An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia

Klaus H. Metzeler,1 Manuela Hummel,2 Clara D. Bloomfield,3 Karsten Spiekermann,1,4 Jan Braess,1 Maria-Cristina Sauerland,5 Achim Heinecke,5 Michael Radmacher,3,6 Guido Marucci,3,6 Susan P. Whitman,3 Kati Maharry,3,6 Peter Paschka,3,7 Richard A. Larson,8 Wolfgang E. Berdel,9 Thomas Büchner,9 Bernhard Wörmann,10 Ulrich Mansmann,2 Wolfgang Hiddemann,1,4 Stefan K. Bohlander,1,4 and Christian Buske,1,4 for Cancer and Leukemia Group B and the German AML Cooperative Group

Patients with cytogenetically normal acute myeloid leukemia (CN-AML) show heterogeneous treatment outcomes. We used gene-expression profiling to develop a gene signature that predicts overall survival (OS) in CN-AML. Based on data from 163 patients treated in the German AMLCG 1999 trial and analyzed on oligonucleotide microarrays, we used supervised principal component analysis to identify 86 probe sets (representing 66 different genes), which correlated with OS, and defined a prognostic score based on this signature. When applied to an independent cohort of 79 CN-AML patients, this continuous score remained a significant predictor for OS (hazard ratio [HR], 1.85; P = .002), event-free survival (HR = 1.73; P = .001), and relapse-free survival (HR = 1.76; P = .025). It kept its prognostic value in multivariate analyses adjusting for age, FLT3 ITD, and NPM1 status. In a validation cohort of 64 CN-AML patients treated on CALGB study 9621, the score also predicted OS (HR = 4.11; P < .001), event-free survival (HR = 2.90; P < .001), and relapse-free survival (HR = 3.14, P < .001) and retained its significance in a multivariate model for OS. In summary, we present a novel gene-expression signature that offers additional prognostic information for patients with CN-AML. (Blood. 2008;112:4193-4201)

Introduction

Acute myeloid leukemia (AML) is a heterogeneous disorder in terms of its genetic basis, its pathophysiology, and its prognosis. In the past, the karyotype of the AML blasts has emerged as the single most important prognostic factor, and patients can be classified into favorable, intermediate, and unfavorable prognostic groups according to the presence of specific chromosomal aberrations.1,2 However, almost half of all adult AML patients present with a normal karyotype at diagnosis. These patients are usually assigned to the intermediate prognostic group. More recently, mutations in specific genes that allow the identification of prognostic subgroups in cytogenetically normal (CN) AML have been described.3 Internal tandem duplications (ITD) of the fms-like tyrosine kinase 3 (FLT3) gene are strong predictors of inferior outcome. Mutations in the nucleophosmin 1 (NPM1) gene lead to cytoplasmic localization of the NPM protein4 and are a favorable prognostic factor, especially in patients lacking FLT3 ITD.3-5 Other prognostically relevant alterations include mutations in the MLL and CEBPA genes as well as overexpression of BAALC, ERG, and EVII.3 However, in approximately 24% of CN-AML cases, none of the aforementioned mutations can be detected.5 These findings illustrate that CN-AML itself constitutes a heterogeneous disease and that there probably are many genetic alterations in CN-AML to be discovered that affect response to treatment and survival.

More recently, microarray technology has made it possible to study deregulated gene expression in AML. Several groups have used unsupervised clustering algorithms to identify subgroups of AML with different prognoses.10-12 Although such clustering results may offer insights into the biology of AML, they cannot be directly used to estimate the prognosis of newly diagnosed individual patients. Bullinger et al used CDNA microarrays and a semisupervised clustering approach to develop an outcome classifier for AML patients,13 using a dataset that contained samples with both normal and aberrant karyotypes. The prognostic utility of this signature was recently validated in an independent set of 64 CN-AML patients, using Affymetrix oligonucleotide microarrays (Affymetrix, Santa Clara, CA).14 Chromosomal abnormalities are a major prognostic factor in AML, and they strongly affect gene-expression profiles.10,14 Thus, some of the genes identified in these studies probably are surrogate markers for cytogenetic aberrations, and they may be of little value for outcome prediction.

The online version of this article contains a data supplement.
in CN-AML. We therefore investigated a cohort of CN-AML patients and present a gene signature, based on the expression of 66 genes, that was specifically developed to predict treatment outcome in patients with a normal karyotype.

