Identification of Gene Expression Profiles That Segregate Patients with Childhood Leukemia

Philip J. Moos, Elizabeth A. Raetz, Marlee A. Carlson, Aniko Szabo, Fiona E. Smith, Cheryl Willman, Qi Wei, Stephen P. Hunger, and William L. Carroll

Center for Children at the Huntsman Cancer Institute [P. J. M., E. A. R., M. A. C., F. E. S., W. L. C.], Division of Pediatric Hematology/Oncology, Departments of Pediatrics [E. A. R., W. L. C.] and Oncological Sciences [P. J. M., A. S., W. L. C.], University of Utah School of Medicine, Salt Lake City, Utah 84112; University of New Mexico, Albuquerque, New Mexico 87131 [C. W.]; and University of Colorado Health Sciences Center [Q. W., S. P. H.] and the Children’s Hospital [S. P. H.], Denver, Colorado 80218

ABSTRACT

To identify genes whose expression correlated with biological features of childhood leukemia, we prospectively analyzed the expression profiles of 4608 genes using cDNA microarrays in 51 freshly processed bone marrow samples from children with acute leukemia, over a 24-month period, at a single institution. Two supervised methods of analysis were used to identify the 20 best discriminating genes between the following cohorts: acute myelogenous leukemia (AML) versus acute lymphoblastic leukemia (ALL); B-lineage versus T-lineage ALL; newly diagnosed B-lineage standard-risk versus high-risk ALL; and B-lineage leukemia harboring the TEL-AML1 fusion versus patients without a molecularly characterized translocation. These methods identified overlapping sets of genes that segregated patients within described subgroups. Cross-validation demonstrated that the majority of patients could be correctly classified based on these genes alone, and hierarchical clustering grouped patients with similar clinical and biological disease features. The potential for select genes to discriminate patients was validated using real-time PCR in samples that were analyzed by microarray profiling and in other formally processed leukemic marrow samples. As expected, microarray technology can successfully segregate patients defined by traditional measures such as immunophenotype and cytogenetic alterations. However, among specific subgroups, this preliminary analysis also suggests that microarrays can identify unanticipated similarities and diversity in individual patients and thus may be useful in augmenting risk-group stratification in the future.

INTRODUCTION

Leukemia is the most common pediatric malignancy in the United States, and ALL accounts for the vast majority (75%) of new cases diagnosed annually (1). The outcome for children with ALL has improved dramatically over the past three decades (2, 3), but despite significant progress in treatment, 25% of children with ALL and 50–60% of children with AML develop recurrent disease and their prognosis remains poor (4, 5).

Although cure rates have improved considerably, curative therapy is often associated with serious side effects. Risk-based therapy attempts to tailor treatment to maximize the chances for cure and minimize side effects (6, 7). Using this approach, patients are segregated into different risk groups based upon a combination of clinical and biological features of their disease, and subsequent therapy is tailored according to their predicted risk for treatment failure. Age, presenting WBC count, and abnormalities in either chromosome number or structure are the most important features presently used to stratify children with ALL into good-, standard-, high-, or very high-risk groups (8–12). Similar approaches have been used in the treatment of childhood AML.

Although risk-group stratification has been very useful in classifying patients and improving their outcomes, variability in clinical course exists among individuals within a single risk group and among those with the same prognostic features. For example, the majority of children who suffer a relapse while being treated on contemporary therapy were originally defined as standard risk. Also, some patients whose blasts express the TEL-AML1 fusion transcript, which has been associated with a very favorable outcome, fail treatment for reasons that are poorly understood (13–19). In fact, Donadieu et al. (20) have found that the most significant prognostic factors in childhood ALL explain no more than 4% of the variability in prognosis, suggesting that undiscovered molecular mechanisms dictate clinical behavior. Discovery of additional molecular prognosticators may therefore enhance current risk-based classification schemes.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 Supported by Translational Research Awards from the Leukemia and Lymphoma Society (to W. L. C. and S. P. H.), Director’s Challenge Grant U01 CA88361 (to W. L. C. and C. W.) from the National Cancer Institute, and NIH Mentored Clinical Scientist Development Award K08 HD01215-01 (to E. R.). Core Facility support was provided by Cancer Center Support Grant 2P30CA42014 and The Huntsman Cancer Foundation.

2 These authors contributed equally to this work.

3 To whom requests for reprints should be addressed, at Combined Division of Pediatric Oncology Mount Sinai and NYU Schools of Medicine, 1 Gustave Levy Place, New York, NY 10029. Phone: (212) 241-7022; Fax: (212) 360-6921.

