Gene expression signature predicts relapse in adult patients with cytogenetically normal acute myeloid leukemia

Christopher J. Walker  
*The Ohio State University*

Geoffrey L. Uy  
*Washington University School of Medicine in St. Louis*

et al.

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Please let us know how this document benefits you.

**Recommended Citation**


This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
Gene expression signature predicts relapse in adult patients with cytogenetically normal acute myeloid leukemia

Christopher J. Walker,1 Krzysztof Mrózek,1 Hatice Gulcin Ozer,1 Deedra Nicolet,1,2 Jessica Kohlschmidt,1,2 Dimitrios Papaioannou,1 Luke K. Genutis,1 Marius Bill,1 Bayard L. Powell,3 Geoffrey L. Uy,4 Jonathan E. Kolitz,5 Andrew J. Carroll,6 Richard M. Stone,7 Ramiro Garzon,1 John C. Byrd,1 Ann-Kathrin Eisfeld,1,* Albert de la Chapelle,1,* and Clara D. Bloomfield1,*

1The Ohio State University Comprehensive Cancer Center, Columbus, OH; 2Alliance Statistics and Data Center, The Ohio State University Comprehensive Cancer Center, Columbus, OH; 3Wake Forest Baptist Comprehensive Cancer Center, Winston-Salem, NC; 4Washington University School of Medicine in St. Louis, Siteman Cancer Center, St. Louis, MO; 5Monter Cancer Center, Zucker School of Medicine at Hofstra/Northwell, Lake Success, NY; 6Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; and 7Dana-Farber/Partners CancerCare, Boston, MA

Key Points
- The weighted expressions of 7 coding and 3 noncoding genes is strongly associated with relapse in CN-AML patients.
- The 10-gene signature is independent from mutations known to associate with outcome in AML patients.

Although ~80% of adult patients with cytogenetically normal acute myeloid leukemia (CN-AML) achieve a complete remission (CR), more than half of them relapse. Better identification of patients who are likely to relapse can help to inform clinical decisions. We performed RNA sequencing on pretreatment samples from 268 adults with de novo CN-AML who were younger than 60 years of age and achieved a CR after induction treatment with standard “7+3” chemotherapy. After filtering for genes whose expressions were associated with gene mutations known to impact outcome (ie, CEBPA, NPM1, and FLT3-internal tandem duplication [FLT3-ITD]), we identified a 10-gene signature that was strongly predictive of patient relapse (area under the receiver operating characteristics curve [AUC], 0.81). The signature consisted of 7 coding genes (GAS6, PSD3, PLCB4, DEXI, JMY, NRP1, C10orf55) and 3 long noncoding RNAs. In multivariable analysis, the 10-gene signature was strongly associated with relapse (P < .001), after adjustment for the FLT3-ITD, CEBPA, and NPM1 mutational status. Validation of the expression signature in an independent patient set from The Cancer Genome Atlas showed the signature’s strong predictive value, with AUC = 0.78. Implementation of the 10-gene signature into clinical prognostic stratification could be useful for identifying patients who are likely to relapse.

Introduction

A major obstacle to improved survival of patients with acute myeloid leukemia (AML) is disease relapse after achievement of complete remission (CR). Prognostic stratification using molecular and cytogenetic markers is useful for the early identification of patients who are likely to be refractory to standard induction chemotherapy regimens and/or have a higher risk for relapse; thus, it is being used for making informed clinical decisions. The 2017 European LeukemiaNet (ELN) genetic risk classification is widely accepted as the standard method for prognostic stratification of AML patients.1 However, the 2017 ELN classification includes only selected gene mutations and cytogenetic abnormalities and does not take into account gene expression data.1

Genomic alterations underlying disease in AML patients are heterogeneous, including diverse transcriptional profiles.2,3 Previous studies have demonstrated that the use of differential expression of
single genes and, more recently, gene expression signatures, are effective tools for risk stratification of AML patients.2-10 Herein, we sought to explore the association between gene expression and disease relapse in first CR in adult patients younger than 60 years of age who were diagnosed with cytogenetically normal acute myeloid leukemia (CN-AML).