Methods

Patients

In the current analysis, 3 independent cohorts of patients were studied. The training cohort consisted of 163 adult patients with CN-AML. A total of 156 (95.7%) of these patients were enrolled in the multicenter AMLCG-1999 trial of the German AML Cooperative Group between 1999 and 2003.10 All patients in the training cohort received intensive double-induction and consolidation chemotherapy. The independent test cohort consisted of 79 adult patients who were diagnosed with CN-AML in 2004. Sixty-two of them were treated according to the AMLCG-1999 protocol, whereas 17 received intensive chemotherapy outside the study. For both cohorts, pretreatment diagnostic evaluation was performed centrally at the Laboratory for Leukemia Diagnostics, University of Munich, and included standard cytomorphology, cytogenetics, fluorescence in situ hybridization, polymerase chain reaction-based testing for standard cytomorphology, cytogenetics, fluorescence in situ hybridization, polymerase chain reaction-based testing for ITD status. The HG-U133 A&B microarrays were supplied by the Laboratory for Leukemia Diagnostics, University of Munich, and included cohorts, pretreatment diagnostic evaluation was performed centrally at the Laboratory for Leukemia Diagnostics, University of Munich, and included standard cytomorphology, cytogenetics, fluorescence in situ hybridization, polymerase chain reaction-based testing for ITD status.

Gene-expression profiling and data preprocessing

Gene-expression profiling in the training cohort was performed using Affymetrix HG-U133 A&B microarrays, and samples from the test cohort were analyzed using Affymetrix HG-U133 plus 2.0 microarrays. Details regarding sample preparation, hybridization, and image acquisition have been published elsewhere.15 Karyotypes were reviewed centrally, and evaluation of FLT3 ITD and NPM1 status and gene-expression profiling were performed centrally at the Ohio State University Comprehensive Cancer Center.14 The AMLCG and CALGB clinical trials were approved by the local institutional review boards of all participating centers, and informed consent was obtained from all patients in accordance with the Declaration of Helsinki. Both the AMLCG and CALGB trials have been registered at www.clinicaltrials.gov (NCT00266136 and NCT00002925).

Gene-expression profiling

Data preprocessing

Overall survival (OS) was defined as time from study entry until death from any cause, and event-free survival (EFS) was defined as time from study entry until removal from study because of failure to achieve complete remission, relapse, or death from any cause. Relapse-free survival (RFS) was defined as time from the date of complete remission (CR) until relapse or death, regardless of cause.23 Patients alive without an event were censored at the time of their last follow-up. Clinical variables were compared between the study cohorts using Fisher exact test for categorical variables, and the Mann-Whitney U test for continuous variables.

Validation of the prognostic gene signature

In contrast to the training population, the 79 samples in the test cohort were analyzed with a newer version of the Affymetrix microarray. Probe set-wise comparison of the expression values between the 2 cohorts showed that in the test cohort (hybridized on HG-U133 plus 2.0 arrays), most probe sets tended to have lower mean signal levels and higher SDs than in the training dataset (analyzed on HG-U133 A&B chips; Figure S1). To achieve a similar distribution of expression values, we adjusted the mean expression value of each probe set in the test dataset to 0 and then scaled the SD of each probe set to the values observed in the training cohort. The 64 patients in the validation cohort were also analyzed on HG-U133 plus 2.0 microarrays. Rescaling of the expression values in the validation cohort was performed as in the test cohort. A continuous risk score was calculated for each patient in the test and validation group on the basis of the 86 selected probe sets and their respective weights, as determined in the training dataset.

Statistical analyses

Overall survival (OS) was defined as time from study entry until death from any cause, and event-free survival (EFS) was defined as time from study entry until removal from study because of failure to achieve complete remission, relapse, or death from any cause. Relapse-free survival (RFS) was defined as time from the date of complete remission (CR) until relapse or death, regardless of cause.23 Patients alive without an event were censored at the time of their last follow-up. Clinical variables were compared between the study cohorts using Fisher exact test for categorical variables, and the Mann-Whitney U test for continuous variables.