4 The abbreviations used are: ALL, acute lymphoblastic leukemia; AML, acute myelogenous leukemia; RT-PCR, reverse transcription-PCR; LOOCV, leave-one-out cross-validation; FAB, French-American-British.
PATIENTS AND METHODS

Patient mRNA Preparation. cDNA microarray technology was used to prospectively determine patterns of gene expression in leukemia samples from 51 children with leukemia collected at the time of initial diagnosis or subsequent relapse through an Institutional Review Board approved protocol from 11/1998 to 11/2000 (Table 1). This cohort represented ~50% of all patients with newly diagnosed leukemia seen at Primary Children’s Medical Center during this time period. Heparinized bone marrow aspirates or, in some cases, peripheral blood samples were obtained from patients undergoing an initial evaluation for leukemia or suspected relapse. Patients were enrolled into the study if additional bone marrow could be aspirated at the time of diagnosis or if they had >50% circulating peripheral blasts. Analyses were performed on 44 marrow samples and 8 peripheral blood samples. All samples were processed into RNA within 1 h of acquisition to avoid artifacts associated with processing and/or storage. Enrichment for leukemic blasts was accomplished using Ficoll-Paque centrifugation. Mononuclear cells were solubilized with Trizol for total RNA purification, and mRNA was purified from DNase-treated total RNA samples using Oligotex oligo-dT beads (Qiagen, Chatsworth, CA). The arbitrary reference control sample used in all comparisons was the HL-60 cell line.

Table 1  Patient characteristics

<table>
<thead>
<tr>
<th>No. of patients</th>
<th>Morphology</th>
<th>Disease status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lymphoid</td>
<td>New diagnosis-SR</td>
</tr>
<tr>
<td></td>
<td>Precursor B-cell</td>
<td>New diagnosis-HR</td>
</tr>
<tr>
<td></td>
<td>T-cell</td>
<td>Infant</td>
</tr>
<tr>
<td></td>
<td>Myeloid</td>
<td>Relapse</td>
</tr>
<tr>
<td></td>
<td>FAB M0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAB M1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAB M2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAB M4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FAB M5b</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Undifferentiated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TMD*</td>
<td></td>
</tr>
</tbody>
</table>

* TMD, transient myeloproliferative disorder; HR, high risk; SR, standard risk; ND, not done.

Molecular Characterization of Leukemia Specimens. Single-round or nested RT-PCR assays were performed to detect fusion transcripts produced by specific nonrandom chromosomal rearrangements in 24 of 25 patients with newly diagnosed B-lineage ALL. RT-PCR assays were performed using methods described previously (25–29). Additionally, the BCR-ABL minor (m) and major (M) transcripts were detected using the following primers: M-BCR-ABL first-round sense, 5'-GGT TGG GGT TCA TTT TCA CTG; M-BCR-ABL antisense, 5'-GGT TGG GGT TCA TTT TCA CTG; M-BCR-ABL second-round sense, 5'-GGT TGG GGT TCA TTT TCA CTG; M-BCR-ABL antisense, 5'-AGA TGC TGA CCA ACT GTG GTG; mBCR-ABL first-round sense, 5'-GAA CTC GCA ACA GTG TTC ACC TTC TAA TAA TGG GGT GGT TCA TTT TCA CTG; m-BCR-ABL antisense, 5'-GTT TGG GGT TCA TTT TCA CTG; m-BCR-ABL second-round sense, 5'-CAC GCC GCAGT GCA CCA; and m-BCR-ABL antisense, 5'-GTT TGG GGT TCA TTT TCA CTG. ALL samples were tested for the presence of BCR-ABL, E2A-PBX1, MLL-AF4, and TEL-AML1 fusion transcripts, whereas AML specimens were tested for AML1-ETO, CBFB-SMMHC, and PML-RARA fusion transcripts. The single case of acute undifferentiated leukemia was
tested with both the ALL and AML panels. Additional molecular studies were performed in 3 infants. MLL gene rearrangements were detected by Southern blot analysis as part of the Children’s Cancer Group protocol, CCG-1953, for infant leukemia. RT-PCR for the MLL/AF4, MLL/AF9, and MLL/ENL gene products was also performed in these patient samples.

