JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

From the New York University Cancer Insti-
tute and Division of Pediatric Hematology/
Oncology, New York University School of
Medicine; and Department of Pediatrics,
Mount Sinai School of Medicine, New
York, NY; Cancer Research and Treatment
Center, University of New Mexico and
Sandia National Laboratories, Albuquerque,
NM; the Department of Paediatric Oncolo-
gy/Haematology, Erasmus Medical Center,
Sophia Children's Hospital, Rotterdam; and
Department of Human Genetics, Leiden
University Medical Center, Leiden, the
Netherlands; Department of Pharmaceuti-
cal Sciences, Colleges of Medicine and
Pharmacy, University of Tennessee,
Memphis, TN; University of Southern Cali-
fornia and Children's Oncology Group,
Arcadia, CA; Department of Pediatrics,
University of Florida College of Medicine
and University of Florida Shands Cancer
Center, Gainesville, FL; Department of
Hematology and Oncology, Children's
National Medical Center, George Washing-
ton University School of Medicine and
Public Health, Washington, DC; and
Research Institute and Clinic of Pediatric
Oncology, University Medical Center,
Hamburg, Germany.

Gene Expression Signatures Predictive of Early Response and Outcome in High-Risk Childhood Acute Lymphoblastic Leukemia: A Children's Oncology Group Study

Deepa Bhojwani, Huining Kang, Renee X. Menezes, Wenjian Yang, Harland Sather, Naomi P. Moskowitz, Dong-Joon Min, Jeffrey W. Potter, Richard Harvey, Stephen P. Hunger, Nita Seibel, Elizabeth A. Raetz, Rob Pieters, Martin A. Horstmann, Mary V. Relling, Monique L. den Boer, Cheryl L. Willman, and William L. Carroll

ABSTRACT

Purpose
To identify children with acute lymphoblastic leukemia (ALL) at initial diagnosis who are at risk for inferior response to therapy by using molecular signatures.

Patients and Methods
Gene expression profiles were generated from bone marrow blasts at initial diagnosis from a cohort of 99 children with National Cancer Institute–defined high-risk ALL who were treated uniformly on the Children's Oncology Group (COG) 1961 study. For prediction of early response, genes that correlated to marrow status on day 7 were identified on a training set and were validated on a test set. An additional signature was correlated with long-term outcome, and the predictive models were validated on three large, independent patient cohorts.

Results
We identified a 24–probe set signature that was highly predictive of day 7 marrow status on the test set ($P = .0061$). Pathways were identified that may play a role in early blast regression. We have also identified a 47–probe set signature (which represents 41 unique genes) that was predictive of long-term outcome in our data set as well as three large independent data sets of patients with childhood ALL who were treated on different protocols. However, we did not find sufficient evidence for the added significance of these genes and the derived predictive models when other known prognostic features, such as age, WBC, and karyotype, were included in a multivariate analysis.

Conclusion
Genes and pathways that play a role in early blast regression may identify patients who may be at risk for inferior responses to treatment. A fully validated predictive gene expression signature was defined for high-risk ALL that provided insight into the biologic mechanisms of treatment failure.

J Clin Oncol 26:4376-4384. © 2008 by American Society of Clinical Oncology

INTRODUCTION

The current management of children with acute lymphoblastic leukemia (ALL) modulates treatment intensity according to the risk of relapse, which thereby maximizes opportunities for cure and minimizes adverse effects.[1]

A number of variables have been shown to be predictive of outcome in childhood ALL, including clinical and laboratory features, cytogenetic characteristics of the blast, and early response to chemotherapy.[2] These variables are routinely used for treatment assignment, but approximately 20% of children unpredictably suffer a relapse.[3]

Global gene expression profiling has facilitated the discovery of biologic subgroups in a variety of cancers.[4,5] This technique has been shown to accurately classify ALL into cohorts that correspond to known biologic subgroups.[6,7] However, it has proved more difficult to identify signatures that are globally predictive of outcome. In the present study, we performed gene expression profiles on leukemic blasts from children who were treated on a single, contemporary Children's Oncology Group (COG) protocol for high-risk (HR) ALL to discover gene expression signatures that are predictive of early response and outcome.

PATIENTS AND METHODS

Diagnostic marrow samples from 99 children (age 1 to 18 years) with National Cancer Institute–defined HR B-precursor ALL (age $\geq$10 years and/or presenting WBC $\geq$ 50,000/μL) who were treated on the COG 1961 protocol were analyzed.[8] We focused on this particular group of patients, because many lack known genetic subtypes predictive of outcome. All patients received a standard four-drug induction and were further classified as slow early responders (SER)—day 7 marrow was M3 ($>$ 25% blasts)—or rapid early responders (RER)—day 7 marrow was M1 ($<$ 5% blasts) or M2 (5% to 25% blasts).

To determine genetic profiles associated with early response to therapy, we analyzed 82 of 99 patients: 42 patients who had M1 marrow on day 7 were compared with 40 patients who had M3 marrow on day 7. Patients with M2 marrow (n = 17) were excluded to maximize the distinction between responders. To study the genes associated with long-term outcome, we analyzed expression profiles of 59 patients who fulfilled the following criteria: 28 patients who remained in complete continuous remission (CCR) for at least 4 years and 31 patients with marrow relapse within the first 3 years of initial diagnosis. Forty-two samples were common to both the early response and outcome analyses. Patient characteristics are listed in Appendix Table A1 (online only).