Methods
Total transcriptome RNA sequencing (RNAseq) was performed using pretreatment blood or bone marrow samples from 268 adult CN-AML patients younger than 60 years who were similarly treated with intensive chemotherapy on Cancer and Leukemia Group B (CALGB) (now part of Alliance for Clinical Trials in Oncology [Alliance]) therapeutic trials, including CALGB 10503 (ClinicalTrials.gov Identifier: NCT00416598), CALGB 10603 (NCT00651261), and CALGB 19808 (NCT00006363) (see supplemental Data) and achieved a CR. The patient cohort did not include patients with AML secondary to antecedent hematologic disorder or patients with therapy-related AML. Targeted sequencing of 80 cancer- and leukemia-associated genes, as well as detection of FLT3-internal tandem duplication (FLT3-ITD) and CEBPA mutations, were performed previously on all patients.11-13 Pretreatment cytogenetic analyses were performed in the CALGB/Alliance-approved institutional laboratories. The presence of a normal karyotype was determined by examination of ≥20 metaphase cells obtained from short-term (24- and/or 48-hour) unstimulated cultures of bone marrow samples and confirmed by central karyotype review in each case.14

RNAseq reads were aligned to hg38 using HISAT2,15 and gene counts were obtained using featureCounts.16 Normalization was performed with DeSeq2,17 which divides counts by sample-specific size factors determined by the median ratio of gene counts relative to geometric mean per gene. Hierarchical clustering was performed using the hclust function in the R (v4.0.1) stats package with Ward’s method, performed on a distance matrix computed using the ClassDiscovery R package with the absolute Pearson metric.18 Random forest models were generated with the randomForest R package, performing 100 iterations with n = 501 and default mtry.19 Expression between groups was assessed using a negative binomial model with DeSeq2 or random forests, as indicated after removing genes with low expression (normalized counts $\leq 10$) and low variability (standard deviation $< 10$). Predictive ability of the random forest model was optimized by first determining the importance of all 539 genes and then iterating through different mtrys to select 539 genes that were input into a random forest model to predict CR (supplemental Table 6). Optimization iterations determined that the maximum predictive power was achieved using a model fit on the expression of the following 10 genes: NRP1, PLCB4, JMY, PSD3, DEXI, GAS6, C10orf55, AC139769.2, AC015712.2, and AL096865.1; these genes were assigned importances from the model based on their ability to predict relapse (Table 1). The AUC of this model was 0.81 (Figure 2A), and the 10-gene signature correctly classified 141 of 165 patients who relapsed and 65 of 104 patients who maintained a CR (Figure 2B). Classifying patients into genetic-risk groups according to the 2017 ELN criteria revealed that the 10-gene signature correctly predicted relapse in 94% of patients in the adverse-risk group, 86% of patients in the intermediate-risk group, and 71% of patients in the favorable-risk group (supplemental Table 7).

The predictive relapse score for each patient generated by the 10-gene signature was input into a multivariable logistic regression model for relapse and a Cox proportional multiple regression model for DFS, which contained all available clinical and demographic variables, gene mutations present in ≥8 patients, and expression of ERG, BAALC, MN1, mir-155, and miR-3151, which were previously shown to be associated with outcome in adults with CN-AML (supplemental Data).23-30 The logistic multivariable regression model showed that the 10-gene predictive score was significantly associated with the risk of patient relapse ($P < .001$; odds ratio, 1.79; 95% confidence interval [CI], 1.52-2.13). Multivariable logistic CEBPA mutations, mutation of NPM1, and FLT3-ITD also remained significant in the same model (Table 2). In the DFS Cox proportional multiple regression model, the 10-gene predictive score was associated with DFS ($P < .001$; hazard ratio, 1.32; 95% CI, 1.22-1.43) after adjusting for biallelic CEBPA mutations, FLT3-ITD, and MN1 expression (Table 2). Together, these data indicate that the 10-gene signature

Results
We performed RNAseq on 268 adult CN-AML patients younger than 60 years of age and then compared gene expressions between patients who relapsed (n = 164) and patients who remained in CR for ≥3 years (n = 104). The mutation status of 18 genes that were found to be mutated in ≥3% of patients and the patients’ pretreatment characteristics, including assignment to genetic-risk groups according to the 2017 ELN classification, are presented in supplemental Table 1. Differential expression analysis using a negative binomial model identified 255 genes that were significantly differentially expressed (adjusted $P$ value $< .001$ and absolute fold change $> 0.667$; supplemental Table 2). Hierarchical clustering was performed using these genes, which separated patients into distinct groups (Figure 1). Although these patient clusters had different rates of relapse, they were strongly associated with mutations known to be associated with AML prognosis, specifically mutations in NPM1, biallelic CEBPA mutations, and FLT3-ITD (Figure 1).

