

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

Roel G. W. Verhaak, Chantal S. Goudswaard, Wim van Putten, Maarten A. Bijl, Mathijs A. Sanders, Wendy Hagens, André G. Uitterlinden, Claudia A. J. Erpelinck, Ruud Delwel, Bob Löwenberg, and Peter J. M. Valk

Mutations in nucleophosmin *NPM1* are the most frequent acquired molecular abnormalities in acute myeloid leukemia (AML). We determined the *NPM1* mutation status in a clinically and molecularly well-characterized patient cohort of 275 patients with newly diagnosed AML by denaturing high-performance liquid chromatography (dHPLC). We show that *NPM1* mutations are significantly underrepresented in patients younger than 35 years. *NPM1* mutations positively correlate with AML with high white blood cell counts, normal karyotypes, and *fms*-like ty-

rosine kinase-3 gene (*FLT3*) internal tandem duplication (ITD) mutations. *NPM1* mutations associate inversely with the occurrence of CCAAT/enhancer-binding protein- α (*CEBPA*) and *NRAS* mutations. With respect to gene expression profiling, we show that AML cases with an *NPM1* mutation cluster in specific subtypes of AML with previously established gene expression signatures, are highly associated with a homeobox gene-specific expression signature, and can be predicted with high accuracy. We demonstrate that patients with intermediate cyto-

netic risk AML without *FLT3* ITD mutations but with *NPM1* mutations have a significantly better overall survival (OS) and event-free survival (EFS) than those without *NPM1* mutations. Finally, in multivariable analysis *NPM1* mutations express independent favorable prognostic value with regard to OS, EFS, and disease-free survival (DFS). (Blood. 2005;106:3747-3754)

© 2005 by The American Society of Hematology

Introduction

Acute myeloid leukemia (AML) is a heterogeneous disease with diverse genetic abnormalities and variable responsiveness to therapy. Cytogenetic analyses and molecular analyses are currently used to risk-stratify AML. For instance, the translocations *inv*(16), *t*(8;21), and *t*(15;17) herald a favorable prognosis, whereas certain other cytogenetic aberrations indicate leukemia with intermediate or high risk of relapse.¹⁻⁵ Nevertheless, the classification of AML on the basis of karyotyping is still far from satisfactory. In recent years extended molecular analyses have yielded novel molecular markers important for proper diagnostics of AML. The internal tandem duplication (ITD) in the *fms*-like tyrosine kinase-3 gene (*FLT3*),^{6,7} partial tandem duplication (PTD)^{8,9} of the mixed lineage leukemia gene (*MLL*), and increased expression of the transcription factor ecotropic virus integration site 1 (*EV11*)¹⁰ are indicative of poor prognosis. In contrast, mutations in the transcription factor CCAAT/enhancer-binding protein- α (*CEBPA*) have been associated with a favorable response to therapy.^{11,12} A recent study showed mutations in exon 12 of the gene encoding nucleophosmin *NPM1* in approximately 35% of cases of *de novo* AML.¹³ Mutations in *NPM1* were found to be mutually exclusive with certain common recurrent chromosomal aberrations and are predominantly seen in AML with normal karyotypes and *FLT3* ITD mutations.

NPM1 is predominantly localized in the nucleolus and is thought to function as a molecular chaperone of proteins, facilitating the transport

of ribosomal proteins through the nuclear membrane.¹⁴⁻¹⁶ Disruption of *NPM1*, either by chromosomal translocation or by mutation, results in the cytoplasmic dislocation of *NPM1*. The high frequency of *NPM1* mutations in AML with normal karyotypes and the observation that cytoplasmic *NPM1* cannot exert its normal functions as binding partner and transporter protein lead to the notion that *NPM1* mutation may be an early event in leukemogenesis.

An important role for *NPM1* in leukemias and lymphomas has been proposed previously. *NPM1* has been found to be part of several fusion proteins that are formed as a result of chromosomal translocation and in which only the *NPM1* N-terminal region is conserved. A *t*(2;5)(p23;q35) chromosomal translocation occurs in approximately 8% of non-Hodgkin lymphomas in children and young adults and results in the chimeric fusion of *NPM1* to *ALK*.¹⁷ In rare cases of acute promyelocytic leukemia (APL), characterized by chromosomal translocations that disrupt the gene encoding the retinoic acid receptor (*RARA*), fusion of *NPM1* to *RARA* was shown.¹⁸ A *t*(3;5)(q25.1;q34) chromosomal translocation, infrequently seen in myelodysplastic syndrome and AML, gives rise to a fusion transcript of *NPM1* and *MLF1*.¹⁹

Gene expression profiling is a powerful way to comprehensively classify individuals with AML and to further resolve the heterogeneous nature of AML.²⁰ Using this technique, new prognostically relevant AML subtypes have been identified, while the

From the Departments of Hematology, Statistics, and Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands.

Submitted June 1, 2005; accepted July 19, 2005. Prepublished online as *Blood* First Edition Paper, August 18, 2005; DOI 10.1182/blood-2005-05-2168.

Supported by grants from the Dutch Cancer Society (Koninkrijk Wilhelmina Fonds) and the Erasmus University Medical Center (Revolving Fund).

The online version of the article contains a data supplement.

An Inside *Blood* analysis of this article appears in the front of this issue.

Reprints: Peter J. M. Valk, Erasmus University Medical Center Rotterdam, Department of Hematology, Ee1391a, Dr Molewaterplein 50, 3015 GE Rotterdam Z-H, The Netherlands; e-mail: p.valk@erasmusmc.nl.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

© 2005 by The American Society of Hematology

presence of recurrent chromosomal abnormalities such as *inv(16)*, *t(15;17)*, and *t(8;21)* as well as other molecular aberrations (eg, C- and N-terminal mutations in *CEBPA*) could be predicted with high accuracy by unique expression patterns.²¹⁻²³ In a recent study, novel subtypes of AML have also been defined based on gene expression profiling; however, the common molecular abnormalities in these AML subtypes are largely unknown.²¹ Because *NPM1* is mutated in approximately one third of AML patients, this molecular abnormality may drive the clustering of these AML subtypes. The effect of mutant *NPM1* has been studied using gene expression profiling and revealed a distinctive signature for *NPM1* mutations.²⁴ Among players in this signature were several homeobox-containing family members of homeobox (*HOX*) transcription factors and *CD34*, both observations being indicative of hematopoietic development.²⁴ However, it is currently not known whether *NPM1* mutations are predictable on the basis of a gene expression signature.

Cytoplasmic *NPM1* has been positively associated with remission rate¹³; however, the relation of mutant *NPM1* with survival outcome parameters remains to be elucidated.

