

# The New England Journal of Medicine

Copyright © 2002 by the Massachusetts Medical Society

VOLUME 346

JUNE 20, 2002

NUMBER 25



## THE USE OF MOLECULAR PROFILING TO PREDICT SURVIVAL AFTER CHEMOTHERAPY FOR DIFFUSE LARGE-B-CELL LYMPHOMA

ANDREAS ROSENWALD, M.D., GEORGE WRIGHT, PH.D., WING C. CHAN, M.D., JOSEPH M. CONNORS, M.D.,  
ELIAS CAMPO, M.D., RICHARD I. FISHER, M.D., RANDY D. GASCOYNE, M.D., H. KONRAD MULLER-HERMELINK, M.D.,  
ERLEND B. SMELAND, M.D., PH.D., AND LOUIS M. STAUDT, M.D., PH.D.,  
FOR THE LYMPHOMA/LEUKEMIA MOLECULAR PROFILING PROJECT

### ABSTRACT

**Background** The survival of patients with diffuse large-B-cell lymphoma after chemotherapy is influenced by molecular features of the tumors. We used the gene-expression profiles of these lymphomas to develop a molecular predictor of survival.

**Methods** Biopsy samples of diffuse large-B-cell lymphoma from 240 patients were examined for gene expression with the use of DNA microarrays and analyzed for genomic abnormalities. Subgroups with distinctive gene-expression profiles were defined on the basis of hierarchical clustering. A molecular predictor of risk was constructed with the use of genes with expression patterns that were associated with survival in a preliminary group of 160 patients and was then tested in a validation group of 80 patients. The accuracy of this predictor was compared with that of the international prognostic index.

**Results** Three gene-expression subgroups — germinal-center B-cell–like, activated B-cell–like, and type 3 diffuse large-B-cell lymphoma — were identified. Two common oncogenic events in diffuse large-B-cell lymphoma, *bcl-2* translocation and *c-rel* amplification, were detected only in the germinal-center B-cell–like subgroup. Patients in this subgroup had the highest five-year survival rate. To identify other molecular determinants of outcome, we searched for individual genes with expression patterns that correlated with survival in the preliminary group of patients. Most of these genes fell within four gene-expression signatures characteristic of germinal-center B cells, proliferating cells, reactive stromal and immune cells in the lymph node, or major-histocompatibility-complex class II complex. We used 17 genes to construct a predictor of overall survival after chemotherapy. This gene-based predictor and the international prognostic index were independent prognostic indicators.

**Conclusions** DNA microarrays can be used to formulate a molecular predictor of survival after chemotherapy for diffuse large-B-cell lymphoma. (N Engl J Med 2002;346:1937-47.)

Copyright © 2002 Massachusetts Medical Society.

**D**IFFUSE large-B-cell lymphoma, the most common type of lymphoma in adults, can be cured by anthracycline-based chemotherapy in only 35 to 40 percent of patients.<sup>1</sup> The multiple unsuccessful attempts to increase this rate<sup>2</sup> suggest that diffuse large-B-cell lymphoma actually comprises several diseases that differ in responsiveness to chemotherapy. Support for this idea comes from a study of gene-expression profiles, which identified two subgroups of diffuse large-B-cell lymphoma that had different outcomes after multiagent chemotherapy.<sup>3</sup> The germinal-center B-cell–like subgroup expressed genes characteristic of normal germinal-center

From the Metabolism Branch, Center for Cancer Research (A.R., L.M.S.), and the Biometric Research Branch, Division of Cancer Treatment and Diagnosis (G.W.), National Cancer Institute, National Institutes of Health, Bethesda, Md.; the Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha (W.C.C.); British Columbia Cancer Center, Vancouver, B.C., Canada (J.M.C., R.D.G.); Hospital Clinic, University of Barcelona, Barcelona, Spain (E.C.); Southwest Oncology Group and James P. Wilmot Cancer Center, University of Rochester School of Medicine, Rochester, N.Y. (R.I.F.); the Department of Pathology, University of Würzburg, Würzburg, Germany (H.K.M.-H.); and the Department of Immunology, Norwegian Radium Hospital, Oslo, Norway (E.B.S.). Address reprint requests to Dr. Staudt at the Metabolism Branch, Center for Cancer Research, National Cancer Institute, Bldg. 10, Rm. 4N114, National Institutes of Health, Bethesda, MD 20892, or at lstaudt@mail.nih.gov.

Other authors were Jena M. Giltman, B.S., Elaine M. Hurt, Ph.D., Hong Zhao, M.S., Lauren Averett, B.A., Liming Yang, Ph.D., Wyndham H. Wilson, M.D., Ph.D., Elaine S. Jaffe, M.D., Richard Simon, D.Sc., and Richard D. Klausner, M.D., National Cancer Institute, National Institutes of Health (NIH), Bethesda, Md.; John Powell, M.S., Center for Information Technology, NIH, Bethesda, Md.; Patricia L. Duffey, R.N., and Dan L. Longo, M.D., National Institute on Aging, NIH Gerontology Research Center, Baltimore; Timothy C. Greiner, M.D., Dennis D. Weisenburger, M.D., Warren G. Sanger, Ph.D., Bhavana J. Dave, Ph.D., James C. Lynch, Ph.D., Julie Vose, M.D., and James O. Armitage, M.D., University of Nebraska Medical Center, Omaha; Emilio Montserrat, M.D., and Armando López-Guillermo, M.D., Hospital Clinic, University of Barcelona, Barcelona, Spain; Thomas M. Grogan, M.D., and Thomas P. Miller, M.D., Southwest Oncology Group and University of Arizona Cancer Center, Tucson; Michel LeBlanc, Ph.D., Southwest Oncology Group and Fred Hutchinson Cancer Research Center, Seattle; German Ott, M.D., University of Würzburg, Würzburg, Germany; and Stein Kvaloy, M.D., Ph.D., Jan Delabie, M.D., Ph.D., Harald Holte, M.D., Ph.D., Peter Krajci, M.D., Ph.D., and Trond Stokke, Ph.D., Norwegian Radium Hospital, Oslo, Norway.

ter B cells and were associated with a good outcome, whereas the activated B-cell–like subgroup expressed genes characteristic of activated blood B cells and were associated with a poor outcome.

