Use of Gene-Expression Profiling to Identify Prognostic Subclasses in Adult Acute Myeloid Leukemia

Lars Bullinger, M.D., Konstanze Döhner, M.D., Eric Bair, Stefan Fröhling, M.D., Richard F. Schlenk, M.D., Robert Tibshirani, Ph.D., Hartmut Döhner, M.D., and Jonathan R. Pollack, M.D., Ph.D.

ABSTRACT

BACKGROUND
In patients with acute myeloid leukemia (AML), the presence or absence of recurrent cytogenetic aberrations is used to identify the appropriate therapy. However, the current classification system does not fully reflect the molecular heterogeneity of the disease, and treatment stratification is difficult, especially for patients with intermediate-risk AML with a normal karyotype.

METHODS
We used complementary-DNA microarrays to determine the levels of gene expression in peripheral-blood samples or bone marrow samples from 116 adults with AML (including 45 with a normal karyotype). We used unsupervised hierarchical clustering analysis to identify molecular subgroups with distinct gene-expression signatures. Using a training set of samples from 59 patients, we applied a novel supervised learning algorithm to devise a gene-expression–based clinical-outcome predictor, which we then tested using an independent validation group comprising the 57 remaining patients.

RESULTS
Unsupervised analysis identified new molecular subtypes of AML, including two prognostically relevant subgroups in AML with a normal karyotype. Using the supervised learning algorithm, we constructed an optimal 133-gene clinical-outcome predictor, which accurately predicted overall survival among patients in the independent validation group (P=0.006), including the subgroup of patients with AML with a normal karyotype (P=0.046). In multivariate analysis, the gene-expression predictor was a strong independent prognostic factor (odds ratio, 8.8; 95 percent confidence interval, 2.6 to 29.3; P<0.001).

CONCLUSIONS
The use of gene-expression profiling improves the molecular classification of adult AML.
Acute Myeloid Leukemia (AML) is the most common acute leukemia in adults. Chemotherapy induces a complete remission in 70 to 80 percent of younger patients (age, 16 to 60 years), but many of them have a relapse and die of their disease. Myeloablative conditioning followed by allogeneic stem-cell transplantation can prevent relapse, but this approach is associated with a high treatment-related mortality. Therefore, accurate predictors of the clinical outcome are needed to determine appropriate treatment for individual patients.

Currently used prognostic indicators include age, cytogenetic findings, the white-cell count, and the presence or absence of an antecedent hematologic disorder (e.g., myelodysplasia). Of these, cytogenetic findings represent the most powerful prognostic factor. The karyotype can be used to classify patients as being at low risk (t(8;21), t(15;17), or inv(16)), intermediate risk (e.g., a normal karyotype or t(9;11)), or high risk (e.g., inv(3), -5/del(5q), -7, or a complex karyotype [three or more aberrations]). Nevertheless, there is substantial heterogeneity within these risk groups. Thirty-five to 50 percent of patients have a normal karyotype, but molecular markers such as mutations in the fms-like tyrosine kinase 3 (FLT3) gene and the mixed-lineage leukemia (MLL) gene have allowed us to begin to subdivide this large group. These markers have been shown to predict the clinical outcome, and they provide potential targets for molecular therapies. Despite these successes, however, there is no consensus as to the appropriate means of risk stratification of patients with AML with a normal karyotype. We therefore used DNA microarrays to explore systematically the molecular variation underlying the biologic and clinical heterogeneity in AML, an approach that has provided insight into diffuse large-B-cell lymphoma and childhood acute lymphoblastic leukemia.

**Methods**

**Samples**

The AML Study Group Ulm (Ulm, Germany) provided 65 peripheral-blood samples and 54 bone marrow specimens from 116 adult patients with AML. Written informed consent was obtained from all patients, and the study was approved by the institutional review board of each participating center. After providing samples, the patients began one of two treatment protocols (AML HD98A and AML HD98B, described in detail in Supplementary Appendix 1, available with the full text of this article at www.nejm.org) between February 1998 and November 2001 and received intensive induction and consolidation therapy. The median duration of follow-up was 334 days (611 days for survivors); during this period, 68 of the 116 patients died and 34 of the 79 patients who had a complete remission relapsed. Conventional cytogenetic banding, fluorescence in situ hybridization, and analysis of MLL and FLT3 for mutations were performed as previously described, at the central reference laboratory for cytogenetic and molecular diagnostics of the AML Study Group Ulm. Detailed clinical, cytogenetic, and molecular cytogenetic information is available at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/, accession number GSE425).