We have studied a well-characterized cohort of 275 cases of de novo AML for the presence of a *NPM1* mutations to (1) validate denaturing high-performance liquid chromatography (dHPLC) as a rapid approach to determine *NPM1* mutations; (2) investigate the relation of *NPM1* mutations with regard to clinical parameters, cytogenetics, and various molecular abnormalities; (3) determine the relation of *NPM1* mutations in subtypes of AML, recently identified by gene expression profiling²¹; (4) derive *NPM1* mutation-specific and predictive gene expression signatures; and (5) determine the prognostic value of mutated *NPM1*.

Patients, materials, and methods

Patients and cell samples

Patients had a diagnosis of primary AML confirmed by cytologic examination of blood and bone marrow (median age, 44 years; range, 15-78 years); median bone marrow blast count, 65% (range, 0% [for APL] to 98%); median white blood cell (WBC) count, $32 \times 10^9/L$ (range, $0.3 \times 10^9/L$ to $263 \times 10^9/L$). All patients had been treated according to the Dutch-Belgian Hemato-Oncology Cooperative Group (HOVON) protocols.²⁵⁻²⁷ After informed consent, bone marrow aspirates or peripheral blood samples were taken at diagnosis. Blasts and mononuclear cells were purified by Ficoll-Hypaque (Nygaard, Oslo, Norway) centrifugation and cryopreserved. The AML samples contained 80% to 100% blast cells after thawing, regardless of the blast count at diagnosis.

PCR, WAVE, and sequence analyses

RNA isolation and cDNA synthesis were performed as described.^{21,28} Complementary DNA prepared from 50 ng RNA was used for all polymerase chain reaction (PCR) amplifications. *NPM1* mutations in exon 12 were determined by cDNA amplification using the primers *NPM1-FOR* 5'-CTTCCGGATGACTGACCAAGAG-3' and primer *NPM1-REV* 5'-CCTGGACAACATTATCAAACACG-3' (25 mM deoxyribonucleoside triphosphate [dNTP], 15 pmol primers, 2 mM MgCl₂, Taq polymerase, and $10 \times$ buffer [Invitrogen Life Technologies, Breda, The Netherlands]). Cycling conditions for *NPM1* mutation detection were as follows: 1 cycle, 5 minutes at 94°C; 30 cycles, 1 minute at 94°C, 1 minute at 58°C, and 1 minute at 72°C; and 1 cycle, 7 minutes at 72°C. PCR products were subsequently subjected to dHPLC using a Transgenomics (Omaha, NE) WAVE dHPLC system.²⁹ Samples were run at 56°C and 58°C. The exact *NPM1* mutant sequence was confirmed for all samples showing an abnormal dHPLC profile. PCR products were purified using the Multiscreen-PCR 96-well system (Millipore, Bedford, MA) followed by direct sequenc-

ing with *NPM1-REV* using an ABI-PRISM3100 genetic analyzer (Applied Biosystems, Foster City, CA).

Sequence analyses for mutations in *FLT3* (ITD and tyrosine kinase domain [TKD] mutation), *N-RAS*, *K-RAS*, and *CEBPA* were performed as described previously.^{12,30,31}

Gene expression profiling and unsupervised cluster analyses

A total of 285 AML cases were analyzed using Affymetrix HGU133A GeneChips (Affymetrix, Santa Clara, CA).²¹ Unsupervised cluster analysis on the basis of the gene expression profiles of the 285 cases of AML was performed using the correlation view tool (version 3.6) of OmniViz (Maynard, MA).²¹ The Pearson correlation values calculated in OmniViz were subsequently imported into the MicroArray Data Explorer (MADEx), which was developed in our laboratory. MADEx was used to visualize the relations between the OmniViz unsupervised clustering results and other parameters, such as clinical and molecular characteristics of the AML patients (Figure 1). MADEx is a database system that stores, mines, and visualizes microarray data in a secure and scalable manner.

Significance analysis of microarrays (SAM)

All supervised analyses were performed using significance analysis of microarrays (SAM; version 1.21).³² A threshold was set for a minimum change in expression of at least 1.5-fold. A false discovery rate (FDR) of 0.01 was used to select the differentially expressed genes.

Prediction analysis of microarrays (PAM)

All supervised class prediction analyses were performed by applying prediction analysis of microarrays (PAM; version 2.0).³³ The gene signature was selected based on the smallest prediction error in the training set and was subsequently tested using the test set. The positive predictive value was calculated as follows: true positives/(true positives + false positives).

Statistical analyses of survival

Cytogenetic abnormalities were categorized in 3 cytogenetic groups for statistical analyses. Patients with *inv(16)/t(16;16)*, *t(8;21)*, and *t(15;17)* abnormalities were considered as being in the favorable-risk category. The unfavorable-risk category was defined by the presence of *-5del(5q)*, *-7del(7q)*, *t(6;9)*, *t(9;22)*, *3q26* abnormality, or complex karyotype (more than 3 abnormalities). All other patients were classified as intermediate risk. Statistical analyses were performed with Stata Statistical Software, Release 7.0 (Stata, College Station, TX). Actuarial probabilities of overall survival (OS) (with failure defined as death due to any cause) and event-free survival (EFS) (with failure defined as not achieving complete remission [set at day 1], relapse, or death in first complete remission) were estimated by the method of Kaplan and Meier. The Cox proportional hazards model was applied to determine the association of *NPM1* mutation with OS, EFS, and disease-free survival (DFS) without and with adjustment for other factors such as cytogenetic risk, age, WBC count, and *FLT3* ITD. All tests were 2-sided, and a *P* of less than .05 was considered statistically significant.

Results

Different *NPM1* variant mutations in AML

The presence of *NPM1* mutations in 275 cases of primary AML was rapidly and reliably detected by dHPLC WAVE. Nucleotide sequencing was performed on those cases with an abnormal dHPLC profile (Table 1). Each *NPM1* mutation variant reveals a specific dHPLC WAVE profile. Thus, each type of *NPM1* mutation could be predicted on the basis of a specific dHPLC WAVE profile.