RNA Extraction and Amplification and DNA Arrays

Total RNA was extracted from cryopreserved blasts from the COG cell bank by using RNeasy Midi kits (Qiagen, Valencia, CA) followed by the MinEluate kit (Qiagen). Fifty nanograms of total RNA were used as template in a double-amplification protocol by using the RiboAmp OA kit (Arcturus, Mountain View, CA) according to the manufacturer's recommendations. In vitro transcription was completed with biotinylated UTP and CTP for labeling by using the Enzo BioArray HighYield RNA Transcript Labeling kit (Enzo Diagnostics, Farmingdale, NJ). Twenty micrograms of labeled cRNA were fragmented and hybridized to Affymetrix U133Plus2.0 microarrays (Affymetrix, Santa Clara, CA). These arrays contain 54,675 probe sets, which represented approximately 38,500 genes.

Screening Analysis for Cytogenetic Risk Group

Patients were tested by reverse transcriptase polymerase chain reaction (RT-PCR) for the presence of each of four common prognostic translocations: t(1;19), t(4;11), t(9;22), and t(12;21). The t(1;19), t(4;11), and t(12;21) fusion products were assayed by qualitative RT-PCR, whereas the t(9;22) analysis was done quantitatively by using TaqMan technology (Applied Biosystems, Foster City, CA). Primers are listed in Appendix Table A2 and methods for the assays detailed in the Appendix (online only).

Rank by Response Type	Training Set P Adjusted for Clinical Covariates	Test Set		Affymetrix Identification	Gene Symbol	Gene Description
		P of t Test	P Adjusted for Clinical Covariates			
RER						
1	.001	.066	.252	219489_s_at	RHBDL2	Rhomboid, veinlet-like 2 (*Drosophila*)
2	< .001	.232	.208	228346_at		Transcribed sequence with strong similarity to protein sp:P00722 (*Escherichia coli*) BGAL_ECOLI β-galactosidase
3	.002	.008*	.035*	225606_at	BCL2L11	*BCL2*-like 11 (apoptosis facilitator)
4	< .001	.300	.885	203588_s_at	TFDP2	Transcription factor Dp-2 (*E2F* dimerization partner 2)
5	< .001	.391	.889	203505_at	ABCA1	ATP-binding cassette, sub-family A (*ABC1*), member 1
6	.001	.001*	.004*	1555372_at	BCL2L11	*BCL2*-like 11 (apoptosis facilitator)
7	.007	.019*	.018*	1569110_x_at	PDCD6	Programmed cell death 6
SER						
1	< .001	.064	.083	227353_at	EVER2	Epidermodysplasia verruciformis 2
2	.001	.158	.670	214255_at	ATP10A	ATPase, Class V, type 10A
3	.004	.173	.740	219667_s_at	BANK1	B-cell scaffold protein with ankyrin repeats 1
4	.001	.023*	.028*	206940_s_at	POU4F1	*POU* domain, class 4, transcription factor 1
5	.002	.346	.863	223562_at	PARVG	Parvin, γ
6	.001	.059	.378	203373_at	SOCS2	Suppressor of cytokine signaling 2
7	< .001	.324	.707	226869_at		Full-length insert cDNA clone ZD77F06
8	.015	.013*	.110	211675_s_at	HIC	I-mfa domain-containing protein
9	< .001	.432	.799	232614_at		MRNA; cDNA DKFZp686K02231 (from clone DKFZp686K02231)
10	.005	.121	.208	242644_at	EVER2	Epidermodysplasia verruciformis 2
11	.017	.292	.590	205290_s_at	BMP2	Bone morphogenetic protein 2
12	.001	.505	.306	223451_s_at	CKLF	Chemokine-like factor
13	.030	.139	.418	207339_s_at	LTB	Lymphotoxin beta (tumor necrosis factor superfamily, member 3)
14	.023	.015*	.043*	229390_at		Full length insert cDNA clone ZA84A12
15	.014	.017*	.087	212070_at	GPR56	G protein–coupled receptor 56
16	.016	.030*	.044*	204198_s_at	RUNX3	Runt-related transcription factor 3
17	.008	.129	.184	227013_at	LATS2	*LATS*, large tumor suppressor, homolog 2 (*Drosophila*)

Table 1. Significant Probe Sets Predictive of Early Response

Abbreviations: RER, rapid early responder; SER, slow early responder.
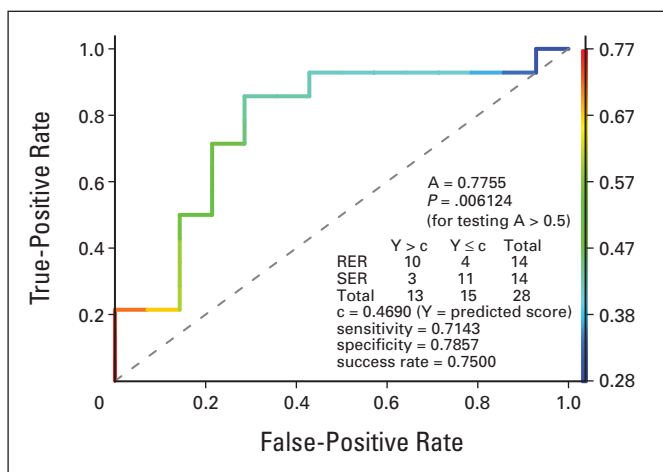*Statistically significant in the test set.

Data Analysis

Data generated from the COG 1961 samples discussed in this publication have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) and are accessible through series accession number GSE7440.

Gene expression values were generated by using Affymetrix MAS 5.0 Software. Expression levels were scaled to an average value of 1,000 per gene chip[9] and were log transformed. In each analysis, the probe sets of 53 nonhuman genes and those that did not receive present calls in at least 30% of the samples were removed from the study.

