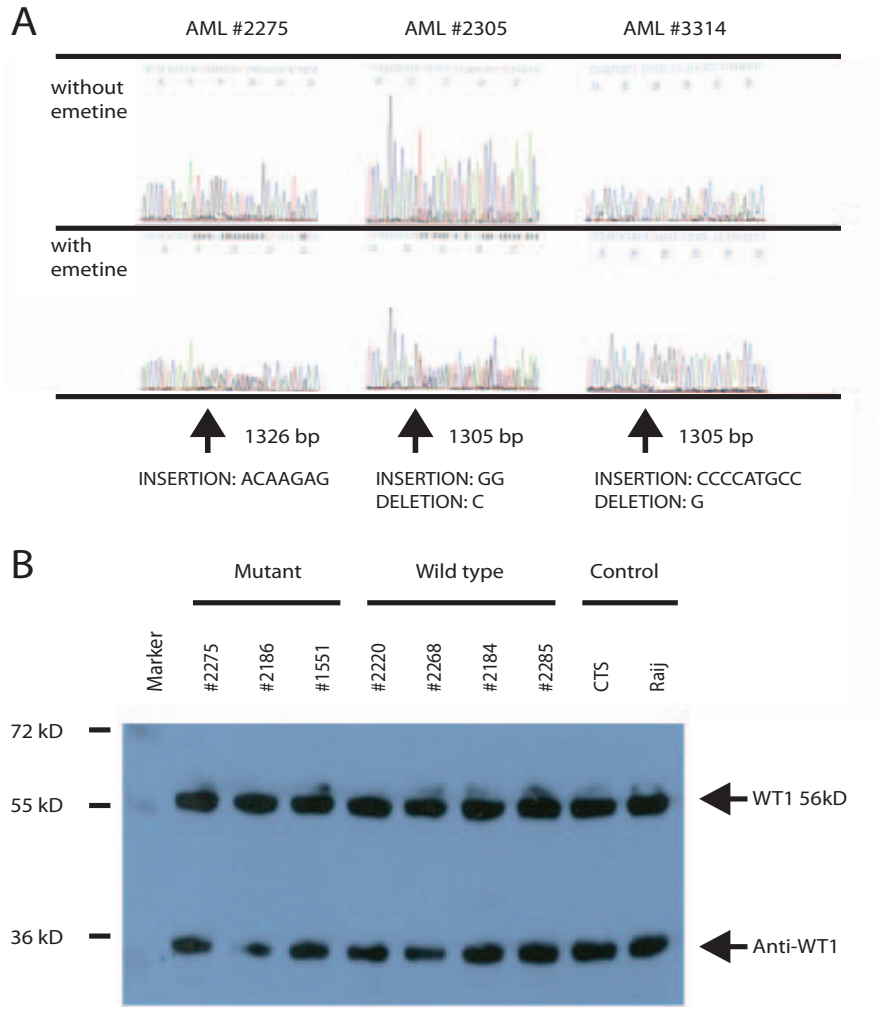


Figure 1 (A) Direct nucleotide sequence analyses of *WT1* RT-PCR products (WT1-FOR3 and WT1-REV4) with primer WT1-REV4 of three cases of AML (no. 2275, no. 2305 and no. 3314) cultured in the absence or presence of the translation inhibitor emetine. **(B)** Immunoprecipitations of *WT1*-mutant (no. 2275, no. 2186, no. 1551) and wild-type (no. 2220, no. 2268, no. 2184, no. 2285) AML cases with an N-terminal antibody against WT1. The cell lines CTS (GA (del) (nt1398) 2 bp del (detectable on gDNA and cDNA)) and Raji were included as mutant and wild-type WT1 controls, respectively.

