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.[1] 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).[2] Of these, cytogenetic findings represent the most powerful prognostic factor.[3,4] 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]).[3,5,6] Nevertheless, there is substantial heterogeneity within these risk groups. Thirty-five to 50 percent of patients have a normal karyotype,[7] but molecular markers such as mutations in the fms-like tyrosine kinase 3 (*FLT3*) gene[8,9] and the mixed-lineage leukemia (*MLL*) gene[10,11] 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.[12] 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[13-15] and childhood acute lymphoblastic leukemia.[16,17]

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,[8,11,18] 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.[19,20] 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)[21] 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.[22] For all subsequent analyses, we included only the 6283 genes on the microarray whose expression was both well measured and highly variable among

---

Glossary

Centroid: The average expression profile across a set of genes; the centroid corresponds to the center of a cluster.

Chromosomal abnormalities

    t(8;21): One of the commonest cytogenetic abnormalities in AML; produces a hybrid gene by fusing *AML1* on the long arm of chromosome 21 with *ETO* on the long arm of chromosome 8.

    inv(16): Inversion of a segment of chromosome 16 that produces the *CBFβ-MYH11* fusion.

    t(15;17): Reciprocal translocation of genetic material between the long arms of chromosomes 15 and 17 that produces the *PML-RARα* fusion gene, typical of acute promyelocytic leukemia.

    11q23: A chromosomal region that becomes rearranged with various partner chromosomal regions in diverse forms of leukemia. Usually involves the *MLL* gene whose multifunctional product regulates gene expression.

    t(6;9): A rare translocation often found in young patients and sometimes associated with basophilia.

    −7(q): Loss of the long arm of chromosome 7 (monosomy 7).

French–American–British (FAB) classification: An internationally agreed-on method of classifying acute leukemia by morphologic means. There are eight subtypes, ranging from M0 (myeloblasts) to M7 (megakaryoblasts).

Gene-clustering technique: A method of grouping genes with similar patterns of expression. These clusters often consist of genes that are mechanistically related (e.g., genes of a particular metabolic cycle).

Hierarchical clustering: A computational method that groups genes (or samples) into small clusters and then groups these clusters into increasingly higher level clusters. As a result, a dendrogram (i.e., tree) of connectivity emerges.

Microarray profiling, also called DNA-microarray or gene expression profiling: Determination of the level of expression of thousands of genes simultaneously by using, for example, a robotically spotted microarray of complementary DNAs or oligonucleotides. Messenger RNA extracted from the test tissue or cells and labeled with a fluorescent dye is tested for its ability to hybridize to the spotted nucleic acid.

Omniviz Correlation View: A commercial multifunctional statistical package used for analysis of microarrays. It allows the visual representation of genes that behave similarly in a given disease.

Pearson's correlation coefficient: A statistical measure of the strength of the linear relationship between variables.

Prediction analysis of microarrays (PAM): A statistical technique that identifies a subgroup of genes that best characterizes a predefined class and uses this gene set to predict the class of new samples.

Probe set: A group of 10 to 20 oligonucleotides; each set corresponds to one gene.

Significance analysis of microarrays (SAM): A statistical method used in microarray analyses that calculates a score for each gene and thus identifies genes with a statistically significant association with an outcome variable such as survival.

Supervised analysis: An analysis of the results of microarray profiling that takes external factors into account.

Unsupervised analysis: An analysis of the results of microarray profiling that does not take external factors such as survival or clinical signs into account.

10-Fold cross-validation: A validation method that works as follows: the model is fitted on 90 percent of the samples, and the class of the remaining 10 percent is then predicted. This procedure is repeated 10 times, with each part playing the role of the test samples and the error of all 10 parts added together to compute the overall error. This error reflects the number of misclassified samples.

---

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 sample or the Cy3-labeled reference sample, in at least 75 percent of the AML samples hybridized. We defined genes that were highly variably expressed as genes whose expression was higher or lower by a factor of at least 4 than the average expression of all AML samples in at least two AML samples. For hierarchical clustering, we applied two-way (genes-against-samples) average-linkage hierarchical clustering[19] and used TreeView to visualize the results.[19] Principal component analysis[23] was performed with the use of the R software package (available at www.r-project.org). For two-class and multiclass supervised analyses, we used the significance analysis of microarrays (SAM) method,[24] which uses a modified t-test statistic (or F-test statistic for multiclass analysis), with sample-label permutations to evaluate statistical significance. The chi-square test, Student's t-test, and Kaplan–Meier survival analysis were 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,[25] 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[26] or the nearest-shrunken-centroid method[25] 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[19] 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.[17] 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[23] (Fig. 2B), suggesting they represent robust classes.

