Genomic and Epigenomic Studies of Acute Myeloid Leukemia with *CEBPA* Abnormalities

Bas Wouters

Genomic and Epigenomic Studies of Acute Myeloid Leukemia with *CEBPA* Abnormalities

Genoombrede genetische en epigenetische studies naar acute myeloïde leukemie met afwijkingen in *CEBPA*

Proefschrift

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

op gezag van de rector magnificus prof.dr. S.W.J. Lamberts
en volgens het besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 20 mei 2009 om 15:45 uur

door
Bastiaan Jacob Wouters
geboren te Nijmegen

ERASMUS UNIVERSITEIT ROTTERDAM

PROMOTIECOMMISSIE

Promotor: Prof.dr. B. Löwenberg

Overige leden: Dr. J. Cools
Prof.dr. J.N.J. Philipsen
Prof.dr. R. Pieters

Co-promotor: Dr. H.R. Delwel

Aan mijn ouders

CONTENTS

CHAPTER 1

General introduction

NORMAL HEMATOPOIESIS AND LEUKEMIA

Hematopoiesis is the process of blood cell formation. In embryonic vertebrate development, the first blood cells appear in the yolk sac together with the developing vasculature (1). After various developmental stages, including an important role for the fetal liver, definitive adult hematopoiesis is established in the bone marrow (1).

Mature blood cells have a limited life span and new cells therefore need to be produced continuously throughout life. The demand for numbers and types of particular blood cells at any given time point may be dependent on external factors such as infections or oxygen need. Hematopoiesis consequently is a tightly regulated process, in which hematopoietic growth factors play a determining role. All hematopoietic cells are derived from multipotent hematopoietic stem cells (HSCs) that have the unique ability to self-renew as well as to give rise to all different types of blood cells (Figure 1) (2). In embryonic development the first HSCs are found in the aorta-gonad-mesonephros region, while in adult hematopoiesis they reside in the bone marrow.

The predominant model of hematopoiesis dictates that hematopoietic development is a highly hierarchical process. In this model, the HSC produces two distinct lines of daughter cells: a common myeloid progenitor cell and a common lymphoid progenitor cell (Figure 1). These cells in turn are exclusively able to give rise to the progeny belonging to the two lineages of blood cells (3,4). While the lymphoid precursor cells generate B cells and T cells as well as natural killer cells, the myeloid lineage produces the other white blood cells (leukocytes), i.e. granulocytes and monocytes/macrophages, as well as red blood cells (erythrocytes) and platelets (Figure 1).

Malignant transformation of normal hematopoietic progenitor cells can result in the accumulation of hematopoietic cells that have lost their physiological capacity to differentiate, proliferate and/or survive. This condition is termed leukemia. Leukemias are broadly divided into acute leukemia and chronic leukemia. Acute leukemias are a heterogeneous group of disorders of the most immature hematopoietic progenitor cells (5). They are further subdivided based on the affected lineage, i.e. acute lymphoblastic leukemia (ALL) (6) and acute myeloid, or myelogenous, leukemia (AML) (7,8). This distinction can in diagnostic practice be made based on differences in cellular morphology and expression of membrane surface antigens that can be detected with specific monoclonal antibodies. Some acute leukemias appear hard to classify as strictly myeloid or lymphoid because they exhibit characteristics, e.g. expression of surface antigens, of both lineages. These rare leukemias are sometimes termed "biphenotypic", although the exact nature and definition of these diseases are subject of dispute (9,10).
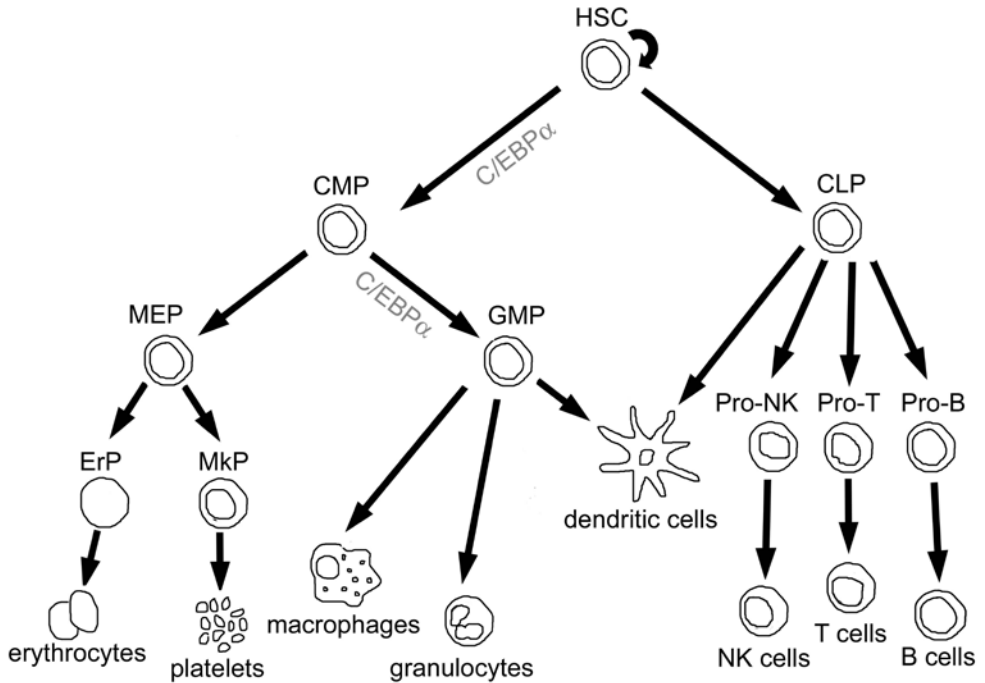
Figure 1. Schematic representation of hematopoiesis
Simplified scheme of hematopoiesis according to the current predominant working model, adapted from Reya et al (2). The hematopoeietic stem cell (HSC) gives rise to two main daughter cells, i.e. the common myeloid progenitor (CMP) and the common lymphoid progenitor (CLP), which in turn are at the basis of all blood cell types. The stages during myeloid differentiation in which C/EBPα is critical are indicated (27,33). Other abbreviations: MEP, macrophage erythrocyte precursor; ErP, erythrocyte precursor; MkP, megakaryocyte precursor; NK, natural killer. For visualization purposes, not all stages of development are depicted. For instance, the granulocyte macrophage precursor (GMP) does not directly give rise to granucolytes, but first passes several intermediate stages, including myeloblast, promyelocyte, myelocyte and metamyelocyte.

ACUTE MYELOID LEUKEMIA (AML)

The focus in this thesis is on AML. Although always characterized by the accumulation of primitive myeloid cells that have lost the ability to differentiate towards mature granulocytic or monocytic cells, AML should be considered a heterogeneous group of disorders with variable underlying abnormalities and clinical behavior including responses to treatment.

Epidemiology and clinical symptoms

The overall prevalence of AML is 3.8 cases per 100000 individuals, with up to 17.9 cases per 100000 people aged over 65 years (figures derived from USA in 2006) (7). The median age of patients at presentation is around 70 years and the disease is slightly more common in men than in women (7). The first clinical symptoms of AML, most notably infections, fatigue

and hemorrhage, are related to bone marrow failure and the absence of functional mature myeloid cells (7,8).

Diagnosis and classification

AML is not one disease, but a group of neoplasms with variable responses to treatment. Classification of the disease is thus important. Cytology has traditionally played a key role in this (8), with a particularly important role for the French-American-British classification system that was introduced in the 1970s (11). The more recent World Health Organization (WHO) classification scheme for myeloid malignancies (12,13) still recognizes the importance of morphologic analysis but additionally takes into account other information thereby aiming to provide a more clinically relevant subdivision. In 2008, the WHO classification was revised (14). The diagnosis of AML based on current WHO criteria requires more than 20% blasts in the bone marrow, which should be of myeloid origin as confirmed by cytochemistry and/or immunophenotyping (12). The criterion of 20% does not apply in case one of the balanced recurrent chromosomal translocations t(8;21), inv(16) or t(15;17) has been detected. Once the diagnosis AML has been established, the WHO now distinguishes five subtypes, including a category of AML with recurrent genetic abnormalities. This category recognizes several cytogenetic abnormalities as distinct entities, and two types of gene mutations, i.e. those in the nucleophosmin (*NPM1*) and in the CCAAT/enhancer binding protein alpha (*CEBPA*) genes, as provisional entities of AML (14). Regular further updates of the WHO classification scheme are anticipated as knowledge about underlying molecular abnormalities increases.

Prognosis

Treatment response and survival of patients diagnosed with AML is highly dependent on prognostic factors at the time of presentation of the disease. Factors that predict response to treatment include patient related characteristics, e.g. age, and disease-related factors. Of particular prognostic impact is the pre-treatment karyotype of the leukemic cells, which provides a means to divide AML into distinct subgroups with different prognosis (Table 1). The sole use of cytogenetics, however, leaves many cases of AML unclassifiable, in particular those with a "normal karyotype", i.e. without detectable cytogenetic abnormalities. These AMLs represent approximately 45% of all cases. Additional prognostic markers are now becoming available through the identification of mutations in, or altered expression of, particular genes that predict response to treatment (see below and Table 2).

Treatment

Treatment of AML consists of two phases. The first phase (induction) aims to achieve complete remission, while the second phase (post-induction) aims to prevent a relapse of the disease. The induction phase has for some decades primarily relied on combinations of cytotoxic drugs, typically a combination of an anthracycline (daunorubicin or idarubicin)

Table 1. Recurrent cytogenetic abnormalities in adult AML

Cytogenetic abnormality	Percentage of cases[#]	Gene(s) involved	Prognostic significance[#]
None (normal cytogenetics)	45	-	Intermediate
complex karyotype (more than 3 abnormalities)	11	-	Unfavorable
+8	9	-[*]	Intermediate
t(15;17)(q22;q21)	8	*PML-RARA*	Favorable
-7/7q-	8	-[*]	Unfavorable
-5/5q-	7	-[*]	Unfavorable
t(8;21)(q22;q22)	6	*AML1-ETO*	Favorable
inv(16)(p13q22) / t(16;16)(p13;q22)	5	*CBFB-MYH11*	Favorable
-Y	4	-[*]	Intermediate
abn(12p)	3	-[*]	Intermediate
+21	3	-[*]	Intermediate
abn(17p)	2	-[*]	Intermediate
del(9q)	2	-[*]	Intermediate
inv(3)(q21q26) / t(3;3)(q21;q26)	2	*EVI1*	Unfavorable
del(11q)	1	-[*]	Intermediate
t/inv(11q23)		*MLL*	Unfavorable
t(9;22)(q34;q11)	1	*BCR-ABL*	Intermediate/Unfavorable
t(6;9)(p23;q34)	1	*DEK-CAN*	Unfavorable

The abnormalities are ranked based on the percentage of AML affected.
[*] Critical gene(s) unknown.
[#] based on Mrozek et al.(84)

and cytarabine (7,8). For post-induction, either chemotherapy or bone marrow transplantation (allogeneic or autologous) are available, at least when patients are of younger age. For patients of older age, frequently defined by an age over 60, treatment options are more limited because of higher treatment-related toxicity.

MOLECULAR GENETICS OF AML

The past years have yielded increasing insights into molecular genetic abnormalities underlying AML (15-18). There are at least two major explanations for the great interest in this subject. First, molecular insights promote a better understanding of the pathobiology of the disease and may be used to design new targeted drugs. The success of imatinib, that was

designed specifically for BCR-ABL positive chronic myeloid leukemia, is a classic example (19). Second, it is hoped that the discovery of molecular abnormalities such as mutations can lead to the identification of new markers that can be used to define prognostically relevant subgroups of AML. As pointed out earlier, such novel markers are particularly welcomed for AMLs lacking prognostically informative karyotypes.

It is assumed that leukemogenesis is a multistep process that requires more than one genetic aberration. This hypothesis is based on the observation that most cases of clinical AML demonstrate multiple genetic aberrations. Furthermore, in mouse models a single mutation appears generally insufficient for overt leukemia to develop (20-23). It has been proposed that at least two types of molecular abnormalities are required for AML to develop: one that impairs normal differentiation, and another hit that provides advantage for survival or proliferation of the leukemic cells (24). Many mutations that have been uncovered in the past years can indeed be classified into either one of these two classes, based on experimental and/or theoretical bases (15,17,18). In support of the proposed two class model, in clinical AML the two types of mutation occur together more often than two mutations within the same class (7,18,25).

Abnormalities leading to enhanced survival or proliferation frequently involve mutations in signal transduction pathways (26). In normal hematopoiesis these pathways are tightly controlled. Mutations in this class mostly lead to constitutive activation, thereby rendering the cell insensitive to growth control (26). According to current knowledge, the most commonly affected gene of this type is the receptor tyrosine kinase *FLT3* (Table 2) (15,17,26).

Transcription factors are proteins that regulate gene expression by activating target genes through DNA interaction (27,28). These factors thereby play an important role in deciding which genes are activated in particular cell types and stages of differentiation. Improper functioning of transcription factors may lead to impaired differentiation as seen in leukemia (27,29). Abrogation of transcription factor activity may involve gene mutations, but can also be due to inhibition of expression or function through other mechanisms. A summary of transcription factors that are involved in subtypes of AML is given in Table 2. A transcription factor that is implicated in multiple subtypes of AML is CCAAT/enhancer binding protein alpha (C/EBPα). C/EBPα takes a central position in several studies in this thesis, and therefore its role in normal and malignant hematopoiesis will be introduced here.

THE TRANSCRIPTION FACTOR C/EBPα IN NORMAL AND MALIGNANT GRANULOPOIESIS

C/EBPα and normal granulopoiesis

C/EBPα is the founding member of the family of C/EBP proteins, a group of basic leucine zipper transcription factors (30,31). It is encoded by a one exon gene that is located on hu-

Table 2. Recurrent molecular abnormalities in adult AML

Gene	Percentage of cases#	Association with cytogenetics	Prognostic significance#	Reference**
Mutation				
Nucleophosmin (*NPM1*)	25-35	Normal	Favorable in absence of *FLT3-ITD*	(85,86)
Fms-related tyrosine kinase 3 (*FLT3*), internal tandem duplication (ITD)	28-33	Normal / t(15;17) / t(6;9)	Unfavorable	(87-89)
Neuroblastoma RAS viral (v-ras) oncogene homolog (*NRAS)*	10-15	Inv(16)	-	(90-92)
Wilms tumor 1 (*WT1*)	10	Normal	Unfavorable?	(93-95)
Tumor protein p53 (*TP53*)	<10	Complex karyotype	Unfavorable	(96,97)
Runt-related transcription factor 1 *(RUNX1; AML1)*	6-11	Normal / Trisomy 21 / others	-	(98,99)
Myeloid/lymphoid or mixed lineage leukemia (*MLL*), partial tandem duplication (PTD)	5-11% (CN-AML)	Trisomy 11 / Normal	Unfavorable	(100,101)
Fms-related tyrosine kinase 3 (*FLT3*), tyrosine kinase domain (TKD)	5-10	Normal / inv(16)	?	(89,102)
CCAAT/enhancer binding protein alpha (*CEBPA*)	5-10	Normal	Favorable	(32,50)
v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog (*KIT*)	2-8	Inv(16) / t(8;21) / Trisomy 4	Unfavorable in CBF AMLs	(25,103)
v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*)	5	Abn3q	-	(91)
Protein tyrosine phosphatase, non-receptor type 11 (*PTPN11; SHP2*)	3	-	-	(104,105)
Janus kinase 2 *(JAK2)*	2	t(8;21) / Inv(16)	-	(106)

Table 2. continued

Gene	Percentage of cases[#]	Association with cytogenetics	Prognostic significance[#]	Reference**
Overexpression				
Brain and acute leukemia gene, cytoplasmic (*BAALC*)	.~50% (CN-AML)[*]	Trisomy 8 / Normal	Unfavorable	(107,108)
Meningioma (disrupted in balanced translocation) 1 (*MN1*)	-50%[*]	Normal	Unfavorable in CN-AML?	(109)
v-ets erythroblastosis virus E26 oncogene homolog (*ERG*)	~25% (CN-AML)[*]	Complex / normal, chromosome 21	Unfavorable	(110,111)
Ecotropic viral integration site 1 (*EVI1)*	~6%[*]	Abn3q / 11q23	Unfavorable	(112,113)

CN-AML: cytogenetically normal AML
[*] In case of overexpression, the percentage is based on the cutoff used in the referenced papers. This may involve simple dichotomization (e.g. *MN1*), resulting in 50% of the cases by definition exhibiting overexpression.
[#] Based on (15-17).
** Due to space limitations, only a selected number of references is given for each abnormality.

man chromosome 19q13. Throughout this thesis, nomenclature will follow Pabst et al (32), i.e. *CEBPA* for the gene, and C/EBPα for the protein.

C/EBPα is expressed in many tissues, including liver, lung, adipose tissue, intestine and placenta (30). In the hematopoietic system its expression is restricted to the myeloid lineage (27). Experiments in conditional knockout mice have shown that C/EBPα is particularly critical in early stages of myelopoiesis, more specifically in the transition from common myeloid progenitor to granulocyte-monocyte progenitor cells (Figure 1)(33). *Cebpa* knockout animals lack mature granulocytes (34). The strong impact of C/EBPα on myeloid lineage development was further underscored by the ability of C/EBPα, when experimentally ectopically expressed, to switch committed lymphoid cells towards the myeloid lineage (35-37).

The C/EBPα protein contains a conserved basic leucine zipper (bZIP) domain at the C-terminus (Figure 2). Through this region, it binds DNA according to a scissor-grip model that involves dimerization with other C/EBPα molecules (homodimerization) or with other proteins, that may include members of the C/EBP family, e.g. C/EBPβ, or other proteins (heterodimerization) (38,39). In the N-terminus, at least two transactivation regions are present (Figure 2), while some researchers recognize three such domains. C/EBPα not only regulates gene expression by DNA binding, but also directly affects proliferation by interacting with
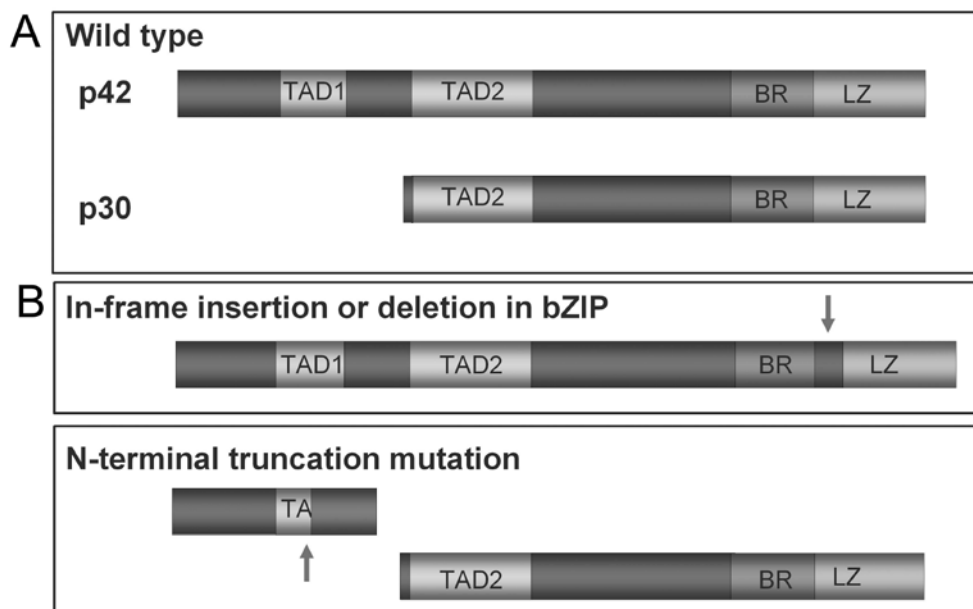
Figure 2. Schematic overview of wild type C/EBPα and effect of prototypical mutations found in human AML
A. The wild type C/EBPα protein is translated in a 42 kDa isoform (p42) and a shorter 30 kDa isoform(p30). The N-terminus contains two transactivation domains (TAD1 and TAD2). The C-terminus encodes the basic leucine zipper region, which consists of the DNA-binding basic region (BR) and the leucine zipper (LZ) involved in dimerization.
B. In AML, two main types of mutations in C/EBPα are found. Mutations in the BR or LZ are in-frame and in most cases introduce one or more amino acids. Mutations in the N-terminal part mostly cause a frameshift and premature stop codon before the second ATG from which p30 is translated. As a result, p30 can still be translated while full length p42 is not produced.

cell cycle control proteins. These include E2F (40) and possibly also p21 (41) and cdk2 and cdk4 (42).

The normal predominant isoform of C/EBPα is a protein of 42 kDa (p42). A shorter 30 kDa isoform is produced from a downstream translation initiation codon in the same open reading frame (Figure 2). While p30 C/EBPα contains a functional bZIP domain, and thus can dimerize as well as bind DNA, it lacks the first transactivation domain and cannot block cell proliferation. The ratio between translation of p42 and p30 is mediated through extracellular signals such as nutrients and growth factors, and thereby provides a possible means to regulate the balance between proliferation and differentiation (43,44).

Several target genes of C/EBPα have been identified. The *Cebpa* knockout mouse model uncovered the G-CSF receptor and IL6-receptor as critical downstream effectors, although not necessarily direct (34,45). Genes that have been implicated as direct targets of C/EBPα include myeloperoxidase (46), lactoferrin (47) and PU.1 (48).

Table 3. Mechanisms of C/EBPα abrogation in human leukemia

Mechanism	Details / Examples	References
Mutation	N-terminal frameshift mutations, C-terminal bZIP insertions; Somatic or germ-line	(32,49,50)
Repression of transcription		
- by oncogenes	*AML1-ETO*	(114)
- by DNA promoter hypermethylation	proximal promoter, distal enhancer	(115,116)
Repression of translation by oncogenes	*AML1-MDS-EVI1*; *BCR-ABL*	(117,118)
Posttranslational inhibition		
- phosphorylation	*FLT3(-ITD)* on serine 21	(119)
Impaired DNA binding	c-Jun	(120)
Overexpression		
- translocation	IGH translocation	(121,122)
Unknown/downregulation	*PML-RARA*	(123,124)

CEBPA mutations in acute myeloid leukemia

Mutations in the *CEBPA* gene in AML were first discovered in 2001 (32) and were subsequently validated by other research groups (49-54). Several concepts have emerged from these studies (43,55,56). *CEBPA* mutations are primarily associated with AMLs without recurrent cytogenetic aberrations. There is a non-random distribution of mutations in the gene, with hot spots in the N-terminus and in the bZIP region, respectively (Figure 2B). Furthermore, the majority of *CEBPA* mutant AMLs presents with two individual mutations, which are typically on different alleles (57). In most studies to date, *CEBPA* mutations were associated with relatively favorable treatment responses (49-51,53,58,59). Germ-line mutations in *CEBPA* are found in familial AML and predispose individuals to development of AML later in life (60-62).

Besides mutations in the *CEBPA* gene itself, other mechanisms of abrogation of its expression or function in leukemia have now been identified (55). These are summarized in Table 3. Yet additional physiological (post-translational) mechanisms of regulation of C/EBPα functioning have recently been recognized, such as phosphorylation at sites other than serine 21 and sumoylation (63). It remains to be determined whether impairment at these levels can play a role in leukemogenesis as well.

GENOME-WIDE APPROACHES TO THE STUDY OF AML

Molecular genetics have thus provided insights into questions related to classification, prognosis and pathogenesis of AML in the past years. The publication of the complete human genome sequence in 2001 (64,65) marks the increasing focus in the field on genome-wide analyses. Instead of classical gene based approaches, several new techniques allow researchers to simultaneously investigate changes in normal and malignant cells on a genome-wide basis. Many of these assays, that frequently involve DNA microarray technology (66-68), have been applied to the study of leukemia in an early stage. The early application in hematological research may at least partly be explained by the relatively easy accessibility and methods available for purification of blood cells.

For most of the specific questions addressed in this thesis, at least one of these newly developed genome-wide techniques was applied. A particularly central role in this thesis is taken by gene expression profiling (GEP), a technique to simultaneously measure the levels of thousands of mRNA transcripts using DNA microarray chips. It was the first of the microarray based assays that was widely and successfully applied in cancer in general, and in AML in particular (69-72). The experience with GEP in the field of AML will be reviewed in the next chapter.

Amplifications or deletions of chromosomal regions may result in gains or losses of critical AML-related genes. Traditionally, cytogenetic analysis has been applied to study these alterations. While valuable to detect chromosomal translocations, this technique has a limited resolution. To study smaller alterations, DNA microarray platforms involving comparative genomic hybridization have been developed. Alternatively, array-based genotyping of single nucleotide polymorphisms (SNPs) can be used to infer gains or losses. The latter technique is also capable of detecting chromosomal regions showing loss of heterozygosity without copy number changes as the result of mitotic recombination, a phenomenon that can be referred to as (segmental) uniparental disomy (UPD). UPD may be functionally relevant in disease development, as it can serve as a second hit after one allele has been mutated or deleted (73). UPD appears relatively frequent in AML, while copy number variations are less common, for instance when compared to ALL (74-78).

In addition to changes in the DNA sequence itself, epigenetic mechanisms are involved in normal control of gene expression and may be involved in cancer when deregulated (79). The accessibility of DNA for transcription is controlled by modifications to DNA binding proteins (histones) as well as to the DNA itself (79). Hypermethylation of DNA may occur on cytosine residues in CG (frequently referred to as CpG) dinucleotides (80). Hypermethylation of CpG-rich elements in gene promoter regions ("CpG islands") is associated with gene silencing (80,81). Several techniques are now available to study epigenetic marks, including DNA methylation, on a genome-wide level (82). In this thesis, we applied one of these techniques to specifically examine a subset of leukemia cases, the rationale being that these

leukemias had previously been found to exhibit extensive promoter hypermethylation of the *CEBPA* gene.

While DNA microarrays can be used to directly examine the abundance of mRNA or methylation levels or to determine genotypes, they can also be incorporated into other experimental procedures. One such approach is chromatin immunoprecipitation (ChIP), the analysis of DNA fragments that have been immunoprecipitated using an antibody against a particular DNA binding protein. The analysis of such immunoprecipitated fragments on DNA microarrays specifically dedicated to promoter regions across the genome (ChIP-chip) has proven to be an interesting tool to study target genes of transcription factors (83).

SCOPE AND AIMS OF THE THESIS

The introduction of DNA microarray based genome-wide analysis techniques has created new opportunities to resolve the large heterogeneity of AML. The studies in this thesis are centered around the use of several of these techniques in AML, with a strong focus on previously established and new subtypes of AML associated with *CEBPA* abnormalities.

The thesis presents work divided into four sections. The first section deals with the use of GEP to predict subtypes in a representative cohort of AML. This question is addressed in **chapter 3,** which follows a review of the experience with GEP in the study of AML (**chapter 2**).

The second section (chapters 4, 5 and 6) centers on the question whether GEP in combination with DNA methylation profiling can be used to identify previously unrecognized subgroups of AML, and whether these analyses can provide insights into pathobiological mechanisms underlying the particular disease subtype. Through mouse modeling experiments, **chapter 4** introduces a novel mechanism of inhibition of C/EBPα that appears to be largely restricted to a particular subtype of AML. This subtype is further investigated in the context of clinical AML cases through GEP in **chapter 5**. Genome-wide DNA methylation analysis is used in **chapter 6** to investigate what the putative role of epigenetic silencing of *CEBPA* is in the latter leukemia subtype.

The third section of the thesis (chapters 7, 8 and 9) is concerned with AML with mutations in the *CEBPA* gene. While most *CEBPA* mutations are found in the N-terminus or in the bZIP domain (Figure 2B), other sequence variations have been described as well. In **chapter 7** the importance of one of these other variations, an in-frame insertion in the second transactivation domain, is assessed by comparing leukemic samples to a collection of non-leukemic samples using a high-throughput mutation screening approach involving denaturing high performance liquid chromatography. In **chapter 8,** we investigate by microarray based genotyping whether homozygosity of *CEBPA* mutations is associated with physical loss of the wild type allele, or with segmental UPD. And in **chapter 9**, we evaluate the biologic and

clinical impact of two mutations in *CEBPA* in direct comparison with a single heterozygous mutation in the gene.

Finally, the fourth section (**chapter 10**) presents experiments regarding the use of ChIP-chip to identify direct target genes of C/EBPα in an inducible myeloid cell line model. These findings are compared to results obtained in cells that express a bZIP mutant C/EBPα protein.

Chapter 11 provides a summary of the findings reported in the thesis and a general discussion of the results and their relationship to future research.

REFERENCES

1. Dzierzak E, Speck NA. Of lineage and legacy: the development of mammalian hematopoietic stem cells. Nat Immunol 2008;9(2):129-36.

2. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature 2001;414(6859):105-11.

3. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature 2000;404(6774):193-7.

4. Kondo M, Weissman IL, Akashi K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell 1997;91(5):661-72.

5. Gilliland DG, Tallman MS. Focus on acute leukemias. Cancer Cell 2002;1(5):417-20.

6. Pui CH, Robison LL, Look AT. Acute lymphoblastic leukaemia. Lancet 2008;371(9617):1030-43.

7. Estey E, Dohner H. Acute myeloid leukaemia. Lancet 2006;368(9550):1894-907.

8. Lowenberg B, Downing JR, Burnett A. Acute myeloid leukemia. N Engl J Med 1999;341(14):1051-62.

9. Bene MC, Castoldi G, Knapp W, Ludwig WD, Matutes E, Orfao A, et al. Proposals for the immunological classification of acute leukemias. European Group for the Immunological Characterization of Leukemias (EGIL). Leukemia 1995;9(10):1783-6.

10. Bagg A. Lineage ambiguity, infidelity, and promiscuity in immunophenotypically complex acute leukemias: genetic and morphologic correlates. Am J Clin Pathol 2007;128(4):545-8.

11. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, et al. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. Br J Haematol 1976;33(4):451-8.

12. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. Blood 2002;100(7):2292-302.

13. Harris NL, Jaffe ES, Diebold J, Flandrin G, Muller-Hermelink HK, Vardiman J, et al. World Health Organization classification of neoplastic diseases of the hematopoietic and lymphoid tissues: report of the Clinical Advisory Committee meeting-Airlie House, Virginia, November 1997. J Clin Oncol 1999;17(12):3835-49.

14. Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, et al., editors. WHO classification of tumours of haematopoietic and lymphoid tissues: IARC: Lyon; 2008.

15. Renneville A, Roumier C, Biggio V, Nibourel O, Boissel N, Fenaux P, et al. Cooperating gene mutations in acute myeloid leukemia: a review of the literature. Leukemia 2008;22(5):915-31.

16. Mrozek K, Marcucci G, Paschka P, Whitman SP, Bloomfield CD. Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification? Blood 2007;109(2):431-48.

17. Dohner K, Dohner H. Molecular characterization of acute myeloid leukemia. Haematologica 2008;93(7):976-82.

18. Frohling S, Scholl C, Gilliland DG, Levine RL. Genetics of myeloid malignancies: pathogenetic and clinical implications. J Clin Oncol 2005;23(26):6285-95.

19. Tallman MS, Gilliland DG, Rowe JM. Drug therapy for acute myeloid leukemia. Blood 2005;106(4):1154-63.

20. Castilla LH, Garrett L, Adya N, Orlic D, Dutra A, Anderson S, et al. The fusion gene Cbfb-MYH11 blocks myeloid differentiation and predisposes mice to acute myelomonocytic leukaemia. Nat Genet 1999;23(2):144-6.

21. Yuan Y, Zhou L, Miyamoto T, Iwasaki H, Harakawa N, Hetherington CJ, 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 2001;98(18):10398-403.

22. Chan IT, Kutok JL, Williams IR, Cohen S, Kelly L, Shigematsu H, et al. Conditional expression of oncogenic K-ras from its endogenous promoter induces a myeloproliferative disease. J Clin Invest 2004;113(4):528-38.

23. Kelly LM, Liu Q, Kutok JL, Williams IR, Boulton CL, Gilliland DG. FLT3 internal tandem duplication mutations associated with human acute myeloid leukemias induce myeloproliferative disease in a murine bone marrow transplant model. Blood 2002;99(1):310-8.

24. Dash A, Gilliland DG. Molecular genetics of acute myeloid leukaemia. Best Pract Res Clin Haematol 2001;14(1):49-64.

25. Care RS, Valk PJ, Goodeve AC, Abu-Duhier FM, Geertsma-Kleinekoort WM, Wilson GA, et al. Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. Br J Haematol 2003;121(5):775-7.

26. Van Etten RA. Aberrant cytokine signaling in leukemia. Oncogene 2007;26(47):6738-49.

27. Rosenbauer F, Tenen DG. Transcription factors in myeloid development: balancing differentiation with transformation. Nat Rev Immunol 2007;7(2):105-17.

28. Tenen DG, Hromas R, Licht JD, Zhang DE. Transcription factors, normal myeloid development, and leukemia. Blood 1997;90(2):489-519.

29. Tenen DG. Disruption of differentiation in human cancer: AML shows the way. Nat Rev Cancer 2003;3(2):89-101.

30. Ramji DP, Foka P. CCAAT/enhancer-binding proteins: structure, function and regulation. Biochem J 2002;365(Pt 3):561-75.

31. Landschulz WH, Johnson PF, Adashi EY, Graves BJ, McKnight SL. Isolation of a recombinant copy of the gene encoding C/EBP. Genes Dev 1988;2(7):786-800.

32. Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, et al. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. Nat Genet 2001;27(3):263-70.

33. Zhang P, Iwasaki-Arai J, Iwasaki H, Fenyus ML, Dayaram T, Owens BM, et al. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. Immunity 2004;21(6):853-63.

34. Zhang DE, Zhang P, Wang ND, Hetherington CJ, Darlington GJ, Tenen DG. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc Natl Acad Sci U S A 1997;94(2):569-74.

35. Xie H, Ye M, Feng R, Graf T. Stepwise reprogramming of B cells into macrophages. Cell 2004;117(5):663-76.

36. Laiosa CV, Stadtfeld M, Xie H, de Andres-Aguayo L, Graf T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. Immunity 2006;25(5):731-44.

37. Fukuchi Y, Shibata F, Ito M, Goto-Koshino Y, Sotomaru Y, Ito M, et al. Comprehensive analysis of myeloid lineage conversion using mice expressing an inducible form of C/EBPalpha. Embo J 2006.

38. Landschulz WH, Johnson PF, McKnight SL. The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science 1989;243(4899):1681-8.

39. Agre P, Johnson PF, McKnight SL. Cognate DNA binding specificity retained after leucine zipper exchange between GCN4 and C/EBP. Science 1989;246(4932):922-6.

40. Porse BT, Pedersen TA, Xu X, Lindberg B, Wewer UM, Friis-Hansen L, et al. E2F repression by C/EBPalpha is required for adipogenesis and granulopoiesis in vivo. Cell 2001;107(2):247-58.

41. Timchenko NA, Harris TE, Wilde M, Bilyeu TA, Burgess-Beusse BL, Finegold MJ, et al. CCAAT/ enhancer binding protein alpha regulates p21 protein and hepatocyte proliferation in newborn mice. Mol Cell Biol 1997;17(12):7353-61.

42. Wang H, Iakova P, Wilde M, Welm A, Goode T, Roesler WJ, et al. C/EBPalpha arrests cell pro-liferation through direct inhibition of Cdk2 and Cdk4. Mol Cell 2001;8(4):817-28.

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

44. Calkhoven CF, Muller C, Leutz A. Translational control of C/EBPalpha and C/EBPbeta isoform expression. Genes Dev 2000;14(15):1920-32.

45. Zhang P, Iwama A, Datta MW, Darlington GJ, Link DC, Tenen DG. Upregulation of interleukin 6 and granulocyte colony-stimulating factor receptors by transcription factor CCAAT enhancer binding protein alpha (C/EBP alpha) is critical for granulopoiesis. J Exp Med 1998;188(6):1173-84.

46. Ford AM, Bennett CA, Healy LE, Towatari M, Greaves MF, Enver T. Regulation of the myeloper-oxidase enhancer binding proteins Pu1, C-EBP alpha, -beta, and -delta during granulocyte-lineage specification. Proc Natl Acad Sci U S A 1996;93(20):10838-43.

47. Khanna-Gupta A, Zibello T, Sun H, Gaines P, Berliner N. Chromatin immunoprecipitation (ChIP) studies indicate a role for CCAAT enhancer binding proteins alpha and epsilon (C/EBP alpha and C/EBP epsilon) and CDP/cut in myeloid maturation-induced lactoferrin gene expres-sion. Blood 2003;101(9):3460-8.

48. Kummalue T, Friedman AD. Cross-talk between regulators of myeloid development: C/EBPal-pha binds and activates the promoter of the PU.1 gene. J Leukoc Biol 2003;74(3):464-70.

49. Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, Meijer J, van Oosterhoud S, van Putten WL, Valk PJ, et al. Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML. Hematol J 2003;4(1):31-40.

50. Frohling S, Schlenk RF, Stolze I, Bihlmayr J, Benner A, Kreitmeier 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 2004;22(4):624-33.

51. Preudhomme C, Sagot C, Boissel N, Cayuela JM, Tigaud I, de Botton S, 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 2002;100(8):2717-23.

52. Gombart AF, Hofmann WK, Kawano S, Takeuchi S, Krug U, Kwok SH, et al. Mutations in the gene encoding the transcription factor CCAAT/enhancer binding protein alpha in myelodysplastic syndromes and acute myeloid leukemias. Blood 2002;99(4):1332-40.

53. Snaddon J, Smith ML, Neat M, Cambal-Parrales M, Dixon-McIver A, Arch R, et al. Mutations of CEBPA in acute myeloid leukemia FAB types M1 and M2. Genes Chromosomes Cancer 2003;37(1):72-8.

54. Lin LI, Chen CY, Lin DT, Tsay W, Tang JL, Yeh YC, et al. Characterization of CEBPA mutations in acute myeloid leukemia: most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells. Clin Cancer Res 2005;11(4):1372-9.

55. Mueller BU, Pabst T. C/EBPalpha and the pathophysiology of acute myeloid leukemia. Curr Opin Hematol 2006;13(1):7-14.

56. Leroy H, Roumier C, Huyghe P, Biggio V, Fenaux P, Preudhomme C. CEBPA point mutations in hematological malignancies. Leukemia 2005;19(3):329-34.

57. Pabst T, Mueller BU. Transcriptional dysregulation during myeloid transformation in AML. Oncogene 2007;26(47):6829-37.

58. Marcucci G, Maharry K, Radmacher MD, Mrozek K, Vukosavljevic T, Paschka P, et al. Prognostic Significance of, and Gene and MicroRNA Expression Signatures Associated With, CEBPA Mutations in Cytogenetically Normal Acute Myeloid Leukemia With High-Risk Molecular Features: A Cancer and Leukemia Group B Study. J Clin Oncol 2008.

59. Schlenk RF, Dohner K, Krauter J, Frohling S, Corbacioglu A, Bullinger L, et al. Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia. N Engl J Med 2008;358(18):1909-18.

60. Smith ML, Cavenagh JD, Lister TA, Fitzgibbon J. Mutation of CEBPA in familial acute myeloid leukemia. N Engl J Med 2004;351(23):2403-7.

61. Sellick GS, Spendlove HE, Catovsky D, Pritchard-Jones K, Houlston RS. Further evidence that germline CEBPA mutations cause dominant inheritance of acute myeloid leukaemia. Leukemia 2005;19(7):1276-8.

62. Pabst T, Eyholzer M, Haefliger S, Schardt J, Mueller BU. Somatic CEBPA Mutations Are a Frequent Second Event in Families With Germline CEBPA Mutations and Familial Acute Myeloid Leukemia. J Clin Oncol 2008.

63. Khanna-Gupta A. Sumoylation and the function of CCAAT enhancer binding protein alpha (C/EBP alpha). Blood Cells Mol Dis 2008;41(1):77-81.

64. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science 2001;291(5507):1304-51.

65. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860-921.

66. Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. Nat Rev Genet 2006;7(3):200-10.

67. Ebert BL, Golub TR. Genomic approaches to hematologic malignancies. Blood 2004;104(4):923-32.

68. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. Nat Genet 1999;21(1 Suppl):33-7.

69. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286(5439):531-7.

70. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

71. Ross ME, Mahfouz R, Onciu M, Liu HC, Zhou X, Song G, et al. Gene expression profiling of pediatric acute myelogenous leukemia. Blood 2004;104(12):3679-87.

72. Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004;350(16):1605-16.

73. Young BD, Debernardi S, Lillington DM, Skoulakis S, Chaplin T, Foot NJ, et al. A role for mitotic recombination in leukemogenesis. Adv Enzyme Regul 2006;46:90-7.

74. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature 2007;446(7137):758-64.

75. Raghavan M, Lillington DM, Skoulakis S, Debernardi S, Chaplin T, Foot NJ, et al. Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. Cancer Res 2005;65(2):375-8.

76. Fitzgibbon J, Smith LL, Raghavan M, Smith ML, Debernardi S, Skoulakis S, et al. Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. Cancer Res 2005;65(20):9152-4.

77. Raghavan M, Smith LL, Lillington DM, Chaplin T, Kakkas I, Molloy G, et al. Segmental uniparental disomy is a commonly acquired genetic event in relapsed acute myeloid leukemia. Blood 2008;112(3):814-21.

78. Rucker FG, Bullinger L, Schwaenen C, Lipka DB, Wessendorf S, Frohling S, et al. Disclosure of candidate genes in acute myeloid leukemia with complex karyotypes using microarray-based molecular characterization. J Clin Oncol 2006;24(24):3887-94.

79. Jones PA, Baylin SB. The epigenomics of cancer. Cell 2007;128(4):683-92.

80. Weber M, Schubeler D. Genomic patterns of DNA methylation: targets and function of an epigenetic mark. Curr Opin Cell Biol 2007;19(3):273-80.

81. Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med 2003;349(21):2042-54.

82. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet 2007;8(4):286-98.

83. Wu J, Smith LT, Plass C, Huang TH. ChIP-chip comes of age for genome-wide functional analysis. Cancer Res 2006;66(14):6899-902.

84. Mrozek K, Heerema NA, Bloomfield CD. Cytogenetics in acute leukemia. Blood Rev 2004;18(2):115-36.

85. Verhaak RG, Goudswaard CS, van Putten W, Bijl MA, Sanders MA, Hugens W, 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 2005;106(12):3747-54.

86. Dohner K, Schlenk RF, Habdank M, Scholl C, Rucker FG, Corbacioglu A, et al. Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. Blood 2005;106(12):3740-6.

87. Frohling S, Schlenk RF, Breitruck J, Benner A, Kreitmeier S, Tobis K, et al. Prognostic significance of activating FLT3 mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML Study Group Ulm. Blood 2002;100(13):4372-80.

88. Nakao M, Yokota S, Iwai T, Kaneko H, Horiike S, Kashima K, et al. Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. Leukemia 1996;10(12):1911-8.

89. 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 2005;19(8):1345-9.

90. Bacher U, Haferlach T, Schoch C, Kern W, Schnittger S. Implications of NRAS mutations in AML: a study of 2502 patients. Blood 2006;107(10):3847-53.

91. Bowen DT, Frew ME, Hills R, Gale RE, Wheatley K, Groves MJ, 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 2005;106(6):2113-9.

92. Valk PJ, Bowen DT, Frew ME, Goodeve AC, Lowenberg B, Reilly JT. Second hit mutations in the RTK/RAS signaling pathway in acute myeloid leukemia with inv(16). Haematologica 2004;89(1):106.

93. King-Underwood L, Pritchard-Jones K. Wilms' tumor (WT1) gene mutations occur mainly in acute myeloid leukemia and may confer drug resistance. Blood 1998;91(8):2961-8.

94. Paschka P, Marcucci G, Ruppert AS, Whitman SP, Mrozek K, Maharry K, et al. Wilms' tumor 1 gene mutations independently predict poor outcome in adults with cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. J Clin Oncol 2008;26(28):4595-602.

95. Virappane P, Gale R, Hills R, Kakkas I, Summers K, Stevens J, et al. Mutation of the Wilms' Tumor 1 Gene Is a Poor Prognostic Factor Associated With Chemotherapy Resistance in Normal Karyotype Acute Myeloid Leukemia: The United Kingdom Medical Research Council Adult Leukaemia Working Party. J Clin Oncol 2008.

96. Fenaux P, Jonveaux P, Quiquandon I, Lai JL, Pignon JM, Loucheux-Lefebvre MH, et al. P53 gene mutations in acute myeloid leukemia with 17p monosomy. Blood 1991;78(7):1652-7.

97. Slingerland JM, Minden MD, Benchimol S. Mutation of the p53 gene in human acute myelogenous leukemia. Blood 1991;77(7):1500-7.

98. Osato M, Asou N, Abdalla E, Hoshino K, Yamasaki H, Okubo T, et al. Biallelic and heterozygous point mutations in the runt domain of the AML1/PEBP2alphaB gene associated with myeloblastic leukemias. Blood 1999;93(6):1817-24.

99. Preudhomme C, Warot-Loze D, Roumier C, Grardel-Duflos N, Garand R, Lai JL, et al. High incidence of biallelic point mutations in the Runt domain of the AML1/PEBP2 alpha B gene in Mo acute myeloid leukemia and in myeloid malignancies with acquired trisomy 21. Blood 2000;96(8):2862-9.

100. Dohner K, Tobis K, Ulrich R, Frohling S, Benner A, Schlenk RF, et al. Prognostic significance of partial tandem duplications of the MLL gene in adult patients 16 to 60 years old with acute myeloid leukemia and normal cytogenetics: a study of the Acute Myeloid Leukemia Study Group Ulm. J Clin Oncol 2002;20(15):3254-61.

101. Caligiuri MA, Strout MP, Oberkircher AR, Yu F, de la Chapelle A, Bloomfield CD. 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 1997;94(8):3899-902.

102. Mead AJ, Linch DC, Hills RK, Wheatley K, Burnett AK, Gale RE. FLT3 tyrosine kinase domain mutations are biologically distinct from and have a significantly more favorable prognosis than FLT3 internal tandem duplications in patients with acute myeloid leukemia. Blood 2007;110(4):1262-70.

103. Beghini A, Peterlongo P, Ripamonti CB, Larizza L, Cairoli R, Morra E, et al. C-kit mutations in core binding factor leukemias. Blood 2000;95(2):726-7.

104. Nomdedeu J, Carricondo MT, Lasa A, Perea G, Aventin A, Sierra J. Low frequency of exon 3 PTPN11 mutations in adult de novo acute myeloid leukemia. Analysis of a consecutive series of 173 patients. Haematologica 2005;90(3):412-3.

105. Johan MF, Bowen DT, Frew ME, Goodeve AC, Wilson GA, Peake IR, et al. Mutations in PTPN11 are uncommon in adult myelodysplastic syndromes and acute myeloid leukaemia. Br J Haematol 2004;124(6):843-4.

106. Levine RL, Loriaux M, Huntly BJ, Loh ML, Beran M, Stoffregen E, 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 2005;106(10):3377-9.

107. Tanner SM, Austin JL, Leone G, Rush LJ, Plass C, Heinonen K, et al. BAALC, the human member of a novel mammalian neuroectoderm gene lineage, is implicated in hematopoiesis and acute leukemia. Proc Natl Acad Sci U S A 2001;98(24):13901-6.

108. Baldus CD, Tanner SM, Ruppert AS, Whitman SP, Archer KJ, Marcucci G, et al. BAALC expression predicts clinical outcome of de novo acute myeloid leukemia patients with normal cytogenetics: a Cancer and Leukemia Group B Study. Blood 2003;102(5):1613-8.

109. Heuser M, Beutel G, Krauter J, Dohner K, von Neuhoff N, Schlegelberger B, et al. High meningioma 1 (MN1) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. Blood 2006;108(12):3898-905.

110. Mrozek K, Heinonen K, Theil KS, Bloomfield CD. Spectral karyotyping in patients with acute myeloid leukemia and a complex karyotype shows hidden aberrations, including recurrent over-representation of 21q, 11q, and 22q. Genes Chromosomes Cancer 2002;34(2):137-53.

111. Baldus CD, Liyanarachchi S, Mrozek K, Auer H, Tanner SM, Guimond M, et al. Acute myeloid leukemia with complex karyotypes and abnormal chromosome 21: Amplification discloses over-expression of APP, ETS2, and ERG genes. Proc Natl Acad Sci U S A 2004;101(11):3915-20.

112. Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, van Putten WL, Valk PJ, van der Poel-van de Luytgaarde S, Hack R, et al. High EVI1 expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients. Blood 2003;101(3):837-45.

113. Lugthart S, van Drunen E, van Norden Y, van Hoven A, Erpelinck CA, Valk PJ, et al. High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression and chromosome 3q26 abnormalities underestimated. Blood 2008;111(8):4329-37.

114. Pabst T, Mueller BU, Harakawa N, Schoch C, Haferlach T, Behre G, et al. AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. Nat Med 2001;7(4):444-51.

115. Hackanson B, Bennett KL, Brena RM, Jiang J, Claus R, Chen SS, et al. Epigenetic modification of CCAAT/enhancer binding protein alpha expression in acute myeloid leukemia. Cancer Res 2008;68(9):3142-51.

116. Chim CS, Wong AS, Kwong YL. Infrequent hypermethylation of CEBPA promotor in acute myeloid leukaemia. Br J Haematol 2002;119(4):988-90.

117. Perrotti D, Cesi V, Trotta R, Guerzoni C, Santilli G, Campbell K, et al. BCR-ABL suppresses C/EBPalpha expression through inhibitory action of hnRNP E2. Nat Genet 2002;30(1):48-58.

118. Helbling D, Mueller BU, Timchenko NA, Hagemeijer A, Jotterand M, Meyer-Monard S, et al. The leukemic fusion gene AML1-MDS1-EVI1 suppresses CEBPA in acute myeloid leukemia by activation of Calreticulin. Proc Natl Acad Sci U S A 2004;101(36):13312-7.

119. Radomska HS, Basseres DS, Zheng R, Zhang P, Dayaram T, Yamamoto Y, et al. Block of C/EBP alpha function by phosphorylation in acute myeloid leukemia with FLT3 activating mutations. J Exp Med 2006;203(2):371-81.

120. Rangatia J, Vangala RK, Singh SM, Peer Zada AA, Elsasser A, Kohlmann A, et al. Elevated c-Jun expression in acute myeloid leukemias inhibits C/EBPalpha DNA binding via leucine zipper domain interaction. Oncogene 2003;22(30):4760-4.

121. Chapiro E, Russell L, Radford-Weiss I, Bastard C, Lessard M, Struski S, et al. Overexpression of CEBPA resulting from the translocation t(14;19)(q32;q13) of human precursor B acute lympho-blastic leukemia. Blood 2006;108(10):3560-3.

122. Akasaka T, Balasas T, Russell LJ, Sugimoto KJ, Majid A, Walewska R, et al. Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). Blood 2007;109(8):3451-61.

123. Truong BT, Lee YJ, Lodie TA, Park DJ, Perrotti D, Watanabe N, et al. CCAAT/Enhancer binding proteins repress the leukemic phenotype of acute myeloid leukemia. Blood 2003;101(3):1141-8.

124. Lee YJ, Jones LC, Timchenko NA, Perrotti D, Tenen DG, Kogan SC. CCAAT/enhancer binding proteins alpha and epsilon cooperate with all-trans retinoic acid in therapy but differ in their antileukemic activities. Blood 2006;108(7):2416-9.

CHAPTER

2

A decade of genome-wide gene expression profiling in acute myeloid leukemia: flashback and prospects

Bas J. Wouters[1], Bob Löwenberg[1] and Ruud Delwel[1]

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

ABSTRACT

The past decade has shown a marked increase in the use of high-throughput assays in clinical research into human cancer, including acute myeloid leukemia (AML). In particular, genome-wide gene expression profiling (GEP) using DNA microarrays has been extensively used for improved understanding of the diagnosis, prognosis and pathobiology of this heterogeneous disease. This review discusses the progress that has been made, it places the technological limitations in perspective, and highlights promising future avenues.

INTRODUCTION

Acute myeloid leukemia (AML) is characterized by a maturation block and accumulation of myeloid progenitor cells (1,2). Clinically, it has been recognized as a heterogeneous disorder (1,2). Laboratory support for that notion has come from various directions. Chromosomal abnormalities and gene mutations are common in AML, many of which are apparent in particular subtypes (3,4).

Classification of AML subtypes is clinically relevant, as particular abnormalities are associated with distinct clinical behavior (2). For instance, recurring reciprocal translocations t(15;17)(q22;q21), t(8;21)(q22;q22) or inv(16)(p13q22)/t(16;16)(p13;q22) (further abbreviated as t(15;17), t(8;21) and inv(16), respectively) predict favorable prognosis, whereas other chromosomal aberrations are associated with inferior outcome (2). Likewise, sequence mutations in certain genes are associated with either favorable or unfavorable response to treatment (4).

While insight into cytogenetic and genetic aberrations is invaluable for diagnosis, it may also allow for better understanding of the pathobiology. Furthermore, it may enable the development and application of specific treatment modalities targeted to underlying oncogenic abnormalities. The efficacy of such drugs as all-trans retinoic acid for the treatment of t(15;17) AML and imatinib for BCR-ABL-positive chronic myeloid leukemia offer well established examples (5).

In spite of great progress, much of the heterogeneity of AML remains to be resolved. A significant proportion of human AML appears cytogenetically and genetically normal, which implies that the underlying molecular abnormalities are still unknown. Furthermore, it is likely that in AMLs carrying recognizable aberrations additional hits await to be uncovered, as one lesion usually does not appear sufficient for full leukemic transformation (6).

In recent years, DNA microarrays, together with the availability of the complete nucleotide sequence of the human genome, have spurred the search for abnormalities in cancer, including AML (7). The accessibility of these tools allows assessment of abnormalities and variations on a genome wide basis, covering various molecular levels (8). Gene expression profiling (GEP) is one of these technologies, in which DNA microarrays containing cDNAs or oligonucleotide probes are used to simultaneously measure levels of many different mRNA transcripts (7-10). In an early landmark study in 1999, Golub and colleagues were able to use GEP to discriminate a collection of AML from acute lymphoblastic leukemia (ALL) specimens (11). Their study suggested three important potential applications of GEP: class discovery, class prediction and class comparison. *Class discovery* refers to the identification of new subgroups while for *class prediction* one uses gene expression data to predict already defined subgroups. These two applications therefore have diagnostic implications. The third proposed application, *class comparison*, refers to the identification of genes that are deregulated in certain subgroups, and may address biological questions. Given these pro-

posed possibilities, investigators have used DNA microarrays to investigate gene expression in clinical AML samples in the past years (12-14).

Here, some ten years later, we shall discuss the results of a decade of experience with GEP as regards diagnosis, prognosis and biology of AML in the context of clinical studies. Is it possible to identify new clinically relevant subgroups of AML using GEP? Does GEP deserve a place in clinical diagnosis? Does GEP allow prediction of prognosis? And is it possible to extract new insights into the pathobiology of AML from clinical GEP data?

CAN GEP IDENTIFY NEW SUBGROUPS OF AML?

A straightforward way of using GEP is to compare expression profiles between cases of AML and to search for similarities and differences. In an *unsupervised* approach this is done in an unbiased way, i.e. without the use of external information such as mutations or karyotypic subtypes. This procedure is therefore representative of *class discovery*. The grouping of cases according to similar gene expression signatures is often referred to as *clustering* (15,16). The underlying assumption is that cases with the same gene expression profiles may carry the same genetic abnormality. Support for this hypothesis has come from observations, now confirmed by several research groups, that particular cytogenetic AML subtypes – e.g., AMLs with t(8;21), t(15;17) and inv(16) – each share distinctive GEP profiles (Figure 1A) (17-23). Likewise, mutations in CCAAT/enhancer binding protein alpha (*CEBPA)*, and to a lesser extent nucleophosmin *(NPM1)* (Figure 1A), correlate with gene expression signatures that appear following unsupervised clustering (22,24).

The potential of GEP to uncover new subgroups has been illustrated for several types of cancer, such as cutaneous malignant melanoma, diffuse large B-cell lymphoma, breast cancer and acute lymphoblastic leukemia (25-28). In AML, several GEP cohort studies have been performed in the last few years. In many of those, new subgroups were discerned. In a cohort of 166 cases of AML, two subgroups of normal karyotypes with distinguishable expression profiles were identified (21). The investigators postulated that this subdivision of the normal karyotype group could be diagnostically relevant, but noted an association of the two sub-groups with known factors – internal tandem duplications (ITDs) in fms-like tyrosine kinase 3 (*FLT3*) and FAB-M4 and M5 monocytic leukemias, respectively. In another data set, of 285 AMLs, 16 subgroups were recognized, of which several lacked previously known denominators (22). A different study divided 170 cases of AML, mostly patients of older age lacking favorable cytogenetic features, into 6 subgroups, of which some appeared novel (29).

Unsupervised analyses have also been performed to uncover heterogeneity within established AML subtypes. In a set of 166 AMLs, the core binding factor (CBF) AMLs, i.e. cases with t(8;21) or inv(16), each could be split into subgroups merely based on the GEP data, which was subsequently reproduced in another study (21,30). These observations suggested

that CBF leukemias in terms of gene expression patterns may represent heterogeneous entities. In 130 cases of pediatric AML, heterogeneity within CBF leukemias was seen as well, particularly within AML inv(16) (23).

In each of the above studies, GEP revealed previously unrecognized heterogeneity of AML. But are these diversities relevant? As cluster algorithms by definition focus on similarities between cases, a newly identified group of AMLs does not necessarily have biological or clinical importance. The biological significance of a newly identified subgroup would be convincingly substantiated by the subsequent discovery of a related underlying defect, or correlation with a characteristic clinical phenotype or treatment response. Alternatively, validation of a signature in one or more independent data sets would provide support that the novel subgroup is stable. Evidence indicating that the detection of a distinctive gene expression subtype indeed can lead to the discovery of a biologically and clinically significant subgroup has recently been demonstrated by the identification of a distinct form of leukemia characterized by epigenetic *CEBPA* silencing and an immature myeloid/T-lymphoid phenotype (Figure 1A) (31). Importantly, these findings could be confirmed in an independent cohort of human AML. Studies like these show how complementing technologies can be employed to make full use of the wealth of GEP data for subgroup discovery. Evidently, for successful subgroup discovery it is important to have access to sufficiently large series of cases that represent the variable subtypes of AML. Compatibility of platforms and data sharing through on-line repositories can facilitate the latter (32,33).

INTERPRETATION OF THE RESULTS OF GEP STUDIES FOR CLASS DISCOVERY

While GEP has been demonstrated to offer robust and reproducible technology in the analysis of cohorts of AML (21-23), one should remain well aware of the factors that may affect the results. We will here discuss some of the most important of those factors in the context of class discovery, although many will also affect class prediction and class comparison.

Differences in study design may determine the probability of the discovery of disease heterogeneity. Variations in the selection of study populations, in terms of size as well as in terms of demographic diversity will have a direct impact on experimental results. Furthermore, various technical differences between studies may influence results, ranging from sample processing and mRNA isolation to microarray hybridization and analysis. Inter-study variations with regard to the bioinformatic and biostatistical approaches, which involves choices regarding data normalization, gene filtering, and clustering procedures, may exert marked effects on outcome of the analysis (10,34). Most cluster algorithms present results in a two-dimensional way, in which slight changes in calculation may move samples from one cluster to another. Such analytical differences will most likely not have a major influence on
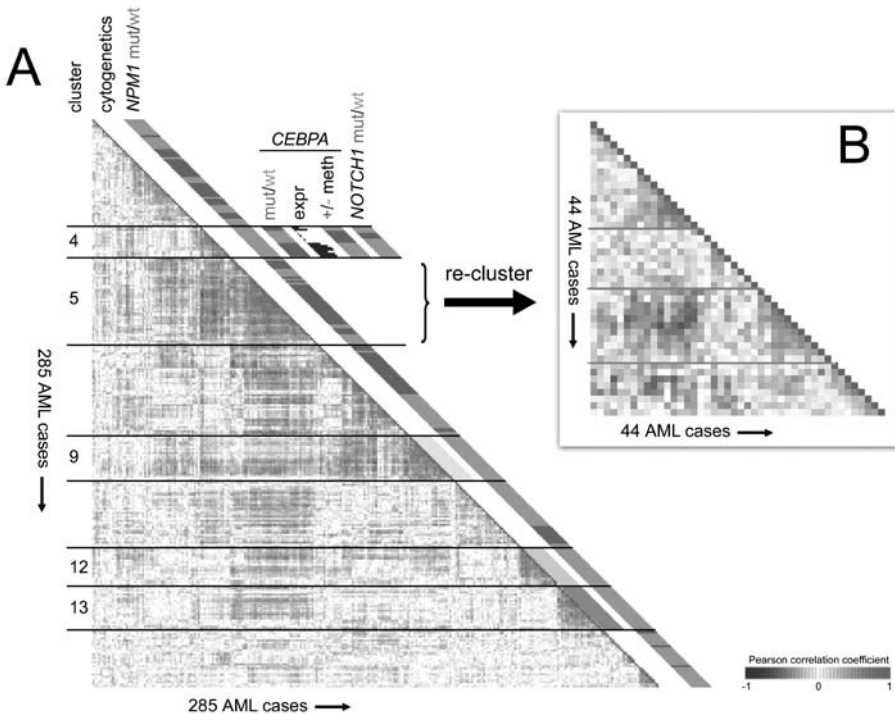
Figure 1. Summary of GEP findings in a cohort of 285 cases of AML.

A. A previous study of 285 cases of AML revealed 16 subgroups (clusters) of cases based on similarities in gene expression profiles (22). In the figure on the left, pair wise correlations between these AML cases are shown. The cells in the visualization are colored by Pearson correlation coefficient values, with deeper colors depicting higher positive (red) or negative (blue) correlations, as indicated by the scale bar. Five of the 16 clusters have been labeled, i.e. clusters #4, #5, #9, #12 and #13.

One finding of the original study was the tight aggregation into distinct clusters of AML cases with cytogenetic abnormalities that predict good risk. For those cases, cytogenetic status is color coded in the *cytogenetics* column, i.e. inv(16) (yellow, next to cluster #9), t(15;17) (orange, cluster #12) and t(8;21) (pink, cluster #13).

A subsequent study in the same patient cohort identified *NPM1* mutations in 95/285 cases. *NPM1* mutational status is depicted next to each case (red = *NPM1* mutant, green = *NPM1* wild type) (24). The figure demonstrates that *NPM1* mutations were not randomly distributed over the 16 previously defined clusters, but enriched in several of them.

Cluster #4 was found to associate with *CEBPA* mutations (red = *CEBPA* mutant). However, a subset of 6 patients in this cluster did not show any *CEBPA* mutation (green = *CEBPA* wild type). It was found that these cases differed in their *CEBPA* mRNA expression as compared to the *CEBPA* mutant AMLs, as indicated by the histograms depicting signal intensity values for the *CEBPA* probe set on the microarray. In fact, whereas *CEBPA* mutant AMLs highly expressed *CEBPA* mRNA, expression was silenced in the cases lacking mutations. This silencing was associated with *CEBPA* DNA promoter hypermethylation (+ = methylation, - = no methylation). In addition, *NOTCH1* mutations were found as common characteristics of this subgroup (red = *NOTCH1* mutation, green = *NOTCH1* wild type) (31).