For prediction of early response, the samples (n = 82) were randomly divided into a training set (28 RER, 26 SER) and a test set (14 RER, 14 SER). A nearest shrunken centroids prediction model with a subset of genes that were best associated with early response (RER v SER) was determined by utilizing Prediction Analysis of Microarrays (PAM)[10] packaged in R (Stanford University Labs, Palo Alto, CA; www.r-project.org/) with a 200×10-fold cross validation procedure on the training data set. This model was used to make predictions on the test set. Logistic regression was utilized to test the significances of the subset of genes and the class predictor when analysis was adjusted for clinical covariates, such as age and presenting WBC.

For long-term outcome prediction, t test and adjusted P value (or false discovery rate [FDR]), as proposed by Benjamini and Hochberg,[11] were utilized to select a subset of probe sets that were statistically associated with outcome. A logistic regression[12] model was used to test whether each of the genes added prognostic value beyond that of known clinical covariates. Logistic regression with various variable selection options[11] was utilized to build the best models for predicting outcome on the basis of clinical covariates and the genes identified by the t test with the adjusted P value (or FDR). Prediction accuracies of these models were estimated by using an unbiased, leave-one-out cross validation (LOOCV).

Three independent microarray data sets of childhood B-precursor ALL were used for validation of the outcome signature: a set of 220 patients treated on Pediatric Oncology Group (POG) trials,[13] 145 patients treated on German Cooperative Study Group for Childhood ALL (COALL) protocols,[14] and 92 patients from Dutch Childhood Oncology Group (DCOG) protocols.[15] The samples of the POG, COALL, and DCOG data sets were hybridized to Affymetrix U95Av2, U133A, and U133Plus2.0 arrays, respectively. Logistic regression was used to determine the association of the significant probe sets in the POG data set, and Cox regression was used in the COALL and DCOG data sets. We next built models for outcome prediction. Of the 47 probe sets identified in the COG 1961 data set, 18 could be matched by 20 probe sets of the U95Av2 microarrays. We constructed logistic regression models with these 20 probe sets and with three clinical covariates (sex, age, and WBC) as predictors of outcome. Briefly, model I (LP1) was based on three genes, model II (LP2) on five genes, and model III (LP3) on four genes. Receiver operating characteristic (ROC) accuracy, t test, Mann-Whitney U test, Cox proportional hazards regression, and logistic regression were used to validate these predictive models on the independent data sets. In addition to the three statistical models mentioned above, we considered a simple linear combination of the expression values of the probe sets that match the 47 probe sets in each of the three validation cohorts (LPV).

RESULTS

Prediction of Early Response

Analysis with PAM on the training set (n = 54) led to a model comprised of 24 probe sets with a minimal average cross validated

Fig 1. Receiver operating characteristic (ROC) curve of the predicted score of early response on the test set. The model that comprised 24 probe sets that were derived from the training set was used for the prediction of early response on the test set. The ROC accuracy (ie, the area under the curve) is A = 0.77, which is significantly larger than that of noninformative prediction (P = .006). By using a threshold of C = 0.4 (determined in training set), the model correctly predicted 21 of 28 patient cases (success rate, 0.75). RER, rapid early responder; SER, slow early responder.

Table 2. Enrichment Analysis of Genes Associated With Early Response

GeneOntology Term	GeneOntology Number	P
Hemocyte development	GO:0007516	.00007
Vesicle targeting	GO:0006903	< .001
Intercellular junction assembly	GO:0007043	.001
Induction of apoptosis	GO:0006917	.003
Cytoplasm organization and biogenesis	GO:0007028	.004
Hemopoiesis	GO:0030097	.005
Membrane fusion	GO:0006944	.005
Hemopoietic or lymphoid organ development	GO:0048534	.005
Nucleotide catabolism	GO:0009166	.00004
Growth	GO:0040007	< .001
Hormone-mediated signaling	GO:0009755	< .001
Cellular morphogenesis	GO:0000902	.002
Leukotriene biosynthesis	GO:0019370	.002
Regulation of neuron differentiation	GO:0045664	.003
Alkene biosynthesis	GO:0043450	.004
Nucleosome assembly	GO:0006334	.006
Cell growth	GO:0016049	.007
Alkene metabolism	GO:0043449	.007
Chromatin assembly	GO:0031497	.009
Phospholipase C activation	GO:0007202	.009

NOTE. The pathways upregulated in rapid early responders are hemocyte development through hematopoietic or lymphoid organ development; those upregulated in slow early responders are nucleotide catabolism through phospholipase C activation.