To find gene expressions associated with relapse that are independent from the aforementioned mutations, we filtered out genes that were significantly differentially expressed between patients with and without NPM1 mutations (2064 genes), biallelic CEBPA mutations (3923 genes), and FLT3-ITD (675 genes; adjusted $P$ value $< .01$ and absolute fold change $> 0.667$; supplemental Tables 3-5). From the remaining 14 741 genes, we used a cutoff of an absolute fold-change difference $> 0.3$ and a $P$ value $< .1$ to select 539 genes that were input into a random forest model to predict CR (supplemental Table 6). Optimization iterations determined that the maximum predictive power was achieved using a model fit on the expression of the following 10 genes: NRP1, PLCB4, JMY, PSD3, DEXI, GAS6, C10orf55, AC139769.2, AC015712.2, and AL096865.1; these genes were assigned importances from the model based on their ability to predict relapse (Table 1). The AUC of this model was 0.81 (Figure 2A), and the 10-gene signature correctly classified 141 of 165 patients who relapsed and 65 of 104 patients who maintained a CR (Figure 2B). Classifying patients into genetic-risk groups according to the 2017 ELN criteria revealed that the 10-gene signature correctly predicted relapse in 94% of patients in the adverse-risk group, 86% of patients in the intermediate-risk group, and 71% of patients in the favorable-risk group (supplemental Table 7).

The predictive relapse score for each patient generated by the 10-gene signature was input into a multivariable logistic regression model for relapse and a Cox proportional multiple regression model for DFS, which contained all available clinical and demographic variables, gene mutations present in ≥8 patients, and expression of ERG, BAALC, MN1, mir-155, and miR-3151, which were previously shown to be associated with outcome in adults with CN-AML (supplemental Data).23-30 The logistic multivariable regression model showed that the 10-gene predictive score was significantly associated with the risk of patient relapse ($P < .001$; odds ratio, 1.79; 95% confidence interval [CI], 1.52-2.13). Multivariable logistic CEBPA mutations, mutation of NPM1, and FLT3-ITD also remained significant in the same model (Table 2). In the DFS Cox proportional multiple regression model, the 10-gene predictive score was associated with DFS ($P < .001$; hazard ratio, 1.32; 95% CI, 1.22-1.43) after adjusting for biallelic CEBPA mutations, FLT3-ITD, and MN1 expression (Table 2). Together, these data indicate that the 10-gene signature
Figure 1. Clustering of patients with CN-AML based on expression of 255 genes associated with relapse. Heatmap shows expressions of genes differentially expressed between patients who relapsed and those who did not relapse for ≥3 years after achieving a CR. Each row of the heatmap represents expression of a gene, and each column represents a patient. Differential expression analysis to determine the 255 genes included was performed using a negative binomial model with the DeSeq2 R package. Shown above the heatmap is the relapse status for each patient, and the mutation statuses of genes mutated in ≥9 patients, as assessed by sequencing 81 genes. Six genes included in the 10-gene relapse signature that we derived in this study are indicated with arrows on the right side of the heatmap.
is a strong predictor of relapse in younger CN-AML patients treated with intensive induction chemotherapy, and it adds predictive value to mutations that are already known to predict relapse.

To independently validate the 10-gene signature in another patient set, we used expression data from The Cancer Genome Atlas (TCGA) for AML. TCGA data contained 32 CN-AML patients younger than 60 years of age who achieved a CR, 22 of whom relapsed in first CR. We calculated the 10-gene predictive relapse score for TCGA patients and found that the model correctly classified 20 of the 22 patients who relapsed and 7 of the 10 who did not, with AUC = 0.78 (Figure 2C-D).

Finally, we sought to examine the association between expression of the genes in the 10-gene relapse signature and germline polymorphisms to identify expression quantitative trait loci (eQTLs) for these genes in AML. Using genotyping data from these patients, we tested for expression associations with single nucleotide polymorphisms (SNPs) in the same regions. Indeed, we found evidence for eQTLs in the JMY gene and 5' of DEXI (Figure 3). In the JMY eQTL, the sentinel SNP, rs6414979, was common (global minor allele frequency, 0.37) and was strongly associated with JMY expression ($P = 9.05 \times 10^{-6}$). Likewise, the strongest associated SNP in the DEXI eQTL, rs3087876, was also common (global minor allele frequency, 0.45) and was associated with DEXI expression ($P = 4.10 \times 10^{-9}$) (supplemental Table 8).