B. In the original analysis of 285 AML cases on the left hand side (panel A), the 44 cases in cluster #5 aggregated very tightly, as indicated by the deep red colors, i.e. positive Pearson correlation coefficients. Most of these 44 cases showed a monocytoid morphology (FAB-M4 or –M5) (22). This raises the possibility that a significant part of the clustering effect was caused by specific up- or downregulation of genes that are important in monocytic differentiation, resulting in a different signature than the remaining, mostly non-monocytoid, cases of AML in the study. To answer whether gene expression profiling would enable identification of potential heterogeneity within this apparently homogeneous subgroup, here (panel B) the 44 cases were re-clustered as an isolated cohort. For this analysis, only probe sets that showed a variable expression within these 44 AML cases were taken into account, as defined by a fold change of 3.5 to the mean in log2 scale in at least one case. The resulting cluster image shows that a number of potentially interesting subgroups can indeed be identified within these 44 AML cases, which have been indicated by gray lines

(A full color version of this figure can be found in the color section.)

subgroups with very distinctive signatures, but may have an effect on the identification of more subtle differences.

It is important to keep in mind that GEP based clustering is driven by similarities and differences in gene expression profiles between samples. The similarities *could* be caused by shared underlying genomic defects – which is what most researchers are primarily interested in – but could also be caused by factors that are not directly related to pathogenetic mechanisms, e.g. similarities in maturation phenotype of the leukemias. This has been observed in AML as well. In two relatively large studies, AMLs with monocytoid morphologies (FAB-M4-M5 leukemias) showed a tendency to aggregate according to similar gene expression patterns (21,22). The differences in predominant maturation stage of the leukemias may also have influenced the aggregation of *CEBPA* mutant cases into two gene expression clusters, one cluster including cytologically more immature leukemias (22,31). Such pitfalls in the analysis of GEP for subgroup discovery may explain why the number of novel subtypes of AML reliably identified by GEP have remained relatively limited.

How can the potential of GEP for discovery of novel subgroups of AML be enhanced? In the search for pathogenetically relevant differences between leukemias, one would wish to avoid the interference of effects of phenotypic differences. One way of achieving this would be to restrict the analysis to relatively homogeneous populations. The feasibility of avoiding unwanted background differences that might obscure interesting pathogenetic differences in a selected predefined AML group was demonstrated by investigators who established two gene expression profiles of non-Down syndrome associated acute megakaryoblastic leukemia (35). Similarly, it has been noted that distinct subgroups can be found within the preselected t(15;17) AML subtype (36). Likewise, the 44 AMLs of a previously established cluster that strongly associated with FAB-M4 and M5 leukemias exhibited notable internal heterogeneity when studied as an isolated population (Figure 1B). Studies dealing with purified progenitor cells will also be instrumental for exclusion of interfering transcriptional background. Accumulating evidence suggests the existence of leukemia stem cells (37,38). Profiling of those cells, instead of total blast populations, may enhance the possibilities of GEP for subgroup discovery. Focusing on the leukemic stem cell may also reveal stronger transcriptional profiles that may be buried only in subclones. However, this approach directly depends on the accepted definition of immunophenotypic markers of leukemic stem cells and suffers from the technical drawback of rare stem cell numbers. Another opportunity for class discovery lies in the application of improved analytical procedures. A notable example is the use of pathway oriented analyses, which allow the discovery of distorted functional networks of genes as opposed to gene-based approaches (39,40).

DO GENOME-WIDE GENE EXPRESSION ANALYSES DESERVE A PLACE IN CLINICAL DIAGNOSIS?

Several techniques are currently used in the initial diagnosis of AML, including cytology, immunophenotyping, karyotyping, polymerase chain reaction (PCR) and fluorescence in-situ hybridization (FISH). Because a GEP-based approach allows detection of many transcripts at the same time, it provides a transcriptional snap shot of the leukemia. Several research groups have investigated the possibility to define specific gene expression *classifiers* (or *class predictors*) for disease subtypes, e.g. a discriminative set of genes for AML with the t(8;21) translocation. This procedure of *class prediction* through the generation of classifiers differs from class discovery, as discussed in the first section, in one important aspect: it makes use of external information – e.g., absence or presence of t(8;21) – to derive a signature that can subsequently be used for prediction of samples of leukemia of which the status is not known yet. This type of approach is therefore often referred to as *supervised*.

AMLs defined by t(8;21), AML inv(16) and acute promyelocytic leukemia with t(15;17) have consistently been found to be predictable using gene expression classifiers with almost 100% accuracy (17,22,23,41). For AMLs with 11q23-rearrangements involving the mixed lineage leukemia *(MLL)* gene, reported prediction accuracy in a study within several types of human leukemia was approximately 90%, while this was 95% in a study on pediatric AML (23,41). Efforts to derive gene expression classifiers for AMLs with other chromosomal abnormalities have as yet met little success. Thus, for instance AML with chromosomal abnormalities involving trisomy 8, complex karyotype and 3q appears not to be accurately predictable by GEP in representative cohorts of AML (20,21,41).

Prediction of mutations in *CEBPA* and *NPM1* has also been pursued (22,24). In one study, correct prediction of most *NPM1* mutation positive cases was possible, albeit at the expense of a significant number of false positives, resulting in a positive predictive value of 70-75% (24). This would create a hurdle when GEP would be used for diagnostic purposes. For *CEBPA* mutations, prediction accuracy may be dependent on the type of mutation. While biallelic mutations, which are relatively frequent, appear to be predictable with a gene expression classifier with high positive and negative predictive values, heterozygous monoallelic mutations may not be accurately predicted (42).

Likewise, abnormalities in signaling molecules such as FLT3 and RAS appear not to be readily predictable within diverse AML cohorts (22). This may not be too surprising taking into account their less direct role in transcriptional modulation. In some studies, reasonably successful prediction of FLT3-ITD status was achieved when cytogenetically normal AMLs were selectively analyzed (43,44). A recently reported classifier for *FLT3*-ITD mutation status in normal karyotype AML showed only a modest performance in predicting *FLT3*-ITD status in a validation cohort of 72 normal karyotype cases, with both a relatively high number of false positives and false negatives (sensitivity 73%, specificity 85%) (45).

Taken together, these observations suggest that, using current methodology, definition of GEP-based classifiers is only feasible for selected AML subtypes in which the underlying molecular abnormality is not too distantly involved in transcriptional modulation. Those abnormalities, i.e. those involving t(8;21), inv(16) and t(15;17), can be identified by rapid and widely used methods such as PCR as well. This may raise the important question what the ultimate clinical utility of GEP classifiers will be. The value of GEP-based classification lies particularly in its comprehensiveness, i.e. the opportunity to perform many tests simultaneously, for instance using specifically designed diagnostic DNA microarrays. Evaluation of larger representative AML patient series will be needed to reveal whether it is possible to define additional classifiers for less frequent cytogenetic or molecular subtypes. Definition of such additional classifiers will enhance the attractiveness of GEP as a diagnostic assay.

IS GENOME-WIDE GENE EXPRESSION ANALYSIS USEFUL FOR PREDICTING PROGNOSIS?

Several attempts have been made to derive prognostic signatures for AML. This approach is similar to class prediction, as discussed in the previous section, but instead of subgroup status (e.g. t(8;21)), outcome (e.g. overall survival) is used as the endpoint to define a *prognostic predictor* (46,47).

In one study, a set of 133 genes was demonstrated to predict survival among adults with normal karyotype AML (21). A second group of investigators converted this signature into a prognostic predictor and confirmed its prognostic ability as regards overall and disease-free survival in another series of normal karyotype AML (48). Because of differences in DNA microarray platforms, the investigators could only verify 81 of the original 133 genes, so that a complete validation of the original signature was not possible (34). A significant part of the prognostic effect was associated with *FLT3*-ITD mutations (48).

For relapse of pediatric AML, a two-gene predictor has been proposed (23). The predictive value of this indicator appeared to be modest in a small validation subset of the pediatric cohort as well as in a set of adult AML cases. The same investigators were not able to confirm the value of a prognostic signature for pediatric AML that had been proposed by a different group (49).

While the above prognostic predictors were constructed without any a priori biological assumption, a hypothesis driven approach was used to construct a predictor consisting of 11 genes associated with a stem-cell like expression pattern. Those genes were chosen because of their relationship to *BMI1* activation in a murine prostate cancer model and in human cancer samples (50). The predictor recognized adverse outcome in several types of human cancer including AML, but its value awaits independent evaluation.

Will these studies lead to the introduction of GEP-based prognostic tests for AML in clinical practice? Experience in the field of breast cancer and diffuse large B-cell lymphoma may exemplify the use of GEP-based prognostication (51-56). For breast cancer, recently two prognostic predictors were reported by independent groups (53,54). Initially these GEP-based predictors met some skepticism because of their minimal overlap in genes. Independent validation studies, however, subsequently confirmed prognostic value for both signatures, independent of other available prognostic markers (52,57). The very limited overlap in genes between the two predictors is most likely explained by the fact that in a typical GEP experiment, many transcripts are more or less similarly correlated to outcome. Consequently, several combinations of genes are equally informative for prognosis (58). The success of the breast cancer predictors has led to the recent approval by the United States Food and Drug Administration of a commercial test that is based on one of them (56,59).

From currently available evidence it appears that GEP predictors for prognosis can probably be established for AML. This will have the advantage of a comprehensive prognostic test that can substitute several currently used prognostic markers. However, it is still necessary to better assess whether such GEP based predictors can add information over currently available cytogenetic and molecular markers for prognosis of AML (46,60). Validation of any prognostic signature in sufficiently large series will be necessary before accepting this as a solid and reliable indicator of prognosis (46,58,61-65).

WHAT CAN GENOME-WIDE GENE EXPRESSION ANALYSES TELL ABOUT THE BIOLOGY OF AML?

An additional quality of genome-wide assessment of mRNA levels of ten thousands of genes is that it may allow for the discovery of pathobiological pathways. Unbiased genome-wide GEP in AML may identify critical downstream targets of known oncogenes or tumor suppressors or identify novel causative genetic abnormalities. This concept has been illustrated to be particularly promising in animal or cell line cancer models in which an oncogene of interest had been introduced. Recent examples include models for acute promyelocytic leukemia (66), mutant *Cebpa* (67) and *Mll-Af9* (68). Indeed, these studies demonstrate that biological information is captured within the thousands of transcripts measured.

Although the search for critical abnormalities in clinical AML sounds straightforward, in practice there may be nontrivial hurdles on the way to the discovery of genomic abnormalities playing a key role in the pathophysiology. There may be numerous differences in gene expression between AML cases, but only few of them may be truly disease-pathogenesis related. How to extract the biologically important transcriptional differences from the large amount of data? One possible approach is the supervised class comparison strategy, which typically involves three steps. First, a subgroup of AML of interest is defined. Next, the gene

expression profiles of cases in this subgroup are compared to those of control cases, yielding a list of differentially expressed genes. And finally, a selection of a number of genes from that list is made for further study, based on their presumed biological impact. For example, in one recent study, gene expression profiles from clinical acute promyelocytic leukemia samples with the t(11;17) translocation expressing both PLZF-RARA and the reciprocal fusion protein RARA-PLZF were compared to those from samples expressing the PLZF-RARA fusion only. This led to the identification of *CRABP1* as a specific target of the reciprocal fusion product. *CRABP1* was subsequently shown to play a functionally important role in retinoid resistance (69). Comparisons of *FLT3*-ITD and *FLT3*-TKD signatures to *FLT3*-wild type signatures have led to the discovery of potentially relevant downstream targets as well, although these remain to be functionally validated (44). While these examples show the potential of this approach, the disadvantage of supervised comparisons with subsequent selection of candidate genes is restriction due to inherent selection bias. One way around the selection bias problem may be employment of pathway oriented bioinformatic analyses. Using specialized software it has become possible to investigate the differential expression of sets of genes known to function in the same biological pathways or to identify genes that have frequently been linked in the literature (39,40,70). These analyses facilitate the identification of the most promising candidate genes and their selection for functional validation in *in vitro* or *in vivo* model systems. At the same time, it should be kept in mind that relative levels of mRNA expression do not necessarily reflect biological activity, as the latter may be highly dependent on other factors, such as post-translational modifications. Nevertheless, a recent report on a novel subgroup of immature leukemias with myeloid and T-lymphoid characteristics demonstrated the value of the application of biological pathway analysis to clinical GEP data (31). In a clinical AML GEP study, these leukemias were found to carry expression profiles similar to AML cases with *CEBPA* mutations, but such mutations were not present (Figure 1A). Subsequent experiments elucidated the likely explanation for this phenomenon, as in the novel subgroup *CEBPA* expression was silenced, frequently through promoter hypermethylation. Pathway analysis then revealed that the leukemias carried both myeloid and T-lymphoid features. Mouse modeling subsequently demonstrated that lack of *CEBPA* expression induced the expression of certain T-cell genes in immature hematopoietic cells. Thus, the analysis of human AML GEP using pathway analysis in combination with experiments in a representative mouse model uncovered part of the pathobiology of a subgroup of leukemia characterized by a myeloid/T-lymphoid phenotype, *CEBPA* silencing and, in fact, also frequent *NOTCH1* mutations.

Another strategy for distilling biologically significant genes from human AML gene expression data makes use of possible associations with large sets of genes involved in animal retroviral insertion leukemogenesis. A study that compared integration sites in retrovirally induced leukemias in mice with human AML data sets demonstrated that mouse cancer genes were frequently deregulated in the human AMLs (71). Moreover, pathway analysis defined

several biological networks that associated with particular AML subsets. Thus, comparisons of human AML GEP data with high-throughput results from dedicated experimental models provide valuable opportunities to identify candidate disease genes. As pointed out before, these experimental approaches could also utilize GEP of cell lines or animal models (25-27), or, alternatively, employ techniques such as chromatin immunoprecipitation on DNA microarray chips (ChIP-chip) (72), ChIP-sequencing (73) and RNA interference libraries to uncover target genes. A successful example of the latter approach comes from diffuse large B-cell lymphoma, for which two previously GEP-defined disease subgroups were functionally investigated using an RNA interference library to search for specific targets inhibiting tumor growth (26,74).

Another viable strategy is based on the correlation of putative disease genes from in vivo or in vitro models to specific human leukemia subtypes. The strength of this approach is that the gene choice is based on experimental data. This allows for a rapid correlation of basic research findings to human AML data sets: is transforming gene X, that induces a leukemic phenotype in an experimental mouse model, differentially expressed in particular forms of human AML? Using this approach, researchers have provided evidence that genes such as *TrkA* and *Trib2*, which are involved in murine leukemia, may also be engaged in t(8;21) AML and *CEBPA* silenced AML, respectively (75,76).

These studies demonstrate the abilities of GEP to resolve questions related to the biology of AML when combined with appropriate other experimental and analytical tools. At the same time, there are various other experimental options to pinpoint key genomic abnormalities through GEP in human AML samples.

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Recent years have shown an increase in high-throughput applications apart from GEP (8). In this respect profiling of microRNA (miRNA) levels, chromosomal copy number changes, epigenetic modifications, and DNA sequencing offer interesting opportunities.

MiRNAs are small non-coding RNAs that play a role in transcriptional or post-transcriptional regulation of genes involved in numerous biological processes, including differentiation and proliferation (77). Profiling of miRNA expression levels in AML cohorts has indicated that, similar to mRNA profiling, distinct miRNA signatures are associated with specific subgroups of AML (78-80). As a single miRNA can play a role in concomitant regulation of multiple genes, closer comparisons between miRNA and mRNA profiles may provide clues for significant defects.

Changes in expression levels of critical genes may be due to small DNA amplifications or deletions undetected by conventional cytogenetics. Platforms to study those small chromosomal aberrations include single nucleotide polymorphism (SNP) arrays and array-based

comparative genomic hybridization (CGH) (81,82). The power of the use of SNP arrays for this type of analysis was recently illustrated in a study of 242 cases of pediatric ALL (83). The investigators identified focal abnormalities in lymphocyte differentiation related genes in 40% of cases, including 30% in the *PAX5* gene. As yet, only studies of limited size have been performed in AML (84-89). A study in which array-CGH (bacterial artificial and P1-derived artificial chromosome clones) was used in a series of 60 cases of AML with complex karyotypes, identified several recurrent lesions (90). It is not clear yet what the overall frequency and distribution of these genomic alterations in AML is. A notable observation from several SNP-array-based studies is the relatively common occurrence of copy-number neutral loss of heterozygosity through segmental uniparental disomy (84,87,91). This phenomenon can lead to homozygous mutations or deletions of leukemia-related genes, including *FLT3, WT1, RUNX1* and *CEBPA* (91,92).

Variations in gene expression may also be caused by yet unknown gene sequence mutations. Whole genome sequencing is emerging as a means to address this issue on a global scale (93,94). A challenge in this regard will be to distinguish functionally relevant mutations from the abundant so-called passenger mutations – unimportant genetic changes caused by genomic instability of cancer cells – that will be picked up at the same time. Studies on tyrosine kinase abnormalities have underscored the need for validation of the biological effects of novel mutations identified by high-throughput nucleotide sequencing (95,96).

In addition to genetic alterations, epigenetic modifications, including DNA and histone methylation, play a pivotal role in gene regulation. Several platforms have been developed to study alterations in these mechanisms on a genome-wide scale (97,98). The first results of genome wide CpG methylation profiling of AML cell lines and primary cells have now been reported, and it is likely that more extensive investigations will follow (99,100). Given the direct association between epigenetics and expression, much is to be expected from this field.

As these fields have just started to emerge, there is still a need for bioinformatics to keep up with the technical developments, and to develop software platforms that allow the integration of the massive amounts of data. It is likely that the combination of GEP with complementing technologies, such as those outlined above, will provide challenging opportunities to address questions that can not be resolved by GEP alone.

ACKNOWLEDGEMENTS

REFERENCES

1. Lowenberg B, Downing JR, Burnett A. Acute myeloid leukemia. N Engl J Med 1999;341(14):1051-62.

2. Estey E, Dohner H. Acute myeloid leukaemia. Lancet 2006;368(9550):1894-907.

3. Mrozek K, Heinonen K, Bloomfield CD. Clinical importance of cytogenetics in acute myeloid leukaemia. Best Pract Res Clin Haematol 2001;14(1):19-47.

4. Mrozek K, Marcucci G, Paschka P, Whitman SP, Bloomfield CD. Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification? Blood 2007;109(2):431-48.

5. Tallman MS, Gilliland DG, Rowe JM. Drug therapy for acute myeloid leukemia. Blood 2005;106(4):1154-63.

6. Dash A, Gilliland DG. Molecular genetics of acute myeloid leukaemia. Best Pract Res Clin Haematol 2001;14(1):49-64.

7. Pollack JR. A perspective on DNA microarrays in pathology research and practice. Am J Pathol 2007;171(2):375-85.

8. Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. Nat Rev Genet 2006;7(3):200-10.

9. Elvidge G. Microarray expression technology: from start to finish. Pharmacogenomics 2006;7(1):123-34.

10. Quackenbush J. Microarray analysis and tumor classification. N Engl J Med 2006;354(23):2463-72.

11. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286(5439):531-7.

12. Bullinger L, Valk PJ. Gene expression profiling in acute myeloid leukemia. J Clin Oncol 2005;23(26):6296-305.

13. Ebert BL, Golub TR. Genomic approaches to hematologic malignancies. Blood 2004;104(4):923-32.

14. Haferlach T, Kohlmann A, Kern W, Hiddemann W, Schnittger S, Schoch C. Gene expression profiling as a tool for the diagnosis of acute leukemias. Semin Hematol 2003;40(4):281-95.

15. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95(25):14863-8.

16. D'Haeseleer P. How does gene expression clustering work? Nat Biotechnol 2005;23(12):1499-501.

17. Schoch C, Kohlmann A, Schnittger S, Brors B, Dugas M, Mergenthaler S, et al. Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. Proc Natl Acad Sci U S A 2002;99(15):10008-13.

18. Vey N, Mozziconacci MJ, Groulet-Martinec A, Debono S, Finetti P, Carbuccia N, et al. Identification of new classes among acute myelogenous leukaemias with normal karyotype using gene expression profiling. Oncogene 2004;23(58):9381-91.

19. Debernardi S, Lillington DM, Chaplin T, Tomlinson S, Amess J, Rohatiner A, 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 2003;37(2):149-58.

20. Virtaneva K, Wright FA, Tanner SM, Yuan B, Lemon WJ, Caligiuri MA, 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 2001;98(3):1124-9.

21. Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004;350(16):1605-16.

22. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

23. Ross ME, Mahfouz R, Onciu M, Liu HC, Zhou X, Song G, et al. Gene expression profiling of pediatric acute myelogenous leukemia. Blood 2004;104(12):3679-87.

24. Verhaak RG, Goudswaard CS, van Putten W, Bijl MA, Sanders MA, Hugens W, 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 2005;106(12):3747-54.

25. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature 2000;406(6795):536-40.

26. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403(6769):503-11.

27. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406(6797):747-52.

28. Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell 2002;1(2):133-43.

29. Wilson CS, Davidson GS, Martin SB, Andries E, Potter J, Harvey R, et al. Gene expression profiling of adult acute myeloid leukemia identifies novel biologic clusters for risk classification and outcome prediction. Blood 2006;108(2):685-96.

30. Bullinger L, Rucker FG, Kurz S, Du J, Scholl C, Sander S, et al. Gene-expression profiling identifies distinct subclasses of core binding factor acute myeloid leukemia. Blood 2007;110(4):1291-300.

31. Wouters BJ, Jorda MA, Keeshan K, Louwers I, Erpelinck-Verschueren CA, Tielemans D, et al. Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. Blood 2007;110(10):3706-14.

32. Brazma A, Parkinson H, Sarkans U, Shojatalab M, Vilo J, Abeygunawardena N, et al. ArrayExpress-a public repository for microarray gene expression data at the EBI. Nucleic Acids Res 2003;31(1):68-71.

33. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res 2002;30(1):207-10.

34. Michiels S, Koscielny S, Hill C. Interpretation of microarray data in cancer. Br J Cancer 2007;96(8):1155-8.

35. Bourquin JP, Subramanian A, Langebrake C, Reinhardt D, Bernard O, Ballerini P, et al. Identification of distinct molecular phenotypes in acute megakaryoblastic leukemia by gene expression profiling. Proc Natl Acad Sci U S A 2006;103(9):3339-44.

36. Marasca R, Maffei R, Zucchini P, Castelli I, Saviola A, Martinelli S, et al. Gene expression profiling of acute promyelocytic leukaemia identifies two subtypes mainly associated with flt3 mutational status. Leukemia 2006;20(1):103-14.

37. Huntly BJ, Gilliland DG. Leukaemia stem cells and the evolution of cancer-stem-cell research. Nat Rev Cancer 2005;5(4):311-21.

38. Wang JC, Dick JE. Cancer stem cells: lessons from leukemia. Trends Cell Biol 2005;15(9):494-501.

39. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102(43):15545-50.

40. Zhu X, Gerstein M, Snyder M. Getting connected: analysis and principles of biological networks. Genes Dev 2007;21(9):1010-24.

41. Haferlach T, Kohlmann A, Schnittger S, Dugas M, Hiddemann W, Kern W, et al. Global approach to the diagnosis of leukemia using gene expression profiling. Blood 2005;106(4):1189-98.

42. Verhaak RG, Wouters BJ, Erpelinck CA, Abbas S, Beverloo HB, Lugthart S, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. Haematologica 2009;94(1):131-4.

43. Lacayo NJ, Meshinchi S, Kinnunen P, Yu R, Wang Y, Stuber CM, et al. Gene expression profiles at diagnosis in de novo childhood AML patients identify FLT3 mutations with good clinical outcomes. Blood 2004;104(9):2646-54.

44. Neben K, Schnittger S, Brors B, Tews B, Kokocinski F, Haferlach T, et al. Distinct gene expression patterns associated with FLT3- and NRAS-activating mutations in acute myeloid leukemia with normal karyotype. Oncogene 2005;24(9):1580-8.

45. Bullinger L, Dohner K, Kranz R, Stirner C, Frohling S, Scholl C, et al. An FLT3 gene-expression signature predicts clinical outcome in normal karyotype AML. Blood 2008;111(9):4490-5.

46. Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. J Clin Oncol 2005;23(29):7332-41.

47. Bovelstad HM, Nygard S, Storvold HL, Aldrin M, Borgan O, Frigessi A, et al. Predicting survival from microarray data--a comparative study. Bioinformatics 2007;23(16):2080-7.

48. Radmacher MD, Marcucci G, Ruppert AS, Mrozek K, Whitman SP, Vardiman JW, et al. Independent confirmation of a prognostic gene-expression signature in adult acute myeloid leukemia with a normal karyotype: a Cancer and Leukemia Group B study. Blood 2006;108(5):1677-83.

49. Yagi T, Morimoto A, Eguchi M, Hibi S, Sako M, Ishii E, et al. Identification of a gene expression signature associated with pediatric AML prognosis. Blood 2003;102(5):1849-56.

50. Glinsky GV, Berezovska O, Glinskii AB. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. J Clin Invest 2005;115(6):1503-21.

51. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 2002;8(1):68-74.

52. Foekens JA, Atkins D, Zhang Y, Sweep FC, Harbeck N, Paradiso A, et al. Multicenter validation of a gene expression-based prognostic signature in lymph node-negative primary breast cancer. J Clin Oncol 2006;24(11):1665-71.

53. Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365(9460):671-9.

54. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415(6871):530-6.

55. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347(25):1999-2009.

56. van't Veer LJ, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. Nature 2008;452(7187):564-70.

57. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst 2006;98(17):1183-92.

58. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 2007;99(2):147-57.

59. Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, et al. Converting a breast cancer microarray signature into a high-throughput diagnostic test. BMC Genomics 2006;7:278.

60. Kattan MW. Judging new markers by their ability to improve predictive accuracy. J Natl Cancer Inst 2003;95(9):634-5.

61. Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, Gehan EA, et al. The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. Nat Rev Cancer 2008;8(1):37-49.

62. Tinker AV, Boussioutas A, Bowtell DD. The challenges of gene expression microarrays for the study of human cancer. Cancer Cell 2006;9(5):333-9.

63. Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. Lancet 2003;362(9394):1439-44.

64. Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. Lancet 2005;365(9458):488-92.

65. Dobbin KK, Zhao Y, Simon RM. How Large a Training Set is Needed to Develop a Classifier for Microarray Data? Clin Cancer Res 2008;14(1):108-14.

66. Yuan W, Payton JE, Holt MS, Link DC, Watson MA, DiPersio JF, et al. Commonly dysregulated genes in murine APL cells. Blood 2007;109(3):961-70.

67. Kirstetter P, Schuster MB, Bereshchenko O, Moore S, Dvinge H, Kurz E, et al. Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. Cancer Cell 2008;13(4):299-310.

68. Chen W, Kumar AR, Hudson WA, Li Q, Wu B, Staggs RA, et al. Malignant transformation initiated by Mll-AF9: gene dosage and critical target cells. Cancer Cell 2008;13(5):432-40.

69. Guidez F, Parks S, Wong H, Jovanovic JV, Mays A, Gilkes AF, et al. RARalpha-PLZF overcomes PLZF-mediated repression of CRABPI, contributing to retinoid resistance in t(11;17) acute promyelocytic leukemia. Proc Natl Acad Sci U S A 2007;104(47):18694-9.

70. Jelier R, Jenster G, Dorssers LC, Wouters BJ, Hendriksen PJ, Mons B, et al. Text-derived concept profiles support assessment of DNA microarray data for acute myeloid leukemia and for androgen receptor stimulation. BMC Bioinformatics 2007;8:14.

71. Erkeland SJ, Verhaak RG, Valk PJ, Delwel R, Lowenberg B, Touw IP. Significance of murine retroviral mutagenesis for identification of disease genes in human acute myeloid leukemia. Cancer Res 2006;66(2):622-6.

72. Wu J, Smith LT, Plass C, Huang TH. ChIP-chip comes of age for genome-wide functional analysis. Cancer Res 2006;66(14):6899-902.

73. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 2007;4(8):651-7.

74. Ngo VN, Davis RE, Lamy L, Yu X, Zhao H, Lenz G, et al. A loss-of-function RNA interference screen for molecular targets in cancer. Nature 2006;441(7089):106-10.

75. Keeshan K, He Y, Wouters BJ, Shestova O, Xu L, Sai H, et al. Tribbles homolog 2 inactivates C/EBPalpha and causes acute myelogenous leukemia. Cancer Cell 2006;10(5):401-11.

76. Mulloy JC, Jankovic V, Wunderlich M, Delwel R, Cammenga J, Krejci O, et al. AML1-ETO fusion protein up-regulates TRKA mRNA expression in human CD34+ cells, allowing nerve growth factor-induced expansion. Proc Natl Acad Sci U S A 2005;102(11):4016-21.

77. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nat Rev Cancer 2006;6(11):857-66.

78. Jongen-Lavrencic M, Sun SM, Dijkstra MK, Valk PJ, Lowenberg B. MicroRNA expression profiling in relation to the genetic heterogeneity of acute myeloid leukemia. Blood 2008;111(10):5078-85.

79. Mi S, Lu J, Sun M, Li Z, Zhang H, Neilly MB, et al. MicroRNA expression signatures accurately discriminate acute lymphoblastic leukemia from acute myeloid leukemia. Proc Natl Acad Sci U S A 2007;104(50):19971-6.

80. Garzon R, Volinia S, Liu CG, Fernandez-Cymering C, Palumbo T, Pichiorri F, et al. MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. Blood 2008;111(6):3183-9.

81. Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet 2007;39(7 Suppl):S16-21.

82. Dutt A, Beroukhim R. Single nucleotide polymorphism array analysis of cancer. Curr Opin Oncol 2007;19(1):43-9.

83. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature 2007;446(7137):758-64.

84. Raghavan M, Lillington DM, Skoulakis S, Debernardi S, Chaplin T, Foot NJ, et al. Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. Cancer Res 2005;65(2):375-8.

85. Suela J, Alvarez S, Cifuentes F, Largo C, Ferreira BI, Blesa D, et al. DNA profiling analysis of 100 consecutive de novo acute myeloid leukemia cases reveals patterns of genomic instability that affect all cytogenetic risk groups. Leukemia 2007;21(6):1224-31.

86. Karnan S, Tsuzuki S, Kiyoi H, Tagawa H, Ueda R, Seto M, et al. Genomewide array-based comparative genomic hybridization analysis of acute promyelocytic leukemia. Genes Chromosomes Cancer 2006;45(4):420-5.

87. Gorletta TA, Gasparini P, D'Elios MM, Trubia M, Pelicci PG, Di Fiore PP. Frequent loss of heterozygosity without loss of genetic material in acute myeloid leukemia with a normal karyotype. Genes Chromosomes Cancer 2005;44(3):334-7.

88. Tyybakinoja A, Elonen E, Piippo K, Porkka K, Knuutila S. Oligonucleotide array-CGH reveals cryptic gene copy number alterations in karyotypically normal acute myeloid leukemia. Leukemia 2007;21(3):571-4.

89. Paulsson K, Heidenblad M, Strombeck B, Staaf J, Jonsson G, Borg A, et al. High-resolution genome-wide array-based comparative genome hybridization reveals cryptic chromosome changes in AML and MDS cases with trisomy 8 as the sole cytogenetic aberration. Leukemia 2006;20(5):840-6.

90. Rucker FG, Bullinger L, Schwaenen C, Lipka DB, Wessendorf S, Frohling S, et al. Disclosure of candidate genes in acute myeloid leukemia with complex karyotypes using microarray-based molecular characterization. J Clin Oncol 2006;24(24):3887-94.

91. Fitzgibbon J, Smith LL, Raghavan M, Smith ML, Debernardi S, Skoulakis S, et al. Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. Cancer Res 2005;65(20):9152-4.

92. Wouters BJ, Sanders MA, Lugthart S, Geertsma-Kleinekoort WM, van Drunen E, Beverloo HB, et al. Segmental uniparental disomy as a recurrent mechanism for homozygous CEBPA mutations in acute myeloid leukemia. Leukemia 2007;21(11):2382-4.

93. Ley TJ, Minx PJ, Walter MJ, Ries RE, Sun H, McLellan M, et al. A pilot study of high-throughput, sequence-based mutational profiling of primary human acute myeloid leukemia cell genomes. Proc Natl Acad Sci U S A 2003;100(24):14275-80.

94. Thomas RK, Baker AC, Debiasi RM, Winckler W, Laframboise T, Lin WM, et al. High-throughput oncogene mutation profiling in human cancer. Nat Genet 2007;39(3):347-51.

95. Loriaux MM, Levine RL, Tyner JW, Frohling S, Scholl C, Stoffregen EP, et al. High-throughput sequence analysis of the tyrosine kinome in acute myeloid leukemia. Blood 2008;111(9):4788-96.

96. Frohling S, Scholl C, Levine RL, Loriaux M, Boggon TJ, Bernard OA, et al. Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. Cancer Cell 2007;12(6):501-13.

97. Khulan B, Thompson RF, Ye K, Fazzari MJ, Suzuki M, Stasiek E, et al. Comparative isoschizomer profiling of cytosine methylation: the HELP assay. Genome Res 2006;16(8):1046-55.

98. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 2005;37(8):853-62.

99. Gebhard C, Schwarzfischer L, Pham TH, Schilling E, Klug M, Andreesen R, et al. Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. Cancer Res 2006;66(12):6118-28.

100. Yuan E, Haghighi F, White S, Costa R, McMinn J, Chun K, et al. A single nucleotide polymorphism chip-based method for combined genetic and epigenetic profiling: validation in decitabine therapy and tumor/normal comparisons. Cancer Res 2006;66(7):3443-51.

CHAPTER

3

Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling

Roel G.W. Verhaak[1,2,*], Bas J. Wouters[1,*], Claudia A.J. Erpelinck[1], Saman Abbas[1], H.Berna Beverloo[3], Sanne Lugthart[1], Bob Löwenberg[1], Ruud Delwel[1] and Peter J.M. Valk[1]

[*] these authors contributed equally to the study

[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 Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands

ABSTRACT

We examined the gene expression profiles of two independent cohorts of patients with acute myeloid leukemia (n=247 and n=214 (≤60 years)) to study the applicability of gene expression profiling as a single assay in prediction of acute myeloid leukemia-specific molecular subtypes. The favorable cytogenetic acute myeloid leukemia subtypes, i.e., acute myeloid leukemia with t(8;21), t(15;17) or inv(16), were predicted with maximum accuracy (positive and negative predictive value: 100%). Mutations in *NPM1* and *CEBPA* were predicted less accurately (positive predictive value: 66% and 100% and negative predictive value: 99% and 97% respectively). Various other characteristic molecular acute myeloid leukemia subtypes, i.e., mutant *FLT3* and *RAS*, abnormalities involving 11q23, -5/5q-, -7/7q-, abnormalities involving 3q (abn3q) and t(9;22), could not be correctly predicted using gene expression profiling. In conclusion, gene expression profiling allows accurate prediction of certain AML subtypes, e.g. those characterized by expression of chimeric transcription factors. However, detection of mutations affecting signaling molecules and numerical abnormalities still requires alternative molecular methods.

INTRODUCTION

Acute myeloid leukemia (AML) is not a single disease but a group of neoplasms with various genetic abnormalities and variable responses to treatment. The pre-treatment karyotype is still essential in therapy decision-making in AML (1-3). In recent years, a number of novel molecular markers has been associated with AML prognostics (2,3). Several attempts have been made to investigate whether genome-wide gene expression profiling (GEP) could be valuable for prediction of certain subtypes of AML (4-12). Although there was concordance in predictive signatures in the various studies, none of those studies validated the derived signatures to predict the recurrent molecular markers using independent representative AML cohorts. The question, therefore, remains whether GEP could substitute current diagnostic techniques and could be applied as a reliable single test to simultaneously detect known cytogenetic and molecular abnormalities. The aim of this study was to validate GEP as preferred single assay to predict prognostically relevant AML subtypes using two large independent cohorts of young adults with AML.

DESIGN AND METHODS

Bone marrow aspirates or peripheral blood samples of two independent representative cohorts of *de novo* AML patients (≤60 years), consisting of 247 and 214 patients, were collected (Table 1). The first cohort represents a subset of 285 patients previously studied (8), while the second cohort has not yet been described.

  Blast cell purification and RNA isolation were carried out as previously described (8). All samples were analyzed using Affymetrix Human Genome U133Plus2.0 GeneChips (Affymetrix, Santa Clara, CA, USA). Labeling, hybridization, scanning and data normalization were performed as previously described (8). The variation between the scaling/normalization factors of the GeneChips in both cohorts was less than 3-fold [cohort1: 0.53(±0.15); cohort2: 0.73(±0.20)]. Also, the percentage of genes present [cohort1: 39.1(±3.1); cohort2: 40.6(±3.7)], GAPDH 3'/5' ratio [cohort1: 1.07(±0,13); cohort2: 1.08(±0.16)] and actin 3'/5' ratio [cohort1: 1.26(±0.21); cohort2: 1.33(±0.29)] were indicative for high overall quality and consistency between both AML sample populations. Mutational analyses to detect recurrent mutations in AML were performed as previously described (13-15). All supervised class prediction analyses were performed with Prediction Analysis for Microarrays (PAM) software version 1.28 in R version 2.1.0 (16). Clinical, cytogenetic and molecular information as well as the gene expression profiles of all primary AML cases is available at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo, accession number GSE6891).

Table 1. Clinical and molecular data

	AML cohort 1 (n=247)	AML cohort 2 (n=214)
Gender		
Male	119	113
Female	128	101
Age (median (range))	43 (15-60)	46 (17-60)
White blood cell count (x 10^9/l)	30 (0-278)	29 (1-349)
Bone marrow blast count (%)	68 (0-98)	64 (0-96)
Platelet count (x 10^9/l)	49 (3-931)	59 (5-998)
FAB		
M0	6	10
M1	55	41
M2	54	52
M3	17	7
M4	43	41
M5	62	42
M6	3	3
not determined	7	18
Cytogenetics*		
normal	99 (41%)	95 (46%)
inv(16)	21 (9%)	16 (7%)
*t(15;17)***	18 (7%)	7 (3%)
t(8;21)	21 (9%)	14 (7%)
t(6;9)	4 (2%)	2 (1%)
abn3q	7 (3%)	9 (4%)
del5(q)	3 (1%)	12 (6%)
del7(q)	17 (7%)	14 (7%)
11q23	13 (5%)	8 (4%)
+8	22 (9%)	11 (5%)
t(9;22)	4 (2%)	1 (<1%)
complex	13 (5%)	21 (10%)
other	63 (26%)	45 (22%)

Table 1 continued

	AML cohort 1 (n=247)	AML cohort 2 (n=214)
Mutations*		
CEBPA	16 (6%)	15 (7%)
NPM1	77 (31%)	63 (29%)
FLT3-ITD	65 (26%)	61 (29%)
FLT3-TKD	30 (12%)	19 (9%)
KRAS	4 (2%)	0 (0%)
NRAS	23 (9%)	22 (10%)

*All patients with a specific abnormality were considered, irrespective of the presence of additional abnormalities. Percentages were calculated based on the total number of cases investigated for the particular abnormality, as also indicated in Table2.

**The overall frequencies of the AML-specific recurrent (cyto)genetic abnormalities in both cohorts is similar, except for the number of AML-M3 cases carrying t(15;17), which is lower in cohort 2. In recent studies these AML t(15;17) patients were enrolled into appropriate alternative clinical protocols.

RESULTS AND DISCUSSION

In this study of 461 clinically and molecularly well-characterized cases of AML (Table 1), we were able to comprehensively validate the application of GEP to predict therapeutically relevant molecular subtypes in AML.

We applied PAM to investigate whether karyotypic and mutational abnormalities with prognostic or therapeutic value in AML were accurately predictable based on GEP. PAM allows the selection of the minimal number of genes required for optimal prediction, which may be beneficial in a diagnostic setting. The AML cohort 1 (n=247) was used as training set to derive predictive signatures that were subsequently validated on AML cohort 2 (n=214). The deduced expression signatures are available in the Supplementary Tables S1-18.

The cytogenetic status of all AML patients with favorable risk, i.e. those with t(8;21), t(15;17) or inv(16) abnormalities, was predicted with 100 percent accuracy (Table 2). In fact, among these predicted AML cases, there were cases with favorable cytogenetics that had previously been missed by routine cytogenetics (4 out of 37 inv(16) and 4 out of 25 t(15;17)). The presence of the translocation-related fusion transcripts in these specific cases was confirmed by real-time quantitative PCR. Thus, GEP is a reliable alternative to discriminate these three AML subtypes (2,3) which represent approximately 20% of all cases (2,3). Prediction of t(15;17) and inv(16) required only few genes, as seen previously (8). For the t(8;21) cases, 76 probe sets were needed to correctly classify all samples. However, as few as two probe sets, including one associated with the *RUNX1T1* (*ETO*) gene, were sufficient to accurately classify all but one t(8;21) cases, which is also consistent with earlier studies (8) (Supplementary Figure S3).

Table 2. Class prediction using Prediction Analysis for Microarrays.

Molecular abnormality	Cross-validation error*		Error validation set**		# probe sets	Supp. Data	Sensitivity***	Specificity***	Predictive value***	
	Neg	Pos	Neg	Pos					Neg	Pos
inv(16)	0/226	0/21	0/198	0/16	2	F1 / T1	100	100	100	100
t(15;17)	0/229	0/18	0/207	0/7	7	F2 / T2	100	100	100	100
t(8;21)	0/226	0/21	0/200	0/14	76	F3 / T3	100	100	100	100
CEBPA	3/231	2/16	0/197	6/15	15	F4 / T4	60	100	97	100
NPM1	30/170	0/77	32/151	1/63	68	F5 / T5	98	79	99	66
FLT3-ITD	14/182	17/65	12/153	20/61	64	F6 / T6	67	92	88	77
FLT3-TKD	89/216	14/30	72/194	5/19	2307	F7 / T7	74	63	96	16
FLT3-ITD and/or - TKD	34/155	25/91	24/135	20/78	407	F8 / T8	74	82	85	71
NPM1 - / FLT3-ITD +	35/219	7/28	30/189	8/25	194	F9 / T9	68	84	95	36
NPM1 + / FLT3-ITD -	56/207	3/40	59/187	1/27	86	F10 / T10	96	68	99	31
NPM1 + / FLT3-ITD +	54/210	1/37	39/178	1/36	50	F11 / T11	97	78	99	47
KRAS	33/241	2/4	32/214	0/0	173	F12 / T12	N/A	85	N/A	N/A
NRAS	65/223	7/23	55/192	9/22	225	F13 / T13	59	71	94	19
t(6;9)	1/240	0/4	0/203	1/2	25	F14 / T14	50	100	100	100
3q	29/237	1/7	20/196	3/9	51	F15 / T15	67	90	98	23
-5(q)	9/240	2/3	7/193	11/12	66	F16 / T16	8	96	94	13
-7(q)	16/226	3/17	16/190	1/14	96	F17 / T17	93	92	99	45
11q23	7/231	4/13	8/197	1/8	40	F18 / T18	88	96	100	47

* The prediction error was calculated by 10-fold cross validation within the training set (cohort 1) (Cross-validation error; indicated in Supplementary Figures 1 to 19).

** The deduced gene expression signature was tested on the independent validation set (cohort 2) (Error validation set).

*** The following calculations were used for evaluation measures: positive predictive value = true positives/(true positives + false positives), negative predictive value = true negatives/(true negatives + false positives), sensitivity = true positives/(true positives + false negatives), and specificity = true negatives/(true negatives + false positives). Values are given as percentages. The overall number of cases in cohort1 and cohort2 vary slightly because in rare instances the molecular abnormality was unknown. Number of probe sets used for prediction. Supp. Data: Number Supplementary Figure (F) and Table (T). For identities of the probe sets and genes see Supplemental Tables.

AML cases with mutations in the transcription factor CCAAT/enhancer binding protein α (*CEBPA*), which are associated with a relatively favorable treatment outcome, were predicted with positive and negative predictive values of 100% and 97% respectively. Six out of 15 *CEBPA* mutant cases were missed in the validation set (sensitivity 60%; Table 2). Of note, the misclassified cases all carried a single heterozygous *CEBPA* mutation, whereas samples with biallelic mutations (either homo- or heterozygous) were all correctly recognized (data not shown). In the training cohort, all but two (14/16) samples carried biallelic mutations (14,17) and in cross-validation in the training cohort the two heterozygous mutants were the only misclassified samples as well.

Previous work has shown that mutations in nucleophosmin (*NPM1*) are strongly associated with a discriminative *HOX*-and *TALE* gene-specific signature (18,19). In this study, AML cases carrying a *NPM1* mutation were indeed recognized with high accuracy based on such a signature (Table 2 and Supplementary Table S5). However, a relatively high number of AML cases without *NPM1* mutations was incorrectly predicted positive (32 out of 151), suggesting the presence of genetic alterations resulting in a similar upregulation of the *HOX*-and TALE genes in those cases. Among these false positives were several AMLs carrying 11q23 abnormalities, which is in line with the role of the mixed lineage leukemia (MLL) protein as an important regulator of *HOX* gene expression (18). Of note, all t(6;9) AML cases in the training and validation cohort (n=6) were predicted to also carry an *NPM1* mutation, raising the possibility that the DEK-CAN fusion protein also induces *HOX*-related gene expression. Interestingly, prediction of t(6;9) translocation was partly feasible using a unique signature (Table 2 and Supplementary Table S14), although these results are based on a relatively low number of cases.

NPM1 mutations are associated with relatively favorable survival parameters in patients with a normal karyotype and standard risk AML (18,20-22). The favorable risk is particularly associated with AMLs lacking internal tandem duplications (*ITD*) in the fms-related tyrosine kinase (*FLT3*) gene (18,20-22). Analyses of AML subsets defined by combined presence or absence of *NPM1* and *FLT3*-ITD abnormalities demonstrated that only patients carrying both mutations could be moderately predicted, whereas the remaining subtypes could not be discriminated (Table 2). Restriction of these analyses to normal karyotype cases only did not result in a significant improvement in prediction accuracy (Supplementary Table S19). Of note, prediction of *NPM1* mutation in preselected normal karyotype samples led to a slightly increased positive predictive value (83 *vs.* 66%), which may be consistent with the lack of interfering 11q23 positive samples. The remaining cytogenetic and molecular subgroups we studied were not associated with strong predictive signatures. Whereas the positive predictive value for *FLT3*-ITD aberrations was relatively high (77%), the high number of false predictions eliminates GEP, with the currently available analyses tools, as a reliable test to determine the *FLT3*-ITD status. Restriction to the normal karyotype group did not lead to a marked improvement (Supplementary Table S19). Likewise, the low positive predictive

values for FLT3 tyrosine kinase domain (TKD) or RAS mutations, abnormalities involving 11q23, -5/5q-, -7/7q- and abn3q, and the translocation t(9;22), disqualify GEP as single detection method for these abnormalities. Similarly, 3q aberrations were not readily predictable. Nevertheless, the most discriminative gene for abn3q abnormalities was the oncogenic transcription factor ecotropic viral integration site 1 (EVI1)(Supplementary Table S15), which is frequently involved in 3q26 abnormalities. Of note, in these predictions we included the cases carrying a cryptic abn3q recently identified by gene expression analyses and fluorescence in-situ hybridization (23).

Classifiers were also deduced using a number of other approaches, i.e. compound covariate predictor, linear discriminant analysis, 1-nearest neighbor and 3-nearest neighbors, nearest centroid and support vector machines (probe set selection at 0.001 significance level). These alternative analyses were carried out in BRB-ArrayTools, version 3.7.0 β2 release, developed by Dr. Richard Simon and Amy Peng Lam. Overall, this comparative analysis yielded highly similar results, i.e. the favorable cytogenetic subclasses were predictable with (close to) 100% accuracy, whereas other subtypes showed a similar prediction pattern as depicted in Table 2 (data not shown). One exception was *NPM1* mutation status, for which prediction accuracy was better using an approach based on support vector machines (positive predictive value 91% with a negative predictive value of 99%).

Several general causes for the inability to predict specific recurrent abnormalities could apply: (i) if different recurrent genetic aberrations affect similar pathways, their GEP signatures may overlap; (ii) Mutations affecting signaling pathways may not result in strong discriminative mRNA expression signatures; (iii) The expression of differentiation-related genes may affect accurate prediction; (iv) Secondary mutations, or bi-allelic versus monoallelic mutations as in the case of *CEBPA*, may prohibit reliable prediction. More specifically, (v) the various partners of the MLL gene may affect reliable prediction of 11q23 abnormalities, and (vi) the numerical changes in (part of) the chromosomes 5 and 7 manuscript may only result in minor changes in gene expression that are insufficient for GEP prediction. Of note, still almost all discriminative genes with decreased expression in the deduced signature for 7(q) abnormalities were located on chromosome 7, including *FASTK, GSTK1, LSM8* and *ZNF746* (Supplementary Table S17).

Together, we conclude that AML cases with favorable cytogenetics are predictable with high accuracy with the currently available genome-wide gene expression technology and analyses tools. All other prognostically and therapeutically known abnormalities in AML still require additional molecular methods for detection.

ACKNOWLEDGEMENTS

REFERENCES

1. Mrozek K, Heerema NA, Bloomfield CD. Cytogenetics in acute leukemia. Blood Rev 2004;18(2):115-36.

2. Mrozek K, Bloomfield CD. Chromosome aberrations, gene mutations and expression changes, and prognosis in adult acute myeloid leukemia. Hematology Am Soc Hematol Educ Program 2006:169-77.

3. Estey E, Dohner H. Acute myeloid leukaemia. Lancet 2006;368(9550):1894-907.

4. Virtaneva K, Wright FA, Tanner SM, Yuan B, Lemon WJ, Caligiuri MA, 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 2001;98(3):1124-9.

5. Schoch C, Kohlmann A, Schnittger S, Brors B, Dugas M, Mergenthaler S, et al. Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. Proc Natl Acad Sci U S A 2002;99(15):10008-13.

6. Debernardi S, Lillington DM, Chaplin T, Tomlinson S, Amess J, Rohatiner A, 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 2003;37(2):149-58.

7. Kohlmann A, Schoch C, Schnittger S, Dugas M, Hiddemann W, Kern W, et al. Molecular characterization of acute leukemias by use of microarray technology. Genes Chromosomes Cancer 2003;37(4):396-405.

8. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

9. Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004;350(16):1605-16.

10. Haferlach T, Kohlmann A, Schnittger S, Dugas M, Hiddemann W, Kern W, et al. Global approach to the diagnosis of leukemia using gene expression profiling. Blood 2005;106(4):1189-98.

11. Vey N, Mozziconacci MJ, Groulet-Martinec A, Debono S, Finetti P, Carbuccia N, et al. Identification of new classes among acute myelogenous leukaemias with normal karyotype using gene expression profiling. Oncogene 2004;23(58):9381-91.

12. Bullinger L, Rucker FG, Kurz S, Du J, Scholl C, Sander S, et al. Gene-expression profiling identifies distinct subclasses of core binding factor acute myeloid leukemia. Blood 2007;110(4):1291-300.

13. Valk PJ, Bowen DT, Frew ME, Goodeve AC, Lowenberg B, Reilly JT. Second hit mutations in the RTK/RAS signaling pathway in acute myeloid leukemia with inv(16). Haematologica 2004;89(1):106.

14. Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, Meijer J, van Oosterhoud S, van Putten WL, Valk PJ, et al. Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML. Hematol J 2003;4(1):31-40.

15. Care RS, Valk PJ, Goodeve AC, Abu-Duhier FM, Geertsma-Kleinekoort WM, Wilson GA, et al. Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. Br J Haematol 2003;121(5):775-7.

16. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002;99(10):6567-72.

17. Wouters BJ, Jorda MA, Keeshan K, Louwers I, Erpelinck-Verschueren CA, Tielemans D, et al. Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. Blood 2007;110(10):3706-14.

18. Verhaak RG, Goudswaard CS, van Putten W, Bijl MA, Sanders MA, Hugens W, 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 2005;106(12):3747-54.

19. Alcalay M, Tiacci E, Bergomas R, Bigerna B, Venturini E, Minardi SP, et al. Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. Blood 2005;106(3):899-902.

20. Schnittger S, Schoch C, Kern W, Mecucci C, Tschulik C, Martelli MF, et al. Nucleophosmin gene mutations are predictors of favorable prognosis in acute myelogenous leukemia with a normal karyotype. Blood 2005;106(12):3733-9.

21. Falini B, Mecucci C, Tiacci E, Alcalay M, Rosati R, Pasqualucci L, et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. N Engl J Med 2005;352(3):254-66.

22. Dohner K, Schlenk RF, Habdank M, Scholl C, Rucker FG, Corbacioglu A, et al. Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. Blood 2005;106(12):3740-6.

23. Lugthart S, van Drunen E, van Norden Y, van Hoven A, Erpelinck CA, Valk PJ, et al. High EVI1 levels predict adverse outcome in acute myeloid leukemia: prevalence of EVI1 overexpression and chromosome 3q26 abnormalities underestimated. Blood 2008;111(8):4329-37.

CHAPTER

4

Tribbles homolog 2 inactivates C/EBPα and causes acute myelogenous leukemia

Karen Keeshan[1], Yiping He[1], Bas J. Wouters[2], Olga Shestova[1], Lanwei Xu[1], Hong Sai[1], Carlos G. Rodriguez[1], Ivan Maillard[1,3], John W. Tobias[4], Peter Valk[2], Martin Carroll[3], Jon C. Aster[5], Ruud Delwel[2] and Warren S. Pear[1]

[1] Department of Pathology and Laboratory Medicine, Abramson Family Cancer Research Institute, Institute for Medicine and Engineering, University of Pennsylvania, Philadelphia, PA, USA

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

[3] Division of Hematology and Oncology, University of Pennsylvania, Philadelphia, PA, USA

[4] University of Pennsylvania Bioinformatics Core, University of Pennsylvania, Philadelphia, PA, USA

[5] Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA

ABSTRACT

Tribbles homolog 2 (Trib2) was identified as a downregulated transcript in leukemic cells undergoing growth arrest. To investigate the effects of Trib2 in hematopoietic progenitors, mice were reconstituted with hematopoietic stem cells retrovirally expressing Trib2. Trib2-transduced bone marrow cells exhibited a growth advantage ex vivo and readily established factor-dependent cell lines. In vivo, Trib2-reconstituted mice uniformly developed fatal transplantable acute myelogenous leukemia (AML). In mechanistic studies, we found that Trib2 associated with and inhibited C/EBPα. Furthermore, Trib2 expression was elevated in a subset of human AML patient samples. Together, our data identify Trib2 as an oncogene that induces AML through a mechanism involving inactivation of C/EBPα.

INTRODUCTION

Acute myelogenous leukemia (AML) is a genetically and phenotypically heterogeneous disease that is characterized by a block in myeloid differentiation, as well as enhanced proliferation and survival (1). AML is frequently associated with chromosomal translocations that target transcription factors such as members of the core binding factor (CBF) family, resulting in fusion proteins that include AML1/ETO [t(8;21)], CBFb/SMMHC [inv(16)], and TEL/AML1 [t(12;21)]. Mutations in transcription factors such as *PU.1*, *CEBPA*, *AML1*, and *GATA*-1 are also associated with AML, as are other types of oncogenic perturbations that lead to functional inactivation of critical transcription factors (2-4). Recurrent involvement of a limited set of transcription factors suggests that disruption of these genes is rate limiting for leukemia development.

Tribbles was first identified in *Drosophila* as a gene that is required for gastrulation, oogenesis, and viability (5-7). Tribbles coordinates cell division during gastrulation by promoting turnover of the cell cycle protein String/CDC25, thereby inhibiting premature mitosis. In this context, loss of Tribbles function was associated with increased proliferation, whereas overexpression of Tribbles slowed the cell cycle (5). Tribbles also promotes the degradation of slbo, the *Drosophila* homolog of the C/EBP family of basic region-leucine zipper transcription factors, during oogenesis (8). Based on amino acid sequence, Tribbles resembles a serine-threonine kinase; however, it has a variant catalytic core that lacks a canonical ATP binding site, and how it functions is unknown.

Mammals have three homologs of *Tribbles* – *Trib1*, *Trib2*, and *Trib3*. During fasting conditions, Trib3 negatively regulates AKT in the liver (9) and binds the E3 ligase COP1 to degrade ACC in adipose tissue (10). It is also a transcriptional target of the nuclear hormone receptor PPARα (11). These findings are consistent with a function downstream of the insulin receptor. Trib3 is also upregulated in tumor cells and in hypoxic conditions (12). Less is known about the other two mammalian Trib family members. When overexpressed, Trib1 can inhibit MEKK-1-mediated activation of AP-1 and stress kinase signaling (13,14). Molecular data for Trib2, originally identified as a phosphoprotein in canine thyroid cells (15), are limited to a study identifying it as a possible autoantigen in autoimmune uveitis (16) and as a differentially expressed gene in kidney mesenchymal cells and in a prostate cancer cell line (17,18).

To investigate Trib2's function in hematopoiesis, we retrovirally expressed it in murine hematopoietic progenitors, which were studied for changes in growth and differentiation in vivo and in vitro. This led to long-term growth of myeloid progenitors in vitro and induction of AML in vivo. Our biochemical studies suggest that an important function of Trib2 in transformation is to inactivate C/EBPα. Furthermore, an analysis of microarray data generated from 285 AML patient samples identified elevated Trib2 expression in a distinct subset of

patients in a cluster with a high frequency of *CEBPA* mutations. Together, these data identify Trib2 as a leukemogenic oncogene with the capacity to dysregulate C/EBPα signaling.

MATERIALS AND METHODS

Constructs and retroviruses

A 1032 bp fragment encoding the entire murine Trib2 cDNA was subcloned (6N terminus FLAG tag) into pcDNA3.1, pcDNA3.1/myc-HIS plasmid, or one of two different versions of MigR1 (one coexpressing GFP, the other truncated human nerve growth factor receptor [tNGFR]). C/EBPα rat cDNA (19) was cloned into the pcDNA3.1/myc-HIS plasmid. The MigR1-C/EBPαER construct was a kind gift of Dr. S. Nimer. The IL-12 p40 promoter containing the B6 genomic fragment 2700 to +54 of the IL-12 p40 gene was cloned into the pGL3-basic vector (Promega), and site-directed mutagenesis of the C/EBP binding site (293 to 289) was performed using the QuickChange kit (Stratagene) according to the manufacturer's instructions (kind gifts from Dr. R Carmody).

32D differentiation assay

32D cells were transduced with C/EBPαER (GFP), Trib2 (tNGFR + FLAG), or both, sorted for GFP and/or tNGFR 48 hr posttransduction, and cultured in DMEM without phenol red (GIBCO) containing 10% charcoal stripped FCS (Gemini Bioproducts) and 5 ng/ml rIL-3. Cells (1×10^5) were plated in 5 ng/ml IL-3 or 25 ng/ml G-CSF ± 1 μM β-estradiol (Sigma, E2257) and assessed for granulocytic differentiation by FACS analysis and morphological criteria.

Methylcellulose clonogenic assays

Sorted GFP+ and GFP+/Lin- (CD3, CD4, CD8, B220, Gr-1, Ter119, CD19, MHCII, IL-7Rα) BM cells from MigR1 and Trib2 transplanted mice were plated in triplicate in methylcellulose media (Methocult M3231, Stem Cell Technologies) supplemented with cytokines (10 ng/ml GM-CSF, 10 ng/ml IL-3, 10 ng/ml IL-6, 50 ng/ml SCF [Peprotech and BD Pharmingen]). Cells (15,000) from primary plates (colonies scored at 9 days) were transferred to secondary (colonies scored at 10 days) and tertiary (colonies scored at 8 days) plates containing the indicated cytokines, then transferred to RPMI liquid culture supplemented with 10% FBS and cytokines. Single colonies were transferred from primary plates to media containing IL-3, IL-6, and SCF in 24-well plates and assessed for growth after 11 days.

Bone marrow transduction and transplantation

C57BL/6 mice (B6) were obtained from Taconic Farms. Experiments were performed according to guidelines from the National Institutes of Health and with an approved protocol from the University of Pennsylvania Animal Care and Use Committee. Retroviral transduc-

tion of B6 BM cells and subsequent transfer of these cells into lethally irradiated recipients were performed as described (20), and details are provided in the Supplemental Experimental Procedures. Seven independent transplant experiments were performed with different Trib2 retroviral constructs (MigR1-Trib2, MigR1FLAG-Trib2, NGFR-FLAG-Trib2); each gave similar results and are summarized together in Figure 2A. Secondary transplants were performed by injecting 2 x 10⁶ nucleated BM or spleen cells from the primary leukemic mice into sublethally irradiated (600 rads) B6 mice. Histology and immunohistochemical staining details are provided in the Supplemental Data.

Flow cytometry
Cell suspensions were stained in PBS/2% FBS after blocking with nonspecific rat/mouse IgG (Sigma). Cells were sorted on a MoFlo (Cytomation) cell sorter. Analytical flow cytometry was performed on a FACS Calibur (Becton Dickinson) and analyzed using FlowJo software (Treestar). Antibodies used are described in the Supplemental Experimental Procedures. Dead cells were excluded by a combination of forward and side scatter properties.

Immunoprecipitation, western blotting, and ³⁵S-methionine pulse-chase
293T cells were transfected with 3 mg pcDNA3.1/myc-HIS-Trib2 and 2 mg MigR1-C/EBPα. After 36 hr, cotransfected cells were treated with 10 mM MG132 for 1 hr and 10 mM N-ethylmaleimide (both Calbiochem) for 30 s, washed in 13 PBS containing N-ethylmaleimide, and lysed with modified RIPA buffer (50 mM Tris [pH 8.0], containing 0.5% NP-40, 0.25% sodium deoxycholate, 150 mM NaCl, 1 mM EDTA, 1 mM PMSF, 1 mM Na3Vo4,1mM NaF, and 20 mM N-ethylmaleimide) supplemented with protease inhibitors (Complete EDTA-free; Roche). Precleared lysates (3 mg) were incubated overnight with 10 ml of protein A beads coated with 5 mg anti-MYC 9E10 antibody. To detect C/EBPα in leukemic samples, cells were lysed directly in 23 SDS buffer. To detect Flag-tagged Trib2 in leukemic samples, protein lysates were prepared with modified RIPA buffer. Antibodies used were anti-C/EBPα (Sc-61, Santa Cruz), anti-HA (HA.11, Covance), anti-FLAG (M2, Sigma), and anti-MYC (9E10). In pulse-chase experiments, U937 cells transduced with MigR1 and Trib2 were starved for 30 min, pulse-labeled for 60 min with ³⁵S-methionine (1 mCi/ml), and then chased for up to 180 min. Precleared protein lysates were incubated overnight with protein G beads coated with 5 mg C/EBPα antibody (SC-61x). Western blotting was performed according to standard procedures.

Luciferase reporter assay
RAW264.7 macrophages were transfected with various combinations of pcDNA control vector, vectors encoding Trib2 (390 ng) or C/EBPα (10 ng), IL-12 p40 (100 ng) wild-type and mutant promoter-luciferase (firefly) constructs, and the internal control Renilla-luciferase plasmid pRL-TK (10 ng; Promega). Firefly and Renilla luciferase activities were measured in

whole-cell lysates using the Dual-Luciferase reporter assay kit (Promega), as per the manufacturer's instructions.

Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSAs) were performed as described (19), except that nuclear extracts were prepared using a Nuclear Extract Kit (Active Motif) as per the manufacturer's instructions. The G-CSF receptor promoter oligonucleotide (C/EBP site is underlined) had the sequence 50-AAGGTGTT<u>GCAAT</u>CCCCAGC-30. The Oct-1 consensus oligonucleotide was obtained from Santa Cruz. Two microliters of C/EBPα (SC-61x) was used for supershift experiments.

Quantitative RT-PCR

RNA, isolated using the RNEasy kit (Qiagen), was digested with DNaseI and used for reverse transcription according to the manufacturer's instructions (Superscript II kit, Invitrogen). Validated human Trib2 and 18s rRNA primer/probe sets with TaqMan Universal PCR Master Mix (Applied Biosystems) and murine Trib2 primers (forward, 50-AGCCCGACT-GTTCTACCAGA-30; reverse, 50-AGCGTCTTCCAAACTCTCCA-30) with SYBR GREEN PCR Master mix (Applied Biosystems) were used for qRT-PCR and analyzed on the ABI Prism 7900 sequence detection system (Applied Biosystems).

DNA analysis

Southern blotting was performed according to standard procedures using QuikHyb (Stratagene) buffer and a ^{32}P-labeled IRES probe.

RNA interference

Human Trib2 siRNA (Silencer Validated siRNA #1060), murine Trib2 siRNA (Silencer Pre-designed siRNA #169360-exon2), and control Silencer Negative Control #1 were obtained from Ambion. U937 parental cells (2×10^6) (expressing human Trib2) and U937 cells that were transduced with Trib2 (murine) in 100 ml of Amaxa solution (Nucleofector Kit V, Amaxa Biosystems) were mixed with 100 nM of siRNA and electroporated using the Amaxa Nucleofector. Trib2 (human and murine) mRNA expression and C/EBPα protein expression were assessed 24 hr later.

RESULTS

Trib2 perturbs myeloid development in vivo

We identified *Trib*2 as a transcript that was downregulated after treatment of Notch-dependent murine T-ALL cell lines with γ secretase inhibitors (unpublished data). As little was known

about Trib2 function, we initiated studies to determine its effects on hematopoiesis. Lethally irradiated C57BL/6 (B6) mice were reconstituted with Trib2 or retroviral control vector (MigR1) transduced hematopoietic progenitors as described (20). At 9–14 weeks post-bone marrow transplant (BMT), MigR1 and Trib2 chimeric mice were similar in terms of bone marrow (BM) and splenic cellularity, the overall percentage of GFP+ cells, and the peripheral white blood cell (WBC) counts (Figures 1A and 1B, left panels). However, granulocytes (CD11b+Gr-1hi) were reduced in the GFP+ BM population of Trib2 chimeric mice (Figures 1A and 1B, center panels), and monocytes (CD11b+F4/80+) were significantly increased in the GFP+ population of both the BM and spleen of Trib2 mice (Figures 1A and 1B, right pan-

Table C:

		MigR1	Trib2
IL-3	Primary	102 +/- 40	208 +/- 5
	Secondary	0	272 +/- 9
	Tertiary	0	++
	Liquid Culture	0	++
GM-CSF	Primary	138 +/- 25	303 +/- 23
	Secondary	0	53 +/- 7
	Tertiary	0	++
	Liquid Culture	0	++

Table D:

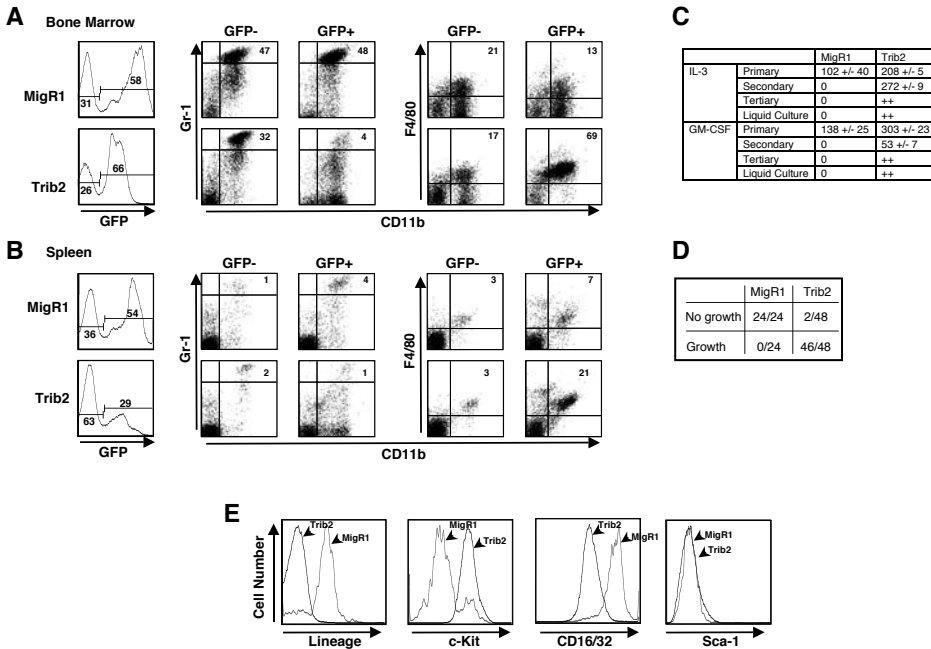|	MigR1	Trib2
No growth	24/24	2/48
Growth	0/24	46/48

Figure 1. Trib2 perturbs myeloid development in vivo and promotes self-renewal in vitro
B6 mice were lethally irradiated and reconstituted with BM cells transduced with control MigR1 vector or Trib2. Flow cytometric analysis of MigR1 and Trib2 chimeric mice at 9–14 weeks post-BMT. A and B: BM (A) and spleen (B) showing GFP positivity and Gr-1hiCD11b+ neutrophils (percentages given) in the GFP+ and GFP- fractions (left panel). In the right panel, F4/80+CD11b+ monocytes (percentages given) in the GFP+ and GFP- fractions are shown. Results are representative of three independent experiments using six mice.
C. GFP+ BM cells (25,000) sorted from MigR1 and Trib2 chimeric mice at week 9 post-BMT were plated in methylcellulose in the presence of the indicated cytokines. Colonies with >50 cells were scored after primary plating (9 days), secondary replating (15,000 cells, 10 days), tertiary replating (15,000 cells, 8 days), and growth in liquid culture (8 days) in triplicate. The mean numbers of colonies, 6SEM, for each condition are tabulated. "++" indicates colony growth or growth in liquid culture.
D. Five thousand sorted GFP+, lin- BM cells from Trib2 and MigR1 chimeras were plated in triplicate in methylcellulose containing IL-3, IL-6, SCF, and GM-CSF. Single colonies from primary MigR1 and Trib2 plates at day 9 were replated in liquid culture plus cytokines (IL-3, IL-6, SCF) in 24-well plates and assessed for growth after 11 days (growth = cell expansion and yellow media).
E. FACS analysis of MigR1 and Trib2 cells from a representative liquid culture in D.

els), suggesting that enforced expression of Trib2 promotes monocytic differentiation and inhibits granulocytic differentiation. Further characterization of myeloid subsets showed the presence of increased numbers of CD11b[+] D11c[+] MHCII[+] dendritic cells (DCs) and CD11b[+] F4/ 80[+] MHCII[+] macrophages in the GFP[+] population of Trib2 chimeric mice as compared to MigR1 control mice (Figures S1A and S1B in the Supplemental Data). In contrast to myeloid development, Trib2-GFP[+] lymphoid cells obtained from the BM, lymph nodes, and thymus did not reveal developmental abnormalities (data not shown). These experiments suggest that enforced Trib2 expression perturbed myeloid development.

Trib2-transduced progenitors exhibit a growth advantage in vitro and establish long-term myeloid progenitor cell lines

To assay the proliferative potential of Trib2-transduced cells in vitro, sorted GFP[+] cells from the BM of MigR1 and Trib2 chimeras at 9–10 weeks post-BMT were plated in methylcellulose in the presence of IL-3 or GM-CSF. Significant differences in colony numbers were observed in both IL-3 and GM-CSF (Figure 1C). When replated, Trib2 cells formed secondary colonies in both IL-3 and GM-CSF, whereas MigR1 primary colonies did not. Cells from Trib2 secondary colonies formed tertiary colonies, and cells transferred from these colonies to liquid culture exhibited factor-dependent (IL-3 and GM-CSF) long-term growth (Figure 1C). Lineage-depleted GFP[+] BM progenitor cells sorted from MigR1 and Trib2 chimeras were then assayed for proliferation. Single colonies were randomly picked from primary methylcellulose plates and replated in liquid media containing IL-3, IL-6, and SCF. Cells from MigR1 colonies did not proliferate in liquid medium, whereas cells from 46/48 Trib2 colonies exhibited factor-dependent long-term growth in liquid cultures (Figure 1D). These Trib2-expressing cell lines remained factor dependent and did not produce leukemia when injected into recipient mice (data not shown). The proliferating Trib2 cells did not express lineage markers (Gr-1, Ter119, CD3, CD8, CD4, B220, CD19, CD11c, CD11b) or Sca-1 but did express c-Kit and CD16/32 (FcgIII/II) as compared to MigR1 control cells (Figure 1E). These data suggest that Trib2 expression in myeloid progenitors promotes replating and self-renewal activities.

Mice reconstituted with Trib2 bone marrow cells develop AML

Trib2 chimeric mice died prematurely at a median of 179 days post-BMT (Figure 2A and Table S1, one outlier mouse of n = 26). These data include mice from multiple independent BMT experiments performed with different Trib2 retroviral constructs (MigR1-Trib2, MigR1-Trib2-Flag-tagged, tNGFR-Trib2-Flag-tagged). The peripheral blood (PB) prior to death exhibited a leukocytosis consisting of blasts and immature myelomonocytic cells (Figure 2C and Table S1). At necropsy, all mice displayed splenomegaly and lymphadenopathy (Figure 2B and Table S1). Morphologically, the infiltrating cells in these tissues resembled those seen in the PB; mature granulocytes were notably absent (Figure 2C). Histologic
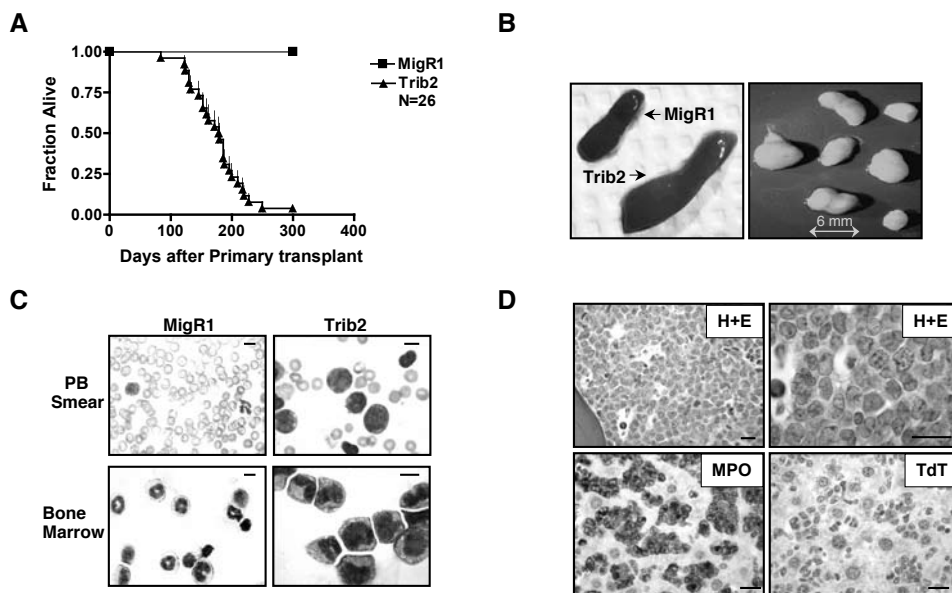
Figure 2. Trib2 induces AML

A. Kaplan-Meier survival curve of mice receiving Trib2-transduced BM compared to MigR1 control. The median survival of Trib2 mice was 179 days. Results are derived from seven independent experiments.

B. Representative photographs of splenomegaly in Trib2 mice compared to control MigR1 spleen, and lymphade-nopathy in Trib2 mice.

C. Wright-Giemsa-stained PB and BM single cell suspensions from MigR1 and leukemic Trib2 mice. Scale bars (upper right) represent 10 µm. The percentage of GFP+ cells in Trib2 BM was approximately 90%–100%.

D. Histopathology of BM sections from Trib2-induced AML. Hematoxylin and eosin (H+E) section showing hyper-cellularity (top left) due to the presence of sheets of immature cells and blasts (top right). The tumor cells stain positively for myeloperoxidase (MPO, bottom left) and negatively for terminal deoxytransferase (TdT, bottom right). Scale bars (lower right) represent 20 µm.

(A full color version of this figure van be found in the color section.)

sections revealed extensive (>80% in all cases) involvement of the BM by myeloblasts and immature myelomonocytic cells (Figure 2D, top panels). Blasts made up >20% of the BM cellularity in all affected animals, supporting the diagnosis of AML. Leukemic cells found in the spleen, liver, lymph nodes, and blood had similar morphologic and immunophenotypic characteristics (data not shown). By immunohistochemical staining, these cells were positive for myeloperoxidase (a marker of myeloblasts and more mature granulocytes) and were negative for terminal deoxytransferase (a marker of lymphoblasts) (Figure 2D, lower panels). Cytochemical stains for nonspecific esterase, a marker of monocytes, were negative (data not shown).

Flow cytometric analysis revealed that the leukemic cells were GFP+ and relatively large in size (as judged by forward scatter, Figure 3A) and expressed intermediate levels of Gr-1 and CD11b (Figure 3B), as well as c-Kit and F4/80 (Figure 3C). The coexpression of Gr-1 and CD11b at low/intermediate levels is characteristic of immature myeloid precursors and
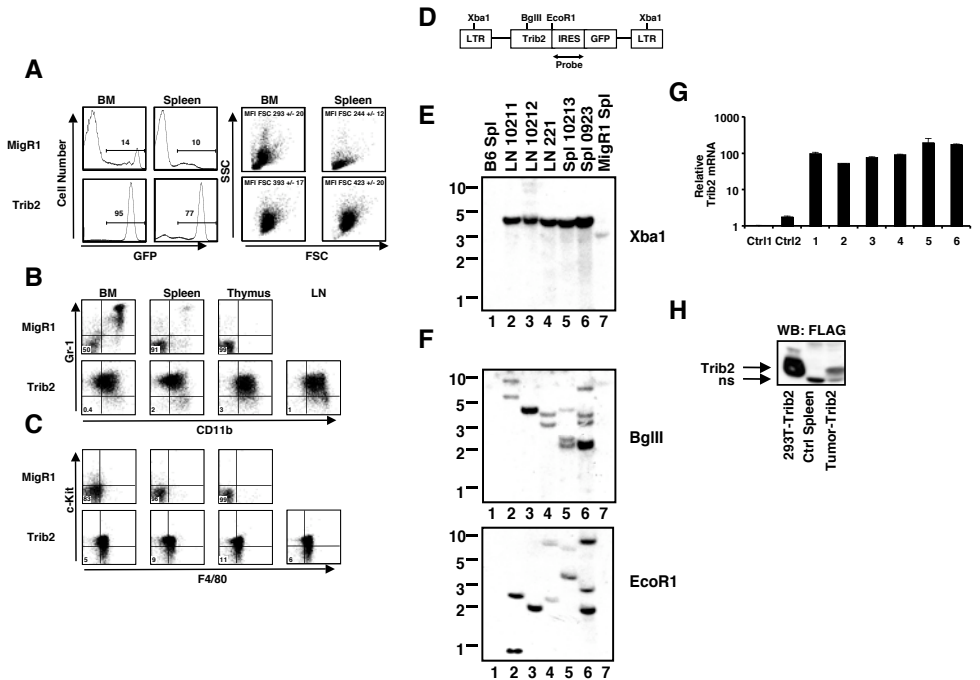
Figure 3. Analysis of Trib2-induced AML

A. Flow cytometric analysis of GFP expression (percentages given), forward scatter (FSC), and side scatter (SSC) in cells obtained from BM and spleens of leukemic mice. Mean fluorescence intensities (MFI) of FSC of the GFP$^+$ cells are shown ± SD.

B and C. Immunophenotypes of primary Trib2-induced AML cells from lymph nodes (LN), thymus, spleen, and BM compared to control MigR1 cells. Flowcytometric analysis of Gr-1 and CD11b expression (B) and c-Kit and F4/80 expression (C), in the GFP$^+$ fraction of MigR1 (top) and Trib2 (bottom panels) mice. The percentages of cells that are Gr-1$^-$CD11b$^-$ are shown in B, and the percentages of cells that are c-Kit$^-$F4/80$^-$ are shown in C. Results are representative of seven independent experiments.

D. Schematic overview of Trib2 provirus and probe used for Southern blotting.

E and F. Southern blot detection of proviruses in Trib2 mice. DNA preparations were digested with either Xba1 to detect intact provirus (w4 kb) (E) or BglII or EcoR1, which cleaved once in the provirus, to detect different proviral integrants (F). All leukemic samples were obtained from different primary mice; lanes 1 and 7 contain control DNAs from B6 mouse spleen (Spl) and MigR1 control spleen, respectively.

G. Real-time RT-PCR analysis of Trib2 expression in cDNAs from six primary Trib2-induced AML tumor spleen samples (1–6), using murine Trib2-specific primers. The results are presented as Trib2 mRNA in AML samples relative to expression in control MigR1 spleen cells (Ctrl1, Ctrl2), normalized for 18S rRNA content. Error bars denote ±SD of each sample measured in triplicate.

H. Analysis of Trib2 protein expression by western blot. Extracts of primary Trib2-induced AML cells obtained from spleen (>90% GFP, MigR1-Trib2-FLAG) were compared to extracts prepared from control MigR1 spleen cells and 293T cells transfected with pcDNA-Trib2-FLAG. Trib2-FLAG polypeptides were detected on the blot with the FLAG antibody (M2, Sigma). ns, nonspecific band.

murine myeloid leukemias (21,22). The overall combination of morphologic and immuno-phenotypic findings is consistent with AML.

To detect provirus, DNA was isolated from the lymph nodes and spleens of chimeric mice and analyzed by digestion with restriction endonucleases (Figure 3D). These studies revealed

that all tumors contained intact provirus (Figure 3E) and that individual tumors contained one to several distinct proviral insertions (Figure 3F). The intensity of the bands indicated that the tumors were either monoclonal (Figure 3F, lanes 2, 3, and 4) or biclonal (Figure 3F, lanes 5 and 6), suggesting that additional genetic events are necessary to generate Trib2-induced AML. In addition to GFP (Figure 3A), the tumor cells expressed both Trib2 mRNA (Figure 3G) and protein (38 kDa, Figure 3H). Because antibodies against Trib2 are not available, the latter was detected in a cohort of tumors induced with epitope-tagged Trib2 retrovirus.

To further establish the malignancy of Trib2 leukemia, cells (2×10^6) from the BM and spleens of primary leukemic mice were transplanted into sublethally irradiated secondary recipients. All secondary recipients developed AML with an average latency of 36 days (Figure 4A and Table S2). Secondary leukemias were associated with leukocytosis and extensive involvement of the BM, liver, spleen, and other tissues (Figure 4B and Table S2). The immunophenotype of the leukemic cells in secondary recipients resembled that of the primary disease, with most cells retaining intermediate expression of Gr-1 and CD11b (Figure 4B) and expression of F4/80 (Figure 4C). In addition, the leukemic cells were c-Kit+, Sca-1-, CD34[intermediate], and CD16/ 32+ (FcgIII/II) (Figure 4D). Together, these data indicate that Trib2 is a potent leukemogen.

Trib2 reduces normal C/EBPα levels and inhibits C/EBPα function

A deficiency of the transcription factor C/EBPα inhibits granulocyte differentiation and is associated with AML (4), features similar to the consequences of Trib2 expression in hematopoietic progenitors. Furthermore, *Drosophila* Tribbles can promote the degradation of Slbo, the *Drosophila* homolog of C/EBP, during oocyte border cell migration (8). To determine whether Trib2 alters C/EBPα protein levels in myeloid cells, western blot analysis was performed on whole-cell extracts prepared from 32D and U937 cells transduced with MigR1 or Trib2. In both cell lines, Trib2 expression reduced levels of the C/EBPαp42 full-length protein and concomitantly increased levels of C/EBPα p30 (Figure 5A), a dominant-negative variant of C/EBPαp42 that is thought to arise from an internal translation initiation site (23).

Human AML is associated with mutations in C/EBPα that lead to decreased levels of normal C/EBPαp42 and increased levels of C/EBPαp30 (24,25). C/EBPαp30 dominantly inhibits C/EBPα p42 function when the ratio of C/EBPα p42 to C/EBPα p30 is less than one (23). In contrast to normal CMPs and GMPs (26), high levels of C/EBPαp30 and low ratios of C/EBPαp42 to C/EBPαp30 were present in all Trib2 primary leukemic samples (Figure 5B, left panel). The inverted ratio of C/EBPαp42 to C/EBPαp30 protein persisted in serially passaged leukemic cells (Figure 5B, right panel). The identity of C/EBPαp30 was confirmed by performing additional western blots with an antibody specific for the N-terminal C/EBPα epitope, which detected only C/EBPαp42 (data not shown).
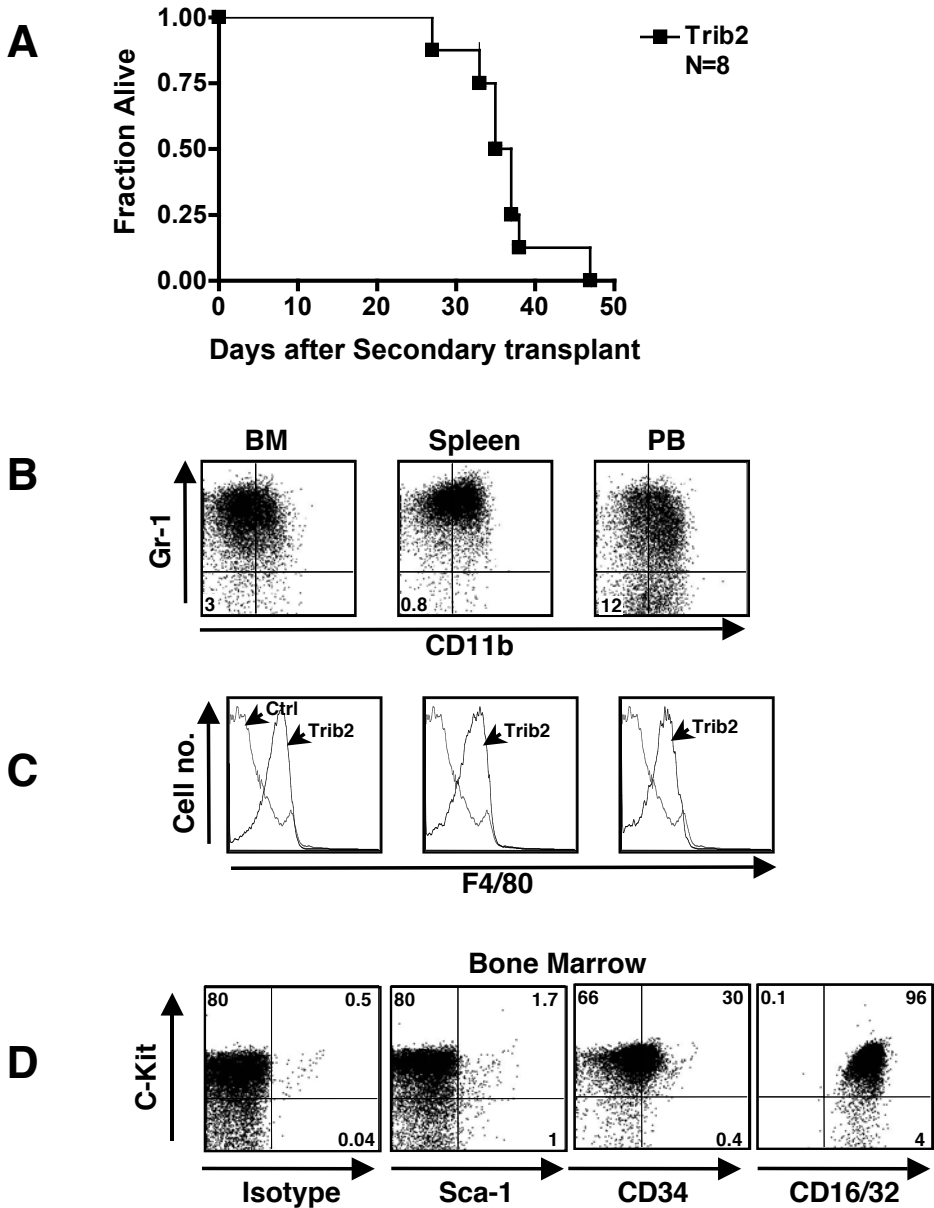
Figure 4. Trib2-induced AML is 100% transplantable

A. Kaplan-Meier survival curve of secondary transplants. Sublethally irradiated mice (600 rads) received 2×10^6 BM or spleen cells from eight primary leukemic mice.

B–D. Immunophenotype of Trib2-induced AML after secondary transplant. Cells from BM, spleen, and PB were assessed by flow cytometry. B: The Gr-1 and CD11b profile of the GFP+ population is shown; percentages given are cells negative for both markers. C: F4/80 profile of the GFP+ Trib2 cells (black line) are overlaid on the normal F4/80 staining profile of B6 control BM cells (gray line). Results shown are representative of n = 8 mice. D: The c-Kit and isotype/Sca-1/CD34/CD16/32 profile of the GFP+ population is shown.

Chapter 4

75

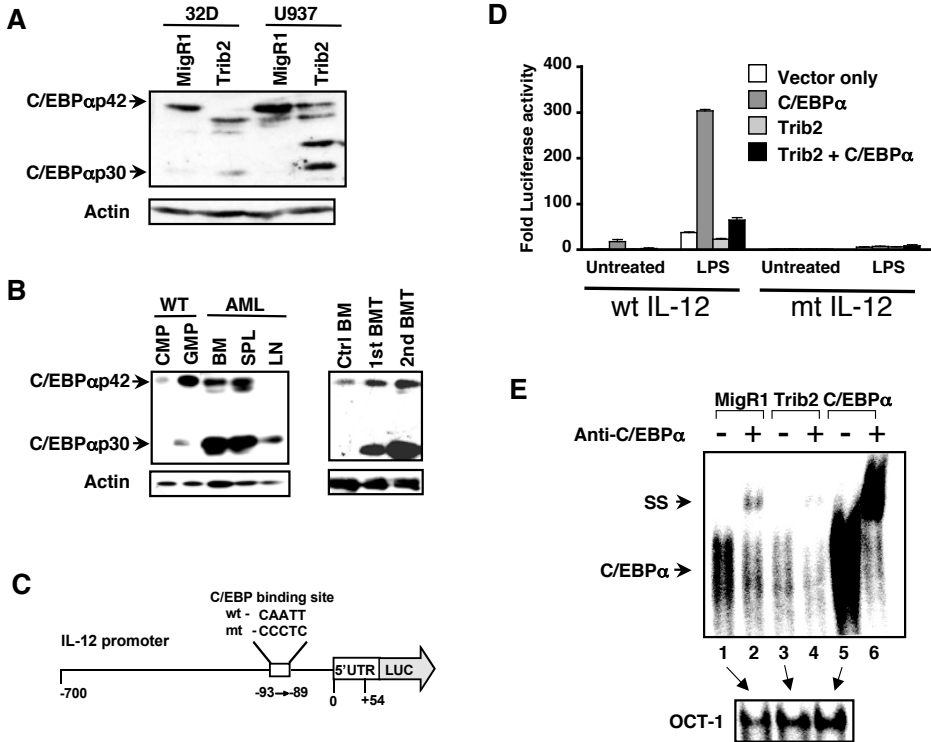**Figure 5. Trib2 reduces normal levels of C/EBPα and inhibits its activity**

A. Sorted GFP+ 32D and U937 cells transduced with either MigR1 or Trib2 were assessed for C/EBPα protein expression by western blot. C/EBPα p42 is the full-length protein, and C/EBPα p30 is the truncated protein. Actin is the protein loading control.

B. Analysis of C/EBPα p42 and p30 protein expression in primary leukemic samples from BM (93% GFP+), spleen (63% GFP+), and lymph node (88% GFP+) compared to normal levels expressed in CMPs and GMPs from B6 BM (left panel). Levels of C/EBPα p42 and p30 proteins were also compared in unfractionated normal B6 mouse BM cells and primary (94% GFP+) and secondary (98% GFP+) leukemic BM samples (right panel).

C. Schematic overview of the IL-12 promoter containing the C/EBPα binding site.

D. RAW264.7 macrophages were cotransfected with (1) the IL-12 promoter firefly luciferase constructs containing the C/EBP wild-type or mutant binding sites; (2) either empty vector, Trib2, C/EBPα, or both Trib2 and C/EBPα; and (3) a pRL-TK Renilla luciferase internal control plasmid. Luciferase activity was measured following LPS (100 ng/ml) treatment for 8 hr, 24 hr posttransfection. Reporter luciferase activity for each sample was normalized to the Renilla luciferase activity of the same sample. Data presented are mean ± SD of triplicate cultures.

E: C/EBPα DNA binding activity was assessed by EMSA using a double-stranded oligonucleotide probe containing the C/EBP binding site from the human G-CSF receptor. Nuclear extracts from sorted GFP+ U937 cells transduced with MigR1 (lanes 1 and 2), Trib2 (lanes 3 and 4), or C/EBPα (lanes 5 and 6) were incubated with ^{32}P-labeled probe. In lanes 2, 4, and 6, 2 ml of C/EBPα antibody was added. "SS" indicates the supershifted C/EBPα complex. The extracts used in lanes 1, 3, and 5 were tested with an OCT-1 probe in a second EMSA assay as a control for the integrity and quantity of nuclear binding proteins.

To show that the relative decrease in C/EBPαp42 translated into diminished functional activity, we investigated whether Trib2 inhibited activation of an IL-12 promoter luciferase reporter by C/EBPα. This reporter contains a single C/EBPα binding site, which is involved

in the induction of IL-12 transcription in RAW cells following LPS exposure (Figure 5C). Trib2 inhibited the stimulation of an IL-12 reporter construct by both C/EBPαp42 and LPS (Figure 5D). In both experiments, the stimulation of reporter gene activity by C/EBPα or LPS required an intact C/EBPα binding site (Figure 5D). Trib2 had no effect on an NFkB-sensitive luciferase reporter gene (Figure S2A), suggesting that the effect of Trib2 was specific for C/EBPα. Furthermore, the IL-6 promoter also contains a functional C/EBPα binding site, and BM-DCs and macrophages derived from Trib2 chimeric mice released less IL-12 and IL-6 in response to LPS than did MigR1 control cells (Figures S2B–S2E). To further characterize the relationship between C/EBPα and Trib2, the C/EBPα DNA binding activity was assayed in extracts prepared from transduced U937 cells. The C/EBPα-DNA binding activity was markedly reduced by Trib2 (Figure 5E, lanes 1 and 3), with the specificity of the observed gel shift being confirmed by production of a "supershift" upon addition of an antibody against C/EBPα (Figure 5E, lane 2 and 4). These data demonstrate that Trib2 expression inhibits the DNA binding function of C/EBPαp42.

As Trib2 overexpression decreased C/EBPαp42, we investigated whether siRNA-mediated knockdown of endogenous Trib2 would increase endogenous C/EBPαp42. Indeed, human U937 cells transfected with siRNA against human Trib2 had reduced Trib2 mRNA expression (Figure 6A) and a concomitant increase in C/EBPα p42 protein (Figure 6B). C/EBPα p30 is not detectable in these cells (see Figure 6D). Moreover, siRNA-mediated knockdown of retro-virally expressed murine Trib2 in U937 cells reduced murine Trib2 expression (Figure 6C) and abrogated its inhibitory effect on C/EBPap42 (Figure 6D). Trib2 knockdown restored the C/EBPαp42:C/EBPαp30 ratio by increasing C/EBPαp42 expression. The failure to observe a concomitant decrease in C/EBPαp30 may stem from a long C/EBPαp30 half-life, residual Trib2 activity, or other unknown factors, and is a point that requires further investigation. Nevertheless, these data complement the Trib2 overexpression studies and further support the hypothesis that Trib2 opposes C/EBPαp42 function.

To further investigate the link between Trib2's inhibitory effects on granulopoiesis (Figure 1A) and C/EBPαp42 activity, we coexpressed Trib2 and C/EBPαp42 in 32D cells, which require IL-3 to survive and undergo C/EBPαp42-dependent granulocytic differentiation in response to G-CSF (27). To control C/EBPαp42 activity, we utilized an estradiol-inducible C/EBPαp42 fusion gene (C/EBPα-ER) cloned in MigR1 (28). In these experiments, Trib2 was expressed in a MSCV retrovirus that also bears a tNGFR gene as a surrogate marker (29). 32D cells were sorted for retroviral expression of GFP (C/EBPα-ER), tNGFR (Trib2), or both, treated with either estradiol or vehicle in the presence of either IL-3 or G-CSF, and then assayed for granulocytic differentiation by both flow cytometry (Figure 6E) and morphology (data not shown). As expected, vehicle-treated cells transduced with C/EBPα-ER differenti-ated only in the presence of G-CSF (induction of endogenous C/EBPα activity), whereas estradiol (induction of exogenous C/EBPα activity) induced the differentiation of these cells in the presence of either G-CSF or IL-3 (Figure 6E). Further, cells transduced with Trib2
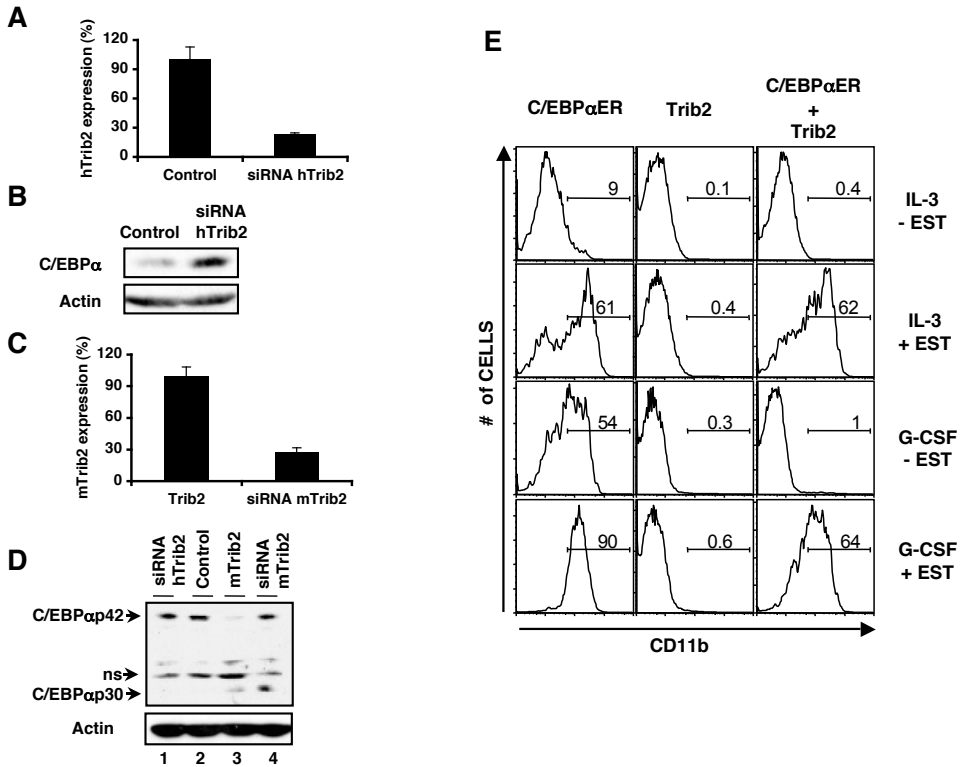
Figure 6. Trib2-dependent inhibition of C/EBPα expression and function is abrogated by siRNAmediated knockdown of Trib2 and overexpression of C/EBPα

A and B. siRNA designed to knock down human Trib2 (hTrib2) was electroporated into U937 cells and analyzed 24 hr later.

A. hTrib2 expression assessed by real-time RT-PCR. Error bars denote ±SEM of each sample measured in triplicate.

B. Western blot of C/EBPα protein expression. Lane 1, negative control siRNA; lane 2, hTrib2 siRNA. C/EBPαp30 is undetectable in parental U937 cells.

C and D. siRNA designed to knock down murine Trib2 (mTrib2) was electroporated into U937 cells retrovirally expressing murine Trib2 and analyzed 24 hr later. C: mTrib2 expression assessed by real-time RT-PCR. Error bars denote ±SEM of each sample measured in triplicate.

D. Western blot of C/EBPαp42 and C/EBPαp30 protein expression. Lanes 1 and 2, U937 cells + human Trib2 siRNA and negative control siRNA, respectively, as in B; lanes 3 and 4, U937-mTrib2 cells + negative control siRNA and siRNA mTrib2, respectively. ns, nonspecific band.

E. 32D cells transduced with C/EBPαER (GFP), Trib2 (tNGFR), or C/EBPαER + Trib2 were sorted for GFP and/or tNGFR expression. Sorted cells were plated in equal numbers (1 x 10^5) (day 0, 48 hr posttransduction) in IL-3 or G-CSF ± 1 μM β-estradiol (EST). CD11b expression was assessed after 2 days. Data are representative of three independent experiments.

failed to differentiate in response to either G-CSF or IL-3, as did cells that were cotransduced with Trib2 and C/EBPα-ER in the presence of vehicle. In contrast, the inhibitory effect of Trib2 in doubly transduced cells was overcome by estradiol treatment, indicating that high doses of C/EBPα overcome the inhibitory effects of Trib2.
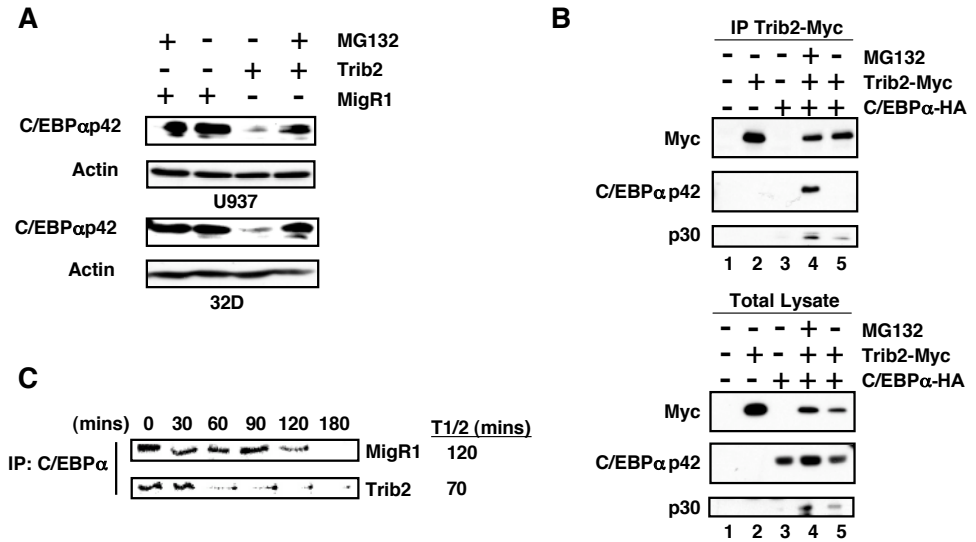
Figure 7. Trib2 forms a complex with C/EBPα and results in its proteasomal-dependent degradation.
A. Sorted GFP+ U937 (top panel) and 32D (lower panel) cells transduced with either MigR1 or Trib2, treated ± 10 mM MG132 for 2 hr, were assessed for C/EBPα expression by western blot.
B. 293T cells were transfected with empty vector (lane 1), myc-tagged Trib2 (lane 2), and HA-tagged C/EBPα (lane 3), or cotransfected with both (lanes 4 and 5), and treated with 10 mM MG132 for 2 hr (lane 4). Trib2 was immunoprecipitated using a Myc 9E10 antibody, and western blotting was performed with HA and Myc antibodies on immunoprecipitates (top panel) and total lysates (lower panel).
C. U937 cells transduced with MigR1 or Trib2 were metabolically labeled with ^{35}S-methionine for 60 min and chased for the indicated times. C/EBPα was immunoprecipitated, and radiolabeled protein was detected by SDS-PAGE. The half-life (T1/2) was calculated using ImageJ software.

Cell proliferation and survival were also assayed in these cultures by cell counting and trypan blue exclusion. 32D-C/EBPα-ER cells proliferated in IL-3, and this proliferation was inhibited by either endogenous or exogenous C/EBPα activity (Figure S3, left bars). 32D-Trib2 cells proliferated in the presence of IL-3 but died in the presence of G-CSF (Figure S3, middle bars), an effect that was similar to transduction of 32D cells with a C/EBPα DNA binding mutant that was unable to induce differentiation (19). Cell numbers/viability were restored following estradiol induction of C/EBPα in the Trib2 + C/EBPα-ER-expressing cells (Figure S3, compare two far right bars). Similar to the CD11b expression data, these results also show that high doses of C/EBPα overcome Trib2's inhibitory effects.

We next sought to determine the mechanism by which Trib2 decreased normal C/EBPp42 levels. In 32D and U937 cells, C/EBPαp42 levels in Trib2-expressing cells were restored by pretreatment with the proteasome inhibitor MG132 (Figure 7A). In coimmunoprecipitation experiments using 293T cells, Trib2 association with C/EBPαp42 was detected only in the presence of MG132, whereas C/EBPαp30 coimmunoprecipitated in the presence and absence of MG132 (Figure 7B, lane 4 and 5). Expression of Trib2 in U937 cells decreased the half-life of endogenous C/EBPα protein as determined by ^{35}S metabolic pulse labeling and

immunoprecipitation chase of labeled C/EBPα (Figure 7C). These data suggest that Trib2 acts by associating with C/EBPαp42 and promoting its degradation via the proteasome.

Elevated Trib2 expression is found in a subset of AML patients exhibiting C/EBPα defects

Since Trib2 is a potent inducer of AML in our experimental model, we investigated *TRIB2* expression in an mRNA expression database derived from 285 AML patient samples. Previous analysis of gene expression in this data set identified 16 groups of AML with distinct gene expression profiles (30). Based on signals from two different *TRIB2* probe sets, we noted that elevated *TRIB2* expression was observed in cluster #4 (Figure 8), which is one of two clusters that are associated with *CEBPA* mutations (30). Although cluster 4 contains AML samples with wild-type (green) or mutated (red) *CEBPA*, *TRIB2* expression was primarily elevated in tumors with wild-type *CEBPA* (Figure 8, enlarged area). Thus, in an unbiased screen of AML samples, high levels of *TRIB2* expression were preferentially found in a cohort associated with C/EBPα defects. This association again links TRIB2 to tumors with altered C/EBPα function and suggests that elevated TRIB2 may have a pathogenic role in a subset of human AML.

Figure 8. *TRIB2* is elevated in a subset of human AML
Correlation view of 285 AML patients (30). Colors of cells relate to Pearson's correlation coefficient values: red indicates higher positive and blue indicates higher negative correlation between samples. Sixteen clusters represented by red blocks along the diagonal can be identified. *CEBPA* mutation status is indicated next to each tumor (red, mutant; green, wild-type). Histograms next to each tumor represent expression levels of the two probe sets for *TRIB2*. Cluster 4, one of the two clusters harboring most patients with *CEBPA* mutations, has a significantly elevated expression of *TRIB2* relative to other clusters
(A full color version of this figure can be found in the color section.)

DISCUSSION

Our data provide strong evidence that dysregulated Trib2 expression contributes to the pathogenesis of AML. Retroviral expression of Trib2 immortalized hematopoietic progenitors in vitro and induced fatal transplantable AMLs in murine recipients with a mean latency of 179 days. In accordance with the central role of Trib2 in the development of these tumors, we have never observed AML in several hundred mice that have received MigR1-transduced BM cells using identical protocols. Microarray analysis of a cohort of primary human AMLs identified elevated *TRIB2* mRNA expression in a cluster of tumors associated with a high frequency of *CEBPA* mutations. Although we have not yet determined the mechanism of TRIB2 upregulation in these leukemias, our data linking Trib2 expression to inhibition of C/EBPα function suggests that TRIB2 overexpression is likely to be an important pathogenic mechanism in a subset of human AML.

Decreased C/EBPαp42 levels are observed in diverse molecular subtypes of human AML (31-35) and are sometimes associated with point mutations in C/EBPα (24). A number of observations suggest that Trib2-mediated alterations of C/EBPα function are likely to contribute to AML induction. We found that Trib2 forms a complex with C/EBPαp42 that results in its destabilization through proteasome-dependent degradation. We further demonstrated an inverse functional relationship between Trib2 and C/EBPα through both gain-and loss-of-function approaches. The effects of C/EBPα on myeloid cell differentiation are opposed by enforced Trib2 expression, and we found high *TRIB2* expression in a specific subset of human AMLs that is associated with *CEBPA* mutations. We hypothesize that Trib2-expressing AMLs lacking *CEBPA* mutations share a similar expression profile due to functional inactivation of C/EBPα. It is unlikely, however, that Trib2 causes AML only through C/EBPα inactivation. Mice bearing knockout and mutated C/EBPα alleles have abnormalities of myeloid differentiation but do not develop AML (36-38). Furthermore, despite its effect on C/EBPα function, Trib2 is not sufficient for AML, as the tumors arising in Trib2 mice were either mono-or biclonal.

In addition to C/EBPα, Trib2 may influence other signaling pathways. In *Drosophila*, Tribbles degrades String/CDC25 to inhibit the cell cycle. Although it seems counterintuitive that decreased String/CDC25 would be oncogenic, antagonism of String by Tribbles is associated with enhanced proliferation in certain contexts in the fly (5). Other Trib family members have been associated with Akt and MAPK signaling (13), and it remains to be determined whether Trib2 influences these pathways in AML cells.

It is also possible that proviral integration by the Trib2expressing retroviral vector created lesions in genes that cooperate with Trib. We have begun to address this by cloning proviral integrations in Trib2 tumors. In at least one Trib2 AML, the provirus was found adjacent to HoxA9, whose mRNA was markedly elevated in this tumor (K.K., unpublished data). HoxA9

has been shown to synergize with other oncogenes to accelerate AML development (39). We are currently investigating the functional significance of this finding.

An additional property of Trib2 is its association with increased C/EBPαp30 levels in cell lines and leukemic samples. *Cebpa* is an intronless gene, and it has been suggested that the two major C/EBPα isoforms stem from alternative translational start sites. Other possibilities exist, however, as the homologous protein C/EBPβ is proteolytically processed into full-length and N-terminally truncated dominant-negative isoforms (40-43). In our coimmunoprecipitation experiments, the C/EBPα expression construct lacked the uORF (upstream open reading frame) and contained a C-terminal HA tag that was detected with a HA antibody and verified with a specific C/EBPα antibody. This suggests that the 30 kDa protein is an N-terminal truncation of the full-length C/EBPα induced by Trib2 expression. However, this construct retains the alternative ATG start codon, and therefore translational control cannot be ruled out. In favor of this possibility, we were unable to detect C/EBPαp30 in the pulse-chase experiment.

Although Trib2 potently induced AML in our BMT model, it is not clear whether the diminution of C/EBPαp42, the enhancement of C/EBPαp30, or both were critical to the phenotypes that we observed. Decreased C/EBPαp42 levels are associated with human AML (2), and mice engineered to express a C/EBPαp42 mutant that fails to block the cell cycle develop a transplantable myeloproliferative disease that progresses to an AML-like phenotype but does not affect survival (44). Murine BM cells transduced with the dominant-negative C/EBPαp30 showed poor long-term engraftment (45). However, in CFU assays, the C/EBPαp30 dominant-negative was shown to block murine and human myeloid differentiation (45-47). Thus, both decreased C/EBPαp42 and increased C/EBPαp30 have been associated with myeloid transformation. Additional studies will be required to determine which of these isoforms is most relevant to Trib2 induced AML.

We initially identified Trib2 by screening for transcripts that were downregulated when Notch1-dependent T cell tumor cell lines were induced to growth arrest by γ secretase inhibitor treatment (48). Although we find that Trib2 expression is modulated by Notch1 via CSL binding sites in the Trib2 promoter (unpublished data), mice that retrovirally express activated Notch1 do not develop AML or exhibit major defects in granulocyte differentiation. Trib2 regulation by Notch1 may be context dependent and nonoperational in myeloid progenitors. However, Notch signaling promotes macrophage/DC differentiation in some assays (49,50), and it will be interesting to determine if this is Trib2 dependent. In addition, the subset of human AMLs associated with increased Trib2 expression are very immature and express T cell markers, such as TCRδ (30). Trib2 is highly expressed during several stages of T cell development (data not shown), and it is possible that Trib2 normally contributes to the expression of genes associated with T cell development.

Our data predict that Trib2 would functionally inactivate C/EBPα by causing its degradation. Thus, one expectation would be that it would associate with tumors in which C/EBPα

is inactivated. Analysis of microarray data obtained from 285 AMLs showed that elevated Trib2 expression preferentially associated with a cluster of AMLs characterized by C/EBPα defects. High Trib2 expression was predominantly seen in tumors without *CEBPA* mutations; however, it was also high in two tumors with these mutations (Figure 8). It may be that induction of this molecular class of AML requires C/EBPα function to be suppressed below a certain threshold, and that this may occur by mutation alone, Trib2 expression, or a combination of these two factors.

In summary, our studies identify Trib2 as an oncoprotein that contributes to the pathogenesis of AML through the inhibition of C/EBPa function. This occurs through the interaction of Trib2 with C/EBPα, resulting in the proteasomal-dependent degradation of C/EBPα. Our observations suggest that multiple leukemogenic mechanisms target C/EBPα in AML. Understanding the precise role of Trib2 in AML may lead to new diagnostic and therapeutic strategies for this aggressive human malignancy.

ACKNOWLEDGEMENTS

REFERENCES

1. Kelly LM, Gilliland DG. Genetics of myeloid leukemias. Annu Rev Genomics Hum Genet 2002;3:179-98.

2. Mueller BU, Pabst T. C/EBPalpha and the pathophysiology of acute myeloid leukemia. Curr Opin Hematol 2006;13(1):7-14.

3. Rosenbauer F, Koschmieder S, Steidl U, Tenen DG. Effect of transcription-factor concentrations on leukemic stem cells. Blood 2005;106(5):1519-24.

4. Tenen DG. Disruption of differentiation in human cancer: AML shows the way. Nat Rev Cancer 2003;3(2):89-101.

5. Mata J, Curado S, Ephrussi A, Rorth P. Tribbles coordinates mitosis and morphogenesis in Drosophila by regulating string/CDC25 proteolysis. Cell 2000;101(5):511-22.

6. Grosshans J, Wieschaus E. A genetic link between morphogenesis and cell division during formation of the ventral furrow in Drosophila. Cell 2000;101(5):523-31.

7. Seher TC, Leptin M. Tribbles, a cell-cycle brake that coordinates proliferation and morphogenesis during Drosophila gastrulation. Curr Biol 2000;10(11):623-9.

8. Rorth P, Szabo K, Texido G. The level of C/EBP protein is critical for cell migration during Drosophila oogenesis and is tightly controlled by regulated degradation. Mol Cell 2000;6(1):23-30.

9. Du K, Herzig S, Kulkarni RN, Montminy M. TRB3: a tribbles homolog that inhibits Akt/PKB activation by insulin in liver. Science 2003;300(5625):1574-7.

10. Qi L, Heredia JE, Altarejos JY, Screaton R, Goebel N, Niessen S, et al. TRB3 links the E3 ubiquitin ligase COP1 to lipid metabolism. Science 2006;312(5781):1763-6.

11. Koo SH, Satoh H, Herzig S, Lee CH, Hedrick S, Kulkarni R, et al. PGC-1 promotes insulin resistance in liver through PPAR-alpha-dependent induction of TRB-3. Nat Med 2004;10(5):530-4.

12. Bowers AJ, Scully S, Boylan JF. SKIP3, a novel Drosophila tribbles ortholog, is overexpressed in human tumors and is regulated by hypoxia. Oncogene 2003;22(18):2823-35.

13. Hegedus Z, Czibula A, Kiss-Toth E. Tribbles: novel regulators of cell function; evolutionary aspects. Cell Mol Life Sci 2006;63(14):1632-41.

14. Kiss-Toth E, Bagstaff SM, Sung HY, Jozsa V, Dempsey C, Caunt JC, et al. Human tribbles, a protein family controlling mitogen-activated protein kinase cascades. J Biol Chem 2004;279(41):42703-8.

15. Wilkin F, Suarez-Huerta N, Robaye B, Peetermans J, Libert F, Dumont JE, et al. Characterization of a phosphoprotein whose mRNA is regulated by the mitogenic pathways in dog thyroid cells. Eur J Biochem 1997;248(3):660-8.

16. Zhang Y, Davis JL, Li W. Identification of tribbles homolog 2 as an autoantigen in autoimmune uveitis by phage display. Mol Immunol 2005;42(11):1275-81.

17. Bisoffi M, Klima I, Gresko E, Durfee PN, Hines WC, Griffith JK, et al. Expression profiles of androgen independent bone metastatic prostate cancer cells indicate up-regulation of the putative serine-threonine kinase GS3955. J Urol 2004;172(3):1145-50.

18. Takasato M, Osafune K, Matsumoto Y, Kataoka Y, Yoshida N, Meguro H, et al. Identification of kidney mesenchymal genes by a combination of microarray analysis and Sall1-GFP knockin mice. Mech Dev 2004;121(6):547-57.

19. Keeshan K, Santilli G, Corradini F, Perrotti D, Calabretta B. Transcription activation function of C/EBPalpha is required for induction of granulocytic differentiation. Blood 2003;102(4):1267-75.

20. Pui JC, Allman D, Xu L, DeRocco S, Karnell FG, Bakkour S, et al. Notch1 expression in early lymphopoiesis influences B versus T lineage determination. Immunity 1999;11(3):299-308.

21. Cozzio A, Passegue E, Ayton PM, Karsunky H, Cleary ML, Weissman IL. Similar MLL-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors. Genes Dev 2003;17(24):3029-35.

22. de Guzman CG, Warren AJ, Zhang Z, Gartland L, Erickson P, Drabkin H, et al. Hematopoietic stem cell expansion and distinct myeloid developmental abnormalities in a murine model of the AML1-ETO translocation. Mol Cell Biol 2002;22(15):5506-17.

23. Calkhoven CF, Muller C, Leutz A. Translational control of C/EBPalpha and C/EBPbeta isoform expression. Genes Dev 2000;14(15):1920-32.

24. Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, et al. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. Nat Genet 2001;27(3):263-70.

25. Leroy H, Roumier C, Huyghe P, Biggio V, Fenaux P, Preudhomme C. CEBPA point mutations in hematological malignancies. Leukemia 2005;19(3):329-34.

26. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature 2000;404(6774):193-7.

27. Wang X, Scott E, Sawyers CL, Friedman AD. C/EBPalpha bypasses granulocyte colony-stimulating factor signals to rapidly induce PU.1 gene expression, stimulate granulocytic differentiation, and limit proliferation in 32D cl3 myeloblasts. Blood 1999;94(2):560-71.

28. Cammenga J, Mulloy JC, Berguido FJ, MacGrogan D, Viale A, Nimer SD. Induction of C/EBPalpha activity alters gene expression and differentiation of human CD34+ cells. Blood 2003;101(6):2206-14.

29. Izon DJ, Punt JA, Xu L, Karnell FG, Allman D, Myung PS, et al. Notch1 regulates maturation of CD4+ and CD8+ thymocytes by modulating TCR signal strength. Immunity 2001;14(3):253-64.

30. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

31. Helbling D, Mueller BU, Timchenko NA, Hagemeijer A, Jotterand M, Meyer-Monard S, et al. The leukemic fusion gene AML1-MDS1-EVI1 suppresses CEBPA in acute myeloid leukemia by activation of Calreticulin. Proc Natl Acad Sci U S A 2004;101(36):13312-7.

32. Helbling D, Mueller BU, Timchenko NA, Schardt J, Eyer M, Betts DR, et al. CBFB-SMMHC is correlated with increased calreticulin expression and suppresses the granulocytic differentiation factor CEBPA in AML with inv(16). Blood 2005;106(4):1369-75.

33. Pabst T, Mueller BU, Harakawa N, Schoch C, Haferlach T, Behre G, et al. AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. Nat Med 2001;7(4):444-51.

34. Perrotti D, Cesi V, Trotta R, Guerzoni C, Santilli G, Campbell K, et al. BCR-ABL suppresses C/EBPalpha expression through inhibitory action of hnRNP E2. Nat Genet 2002;30(1):48-58.

35. Westendorf JJ, Yamamoto CM, Lenny N, Downing JR, Selsted ME, Hiebert SW. The t(8;21) fusion product, AML-1-ETO, associates with C/EBP-alpha, inhibits C/EBP-alpha-dependent transcription, and blocks granulocytic differentiation. Mol Cell Biol 1998;18(1):322-33.

36. Heath V, Suh HC, Holman M, Renn K, Gooya JM, Parkin S, et al. C/EBPalpha deficiency results in hyperproliferation of hematopoietic progenitor cells and disrupts macrophage development in vitro and in vivo. Blood 2004;104(6):1639-47.

37. Zhang DE, Zhang P, Wang ND, Hetherington CJ, Darlington GJ, Tenen DG. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc Natl Acad Sci U S A 1997;94(2):569-74.

38. Zhang P, Iwasaki-Arai J, Iwasaki H, Fenyus ML, Dayaram T, Owens BM, et al. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. Immunity 2004;21(6):853-63.

39. Kawagoe H, Grosveld GC. Conditional MN1-TEL knock-in mice develop acute myeloid leukemia in conjunction with overexpression of HOXA9. Blood 2005;106(13):4269-77.

40. Welm AL, Timchenko NA, Darlington GJ. C/EBPalpha regulates generation of C/EBPbeta isoforms through activation of specific proteolytic cleavage. Mol Cell Biol 1999;19(3):1695-704.

41. Baer M, Johnson PF. Generation of truncated C/EBPbeta isoforms by in vitro proteolysis. J Biol Chem 2000;275(34):26582-90.

42. Lin FT, MacDougald OA, Diehl AM, Lane MD. A 30-kDa alternative translation product of the CCAAT/enhancer binding protein alpha message: transcriptional activator lacking antimitotic activity. Proc Natl Acad Sci U S A 1993;90(20):9606-10.

43. Ossipow V, Descombes P, Schibler U. CCAAT/enhancer-binding protein mRNA is translated into multiple proteins with different transcription activation potentials. Proc Natl Acad Sci U S A 1993;90(17):8219-23.

44. Porse BT, Bryder D, Theilgaard-Monch K, Hasemann MS, Anderson K, Damgaard I, et al. Loss of C/EBP alpha cell cycle control increases myeloid progenitor proliferation and transforms the neutrophil granulocyte lineage. J Exp Med 2005;202(1):85-96.

45. Schwieger M, Lohler J, Fischer M, Herwig U, Tenen DG, Stocking C. A dominant-negative mutant of C/EBPalpha, associated with acute myeloid leukemias, inhibits differentiation of myeloid and erythroid progenitors of man but not mouse. Blood 2004;103(7):2744-52.

46. Iwama A, Osawa M, Hirasawa R, Uchiyama N, Kaneko S, Onodera M, et al. Reciprocal roles for CCAAT/enhancer binding protein (C/EBP) and PU.1 transcription factors in Langerhans cell commitment. J Exp Med 2002;195(5):547-58.

47. Wang QF, Friedman AD. CCAAT/enhancer-binding proteins are required for granulopoiesis independent of their induction of the granulocyte colony-stimulating factor receptor. Blood 2002;99(8):2776-85.

48. Weng AP, Nam Y, Wolfe MS, Pear WS, Griffin JD, Blacklow SC, et al. Growth suppression of pre-T acute lymphoblastic leukemia cells by inhibition of notch signaling. Mol Cell Biol 2003;23(2):655-64.

49. Cheng P, Nefedova Y, Miele L, Osborne BA, Gabrilovich D. Notch signaling is necessary but not sufficient for differentiation of dendritic cells. Blood 2003;102(12):3980-8.

50. Ohishi K, Varnum-Finney B, Serda RE, Anasetti C, Bernstein ID. The Notch ligand, Delta-1, inhibits the differentiation of monocytes into macrophages but permits their differentiation into dendritic cells. Blood 2001;98(5):1402-7.

CHAPTER

5

Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced *CEBPA* and mutations in *NOTCH1*

Bas J. Wouters[1], Meritxell Alberich Jordà[2], Karen Keeshan[3], Irene Louwers[1], Claudia A.J. Erpelinck-Verschueren[1], Dennis Tielemans[4], Anton W. Langerak[4], Yiping He[3], Yumi Yashiro-Ohtani[3], Pu Zhang[2], Christopher J. Hetherington[2], Roel G.W. Verhaak[1], Peter J.M. Valk[1], Bob Löwenberg[1], Daniel G. Tenen[2], Warren S. Pear[3], and Ruud Delwel[1]

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

[2] Harvard Institutes of Medicine, Boston, MA, USA

[3] Abramson Family Cancer Research Institute and Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA

[4] Department of Immunology, Erasmus University Medical Center, Rotterdam, The Netherlands

ABSTRACT

Gene expression profiling of acute myeloid leukemia (AML) allows the discovery of previously unrecognized molecular entities. Here, we identified a specific subgroup of AML, defined by an expression profile resembling that of AMLs with mutations in the myeloid transcription factor CCAAT/enhancer binding protein alpha (C/EBPα), while lacking such mutations. We found that in these leukemias, the CEBPA gene was silenced, which was associated with frequent promoter hypermethylation. The leukemias phenotypically showed aberrant expression of T-cell genes, of which CD7 was most consistent. We identified two mechanisms that may contribute to this phenotype. First, absence of Cebpa led to upregulation of specific T-cell transcripts, i.e. Cd7 and Lck, in hematopoietic stem cells isolated from conditional Cebpa knockout mice. Second, the enhanced expression of TRIB2, which we identify here as a direct target of the T-cell commitment factor NOTCH1, suggested aberrantly activated Notch signaling. Putatively activating NOTCH1 mutations were found in several specimens of the newly identified subgroup, while a large set of control AMLs was mutation negative. A gene expression prediction signature allowed the detection of similar cases of leukemia in independent series of AML.

INTRODUCTION

Acute myeloid leukemia (AML) is a heterogeneous disease with respect to its underlying genetic abnormalities and clinical biology (1). It has therefore been postulated that AML, although always characterized by a malignant accumulation of myeloid progenitor cells, in fact represents not a single disease, but a group of neoplasms. Cytogenetic and molecular analyses, studying specific chromosomal translocations and mutations, are used to identify subgroups of AML with distinct biological and clinical behavior. Recent developments in microarray research have resulted in further improvements in the characterization of the molecular heterogeneity of AML (2), and gene expression profiling (GEP) studies have demonstrated that specific chromosomal aberrations, such as the common translocations t(8;21), t(15;17) and inv(16), correlate with unique expression signatures (3-5).