	t Test			Likelihood Ratio Test (P)				
Rank	P	FDR	Response Associated With High Expression	Significance of Gene	Significance of Clinical Variables	Affymextrix Probe Set Identification	Gene Name	Gene Description
1	< .001	.013	CCR	< .001	< .001	35666_at	SEMA3F	SEMA domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphoring) 3F
2	< .001	.015	Fail	< .001	< .001	227877_at		Similar to annexin II receptor (LOC389289), mRNA
3	< .001	.015	CCR	< .001	< .001	227131_at	MAP3K3	Mitogen-activated protein kinase kinase kinase 3
4	< .001	.015	Fail	.004	.001	205401_at	AGPS	Alkylglycerone phosphate synthase
5	< .001	.028	Fail	.001	< .001	208687_x_at	HSPA8	Heat shock 70 kDa protein 8
6	< .001	.028	CCR	.002	< .001	212229_s_at	FBXO21	F-box only protein 21
7	< .001	.028	CCR	.001	< .001	212576_at	MGRN1	Mahogunin, ring finger 1
8	< .001	.028	CCR	.004	< .001	225446_at	C21orf107	Chromosome 21 open reading frame 107
9	< .001	.031	CCR	.004	< .001	224793_s_at	TGFBR1	Transforming growth factor, β receptor I (activin A receptor type II–like kinase, 53 kDa)
10	< .001	.033	CCR	.001	< .001	221840_at	PTPRE	Protein tyrosine phosphatase, receptor type, E
11	< .001	.033	CCR	< .001	< .001	203514_at	MAP3K3	Mitogen-activated protein kinase kinase kinase 3
12	< .001	.034	CCR	.003	< .001	1559018_at	PTPRE	Protein tyrosine phosphatase, receptor type, E
13	< .001	.034	Fail	< .001	< .001	217499_x_at	OR7E47P	Olfactory receptor, family 7, subfamily E, member 47 pseudogene
14	< .001	.034	Fail	.007	< .001	224187_x_at	HSPA8	Heat shock 70 kDa protein 8
15	< .001	.034	Fail	.007	< .001	221891_x_at	HSPA8	Heat shock 70 kDa protein 8
16	< .001	.034	CCR	< .001	< .001	201642_at	IFNGR2	Interferon γ receptor 2 (interferon gamma transducer 1)
17	< .001	.034	CCR	.002	< .001	218418_s_at	ANKRD25	Ankyrin repeat domain 25
18	< .001	.034	Fail	< .001	< .001	242305_at		cDNA FLJ42757 fis, clone BRAWH3001712
19	< .001	.034	CCR	.002	< .001	216035_x_at	TCF7L2	Transcription factor 7-like 2 (T-cell specific, high mobility group box)
20	< .001	.034	CCR	.001	< .001	1556321_a_at		MRNA full length insert cDNA clone EUROIMAGE 283668
21	< .001	.034	Fail	< .001	< .001	235014_at	LOC147727	Hypothetical protein LOC147727
22	< .001	.034	CCR	.002	< .001	208820_at		PTK2 protein tyrosine kinase 2
23	< .001	.034	CCR	.004	< .001	212231_at	FBXO21	F-box only protein 21
24	< .001	.034	CCR	.004	< .001	229618_at	SNX16	Sorting nexin 16
25	< .001	.034	CCR	.005	< .001	209033_s_at	DYRK1A	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1A
26	< .001	.034	CCR	< .001	< .001	200641_s_at	YWHAZ	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide
27	< .001	.034	Fail	.015	< .001	202657_s_at	SERTAD2	SERTA domain containing 2
28	< .001	.034	CCR	.015	< .001	201099_at	USP9X	Ubiquitin specific protease 9, X-linked (fat facets-like, Drosophila)
29	< .001	.034	CCR	.004	< .001	201542_at	SARA1	SAR1a gene homolog 1 (S. cerevisiae)
30	< .001	.034	CCR	.005	< .001	227068_at	PGK1	Phosphoglycerate kinase 1
31	< .001	.037	CCR	< .001	< .001	213944_x_at	GNA11	Guanine nucleotide binding protein (G protein), α-11 (Gq class)
32	< .001	.039	CCR	.015	< .001	201472_at	VBP1	von Hippel-Lindau binding protein 1
33	< .001	.039	CCR	.018	< .001	202806_at	DBN1	Drebrin 1
34	< .001	.039	CCR	.022	< .001	221918_at	PCTK2	PCTAIRE protein kinase 2
35	< .001	.039	Fail	< .001	< .001	214585_s_at	VPS52	Vacuolar protein sorting 52 (yeast)
36	< .001	.039	Fail	< .001	< .001	219078_at	GPATC2	G patch domain containing 2
37	< .001	.039	Fail	< .001	< .001	219133_at	FLJ20604	Hypothetical protein FLJ20604
38	< .001	.040	Fail	< .001	< .001	1558111_at	MBNL1	Muscleblind–like (Drosophila)
39	< .001	.040	CCR	.001	< .001	221773_at		ELK3, ETS-domain protein (SRF accessory protein 2)
40	< .001	.040	CCR	.003	< .001	1558732_at	MAP4K2	mitogen-activated protein kinase kinase kinase kinase 4
41	< .001	.043	CCR	.032	< .001	212441_at	KIAA0232	KIAA0232 gene product
42	< .001	.045	CCR	.015	< .001	226775_at	e(y)2	e(y)2 protein
43	< .001	.045	Fail	.001	< .001	208498_s_at	AMY2B	Amylase, α 2B; pancreatic
44	< .001	.049	CCR	.016	< .001	201121_s_at	PGRMC1	Progesterone receptor membrane component 1
45	< .001	.049	CCR	.004	< .001	202984_s_at	BAG5	BCL2-associated athanogene 5
46	< .001	.049	Fail	.032	< .001	210338_s_at	HSPA8	Heat shock 70 kDa protein 8
47	< .001	.049	CCR	.002	< .001	206548_at	FLJ23556	Hypothetical protein FLJ23556

Table 3. Probe Sets Differentially Expressed Between Patients in CCR and Those Who Experienced Relapse

Abbreviations: CCR, complete continuous remission; fail, relapse.

error rate of 0.38 that best characterized early response (FDR, 3.6%). The Affymetrix probe set identifications and gene descriptions in rank order can be found in Table 1.

To validate the significance of the 24 probe sets, we performed t test and logistic regression analyses on the expression values in the test data set (n = 28). Although there was a positive trend of association between all probes in the test and training sets, eight reached statistical significance. The estimated ROC accuracy of the predicted score on the test set was 0.7755 (P = .0061; Fig 1). The overall misclassification rate was 0.25 (sensitivity = .7143 and specificity = .7857). The observed and predicted early responses significantly correlated with each other (odds ratio, 8.33; P (one-sided Fisher's exact test) = .011).