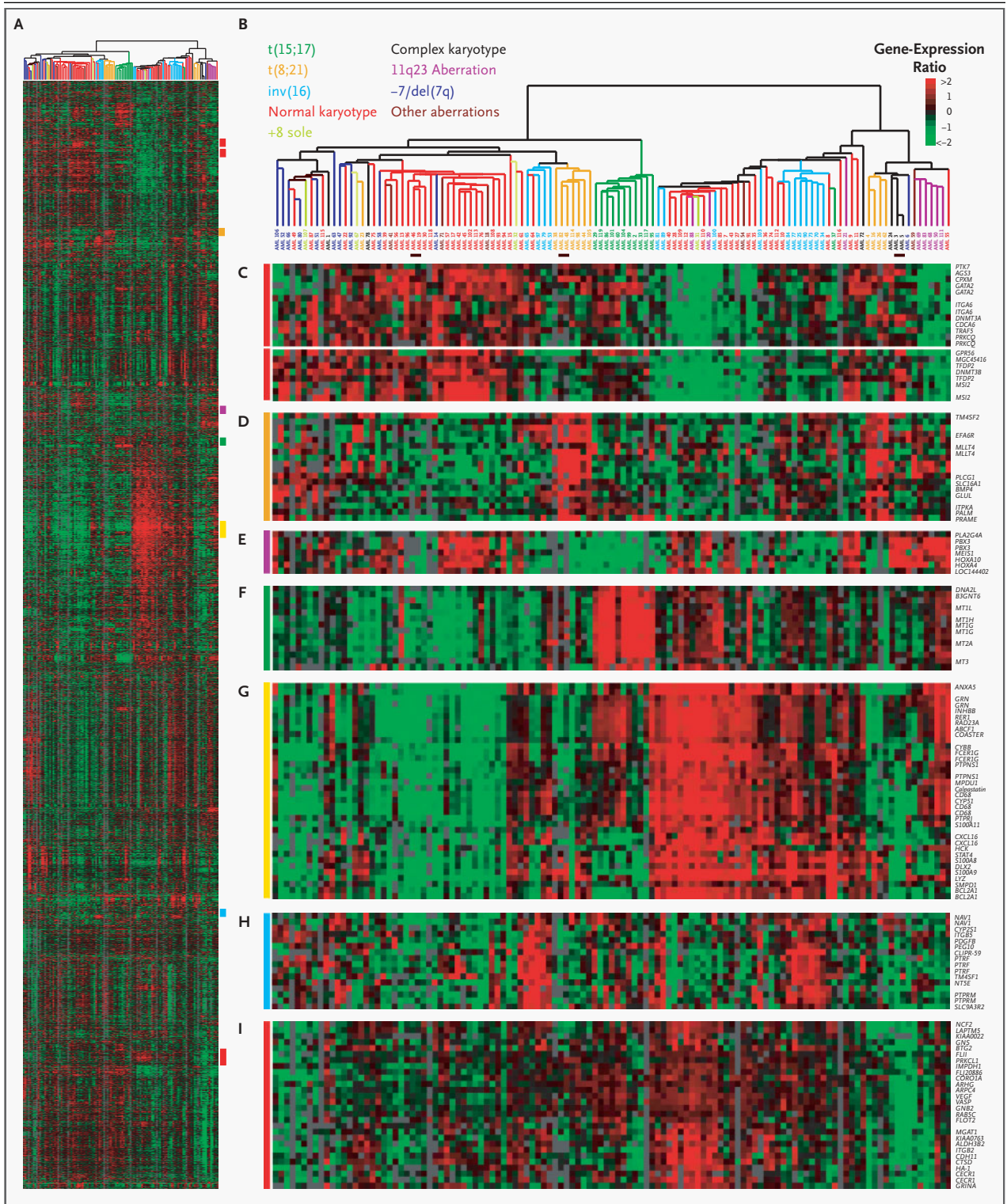To gain further insight into the importance of these newly identified subtypes, we examined the