In a GEP study of 285 cases of *de novo* AML, we found that specific expression profiles are associated with other recurrent genetic aberrations such as mutations in *NPM1*, alterations involving *MLL*, and with overexpression of the proto-oncogene *EVI1* as well (4,6). Some AMLs lacking these particular abnormalities were found to express the same characteristic profiles, suggesting that deregulation of the same pathways had occurred through yet unknown mechanisms. This study also showed that two distinct expression clusters (#4 and #15) consisted primarily of AML cases with mutations in *CEBPA*, the gene encoding the basic leucine zipper transcription factor CCAAT/enhancer binding protein alpha (C/EBPα). C/EBPα is a critical regulator of hematopoietic stem cell homeostasis and myeloid differentiation (7-10), and multiple studies have shown that C/EBPα function is frequently abrogated in AML. Besides mutations in the *CEBPA* gene itself, C/EBPα inhibition may result from functional inactivation due to dysregulation by oncogenes, such as *AML1-ETO*, *BCR-ABL* or *FLT3*-ITD (11-14). While both *CEBPA* clusters appeared sharply defined, a subset of AMLs in cluster #4 did not carry *CEBPA* mutations, nor were any other common genetic aberrations found, which was suggestive for C/EBPα interference by other mechanisms.

Recently, we found that Tribbles homolog 2 (*TRIB2)* was highly expressed in several of the cluster #4 leukemias lacking *CEBPA* mutations (15). *Trib2* causes fatal transplantable AML when introduced in murine hematopoietic stem cells *in vivo* (15). Significantly, the leukemogenic potential of Trib2 was associated with its ability to promote degradation of C/EBPα, thus providing further evidence suggesting that disruption of normal C/EBPα function was an important characteristic of this leukemia subgroup. *TRIB2* had been initially identified in an unbiased search for downstream effectors of the *NOTCH1* pathway (15). *NOTCH1* encodes a membrane receptor and transcriptional regulator that plays a critical role in T-cell development (16). Activating *NOTCH1* mutations are observed in approximately 50% of individuals with T-cell acute lymphoblastic leukemia (T-ALL), ranking them among the most frequent abnormalities observed in this type of malignancy (17,18). In contrast, *NOTCH1* mutations are rare in AML and other malignancies (19-21). The selectively high mRNA levels

of the downstream *NOTCH1* target *TRIB2* pointed to a possible role for activated Notch signaling in these AMLs.

Here, we present the results of experiments which indicate that the leukemias in cluster #4 lacking *CEBPA* mutations represent a previously unidentified subset of AML, of which *CEBPA* silencing is a key hallmark. Furthermore, we found in these leukemias a strong association with putatively activating mutations in *NOTCH1*. We established a link between these genetic characteristics and the partially T-lymphoid phenotype of the AMLs, and describe an approach to identify similar leukemia cases in a second cohort of AMLs using gene expression data.

MATERIALS AND METHODS

AML patients, gene expression profiling and data analysis

We made use of the leukemic cell specimens of two independent and representative series of AML patients. The first cohort of *de novo* AML patients has been described previously (4). The second series is comparable to the first cohort with respect to mean age, white blood cell and bone marrow blast counts, French-American-British (FAB) classification (with the exception of FAB-M3, which is underrepresented in this second cohort), distribution of the main cytogenetic abnormalities (with the exception of t(15;17)), and mutations in *FLT3, CEBPA*, *NPM1* and *NRAS* (22). Gene expression profiling of 285 AML specimens from the first cohort using Affymetrix HGU133A GeneChips has been described previously (4). Further details on data analysis are available in Supplementary Materials and Methods.

Nucleotide sequencing of *NOTCH1* and *CEBPA*

RNA isolation and synthesis of complementary DNA (cDNA) were performed as described previously (4). The strategy for determining *CEBPA* mutations in AML has been described (23). Regions of *NOTCH1* cDNA in the heterodimerization domain (HD, exons 26 and 27) and proline-glutamate-serine-threonine-rich domain (PEST, exon 34) found mutated in T-ALL previously (17) were analyzed for the presence of mutations (Supplementary Material and Methods).

Immunophenotyping and T-cell receptor (TCR) gene rearrangement analysis

Immunophenotyping was performed by multicolor flowcytometry using an LSR II flowcytometer (BD Biosciences, San Jose, CA, USA). Details on labeling combinations are given in Supplementary Material and Methods. Data analysis was performed using FACSDiva software, version 4.1.2 (BD Biosciences). PCR for TCR gene rearrangement analysis were performed in an automated thermocycler (model ABI 9600/2700; Applied Biosystems) using primers and conditions as defined in the BIOMED-2 multiplex PCR protocol (24). All

BIOMED-2 multiplex PCR kits were obtained from InVivoScribe Technologies (San Diego, CA, USA; www.invivoscribe.com). After PCR amplification of *TCRG* and *TCRD* gene rearrangements, products were subjected to either heteroduplex analysis (*TCRD*) or GeneScan analysis (*TCRG*) for identification of monoclonal gene rearrangements (24).

Cebpa^{F/F} mice and poly I:C treatment, and *s*taining and sorting of murine bone marrow HSCs

Adult *Cebpa*^F/F *Mx1-Cre* conditional knockout mice and *Cebpa*^wt/wt *Mx1-Cre* control mice were treated with polyinosinic-polycytidylic acid (poly I:C) (Sigma, Steinheim, Germany). 450 μg of poly I:C were administrated by intraperitoneal injection every other day for a total of seven injections. *Cebpa* excision was determined as previously described (7). Hematopoietic stem cells were isolated from murine bone marrow as previously described (25). Briefly, bone marrow cells were stained with rat antibodies specific for lineage markers, i.e. CD11b, CD3, CD4, CD8, Gr1, B220 and CD19 (Caltag Laboratories, Carlsbad, CA, USA). Lineage positive cells were partially removed with sheep anti-rat IgG-conjugated immunomagnetic beads (Dynabeads M-450, Dynal A.S., Oslo, Norway). The remaining bone marrow cells were stained with TC-conjugated lineage markers (Caltag Laboratories), APC-conjugated c-Kit and PE-conjugated Sca-1 (BD Biosciences Pharmingen, San Diego, CA, USA). Lin- c-Kit+ Sca1+ (KSL) cells were sorted in PBS and resorted in RTL buffer (Qiagen, Valencia, CA, USA) for RNA isolation or culture medium for retroviral transduction.

Cebpa expression construct and retroviral transduction

Murine *Cebpa* was cloned into the MSCV-IRES-GFP retroviral vector. Bosc23 cells were co-transfected using lipofectamin with either empty vector or MSCV-*Cebpa*-IRES-GFP, a Gag-Pol construct and an ecotropic Env construct. Virus containing supernatants were collected at 48 and 72 hours after transfection, filtered though 0.45 μm filter and centrifuged at 22.000 r.p.m. for 2 hours at 4°C in a Beckman Coulter XL-90 ultracentrifuge (Beckman Coulter, Fullerton, CA, USA) using an SW28 rotor. Pellets were resuspended in PBS containing 0.1% BSA and stored at -80°C. For retroviral transduction, sorted KSL cells were pre-stimulated in CellGenix SCGM medium (CellGenix, Freiburg, Germany) supplemented with murine Flt3 (100 ng/ml), murine IL3 (20 ng/ml), human IL6 (20 ng/ml), murine Scf (100 ng/ml) and murine Tpo (100 ng/ml) (PeproTech, Rocky Hill, NJ, USA) for 24 hours. Transductions were performed in culture dishes (Falcon 1008, Becton Dickinson) coated with 12 μg/ml retronectin during two consecutive days. Transduced cells were sorted twice for GFP expression.

RNA isolation from murine HSCs and real-time quantitative (RQ)-PCR

RNA was extracted with an RNeasy Micro kit (Qiagen) and reverse transcribed with Super-ScriptII RNase H-Reverse Transcriptase (Invitrogen). Complete absence of *Cebpa* expression was confirmed by RQ-PCR using TaqMan technology on an ABI Prism 7700 Sequence

Detector (Applied Biosystems) with primers, probe and conditions as described previously (26). For *Cd7* and *Lck*, RQ-PCR was performed with SYBR Green (Applied Biosystems) with parameters: 50ºC, 2 min; 95ºC, 10 min followed by 40 cycles of 95ºC, 15 sec and 60ºC for 60 sec. Expression levels were normalized to *Gapdh*. In the PCRs involving *Cd7*, the normalized expression levels for samples from same groups (e.g. F/F cre+, or wt cre+) were then directly combined for subsequent analysis. In the *Lck* PCRs, additional normalization to a calibrator sample was carried out according to the ΔΔCt method (23) to correct for PCR-dependent variability. Student's t-test (two-tailed) was performed using Microsoft Excel 2002 (Microsoft, Redmond, WA, USA) to compare expression levels.

NOTCH1 expression construct and retroviral transduction
The intracellular domain of human *NOTCH1* (ICN1, amino acids 1760-2555) was subcloned into a MSCV-based retroviral vector that co-expresses truncated human nerve growth factor receptor (tNGFR) as a surrogate marker. Scid.adh cells were grown in RPMI 1640 (Invitrogen, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum, 2 mM L-glutamine, 1 mM sodium pyruvate and 1 mM MEM Non-essential Amino acids. For gamma secretase inhibitor (GSI) rescue experiments, 5×10^5 Scid.adh cells were transduced with NGFR-normalized retroviral supernatants obtained following transfection of 293T cells with either NGFR or NGFR-ICN1 and sorted 48 hours later. Cells were subsequently treated with 1 μM GSI (compound E, synthesized in the laboratory of M.S. Wolfe, Harvard Medical School and Brigham and Women's Hospital, Boston, MA, USA (27)) or DMSO vehicle control for 15 hours. RNA, isolated using the RNEasy kit (Qiagen), was digested with DNaseI and used for reverse transcription according to the manufacturers instructions (Superscript IITM kit, Invitrogen). Standard semi-quantitative polymerase chain reaction (PCR) techniques were used, and murine *Hprt* was used as an internal control.

Chromatin immunoprecipitation (ChIP) assay
ChIP was performed using ChIP assay kits (Upstate Biotechnology, Lake Placid, NY, USA). 2.5×10^7 Scid.adh cells were fixed with 1% paraformaldehyde at room temperature for 10 minutes, washed, and lysed with SDS lysis buffer (50mM Tris-HCL, 1% SDS, 10mM EDTA, protease inhibitor cocktail, Sigma, St. Louis, MO, USA). The lysates were sonicated for 30 seconds x 5 at 50% output. The soluble fraction was diluted and precleared with salmon sperm DNA/protein A-agarose. The precleared lysate was split and incubated with either Notch1, acetylated histone 4 or Myc antibodies. The immune complexes were then precipitated with protein A-agarose, washed and eluted with elution buffer. The eluate was reverse cross-linked and treated with proteinase K (20 μg/ml). DNA was purified using a PCR purification kit (Qiagen) and eluted in water (5×10^6 cell equivalents/50 μl). (Semi-)Quantitative PCR was performed (See Supplementary Materials and Methods for primers). For RQ-PCR, the precleared lysate was split and incubated with either Notch1 or control normal rabbit IgG.

DNA sequences were quantified using SYBR green. Not immunoprecipitated input DNA was serially diluted and used as a standard curve to quantify levels of DNA recovered after IP.

RESULTS

CEBPA silencing and promoter hypermethylation are associated with AMLs sharing a *CEBPA* mutant gene expression signature

Unsupervised cluster analysis previously revealed 16 subclasses of AML with distinctive gene expression signatures (4). AMLs with mutations in the *CEBPA* gene were predominantly found in 2 clusters (Figure 1A and Supplementary Table 1) (4). Among cases in cluster #4, a subset of 6 AMLs without *CEBPA* mutations was apparent (Figure 1A). We applied Significance Analysis of Microarrays (SAM) (28) to determine gene array probe sets distinguishing these 6 cases from the specimens with *CEBPA* mutations in the same cluster. One of the strongest discriminating genes was *CEBPA* itself (Supplementary Table 2). Whereas *CEBPA* levels were very high in *CEBPA* mutant cases, expression was minimal or non-existent in the non-mutant specimens (Figure 1A and Supplementary Table 3A). *CEBPA* expression levels were confirmed by real-time quantitative polymerase chain reaction (RQ-PCR) (data not shown). Bisulfite genomic sequencing revealed that in 4 of the 6 patient samples, a proximal region in the *CEBPA* promoter was densely CpG hypermethylated (Figure 1B). In contrast, no methylation of the same region was observed in 38 control AMLs from variable clusters (Supplementary Table 3B).

AML cases with silenced *CEBPA* in GEP cluster #4 express myeloid as well as T-lymphoid genes

Further comparison of *CEBPA* silenced leukemias to *CEBPA* mutant cases in cluster #4 revealed that a considerable number of significantly overexpressed genes in the first group was related to T-cells and T-lymphoid development (Supplementary Tables 2, 4 and 5 and Supplementary Figure 1). In Figure 2A, the expression of a selection of those significantly elevated T-lymphoid genes is depicted as a snapshot of the GEP correlation plot of 285 AML cases (4), and illustrates expression of *CD7*, *CD3D*, *CD3Z*, *CD3G*, the T-cell specific Src-family kinase *LCK*, and *TRD@*, encoding the T-cell receptor (TCR) delta locus, and *NOTCH1*. Multicolor flowcytometric analysis demonstrated that blast cells of the 6 cases co-expressed myeloid and T-lymphoid lineage markers (Table 1 and Figure 2B). CD7 was consistently highly expressed in all 6 specimens, while in 4 cases the expression of one or more additional T-cell marker(s) was seen. CD3 protein surface expression was only observed weakly in one specimen, which contrasted with high mRNA expression in multiple cases. Rearrangements in T-cell receptor genes, i.e. *TCRD* and/or *TCRG*, were detected in 4/6 cases (Table 1). Importantly, all 6 specimens with silenced *CEBPA* showed expression of several myeloid lineage markers, including
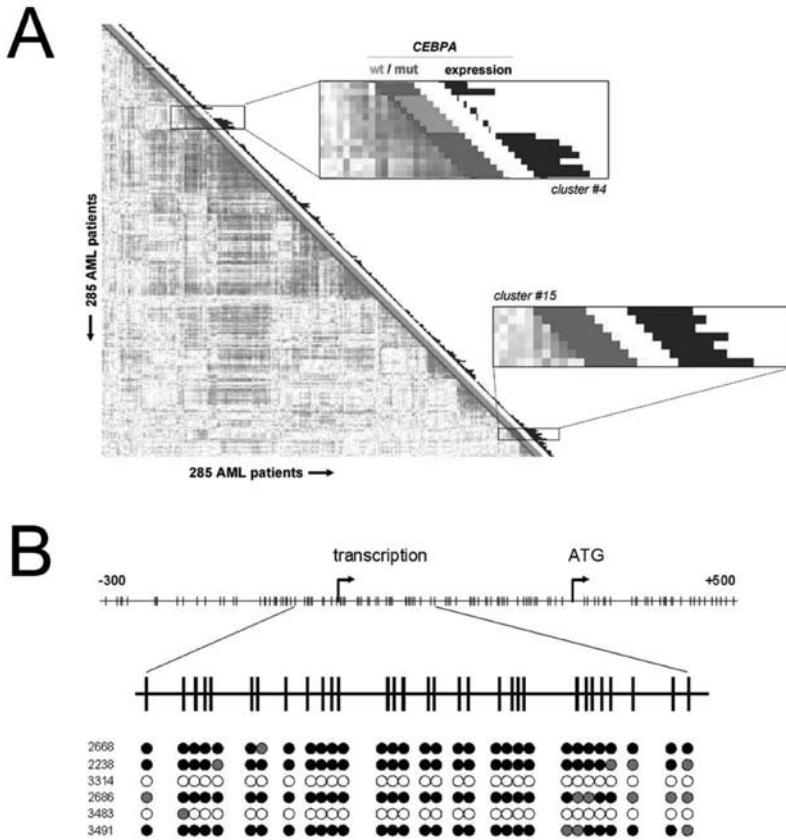
Figure 1. *CEBPA* **silencing and promoter hypermethylation are associated with AMLs sharing a** *CEBPA* **mutant gene expression signature in GEP cluster #4.**

A. Pair wise correlations between gene expression profiles of 285 AML samples calculated on the basis of 2856 probe sets are displayed as described (4). Colors of boxes visualize Pearson correlation coefficient: deeper red indicates higher positive correlation, deeper blue indicates higher negative correlation. Sixteen distinct clusters were previously distinguished, which can be recognized by the red blocks showing high correlation along the diagonal (4). Cluster #4 and cluster #15, associated with *CEBPA* mutations, are enlarged. The bar and histogram next to each patient represent *CEBPA* mutation status and *CEBPA* expression level, respectively. *CEBPA* mutation status: presence ("mut", red) or absence ("wt", green) of mutations in bZIP region and/or N-terminus. These data indicate that in leukemias lacking mutations, *CEBPA* expression is low or absent. In cluster #4, the order of samples, from top to bottom, is: #3327, #2242 (both *CEBPA* mutant), #2668, #2238, #3314, #2686, #3483, #3491 (all 6 without *CEBPA* mutation), #2218, #1316, #2273, #2545, #2169, #2753, and #2192 (all 7 *CEBPA* mutant).

B. Upper part: schematic representation of the chromosomal region surrounding the transcriptional start of the *CEBPA* gene. Numbers indicate position relative to *CEBPA* transcriptional start. Vertical lines represent CpG dinucleotides, "transcription" stands for transcriptional start, "ATG" stands for translational start site. Lower part: level of cytosine methylation in the region surrounding the *CEBPA* transcriptional start site of the 6 AML cases in cluster #4 with low *CEBPA* expression, with patient numbers on the left. Every cytosine in a CpG dinucleotide is depicted as a circle. For each of these cytosines, the fraction of methylated residues was determined, which is visualized by the color of the circle: methylated (>75% of all cytosines methylated, black), partly methylated (25-75% methylated, grey), unmethylated (<25% methylated, white).

(A full color version of this figure can be found in the color section.)

Figure 2. AML cases with silenced *CEBPA* in cluster #4 simultaneously express myeloid and T-lymphoid specific genes and lineage markers.

A. Pair wise correlations between samples are displayed as explained in the legend to Figure 1, and GEP cluster #4 is enlarged in the box on the right. Histograms next to each patient display expression of selected genes with significantly elevated expression in cluster #4 cases with silenced *CEBPA*. Expression levels for probe sets of the following genes are visualized: *CD7, CD3D, CD3Z, CD3G, LCK, TRD@* and *NOTCH1*. Corresponding expression levels and Affymetrix probe set identifiers are depicted in Supplementary Table 4.

B. Representative dot plot images from flowcytometric analysis of samples obtained from 2 individual patients, i.e. patients #2238 and #3491, demonstrating that the majority of cells from these patients simultaneously express CD34 and CD13, CD7 and terminal deoxynucleotidyltransferase (TdT). The tumor population was identified by weak expression of CD45, depicted in black, whereas CD45[high] cells, which are considered to be mature lymphocytes, are colored in grey.

(A full color version of this figure can be found in the color section.)

CD13 and CD33. The immature stage of the myeloid/T-lymphoid leukemias was confirmed by membrane presence of CD34 protein.

Excision of *CEBPA* in hematopoietic stem cells induces expression of CD7 and LCK

To investigate whether lack of the myeloid master regulator *Cebpa* could be sufficient to induce T-lymphocyte associated transcripts, we made use of a conditional *Cebpa* knockout mouse model.(7) *Cebpa*F/F (*Cebpa*Flox/Flox) Mx1-Cre conditional knockout mice and *Cebpa*wt/wt (*Cebpa*wild type/wild type) Mx1-Cre control mice were treated with polyinosinic-polycytidylic acid (poly I:C), and Lin- c-Kit+ Sca1+ hematopoietic stem cells (HSCs) were

Figure 3. Absence of *Cebpa* in murine HSCs upregulates *Cd7* and *Lck*.
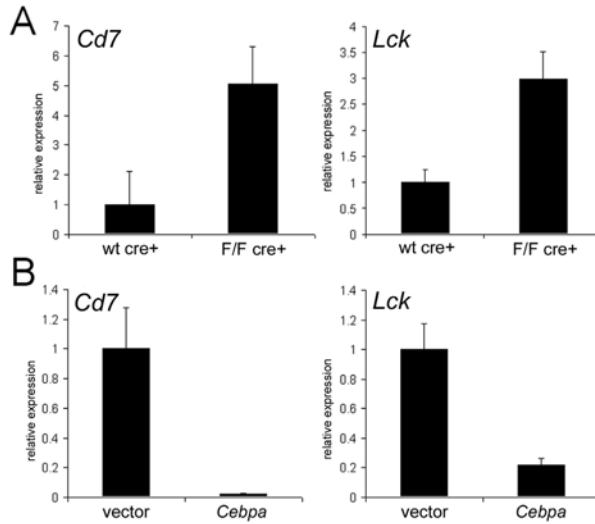A. Lin- c-Kit+ Sca1+ hematopoietic stem cells (HSCs) were isolated and purified from poly I:C treated *Cebpa*[wt/wt] *Mx1-Cre* mice (wt cre+) and *Cebpa*[F/F] *Mx1-Cre* knockout mice (F/F cre+). mRNA expression levels of *Cd7* and *Lck* were determined by RQ-PCR. Data are presented relative to expression of wt cre+ animals as mean plus standard deviation of least 3 individual mice in both groups.
B. Purified HSCs from poly I:C treated *Cebpa*[F/F] *Mx1-Cre* mice were transduced with either empty vector (vector) or *Cebpa* expression construct (*Cebpa*). *Cd7* and *Lck* mRNA expression levels were determined in EGFP+ infected cells 1 day after transduction. Data are presented relative to expression of empty vector transduced cells as mean plus standard deviation of 4 reactions.

isolated from bone marrow 2 weeks after the last injection. We investigated expression levels of 4 particular T-cell associated genes by RQ-PCR. Excision of *Cebpa* resulted in upregulation of *Cd7* and *Lck* compared to control animals ($p<0.01$ and $p<0.05$, respectively) (Figure 3A). Moreover, re-introduction of *Cebpa* in HSCs isolated from poly I:C-treated *Cebpa*F/F Mx1-Cre animals strongly reduced mRNA levels of both genes (Figure 3B), further demonstrating that *Cebpa* plays a role in repression of these T-lymphocyte associated transcripts. However, transcript levels of *Cd3d* and *Cd3g* did not change in *Cebpa* knockout HSCs (data not shown), suggesting that additional T-cell regulatory pathways had been dysregulated in AML cases with silenced *CEBPA* in GEP cluster #4.

TRIB2 overexpression points to aberrant Notch activation in AMLs with silenced CEBPA, and associates with NOTCH1 mutations

The leukemias with silenced *CEBPA* in cluster #4 frequently showed high expression of the myeloid oncogene *TRIB2* and *NOTCH1* (Figure 2A), suggesting potential activation of Notch signaling. Treatment of Notch1-dependent murine cell lines T6E (29) and Scid.adh (15,30) with gamma secretase inhibitors (GSI) led to *Trib2* downregulation (Figure 4A and not shown). Moreover, *Trib2* expression was restored when these cells were transduced with a retroviral vector expressing intracellular NOTCH1 (ICN1), which is GSI resistant (Figure 4A).
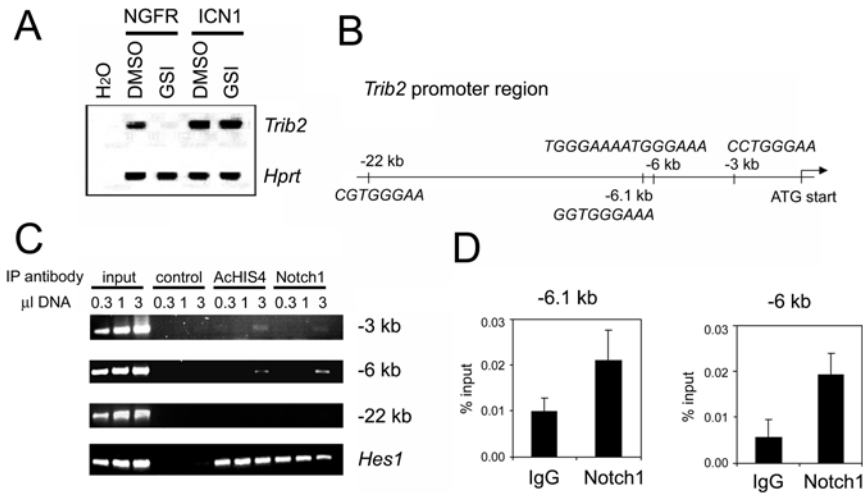
Figure 4. *Trib2* is a direct target gene of Notch1.
A. Reverse transcriptase (RT)-PCR analysis of *Trib2* expression in Scid.adh cells treated with GSI (+) or DMSO as vehicle control (-). Cells were transduced with an empty vector only expressing the NGFR selection marker (NGFR) or a vector expressing ICN1 (ICN1). *Trib2* mRNA expression was assessed. *Hprt* is an internal loading control. Results are representative of triplicate experiments.
B. Schematic representation of the *Trib2* promoter region. Four putative CSL binding sites were identified in the region shown, including a tandem CSL site at -6kb and 3 single binding sites at -22kb, -6.1kb and -3kb relative to the translational start site.
C. Notch1 binds to CSL binding sites in the *Trib2* promoter. Chromatin immunoprecipitates were performed on cross-linked fragmented DNAs prepared from Scid.adh cells. Immunoprecipitations were carried out with antibodies against Myc as a negative control (control), acetylated Histone 4 (AcHIS4) and Notch1. PCR was performed using primers directed against indicated CSL binding site regions at -22, -6 and -3kb from the ATG translational start site of *Trib2*. PCR for the *Hes1* promoter region is shown as a positive control. The figure shown is representative of duplicate experiments. For the -6 kb region, which shows the strongest enrichment, results are representative of duplicate samples and triplicate experiments.
D. After ChIP as described in panel C, using either Notch1 or normal rabbit control IgG antibody, RQ-PCR was performed using primer sets flanking putative CSL binding sites in the *Trib2* promoter. RQ-PCR was performed using the primers at -6 kb to quantify enrichment in this region, and also using primers flanking a conserved CSL binding site at -6.1 kb. Graphs represent mean of the ratios of the amount of IP DNA/input from values of duplicate wells +/- standard deviation. Data are representative of 3 independent experiments.

Likewise, introduction of ICN1 into the myeloid leukemia cell line U937 led to upregulation of *TRIB2* levels (Supplementary Figure 2). To gain further insight, we investigated whether *TRIB2* could be directly regulated by *NOTCH1*. ICN1 binds the transcriptional repressor protein CSL, converting it into an activator. Several putative CSL binding sites were identified, at -22 kb, -6.1kb, -6 kb and -3 kb from the *Trib2* translational start site, respectively (Figure 4B). Chromatin immunoprecipitation (ChIP) showed binding of the ICN1/CSL complex to both the –6.1kb/-6 kb region, containing a single conserved as well as a tandem CSL binding site, and to a weaker extent to the single -3 kb site (Figure 4C and 4D). Both regions were also immunoprecipitated by antibodies against acetylated Histone 4 (AcHIS4), consistent with transcriptional activity. Together, these results indicate that *Trib2* is a direct transcriptional target of Notch1, and further suggest that AML cases with selectively increased levels of *TRIB2*
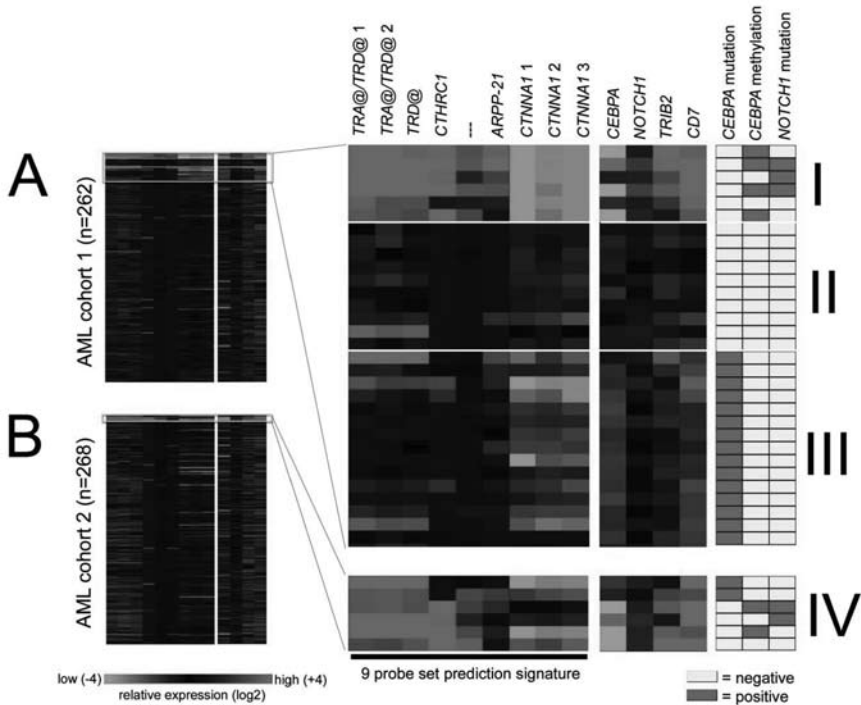
Figure 5. A gene expression prediction signature identifies new leukemias with the silenced *CEBPA* phenotype in an independent cohort of AML.

A. For 262 samples analyzed on Affymetrix HGU133Plus2.0 GeneChips, log transformed (base 2) and mean centered expression levels for 13 probe sets are depicted (left panel) for an arbitrary range from -4 to +4 (corresponding to 16-fold lower to 16-fold higher expression relative to the mean, respectively). The ordering of patients in the figure is arbitrary. In the right panel, these data are enlarged for 31 of these 262 leukemias, representing 3 groups: (I) the 6 cases with silenced *CEBPA* previously identified, with from top to bottom cases #2668, #2238, #3314, #2686, #3483, and #3491; (II) a variable selection of 10 AMLs from distinct GEP clusters for which also *NOTCH1* mutational analysis and *CEBPA* promoter bisulfite sequencing were performed, and (III) 15 AMLs with *CEBPA* mutations, originating from either GEP cluster #4 (upper 9 samples) or cluster #15 (lower 6 samples). The 9 probe sets on the left side constitute the most predictive gene expression signature for group I, as determined by PAM: 216191_s_at (*TRA@/ TRD@* 1), 217143_s_at (*TRA@/TRD@* 2), 213830_at (*TRD@*), 225681_at (*CTHRC1*), 1565809_x_at (no annotation), 1560018_at (*ARPP-21*), 210844_x_at (*CTNNA1* 1), 200764_s_at (*CTNNA1* 2), and 200765_x_at (*CTNNA1* 3). To the right, 4 additional probe sets are indicated, i.e. 204039_at (*CEBPA*), 218902_at (*NOTCH1*), 202478_at (*TRIB2*) and 214551_s_at (*CD7*). Mutational data for *NOTCH1* and *CEBPA*, and methylation status of the *CEBPA* promoter are depicted next to the normalized hybridization intensities of the probe sets.

B. 268 samples obtained from a second cohort of AML were hybridized to HGU133Plus2.0 GeneChips. The 9 probe set signature was used to identify leukemias with a profile similar to group I (A), resulting in the detection of group IV (from top to bottom cases #6376, #6735, #6947, #7053, #7076 and #7120).

(A full color version of this figure can be found in the color section.)

may be subject to aberrant induction of Notch signaling. Activating mutations in *NOTCH1*, as found in T-ALL, target two regions of the gene: the heterodimerization (HD) and the proline-glutamate-serine-threonine-rich (PEST) domain. Nucleotide sequencing showed 4 similar, putatively activating mutations in cDNA and genomic DNA in 3 of the 6 AMLs with high *TRIB2* expression (Table 2). In one case, an HD and a PEST mutation were found *in cis*.

Table 1. Immunophenotypes and T-cell receptor rearrangements of AMLs with silenced *CEBPA* in GEP cluster #4.

	2668	2238	3314	2686	3483	3491
T-lymphoid antigens						
CD7	+	+	+	+	+	+
SmCD3	-	-	-	+/-	-	-
CyCD3	+/-	+	-	+/-	-	-
CD2	+/-	-	-	+/-	-	-
TdT	-	+	-	-	-	+
Myeloid antigens						
CD13	+	+ partly	+/-	+/-	+	+
CD33	+/-	+/-	+	+/-	+/-	+
CD14	+/-	-	-	+/-	-	-
CD64	-	+	-	-	-	+
CD11b	+	-	+	+	-	-
MPO	-	-	+	+	+	+
Stem cell antigens						
CD34	+	+	+	+/-	+	+
CD117	+	-	+/-	+/-	+	+
CD133	-	-	-	-	+/-	+/-
Other antigens						
CD45	+/-	+/-	+/-	+/-	+/-	+/-
CD56	+/- partly	ND	+	+/-	+	ND
DNAs						
TCRD rearrangement	+	+	+	-	+	-
TCRG rearrangement	+	+	-	-	+	-

+ indicates positive; +/-, weakly positive; -, negative; partly, part of population positive; and ND, not determined.
The following lineage markers were also analyzed, and found negative on all 6 samples: TCRα/β, TCRγ/δ, CD4, CD8, CD10, CD1a, CD65, and CD15. CD5, CD19 and CD22 were tested in 4 samples (all except for #2238 and #3491), and were negative in all these 4 cases.
* As a control, 6 GEP cluster #4 cases with *CEBPA* mutations were analyzed for T-cell receptor rearrangements as well. Patients #1316, #2242, #2273, #2545 and #3327 were negative for *TCRD* and *TCRG* rearrangements. Patient #2218 was positive for *TCRG* rearrangement and negative for *TCRD* rearrangement.

In contrast, we did not identify *NOTCH1* mutations in 100 control AML samples, including several AMLs from each GEP cluster, and the *CEBPA* mutant AML cases from clusters #4 and #15 (Supplementary Table 3B and 3C).

Gene expression prediction signature allows identification of similar leukemias in an independent cohort of AMLs

To validate our findings, we determined a gene expression prediction signature to identify additional leukemias with the same characteristics. We made use of Affymetrix HGU133 Plus2.0 GeneChips, and based prediction analysis on data obtained from 262/285 specimens, including all cases in cluster #4. A signature of 9 probe sets was defined by the Prediction Analysis of Microarrays (PAM) class prediction algorithm (31). This signature accurately detected all 6 *CEBPA* silenced cases in cross-validation experiments, while the average number of false positives was minimal using the same detection criteria (2/256, 0.8%) (Figure 5A). In a second cohort of 268 AML cases, this predictor detected 6 new cases sharing the same profile (Figure 5B). Four of these 6 new leukemias showed no or low amounts of *CEBPA* as well as elevated levels of *TRIB2* and *NOTCH1* mRNA, consistent with our prior findings - in spite of the fact that the prediction signature itself did not contain probe sets for these genes (Figure 5). Analysis of available diagnostic flowcytometric data for these 4 leukemias revealed myeloid lineage marker expression in combination with CD7 in all, while 2/4 leukemias expressed additional T-lymphoid markers (Supplementary Table 6). *NOTCH1* mutations were found in 2 of these leukemias, and hypermethylation of the *CEBPA* promoter was apparent in 2 cases (Table 2; Figure 5B). In contrast, elevated levels of *CEBPA* mRNA were observed in the remaining 2/6 leukemias predicted in the second series of AML (Figure 5B). Both cases showed mutations in the *CEBPA* coding region, explaining the similar expression profile.

DISCUSSION

In the present study, we define a specific subgroup of AML with a discriminative gene expression profile and expression of T-cell genes. These specific AMLs associate with silencing of *CEBPA* and aberrant activation of *NOTCH1* by mutations. Our data provide evidence for the involvement of C/EBPα in the active repression of several of the T-cell associated genes, supporting a cooperating role for *CEBPA* silencing in development of the leukemic phenotype. Using gene expression data we determined a predictor of 9 probe sets by which we have been able to identify similar leukemia cases in a second cohort of AMLs.

We initially studied these specific AMLs because their gene expression profiles were similar to leukemias with mutations in the transcription factor *CEBPA*. This similarity suggested another mechanism of *CEBPA* deregulation. Indeed, all 6 cases revealed complete or near-complete absence of *CEBPA* mRNA. Importantly, *CEBPA* mutant cases in this cluster either carried homozygous mutations or mutations in both *CEBPA* N-terminus and basic leucine zipper region, which are typically biallelic (Supplementary Table 1) (10,23). Therefore, none of these mutant cases was predicted to express wild type *CEBPA*, and lack of wild type *CEBPA* appeared to be the common hallmark of all specimens in GEP cluster #4.

Downregulation or functional inactivation of *CEBPA* has been implicated in development of myeloid malignances through several mechanisms previously (10-12). Aberrant CpG methylation of the *CEBPA* promoter is involved in carcinogenesis of solid tissue such as lung (32), and has recently been described sporadically in AML as well (33,34). Because of lack of other recurrent molecular lesions, we addressed this particular mechanism of gene silencing. Indeed, extensive *CEBPA* promoter hypermethylation associated with 4/6 of these cases and with 2/4 cases of the second cohort. Similar experiments in a set of control AMLs by us and experiments by other groups suggest that *CEBPA* promoter hypermethylation is generally not very common in AML (33,34), and no other clusters appeared to be characterized by unified *CEBPA* silencing (Figure 1A). Although we can not rule out that *CEBPA* promoter hypermethylation may occur in other sporadic cases, these observations suggest that this mechanism is significantly associated with this specific subset. Interestingly, in 4 of the 10 cases identified in the 2 cohorts, silencing was not associated with hypermethylation, suggesting a possible other, yet unknown, mechanism of *CEBPA* mRNA repression in those.

The leukemias with silenced *CEBPA* simultaneously expressed myeloid and T-lymphoid genes. Mixed myeloid/T-lymphoid leukemias have been described previously and are thought to represent a variety of diseases with a large genetic heterogeneity, ranging from myeloid leukemias with aberrant expression of only one or few T-cell markers, to "true" biphenotypic leukemias (35-38). The T-cell gene most consistently associated with the patient group described here was *CD7*, which is known to be expressed by a considerable proportion of immature AMLs (39-41), suggesting that the precursor cell in which transformation occurred already had undergone some initial myeloid differentiation. Interestingly, *CD7* expression is frequently associated with AMLs carrying *CEBPA* mutations (41,42). Expression of additional T-cell lineage markers was apparent in 4/6 cases of the subgroup with silenced *CEBPA*. A previously proposed scoring system based on immunophenotype would classify the majority of the cases identified by us as myeloid malignancies with aberrant T-cell expression, rather than as "true" biphenotypic leukemias (43). In line with previous observations, most leukemias in the subgroup described here showed several cytogenetic abnormalities, however, none of these appeared common to all (Supplementary Tables 1 and 6). Our observations classify this specific heterogeneous population as a homogeneous subset based on their gene expression profiles.

We found that in a *Cebpa* conditional knockout model, transcript levels in HSCs of *Lck* and *Cd7* appeared to be dependent on the absence of C/EBPα, suggesting that C/EBPα is directly involved in negative regulation of these genes. These observations mirror those from previous studies that showed that C/EBPα has lineage switch potential when expressed in lymphoid cells (44-48). Although the role of C/EBPα in lymphoid cell fate decision may not be fully understood yet, as highlighted by recent observations indicating that overexpression of C/EBPα due to chromosomal translocations may be involved in B-cell ALLs (49,50), our

Table 2. *NOTCH1* mutations identified in patients from two independent AML cohorts

Patient	AML cohort	NOTCH1 mutation*	Predicted amino acid change
2238	First	4741-4743del	In-frame deletion of M1581
3314	First	5036T>A,	L1679Q,
		7376ins37bp	L2458fsX2463
2686	First	7552ins13bp	P2518fsX2523
6947	Second	4790T>C	L1597P
7053	Second	4820del22/ins13	Mutation at aa 1607 FKRDAHGQ into SGRRP

Nucleotide numbering according to Entrez nucleotide accession number NM_017617.2 The mutations in patients #2238, #6947 and #7053 are located in the *NOTCH1* HD domain (exons 26 and 27). The mutation in patient #2686 is located in the *NOTCH1* PEST domain (exon 34). Patient #3314 harbors *in cis* mutations both in the HD-domain as well as the PEST-domain. All mutations are heterozygous.
The first AML cohort represents the initial series of 285 cases. The second cohort represents the independent series of 268 AMLs, which was interrogated using a gene expression prediction signature.

studies show that decreased expression of C/EBPα in hematopoietic stem cells can be sufficient to induce T-cell transcripts.

In addition to loss of C/EBPα function, the experiments performed in murine HSCs also suggested dysregulation of additional pathways leading to aberrant T-cell commitment in the clinical leukemias, as *Cebpa* excision failed to explain the expression of all T-cell transcripts. As an additional explanation for the phenotype observed, we found evidence for aberrant Notch activation in several of these leukemias. In the present study, we identified *TRIB2* as a direct target of *NOTCH1*, which led us to investigate whether activating *NOTCH1* mutations had occurred in the AMLs. Several cases exhibited mutations in *NOTCH1* PEST and/or HD domains, which were all predicted to result in enhanced NOTCH1 signaling (Table 2) (51,52). In contrast, we did not detect similar mutations in a large cohort of AML control samples, concordant with recent literature (19,20). Our data therefore suggest that in AML, *NOTCH1* mutations are largely restricted to this particular subset of myeloid/T-lymphoid leukemias. At present, it is not known what the precise mechanism of transformation in these leukemias is, nor why *NOTCH1* mutations specifically associate with this type of AML. Expression of activated *Notch1* in murine bone marrow cells *in vivo* leads exclusively to T-cell neoplasms (29), whereas the tumors described in the current study had clear myeloid features, possibly suggesting that the leukemia initiating cell is a committed myeloid progenitor. Interestingly, overexpression of *Trib2* in murine HSCs induced AML (15), suggesting an alternative scenario in which *TRIB2* plays an important role in leukemic transformation. Trib2 inactivates C/EBPα p42, implying that high TRIB2 expression may be an alternative mechanism to interfere with C/EBPα function in AMLs lacking hypermethylation of the *CEBPA* promoter. TRIB proteins may also enhance ERK phosphorylation, which might be suggestive of an additional mechanism involved in transformation (15,53). Several leukemias of this subgroup did not reveal *NOTCH1* HD and PEST mutations. It is possible that in

these cases aberrant Notch activation occurred by other mechanisms, explaining the elevated expression of *TRIB2* in those tumors. This could involve other Notch receptors, ligands, or other mutations in *NOTCH1*. Cytogenetic analysis did not reveal translocations involving the *NOTCH1* locus on chromosome 9q34.3 (Supplementary Table 1) (18), neither did we observe overexpression of NOTCH1 ligand coding genes (data not shown).

We evaluated our observations in an independent representative series of AML. We successfully defined a gene expression predictor signature for the leukemias with silenced *CEBPA*, and identified 4 new cases with highly similar molecular and phenotypic characteristics. The 6 classifying genes, represented by 9 probe sets, have previously been associated with immature thymocytes (T-cell receptor genes *TRA@* and *TRD@* and the cyclic AMP-regulated phosphoprotein *ARPP-21* (54,55)), or with malignant transformation (alpha-E-catenin (*CTNNA1*) (56-58) and collagen triple helix repeat containing 1 (*CTHRC1*) (59)), while another probe set (1565809_x_at) maps to an uncharacterized gene. We propose that a simple algorithm to diagnose this subclass of leukemias could make use of a combination of this prediction signature with subsequent *CEBPA* mutational analysis. Our data do not rule out the possibility that some leukemias with silenced *CEBPA* and/or *NOTCH1* mutations were not detected by the predictor. Although the low prevalence of *NOTCH1* mutations and *CEBPA* promoter hypermethylation in AML argues against many overlooked cases, additional studies are warranted to test these findings in a more diagnostic setting. Additional studies will also need to address whether our preliminary observations concerning clinical outcome, which suggest a poor response to treatment (Supplementary Table 7), will hold in larger series.

In conclusion, our data delineate a specific subset of AMLs within the population of myeloid malignancies with aberrant expression of T-cell genes. This subtype of leukemia is primarily defined by its gene expression profile, which is likely to be determined by silencing of the *CEBPA* gene, and is associated with recurring mutations in *NOTCH1*. Importantly, we have demonstrated that these leukemias can be predicted in independent series of AML using expression data from a limited set of genes. Together, these observations emphasize the applicability and power of in depth exploration of GEP data for characterization of previously unrecognized subgroups of AML.

ACKNOWLEDGEMENTS

REFERENCES

1. Lowenberg B, Downing JR, Burnett A. Acute myeloid leukemia. N Engl J Med 1999;341(14):1051-62.

2. Valk PJ, Delwel R, Lowenberg B. Gene expression profiling in acute myeloid leukemia. Curr Opin Hematol 2005;12(1):76-81.

3. Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004;350(16):1605-16.

4. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

5. Ross ME, Mahfouz R, Onciu M, Liu HC, Zhou X, Song G, et al. Gene expression profiling of pediatric acute myelogenous leukemia. Blood 2004;104(12):3679-87.

6. Verhaak RG, Goudswaard CS, van Putten W, Bijl MA, Sanders MA, Hugens W, 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 2005;106(12):3747-54.

7. Zhang P, Iwasaki-Arai J, Iwasaki H, Fenyus ML, Dayaram T, Owens BM, et al. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. Immunity 2004;21(6):853-63.

8. Zhang DE, Zhang P, Wang ND, Hetherington CJ, Darlington GJ, Tenen DG. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc Natl Acad Sci U S A 1997;94(2):569-74.

9. Radomska HS, Huettner CS, Zhang P, Cheng T, Scadden DT, Tenen DG. CCAAT/enhancer binding protein alpha is a regulatory switch sufficient for induction of granulocytic development from bipotential myeloid progenitors. Mol Cell Biol 1998;18(7):4301-14.

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

11. Pabst T, Mueller BU, Harakawa N, Schoch C, Haferlach T, Behre G, et al. AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. Nat Med 2001;7(4):444-51.

12. Perrotti D, Cesi V, Trotta R, Guerzoni C, Santilli G, Campbell K, et al. BCR-ABL suppresses C/EBPalpha expression through inhibitory action of hnRNP E2. Nat Genet 2002;30(1):48-58.

13. Rosenbauer F, Tenen DG. Transcription factors in myeloid development: balancing differentiation with transformation. Nat Rev Immunol 2007;7(2):105-17.

14. Zheng R, Friedman AD, Levis M, Li L, Weir EG, Small D. Internal tandem duplication mutation of FLT3 blocks myeloid differentiation through suppression of C/EBPalpha expression. Blood 2004;103(5):1883-90.

15. Keeshan K, He Y, Wouters BJ, Shestova O, Xu L, Sai H, et al. Tribbles homolog 2 inactivates C/EBPalpha and causes acute myelogenous leukemia. Cancer Cell 2006;10(5):401-11.

16. Radtke F, Wilson A, Mancini SJ, MacDonald HR. Notch regulation of lymphocyte development and function. Nat Immunol 2004;5(3):247-53.

17. Weng AP, Ferrando AA, Lee W, Morris JPt, Silverman LB, Sanchez-Irizarry C, et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. Science 2004;306(5694):269-71.

18. Grabher C, von Boehmer H, Look AT. Notch 1 activation in the molecular pathogenesis of T-cell acute lymphoblastic leukaemia. Nat Rev Cancer 2006;6(5):347-59.

19. Fu L, Kogoshi H, Nara N, Tohda S. NOTCH1 mutations are rare in acute myeloid leukemia. Leuk Lymphoma 2006;47(11):2400-3.

20. Palomero T, McKenna K, J ON, Galinsky I, Stone R, Suzukawa K, et al. Activating mutations in NOTCH1 in acute myeloid leukemia and lineage switch leukemias. Leukemia 2006.

21. Leong KG, Karsan A. Recent insights into the role of Notch signaling in tumorigenesis. Blood 2006;107(6):2223-33.

22. Verhaak RG, Wouters BJ, Erpelinck CA, Abbas S, Beverloo HB, Lugthart S, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. Haematologica 2009;94(1):131-4.

23. Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, Meijer J, van Oosterhoud S, van Putten WL, Valk PJ, et al. Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML. Hematol J 2003;4(1):31-40.

24. van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 2003;17(12):2257-317.

25. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature 2000;404(6774):193-7.

26. Hirai H, Zhang P, Dayaram T, Hetherington CJ, Mizuno S, Imanishi J, et al. C/EBPbeta is required for 'emergency' granulopoiesis. Nat Immunol 2006;7(7):732-9.

27. Weng AP, Millholland JM, Yashiro-Ohtani Y, Arcangeli ML, Lau A, Wai C, et al. c-Myc is an important direct target of Notch1 in T-cell acute lymphoblastic leukemia/lymphoma. Genes Dev 2006;20(15):2096-109.

28. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98(9):5116-21.

29. Pear WS, Aster JC, Scott ML, Hasserjian RP, Soffer B, Sklar J, et al. Exclusive development of T cell neoplasms in mice transplanted with bone marrow expressing activated Notch alleles. J Exp Med 1996;183(5):2283-91.

30. Carleton M, Ruetsch NR, Berger MA, Rhodes M, Kaptik S, Wiest DL. Signals transduced by CD3epsilon, but not by surface pre-TCR complexes, are able to induce maturation of an early thymic lymphoma in vitro. J Immunol 1999;163(5):2576-85.

31. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002;99(10):6567-72.

32. Tada Y, Brena RM, Hackanson B, Morrison C, Otterson GA, Plass C. Epigenetic modulation of tumor suppressor CCAAT/enhancer binding protein alpha activity in lung cancer. J Natl Cancer Inst 2006;98(6):396-406.

33. Chim CS, Wong AS, Kwong YL. Infrequent hypermethylation of CEBPA promotor in acute myeloid leukaemia. Br J Haematol 2002;119(4):988-90.

34. Agrawal S, Hofmann WK, Tidow N, Ehrich M, van den Boom D, Koschmieder S, et al. The C/EBP{delta} tumor suppressor is silenced by hypermethylation in acute myeloid leukemia. Blood 2007.

35. Owaidah TM, Al Beihany A, Iqbal MA, Elkum N, Roberts GT. Cytogenetics, molecular and ultra-structural characteristics of biphenotypic acute leukemia identified by the EGIL scoring system. Leukemia 2006;20(4):620-6.

36. Rubio MT, Dhedin N, Boucheix C, Bourhis JH, Reman O, Boiron JM, et al. Adult T-biphenotypic acute leukaemia: clinical and biological features and outcome. Br J Haematol 2003;123(5):842-9.

37. Killick S, Matutes E, Powles RL, Hamblin M, Swansbury J, Treleaven JG, et al. Outcome of biphe-notypic acute leukemia. Haematologica 1999;84(8):699-706.

38. Cross AH, Goorha RM, Nuss R, Behm FG, Murphy SB, Kalwinsky DK, et al. Acute myeloid leukemia with T-lymphoid features: a distinct biologic and clinical entity. Blood 1988;72(2):579-87.

39. Drexler HG, Thiel E, Ludwig WD. Acute myeloid leukemias expressing lymphoid-associated antigens: diagnostic incidence and prognostic significance. Leukemia 1993;7(4):489-98.

40. Khalidi HS, Medeiros LJ, Chang KL, Brynes RK, Slovak ML, Arber DA. The immunophenotype of adult acute myeloid leukemia: high frequency of lymphoid antigen expression and comparison of immunophenotype, French-American-British classification, and karyotypic abnormalities. Am J Clin Pathol 1998;109(2):211-20.

41. Bienz M, Ludwig M, Leibundgut EO, Mueller BU, Ratschiller D, Solenthaler M, et al. Risk as-sessment in patients with acute myeloid leukemia and a normal karyotype. Clin Cancer Res 2005;11(4):1416-24.

42. Lin LI, Chen CY, Lin DT, Tsay W, Tang JL, Yeh YC, et al. Characterization of CEBPA mutations in acute myeloid leukemia: most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells. Clin Cancer Res 2005;11(4):1372-9.

43. Bene MC, Castoldi G, Knapp W, Ludwig WD, Matutes E, Orfao A, et al. Proposals for the im-munological classification of acute leukemias. European Group for the Immunological Charac-terization of Leukemias (EGIL). Leukemia 1995;9(10):1783-6.

44. Hsu CL, King-Fleischman AG, Lai AY, Matsumoto Y, Weissman IL, Kondo M. Antagonistic ef-fect of CCAAT enhancer-binding protein-alpha and Pax5 in myeloid or lymphoid lineage choice in common lymphoid progenitors. Proc Natl Acad Sci U S A 2006;103(3):672-7.

45. Heavey B, Charalambous C, Cobaleda C, Busslinger M. Myeloid lineage switch of Pax5 mutant but not wild-type B cell progenitors by C/EBPalpha and GATA factors. Embo J 2003;22(15):3887-97.

46. Xie H, Ye M, Feng R, Graf T. Stepwise reprogramming of B cells into macrophages. Cell 2004;117(5):663-76.

47. Kondo M, Scherer DC, Miyamoto T, King AG, Akashi K, Sugamura K, et al. Cell-fate con-version of lymphoid-committed progenitors by instructive actions of cytokines. Nature 2000;407(6802):383-6.

48. Fukuchi Y, Shibata F, Ito M, Goto-Koshino Y, Sotomaru Y, Ito M, et al. Comprehensive analysis of myeloid lineage conversion using mice expressing an inducible form of C/EBPalpha. Embo J 2006.

49. Chapiro E, Russell L, Radford-Weiss I, Bastard C, Lessard M, Struski S, et al. Overexpression of CEBPA resulting from the translocation t(14;19)(q32;q13) of human precursor B acute lympho-blastic leukemia. Blood 2006;108(10):3560-3.

50. Akasaka T, Balasas T, Russell LJ, Sugimoto KJ, Majid A, Walewska R, et al. Five members of the CEBP transcription factor family are targeted by recurrent IGH translocations in B-cell precursor acute lymphoblastic leukemia (BCP-ALL). Blood 2007;109(8):3451-61.

51. Malecki MJ, Sanchez-Irizarry C, Mitchell JL, Histen G, Xu ML, Aster JC, et al. Leukemia-asso-ciated mutations within the NOTCH1 heterodimerization domain fall into at least two distinct mechanistic classes. Mol Cell Biol 2006;26(12):4642-51.

52. Chiang MY, Xu ML, Histen G, Shestova O, Roy M, Nam Y, et al. Identification of a conserved negative regulatory sequence that influences the leukemogenic activity of NOTCH1. Mol Cell Biol 2006;26(16):6261-71.

53. Jin G, Yamazaki Y, Takuwa M, Takahara T, Kaneko K, Kuwata T, et al. Trib1 and Evi1 cooperate with Hoxa and Meis1 in myeloid leukemogenesis. Blood 2007;109(9):3998-4005.

54. Kisielow J, Nairn AC, Karjalainen K. TARPP, a novel protein that accompanies TCR gene rear-rangement and thymocyte education. Eur J Immunol 2001;31(4):1141-9.

55. Kim J, Lee J, Yadav N, Wu Q, Carter C, Richard S, et al. Loss of CARM1 results in hypomethyla-tion of thymocyte cyclic AMP-regulated phosphoprotein and deregulated early T cell develop-ment. J Biol Chem 2004;279(24):25339-44.

56. Kobielak A, Fuchs E. Alpha-catenin: at the junction of intercellular adhesion and actin dynamics. Nat Rev Mol Cell Biol 2004;5(8):614-25.

57. Drees F, Pokutta S, Yamada S, Nelson WJ, Weis WI. Alpha-catenin is a molecular switch that binds E-cadherin-beta-catenin and regulates actin-filament assembly. Cell 2005;123(5):903-15.

58. Liu TX, Becker MW, Jelinek J, Wu WS, Deng M, Mikhalkevich N, et al. Chromosome 5q dele-tion and epigenetic suppression of the gene encoding alpha-catenin (CTNNA1) in myeloid cell transformation. Nat Med 2007;13(1):78-83.

59. Tang L, Dai DL, Su M, Martinka M, Li G, Zhou Y. Aberrant expression of collagen triple helix repeat containing 1 in human solid cancers. Clin Cancer Res 2006;12(12):3716-22.

CHAPTER 6

Genome wide epigenetic analysis delineates a biologically distinct immature acute leukemia with myeloid/T-lymphoid features

Maria E. Figueroa[1,*], Bas J. Wouters[2,*], Lucy Skrabanek[3], Jacob Glass[4], Yushan Li[1], Claudia A.J. Erpelinck-Verschueren[2], Anton W. Langerak[5], Bob Löwenberg[2], Melissa Fazzari[6], John M Greally[4,7], Peter J.M. Valk[2], Ari Melnick[1] and Ruud Delwel[2]

* These authors contributed equally to the study

[1] Department of Medicine (Hematology Oncology Division), Weill Cornell Medical College, New York, NY, USA

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

[3] Department of Physiology and Biophysics and HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY, USA

[4] Department of Molecular Genetics, Albert Einstein College of Medicine, New York, NY, USA

[5] Department of Immunology, Erasmus University Medical Center, Rotterdam, The Netherlands

[6] Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, NY, USA

[7] Department of Medicine, Albert Einstein College of Medicine, New York, NY, USA

ABSTRACT

Acute myeloid leukemia (AML) is a heterogeneous disease from the molecular and biological standpoints, and even patients with a specific gene expression profile may present clinical and molecular heterogeneity. We studied the epigenetic profiles of a cohort of patients that shared a common gene expression profile but differed in that only half of them harbored mutations of the *CEBPA* locus, while the rest presented with silencing of this gene and co-expression of certain T cell markers. DNA methylation studies revealed that these two groups of patients could be readily segregated in an unsupervised fashion based on their DNA methylation profiles alone. Furthermore, *CEBPA* silencing was associated with the presence of an aberrant DNA hypermethylation signature, which was not present in the *CEBPA* mutant group. This aberrant hypermethylation occurred more frequently at sites within CpG islands. *CEBPA* silenced leukemias also displayed marked hypermethylation when compared with normal CD34+ hematopoietic cells, while *CEBPA* mutant cases showed only mild changes in DNA methylation when compared to these normal progenitors. Biologically, *CEBPA* silenced leukemias presented with a decreased response to myeloid growth factors in vitro.

INTRODUCTION

Aberrant transcriptional programming is a hallmark of leukemogenesis (1). While it is well accepted that transcription factors play a crucial role in directing lineage decisions during normal hematopoiesis, it is becoming increasingly clear that epigenetic modifications also have a profound influence on regulation of gene transcription (2). It is likely that these levels of regulation are not mutually exclusive – for instance, it has been shown that certain lineage-specific transcription factors such as PU.1 can guide the locations in the genome that become methylated (3). Still, it remains largely unknown how DNA methylation patterns are determined.

The basic leucine zipper transcription factor CCAAT/enhancer binding protein-alpha (C/EBPα) is a key regulator of hematopoietic stem cell homeostasis and normal myeloid development, and disruption of its normal function has been implicated in myeloid malignancies (4,5). The *CEBPA* gene is reported to be mutated in 5-15% of all cases of acute myeloid leukemia (AML) (e.g. (6-9)). Furthermore, it can also be aberrantly repressed by oncogenes such as AML1-ETO (10), or modified at the post-translational level through the actions of FLT3 and Ras (11,12). Methylation of the *CEBPA* promoter has also been reported in AML as well as in other malignancies (13-16).

In a genome-wide gene expression profiling study of 285 cases of de novo AML, it was found that AMLs with mutations in *CEBPA (CEBPA^mut)* featured gene expression signatures that clearly discriminated them from other AML subgroups (17). However, a small subset of leukemias without *CEBPA* gene aberrations was found to express a similar gene expression signature. It was found that a frequent characteristic in this subset of leukemias was CpG hypermethylation of the *CEBPA* proximal promoter, which was associated with silencing of *CEBPA* expression (*CEBPA^sil*). Subsequent analyses revealed an immature mixed myeloid/T-lymphoid immunophenotype and frequent *NOTCH1* mutations as additional characteristics of these leukemias (16).

Although these studies suggested a role for epigenetic silencing of *CEBPA* in the leukemic phenotype, they also raised several questions: 1) is silencing of *CEBPA* through hypermethylation an isolated event, or rather one aspect of broader epigenetic modifications, 2) are particular biological pathways specifically involved in the disease phenotype, and 3) are these leukemias more similar to T-cell acute lymphoblastic leukemia (T-ALL) than AML? In order to address these questions and further investigate the role of DNA methylation in this subgroup of leukemias we analyzed DNA methylation on a genome-wide scale in *CEBPA^mut* and *CEBPA*^sil patients and also in T-ALL. We found that *CEBPA* methylation was associated with aberrant hypermethylation of many genes as compared to *CEBPA^mut* AMLs or to normal CD34+ hematopoietic progenitor cells. These hypermethylated genes included many factors involved in transcriptional regulation and mesenchymal cell differentiation and harbored several common genomic features. The epigenetic signature of this form of leukemia was

distinct from T-ALL. Taken together, these data suggest the existence of a subtype of im-mature myeloid/T-lymphoid leukemia in which aberrant epigenetic programming results in a disease that is refractory to normal environmental signals and anti-leukemia therapy.

MATERIALS AND METHODS

Patient samples

Twenty-five acute leukemia cases were studied. They included 14/15 leukemias from the original gene expression cluster #4 described in Valk et al (17), and represented 8 cases with *CEBPA* mutations and 6 cases with silenced *CEBPA* characteristics. Two additional cases of *CEBPA* silenced leukemias, which had been identified in a second AML cohort and described in our previous work (16), were also studied. Nine adult T-ALLs were selected, represent-ing various stages of maturation. Characteristics of these T-ALL cases are given in Table S1. Normal CD34+ progenitor cells were purified from bone marrow specimens from 8 healthy donors: 4 acquired from the Translational Trials Development and Support Laboratory, Cincinnati Children's Hospital and 4 purchased from Allcells (Emerville, CA, USA).

DNA methylation analysis by HELP

High-molecular-weight DNA was isolated from white cell fractions using a standard high salt procedure and the HpaII tiny fragment enrichment by ligation mediated PCR (HELP) assay was carried out as previously described (18,19). All samples for microarray hybridization were processed at the Roche-NimbleGen Service Laboratory. Samples were labeled using Cy-labeled random primers (9mers) and then hybridized onto a human HG17 custom-designed oligonucleotide array (50-mers) covering 25,626 HpaII amplifiable fragments (HAF) located at gene promoters and imprinted regions. HpaII amplifiable fragments are defined as ge-nomic sequences contained between two flanking HpaII sites found within 200-2,000 bp from each other. Each HAF on the array is represent by 15 individual probes, randomly distributed across the microarray slide. Scanning was performed using a GenePix 4000B scanner (Axon Instruments) as previously described (20). Quality control and data analysis of HELP microarrays was performed as described (21). DNA methylation was measured as the log(HpaII/MspI) ratio, where HpaII reflects the hypomethylated fraction of the genome and MspI represents the whole genome reference. All microarray data have been submitted to the GEO repository (accession numbers GSE14417 and GSE14479).

Gene expression microarrays

Detailed protocols provided as supplementary methods.