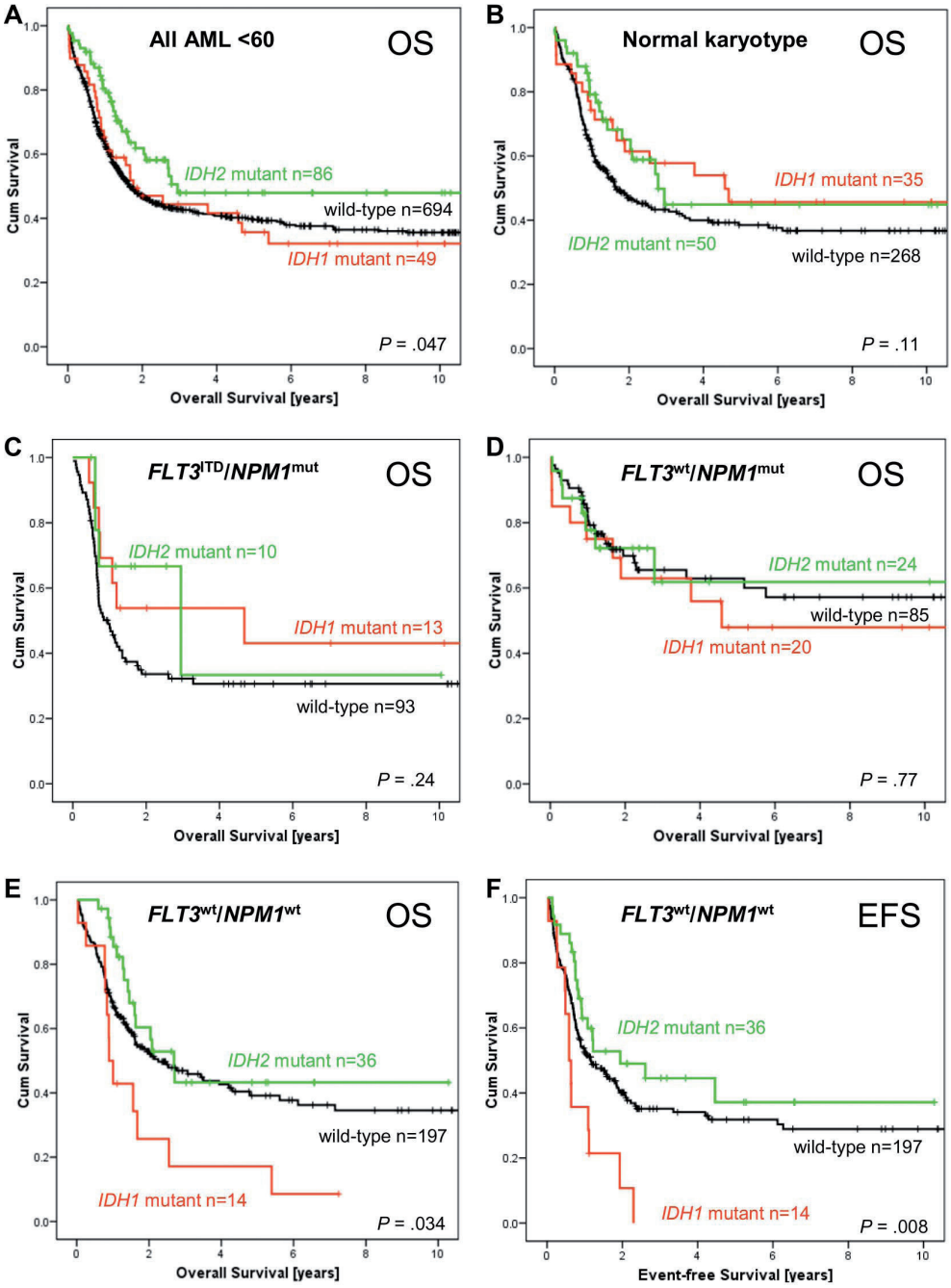


Figure 1. Survival analyses of patients of AML with or without *IDH1* and *IDH2* mutations. (A) Kaplan-Meier estimates of OS for all AML patients. (B) OS for AML patients with normal karyotypes. (C) OS for patients with intermediate-risk AML and *FLT3*^{mutant} and *NPM1*^{mutant}. (D) OS for patients with intermediate-risk AML and *FLT3*$^{wild\text{-}type}$ and *NPM1*^{mutant}. (E) OS for patients with intermediate-risk AML and *FLT3*$^{wild\text{-}type}$ and *NPM1*$^{wild\text{-}type}$. (F) EFS for patients with intermediate-risk AML and *FLT3*$^{wild\text{-}type}$ and *NPM1*$^{wild\text{-}type}$. Survival curves in red represent cases with *IDH1*^{mutant}; those in green, *IDH2*^{mutant}; and those in black, cases with *IDH1*$^{wild\text{-}type}$ and *IDH2*$^{wild\text{-}type}$, respectively. The log-rank *P* value is indicated per Kaplan-Meier analysis.