**Gene Expression Profiling.** cDNA hybridization probes were prepared by reverse transcription of poly(A) RNA with end-labeling with dCTP conjugated to fluorescent dyes. The hybridization probes consisted of reference cell (HL-60) cDNA labeled with Cy5-dCTP and patient cell cDNA labeled with Cy3-dCTP. The Cy3- and Cy5-cDNA hybridization probes were diluted in buffer containing 10% (w/v) dextran sulfate, 50% (v/v) formamide, 0.3 mM NaCl, 10 mM Tris (pH 8.0), 1 mM EDTA, 1 X Denhardt’s, 10 mM DTT, and 0.5 mg/ml nonhomologous DNA. Probes generated from patient marrow samples and from the reference cell line were hybridized simultaneously to a glass slide containing 4608 separate cDNAs deposited in duplicate using a Molecular Dynamics Gen III robotic spotter. Eight hundred of the cDNA clones were selected by investigators and included primarily genes involved in transcription, apoptosis, and development. The remaining cDNAs were randomly selected from IMAGE consortium clones. All differentially expressed cDNAs identified in our comparative analyses were independently sequence verified. After hybridization at 42°C for 36 h in a humidified chamber, nonbound labeled cDNA hybridization probes were removed by washing with 2X SSC, 0.1% SDS at room temperature, followed by washing with 0.5X SSC, 0.1% SDS at 45°C. Fluorescence from dried microarray slides was detected with a Molecular Dynamics Array Scanner at two respective emission wavelengths.

**Microarray Analysis.** Gene expression for each channel was quantified by ArrayVision software. Therefore, a single channel (patient) analysis was performed as follows. Within each array, the minimum value was subtracted from each element to adjust for hybridization, Cy-dye incorporation, and background, and all arrays were grid normalized to account for spotting differences for each pin (30). The average of the duplicate spotting for each array element was then log2 transformed before analysis to stabilize variance and improve normality. Cohorts were compared using a t test (two tailed and homoscedastic) and the entropy-dependent statistic Infoscore (31). The 20 genes that best discriminated the cohorts by each method were selected for further analysis. We evaluated the gene selection procedure through a cross-validation strategy. LOOCV was used to assess the quality of the genes for discriminating leukemia cohorts by measuring the Euclidean distance metric with a k-nearest neighbor (k-NN) approach (k = 9; Ref. 32). In other words, the Euclidean distance between each individual patient sample was calculated to all of the other samples using these genes. In most comparisons, the number of samples between the two cohorts was not equal, and there could be a disproportionate number of votes for the larger cohort. Still, we used the k-NN method because it was the least parametric and required the fewest number of assumptions with the following modification. With unequal sample sizes, m < n, to be assigned to the larger cohort (n), the proportion of k neighbors should be at least: \( P_a = n/(m + n) \), i.e., more than expected by random chance. In other words, the proportion of votes required to be assigned to the larger cohort has to be greater than the fraction of samples of the larger cohort relative to all samples being compared. We estimated the random prediction rate by randomly permuting the samples labels (five times), used the same algorithm, and averaged the results to determine a cross-validation prediction rate for cohorts without differences. Hierarchical clusters were generated using Spotfire to replace the relative expression values with trimmed mean, Z-scores to accentuate the patterns of expression, and the Stanford University Microarray package that groups samples and target genes based upon similarity of gene expression (33).

**Verification of Gene Expression Using Real-Time PCR.** The results from array analysis were verified using quantitative PCR for a limited number of target genes. Real-time PCR was performed using the following sense and antisense primer pairs: 5’-ATT TCG ACC GCG ATG ATG TGG and 5’-GAA CCC AGG GCA TGA AGA TTC for ferritin light chain (Hs.111334, M11147); 5’-CTT CCA GGT GAT TCC AAC AAG and 5’-AAT CCA TGT TGA CCC AAG TCC for proteoglycan (Hs.1908, W25334); 5’-CAA CAT GAC GAT CTG GAC CAC and 5’-CTT TAG TAC CAG GGT CAT GGC for endoglin (Hs.76753, W24164); 5’-AAT CCT CCT TCT TGA GGG GTC and 5’-GGA ACA AAT GAG TGC CCA TTA for MHC (MHC) class I (Hs.1845, AA149083); 5’-AGG CCC GCA ACA AGT TCG AG and 5’-TGG CAC TGA ACT TGT CGG for protein kinase C substrate 80-K-H (Hs.1432, NM_002743); and 5’-CTT CAG AGA ATC ACA CCT TCT C and 5’-ACC CTC TCC TGA CCC TGA AG for glyoxylate reductase/hydroxy- pyruvate reductase (Hs.155742, W39164).

