

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

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

<sup>1</sup>The Ohio State University Comprehensive Cancer Center, Columbus OH; <sup>2</sup>Alliance Statistics and Data Center, The Ohio State University Comprehensive Cancer Center, Columbus, OH; <sup>3</sup>Wake Forest Baptist Comprehensive Cancer Center, Winston-Salem, NC; <sup>4</sup>Washington University School of Medicine in St. Louis, Siteman Cancer Center, St. Louis, MO; <sup>5</sup>Monter Cancer Center, Zucker School of Medicine at Hofstra/Northwell, Lake Success, NY; <sup>6</sup>Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; and <sup>7</sup>Dana-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.<sup>1</sup> However, the 2017 ELN classification includes only selected gene mutations and cytogenetic abnormalities and does not take into account gene expression data.<sup>1</sup>

Genomic alterations underlying disease in AML patients are heterogeneous, including diverse transcriptional profiles.<sup>2,3</sup> Previous studies have demonstrated that the use of differential expression of

Submitted 30 October 2020; accepted 30 December 2020; published online 8 March 2021. DOI 10.1182/bloodadvances.2020003727.

\*A.-K.E., A.d.l.C., and C.D.B. contributed equally to this work.

The data reported in this article have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus database (accession GSE165430).

Data sharing requests should be sent to Christopher J. Walker (christopher.walker@osumc.edu).

The full-text version of this article contains a data supplement.

© 2021 by The American Society of Hematology

single genes and, more recently, gene expression signatures, are effective tools for risk stratification of AML patients.<sup>2-10</sup> 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.<sup>11-13</sup> Pretreatment cytogenetic analyses were performed in the CALGB/Alliance-approved institutional laboratories. The presence of a normal karyotype was determined by examination of  $\geq 20$  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.<sup>14</sup>

RNAseq reads were aligned to hg38 using HISAT2,<sup>15</sup> and gene counts were obtained using featureCounts.<sup>16</sup> Normalization was performed with DeSeq2,<sup>17</sup> 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.<sup>18</sup> Random forest models were generated with the randomForest R package, performing 100 iterations with  $n = 501$  and default *mtry*.<sup>19</sup> 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  $< 10$ ) 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 numbers of genes ( $n = 2-20, 25, 30, 35, 40, 45, 50, 75, 100, 200, 500$ ), starting with the most important, to determine the number that produces the highest area under the receiver operating characteristics curve (AUC). Multivariable logistic and proportional hazards regression models used a backward selection technique to build the final models for relapse and disease-free survival (DFS) that included relapse prediction score, clinical variables, mutation status, and indicated gene expressions associated with relapse at a level of  $P < .2$  from univariable analyses.