### Discussion

Our study identified a 10-gene expression signature present at the time of diagnosis that can predict relapse during first CR, independent from known prognostic markers in CN-AML. Although identification of molecular markers that predict outcome for adult patients with CN-AML treated with intensive chemotherapy is a relatively well-researched area, our study is unique in that we focused on gene expressions independent from known prognostic mutations. It was not surprising that, in our initial differential expression analysis comparing gene expressions between patients who relapsed and patients who maintained CR, clustering was driven by biallelic mutations in CEBA and FLT3-ITDs, because these are known to be associated with outcome in CN-AML and have distinct expression profiles. Removing the genes differentially expressed between distinct CEBA, FLT3-ITD, and NPM1 clusters allowed us to discover an expression signature that was independent from these known prognostic markers.

Early genome-wide investigations of gene expression in AML include work by Bullinger et al6 and Valk et al, who conducted seminal studies using microarrays that revealed the transcriptional heterogeneity between cytogenetic subsets of patients. These studies also offered the first insights into the relevance of transcriptional signatures for predicting patient outcome, by describing associations between expression-defined patient clusters and survival.

However, although gene-expression profiling is capable of providing prognostic information that is independent from other genetic risk factors, reproducibility issues have largely prevented its use in clinical practice. Limiting factors include lack of standardization of laboratory procedures and implementation of quality controls among various institutions, normalization and quantification of RNAseq data, and differences in probe content of microarrays. Recently, strides have been made to overcome these issues by implementing standard procedures for the use of commercially available tests suitable for clinical use in individual patients, which rely on highly reproducible multiplexed quantitative polymerase chain reactions assays or, less frequently, NanoString nCounter technology. Continued optimization and rigorous scrutiny of these methods may lead to routine use of RNA expression in some circumstances in the near future, similar to the currently accepted use of protein expression, as determined by immunohistochemistry, as diagnostic and predictive markers.

More recent work with RNAseq has been conducted to specifically identify coding and noncoding RNA signatures predictive of outcome in AML patients, including patients with CN-AML. The 10 genes that make up our predictive expression signature have not been included in any of the more notable gene expression signatures that are predictive of AML prognosis, including a long noncoding RNA signature described by our group. We speculate that this might be due to our exclusion of genes associated with biallelic CEBA mutations, NPM1 mutations, and FLT3-ITDs.

Changes in the expression of 3 of the 10 genes constituting our gene expression signature (GAS6, PLCB4, and NRP1) have previously been shown to associate with outcomes of patients with AML in single-gene studies. The other coding genes in the 10-gene signature have been described to play roles in cancer as diagnostic and predictive markers.

#### Table 1. Importance of the expression of the 10 genes in the relapse prediction signature

<table>
<thead>
<tr>
<th>Gene</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEXI</td>
<td>14.81542</td>
</tr>
<tr>
<td>C10orf55</td>
<td>13.75139</td>
</tr>
<tr>
<td>PS3</td>
<td>13.68804</td>
</tr>
<tr>
<td>AC139769.2</td>
<td>13.56775</td>
</tr>
<tr>
<td>GAS6</td>
<td>13.0606</td>
</tr>
<tr>
<td>AC015112.2</td>
<td>12.65445</td>
</tr>
<tr>
<td>JMY</td>
<td>12.5374</td>
</tr>
<tr>
<td>PLCB4</td>
<td>11.54397</td>
</tr>
<tr>
<td>AL096865.1</td>
<td>11.15018</td>
</tr>
<tr>
<td>NRP1</td>
<td>10.59661</td>
</tr>
</tbody>
</table>

Importance based on the Gini impurity index was used for the calculation of splits during training.
Our findings were validated using publicly available data from the TCGA, which, despite a relatively small number of patients, showed that the 10-gene signature was strongly predictive of relapse in adult CN-AML patients from this study. Although corroboration of our findings in another large set of patients with CN-AML is still desirable, we believe that addition of this signature to standard clinical practice could be beneficial.  

Figure 2. Gene expression signature is predictive of relapse in patients with CN-AML. (A) Receiver operating characteristic (ROC) curve shows the sensitivity and specificity of 10-gene expression signature for predicting relapse in 268 adult CN-AML patients younger than 60 years. (B) Predicted relapse probability for the 268 patients determined using the 10-gene signature. Each bar represents a patient, colored according to actual relapse status. (C) Predicted relapse probability for a validation set of 32 adult patients with CN-AML younger than 60 years included in the TCGA database determined using the 10-gene signature. (D) ROC curve showing the sensitivity and specificity of the 10-gene expression signature for predicting relapse in the 32 TCGA patients with CN-AML. Maintain CR denotes CR maintained for ≥3 years.