Age, Eastern Cooperative Oncology Group (ECOG) performance status, tumor stage, lactate dehydrogenase level, and the number of sites of extranodal disease also have prognostic value in diffuse large-B-cell lymphoma, and they are included in the international prognostic index.<sup>4</sup> Although the index is of value, it has not been used successfully to stratify patients for therapeutic trials.<sup>5</sup>

We hypothesized that gene-expression profiles of diffuse large-B-cell lymphoma could be used independently of the international prognostic index to predict the outcome of chemotherapy. Since the outcomes for patients within the same subgroup of diffuse large-B-cell lymphoma vary,<sup>3</sup> we sought to identify individual genes that could influence survival within these subgroups.

## METHODS

Tumor-biopsy specimens and clinical data were obtained retrospectively from 240 patients with untreated diffuse large-B-cell lymphoma who had no previous history of lymphoma, according to a protocol approved by the National Cancer Institute institutional review board. A panel of eight hemopathologists confirmed the diagnosis of diffuse large-B-cell lymphoma in all patients and were able to assign a histologic subtype to 236. Patients were selected on the basis of the availability of tumor-biopsy specimens, without regard to the clinical outcome. All patients had received anthracycline-based chemotherapy, in most cases a regimen of cyclophosphamide, doxorubicin, vincristine, and prednisone or similar regimens. Median follow-up was 2.8 years overall (7.3 years for survivors), and 57 percent of patients died during this period. The median age of the patients was 63 years, and 56 percent were men. According to the Ann Arbor classification, 15 percent of patients had stage I disease, 31 percent had stage II, 20 percent had stage III, and 34 percent had stage IV.

“Lymphochip” DNA microarrays<sup>6</sup> are composed of genes whose products are preferentially expressed in lymphoid cells and genes thought or confirmed to play a part in cancer or immune function. These microarrays were constructed from 12,196 clones of complementary DNA (i.e., microarray features) and were used to quantify the expression of messenger RNA in the tumors.<sup>3</sup> Established procedures were followed to detect the amplification of *c-rel*<sup>7</sup> and the translocation of *bcl-2*.<sup>8</sup>

Hierarchical clustering<sup>9</sup> was used to define subgroups of diffuse large-B-cell lymphoma. To create an outcome variable based on gene-expression studies, we divided the patients into two groups: the preliminary group comprised 160 patients, and the validation group comprised 80 patients. Within the preliminary group, the significance of the correlation between outcome (overall survival after chemotherapy) and gene-expression data from individual microarray features was determined with use of the Cox proportional-hazards model. Genes that were associated with a good or a bad outcome at a significance level of  $P < 0.01$  were assigned to gene-expression signatures, as described previously.<sup>10</sup> These representative genes were chosen because of their high variance in expression (i.e., they were among the top third of genes in variance of gene-expression levels in the preliminary group). We used the average value for each signature and the value for the *BMP6* gene, a member of the transforming growth factor  $\beta$  superfamily of genes, to construct a multivariate Cox survival model according to the following formula: outcome-predictor score =  $(0.241 \times \text{the average value of the proliferation signature}) + (0.310 \times \text{value for } BMP6) - (0.290 \times \text{the average value of the germinal-center B-cell signature}) - (0.311 \times \text{the average value of the major-histocompatibility-complex [MHC] class II signature}) - (0.249 \times \text{the average value of the lymph-node signature})$ . The coefficients in this formula were derived from the Cox model, and a high score indicates a poor outcome. We used the Wald test to determine the significance of the association between this model and the outcome in the preliminary group, the validation group, and the group as a whole and to assess the independence of the risk groups defined by the international prognostic index and the outcome predicated on gene-expression profiles. All t-tests were two-sided except those used for the validation group; the results of one-sided t-tests are reported for this group, since the analysis of the preliminary group indicated the direction of the effect. The data set used to determine the outcome predictor and a detailed description of the statistical methods used are available as Supplementary Appendix 1 with the full text of this article at <http://www.nejm.org> and at <http://llmpp.nih.gov/DLBCL>.

## RESULTS

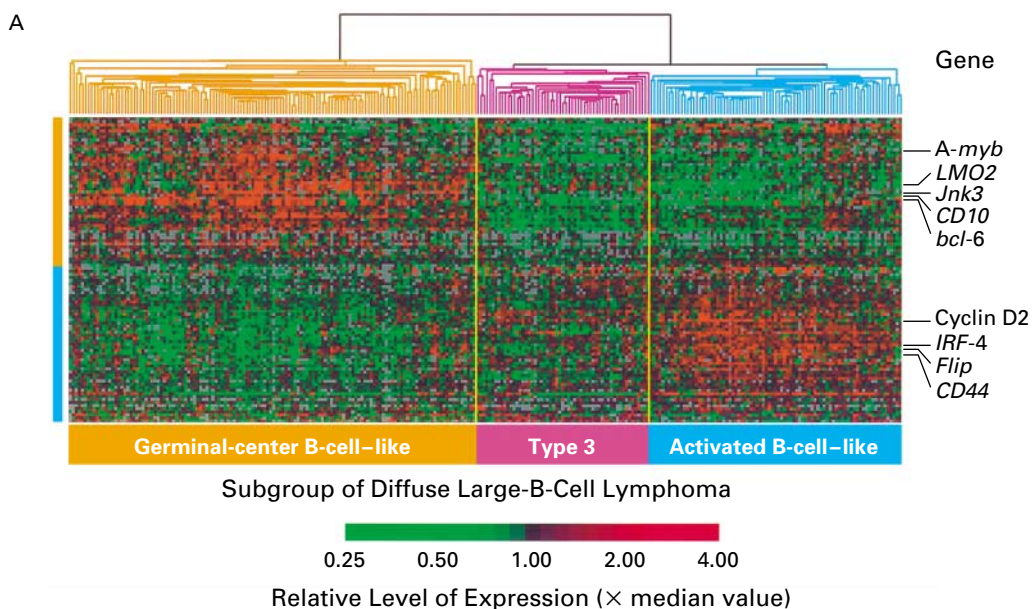
### Molecular, Pathological, and Clinical Features of the Subgroups