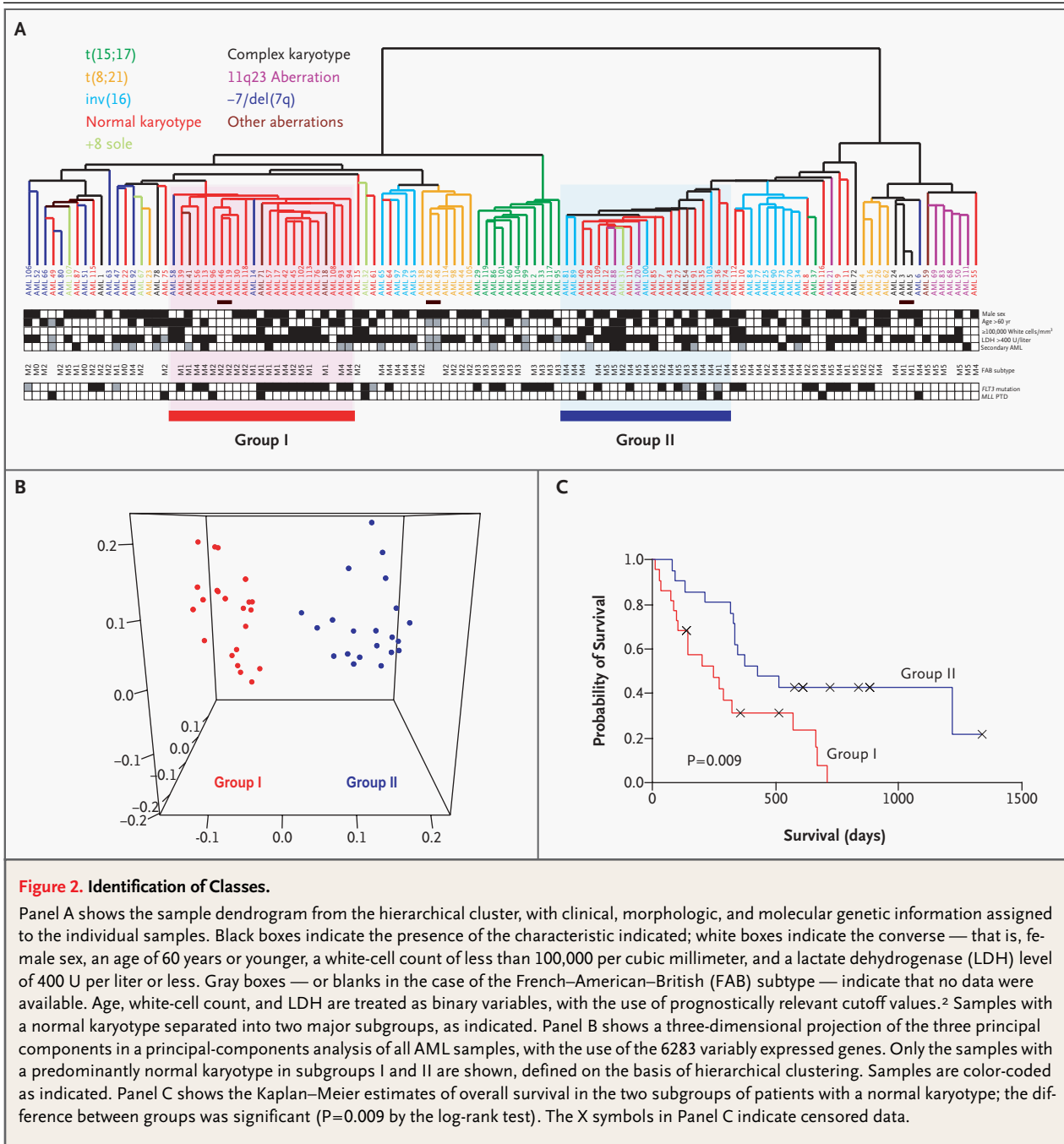**Figure 1 (facing page). Hierarchical Cluster Analysis of Diagnostic AML Samples.**