First-strand cDNA was generated in a 40-μl reverse transcription reaction and was purified using a QiAquick PCR purification kit (Qiagen, Chatsworth, CA). Real-time PCR was performed in a fluorescence temperature cycler (LightCycler LC24; Idaho Technology, Idaho Falls, ID; Refs. 34, 35). Individual 10-μl reactions contained 3 mM MgCl2, 0.25 mg/ml BSA, 50 mM Tris (pH 8.3), 200 μM deoxyribonucleotide triphosphates, 5 μM each primer, 1:30,000 dilution in TE (10 mM Tris 1 mM EDTA, pH 8) of SYBR Green I (Molecular Probes), and 1 μl of a 1:1:8 mix of Taq DNA Polymerase, TaqStart Antibody (Clontech Laboratories, Palo Alto, CA), and Taq Enzyme Diluent (Idaho Technology) and template. Each reaction contained 50 ng of cDNA purified from individual patient RT reactions as template. After an initial 95°C denaturation, the amplification program consisted of 40 four-step cycles of heating at 20°C/s to 95°C with a 0-s hold, cooling at 20°C/s to 55°C with a 0-s hold, heating at 10°C/s to 72°C with a 15-s hold, and heating at 20°C/s to 85°C with a 0-s hold. Fluorescent product detection was undertaken at the last step of each cycle. After amplification, a melting profile was acquired by first cooling at 20°C/s to 95°C with fluorescence collection at 0.2°C intervals.

Standards for determining PCR product copy number were prepared from diluted PCR products. PCR products were purified using a QiAquick PCR purification kit (Qiagen). The copy number (copies/μl) for each standard was calculated based on the size of the PCR product and absorbance at 260 nm. Tran-
script copy numbers for patient samples were calculated by comparing the fluorescence of PCR products of unknown concentration with the fluorescence of serial dilutions of the external standard. Only fluorescence values measured in the log-linear phase of amplification were used for sample quantification. A "best-fit" log-linear fluorescence curve was calculated for each sample, and these curves in turn were used to estimate the number of transcripts in individual patient samples.

All samples were run in triplicate and analyzed as follows. The real-time PCR data were log-transformed to stabilize variance and improve normality. The statistical assessment to evaluate the real-time PCR was a repeated measures analysis that was performed separately for the samples on and off the slides (SAS). The effect size estimates and confidence intervals were transferred back to the original scale through exponentiation; therefore, they represent multiplicative effects.

RESULTS

Bone Marrow versus Peripheral Blood. Eight peripheral blood samples and 43 patient marrow samples were used for this analysis. To determine the similarity between blood and marrow samples, gene expression analyses were compared in paired samples in several patients (nos. 43, 44, and 49). In all cases, peripheral blood samples clustered next to the appropriate marrow sample when arranging the samples by gene expression correlation coefficients, illustrating the homogeneity of blasts in the peripheral blood and the marrow. Fig. 1 shows the Euclidean distance matrix for these samples.

Microarray Analysis of Established Patient Subgroups. The ability of two approaches, t-tests and Infoscore, were tested to identify individual genes that discriminate patients with AML and ALL. This initial analysis included 32 patients with B-lineage ALL, 7 patients with T-lineage ALL, and 10 patients with AML. A LOOCV approach was used to determine the error rate of selecting the top 20 discriminating genes between ALL and AML cohorts. These methods could predict the clinical classification at overall rates of 87.7 and 75%, respectively (Table 2). ALL patients were predicted correctly nearly 95% of the time when using t-tests but only 77% for Infoscore, but AML patients were misclassified more frequently with t-tests than Infoscore. Still, the same 19 genes were found in 50% of the comparisons. Perhaps the reduced prediction rate for AML is partially attributable to the heterogeneous FAB classifications of AML patients (Table 1).