We first determined whether we could identify the previously described gene-expression subgroups<sup>3</sup> in the group of diffuse large-B-cell lymphomas that we analyzed. We applied a hierarchical-clustering algorithm to group the lymphomas according to the expression of 100 genes that distinguished between germinal-center B-cell–like and activated B-cell–like diffuse large-B-cell lymphomas at a significance level of  $P < 0.001$  in the previous analysis<sup>3</sup> and found three

**Figure 1 (facing page).** Subgroups of Diffuse Large-B-Cell Lymphoma According to Gene-Expression Profiles.

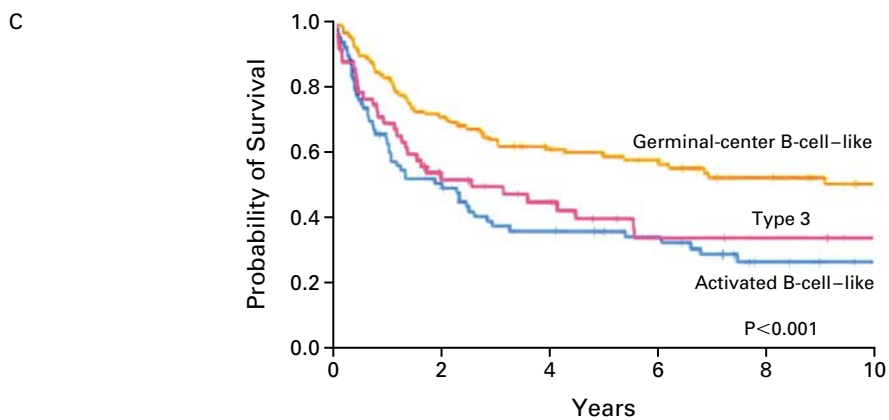
Panel A shows the hierarchical clustering of diffuse large-B-cell lymphomas from 240 patients with untreated disease and 34 patients who had previously been treated or who had a preexisting low-grade lymphoma, according to the level of expression of 100 genes. Red areas indicate increased expression, and green areas decreased expression. Each column represents a single diffuse large-B-cell lymphoma, and each row represents a single gene. Genes that are characteristically expressed in germinal-center B-cell–like diffuse large-B-cell lymphomas or activated B-cell–like diffuse large-B-cell lymphomas are indicated. The dendrogram at the top shows the degree to which each diffuse large-B-cell lymphoma is related to the others with respect to gene expression. Panel B shows the number of samples with amplification of the *c-rel* locus and *bcl-2* translocations in subgroups of diffuse large-B-cell lymphoma. The ratio of genomic copy number for the *c-rel* and  $\beta_2$ -microglobulin loci was determined by a quantitative polymerase-chain-reaction assay, and ratios greater than 2 were considered to indicate *c-rel* amplification. The *bcl-2* translocations were detected with the use of a polymerase-chain-reaction assay for the main break-point cluster region that is frequently involved in the t(14;18) translocation. Data are from patients who had untreated diffuse large-B-cell lymphomas without preexisting cancer. Panel C shows Kaplan–Meier estimates of overall survival after chemotherapy among the 240 previously untreated patients, according to the gene-expression subgroup.

MOLECULAR PROFILING OF SURVIVAL IN DIFFUSE LARGE-B-CELL LYMPHOMA



**B**

Oncogenic Abnormality	Germinal-center B-cell-like	Type 3	Activated B-cell-like
	no. of samples		
<i>c-rel</i> amplification	17	0	0
<i>bcl-2</i> t(14;18)	26	0	0



No. AT RISK

	0	2	4	6	8	10
Germinal-center B-cell-like	115	81	60	46	32	19
Type 3	52	24	18	10	8	5
Activated B-cell-like	73	35	23	19	8	5

large subgroups (Fig. 1A). One had a high level of expression of the genes characteristic of germinal-center B-cell–like diffuse large-B-cell lymphoma and normal germinal-center B cells, another expressed genes characteristic of activated B-cell–like diffuse large-B-cell lymphoma and mitogenically activated blood B cells, and the third, termed type 3 diffuse large-B-cell lymphoma, did not express either set of genes at a high level. The heterogeneity of expression within this third subgroup indicates that it may consist of more than one type of diffuse large-B-cell lymphoma.

The subgroups differed substantially with respect to two recurrent oncogenic events. The t(14;18) translocation involving the *bcl-2* gene and the amplification of the *c-rel* locus on chromosome 2p occurred exclusively in germinal-center B-cell–like diffuse large-B-cell lymphomas (Fig. 1B). These findings support the view that the various subgroups represent different diseases that arise as a result of distinct mechanisms of malignant transformation.<sup>3,11</sup>

The clinical and pathological features of the three subgroups are shown in Table 1. The most common histologic type of diffuse large-B-cell lymphoma — centroblastic monomorphic — was found in 66 percent of the germinal-center B-cell–like subgroup but also in 32 percent of the activated B-cell–like subgroup and 27 percent of the type 3 subgroup. Centroblastic polymorphic and immunoblastic subtypes

were more common in activated B-cell–like and type 3 diffuse large-B-cell lymphomas, but they were also observed in the germinal-center B-cell–like subgroup. Thus, these three subgroups were not strictly related to a specific histologic subtype. With respect to clinical features, a significantly higher proportion of patients in the activated B-cell–like subgroup than in the other two groups were older than 60 years of age ( $P=0.05$ ) and had an ECOG performance status of more than 1 ( $P=0.03$ ), but the tumor subgroup did not correlate with the risk groups defined on the basis of the international prognostic index ( $P=0.44$ ).

Overall survival after anthracycline-based chemotherapy differed significantly among the three subgroups ( $P<0.001$ ) (Fig. 1C). Patients with germinal-center B-cell–like diffuse large-B-cell lymphoma had a five-year survival rate of 60 percent, as compared with a rate of 39 percent for patients with diffuse type 3 large-B-cell lymphoma and 35 percent for those with activated B-cell–like diffuse large-B-cell lymphoma.