**Gene-expression profiling**

We isolated total RNA from stored, frozen mononuclear AML-cell pellets using Trizol reagent (Invitrogen) according to the manufacturer’s recommendations and assessed RNA quality by means of gel electrophoresis. We hybridized Cy5-labeled total RNA from AML samples, along with Cy3-labeled common reference messenger RNA (mRNA) (pooled from 11 cell lines), on microarrays of complementary DNA (cDNA) (manufactured by the Stanford Functional Genomics Facility) that contain 39,711 nonredundant cDNA clones, representing 26,260 unique UniGene clusters (i.e., genes). Details of cDNA-microarray fabrication, prehybridization array processing, and RNA-sample labeling and hybridization have been described elsewhere. We imaged arrays using an Axon GenePix 4000B scanner (Axon Instruments), determined fluorescence ratios (ratio of the specimen value to the reference value) using the GenePix software, and entered data into a data base (Stanford Microarray Database) for subsequent analysis. The complete-microarray data set is also available at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/, accession number GSE425).

**Statistical analysis**

We normalized fluorescence ratios by mean-centering genes for each array and then mean-centering each gene across all arrays within each of three array print runs, to minimize potential print-run–specific bias. For all subsequent analyses, we included only the 6283 genes on the microarray whose expression was both well measured and highly variable among
samples (a list is available at www.ncbi.nlm.nih.gov/geo/). We defined well-measured genes as genes that had a ratio of signal intensity to background noise of more than 2, for either the Cy5-labeled AML samples (a list is available at www.ncbi.nlm.nih.gov/microarrays (SAM) method, for outcome prediction, we randomly divided samples that had been prestratified to ensure that a similar number of samples in each group were from patients who had died into a separate training set (59 samples) and test set (57 samples); in the case of paired peripheral-blood and bone marrow samples (obtained from three patients), only 1 sample was used. In the training set, we used the SAM method, which involved a modified Cox proportional-hazards analysis was performed with the use of the R software package. Multivariate proportional-hazards analysis was performed with the use of WinStat software (R. Fitch Software). Multivariate proportional-hazards analysis was performed with the use of the R software package.

For outcome prediction, we randomly divided samples that had been prestratified to ensure that a similar number of samples in each group were from patients who had died into a separate training set (59 samples) and test set (57 samples); in the case of paired peripheral-blood and bone marrow samples (obtained from three patients), only 1 sample was used. In the training set, we used the SAM method, which involved a modified Cox proportional-hazards maximum-likelihood score, to identify genes whose expression correlated with the duration of survival. We used this set of SAM genes in k-means cluster analysis to identify two subgroups of samples in the training set. We used Kaplan–
Meier survival analysis to determine the prognostic relevance of the two subgroups — and to assign good-outcome and poor-outcome labels to each subgroup — and the prediction analysis for microarrays method, using the “nearest shrunken centroid” approach to identify a 10-fold cross-validated gene-expression predictor (analogous to the leave-one-out method) for these cluster-defined outcome classes. Taking into account both the P value on the log-rank test and the cross-validation error rate (see Supplementary Appendix 2, available with the full text of this article at www.nejm.org), we selected a set of 133 predictive genes (represented by 149 cDNAs), which we used for all subsequent analyses. We used cluster analysis or the nearest-shrunken-centroid method to determine the prognostic accuracy of outcome classes in the test set.

RESULTS

IDENTIFICATION OF CLASSES

Our sample set included the most common cytogenetic subtypes of AML and reflected the spectrum of cytogenetic aberrations in AML.