To determine the usefulness of the top discriminating genes when using all AML and ALL patients in this analysis for predicting cohort status, the 20 best discriminating genes for each methodology were identified with high significance ranging from $P = 0.000003$ to 0.0005 (Fig. 2). These 20 discriminating genes were used as the inputs for hierarchical clustering to assess the reliability of this method as a classification tool for known subtypes of leukemia (Fig. 2). After the selection of the top 20 discriminators, 1 patient sample each with acute undifferentiated leukemia and transient myeloproliferative disorder were added to the AML versus ALL comparison. In determining gene profiles that distinguished lymphoid from myeloid disease, the patients
Fig. 2  A, hierarchical cluster analysis of the 20 best array elements discriminating AML and ALL as determined by \(t\) test. B, hierarchical cluster analysis of the 20 best array elements discriminating AML and ALL as determined by Infoscore. Genes found in common among both methods are represented in bold. C, real-time PCR validation of select genes in common between A and B. The red bar represents the geometric mean of the number of transcripts. AUL, acute undifferentiated leukemia; PRG, proteoglycan (W25334); Ferritin L chain, M11147.
clustered on more explicit branches when using the best 20 discriminators by \( t \) test. All AML patients segregated on one major branch. When clustering using the best 20 genes by Infoscore, most of the AML patients segregated in one branch. Eleven genes were commonly identified by both methods (Fig. 2).

The microarray data were validated independently by real-time PCR using the actual patient samples from the arrays, as well as other genetically processed, leukemic marrow samples not included in this initial array analysis. The target genes selected for validation were those that were identified as discriminators by \( t \) test and Infoscore, independently, e.g., hematopoietic proteoglycan core protein (Hs. 1908) and the ferritin light chain (Hs.111335; Fig. 2C). In the case of hematopoietic proteoglycan core protein, expression profiles were validated in 15 AML patient samples, with 9 from the arrays and 6 marrow samples not on the arrays. Expression profiles were similarly validated in 29 ALL patient samples, with 17 from the arrays and 12 marrow samples not on the arrays (Fig. 2C). Real-time PCR indicated that patient samples that were analyzed previously with the microarray affirmed distinct expression, and an independent test confirmed that this gene was differentially expressed (\( P = 0.006 \)) in additional samples. In the case of ferritin light chain, expression profiles were validated in 13 AML patient samples, with 9 from the arrays and 4 banked samples not on the arrays. Expression profiles were similarly validated in 28 ALL patient samples, with 16 from the arrays and 12 banked samples (Fig. 2C). Once again, real-time PCR demonstrated that patient samples that were analyzed previously with the microarray were differentially expressed, and an independent test confirmed that this gene was expressed more highly in AML than ALL marrow samples (\( P = 0.033 \)).

Within the ALL cohort, the genes that best discriminated T-lineage ALL versus B-lineage ALL were identified in a similar manner. This initial analysis included the 32 patients with B-lineage ALL and 7 patients with T-lineage ALL. Both methods were tested using LOOCV and predicted the correct classification at a rate of 89.7% and 89.5%, respectively (Table 3). The \( t \) test predicted T-lineage correctly 100% of the time, whereas Infoscore correctly predicted T-lineage at a rate of 71.4%. It is noteworthy that this methoderrred in predicting 2 T-cell ALL patients who had very immature immunophenotypes. The 20 best discriminating genes for each methodology were identified with high significance ranging from \( P = 0.000000002 \) to 0.0006. Hierarchical clustering of the patients demonstrated clear distinctions of T-lineage patients from the B-lineage patients using either methodology (Fig. 3). The majority (16 unique genes) of the top 20 discriminating genes were identified by both \( t \) test and Infoscore. The genes identified included those predicted to show differential expression, such as the T-cell specific surface marker, CD3. Additionally, as expected, MHC molecules displayed increased expression in B-lineage blasts. We again selected a subset of genes that were independently identified by both methods for validation. In the case of MHC class I, expression profiles were validated in 7 T-cell patient samples (4 from the arrays; 3 banked) and in 23 B-lineage samples (13 from the arrays; 10 banked). Real-time PCR affirmed that patient samples among the two groups that were analyzed with the microarrays were different, and additional testing in off array samples confirmed this differential gene expression (\( P = 0.023 \)) in an independent sample set. In the case of endoglin, expression profiles were validated in 6 T-cell patient samples (4 from the arrays; 2 banked) and in 18 B-lineage samples (12 from the arrays; 6 banked). Again, real-time PCR demonstrated that patient samples that were analyzed previously with the microarray were distinct; however, an independent test did not confirm that this gene was differentially expressed (\( P = 0.28 \)). However, the test set only contained two T-cell samples resulting in a very wide confidence interval.