In addition, 3 novel *NPM1* mutant variants were identified (*NPM1* mutants I to K [Table 1]). These rare variants have comparable 4 bp insertions, like *NPM1* variant mutations A to D,¹³

Table 1. NPM1 mutation frequencies in 275 cases of de novo AML

NPM1 variant	Nucleotide sequence*		Protein sequence†	No. of NPM1 mutants (%)
WT	GATCTCTG	GCAGTGGAGGAAGTCTCTTTAAGAAAATAG	-DLWQWRKSL	NA
Mutant A	GATCTCTG <u>TCTGG</u> CAGTGGAGGAAGTCTCTTTAAGAAAATAG		-DLCLAVEEVSLRK	72 (26)
Mutant B	GATCTCTG <u>CATGG</u> CAGTGGAGGAAGTCTCTTTAAGAAAATAG		-DLCMAVEEVSLRK	12 (4)
Mutant D	GATCTCTG <u>CCTGG</u> CAGTGGAGGAAGTCTCTTTAAGAAAATAG		-DLCLAVEEVSLRK	4 (1)
Mutant I‡	GATCTCTG <u>CAGAG</u> CAGTGGAGGAAGTCTCTTTAAGAAAATAG		-DLCRAVEEVSLRK	1 (< 1)
Mutant J‡	GATCTCTG <u>CTTGG</u> CAGTGGAGGAAGTCTCTTTAAGAAAATAG		-DLCLAVEEVSLRK	1 (< 1)
Mutant K‡	GATCTCTG <u>TATGG</u> CAGTGGAGGAAGTCTCTTTAAGAAAATAG		-DLCMAVEEVSLRK	1 (< 1)

Number of NPM1 mutants with undetermined variants, 4 (1%); total number of NPM1 mutants, 95 (35%).

*Inserted nucleotides are underlined and italicized.

†The C-terminal tryptophane residues (W) are underlined for wild-type (WT) NPM1.

‡NPM1 mutants I, J, and K are novel variants identified in addition to the 6 known NPM1 variants.¹³

resulting in a frame shift and replacement of the 7 C-terminal amino acids of the NPM1 protein by 11 different residues (Table 1).

NPM1 mutations in relation to clinical and molecular features in AML

The NPM1 mutation frequencies of the 275 cases of primary AML with regard to clinical parameters, morphology, cytogenetics, and molecular characteristics are shown in Table 2. NPM1 mutations are significantly less frequently present in patients of younger age (less than 35 years, $P < .001$). The mean age of patients with AML and NPM1 mutations is 47.3 years (± 10.7 years), whereas the mean age of patients with AML and wild-type NPM1 is 39.7 years (± 13.3 years). NPM1 mutations are seen in AML French-American-British (FAB) subtypes M1 to M6 but are absent in AML FAB M0. The mutations are relatively frequently found in AML FAB M5 as well as in all 3 cases of AML FAB M6. In AML with recurrent translocations (ie, t(8;21), inv(16), and t(15;17)) no NPM1 mutations were demonstrated. In cases with various other cytogenetic abnormalities, mutations in NPM1 were also rare. As a result, there appears to be a positive correlation between NPM1 mutations and AML with normal karyotypes ($P < .001$). The analysis of NPM1 mutations reveals interesting relationships with particular common molecular abnormalities. The presence of NPM1 mutations significantly correlates with the presence of FLT3 ITD mutations ($P < .001$). A correlation between NPM1 mutations and FLT3 TKD mutations is not apparent. AML with NRAS mutations showed a significantly lower percentage of NPM1 mutations ($P = .024$). Among 5 of 8 cases of AML with NPM1 mutations, KRAS mutations were found. No mutations in NPM1 were found in cases of AML with CEBPA mutations ($P = .001$).

NPM1 mutations occur within specific AML subtypes defined by gene expression profiling

Of the cohort of 285 cases of primary AML that had previously been profiled using the Affymetrix HGU133A GeneChip²¹ and for which 16 distinct expression signatures had been defined following unsupervised cluster analyses, we have now examined 275 cases for the presence of NPM1 mutations. Among these pre-established signatures, the AML cases with NPM1 mutations aggregate within particular clusters (Figure 1). Most AML cases in cluster nos. 2, 3, 5, and 7 carry NPM1 mutations. In fact, all cases of AML of cluster nos. 6 (100% FLT3 ITD) and 11 (78% normal karyotypes) carried mutations in NPM1. Although cluster nos. 7 and 8 have comparable expression profiles (Figure 1), 13 of the 18 AML cases in cluster no. 7 (72%) and only 1 of 12 cases of cluster no. 8 (8%) reveal NPM1 mutations. The clusters merely consisting of AML with inv(16) (no. 9), t(15;17) (no. 12), or t(8;21) (no. 13) as well as the clusters predominantly containing cases with CEBPA mutations

(nos. 4 and 15) all lack NPM1 mutations. The subset of AML patients in cluster no. 10, with adverse prognosis and an expression profile comparable to CD34⁺ cells,²¹ did not present with NPM1 mutations.

Falini and colleagues¹³ had shown a negative correlation between the presence of NPM1 mutations and CD34 expression levels. In fact, by plotting the CD34 mRNA expression levels of the 285 AML cases, as determined by the GeneChip analyses, along with the NPM1 mutation status (Figure 1), a distinguishable association of CD34 mRNA expression and NPM1 mutation is apparent (ie, CD34 mRNA expression is low or absent in cases of AML with NPM1 mutations, while CD34 mRNA levels are high in AML cases without NPM1 mutations).

HOX gene-specific gene expression signature of NPM1 mutant cases

To identify genes with significant differential expression between primary AML samples with NPM1 mutations (n = 95) and samples without NPM1 mutations (n = 180) the SAM algorithm was used.³² A fold change threshold of 1.5 for up-regulation of gene expression and 0.667 for down-regulation of gene expression was applied. With an FDR of 0.01, 569 probe sets representing 440 unique genes were identified as being significantly differentially expressed (Figure S1 and Table S1, available on the Blood website; see the Supplemental Materials link at the top of the online article). The identity of the top 50 genes, in some cases represented by multiple probe sets, is depicted in Table 3.

A dominant homeobox (HOX) gene-specific signature is strongly associated with AML carrying an NPM1 mutation. Moreover, the expression of members of the HOXA and HOXB gene families, but also the HOX gene-related three-amino acid loop extension (TALE) genes, PBX3 and MEIS1, is increased. In contrast, the CD34 gene is the strongest significantly down-regulated gene with regard to NPM1 mutation in the AML patient cohort.