To explore the relationship among samples, as well as the underlying patterns of gene expression, we performed an unsupervised two-way, hierarchical cluster analysis using the 6283 genes whose expression varied most across samples (Fig. 1A). For patients for whom we had samples from both peripheral blood and bone marrow, we found that the expression profiles were highly correlated (Fig. 1B), as has been reported elsewhere. Of the cytogenetic groups, samples with t(15;17) had a highly correlated pattern of expression, whereas samples with t(8;21) or inv(16) were less well correlated, with each group being divided into separate clusters (Fig. 1B). Interestingly, AML specimens with a normal karyotype (as determined by conventional chromosome-banding and fluorescence in situ hybridization analysis) also segregated mainly into two distinct groups, each of which included a small number of AML specimens from other classes (Fig. 2A).

These newly defined subgroups were identified with the use of a variety of preclustering data-filtering criteria (Supplementary Appendix 3, available with the full text of this article at www.nejm.org) and were also evident by means of principal-component analysis (Fig. 2B), suggesting they represent robust classes.

To gain further insight into the importance of these newly identified subtypes, we examined the distribution of prognostically relevant clinical and molecular genetic variables among samples (Fig. 2A). The two subclasses in which a normal karyotype predominated were similar with respect to the patients’ sex, age, white-cell count, serum lactate dehydrogenase level, and presence or absence of an antecedent hematologic disorder (described at www.ncbi.nlm.nih.gov/geo/). FLT3 aberrations were more prevalent in group I (P=0.005 by the chi-square test), and French–American–British (FAB) morphologic subtype M1 or M2 was significantly more common in group I than in group II, whereas FAB subtype M4 or M5 was more common in group II (P=0.013 by the chi-square test). It is noteworthy that Kaplan–Meier analysis identified a significant difference in overall survival between the two subclasses (P=0.009 by the log-rank test) (Fig. 2C). Within our sample set, no significant differences in clinical and laboratory variables were identified between gene-expression subgroups for either t(8;21) or inv(16).

BIOLGIC INSIGHTS

Within the unsupervised hierarchical cluster, we found gene-expression signatures characterizing known cytogenetic groups, as well as newly identified subtypes (Fig. 1). Group signatures could also be identified with the use of supervised analyses, such as the SAM method. In both supervised and unsupervised analyses, gene-expression signatures were identified for groups with t(15;17), t(8;21), inv(16), 11q23 aberrations, del(7q)/−7, a normal
karyotype, and FLT3 mutations, as well as for the newly defined subgroups within the t(8;21), inv(16), and normal-karyotype groups (Fig. 1 and Supplementary Appendix 4, available with the full text of this article at www.nejm.org; and at www.ncbi.nlm.nih.gov/geo). In contrast, we identified no such characteristic signatures for AML specimens with a complex karyotype, MLL partial tandem duplications, and trisomy 8 (whose molecular heterogeneity has been reported), though this may reflect our limited statistical power owing to the small sizes of the groups.

Among the group-specific signatures, we found genes located at translocation breakpoints defining cytogenetic classes, including ETO in t(8;21) and MYH11 in inv(16) (described at www.ncbi.
also known as AML. It has been associated with resistance to cytarabine in AML specimens in group II were characterized in part by a prominent gene-expression feature (Fig. 1C). The transcriptional regulator GATA2 is required for NOTCH1 signaling-induced inhibition of hematopoietic differentiation. Consistent with this finding, many group I specimens also had elevated NOTCH1 expression. The high level of expression of DNMT3A and DNMT3B among group I specimens also suggests a potential role of aberrant patterns of methylation in the pathogenesis of this subtype.

AML specimens in group II were characterized in part by a prominent gene-expression feature (Fig. 1C) associated with granulocytic or monocytic differentiation and the immune response. A candidate pathogenic gene within this subgroup was the gene for vascular endothelial growth factor (VEGF), which is involved in the regulation of hematopoietic-stem-cell survival and in the progression of AML.

Likewise, the t(8;21) signature (Fig. 1D) included MLLT4 (also known as AF6), a recurrent fusion partner of MLL in leukemias with t(6;11), suggesting a possible shared mechanism contributing to leukaemogenesis. In specimens with inv(16), we found high levels of expression of NTSE (5’ nucleotidase, also known as 5NT or CD73) (Fig. 1H), which has been associated with resistance to cytarabine in AML. This finding is somewhat surprising, since AML with inv(16) is clinically quite sensitive to cytarabine. In contrast to childhood acute lymphoblastic leukemia, in AML, the expression of putative pathogenic homeobox genes, including HOXA4, HOXA9, HOXA10, PBX3, and MEIS1 (some of which are shown in Fig. 1E), was not limited to specimens with MLL translocations but also characterized many specimens with normal and complex karyotypes.