Panel A shows a thumbnail overview of the two-way (genes against samples) hierarchical cluster of 119 numbered samples of AML (columns) and 6283 genes with variable levels of expression (rows). Mean-centered ratios of gene expression are depicted by a log-transformed (on a base 2 scale) pseudocolor scale. Gray areas indicate poorly measured genes (genes with a ratio of signal intensity to background noise of 2 or less). Panel B shows an enlarged view of the sample dendrogram. Samples are color-coded according to the prognostically relevant cytogenetic groups, determined on the basis of conventional chromosome-banding and fluorescence in situ hybridization analysis. Three paired samples of peripheral blood and bone marrow from three patients are indicated by horizontal black bars. Panels C, D, E, F, G, H, and I show selected gene-expression features whose locations are indicated by the vertical colored bars. Owing to space limitations, only named genes (and not expressed-sequence tags) are indicated.

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).

BIOLOGIC 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.[24] 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

Figure 2. Identification of Classes.

Panel A shows the sample dendrogram from the hierarchical cluster, with clinical, morphologic, and molecular genetic information assigned to the individual samples. Black boxes indicate the presence of the characteristic indicated; white boxes indicate the converse — that is, female sex, an age of 60 years or younger, a white-cell count of less than 100,000 per cubic millimeter, and a lactate dehydrogenase (LDH) level of 400 U per liter or less. Gray boxes — or blanks in the case of the French–American–British (FAB) subtype — indicate that no data were available. Age, white-cell count, and LDH are treated as binary variables, with the use of prognostically relevant cutoff values.[2] Samples with a normal karyotype separated into two major subgroups, as indicated. Panel B shows a three-dimensional projection of the three principal components in a principal-components analysis of all AML samples, with the use of the 6283 variably expressed genes. Only the samples with a predominantly normal karyotype in subgroups I and II are shown, defined on the basis of hierarchical clustering. Samples are color-coded as indicated. Panel C shows the Kaplan–Meier estimates of overall survival in the two subgroups of patients with a normal karyotype; the difference between groups was significant (P=0.009 by the log-rank test). The X symbols in Panel C indicate censored data.

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 duplica-

tions, and trisomy 8 (whose molecular heterogeneity has been reported[27]), though this may reflect our limited statistical power owing to the small sizes of the groups.
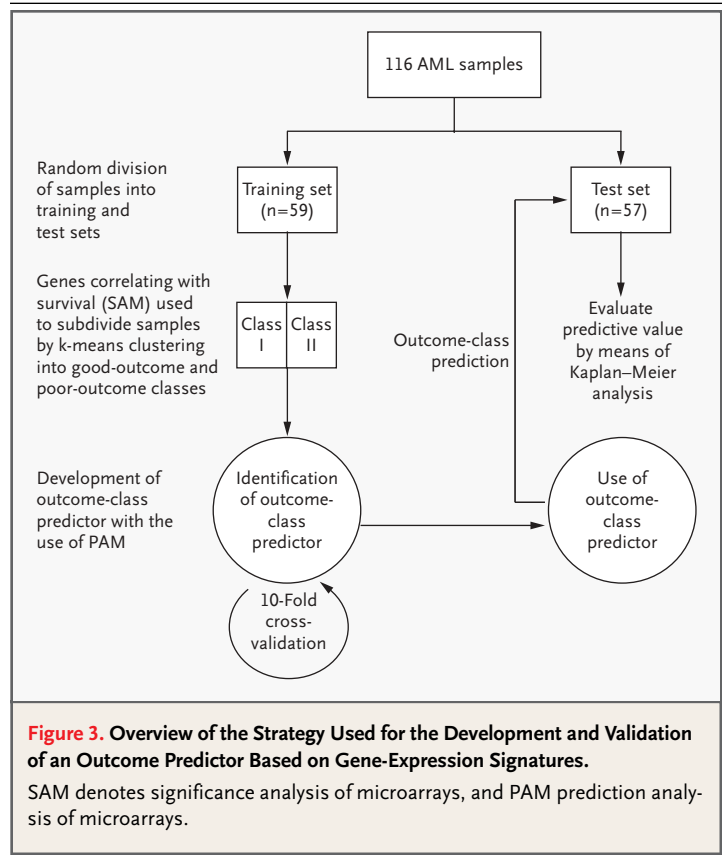
Among the group-specific signatures, we found genes located at translocation breakpoints defining cytogenetic classes, including *ETO* in t(8;21) and *MYH11* in inv(16) (described at www.ncbi.