Functional Analysis of Genes Related to Early Response

A list of 188 differentially expressed probe sets (RER v SER) that were selected by PAM on the entire data set (N = 82; FDR ≤ 10%) was used for the detection of the relative enrichment of genes according to GeneOntology[15a] terms with the help of the L2L tool.[16] Genes significantly over-represented in RER patients included those involved with induction of apoptosis and hematopoeitic development, whereas genes involved with cell growth and metabolism were over-represented in SER patients (Table 2).

Prediction of Long-Term Outcome

Gene expression profiles from 59 patients (28 CCR; 31 relapse) were analyzed to identify genes related to long-term outcome. By using a threshold FDR of 5%, we identified 47 probe sets (which represented 41 unique genes) that were significantly associated with outcome. The Affymetrix identifications, which are descriptions for the genes, are listed in rank order in Table 3 with t test P values that compare CCR and relapse on the selected genes. Figure 2 represents the heatmap of the expression values.

To get an unbiased estimate for the prediction accuracy of each of the three models (LP1 through 3), we performed LOOCV. The misclassification rates for the three models were 0.2542, 0.3051, and 0.2881, respectively (sensitivity = 0.643, 0.642, and 0.643, respectively; specificity = 0.839, 0.742, and 0.774, respectively). The ROC accuracies were 0.8065, 0.7154, and 0.7316 (P < .0001, < .002, and < .001, respectively). These LOOCV results indicated that the three models were significantly predictive of outcome.

Validation of Outcome Prediction Models on Independent Patient Cohorts

Three large patient cohorts—POG, COALL, and DCOG—were used as independent sets for validation of the 47–probe set signature. Notably, the trend in the DCOG and COALL data sets of the association of the matched probe sets all agree with that observed in the 1961 data set, and this was also true of the POG data set with three exceptions (Appendix Table A2, online only).

The POG data set consisted of 220 patient cases (4-year CCR, n = 95; relapse, n = 125) of childhood B-precursor ALL. The estimated ROC accuracies for the three prediction models were 0.6119, 0.5820, and 0.5674, respectively. By using the one-sided Mann-Whitney U test, the P values were .00226, .0187, and .0436, respectively, which indicated that each of the predicted LP values of the three models were significantly predictive of outcome in the independent POG set. To further validate the predictive value of the three models,

we fit the univariate and multivariate logistic regression models (Table 4). The LPs of all the three models were significantly associated with outcome (P < .05 for all). However, we did not find statistical evidence for the prognostic significance of the majority of the models when analysis was adjusted for age and WBC or for karyotype. Only model I retained prognostic significance when age, WBC, and karyotype were considered. Models II and III were significant after analysis was adjusted for karyotype but not for age and WBC. Similar results were obtained when only the HR subset of patients was analyzed (data not shown). Logistic regression with LPV (ie, the weighted sum of expression values of 20 probe sets common in the COG 1961 and POG datasets) as the explanatory variable indicated that LPV also was associated with a good outcome; P (one-tailed Wald test) = .007.

We next validated the three prediction models by using COALL data with Cox proportional hazards regression (Table 5). We again noted that the predicted LP values of all three models were significantly associated with outcome (P < .05), and they remained significant after analysis was adjusted for age and WBC but not for karyotype. Cox regression with LPV was significantly associated with outcome (P = .0002). The DCOG data set comprised of 92 (4-year CCR, n = 67; relapse, n = 25) B-lineage diagnostic samples. By using the one-sided Wilcoxon rank sum test, the P values were .030, .020, and .0635, respectively, which provided a significant or marginal association with outcome. Cox PH regression was performed to additionally validate the association of the predicted values with outcome (Table 6). We noted again that the hazard ratios were all less than 1, which indicated a consistent trend that the high predicted values were associated with good outcome. In the DCOG data set, the three models were statistically significant when considered on their own (univariate analysis) but were not after analysis was adjusted for WBC, age, and karyotype. Logistic regression yielded similar results (Appendix Table A3, online only). LPV with all 47 probe sets was significant (P = .02, Appendix Table A5, online only).

DISCUSSION

The goal of our study was to identify gene expression signatures in diagnostic samples that are predictive of early response to therapy and overall outcome in children with National Cancer Institute–defined HR ALL. All samples studied in these experiments were from patients who were treated on a single, contemporary protocol and who received intensified therapy according to a COG-modified Berlin-Frankfurt-Munster backbone, which thus minimized the effects of treatment variables.