NPM1 mutant cases are predicted with high accuracy based on their gene expression signature

NPM1 mutation prediction analyses were performed using the PAM algorithm.³³ All 275 primary AML samples were randomly assigned to a training set, consisting of 122 samples without NPM1 mutations and 62 samples with NPM1 mutations, and a validation series, consisting of 58 samples lacking the NPM1 mutation and 33 samples with mutations in NPM1. Cross-validation to predict the mutation status of NPM1 on the training set resulted in 100% correct predictions of presence of mutation (sensitivity) and 80.3% correct predictions of absence of mutation (specificity) (Table 4). Prediction of an independent validation set also resulted in 100% correct prediction of presence of mutation and 82.7% correct

Table 2. *NPM1* mutation frequencies in relation to clinical parameters, morphology, cytogenetics, and molecular characteristics of the 275 patients with de novo AML

	No.	No. of <i>NPM1</i> mutants (%)	<i>P</i>
Sex			
Male	135	40 (30)	.100
Female	140	55 (39)	
Age, y			
Younger than 35	74	11 (15)	< .001
35 to 60	169	69 (41)	
60 and older	32	15 (47)	
WBC count, × 10⁹/L			
20 or below	113	28 (25)	< .001
Above 20	157	66 (42)	
ND	5	1 (20)	
FAB			
M0	6	0 (0)	ND
M1	62	21 (34)	> .999
M2	63	18 (29)	.296
M3	17	1 (6)	.008
M4	49	15 (31)	.062
M5	65	32 (49)	.007
M6	3	3 (100)	ND
ND	10	5 (50)	
Cytogenetic abnormalities*			
t(15;17)	16	0 (0)	.002
t(8;21)	21	0 (0)	< .001
inv(16)/t(16;16)	17	0 (0)	.001
+8	24	5 (21)	ND
+11	5	0 (0)	ND
+21	1	1 (100)	ND
-5	2	0 (0)	ND
-5(q)	1	0 (0)	ND
-7	13	0 (0)	.005
-7(q)	7	0 (0)	ND
3q	5	1 (20)	ND
t(6;9)	4	0 (0)	ND
t(9;22)	2	1 (50)	ND
t(11q23)	16	1 (6)	.014
Complex; more than 3 abnormalities	11	0 (0)	.018
Other	55	10 (18)	.004
Normal	116	74 (64)	< .001
ND	10	6 (60)	
Molecular abnormalities			
<i>FLT3</i> ITD	78	47 (60)	< .001
<i>FLT3</i> TKD	32	14 (44)	.243
<i>NRAS</i>	25	3 (12)	.024
<i>KRAS</i>	8	5 (63)	.130
<i>CEBPA</i>	17	0 (0)	.001

P values were calculated using the Fisher exact test (2-tailed; ND indicates not determined).

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

prediction of absence of mutation. The positive predictive value in this cohort is 72% in the training set and 74% overall. As expected, the genes included in the PAM gene signature were among the most significant differentially expressed genes as determined by SAM, thereby validating both algorithms. Of note, *NPM1* mRNA expression did not correlate with mutation status (Figure 2). The *NPM1* mutant AML cases have a distinct signature with regard to the 18 selected genes (Figure 2) and are therefore predicted with high accuracy (Table 4). However, these 18 genes seem to be expressed at comparable levels in a subset of AML cases with wild-type *NPM1* (Figure 2).

NPM1 mutation is an independent favorable prognostic marker

To investigate the prognostic value of *NPM1* mutation, 252 AML patients with long-term follow-up following therapy completion were included for survival analysis. The EFS, OS, and probability of relapse at 60 months for the AML patients with or without *NPM1* mutations were similar (Figure 3A-B). Likewise, among the subgroup with cytogenetics of intermediate prognostic risk, EFS, OS, and probability of relapse at 60 months were not different (Figures 3C-D), although there appears a trend for more favorable outcome for patients with AML with *NPM1* mutations. Because *NPM1* mutations are significantly associated with both normal karyotype and presence of a *FLT3* ITD (Table 2), we investigated the prognostic value of *NPM1* mutations within the intermediate cytogenetic risk group in relation to *FLT3* ITD status. Patients in the intermediate cytogenetic risk group without *FLT3* ITD mutations but with *NPM1* mutations have a significantly better OS and EFS than those without *NPM1* mutations (*P* = .05) (Figure 4A-C). In intermediate cytogenetic risk AML with *FLT3* ITD mutations, *NPM1* mutations do not significantly distinguish prognosis (Figure 4B,D).

NPM1 mutations are asynchronously associated with particular karyotypes and molecular abnormalities that might express additional positive or negative prognostic value and therefore might obscure the prognostic significance of *NPM1* mutations (Table 2; Figure 1). Therefore we investigated the prognostic value of *NPM1* mutations in both univariable and multivariable analyses. Univariable and multivariable Cox regression analyses were applied to assess the prognostic significance of *NPM1* mutation, cytogenetic risk class, WBC count below or above $20 \times 10^9/L$, age, and *FLT3*

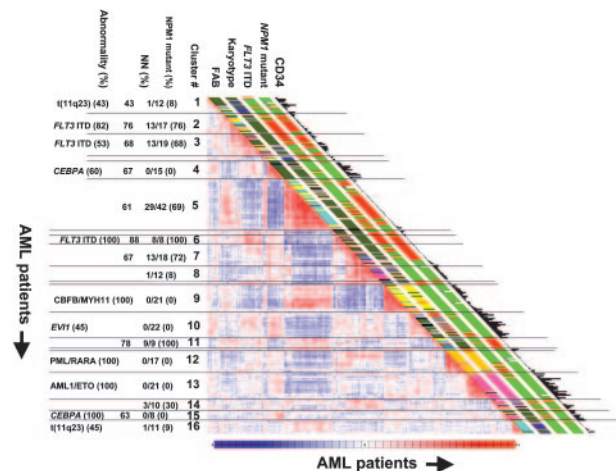


Figure 1. OmniViz correlation view of 285 AML patients. The correlation view displays pairwise correlations between AML patients. The cells in the visualization are colored by Pearson correlation coefficient values, with deeper colors indicating higher positive (red) or negative (blue) correlations. The scale bar indicates maximum positive correlation (red) toward maximum negative correlation (blue). The 16 clusters identified in the cohort of 285 AML patients using 2856 probe sets on the basis of the correlation view are indicated (1 to 16).²¹ Clinical and molecular data are depicted in the columns along the original diagonal of the correlation view.²¹ FAB classification and karyotype based on cytogenetics are depicted in the first 2 columns (FAB M0, red; M1, green; M2, purple; M3, orange; M4, yellow; M5, blue; M6 gray; and karyotype: normal, green; inv(16), yellow; t(8;21), purple; t(15;17), orange; 11q23 abnormalities, blue; 7(q) abnormalities, red; +8, pink; complex, black; other, gray). *FLT3* ITD and *NPM1* mutations are depicted in the same set of columns (red bar, positive; green bar, negative). The expression levels of *CD34* (probe set: 209543_s_at) in the 285 AML patients are plotted in the last column (bars are proportional to the level of expression). The percentages of the most common (more than 40%) abnormalities, *NPM1* mutations, as well as normal karyotypes (NN) for each cluster are indicated.