Microarray data analysis

Unsupervised clustering of HELP and gene expression data by principal component analysis was performed using the statistical software R 2.6.2 and the BioConductor package MADE4 (22). Supervised analysis was carried out using a moderated T-test with a significance level of p <0.001 (all p values were still significant at p<0.05 after correcting for multiple testing using the Benjamini-Hochberg approach) and fold change > 2 for gene expression. In the case of the methylation data, an absolute difference in methylation > 2 between the means of the two populations was required in order to increase the likelihood of detection of biologically significant changes in methylation levels.

Quantitative DNA methylation analysis by MassARRAY EpiTyping

Described in supplementary methods.

CEBPA status classifier

The DNA methylation status for each HpaII amplifiable fragment was categorized as follows: fragments having log2 (HpaII/MspI) < 0 were considered methylated, and those with log2 ratios > 0 were considered hypomethylated. AML samples were categorized according to their *CEBPA* locus mutation status, and a Fisher's exact test was performed at each locus to rank its predictive ability. Cutoffs for perfect, intermediate and poor classifiers were set at p-values of 1.55e-4, (which corresponds to Fishers Exact p-value based on perfect classification for this sample size), 0.2 and > 0.2, respectively. The use of 0.20 as the threshold was arbitrarily set, but represents an acceptable p-value range for intermediate classification ability given the sample size.

Sequence analysis

Sequence retrieval, Repeat element analysis and Motif analysis are described in the supplementary methods section.

In vitro cultures and thymidine incorporation assay

Described in Supplementary materials section.

RESULTS

CEBPA silencing is associated with an aberrant DNA methylation signature

In order to determine whether DNA methylation of the *CEBPA* promoter was an isolated event or part of widespread aberrant epigenetic gene silencing, we carried out a genome-wide DNA methylation study. For this, we used the HpaII tiny fragment enrichment by Ligation-mediated PCR (HELP) assay (19), a method that accurately identifies the DNA methylation

Figure 1. A unique methylation profile distinguishes *CEBPA^sil* from *CEBPA^mut* AML.
A. Principal component analysis of DNA methylation data using the HELP assay on 8 *CEBPA^sil* and 8 *CEBPA^mut* AML cases revealed that the cases were readily segregated into two clusters, which matched exactly with *CEBPA* status.
B. heatmap representation of the four probe sets annotated to the *CEBPA* locus on the HELP microarray; cases are clustered according their methylation status. *CEBPA^sil* cases cluster together (left node) and all show higher levels of methylation for at least 3 or the 4 probe sets.
C. Representation of the positioning of the four probe sets relative to the genomic localization of the *CEBPA* locus and its CpG island on chromosome 19. HELP methylation values for each leukemia case are represented in one row; the y axis represents centered log2 (HpaII/MspI) ratios. Positive values correspond to hypomethylated fragments, while a negative deflection reflects a methylated fragment. The first 8 rows correspond to the *CEBPA^sil* cases (in red) and the remaining rows to the *CEBPA^mut* cases (in blue).
D. Heatmap representing the DNA methylation status at five different regions of the *CEBPA* locus. Percent cytosine methylation was determined at these regions for all cases using MassARRAY EpiTyper.
E. Two-dimensional hierarchical clustering of genes differentially methylated between the two leukemia subgroups, illustrated by a heatmap. Supervised analysis identified 567 HpaII amplifiable fragments (p <0.001 and absolute difference in methylation >2). Cases are represented in the columns and probe sets in the rows. *CEBPA^sil* cases are clustered in the left node, and display high methylation levels for 563 HpaII amplifiable fragments. *To the right* heatmap representation of the four probe sets that displayed the opposite behavior i.e. relative hypomethylation in *CEBPA^sil* leukemia.
F. A plot of methylation difference between *CEBPA^sil* and *CEBPA^mut* cases (x axis) vs. statistical significance (y axis) shows the marked asymmetry of the two branches, illustrating the overall tendency to higher methylation levels in the *CEBPA^sil* cases. Red points demark probe sets that reached both criteria for differential methylation on our analysis (p<0.001 and absolute methylation difference >2)
(A full color version of this figure can be found in the color section.)

level of 25,626 HpaII amplifiable fragments annotated to ~14,000 human gene promoters. All of the methylation profiles passed a rigorous quality control and normalization procedure as described (21). The accuracy in detecting variance in DNA methylation was validated by performing single locus quantitative EpiTyping on a panel of ten genes (Figure S1).

We first assessed the underlying variability in DNA methylation of 8 cases with *CEBPA^sil* and 8 cases with *CEBPA^mut* using principal component analysis (PCA). This analysis demonstrated separate clustering of the two groups, indicating that these groups of patients are epigenetically distinct (Figure 1A). The *CEBPA* locus itself is represented on the HELP microarray by four HpaII amplifiable fragments. These fragments correspond to nucleotides +1334 to +1980, +857 to +1113, -1569 to -1819, and -2247 to -2863, relative to the transcription start site. Analysis of these fragments confirmed our previous observations: while the region was unmethylated in all *CEBPA^mut* cases, it was highly methylated in all *CEBPA^sil* cases (two-tailed T test: p <1e-9) (Figure 1B and 1C). DNA methylation of the *CEBPA* locus was confirmed using MassARRAY EpiTyping (Figure 1D).

In order to identify the genes other than *CEBPA* that were differentially methylated between the two groups we performed a supervised analysis. This comparison identified 567 HpaII amplifiable fragments (474 genes), which included *CEBPA*. Strikingly, almost all (n=470) of these genes were hypermethylated in the *CEBPA^sil* leukemias. Hypermethylation is thus a specific biological feature of these leukemias that distinguishes them from *CEBPA^mut* cases. Only 4 probe sets were hypomethylated in *CEBPA^sil* leukemias, which included *LCK*

Table 1. Genes in the top 5 networks generated by the differentially methylated signature between *CEBPA*[sil] and *CEBPA*[mut] cases.

#	Molecules in Network	Focus molecules
1	ADAMTS9, BARX1, CaMKII, CD180, CDH8, CHAT (includes EG:1103), DRD1, ELAVL3, FGF5, GNAS, GRIN2B, Gs-coupled receptor, HTR7, IGHM, IGKC, ISL1, KRT17, KRT19, LHX3, MAP6, MEOX2, MNX1, NEFL, NEFM, NEUROD1, NEUROG3, NFkB, NKX2-2, PAX1, PENK, PTPN13, RASGRF1, SATB2, TLR1, TLR8	32
2	CCK, CDKN2B, CDX2, CEBPA, COL4A1, Cyclin A, E2f, EID2, ESX1, FOXG1, HGF, HIST1H2BJ, Histone h3, Insulin, IRF8, JTV1, LRAT, Mapk, MARCKS (includes EG:4082), PGR, PITX2, POU2F3, POU3F2, PRLR, PTGFR, Rap1, Ras, RASA2, Rb, RELN, Rsk, Smad1/5/8, SMPDL3A, TFAP2C, Vegf	24
3	Alkaline Phosphatase, ALPL, APOB, ATOH1, BMP, BMP3, BMP6, BMP7, FCGR2B, FOXA2, FOXC2, GSC, IL1, LDL, LPL, LRP6, MHC Class II, Mmp, MSX2, NELL1, ONECUT2, OSBPL6, P38 MAPK, PCDHGC3, PDGF BB, PLC gamma, PTHLH, SFRP1, SOX9, Tgf beta, UCP2, Wnt, WNT2, WNT10B, ZAK	24
4	ACTN2, ADCY1, Adenylate Cyclase, AFAP1, Alpha Actinin, BDNF, Cacna1, CACNA1A, CACNA1E, CACNA1I, CACNA2D1, CALML5, Calmodulin, F Actin, FOXL1, GAL, GALR1, GRIN3A, GRM1, Jnk, KCNN2, MARCKSL1, MBP, N-type Calcium Channel, NMDA Receptor, OCLN, OPRK1, PDLIM3, Pkc(s), PP2A, SLC6A2, SLC6A5, SNAP25, TAC1, Voltage Gated Calcium Channel	24
5	ADAM12, Akt, BRCA2, Cofilin, Creatine Kinase, Creb, CYR61, EFNA5, EPHA, EPHA2, EPHA3, EPHA5, EPHA7, EPHA/B, GABBR2, GDNF, GFRA1, GREM1, HABP4, HSPB8, HTR1B, IRS4, MAGI1, N-cor, NAE1, p70 S6k, Pdgf, PDGFC, PI3K, PIK3R1, PRKCZ, Rab5, RET, Rock, TOP2B	23

and *TRBC1*, T cell genes which we previously reported to be overexpressed in this group of patients (16)(Figure 1E and 1F; Table S2).

Using Ingenuity Pathway Analysis (IPA) software we carried out a network analysis to investigate whether genes in this differentially methylated signature interrelated with each other. Interestingly, the two top scoring networks were centered around *CEBPA* and *NFkB* and could almost exclusively be generated using genes that were differentially methylated (32/35 and 24/35 genes in the networks, respectively). The fact that the majority of genes in the network were affected by aberrant methylation is highly suggestive of the biological importance of this observation. An additional network involved several BMP factors (in which 25/34 genes were involved), suggesting potential involvement of TGFB signaling in these tumors as well. The genes involved in the three top scoring networks are listed in Table 1 (Figure S2). Among the most represented gene ontology terms within the hypermethylated

Table 2. GO terms highly represented among the differentially methylated genes between the two leukemia subgroups

Go Term	P value	Benjamini	FDR
Mesenchymal cell development	4.1e-5	1.9 e-2	0.1
Neural crest cell differentiation	3.5e-4	5.1e-2	0.6
Regulation of transcription	1.4e-4	4.6e-2	0.2
Transcription DNA-dependent	2.8e-4	4.7e-2	0.5
RNA biosynthetic process	2.9e-4	4.6e-2	0.5

signature were regulation of transcription and mesenchymal cell development. (Table 2) Despite the association of some *CEBPA^sil* cases with *NOTCH1* activating mutations, we did not find a significant overlap between this hypermethylated gene signature and the *NOTCH1* signature reported by Palomero et al (23).

CEBPA^sil and *CEBPA^mut* leukemia cells are epigenetically distinct from normal bone marrow CD34+ cells

Does hypermethylation in *CEBPA^sil* leukemic blasts represent aberrant epigenetic programming when compared to normal CD34+ cells? In a supervised comparison of *CEBPA^sil* leukemia cells with CD34+ healthy donor bone marrow fractions (n=8) a large number of differentially methylated loci (1035 HpaII amplifiable fragments, 876 unique genes) was found. Similarly as in the previous analyses, most of the genes, i.e. 841/876, were hypermethylated in the *CEBPA*^sil leukemia subgroup (Figure 2A and B). Moreover, 281 of these 876 genes appeared to be hypermethylated both when compared to normal CD34+ cells and to *CEBPA^mut* leukemias, indicating that aberrant hypermethylation of these genes occurred exclusively in *CEBPA^sil* cases. These included among others the tumor suppressors *CDKN2B* and *IRF8,* the cell cycle regulator *MyoD1,* the WNT signaling antagonist *SFRP1.* Loss of all these genes has been previously implicated in tumorigenesis (24-28). Furthermore, hypermethylation of the promoter regions of *CDKN2B, MyoD1* and *SFRP1* have been described in association with several forms of leukemia (25,29-34). A complete list of the genes exclusively methylated in the *CEBPA^sil* subgroup vs. normal CD34+ cells is provided in Table S3. Both of the proximal HpaII amplifiable fragments for the *CEBPA* locus reached statistical significance and showed hypermethylation in the *CEBPA^sil* group vs. the normal CD34+ cells, with an absolute methylation change slightly below our set threshold of 2 (change -1.63 and 1.99) (Figure S3). Only 45 of the 1035 HpaII amplifiable fragments were hypomethylated in the *CEBPA^sil* cases while hypermethylated in normal CD34+ cells. These loci included the T cell genes *CD3D* and *TRBC1.*

In the *CEBPA^mut* group, abnormal methylation was less abundant, with only 322 HpaII amplifiable fragments (286 unique genes) detected as differentially methylated with respect to normal CD34+ cells. Moreover, the difference in DNA methylation was distributed evenly,

Figure 2. Aberrant hypermethylation is a feature of *CEBPA^sil* leukemia.

A. Two-dimensional hierarchical clustering of genes differentially methylated between the *CEBPA^sil* cases and normal CD34+ hematopoietic progenitors, illustrated by a heatmap. Supervised analysis identified 1035 HpaII amplifiable fragments (p <0.001 and absolute difference in methylation >2). Cases are represented in the columns and probe sets in the rows. *CEBPA^sil* cases are clustered in the left node, and display marked hypermethylation when compared to the normals, as illustrated by the predominance of probe sets with low log2 (HpaII/MspI) ratios.

B. Methylation difference between *CEBPA^sil* and normal CD34+ cells (x axis) vs. statistical significance (y axis) plot with marked asymmetry of the two branches, reflecting the tendency to higher methylation levels in this subgroup. Red points demark probe sets that reached both criteria for differential methylation in our analysis.

C. Two-dimensional hierarchical clustering of genes differentially methylated between the *CEBPA^mut* AML and normal CD34+ hematopoietic progenitors, illustrated by a heatmap. Supervised analysis identified 322 probe sets (286 genes). *CEBPA^mut* cases are clustered in the left node, and display equal components of *hyper* and *hypo*methylation when compared to CD34+ normal cells.

D. A plot of methylation difference between mutant *CEBPA* and normal CD34+ cells (x axis) vs. statistical significance (y axis) shows symmetric branches and less pronounced differences in methylation than in the case of the silenced *CEBPA* subgroup. Red points demark probe sets that reached both criteria for differential methylation on our analysis.

(A full color version of this figure can be found in the color section.)
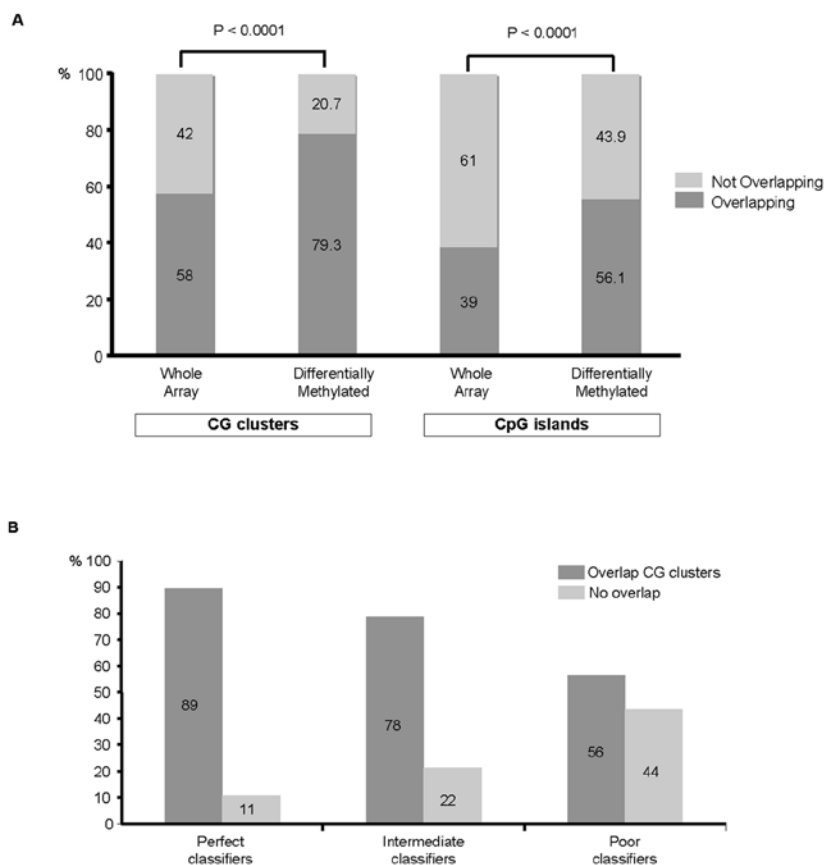
Figure 3. Aberrant hypermethylation colocalizes to CpG islands and CG clusters.
A. The genomic position of every HpaII amplifiable fragment on the HELP array was compared to the location of known CpG islands and CG clusters, and the fragments on the array were divided into two categories: those overlapping with either one of these genomic elements and those not overlapping. In order to determine if the differentially methylated genes between *CEBPA^sil^* and *CEBPA^mut^* leukemias were enriched for either one of these types of elements, a proportions test was used to compare the relative proportion of the two types of HpaII fragments in the signature (overlapping vs. not overlapping) to the relative proportion on the array. Stacking bars are used to illustrate the finding of a significant enrichment for HpaII amplifiable fragments overlapping with CpG islands (right) and CG clusters (left) in the hypermethylated signature of *CEBPA^sil^* leukemia, as it compares to the genomic localization of all HpaII amplifiable fragments on the HELP array.
B. Each HpaII amplifiable fragment represented on the HELP array was also categorized according to its ability to discriminate between *CEBPA^sil^* and *CEBPA^mut^* leukemias as perfect, intermediate or poor classifiers and the proportional amount of fragments overlapping with CG clusters were calculated for each group of classifiers. Better classifiers were more frequently associated with CG clusters.

where approximately half of these probe sets were hypermethylated and the other half hypo methylated (Figure 2C and D). This finding indicates that the markedly increased prevalence of abnormal promoter hypermethylation is highly specific to *CEBPA^sil^* cases among the *CEBPA* gene expression signature AML patients.

When we compared both subgroups of leukemias with normal CD34+ cells, 100 unique genes were found to be aberrantly methylated in the leukemia samples. These included *CDKN2A*, *CCR7, MT2A* and *NOS3*, and were all linked together in a cancer/cell signaling network through IPA (Table S4). This finding suggests that aberrant hypermethylation of these genes is a common event in both these leukemogenic processes, and therefore deserves further study.

Hypermethylation in *CEBPA^sil* leukemias is mainly localized to CpG islands and CG clusters

To obtain further insight into the differences in methylation profiles between *CEBPA^sil* leukemias and *CEBPA^mut* AMLs, we next asked whether the observed differences in DNA methylation were found at either CpG islands or CG clusters (35) more frequently than expected based on the distribution of all fragments on the arrays. For this analysis, we categorized all HpaII amplifiable fragments on the HELP array into two categories: those that overlapped with a known CpG island or CG cluster, and those that did not. We found 15,043/25,626 (58%) fragments to be completely located within, or partly overlapping with CG clusters. Furthermore, 10,006/25,626 (39%) fragments were located in or partly overlapped with CpG islands. Of the 567 fragments identified as differentially methylated between the two leukemia groups, 450 (79.3%) overlapped with at least one CG cluster, while 318 (56.1%) overlapped with a CpG island (Figure 3A). Thus, there was a significant enrichment in the signature for fragments overlapping with either CpG islands (56% versus 39%) as well as with CG clusters (79% versus 58%) (proportion test: p <0.0001 for both), which is consistent with their proposed regulatory functions. On the other hand, these analyses show that a subset of the differentially methylated HpaII amplifiable fragments did not overlap with CpG islands or CG clusters, underscoring the importance of extending DNA methylation studies beyond those areas of the genome containing CpG islands and CG clusters.

Next we categorized each HpaII amplifiable fragment into methylated or hypomethylated and ranked them by their ability to discriminate between *CEBPA^mut* and *CEBPA^sil* cases (summarized by Fisher's exact test p-value). Fragments were then grouped into perfect, intermediate or poor classifiers. 66 fragments were categorized as perfect classifiers. Of these, 59 (89%) were located in CG clusters (52 of which were also annotated as CpG islands). All 66 fragments exhibit consistent methylation in *CEBPA^sil* cases and hypomethylation in *CEBPA^mut* cases. Intermediate classifiers showed 78% of the fragments overlapping with CG clusters and only 56% of the poor classifiers overlapped with a CG cluster (Figure 3B). Taken together, these data suggest that among differentially methylated genes, those that contain CpG islands or clusters are most critical for defining the epigenetic signature of *CEBPA^sil* leukemia cases.

Hypermethylated promoters in *CEBPA*^{sil} display specific genomic features

We next examined whether genes susceptible to becoming methylated in *CEBPA*^{sil} leukemias (as compared to *CEBPA*^{mut}) shared any common genomic features. The regions spanning the 2 kb sequence upstream of the transcription start site of the 327 RefSeq genes assigned to the differentially methylated 567 HpaII amplifiable fragments were retrieved. Next, these sequences were compared to 2422 RefSeq gene promoter sequences (corresponding to 2672 HpaII amplifiable fragments) that were randomly selected from the pool of HpaII amplifiable fragments that did not show any methylation difference between *CEBPA*^{sil} and *CEBPA*^{mut}. Remarkably, hypermethylated genes were significantly less likely to contain an Alu sequence than the control sequences. (Genes containing Alu sequences: 60% (1453/2422) in the control sequences vs. 26% (85/327) in the differentially methylated sequences; Fisher Exact test p-value <2.2e-16) (Table 3). This was the only repeat element that showed a statistically significant distribution difference between the *CEBPA*^{sil} signature and the randomly selected control group. All other repeat sequences (e.g., SINE/MIRs, LINEs or MERs) were present in approximately equal proportions. We also looked at the promoter regions from 780 RefSeq genes annotated to the 1035 HpaII amplifiable fragments differentially methylated between *CEBPA*^{sil} leukemias vs. normal CD34+ bone marrow cells. We compared these to

Table 3. Alu elements are depleted from RefSeq promoters that become hypermethylated in *CEBPA*^{sil} leukemia.

	Silenced RefSeq (n=327)	Random RefSeq controls (n=2422)	Fisher Exact Test
Alu	85 (26%)	1453 (60%)	p < 2.2e-16
No Alu	242 (74%)	969 (40%)	

3086 randomly selected RefSeq promoters (corresponding to 4470 control HpaII amplifiable fragments). Once again we found that differentially methylated sequences were less likely to contain Alu elements (60% in control sequences (1865/3086) vs. 35% in the differentially methylated sequences (277/780); Fisher Exact Test p-values <2.2e-16).

In order to determine whether a specific DNA sequence was associated with the differentially methylated genes, we used the Finding Informative Regulatory Elements (FIRE) algorithm (36) to look at a region containing 2 kb of sequence upstream of the reported transcription start site. This analysis identified a non-degenerate 5'-CGCGCTC-3' motif, which was significantly enriched in the differentially methylated sequences compared to the control sequences. The non-degenerate motif could be detected in 45% of the differentially methylated sequences (147/327), while it was found in less than 24% of the control sequences (574/2422) (Fisher Exact Test p-value <3.3e-13). An independent FIRE analysis was carried out to study the promoter sequences of the RefSeq genes annotated to the probe sets in the *CEBPA*^{sil} vs. normal CD34+ cells signature. Although this analysis discovered the same non-
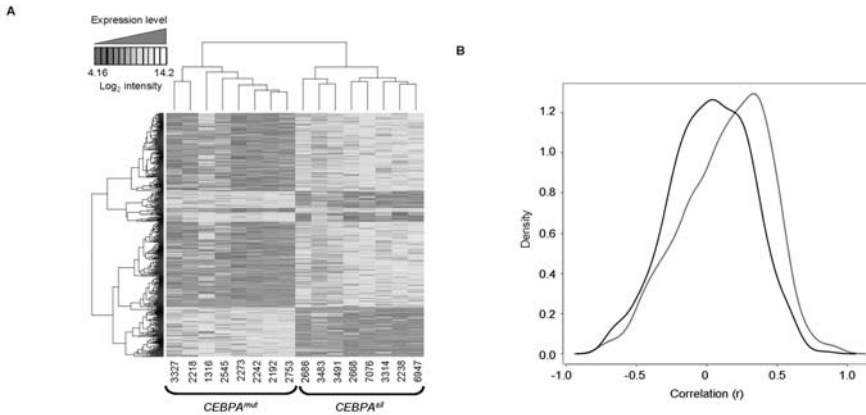
Figure 4. DNA methylation and gene expression capture complementary information.
A. Two-dimensional hierarchical clustering of genes differentially expressed between the two leukemia subgroups, illustrated by a heatmap. Supervised analysis identified 587 probe sets (415 genes) at a p <0.001 and fold change >2. Cases are represented in the columns and probe sets in the rows. *CEBPA^sil* cases are clustered in the right node, and *CEBPA^mut* cases are clustered on the right node.
B. Density (y axis) plot for the gene-by-gene correlations (x axis) between gene expression log intensity and log(HpaII/MspI) values. A positive correlation between these two measures translates into a negative biological correlation, i.e. hypermethylation in combination with lower expression levels, or hypomethylation in combination with higher expression levels. In black, density plot for the correlation between expression and methylation for a set of 600 randomly selected probe sets. In red, density plot for the correlation between expression and methylation measured by the 567 probe sets in the genes differentially methylated between the two subgroups. The shifting of the density plot to the right reflects a tendency to a stronger correlation of DNA methylation with gene expression levels in this subset of genes. The figure is representative of five analyses; each time using a different set of 600 randomly selected HpaII amplifiable fragments for the calculation of correlations.
(A full color version of this figure can be found in the color section.)

degenerate motif as before, it was slightly less successful at discriminating between the two groups (present in 32% of the differentially methylated sequences (251/780) vs. 20% of the control sequences (628/3086); Fisher Exact test p-value <8.7e-12). Although this motif does not correspond to a known transcription factor binding site, FIRE annotated sequences containing this motif as being associated with the Gene Ontology term GO:0043565, involved in sequence-specific DNA binding (p-value <1e-07). The enrichment of this element within the hypermethylated signature is suggestive of the existence of a putative DNA binding protein that might participate in establishing this epigenetic signature.

Methylated genes in *CEBPA*^sil leukemias tend to be expressed at low levels
To determine the correlation of DNA methylation with gene expression, we examined genome wide mRNA transcript abundance of the same *CEBPA^sil* and *CEBPA^mut* cases using Affymetrix HGU133 Plus 2.0 microarrays. An unsupervised analysis using PCA indicated that, in contrast to DNA methylation profiles, expression data did not result in segregation along the major axes (Figure S4). Nevertheless, a supervised analysis identified 587 probe sets, corresponding to 415 unique genes, as differentially expressed between the two groups
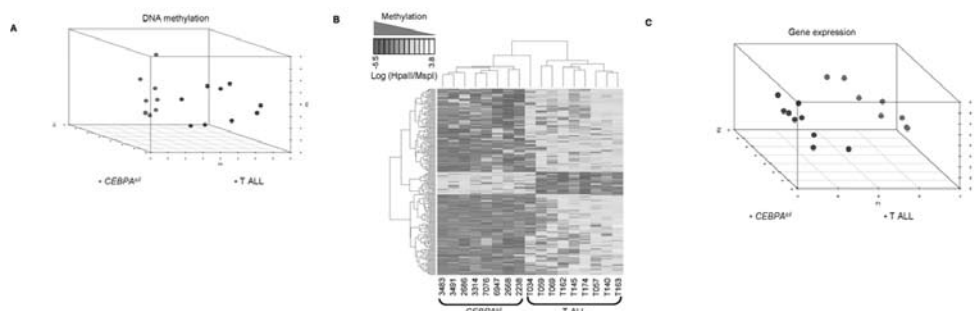
Figure 5. *CEBPA^{sil}* leukemias differ genetically and epigenetically from T-ALL.

A. Principal component analysis of DNA methylation data comparing 8 immature acute myeloid/T lymphoid *CEBPA^{sil}* cases to a selection of 9 T-ALL cases representing a spectrum of maturation stages, showing separate clustering of the two groups of leukemias.

B. Two-dimensional hierarchical clustering of genes differentially methylated between the *CEBPA^{sil}* leukemias and the T-ALL cases, illustrated by a heatmap. Cases are represented in the columns and probe sets in the rows. Supervised analysis identified 213 differentially methylated probe sets (199 genes). *CEBPA^{sil}* cases are clustered in the left node, and display a predominance of hypomethylated probe sets.

C. Principal component analysis of gene expression data for the same cases also demonstrates separate clustering of the two groups of leukemias, indicating that these two groups display distinct expression profiles.

(A full color version of this figure can be found in the color section.)

of leukemias, which is in line with our previous findings (16) (Figure 4A). Overlap between the differentially methylated and differentially expressed genes was only minimal (12 unique genes, including *CEBPA* itself). This is consistent with our previous report showing that gene expression and methylation profiling tend to capture biological variance in different sets of genes (18). Interestingly, signal intensities of 62.5% of the genes detected as differentially hypermethylated were measured at low abundance (intensity $<\log_2(100)$) in both the *CEBPA^{sil}* and the *CEBPA^{mut}* groups. This may imply that these genes are epigenetically silenced by DNA methylation in one group and repressed through a different mechanism in the other. Along these lines, expression levels for most of these genes were detected at similarly low levels in an independent cohort of 400 AML cases of other subtypes (data not shown). Methylation and gene expression were validated by MassArray on 15 genes for 3 *CEBPA^{sil}* and 4 *CEBPA^{mut}* randomly selected cases, and by qRT-PCR on 10 of these genes on 2-4 randomly selected cases from each group) (Figure S5).

We next examined whether differential methylation in the genes contained within the *CEBPA^{sil}* vs. *CEBPA^{mut}* subtypes was more correlated to expression levels than methylation occurring in randomly selected subsets of genes. For this we determined the correlation between transcript abundance and DNA methylation status for the 474 genes differentially methylated between the two leukemia subgroups, and compared it to that of genes annotated to 600 randomly selected probe sets from the HELP microarray. This comparison was repeated for five different randomly selected groups of control probe sets. In every instance the differentially methylated genes showed a slightly stronger correlation with gene expression
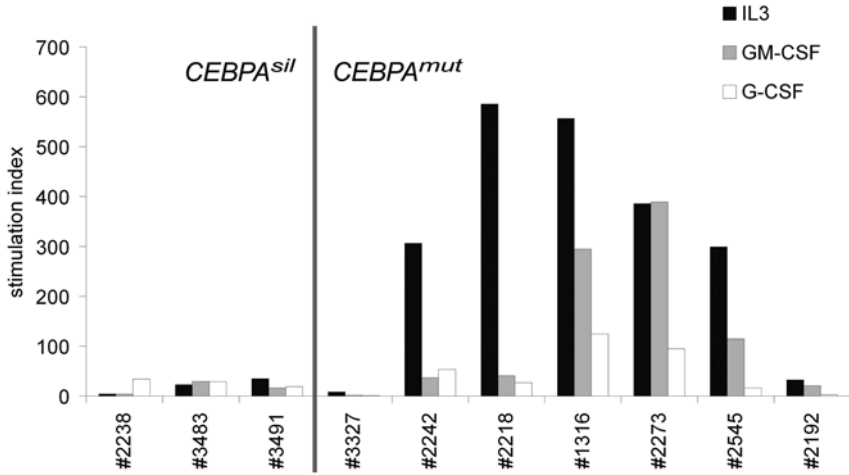
Figure 6. *CEBPA*^sil leukemia cells are resistant to myeloid growth.
Tritiated thymidine incorporation experiment of three *CEBPA*^sil and seven *CEBPA*^mut leukemia samples. Response of leukemia cells to IL3, GM-CSF and G-CSF is shown relative to irradiated non-stimulated cells, i.e. stimulation index, based on the mean of three measurements.

levels (corresponding to a negative biological correlation between DNA methylation and gene expression) than the randomly selected genes (Figure 4B).

CEBPA^sil leukemias are epigenetically distinct from T-ALL

We hypothesized that *CEBPA*^sil leukemias, although expressing certain T-cell genes, are biologically and epigenetically not only different from other AMLs, but also from T-ALL. To test this hypothesis, we generated HELP profiles from a selection of 9 T-ALL cases from various maturation stages, and compared those to the methylation data of the *CEBPA*^sil leukemias. Unsupervised analysis of HELP using PCA resulted in separate clustering of the *CEBPA*^sil leukemias and the T-ALL cases (Figure 5A). A supervised analysis identified 164 genes (199 HpaII amplifiable fragments) differentially methylated between the *CEBPA*^sil leukemias and T-ALLs (Figure 5B and Table S5). Specific analysis of the *CEBPA* locus HpaII fragments indicated a mixed pattern in the T-ALL samples. While the majority of cases (n=6) showed hypermethylation of the *CEBPA* proximal promoter, the remaining 3 did not (Figure S6). A similar separation of *CEBPA*^sil leukemias and T-ALL samples was observed based on gene expression data (Figure 5C), indeed suggesting that *CEBPA*^sil leukemias are biologically and epigenetically distinct from T-ALL.

CEBPA^sil leukemia cells are resistant to myeloid growth factors

To study whether silencing by methylation of the myeloid specific transcription factor *CEBPA* corresponded with altered biological response to natural growth stimuli, we compared growth factor stimulated ³H-TdR incorporation data obtained from pre-treatment

CEBPA^sil and *CEBPA^mut* AML samples. Remarkably, *CEBPA^sil* leukemia cells showed little or no response to the myeloid growth stimuli IL3, GM-CSF or G-CSF, whereas *CEBPA^mut* samples were particularly sensitive to IL3, but in certain cases also to GM-CSF (Figure 6 and Table S6). While two of the *CEBPA^mut* cases (#3327 and #2192) showed comparable results to the *CEBPA^sil* cases, this did not appear to be explained by their methylation profiles, and we believe that it is a reflection of the biological heterogeneity of the group. In contrast, both *CEBPA^sil* and *CEBPA^mut* patient cells showed weak but significant response to factors that stimulate primitive cells, in particular TPO or SCF (Table S6). These data are in agreement with previous data showing loss of the GMP fraction and enrichment of HSC population in marrow of conditional *Cebpa* knock-out animals (37). Interestingly, we did not observe promoter hypermethylation of the genes encoding the receptors for IL3, GM-CSF or G-CSF, suggesting other mechanisms of altered growth factor response. Several of the growth factor receptors appeared to be expressed as determined by gene expression profiling analysis, although we observed significantly lower expression of *CSF2RB,* the gene encoding the common beta chain of the receptors of IL3 and GM-CSF (p=0.02, Wilcoxon rank sum test) in *CEBPA^sil* patients as compared to *CEBPA^mut* AML samples. Together, these data suggest that *CEBPA^sil* cases reflect a more primitive hematopoietic cell that does not enter cell cycle in response to myeloid growth factors, an event that could be associated with chemo-resistance. In agreement, although the numbers of patients were too small to rule out the influence of co-variables, the *CEBPA^sil* patient outcomes (5 year overall survival (OS) 25%) were considerably worse than that of the *CEBPA^mut* cases (5 year OS 88%, log rank test p <0.0027).

DISCUSSION

The subtype of immature myeloid/T-lymphoid leukemia reported in this study was first captured as belonging to a larger group of patients all sharing a specific gene expression signature (originally designated as cluster #4 in Valk et al (17)). A subset of these patients was subsequently observed to harbor hypermethylation and silencing of the *CEBPA* gene along with expression of T-cell genes (16). In the current study we applied integrated epigenetic profiling to explore in greater depth the nature of this disease, and demonstrate that this is in fact a biological entity that is distinct from both AML and T-ALL and has unique epigenetic, transcriptional, and biological features.

Given the similarity in gene expression profiles of *CEBPA^sil* and *CEBPA^mut* AML found initially, we were primarily interested in determining whether DNA methylation profiles might more readily discriminate these two groups. Along these lines, we have previously shown that integrating information obtained from DNA methylation profiling and gene expression profiling in human leukemia specimens could yield more in depth insight into biological differences between patients (18). We find that PCA, an unbiased method to discriminate

biological differences, showed a marked separation between epigenetic profiles of patients with mutant or silenced *CEBPA*. This indicates that these are two biologically distinct entities, which is further supported by the differences in growth factor responses between the two leukemia subtypes. In contrast, gene expression was not as obviously distinct, demonstrating for the first time in a practical manner that these two methods of interpreting transcriptional programming are complementary and when combined can most robustly identify new disease subtypes.

One remarkable finding in the *CEBPA^{sil}* cases was that their unique DNA methylation signature consisted almost entirely in *hyper*methylation of hundreds of loci when compared to their *CEBPA^{mut}* counterparts. By chance, one would expect a more or less balanced distribution between hyper- and hypomethylation, as observed in the comparison between *CEBPA^{mut}* cases and normal CD34+ cells. The greater tendency towards *hyper*methylation of these leukemias with silenced *CEBPA* was further confirmed by the comparison of the *CEBPA^{sil}* subgroup to normal CD34+ hematopoietic cells. Together, this suggests epigenetic deregulation to play a more critical role in leukemogenesis of those cases than of *CEBPA^{mut}* cases. Such a hypothesis is supported by our observation of involvement of a majority of the members of gene regulatory networks centered around the *CEBPA*, *NFkB* and *BMP* pathways, all implicated in the normal homeostasis of hematopoietic stem cells and the myeloid lineage.

When DNA methylation profiles and gene expression profiles were queried by supervised analysis, 474 and 415 genes were identified as significantly differentially methylated or differentially expressed, respectively. Yet there was only minimal overlap between these two gene lists. This is consistent with our previous observation that gene expression and DNA methylation array studies capture different sets of genes, and thus are complementary to each other when collecting information on biological variance (18). The fact that expression of more than 60 percent of the genes in the hypermethylated signature was not detected in both groups partly explains this finding, since differences in low abundance transcripts may be hard to detect by expression arrays. In spite of all these technical limitations, we still observed a greater tendency to overall inverse correlation between gene expression and DNA methylation status among those genes whose methylation status varied between the two groups. This result suggests that the biological impact of epigenetic modifications may vary between gene subsets. While this has been shown for specific histone modifications such as the association between HOX cluster genes and methylation of lysine 27 on histone 3 (38,39), a similar situation has not been previously shown for cytosine methylation beyond that of its distribution relative to CpG islands.

While the existence of a hypermethylated phenotype in *CEBPA^{sil}* immature acute myeloid/T-lymphoid leukemia could be related to a relatively more immature cell type involved in the leukemic transformation, an alternative explanation could lie in leukemia-specific mechanisms. The latter could involve either a global upregulation of DNA methyla-

tion machinery, leading to more or less uncontrolled hypermethylation of multiple loci, or, alternatively, could be a specific effect of targeted methylation of selected genes. We took several approaches to investigate the various possibilities. Comparative analyses between the two leukemia subgroups and immature CD34+ control samples from healthy individuals indicated in fact a greater state of hypermethylation in the *CEBPA^sil* leukemia blasts than in normal CD34+ cells. This is suggestive for the idea that hypermethylation is not primarily a read-out for cellular maturation status. In contrast, *CEBPA^mut* AMLs had markedly fewer hypermethylated genes compared to CD34+, indicating that an increase in DNA methylation is not a general feature of all AMLs. Interestingly, mRNA expression levels of the *de novo* DNA methyltransferase *DNMT3B* were significantly increased in *CEBPA^sil* versus *CEBPA^mut* leukemias, but showed similarly high levels in normal CD34+ cells (data not shown). Thus, while expression of methylating enzymes may be high in healthy immature cells as well, this does not necessarily lead to a similar hypermethylated signature as found in the leukemic subgroup. There were no differences in expression of other DNMTs (data not shown).

Specific targeting of methylation to certain genes by complexes involving DNMTs and transcription factors has been shown previously (40). Inversely, it could be hypothesized that loss of protective mechanisms, such as DNA binding proteins and nucleosome positioning, could render certain promoters susceptible to the targeting by the DNA methylation machinery (41). A common characteristic of the leukemia subgroup studied here was silencing of the transcription factor C/EBPα. Using algorithms for motif analysis comparing significantly hypermethylated genes to control genes, we were unable to detect a significant enrichment for C/EBP binding sites among these genes, arguing against a direct role for the absence of C/EBPα. These analyses did however yield two potentially relevant findings. First, we identified enrichment of a novel sequence motif. This sequence could possibly play a role in either directing the binding certain protein complexes to DNA or in helping determine nucleosome positioning. Second, a significant depletion of Alu repeats was detected in the promoters of differentially methylated RefSeq genes. While the causes behind such a finding can only be speculated on, similar observation have been made in the past, suggesting that this is unlikely to be a random finding (42).

CEBPA^sil leukemias were previously found to have a mixed myeloid/T-lymphoid phenotype. We were therefore interested to see how their DNA methylation profiles related to those from T-ALL samples. Comparisons of the DNA methylation and expression profiles of a representative collection of 9 T-ALL cases –ranging from very immature to mature– to those of the *CEBPA^sil* leukemias revealed significant differences between the two types of leukemias. This is consistent with the idea that *CEBPA^sil* leukemias represent an entity not entirely myeloid but also not entirely T-lymphoid and may represent a very immature myeloid/T-lymphoid subtype that is in current clinical practice sometimes diagnosed as AML and sometimes as T-ALL.

The discovery of a hypermethylation profile in a specific leukemia subgroup is particularly interesting in light of the recent developments of demethylating drugs. Notably, the *CEBPA*^{sil} leukemia subgroup showed a markedly worse treatment response than *CEBPA*^{mut} AML (16). Interestingly, this is in contrast with the previous report by Hackanson *et al*, who found that methylation of *CEBPA* was associated with favorable prognostic groups in AML (13). However, the cases reported by that group all displayed methylation restricted to an upstream region of the *CEBPA* locus, while the methylation in our samples was more extensive, covering both proximal and distal regions. Although the numbers are small and multivariable analysis with additional covariates such as cytogenetic and molecular subgroups is warranted, these results identify *CEBPA*^{sil} leukemia cases as attractive candidates for investigative treatment with such demethylating agents. It is interesting to speculate that restoring the normal programming of *CEBPA*, *NFkB* and *BMP* pathways could facilitate targeting of these tumors with either standard anti-leukemia regimens or with specific targeted therapy agents.

Taken together, we show that integrated epigenetic and gene expression analyses of leukemia can distinguish and illustrate phenotypes with biological and potential clinical significance. Analysis of the genes involved in aberrant epigenetic programming allows for generation of hypotheses towards dissecting previously unrecognized mechanisms of leukemogenesis. A broader application of integrated gene expression and epigenetic profiling to acute leukemia might allow many other such entities to be resolved.

ACKNOWLEDGEMENTS

REFERENCES

1. Rosenbauer F, Tenen DG. Transcription factors in myeloid development: balancing differentiation with transformation. Nat Rev Immunol 2007;7(2):105-17.

2. Rice KL, Hormaeche I, Licht JD. Epigenetic regulation of normal and malignant hematopoiesis. Oncogene 2007;26(47):6697-714.

3. Suzuki M, Yamada T, Kihara-Negishi F, Sakurai T, Hara E, Tenen DG, et al. Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. Oncogene 2006;25(17):2477-88.

4. Pabst T, Mueller BU. Transcriptional dysregulation during myeloid transformation in AML. Oncogene 2007;26(47):6829-37.

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

6. Frohling S, Schlenk RF, Stolze I, Bihlmayr J, Benner A, Kreitmeier 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 2004;22(4):624-33.

7. Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, et al. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. Nat Genet 2001;27(3):263-70.

8. Preudhomme C, Sagot C, Boissel N, Cayuela JM, Tigaud I, de Botton S, 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 2002;100(8):2717-23.

9. Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, Meijer J, van Oosterhoud S, van Putten WL, Valk PJ, et al. Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML. Hematol J 2003;4(1):31-40.

10. Pabst T, Mueller BU, Harakawa N, Schoch C, Haferlach T, Behre G, et al. AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. Nat Med 2001;7(4):444-51.

11. Behre G, Singh SM, Liu H, Bortolin LT, Christopeit M, Radomska HS, et al. Ras signaling enhances the activity of C/EBP alpha to induce granulocytic differentiation by phosphorylation of serine 248. J Biol Chem 2002;277(29):26293-9.

12. Radomska HS, Basseres DS, Zheng R, Zhang P, Dayaram T, Yamamoto Y, et al. Block of C/EBP alpha function by phosphorylation in acute myeloid leukemia with FLT3 activating mutations. J Exp Med 2006;203(2):371-81.

13. Hackanson B, Bennett KL, Brena RM, Jiang J, Claus R, Chen SS, et al. Epigenetic modification of CCAAT/enhancer binding protein alpha expression in acute myeloid leukemia. Cancer Res 2008;68(9):3142-51.

14. Tada Y, Brena RM, Hackanson B, Morrison C, Otterson GA, C. P. Epigenetic modulation of tumor suppressor CCAAT/enhancer binding protein alpha activity in lung cancer. J Natl Cancer Inst 2006;98(6):396-406.

15. Bennett KL, Hackanson B, Smith LT, Morrison CD, Lang JC, Schuller DE, et al. Tumor suppressor activity of CCAAT/enhancer binding protein alpha is epigenetically down-regulated in head and neck squamous cell carcinoma. Cancer Res 2007;67(10):4657-64.

16. Wouters BJ, Jorda MA, Keeshan K, Louwers I, Erpelinck-Verschueren CA, Tielemans D, et al. Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. Blood 2007;110(10):3706-14.

17. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

18. Figueroa ME, Reimers M, Thompson RF, Ye K, Li Y, Selzer RR, et al. An integrative genomic and epigenomic approach for the study of transcriptional regulation. PLoS ONE 2008;3(3):e1882.

19. Khulan B, Thompson R, Ye K, Fazzari Mj, Suzuki M, Stasiek E, et al. Comparative isoschizomer profiling of cytosine methylation: the HELP assay. Genome Research 2006;16(8):1046-55.

20. Selzer RR, Richmond TA, Pofahl NJ, Green RD, Eis PS, Nair P, et al. Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. Genes Chromosomes Cancer 2005;44(3):305-19.

21. Thompson RF, Reimers M, Khulan B, Gissot M, Richmond TA, Chen Q, et al. An analytical pipeline for genomic representations used for cytosine methylation studies. Bioinformatics 2008;24(9):1161-7.

22. Culhane AC, Thioulouse J, Perriere G, Higgins DG. MADE4: an R package for multivariate analysis of gene expression data. Bioinformatics 2005;21(11):2789-90.

23. Palomero T, Lim WK, Odom DT, Sulis ML, Real PJ, Margolin A, et al. NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. Proc Natl Acad Sci U S A 2006;103(48):18261-6.

24. Stock W, Tsai T, Golden C, Rankin C, Sher D, Slovak ML, et al. Cell cycle regulatory gene abnormalities are important determinants of leukemogenesis and disease biology in adult acute lymphoblastic leukemia. Blood 2000;95(7):2364-71.

25. Konieczna I, Horvath E, Wang H, Lindsey S, Saberwal G, Bei L, et al. Constitutive activation of SHP2 in mice cooperates with ICSBP deficiency to accelerate progression to acute myeloid leukemia. J Clin Invest 2008;118(3):853-67.

26. Turcotte K, Gauthier S, Tuite A, Mullick A, Malo D, Gros P. A mutation in the Icsbp1 gene causes susceptibility to infection and a chronic myeloid leukemia-like syndrome in BXH-2 mice. J Exp Med 2005;201(6):881-90.

27. Dahl E, Wiesmann F, Woenckhaus M, Stoehr R, Wild PJ, Veeck J, et al. Frequent loss of SFRP1 expression in multiple human solid tumours: association with aberrant promoter methylation in renal cell carcinoma. Oncogene 2007;26(38):5680-91.

28. Huang J, Zhang YL, Teng XM, Lin Y, Zheng DL, Yang PY, et al. Down-regulation of SFRP1 as a putative tumor suppressor gene can contribute to human hepatocellular carcinoma. BMC Cancer 2007;7:126.

29. Toyota M, Kopecky KJ, Toyota MO, Jair KW, Willman CL, Issa JP. Methylation profiling in acute myeloid leukemia. Blood 2001;97(9):2823-9.

30. Cameron EE, Baylin SB, Herman JG. p15(INK4B) CpG island methylation in primary acute leukemia is heterogeneous and suggests density as a critical factor for transcriptional silencing. Blood 1999;94(7):2445-51.

31. Tessema M, Langer F, Dingemann J, Ganser A, Kreipe H, Lehmann U. Aberrant methylation and impaired expression of the p15(INK4b) cell cycle regulatory gene in chronic myelomonocytic leukemia (CMML). Leukemia 2003;17(5):910-8.

32. Liu TH, Raval A, Chen SS, Matkovic JJ, Byrd JC, Plass C. CpG island methylation and expression of the secreted frizzled-related protein gene family in chronic lymphocytic leukemia. Cancer Res 2006;66(2):653-8.

33. Roman-Gomez J, Cordeu L, Agirre X, Jimenez-Velasco A, San Jose-Eneriz E, Garate L, et al. Epigenetic regulation of Wnt-signaling pathway in acute lymphoblastic leukemia. Blood 2007;109(8):3462-9.

34. Shimamoto T, Ohyashiki JH, Ohyashiki K. Methylation of p15(INK4b) and E-cadherin genes is independently correlated with poor prognosis in acute myeloid leukemia. Leuk Res 2005;29(6):653-9.

35. Glass J, Thompson R, Khulan B, Figueroa M, Olivier E, Oakley E, et al. CG dinucleotide clustering is a species-specific property of the genome. Nucleic Acids Res 2007.

36. Elemento O, Slonim N, Tavazoie S. A universal framework for regulatory element discovery across all genomes and data types. Mol Cell 2007;28(2):337-50.

37. Zhang P, Iwasaki-Arai J, Iwasaki H, Fenyus ML, Dayaram T, Owens BM, et al. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. Immunity 2004;21(6):853-63.

38. Agger K, Cloos PA, Christensen J, Pasini D, Rose S, Rappsilber J, et al. UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature 2007;449(7163):731-4.

39. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, et al. Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 2002;298(5595):1039-43.

40. Di Croce L, Raker VA, Corsaro M, Fazi F, Fanelli M, Faretta M, et al. Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. Science 2002;295(5557):1079-82.

41. Hinshelwood RA, Huschtscha LI, Melki J, Stirzaker C, Reddel RR, Clark SJ. The relationship between nucleosome positioning and aberrant p16-INK4A DNA methylation. Proceedings of the AACR special conference in Cancer Research: Cancer Epigenetics 2008:PR9.

42. Feltus FA, Lee EK, Costello JF, Plass C, Vertino PM. DNA motifs associated with aberrant CpG island methylation. Genomics 2006;87(5):572-9.

CHAPTER 7

A recurrent in-frame insertion in a *CEBPA* transactivation domain is a polymorphism rather than a mutation that does not affect gene expression profiling-based clustering of AML

Bas J. Wouters[1], Irene Louwers[1], Peter J.M. Valk[1], Bob Löwenberg[1] and Ruud Delwel[1]

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

Mutations in *CEBPA*, the gene encoding the transcription factor CCAAT/enhancer binding protein alpha (C/EBPα), have been reported in multiple studies, and are found in approximately 8% of patients with acute myeloid leukemia (AML) (1,2). Specific regions of the gene tend to be most commonly mutated, i.e. 1) in-frame insertions in the basic/leucine zipper (bZIP) region and 2) truncating out-of-frame insertions or deletions in the N-terminus (1,2). Although mutations are most frequently found in these two regions, other abnormalities have been described as well (2). Fröhling et al reported in 6 out of 236 AML cases the existence of an in-frame insertion mutation of 6 nucleotides (3). This insertion is predicted to result in a histidine-proline duplication (HP196-197ins) in a transactivation domain of C/EBPα. *In vitro* studies have suggested that this proline-histidine-rich region may play a role in anti-proliferative control, although this notion has not been supported by *in vivo* experiments (4,5). Remarkably, in none of the other initial *CEBPA* mutation studies, the insertion was
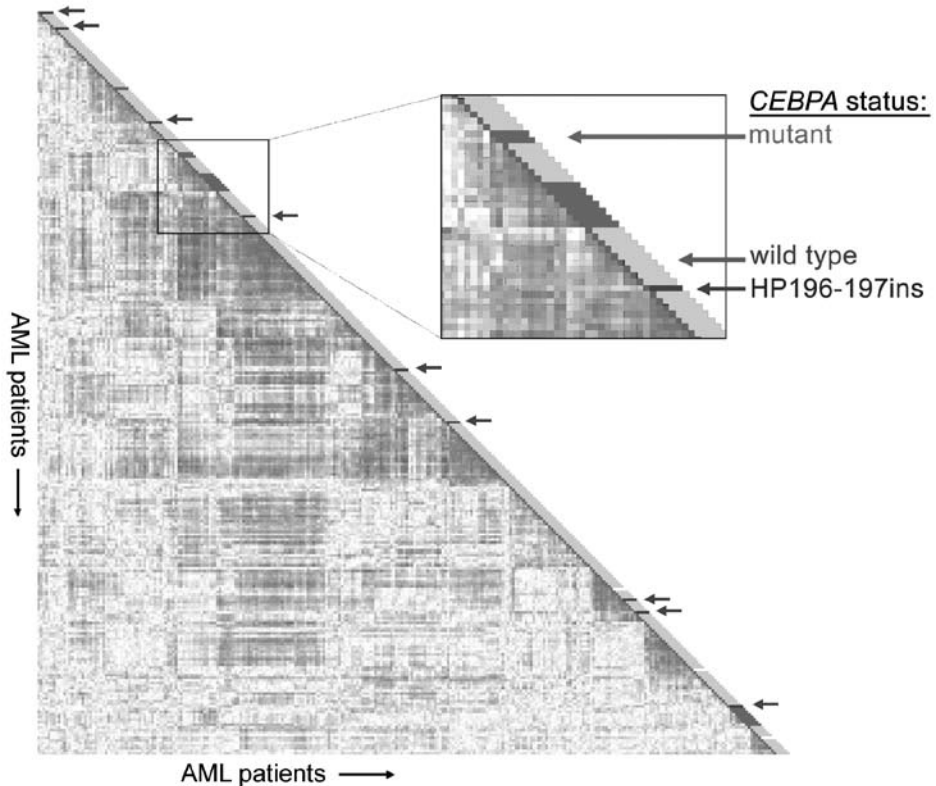
Figure 1. Correlation view of 285 AML cases including *CEBPA* status.
Pairwise correlations between samples are displayed using 2856 probe sets as described.(8) Colors of boxes visualize Pearson's correlation coefficient: red indicates higher positive correlation, blue indicates higher negative correlation. Bars next to each sample represent *CEBPA* status: mutation in bZIP region and/or N-terminus (red), presence of HP196-197ins (blue), or neither (green). For three specimens, depicted in white, no material was available for dHPLC analysis. Figure generated using HeatMapper software (10).
(A full color version of this figure can be found in the color section.)

reported as either a mutation or polymorphism, notwithstanding the fact that investigators frequently applied SSCP or nucleotide sequenced the complete *CEBPA* cDNA (see Leroy et al (2) for references). More recently, one other group described the HP-duplication in 20 out of 100 AML samples (20%) (6). In this study, the insertion was reported in 7/19 healthy volunteers (39%) as well, questioning its role in AML (6).

In a cohort of 285 AML cases, we previously selectively screened for the two major mutation types and identified 17 patients with mutations (7,8). Here, we asked whether this cohort also included cases with HP196-197ins. By means of a dHPLC (denaturing high-performance liquid chromatography) approach (9) and subsequent nucleotide sequencing, we identified the heterozygous HP196-197ins in 9 patients (3.2% of 282 available samples). We also screened an independent second cohort of 305 AML cases, and again found 12 cases (3.9%) to present with this duplication. Finally, we analyzed a series of 274 non-leukemic blood samples and found 22 individuals (8.0%) to carry the same insertion.

Cases with *CEBPA* mutations were predominantly found in two distinct gene expression clusters (8) (Figure 1). We asked whether cases with HP196-197ins associated with specific gene expression clusters as well. The 9 specimens carrying HP196-197ins in the first cohort of 285 AML cases did not cluster with *CEBPA* mutant cases. Moreover, they did not belong to one single previously defined cluster of AML, but were spread out over several subgroups instead (Figure 1).

We conclude that HP196-197ins represents a common *CEBPA* polymorphism, rather than a mutation, that does not influence gene expression profiling-based clustering of AML specimens. Whether the higher percentage of HP196-197ins observed in non-leukemic samples compared to AML cases is due to chance, or represents an important difference, remains to be elucidated in larger series.

REFERENCES

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

2. Leroy H, Roumier C, Huyghe P, Biggio V, Fenaux P, Preudhomme C. CEBPA point mutations in hematological malignancies. Leukemia 2005;19(3):329-34.

3. Frohling S, Schlenk RF, Stolze I, Bihlmayr J, Benner A, Kreitmeier 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 2004;22(4):624-33.

4. Wang H, Iakova P, Wilde M, Welm A, Goode T, Roesler WJ, et al. C/EBPalpha arrests cell proliferation through direct inhibition of Cdk2 and Cdk4. Mol Cell 2001;8(4):817-28.

5. Porse BT, Pedersen TA, Hasemann MS, Schuster MB, Kirstetter P, Luedde T, et al. The proline-histidine-rich CDK2/CDK4 interaction region of C/EBPalpha is dispensable for C/EBPalpha-mediated growth regulation in vivo. Mol Cell Biol 2006;26(3):1028-37.

6. Lin LI, Chen CY, Lin DT, Tsay W, Tang JL, Yeh YC, et al. Characterization of CEBPA mutations in acute myeloid leukemia: most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells. Clin Cancer Res 2005;11(4):1372-9.

7. Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, Meijer J, van Oosterhoud S, van Putten WL, Valk PJ, et al. Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML. Hematol J 2003;4(1):31-40.

8. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

9. Verhaak RG, Goudswaard CS, van Putten W, Bijl MA, Sanders MA, Hugens W, 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 2005;106(12):3747-54.

10. Verhaak RG, Sanders MA, Bijl MA, Delwel R, Horsman S, Moorhouse M, et al. HeatMapper: Powerful combined visualization of gene expression profile correlations, genotypes, phenotypes and sample characteristics. BMC Bioinformatics 2006;7(1):337.

CHAPTER 8

Segmental uniparental disomy as a recurrent mechanism for homozygous *CEBPA* mutations in acute myeloid leukemia

Bas J. Wouters[1], Mathijs A. Sanders[1], Sanne Lugthart[1], Wendy M. C. Geertsma-Kleinekoort[1], Ellen van Drunen[2], H. Berna Beverloo[2], Bob Löwenberg[1], Peter J. M. Valk[1] and Ruud Delwel[1]

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

[2] Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands

The basic leucine zipper (bZIP) transcription factor CCAAT/enhancer-binding protein-α (C/EBPα), encoded by the *CEBPA* gene, is essential for myeloid development (1). In acute myeloid leukemia (AML), interference with C/EBPα function may occur through several mechanisms (2). Mutations in the *CEBPA* coding region itself are observed in approximately 8% of AML. These mutations either involve N-terminal frameshift abnormalities, leading to increased translation of a 30 kDa isoform with dominant-negative properties, or in-frame insertions or deletions disrupting the bZIP domain located in the C-terminus (2). In many cases of AML, both types of mutations are found in the same leukemic cell, frequently involving both alleles, that is N-terminal mutation of one allele and C-terminal disruption of the other. In a previous gene expression profiling (GEP) study of 285 cases of de novo AML, two expression clusters were highly associated with *CEBPA* mutations (3). The majority of *CEBPA* mutant tumors in these clusters carried both an N-terminal and a bZIP mutation. Interestingly, in three cases only one of the two types of mutations was present, while a wild-type allele appeared to be absent. Here, we combined fluorescence in situ hybridization (FISH) experiments with single nucleotide polymorphism (SNP) array hybridization to investigate two potential mechanistic explanations for the existence of homozygous *CEBPA* mutations in these particular cases: deletion of the wild-type allele, or uniparental disomy (UPD) as a result of mitotic recombination (4).

CEBPA mutational analysis by PCR and nucleotide sequencing has been described (3). These analyses revealed 17 mutant specimens, of which 3 carried homozygous mutations, in the cohort of 285 cases of AML (3). Two leukemias harbored homozygous mutations in the *CEBPA* bZIP region, whereas in the other case the mutation was found in the N-terminus (Table 1).

To investigate whether homozygous *CEBPA* mutations correlated with loss of the wild-type allele through deletion, we carried out dual color FISH using three BAC clones RP11270I13, RP11-547I3 and RP11-1150B17; Supplementary Figure 1) located on 19q13.11 as probes. In all three specimens with homozygous mutations, signals for both homologues were detected in more than 95% of cells (data not shown), excluding deletion as the underlying mechanism for homozygosity.

Table 1. Characteristics of AML cases with homozygous *CEBPA* mutations

Patient #	FAB subtype	Karyotype	Homozygous *CEBPA* mutation
1	M2	46, XY[35]	361–262ins2bp
2	M1	46, XY[67]	33-bp duplication of bp 1062–1094
3	M1	46, XX[27]	1104–1115del

Abbreviations: AML, acute myeloid leukemia.
FAB: morphologic classification according to the French–American–British system.
Nucleotide numbering according to Entrez accession number XM_009180. Patient #1 carries a mutation in the region of *CEBPA* encoding the N-terminus. Patients #2 and #3 carry mutations in the basic leucine zipper (bZIP) region.

We next used Affymetrix (Santa Clara, CA, USA) Mapping 250K NspI SNP arrays to assess the CEBPA locus in more detail. We investigated 14/17 AML cases with *CEBPA* mutations, including the 3 AMLs with homozygous CEBPA abnormalities.

Data were imported into SNPExpress, a software tool developed in our department (5). Mapping of probe sets to genes was done according to data provided by Affymetrix. On the arrays used, none of the probe sets mapped directly to the *CEBPA* gene. Probe set SNP_A-2236362, associated with neighboring gene *CEBPG* (Supplementary Figure 1), was therefore used as a marker for the *CEBPA* locus. We first determined copy numbers of chromosome 19q13.11 for each patient based on these SNP arrays using dChipSNP. This analysis confirmed the results obtained by FISH, as the calculated copy number of the region containing the *CEBPA* gene fluctuated closely around 2 in all 14 leukemias (Figure 1). We subsequently assessed potential regions of loss of heterozygosity (LOH) using a hidden Markov model. Strikingly, we found extensive regions of LOH of chromosome 19, including the *CEBPA* locus, in the three samples with homozygous *CEBPA* aberrations. In all three leukemias, LOH included the majority of the 19q arm, and stretched until the telomeres (Figure 1). In contrast, no LOH of the *CEBPA* locus was apparent in the 11 samples carrying heterozygous *CEBPA* mutations. These results demonstrate that copy number neutral LOH of 19q13.11 is associated with homozygous *CEBPA* mutations, which is indicative of UPD through mitotic recombination in this region.

UPD has recently emerged as a mechanism involved in malignancy, including AML (6-8). Genome-wide studies using high-resolution SNP arrays have indicated that approximately 20% of AMLs with a normal karyotype show either interstitial or terminal UPD (4,6-9). Acquired UPD is associated with removal of wild-type alleles of genes commonly mutated in AML, including *RUNX1*, *FLT3* and *WT1* (6). Previously, a single case of AML with a homozygous *CEBPA* bZIP mutation and UPD has been reported (6). Our results imply that terminal UPD involving mutant *CEBPA* alleles is a recurrent mechanism.

In a previous unsupervised GEP study, the three AMLs described here clustered together with other *CEBPA* mutant AML cases in two distinct expression clusters (3). The common finding in these clusters was that the *CEBPA* mutant AMLs carried mutations in both alleles and consequently did not express wild-type *CEBPA*. Although the functional implication of replacement of wild-type *CEBPA* remains elusive, its absence, resulting either from biallelic mutations or from homozygous mutations, seems to be critical for leukemic development of those cases. It is currently unclear whether homozygous *CEBPA* mutations have a specific differential significance as opposed to biallelic heterozygous aberrations. It will be interesting to investigate whether loss of wild-type alleles of additional genes on chromosome 19 plays a role in leukemogenesis in these cases.