### Patterns of Gene Expression in Leukemic Patients in Different Risk Strata.

The transcript profiles of 13 standard-risk (by National Cancer Institute criteria) and 10 high-risk newly diagnosed B-lineage leukemic patients were also analyzed. Here, \( t \) test and Infoscore were again used to independently identify groups of genes that discriminated patients based on their predicted risk of relapse. The error rates, based on cross-validation, were higher for these cohorts, resulting in lower overall prediction rates of 61 and 65%, respectively (Table 4). Here the high-risk group was poorly predicted by both methods, at only 50%. Interestingly, the same high-risk patient samples were misclassified as standard risk by both methods. The same was found for those standard risk that were misclassified as high risk. This may indicate that the clinical criteria for segregating these cohorts are less well defined, as determined by this molecular analysis. A larger cohort comparison is necessary to determine whether this is the primary reason for the lower prediction rates. The 20 best discriminating genes for each methodology were identified with high significance ranging from \( P = 0.00002 \) to 0.0003. Cross-validation indicated good discrimination of patients based on clinical risk of relapse. The patients that were misclassified by the clinical criteria and the prediction strength are listed in Table 4. However, only 5 genes were commonly identified using a combination of \( t \) test and Infoscore (Fig. 4). The majority of the gene sequences identified were not annotated. Several patient groups emerged after hierarchical clustering, some showing an admixing of standard-risk and high-risk patients (Fig. 4). We again selected genes for validation in patient samples that were independently identified by both methods (Fig. 4C). The expression of protein kinase C substrate (Hs.1432) was
Fig. 3 A, hierarchical cluster analysis of the 20 best array elements discriminating T-lineage and B-lineage ALL as determined by \( t \) test. B, hierarchical cluster analysis of the 20 best array elements discriminating T-lineage and B-lineage ALL as determined by Infoscore. C, real-time PCR validation of select genes in common between A and B. The red bar represents the geometric mean of the number of transcripts. MHC, AA149083; Endoglin, W24164.
validated in 22 standard-risk patient samples (9 from the arrays; 13 banked) and in 21 high-risk samples (10 from the arrays; 11 banked). Real-time PCR affirmed that patient samples that were analyzed previously with the microarray were distinct; however, an independent test did not confirm that this gene was differentially expressed \((P = 0.86)\).

**Gene Expression Analyses of Patients with Defined Cytogenetic Abnormalities.** The \(t(12;21)\) translocation is a defined cytogenetic abnormality that generally has been associated with a favorable outcome. We compared 8 patient samples harboring the TEL-AML1 fusion transcript by PCR to 11 samples that were PCR negative using both methods. Cross-validation analysis indicated that both methods could predict the correct cohort at rates of 73.7 and 90\%, respectively (Table 5). The 20 best discriminating genes for each methodology were identified with high significance ranging from \(P = 0.00001\) to 0.0002. Eleven of the genes were common to both methodologies, and most of these are not annotated (Fig. 5).

Hierarchical clustering indicated clear classifications of the patients with and without the translocation. The only exception was #102, a sample that also failed cross-validation by both methodologies (Fig. 5). Otherwise, all patients with the TEL-AML1 translocation clustered together in one branch. The expression of glyoxylate reductase/hydroxypropionate reductase (Hs. 155742) was validated in 4 Tel-AML1-positive patient samples and in 7 patient samples where no translocation was identified by molecular means (Fig. 5C).

Here, real-time PCR did not affirm that patient samples that were analyzed previously with the microarray were differentially expressed. This could indicate that there could be some other property of these samples such that they segregated when using multiple genes in the clustering analysis, or that the primer pairs used in real-time PCR did not have sufficient fidelity to faithfully produce the product in question, or that the sample size was too small to differentiate these cohorts. It is noteworthy that a study published recently, using a large number of patient samples, was capable of distinguishing all of the common chromosomal translocations observed in pediatric ALL (36).

**DISCUSSION**

Traditionally, childhood leukemia has been classified according to a host of clinical parameters and biological features that correlate with prognosis (37, 38). This classification system has been used for risk-based treatment assignment. Universally accepted criteria for leukemia classification are age and initial WBC count at the time of diagnosis (8). Other factors that are also used to determine risk of relapse are the rapidity of response to therapy (39, 40) and cytogenetic abnormalities that include both alterations in chromosome number (ploidy; Refs. 41–43) and structure (translocations; Ref. 10). Although the prognostic importance of these features has been well established, unexplained variability in clinical course still exists among some individuals within defined risk-group strata. Differences in the molecular constitution of malignant cells within subgroups may help to explain this variability.