Among the subtypes in which the normal karyotype predominated, group I was characterized by a high level of expression of GATA2, DNMT3A, and DNMT3B (Fig. 1C). The transcriptional regulator GATA2 is required for NOTCH1 signaling-induced inhibition of hematopoietic differentiation. Consistent with this finding, many group I specimens also had elevated NOTCH1 expression. The high level of expression of DNMT3A and DNMT3B among group I specimens also suggests a potential role of aberrant patterns of methylation in the pathogenesis of this subtype.

AML specimens in group II were characterized in part by a prominent gene-expression feature (Fig. 1C) associated with granulocytic or monocytic differentiation and the immune response. A candidate pathogenic gene within this subgroup was the gene for vascular endothelial growth factor (VEGF), which is involved in the regulation of hematopoietic-stem-cell survival and in the progression of AML.
A strong correspondence was observed between all survival (P=0.046 by the log-rank test) (Fig. 4D). Notably, when we applied the same procedure to the subgroup of 22 samples with the good-outcome signature (P=0.006 by the log-rank test) (Fig. 4C), there was a significant difference between groups (P<0.001 by the chi-square test).

Therefore, we instead devised a strategy for outcome prediction that combined the strengths of supervised and unsupervised approaches (Fig. 3). The idea was to try to identify the prognostically relevant, underlying biologic subclases. First, we randomly classified AML samples into separate training and test sets. In the training set, we used a supervised analysis (the SAM method) to identify genes whose expression correlated with the duration of survival. Next, we used these genes in an unsupervised cluster analysis to determine the underlying, prognostically relevant AML classes (i.e., good and poor outcomes) in the training set. We then devised a cross-validated gene-expression predictor for these cluster-defined outcome classes, using the prediction analysis of microarrays (PAM) method based on nearest shrunken centroids. We then validated this class predictor, comprising 133 unique genes (represented by 149 cDNAs) (Fig. 4A and www.ncbi.nlm.nih.gov/geo/), by using it to predict which outcome class samples in the independent test set would be included.

To predict outcome class in the test set, we performed hierarchical clustering using the 133 predictive genes, which yielded a cluster of samples with gene-expression profiles that were highly correlated with the good-outcome group and a cluster with profiles that were highly correlated with the poor-outcome group in the training set (P<0.001) (Fig. 4B and Supplementary Appendix 6, available with the full text of this article at www.nejm.org). The cluster-defined subgroup of samples having the poor-outcome signature was associated with significantly shorter survival than was the subgroup of samples with the good-outcome signature (P=0.006 by the log-rank test) (Fig. 4C). Notably, when we applied the same procedure to the subgroup of 22 AML samples with a normal karyotype, it also identified good-outcome and poor-outcome classes associated with significant differences in overall survival (P=0.046 by the log-rank test) (Fig. 4D). A strong correspondence was observed between samples represented in our group I and group II subtypes and samples predicted to have a poor and a good outcome, respectively (P<0.001 by the chi-square test).

The preceding method required a group of test samples in order to predict, by means of cluster analysis, the outcome class for individual patients. Because it is useful clinically to predict the outcome for individual patients who are not part of a test group, we also evaluated a procedure to predict the outcome class of individual test samples, based on the PAM method of nearest shrunken centroids. Each test-set sample was individually assigned to an outcome class by determining whether its gene-expression signature across the 133 predictive genes was more highly correlated with the average (centroid) good-outcome signature or with the average poor-outcome signature in the training set. With the use of this procedure, the subgroup of samples predicted to have a poor outcome was associated with significantly shorter survival than the subgroup of samples predicted to have a good outcome (P=0.034 by the log-rank test) (Supplementary Appendix 7, available with the full text of this article at www.nejm.org). However, when we used this method on AML samples with a normal karyotype, we found...
no significant difference in overall survival (P=0.65 by the log-rank test), which may reflect the relatively small sample or an inherently poorer performance of this alternative approach to outcome prediction.