nlm.nih.gov/geo/). We also identified numerous other group-specific named genes and expressed-sequence tags; the function of known genes suggested plausible pathogenetic roles. For example, the t(15;17) signature (partially shown in Fig. 1F) included genes associated with abnormalities in hemostasis (*PLAU*, *SERPING1*, *ANXA8*, and *PLAUR*), resistance to apoptotic stimuli (*TNFRSF4*, *AVEN*, and *BIRC5*), and impairment of retinoic acid–induced cell differentiation (*TBLX1*, *CALR*, and *RARRES3*), as well as detoxification of chemical compounds and resistance to chemotherapy (*CYP2E1*, *EPHX1*, *MT1G*, *MT1H*, *MT1L*, *MT2A*, and *MT3*).

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),[28] suggesting a possible shared mechanism contributing to leukemogenesis. In specimens with inv(16), we found high levels of expression of *NT5E* (5′ nucleotidase, also known as *5NT* or *CD73*) (Fig. 1H), which has been associated with resistance to cytarabine in AML.[29] This finding is somewhat surprising, since AML with inv(16) is clinically quite sensitive to cytarabine.[4] In contrast to childhood acute lymphoblastic leukemia,[17] 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.[30] 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[31] in the pathogenesis of this subtype.

AML specimens in group II were characterized in part by a prominent gene-expression feature (Fig. 1G) associated with granulocytic or monocytic differentiation and the immune response. A candidate pathogenetic gene within this subgroup was the gene for vascular endothelial growth factor (*VEGF*), which is involved in the regulation of hematopoietic-stem-cell survival[32] and in the progression of AML[33] (Fig. 1I).

Figure 3. Overview of the Strategy Used for the Development and Validation of an Outcome Predictor Based on Gene-Expression Signatures.

SAM denotes significance analysis of microarrays, and PAM prediction analysis of microarrays.

OUTCOME PREDICTION

Having demonstrated the presence at diagnosis of gene-expression signatures correlating with the clinical outcome (Fig. 2C), we next sought to construct a gene-expression–based outcome predictor for AML. Both supervised and unsupervised strategies have been proposed as means of identifying outcome predictors with the use of DNA-microarray data. Unsupervised cluster analysis based on genes whose expression varies (or on a subgroup of gene-expression features) has been used to define prognostically relevant tumor subtypes that might form the basis for outcome prediction.[13,34] In our AML data set, however, the clustering of samples was driven in large part by underlying cytogenetic aberrations, and thus except for the normal-karyotype subgroups, such a gene-expression–based outcome predictor would be unlikely to provide additional information independent of cytogenetic findings. Supervised analyses have also been used to identify genes whose expression correlates with the likelihood of recurrent disease or survival (as binary out-

come variables)[15,17,35,36] or the duration of survival.[37] However, the likelihood and the duration of survival are likely to be fairly crude surrogates for the underlying biologic characteristics distinguishing prognostically relevant tumor subclasses (see Supplementary Appendix 5, available with the full text of this article at www.nejm.org), and indeed this approach was not very accurate in predicting the clinical outcome in our data set (not shown).

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 subclasses. 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)[25] 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

Figure 4 (facing page). Outcome Prediction.

In Panel A, columns represent AML samples in the training set ordered according to k-means clustering (a nonhierarchical computational method of organizing clusters); rows represent the 149 predictive complementary DNAs (cDNAs), ordered according to hierarchical clustering. Mean-centered ratios of gene expression are depicted by a log-transformed (on a base 2 scale) pseudocolor scale; gray denotes poorly measured genes. Good-outcome and poor-outcome subgroups were identified by means of Kaplan–Meier analysis. In Panel B, columns represent AML samples in the test set ordered according to hierarchical clustering; rows represent the 149 predictive complementary DNAs, ordered according to hierarchical clustering. Good-outcome and poor-outcome subgroups were defined by correlating gene-expression signatures with those in the training set (see text). Vertical bar (left) indicates genes that were expressed in the good-outcome subgroup (blue) or the poor-outcome subgroup (red) in the training set. Panel C shows Kaplan–Meier survival estimates in the cluster-defined poor-outcome and good-outcome subgroups of samples; there was a significant difference between groups (P=0.006 by the log-rank test). Panel D shows the same Kaplan–Meier analysis as shown in Panel C, except the analysis is restricted to AML samples in the test set with a normal karyotype. The X symbols in Panels C and D indicate censored data.