Early response to therapy has proven to be one of the strongest predictors of outcome and now is routinely used to stratify patients according to the risk of relapse.[17] We were able to identify and validate a gene expression signature that correlated with the kinetics of regression of tumor burden, as assessed by the bone marrow blast content on day 7. Apoptosis-facilitating genes, such as *BIM* and *PDCD6*, were upregulated in RER patients, whereas multiple genes involved in cell adhesion (eg, *GPR56, PARVG*), cell proliferation (eg, *CKLF, BMP2*), and antiapoptosis (eg, *BCL2, SOCS2*) were upregulated in SER patients. If this signature is validated with additional research, more rapid approaches to assessment of gene expression could be used so that augmented therapy might be deployed early—within the first few days of diagnosis—to overcome slow response and possibly the emergence of drug-resistant clones and, ultimately, to improve outcome.
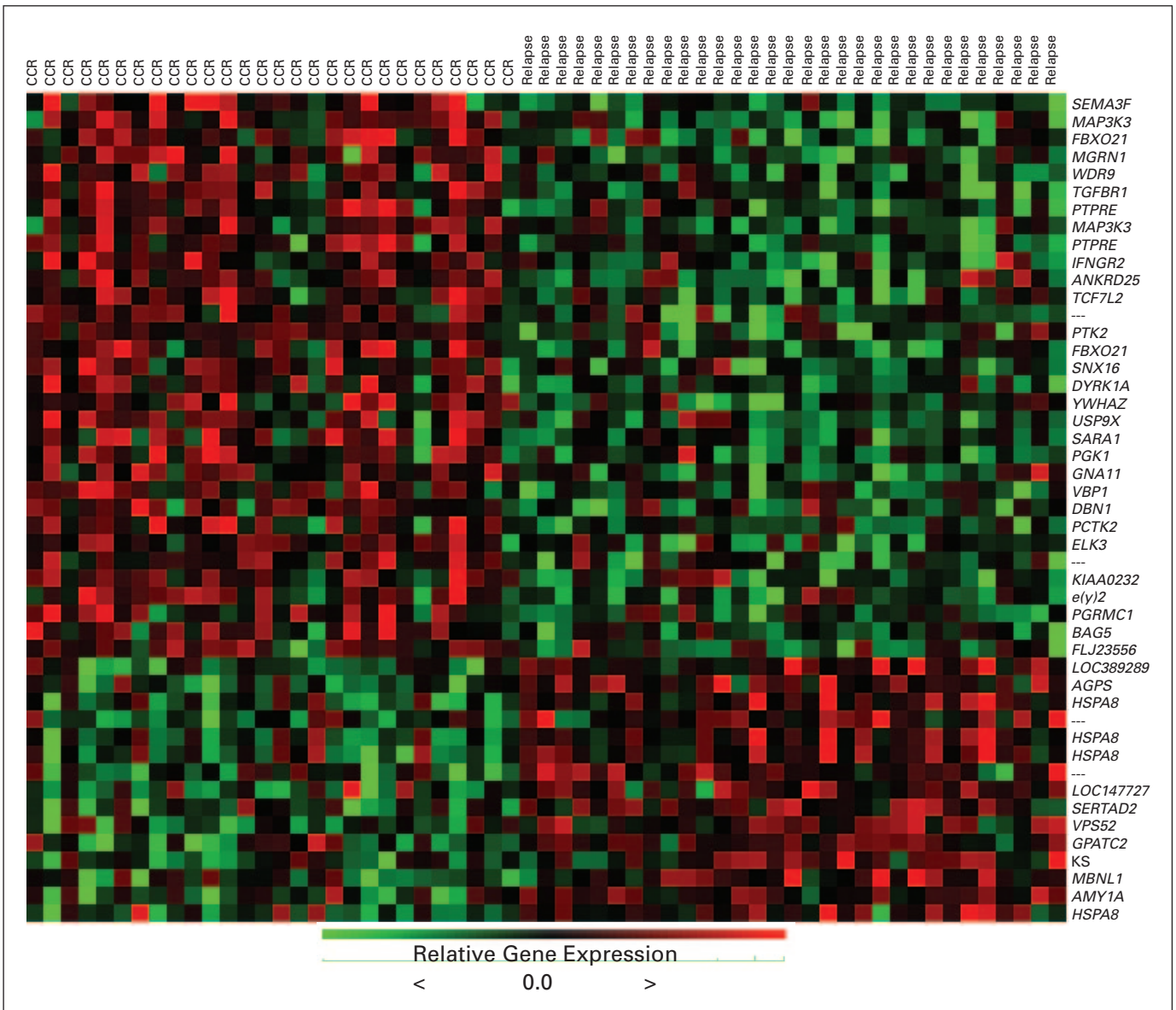
Fig 2. Genes differentially expressed in patients that remained in complete continuous remission *v* in those that relapsed. Heatmap of the 47–probe set signature that was predictive of outcome (which represented 41 unique genes).

Other investigators also have sought to identify gene expression profiles associated with early response to therapy. Two recent publications from Flotho et al[18,19] have portrayed signatures that correlated with minimal residual disease at day 19[18] and at day 46[19] of induction. Though only five of 44 probe sets from the day 19 signature reached statistical significance in our data set of day 7 response, the trend of association for all the probe sets was remarkably strong. Not surprisingly, this trend was not observed with the day 46 signature (data not shown). Previous studies show that the kinetics of blast reduction is quite steep in the first 2 weeks of induction and is much slower thereafter.[20] Thus, although day 7 bone marrow morphology and end induction minimal residual disease may correlate,[21] it is likely that fundamental differences exist in the mechanisms of leukemia cell death that occurs in early compared with late induction.

Though various groups have performed microarray experiments on childhood ALL samples, it has proved difficult to identify a prog-

nostic signature at diagnosis. For example, Yeoh et al[7] were able to detect distinct expression profiles that predicted relapse in T-cell acute lymphoblastic leukemia and hyperdiploid ALL but not in other subtypes.[7] Although expression of *OPAL1* predicts ALL in some studies, it has not been validated in others, which suggests that differences in treatment may influence the prognostic impact of expression profiles.[22] Other investigators have correlated gene expression signatures with in vitro drug response.[14,23] However, this drug resistance profile was not selected for its prognostic value and, hence, may not represent the best selection of outcome-predictive genes. Despite these challenges, we have identified a gene expression signature that was predictive of long-term outcome and was validated in three independent cohorts of diagnostic samples from children who were treated on different protocols, which thus yielded an accurate perspective on the validity and reproducibility of the results.