One-way analysis of variance was used to assess the association of baseline clinical and molecular characteristics with the continuous microarray risk score, and Cox proportional hazards models were used to test the prognostic value of the gene-expression score. Hazard ratios (HRs) were computed for a difference in the score equal to its interquartile range. Therefore, the HRs for the microarray-based score represent the increase in risk associated with a change in score equal to the difference between the 25th and 75th percentiles of the score values.

To assess the prognostic impact of our microarray-based score after adjusting for the influence of other clinical and molecular risk factors,
multivariate Cox regression models were constructed. The following variables, in addition to the gene-expression risk score, were included in the multivariate model based on their established prognostic relevance in CN-AML patients and the availability of complete data: age (as a continuous variable), NPM1 mutation status, and FLT3 ITD. In the test cohort, Genescan analysis of the FLT3 ITD/wt allelic ratio was available for 78 of the 79 patients, and patients were thus divided into 3 groups: those with no detectable FLT3 ITD, those with detectable FLT3 ITD and a low FLT3 ITD/wt ratio (ie, a value of ≤ 0.8), and those with a high FLT3 ITD/wt ratio (ie, a value of > 0.8). Information on FLT3 ITD/wt ratio was not available for the validation cohort; therefore, FLT3 ITD status was used as a binary variable. Because we sought to assess the prognostic value of the gene-expression score after adjustment for possible confounders, no variable selection algorithm was used, and all 4 variables were retained in the Cox regression model. The proportional hazards assumption was checked by examining Schoenfeld residuals for each variable individually. To account for possibly important unobserved covariates, we also considered models including an additional random effect (a so-called frailty term). The results were very similar to those of the models without a frailty term (data not shown).

For the purpose of visualization, we also derived a dichotomized version of the predictor. We performed 2-means clustering (using a k-means clustering algorithm) on the values of the continuous predictor for the 163 patients in the training set. This method of dichotomization does not make use of information about survival times. The mean of the 2 cluster centers (ie, a score value of 0.1703) was then chosen as the cutoff threshold, which was subsequently used to classify patients in the test and validation cohorts into predicted “good” and “poor” outcome groups (Figure 1). The Kaplan-Meier method was used to generate survival curves for patient subgroups identified by dichotomization of the gene-expression score. All statistical analyses were performed using the R 2.7.0 software package and routines from the bioinformatics software repository Bioconductor.

Results

Characteristics of the patient cohorts

The clinical characteristics of the 3 patient cohorts at the time of the initial diagnostic evaluation are shown in Table 1. The patients in the training cohort were younger and presented with higher leukocyte counts and lower hemoglobin levels than those in the test cohort. The incidence of FLT3 ITD was higher in the training cohort than in the test cohort (P = .005), whereas the proportion of patients with NPM1 mutations was similar in both groups. The prognostically favorable combination of an NPM1 mutation in the absence of FLT3 ITD (FLT3 ITD−/NPM1+) was present in 23% of the patients in the training cohort and in 34% of subjects in the test cohort (P = .09). Because of the inclusion criteria of the CALGB 9621 study (with an upper age limit of 59 years), the 64 patients in the validation cohort were significantly younger than those in the training and test cohorts. A total of 71% of the patients had an NPM1 mutation, a significantly higher percentage than in the training cohort (P = .01). A total of 35% of the patients in the validation cohort were FLT3 ITD−/NPM1+, similar to the other 2 cohorts.

Identification of prognostic genes

Using univariate Cox regression, we identified 86 probe sets on the Affymetrix microarrays that were significantly associated with OS in the training cohort, with an absolute Cox score of more than 2.9 (Figure 1; Table S1). These 86 probe sets corresponded to 60 annotated genes and 6 expressed sequence tag clones, and 13 genes were represented by more than one probe set each. We then used principal component analysis to assign a weight to each
of these probe sets, and a linear combination of the weighted expression values of the 86 probe sets was tested for its ability to serve as a “risk score.”