**Correlation between Expression of Individual Genes and Outcome**

Although the three subgroups may be viewed as distinct diseases, these divisions did not fully reflect the differences in survival among patients with diffuse large-B-cell lymphoma. For example, patients with germinal-center B-cell–like diffuse large-B-cell lymphoma

**TABLE 1. CHARACTERISTICS OF PATIENTS WITH DIFFUSE LARGE-B-CELL LYMPHOMA.\***

CHARACTERISTIC	TOTAL (N=240)	percent			P VALUE
		GERMINAL-CENTER B-CELL-LIKE (N=115)	TYPE 3 (N=52)	ACTIVATED B-CELL-LIKE (N=73)	
International-prognostic-index component					
Lactate dehydrogenase >1× normal	57	56	50	65	0.3
Age >60 yr	55	48	58	66	0.05
Ann Arbor stage >II	55	53	53	59	0.68
No. of extranodal sites >1	20	20	25	15	0.34
ECOG performance status >1	22	19	16	33	0.03
Risk group					0.44
Low (score, 0–1)	37	40	41	29	
Intermediate (score, 2–3)	49	49	43	51	
High (score, 4–5)	14	11	15	19	
Histologic subtype					
Centroblastic monomorphic	47	66	27	32	<0.001
Centroblastic polymorphic	19	9	27	28	0.009
Immunoblastic	8	3	12	15	0.007
Burkitt-like	4	3	2	8	0.31
Plasmablastic	5	3	8	6	0.71
T-cell rich	1	0	4	0	0.05
Anaplastic	1	0	4	0	0.05
Unclassified	15	18	16	11	0.49

\*Because of rounding, percentages may not total 100. ECOG denotes Eastern Cooperative Oncology Group.

**TABLE 2.** USE OF GENE-EXPRESSION PROFILES TO PREDICT OUTCOME FOR PATIENTS WITH DIFFUSE LARGE-B-CELL LYMPHOMA.\*

GENE-EXPRESSION VARIABLE	NO. OF MICROARRAY FEATURES IN SIGNATURE (N = 7399)	NO. OF MICROARRAY FEATURES SIGNIFICANTLY ASSOCIATED (P < 0.01) WITH A GOOD OUTCOME (N = 162)	NO. OF MICROARRAY FEATURES SIGNIFICANTLY ASSOCIATED (P < 0.01) WITH A BAD OUTCOME (N = 508)	REPRESENTATIVE GENES	GENEBANK ACCESSION No.	P VALUE		RELATIVE RISK OF DEATH AMONG ALL PATIENTS (95% CI)	OUTCOME	
						PRELIMINARY GROUP	VALIDATION GROUP			
Germinal-center B-cell signature	151	15	0	<i>bcl-6</i> IMAGE 1334260† IMAGE 814622†	U00115 AA805575 AA236080	<0.001	0.009	<0.001	0.69 (0.59–0.81)	Favorable
MHC class II signature	37	35	0	HLA-DPα HLA-DQα HLA-DRα HLA-DRβ	X00457 X00452 K01171 M20430	<0.001	0.11	<0.001	0.69 (0.57–0.82)	Favorable
Lymph-node signature	357	30	2	α-Actinin Collagen type III α 1 Connective-tissue growth factor Fibronectin KIAA0233	X15804 X14420 M54995 X02761 D87071	<0.001	0.04	<0.001	0.72 (0.62–0.85)	Favorable
Proliferation signature	1333	6	287	Urokinase plasminogen activator <i>c-myc</i> <i>E2F3</i> <i>NPM3</i>	D00244 V00568 NM_014566 AF081280	<0.001	0.05	<0.001	1.63 (1.27–2.09)	Poor
Other	5521	76	219	<i>BMP6</i>	M60315	0.005	0.08	0.003	1.36 (1.11–1.65)	Poor

\*P values indicate the significance of the association between the average value of the gene-expression variable and overall survival. Relative risk indicates the change in risk associated with a change by a factor of 2 in the average value for a given gene-expression signature. CI denotes confidence interval, and MHC major histocompatibility complex.

†IMAGE refers to an Integrated Molecular Analysis of Genomes and Their Expression consortium complementary DNA clone on the Lymphochip microarray.

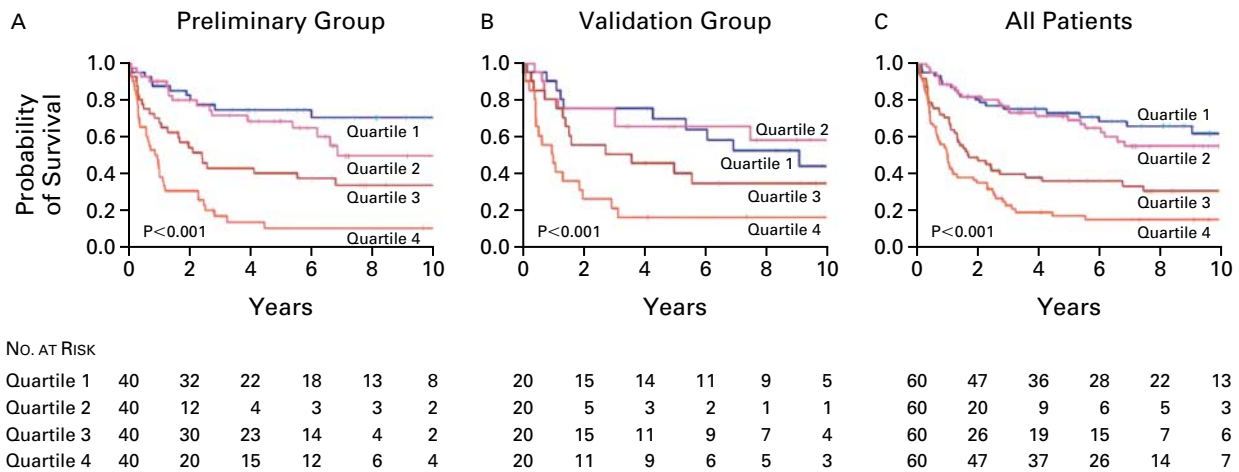
had the best prognosis, but these patients still had a 36 percent risk of death within three years after treatment. The patients with activated B-cell–like diffuse large-B-cell lymphoma, by contrast, had the worst prognosis, but the five-year survival rate of 35 percent suggests that some patients in this subgroup may be cured by chemotherapy.