Table 3. *NPM1* mutation-associated gene expression in 275 patients with de novo AML

Probe set ID	Gene symbol	Fold change
Up-regulated in AML with mutant <i>NPM1</i>		
213844_at	<i>HOXA5</i>	4.2
205366_s_at	<i>HOXB6</i>	2.6
208414_s_at	<i>HOXB3</i>	1.8
204082_at	<i>PBX3</i>	2.8
205600_x_at	<i>HOXB5</i>	2.1
206289_at	<i>HOXA4</i>	2.1
205453_at	<i>HOXB2</i>	2.1
213150_at	<i>HOXA10</i>	2.6
204069_at	<i>MEIS1</i>	2.6
209905_at	<i>HOXA9</i>	2.8
201664_at	<i>SMC4L1</i>	2.1
209439_s_at	<i>PHKA2</i>	1.6
206847_s_at	<i>HOXA7</i>	1.5
63825_at	<i>ABHD2</i>	1.6
219304_s_at	<i>PDGFD</i>	1.9
212820_at	<i>RC3</i>	2.4
213110_s_at	<i>COL4A5</i>	2.9
207111_at	<i>EMR1</i>	2.1
208557_at	<i>HOXA6</i>	1.6
203471_s_at	<i>PLEK</i>	1.6
203680_at	<i>PRKAR2B</i>	2.3
202729_s_at	<i>LTBP1</i>	3.6
210145_at	<i>PLA2G4A</i>	1.6
220162_s_at	<i>CARD9</i>	1.5
206298_at	<i>ARHGAP22</i>	2.0
219602_s_at	<i>FAM38B</i>	1.9
Down-regulated in AML with mutant <i>NPM1</i>		
209543_s_at	<i>CD34</i>	-5.4
206896_s_at	<i>GNG7</i>	-1.7
209583_s_at	<i>MOX2</i>	-3.1
200953_s_at	<i>CCND2</i>	-2.1
221004_s_at	<i>ITM2C</i>	-3.0
205330_at	<i>MN1</i>	-4.3
200602_at	<i>APP</i>	-2.6
200665_s_at	<i>SPARC</i>	-3.8
201015_s_at	<i>JUP</i>	-3.1
218899_s_at	<i>BAALC</i>	-4.3
209679_s_at	<i>LOC57228</i>	-1.9
219694_at	<i>FLJ11127</i>	-2.1
206042_x_at	<i>SNRPN</i>	-2.5
211535_s_at	<i>FGFR1</i>	-2.6
214582_at	<i>PDE3B</i>	-1.8
221523_s_at	<i>RRAGD</i>	-1.8
213618_at	<i>CENTD1</i>	-1.7
218589_at	<i>P2RY5</i>	-3.3
202016_at	<i>MEST</i>	-2.9
208116_s_at	<i>MAN1A1</i>	-3.1
205240_at	<i>GPSM2</i>	-1.9
202747_s_at	<i>ITM2A</i>	-4.1
206622_at	<i>TRH</i>	-9.1
206726_at	<i>PGDS</i>	-4.7

The top 50 unique most discriminating genes and fold change in expression with regard to *NPM1* mutation as determined by SAM (up-regulated: increased expression in *NPM1* mutant cases; down-regulated: decreased expression in *NPM1* mutant cases).

ITD mutation for OS, DFS, and EFS (Table 5). In univariable analyses *NPM1* mutation showed no statistical significance with respect to the end points. However, in multivariable analyses mutated *NPM1* was significantly associated with a much lower hazards ratio (HR), which was statistically significant for all end points (EFS: HR = 0.59, *P* = .005; DFS: HR = 0.52, *P* = .003; and OS: HR = 0.49, *P* < .001). Mutant *NPM1* appeared as an

Table 4. *NPM1* mutation prediction by using PAM

Genotype	Predicted genotype			
	10-fold CV error*		Error validation set*	
	WT	Mutant	WT	Mutant
WT <i>NPM1</i>	98	24	48	10
Mutant <i>NPM1</i>	0	62	0	33

Most optimal result for *NPM1* mutation prediction using a cohort of 275 cases of AML divided in a training and test set.²¹ The number of probe sets used in this prediction was 22, which represents 18 unique genes. For the 10-fold CV error set, n = 184; for the error validation set, n = 91.

*Ten-fold cross-validation (CV) prediction error on training set (n = 184); prediction error on validation set (n = 91). In 10-fold cross-validation, the model is fitted on 90% of the samples and the class of the remaining 10% is predicted. This procedure is repeated 10 times, with each part playing the role of the test samples and the error of all 10 parts added together to compute the overall error. The error within the validation set reflects the number of samples wrongfully predicted in this set.

independent prognostic marker in addition to cytogenetic risk, age, and *FLT3* ITD mutations.

Discussion

Recently, we established a comprehensive classification of AML based on gene expression profiling.²¹ In this classification several clusters of AML signatures correlated with distinct (cyto)genetic abnormalities, such as t(8;21), t(15;17), inv(16), and C- and N-terminal mutations in CEBPA. However, the common underlying molecular abnormality for the other subtypes of AML was unknown. In the present study we show that 2 clusters consist entirely of AML cases with *NPM1* mutations (ie, cluster nos. 6 and 11), whereas cluster nos. 2, 3, 5, and 7 predominantly include patients with *NPM1* mutations. Thus, we identify mutant *NPM1* as a common molecular abnormality in these subtypes of AML.

Falini and colleagues¹³ have shown that mutations in *NPM1* are present in 35% of patients with AML. In this study, we confirm that

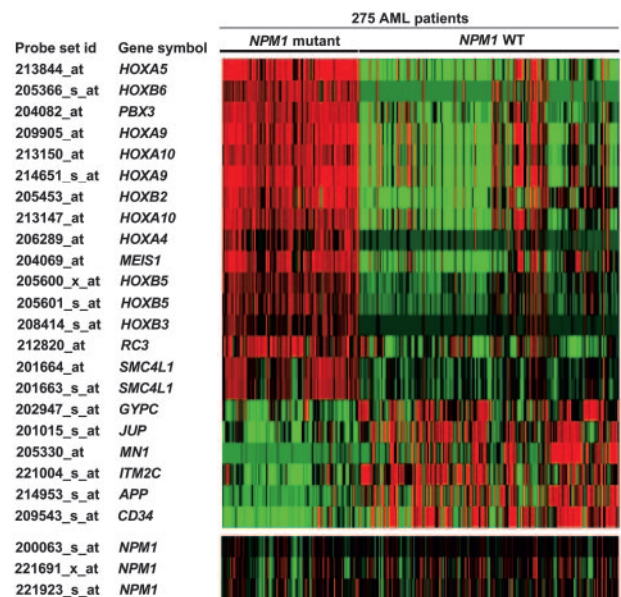


Figure 2. The most predictive molecular signature with regard to *NPM1* mutation in AML assessed with 22 probe sets representing 18 unique genes. The levels of expression of the probe sets in the 275 cases of AML are depicted (scale bar indicates an increase [red] or a decrease [green] in the level of expression of at least 4 relative to the geometric mean).²¹ The 3 probe sets representing the *NPM1* gene are also depicted.