**Association of the gene-expression score with baseline patient characteristics and clinical outcomes**

Our test cohort consisted of 79 patients treated on the same therapeutic protocol as the patients in the training cohort. A continuous risk score was calculated for each patient using the probe sets and weights identified in the training cohort. The presence of a FLT3 ITD was associated with higher values of the gene-expression–based continuous risk score ($P < .001$), and the score also showed a positive association with higher leukocyte counts ($P = .001$; Table S2). In an analysis of treatment outcomes, a higher continuous risk score was a significant negative predictor of both OS (HR = 1.85; 95% CI, 1.25-2.74; $P = .002$) and EFS (HR = 1.73; 95% CI, 1.24-2.42; $P = .001$). Information on the response to induction treatment was available for 77 of the 79 patients, and 50 of them (65%) had a CR. In a logistic regression model, a higher gene-expression score was associated with a lower chance of reaching CR ($P = .045$). Among patients in CR, a higher gene-expression score predicted shorter RFS (HR = 1.76; 95% CI, 1.07-2.89; $P = .025$). For visualization of these results, survival curves generated by dichotomization of the continuous risk score are shown in Figure 2.

To further corroborate our results, we assessed our prognostic signature in an entirely independent validation cohort of patients from the CALGB study group. For these 64 patients, we used the same probe sets and weights as in our original training and test cohorts to calculate the microarray-based risk score. Similar to the test cohort, we found a positive association of the continuous risk score with the presence of FLT3 ITD ($P < .001$). No significant association of the gene-expression score and other clinical variables was observed. In the validation cohort, the continuous risk score showed a significant negative correlation with both OS (HR = 4.11; 95% CI, 2.10-8.03; $P < .001$) and EFS (HR = 2.90; 95% CI, 1.63-5.18; $P < .001$). Fifty-four patients (84%) had a CR after induction chemotherapy, but the continuous risk score did not

### Table 1. Pretreatment patient characteristics in the training, test, and validation cohorts

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Training cohort (n = 163)</th>
<th>Test cohort (n = 79)</th>
<th>Validation cohort (n = 64)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female sex, no. (%)</td>
<td>88 (54)</td>
<td>46 (58)</td>
<td>29 (45)</td>
</tr>
<tr>
<td>Median age, y (range)</td>
<td>58 (17-83)</td>
<td>62 (18-85)</td>
<td>45 (21-59)</td>
</tr>
<tr>
<td>AML type, no.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>De novo AML</td>
<td>156</td>
<td>71</td>
<td>64</td>
</tr>
<tr>
<td>s-AML</td>
<td>6</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>t-AML</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>FLT3 ITD status, no. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLT3 ITD−</td>
<td>86 (53)</td>
<td>57 (72)</td>
<td>38 (59)</td>
</tr>
<tr>
<td>FLT3 ITD+</td>
<td>77 (47)</td>
<td>22 (28)</td>
<td>26 (41)</td>
</tr>
<tr>
<td>NPM1 status, * no. (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPM1+</td>
<td>77 (47)</td>
<td>37 (47)</td>
<td>18 (29)</td>
</tr>
<tr>
<td>NPM1−</td>
<td>86 (53)</td>
<td>42 (53)</td>
<td>45 (71)</td>
</tr>
<tr>
<td>FLT3 ITD−/NPM1+, * no. (%)</td>
<td>38 (23)</td>
<td>27 (34)</td>
<td>22 (35)</td>
</tr>
<tr>
<td>FAB classification, † no.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>M1</td>
<td>45</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>M2</td>
<td>45</td>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>M4</td>
<td>42</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>M5</td>
<td>19</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>M6</td>
<td>6</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>RAEB</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Median leukocyte count, × 109/L (range)</td>
<td>36.9 (0.85-486)</td>
<td>15.9 (1-440.3)</td>
<td>36.9 (2.1-295)</td>
</tr>
<tr>
<td>Median hemoglobin level, g/L (range)</td>
<td>91 (40-142)</td>
<td>93.5 (60-147)</td>
<td>92 (60-129)</td>
</tr>
<tr>
<td>Median platelet count, × 109/L (range)</td>
<td>56 (6-471)</td>
<td>64 (9-239)</td>
<td>52 (12-378)</td>
</tr>
<tr>
<td>Median BM blasts, % (range)</td>
<td>85 (17-100)</td>
<td>80 (11-97)</td>
<td>64 (30-90)</td>
</tr>
<tr>
<td>Median follow-up for surviving patients, mo (range)</td>
<td>30 (1.6-79)</td>
<td>39 (4.6-50)</td>
<td>56 (47-82)</td>
</tr>
<tr>
<td>Overall survival</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median, mo</td>
<td>9.7</td>
<td>17.7</td>
<td>26.2</td>
</tr>
<tr>
<td>Estimated OS at 2 y, %</td>
<td>37</td>
<td>44</td>
<td>53</td>
</tr>
<tr>
<td>Event-free survival, † months</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median, mo</td>
<td>6.3</td>
<td>8.0</td>
<td>9.9</td>
</tr>
<tr>
<td>Estimated EFS at 2 y, %</td>
<td>24</td>
<td>26</td>
<td>38</td>
</tr>
<tr>
<td>Relapse-free survival, ‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median, mo</td>
<td>12.4</td>
<td>11.8</td>
<td>19.9</td>
</tr>
<tr>
<td>Estimated RFS at 2 y, %</td>
<td>39</td>
<td>38</td>
<td>44</td>
</tr>
<tr>
<td>CR rate, ‡ %</td>
<td>62</td>
<td>65</td>
<td>84</td>
</tr>
</tbody>
</table>