For these reasons, we used a Cox proportional-hazards model to identify individual genes whose expression correlated with the outcome. Data from 670 of 7399 microarray features were significantly associated with a good or a bad outcome in the preliminary group ( $P < 0.01$ ); this number is greater than would be expected by chance ( $P = 0.005$  with the use of a permutation test) (Table 2).

To classify the genes that were correlated with outcome, we used hierarchical clustering to group them into gene-expression signatures.<sup>3</sup> A gene-expression signature is a group of genes expressed in a specific cell lineage or stage of differentiation or during a particular biologic response. Many of the genes we identified fell within previously described gene-expression sig-

natures (Table 2) (see Supplementary Appendix 1 at <http://www.nejm.org>). Among the 162 microarray features associated with a favorable outcome, 15 belonged to the signature that characterizes normal germinal-center B cells,<sup>3</sup> 30 were in the lymph-node signature of reactive nonmalignant cells in biopsy specimens of diffuse large-B-cell lymphoma,<sup>3</sup> and 35 were in the MHC class II signature. In the proliferation signature, which includes genes that are highly expressed in dividing cells,<sup>3</sup> 287 of 1333 microarray features were associated with a poor outcome.

Since genes within the same gene-expression signature are probably associated with similar biologic aspects of a tumor, we combined the genes that were significantly associated with survival ( $P < 0.01$ ) within each signature. To minimize the number of genes in the outcome predictor, we selected 16 genes that were highly variable in expression — 3 germinal-center B-cell genes, 4 MHC class II genes, 6 lymph-node genes, and 3 proliferation genes — and averaged the expression values for genes belonging to the same signature (Table 2). In a univariate analysis, these four

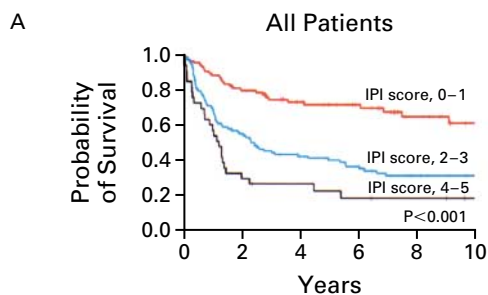


**Figure 2.** Kaplan–Meier Estimates of Overall Survival among Patients with Diffuse Large-B-Cell Lymphoma in the Preliminary Group (Panel A), the Validation Group (Panel B), and All Patients (Panel C).

**Figure 3 (facing page).** Kaplan–Meier Estimates of Survival According to the International Prognostic Index (IPI) Alone (Panel A) and to the International Prognostic Index and the Gene-Expression Profile (Panel B).

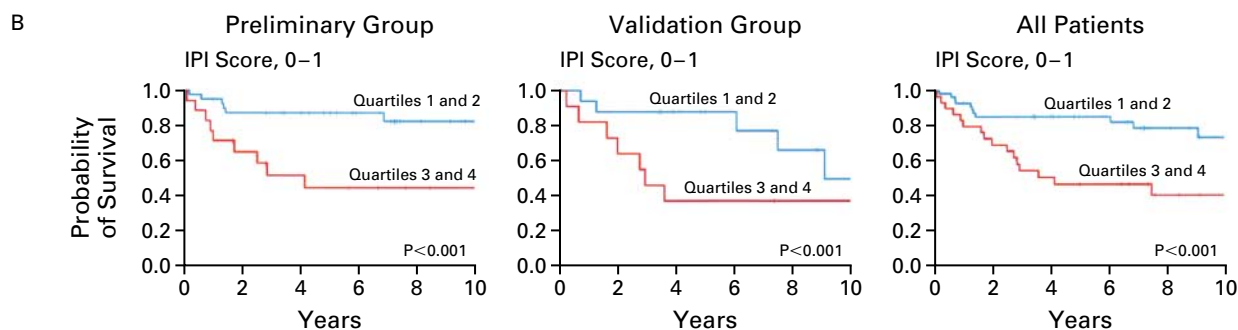
Panel A shows overall survival among patients in the low-risk group according to the international prognostic index score (a score of 0 to 1), those in the intermediate-risk group (indicated by a score of 2 to 3), and those in the high-risk group (indicated by a score of 4 to 5). Panel B shows overall survival among patients in the various IPI risk groups in the preliminary group, the validation group, and the group as a whole, stratified according to the quartile of the gene-expression–based outcome-predictor scores; quartiles 1 and 2 and quartiles 3 and 4 were merged. Higher quartiles indicate a poorer outcome. A few patients in the preliminary and validation groups were assigned to different gene-expression–based outcome-predictor groups when the groups were combined owing to slight differences in the cutoff points for each quartile within each group of patients.

MOLECULAR PROFILING OF SURVIVAL IN DIFFUSE LARGE-B-CELL LYMPHOMA



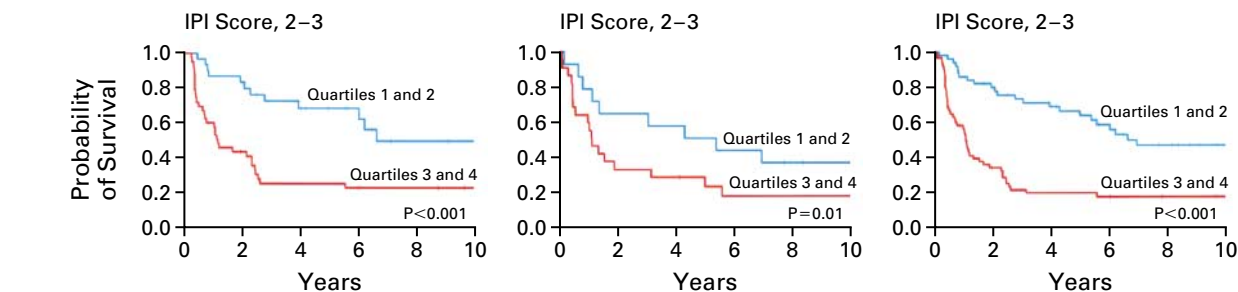
No. AT RISK

IPI score, 0-1	82	62	50	37	22	13
IPI score, 2-3	108	56	38	28	19	13
IPI score, 4-5	32	10	6	4	3	1