Another interesting question is whether the initially mutated allele in the AMLs with homozygous *CEBPA* mutations was acquired, or was already present in the germline. This would particularly be of interest because familial N-terminal mutations have been described.

Figure 1. Copy number neutral loss of heterozygosity (LOH) in three cases of acute myeloid leukemia (AML) with homozygous *CEBPA* mutations.

Affymetrix Mapping 250K NspI arrays were used to investigate chromosomal copy numbers and genotypes of AML cases with *CEBPA* mutations. Six cases are shown in this figure: patients #1, #2 and #3 harbor homozygous *CEBPA* mutations, and cases #4, #5 and #6 carry heterozygous *CEBPA* mutations. A region on chromosome 19q13.11, including the *CEBPA* locus (*CEBPA*), is depicted at larger magnification. For each individual single nucleotide polymorphism (SNP) in this region, copy number was calculated, and is indicated as deflection from the green midline, which represents the presence of two chromosomes (deflection to the left indicates lower copy number, deflection to the right indicates higher copy number). Genotypes for each SNP are color-coded: homozygous AA (red) or BB (yellow), and heterozygous AB (blue). A hidden Markov model was employed to assess potential LOH, and revealed LOH of the region shown in patients #1, #2 and #3, illustrated by the fact that the large majority of calls is homozygous. The infrequent remaining heterozygous (blue) signals are considered false calls according to this model. The extents of the regions of LOH are indicated as red bars next to the schematic representation of chromosome 19 ("region of LOH") for each of these three cases.

(A full color version of this figure can be found in the color section.)

Unfortunately, we were not able to address this issue, as normal or remission control material was not available for any of the three patients.

In conclusion, we describe the association of homozygous *CEBPA* mutations in three cases of AML with mitotic recombination involving the *CEBPA* locus. Our findings are consistent with recent reports suggesting the frequent occurrence of UPD in AML, and imply that this mechanism is responsible for a substantial proportion of cases with homozygous *CEBPA* aberrations.

ACKNOWLEDGEMENTS

REFERENCES

1. Zhang DE, Zhang P, Wang ND, Hetherington CJ, Darlington GJ, Tenen DG. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc Natl Acad Sci U S A 1997;94(2):569-74.

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

3. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

4. Young BD, Debernardi S, Lillington DM, Skoulakis S, Chaplin T, Foot NJ, et al. A role for mitotic recombination in leukemogenesis. Adv Enzyme Regul 2006;46:90-7.

5. Sanders MA, Verhaak RG, Geertsma-Kleinekoort WM, Abbas S, Horsman S, van der Spek PJ, et al. SNPExpress: integrated visualization of genome-wide genotypes, copy numbers and gene expression levels. BMC Genomics 2008;9:41.

6. Fitzgibbon J, Smith LL, Raghavan M, Smith ML, Debernardi S, Skoulakis S, et al. Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. Cancer Res 2005;65(20):9152-4.

7. Gorletta TA, Gasparini P, D'Elios MM, Trubia M, Pelicci PG, Di Fiore PP. Frequent loss of heterozygosity without loss of genetic material in acute myeloid leukemia with a normal karyotype. Genes Chromosomes Cancer 2005;44(3):334-7.

8. Raghavan M, Lillington DM, Skoulakis S, Debernardi S, Chaplin T, Foot NJ, et al. Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. Cancer Res 2005;65(2):375-8.

9. Stephens K, Weaver M, Leppig KA, Maruyama K, Emanuel PD, Le Beau MM, et al. Interstitial uniparental isodisomy at clustered breakpoint intervals is a frequent mechanism of NF1 inactivation in myeloid malignancies. Blood 2006;108(5):1684-9.

CHAPTER

9

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

Bas J. Wouters[1], Bob Löwenberg[1], Claudia A.J. Erpelinck-Verschueren[1], Wim L.J. van Putten[2], Peter J.M. Valk[1] and Ruud Delwel[1]

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

ABSTRACT

Mutations in CCAAT/enhancer binding protein alpha (*CEBPA*) are seen in 5-14% of acute myeloid leukemia (AML) and have been associated with a favorable clinical outcome. Most AMLs with *CEBPA* mutations simultaneously carry two mutations (*CEBPA*^double-mut), usually biallelic, while single heterozygous mutations (*CEBPA*^single-mut) are less frequently seen. Using denaturing high performance liquid chromatography and nucleotide sequencing we identified among a cohort of 598 newly diagnosed AMLs a subset of 41 *CEBPA* mutant cases, i.e. 28 *CEBPA*^double-mut and 13 *CEBPA*^single-mut cases. *CEBPA*^double-mut associated with a unique gene expression profile as well as favorable overall and event-free survival, retained in multivariable analysis that included cytogenetic risk, *FLT3*-ITD and *NPM1* mutation, white blood cell count and age. In contrast, *CEBPA*^single-mut AMLs did not express a discriminating signature and could not be distinguished from wild type cases as regards clinical outcome. These results demonstrate significant underlying heterogeneity within *CEBPA* mutation positive AML with prognostic relevance.

INTRODUCTION

Mutations in the transcription factor CCAAT/enhancer binding protein alpha (*CEBPA*) are found in 5-14% of acute myeloid leukemia (AML) (1-9). *CEBPA* mutations have been associated with a relatively favorable outcome, and have therefore gained interest as a prognostic marker (4-6,10). While variable sequence variations have been described, two prototypical classes of mutations are most frequent. N-terminal mutations are located between the major translational start codon and a second ATG in the same open reading frame. These mutations introduce a premature stop of translation of the p42 C/EBPα protein while preserving translation of a p30 isoform that has been reported to inhibit the function of full length protein (2). Mutations in the C-terminal basic leucine zipper (bZIP) region, in contrast, are in-frame, and may impair DNA binding and/or homo- and heterodimerization (8). The remaining mutations are mostly found between the N-terminus and bZIP region (11).

Most *CEBPA* mutant AMLs exhibit two mutations, which most frequently involves a combination of an N-terminal and a bZIP gene mutation (7,8,11,12). In AMLs with two *CEBPA* mutations, the mutations are typically on different alleles (11). Hence, in these cases no wild type C/EBPα protein is expressed. A similar condition is found in AMLs carrying a homozygous *CEBPA* mutation (13). However, there are also AMLs that only show one single heterozygous mutation, and thus retain expression of a wild type allele (7,11,12).

To obtain better insight into the distribution of the various types of *CEBPA* mutations in *de novo* adult AML and their impact on clinical outcome, we examined a cohort of 598 cases. Following denaturing high performance liquid chromatography (dHPLC) and nucleotide sequencing, we distinguished cases with two different mutations or one homozygous mutation (further referred to as double mutations; *CEBPA*^double-mut) as well as cases with only one single heterozygous mutation (*CEBPA*^single-mut). Genome-wide gene expression profiling revealed that *CEBPA*^double-mut AMLs expressed a highly characteristic signature, while *CEBPA*^single-mut cases did not. In addition, favorable prognosis appeared uniquely associated with *CEBPA*^double-mut AML.

MATERIALS AND METHODS

Patient characteristics and molecular analyses

Bone marrow aspirates or peripheral blood samples of 598 cases of *de novo* AML were collected, blast cells were purified, and mRNA was isolated as reported (14). Detailed clinical and molecular characteristics were available for 524/598 cases (Table S4). These 524 patients were enrolled in the Dutch-Belgian Hemato-Oncology Cooperative Group (HOVON)–04, -10, -12, -29, -32, -42, or -43 protocols (available at http://www.hovon.nl). Reverse-transcription

polymerase chain reaction (RT-PCR) and sequence analyses for *FLT3*-ITD, *FLT3*-TKD, *NPM1*, N-*RAS*, and K-*RAS*, mutations were performed as described previously (15-17).

Detection of *CEBPA* mutations

Complementary DNA (cDNA) was generated from 1 μg of mRNA using SuperScript reverse transcriptase (Invitrogen, Carlsbad, CA, USA). The *CEBPA* coding region was divided into three overlapping amplicons (Figure 1A). Primers for the three fragments (A, B and C) are indicated below.

Primer name	Location relative to XM_009180.3 (relative to main translational start site)	Sequence (5' to 3')
A fw	122 – 141 (-28 – -9)	CGCCATGCCGGGAGAACTCT
A rev	420 – 400 (270 – 250)	CTTCTCCTGCTGCCGGCTGT
B fw	361 – 380 (211 – 230)	GCCGCCTTCAACGACGAGTT
B rev	663 – 644 (513 – 494)	CTTGGCTTCATCCTCCTCGC
C fw	616 – 633 (466 – 483)	CGGCCGCTGGTGATCAAG
C rev	1254 – 1236 (1104 – 1086)	CCCAGGGCGGTCCCACAGC

PCR amplification for all three fragments was carried out using 2 μl of cDNA in mixes containing 0.5 mM dNTPs, 10% DMSO, 2 mM $MgCl_2$, 0.4 μM of forward and reverse primer, 1X PCR buffer and 2.5 units of Taq polymerase (Invitrogen), in a total volume of 50 μl. Thermal cycling conditions for the three reactions were equal, i.e. denaturation at 94°C for 5 minutes, followed by 35 cycles of 94°C for 1 minute, 56°C for 1 minute and 72°C for 1 minute, and a final 5-minute elongation step of 72°C. After PCR amplification, 10 μl of PCR product was mixed with 10 μl of corresponding PCR product obtained from NB4 cell line cDNA. Heteroduplexes were allowed to form in an Applied Biosystems (Foster City, CA, USA) GeneAmp PCR System 9700 (2 cycles of 95°C for 3 minutes, cooled to 20°C with a ramp of 5%, and maintained at 20°C for 5 minutes). The samples were then subjected to denaturing high-performance liquid chromatography (dHPLC) analysis on a Transgenomics (Omaha, NE, USA) WAVE device, using temperatures of 65.4°C, 66.4°C and 65.5°C, respectively. Data were analyzed using Transgenomics software, and aberrant peaks were independently scored by two investigators. Samples with aberrant peaks were subjected to direct nucleotide sequencing on an Applied Biosystems 3100 device using the forward and reverse primers. In case a mutation was found, a second analysis on new input material was performed to rule out PCR-induced artifacts.

In AML cases for which dHPLC had revealed one single heterozygous mutation, the *CEBPA* coding region was fully sequenced to exclude the possibility that a second mutation had gone unnoticed. In three cases with an N-terminal mutation (#4336, #5362 and #5364), this extra analysis revealed an additional mutation, two of which were point mutations in the bZIP region.

Cases that appeared negative by dHPLC were additionally screened as follows. The *CEBPA* N-terminal part was nucleotide sequenced using previously described primers 2 and 10 (6). Insertions or deletions in the basic leucine zipper domain were detected using a previously described ethidium bromide agarose gel electrophoresis approach and subsequent nucleotide sequencing (primers 4 and 8) in cases with apparent abnormalities (6).

Statistical analysis

Statistical analyses were performed in Statistical Package for the Social Sciences (SPSS, Chicago, IL, USA) software, version 16.0. All patients received induction therapy and were included in the survival analysis. Actuarial probabilities of overall survival (OS, with death due to any cause) and event-free survival (EFS, with failure in case of no complete remission at day 1 [CR1] or relapse or death) were estimated by the method of Kaplan and Meier, and significance was assessed with the log rank test. Detailed definitions of all clinical endpoints for the various HOVON studies may be found in the study protocols which are accessible through http://www.hovon.nl. In brief, CR was defined as a bone marrow blast percentage less than 5% with peripheral blood recovery, without signs of extramedullary disease. Relapse was defined as recurrence of more than 5% blasts in the bone marrow (excluding increased blasts in the context of regenerating marrow) and recurrence of leukemia in peripheral blood/ at extramedullary sites. Cox's proportional hazards models were fitted for multivariable analysis. The choice of variables to include was based on their a priori presumed prognostic impact. No further forward or backward selection procedure was applied, so all variables were included in the final models. The proportional hazard assumption for *CEBPA* status was visually tested by plotting –ln(-ln(survival probability)) adjusted for the other covariates versus ln(analysis time). The three *CEBPA* curves were roughly parallel and there was no indication of non-proportionality.

Cytogenetic risk groups (favorable, intermediate, or poor) were defined as described (15). Briefly, patients with inv(16)/t(16;16), t(8;21), and t(15;17) abnormalities, irrespective of the presence of additional cytogenetic aberrations, were considered as being in the favorable-risk category. These included a small number of cases in which the abnormality had been identified by RQ-PCR, despite normal cytogenetics. The poor-risk category was defined by the presence of –5/del(5q), –7del(7q), t(6;9), t(9;22), 3q26 abnormality, or complex karyotype (more than 3 abnormalities) in the absence of good risk cytogenetic characteristics. All other patients were classified as intermediate risk. All tests were 2 tailed, and a *P* value of less than 0.05 was considered statistically significant.

Gene expression profiling analysis

Gene expression profiles of 524 cases of AML were derived using Affymetrix (Santa Clara, CA, USA) HGU133Plus2.0 GeneChips. Sample processing and quality control were carried out as described previously (14). Raw microarray data were processed using Affymetrix Microarray
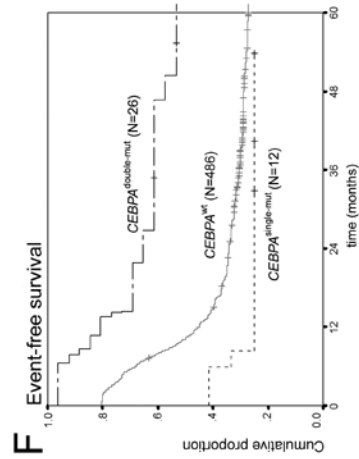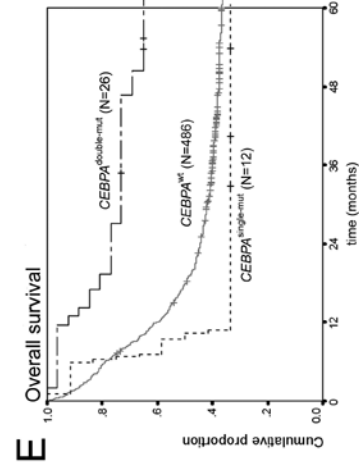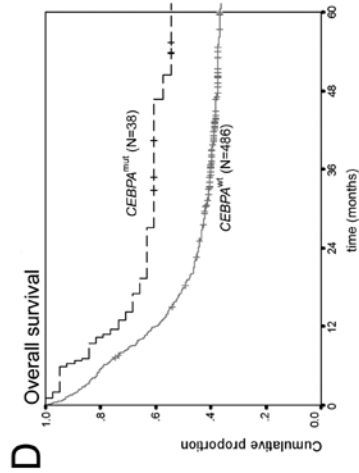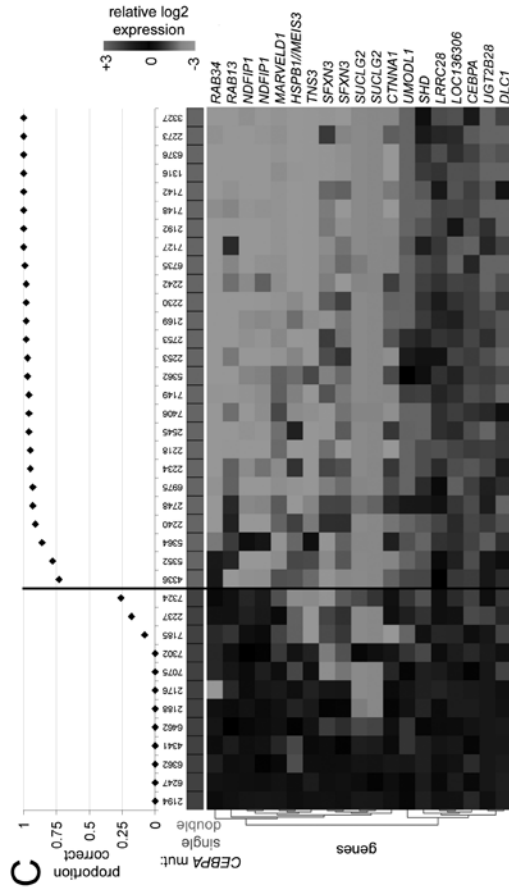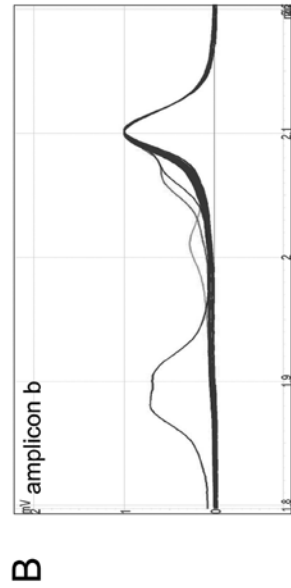
Figure 1. Schematic overview of dHPLC analysis, gene expression profiling analysis and survival estimates.
A. Schematic representation of the *CEBPA* gene and location of amplicons a, b and c for PCR, used for dHPLC analysis. Functional regions are depicted, i.e. two transactivation domains (TAD1 and TAD2) in the N-terminal part, and the basic leucine zipper (bZIP) region in the C-terminal part. Nucleotide (nt) position is indicated relative to the main translation start site. Amino acid (aa) numbering and the alternative translation start site at position nt 358 (aa 120) are also depicted.
B. Representative profiles of dHPLC analysis of one of the three investigated fragments, i.e. amplicons b, in a random selection of 90 samples. Heteroduplexes (various colors) are released earlier than homoduplexes (green), and can therefore be recognized as distinct peaks. Time is depicted on the x-axis, and absorbance on the y-axis.
C. A gene expression prediction signature for *CEBPA*[mut] AML (irrespective of single or double mutant status) was derived in a data set of 524 AMLs, including 38 *CEBPA*[mut] cases. Prediction accuracy for each of the 38 *CEBPA*[mut] cases was estimated using repeated 10-fold cross-validation as detailed in Supplementary Materials and Methods. The proportion of correct predictions for the selected 38 *CEBPA*[mut] specimens is indicated (upper panel). Mutation status is color coded, i.e *CEBPA*[single-mut] (blue) or *CEBPA*[double-mut] (red). The heat map in the lower panel depicts the 19 probe sets in the resulting *CEBPA*[mut] gene expression classifier (see Table S2 for probe set information). Intensity values (log2) were mean centered over the cohort of 524 AML cases and for visualization purposes the genes were hierarchically clustered (Euclidian distance, average linkage). Cells represent relative log2 expression values, and have been color coded on a scale ranging from bright green (-3) to bright red (+3), with black indicating no change relative to the mean.
D. Kaplan Meier estimates of overall survival (OS) among *CEBPA*[mut] and *CEBPA*[wt] AML, log rank test P=0.027.
E. OS among *CEBPA*[double-mut] versus *CEBPA*[wt] AML, P=0.004, and versus *CEBPA*[single-mut] AML, P=0.005; pooled P=0.012 F. Event-free survival (EFS) among *CEBPA*[double-mut] and *CEBPA*[wt] AML, P=0.005, and versus *CEBPA*[single-mut] AML, P=0.004; pooled P=0.008. The cumulative proportion of survival at the intercept, i.e. the point where a line crosses the Y-axis, reflects the proportion of patients reaching complete remission. Analyses similar to those depicted in panels D-F were performed after splitting the group of *CEBPA*[wt] AMLs into those with favorable cytogenetics and those with other cytogenetics. These additional analyses can be found in Figure S4.
(A full color version of this figure can be found in the color section.)

Suite 5 (MAS5) to target intensity values of 100. Intensity values lower than 30 were set at 30, and subsequently all data were log2 transformed. Gene expression data are available at the NCBI Gene Expression Omnibus (accession number GSE14468). Gene expression classifiers for *CEBPA*[mut] and *CEBPA*[double-mut] were derived using Prediction Analysis for Microarrays (PAM)(18) version 1.28 in R version 2.1.0. The method of the nearest shrunken centroids identifies a subgroup of genes that best characterizes a predefined class. In accordance with good practice guidelines (19,20), all available data were used for classifier construction, and predictive performances were estimated based on cross-validation as follows. PAM was first used to train a classifier based on the entire data set of 524 AML cases. Next, selection of a shrinkage factor (in order to only use the most informative genes) as well as estimation of classifier performance were carried out using 10-fold cross-validation, involving a random split of the data into 10 folds which was balanced with respect to mutation status. Each fold was once used as an independent validation set for a classifier that has been trained on the remaining 9 folds. The minimum number of misclassified cases was subsequently deter-mined, and the corresponding shrinkage threshold was recorded. Furthermore, sensitivity and specificity were calculated. This entire procedure of 10-fold random cross-validation was repeated 100 times. Reported final classifiers represent the probe sets that remained after shrinkage using the median threshold over the 100 rounds of cross-validation. Reported final sensitivities and specificities represent the averages over the 100 rounds of cross-validation.

Criterion for the *CEBPA*^mut classifier was minimum total misclassification rate (i.e. minimum false positives + false negatives). Criterion for the reported *CEBPA*^double-mut classifier was minimum misclassification of double mutant specimens (i.e. minimum false negatives). We also assessed the possibility to derive a *CEBPA*^mut classifier that minimized the number of false negatives. However, it was not possible to find such a classifier that correctly predicted all mutant specimens. Furthermore, at the minimum misclassification rate (average of 6/38 misclassified; average sensitivity 84%), an unacceptably high number of false positives was found (i.e. an average of 85 cases; average specificity 83%).

Principal component analysis was performed using Spotfire Decision Site (Spotfire, Inc., Somerville, MA, USA). Before the analysis, data for all probe sets were mean-centered.

RESULTS AND DISCUSSION

In a cohort of 598 cases of adult de novo AML we identified 65 cases with an aberrant profile in at least one of the three investigated amplicons of the *CEBPA* coding sequence (Figure 1A-B). The presence of a *CEBPA* sequence variation was confirmed by nucleotide sequencing. Cases that only carried an insertion polymorphism (21,22) or variation(s) that did not lead to amino acid changes were considered wild type. Two additional specimens were not considered in further analysis because they carried in-frame variations of unknown significance in the N-terminus (Table 1). As a result, 41/598 unambiguous *CEBPA*^mut AML cases (6.9%) were considered. These included 13 *CEBPA*^single-mut cases and 28 *CEBPA*^double-mut cases. Four of the *CEBPA*^double-mut cases carried homozygous mutations whereas the remaining 24 cases showed 2 heterozygous mutations (Table 1). Additional screening of the remaining AML cases using a combination of agarose gel analysis and nucleotide sequencing as described (6) did not reveal mutations that had been missed by dHPLC.

To investigate whether *CEBPA* mutations related to gene expression, we examined genome-wide gene expression data of 524 AML cases, that included 26 *CEBPA*^double-mut and 12 *CEBPA*^single-mut cases. Clinical and molecular characteristics of the AML cases are reported in Table 2. Using Prediction Analysis for Microarrays (PAM) (18) according to a supervised approach, we derived a 19-probe set signature predictive of *CEBPA* mutations (Figure 1C, Table S2). This classifier showed a high specificity (99%), but a limited sensitivity (67%) in cross-validation, indicating a limited ability to recognize all *CEBPA*^mut specimens. Strikingly, misclassification was almost entirely due to *CEBPA*^single-mut cases, whereas *CEBPA*^double-mut AMLs were predicted with an accuracy that was near perfect (Figure 1C, S1). In line with this, we were able to derive a specific 21-probe set classifier for *CEBPA*^double-mut AMLs within the entire AML cohort with a cross-validated sensitivity of 100% (specificity 98%) (Table S3). In further support, unsupervised analysis of the expression data derived from the *CEBPA*^mut subset

indicated an underlying variability in gene expression that correlated with either double or single mutation status (Figure S2).

We next assessed how these differences between *CEBPA*double-mut and *CEBPA*single-mut related to clinical outcome. In line with previous data, overall survival and event-free survival were significantly better for *CEBPA*mut cases compared to cases with wild type *CEBPA* (*CEBPA*wt) (Figure 1D and not shown). Separate analyses for the *CEBPA*double-mut and *CEBPA*single-mut subgroups, however, revealed a favorable outcome that was specific for *CEBPA*double-mut cases. We failed to find a favorable prognostic effect in relation to the *CEBPA*single-mut cases. In fact, *CEBPA*single-mut AMLs showed a significantly worse outcome than *CEBPA*double-mut cases, including a poor rate of complete remission (Figure 1E-F). These findings were also apparent in multivariable analysis (Table 3). When only patients less than 60 years of age or only patients with normal cytogenetics were considered, similar results were found, although in the latter subgroup with smaller numbers only the pair-wise comparison for overall survival between *CEBPA*double-mut and *CEBPA*single-mut reached statistical significance (Figure S3, Table S6).

Based on our previous analyses (6) and based on the literature (11) it is likely that in the vast majority of the *CEBPA*double-mut AML studied, both *CEBPA* alleles were affected. A plausible hypothesis is therefore that absence of wild type *CEBPA* mRNA is directly involved in the *CEBPA*double-mut gene expression profile. This may be further supported by our previous and current observations that indicate a high degree of similarity between the profiles of *CEBPA*double-mut AML and a specific subgroup of leukemias characterized by epigenetic *CEBPA* silencing (Figure S1) (23). It is possible that analysis of larger patient series will lead to further refinement of this subclassification, for instance based on the location of the mutations. For example, our data indicated a tendency of *CEBPA*single-mut cases with mutations in the bZIP region to be potentially less distinct from the *CEBPA*double-mut AMLs (cases #7185, #7324, #2237; Figure 1C, S2). Of note, a subset of the *CEBPA*mut AMLs studied here was included in the cohort of 285 cases of AML that we previously investigated by gene expression profiling (14). In that study, all *CEBPA*double-mut AMLs were found in two particular clusters, while *CEBPA*single-mut AMLs did not specifically aggregate (14,23).

Studies to date have associated *CEBPA* mutations with outcome (4-6,9), but have not applied subdivisions into single and double mutants. It is unclear why *CEBPA*double-mut AMLs would have a better outcome than those with a single heterozygous mutation. One explanation could be that a single mutant *CEBPA* allele is not sufficient for leukemogenesis, and requires cooperating mutations which may be in *CEBPA* itself or in other genes. Of note, recent data indicate that germ line *CEBPA* mutations predispose to AML and the acquisition of a second, somatic *CEBPA* mutation may then contribute to AML development (24). In fact, we found a tendency towards more *FLT3*-ITD, *FLT3*-TKD and *NPM1* mutations in *CEBPA*single-mut compared to *CEBPA*double-mut cases (Table 2). Yet unknown abnormalities may associate with *CEBPA*single-mut AML as well and predispose to inferior outcome. It is however

Table 1. Mutations detected in *CEBPA* coding sequence.

Patient	Category	Nucleotide change	Amino acid change	Comment
1316	double	357del	D69fsX159	N-terminal stop
		1106_1107insACT	S319delinsRL	In-frame insertion in bZIP
2169	double	213del	P22fsX159	N-terminal stop
		1066_1067insGCAACGTGGACAAGCAGC	V308_E309insDKQRNV	In-frame insertion in bZIP
2192	double	ins396GG	L81fsX160	N-terminal stop
		1060_1062dup	K304dup	In-frame duplication in bZIP
2218	double	1062_1094dup (homozygous)	Q305_L315dup	In-frame duplication in bZIP
2230	double	381del	F57fsX159	N-terminal stop
		1076_1078dup	E309_T310insK	In-frame insertion in bZIP
2234	double	392_395dup	F82fsX108	N-terminal stop
		1084_1089dup, 1090G>A	V314delinsQKM	In-frame insertion in bZIP
2240	double	332del	S161fsX159	N-terminal stop
		1076_1078dup	E309_T310insK	In-frame insertion in bZIP
2242	double	252_261del	P34fsX156	N-terminal stop
		1064_1066dup	Q305dup	In-frame duplication in bZIP
2253	double	213del	P22fsX159	N-terminal stop
		1051_1052insGACAAGGCCAAGCAGCGCAACGTGGAGACGCAGCAC	T310_Q311insQHHKAKQRNVET	In-frame insertion in bZIP
2273	double	472_473insT	D107fsX169	N-terminal stop
		1062_1079dup	R306_Q311dup	In-frame duplication in bZIP
2545	double	474C>G	P108X	N-terminal stop

Table 1 continued

Patient	Category	Nucleotide change	Amino acid change	Comment
2748	double	813del	L220fsX317	Frameshift between TAD2 and bZIP, stop in bZIP
		424_425insA	K92fsX107	N-terminal stop
		1057_1058insTTG	K302_A303insV	In-frame insertion in bZIP
2753	double	302_317del	P51fsX154	N-terminal stop
		1085_1087dup	K313dup	In-frame duplication in bZIP
3101*	double	362_363insCC (homozygous)	A71fsX160	N-terminal stop
3117*	double	397del	F82fsX159	N-terminal stop
		1087_1089dup	K313dup	In-frame duplication in bZIP
3327	double	1104_1115del (homozygous)	S319_D322del	In-frame deletion in bZIP
4336	double	218_219insC	P23fsX107	N-terminal stop
		1016G>C	R289P	Substitution in bZIP
5352	double	311del	G53fsX159	N-terminal stop
		679_691dup	P180fsX324	Frameshift between TAD1 and TAD2, stop in bZIP
5362	double	349_359del	A66fsX103	N-terminal stop
		1033G>C	A295P	Substitution in bZIP
5364	double	376_377insG	D75fsX107	N-terminal stop
		486_522dup	A124fsX181	Frameshift, stop in TAD2
6376	double	1091_1092insTGCTGGAGCTGCAGCGCAACGTGGAGACGCAGCAGAAGG (homozygous)	Q305_L1317dup	In-frame duplication in bZIP
6735	double	377_389del	D75fsX155	N-terminal stop
		1084_1086dup	Q312dup	In-frame duplication in bZIP

Table 1 continued

Patient	Category	Nucleotide change	Amino acid change	Comment
6975	double	397C>T	Q83X	N-terminal stop
		1073_1075dup	V308dup	In-frame duplication in bZIP
7127	double	437del	V95fsX159	N-terminal stop
		1087_1089dup	K313dup	In-frame duplication in bZIP
7142	double	354delCG	I68fsX106	N-terminal stop
		1087_1089dup	K313dup	In-frame duplication in bZIP
7148	double	248_249del	G32fsX106	N-terminal stop
		1072_1073insGTGGAGACGCAGCACCTAAAATCG	Q311_Q312insHLKSVETQ	In-frame insertion in bZIP
7149	double	382del	F77fsX159	N-terminal stop
		1087_1089dup	K313dup	In-frame duplication in bZIP
7406	double	406_412del	R86fsX157	N-terminal stop
		1084_1086dup	Q312dup	In-frame duplication in bZIP
2176	single	218_219insC	P23fsX107	N-terminal stop
2188	single	852_858dup	P239fsX322	Frameshift between TAD2 and bZIP, stop in bZIP
2194	single	468_469insAACC	D105fsX108	N-terminal stop
2237	single	1188_1189insAGCGCAACGTGGAGACGCAG CAGAAGGTGCTGGAGCTGACCAGTGACAA TGACCGCCTGCGCAAGCGGGTGGAACAGCT- GAGCCGCGAACTGGACACGCTGCGGGGC ATCTTCCGCCAGCTGCCA	P346fsX359	Frameshift and stop in bZIP
3096*	single	1114_1134del	D322_V328del	In-frame deletion in bZIP

Table 1 continued

Patient	Category	Nucleotide change	Amino acid change	Comment
4341	single	309_311delinsTT	G53fsX159	N-terminal stop
6247	single	564_565insTA	Y138fsX160	N-terminal stop
6362	single	648_649insG	E166fsX169	N-terminal stop
6462	single	445_446insCCAA	T98fsX108	N-terminal stop
7075	single	218_219insC	P22fsX107	N-terminal stop
7185	single	1029_1030insGGACCC	N293delinsKDP	In-frame insertion in bZIP
7302	single	505del	A118fsX159	N-terminal stop
7324	single	1075_1203dup	V351fsX402	Frameshift and stop in bZIP
2183#	not included	722-736dup	H191_H195dup	Duplication in TAD2
5359#	not included	575G>C	R142T	Substitution in TAD2

bZIP indicates basic leucine zipper region; TAD1, first transactivation domain; and, TAD2, second transactivation domain. Column "category" indicates whether a case was considered double or single *CEBPA* mutant AML in gene expression profiling and survival analyses. Nucleotide numbering is according to NCBI Entrez accession no. XM_009180.3, in which the major translational start codon starts at nucleotide position 151. All mutations are heterozygous, unless "homozygous" is specifically indicated. The locations of functional domains are derived from Mueller and Pabst (7).

* AML case not included in gene expression profiling and survival analyses because data not available.

AML case considered to be non-mutant in gene expression profiling and survival analysis; see main text.

Table 2. Clinical and molecular characteristics of *CEBPA*[wt], *CEBPA*[single-mut] and *CEBPA*[double-mut] AML cases included in gene expression profiling and survival analyses.

	CEBPA[wt] (N=486)	*CEBPA*[single-mut] (N=12)	*CEBPA*[double-mut] (N=26)	P*
Sex				>0.99
Male	240 (49)	7 (58)	16 (62)	
Female	246 (51)	5 (42)	10 (38)	
Age	46 (15 – 77)	52 (20 - 70)	45 (16 - 75)	0.27
Age group				0.58
< 60	426 (88)	10 (83)	24 (92)	
> 60	60 (12)	2 (17)	2 (8)	
HOVON protocol				0.091
04/04A	89 (18)	4 (33)	6 (23)	
29	200 (41)	1 (8)	12 (46)	
32	4 (1)	0 (0)	1 (4)	
42	137 (28)	5 (42)	6 (23)	
43	56 (12)	2 (17)	1 (4)	
Type of stem cell transplant				0.25
None	281 (58)	8 (67)	10 (38)	
Autologous	62 (13)	1 (8)	7 (27)	
Allogeneic	139 (29)	3 (25)	8 (31)	
Unknown	4 (1)	0 (0)	1 (4)	
WBC (x 10^9/L)	28 (0 – 510)	11 (3 – 263)	36 (3 – 174)	0.65
Blasts in bone marrow (%)	64 (0 – 98)	70 (16 – 92)	65 (25 – 94)	0.66
Platelets (x 10^9/L)	55 (3 – 998)	89 (18 – 174)	51 (8 – 265)	0.087
FAB				0.18
M0	18 (4)	0 (0)	0 (0)	
M1	87 (18)	3 (25)	15 (58)	
M2	116 (24)	4 (33)	7 (27)	
M3	25 (5)	0 (0)	0 (0)	
M4	82 (17)	3 (25)	2 (8)	
M5	113 (23)	0 (0)	1 (4)	
M6	7 (1)	0 (0)	0 (0)	
RAEB-t	17 (3)	1 (8)	1 (4)	
RAEB	4 (1)	0 (0)	0 (0)	
M4E	5 (1)	0 (0)	0 (0)	
Unknown	12 (2)	1 (8)	0 (0)	
Cytogenetic risk group				0.54

Table 2. continued

	CEBPA[wt] (N=486)	*CEBPA*[single-mut] (N=12)	*CEBPA*[double-mut] (N=26)	P*
Good	103 (21)	0 (0)	0 (0)	
Intermediate	269 (53)	11 (92)	25 (96)	
Poor	102 (21)	1 (8)	1 (4)	
Unknown	12 (2)	0 (0)	0 (0)	
Cytogenetics[†]				
Normal cytogenetics	187 (38)	7 (58)	20 (77)	0.27
t(15;17)	25 (5)	0 (0)	0 (0)	NA
t(8;21)	37 (8)	0 (0)	0 (0)	NA
Inv(16)	41 (8)	0 (0)	0 (0)	NA
11q23	22 (5)	0 (0)	0 (0)	NA
3q	20 (4)	0 (0)	0 (0)	NA
-5(q)	15 (3)	0 (0)	1 (4)	>0.99
-7(q)	35 (7)	1 (8)	1 (4)	0.54
+chr8	35 (7)	2 (17)	0 (0)	0.094
Complex karyotype	32 (7)	1 (8)	0 (0)	0.32
Other cytogenetic abnormality	123 (25)	2 (17)	5 (19)	>0.99
Molecular abnormalities[†]				
FLT3-ITD	133 (27)	5 (42)	3 (12)	0.081
FLT3-TKD	51 (10)	2 (17)	0 (0)	0.094
NPM1	154 (32)	3 (25)	0 (0)	0.026*
NRAS	46 (9)	1 (8)	3 (12)	>0.99
KRAS	5 (1)	0 (0)	0 (0)	NA

WBC indicates white blood cell count; FAB, French-American-British classification; and NA, not applicable.
Number of cases (percentage) or median (range) are depicted where appropriate.
P values are indicated for the comparison between *CEBPA*[double-mut] and *CEBPA*[single-mut] AML using the two-sided Fisher exact test (categorical variables) and Mann-Whitney test (continuous variables). A *P* value lower than 0.05 has been marked (*).
[†] Each abnormality was taken into account, irrespective of the presence of other abnormalities.

evident that these findings and their clinical significance warrant confirmation in independent cohorts of AML.

In summary, the data presented here indicate that *CEBPA*[mut] AML should at least be distinguished according to the presence of *CEBPA*[double-mut] and *CEBPA*[single-mut]. Screening using dHPLC, followed by nucleotide sequencing, appears useful for rapidly identifying mutant cases. In addition, gene expression based classification, for instance using the classifiers described here, enables the accurate identification of *CEBPA*[double-mut] AML cases.

Table 3. Multivariable analysis of *CEBPAdouble-mut* and *CEBPAsingle-mut* as prognostic markers for overall survival and event-free survival.

	Overall survival		Event-free survival	
	HR (95% CI)	P*	HR (95% CI)	P*
CEBPAsingle-mut [†]	1.18 (0.58 - 2.40)	0.65	1.61 (0.82 - 3.17)	0.16
CEBPAdouble-mut [†]	0.32 (0.17 - 0.61)	<0.001*	0.35 (0.20 - 0.62)	<0.001*
Intermediate[‡]	2.21 (1.52 - 3.22)	<0.001*	2.05 (1.46 - 2.87)	<0.001*
Poor[‡]	3.35 (2.27 - 4.94)	<0.001*	2.85 (2.00 - 4.06)	<0.001*
Age [decades]	1.17 (1.08 - 1.28)	<0.001*	1.10 (1.02 - 1.19)	0.014*
WBC[§]	1.33 (1.05 - 1.68)	0.019*	1.29 (1.03 - 1.62)	0.025*
FLT3-ITD[|]	1.56 (1.20 - 2.03)	<0.001*	1.46 (1.14 - 1.89)	0.003*
NPM1[|]	0.55 (0.41 - 0.74)	<0.001*	0.51 (0.39 - 0.67)	<0.001*

Complete data for multivariable analysis were available for 511 cases.
HR indicates hazard ratio; CI, confidence interval; WBC, white blood cell count; *FLT3*, fms-related tyrosine kinase 3; ITD, internal tandem duplication; and, *NPM1,* nucleophosmin.
* P value < 0.05
[†] *CEBPA* status versus *CEBPAwt*
[‡] Cytogenetic risk versus cytogenetic good risk
[§] WBC higher than 20 x 10^9/L versus lower than 20 x 10^9/L
[|] *FLT3*-ITD versus no *FLT3*-ITD
[|] *NPM1* mutation versus no *NPM1* mutation

ACKNOWLEDGEMENTS

REFERENCES

1. Snaddon J, Smith ML, Neat M, Cambal-Parrales M, Dixon-McIver A, Arch R, et al. Mutations of CEBPA in acute myeloid leukemia FAB types M1 and M2. Genes Chromosomes Cancer 2003;37(1):72-8.

2. Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, et al. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. Nat Genet 2001;27(3):263-70.

3. Gombart AF, Hofmann WK, Kawano S, Takeuchi S, Krug U, Kwok SH, et al. Mutations in the gene encoding the transcription factor CCAAT/enhancer binding protein alpha in myelodysplastic syndromes and acute myeloid leukemias. Blood 2002;99(4):1332-40.

4. Preudhomme C, Sagot C, Boissel N, Cayuela JM, Tigaud I, de Botton S, 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 2002;100(8):2717-23.

5. Frohling S, Schlenk RF, Stolze I, Bihlmayr J, Benner A, Kreitmeier 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 2004;22(4):624-33.

6. Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, Meijer J, van Oosterhoud S, van Putten WL, Valk PJ, et al. Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML. Hematol J 2003;4(1):31-40.

7. Mueller BU, Pabst T. C/EBPalpha and the pathophysiology of acute myeloid leukemia. Curr Opin Hematol 2006;13(1):7-14.

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

9. Bienz M, Ludwig M, Leibundgut EO, Mueller BU, Ratschiller D, Solenthaler M, et al. Risk assessment in patients with acute myeloid leukemia and a normal karyotype. Clin Cancer Res 2005;11(4):1416-24.

10. Schlenk RF, Dohner K, Krauter J, Frohling S, Corbacioglu A, Bullinger L, et al. Mutations and treatment outcome in cytogenetically normal acute myeloid leukemia. N Engl J Med 2008;358(18):1909-18.

11. Pabst T, Mueller BU. Transcriptional dysregulation during myeloid transformation in AML. Oncogene 2007;26(47):6829-37.

12. Leroy H, Roumier C, Huyghe P, Biggio V, Fenaux P, Preudhomme C. CEBPA point mutations in hematological malignancies. Leukemia 2005;19(3):329-34.

13. Wouters BJ, Sanders MA, Lugthart S, Geertsma-Kleinekoort WM, van Drunen E, Beverloo HB, et al. Segmental uniparental disomy as a recurrent mechanism for homozygous CEBPA mutations in acute myeloid leukemia. Leukemia 2007;21(11):2382-4.

14. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

15. Verhaak RG, Goudswaard CS, van Putten W, Bijl MA, Sanders MA, Hugens W, 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 2005;106(12):3747-54.

16. Valk PJ, Bowen DT, Frew ME, Goodeve AC, Lowenberg B, Reilly JT. Second hit mutations in the RTK/RAS signaling pathway in acute myeloid leukemia with inv(16). Haematologica 2004;89(1):106.

17. Care RS, Valk PJ, Goodeve AC, Abu-Duhier FM, Geertsma-Kleinekoort WM, Wilson GA, et al. Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukaemias. Br J Haematol 2003;121(5):775-7.

18. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002;99(10):6567-72.

19. Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. J Clin Oncol 2005;23(29):7332-41.

20. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 2007;99(2):147-57.

21. Wouters BJ, Louwers I, Valk PJ, Lowenberg B, Delwel R. A recurrent in-frame insertion in a CEBPA transactivation domain is a polymorphism rather than a mutation that does not affect gene expression profiling-based clustering of AML. Blood 2007;109(1):389-90.

22. Lin LI, Chen CY, Lin DT, Tsay W, Tang JL, Yeh YC, et al. Characterization of CEBPA mutations in acute myeloid leukemia: most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells. Clin Cancer Res 2005;11(4):1372-9.

23. Wouters BJ, Jorda MA, Keeshan K, Louwers I, Erpelinck-Verschueren CA, Tielemans D, et al. Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. Blood 2007;110(10):3706-14.

24. Pabst T, Eyholzer M, Haefliger S, Schardt J, Mueller BU. Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia. J Clin Oncol 2008;26(31):5088-93.

CHAPTER 10

CCAAT/enhancer binding protein alpha target genes in myeloid progenitor cells uncovered by chromatin immunoprecipitation on DNA promoter microarrays (ChIP-chip)

Bas J. Wouters[1], Claudia A.J. Erpelinck-Verschueren[1], Diana Hamer[1], and Ruud Delwel[1]

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

Work in progress

ABSTRACT

CCAAT/enhancer binding protein alpha (C/EBPα) is a lineage-specific transcription factor that directs development of granulocytes. Mutations in its gene are common in acute myeloid leukemia. One of the most frequent types of mutations affects the basic leucine zipper (bZIP) domain, which is involved in dimerization, DNA interaction, and protein interaction. To obtain insights into the genome-wide transcriptome of wild type as well as bZIP-mutant C/EBPα, we have performed chromatin immunoprecipitation on DNA promoter microarrays (ChIP-chip) as well as genome-wide gene expression profiling in the 32D murine myeloid progenitor cell line. These cells expressed an estradiol-inducible form of either wild-type C/EBPα or bZIP-mutant C/EBPα. Our experiments have yielded a collection of putative novel target genes of C/EBPα that may be important in granulocytic development. Furthermore, they suggest an overall loss-of-function of bZIP mutant C/EBPα as regards DNA binding.

INTRODUCTION

The basic leucine zipper (bZIP) transcription factor CCAAT/enhancer binding protein alpha (C/EBPα) is one of the master regulators of myeloid differentiation (1,2). C/EBPα plays a crucial role in regulating the balance between proliferation and differentiation in myeloid progenitor cells. The importance of normal C/EBPα functioning was underscored by the recent discoveries of various mechanisms of deregulation of C/EBPα in acute myeloid leukemia (AML) (1,3).

Two types of mutations in the gene encoding C/EBPα have frequently been reported. Out of frame N-terminal mutations cause the relative overrepresentation of a short isoform of C/EBPα lacking the first transactivation domain. This p30 protein may act as a dominant negative inhibitor of the wild type (wt) 42kD C/EBPα transcriptional activator (4). Mutations in the C-terminal bZIP region are in-frame. They may impair dimerization and/or DNA binding of C/EBPα and possibly other C/EBP members by dimerizing with those (3). The exact functional effects of C/EBPα bZIP mutations, however, remain to be investigated.

A few genes that are regulated by C/EBPα under normal granulocytic differentiation have been identified. These include genes encoding the receptors for the myeloid growth factors G-CSF and GM-CSF and IL-6, and genes that encode specific granulocytic markers such as myeloperoxidase (2,5-8). Those previous studies have provided important insights, but a genome-wide picture of the transcriptome of C/EBPα has not been generated thus far. The rationale behind performing such a study is that many functionally important target genes in normal and/or malignant myeloid development may still remain to be identified.

To establish a genome-wide collection of C/EBPα target genes and to study the effect of mutations in the bZIP region of the protein, we have performed chromatin immunoprecipitation (ChIP) experiments in a murine myeloid progenitor cell line model in which the activity of C/EBPα is inducible. We have analyzed the resulting DNA fragments of wt and C-terminal mutant expressing cells on DNA promoter microarrays. Integrative analyses with genome-wide gene expression data indicated a significant enrichment in genes exhibiting binding of the fusion protein and upregulation on the mRNA level for wt-C/EBPα. Experiments using a C-terminal mutant indicated reduced DNA binding.

MATERIAL AND METHODS

Plasmids

A pBabe-Cebpa-ER fusion construct and pBabe-ER construct (9) were kindly provided by Dr. Daniel Tenen (Harvard Institutes of Medicine, Boston, MA, USA). The Cebpa-ER fusion gene was re-cloned into the pLNCX-neo vector using polymerase chain reaction (PCR) amplification with primers that included appropriate restriction sites, i.e. *Hpa*I and *Cla*I.

Primer sequences were: fw 5'-GTA CGT TAA CAG GAA TTC GCG CCA CCA TGG A-3' and rev 5'-AGG AAT CGA TCT CTC AGA CTG TGG CAG GGA A-3'. A mutant construct was generated using the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA, USA), using the following oligos (insertion mutation is underlined): 5'-GAT AAA GCC AAA CAA CGT AAT GTG GAC AAG CAG CGC AAC GTG GAG-3' (sense) and 5'-CTC CAC GTT GCG CTG CTT GTC CAC ATT ACG TTG TTT GGC TTT ATC-3' (anti-sense). A pLNCX-ER construct was generated with primers fw 5'-TAT GGT TAA CAT CTG CTG GAG ACA TGA GAG CT-3' and the reverse primer used for the Cebpa-ER construct. The cloned constructs and site of inserted mutation were nucleotide-sequenced.

Cell lines and transduction

Phoenix-E cells were transfected with either ER or Cebpa-ER using a calcium phosphate protocol. 32D cells, derived from murine myeloid progenitor cells, were used for cell culture experiments. We used a subclone that stably expresses the human G-CSF receptor (10). Cells were maintained in RPMI 1640 medium, supplemented by 10% fetal calf serum, penicillin (100 IU/ml), streptomycin (100 ng/ml) and murine interleukin 3 (IL-3). IL-3 was obtained from an IL-3-producing Chinese hamster ovary (CHO) cell line. Cells were infected using either ER or Cebpa-ER virus. Stable clones were selected using limiting dilution under G418 selection (0.8 mg/mL; Gibco, Breda, The Netherlands). After selection, expression of ER or Cebpa-ER was confirmed by Western blotting.

Cytospins

Morphologic analysis was performed by microscopy on May-Grünwald-Giemsa-stained cytospins (Shandon Holland, Amsterdam, The Netherlands) using a Zeiss Axioskop microscope with 63 x plan-apochromat objective. Pictures were taken with a Leica DC 500 camera.

Western blot analysis

Protein lysates were prepared according to standard methods and Western blot analysis was performed using antibodies directed against ERα (rabbit polyclonal IgG, HC-20, sc543, Santa Cruz (Santa Cruz, CA, USA)).

Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) was carried out according to a protocol from Affymetrix (Santa Clara, CA, USA), modified as follows. Fifty to 100 million cells were cultured in the presence of 1 μM 17β-estradiol (E$_2$) (Sigma-Aldrich, St. Louis, MO, USA) for 4 hours. Protein and DNA were cross-linked by incubation with formaldehyde (Sigma-Aldrich; final concentration 1%) for 10 minutes, followed by quenching with glycine (Sigma-Aldrich; final concentration 0.125 M) for 5 minutes. Cells were pelleted, washed twice with ice cold phosphate buffered saline (PBS) and three times with lysis buffer (10 mM Tris-HCl, 10 mM

NaCl, 3 mM MgCl$_2$, 0.5% IGEPAL (Sigma-Aldrich), 1 mM PMSF), and then resuspended in pre-IP dilution buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM MgCl$_2$, 1 mM CaCl$_2$, 4% IGEPAL, 1 mM phenylmethanesulfonyl fluoride solution (PMSF; Sigma-Aldrich)). Material was sonicated in a total volume of 1 ml using a Soniprep 150 device (Machine Shop Equipment (MSE), London, UK) for 6 cycles of 20 seconds pulse/90 seconds rest at an amplitude of 6 microns. An aliquot was de-crosslinked, purified and subjected to ethidium bromide agarose gel electrophoresis to ensure the predominant presence of fragments of 100-1000 nucleotides. Cellular debris was removed by pelleting. Ten % of the supernatant was kept aside as input control, and the remainder was used for immunoprecipitation (IP) after addition of 5 volumes of IP dilution buffer (20 mM Tris-Hcl, 2 mM EDTA, 1% Triton X-100 (Roche, Basel, Switzerland), 150 mM NaCl and fresh protease inhibitor (Roche)). Chromatin was pre-cleared by incubation for 30 minutes at 4°C with pre-equilibrated protein G-coupled magnetic beads (Dynabeads; Invitrogen, Carlsbad, CA, USA). Immunoprecipitation was carried out overnight at 4°C using 20 µg rabbit polyclonal IgG anti-ERα (Santa Cruz sc543X). Samples were incubated for 2.5 hours at room temperature in the presence of two hundred µl of pre-equilibrated magnetic beads. The beads were washed twice with ChIP wash 1 (20 mM Tris-HCl, 2 mM EDTA, 1% Triton X-100, 150 mM NaCl and 1 mM PMSF), once with ChIP wash 2 (20 mM Tris-HCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS, 500 mM NaCl and 1 mM PMSF), once with ChIP wash 3 (10 mM Tris-HCl, 1 mM EDTA, 0.25 M LiCl, 0.5% IGEPAL and 0.5% deoxycholate), and twice with TE. Protein/DNA complexes were eluted into 400 µl elution buffer (25 mM Tris-HCl, 10 mM EDTA, 0.5% SDS) by heating at 65°C for 30 minutes. Cross-links were reversed overnight at 65°C in the presence of 20 µl proteinase K (20 mg/ml; New England Biolabs, Ipswich, MA, USA). The non-immunoprecipitated 10% input control sample was processed similarly. De-crosslinked material was purified using a Qiagen (Valencia, CA, USA) PCR Purification Kit and eluted into 20 µl buffer EB.

Whole genome amplification and promoter array hybridization
Ten µl was next used for amplification using a Whole Genome Amplification (WGA2) kit (Qiagen) according to the manufacturer's recommendations. Five to 10 µg of amplified DNA was fragmented following the protocol of Affymetrix SNP array kits, which was confirmed on an Agilent (Santa Clara, CA, USA) Bioanalyzer. Labeling, hybridization to Affymetrix Mouse Promoter 1.0R arrays, staining, washing and scanning were performed according to the manufacturer's recommendations.

Promoter array data analysis
Raw promoter array data were processed using Model based Analysis of Tiling arrays (MAT) software (11). MAT models the baseline probe behavior by considering probe sequence and copy number on each array. It standardizes the probe value through the probe model, eliminating the need for sample normalization. The advantage of MAT is, that it can detect ChIP

regions from a single ChIP sample. In these analyses, 2 ER samples, 2 Cebpamut-ER samples and 3 Cebpa-ER samples were included. Enriched fragments at the 0.0001 significance level were considered to be target regions at a bandwidth of 300, maxgap of 300 and minprobe of 8. The enriched chromosomal regions were mapped to NCBI murine Genome Build 36 (February 2006) and annotated using Cis-regulatory Element Annotation System (CEAS) (http://ceas.cbi.pku.edu.cn/) (12). CEAS was also used to search for transcription factor binding motifs from the TRANSFAC and JASPAR databases (12-14). Visualization of enriched regions was carried out using Affymetrix Integrated Genome Browser (IGB) software.

Gene expression profiling

32D-Cebpa-ER, 32D-CebpaMut-ER and 32D-ER clones were treated with 1 μM E_2. mRNA was isolated using a cesium chloride protocol. Gene expression profiles were obtained using Affymetrix Mouse Genome 430_2 GeneChips. Quality control was performed by assessing scaling factor and plots depicting relative log expression (RLE) and normalized unscaled standard errors (NUSE) (Bioconductor AffyPLM package). Data were processed and normalized with Robust Multi Averaging (RMA) (15) and transformed to $\log_2$ scale. Filtering was performed to include in the analysis only probe sets that showed a variable expression across all samples (interquartile range > 0.10), and expressed at an intensity level greater than $\log_2(100)$ in at least 5% of the samples. Differential expression between groups of samples was assessed using Linear Models for Microarray Analysis (limma) (16), in which a factor was included to adjust for experiment batches. Enrichment of immunoprecipitated genes in the list of differentially expressed probe sets was assessed using the option to investigate pre-ranked gene lists in the java-version of Gene Set Enrichment Analysis (GSEA) (17). The limma-derived batch-effect-adjusted list of probe sets at the 2 hour time point was used as input list for this, ordered according to the moderated t-statistic. These analyses were performed on the probe set level without collapsing to gene name. All bioinformatic analyses were performed in R. To map between various identifiers (murine Affy ID, Refseq ID, human HGNC symbols for GSEA), the biomaRt package from Bioconductor was used.

PCR for the *Il6ra* promoter

Quantitative RT-PCR for the *Il6ra* promoter was carried out with SYBRGreen PCR master mix (Applied Biosystems, Foster City, CA, USA) on an Applied Biosystems 7900HT Sequence Detection System using primers 5'-GCT TCA GCA CCG GTT CCT CA-3' (fw) and 5'-CTC CAC ACT CAG CTC GCA CT-3' (rev). A standard curve was made to confirm efficiency of the reaction. Expression values relative to input sample were calculated by subtracting threshold cycle (Ct) values, i.e. using the delta Ct method.

RESULTS

32D-Cebpa-ER cells differentiate upon E$_2$ treatment while 32D-Cebpamut-ER cells show impaired differentiation

Stable 32D clones expressing a fusion protein of murine C/EBPα and the ligand-binding domain (LBD) of the human estrogen receptor alpha (ERα) (32D-Cebpa-ER) were generated. 32D cells proliferate when stimulated with IL-3 and differentiate into granulocytes when induced by G-CSF (10). Upon addition of E$_2$, 32D-Cebpa-ER cells morphologically differentiated into mature granulocytes even in the presence of IL-3 (Figure 1A), while their proliferation rate dropped (Figure 1B). This effect appeared even more pronounced when the cells were treated with G-CSF plus E$_2$. (Figure 1C-D). As expected, control cells only expressing the LBD of ERα (32D-ER) did not show a response to E$_2$ (Figure 1B and data not shown).

32D cells expressing a mutant Cebpa-ER fusion with an insertion of 6 amino acids in the bZIP domain (Cebpamut-ER) previously found in a human AML patient (18) showed less pronounced inhibition of proliferation upon treatment with E$_2$ in the presence of IL3. Mor-

Figure 1. 32D-Cebpa-ER cells differentiate upon E2 treatment while 32D-Cebpamut-ER cells show impaired differentiation

A. Representative cytospins of 32D-Cebpa-ER cells (clone C-E.4) and 32D-Cebpamut-ER cells (clone Cmut-E.9) in the absence (-) and presence (+) of E$_2$ for 3 or 7 days. The cells were cultured in the presence of IL3.
B. Proliferation curves of 32D-ER cells (ER), 32D-Cebpa-ER cells (CEBPA) and 32D-Cebpamut-ER cells (mut4) stimulated with IL3 alone (IL3 -) or IL3 in combination with E$_2$ (IL3 +est). The graph represents the average and standard deviations for 4 individual clones for each of the 3 cell types.
C. Representative cytospins of 32D-Cebpa-ER cells (clone C-E.4) and 32D-Cebpamut-ER cells (clone Cmut-E.9) in the absence (-) and presence (+) of E$_2$ for 3 or 7 days. The cells were cultured in the presence of G-CSF.
D. Proliferation curves of 32D-ER cells (ER), 32D-Cebpa-ER cells (CEBPA) and 32D-Cebpamut-ER cells (mut4) stimulated with G-CSF alone (G-CSF -) or IL3 in combination with E$_2$ (G-CSF +est). The graph represents the average and standard deviations for 4 individual clones for each of the 3 cell types.
(A full color version of this figure can be found in the color section.)

phological differentiation of these cells could hardly be detected under IL3 plus E_2 conditions (Figure 1A-B). In the presence of G-CSF, 32D-Cebpamut-ER cells demonstrated delayed differentiation with the suggestion of a partial block (Figure 1C).

Discovery of putative novel and previously identified C/EBPα interacting loci in 32D-Cebpa-ER cells applying chromatin immunoprecipitation

32D-Cebpa-ER cells and 32D-ER control cells were treated with E_2 for 4 hours, and ChIP was carried out using an antibody directed against ERα. The immunoprecipitated fragments were analyzed on DNA microarrays covering over 25.000 murine promoter regions. 4826 chromosomal regions, sized between 310 and 7399 nucleotides (median 1084), were enriched in 32D-Cebpa-ER cells compared to 32D-ER control cells at $P<1x10^{-5}$ (false discovery rate 2.3%). The transcriptional start site of 1064 unique RefSeq genes was located within 1 kb downstream of one or more of the 4826 fragments. Among those were genes that have previously been reported to be directly or indirectly regulated by C/EBPs, such as *Mpo* (8), *Hp* (19), *C3* (19), and *Sfpi1 (PU.1)* (20). The analysis also indicated direct binding to the promoter of *Il6ra,* the gene encoding the IL-6 receptor alpha, which has previously been identified as a critical downstream target of C/EBPα (6). The direct interaction of C/EBPα to the *Il6ra* locus in the ChIP experiments was confirmed by PCR (Figure 2A). These latter findings give confidence, that among the the putative novel C/EBPα target genes critical players may be hidden, important in neutrophilic development and consequently in myeloid transformation.

Enriched fragments share putative binding sites

We next investigated which known transcription factor binding motifs could be detected in the 4826 immunoprecipitated DNA fragments. C/EBP binding motifs were among the most highly enriched sequences (Table 1). Ninety-seven other binding sites were also enriched (fold change over 1.5 and $P<1x10^{-5}$). These included motifs associated with known interactors of C/EBP, such as E2F (21), Sp1 (22-25) and Ets factors (3). On the other hand, binding motifs of multiple potentially novel C/EBPα interacting factors were found, such as AP2α and AP2γ (Table 1).

Immunoprecipitated DNA fragments are associated with differential expression of nearby genes

Genome-wide gene expression profiles were generated at steady state proliferation and at two time points following E_2 treatment. Principal component analysis of the gene expression data showed that E_2 treatment of 32D-Cebpa-ER clones resulted in a shift in genome-wide expression patterns. These effects were already apparent as early as 2 hours of E_2 stimulation and were time dependent, with the profiles at time point 8 hours being distinct from those

Table 1. Transcription factor binding sites enriched in DNA fragments immunoprecipitated in 32D-Cebpa-ER cells compared to 32D-ER cells.

Motif	Hits	Fold change
AP2alpha	9606	1.59
cEBP	9510	1.51
Elk-1	5313	1.69
M00109: CEBPbeta	6122	1.61
M00470: AP-2gamma	5839	1.62
M00695: ETF	3412	1.87
M00446: Spz1	4789	1.65
M00025: Elk-1	2917	1.92
M00008: Sp1	5085	1.62
M00771: ETS	3719	1.67
E74A	3585	1.68
M00427: E2F	2269	1.89
M00189: AP-2	1454	2.24
M00932: Sp-1	1230	2.40
M00940: E2F-1	1437	2.17
M00931: Sp-1	1310	2.23
M00933: Sp-1	1404	2.12
M00918: E2F	1184	2.27
M00373: Pax-4	1290	2.18
M00067: Hairy	1266	2.18
M00939: E2F-1	1137	2.27
SP1	3600	1.54
M00196: Sp1	775	2.62
M00032: c-Ets-1(p54)	2551	1.64
HLF	2472	1.64
M00380: Pax-4	2242	1.68
Ahr-ARNT	2861	1.56
M00426: E2F	1550	1.86
M00915: AP-2	1324	1.96
M00920: E2F	725	2.48
M00919: E2F	844	2.29
M00736: E2F-1DP-1	1496	1.82
M00322: c-MycMax	1879	1.67
M00327: Pax-3	1591	1.74

Table 1. continued

Motif	Hits	Fold change
Pax-4	1871	1.63
M00737: E2F-1DP-2	846	2.13
M00739: E2F-4DP-2	762	2.04
M00260: HLF	2071	1.50
M00778: AhR	1159	1.74
M00016: E74A	495	2.37
M00227: v-Myb	1384	1.60
M00800: AP-2	602	2.07
M00113: CREB	954	1.72
M00339: c-Ets-1	667	1.93
M00516: E2F	463	2.19
M00916: CREB	1091	1.62
Dorsal_1	718	1.79
M00938: E2F-1	299	2.58
M00769: AML	884	1.68
Bsap	575	1.88
M00649: MAZ	934	1.58
CREB	562	1.82
M00740: RbE2F-1DP-1	358	2.14
M00984: PEBP	759	1.65
NRF-2	370	2.10
SAP-1	368	2.10
M00947: CP2LBP-1cLSF	784	1.61
M00774: NF-kappaB	753	1.58
M00039: CREB	716	1.59
M00965: LXR	744	1.57
M00738: E2F-4DP-1	286	2.13
M00178: CREB	482	1.76
M00017: ATF	623	1.62
M00280: RFX1	679	1.58
M00177: CREB	463	1.72
M01045: AP-2alphaA	305	1.94
M00235: AhRArnt	305	1.89
E2F	223	2.10
M00050: E2F	214	2.12

Table 1. continued

Motif	Hits	Fold change
M00236: Arnt	607	1.53
M00453: IRF-7	472	1.61
M00982: KROX	150	2.43
M00929: MyoD	562	1.53
M00341: GABP	142	2.44
M01036: COUPTF	327	1.73
M00281: RFX1	558	1.50
M00797: HIF-1	318	1.72
M00040: CRE-BP1	348	1.63
M00181: E2	306	1.66
M00466: HIF-1	358	1.59
M01047: AP-2alphaA	74	3.01
M00539: Arnt	181	1.84
M00171: Adf-1	99	2.31
M00114: TaxCREB	61	2.62
M00107: E2	253	1.55
Irf-1	187	1.62
M00928: E2	122	1.81
M00985: Stra13	147	1.65
M00460: STAT5A	99	1.79
M00966: VDR	84	1.89
M00259: STAT	58	2.13
Hen-1	122	1.63
M00043: dl	98	1.70
M00062: IRF-1	41	2.29
p50	84	1.70
M00284: TCF11MafG	65	1.83
M00058: HEN1	13	4.31
M00068: HEN1	25	2.63
Androgen	94	1.60

Enrichment for binding sites was assessed by CEAS software, which queries the TransFac and JASPAR databases. Enriched sites matching the TransFac database are denoted by their entry number (e.g. M00109 for the C/EBPbeta binding motif). Motifs with a *P*-value < 10^{-5} are shown, ranked based on increasing *P*-value, so that the most significant motif is on top.

generated at 2 hours (Figure 2B). Treatment of 32D-ER cells did not result in major changes in gene expression levels.