*P* values are for the comparison between the test and training or the validation and training cohort.

s-AML indicates secondary AML after antecedent hematologic disorder; t-AML, therapy-related AML after previous chemotherapy or radiotherapy; CR, complete remission; BM, bone marrow; PB, peripheral blood.

*Information on NPM1 status was missing for one patient in the validation cohort, who was FLT3 ITD negative.

†FAB classification was unknown for 3 patients in the validation cohort.

‡Information on remission status was missing for 2 patients in the test cohort.
predict the chance of reaching CR ($P = .21$). However, among patients who had a CR after induction therapy, a higher gene-expression score predicted a shorter RFS (HR = 3.14; 95% CI, 1.62-6.07; $P < .001$). Survival curves generated by dichotomization of the continuous risk score are displayed in Figure 3.

**Performance of the gene-expression score in the context of FLT3 ITD, NPM1 mutations, and age**

Because we observed an association of the continuous gene-expression score with FLT3 ITD status, Cox proportional hazards models were used to determine whether our gene-expression signature provided additional prognostic information after adjustment for FLT3 and other known risk factors. In multivariate analyses, after adjusting for age, FLT3 ITD/wt allelic ratio and NPM1 mutation status, the gene-expression score remained a significant predictor for OS in the test cohort ($P = .037$; Table 2). Along with the continuous risk score, higher age was a negative prognostic factor for OS, and the presence of an NPM1 mutation was of borderline significance. A higher gene-expression score also was significantly associated with shorter EFS ($P = .007$) and wild-type NPM1 predicted a worse OS, whereas age was not significant. In a multivariate analysis of EFS, the gene-expression score was of borderline significance in the smaller validation cohort ($P = 0.09$; Table 3).

Patients with an isolated NPM1 mutation in the absence of FLT3 ITD (FLT3 ITD−/NPM1+) constitute a subgroup of CN-AML with favorable treatment outcomes. On the other hand, the presence of FLT3 ITD or the FLT3 ITD+/NPM1− genotype are commonly regarded as high-risk molecular features in CN-AML.5-9 We therefore tested whether microarray profiling provided additional prognostic information for the latter subgroup of patients. A higher value of the gene-expression score was associated with inferior OS of such “molecular high risk” patients in both the test cohort (HR = 1.71; 95% CI, 1.14-2.58; $P = .009$) and the validation cohort (HR = 2.07; 95% CI, 1.21-3.53; $P = .008$; Figure 4).

**Discussion**

We present a prognostic gene signature, based on the expression levels of 66 genes and expressed sequence tags represented by 86 oligonucleotide microarray probe sets, which is able to predict OS, EFS, and RFS in AML patients with a normal karyotype. Because generating a prognostic score based on thousands of expression values per patient involves the risk of “overfitting,” assessing the performance of a prognostic gene signature on independent datasets is of critical relevance.26 In this respect, it is important to note that we verified the prognostic value of our signature in 2 independent and diverse patient cohorts from different study groups.

