ORIGINAL ARTICLE

# Prediction of Survival in Diffuse Large-B-Cell Lymphoma Based on the Expression of Six Genes

Izidore S. Lossos, M.D., Debra K. Czerwinski, B.A., Ash A. Alizadeh, M.D., Ph.D., Mark A. Wechser, Ph.D., Rob Tibshirani, Ph.D., David Botstein, Ph.D., and Ronald Levy, M.D.

## ABSTRACT

### BACKGROUND

Several gene-expression signatures can be used to predict the prognosis in diffuse large-B-cell lymphoma, but the lack of practical tests for a genome-scale analysis has restricted the use of this method.

### METHODS

We studied 36 genes whose expression had been reported to predict survival in diffuse large-B-cell lymphoma. We measured the expression of each of these genes in independent samples of lymphoma from 66 patients by quantitative real-time polymerase-chain-reaction analyses and related the results to overall survival.

### RESULTS

In a univariate analysis, genes were ranked on the basis of their ability to predict survival. The genes that were the strongest predictors were *LMO2*, *BCL6*, *FN1*, *CCND2*, *SCYA3*, and *BCL2*. We developed a multivariate model that was based on the expression of these six genes, and we validated the model in two independent microarray data sets. The model was independent of the International Prognostic Index and added to its predictive power.

### CONCLUSIONS

Measurement of the expression of six genes is sufficient to predict overall survival in diffuse large-B-cell lymphoma.

From the Division of Oncology, Department of Medicine (I.S.L., D.K.C., R.L.), the Department of Genetics (A.A.A., D.B.), and the Departments of Health Research and Policy and Statistics (R.T.), Stanford University Medical Center, Stanford, Calif.; the Division of Hematology–Oncology, Department of Medicine, University of Miami, Miami (I.S.L.); and Applied Biosystems, Foster City, Calif. (M.A.W.). Address reprint requests to Dr. Levy at Stanford University School of Medicine, Division of Oncology, Rm. 1105, Stanford, CA 94305-5151, or at levy@stanford.edu.

N Engl J Med 2004;350:1828-37.

THE MOST COMMON TYPE OF LYMPHOMA in adults, diffuse large-B-cell lymphoma, has an annual incidence in the United States of more than 25,000 cases and accounts for 30 to 40 percent of cases of non-Hodgkin's lymphomas.[1] Combination chemotherapy has transformed diffuse large-B-cell lymphoma from a universally fatal disease to a potentially curable one, but less than half of all patients are cured.[2] The International Prognostic Index (IPI), a well-established predictor of outcome in diffuse large-B-cell lymphoma, is based on five clinical characteristics (age, tumor stage, serum lactate dehydrogenase concentration, performance status, and number of extranodal disease sites).[3] However, the outcome in patients with diffuse large-B-cell lymphoma who have identical IPI values varies considerably. New molecular methods may make risk-adjusted therapies possible for diffuse large-B-cell lymphoma in a way similar to the current practice in acute leukemia.

The relation between prognosis and the molecular features of diffuse large-B-cell lymphoma has been investigated with the use of genome-scale expression profiles assessed by DNA microarrays.[4-6] There are a variety of techniques for analyzing microarray data, but the two general types are unsupervised and supervised. With the unsupervised approach, microarray data are analyzed without the use of external information such as clinical data or survival time. In contrast, with the supervised approach, the aim is to identify genes whose expression correlates with some external variables. With both unsupervised[4] and supervised[5,6] methods, microarray studies of diffuse large-B-cell lymphomas showed that gene-expression signatures were associated with clinical outcomes.

Alizadeh et al.,[4] with lymphochip complementary DNA (cDNA) microarrays, showed that overall survival after chemotherapy was significantly longer among patients with diffuse large-B-cell lymphoma that had high levels of expression of genes characteristic of normal germinal-center B cells than among patients whose tumors had low levels of expression of these same genes. Two genes specifically expressed in the germinal-center B cell, *BCL6* and *HGAL*, have been shown to predict overall survival, independently of the IPI, in unrelated groups of patients studied with the use of other methods.[7-9] However, another germinal-center B-cell marker, *CD10*, did not predict survival in diffuse large-B-cell lymphoma, suggesting that the outcome is associated with the expression of only some genes in germinal-center B-cell signatures.[8]