Because the biological behavior of a cancer cell is governed, in part, by its expressed genetic repertoire, the development of techniques to determine relative levels of RNA expression offer powerful insight into the molecular mechanisms of disease. Microarray methodology is a particularly useful tool for these analyses and has been used successfully to study many hematopoietic and solid tumors. Golub et al. (44) have shown that array technology can reliably classify leukemias and can specifically distinguish acute myeloid from lymphoid leukemias and B- and T-lineage ALL. More recently, a large study was conducted that could distinguish T-lineage ALL and all of the common B-lineage chromosomal translocations with specific transcript profiles (36). Molecular classification schemes have also been applied to human B-cell lymphomas and B-lineage ALL and have uncovered previously unrecognized genetic heterogeneity among patient subgroups, some of which correlate with prognosis (36, 45). Variation in gene expression has also been characterized recently in human breast cancers (46) and correlated with the risk of metastasis in melanomas (47).

In our initial profiling of 51 pediatric bone marrow samples, we have used microarray technology to both segregate patient subgroups and to uncover genetic diversity among patients that fall within the same traditional risk groups. We used two complementary methods of supervised classification for our analyses that differ in their assumptions of the underlying data structure. \(t\) tests assume normally distributed data, and significance is defined by how distinct the means of the cohorts are, given the variability of the data. In contrast, Infoscore does not assume anything about the data distribution, but expression data with little dynamic range could result in the identification of genes whose expression level between cohorts may not be biologically significant. Significant overlap in the top 20 genes identified by these two methods was seen. Still, we hypothesize that the best discriminators are those differentially expressed using a combination of the two methods. To test this hypothesis, the expression of potential gene discriminators was validated by real-time PCR in both the samples on the arrays and in other uniformly processed leukemic marrow samples not included on the arrays.

Independent validation of the array data were seen in the AML versus ALL and T-cell versus B-lineage cohorts. Validation of the expression of protein kinase C substrate,
Fig. 4  
Table 5  TEL-AML1 versus none LOOCV: LOOCV of gene selection method (Pc = 0.579)

<table>
<thead>
<tr>
<th></th>
<th>Predicted TEL-AML1 % correct</th>
<th>Predicted none % correct</th>
<th>Overall % predicted correctly</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-test</td>
<td>Observed</td>
<td>62.5</td>
<td>81.8</td>
</tr>
<tr>
<td></td>
<td>Randomly permuted</td>
<td>37</td>
<td>53</td>
</tr>
<tr>
<td>Infoscore</td>
<td>Observed</td>
<td>88</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>Randomly permuted</td>
<td>48</td>
<td>44</td>
</tr>
</tbody>
</table>

*Observed, prediction rates according to clinical pathologic diagnoses; Randomly permuted, average after randomizing data labels.

however, was not seen in the standard-risk and high-risk ALL samples that were not on the arrays. This was not unexpected because the clinical distinctions among patients within different risk strata of a uniform disease type may be less than the differences among entirely different types of leukemia or different lineages. Also, a smaller sample number limited the standard-risk/high-risk comparison as well as the expected regression to the mean effect. A group of genes, rather than an individual gene, may more accurately predict risk-group assignment. In the future, we plan to use the same methods, in larger numbers of banked samples, to identify a set of genes that reliably differentiate standard-risk from high-risk ALL. Once a discriminating gene set for standard-risk and high-risk ALL is validated, patients could potentially undergo testing for gene expression at the time of initial diagnosis. This could be done efficiently using techniques such as quantitative PCR with the ultimate objective of using this information prospectively to augment risk classification schemes.

A unique aspect of our study was the exclusive use of freshly harvested bone marrow and uniform RNA processing within 1 h of sample acquisition. It is currently unknown whether differences in expression profiles are seen as a result of differences in sample banking or processing. A limitation to this method was small patient sample number, because we analyzed only samples from our own institution and, furthermore, only those in which enough additional bone marrow could be obtained at the time of initial diagnosis to perform an array without amplification. The overall goal of this initial study, however, was to pilot a method of analysis that could be used in larger sample numbers in the future.