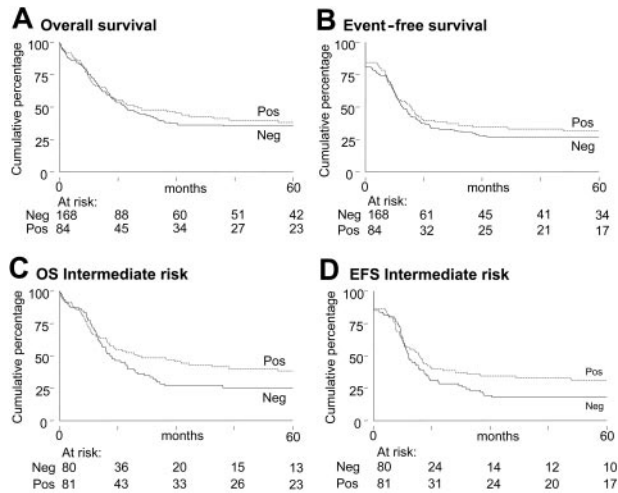


Figure 3. Survival analyses of AML patients with and without *NPM1* mutations. Kaplan-Meier estimates of OS (A) and EFS (B) among patients with AML and OS (C) and EFS (D) among patients with AML with intermediate-risk karyotypes.

NPM1 mutations are frequently present in AML (ie, in 35% of all cases). *NPM1* mutant is less frequently represented in patients with AML of age younger than 35 years. This seems consistent with the tendency of the *NPM1* mutation to be more frequently present in older children with AML.³⁴ Furthermore, *NPM1* mutations are significantly associated with AML with high WBC count.

We detected the various mutation variants of *NPM1* at similar frequencies as was described recently¹³ and also identified 3 novel mutant variants. These novel variants carry, like the other *NPM1* mutant variants in our study, an insertion of 4 bp, resulting in a protein with an altered C-terminus (Table 1). Falini and colleagues¹³ suggested that the disruption of 1 of the 2 C-terminal tryptophan residues and the last 5 residues (ie, VSLRK) are important for *NPM1* mutant function. Our study suggests that the final 9 amino acids (ie, AVEEVSLRK) may in general be required for mutant *NPM1* function.

NPM1 mutations often coincide with mutations in *FLT3*, particularly with the ITD-type mutations. Our data may suggest an association of mutations in *KRAS* and *NPM1*, but the study has limited power because of the small number of mutant *KRAS* AMLs. In contrast, mutations in *NRAS* were not associated with mutant *NPM1*, because *NRAS* mutations are found in AML with inv(16),^{21,30} a subclass of AML that lacks *NPM1* mutations. In addition, we did not find *NPM1* mutations in clusters of AML patients, previously identified by gene expression profiling,²¹ characterized by C- and N-terminal mutations in CEBPA. These observations might perhaps suggest that constitutive active *FLT3* or K-RAS provides the proliferative signal, whereas mutant *NPM1* might serve to impair differentiation in the multistep pathogenesis model of AML.³⁵

Notably, the discriminative genes identified by SAM and PAM revealed a strong *HOX* and TALE gene-specific signature associated with AML cases with mutant *NPM1*, as was published recently.²⁴ Thus, although CD34 has generally been used as marker for immature hematopoietic progenitor cells, the *NPM1* mutant CD34⁻ cells display a molecular signature similar to that of hematopoietic stem cells (HSCs). Recent studies have shown that CD34⁻ HSCs exist as well.^{36,37} These CD34⁻ cells also possess HSC-specific characteristics, including the ability for hematopoietic engraftment.^{36,37} This brings up the question as to whether the *NPM1* mutant AML cells in fact represent a more primitive population of HSCs with a *HOX* gene signature.

The sharp distinction between CD34 positivity and *NPM1* mutations in AML cases based on the *HOX* gene signature is notable because 22 of the 39 *HOX* genes are expressed in human CD34⁺ cells, whereas during normal differentiation *HOX* gene expression declines.³⁸ Extensive studies have demonstrated for a number of *HOX* genes that sustained overexpression in murine bone marrow results in perturbations in the stem cell pools, and coexpression of certain *HOX* gene family members with their protein binding partner, such as MEIS1, results in leukemia.³⁸ Thus, in AML with *NPM1* mutations the hematopoietic progenitor cells may have arrested at a differentiation stage with endogenous coexpression of the *HOX* genes (ie, *HOXA5*, *-A9*, *-A10*, *-B2*, *-B3*, *-B5*, and *-B6*) and their TALE partner genes (ie, *MEIS1* and *PBX3*) or as a result of the increased expression of these genes.

SAM and PAM analyses were highly concordant for the genes identified with differential expression between AML with mutant and AML with wild-type *NPM1*, where, in both cases, *CD34* was the most discriminating gene, down-regulated in *NPM1* mutants. Patients harboring an *NPM1* mutation can be predicted with high accuracy; however, a subset of patients with wild-type *NPM1* is wrongly predicted. Interestingly, within this subgroup the percentage of patients with 11q23 abnormalities is significantly increased. In fact, this may not be surprising because MLL has been implicated as a *HOX* gene regulator³⁹ and selective expression of *HOX* genes in ALL cases with mutant MLL has been shown.⁴⁰

By univariate analyses we show that there is a tendency that the presence of an *NPM1* mutation, and concomitant low *CD34* mRNA expression, is a favorable marker for clinical outcome (Figure 3). This is in agreement with CD34 expression as an indicator for poor response to induction therapy.⁴¹⁻⁴³ In addition, we demonstrate by multivariate analyses that *NPM1* mutations are a strong independent favorable predictive marker for EFS, DFS, and OS in AML. The finding that the effect of *NPM1* mutation is much more pronounced in multivariable than in univariable analyses can be explained by the strong correlation between *NPM1* and *FLT3* ITD mutations and additionally by the association with high WBC count and age. *NPM1* mutations are more frequently present in AML patients with *FLT3* ITD, high WBC count, and older age. These factors are unfavorable determinants of prognosis,⁴⁴ while *NPM1* mutation seems to express a favorable prognostic