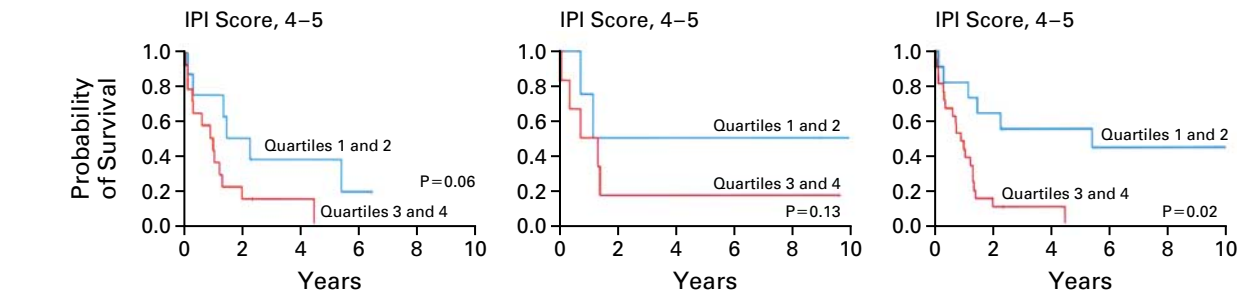
No. AT RISK

Quartiles 1 and 2	38	31	28	20	11	7	16	14	11	8	6	3	53	43	37	27	17	10
Quartiles 3 and 4	17	10	7	5	3	2	11	7	4	4	2	1	29	19	13	10	5	3



No. AT RISK

Quartiles 1 and 2	30	24	15	11	6	3	14	9	8	6	4	3	49	38	28	20	13	8
Quartiles 3 and 4	42	16	9	8	6	4	22	7	6	3	3	3	59	18	10	8	6	5



No. AT RISK

Quartiles 1 and 2	8	4	2	1	0	0	4	2	2	2	2	1	11	7	5	4	3	1
Quartiles 3 and 4	14	3	1	0	0	0	6	1	1	1	1	0	21	3	1	0	0	0

signatures were found to predict survival in the preliminary group, in the validation group, and in the group of all patients (Table 2).

Individual genes that were not in these four signatures but that predicted the likelihood of survival in a univariate analysis of the preliminary group ( $P < 0.01$ ) were evaluated to determine whether these variables increased the predictive value of the test. The only gene that significantly increased the predictive value ( $P = 0.005$ ) was *BMP6*, which was associated with a poor outcome.

The final model combined the four gene-expression signatures and *BMP6*. Each case of diffuse large-B-cell lymphoma was assigned a score that was calculated as the weighted sum of these components, optimized by a Cox proportional-hazards model of overall survival within the preliminary group. The score, expressed as a continuous variable, correlated significantly with the clinical outcome in both the preliminary group ( $P < 0.001$ ) and the validation group ( $P < 0.001$ ), indicating that the results are reproducible. The score ranged from  $-1.7$  to  $2.4$ , with a standard deviation of  $0.72$ , and each unit increase in the score induced an increase in the relative risk of death by a factor of  $2.7$  (95 percent confidence interval,  $2.11$  to  $3.51$ ).

The patients were ranked according to their score and divided into quartiles (from highest to lowest scores). Kaplan–Meier plots of overall survival showed distinct differences in the five-year survival rates in the various quartiles in both the preliminary and validation groups (Fig. 2). In the group as a whole, the five-year survival rates were 73 percent in quartile 1, 71 percent in quartile 2, 34 percent in quartile 3, and 15 percent in quartile 4 (Fig. 2C).

#### Comparison of the Gene-Expression–Based Outcome Predictor and the International Prognostic Index

The international-prognostic-index scores (Fig. 3A) and the gene-expression–based scores were independent predictors of survival in the preliminary group ( $P < 0.001$ ) and the validation group ( $P = 0.002$ ). In a multivariate Cox model that combined both the international-prognostic-index scores and the gene-expression–based scores, each unit increase in the latter score increased the relative risk of death by a factor of  $2.6$  (95 percent confidence interval,  $2.02$  to  $3.48$ ).

Kaplan–Meier plots of overall survival showed the independence of the international-prognostic-index score and the gene-expression–based score (Fig. 3B). For these plots, we combined quartiles 1 and 2 into one group and quartiles 3 and 4 into a second group. These two groups had significantly different outcomes in the analysis of patients with low or intermediate risk according to their international-prognostic-index scores, and this difference was observed in both the preliminary and the validation groups. The gene-

expression–based method also identified the few patients in the high-risk group according to the international-prognostic-index score who were long-term survivors (Fig. 3B).

#### Relation between Gene-Expression–Based Score and Subtype of Diffuse Large-B-Cell Lymphoma

Since overall survival differed in the three subgroups of diffuse large-B-cell lymphoma (Fig. 1C), we investigated whether the components of the outcome predictor were differentially expressed by these subgroups. As expected, the germinal-center B-cell signature was much more highly expressed in the germinal-center B-cell-like subgroup than in the other two subgroups (Fig. 4A). The activated B-cell–like subgroup had the highest level of expression of the proliferation signature and *BMP6* but the lowest level of expression of the lymph-node signature. On the other hand, the level of expression of the MHC class II signature was similar among the three subgroups. The gene-expression–based score was highest in the activated B-cell–like group, intermediate in the type 3 subgroup, and lowest in the germinal-center B-cell–like subgroup (Fig. 4A), demonstrating that this approach incorporates the clinical distinctions among the subgroups of diffuse large-B-cell lymphoma.

Nonetheless, the components of the outcome predictor were differentially expressed within each subgroup (Fig. 4B), and the predictor score could be used to subdivide the patients within each subgroup into distinct risk groups (Fig. 4C). This feature accounts for the predictor's greater prognostic power as compared with that derived from the use of subgroups of diffuse large-B-cell lymphoma.