As reported by Golub et al. (44), we found that no single transcript could unilaterally distinguish subgroups, even between seemingly disparate subgroups such as ALL versus AML. However, we did identify and validate certain genes that did distinguish between ALL versus ALL and T-lineage versus B-lineage. Furthermore, several predicted patterns of differential gene expression were observed. Class II HLA molecules are located predominantly on B cells and macrophages, playing a major role in antigen presentation. Consistent with known endogenous patterns of expression, MHC molecules were consistently overexpressed in B-lineage ALL when compared with T-ALL. Of the top 20 genes discriminating T-lineage versus B-lineage ALL, 5 were class II MHC molecules using both methods of analysis. Similarly, the delta subunit of the T-cell antigen receptor, CD3, was highly expressed in T-ALL samples. Although the majority of the 50 genes reported to differentiate AML versus ALL by oligonucleotide array in an analysis by Golub et al. were not on our cDNA arrays, we observed some similar patterns of gene expression in our comparison of ALL versus AML samples, i.e., SNF2 was more highly expressed in ALL samples, and proteoglycan and cathepsin D were more highly expressed in AML patient samples. Differentially regulated genes from broad functional categories were also identified. For example, cell surface proteins (CD14 antigen), genes involved in cell cycle control (S-100 proteins, rhoG), transcription factors (SRY-box 4), oncogenes (v-jgr), cell adhesion molecules (integrin β2), genes involved in metabolism and protein degradation (phosphodiesterase, cathepsin D), and genes involved in chromatin remodeling (SNF2b) were among those that were differentially expressed within the cohorts of leukemia patients.

Although limited by the small sample number for certain disease subtypes, interesting clinical correlates were suggested by this preliminary analysis. Some admixing of the gene expression profiles of newly diagnosed B-lineage ALL standard-risk and high-risk patients was seen. Of note, the three standard-risk patients that were misclassified as high-risk by both methods all showed a slow initial marrow response with day 7 marrow blast percentages ranging from 34 to 92%. Indeed, one patient has died from an early marrow relapse. In contrast, the four high-risk patients misclassified as standard risk all demonstrated rapid early marrow responses, with day 7 marrow blast percentages of <25%. Although the prognostic significance of slow early marrow response remains to be seen in standard-risk patients, it has been associated with inferior outcomes in patients with higher risk disease, unless therapy is intensified (48, 49). The admixing of standard-risk and high-risk patient expression profiles was not unexpected because approximately one quarter of standard-risk patients develop disease recurrence, and conversely, a significant subset of high-risk patients respond very favorably to less intensive therapy. Upon further validation of the discriminating gene set in a larger sample size, future goals include prospective testing of patient marrow samples at the time of diagnosis.

In summary, we have demonstrated that microarray technology can be successfully applied to the study of childhood leukemia, broadly segregating patients defined by traditional measures such as clinical features, immunophenotype, and cytogenetic alterations. This method can also uncover previously unrecognized diversity among individual patients in disease subgroups. This additional diversity may explain differences in clinical behavior and may help to further refine risk-group allocation. Future direct clinical applications of DNA array technology include the prospective identification of specific patterns of gene expression that predict clinical outcome more precisely. Finally, these studies of childhood leukemia may provide insight into relevant pathways for the development of novel therapeutics.
Fig. 5  A, hierarchical cluster analysis of the 20 best array elements discriminating patients with TEL-AML1 translocation from patients without a molecularly classified translocation as determined by t test. B, hierarchical cluster analysis of the 20 best array elements discriminating patients with TEL-AML1 translocation from patients without a molecularly classified translocation as determined by Infoscore. C, real-time PCR analysis of select genes in common between A and B. The red bar represents the geometric mean of the number of transcripts. Glyoxylate reductase, glyoxylate reductase/hydroxypropionate reductase (W39164).
ACKNOWLEDGMENTS

We acknowledge all of the patients and their families who participated in this study. In addition, we thank members of the Division of Pediatric Hematology-Oncology at Primary Children’s Medical Center for their help in obtaining marrow samples. We would also like to acknowledge the Microarray Core Facility and the Biostatistics Resource at the Huntsman Cancer Institute and Drs. Stephen Prescott, Joseph Yost, Frank Fitzpatrick, and David Jones for their helpful comments and critical review of the manuscript. We would also like to thank Drs. Bruce Camitta and Paul Gaynon of the Children’s Oncology Group ALL Strategy Committee for their encouragement and support of these studies.

REFERENCES

Identification of Gene Expression Profiles That Segregate Patients with Childhood Leukemia

Philip J. Moos, Elizabeth A. Raetz, Marlee A. Carlson, et al.


Updated version Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/8/10/3118

Cited articles This article cites 48 articles, 20 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/8/10/3118.full.html#ref-list-1

Citing articles This article has been cited by 14 HighWire-hosted articles. Access the articles at:
/content/8/10/3118.full.html#related-urls

E-mail alerts Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.