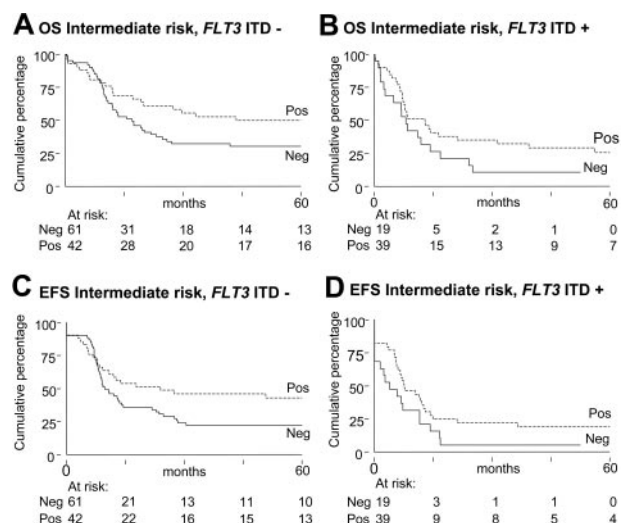


Figure 4. Survival analyses of intermediate risk AML patients with and without *FLT3* ITD and/or *NPM1* mutations. Kaplan-Meier estimates of OS (panels A and B) and EFS (panels C and D) for patients with intermediate-risk AML and *FLT3* ITD mutations (*FLT3* ITD positive [panels B and D]) versus those with intermediate-risk AML without *FLT3* ITD mutations (*FLT3* ITD negative [panels A and C]).

Table 5. Univariable and multivariable analyses of *NPM1* mutation as prognostic factor for EFS, DFS, and OS in AML

	EFS		DFS		OS	
	HR (95% CI)	P*	HR (95% CI)	P*	HR (95% CI)	P*
Univariable						
Intermediate†	2.06 (1.33-3.20)	.001*	2.82 (1.63-4.87)	< .001*	2.67 (1.62-4.41)	< .001*
Poor†	3.88 (2.29-6.56)	< .001*	5.09 (2.61-9.91)	< .001*	4.44 (2.49-7.93)	< .001*
WBC count‡	1.14 (0.84-1.53)	.39	1.18 (0.84-1.68)	.34	1.19 (0.87-1.62)	.29
Age, decades	1.05 (0.94-1.19)	.38	1.06 (0.92-1.22)	.41	1.17 (1.03-1.32)	.014*
<i>FLT3</i> ITD§	1.76 (1.28-2.41)	.001*	1.71 (1.17-2.50)	.006*	1.89 (1.36-2.62)	< .001*
<i>NPM1</i> mutation	0.88 (0.64-1.21)	.43	0.93 (0.64-1.34)	.69	0.90 (0.65-1.27)	.57
Multivariable						
Intermediate†	2.18 (1.36-3.50)	.001*	3.11 (1.74-5.55)	< .001*	2.87 (1.69-4.86)	< .001*
Poor†	3.96 (2.34-6.74)	< .001*	5.73 (2.93-11.22)	< .001*	4.68 (2.60-8.43)	< .001*
WBC count‡	1.06 (0.78-1.44)	.72	1.10 (0.77-1.58)	.60	1.18 (0.85-1.63)	.32
Age, decades	1.07 (0.95-1.21)	.27	1.10 (0.95-1.27)	.21	1.19 (1.05-1.35)	.008*
<i>FLT3</i> ITD§	1.96 (1.39-2.76)	< .001*	2.00 (1.31-3.06)	.001*	2.13 (1.49-3.05)	< .001*
<i>NPM1</i> mutation	0.59 (0.41-0.85)	.005*	0.52 (0.34-0.80)	.003*	0.49 (0.33-0.72)	< .001*

CI indicates confidence interval.

*P values < .05.

†Cytogenetic risk versus cytogenetic good risk.

‡More than 20 × 10⁹/L versus less than 20 × 10⁹/L.

§*FLT3* ITD versus no *FLT3* ITD.

||*NPM1* mutation versus no *NPM1* mutation.

impact. This implies that in univariable analysis the positive effect of *NPM1* mutations, as measured by the method of Kaplan and Meier or by the HR, is less pronounced, because the adverse effects of *FLT3* ITD, high WBC count, and age might mask this effect. In fact, *NPM1* mutations distinguish a favorable subgroup among intermediate cytogenetic risk *FLT3* ITD-negative AML that shows comparatively better OS and EFS (Figure 4). In addition, in the multivariable analyses the effect of *NPM1* mutations is more pronounced than is apparent in the univariable analyses. This is not only true for *NPM1* mutations but also for the independent prognostic effect of *FLT3* ITD mutations, age, and cytogenetic risk (Table 5).

The data presented here demonstrate that the frequent C-terminal insertion mutations in *NPM1* correlate with favorable

outcome for patients with AML. Because *NPM1* mutations are predominantly found in patients with standard-risk AML, determination of the *NPM1* mutation status, in combination with other prognostically relevant markers, will be useful for further risk stratification of adult patients with de novo AML.

Acknowledgments

We are indebted to our colleagues from the bone marrow transplantation group and molecular diagnostics laboratory of the Department of Hematology at the Erasmus Medical Center for storage of the samples and molecular analyses, respectively.

