

Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling

Eng-Juh Yeoh,^{1,7,11} Mary E. Ross,^{2,11} Sheila A. Shurtleff,¹ W. Kent Williams,¹ Divyen Patel,⁶ Rami Mahfouz,¹ Fred G. Behm,¹ Susana C. Raimondi,¹ Mary V. Relling,³ Anami Patel,¹ Cheng Cheng,⁴ Dario Campana,^{1,2} Dawn Wilkins,⁸ Xiaodong Zhou,⁸ Jinyan Li,⁹ Huiqing Liu,⁹ Ching-Hon Pui,² William E. Evans,³ Clayton Naeve,⁶ Limsoon Wong,⁹ and James R. Downing^{1,5,10}

¹Department of Pathology

²Department of Hematology-Oncology

³Department of Pharmaceutical Sciences

⁴Department of Biostatistics

⁵Department of Tumor Cell Biology

⁶Hartwell Center for Bioinformatics and Biotechnology

St. Jude Children's Research Hospital, Memphis, Tennessee 38105

⁷The Department of Pediatrics, National University of Singapore, National University Hospital, 5 Lower Kent Ridge Road, Singapore 119074

⁸The Department of Computer and Information Sciences, University of Mississippi, Oxford, Mississippi

⁹Laboratories for Information Technologies, 21 Heng Mui Keng Terrace, Singapore 119613

¹⁰Correspondence: jim.downing@stjude.org

¹¹These authors contributed equally to this work.

Summary

Treatment of pediatric acute lymphoblastic leukemia (ALL) is based on the concept of tailoring the intensity of therapy to a patient's risk of relapse. To determine whether gene expression profiling could enhance risk assignment, we used oligonucleotide microarrays to analyze the pattern of genes expressed in leukemic blasts from 360 pediatric ALL patients. Distinct expression profiles identified each of the prognostically important leukemia subtypes, including T-ALL, *E2A-PBX1*, *BCR-ABL*, *TEL-AML1*, *MLL* rearrangement, and hyperdiploid >50 chromosomes. In addition, another ALL subgroup was identified based on its unique expression profile. Examination of the genes comprising the expression signatures provided important insights into the biology of these leukemia subgroups. Further, within some genetic subgroups, expression profiles identified those patients that would eventually fail therapy. Thus, the single platform of expression profiling should enhance the accurate risk stratification of pediatric ALL patients.

Introduction

Pediatric acute lymphoblastic leukemia is one of the great success stories of modern cancer therapy, with contemporary treatment protocols achieving overall long-term event-free survival rates approaching 80% (Schrappe et al., 2000; Silverman et al., 2001; Pui and Evans, 1998). This success has been achieved, in part, by using risk-adapted therapy that involves tailoring the intensity of treatment to each patient's risk of relapse. This

approach was developed following the realization that pediatric ALL is a heterogeneous disease consisting of various leukemia subtypes that differ markedly in their response to chemotherapy (Pui and Evans, 1998). By tailoring the intensity of treatment to a patient's relative risk of relapse, patients are neither undertreated or overtreated and are thus afforded the highest chance for a cure.

Critical to the success of this approach has been the accurate assignment of individual patients to specific risk groups.

SIGNIFICANCE

Acute lymphoblastic leukemia is a heterogeneous disease, with individual leukemia subtypes differing in their response to chemotherapy. Identifying prognostically important leukemia subtypes is an imprecise process and is labor intensive, requiring the combined expertise of hematologist/oncologist, pathologist, and cytogeneticist. Here we report results of expression profiling of leukemic blasts from a large cohort of pediatric ALL patients. Our results demonstrate that expression profiling can not only accurately identify the known prognostically important leukemia subtypes, but can further enhance our ability to assess a patient's risk of failing therapy. In addition, the identified expression profiles were found to include new diagnostic and subclassification markers, as well as candidates against which novel therapeutics may be developed. Lastly, the analysis resulted in the identification of a new leukemia subtype. These data suggest that in the near future, expression profiling will become an important diagnostic tool for the evaluation of pediatric ALL patients.

Although risk assignment is influenced by a variety of clinical and laboratory parameters, the genetic alterations that underlie the pathogenesis of individual leukemia subtypes figure prominently in most classification schemes (Silverman et al., 2001; Pui and Evans, 1998). Through systematic immunophenotyping and cytogenetic analysis and the subsequent molecular cloning of the genes targeted by the identified chromosomal rearrangements, a number of genetically distinct leukemia subtypes have been defined. These include B lineage leukemias that contain t(9;22)[*BCR-ABL*], t(1;19)[*E2A-PBX1*], t(12;21)[*TEL-AML1*], rearrangements in the *MLL* gene on chromosome 11, band q23, or a hyperdiploid karyotype (i.e., >50 chromosomes), and T lineage leukemias (T-ALL; Silverman et al., 2001; Pui and Evans, 1998). The underlying genetic lesions in these leukemia subtypes influence the response to cytotoxic drugs. For example, leukemias that express the E2A-PBX1 fusion protein respond poorly to conventional antimetabolite-based treatment but have cure rates approaching 80% when treated with more intensive therapies (Raimondi et al., 1990; Hunger, 1996). Similarly, *BCR-ABL*-expressing ALLs or infants with *MLL* rearrangements have exceedingly poor cure rates with conventional chemotherapy, and allogeneic hematopoietic stem cell transplantation with an HLA-matched sibling donor has recently been shown to improve outcome for patients with the former leukemia subtype (Pui et al., 1991; Heerema et al., 1999; Arico et al., 2000; Biondi et al., 2000).

Unfortunately, the accurate assignment of patients to specific risk groups is a difficult and expensive process, requiring intensive laboratory studies including immunophenotyping, cytogenetics, and molecular diagnostics (Pui and Evans, 1998; Pui et al., 2001). Moreover, these diagnostic approaches require the collective expertise of a number of professionals, and although this expertise is available at most major medical centers, it is generally unavailable in developing countries. With the recent development of DNA microarrays, it is now possible to take a genome-wide approach to leukemia classification (Perou et al., 1999; Golub et al., 1999; Alizadeh et al., 2000). To determine whether the single platform of gene expression profiling of leukemic blasts could replace conventional laboratory approaches while simultaneously enhancing prognostic criteria, we utilized oligonucleotide microarrays to analyze the expression of over 12,600 genes in diagnostic leukemic blasts from 360 pediatric ALL patients. These studies demonstrate that expression profiling is not only a robust approach for the accurate identification of known lineage and molecular subtypes of ALL, but also provides new insights into their underlying biology. In addition, gene expression profiling allows the accurate identification of some patients who are at a high risk for failing conventional therapeutic approaches.

Results

Expression profiling of pediatric ALL—Biologic insights

To determine if gene expression profiling of leukemic cells could identify known biologic ALL subgroups, we analyzed 327 diagnostic bone marrow (BM) samples with Affymetrix oligonucleotide microarrays containing 12,600 probe sets. The distribution of the individual prognostic subgroups within this data set is detailed in the Supplemental Data at <http://www.stjudersearch.org/data/ALL1>.