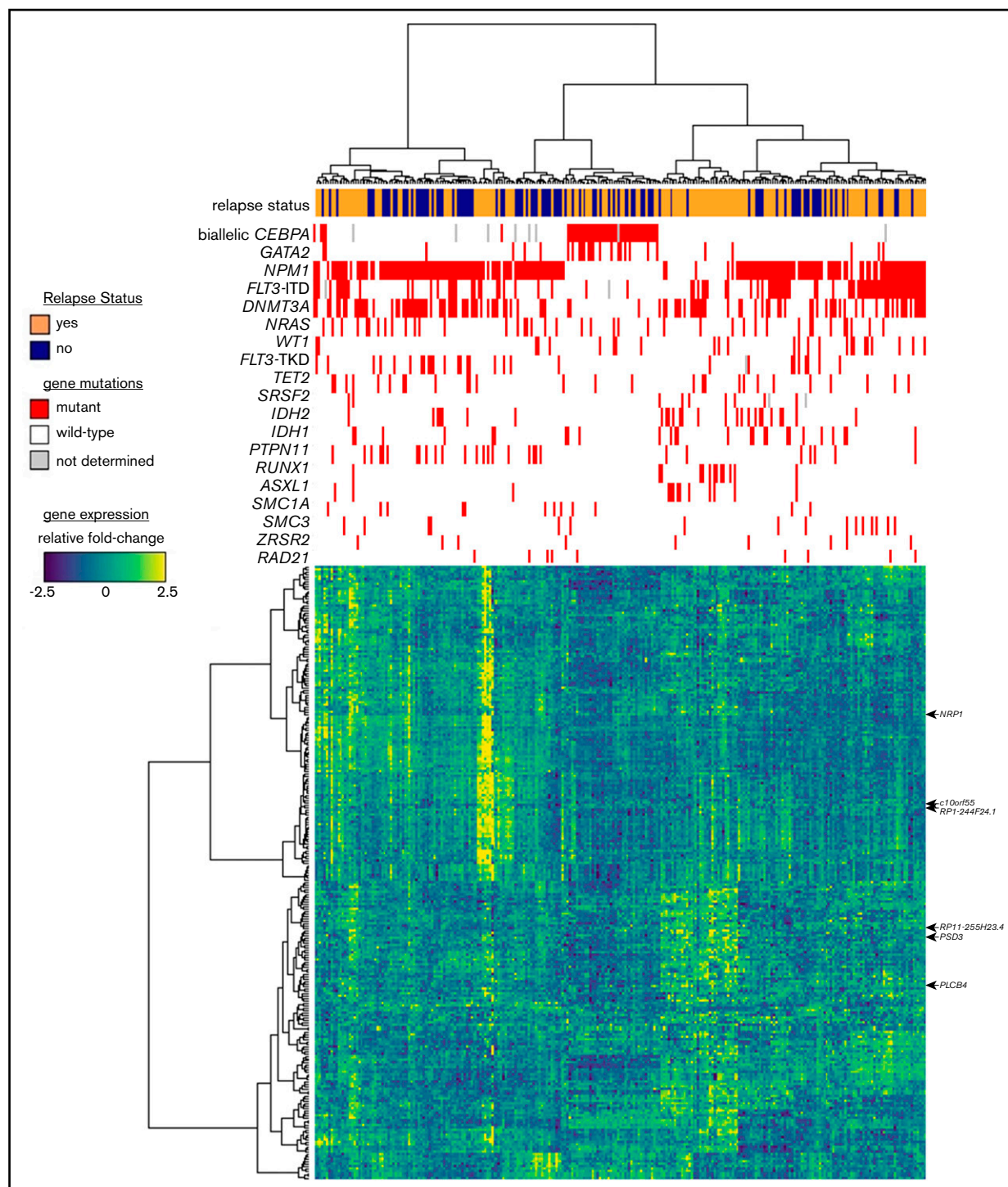
Genotyping of germline polymorphisms was performed previously on all patients, as described, using Infinium HumanOmni1-Quad BeadChip arrays (Illumina, San Diego, CA).<sup>20</sup> Imputation was performed using the haplotype reference consortium,<sup>21</sup> and testing for associations between germline polymorphisms and genes expressions was done with Matrix eQTL.<sup>22</sup>

## 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  $\geq 3$  years ( $n = 104$ ). The mutation status of 18 genes that were found to be mutated in  $\geq 3\%$  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  $\geq 8$  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).<sup>23-30</sup> 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). Biallelic *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  $\geq 3$  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  $\geq 9$  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.

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

Gene	Importance
<i>DEXI</i>	14.81542
<i>C10orf55</i>	13.75139
<i>PSD3</i>	13.68804
<i>AC139769.2</i>	13.56775
<i>GAS6</i>	13.0605
<i>AC015712.2</i>	12.65445
<i>JMY</i>	12.5374
<i>PLCB4</i>	11.54357
<i>AL096865.1</i>	11.15018
<i>NRP1</i>	10.59661

Importance based on the Gini impurity index was used for the calculation of splits during training.

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.<sup>31</sup> TCGA data contained 32 CN-AML patients younger than 60 years of age who achieved a CR, 22 of whom relapsed in first CR.<sup>31</sup> 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 *CEBPA* and *FLT3-ITDs*, because these are known to be associated with outcome in CN-AML and have distinct expression profiles.<sup>5,32-34</sup> Removing the genes

differentially expressed between distinct *CEBPA*, *FLT3-ITD*, and *NPM1* clusters allowed us to discover an expression signature that was independent from these known prognostic markers.