Table 2. Multivariable analyses for outcome

<table>
<thead>
<tr>
<th>Variable</th>
<th>Categories</th>
<th>P</th>
<th>Odds/hazards ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Logistic regression model for relapse</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-gene signature</td>
<td>Continuous, 10% increase</td>
<td>&lt;.001</td>
<td>1.79 (1.52-2.13)</td>
</tr>
<tr>
<td>Biallelic CEBPA mutation status</td>
<td>Mutated vs wild-type</td>
<td>.007</td>
<td>0.21 (0.07-0.66)</td>
</tr>
<tr>
<td>FLT3-ITD</td>
<td>Present vs absent</td>
<td>.04</td>
<td>2.14 (1.02-4.50)</td>
</tr>
<tr>
<td>NPM1 mutation status</td>
<td>Mutated vs wild-type</td>
<td>.03</td>
<td>0.33 (0.12-0.90)</td>
</tr>
<tr>
<td><strong>Cox proportional hazards regression model for DFS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-gene signature</td>
<td>Continuous, 10% increase</td>
<td>&lt;.001</td>
<td>1.32 (1.22-1.43)</td>
</tr>
<tr>
<td>Biallelic CEBPA mutation status</td>
<td>Mutated vs wild-type</td>
<td>.01</td>
<td>0.53 (0.32-0.87)</td>
</tr>
<tr>
<td>FLT3-ITD</td>
<td>Present vs absent</td>
<td>.005</td>
<td>1.62 (1.16-2.26)</td>
</tr>
<tr>
<td>MN1 expression</td>
<td>High vs low (median)</td>
<td>.008</td>
<td>1.58 (1.13-2.22)</td>
</tr>
</tbody>
</table>
Figure 3. eQTLs regional association plots. Plots show SNPs associated with expression of DEXI (A) and JMY (B). The top track indicates negative log10 P values for associations between SNPs and expression of DEXI (A) or JMY (B). SNPs (represented by triangles) are colored according to linkage disequilibrium (LD), with the sentinel SNP (blue triangle) that showed the most significant association with expression. The middle track (green horizontal lines) shows the location and transcriptional direction of all coding genes in the displayed regions. The bottom track (blue lines) indicates genetic regions containing known regulatory elements annotated using the Ensembl database (microRNA target sites, promoters, enhancers, and ENCODE feature clusters that can be associated with transcription factor binding motifs). Plots were made using SNiPA, a SNPs annotator.
signature to the current molecular prognostication guidelines, especially if expression of the genes constituting the novel signature we report herein can be assessed using a clinically suitable method, will allow more accurate prediction of relapse in CN-AML patients who have achieved a CR.

Acknowledgments

The authors thank the patients who participated in clinical trials, Christopher Manring and the CALGB/Alliance Leukemia Tissue Bank at The Ohio State University Comprehensive Cancer Center for sample processing and storage services, and Lisa J. Sterling for data management.

This work was supported by National Cancer Institute, National Institutes of Health awards U10CA180821, U10CA180882, and U24CA196171 (Alliance for Clinical Trials in Oncology), U10CA180861, UG1CA233180, UG1CA233331, UG1CA233338, UG1CA233339, P30CA016058, and P50CA140158; the Leukemia Clinical Research Foundation; the Warren D. Brown Foundation; the Pelotonia Fellowship Program; and by an allocation of computing resources from The Ohio Supercomputer Center. It was also supported in part by funds from Novartis (CALGB 10603). Support to Alliance for Clinical Trials in Oncology and Alliance Foundation Trials programs is listed at https://acknowledgments.alliancefound.org.

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

This article is dedicated to celebrating the lives and accomplishments of Clara D. Bloomfield and Albert de la Chapelle.

Authorship

Conflict-of-interest disclosure: C.J.W. has acted as a consultant for Vigeo Therapeutics, is employed by Karyopharm Therapeutics, and has ownership interests in Karyopharm Therapeutics and Bristol Myers Squibb. The remaining authors declare no competing financial interests.

Albert de la Chapelle died on 10 December 2020.

Clara D. Bloomfield died on 1 March 2020.

ORCID profiles: D.N., 0000-0001-7796-0251; M.B., 0000-0002-1175-2406; G.L.U., 0000-0002-7809-0996.

Correspondence: Christopher J. Walker, The Ohio State University Comprehensive Cancer Center, 460 West 12th Ave, Columbus, OH 43210-1228; e-mail: christopher.walker@osumc.edu.

References