Our initial test cohort consisted of patients from a German multicenter treatment study (AMLCG-1999). The clinical characteristics of these patients were generally similar to the training cohort, which included patients from the same trial. To corroborate our findings, we also tested our predictor in a second group of 64 patients from a US multicenter trial (CALGB 9621). There were substantial differences between the CALGB validation cohort and the German patients in the test and training datasets: The German
AMLCG trial compared 2 double-induction chemotherapy regimens using either one or 2 cycles of high-dose cytarabine, followed by consolidation chemotherapy and either autologous stem cell transplantation or prolonged monthly maintenance chemotherapy. In contrast, the CALGB trial tested induction therapy with variable doses of daunorubicin and etoposide along with conventional-dose cytarabine, with or without multidrug resistance modulation, followed by autologous stem cell transplantation or an alternative intensification regimen. Furthermore, the CALGB trial did not include patients with secondary or treatment-related AML. The median age in the validation cohort was much lower than in the training and test cohorts because the CALGB study only enrolled patients younger than 60 years; thus, the baseline risk of this population was lower. Moreover, the frequency of important molecular markers, namely, FLT3 ITD and NPM1 mutations, varied among the cohorts. Despite these differences, the gene-expression score was a significant predictor for OS, EFS, and RFS in both populations. We used multivariate models to adjust for the differences in age and the frequencies of FLT3 ITD and NPM1 mutations, and the results of these analyses are remarkably consistent between the test and the validation cohort. In summary, our results indicate that the performance of our gene-expression predictor is robust in patient populations with different baseline characteristics.

Another important difference between the study cohorts is the predominant use of BM specimens for gene-expression profiling in the training and test cohorts, whereas PB samples were analyzed in

Table 2. Results of multivariate Cox regression analyses of OS and EFS in the test cohort

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall survival (n = 78)</th>
<th>Event-free survival (n = 77)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td>Age‡</td>
<td>1.43 (1.10-1.87)</td>
<td>.009</td>
</tr>
<tr>
<td>NPM1 mutation</td>
<td>0.54 (0.29-1.02)</td>
<td>.06</td>
</tr>
<tr>
<td>FLT3 ITD§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low FLT3 ITD/wt ratio</td>
<td>1.38 (0.64-2.97)</td>
<td>.41</td>
</tr>
<tr>
<td>High FLT3 ITD/wt ratio</td>
<td>1.26 (0.42-3.82)</td>
<td>.61</td>
</tr>
<tr>
<td>Microarray-based continuous risk score‡</td>
<td>1.64 (1.07-2.54)</td>
<td>.037</td>
</tr>
</tbody>
</table>

*One patient was not included because of missing data on FLT3 ITD/wt allelic ratio.
†Data on EFS were available for 77 of the 79 patients in the test dataset.
‡The hazard ratio is given for an age difference of 10 years.
§Hazard ratios are for the comparison with patients without detectable FLT3 ITD (ie, a FLT3 ITD/wt allelic ratio of 0).
<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall survival (n = 63)</th>
<th>Event-free survival (n = 63)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td>Age†</td>
<td>0.87 (0.66-1.13)</td>
<td>.30</td>
</tr>
<tr>
<td>NPM1 mutation</td>
<td>0.43 (0.18-1.01)</td>
<td>.052</td>
</tr>
<tr>
<td>FLT3 ITD</td>
<td>1.75 (0.63-4.81)</td>
<td>.28</td>
</tr>
<tr>
<td>Microarray-based continuous risk score‡</td>
<td>3.40 (1.40-8.29)</td>
<td>.007</td>
</tr>
</tbody>
</table>

*One patient in the validation dataset was not included because of missing data on NPM1 mutational status.
†The hazard ratio is given for an age difference of 10 years.
‡The hazard ratio is given for a change of the score equal to its interquartile range.
the validation cohort. The performance of our predictor did not appear to be compromised by the different starting materials. Similarly, in the study by Radmacher et al.,14 PB specimens were used to reproduce the gene signature by Bullinger et al13 that had been developed from a mixed PB/BM dataset. Other studies of gene-expression in acute leukemia have included both BM and PB samples, and in unsupervised analyses, no clustering of samples according to starting material was reported.10,11,13 There are very few studies that directly compared expression patterns in paired PB and BM samples in AML, but the available reports indicate a high degree of correlation between both sources.13,27

Despite the important differences between the patient populations and the analytical methods, our gene-expression score was a significant predictor of survival in both the test and validation cohorts. Assessing the generalizability of a prognostic marker requires testing both its reproducibility in an independent patient sample as well as its transportability to patient cohorts with different geographic and methodologic background.28 In this respect, the fact that we tested our predictor in 2 patient cohorts from different study groups from Europe and the United States strongly supports the validity of our gene signature.