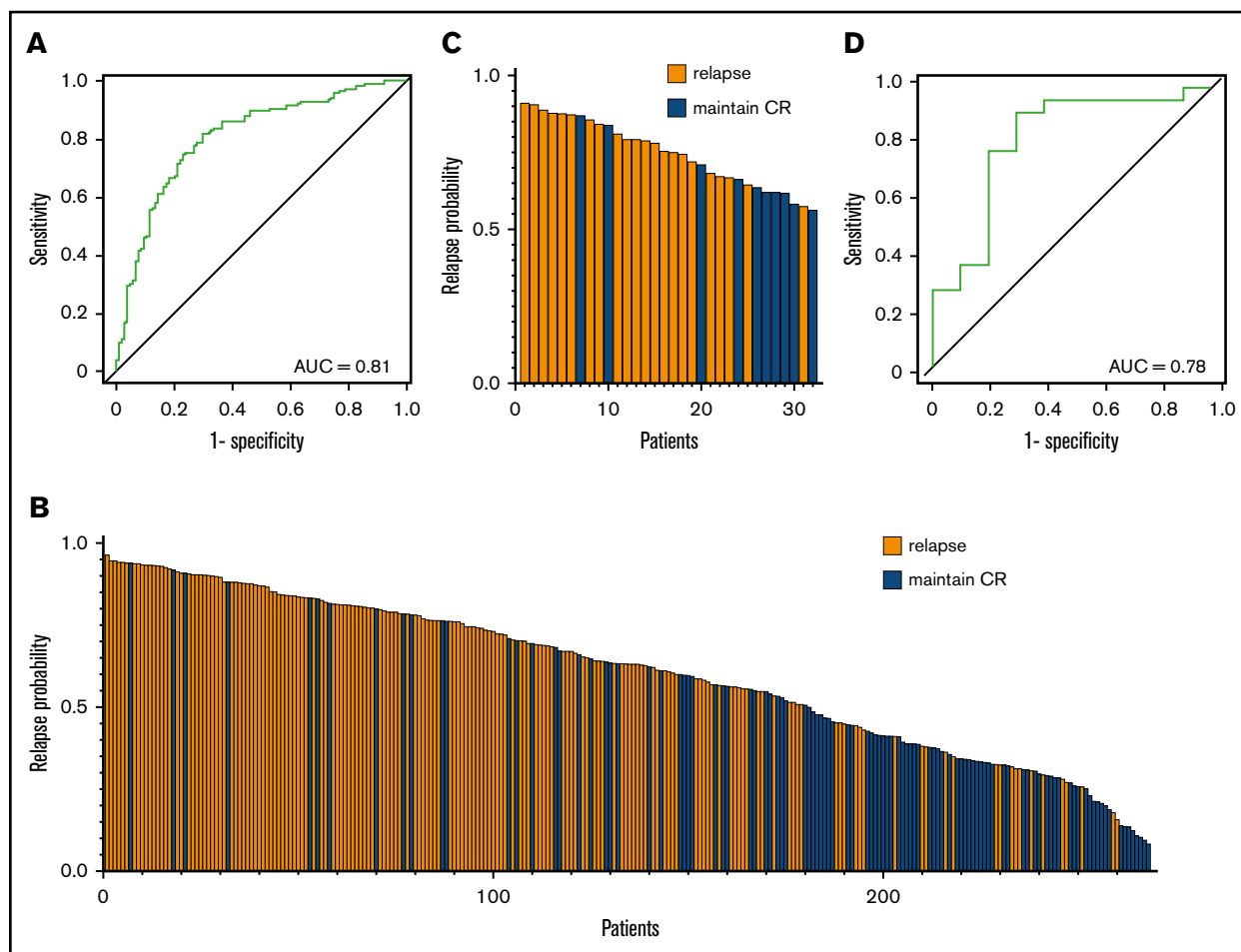
Early genome-wide investigations of gene expression in AML include work by Bullinger et al<sup>6</sup> and Valk et al,<sup>5</sup> who conducted seminal studies using microarrays that revealed the transcriptional heterogeneity between cytogenetic subsets of patients. These studies also offered the first 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,<sup>2-7</sup> 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.<sup>35,36</sup> 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,<sup>5,6,8,9,37,38</sup> including a long noncoding RNA signature described by our group.<sup>39</sup> We speculate that this might be due to our exclusion of genes associated with biallelic *CEBPA* mutations, *NPM1* mutations, and *FLT3-ITDs*.