To integrate the ChIP and expression data, we used gene set enrichment analysis (GSEA) (17). GSEA detected a significant association between genes upregulated in the Cebpa-ER

Figure 2. ChIP-chip identifies promoter regions occupied by the Cebpa-ER fusion protein that are associated with gene expression changes.

A. RQ-PCR validation of ChIP of *Il6ra* promoter. Three independent ChIP experiments were performed using 32D-Cebpa-ER cells (black bars) or 32D-ER cells (white bars). RQ-PCR was performed in the *Il6ra* promoter, and percentage of amount of input material is indicated in log10 scale. In each of the three experiments, enrichment in Cebpa-ER clones is apparent.

B. Principal component analysis of gene expression data from 32D-Cebpa-ER cells (reddish) and 32D-ER cells (bluish). Color intensities reflect time points, ranging from dark (time point 0), brighter (2 hours of E_2 treatment) and brightest (8 hours of E_2 treatment). Separation of samples according to E_2 treatment over the first principal component (PCA1) is apparent for the Cebpa-ER clones only. An additional separation over the second principal component (PCA2) correlates with time point (2 versus 8 hours of E_2 treatment).

C. Gene set enrichment analysis results. Gene expression profiles from 32D-Cebpa-ER cells (N=4) and 32D-ER cells (N=4) were compared after 2 hours of E_2 treatment. All probe sets that passed the variability and intensity filter were ordered according to their higher expression in either Cebpa-ER (left) or ER (right) cells. Genes that were immunoprecipitated are indicated in the bar as vertical lines. There is a significant (P<0.0001) enrichment of immunoprecipitated genes in the left part of the graph, indicated by the positive deflection of the enrichment score (upper part).

(A full color version of this figure can be found in the color section.)

Table 2. Core enrichment signature for 32D-Cebpa-ER cells versus 32D-ER cells.

Probe set	Gene symbol	Also in mutant	Probe set	Gene symbol	Also in mutant
1417566_at	Abhd5		1415693_at	Derl1	
1420591_at	Gpr84		1425037_at	Fgd4	
1448302_at	Kctd20		1427747_a_at	Lcn2	
1452521_a_at	Plaur		1435450_at	Cpne3	
1418337_at	Rpia		1435040_at	Irak3	
1448407_at	4632428N05Rik		1424725_at	D16Ertd472e	+
1429568_x_at	EG432649		1418932_at	Nfil3	
1421408_at	Igsf6		1451466_at	D16Ertd472e	+
1433563_s_at	Derl1		1427155_at	Fchsd1	
1423298_at	Add3		1456796_at	Snai3	
1448881_at	Hp		1416252_at	Stk38	
1416619_at	4632428N05Rik		1416157_at	Vcl	
1418295_s_at	Dgat1		1421560_at	Snai3	
1424032_at	Hvcn1		1451218_at	Edem1	
1450275_x_at	EG432649		1436088_at	0910001A06Rik	
1424724_a_at	D16Ertd472e	+	1423223_a_at	Prdx6	+
1423297_at	Add3		1452096_s_at	D230025D16Rik	
1426574_a_at	Add3		1427994_at	Cd300lf	
1416592_at	Glrx		1417632_at	Atp6v0a1	+
1416156_at	Vcl		1452814_at	Cpne3	
1426521_at	D230025D16Rik		1417811_at	Slc24a6	
1432426_a_at			1416996_at	Tbc1d8	
1423954_at	C3		1443536_at	Slc7a11	
1426545_at	Tnrc6b		1422742_at	Hivep1	+
1416593_at	Glrx		1448377_at	Slpi	+
1423871_at	Tmem63a		1418162_at	Tlr4	
1434042_s_at	Mtmr3		1460650_at	Atp6v0a1	+
1429897_a_at	D16Ertd472e	+	1429487_at	Ppp1r12a	
1425227_a_at	Atp6v0a1	+	1452669_at	2810012G03Rik	
1450513_at	Cd33		1430704_at	Irak3	
1452070_at	Dedd2		1420088_at	Nfkbia	
1420413_at	Slc7a11		1448830_at	Dusp1	
1451416_a_at	Tgm1		1430878_at	Igsf6	
1428514_at	Cpne3		1434037_s_at	Pcaf	
1427117_at	Mtmr3		1424301_at	Zfp219	
1451021_a_at	Klf5		1433968_a_at	Megf9	
1419132_at	Tlr2		1429321_at	Rnf149	+
1437723_s_at	Derl1		1423389_at	Smad7	

Probe set	Gene symbol	Also in mutant
1453839_a_at	Pi16	
1455683_a_at	Tbc1d8	
1435743_at	Klhl23	
1450964_a_at	Osbpl9	+
1448213_at	Anxa1	
1426223_at	2810439F02Rik	
1436982_at	Tnrc6b	
1433934_at	Sec24a	
1460670_at	Riok3	
1416378_at	Pnkp	+
1454713_s_at	Hdc	
1451206_s_at	Pscdbp	
1449519_at	Gadd45a	
1435697_a_at	Pscdbp	
1426299_at	Snx20	
1436186_at	E2f8	
1418747_at	Sfpi1	
1422650_a_at	Riok3	
1420331_at	Clec4e	+
1418645_at	Hal	
1448306_at	Nfkbia	
1423707_at	Tmem50b	
1435114_at	Wdhd1	
1451796_s_at	Hdc	
1433639_at	5730593F17Rik	
1418163_at	Tlr4	
1416654_at	Slc31a2	
1423306_at	2010002N04Rik	+
1449731_s_at	Nfkbia	
1448413_at	2410016O06Rik	
1416440_at	Cd164	
1422603_at	Rnase4	
1427032_at	Herc4	
1436511_at	BC031781	
1448558_a_at	Pla2g4a	
1417750_a_at	Slc25a37	
1431085_a_at	Pcmt1	+
1443859_at	Rsbn1	+
1420330_at	Clec4e	+
1460273_a_at	Naip2	

Probe set	Gene symbol	Also in mutant
1437741_at	Rab21	+
1434885_at	Spty2d1	
1450821_at	Pcaf	
1418219_at	Il15	
1423383_a_at	Osbpl9	+
1438157_s_at	Nfkbia	
1449538_a_at	Gcnt1	
1421098_at	Stap1	+
1417329_at	Slc23a2	
1433506_at	Lrrc8d	
1433505_a_at	Lrrc8d	
1453163_at	Ppp1r12a	
1434082_at	Pctk2	
1417179_at	Tspan5	+
1427568_a_at	Ift80	
1421194_at	Itga4	
1456544_at	Tmem38b	
1423082_at	Derl1	
1423943_at	Dus1l	
1421340_at	Map3k5	+
1415834_at	Dusp6	+
1422286_a_at	Tgif1	
1418135_at	Aff1	
1416454_s_at	Acta2	
1451340_at	Arid5a	
1426445_at	Ctage5	+
1440865_at	Ifitm6	
1460024_at	Tnrc6b	
1449006_at	Gla	
1422665_a_at	Pcmt1	+
1434015_at	Slc2a6	
1453721_a_at	Slc31a2	
1426798_a_at	Ppp1r15b	
1424065_at	Edem1	
1427102_at	Slfn4	
1416610_a_at	Clcn3	
1436545_at	Dtx4	
1419714_at	Cd274	
1437864_at	Adipor2	
1452425_at	Tnfrsf14	

Probe set	Gene symbol	Also in mutant
1417124_at	Dstn	
1419665_a_at	Nupr1	
1425837_a_at	Ccrn4l	
1426046_a_at	Rabggta	
1415747_s_at	Riok3	
1419666_x_at	Nupr1	
1425713_a_at	Rnf146	
1419262_at	Acad8	
1424700_at	Tmem38b	
1454775_at	Hdac10	
1459522_s_at	Gyg	
1456388_at	Atp11a	
1435734_x_at	Dus1l	
1426457_at	Slmap	
1417136_s_at	Srpk2	
1421291_at	Il18rap	
1422051_a_at	Gabbr1	
1416061_at	Tbc1d15	
1416062_at	Tbc1d15	
1448429_at	Gyg	
1421623_at	Il12rb2	+
1428329_a_at	Ift80	
1429614_at	Prpf18	
1426603_at	Rnasel	
1416737_at	Gys3	
1423518_at	Csk	+
1417588_at	Galnt3	
1419247_at	Rgs2	
1427481_a_at	Atp1a3	
1424194_at	Rcsd1	
1433486_at	Clcn3	
1455764_at	Vprbp	
1460674_at	Paqr7	
1417266_at	Ccl6	+
1450241_a_at	Evi2a	
1456498_at	Itga4	
1425640_at	Aff1	
1416488_at	Ccng2	
1450155_at	Itga4	
1451208_at	Etf1	

Probe set	Gene symbol	Also in mutant
1416010_a_at	Ehd1	+
1459853_x_at	BC031781	
1424917_a_at	Wipi1	
1419261_at	Acad8	
1448721_at	D1Ertd622e	
1422124_a_at	Ptprc	+
1418432_at	Cab39	
1426098_a_at	Cast	
1417135_at	Srpk2	
1448175_at	Ehd1	+
1416011_x_at	Ehd1	+
1449324_at	Ero1l	
1451376_at	5730596K20Rik	
1433685_a_at	Hjurp	
1417245_at	Gpr180	
1423829_at	0910001A06Rik	
1431055_a_at	Snx10	
1450478_a_at	Ptpn12	
1417330_at	Slc23a2	
1456604_a_at	Pcmt1	+
1448318_at	Adfp	
1448489_at	Pafah2	
1423622_a_at	Ccnl1	
1426461_at	Ugp2	
1455332_x_at	Fcgr2b	
1451742_a_at	Ugp2	
1422045_a_at	Ptpn12	
1444319_at	E2f8	
1422736_at	Ranbp9	
1419030_at	Ero1l	
1435560_at	Itgal	
1453567_s_at	2810441K11Rik	
1451361_a_at	Pnpla7	
1451746_a_at	Atg12	
1424320_a_at	Traf7	
1434485_a_at	Ugp2	
1452001_at	Nfe2	
1431530_a_at	Tspan5	
1448364_at	Ccng2	
1418096_at	Dok3	

Probe set	Gene symbol	Also in mutant
1427692_a_at	Cask	
1455105_at	Ptpn12	
1457287_at	Edem1	
1435477_s_at	Fcgr2b	
1435241_at	Heatr5a	
1417810_a_at	Pacsin2	
1435476_a_at	Fcgr2b	
1436625_at	Fcgr1	+
1425269_at	Apbb1ip	
1451941_a_at	Fcgr2b	
1425367_at	Itgal	
1417778_at	Zfp35	
1450729_at	Hs2st1	
1435890_at	5730596K20Rik	
1416060_at	Tbc1d15	
1419609_at	Ccr1	
1434329_s_at	Adipor2	
1449988_at	Gimap1	
1419219_at	Cyp4f18	
1439060_s_at	Wipi1	
1438855_x_at	Tnfaip2	
1420249_s_at	Ccl6	+
1418433_at	Cab39	
1425641_at	Aff1	
1418248_at	Gla	
1422191_at	Cd200r1	
1415983_at	Lcp1	
1421853_at	Psen1	
1456199_x_at	Ranbp9	
1435174_at	Rsbn1	+
1422469_at	Tbk1	
1447584_s_at	Myct1	
1422046_at	Itgam	
1420897_at	Snap23	
1450196_s_at	Gys3	
1459902_at	2700007P21Rik	
1417471_s_at	D1Ertd622e	
1450697_at	Slc30a7	
1417995_at	Ptpn22	+
1460315_s_at	Tbk1	

Probe set	Gene symbol	Also in mutant
1434185_at	Acaca	
1452045_at	Zfp281	
1425333_at	Rab43	
1449317_at	Cflar	
1429352_at	Mocos	
1437211_x_at	Elovl5	
1438397_a_at	Rbm39	
1426433_at	Myct1	
1424011_at	Aqp9	
1426458_at	Slmap	
1425129_a_at	Taldo1	
1459981_s_at	Rsbn1	+
1438562_a_at	Ptpn2	
1451639_at	Cebpg	
1426217_at	2810441K11Rik	
1452072_at	Myct1	
1453731_a_at	Tmem77	
1427540_at	Zwint	
1417523_at	Plek	
1417134_at	Srpk2	
1451413_at	Cast	
1421947_at	Gng12	
1438366_x_at	Clcn3	
1455711_at	Dtx4	
1437514_at	B430306N03Rik	
1455089_at	Gng12	
1417140_a_at	Ptpn2	
1453120_at	Txndc13	
1438048_at	Myct1	
1431893_a_at	Pdss1	
1435694_at	Arhgap26	
1422518_at	Cask	
1429012_at	Arhgef6	
1454702_at	4930503L19Rik	
1448603_at	Srpk2	
1420024_s_at	Etf1	
1420502_at	Sat1	
1460314_s_at	NM_178216.1	
1425546_a_at	Trf	
1452058_a_at	Rnf11	

Probe set	Gene symbol	Also in mutant
1419048_at	Pcnx	
1448404_at	Scamp2	
1436037_at	Itga4	
1434405_at	Fnip1	
1454646_at	Tcp11l2	
1438999_a_at	Nfat5	
1419029_at	Ero1l	
1437742_at	Rab21	+
1450764_at	Aoah	
1421215_a_at	Slmap	
1428306_at	Ddit4	
1448160_at	Lcp1	
1450745_at	C1galt1	
1426460_a_at	Ugp2	
1419561_at	Ccl3	
1435402_at	Gramd1b	
1418760_at	Rdh11	
1418127_a_at	Aifm1	
1449677_s_at	Tmem38b	
1437232_at	Bpil2	
1438401_at	Ubn1	
1445687_at	Gm885	
1422886_a_at	Clk4	
1421139_a_at	Zfp386	
1453052_at	Tmem192	
1425806_a_at	Med21	+
1421167_at	Atp11a	
1449983_a_at	Nqo2	
1437289_at	Impad1	
1451275_at	Uhrf1bp1l	
1448749_at	Plek	
1416525_at	Spop	
1448898_at	Ccl9	
1439478_at	Acot2	
1449591_at	Casp4	
1420497_a_at	Cebpz	
1425139_at	Sesn2	
1425173_s_at	Golph3l	
1429881_at	Arhgap15	
1455446_x_at	Acadsb	

Probe set	Gene symbol	Also in mutant
1426440_at	Dhrs7	
1426011_a_at	Ggnbp2	
1416273_at	Tnfaip2	
1448748_at	Plek	
1456341_a_at	Klf9	
1420896_at	Snap23	
1415741_at	Tmem165	
1448148_at	Grn	
1434486_x_at	Ugp2	
1419047_at	Pcnx	
1451619_at	Golph3l	
1428735_at	Cd69	+
1427022_at	Ddx42	
1416611_at	Scamp2	
1420898_at	Snap23	
1417936_at	Ccl9	
1426454_at	Arhgdib	
1436996_x_at	Lyz2	
1439426_x_at	Lyz2	
1435689_at	9030025P20Rik	+
1428615_at	P2ry5	
1453818_a_at	9030025P20Rik	+
1421605_a_at	Aqp9	
1448106_at	Necap1	
1428484_at	Osbpl3	
1460338_a_at	Crlf3	
1449668_s_at	Fnip1	
1418397_at	Zfp275	
1448390_a_at	Dhrs3	+
1424732_s_at	Tmem192	
1444090_at	Pram1	
1434031_at	Zfp692	
1448568_a_at	Slc20a1	
1450018_s_at	Slc25a30	
1425646_at	BC016495	
1448704_s_at	H47	
1434216_a_at	Nudt19	
1434609_at	Vprbp	
1421191_s_at	Gopc	
1437288_at	Impad1	

Probe set	Gene symbol	Also in mutant
1427539_a_at	Zwint	
1451997_at	Zfp426	
1432332_a_at	Nudt19	
1452601_a_at	Acbd6	
1418992_at	F10	
1449164_at	Cd68	
1432262_at	4930504E06Rik	
1423547_at	Lyz2	
1453787_at	Txndc13	
1424083_at	Rod1	+
1460179_at	Dnaja1	
1428229_at	Prkd3	
1426084_a_at	Tor1aip1	
1419647_a_at	Ier3	+
1416265_at	Capn10	
1450479_x_at	Ptpn12	
1428892_at	Ppil1	
1426404_a_at	Rnf11	
1450399_at	Psen1	
1448304_a_at	Rab6	
1426692_at	Ccdc97	
1440825_s_at	Ccdc28a	
1448620_at	Fcgr3	
1422229_at	Dub2a	
1456567_x_at	Grn	
1427595_at	Acaca	
1434387_at	Itfg3	
1455156_at	Strn	
1418367_x_at	Hist2h2aa2	
1460552_at	Ascc3l1	
1417426_at	Srgn	
1428589_at	Mrpl41	
1451390_s_at	Zfand2b	
1417777_at	Ltb4dh	
1428588_a_at	Mrpl41	
1420652_at	Ate1	
1451599_at	Sesn2	
1437290_at	Impad1	
1427663_a_at	Clk4	
1449507_a_at	Cd47	

Probe set	Gene symbol	Also in mutant
1424084_at	Rod1	+
1418427_at	Kif5b	
1451421_a_at	Rogdi	
1452885_at	Sfrs2ip	
1429618_at	Cyld	
1416768_at	1110003E01Rik	
1418366_at	Hist2h2aa2	
1449209_a_at	Rdh11	
1454686_at	Hjurp	
1450056_at	Apc	
1416767_a_at	1110003E01Rik	
1434684_at	Rin3	
1449856_at	Rgs18	
1433780_at	Ubn1	
1426671_a_at	Rbm39	
1437287_at	1110020G09Rik	
1426377_at	Zfp281	
1428482_at	Akap10	
1429110_a_at	Nsun4	
1435959_at	Arhgap15	+
1419481_at	Sell	
1438629_x_at	Grn	
1418635_at	Etv3	
1428101_at	Rnf38	
1428130_at	Lman1	
1429326_at	Cenpl	
1420572_at	Ms4a3	
1447284_at	Trem1	
1460401_at	Edem3	+
1418199_at	Hemgn	
1415960_at	Mpo	
1420773_at	Dub2	
1428288_at	Klf9	
1420651_at	Ate1	
1418262_at	Syk	
1448020_at	Rap1a	
1424831_at	Cpne2	
1419004_s_at	Bcl2a1d	
1418261_at	Syk	
1416706_at	Rpe	

Probe set	Gene symbol	Also in mutant
1450730_at	Hs2st1	
1452481_at	Plcb2	
1451989_a_at	Mapre2	
1417876_at	Fcgr1	+
1418375_at	Mbd6	
1423837_at	2400003C14Rik	
1452034_at	Prepl	
1454707_at	2310035C23Rik	
1438040_a_at	Hsp90b1	
1418722_at	Ngp	
1424766_at	Ercc6l	
1423018_at	Kcna3	
1450547_x_at	Dub2a	
1424549_at	Degs2	
1424637_s_at	Ccdc47	
1451420_at	Ccdc47	
1455081_at	Txnl4b	
1450934_at	Eif4a2	
1431293_a_at	Cldnd1	
1424839_a_at	Nsun4	
1434653_at	Ptk2b	
1427879_at	1810031K17Rik	
1420519_a_at	Eral1	
1428771_at	2410127E18Rik	
1433659_at	Tubgcp4	
1429318_a_at	Qk	
1452504_s_at	Ctbs	
1417357_at	Emd	
1448131_at	Mfn2	
1449265_at	Casp1	
1426345_at	Prepl	
1422264_s_at	Klf9	
1452122_at	AI314180	
1419480_at	Sell	
1424779_at	Reep3	
1428230_at	Prkd3	
1435770_at	Txndc13	
1433697_at	Patl1	
1420820_at	2900073G15Rik	
1418929_at	Ift57	

Probe set	Gene symbol	Also in mutant
1421357_at	Gtf2a1	
1452648_at	Tbrg1	
1427951_s_at	Ccdc28a	
1449184_at	Pglyrp1	
1423388_at	Ap1g1	
1426346_at	Prepl	
1438383_x_at	Ppp2r1a	
1449537_at	Msh5	
1451738_at	Ogt	
1416679_at	Abcd3	
1421121_at	Akap10	
1444024_at	Ercc6l	
1423241_a_at	Tfdp1	
1420398_at	Rgs18	
1434790_a_at	Lta4h	
1434418_at	Lass6	
1450967_at	Ptplad2	+

Ordering is based on rank in GSEA analysis. Probe sets that are also present in the core enrichment signature of 32D-Cebpamut-ER cells are demarked in the third column.

Table 3. Transcription factor binding sites enriched in DNA fragments immunoprecipitated in 32D-Cebpa-mut-ER cells compared to 32D-ER cells.

Motif	Hits	Fold Change
M00771: ETS	514	1.84
M00028: HSF	1072	1.50
M00117: CEBPbeta	834	1.57
M00003: v-Myb	787	1.50
Elk-1	608	1.55
AML-1	420	1.70
E74A	442	1.65
M00025: Elk-1	334	1.75
M00339: c-Ets-1	112	2.59
M00032: c-Ets-1(p54)	322	1.65
Pax-4	248	1.73
M00769: AML	138	2.09
c-MYB_1	379	1.52
M00380: Pax-4	269	1.61
M00984: PEBP	113	1.96
M00227: v-Myb	177	1.64
M00016: E74A	59	2.26
M00259: STAT	17	4.98
Dorsal_1	90	1.80
M00249: CHOPCEBPalpha	104	1.66
M00457: STAT5A	66	1.86
M00772: IRF	105	1.59
M00281: RFX1	79	1.70

Enrichment for binding sites was assessed by CEAS software, which queries the TransFac and JASPAR databases. Enriched sites matching the TransFac database are denoted by their entry number (e.g. M00109 for the C/EBPbeta binding motif). Motifs with a P-value <10^{-5} are shown, ranked based on increasing P-value so that the most significant motif is on top.

clones and genes of which the promoter was occupied by the Cebpa-ER fusion (nominal P<0.001) (Figure 2B). Of those, 533 probe sets (367 unique gene names) contributed most to the association (core enrichment signature) (Table 2).

An insertion in the C/EBPα bZIP region leads to changes in DNA binding properties
We hypothesized that the core enrichment signature included a number of C/EBPα targets that are critical for normal neutrophilic development, and that one way to validate this would be a comparison with the leukemia-associated bZIP mutant. To this end, we investigated whether 32D-Cebpamut-ER cells differed in their DNA binding profile. We used the same protocol to derive ChIP profiles of 32D-Cebpa-mut-ER cells and again compared those to the profiles obtained from 32D-ER control cells. Only 787 fragments were enriched in

the stimulated 32D-Cebpamut-ER cells at $P<1\text{x}10^{-5}$ (FDR 13%). In these 787 fragments, a C/EBPβ binding site was enriched but many sites that had been found to be significantly overrepresented in the previous experiments in wt 32D-Cebpa-ER cells could not be detected (Table 3). Mapping of the 787 fragments to neighboring genes identified 135 RefSeq genes based on their location within 1 kb of the 5' start sites. The majority of these (110; 81%) overlapped with those found for wild type Cebpa-ER.

To answer whether in 32D-Cebpamut-ER cells target gene binding correlated with gene expression, we generated gene expression profiles. Integration of ChIP and gene expression data using GSEA indicated, similar to the situation in the wild type clones, a significant ($P<0.001$) association between expression changes and promoter occupancy, indicating that also in the Cebpamut-ER cells transcriptional activation of genes bound by the fusion took place. The corresponding core enrichment signature consisted of 75 expression probe sets

Table 4. Core enrichment genes unique for 32D-Cebpamut-ER cells.

Probe set	Gene symbol
1425332_at	Zfp106
1420174_s_at	Tax1bp1
1448399_at	Tax1bp1
1418967_a_at	St7
1433979_at	Rbms2
1422707_at	Pik3cg
1424733_at	P2ry14
1421198_at	Itgav
1432296_a_at	Itgav
1451584_at	Havcr2
1424927_at	Glipr1
1426905_a_at	Dnajc10
1426904_s_at	Dnajc10
1452230_at	Dnajc10
1421000_at	Cnot4
1420999_at	Cnot4
1420804_s_at	Clec4d
1456328_at	Bank1
1435230_at	Ankrd12
1443867_at	Ankrd12
1440193_at	Ankrd12
1420808_at	AC154532.3-202
1450006_at	AC154532.3-202

Ordering is based on rank in GSEA analysis.

(50 unique gene names), of which the majority (52 probe sets; 69%) overlapped with those in the core enrichment signature of wild type Cebpa-ER (Table 2). The 23 probe sets (15 unique gene names) that were exclusively found to associate with Cebpamut-ER, but not with wt Cebpa-ER, are listed in Table 4.

DISCUSSION

This study describes the application of ChIP-chip in a C/EBPα inducible myeloid cell line model. In the first part of the study, we investigated the genome-wide DNA binding profile of wild type C/EBPα, while we followed the same approach in the second part to study a particular AML associated C/EBPα mutant.

The experiments involving wild type C/EBPα have yielded a set of genes that we propose to be candidate C/EBPα target genes. Several observations suggest the general validity of the list we derived. First, binding of Cebpa-ER upon E_2 treatment revealed several known targets of C/EBPα. Also known targets of other C/EBP family members were found, which is in agreement with the fact that all C/EBP family members (with the exception of C/EBPζ (26)) bind similar sequence motifs (27). Second, computational analysis indicated a very significant enrichment of C/EBP binding sites. Finally, analysis of genome-wide gene expression profiles indicated that genes that were occupied by the fusion gene were significantly more likely to be upregulated than other genes. We used this integrative analysis of gene expression and ChIP-chip data as a filter to compile a final list of bona fide candidate targets (Table 2).

It is increasingly recognized that transcription factors act in complexes with other proteins. Genome-wide ChIP-chip offers the opportunity to analyze enrichment of transcription factor binding motifs, generating hypotheses regarding interacting proteins. Using these computational approaches followed by experimental validation, novel interactors of C/EBPβ and ER were recently identified (28-30). In our study, several DNA binding motifs that were enriched in the immunoprecipitated DNA fragments belong to proteins that have so far to our knowledge not been associated with C/EBPα. These included factors from the AP-2 helix-turn-helix family (AP-2γ, AP-2α)(31) and Spz1, a factor downstream of MAPK (32). Thus, these proteins represent interesting candidates for validation by assays that can demonstrate protein-protein interactions.

In the second part of the study, we studied a specific AML-related mutation in the C/EBPα bZIP domain. The effect of such mutations could be loss of normal C/EBPα function. However, it has not been studied yet what the overall impact on genome-wide DNA binding is. An alternative hypothesis could be that the mutant binds specific sequences that are not occupied by wild type, thereby providing a "gain of function". Broadly, our results appear more in support of the first rather than the second hypothesis. The majority of genes that were occupied by the mutant were also immunoprecipitated when analyzing the wild type protein

(e.g., 81% when only taking into account the 1 kb promoter associated genes). In addition, in both situations an association between occupancy and expression was found. Likewise, the transcription factor binding profile of mutant versus wild type primarily differed quantitatively, i.e. less sites, rather than qualitatively. Interestingly, one of these binding sites that was lost belongs to E2F. E2F is indeed known to directly bind C/EBPα in the mutated bZIP region (3). Of note, Western blot analysis indicated roughly equal expression levels of wild type and mutant protein in the respective 32D clones, suggesting that these differences could not be accounted for by differences in amounts of protein (data not shown).

It should be noted, however, that we did identify a small subset of genes that appeared specific for Cebpamut-ER (Table 4). In spite of the overall overlap in targeted genes, it is possible that this small subset of mutant-specific genes is important. This remains to be evaluated. Likewise, there was a small collection of mutant-specific binding sites.

An attractive approach to investigate the relevance of altered binding of mutant Cebpa would be proper studies in primary AML samples. We have made some initial attempts to correlate our lists of putative target genes with gene expression profiling data of a cohort of AML specimens. In particular, we hypothesized that the comparison of AMLs with *CEBPA* mutations versus AMLs with wild type *CEBPA* should identify deregulated direct targets showing overlap with our list of core enriched genes. Thus far, we have not succeeded in finding such correlations (data not shown). This may partly be due to the fact that leukemia is a multistep process, implying that the human leukemic blasts carry other genetic abnormalities that disturb the profile. In the second place, disruption of C/EBPα may in fact be common to the majority of AMLs. For instance, it may be transcriptionally or translationally downregulated, or its activity may be otherwise inhibited (33-35). The direct comparison of our bZIP profiles to AML is not straightforward either, as the majority of *CEBPA* mutant AMLs carries not only a bZIP abnormality, but also an N-terminal frameshift mutation on the other allele (3,34). These phenomena may thus hinder proper comparisons. It would be worthwhile to invest in performing not only GEP, but also ChIP-chip in the primary samples, to analyze directly which promoters are occupied.

It is important to realize that a number of limitations may apply to our model. First, the LBD of ERα can interact with co-activators as well as with other transcription factors (36). Induction of ERα with E_2 specifically recruits co-activators (36). Thus, our model may be overly biased towards transcriptional activation of target genes. Second, although the ERα LBD itself lacks DNA binding capacity, it may directly interact with DNA binding transcription factors (36), including C/EBPβ (37) and Sp1 (36,38,39). We used control cells that expressed the ERα LBD, but this protein lacked a nuclear localization signal.

In summary, these data support the applicability of ChIP-chip studies for the identification for target genes of C/EBPα, and present a list of candidates based on the integration of ChIP data with expression data. Validation of these findings in other models, such as alternative inducible models, or *Cebpa* knockout models (5,40) will now be required. Adjustments to

the model we have used, including the use of an ERα LBD control that carries a nuclear localization signal and use of other agents for stimulation of the fusion gene instead of E$_2$, e.g. tamoxifen, will offer opportunities to fine tune and optimize these results. To obtain a better understanding of C/EBPα bZIP mutants, it will be interesting to perform similar experiments with additional mutants in different positions of the bZIP domain, in agreement with what has been found in human AML.

ACKNOWLEDGEMENTS

REFERENCES

1. Rosenbauer F, Tenen DG. Transcription factors in myeloid development: balancing differentiation with transformation. Nat Rev Immunol 2007;7(2):105-17.

2. Friedman AD. Transcriptional control of granulocyte and monocyte development. Oncogene 2007;26(47):6816-28.

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

4. Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, et al. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. Nat Genet 2001;27(3):263-70.

5. Zhang DE, Zhang P, Wang ND, Hetherington CJ, Darlington GJ, Tenen DG. Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc Natl Acad Sci U S A 1997;94(2):569-74.

6. Zhang P, Iwama A, Datta MW, Darlington GJ, Link DC, Tenen DG. Upregulation of interleukin 6 and granulocyte colony-stimulating factor receptors by transcription factor CCAAT enhancer binding protein alpha (C/EBP alpha) is critical for granulopoiesis. J Exp Med 1998;188(6):1173-84.

7. Tenen DG, Hromas R, Licht JD, Zhang DE. Transcription factors, normal myeloid development, and leukemia. Blood 1997;90(2):489-519.

8. Wang W, Wang X, Ward AC, Touw IP, Friedman AD. C/EBPalpha and G-CSF receptor signals cooperate to induce the myeloperoxidase and neutrophil elastase genes. Leukemia 2001;15(5):779-86.

9. Ross SE, Radomska HS, Wu B, Zhang P, Winnay JN, Bajnok L, et al. Phosphorylation of C/EBPalpha inhibits granulopoiesis. Mol Cell Biol 2004;24(2):675-86.

10. Dong F, van Buitenen C, Pouwels K, Hoefsloot LH, Lowenberg B, Touw IP. Distinct cytoplasmic regions of the human granulocyte colony-stimulating factor receptor involved in induction of proliferation and maturation. Mol Cell Biol 1993;13(12):7774-81.

11. Johnson WE, Li W, Meyer CA, Gottardo R, Carroll JS, Brown M, et al. Model-based analysis of tiling-arrays for ChIP-chip. Proc Natl Acad Sci U S A 2006;103(33):12457-62.

12. Ji X, Li W, Song J, Wei L, Liu XS. CEAS: cis-regulatory element annotation system. Nucleic Acids Res 2006;34(Web Server issue):W551-4.

13. Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res 2004;32(Database issue):D91-4.

14. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 2003;31(1):374-8.

15. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4(2):249-64.

16. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:Article3.

17. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102(43):15545-50.

18. Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, Meijer J, van Oosterhoud S, van Putten WL, Valk PJ, et al. Biallelic mutations in the CEBPA gene and low CEBPA expression levels as prognostic markers in intermediate-risk AML. Hematol J 2003;4(1):31-40.

19. Ramji DP, Foka P. CCAAT/enhancer-binding proteins: structure, function and regulation. Biochem J 2002;365(Pt 3):561-75.

20. Kummalue T, Friedman AD. Cross-talk between regulators of myeloid development: C/EBPalpha binds and activates the promoter of the PU.1 gene. J Leukoc Biol 2003;74(3):464-70.

21. Porse BT, Pedersen TA, Xu X, Lindberg B, Wewer UM, Friis-Hansen L, et al. E2F repression by C/EBPalpha is required for adipogenesis and granulopoiesis in vivo. Cell 2001;107(2):247-58.

22. Khanna-Gupta A, Zibello T, Simkevich C, Rosmarin AG, Berliner N. Sp1 and C/EBP are necessary to activate the lactoferrin gene promoter during myeloid differentiation. Blood 2000;95(12):3734-41.

23. Lopez-Rodriguez C, Botella L, Corbi AL. CCAAT-enhancer-binding proteins (C/EBP) regulate the tissue specific activity of the CD11c integrin gene promoter through functional interactions with Sp1 proteins. J Biol Chem 1997;272(46):29120-6.

24. Richer E, Campion CG, Dabbas B, White JH, Cellier MF. Transcription factors Sp1 and C/EBP regulate NRAMP1 gene expression. Febs J 2008.

25. Lee YH, Williams SC, Baer M, Sterneck E, Gonzalez FJ, Johnson PF. The ability of C/EBP beta but not C/EBP alpha to synergize with an Sp1 protein is specified by the leucine zipper and activation domain. Mol Cell Biol 1997;17(4):2038-47.

26. Ron D, Habener JF. CHOP, a novel developmentally regulated nuclear protein that dimerizes with transcription factors C/EBP and LAP and functions as a dominant-negative inhibitor of gene transcription. Genes Dev 1992;6(3):439-53.

27. Friedman AD. Transcriptional regulation of granulocyte and monocyte development. Oncogene 2002;21(21):3377-90.

28. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, et al. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 2005;122(1):33-43.

29. Reed BD, Charos AE, Szekely AM, Weissman SM, Snyder M. Genome-wide occupancy of SREBP1 and its partners NFY and SP1 reveals novel functional roles and combinatorial regulation of distinct classes of genes. PLoS Genet 2008;4(7):e1000133.

30. Zhang Y, Liu T, Yan P, Huang T, Dewille J. Identification and characterization of CCAAT/Enhancer Binding proteindelta (C/EBPdelta) target genes in G0 growth arrested mammary epithelial cells. BMC Mol Biol 2008;9(1):83.

31. Eckert D, Buhl S, Weber S, Jager R, Schorle H. The AP-2 family of transcription factors. Genome Biol 2005;6(13):246.

32. Hsu SH, Hsieh-Li HM, Huang HY, Huang PH, Li H. bHLH-zip transcription factor Spz1 mediates mitogen-activated protein kinase cell proliferation, transformation, and tumorigenesis. Cancer Res 2005;65(10):4041-50.

33. Mueller BU, Pabst T. C/EBPalpha and the pathophysiology of acute myeloid leukemia. Curr Opin Hematol 2006;13(1):7-14.

34. Pabst T, Mueller BU. Transcriptional dysregulation during myeloid transformation in AML. Oncogene 2007;26(47):6829-37.

35. Koschmieder S, Halmos B, Levantini E, Tenen DG. Dysregulation of the C/EBP{alpha} Differentiation Pathway in Human Cancer. J Clin Oncol 2008.

36. Klinge CM. Estrogen receptor interaction with co-activators and co-repressors. Steroids 2000;65(5):227-51.

37. Dong J, Tsai-Morris CH, Dufau ML. A novel estradiol/estrogen receptor alpha-dependent transcriptional mechanism controls expression of the human prolactin receptor. J Biol Chem 2006;281(27):18825-36.

38. Porter W, Saville B, Hoivik D, Safe S. Functional synergy between the transcription factor Sp1 and the estrogen receptor. Mol Endocrinol 1997;11(11):1569-80.

39. Safe S, Kim K. Nonclassical Genomic ER/Sp and ER/AP-1 Signaling Pathways. J Mol Endocrinol 2008.

40. Zhang P, Iwasaki-Arai J, Iwasaki H, Fenyus ML, Dayaram T, Owens BM, et al. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. Immunity 2004;21(6):853-63.

CHAPTER 11

Summary and general discussion

SUMMARY

Since the introduction of molecular genetics, our knowledge about the abnormalities underlying human AML has increased tremendously. This is well reflected in the evolution that classification of the disease has undergone, from cytomorphology (according to the FAB system, introduced in the 1970s) to the recognition of entities of AML defined by gene mutations (WHO classification 2008). The main aim of this thesis was to further progress the molecular dissection of AML. Special attention throughout this thesis was given to abnormalities affecting *CEBPA*, the gene encoding the transcription factor CCAAT/enhancer binding protein alpha, one of the master regulators of normal myeloid differentiation. To address the research questions posed in the various studies, several genome-wide research techniques were applied, with a central role for GEP.

In chapter 2, we reviewed the use of genome-wide GEP in AML and described some of the common research questions that investigators have addressed using this technique, i.e. those related to class discovery, class prediction, outcome prediction and dissection of pathobiology. We also discussed limitations of the technology and we described promising future developments. Chapter 3 focused on the diagnostic potential of GEP-based class prediction of cytogenetically and genetically defined AML subgroups. These analyses indicated, in agreement with previous observations, that the chromosomal defects t(8;21), t(15;17) and inv(16) can be predicted with high accuracy, but that many other abnormalities do not result in distinctive classifiers. For *CEBPA* mutations, we observed two groups of AML cases: one that could be well predicted, and one that could not. This latter finding was studied in further detail in chapter 9. Together, these results confirmed that GEP can be applied in a diagnostic setting for AML, albeit at the moment only for a selected and defined number of subtypes.

In chapters 4-6, we investigated the molecular characteristics and pathobiology of a novel subset of acute leukemia. Chapter 4 described, through mouse modeling, a novel mechanism of inhibition of C/EBPα causing AML, i.e. elevated expression of *Trib2*. Trib2 was shown to bind C/EBPα protein, inducing its degradation. Analysis of our previously established gene expression data base of human AML samples highlighted an association of high *TRIB2* expression with a particular cluster characterized by *CEBPA* mutations. In contrast to the other cases in this cluster, the 6 leukemias that appeared particularly associated with high *TRIB2* levels did not exhibit *CEBPA* mutations. Additional experiments, described in chapters 5 and 6, uncovered distinctive characteristics of these specific leukemias, including, most notably, epigenetic silencing of *CEBPA* expression and an overall DNA methylation profile characterized by hypermethylation. These molecular features were accompanied by an immature, mixed myeloid/T-lymphoid immunophenotype.

In chapters 7 through 9, we studied AMLs with mutations in the *CEBPA* gene. In chapter 7, it was found that an insertion in the second transactivation domain of C/EBPα (HP196-197ins) is a common polymorphism and not an AML-specific mutation. Chapter 8 described

that homozygosity of *CEBPA* mutations is frequently associated with copy number neutral loss-of-heterozygosity of the chromosomal region that includes *CEBPA*, presumably through mitotic recombination. Chapter 9 presented previously unrecognized heterogeneity of *CEBPA* mutant AML. While it was already known that there are AMLs in which both *CEBPA* alleles are affected as well as AMLs with only a single heterozygous aberration, it was not known whether the reported prognostic impact of *CEBPA* mutations, i.e. favorable, was equal in both groups. Our analyses suggested that only AML patients with two mutations in the *CEBPA* gene, presumably affecting both alleles, should be considered to have favorable risk. In support of the prognostic differences, single and double mutant specimens also strikingly differed in genome-wide transcriptional profiles.

In the final chapter, chapter 10, we identified putative novel target genes of C/EBPα in a myeloid cell line model using ChIP-chip. We performed similar experiments with a variant C/EBPα protein that carries a C-terminal insertion mutation in the bZIP region resembling human AML. Our preliminary analysis of these data demonstrated impaired DNA binding of the mutant, providing possible leads to an understanding of the mechanism involved in human leukemogenesis by the mutant protein.

GENERAL DISCUSSION AND DIRECTIONS FOR FUTURE RESEARCH

The work presented in the thesis was divided into four sections, each covering one or more chapters. The results described in the four sections will here be discussed in an integrated way, with a particular focus on possible directions for future research.

I - Gene expression profiling in acute myeloid leukemia

It has been shown in several previous studies that genome-wide GEP can be used to predict molecularly defined subtypes of AML (1-3). A pending question of clinical importance is whether GEP-based classifiers can truly replace existing diagnostic techniques, such as cytogenetics or PCR-based techniques, or whether GEP analyses remain mainly of scientific interest. To be able to make such a statement, proper validation in independent cohorts of AML is required. In chapter 3, we performed one such validation study in retrospectively analyzed samples. Predictive gene expression signatures for the most common recurrent (cyto) genetic abnormalities in AML were derived on a training cohort, and validated on a large independent patient group. The results confirmed previous observations that three specific chromosomal subtypes of AML, i.e. those characterized by t(8;21), inv(16) and t(15;17), are predictable with near-perfect accuracy. Mutations in *CEBPA* (see also chapter 9) and *NPM1* were also predictable, but more mistakes were made. The other aberrations under study were hardly associated with distinctive gene expression profiles. Perhaps not unexpectedly, results obtained by us and others show that abnormalities directly involving transcriptional

regulators are generally associated with more recognizable profiles than other defects. For instance, mutations in genes encoding signaling factors (*RAS, FLT3*) do not associate with very "hard" signatures. Still it is possible that better classifiers can be derived for the latter group, using improved technical and/or analytical approaches. Some of these improvements were discussed in our review of GEP in chapter 2. Generally, a next step before incorporating GEP as a diagnostic approach for AML will be to validate our and others findings in prospective studies.

Of note, there has also been interest in the oncology field to derive gene expression signatures that predict patient outcome (4). Such analyses can be performed on outcome as a binary indicator – for instance: "remission versus no remission", or "dead versus alive". A frequently used alternative is to choose time, e.g. overall survival time, as an endpoint. In that case, adjusted algorithms need to be applied that can handle censored time, for instance modified Cox's regression algorithms (5). Our preliminary results, not presented in the current thesis, indicate that it is indeed possible to derive predictive gene expression signatures for outcome in AML (data not shown). This is consistent with observations from other researchers (6-8). An important clinical consideration is whether such predictive classifiers add prognostic information to already known prognostic parameters. Aren't these gene expression predictors just fancy and expensive ways to detect information that is already reflected in clinical (e.g. age) or molecular (e.g. mutations) parameters? Indeed, it seems that a considerable amount of the prognostic information in such predictors is highly associated with, for instance, the presence of *FLT3*-ITD (7,8). This aspect will clearly need to be further studied. Particular efforts should be directed towards developing predictive algorithms that combine the best of both worlds, i.e. established and novel markers. It is important, however, to realize that even if GEP-based predictors do not "add" information, they may in the end be preferable over currently used methods, as GEP classifiers can combine several tests into one.

II – Identification of a new subtype of leukemia

Another application of GEP is the identification of previously unrecognized subgroups. The particular challenge is to identify new subtypes of AML that are "true" and not artifacts induced by the inherent tendency of cluster algorithms to find similarities between samples. Thus, one would like to find common features apart from the GEP data (e.g., common molecular aberrations or clinical responses) in order to truly designate a group of samples with a similar gene expression profile as "subtype". Our experiments in chapters 4 through 6 were in agreement with this additional requirement, as we not only found a specific gene expression profile, but also certain immunophenotypical, genetic and epigenetic attributes. In fact, the leukemias demonstrated at least 6 frequent molecular characteristics: 1) gene expression profiles resembling those of *CEBPA* mutant AML, 2) *CEBPA* mRNA silencing associated with DNA promoter methylation, 3) a genome-wide pattern of DNA hypermethylation when compared to sorted normal CD34+ control cells as well as to *CEBPA* mutant AML, 4) frequent

mutations in *NOTCH1*, 5) overexpression of the *TRIB2* gene, and 6) an immunophenotype characterized by immature hematopoietic cell characteristics (CD34) in combination with myeloid as well as (T-)lymphoid antigens.

There are questions still to be addressed. In the first place: which of the above characteristics are driving causes, and which are downstream effects? Is *CEBPA* silencing a defining event in transformation? In spite of the fact that the leukemias were initially picked up as displaying a "*CEBPA* mutant" expression profile, it is possible that *CEBPA* silencing, associated with promoter methylation, is just one of many results of genome-wide DNA hypermethylation. A possibility to investigate whether the absence of *CEBPA* mRNA truly can induce the leukemic characteristics is to utilize *Cebpa* knockout mouse models or to perform knockdown experiments in vivo or in vitro. In chapter 5, we found that knock out of *Cebpa* from murine hematopoietic stem cells is sufficient to upregulate certain T-cell transcripts. This implies that the partial T-lymphoid phenotype is at least to a certain extent the result of *CEBPA* silencing. Genome-wide methylation studies in the same mouse model are required to elucidate whether removal of *Cebpa* can also induce global DNA methylation changes.

Likewise, it remains to be addressed whether the mixed myeloid/T-lymphoid phenotype of these leukemia cases is the result of aberrant T-cell differentiation in a committed myeloid progenitor cell – due to activating *NOTCH1* mutations and/or silencing of *CEBPA* – or (also) a reflection of an immature progenitor cell that has been transformed. Of note, the last option would fit with recent reports describing the existence of T-lymphoid progenitor cells that have retained myeloid potential (9). Thus, B-lineage potential may in certain cases be lost before myeloid potential, giving rise to a common myeloid/T-lymphoid progenitor (10,11).

One could hypothesize that there is a common underlying aberration that we have not been able to identify yet. High-resolution SNP-based assays, or alternatively, a candidate-based approach, could be used to find such an abnormality. One such candidate is the *PAX5* gene, which plays a crucial role in normal B-cell development. Heterozygous disruption of this gene has recently been implicated as a frequent abnormality in B-ALL (12). Interestingly, there is experimental evidence that homozygous deletion of the *Pax5* gene in committed B-lymphoid progenitor cells in mice leads to a complete block of B-cell development, and in fact induces dedifferentiation to uncommitted progenitors (13). Strikingly, these cells may already have undergone immunoglobulin heavy- and light-chain gene rearrangements, which can still be detected after dedifferentiation (13). These findings underscore the increasingly recognized plasticity of hematopoietic progenitors that can be influenced by deletion or ectopic expression of lineage instructive transcription factors (14).

From a clinical perspective, a question that we have not been able to address yet is how frequent these leukemias are and whether these patients have a specific prognostic profile. For such analyses, more patients will need to be analyzed. In the light of the mixed immunophenotypes, it is certainly a possibility that one should not restrict to cohorts of AML, but also study T-ALL cases. We therefore propose to interrogate T-ALL gene expression databases for

these analyses as well. Similarly, investigation of additional series of AML is warranted. Comparisons between AML and T-ALL derived samples may shed light on the question whether these leukemia patients are best treated according to AML or ALL protocols.

III – *CEBPA* mutations in AML

In three chapters (7 through 9), we addressed questions related to nucleotide sequence mutations in the *CEBPA* gene. In 2001, the first study on *CEBPA* mutations was published, and reported a predominant presence of truncating mutations in the N-terminus (15). Studies since then have revealed that N-terminal and C-terminal bZIP mutations in *CEBPA* are approximately equally prevalent (16,17). However, in some studies sequence variations in other parts of the gene were described as well (18). These alternative sequence variations included an in-frame insertion in the second transactivation domain, leading to a two-amino-acid duplication in a histidine-proline rich stretch of the protein (HP196-197ins) (18,19). As this variation does not cause changes in the protein that are typically associated with AML, i.e. either truncation in the N-terminus or in-frame changes in the DNA binding/dimerization domain, we hypothesized that it might represent a functionally irrelevant variation. Indeed, in one earlier study, the insertion was found in genomic DNA from a number of healthy volunteers as well, suggesting it to be a polymorphism (20). Results presented in chapter 7 confirmed that HP196-197ins should not be considered a mutation, but a polymorphism, as the variation was found in cDNA from several individuals without leukemia. This is of particular interest because *CEBPA* mutations, but not polymorphisms, are regarded of prognostic value. The validity of this conclusion was further corroborated by other research groups (21-24). These findings emphasize that not all nucleotide sequence variations in cancer-related genes are necessarily mutations, even if they cause changes in the predicted protein sequence. It seems of increasing importance to be aware of this phenomenon. Genome-wide sequencing is now starting to be widely performed for the study of malignancies, including AML (25-28). The goal of these efforts is to identify new cancer genes. Performing such massive parallel sequencing, a challenge will be to discriminate functionally important changes from unimportant sequence variations that could either be polymorphism or functionally irrelevant "passenger" mutations (29). This discrimination can be made by comparison to germ line DNA from the same individual or, similar to our approach, to genomic DNA from sufficient healthy controls. A perhaps more elegant, but obviously much more time-consuming and expensive approach is to perform functional studies in vivo or in vitro to test the oncogenic potential of newly identified variants (30).

In chapter 8, we found loss of heterozygosity of the *CEBPA* locus, without changes in copy number, in 3 AML patients with homozygous *CEBPA* mutations. This study confirmed a previous observation in a single AML case (31) that UPD of *CEBPA* mutations can be found in AML. More broadly, UPD of several AML related genes has now been documented and may be found as acquired abnormality in relapsed disease (31-34). Of note, for *CEBPA* we

analyzed a fourth AML sample with a homozygous mutation (patient #6376, which can also be found in chapter 9) but did not detect LOH nor deletion (FISH). It remains elusive what the mechanism of homozygosity in this patient is. Another question that needs to be addressed is whether the initial heterozygous mutation in patients that later show UPD is already present in the germ line. *CEBPA* mutations have now been reported as germ line abnormalities by several groups (35-38), and a second mutation in *CEBPA* may contribute to the development of leukemia (38). It is conceivable that UPD provides an opportunity to the (pre-)leukemic cell to gain a second hit by removal of the wild type allele.

Chapter 9 provided new insights into the prognostic implications of *CEBPA* mutations for AML patients. These findings may have important clinical consequences once validated. In 2008, an updated WHO classification of hematological and lymphoid malignancies was released (39). In the new AML classification, *CEBPA* mutant AML has been added as a provisional entity. A major rationale behind this is the observation, also reproduced by us, that *CEBPA* mutations appear virtually mutually exclusive with other abnormalities that are recognized as distinct entities by the WHO, such as the three major good risk translocations (39). Thus, *CEBPA* mutant AML may be regarded a distinct genetic AML entity, according to the WHO. A second important rationale for the WHO to recommend *CEBPA* mutation testing is the favorable prognostic impact of these mutations. It will be important to assess whether our findings reported in chapter 9, i.e. significant differences in outcome as well as gene expression profiles between *CEBPA^single-mut* and *CEBPA^double-mut* AMLs, can be reproduced in independent patient series. If that will be the case, we suggest to refine the recent WHO guidelines for classification, and to consider only patients with double *CEBPA* mutations as a distinctive AML class. It is obvious that it will also need to be assessed in larger series what is the relative prognostic influence of additional mutations, such as *FLT3*-ITD, in *CEBPA* mutant AML.

Why would two mutations in *CEBPA* provide a relatively curable disease, while one heterozygous mutation – retaining the wild type allele – would not? This question is especially intriguing taking the indispensable role of C/EBPα in myelopoiesis in consideration. One could argue, according to the two-hit leukemogenesis model, that a single *CEBPA* mutation requires a cooperating hit, and that this second hit – e.g. *FLT3*-ITD – is the one that determines the final, frequently poor outcome. This conceivable hypothesis does not fully explain why a second *CEBPA* mutation would be more favorable than an alternative additional hit such as *FLT3*-ITD. Whether this is something related to the differentiation stage during malignant transformation or whether other factors are involved remains elusive. Another conclusion we derived from chapter 9 is that, besides testing for the mutations themselves, a highly sensitive and specific alternative is the use of a gene expression classifier. The advantage of the latter approach is that it can be easily incorporated into diagnostic gene expression chips that are currently being developed, and that allow the concurrent assessment of various abnormalities – see also chapter 3.

IV – Identification of target genes of C/EBPα in myeloid cells

In this thesis, genome-wide research tools were predominantly used to study genomic DNA or mRNA directly derived from AML patient material. In contrast, chapter 10 described the use of genome-wide DNA microarrays in an experimental setting, i.e. chromatin immuno-precipitation on microarrays (ChIP-chip) in an in vitro model. The main research question addressed in this chapter was: which genes are directly bound and regulated by C/EBPα? A second question was whether, and to what extent, the derived profiles differed between wild-type and bZIP mutant C/EBPα. To this end, ChIP-chip was used in an adapted inducible cell line model for C/EBPα that has previously been successfully used in many other laboratories (40,41). The model relies on a fusion between C/EBPα and the ligand binding domain of the estrogen receptor alpha. The fusion protein resides in the cytoplasm and only translocates to the nucleus once a proper ligand, e.g. beta-estradiol, is bound to ER, which can be visualized by immunohistochemistry (data not shown). We used this model to identify direct target genes of C/EBPα. An overlay of the ChIP-chip derived data with gene expression profiling data performed in the same cell line model indicated a significant overlap in genes, which we believe are strong candidates for validation as critical C/EBPα targets. In contrast to the wild type protein, the mutant we studied showed a strongly reduced ability to bind DNA as well as to induce gene transcription.

It will be important to assess the relevance of these findings for human *CEBPA* mutant AML. To this end, bioinformatical comparisons between expression data from human AML and our list of candidate target genes seem appropriate. Our initial efforts did not pinpoint a very strong association, however, which may be due to the more complex nature of primary leukemias as opposed to model systems. Specifically, cooperating mutations may affect the analysis. Of possible further importance, it has recently been suggested that there are func-tional differences between the roles of C/EBPα and mutations in its gene in human versus murine development and leukemia (42,43). Thus, when assessing the results obtained in in vitro models such as the one we have used or mouse models of other investigators (44), it re-mains of importance to be aware of the fact that the human situation may not be completely similar.

Nevertheless, the significant progress that has been made in the understanding of the pathobiology of AML has been derived to an important extent from such modeling studies. The proper use of techniques that allow genome-wide measurements in these models, such as ChIP-chip, will undoubtedly contribute to further insights. As these approaches can in principle speed up the identification of critical leukemogenic hits, they may be instrumental in paving the way towards the development of novel targeted therapies for specific subtypes of AML.

REFERENCES

1. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350(16):1617-28.

2. Haferlach T, Kohlmann A, Schnittger S, Dugas M, Hiddemann W, Kern W, et al. Global approach to the diagnosis of leukemia using gene expression profiling. Blood 2005;106(4):1189-98.

3. Ross ME, Mahfouz R, Onciu M, Liu HC, Zhou X, Song G, et al. Gene expression profiling of pediatric acute myelogenous leukemia. Blood 2004;104(12):3679-87.

4. van't Veer LJ, Bernards R. Enabling personalized cancer medicine through analysis of gene-expression patterns. Nature 2008;452(7187):564-70.

5. Bovelstad HM, Nygard S, Storvold HL, Aldrin M, Borgan O, Frigessi A, et al. Predicting survival from microarray data--a comparative study. Bioinformatics 2007;23(16):2080-7.

6. Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004;350(16):1605-16.

7. Metzeler KH, Hummel M, Bloomfield CD, Spiekermann K, Braess J, Sauerland MC, et al. An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. Blood 2008;112(10):4193-201.

8. Radmacher MD, Marcucci G, Ruppert AS, Mrozek K, Whitman SP, Vardiman JW, et al. Independent confirmation of a prognostic gene-expression signature in adult acute myeloid leukemia with a normal karyotype: a Cancer and Leukemia Group B study. Blood 2006;108(5):1677-83.

9. Wada H, Masuda K, Satoh R, Kakugawa K, Ikawa T, Katsura Y, et al. Adult T-cell progenitors retain myeloid potential. Nature 2008;452(7188):768-72.

10. Graf T. Immunology: blood lines redrawn. Nature 2008;452(7188):702-3.

11. Bell JJ, Bhandoola A. The earliest thymic progenitors for T cells possess myeloid lineage potential. Nature 2008;452(7188):764-7.

12. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature 2007;446(7137):758-64.

13. Cobaleda C, Jochum W, Busslinger M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. Nature 2007;449(7161):473-7.

14. Cobaleda C, Busslinger M. Developmental plasticity of lymphocytes. Curr Opin Immunol 2008;20(2):139-48.

15. Pabst T, Mueller BU, Zhang P, Radomska HS, Narravula S, Schnittger S, et al. Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. Nat Genet 2001;27(3):263-70.

16. Pabst T, Mueller BU. Transcriptional dysregulation during myeloid transformation in AML. Oncogene 2007;26(47):6829-37.

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

18. Leroy H, Roumier C, Huyghe P, Biggio V, Fenaux P, Preudhomme C. CEBPA point mutations in hematological malignancies. Leukemia 2005;19(3):329-34.

19. Frohling S, Schlenk RF, Stolze I, Bihlmayr J, Benner A, Kreitmeier 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 2004;22(4):624-33.

20. Lin LI, Chen CY, Lin DT, Tsay W, Tang JL, Yeh YC, et al. Characterization of CEBPA mutations in acute myeloid leukemia: most patients with CEBPA mutations have biallelic mutations and show a distinct immunophenotype of the leukemic cells. Clin Cancer Res 2005;11(4):1372-9.

21. Resende C, Regalo G, Duraes C, Carneiro F, Machado JC. Genetic changes of CEBPA in cancer: mutations or polymorphisms? J Clin Oncol 2007;25(17):2493-4; author reply 4-5.

22. Biggio V, Renneville A, Nibourel O, Philippe N, Terriou L, Roumier C, et al. Recurrent in-frame insertion in C/EBPalpha TAD2 region is a polymorphism without prognostic value in AML. Leukemia 2008;22(3):655-7.

23. Leecharendkeat A, Tocharoentanaphol C, Auewarakul CU. CCAAT/enhancer binding protein-alpha polymorphisms occur more frequently than mutations in acute myeloid leukemia and exist across all cytogenetic risk groups and leukemia subtypes. Int J Cancer 2008;123(10):2321-6.

24. Fuchs O, Provaznikova D, Kocova M, Kostecka A, Cvekova P, Neuwirtova R, et al. CEBPA polymorphisms and mutations in patients with acute myeloid leukemia, myelodysplastic syndrome, multiple myeloma and non-Hodgkin's lymphoma. Blood Cells Mol Dis 2008;40(3):401-5.

25. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 2008;456(7218):66-72.

26. Campbell PJ, Stephens PJ, Pleasance ED, O'Meara S, Li H, Santarius T, et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. Nat Genet 2008;40(6):722-9.

27. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature 2008;455(7216):1069-75.

28. Ley TJ, Minx PJ, Walter MJ, Ries RE, Sun H, McLellan M, et al. A pilot study of high-throughput, sequence-based mutational profiling of primary human acute myeloid leukemia cell genomes. Proc Natl Acad Sci U S A 2003;100(24):14275-80.

29. Futreal PA. Backseat drivers take the wheel. Cancer Cell 2007;12(6):493-4.

30. Frohling S, Scholl C, Levine RL, Loriaux M, Boggon TJ, Bernard OA, et al. Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. Cancer Cell 2007;12(6):501-13.

31. Fitzgibbon J, Smith LL, Raghavan M, Smith ML, Debernardi S, Skoulakis S, et al. Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. Cancer Res 2005;65(20):9152-4.

32. Gorletta TA, Gasparini P, D'Elios MM, Trubia M, Pelicci PG, Di Fiore PP. Frequent loss of heterozygosity without loss of genetic material in acute myeloid leukemia with a normal karyotype. Genes Chromosomes Cancer 2005;44(3):334-7.

33. Raghavan M, Smith LL, Lillington DM, Chaplin T, Kakkas I, Molloy G, et al. Segmental uniparental disomy is a commonly acquired genetic event in relapsed acute myeloid leukemia. Blood 2008;112(3):814-21.

34. Raghavan M, Lillington DM, Skoulakis S, Debernardi S, Chaplin T, Foot NJ, et al. Genome-wide single nucleotide polymorphism analysis reveals frequent partial uniparental disomy due to somatic recombination in acute myeloid leukemias. Cancer Res 2005;65(2):375-8.

35. Smith ML, Cavenagh JD, Lister TA, Fitzgibbon J. Mutation of CEBPA in familial acute myeloid leukemia. N Engl J Med 2004;351(23):2403-7.

36. Renneville A, Mialou V, Philippe N, Kagialis-Girard S, Biggio V, Zabot MT, et al. Another pedigree with familial acute myeloid leukemia and germline CEBPA mutation. Leukemia 2008.

37. Sellick GS, Spendlove HE, Catovsky D, Pritchard-Jones K, Houlston RS. Further evidence that germline CEBPA mutations cause dominant inheritance of acute myeloid leukaemia. Leukemia 2005;19(7):1276-8.