References

- Lowenberg B, Downing JR, Burnett A. Acute myeloid leukemia. *N Engl J Med*. 1999;341:1051-1062.
- Slovak ML, Kopecky KJ, Cassileth PA, et al. Karyotypic analysis predicts outcome of pre-remission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. *Blood*. 2000;96:4075-4083.
- Byrd JC, Mrozek K, Dodge RK, et al. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood*. 2002;100:4325-4336.
- Grimwade D, Walker H, Oliver F, et al. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood*. 1998;92:2322-2333.
- Grimwade D, Walker H, Harrison G, et al. The predictive value of hierarchical cytogenetic classification in older adults with acute myeloid leukemia (AML): analysis of 1065 patients entered into the United Kingdom Medical Research Council AML11 trial. *Blood*. 2001;98:1312-1320.
- Gilliland DG, Griffin JD. The roles of *FLT3* in hematopoiesis and leukemia. *Blood*. 2002;100:1532-1542.
- Levis M, Small D. *FLT3*: ITDoes matter in leukemia. *Leukemia*. 2003;17:1738-1752.
- Shiah HS, Kuo YY, Tang JL, et al. Clinical and biological implications of partial tandem duplication of the *MLL* gene in acute myeloid leukemia without chromosomal abnormalities at 11q23. *Leukemia*. 2002;16:196-202.
- Dohner K, Tobis K, Ulrich R, et al. Prognostic significance of partial tandem duplications of the *MLL* gene in adult patients 16 to 60 years old with acute myeloid leukemia and normal cytogenetics: a study of the Acute Myeloid Leukemia Study Group Ulm. *J Clin Oncol*. 2002;20:3254-3261.
- Barjesteh van Waalwijk van Doorn-Khosrovani S, Erpelinck C, van Putten WL, et al. High *EV11* expression predicts poor survival in acute myeloid leukemia: a study of 319 de novo AML patients. *Blood*. 2003;101:837-845.
- Preudhomme C, Sagot C, Boissel N, 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;100:2717-2723.
- van Waalwijk van Doorn-Khosrovani SB, Erpelinck C, Meijer J, et al. Biallelic mutations in the *CEBPA* gene and low *CEBPA* expression levels as prognostic markers in intermediate-risk AML. *Hematol J*. 2003;4:31-40.
- Falini B, Mecucci C, Tiacci E, et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med*. 2005;352:254-266.
- Dumbar TS, Gentry GA, Olson MO. Interaction of nucleolar phosphoprotein B23 with nucleic acids. *Biochemistry*. 1989;28:9495-9501.
- Cordell JL, Pulford KA, Bigerna B, et al. Detection of normal and chimeric nucleophosmin in human cells. *Blood*. 1999;93:632-642.
- Borer RA, Lehner CF, Eppenberger HM, Nigg EA. Major nucleolar proteins shuttle between nucleus and cytoplasm. *Cell*. 1989;56:379-390.
- Morris SW, Kirstein MN, Valentine MB, et al. Fusion of a kinase gene, *ALK*, to a nucleolar protein gene, *NPM*, in non-Hodgkin's lymphoma. *Science*. 1994;263:1281-1284.
- Redner RL, Rush EA, Faas S, Rudert WA, Corey SJ. The t(5;17) variant of acute promyelocytic leukemia expresses a nucleophosmin-retinoic acid receptor fusion. *Blood*. 1996;87:882-886.
- Yoneda-Kato N, Look AT, Kirstein MN, et al. The t(3;5)(q25.1;q34) of myelodysplastic syndrome and acute myeloid leukemia produces a novel fusion gene, *NPM-MLF1*. *Oncogene*. 1996;12:265-275.
- Valk PJ, Delwel R, Lowenberg B. Gene expression profiling in acute myeloid leukemia. *Curr Opin Hematol*. 2005;12:76-81.
- Valk PJ, Verhaak RG, Beijien MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med*. 2004;350:1617-1628.

22. Bullinger L, Dohner K, Bair E, et al. Use of gene-expression profiling to identify prognostic sub-classes in adult acute myeloid leukemia. *N Engl J Med*. 2004;350:1605-1616.
23. Ross ME, Mahfouz R, Onciu M, et al. Gene expression profiling of pediatric acute myelogenous leukemia. *Blood*. 2004;104:3679-3687.
24. Alcalay M, Tiacci E, Bergomas R, 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;106:899-902.
25. Lowenberg B, Boogaerts MA, Daenen SM, et al. Value of different modalities of granulocyte-macrophage colony-stimulating factor applied during or after induction therapy of acute myeloid leukemia. *J Clin Oncol*. 1997;15:3496-3506.
26. Lowenberg B, van Putten W, Theobald M, et al. Effect of priming with granulocyte colony-stimulating factor on the outcome of chemotherapy for acute myeloid leukemia. *N Engl J Med*. 2003;349:743-752.
27. Ossenkoppele GJ, Graveland WJ, Sonneveld P, et al. The value of fludarabine in addition to ARA-C and G-CSF in the treatment of patients with high-risk myelodysplastic syndromes and AML in elderly patients. *Blood*. 2004;103:2908-2913.
28. Van der Reijden BA, de Wit L, van der Poel S, et al. Identification of a novel CBFB-MYH11 transcript: implications for RT-PCR diagnosis. *Hematol J*. 2001;2:206-209.
29. Choy YS, Dabora SL, Hall F, et al. Superiority of denaturing high performance liquid chromatography over single-stranded conformation and conformation-sensitive gel electrophoresis for mutation detection in TSC2. *Ann Hum Genet*. 1999;63(pt 5):383-391.
30. Valk PJM, Bowen DT, Frew ME, Goodeve AC, Löwenberg B, Reilly JT. Second hit mutations in the RTK/RAS signalling pathway in acute myeloid leukaemia and inv(16) [letter]. *Haematologica*. 2004;89:106.
31. Care RS, Valk PJ, Goodeve AC, et al. Incidence and prognosis of c-KIT and FLT3 mutations in core binding factor (CBF) acute myeloid leukemias. *Br J Haematol*. 2003;121:775-777.
32. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*. 2001;98:5116-5121.
33. 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;99:6567-6572.
34. Cazzaniga G, Dell'oro MG, Mecucci C, et al. Nucleophosmin mutations in childhood acute myelogenous leukemia with normal karyotype. *Blood*. 2005;106:1419-1422.
35. Gilliland DG. Molecular genetics of human leukemias: new insights into therapy. *Semin Hematol*. 2002;39:6-11.
36. Engelhardt M, Lubbert M, Guo Y. CD34(+) or CD34(-): which is the more primitive? *Leukemia*. 2002;16:1603-1608.
37. Guo Y, Lubbert M, Engelhardt M. CD34- hematopoietic stem cells: current concepts and controversies. *Stem Cells*. 2003;21:15-20.
38. Grier DG, Thompson A, Kwasniewska A, McGonigle GJ, Halliday HL, Lappin TR. The pathophysiology of HOX genes and their role in cancer. *J Pathol*. 2005;205:154-171.
39. Daser A, Rabbitts TH. The versatile mixed lineage leukaemia gene MLL and its many associations in leukaemogenesis. *Semin Cancer Biol*. 2005;15:175-188.
40. Armstrong SA, Staunton JE, Silverman LB, et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat Genet*. 2002;30:41-47.
41. Chang H, Salma F, Yi QL, Patterson B, Brien B, Minden MD. Prognostic relevance of immunophenotyping in 379 patients with acute myeloid leukemia. *Leuk Res*. 2004;28:43-48.
42. Myint H, Lucie NP. The prognostic significance of the CD34 antigen in acute myeloid leukaemia. *Leuk Lymphoma*. 1992;7:425-429.
43. Raspadori D, Lauria F, Ventura MA, et al. Incidence and prognostic relevance of CD34 expression in acute myeloblastic leukemia: analysis of 141 cases. *Leuk Res*. 1997;21:603-607.
44. Lowenberg B. Prognostic factors in acute myeloid leukaemia. *Best Pract Res Clin Haematol*. 2001;14:65-75.