Supervised analysis of gene-expression data in relation to overall survival has also made possible the construction of models to predict the outcome in diffuse large-B-cell lymphoma. Shipp et al.[5] derived a 13-gene predictive model, which was independent of the IPI, from a cohort of 58 patients whose lymphomas were analyzed by oligonucleotide microarrays. Only 3 of these 13 genes were present in the data analyzed by Alizadeh and colleagues,[4] and of those 3, only 2 were associated with survival. Rosenwald et al.[6] used supervised analysis of gene-array data from 160 patients with diffuse large-B-cell lymphoma to derive a predictive model based on the expression of 17 genes and applied this model to a set of such lymphomas from 80 other patients.

There is no overlap among the genes in the models derived by Shipp et al. and Rosenwald et al.[5,6] Technical differences, the composition of the microarrays used, and different algorithms used for constructing predictive models may underlie this disparity. In addition, every predictive model must be validated in an independent cohort of patients to confirm that it works generally and not just for the group of patients from which it was derived.[10,11] Therefore, it remains unclear which method and which model best capture the molecular, histopathological, and clinical heterogeneity of diffuse large-B-cell lymphoma. Also, since microarrays are not yet readily available in clinical laboratories, more practical assays for gene expression are needed.

We used quantitative reverse-transcriptase polymerase chain reaction (RT-PCR) to measure the expression of 36 genes in diffuse large-B-cell lymphomas from 66 patients. We then built a predictive model based on the genes that were correlated with overall survival, either positively or negatively, and validated the model by applying it to the microarray data from Shipp et al.[5] and Rosenwald et al.[6] in order to determine whether it had predictive value that was independent of the method of measuring gene expression (i.e., quantitative RT-PCR, cDNA microarrays, or oligonucleotide microarrays). Our goal was to devise a model that was technically simple and applicable for routine clinical use.

## METHODS

### TUMOR SPECIMENS

During diagnostic procedures at Stanford University Medical Center from 1975 to 1995, we obtained tumor specimens from patients with newly diagnosed diffuse large-B-cell lymphoma. Specimens

**Table 1. Sources of Evidence for a Panel of 36 Genes Whose Expression Predicts Survival in Diffuse Large-B-Cell Lymphoma.***

| Source of Evidence | Genes |
|---|---|
| Reports identifying single prognostic genes | *ICAM1/CD54*,[14] *PAX5*,[15] *Ki-67*,[16] *CD44*,[17] *P53*,[18,19] *BCL2*,[20-23] *BIRC5* (survivin),[24] *BCL6*,[8,9] *HGAL*,[7-9] *PRDM1*,[25] *SCYA3*,[25] *CCND1*,[26] *CCND2*[25]† |
| Alizadeh et al.[4]‡ | *LMO2*, *LRMP*, *CD10*, *MYBL1/A-MYB*, *BCL7A*, *PIK3CG*, *CR2*, *CD38*, *SLAM*, *WASPIP*, *CFLAR*, *SLA*, *IRF4*, *PMS1*, *HGAL*, *BCL6*, *BCL2* |
| Shipp et al.[5]§ | *NR4A3*, *PDE4B* |
| Rosenwald et al.[6]¶ | *FN1*, *PLAU*, *HLA-DQA1*, *HLA-DRA*, *EEF1A1L4*, *NPM3*, *MYC*, *BCL6*, *HGAL* |

* *BCL2*, *BCL6*, and *HGAL* are present in more than one source.
† Given the prominence of *BCL6* in diffuse large-B-cell lymphoma, we also included three genes that Shaffer et al.[25] have shown to be targets of *BCL6* (*PRDM1*, *SCYA3*, and *CCND2*).
‡ In addition to representatives from the 71 or so genes used by Alizadeh et al.,[4] we included genes identified during a reanalysis of the data set.
§ Of the 13 genes in the model of Shipp et al., 2 genes could be assessed independently in the data set used by Alizadeh et al. and showed significant correlation with survival, though no adjustment was made for multiple-hypothesis testing.
¶ To derive their predictive model, Rosenwald et al.[6] used 17 genes. Only the nine that were associated with survival in independent data analyses (by significance analysis of microarrays) were included in the study.