38. Pabst T, Eyholzer M, Haefliger S, Schardt J, Mueller BU. Somatic CEBPA mutations are a frequent second event in families with germline CEBPA mutations and familial acute myeloid leukemia. J Clin Oncol 2008;26(31):5088-93.

39. Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, et al., editors. WHO classification of tumours of haematopoietic and lymphoid tissues: IARC: Lyon; 2008.

40. Umek RM, Friedman AD, McKnight SL. CCAAT-enhancer binding protein: a component of a differentiation switch. Science 1991;251(4991):288-92.

41. Ross SE, Radomska HS, Wu B, Zhang P, Winnay JN, Bajnok L, et al. Phosphorylation of C/EBPalpha inhibits granulopoiesis. Mol Cell Biol 2004;24(2):675-86.

42. Schwieger M, Lohler J, Fischer M, Herwig U, Tenen DG, Stocking C. A dominant-negative mutant of C/EBPalpha, associated with acute myeloid leukemias, inhibits differentiation of myeloid and erythroid progenitors of man but not mouse. Blood 2004;103(7):2744-52.

43. Niebuhr B, Iwanski GB, Schwieger M, Roscher S, Stocking C, Cammenga J. Investigation of C/EBPalpha function in human (versus murine) myelopoiesis provides novel insight into the impact of CEBPA mutations in acute myelogenous leukemia (AML). Leukemia 2008.

44. Kirstetter P, Schuster MB, Bereshchenko O, Moore S, Dvinge H, Kurz E, et al. Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. Cancer Cell 2008;13(4):299-310.

NEDERLANDSE SAMENVATTING

Door de grote vorderingen die de afgelopen jaren geboekt zijn op het gebied van de moleculaire genetica is onze kennis over de afwijkingen die samenhangen met acute myeloïde leukemie (AML) sterk gegroeid. Een goed voorbeeld daarvan is onze veranderende kijk op de beste manier om de ziekte klinisch te classificeren. Tientallen jaren vormde vooral microscopische beoordeling van leukemiecellen (morfologie) de basis voor classificatie, naar het systeem dat werd geïntroduceerd door de French-American-British (FAB) groep in de jaren zeventig. In 2008 kwam de World Health Organization (WHO) met een nieuwe versie van haar classificatiesysteem, en daarin staan voor het eerst ook bepaalde subtypes van AML die primair gedefinieerd worden door genmutaties. Het doel van dit proefschrift was om de moleculaire kennis over humane AML verder te vergroten. Daarbij stonden vooral afwijkingen in *CEBPA* centraal, het gen dat codeert voor de myeloïde transcriptiefactor CCAAT/enhancer binding protein alpha. Bij de beantwoording van de diverse onderzoeksvragen gebruikten we met name onderzoekstechnieken waarmee bepalingen op genoombreed niveau gedaan kunnen worden.

In hoofdstuk 2 gaven we een overzicht van het gebruik van genexpressie-analyses op genoombreed niveau (in het Engels "gene expression profiling", verder afgekort als GEP), door middel van zogenaamde DNA microarray chips. We beschreven de meest voorkomende redenen om deze methodiek te gebruiken: het ontdekken van nieuwe subtypes van AML, het voorspellen van bekende subgroepen, het voorspellen van prognose en het ophelderen van biologische vraagstukken. We bespraken ook beperkingen van de techniek en gaven een beschouwing over mogelijke toekomstige ontwikkelingen op dit gebied. Hoofdstuk 3 was gericht op het diagnostisch gebruik van GEP, oftewel voorspelling van bekende chromosomale of genetische afwijkingen op basis van genexpressieprofielen. De resultaten in dat hoofdstuk bevestigden enerzijds de bevindingen uit eerdere studies dat de chromosomale translocaties t(8;21), t(15;17) en inv(16) met grote nauwkeurigheid kunnen worden voorspeld. Anderzijds concludeerden we dat veel andere afwijkingen niet geassocieerd zijn met karakteristieke genexpressieprofielen. De groep met *CEBPA* mutaties viel in tweeën uiteen: een deel kon zeer goed voorspeld worden, een ander deel juist niet. Dit gegeven hebben we verder onderzocht in hoofdstuk 9. Al met al bevestigde deze studie dat GEP gebruikt kan worden voor de diagnostiek van AML, maar hij toonde ook aan dat dat op dit moment alleen geldt voor een zeer selecte groep van AML subtypes.

In hoofdstukken 4, 5 en 6 onderzochten we de moleculaire kenmerken en afwijkingen van een niet eerder beschreven subtype van acute leukemie. In hoofdstuk 4 gebruikten we een muizenmodel om aan te tonen dat verhoogde expressie van het *Trib2* gen leidt tot remming van C/EBPα en uiteindelijk vaak tot AML. Dit bleek veroorzaakt te worden door directe binding van Trib2 aan C/EBPα, dat vervolgens afgebroken wordt. In onze eerder opgebouwde database met genexpressiedata van 285 AML patienten bleek hoge expressie van *TRIB2*

vooral geassocieerd te zijn met een specifiek cluster van AMLs waarin veel *CEBPA* mutaties voorkomen. Opvallend genoeg hing *TRIB2* expressie juist vooral samen met de paar gevallen in deze groep waarin géén *CEBPA* mutaties gevonden waren. Aanvullende experimenten, die werden beschreven in hoofdstukken 5 en 6, lieten zien dat deze leukemieën specifieke kenmerken hebben. De meest in het oog springende daarvan zijn uitschakeling van *CEBPA* door DNA methylering en ook op veel andere plaatsen in het genoom een DNA methyleringspatroon dat wordt gekenmerkt door verhoogde methylering. Daarnaast brengen de leukemiecellen zowel myeloïde als T-lymphoïde markers tot expressie.

In de hoofdstukken 7 tot en met 9 onderzochten we mutaties in het *CEBPA* gen. Hoofdstuk 7 beschreef een insertie van 6 nucleotiden in het deel van het gen dat codeert voor het tweede transactivatiedomein. In tegenstelling tot sommige eerdere studies vonden we dat deze variatie geen echte mutatie is, maar een polymorfisme dat ook bij mensen zonder leukemie gevonden kan worden. In hoofdstuk 8 beschreven we dat homozygote *CEBPA* mutaties vaak geassocieerd zijn met verlies van heterozygositeit van een groot gedeelte van het omliggende chromosoom. Dat is een voorbeeld van "segmental uniparental disomy", een fenomeen dat meestal optreedt door recombinatie tijdens mitotische delingen. In hoofdstuk 9 ontdekten we dat er binnen de groep van AML patienten met *CEBPA* mutaties onderscheid gemaakt moet worden tussen gevallen waarin slechts één heterozygote mutatie aanwezig is (enkelmutanten), en gevallen waarbij twee mutaties of homozygote mutaties gevonden zijn (dubbelmutanten). We concludeerden dat op basis van het feit dat in onze patiëntengroep alleen dubbelmutanten geassocieerd waren met een gunstig ziektebeloop. Daarnaast bleek die groep, die het grootste deel van de *CEBPA* mutante AMLs omvat, een karakteristiek genexpressiepatroon te vertonen, terwijl dat niet het geval was voor de enkelmutanten.

In het laatste hoofdstuk, hoofdstuk 10, voerden we een studie uit om nieuwe genen te vinden die direct, op DNA-niveau, gebonden worden door C/EBPα en daardoor gereguleerd worden. We deden dit door middel van de ChIP-chip techniek in een cellijnmodel waarin C/EBPα desgewenst geactiveerd kan worden. Daarnaast voerden we soortgelijke experimenten uit met een bZIP mutant C/EBPα. Onze data wekken de suggestie dat de mutant minder sterk bindt aan DNA dan het fysiologische eiwit. Mogelijk biedt dit aanknopingspunten om beter te begrijpen hoe zo'n mutatie bijdraagt aan het ontstaan van AML.

DANKWOORD

Tijdens het afronden van een proefschrift is er stress, maar er is ook een zekerheid: uiteindelijk wordt alleen het dankwoord echt gelezen. Dat levert dan overigens wel weer nieuwe onrust op, want het dankwoord moet dus wel compleet zijn! Er hebben de afgelopen jaren echter zo veel mensen bijgedragen aan dit boekwerk, dat het ondoenlijk is om een ieder te noemen. Hier een poging om een aantal mensen specifiek te bedanken.

De meeste dank ben ik verschuldigd aan mijn promotor en co-promotor. Bob Löwenberg, beste Bob, dank voor de gelegenheid om te werken op zo'n prettige en succesvolle afdeling, voor het vermogen om altijd kritisch te blijven en voor de hulp bij het uitstippelen van toekomstplannen. Ruud Delwel, beste Ruud, het was een voorrecht om ruim vier jaar bij je te mogen werken. Je combineert een continue stortvloed aan creatieve ideeën met een prettige manier van omgaan met mensen en een groot gevoel voor humor waar je gelukkig ook zelf hard om kunt lachen. Ik kijk er naar uit om te blijven samenwerken.

Professor Sjaak Philipsen, professor Rob Pieters en dr. Jan Cools dank ik voor het completeren van de kleine commissie.

Peter Valk is een van de grondleggers van de AML microarray dataset die ik naar hartelust heb geplunderd, en bovendien lid van de grote commissie. Peter, je hebt een zeer grote bijdrage geleverd aan dit proefschrift, bedankt! Als er zoiets bestaat als een ideale analiste, dan is Claudia Erpelinck het – dit boekje was aanzienlijk leger geweest zonder haar. Er heerste altijd een goede werksfeer in de Delwelgroep, waarvoor ik oud- en nieuwgedienden Fokke Lindeboom, Dominik Spensberger, Eric van den Akker, Antoinette van Hoven, Irene Louwers, Eric Bindels, Sanne Lugthart, Erdogan Taskesen, Marije Havermans en Lucy Solerno dank. Ook verschillende studenten hebben aan die goede sfeer bijgedragen en bovendien meegeholpen aan deelprojecten: Terence Tsie Chin Jong, Clara Koss, Timurs Maculins en Diana Hamer. Roel Verhaak, GEPper van het eerste uur, begeleidde mijn eerste schreden in de wondere wereld van de bioinformatica. Ook Mathijs Sanders was van waarde bij diverse bioinformatische exercities.

Naarmate het onderzoek vorderde, verplaatste ik mij meer richting computer en dus ook richting kamergenoten in Ee1330B. Karishma Palande, Andrzej Nieradka, Judith Oldenampsen en Saman Abbas hebben dit met verve doorstaan, zelfs toen die computer in toenemende mate onduidelijke herrie begon te verspreiden.

Het is moeilijk om Ivo Touw hier te vergeten. Dat komt weliswaar gedeeltelijk door zijn vocale aanwezigheid naast kamer Ee1330B, maar zeker ook door zijn niet aflatende enthousiasme en kritische noten bij werkbesprekingen. Ivo, het is een eer om ook jou in de grote commissie te hebben. Nog niet genoemde (oud-)collega's van de 13e verdieping, met name Marieke von Lindern, Mojca Jongen-Lavrencic, Stefan Erkeland, Suming Sun, Menno Dijkstra, Renee Beekman, Justine Peeters, Mahban Irandoust, Bart Aarts, Judith Gits, Marijke Valkhof, Onno Roovers en dames en heren van moleculaire diagnostiek en BMT, bedankt voor jullie hulp en

voor de leuke tijd in het lab! De prima secretariële ondersteuning op de afdeling hematologie mag niet onvermeld blijven – vooral Ans Mannens, Jeanne Vlasveld en Nathalie van Dommelen waren erg behulpzaam tijdens de laatste periode. Egied Simons verrichtte mooi werk door een verzameling losse tekstbestanden om te vormen tot dit boekje.

Working for Ruud means: collaboration time! It was a pleasure to work together with Ken Figueroa and Ari Melnick (Cornell, New York) and it is a great honor to have Ari in the thesis committee. Karen Keeshan and Warren Pear (University of Pennsylvania, Philadelphia) performed the initial TRIB2 experiments. Anton Langerak en Dennis Tielemans (afdeling Immunologie) deden typeringen van AMLs en T-ALLs. Wim van Putten (afdeling Trials en Statistiek) zorgde voor altijd correcte en up-to-date patiënteninformatie en was een betrouwbaar statistisch baken in hoofdstuk 9. Die patiënteninformatie was afkomstig van alle deelnemende HOVON-centra – zonder hen hadden weinig van de studies in dit proefschrift verricht kunnen worden. Wessel van Wieringen (VU Amsterdam) leerde me veel over statistiek, waaronder de kolder van het begrip "borderline significance". Without Daniel Tenen (Harvard Institutes of Medicine, Boston), who indoctrinated me when I was a helpless student, this thesis would most likely not be about *CEBPA*. It was great to continue the collaboration during the past years. Meritxell Alberich Jordà, former Delwel-student and current Tenen-postdoc, taught me the absolute basics of pipetting – it is hard to imagine a better tutor.

Met twee paranimfen die het klappen van de zweep als promovendus kennen, moet het wel goed komen. Kees Gerestein and Karishma Palande, thanks for spending the afternoon next to me.

Na een dag werk bleek er ook altijd nog een wereld buiten het lab te bestaan. Al meer dan dertig jaar heb ik het grote geluk van een stabiele thuisbasis. Pappa, mamma en Noor, bedankt voor dit uiterst succesvolle samenwerkingsverband. Laura, met jou is het elk weekend vakantie, in Groningen of in Rotterdam. Ik hoop dat het voorlopig vakantie blijft worden.

CURRICULUM VITAE

De auteur van dit proefschrift werd op 8 april 1978 geboren in Nijmegen. Na het afronden van het Erasmiaans Gymnasium in Rotterdam begon bij in september 1996 aan de studie geneeskunde aan de Erasmus Universiteit (nu: Erasmus MC) in Rotterdam. Na een kort doch leervol uitstapje naar de studie econometrie aan dezelfde universiteit (propedeuse 2000) behaalde hij zijn doctoraalexamen geneeskunde in 2001. Als doctoraalstudent deed hij 9 maanden onderzoek naar de rol van afwijkingen in CCAAT/enhancer binding protein alpha in longkanker in het laboratorium van dr. Daniel Tenen aan de Harvard Institutes of Medicine in Boston (VS). Na het behalen van het artsexamen begon hij in juni 2004 als promovendus in de onderzoeksgroep van dr. Ruud Delwel op de afdeling hematologie van het Erasmus MC (promotor prof. dr. Bob Löwenberg). Aldaar vond het onderzoek beschreven in dit proefschrift plaats. In januari 2009 begon hij aan de opleiding tot internist in het Haven-ziekenhuis in Rotterdam (opleider dr. P.J. Wismans).

PUBLICATIONS

- **Wouters BJ**, Löwenberg B, Erpelinck-Verschueren CA, van Putten WL, Valk PJ, Delwel R. Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood. 2009;113(13):3088-3091.*

- Figueroa ME[(*)], **Wouters BJ**[(*)], Skrabanek L, Glass J, Li Y, Erpelinck-Verschueren CA, Langerak AW, Löwenberg B, Fazzari M, Greally JM, Valk PJ, Melnick A, Delwel R. Genome wide epigenetic analysis delineates a biologically distinct immature acute leukemia with myeloid/T-lymphoid features. *Blood. 2009;113(12):2795-2804. (*) these authors contributed equally to the study*

- Verhaak RGW[(*)], **Wouters BJ** [(*)] , Erpelinck CA, Abbas S, Beverloo HB, Lugthart S, Löwenberg B, Delwel R, Valk PJM. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica. 2009;94(1):131-4. (*) these authors contributed equally to the study*

- **Wouters BJ**, Löwenberg B, Delwel R. A decade of genome-wide gene expression profiling in acute myeloid leukemia: flashback and prospects. *Blood. 2009;113(2):291-8.*

- **Wouters BJ**, Koss C, Delwel R. Gene expression profiling for improved dissection of acute leukemia: A recently identified immature myeloid/T-lymphoid subgroup as an example. *Blood Cells Mol Dis. 2008;40(3):395-400.*

- **Wouters BJ**, Alberich Jorda M, Keeshan K, Louwers I, Erpelinck-Verschueren CAJ, Tielemans D, Langerak AW, He Y, Hetherington C, He Y, Yashiro-Ohtani Y, Zhang P, Hetherington CJ, Verhaak RG, Valk PJ, Löwenberg B, Tenen DG, Pear WS, Delwel R. Distinct gene expression profiles of myeloid/T-lymphoid leukemia with silenced CEBPA and mutations in NOTCH1. *Blood.* 2007;110(10):3706-14

- **Wouters BJ**, Sanders MA, Lugthart S, Geertsma W, Van Drunen E, Beverloo HB, Löwenberg B, Valk PJ, Delwel R. Segmental uniparental disomy as a recurrent mechanism for homozygous CEBPA mutations in acute myeloid leukemia. *Leukemia. 2007;21(11):2382-4*

- Jelier R, Jenster G, Dorssers LC, **Wouters BJ**, Hendriksen PJ, Mons B, Delwel R, Kors JA. Text-derived concept profiles support assessment of DNA microarray data for acute myeloid leukemia and for androgen receptor stimulation. *BMC Bioinformatics. 2007;8:14.*

- **Wouters BJ**, Louwers I, Valk PJ, Löwenberg B, Delwel R. A recurrent in-frame insertion in a CEBPA transactivation domain is a polymorphism rather than a mutation that does not affect gene expression profiling-based clustering of AML. *Blood. 2007;109(1):389-90.*

- Keeshan K, He Y, **Wouters BJ**, Shestova O, Xu L, Sai H, Rodriguez CG, Maillard I, Tobias JW, Valk P, Carroll M, Aster JC, Delwel R, Pear WS. Tribbles homolog 2 inactivates C/EBPalpha and causes acute myelogenous leukemia. *Cancer Cell. 2006;10(5):401-11.*

- Costa DB, Dayaram T, D'Alo F, **Wouters BJ**, Tenen DG, Meyerson M, Tsao MS, Halmos B. C/EBP alpha mutations in lung cancer. *Lung Cancer. 2006;53(2):253-4.*

- Halmos B, Basseres DS, Monti S, D'Alo F, Dayaram T, Ferenczi K, **Wouters BJ**, Huettner CS, Golub TR, Tenen DG. A transcriptional profiling study of CCAAT/enhancer binding protein targets identifies hepatocyte nuclear factor 3 beta as a novel tumor suppressor in lung cancer. *Cancer Res. 2004;64(12):4137-47.*

ABBREVIATIONS

AML	Acute myeloid leukemia
AML1	Acute myeloid leukemia 1 (gene)
ALL	Acute lymphoblastic leukemia
BAALC	Brain and acute leukemia, cytoplasmic (gene)
BAC	Bacterial artificial chromosome
BM	Bone marrow
BMT	Bone marrow transplantation
bZIP	Basic leucine zipper
CEBPA	CCAAT/enhancer binding protein alpha (gene)
C/EBPα	CCAAT/enhancer binding protein alpha (protein)
CBF	Core binding factor
CBFB	Core binding factor beta (gene)
CD	Cluster of differentiation / Cluster designation
CGH	Comparative genomic hybridization
ChIP	Chromatin immunoprecipitation
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
dHPLC	Denaturing high performance liquid chromatography
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
E_2	β-estradiol
EFS	Event-free survival
ER	Estrogen receptor
ERG	v-ets erythroblastosis virus E26 oncogene homolog (gene)
ETO	Eight twenty one (gene)
EVI1	Ecotropic viral integration site 1 (gene)
FAB	French American British
FISH	Fluorescence in situ hybridization
FLT3	FMS-like tyrosine kinase 3 (gene)
G-CSF	Granulocyte colony stimulating factor
GEP	Gene expression profiling
GFP	Green fluorescent protein
GM-CSF	Granulocyte-macrophage colony stimulating factor
GMP	Granulocyte monocyte progenitor
GSEA	Gene set enrichment analysis
HD	Heterodimerization
HELP	HpaII tiny fragment enrichment by ligation mediated PCR

HSC	Hematopoietic stem cell
ICN1	Intracellular NOTCH1
IL-3 / -6	Interleukin 3 / -6
IP	Immunoprecipitation
ITD	Internal tandem duplication
LBD	Ligand binding domain
Limma	Linear models for microarray data
MAS	MicroArray Suite
MAT	Model based analysis of tiling-arrays
miRNA	MicroRNA
MLL	Mixed lineage leukemia (gene)
mRNA	Messenger RNA
MYH11	Myosin, heavy chain 11 (gene)
NOTCH1	Notch homolog 1 (gene)
NPM1	Nucleophosmin (gene)
OS	Overall survival
PAM	Prediction analysis for microarrays
PCR	Polymerase chain reaction
PEST	Proline (P), glutamic acid (E), serine (S), and threonine (T)
PML	Promyelocytic leukemia (gene)
PTD	Partial tandem duplication
RARA	Retinoic acid receptor, alpha (gene)
RMA	Robust multi averaging
RNA	Ribonucleic acid
RQ-PCR	Quantitative real-time reverse transcription PCR
RT-PCR	Reverse transcription PCR
RUNX1	Runt-related transcription factor 1 (gene)
SAM	Significance analysis of microarrays
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
TAD	Transactivation domain
TCR	T-cell receptor
TKD	Tyrosine kinase domain
TRIB2	Tribbles homolog 2 (gene)
UPD	Uniparental disomy
WBC	White blood cell
WHO	World Health Organization

Chapter 2

Figure 1. Summary of GEP findings in a cohort of 285 cases of AML.

A. A previous study of 285 cases of AML revealed 16 subgroups (clusters) of cases based on similarities in gene expression profiles (22). In the figure on the left, pair wise correlations between these AML cases are shown. The cells in the visualization are colored by Pearson correlation coefficient values, with deeper colors depicting higher positive (red) or negative (blue) correlations, as indicated by the scale bar. Five of the 16 clusters have been labeled, i.e. clusters #4, #5, #9, #12 and #13.

One finding of the original study was the tight aggregation into distinct clusters of AML cases with cytogenetic abnormalities that predict good risk. For those cases, cytogenetic status is color coded in the *cytogenetics* column, i.e. inv(16) (yellow, next to cluster #9), t(15;17) (orange, cluster #12) and t(8;21) (pink, cluster #13).

A subsequent study in the same patient cohort identified *NPM1* mutations in 95/285 cases. *NPM1* mutational status is depicted next to each case (red = *NPM1* mutant, green = *NPM1* wild type) (24). The figure demonstrates that *NPM1* mutations were not randomly distributed over the 16 previously defined clusters, but enriched in several of them.

Cluster #4 was found to associate with *CEBPA* mutations (red = *CEBPA* mutant). However, a subset of 6 patients in this cluster did not show any *CEBPA* mutation (green = *CEBPA* wild type). It was found that these cases differed in their *CEBPA* mRNA expression as compared to the *CEBPA* mutant AMLs, as indicated by the histograms depicting signal intensity values for the *CEBPA* probe set on the microarray. In fact, whereas *CEBPA* mutant AMLs highly expressed *CEBPA* mRNA, expression was silenced in the cases lacking mutations. This silencing was associated with *CEBPA* DNA promoter hypermethylation (+ = methylation, - = no methylation). In addition, *NOTCH1* mutations were found as common characteristics of this subgroup (red = *NOTCH1* mutation, green = *NOTCH1* wild type) (31).

B. In the original analysis of 285 AML cases on the left hand side (panel A), the 44 cases in cluster #5 aggregated very tightly, as indicated by the deep red colors, i.e. positive Pearson correlation coefficients. Most of these 44 cases showed a monocytoid morphology (FAB-M4 or –M5) (22). This raises the possibility that a significant part of the clustering effect was caused by specific up- or downregulation of genes that are important in monocytic differentiation, resulting in a different signature than the remaining, mostly non-monocytoid, cases of AML in the study. To answer whether gene expression profiling would enable identification of potential heterogeneity within this apparently homogeneous subgroup, here (panel B) the 44 cases were re-clustered as an isolated cohort. For this analysis, only probe sets that showed a variable expression within these 44 AML cases were taken into account, as defined by a fold change of 3.5 to the mean in log2 scale in at least one case. The resulting cluster image shows that a number of potentially interesting subgroups can indeed be identified within these 44 AML cases, which have been indicated by gray lines
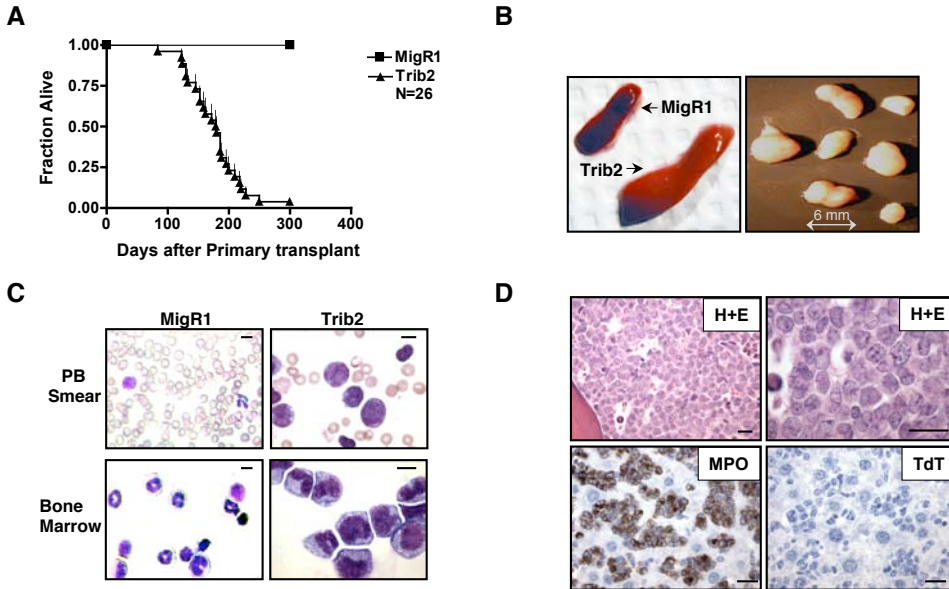
Figure 2. Trib2 induces AML

A. Kaplan-Meier survival curve of mice receiving Trib2-transduced BM compared to MigR1 control. The median survival of Trib2 mice was 179 days. Results are derived from seven independent experiments.

B. Representative photographs of splenomegaly in Trib2 mice compared to control MigR1 spleen, and lymphadenopathy in Trib2 mice.

C. Wright-Giemsa-stained PB and BM single cell suspensions from MigR1 and leukemic Trib2 mice. Scale bars (upper right) represent 10 μm. The percentage of GFP+ cells in Trib2 BM was approximately 90%–100%.

D. Histopathology of BM sections from Trib2-induced AML. Hematoxylin and eosin (H+E) section showing hypercellularity (top left) due to the presence of sheets of immature cells and blasts (top right). The tumor cells stain positively for myeloperoxidase (MPO, bottom left) and negatively for terminal deoxytransferase (TdT, bottom right). Scale bars (lower right) represent 20 μm.

Figure 8. *TRIB2* **is elevated in a subset of human AML**
Correlation view of 285 AML patients (30). Colors of cells relate to Pearson's correlation coefficient values: red indicates higher positive and blue indicates higher negative correlation between samples. Sixteen clusters represented by red blocks along the diagonal can be identified. *CEBPA* mutation status is indicated next to each tumor (red, mutant; green, wild-type). Histograms next to each tumor represent expression levels of the two probe sets for *TRIB2*. Cluster 4, one of the two clusters harboring most patients with *CEBPA* mutations, has a significantly elevated expression of *TRIB2* relative to other clusters

Figure 1. *CEBPA* silencing and promoter hypermethylation are associated with AMLs sharing a *CEBPA* mutant gene expression signature in GEP cluster #4.

A. Pair wise correlations between gene expression profiles of 285 AML samples calculated on the basis of 2856 probe sets are displayed as described (4). Colors of boxes visualize Pearson correlation coefficient: deeper red indicates higher positive correlation, deeper blue indicates higher negative correlation. Sixteen distinct clusters were previously distinguished, which can be recognized by the red blocks showing high correlation along the diagonal (4). Cluster #4 and cluster #15, associated with *CEBPA* mutations, are enlarged. The bar and histogram next to each patient represent *CEBPA* mutation status and *CEBPA* expression level, respectively. *CEBPA* mutation status: presence ("mut", red) or absence ("wt", green) of mutations in bZIP region and/or N-terminus. These data indicate that in leukemias lacking mutations, *CEBPA* expression is low or absent. In cluster #4, the order of samples, from top to bottom, is: #3327, #2242 (both *CEBPA* mutant), #2668, #2238, #3314, #2686, #3483, #3491 (all 6 without *CEBPA* mutation), #2218, #1316, #2273, #2545, #2169, #2753, and #2192 (all 7 *CEBPA* mutant).

B. Upper part: schematic representation of the chromosomal region surrounding the transcriptional start of the *CEBPA* gene. Numbers indicate position relative to *CEBPA* transcriptional start. Vertical lines represent CpG dinucleotides, "transcription" stands for transcriptional start, "ATG" stands for translational start site. Lower part: level of cytosine methylation in the region surrounding the *CEBPA* transcriptional start site of the 6 AML cases in cluster #4 with low *CEBPA* expression, with patient numbers on the left. Every cytosine in a CpG dinucleotide is depicted as a circle. For each of these cytosines, the fraction of methylated residues was determined, which is visualized by the color of the circle: methylated (>75% of all cytosines methylated, black), partly methylated (25-75% methylated, grey), unmethylated (<25% methylated, white).

Figure 2. AML cases with silenced *CEBPA* in cluster #4 simultaneously express myeloid and T-lymphoid specific genes and lineage markers.

A. Pair wise correlations between samples are displayed as explained in the legend to Figure 1, and GEP cluster #4 is enlarged in the box on the right. Histograms next to each patient display expression of selected genes with significantly elevated expression in cluster #4 cases with silenced *CEBPA*. Expression levels for probe sets of the following genes are visualized: *CD7*, *CD3D*, *CD3Z, CD3G, LCK, TRD@* and *NOTCH1.* Corresponding expression levels and Affymetrix probe set identifiers are depicted in Supplementary Table 4.

B. Representative dot plot images from flowcytometric analysis of samples obtained from 2 individual patients, i.e. patients #2238 and #3491, demonstrating that the majority of cells from these patients simultaneously express CD34 and CD13, CD7 and terminal deoxynucleotidyltransferase (TdT). The tumor population was identified by weak expression of CD45, depicted in black, whereas CD45[high] cells, which are considered to be mature lymphocytes, are colored in grey.

Figure 5. A gene expression prediction signature identifies new leukemias with the silenced *CEBPA* phenotype in an independent cohort of AML.

A. For 262 samples analyzed on Affymetrix HGU133Plus2.0 GeneChips, log transformed (base 2) and mean centered expression levels for 13 probe sets are depicted (left panel) for an arbitrary range from -4 to +4 (corresponding to 16-fold lower to 16-fold higher expression relative to the mean, respectively). The ordering of patients in the figure is arbitrary. In the right panel, these data are enlarged for 31 of these 262 leukemias, representing 3 groups: (I) the 6 cases with silenced *CEBPA* previously identified, with from top to bottom cases #2668, #2238, #3314, #2686, #3483, and #3491; (II) a variable selection of 10 AMLs from distinct GEP clusters for which also *NOTCH1* mutational analysis and *CEBPA* promoter bisulfite sequencing were performed, and (III) 15 AMLs with *CEBPA* mutations, originating from either GEP cluster #4 (upper 9 samples) or cluster #15 (lower 6 samples). The 9 probe sets on the left side constitute the most predictive gene expression signature for group I, as determined by PAM: 216191_s_at (*TRA@/TRD@* 1), 217143_s_at (*TRA@/TRD@* 2), 213830_at (*TRD@*), 225681_at (*CTHRC1*), 1565809_x_at (no annotation), 1560018_at (*ARPP-21*), 210844_x_at (*CTNNA1* 1), 200764_s_at (*CTNNA1* 2), and 200765_x_at (*CTNNA1* 3). To the right, 4 additional probe sets are indicated, i.e. 204039_at (*CEBPA*), 218902_at (*NOTCH1*), 202478_at (*TRIB2*) and 214551_s_at (*CD7*). Mutational data for *NOTCH1* and *CEBPA*, and methylation status of the *CEBPA* promoter are depicted next to the normalized hybridization intensities of the probe sets.

B. 268 samples obtained from a second cohort of AML were hybridized to HGU133Plus2.0 GeneChips. The 9 probe set signature was used to identify leukemias with a profile similar to group I (A), resulting in the detection of group IV (from top to bottom cases #6376, #6735, #6947, #7053, #7076 and #7120).
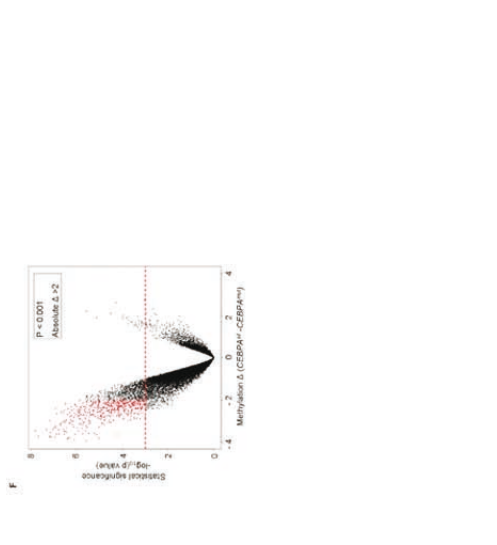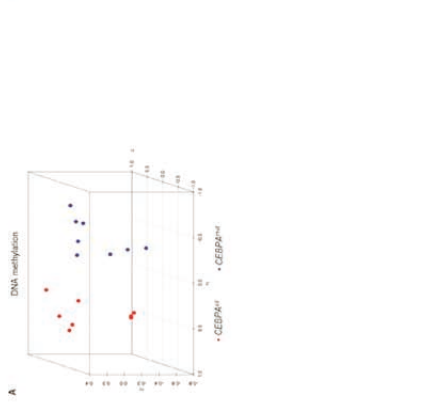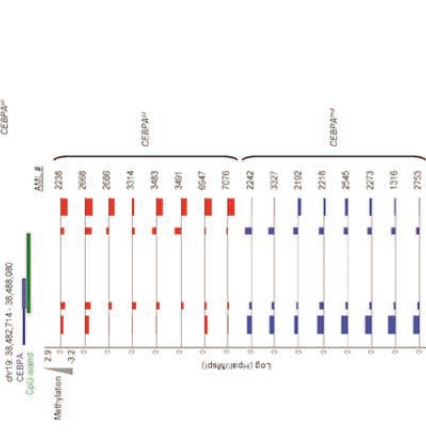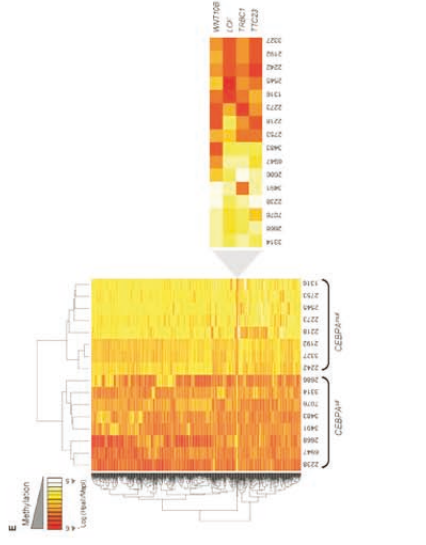
Figure 1. A unique methylation profile distinguishes *CEBPA^sil* from *CEBPA^mut* AML.

A. Principal component analysis of DNA methylation data using the HELP assay on 8 *CEBPA^sil* and 8 *CEBPA^mut* AML cases revealed that the cases were readily segregated into two clusters, which matched exactly with *CEBPA* status.

B. heatmap representation of the four probe sets annotated to the *CEBPA* locus on the HELP microarray; cases are clustered according their methylation status. *CEBPA^sil* cases cluster together (left node) and all show higher levels of methylation for at least 3 or the 4 probe sets.

C. Representation of the positioning of the four probe sets relative to the genomic localization of the *CEBPA* locus and its CpG island on chromosome 19. HELP methylation values for each leukemia case are represented in one row; the y axis represents centered log2 (HpaII/MspI) ratios. Positive values correspond to hypomethylated fragments, while a negative deflection reflects a methylated fragment. The first 8 rows correspond to the *CEBPA^sil* cases (in red) and the remaining rows to the *CEBPA^mut* cases (in blue).

D. Heatmap representing the DNA methylation status at five different regions of the *CEBPA* locus. Percent cytosine methylation was determined at these regions for all cases using MassARRAY EpiTyper.

E. Two-dimensional hierarchical clustering of genes differentially methylated between the two leukemia subgroups, illustrated by a heatmap. Supervised analysis identified 567 HpaII amplifiable fragments (p <0.001 and absolute difference in methylation >2). Cases are represented in the columns and probe sets in the rows. *CEBPA^sil* cases are clustered in the left node, and display high methylation levels for 563 HpaII amplifiable fragments. *To the right* heatmap representation of the four probe sets that displayed the opposite behavior i.e. relative hypomethylation in *CEBPA^sil* leukemia.

F. A plot of methylation difference between *CEBPA^sil* and *CEBPA^mut* cases (x axis) vs. statistical significance (y axis) shows the marked asymmetry of the two branches, illustrating the overall tendency to higher methylation levels in the *CEBPA^sil* cases. Red points demark probe sets that reached both criteria for differential methylation on our analysis (p<0.001 and absolute methylation difference >2)
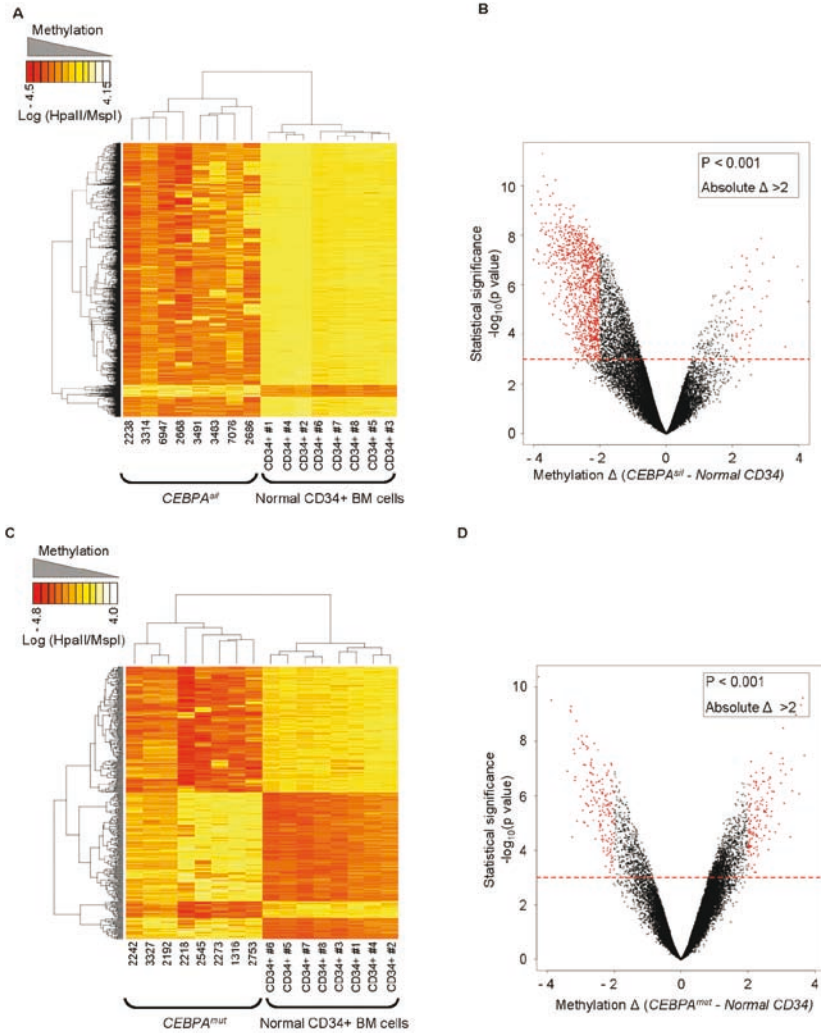
Figure 2. Aberrant hypermethylation is a feature of *CEBPA^sil* leukemia.

A. Two-dimensional hierarchical clustering of genes differentially methylated between the *CEBPA^sil* cases and normal CD34+ hematopoietic progenitors, illustrated by a heatmap. Supervised analysis identified 1035 HpaII amplifiable fragments (p <0.001 and absolute difference in methylation >2). Cases are represented in the columns and probe sets in the rows. *CEBPA^sil* cases are clustered in the left node, and display marked hypermethylation when compared to the normals, as illustrated by the predominance of probe sets with low log2 (HpaII/MspI) ratios.

B. Methylation difference between *CEBPA^sil* and normal CD34+ cells (x axis) vs. statistical significance (y axis) plot with marked asymmetry of the two branches, reflecting the tendency to higher methylation levels in this subgroup. Red points demark probe sets that reached both criteria for differential methylation in our analysis.

C. Two-dimensional hierarchical clustering of genes differentially methylated between the *CEBPA^mut* AML and normal CD34+ hematopoietic progenitors, illustrated by a heatmap. Supervised analysis identified 322 probe sets (286 genes). *CEBPA^mut* cases are clustered in the left node, and display equal components of *hyper* and *hypo*methylation when compared to CD34+ normal cells.

D. A plot of methylation difference between mutant *CEBPA* and normal CD34+ cells (x axis) vs. statistical significance (y axis) shows symmetric branches and less pronounced differences in methylation than in the case of the silenced *CEBPA* subgroup. Red points demark probe sets that reached both criteria for differential methylation on our analysis.
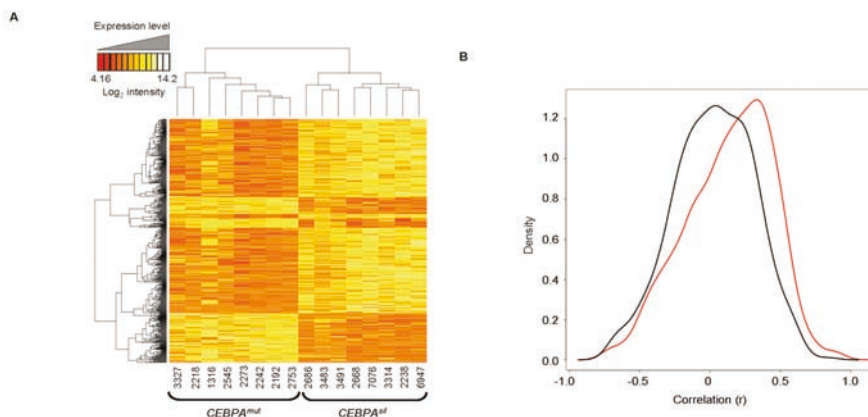
Figure 4. DNA methylation and gene expression capture complementary information.
A. Two-dimensional hierarchical clustering of genes differentially expressed between the two leukemia subgroups, illustrated by a heatmap. Supervised analysis identified 587 probe sets (415 genes) at a p <0.001 and fold change >2. Cases are represented in the columns and probe sets in the rows. *CEBPA^sil* cases are clustered in the right node, and *CEBPA^mut* cases are clustered on the right node.
B. Density (y axis) plot for the gene-by-gene correlations (x axis) between gene expression log intensity and log(HpaII/MspI) values. A positive correlation between these two measures translates into a negative biological correlation, i.e. hypermethylation in combination with lower expression levels, or hypomethylation in combination with higher expression levels. In black, density plot for the correlation between expression and methylation for a set of 600 randomly selected probe sets. In red, density plot for the correlation between expression and methylation measured by the 567 probe sets in the genes differentially methylated between the two subgroups. The shifting of the density plot to the right reflects a tendency to a stronger correlation of DNA methylation with gene expression levels in this subset of genes. The figure is representative of five analyses; each time using a different set of 600 randomly selected HpaII amplifiable fragments for the calculation of correlations.
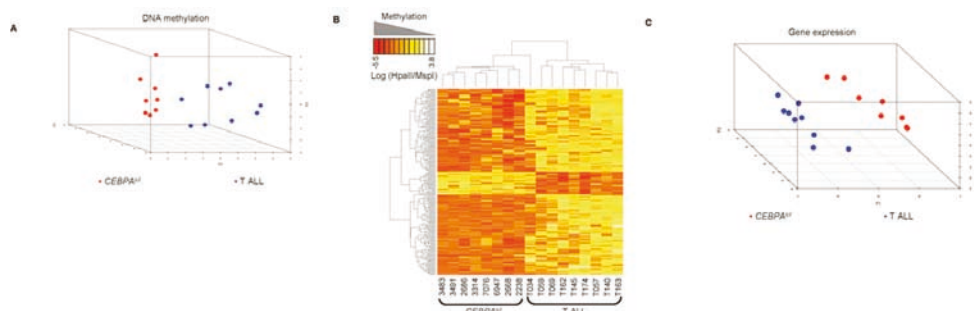

Figure 5. *CEBPA^sil* leukemias differ genetically and epigenetically from T-ALL.
A. Principal component analysis of DNA methylation data comparing 8 immature acute myeloid/T lymphoid *CEBPA^sil* cases to a selection of 9 T-ALL cases representing a spectrum of maturation stages, showing separate clustering of the two groups of leukemias.
B. Two-dimensional hierarchical clustering of genes differentially methylated between the *CEBPA^sil* leukemias and the T-ALL cases, illustrated by a heatmap. Cases are represented in the columns and probe sets in the rows. Supervised analysis identified 213 differentially methylated probe sets (199 genes). *CEBPA^sil* cases are clustered in the left node, and display a predominance of hypomethylated probe sets.
C. Principal component analysis of gene expression data for the same cases also demonstrates separate clustering of the two groups of leukemias, indicating that these two groups display distinct expression profiles.
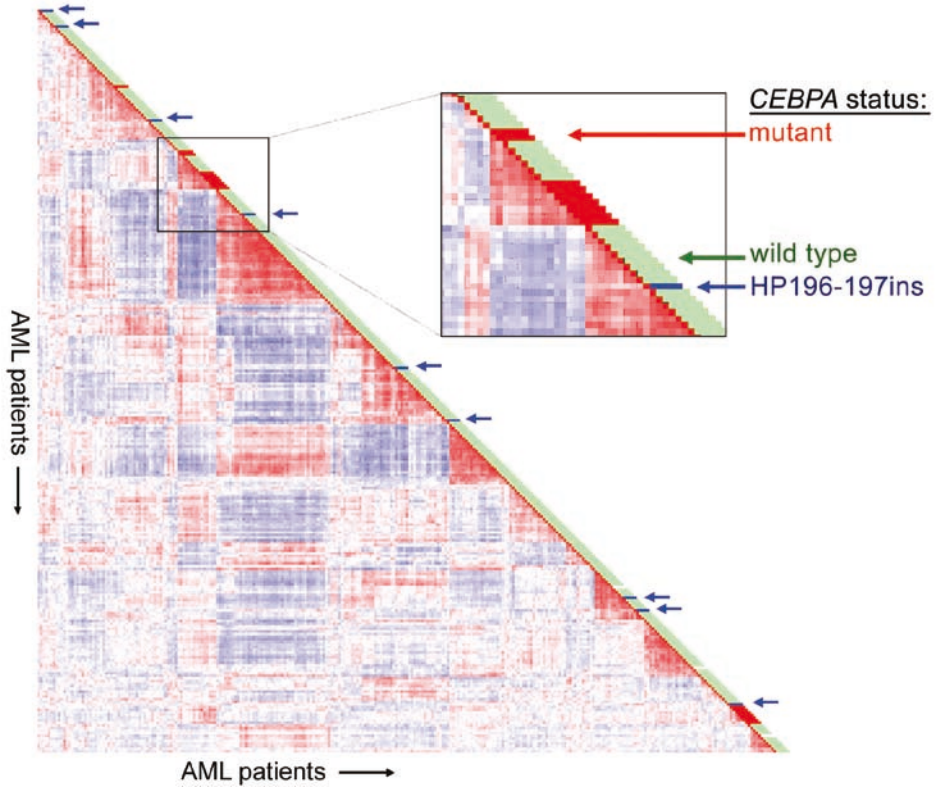
Figure 1. Correlation view of 285 AML cases including *CEBPA* status.
Pairwise correlations between samples are displayed using 2856 probe sets as described.(8) Colors of boxes visualize Pearson's correlation coefficient: red indicates higher positive correlation, blue indicates higher negative correlation. Bars next to each sample represent *CEBPA* status: mutation in bZIP region and/or N-terminus (red), presence of HP196-197ins (blue), or neither (green). For three specimens, depicted in white, no material was available for dHPLC analysis. Figure generated using HeatMapper software (10).
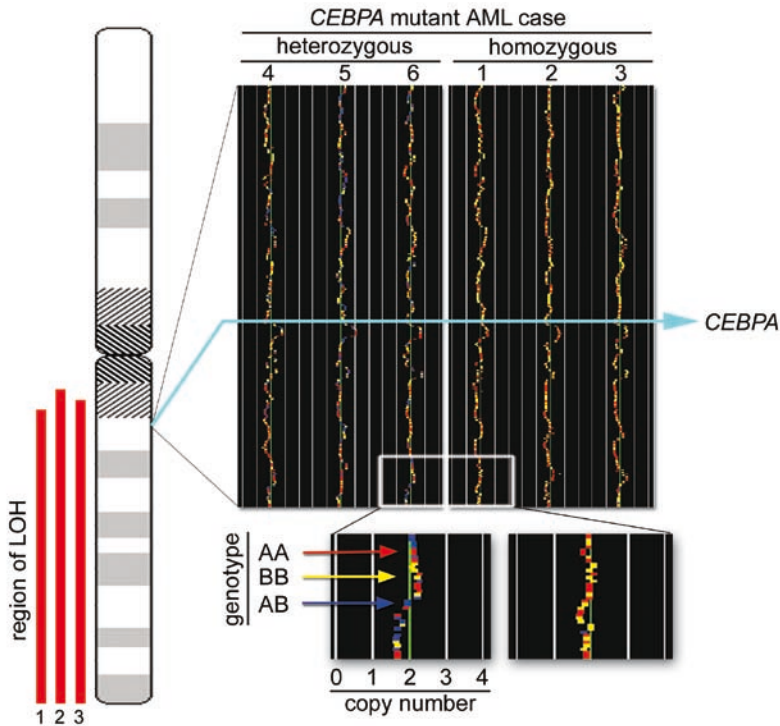
Figure 1. Copy number neutral loss of heterozygosity (LOH) in three cases of acute myeloid leukemia (AML) with homozygous *CEBPA* mutations.

Affymetrix Mapping 250K NspI arrays were used to investigate chromosomal copy numbers and genotypes of AML cases with *CEBPA* mutations. Six cases are shown in this figure: patients #1, #2 and #3 harbor homozygous *CEBPA* mutations, and cases #4, #5 and #6 carry heterozygous *CEBPA* mutations. A region on chromosome 19q13.11, including the *CEBPA* locus (*CEBPA*), is depicted at larger magnification. For each individual single nucleotide polymorphism (SNP) in this region, copy number was calculated, and is indicated as deflection from the green midline, which represents the presence of two chromosomes (deflection to the left indicates lower copy number, deflection to the right indicates higher copy number). Genotypes for each SNP are color-coded: homozygous AA (red) or BB (yellow), and heterozygous AB (blue). A hidden Markov model was employed to assess potential LOH, and revealed LOH of the region shown in patients #1, #2 and #3, illustrated by the fact that the large majority of calls is homozygous. The infrequent remaining heterozygous (blue) signals are considered false calls according to this model. The extents of the regions of LOH are indicated as red bars next to the schematic representation of chromosome 19 ("region of LOH") for each of these three cases.
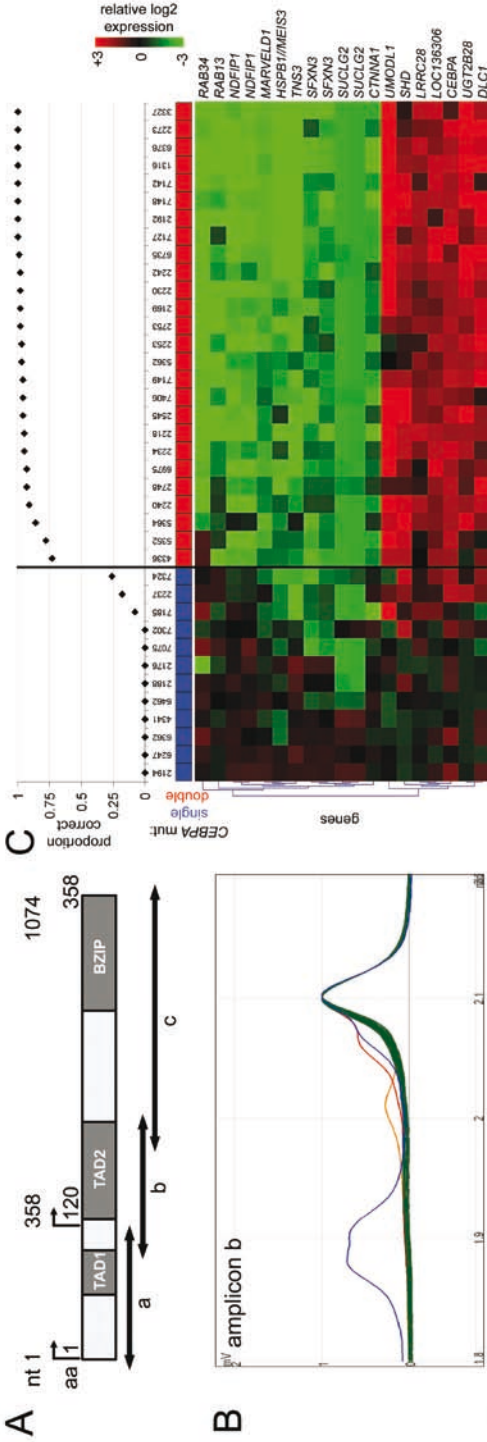
Figure 1. Schematic overview of dHPLC analysis, gene expression profiling analysis and survival estimates.

A. Schematic representation of the *CEBPA* gene and location of amplicons a, b and c for PCR, used for dHPLC analysis. Functional regions are depicted, i.e. two transactivation domains (TAD1 and TAD2) in the N-terminal part, and the basic leucine zipper (bZIP) region in the C-terminal part. Nucleotide (nt) position is indicated relative to the main translation start site. Amino acid (aa) numbering and the alternative translation start site at position nt 358 (aa 120) are also depicted.

B. Representative profiles of dHPLC analysis of one of the three investigated fragments, i.e. amplicons b, in a random selection of 90 samples. Heteroduplexes (various colors) are released earlier than homoduplexes (green), and can therefore be recognized as distinct peaks. Time is depicted on the x-axis, and absorbance on the y-axis.

C. A gene expression prediction signature for *CEBPA*^{mut} AML (irrespective of single or double mutant status) was derived in a data set of 524 AMLs, including 38 *CEBPA*^{mut} cases. Prediction accuracy for each of the 38 *CEBPA*^{mut} cases was estimated using repeated 10-fold cross-validation as detailed in Supplementary Materials and Methods. The proportion of correct predictions for the selected 38 *CEBPA*^{mut} specimens is indicated (upper panel). Mutation status is color coded, i.e *CEBPA*$^{single-mut}$ (blue) or *CEBPA*$^{double-mut}$ (red). The heat map in the lower panel depicts the 19 probe sets in the resulting *CEBPA*^{mut} gene expression classifier (see Table S2 for probe set information). Intensity values (log2) were mean centered over the cohort of 524 AML cases and for visualization purposes the genes were hierarchically clustered (Euclidian distance, average linkage). Cells represent relative log2 expression values, and have been color coded on a scale ranging from bright green (-3) to bright red (+3), with black indicating no change relative to the mean.

D. Kaplan Meier estimates of overall survival (OS) among *CEBPA*^{mut} and *CEBPA*^{wt} AML, log rank test P=0.027.

E. OS among *CEBPA*$^{double-mut}$ versus *CEBPA*^{wt} AML, P=0.004, and versus *CEBPA*$^{single-mut}$ AML, P=0.005; pooled P=0.012 F. Event-free survival (EFS) among *CEBPA*$^{double-mut}$ and *CEBPA*^{wt} AML, P=0.005, and versus *CEBPA*$^{single-mut}$ AML, P=0.004; pooled P=0.008. The cumulative proportion of survival at the intercept, i.e. the point where a line crosses the Y-axis, reflects the proportion of patients reaching complete remission. Analyses similar to those depicted in panels D-F were performed after splitting the group of *CEBPA*^{wt} AMLs into those with favorable cytogenetics and those with other cytogenetics. These additional analyses can be found in Figure S4.
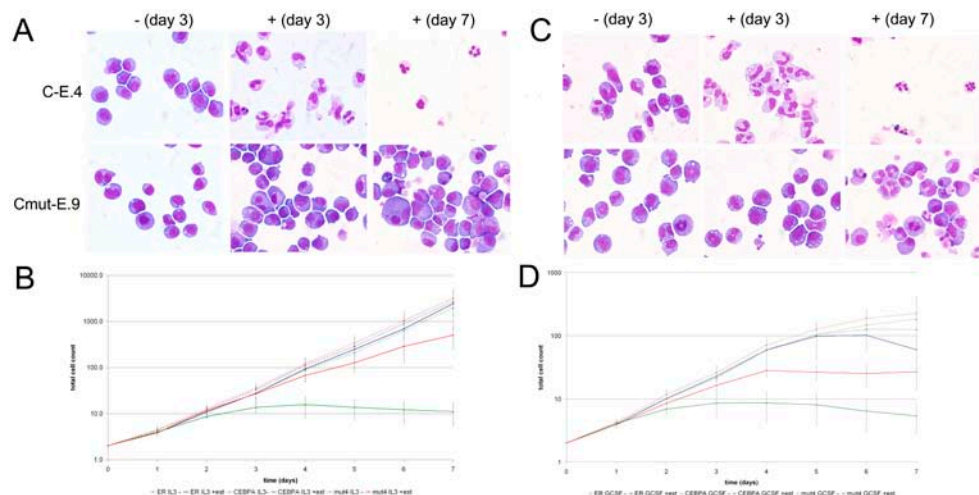
Chapter 10

Figure 1. 32D-Cebpa-ER cells differentiate upon E2 treatment while 32D-Cebpamut-ER cells show impaired differentiation

A. Representative cytospins of 32D-Cebpa-ER cells (clone C-E.4) and 32D-Cebpamut-ER cells (clone Cmut-E.9) in the absence (-) and presence (+) of E$_2$ for 3 or 7 days. The cells were cultured in the presence of IL3.

B. Proliferation curves of 32D-ER cells (ER), 32D-Cebpa-ER cells (CEBPA) and 32D-Cebpamut-ER cells (mut4) stimulated with IL3 alone (IL3 -) or IL3 in combination with E$_2$ (IL3 +est). The graph represents the average and standard deviations for 4 individual clones for each of the 3 cell types.

C. Representative cytospins of 32D-Cebpa-ER cells (clone C-E.4) and 32D-Cebpamut-ER cells (clone Cmut-E.9) in the absence (-) and presence (+) of E$_2$ for 3 or 7 days. The cells were cultured in the presence of G-CSF.

D. Proliferation curves of 32D-ER cells (ER), 32D-Cebpa-ER cells (CEBPA) and 32D-Cebpamut-ER cells (mut4) stimulated with G-CSF alone (G-CSF -) or IL3 in combination with E$_2$ (G-CSF +est). The graph represents the average and standard deviations for 4 individual clones for each of the 3 cell types.
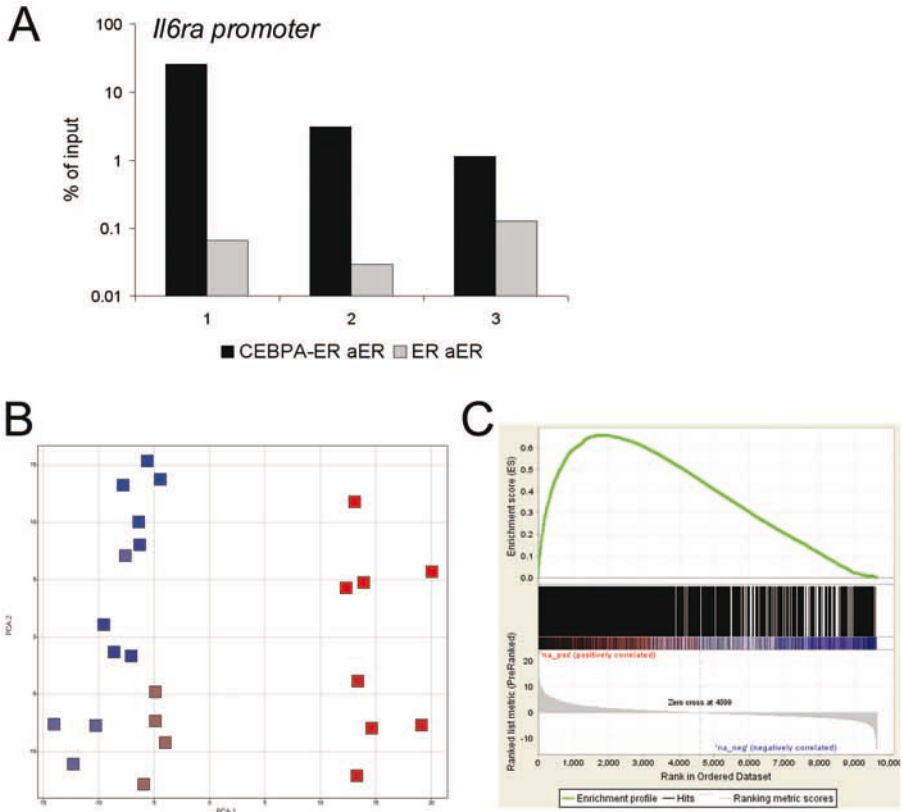
Figure 2. ChIP-chip identifies promoter regions occupied by the Cebpa-ER fusion protein that are associated with gene expression changes.

A. RQ-PCR validation of ChIP of *Il6ra* promoter. Three independent ChIP experiments were performed using 32D-Cebpa-ER cells (black bars) or 32D-ER cells (white bars). RQ-PCR was performed in the *Il6ra* promoter, and percentage of amount of input material is indicated in log10 scale. In each of the three experiments, enrichment in Cebpa-ER clones is apparent.

B. Principal component analysis of gene expression data from 32D-Cebpa-ER cells (reddish) and 32D-ER cells (bluish). Color intensities reflect time points, ranging from dark (time point 0), brighter (2 hours of E_2 treatment) and brightest (8 hours of E_2 treatment). Separation of samples according to E_2 treatment over the first principal component (PCA1) is apparent for the Cebpa-ER clones only. An additional separation over the second principal component (PCA2) correlates with time point (2 versus 8 hours of E_2 treatment).

C. Gene set enrichment analysis results. Gene expression profiles from 32D-Cebpa-ER cells (N=4) and 32D-ER cells (N=4) were compared after 2 hours of E_2 treatment. All probe sets that passed the variability and intensity filter were ordered according to their higher expression in either Cebpa-ER (left) or ER (right) cells. Genes that were immunoprecipitated are indicated in the bar as vertical lines. There is a significant (P<0.0001) enrichment of immunoprecipitated genes in the left part of the graph, indicated by the positive deflection of the enrichment score (upper part).