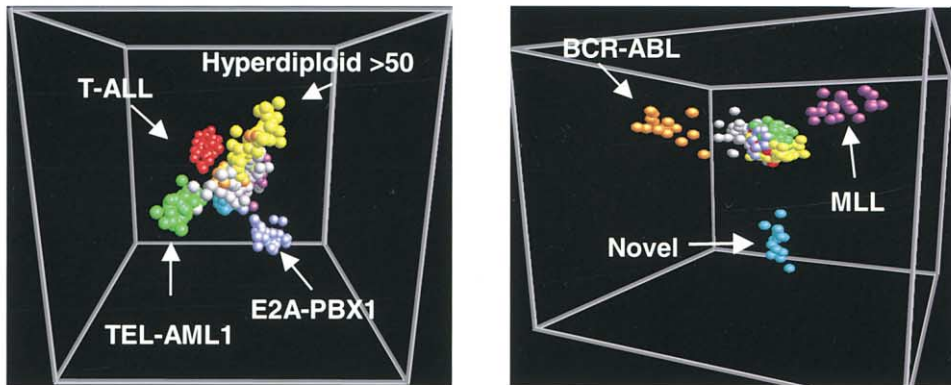
In an initial analysis of the gene expression data set (12,600

probe sets in 327 leukemia samples; greater than 4×10^6 data elements), we used an unsupervised two-dimensional hierarchical clustering algorithm to group genes on the basis of similarity in their pattern of expression over the samples. The same clustering method was also used to group the leukemia samples on the basis of similarities in their pattern of genes expressed (see Supplemental Data at <http://www.stjudersearch.org/data/ALL1>). Remarkably, this analysis clearly identified six major leukemia subtypes that corresponded to T-ALL, hyperdiploid with >50 chromosomes, *BCR-ABL*, *E2A-PBX1*, *TEL-AML1*, and *MLL* gene rearrangement. Moreover, within the heterogeneous collection of leukemias that were not assigned to one of these subtypes, a subgroup of 14 cases was identified that had a distinct gene expression profile. These cases had normal, pseudodiploid, or hyperdiploid karyotypes and lacked any consistent cytogenetic abnormality. The separation of the seven leukemia subgroups was also seen using the multidimensional scaling procedure of discriminant analysis with variance (DAV), in which the data are reduced into component dimensions consisting of linear combinations of discriminating genes (Figure 1A). For example, using the three component dimensions that accounted for 72.8% of the variance of gene expression among the subgroups, we were able to separate T-ALL (43 cases), *E2A-PBX1* (27 cases), *TEL-AML1* (79 cases), and hyperdiploid >50 (64 cases) from the remaining ALL subtypes (114 cases). Similarly, using three different components that account for an additional 16.1% of the variance in gene expression, we could discriminate cases with *BCR-ABL* (15 cases), *MLL* gene rearrangement (20 cases), and the novel subgroup of ALL (14 cases).

We next used statistical methods to identify those genes that best define the individual groups. The expression profiles obtained using the top 40 genes per subgroup selected by a chi-square metric are illustrated in Figure 1B, using the two-dimensional hierarchical clustering algorithm. The chi-square metric is a statistical test of association and provides a rank-ordered list of genes for each genetic subgroup. In this figure, each column corresponds to a single leukemia sample and each row represents the expression level of a single gene across the sample set. The expression level of each gene relative to the mean expression level across all samples is represented by a color, with red representing expression above the mean and green representing expression below the mean, and the intensity of the color corresponds to the magnitude of the deviation from the mean. As shown, distinct groups of genes distinguish cases defined by *E2A-PBX1*, *MLL*, T-ALL, hyperdiploid >50, *BCR-ABL*, the novel subgroup, and *TEL-AML1*. In addition to these specific subgroups, 65 cases (20% of the total) were identified that did not cluster into any of the leukemia subtypes. The expression profiles of these latter cases varied markedly, suggesting that they represent a heterogeneous group of leukemias. Nearly identical results were obtained when the hierarchical clustering was performed with genes selected by other statistical metrics (see Supplemental Figures S14–S18 at <http://www.stjudersearch.org/data/ALL1>).

For T-ALL, two gene clusters were identified, one expressed at high levels and one at low levels, that discriminated this subtype from B lineage cases. By contrast, for each of the other leukemia subtypes, the top-ranked discriminating genes primarily consisted of genes that were overexpressed within the specific leukemia subtype. It is important to emphasize that, with the exception of T-ALL, the identified expression profiles

A



B

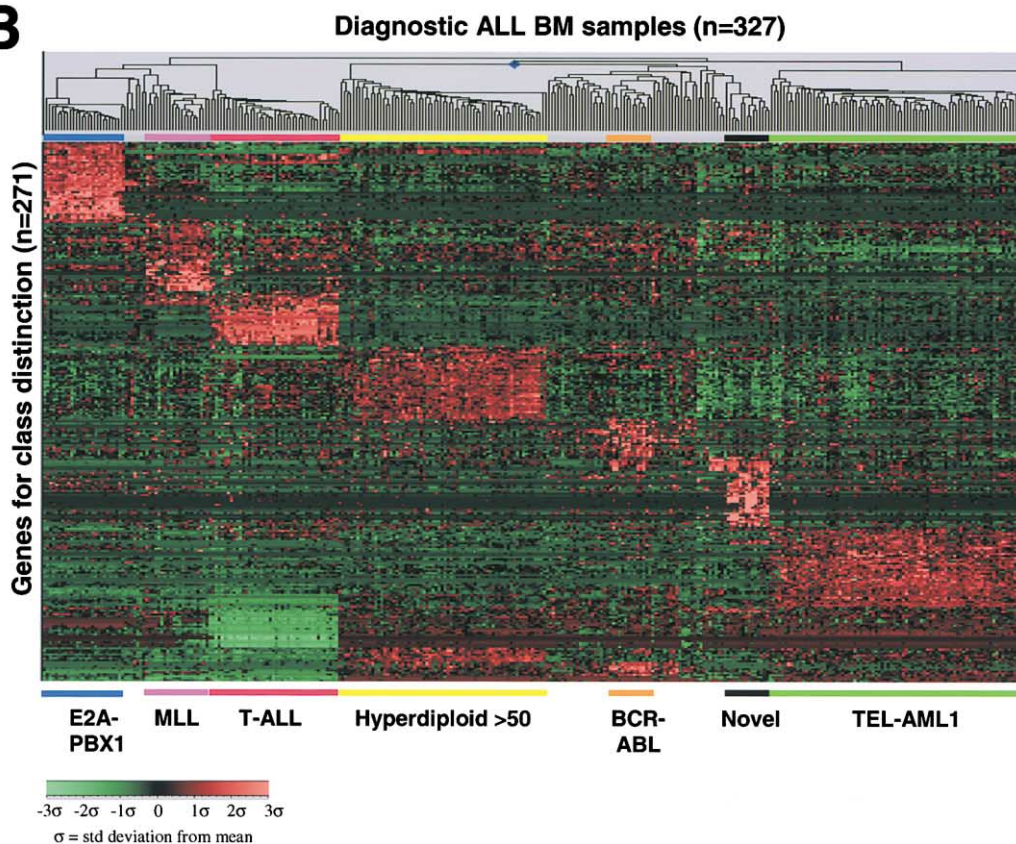


Figure 1. Expression profiles of diagnostic bone marrow ALL blasts

A: Multidimensional scaling plot using DAV of the gene expression data from 327 diagnostic BM samples generated using the 10,991 probe sets that passed the variance filter. Each case is represented by a sphere and is color-coded to indicate the specific genetic subgroup to which it belongs. Cases are displayed in gene space with each component dimension consisting of a linear combination of genes that showed the greatest variance across the data set. In the panel on the left, the space represents expression values of discriminant genes that correspond to 72.8% of the variance across the dataset, whereas the panel on the right corresponds to three separate components that represents 16.1% of the total variance. **B:** Hierarchical clustering of 327 diagnostic ALL samples (columns) versus 271 genes (rows). The genes used in this analysis are the top 40 genes chosen by a chi-square statistic that are most highly correlated with the seven specific class distinctions. Nine genes were identified as useful in discriminating more than one class, but each is used only once in this analysis. The normalized expression value for each gene is indicated by a color, with red representing high expression and green representing low expression, with the scale shown at the lower left.

do not represent a specific differentiation stage of the leukemic blasts. For example, although *E2A-PBX1* is almost exclusively found in ALLs with a pre-B cell immunophenotype (Hunger, 1996), the identified expression profile was specific for the *E2A-*

PBX1 genetic lesion and not the pre-B immunophenotype. Moreover, we were unable to define expression profiles that were specific for the immunophenotypically defined differentiation stages of the B lineage ALLs, including early pre-B, transi-