were stored frozen, as previously reported.[7,8] The diagnosis of diffuse large-B-cell lymphoma according to the revised European–American lymphoma classification[12] was confirmed on reevaluation of all specimens before their inclusion in this study. All the tumors had the histologic appearance of centroblastic large-B-cell lymphomas with a diffuse pattern and no residual follicles. All patients were treated with a regimen that included an anthracycline (cyclophosphamide, doxorubicin, vincristine, and prednisone [CHOP] or CHOP-like regimens) and were followed up at Stanford University Hospital. Primary diffuse large-B-cell lymphoma specimens from a total of 66 patients fulfilled the criteria for inclusion in the study. Information on the tumor stage was obtained for all the patients according to the Ann Arbor system of staging lymphomas. We were able to determine the IPI score for 58 of the patients. Written informed consent was obtained from all patients, and the study was approved by the institutional review board of Stanford University Medical Center.

### RNA ISOLATION AND REAL-TIME PCR

Isolation of RNA, its quantification, and the RT reactions were performed according to established methods.[7,13] Expression of messenger RNA (mRNA) for 36 genes that we tested (Table 1, as well as Table A in the Supplementary Appendix [available with the full text of this article at www.nejm.org]) and 2 endogenous control genes was measured in each biopsy specimen of diffuse large-B-cell lymphoma by real-time PCR (with TaqMan Gene Expression Assays products on an ABI PRISM 7900 HT Sequence Detection System, Applied Biosystems).[13] For each gene, two to four sets of TaqMan probes and primers were tested. The probes contain a 6-carboxy-fluorescein phosphoramidite (FAM dye) label at the 5′ end of the gene and a minor groove binder and nonfluorescent quencher at the 3′ end and are designed to hybridize across exon junctions. The assays are supplied with primers and probe concentrations of 900 nM and 250 nM, respectively. For each gene, the assay with the highest amplification efficiency was selected for this study; the TaqMan probes and primer sequences are presented in Table A in the Supplementary Appendix. No fluorescent signal was generated by these assays when genomic DNA was used as a substrate, which confirms that the assays measured only mRNA.

Phosphoglycerate kinase 1 (*PGK1*) and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) were used as the endogenous RNA and cDNA quantity controls (P/N 4326318E and P/N 4326317E, respectively; Applied Biosystems). We chose *PGK1* and *GAPDH* on the basis of an analysis of their relatively constant expression in diffuse large-B-cell lymphoma.[13] Since the normalization to the endogenous control genes *PGK1* and *GAPDH* led to similar results and conclusions, we present only the data normalized to *PGK1* expression. For calibration and generation of standard curves, we used cDNA derived from the Raji cell line of human B-cell lymphoma, cDNA prepared from Universal Human Reference RNA (Stratagene), or both. The cDNA prepared from Universal Human Reference RNA was used for genes that were not abundant in the Raji cell line (*CCND1*, *CCND2*, *SLA*, *NR4A3*, *CD44*, *PLAU*, and *FN1*). To control for possible variations among PCR runs performed on different days, the expression of all the analyzed and endogenous control genes was assessed in the Raji cell line before, midway through, and on completion of the analysis of all the specimens of diffuse large-B-cell lymphoma. The assays were highly reproducible, with coefficient of variation less than 0.16 among these three runs for all the genes assessed in the Raji cell line.