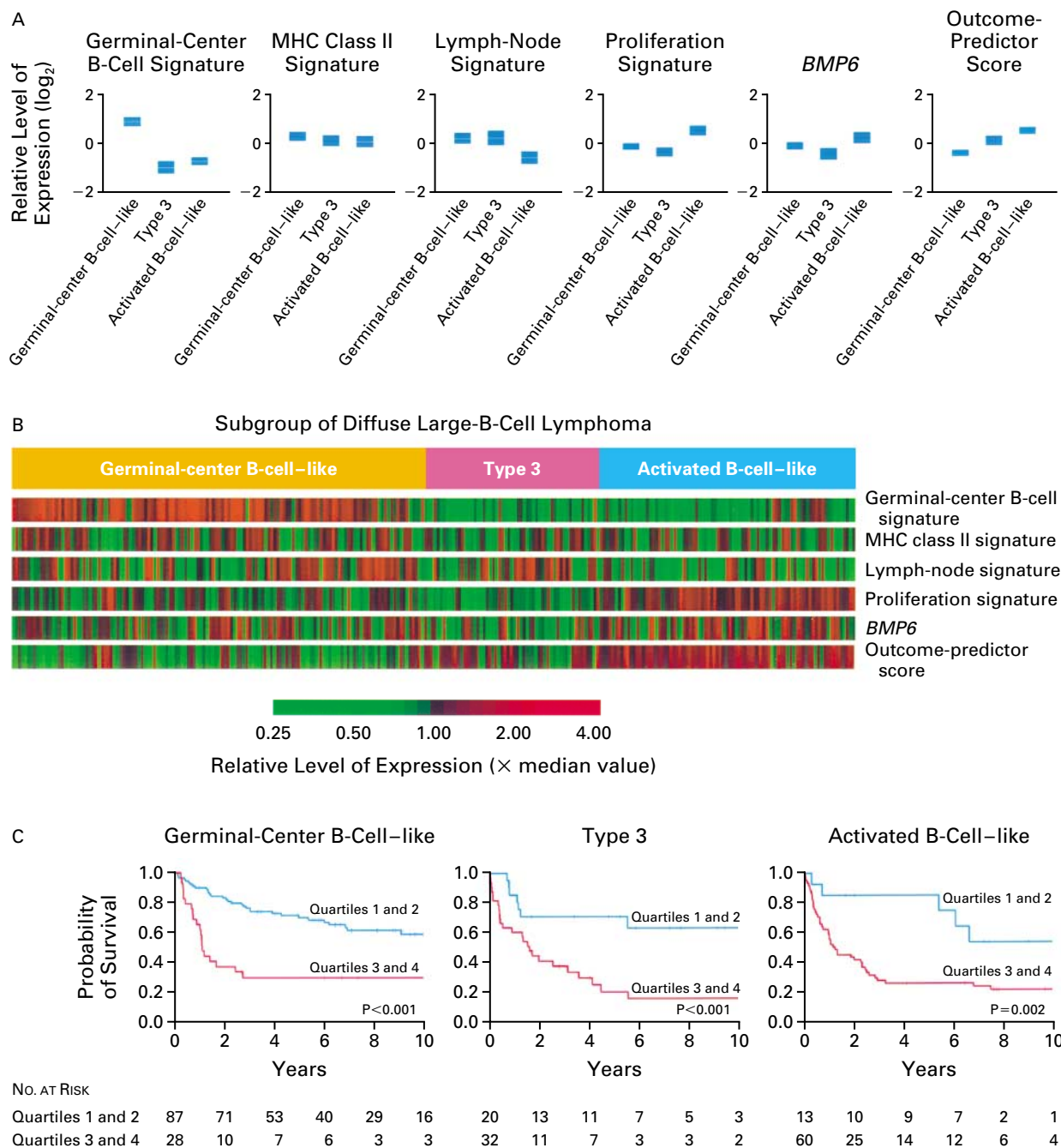
## DISCUSSION

We have developed a method to predict the likelihood of survival after chemotherapy for diffuse large-B-cell lymphoma that is based on patterns of gene expression in biopsy specimens of the lymphoma. This molecular method and the clinically based international prognostic index were found to be independent predictors of the prognosis. The international prognostic index has not proved effective in stratifying patients with diffuse large-B-cell lymphoma for therapeutic trials,<sup>5</sup> but we think that the gene-expression–based method might serve this purpose.

Our analysis of gene-expression profiles integrated two complementary approaches to outcome prediction. In the first approach, we used hierarchical clustering to identify subgroups that differed with respect to the expression of hundreds of genes. With the use of this clustering method, two recurrent genetic abnormalities — the  $t(14;18)$  translocation and amplification of the *c-rel* locus — were found exclusively in the germinal-center B-cell–like subgroup. This finding sup-



MOLECULAR PROFILING OF SURVIVAL IN DIFFUSE LARGE-B-CELL LYMPHOMA



**Figure 4.** Relation between the Gene-Expression–Based Outcome-Predictor Score and the Subgroup of Diffuse Large-B-Cell Lymphoma.

Panel A shows the mean ( $\pm$ SE) expression value (after  $\log_2$  transformation) of each outcome-predictor component and the score. Panel B shows the level of expression of variables in the outcome predictor and the scores in the three subgroups of lymphoma. Panel C shows Kaplan–Meier plots of overall survival in the subgroups, stratified according to the quartile of risk reflected by the gene-expression–based score. Quartiles 1 and 2 and quartiles 3 and 4 were merged. Higher quartiles indicate a poorer outcome. MHC denotes major histocompatibility complex.

ports our view that the subgroups of diffuse large-B-cell lymphoma represent distinct disease entities.<sup>3</sup> Further support for this idea is supplied by the finding that the germinal-center B-cell–like subgroup had a significantly greater likelihood of survival after chemotherapy than did the activated B-cell–like and type 3 subgroups.

In the second approach, we used clinical and gene-expression data to identify individual genes that predicted the outcome and then combined these variables into a multivariate model. This model incorporated differences in the levels of gene expression among the subgroups of diffuse large-B-cell lymphoma that influenced the outcome, as well as other differences in gene expression that were associated with the likelihood of survival.