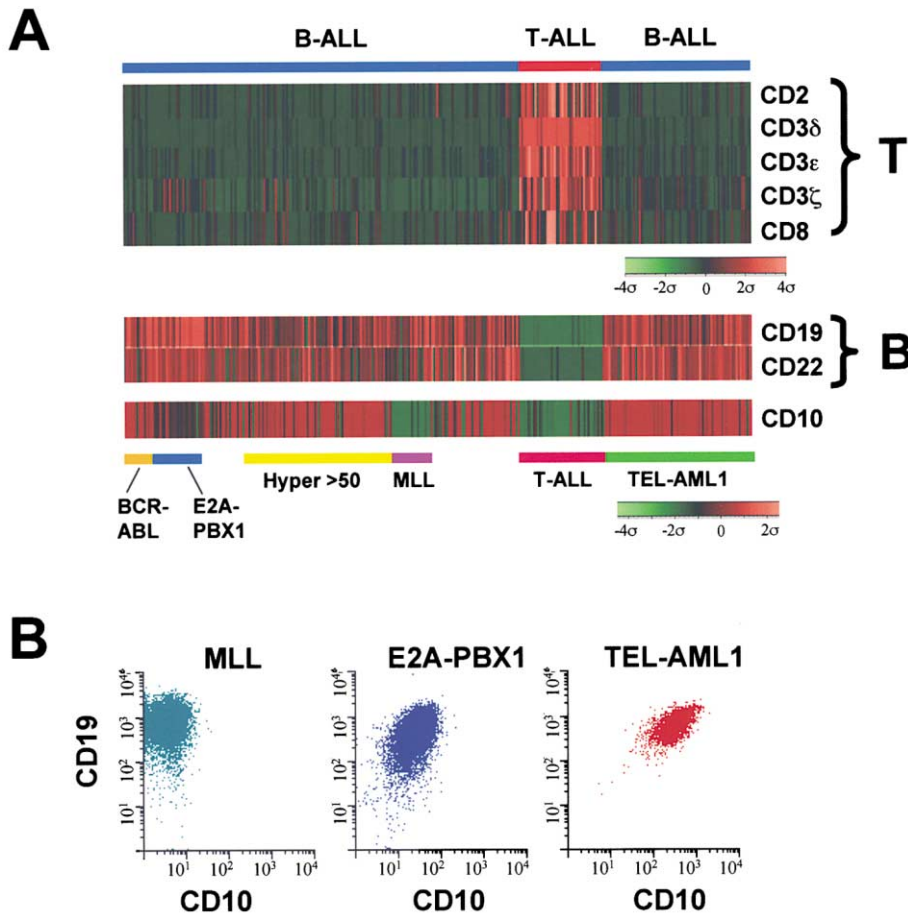


Figure 2. Correlation of gene expression analysis with immunophenotyping of ALL

A: The expression profiles of the probe sets corresponding to the immunophenotypically determined T cell-associated antigens CD2, CD3, and CD8 and B cell-associated antigens CD19, CD22, and CD10 across the 327 diagnostic BM samples. T- and B-ALL cases are indicated at the top of the figure, and the genetic subtypes are indicated at the bottom. The color-coded scales for the normalized expression values are indicated on the bottom right of each panel. **B:** Representative results from immunophenotyping using multicolor flow cytometry for detection of the CD10 and CD19 cell surface antigens. In agreement with the results of expression profiling, each B lineage leukemia expressed a high level of CD19, whereas expression of CD10 varied between subtypes, with no expression detected in MLL, intermediate levels in E2A-PBX1, and high level in TEL-AML1.

tional pre-B, and pre-B (see Supplemental Data, Tables S19–S21 and Figure S20 at <http://www.stjuderesearch.org/data/ALL1>). Rather, the gene expression profiles of the specific genetic subgroups always predominated.

To confirm that the microarray analysis provided an accurate reflection of gene expression levels, we compared the microarray data with results for RNA levels obtained by real-time RT-PCR (five genes) and with the corresponding protein levels as assessed by immunophenotype analysis performed by flow cytometry (nine specific cell surface antigens). As shown in the Supplemental Data (Figures S7–S12 and Tables S4–S9 at <http://www.stjuderesearch.org/data/ALL1>), a very high degree of correlation was observed between the levels of RNA expression detected by quantitative RT-PCR and microarray analysis. Similarly, in agreement with results from immunophenotyping, T lineage-restricted RNA expression was observed for CD2, CD3, and CD8, whereas B lineage-restricted expression was observed for CD19 and CD22 (Figure 2A and Supplemental Data, Figure S13). In addition, the level of CD10 RNA expression closely correlated with protein levels, with high expression detected in TEL-AML1 leukemias, intermediate levels in E2A-PBX1, and low to undetectable expression in cases with rearrangements of MLL (Figure 2B). Thus, microarray analysis provides an accurate reflection of expression levels for most genes and can be used to accurately detect the expression of the more common surface antigens used in the diagnostic evaluation of pediatric ALL patients.

The majority of the leukemia subtype-specific genes identified through this study were not previously known to have a restricted pattern of expression (Figure 3; the list of genes selected by each metric are provided in the Supplemental Data at <http://www.stjuderesearch.org/data/ALL1>). Besides having the potential to be used as new diagnostic and subclassification markers, these genes provide unique insights into the underlying biology of the different leukemia subtypes. For example, E2A-PBX1 leukemias were characterized by high expression of the C-MER receptor tyrosine kinase (MERTK), a known transforming gene (Graham et al., 1994; Georgescu et al., 1999), suggesting that C-MER may be involved in the abnormal growth of these cells. Similarly, HOXA9 and MEIS1 were exclusively expressed in cases having MLL rearrangements, indicating that they may be directly involved in MLL-mediated alterations in the growth of the leukemic cells. Interestingly, high expression of MTG16, a homolog of ETO (Gamou et al., 1998), was found in TEL-AML1 cases. Alteration of ETO family members in both t(8;21) acute myeloid leukemia (by translocation) (Downing, 1999) and TEL-AML1 (by altered expression) suggests that alteration in the biologic function of ETO genes may be mechanistically involved in these leukemias.

Little is known about the underlying molecular pathogenesis of hyperdiploid ALL >50 chromosomes, which clinically is distinct from hyperdiploid cases having 47–50 chromosomes. This distinction is supported by the marked differences in gene expression profiles between these two subgroups. Although