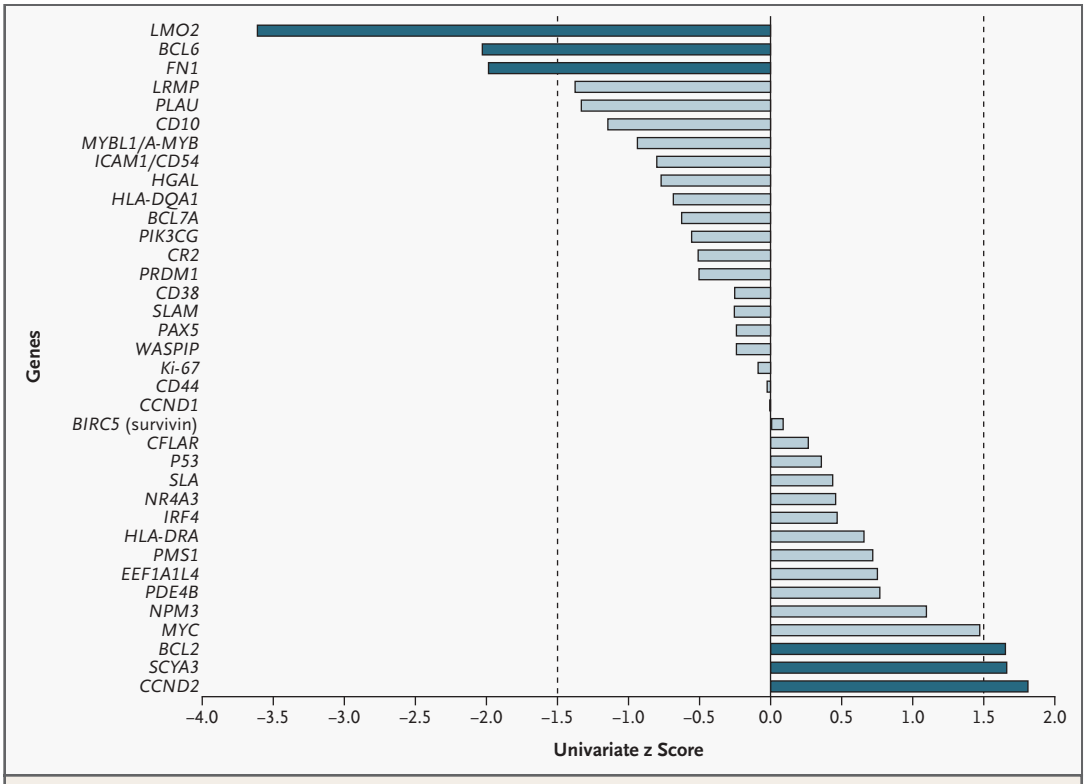


**Figure 1. Univariate Analysis of Expression of 36 Genes with Overall Survival as a Dependent Variable.**

The genes are ranked on the basis of their predictive power (univariate z score), with a negative score associated with longer overall survival and a positive score associated with shorter overall survival. The dashed lines represent an absolute univariate z score of ±1.5. The prediction model is based on the weighted expression of six genes and is expressed by the following equation: mortality-predictor score = (−0.0273 × *LMO2*) + (−0.2103 × *BCL6*) + (−0.1878 × *FN1*) + (0.0346 × *CCND2*) + (0.1888 × *SCYA3*) + (0.5527 × *BCL2*).

### STATISTICAL ANALYSIS

The normalized gene-expression values were log-transformed (on a base 2 scale), in a manner similar to the transformation of array-based hybridization data. Overall survival time was calculated from the date of diagnosis until death or the last follow-up contact. We estimated survival curves by the Kaplan–Meier product-limit method and compared them using the log-rank test. To construct a model for the prediction of survival, univariate Cox proportional-hazards analysis was performed, with overall survival as the dependent variable.[27] Subsequently, genes with an absolute univariate z score greater than 1.5 or less than −1.5 were analyzed in a multivariate Cox proportional-hazards regression model, with overall survival as the dependent variable. The individual components of the IPI and the overall score were included in the model. Two-sided P values of less than 0.05 were considered to indicate statistical significance. In the final model for the prediction of survival, we multiplied the log-transformed normalized expression value measured for each gene by a factor of z, a score derived from the multivariate analysis (see the Supplementary Appendix for a description of this method).

To validate the usefulness of this model, we applied it to two independent, previously published sets of gene-expression data for diffuse large-B-cell lymphoma that were derived from DNA-microarray analysis[5,6] (see the Supplementary Appendix). These data sets were compared without shifting of the means or other scaling of the raw gene-expression data.