The prognostic value of our signature was conserved after adjustment for age and NPM1 and FLT3 ITD mutations, and it also allowed prognostic stratification of the high-risk subgroup of patients who were either FLT3 ITD+ or FLT3 ITD/NPM1-. Patients with these genotypes have similar survival according to many,5-9 but not all,29 studies; therefore, additional prognostic factors in addition to NPM1 and FLT3 would be useful. Moreover, the analyses of RFS, although limited by the smaller patient numbers, suggest that the gene-expression signature may identify patients at increased risk for relapse after CR. This would be helpful for guiding risk-adapted postremission therapy. Taken together, these findings underscore the value of our signature, which apparently works well in diverse populations of patients with CN-AML.

The 64 patients in the validation cohort have been used previously to confirm the prognostic value of a gene signature published by Bullinger et al.13 That signature was originally developed in a cohort of patients with both normal and abnormal karyotypes using cDNA microarrays and comprises 133 genes. Subsequently, Radmacher et al constructed a prognostic classifier using 81 of the genes from the original signature (represented by 157 Affymetrix probe sets) and tested it in a group of 64 CN-AML patients.14 Interestingly, the Radmacher classifier also showed a strong correlation with patients’ FLT3 ITD status. It predicted OS and disease-free survival in univariate analyses, but it did not provide additional prognostic information in multivariate models that adjusted for FLT3 status. Of note, only 7 probe sets in that classifier, corresponding to 5 genes, overlap with our prognostic gene signature.

Similar to the study of Radmacher et al.,14 there was a strong correlation between the risk score derived from our prognostic gene signature and FLT3 ITD status. This finding is not surprising, considering that the presence of a FLT3 length mutation is a strong negative risk factor, and that the genes in our signature were selected according to their correlation with OS in the training cohort, without adjustment for FLT3 status. However, our results suggest that our gene signature is not merely a surrogate marker for FLT3 ITD because in multivariate analyses the gene-expression score turned out to be the stronger prognostic factor. We therefore hypothesize that the genes in our signature do not simply mirror the presence of a FLT3 ITD but are also influenced by other (and maybe unknown) genetic alterations associated with poor survival in CN-AML. In this respect, our results are similar to a recent report on an expression signature developed to predict FLT3 ITD in CN-AML.30 Although this signature misclassified approximately 20% of cases in an independent test set, the gene signature in fact was a stronger predictor of survival than FLT3 status itself. Six of the 20 genes in the FLT3 ITD prediction signature overlap with our prognostic signature. Neben et al examined differential gene expression in CN-AML patients with and without FLT3 ITD, and published a list containing the 11 top-ranking genes.31 Of note, there is no overlap between the genes represented in our survival predictor and the FLT3 target genes described in that study.

It will be interesting to explore how the genes in our risk signature are linked to cellular pathways that are involved in the pathogenesis of the disease and to investigate how these pathways influence a patient’s risk of death. We therefore compared our results with other reports on deregulated gene expression in AML. A recent study examined the gene-expression profiles of cytarine-sensitive and cytarine-resistant murine cell lines.32 The homolog of a gene that was strongly up-regulated in chemo-resistant cell lines, Whp5, is also part of our predictor. Another gene from our prognostic signature, TCF4, is part of a gene set that defines chemo-resistance in childhood acute lymphoblastic leukemia.33 Finally, Heuser et al35 analyzed samples from adult AML patients with resistance to induction chemotherapy. Three genes contained in our signature (FHL1, CD109, and SPARC) were associated with a poor response to chemotherapy in that study.35 Such overlap between studies performed in various patient populations and in animal models might point toward genes that are functionally relevant.
Besides **NPM1** mutations and **FLT3 ITD**, several other molecular prognostic markers have been described in CN-AML, including **CEBPα** mutations, **FLT3 TKD** mutations, **MLL** PTD, and overexpression of **ERG**, **BAALC**, and **EVI1**.\(^3\) Currently, little information is available on the interactions and relative importance of these risk factors. In our present study, we were unable to assess the performance of our gene-expression score in the context of these novel markers due to missing data and the relatively small sample sizes in our study, which made it difficult to capture the effects of low-frequency markers, such as **CEBPα** mutations. Therefore, our microarray-based signature needs to be tested together with other new risk markers in a sufficiently large patient cohort before definitive conclusions about their relative importance can be drawn.