To determine whether the gene-expression outcome predictor added prognostic information over and above that provided by known prognostic factors, we performed multivariate proportional-hazards analysis. Using the cluster-defined outcome-class labels (Fig. 4B and www.ncbi.nlm.nih.gov/geo/), we found that the gene-expression predictor provided significant prognostic information (odds ratio, 8.8; 95 percent confidence interval, 2.6 to 29.3; P<0.001) that was independent of other risk factors determined to be significant in the model: antecedent hematologic disorder (odds ratio, 10; 95 percent confidence interval, 2.8 to 37.2; P<0.001), combined intermediate- and high-risk cytogenetics (P=0.004), and FLT3 mutations (odds ratio, 3.0; 95 percent confidence interval, 1.2 to 7.7; P=0.03). Using the nearest centroid-defined class labels, we obtained similar results (available at www.ncbi.nlm.nih.gov/geo/). When samples with a normal karyotype were excluded, the gene-expression predictor was still a significant variable, demonstrating that it is not only capturing the survival distinction among AML specimens with a normal karyotype, but also providing additional prognostic information for specimens with a non-normal karyotype (data not shown).

The 133-gene outcome predictor included several named genes with potential pathogenic relevance. Genes associated with favorable outcome included the forkhead box O1A gene (FOXO1A, also known as FKHR), which is involved in the arrest of the cell cycle and the regulation of apoptosis. Interestingly, other members of the forkhead family have been identified as pathogenic translocation partners with MLL in acute leukemias, and a synthetic fusion of MLL with FOXO1A has recently been shown to transform hematopoietic progenitor cells in vitro.

Notably, among the genes associated with a poor outcome, several (e.g., MAP7, GUCY1A3, TCF4, and MS12) (some of which are shown in Fig. 1C) were coexpressed within a single-gene expression feature in our unsupervised hierarchical cluster, suggesting the possibility of a coregulated physiological process or pathway with pathogenetic relevance. The association of the overexpression of HOXB2, HOXB5, PBX3, HOXA4, and HOXA10 with a poor outcome supports the concept that homeobox-gene dysregulation has a role in leukemogenesis. Indeed, overexpression of HOXA10 has been shown to perturb myeloid and lymphoid differentiation profoundly in hematopoietic cells in mice and to lead to AML. Interestingly, elevated expression of FLT3 was also associated with a poor outcome. Activating FLT3 mutations are predictive of a poor outcome in AML, but we found no correlation between the levels of FLT3 expression and FLT3 mutational status in our AML sample set (P=0.57 by Student’s t-test). This finding suggests that increased expression of wild-type FLT3 may functionally mimic mutational activation and contribute to the pathogenesis of poor-outcome AML.

**Discussion**

We found that AML samples with a normal karyotype separated into two subgroups based on distinct patterns of gene expression revealed by unsupervised hierarchical clustering and principal component analysis. The unequal distribution of FLT3 mutations and FAB morphologic subtypes between groups with different outcomes supports the concept that distinct biologic changes underlie the clinical phenotype. The identification of these new subgroups suggests that the use of gene-expression profiling can improve the accuracy of the molecular classification of AML and that the study of the genes that are differentially expressed in the two subgroups will help identify the distinct pathways involved in the molecular pathogenesis of AML with a normal karyotype.

Using hierarchical clustering, we also found that samples with t(8;21) and inv(16) each separate into different subgroups. Since the primary translocation events themselves are not sufficient for leukemogenesis, the distinct patterns of gene expression found within each of these cytogenetic groups may lead to the identification of cooperating mutations and dysregulated pathways that eventuate in transformation. Analysis of additional samples will be required to determine the biologic and clinical relevance of these putative subgroups. Nevertheless, the value of unsupervised analytic methods is worth noting, since this molecular heterogeneity was not apparent in the supervised analysis.