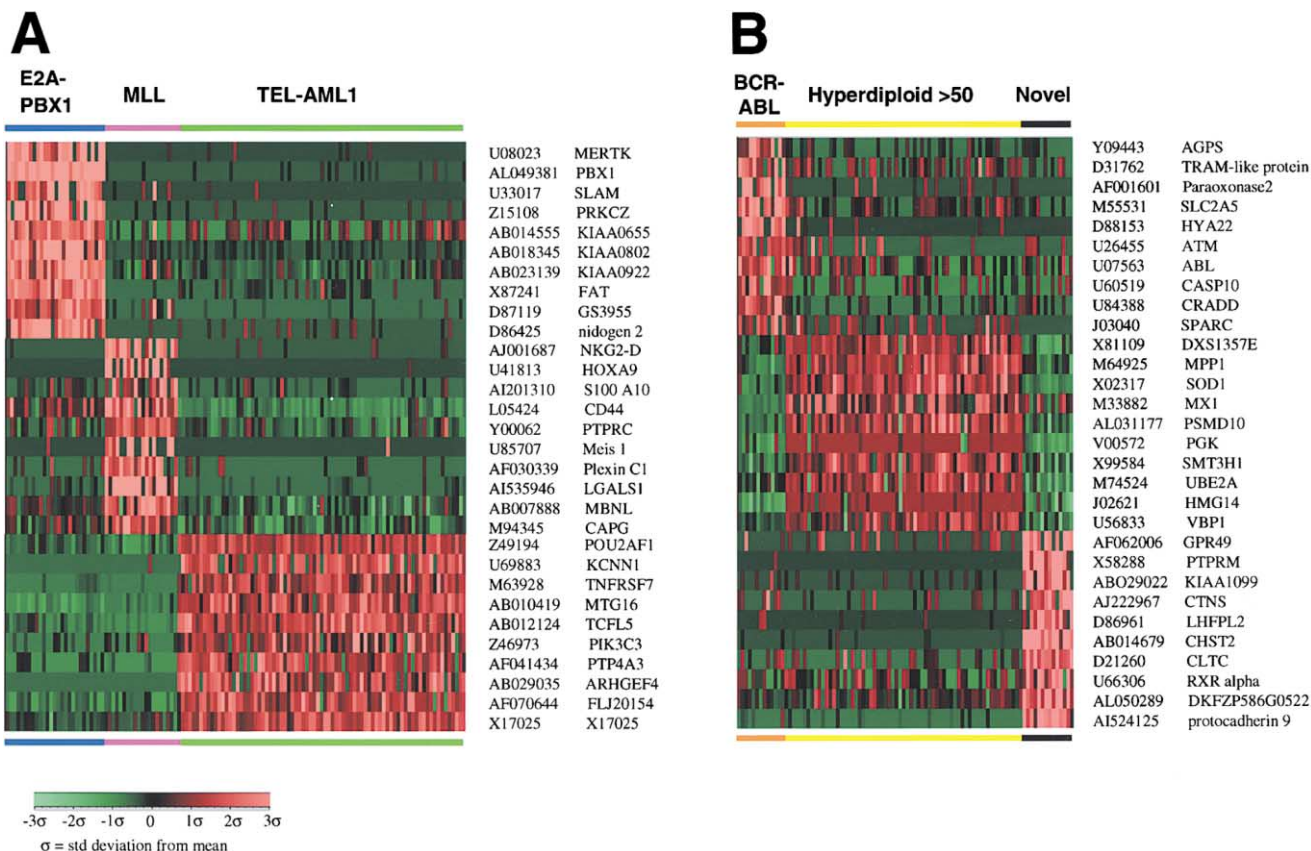


Figure 3. Class-defining genes for the individual leukemia subtypes

A and B: Shown are 10 of the 40 genes chosen by a chi-square statistic that are most highly correlated with each of the individual leukemia classes indicated at the top of the panel. The GenBank accession number and the gene symbol or DNA sequence name are listed on the right side of each panel. The color-coded scale for the normalized expression values is indicated on the bottom left. The discriminating genes for T-ALL are not shown in this figure but are provided in the Supplemental Data at <http://www.stjuderesearch.org/data/ALL1>.

hyperdiploid >50 ALLs have an excellent prognosis, the specific genetic lesions responsible for the aberrant proliferation in these cases remains poorly understood. Interestingly, almost 70% of the genes that defined this subgroup were localized to either chromosome X or 21. Moreover, the class-defining genes on chromosome X were overexpressed in the hyperdiploid >50 chromosomes ALLs irrespective of whether the leukemic blasts had a trisomy of this chromosome (data not shown). Detailed analysis will be required to determine the specific signaling pathways that are disrupted as a result of the altered expression of these genes. Lastly, the novel subgroup of ALL was defined by high expression of a group of genes, including the receptor phosphatase *PTPRM* and *LHFPL2*, a gene that is a part of the *LHFP*-like gene family—the founding member of which was identified as the target of a lipoma-associated chromosomal translocation (Petit et al., 1999). Whether the *LHFPL2* gene is altered by a chromosomal rearrangement in these leukemias remains to be determined.

Expression profiling as a diagnostic tool

A major goal of this study was to determine if the single platform of expression profiling could accurately identify the known prognostically important leukemia subtypes. We formally tested this issue by using computer-assisted learning algorithms to de-

velop an expression-based leukemia classification. Through a reiterative process of error minimization, these supervised learning algorithms learn to recognize the optimal gene expression patterns for a specific subtype. Classification was approached using a decision tree format, in which the first decision was T-ALL versus B lineage (non-T-ALL) and then within the B lineage subset, cases were sequentially classified into the known risk groups characterized by the presence of *E2A-PBX1*, *TEL-AML1*, *BCR-ABL*, *MLL* chimeric genes, and lastly hyperdiploid with >50 chromosomes. Cases not assigned to one of these classes were left unassigned (see Supplemental Data, Figure S19 at <http://www.stjuderesearch.org/data/ALL1>). Classification was performed using the supervised learning algorithm, support vector machine (SVM), with a set of discriminating genes selected by a correlation-based feature selection (CFS) or, if this method selected >20 genes for a particular class, using the top 20 ranked genes selected by a chi-square metric or one of the other metrics detailed in the Supplemental Data. As shown in Table 1, this approach resulted in exceptionally accurate class prediction in a randomly selected training set that consisted of two-thirds of the total cases (215 cases). When this classification model was then applied to a blinded test set consisting of the remaining 112 samples, an overall accuracy of 96% was achieved for class assignment (Table 1 and Supple-

Table 1. ALL subgroup prediction accuracies using support vector machine (SVM)

Subgroups	Training set ^a		Test set ^b	
	Apparent accuracy ^c	True accuracy ^d	Sensitivity ^e	Specificity ^f
T-ALL ^g	100%	100%	100%	100%
E2A-PBX1	100%	100%	100%	100%
TEL-AML1	98%	99%	100%	98%
BCR-ABL	96%	97%	83%	98%
MLL rearrangement	100%	100%	100%	100%
Hyperdiploid >50	93%	96%	100%	93%

^aThe training set consisted of 215 samples.

^bThe blinded test set consisted of 112 samples.

^cApparent accuracy was determined by leave-one-out crossvalidation.

^dTrue accuracy was determined by class prediction on the blinded test set.

^eSensitivity = (the number of positive samples predicted)/(the number of true positives)

^fSpecificity = (the number of negative samples predicted)/(the number of true negatives)

^gThe distribution of cases in the training and test sets are: T-ALL (28 cases, 15 cases); E2A-PBX1 (18, 9); TEL-AML (52, 27); BCR-ABL (9, 6); MLL (14, 6); hyperdiploid >50 (42, 22).

mental Data, Tables S16–S18). The number of genes required for optimal class assignment varied between classes. A single gene was sufficient to give 100% accuracy for both T-ALL (*CD3D*) and *E2A-PBX1* (*PBX1*), whereas 7–20 genes were required for prediction of the other classes. Only slight differences were observed in the prediction accuracies of individual classes when the process was repeated using genes selected by a number of other metrics, including T statistics, a novel metric referred to as Wilkins', or genes selected by a combination of self-organizing maps (SOM) and DAV (see Supplemental Data, Tables S13–S15). Moreover, nearly identical results were obtained when the various sets of selected genes were used in a number of different supervised learning algorithms, including κ -nearest neighbor (κ -NN), artificial neural network (ANN), and prediction by collective likelihood of emerging patterns (PCL) (see Supplemental Data, Tables S16–S18).

Importantly, the rare cases that were misclassified by gene expression analysis were highly informative. For example, four cases were apparently misclassified as *TEL-AML1* by gene expression analysis since they lacked a detectable chimeric transcript by RT-PCR. However, on further analysis, one case was shown by FISH analysis to have a *TEL-AML1* fusion, presumably a variant rearrangement that could not be detected with the amplification primers used for the *TEL-AML1* RT-PCR assay (see Supplemental Figure S21 at <http://www.stjuderresearch.org/data/ALL1>). In the other three cases, reexamination of the karyotypes revealed translocations involving the p arm of chromosome 12 in each case. By FISH analysis, two of these cases had deletion of one *TEL* allele, whereas the remaining case had a partial deletion of one *TEL* allele (see Supplemental Data, Figure S21). Thus, the identified expression profiles appear to reflect an abnormality of the *TEL* transcription factor and may provide a more accurate means of identifying a specific leukemia subtype defined by its underlying biology. Collectively, these data suggest that the single platform of gene expression profiling can accurately identify the known prognostic subtypes of ALL.

Use of expression profiles to identify patients at high risk of treatment failure

Relapse and the development of therapy-induced acute myeloid leukemia (AML) are the major causes of treatment failure in

pediatric ALL. Despite our success in identifying distinct leukemia subtypes that have either a very high or low risk of treatment failure, risk assignment remains an imprecise process. To determine if expression profiling might further enhance our ability to identify those patients who are likely to relapse, we compared the expression profiles of four groups of leukemic samples: (1) diagnostic samples of patients that develop hematological relapses ($n = 32$); (2) diagnostic samples from patients who remained in continuous complete remission (CCR) ($n = 201$); (3) diagnostic samples from patients who develop therapy-induced AML ($n = 16$); and (4) leukemic samples collected at the time of ALL relapse ($n = 25$). Using DAV, distinct gene expression profiles were identified for each of these groups (Figure 4).