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.[25] 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
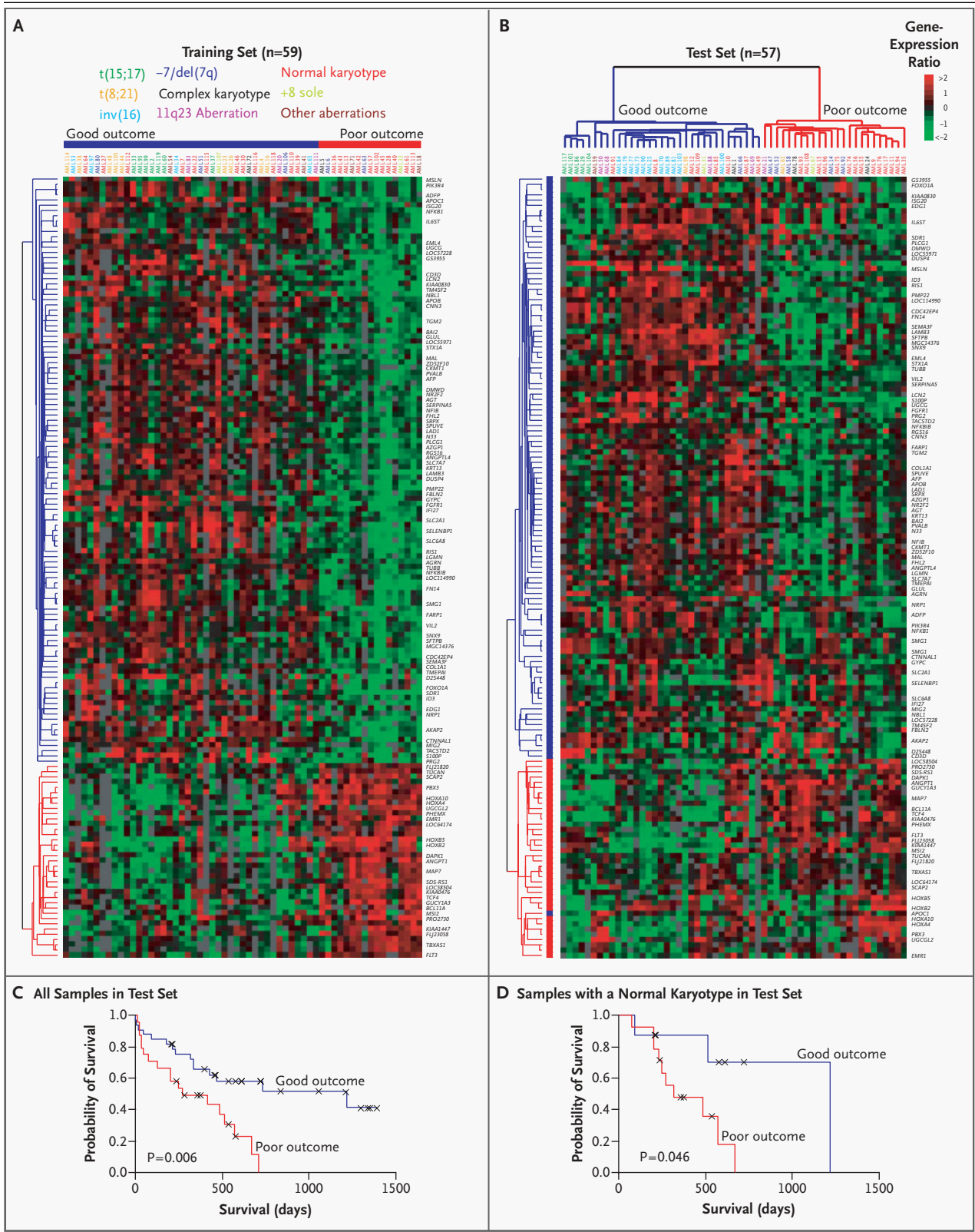
no significant difference in overall survival (P=0.65 by the log-rank test), which may reflect the relatively small sample or an inherently poorer performance of this alternative approach to outcome prediction.