Changes in the expression of 3 of the 10 genes constituting our gene expression signature (*GAS6*,<sup>40,41</sup> *PLCB4*,<sup>42</sup> and *NRP1*<sup>43-45</sup>) 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 well. Although *DEXI* has not been studied in leukemogenesis, the calcium binding protein-encoding gene has been identified as a fusion partner of *CIITA* in CN-AML, suggesting *DEXI* is a particularly interesting candidate for future studies.<sup>46</sup> *JMY* encodes a known cofactor of EP300, which serves as an activator of the tumor suppressor TP53.<sup>47</sup> *PSD3* expression is associated with breast cancer metastasis and glioma progression.<sup>48,49</sup> The 3 noncoding genes in the signature have not been well characterized, but our results suggest that they merit further investigation.

Interestingly, our incorporation of genome-wide genotyping data revealed eQTLs regulating the expression, in AML cells, of 2 of the genes in the 10-gene signature: *JMY* and *DEXI*. These results imply that germline polymorphisms are at least one of the many factors that likely contribute to the expression of these genes, which are associated with an increased likelihood of disease relapse.



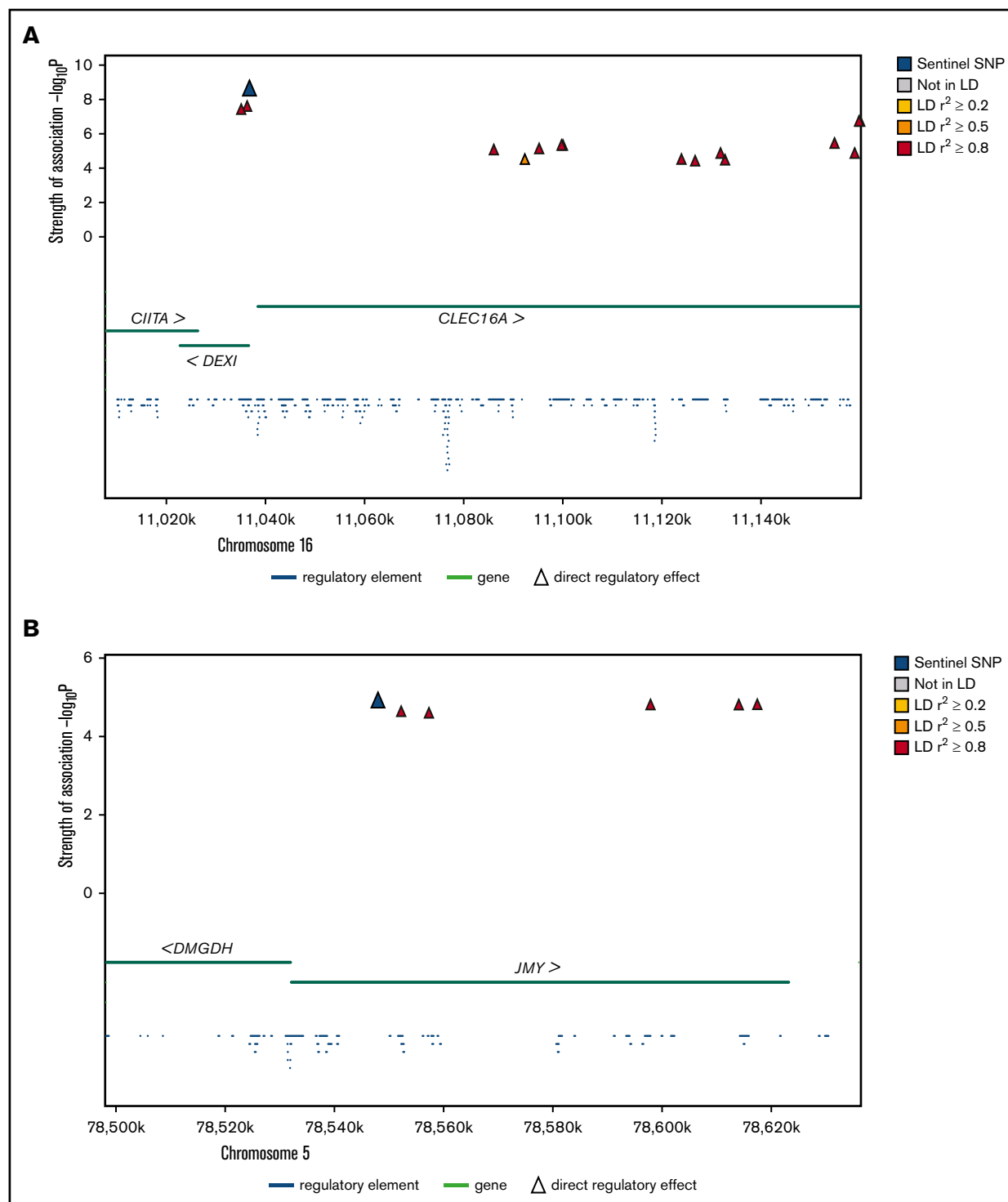
**Figure 2. Gene expression signature is predictive of relapse in patients with CN-AML.** (A) Receiver operating characteristic (ROC) curve shows the sensitivity and specificity of 10-gene expression signature for predicting relapse in 268 adult CN-AML patients younger than 60 years. (B) Predicted relapse 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,<sup>31</sup> 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  $\geq 3$  years.