To further assess the predictive power of the different gene expression profiles, we again used supervised learning algorithms. Because of the overwhelming differences in the expression profiles of the different leukemia subtypes, we were unable to identify a single expression signature that would predict relapse irrespective of the genetic subtype. However, within individual leukemic subtypes, distinct expression profiles could be defined that predicted relapse. Class assignment was performed using a SVM supervised learning algorithm with discriminating genes selected by CFS or, if this method returned >20 genes, we used the top 20 genes selected by T statistics (Supplemental Tables S22–S24 at <http://www.stjuderresearch.org/data/ALL1>). As shown, for both the T-ALL and hyperdiploid >50 subgroups, expression profiles identified those cases that went on to relapse with an accuracy of 97% and 100%, respectively, by crossvalidation. Moreover, these prediction accuracies were statistically significant when compared to results from an analysis of 1000 random permutations of the specific patient data set (Figure 5A and Supplemental Data). Similarly, expression profiles predictive of relapse were identified for *TEL-AML*, *MLL*, or cases that lacked any of the known genetic risk features (Supplemental Data, Table S25). However, although the predictive accuracies of these latter expression profiles were very high by crossvalidation, they did not reach statistical significance when compared to results from an analysis of 1000 random permutations of the same patient data set, likely secondary to the limited number of cases. The expression signatures predictive of relapse for T-ALL and hyperdiploid >50 ALLs are shown in Figures 5B and 5C. A key point is that no single gene

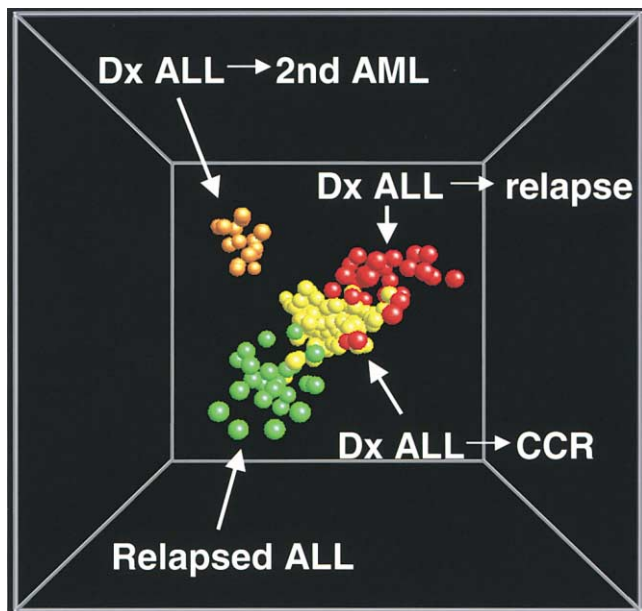


Figure 4. The expression profiles of diagnostic blasts from patients that are cured versus those that relapse or develop secondary AML are distinct

Multidimensional scaling plot of the gene expression data from 16 diagnostic (Dx) samples from patients who developed secondary AML (2nd AML), 32 Dx samples from patients who developed a hematological relapse, 201 Dx samples from patients who remained in continuous complete remission (CCR), and 25 BM or PB samples at the time of ALL relapse (relapsed ALL). Each case is represented by a sphere and is color-coded as indicated. The individual dimensions represent linear combinations of genes. The DAV was performed using all 11,322 probe sets that passed the variation filter and the displayed gene space represents the total variance within this data set.

can be used to predict the risk of relapse. Rather, patterns of expression for a combination of genes were found to be predictive. Since few known risk-stratifying biologic features have been previously identified for either T-ALL or hyperdiploid >50 ALL, our results suggest that the identified expression profiles provide independent risk-stratifying information.

A provocative observation was the identification of a distinct expression profile in the ALL blasts from those patients who developed therapy-induced AML. Because secondary AML is thought to arise from a hematopoietic stem cell that is distinct from that giving rise to the primary leukemia (Figure 6A), it is difficult to understand how the biology of the original ALL blasts could predict the risk of developing a therapy-induced complication. Nevertheless, we formally evaluated the accuracy of expression profiling in identifying these patients. Again, no single expression profile was identified that worked across the different leukemia subgroups. However, within the *TEL-AML1* subgroup, a distinct expression signature consisting of 20 genes was defined that identified, with 100% accuracy in crossvalidation, all patients who developed secondary AML, with a p value of 0.031 as assessed by comparison to results from an analysis of 1000 random permutations of the patient data set (Figure 6B). Genes within this signature included *RSU1*, a suppressor of the RAS signaling pathway, and *MSH3*, a mismatch repair enzyme (Figure 6C and Supplemental Data, Tables S26 and S27 at <http://www.stjudersearch.org/data/ALL1>). Whether the

overexpression of these genes is mechanistically involved in the increased risk of therapy-induced AML or is only a chance association remains to be determined. Formal proof of the predictive value of this identified expression signature will require confirmation in an independent group of patients.

Discussion

Contemporary approaches to the diagnosis of pediatric ALL requires an extensive range of procedures including morphology, immunophenotyping, cytogenetics, and molecular diagnostics. Using gene expression profiling, we now demonstrate that the single platform of microarray expression analysis can accurately identify each of the known prognostically and therapeutically relevant subgroups of childhood ALL. Distinct gene expression profiles were identified for ALL blasts with T lineage, hyperdiploid >50 chromosomes, *BCR-ABL*, *E2A-PBX1*, *TEL-AML1*, and *MLL* gene rearrangement. In addition, using a variety of computer-assisted supervised learning algorithms, overall diagnostic accuracies of 96% were achieved. This level of accuracy exceeds that typically achieved using contemporary diagnostic approaches in most medical centers. Moreover, the assignment of a leukemic sample to a specific biologic subgroup may be more accurately reflected by its gene expression profile than by the presence or absence of a specific genetic lesion. This is best exemplified by four cases that had expression profiles classified as *TEL-AML1* despite lacking a *TEL-AML1* chimeric message by RT-PCR. As noted, each of these cases was found to have an alteration in *TEL*, suggesting a common underlying biology. Thus, from a technical viewpoint, gene expression profiling should be a viable alternative to standard diagnostic approaches. Whether gene expression profiling will become a practical diagnostic alternative remains to be determined. It is important to stress, however, that once a diagnostic algorithm using a defined set of genes is established, its routine use in a clinical setting will require only minimal expertise. As the cost of gene expression profiling decreases, this type of analysis will likely become highly competitive when compared to the cumulative cost of the various diagnostic studies that are presently used.

One of the most surprising observations from this study was the remarkable difference in the expression profiles of the individual leukemia subtypes. Despite having relatively homogeneous morphology and limited variability in the extent of T or B cell differentiation, each leukemia subtype had a distinct expression profile that involved a large number of genes. These observations are in agreement with a more limited study in which the expression profiles of ALLs with *MLL* rearrangements were shown to differ from those of other acute leukemias (Armstrong et al., 2001). Remarkably, the expression differences between individual ALL subtypes were more robust than expression differences between either lung adenocarcinoma (Su et al., 2001) or melanoma (Ramaswamy et al., 2001) and bladder transitional carcinoma, tumors that conceptually would be considered more divergent. Thus, our data supports the interpretation that these leukemic subtypes are distinct biological and clinical entities. For subgroups defined by either translocation-encoded chimeric transcription factors or altered signaling proteins such as *BCR-ABL*, the presence of a distinct gene expression profile is not completely unexpected. By contrast, the identification of a unique expression profile for novel ALL cases with >50 chromo-

A Gene expression profiles as predictors of relapse

Genetic Subtype	Apparent Accuracy	Support Vector Machine $p =$ (Permutation Test)
T-ALL	97%	0.007
Hyperdiploid > 50	100%	0.018