Table 4. Validation of the Outcome Signature on POG Data Set

| | Analyses | | | | | |
| | Univariate | | Multivariate Adjusted for Age and WBC | | Multivariate Adjusted for ALL Subtype* | |
Model	Odds Ratio	P	Odds Ratio	P	Odds Ratio	P
I (LP1)	1.468	.004	1.311	.035	1.314	.033
II (LP2)	1.363	.016	1.242	.070	1.346	.026
III (LP3)	1.312	.029	1.202	.105	1.281	.048

NOTE. Total number of patients in data set = 220. Data were analyzed with the U95Av2 microarray.
Abbreviations: POG, Pediatric Oncology Group; ALL, acute lymphoblastic leukemia.
*Analysis was adjusted for TEL/AML1.

Almost all of the genes that comprised our predictive signature were not identified in the studies mentioned above that looked at drug resistance and/or outcome. However, studies that have used microarray methodology to discover predictive signatures in other cancers also have shown little overlap in gene lists. Although these gene lists may not always be concordant between data sets, each signature still may be significantly predictive across the data sets. For example, five recently published predictive gene sets for outcome in breast cancer showed little overlap between sets.[24] However, four of five were predictive of outcome in a single data set of 295 women, which emphasizes that, despite the lack of overlap, the signatures are reflective of common biologic subsets. This is consistent with our findings that demonstrated the ability of individual gene expression signatures and the derived models by using the COG samples to predict outcome on three different cohorts of patients.

The utilization of predictive signatures in clinical cancer trials is eagerly awaited. The application of array technology to define additional patients with ALL who have a poor outcome may be more difficult given the high cure rate of ALL and the elucidation of many well-established risk factors to date. One of the most crucial findings of our study was that, although gene expression signatures correlated with outcome in univariate analyses in multiple data sets, they lost much significance when well-known outcome predictors, like age, initial WBC, and genotype, were taken into account. A logical interpretation of these findings is that the most important variables associated with treatment failure in ALL have been identified already. However, the inability to accurately predict outcome uniformly

by using these conventional variables may be related in many instances to host factors. In addition, measurements of gene expression do not take into account important events, such as post-translational modifications. Another explanation is that prognostic signatures may exist within biologic subtypes of ALL only. It has been established that gene expression profiles correlate with ALL cohorts defined by molecular changes, such as translocations and ploidy. We specifically focused our efforts on National Cancer Institute–defined HR ALL, because known genetic subtypes account for only a minority of patients in this cohort, and we sought to identify novel biologic subtypes associated with outcome by using gene expression profiling. Our inability to define such a group might reflect the existence of smaller biologic subsets within this population that may not be possible to detect with the number of patient cases studied here. However, our study and similar ones by others, even if not predictive in multivariate analysis, are likely to lead to a biologic understanding of why certain clinical and laboratory variables are associated with clinical outcome. Such information is essential to derive more effective, tumor-specific therapies.

In summary, we have identified a gene expression signature that is significantly predictive of outcome in childhood ALL, but it does not seem to provide additional information beyond that contained in already established prognostic variables. The analysis of a larger number of samples may allow investigators to discover gene signatures that provide additional prognostic information. Strict adherence to uniform protocols for sample acquisition, processing, and array experimentation may facilitate comparison between data sets.[25] In addition, analysis of gene expression profiles

Table 5. Validation of the Outcome Signature on COALL Data Set

| | Analyses | | | | | |
| | Univariate | | Multivariate Adjusted for Age and WBC | | Multivariate Adjusted for ALL Subtype* | |
Model	Hazard Ratio	P	Hazard Ratio	P	Hazard Ratio	P
I (LP1)	0.898	.015	0.917	.041	0.934	.135
II (LP2)	0.993	.009	0.994	.017	0.997	.17
III (LP3)	0.961	.020	0.967	.036	0.983	.23

NOTE. Total number of patients in data set = 145. Data were analyzed with the U133A microarray.
Abbreviations: COALL, German Cooperative Study Group for Childhood Acute Lymphoblastic Leukemia; ALL, acute lymphoblastic leukemia; MLL, mixed lineage leukemia; BCR, break point cluster region; ABL, Abelson murine leukemia viral oncogene; TEL, translocation ETS leukemia, AML, acute myeloid leukemia.
*Analysis was adjusted for MLL, BCRABL, TEL/AML1, hyperdiploid, and E2A subtypes.

Table 6. Validation of the Outcome Signature on DCOG Data Set

	Analyses					
	Univariate		Multivariate Adjusted for Age and WBC		Multivariate Adjusted for ALL Subtype*	
Model	Hazard ratio	P	Hazard ratio	P	Hazard ratio	P
I (LP1)	0.836	.010	0.919	.155	0.981	.415
II (LP2)	0.987	.016	0.991	.090	0.991	.100
III (LP3)	0.938	.047	0.968	.215	0.981	.350

NOTE. Total number of patients in data set = 92. Data were analyzed with the U133Plus2.0 microarray.
Abbreviations: COALL, German Cooperative Study Group for childhood acute lymphoblastic leukemia; ALL, acute lymphoblastic leukemia; *MLL*, mixed lineage leukemia; *BCR*, break point cluster region; *ABL*, Abelson murine leukemia viral oncogene; *TEL*, translocation ETS leukemia; *AML*, acute myeloid leukemia.
*Analysis was adjusted for *MLL, BCRABL, TEL/AML1*, hyperdiploid, and *E2A* subtypes.

may lead to a biologic understanding of why clinical and laboratory variables are associated with outcome, and this information potentially may be exploited therapeutically.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The authors indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: William L. Carroll, Stephen P. Hunger
Financial support: William L. Carroll, Cheryl L. Willman