Our findings were validated using publicly available data from the TCGA,<sup>31</sup> 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

**Table 2. Multivariable analyses for outcome**

Variable	Categories	P	Odds/hazards ratio (95% CI)
<b>Logistic regression model for relapse</b>			
10-gene signature	Continuous, 10% increase	<.001	1.79 (1.52-2.13)
Biallelic <i>CEBPA</i> mutation status	Mutated vs wild-type	.007	0.21 (0.07-0.66)
<i>FLT3</i> -ITD	Present vs absent	.04	2.14 (1.02-4.50)
<i>NPM1</i> mutation status	Mutated vs wild-type	.03	0.33 (0.12-0.90)
<b>Cox proportional hazards regression model for DFS</b>			
10-gene signature	Continuous, 10% increase	<.001	1.32 (1.22-1.43)
Biallelic <i>CEBPA</i> mutation status	Mutated vs wild-type	.01	0.53 (0.32-0.87)
<i>FLT3</i> -ITD	Present vs absent	.005	1.62 (1.16-2.26)
<i>MN1</i> expression	High vs low (median)	.008	1.58 (1.13-2.22)



**Figure 3. eQTLs regional association plots.** Plots show SNPs associated with expression of *DEXI* (A) and *JMY* (B). The top track indicates negative  $\log_{10} P$  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.

## References

1. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424-447.
2. Mrózek K, Radmacher MD, Bloomfield CD, Marcucci G. Molecular signatures in acute myeloid leukemia. *Curr Opin Hematol*. 2009;16(2):64-69.
3. Theilgaard-Mönch K, Boultonwood J, Ferrari S, et al. Gene expression profiling in MDS and AML: potential and future avenues. *Leukemia*. 2011;25(6):909-920.
4. Wouters BJ, Löwenberg B, Delwel R. A decade of genome-wide gene expression profiling in acute myeloid leukemia: flashback and prospects. *Blood*. 2009;113(2):291-298.
5. Valk PJM, Verhaak RGW, Beijen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med*. 2004;350(16):1617-1628.
6. Bullinger L, Döhner K, Bair E, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *N Engl J Med*. 2004;350(16):1605-1616.
7. Radmacher MD, Marcucci G, Ruppert AS, et al. Independent confirmation of a prognostic gene-expression signature in adult acute myeloid leukemia with a normal karyotype: a Cancer and Leukemia Group B study. *Blood*. 2006;108(5):1677-1683.
8. Metzeler KH, Hummel M, Bloomfield CD, et al; Cancer and Leukemia Group B; German AML Cooperative Group. An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. *Blood*. 2008;112(10):4193-4201.
9. Ng SW, Mitchell A, Kennedy JA, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature*. 2016;540(7633):433-437.
10. Bill M, Nicolet D, Kohlschmidt J, et al. Mutations associated with a 17-gene leukemia stem cell score and the score's prognostic relevance in the context of the European LeukemiaNet classification of acute myeloid leukemia. *Haematologica*. 2020;105(3):721-729.
11. Eisfeld A-K, Mrózek K, Kohlschmidt J, et al. The mutational oncoprint of recurrent cytogenetic abnormalities in adult patients with *de novo* acute myeloid leukemia. *Leukemia*. 2017;31(10):2211-2218.
12. Whitman SP, Archer KJ, Feng L, et al. Absence of the wild-type allele predicts poor prognosis in adult *de novo* acute myeloid leukemia with normal cytogenetics and the internal tandem duplication of *FLT3*: a Cancer and Leukemia Group B study. *Cancer Res*. 2001;61(19):7233-7239.
13. Marcucci G, Mahary K, Radmacher MD, et al. Prognostic significance of, and gene and microRNA expression signatures associated with, *CEBPA* mutations in cytogenetically normal acute myeloid leukemia with high-risk molecular features: a Cancer and Leukemia Group B Study. *J Clin Oncol*. 2008;26(31):5078-5087.