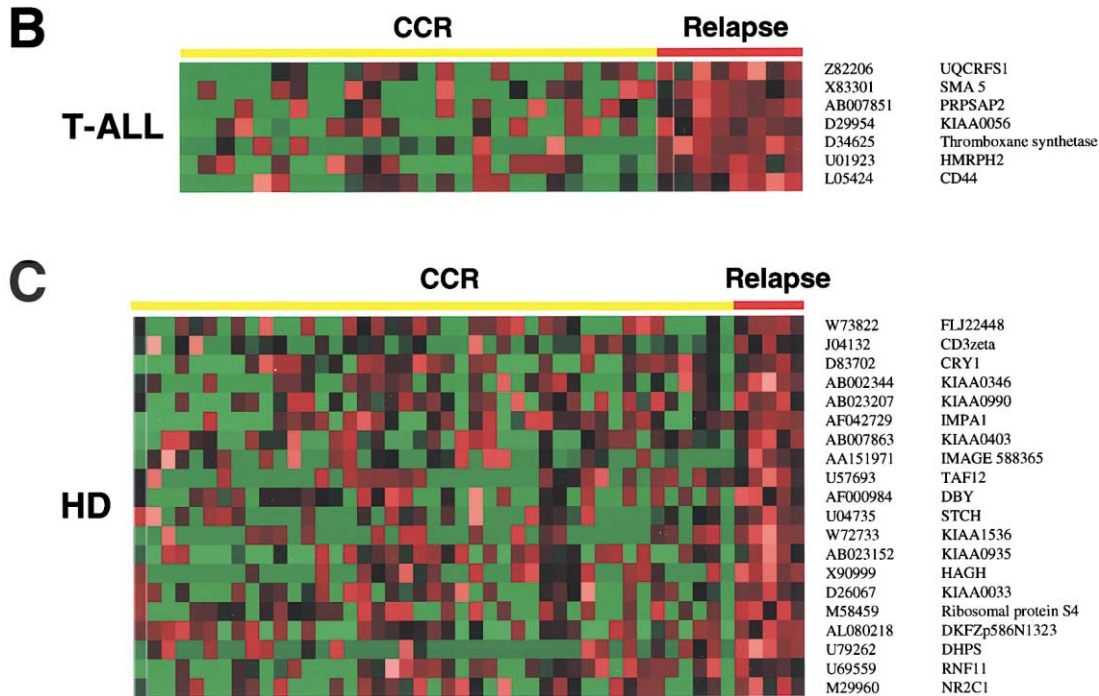


Figure 5. Gene expression profiles as predictors of relapse

A: Genes predictive of the class distinction relapse versus continuous complete remission (CCR) for either T-ALL or hyperdiploid >50 ALL (HD) were chosen by CFS or T statistics. The selected genes were then used in a SVM supervised learning algorithm, and performance was assessed by a crossvalidation experiment. The apparent accuracies of prediction are indicated. The significance of the prediction accuracy was determined by performing 1000 permutation experiments for each subtype-specific group (see Supplemental Data at <http://www.stjuderesearch.org/data/ALL1>). The percentage of these 1000 random partitions that gave a prediction accuracy equal to or better than that for the relapse prediction was taken as a p value. **B and C:** The expression pattern of the 7 and 20 genes that were selected as discriminators of relapse versus CCR in T-ALL and HD cases, respectively. The GenBank accession number and the gene symbol or DNA sequence name are listed on the right side of each panel.

somes is surprising. Examination of the expression profiles of hyperdiploid >50 and the ALL subgroup identified here should provide important new insights into the underlying pathogenesis of these leukemic subtypes. The expression profiles of these and the other leukemia subtypes has not only provided insights into their biology but has also resulted in the identification of a number of genes that should prove useful as markers to monitor patients for minimal residual disease. In addition, some of the identified genes, such as the *C-MER* receptor tyrosine kinase in *E2A-PBX1*, may prove to be useful targets against which novel therapeutic agents could be developed.

The marked differences seen in the expression profiles of the various ALL subtypes suggest that transformation may occur through distinct pathways. Although only a limited number of growth control mechanisms need to be subverted to result in cellular transformation (Hanahan and Weinberg, 2000), our data suggest that in pediatric ALL, multiple pathways may exist and

ultimately converge on these critical functions. Clearly, some of the identified expression differences result from variations in lineage or stage of differentiation. What proportion of the remaining expression changes are mechanistically involved in transformation remains to be determined. Similarly, determining how the identified expression differences are involved in the unique clinical biology of the leukemias, including their distinctive responses to particular types of chemotherapy, remains to be defined.

One of the most promising aspects of gene expression profiling is the hope that it will improve the ability to accurately identify those patients who are at a high risk of failing conventional therapy. Strikingly, our results demonstrate that the gene expression profiles differ between the diagnostic samples of patients who relapse and those who remain in continuous complete remission. Moreover, specific expression profiles in the diagnostic samples of either T-ALL or hyperdiploid ALL appear

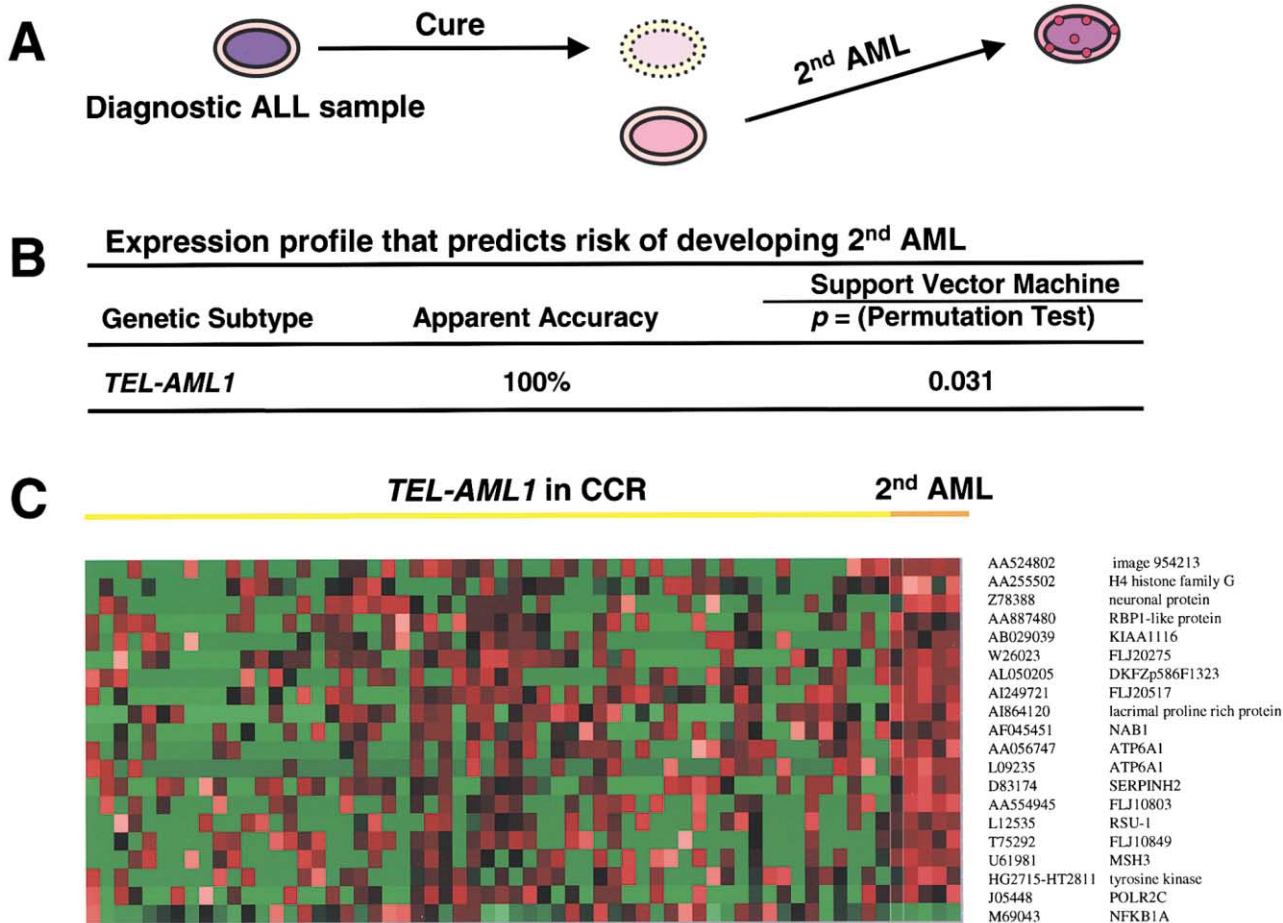


Figure 6. A gene expression profile that predicts the development of secondary AML in *TEL-AML1* ALL