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 groups (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.[38] Interestingly, other members of the forkhead family have been identified as pathogenic translocation fusion 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.[39]

Notably, among the genes associated with a poor outcome, several (e.g., *MAP7*, *GUCY1A3*, *TCF4*, and *MSI2*) (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.[40] Indeed, overexpression of *HOXA10* has been shown to perturb myeloid and lymphoid differentiation profoundly in hematopoietic cells in mice and to lead to AML.[41] Interestingly, elevated expression of *FLT3* was also associated with a poor outcome. Activating *FLT3* mutations are predictive of a poor outcome in AML,[8,9] 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,[42] 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.[43]

Our gene-expression study has provided numerous insights into the pathogenesis of AML, including, for example, the role of homeobox-gene dysregulation.[40] 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,[44] has also recently been observed in children with acute lymphoblastic leukemia with *MLL* rearrangements,[16,17] 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.

APPENDIX

The following AML Study Group Ulm institutions and investigators participated in the study: Universitätsklinikum Bonn, Bonn, Germany, A. Glasmacher; Universitätsklinikum Düsseldorf, Düsseldorf, Germany, U. Germing; Universitätsklinikum Giessen, Giessen, Germany, H. Pralle; Universitätsklinikum Göttingen, Göttingen, Germany, D. Haase; Allgemeines Krankenhaus Altona, Hamburg, Germany, H. Salwender; Universitätskliniken des Saarlandes, Homburg, Germany, F. Hartmann; Universitätsklinikum Innsbruck, Innsbruck, Austria, G. Gastl; Städtisches Klinikum, Karlsruhe, Germany, J.T. Fischer; Universitätsklinikum 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. Grimminger; Katharinenhospital, Stuttgart, Germany, H.G. Mergenthaler; Krankenhaus der Barmherzigen Brüder, Trier, Germany, W. Weber; and Hanusch-Krankenhaus, Vienna, E. Koller.

REFERENCES

1. Giles FJ, Keating A, Goldstone AH, Avivi I, Willman CL, Kantarjian HM. Acute myeloid leukemia. Hematology (Am Soc Hematol Educ Program) 2002:73-110.

2. Lowenberg B. Prognostic factors in acute myeloid leukaemia. Best Pract Res Clin Haematol 2001;14:65-75.

3. Grimwade D, Walker H, Oliver F, et al. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. Blood 1998;92:2322-33.

4. Bloomfield CD, Lawrence D, Byrd JC, et al. Frequency of prolonged remission duration after high-dose cytarabine intensification in acute myeloid leukemia varies by cytogenetic subtype. Cancer Res 1998;58:4173-9.

5. Byrd JC, Mrozek K, Dodge RK, et al. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). Blood 2002;100:4325-36.

6. Slovak ML, Kopecky KJ, Cassileth PA, et al. Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. Blood 2000;96:4075-83.

7. Mrozek K, Heinonen K, Bloomfield CD. Clinical importance of cytogenetics in acute myeloid leukaemia. Best Pract Res Clin Haematol 2001;14:19-47.

8. Frohling S, Schlenk RF, Breitruck J, et al. Prognostic significance of activating FLT3

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.

9. Kottaridis PD, Gale RE, Frew ME, et al. The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. Blood 2001;98:1752-9.

10. Caligiuri MA, Strout MP, Lawrence D, et al. Rearrangement of ALL1 (MLL) in acute myeloid leukemia with normal cytogenetics. Cancer Res 1998;58:55-9.

11. Dohner K, Tobis K, Ulrich R, et al. Prognostic significance of partial tandem duplications of the MLL gene in adult patients 16 to 60 years old with acute myeloid leukemia and normal cytogenetics: a study of the Acute Myeloid Leukemia Study Group Ulm. J Clin Oncol 2002;20:3254-61.

12. Gilliland DG, Griffin JD. The roles of FLT3 in hematopoiesis and leukemia. Blood 2002;100:1532-42.

13. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503-11.

14. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 2002; 346:1937-47.

15. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 2002;8:68-74.