Our gene-expression study has provided numerous insights into the pathogenesis of AML, including, for example, the role of homeobox-gene dysregulation. Our finding that HOXA4, HOXA9, HOXA10, PBX3, and MEIS1 are coexpressed across
diverse cytogenetic groups (e.g., AML specimens with 11q23 aberrations, specimens with a normal karyotype, and specimens with a complex karyotype) suggests a coregulated pathway with pathogenetic relevance in a subgroup of AML. Coexpression of HOXA9 and MEIS1, which is sufficient for the transformation of bone marrow cells in mice, has also recently been observed in children with acute lymphoblastic leukemia with MLL rearrangements, suggesting a possible shared pathogenic mechanism in acute myeloid and lymphoid leukemias. The pathogenetic relevance of the expression of the homeobox genes, as well as numerous other genes, in the data set remains to be explored.

We also developed an algorithm combining supervised and unsupervised approaches to identify a clinical outcome predictor based on gene expression, which we validated in an independent set of AML samples. The gene-expression predictor defined good-outcome and poor-outcome subgroups with significant differences in overall survival, whether they were applied to AML samples encompassing all cytogenetic groups or (for the cluster-derived classes) only to AML samples with a normal karyotype. The latter finding suggests the prognostic usefulness of the approach in this important class of intermediate-risk patients.

In multivariate analysis we found that the gene-expression outcome-class predictor provided prognostic information over and above that provided by known prognostic indicators. Therefore, our data suggest that outcome prediction can be optimized through the use of a combination of prognostic markers, including a gene-expression–based predictor. Although our patients were treated according to two distinct protocols involving various treatments, the protocols were based on a state–of–the–art strategy of intensive treatment, and it is therefore reasonable to expect that our findings can be extrapolated to current treatment protocols. Of course, it will be important to refine and validate our gene-expression predictor in a larger, independent set of AML samples and in a prospective cohort of patients before its routine implementation in clinical practice. Further studies will also be required to determine the ability of this predictor to classify risk in individual patients with AML. Nonetheless, our data support the theory that prognostic gene-expression signatures are present at diagnosis in the bulk population of leukemic cells and that the use of gene-expression profiling will improve molecular classification and outcome prediction in adult AML.

Supported by funds from the Stanford University Department of Pathology. Dr. Bullinger was supported in part by the Deutsche Forschungsgemeinschaft, Bonn, Germany (Forschungsstipendium BU 1339/1).

We are indebted to the members of the AML Study Group Ulm for providing leukemia specimens, to Mike Fero and the staff of the Stanford Functional Genomics Facility for providing high-quality cDNA microarrays, and to Gavin Sherlock and the staff of the Stanford Microarray Database group for providing outstanding database support.

APPENDIX

The following AML Study Group Ulm institutions and investigators participated in the study: Universitätssklinikum Bonn, Bonn, Germany, A. Glasmacher; Universitätssklinikum Düsseldorf, Düsseldorf, Germany, U. Germing; Universitätssklinikum Giessen, Giessen, Germany, H. Pralle; Universitätssklinikum Göttingen, Göttingen, Germany, D. Haase; Allgemeines Krankenhaus Altona, Hamburg, Germany, H. Salwender; Universitätsskliniken des Saarlandes, Homburg, Germany, F. Hartmann; Universitätssklinikum Innsbruck, Innsbruck, Austria, G. Gastl; Städtisches Klinikum, Karlsruhe, Germany, J.T. Fischer; Universitätssklinikum Kiel, Kiel, Germany, M. Kneba; Klinikum Rechts der Isar der Technischen Universität, Munich, K. Götze; Städtisches Krankenhaus, München-Schwabing, Germany, C. Waterhouse; Städtisches Klinikum, Oldenburg, Germany, F. del Valle; CaritasKlinik St. Theresia, Saarbrücken, Germany, J. Preiß; Bürgerhospital, Stuttgart, Germany, W. Griminger; Katharinenhospital, Stuttgart, Germany, H.G. Mergethaler; Krankenhaus der Barmherzigen Brüder, Trier, Germany, W. Weber; and Hanusch-Krankenhaus, Vienna, E. Koller.

REFERENCES


N ENGL J MED 350:16 WWW.NEJM.ORG APRIL 15, 2004

Copyright © 2004 Massachusetts Medical Society. All rights reserved.
mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML Study Group Ulm. Blood 2002;100:4372-80.

Copyright © 2004 Massachusetts Medical Society.