A: Schematic illustration showing that therapy-induced AML (2nd AML) is believed to arise from a hematopoietic progenitor that is distinct from the one that gave rise to the original ALL leukemic clone. **B:** Genes within the diagnostic ALL BM samples that were predictive of development of therapy-induced AML were selected by their MIT score and then used in a SVM supervised learning algorithm. Performance was assessed by a crossvalidation experiment, and the results are indicated as apparent accuracy. The significance of the prediction accuracy was determined by performing 1000 random permutation experiments, with the p value indicating the percentage of these random experiments that gave a prediction accuracy equal to or better than the secondary AML prediction. **C:** The expression profile of a subset of the genes selected as predictors of the development of secondary AML in *TEL-AML1*-positive ALLs. The GenBank accession number and the gene symbol or DNA sequence name are listed on the right of the panel.

to be accurate predictors of relapse. Although these data will need to be validated in prospective studies, these findings raise the expectation that, in the future, this type of analysis will be used to make therapeutic decisions. In addition, the identified expression profiles should provide critical insights into the underlying mechanisms that contribute to relapse. The observation that we could not identify a common expression profile that predicted relapse irrespective of the genetic subtype suggests that a unifying mechanism of relapse may not exist. Rather, mechanisms of relapse or drug resistance may differ among leukemia subtypes. Alternatively, the identified expression profiles may consist of genes that are chance associations with pending relapse and not genes directly involved in the underlying biology. It is important to keep in mind that the present analysis falls far short of a total transcriptional profile of the leukemic blasts. Although 12,600 probe sets are present on the microarray, the total number of genes that this represents accounts for less than 20% of the estimated number of genes

within the human genome (Hogenesch et al., 2001). In the future, the use of higher-density chips should not only further enhance our ability to accurately identify those patients who will relapse, but should also provide a clearer view of the underlying biology.

The provocative finding of an expression profile in the diagnostic samples that identifies patients who subsequently developed therapy-induced AML will need to be validated in an independent cohort of patients. Although a distinct expression profile was defined that identified those *TEL-AML1* ALL patients who developed secondary AML, it remains to be determined if the profile represents genes that are mechanistically involved in the enhanced risk or are only statistical associations. Reassessment of these samples using higher-density chips will allow a significantly broader view of the genes that characterize the ALL blasts of patients who develop secondary AML and may thereby help to answer this question. Despite these caveats, these findings suggest the concept that expression profiling of leukemic blasts, and possibly nonmalignant hematopoietic

cells, may enhance our ability to identify patients who are at a high risk of developing therapy-induced complications, including secondary malignancies, severe organ toxicities, and infections.

Two recent studies have presented results from a more limited analysis of the expression profiles of pediatric ALLs (Armstrong et al., 2001; Ferrando et al., 2002). Although there is a high degree of correlation between our results and those presented in these other studies, subtle differences are evident. Foremost is the observation that not all *MLL* discriminating genes identified in the Armstrong paper were found to distinguish this ALL subtype in our analysis. Using a much larger number of cases, we find that some of the genes that were originally found to correlate with *MLL* in fact have a broader expression pattern than was originally appreciated. Similarly, in T-ALL, by using a much larger number of genes to assess the expression profiles, we now find prognostic markers that were not identified in the study of Ferrando.

In summary, contemporary risk stratification requires a combination of methodologies and fails to identify many patients who are at high risk of drug-induced toxicities. The data presented here suggest that the single platform of gene expression profiling provides a robust and accurate approach for the diagnosis and risk stratification of pediatric ALL patients. Moreover, this approach should enhance our ability to identify patients who are at a high risk of developing marrow relapse and drug-related toxicities. In the future, development of custom diagnostic chips containing those genes that define both prognostically important leukemia subtypes as well as a patient's relative risks to relapse or develop therapy-induced AML would significantly advance our ability to individualize therapy so that each patient has the highest chance for cure. Lastly, the generated database of comprehensive gene expression profiles coupled with detailed immunophenotype, cytogenetic, molecular diagnostic, and treatment outcome data should be an invaluable resource for studies of pediatric leukemia.

Experimental procedures

Tumor samples

The diagnosis of ALL was based on the morphologic evaluation of the bone marrow and on the pattern of reactivity of the leukemic blasts with a panel of monoclonal antibodies directed against lineage-associated antigens. A total of 389 pediatric acute leukemia samples were analyzed in this study, from which high-quality gene expression data was obtained on 360 (93%). The successfully analyzed samples included: 332 diagnostic BM, 3 diagnostic peripheral bloods (PB), and 25 relapsed ALL samples from BM or PB. All relapse samples and 264 (79%) of the diagnostic ALL BM samples were from patients enrolled on St. Jude Children's Research Hospital Total Therapy Studies XIII A or XIII B and corresponded to 64% of the patients treated on these protocols. The details of these protocols have been previously published (Pui et al., 2000). The remaining samples were obtained from patients treated on St. Jude Total Therapy Studies XI, XII, XIV, XV, or by best clinical management. All protocols and consent forms were approved by the hospital's institutional review board, and informed consent was obtained from parents, guardians, or patients (as appropriate). The composition of the data sets used for the identification of gene expression profiles predictive of specific genetic subtypes, hematological relapse, and risk of developing secondary AML are detailed in the Supplemental Data at <http://www.stjude-research.org/data/ALL1>.

Gene expression profiling

RNA was extracted from cryopreserved mononuclear cell suspensions from diagnostic BM aspirates or PB samples using the Trizol reagent, and the RNA integrity was assessed by using an Agilent 2100 Bioanalyzer (Agilent,

Palo Alto, CA). cDNA was synthesized using a T-7 linked oligo-dT primer, and cRNA was then synthesized with biotinylated UTP and CTP. The labeled RNA was then fragmented and hybridized to HG_U95Av2 oligonucleotide arrays (Affymetrix Incorporated, Santa Clara, CA) according to Affymetrix protocols.

Arrays were scanned using a laser confocal scanner (Agilent), and the expression value for each gene was calculated using Affymetrix Microarray software v.4.0. The average intensity difference (AID) values were normalized across the sample set, and minimum quality control standards were established for including a sample's hybridization data in the study (see Supplemental Data at <http://www.stjude-research.org/data/ALL1>). To ensure consistency of data acquisition throughout the study, 10% of samples were run in duplicate. An exceedingly high reproducibility was observed between replicate samples, with less than 1% of genes having a variation in average intensity difference (AID) of >2-fold. The primary hybridization data are available at our website (<http://www.stjude-research.org/ALL1>).

Statistical analysis

Unsupervised hierarchical clustering, principal component analysis (PCA), discriminant analysis with variance (DAV), and self-organizing maps (SOM) were performed using GeneMaths software (v.1.5, Applied Maths, Belgium). Data reduction to define the genes most useful in class distinction was performed using a variety of metrics as detailed in the Supplemental Data at <http://www.stjude-research.org/data/ALL1>. Genes selected by the various metrics were used in supervised learning algorithms to build classifiers that could identify the specific genetic or prognostic subgroups. Algorithms used included *k*-nearest neighbors (*k*-NN), support vector machine (SVM), prediction by collective likelihood of emerging patterns (PCL), an artificial neural network (ANN), and weighted voting. Performance of each model was initially assessed by leave-one-out crossvalidation on a randomly selected stratified training set consisting of two-thirds of the total cases. True error rates of the best-performing classifiers were then determined using the remaining third of the samples as a blinded test group. Details of the individual metrics and supervised learning algorithms are described in the Supplemental Data.