16. Armstrong SA, Staunton JE, Silverman LB, et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet 2002; 30:41-7.

17. Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell 2002;1:133-43.

18. Frohling S, Skelin S, Liebisch C, et al. Comparison of cytogenetic and molecular cytogenetic detection of chromosome abnormalities in 240 consecutive adult patients with acute myeloid leukemia. J Clin Oncol 2002;20:2480-5.

19. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95:14863-8.

20. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747-52.

21. Gollub J, Ball CA, Binkley G, et al. The Stanford Microarray Database: data access and quality assessment tools. Nucleic Acids Res 2003;31:94-6.

22. Nielsen TO, West RB, Linn SC, et al. Molecular characterisation of soft tissue tumours: a gene expression study. Lancet 2002;359:1301-7.

23. Raychaudhuri S, Stuart JM, Altman RB. Principal components analysis to summarize microarray experiments: application to sporulation time series. Pac Symp Bicomput 2000;5:455-66.

24. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116-21. [Erratum, Proc Natl Acad Sci U S A 2001;98:10515.]

25. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 2002;99:6567-72.

26. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nat Genet 2003;33:49-54.

27. Virtaneva K, Wright FA, Tanner SM, et al. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc Natl Acad Sci U S A 2001;98:1124-9.

28. Tanabe S, Zeleznik-Le NJ, Kobayashi H, et al. Analysis of the t(6;11)(q27;q23) in leukemia shows a consistent breakpoint in AF6 in three patients and in the ML-2 cell line. Genes Chromosomes Cancer 1996;15:206-16.

29. Galmarini CM, Thomas X, Calvo F, et al. In vivo mechanisms of resistance to cytarabine in acute myeloid leukaemia. Br J Haematol 2002;117:860-8.

30. Kumano K, Chiba S, Shimizu K, et al. Notch1 inhibits differentiation of hematopoietic cells by sustaining GATA-2 expression. Blood 2001;98:3283-9.

31. Mizuno S-I, Chijiwa T, Okamura T, et al. Expression of DNA methyltransferases DNMT1, 3A, and 3B in normal hematopoiesis and in acute and chronic myelogenous leukemia. Blood 2001;97:1172-9.

32. Gerber HP, Malik AK, Solar GP, et al.

VEGF regulates haematopoietic stem cell survival by an internal autocrine loop mechanism. Nature 2002;417:954-8.

33. Schuch G, Machluf M, Bartsch G Jr, et al. In vivo administration of vascular endothelial growth factor (VEGF) and its antagonist, soluble neuropilin-1, predicts a role of VEGF in the progression of acute myeloid leukemia in vivo. Blood 2002;100:4622-8.

34. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869-74.

35. Pomeroy SL, Tamayo P, Gaasenbeek M, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature 2002;415:436-42.

36. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415:530-6.

37. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816-24.

38. Burgering BM, Kops GJ. Cell cycle and death control: long live Forkheads. Trends Biochem Sci 2002;27:352-60.

39. So CW, Cleary ML. Common mechanism for oncogenic activation of MLL by forkhead family proteins. Blood 2003;101: 633-9.

40. van Oostveen J, Bijl J, Raaphorst F, Walboomers J, Meijer C. The role of homeobox genes in normal hematopoiesis and hematological malignancies. Leukemia 1999;13: 1675-90.

41. Thorsteinsdottir U, Sauvageau G, Hough MR, et al. Overexpression of HOXA10 in murine hematopoietic cells perturbs both myeloid and lymphoid differentiation and leads to acute myeloid leukemia. Mol Cell Biol 1997;17:495-505.

42. Kelly LM, Gilliland DG. Genetics of myeloid leukemias. Annu Rev Genomics Hum Genet 2002;3:179-98.

43. Schoch C, Kohlmann A, Schnittger S, et al. Acute myeloid leukemias with reciprocal rearrangements can be distinguished by specific gene expression profiles. Proc Natl Acad Sci U S A 2002;99:10008-13.

44. Kroon E, Krosl J, Thorsteinsdottir U, Baban S, Buchberg AM, Sauvageau G. *Hoxa9* transforms primary bone marrow cells through specific collaboration with *Meis1a* but not *Pbx1b*. EMBO J 1998;17:3714-25.

Copyright © 2004 Massachusetts Medical Society.