Administrative support: Harland Sather, Stephen P. Hunger, William L. Carroll
Provision of study materials or patients: Stephen P. Hunger, Nita Seibel, Rob Pieters, Monique L. den Boer, Martin A. Horstmann, Cheryl L. Willman
Collection and assembly of data: Deepa Bhojwani, Huining Kang, Harland Sather, Wenjian Yang, Monique L. den Boer, Renee X. Menezes, Jeffrey W. Potter
Data analysis and interpretation: Deepa Bhojwani, Huining Kang, Monique L. den Boer, Renee X. Menezes, Wenjian Yang, Naomi P. Moskowitz, Dong-Joon Min, Richard Harvey
Manuscript writing: Deepa Bhojwani, Huining Kang, Monique L. den Boer, Elizabeth A. Raetz, Mary V. Relling, Stephen P. Hunger, William L. Carroll
Final approval of manuscript: Deepa Bhojwani, Huining Kang, Stephen P. Hunger, Monique L. den Boer, Rob Pieters, Mary V. Relling, Cheryl L. Willman, William L. Carroll

REFERENCES

1. Pui CH, Evans WE: Treatment of acute lymphoblastic leukemia. N Engl J Med 354:166-178, 2006

2. Schultz KR, Pullen DJ, Sather HN, et al: Risk and response-based classification of childhood B-precursor acute lymphoblastic leukemia: A combined analysis of prognostic markers from the Pediatric Oncology Group (POG) and Children's Cancer Group (CCG). Blood 109:926-935, 2007

3. Gaynon PS: Childhood acute lymphoblastic leukaemia and relapse. Br J Haematol 131:579-587, 2005

4. Alizadeh AA, Eisen MB, Davis RE, et al: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503-511, 2000

5. Hayes DN, Monti S, Parmigiani G, et al: Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. J Clin Oncol 24:5079-5090, 2006

6. Golub TR, Slonim DK, Tamayo P, et al: Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 286:531-537, 1999

7. Yeoh EJ, Ross ME, Shurtleff SA, et al: Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell 1:133-143, 2002

8. Siebel NL, Steinherz PG, Sather HN, et al: Early postinduction intensification therapy improves

surivival for children and adolescents with high-risk acute lymphoblastic leukemia: A report from the Children's Oncology Group. Blood 111:2548-2555, 2008

9. Helman P, Veroff R, Atlas SR, et al: A Bayesian network classification methodology for gene expression data. J Comput Biol 11:581-615, 2004

10. Tibshirani R, Hastie T, Narasimhan B, et al: Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A 99:6567-6572, 2002

11. Benjamini Y, Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Statist Soc B 57:289-300, 1995

12. Hosmer D, Lemeshow S: Applied Logistic Regression (ed 2). Hoboken, NJ, John Wiley and Sons Inc, 2000

13. Martin SB, Mosquera-Caro MP, Potter JW, et al: Gene expression overlap affects karyotype prediction in pediatric acute lymphoblastic leukemia. Leukemia 21:1341-1344, 2007

14. Holleman A, Cheok MH, den Boer ML, et al: Gene-expression patterns in drug-resistant acute lymphoblastic leukemia cells and response to treatment. N Engl J Med 351:533-542, 2004

15. Kamps WA, Bokkerink JP, Hakvoort-Cammel FG, et al: BFM-oriented treatment for children with acute lymphoblastic leukemia without cranial irradiation and treatment reduction for standard risk patients: Results of DCLSG protocol ALL-8 (1991-1996). Leukemia 16:1099-1111, 2002

15a. Ashburner M, Ball CA, Blake JA, et al: Gene ontology: Tool for the unification of biology—The

Gene Ontology Consortium. Nature Genetics 25:25-29, 2000

16. Newman JC, Weiner AM: L2L: A simple tool for discovering the hidden significance in microarray expression data. Genome Biol 6:R81, 2005

17. Nachman JB, Sather HN, Sensel MG, et al: Augmented post-induction therapy for children with high-risk acute lymphoblastic leukemia and a slow response to initial therapy. N Engl J Med 338:1663-1671, 1998

18. Flotho C, Coustan-Smith E, Pei D, et al: A set of genes that regulate cell proliferation predicts treatment outcome in childhood acute lymphoblastic leukemia. Blood 110:1271-1277, 2007

19. Flotho C, Coustan-Smith E, Pei D, et al: Genes contributing to minimal residual disease in childhood acute lymphoblastic leukemia: Prognostic significance of CASP8AP2. Blood 108:1050-1057, 2006

20. Brisco MJ, Sykes PJ, Dolman G, et al: Early resistance to therapy during induction in childhood acute lymphoblastic leukemia. Cancer Res 60:5092-5096, 2000

21. Borowitz MJ, Pullen DJ, Shuster JJ, et al: Minimal residual disease detection in childhood precursor-B-cell acute lymphoblastic leukemia: Relation to other risk factors—A Children's Oncology Group study. Leukemia 17:1566-1572, 2003

22. Holleman A, den Boer ML, Cheok MH, et al: Expression of the outcome predictor in acute leukemia 1 (OPAL1) gene is not an independent prognostic factor in patients treated according to COALL or St Jude protocols. Blood 108:1984-1990, 2006

23. Lugthart S, Cheok MH, den Boer ML, et al: Identification of genes associated with chemotherapy cross resistance and treatment response in childhood acute lymphoblastic leukemia. Cancer Cell 7:375-386, 2005

24. Fan C, Oh DS, Wessels L, et al: Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 355:560-569, 2006

25. Staal FJ, Cario G, Cazzaniga G, et al: Consensus guidelines for microarray gene expression analyses in leukemia from three European leukemia networks. Leukemia 20:1385-1392, 2006

■ ■ ■

Appendix

The Appendix is included in the full-text version of this article, available online at www.jco.org. It is not included in the PDF version (via Adobe® Reader®).