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

## Authorship

Contribution: C.J.W. and A.-K.E. conceived and designed the study; C.J.W., H.G.O., J.K., D.N., L.K.G., and M.B. performed bioinformatics and biostatistics analyses; D.P. assisted with RNAseq; K.M. assisted with manuscript writing and preparation and cytogenetics review; A.J.C. performed cytogenetics review; B.L.P., G.L.U., J.E.K., R.M.S., R.G., and J.C.B. treated patients and collected samples and clinical data; A.d.l.C. and C.D.B. supervised the project; and all authors read the manuscript and approved its final version.

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](mailto:christopher.walker@osumc.edu).

14. Mrózek K, Carroll AJ, Maharry K, et al. Central review of cytogenetics is necessary for cooperative group correlative and clinical studies of adult acute leukemia: the Cancer and Leukemia Group B experience. *Int J Oncol.* 2008;33(2):239-244.
15. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907-915.
16. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014; 30(7):923-930.
17. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
18. Wang J, Wen S, Symmans WF, Pusztai L, Coombes KR. The bimodality index: a criterion for discovering and ranking bimodal signatures from cancer gene expression profiling data. *Cancer Inform.* 2009;7:199-216.
19. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5-32.
20. Walker CJ, Oakes CC, Genutis LK, et al. Genome-wide association study identifies an acute myeloid leukemia susceptibility locus near *BICRA*. *Leukemia.* 2019;33(3):771-775.
21. McCarthy S, Das S, Kretzschmar W, et al; Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279-1283.
22. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics.* 2012;28(10):1353-1358.
23. Marcucci G, Baldus CD, Ruppert AS, et al. Overexpression of the ETS-related gene, *ERG*, predicts a worse outcome in acute myeloid leukemia with normal karyotype: a Cancer and Leukemia Group B study. *J Clin Oncol.* 2005;23(36):9234-9242.
24. Marcucci G, Maharry K, Whitman SP, et al. High expression levels of the ETS-related gene, *ERG*, predict adverse outcome and improve molecular risk-based classification of cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B Study. *J Clin Oncol.* 2007;25(22): 3337-3343.
25. Langer C, Radmacher MD, Ruppert AS, et al. High *BAALC* expression associates with other molecular prognostic markers, poor outcome and a distinct gene-expression signature in cytogenetically normal patients younger than 60 years with acute myeloid leukemia: a Cancer and Leukemia Group B (CALGB) study. *Blood.* 2008;111(11):5371-5379.
26. Schwind S, Marcucci G, Maharry K, et al. *BAALC* and *ERG* expression levels are associated with outcome and distinct gene and microRNA expression profiles in older patients with de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood.* 2010;116(25): 5660-5669.
27. Heuser M, Beutel G, Krauter J, et al. High meningioma 1 (*MN1*) expression as a predictor for poor outcome in acute myeloid leukemia with normal cytogenetics. *Blood.* 2006;108(12):3898-3905.
28. Schwind S, Marcucci G, Kohlschmidt J, et al. Low expression of *MN1* associates with better treatment response in older patients with de novo cytogenetically normal acute myeloid leukemia. *Blood.* 2011;118(15):4188-4198.
29. Marcucci G, Maharry KS, Metzeler KH, et al. Clinical role of microRNAs in cytogenetically normal acute myeloid leukemia: *miR-155* upregulation independently identifies high-risk patients. *J Clin Oncol.* 2013;31(17):2086-2093.
30. Eisfeld A-K, Marcucci G, Maharry K, et al. *miR-3151* interplays with its host gene *BAALC* and independently affects outcome of patients with cytogenetically normal acute myeloid leukemia. *Blood.* 2012;120(2):249-258.
31. Ley TJ, Miller C, Ding L, et al; Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med.* 2013;368(22):2059-2074.
32. Wouters BJ, Löwenberg B, Erpelinck-Verschueren CAJ, van Putten WLJ, Valk PJM, Delwel R. Double *CEBPA* mutations, but not single *CEBPA* mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. *Blood.* 2009;113(13):3088-3091.
33. Taskesen E, Bullinger L, Corbacioglu A, et al. Prognostic impact, concurrent genetic mutations, and gene expression features of AML with *CEBPA* mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for *CEBPA* double mutant AML as a distinctive disease entity. *Blood.* 2011;117(8):2469-2475.
34. Scholl S, Melle C, Bleul A, et al. Specific pattern of protein expression in acute myeloid leukemia harboring FLT3-ITD mutations. *Leuk Lymphoma.* 2007; 48(12):2418-2423.
35. Papaioannou D, Nicolet D, Ozer HG, et al. Prognostic and biologic relevance of clinically applicable long non-coding RNA profiling in older patients with cytogenetically normal acute myeloid leukemia. *Mol Cancer Ther.* 2019;18(8):1451-1459.
36. Narrandes S, Xu W. Gene expression detection assay for cancer clinical use. *J Cancer.* 2018;9(13):2249-2265.
37. Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA.* 2010;304(24):2706-2715.
38. Elsayed AH, Rafiee R, Cao X, et al. A six-gene leukemic stem cell score identifies high risk pediatric acute myeloid leukemia [published correction appears in *Leukemia.* 2020;34(10):2821]. *Leukemia.* 2020;34(3):735-745.
39. Papaioannou D, Nicolet D, Volinia S, et al. Prognostic and biologic significance of long non-coding RNA profiling in younger adults with cytogenetically normal acute myeloid leukemia. *Haematologica.* 2017;102(8):1391-1400.
40. Whitman SP, Kohlschmidt J, Maharry K, et al. *GAS6* expression identifies high-risk adult AML patients: potential implications for therapy. *Leukemia.* 2014; 28(6):1252-1258.