Supplemental data

Additional information on the samples, methods, statistical analysis, and results from the comparison of microarray gene expression levels with mRNA levels determined by real time RT-PCR or antigen levels determined by flow cytometry are available in the Supplemental Data at <http://www.stjude-research.org/data/ALL1>.

Acknowledgments

The authors thank the staff of the Molecular Pathology Laboratory and the Hartwell Center for Bioinformatics and Biotechnology at St. Jude Children's Research Hospital (SJCRH) for outstanding technical support and the clinicians for providing excellent medical care to the patients. We also thank Louxin Zhang and Zhuo Zhang for their help with preliminary data analysis, Susan Mathew for assistance with FISH analysis, Kevin Girtman for assistance with the real-time RT-PCR assays, Michael Jaynes for help with obtaining cryopreserved BM and PB samples, and John Cleveland for critical reading of the manuscript. This work was in part supported by National Institutes of Health grants P01 CA71907-06 (J.R.D.), CA51001 (M.V.R. and C.-H.P.), CA36401 (W.E.E., M.V.R., and C.-H.P.), CA78224 (W.E.E., M.V.R., and C.-H.P.), and Cancer Center CORE Grant CA-21765 (to SJCRH). Additional support was provided by a National Science Foundation grant EIA-0074869 (D.W.), the Singapore Agency for Science, Technology and Research (J.L., H.L., and L.W.), the National Medical Research Council of Singapore (E.-J.Y.), and the American Lebanese and Syrian Associated Charities (ALSAC) of SJCRH.

Received: February 12, 2002

Revised: March 1, 2002

References

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. (2000). Distinct types of

- diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503–511.
- Arico, M., Valsecchi, M.G., Camitta, B., Schrappe, M., Chessells, J.M., Baruchel, A., Gaynon, P.S., Silverman, L., Janka-Schaub, G., Kamps, W., et al. (2000). Outcome of treatment in children with Philadelphia chromosome-positive acute lymphoblastic leukemia. *N. Engl. J. Med.* 342, 998–1006.
- Armstrong, S.A., Staunton, J.E., Silverman, L.B., Pieters, R., den Boer, M.L., Minden, M.D., Sallan, S.E., Lander, E.S., Golub, T.R., and Korsmeyer, S.J. (2002). MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nat. Genet.* 30, 41–47.
- Biondi, A., Cimino, G., Pieters, R., and Pui, C.H. (2000). Biological and therapeutic aspects of infant leukemia. *Blood* 96, 24–33.
- Downing, J.R. (1999). The AML1-ETO chimaeric transcription factor in acute myeloid leukaemia: biology and clinical significance. *Br. J. Haematol.* 106, 296–308.
- Ferrando, A.A., Neuberger, D.S., Staunton, J., Loh, M.L., Haurd, C., Raimondi, S.C., Behm, F.G., Pui, C.-H., Downing, J.R., Gilliland, D.G., et al. (2002). Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* 1, 75–87.
- Gamou, T., Kitamura, E., Hosoda, F., Shimizu, K., Shinohara, K., Hayashi, Y., Nagase, T., Yokoyama, Y., and Ohki, M. (1998). The partner gene of *AML1* in t(16;21) myeloid malignancies is a novel member of the *MTG8 (ETO)* family. *Blood* 91, 4028–4037.
- Georgescu, M.M., Kirsch, K.H., Shishido, T., Zong, C., and Hanafusa, H. (1999). Biological effects of c-Mer receptor tyrosine kinase in hematopoietic cells depend on the Grb2 binding site in the receptor and activation of NF- κ B. *Mol. Cell. Biol.* 19, 1171–1181.
- Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531–537.
- Graham, D.K., Dawson, T.L., Mullaney, D.L., Snodgrass, H.R., and Earp, H.S. (1994). Cloning and mRNA expression analysis of a novel human protooncogene, c-mer. *Cell Growth Differ.* 5, 647–657.
- Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* 100, 57–70.
- Heerema, N.A., Sather, H.N., Ge, J., Arthur, D.C., Hilden, J.M., Trigg, M.E., and Reaman, G.H. (1999). Cytogenetic studies of infant acute lymphoblastic leukemia: poor prognosis of infants with t(4;11)—a report of the Children's Cancer Group. *Leukemia* 13, 679–686.
- Hogenesch, J.B., Ching, K.A., Batalov, S., Su, A.I., Walker, J.R., Zhou, Y., Kay, S.A., Schultz, P.G., and Cooke, M.P. (2001). A comparison of the Celera and Ensembl predicted gene sets reveals little overlap in novel genes. *Cell* 106, 413–415.
- Hunger, S.P. (1996). Chromosomal translocations involving the *E2A* gene in acute lymphoblastic leukemia: clinical features and molecular pathogenesis. *Blood* 87, 1211–1224.
- Perou, C.M., Jeffrey, S.S., van de Rijn, M., Rees, C.A., Eisen, M.B., Ross, D.T., Pergamenschikov, A., Williams, C.F., Zhu, S.X., Lee, J.C., et al. (1999). Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl. Acad. Sci. USA* 96, 9212–9217.
- Petit, M.M., Schoenmakers, E.F., Huysmans, C., Geurts, J.M., Mandahl, N., and Van de Ven, W.J. (1999). LHFP, a novel translocation partner gene of HMGIC in a lipoma, is a member of a new family of LHFP-like genes. *Genomics* 57, 438–441.
- Pui, C.H., and Evans, W.E. (1998). Acute lymphoblastic leukemia. *N. Engl. J. Med.* 339, 605–615.
- Pui, C.H., Frankel, L.S., Carroll, A.J., Raimondi, S.C., Shuster, J.J., Head, D.R., Crist, W.M., Land, V.J., Pullen, D.J., and Steuber, C.P. (1991). Clinical characteristics and treatment outcome of childhood acute lymphoblastic leukemia with the t(4;11)(q21;q23): a collaborative study of 40 cases. *Blood* 77, 440–447.
- Pui, C.H., Boyett, J.M., Rivera, G.K., Hancock, M.L., Sandlund, J.T., Ribeiro, R.C., Rubnitz, J.E., Behm, F.G., Raimondi, S.C., Gajjar, A., et al. (2000). Long-term results of Total Therapy studies 11, 12 and 13A for childhood acute lymphoblastic leukemia at St Jude Children's Research Hospital. *Leukemia* 14, 2286–2294.
- Pui, C.H., Campana, D., and Evans, W.E. (2001). Childhood acute lymphoblastic leukaemia—current status and future perspectives. *Lancet Oncol.* 2, 597–607.
- Raimondi, S.C., Behm, F.G., Roberson, P.K., Williams, D.L., Pui, C.H., Crist, W.M., Look, A.T., and Rivera, G.K. (1990). Cytogenetics of pre-B-cell acute lymphoblastic leukemia with emphasis on prognostic implications of the t(1;19). *J. Clin. Oncol.* 8, 1380–1388.
- Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C.H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J.P., et al. (2001). Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl. Acad. Sci. USA* 98, 15149–15154.
- Schrappe, M., Reiter, A., Ludwig, W.D., Harbott, J., Zimmermann, M., Hidde-mann, W., Niemeier, C., Henze, G., Feldges, A., Zintl, F., et al. (2000). Improved outcome in childhood acute lymphoblastic leukemia despite reduced use of anthracyclines and cranial radiotherapy: results of trial ALL-BFM 90. *Blood* 95, 3310–3322.
- Silverman, L.B., Gelber, R.D., Dalton, V.K., Asselin, B.L., Barr, R.D., Clavell, L.A., Hurwitz, C.A., Moghrabi, A., Samson, Y., Schorin, M.A., et al. (2001). Improved outcome for children with acute lymphoblastic leukemia: results of Dana-Farber Consortium Protocol 91-01. *Blood* 97, 1211–1218.
- Su, A.I., Welsh, J.B., Sapinoso, L.M., Kern, S.G., Dimitrov, P., Lapp, H., Schultz, P.G., Powell, S.M., Moskaluk, C.A., Frierson, H.F., Jr., and Hampton, G.M. (2001). Molecular classification of human carcinomas by use of gene expression signatures. *Cancer Res.* 61, 7388–7393.