Gene-expression profiling currently is not part of the standard diagnostic workup of AML. However, previous studies have proven that microarrays can serve as a comprehensive tool for the diagnosis and classification of leukemia.\(^4\) Based on these results, diagnostic microarray platforms for routine clinical use are being developed. In this context, a gene signature that offers prognostic information for a large subgroup of AML cases (ie, CN-AML) would be very useful. Another potential advantage of gene-expression profiling is its ability to capture the effects of multiple genetic alterations simultaneously; thus, a gene-expression signature may summarize the prognostic implications of multiple "conventional" risk markers into one single score. Further prospective studies are necessary to determine whether the signature identified herein is clinically useful to stratify patients to risk-adapted remission induction or postremission treatments beyond the current capacities of conventional single molecular marker screening. Such studies could investigate, for example, whether or not this signature identifies patients with high **FLT3 ITD**/wt allelic ratio who do not relapse, or patients with a favorable **NPM1/FLT3** profile who do ultimately relapse. Moreover, although the current study focused on deriving a signature from CN-AML and assessing its prognostic capability within this cytogenetic subset, we do not know whether this signature might have clinical utility for patients with AML and aberrant karyotypes. Further investigations should address this possibility.

In conclusion, we present a prognostic gene signature that was developed to predict outcome in patients with CN-AML. A predictor based on this signature was associated with OS, EFS, and RFS in 2 different and independent patient cohorts. Although our gene-expression score was strongly influenced by the presence of a **FLT3 ITD**, it provided prognostic information beyond the effects of **FLT3 ITD** and **NPM1** mutation status in multivariate analyses. We think that, given the heterogeneity of CN-AML, a predictor that allows for refined risk assessment would be of value to identify patients who may or may not benefit from intensive therapeutic strategies, such as allogeneic stem cell transplantation.

**Acknowledgments**

The authors thank Eric Bair for helpful comments on the supervised principal components algorithm; Gudrun Mellert, Evelin Zellmeier, and Marlene Seibl for expert technical assistance; and all the patients and local investigators who participated in the AMLCG and CALGB trials and thereby made this work possible.

This work was supported by grants from the Bundesministerium für Bildung und Forschung (German National Genome Research Network; grant 01 GR 0459, M.H., U.M.; and grant 01 GS 0448, C.B., S.K.B.) and the Deutsche Forschungsgemeinschaft (DFG-German Research Council) SFB 684 projects A6 and A8 (S.K.B. and C.B.), by the National Institutes of Health (grants CA101140, CA114725, CA077658, and CA016058), and the Coleman Leukemia Research Foundation (C.D.B.).

**Authorship**

Contribution: K.H.M. designed research, analyzed and interpreted data, and drafted the manuscript; M.H. performed statistical analyses and cowrote the manuscript; K.S., J.B., G.M., S.P.W., P.P., R.A.L, W.E.B., T.B., and B.W. provided data and were involved in patient care; M.-C.S., A.H., M.R., and K.M. performed statistical analyses; U.M. supervised statistical analysis and provided funding; C.D.B., W.H., S.K.B., and C.B. supervised research, provided funding; and all coauthors reviewed and discussed the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Christian Buske, Department of Internal Medicine III, University Hospital Großhadern, Ludwig-Maximilians Universität, Campus Großhadern, Marchioninstr 15, 81377 München, Germany; e-mail: christian.buske@med.uni-muenchen.de; or Stefan K. Bohlander, Department of Internal Medicine III, University Hospital Großhadern, Ludwig-Maximilians Universität, Campus Großhadern, Marchioninstr 15, 81377 München, Germany; e-mail: stefan.bohlander@med.uni-muenchen.de.

**References**


An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia

Klaus H. Metzeler, Manuela Hummel, Clara D. Bloomfield, Karsten Spiekermann, Jan Braess, Maria-Cristina Sauerland, Achim Heinecke, Michael Radmacher, Guido Marcucci, Susan P. Whitman, Kati Maharry, Peter Paschka, Richard A. Larson, Wolfgang E. Berdel, Thomas Büchner, Bernhard Wörmann, Ulrich Mansmann, Wolfgang Hiddemann, Stefan K. Bohlander and Christian Buske