41. Yang X, Shi J, Zhang X, et al. Expression level of GAS6-mRNA influences the prognosis of acute myeloid leukemia patients with allogeneic hematopoietic stem cell transplantation. *Biosci Rep*. 2019;39(5):BSR20190389.
42. Wu S, Zhang W, Shen D, Lu J, Zhao L. PLCB4 upregulation is associated with unfavorable prognosis in pediatric acute myeloid leukemia. *Oncol Lett*. 2019;18(6):6057-6065.
43. Kreuter M, Woelke K, Bieker R, et al. Correlation of neuropilin-1 overexpression to survival in acute myeloid leukemia. *Leukemia*. 2006;20(11):1950-1954.
44. Sallam TH, El Telbany MASE, Mahmoud HM, Iskander MA. Significance of neuropilin-1 expression in acute myeloid leukemia. *Turk J Haematol*. 2013;30(3):300-306.
45. Zhao J, Gu L, Li C, Ma W, Ni Z. Investigation of a novel biomarker, neuropilin-1, and its application for poor prognosis in acute myeloid leukemia patients. *Tumour Biol*. 2014;35(7):6919-6924.
46. Wen H, Li Y, Malek SN, et al. New fusion transcripts identified in normal karyotype acute myeloid leukemia. *PLoS One*. 2012;7(12):e51203.
47. Adighibe O, Turley H, Leek R, et al. JMY protein, a regulator of P53 and cytoplasmic actin filaments, is expressed in normal and neoplastic tissues. *Virchows Arch*. 2014;465(6):715-722.
48. Thomassen M, Tan Q, Kruse TA. Gene expression meta-analysis identifies chromosomal regions and candidate genes involved in breast cancer metastasis [published correction appears in *Breast Cancer Res Treat*. 2009;113(2):251-252]. *Breast Cancer Res Treat*. 2009;113(2):239-249.
49. van den Boom J, Wolter M, Blaschke B, Knobbe CB, Reifenberger G. Identification of novel genes associated with astrocytoma progression using suppression subtractive hybridization and real-time reverse transcription-polymerase chain reaction. *Int J Cancer*. 2006;119(10):2330-2338.