The predictive genes fell within four biologic groups defined on the basis of gene-expression signatures. The proliferation signature was the best predictor of an adverse outcome — a finding that is consistent with those of previous analyses of tumor-cell proliferation in diffuse large-B-cell lymphoma.<sup>12,13</sup> Two of the gene-expression signatures that were associated with a good outcome suggest that the immune response to the tumor cells may be a crucial determinant of survival after chemotherapy. The MHC class II gene-expression signature correlated with a good outcome, suggesting that antigen presentation to the immune system has a role in therapeutic responses.<sup>14</sup> The genes in the lymph-node signature that were associated with a good outcome encode components of the extracellular matrix and connective-tissue growth factor, a mediator of fibrosis that promotes the synthesis of the matrix.<sup>15</sup> Sclerotic reactions occur in some cases of diffuse large-B-cell lymphoma, and the lymph-node signature may reflect these reactions. Other genes in the lymph-node signature are characteristically expressed in macrophages and natural killer cells, again suggesting that an antitumor immune response improves survival after chemotherapy.

The prognostically favorable germinal-center B-cell gene-expression signature is notable because cell lines derived from germinal-center B-cell–like diffuse large-B-cell lymphomas have decreased activity of the nuclear factor  $\kappa$ B signaling pathway.<sup>16</sup> This protective pathway interferes with the apoptotic effect of chemotherapy.<sup>17</sup> In contrast, in activated B-cell–like diffuse large-B-cell lymphomas, there is constitutive activation of this pathway,<sup>16</sup> which, in principle, could block the apoptosis induced by chemotherapy and thus account for the relatively poor outcome in this subgroup.

Our outcome predictor may help identify patients with diffuse large-B-cell lymphoma who are unlikely to be cured by conventional therapy. On the basis of their predictor scores, one quarter of the 240 patients in this study could be assigned to a risk group with a

five-year survival rate of 15 percent. Among patients at intermediate risk according to the international-prognostic-index score, use of the outcome predictor indicated that 55 percent of these patients had a five-year survival rate of 18 percent, whereas this group of patients had a five-year survival rate of 41 percent overall. Even among patients with a low risk according to the international-prognostic-index score, use of the outcome predictor showed that 16 percent of the patients in this group had a five-year survival rate of only 28 percent (data not shown).

One of the features of our analysis is that the outcome predictor involves a small number of genes and thus multiplexed quantitative reverse-transcriptase–polymerase-chain-reaction assays or customized DNA microarrays could easily be developed for clinical application. Regardless of the eventual choice of technique, our study highlights the need to use molecular diagnosis in patients with diffuse large-B-cell lymphoma, since these patients have molecularly distinct diseases that may require individualized and molecularly targeted therapies.

Supported by a Director's Challenge grant (UO1-CA84967) from the National Cancer Institute, by the Center for Cancer Research of the National Cancer Institute, and by the Cancer Genome Anatomy Project. Dr. Rosenwald was supported in part by the Deutsche Krebshilfe, Bonn, Germany. Ms. Averett was a Howard Hughes Medical Institute Research Scholar at the National Institutes of Health during the study.

*We are indebted to Bob Strausberg, Narajan Bhat, and Butch Hopkins for preparing the messenger RNA reference pool and for helpful discussions, and to Ken Buetow and Jenny Kelly for verifying the sequence of Lymphochip clones.*

## REFERENCES

1. Coiffier B. Diffuse large cell lymphoma. *Curr Opin Oncol* 2001;13:325-34.
2. Fisher RI, Gaynor ER, Dahlborg S, et al. Comparison of a standard regimen (CHOP) with three intensive chemotherapy regimens for advanced non-Hodgkin's lymphoma. *N Engl J Med* 1993;328:1002-6.
3. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;403:503-11.
4. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. A predictive model for aggressive non-Hodgkin's lymphoma. *N Engl J Med* 1993;329:987-94.
5. Shipp MA, Abelloff MD, Antman KH, et al. International consensus conference on high-dose therapy with hematopoietic stem cell transplantation in aggressive non-Hodgkin's lymphomas: report of the jury. *J Clin Oncol* 1999;17:423-9.
6. Alizadeh A, Eisen M, Davis RE, et al. The Lymphochip: a specialized cDNA microarray for the genomic-scale analysis of gene expression in normal and malignant lymphocytes. *Cold Spring Harbor Symp Quant Biol* 1998;64:71-8.
7. Goff LK, Neat MJ, Crawley CR, et al. The use of real-time quantitative polymerase chain reaction and comparative genomic hybridization to identify amplification of the REL gene in follicular lymphoma. *Br J Haematol* 2000;111:618-25.
8. Stetler-Stevenson M, Raffeld M, Cohen P, Cossman J. Detection of occult follicular lymphoma by specific DNA amplification. *Blood* 1988;72:1822-5.
9. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 1998;95:14863-8.
10. Shaffer AL, Rosenwald A, Hurt EM, et al. Signatures of the immune response. *Immunity* 2001;15:375-85.

- 11.** Huang JZ, Sanger WG, Greiner TC, et al. The t(14;18) defines a unique subset of diffuse large B-cell lymphoma with a germinal center B-cell gene expression profile. *Blood* 2002;99:2285-90.
- 12.** Grogan TM, Lippman SM, Spier CM, et al. Independent prognostic significance of a nuclear proliferation antigen in diffuse large cell lymphomas as determined by the monoclonal antibody Ki-67. *Blood* 1988;71:1157-60.
- 13.** Silvestrini R, Costa A, Boracchi P, Giardini R, Rilke F. Cell proliferation as a long-term prognostic factor in diffuse large-cell lymphomas. *Int J Cancer* 1993;54:231-6.
- 14.** Miller TP, Lippman SM, Spier CM, Slymen DJ, Grogan TM. HLA-DR (Ia) immune phenotype predicts outcome for patients with diffuse large cell lymphoma. *J Clin Invest* 1988;82:370-2.
- 15.** Grotendorst GR. Connective tissue growth factor: a mediator of TGF-beta action on fibroblasts. *Cytokine Growth Factor Rev* 1997;8:171-9.
- 16.** Davis RE, Brown KD, Siebenlist U, Staudt LM. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J Exp Med* 2001;194:1861-74.
- 17.** Baldwin AS. Control of oncogenesis and cancer therapy resistance by the transcription factor NF-kappaB. *J Clin Invest* 2001;107:241-6.

Copyright © 2002 Massachusetts Medical Society.