## RESULTS

### SELECTION OF A PANEL OF GENES FOR QUANTITATIVE RT-PCR

We selected a group of 36 genes for this study (Table 1). The expression of each of these genes, measured either individually or in large data sets derived

**Table 2. Clinical Characteristics of Patients with Diffuse Large-B-Cell Lymphoma.***

| Characteristic | Low-Risk Group (N=20) | Medium-Risk Group (N=18) | High-Risk Group (N=20) | Total (N=58) |
|---|---|---|---|---|
| Age (yr) | | | | |
| Median | 40 | 50 | 50 | 47 |
| Range | 21–65 | 23–74 | 18–71 | 18–74 |
| Ann Arbor stage (no.) | | | | |
| I | 1 | 1 | 5 | 7 |
| II | 8 | 8 | 8 | 24 |
| III | 0 | 3 | 1 | 4 |
| IV | 11 | 6 | 6 | 23 |
| ECOG performance status (no.) | | | | |
| 0–1 | 15 | 17 | 19 | 51 |
| ≥2 | 5 | 1 | 1 | 7 |
| B symptoms (no.)† | | | | |
| Yes | 6 | 6 | 4 | 16 |
| No | 14 | 12 | 16 | 42 |
| Lactate dehydrogenase (no.) | | | | |
| High | 9 | 12 | 8 | 29 |
| Low | 11 | 6 | 12 | 29 |
| No. of extranodal sites (no.) | | | | |
| ≤1 | 13 | 15 | 11 | 39 |
| > 1 | 7 | 3 | 9 | 19 |
| International Prognostic Index (no.) | | | | |
| 0–2 | 14 | 13 | 17 | 44 |
| 3–5 | 6 | 5 | 3 | 14 |

* Patients in the low-risk group had survival-predictor scores of less than –0.063; those in the medium-risk group, –0.063 to <0.093; and those in the high-risk group, ≥0.093. ECOG denotes Eastern Cooperative Oncology Group. A score of 0 on the International Prognostic Index indicates the absence of all prognostic factors, and a score of 5 the presence of all prognostic factors.
† B symptoms are fever, unintentional weight loss, or night sweats.

from microarrays, has been found to predict survival in diffuse large-B-cell lymphoma. We applied significance analysis of microarrays[28] — a supervised method for the identification of genes with a statistically significant association with survival — to the data set of Alizadeh et al.[4] in order to identify genes that may have been missed in the unsupervised analyses.

We measured the expression of each of the 36 genes and of the 2 internal-control genes for input mRNA (*PGK1* and *GAPDH*) by quantitative RT-PCR. We determined the expression of each gene in each of the 66 specimens of lymphoma relative to its expression in a sample of RNA used as a reference.[13] (For raw data, see the Supplementary Appendix. The primer and probe sets for each of the genes in this study are shown in Table A.)

### A MODEL FOR PREDICTING SURVIVAL IN DIFFUSE LARGE-B-CELL LYMPHOMA

We first performed a univariate analysis of the expression data for the 36 genes, with overall survival as a dependent variable (Fig. 1). We ranked the genes on the basis of their predictive power (univariate z score). A negative z score was associated with longer overall survival, and a positive z score was associated with shorter overall survival. From this ranking, we selected an optimal number of genes for use in constructing a predictive model. Including more genes, even with some redundancy, would have tended to make the predictive model perform better in independent validation analyses, but a smaller number of genes would make the model more practical. By inspection, the conventional cutoff value for z of ±2.0 ($P<0.05$) would have yielded only one gene, *LMO2* (Fig. 1). Therefore, we picked the z value of ±1.5 ($P=0.13$), which allowed the six genes to be included. (Other z values were tried after the fact, and the results are shown for comparison in the Supplementary Appendix.) The six genes that exceeded the z value of 1.5 in the univariate analysis were *LMO2*, *BCL6*, *FN1*, *CCND2*, *SCYA3*, and *BCL2*.

When we performed a multivariate Cox regression analysis with overall survival as a dependent variable, none of these genes independently predicted overall survival at a statistically significant level. This is not surprising, since the genes are interrelated (e.g., *BCL6* is known to down-regulate the expression of *CCND2* and *SCYA3*).[25] Another multivariate Cox regression analysis was then performed, which included the six genes as well as each of the components of the IPI. This analysis showed that only the serum lactate dehydrogenase concentration was an independent predictor of overall survival ($P=0.004$).

Since we intended to construct a model that would be independent of and not overlap the IPI, we did not use the serum lactate dehydrogenase concentration in the model, because it is already a component of the IPI score. Instead, we constructed a model that is based on the relative contributions of each of the six genes in the multivariate analysis, as described in the following equation: mortality-predictor score $=(-0.0273\times LMO2)+(-0.2103\times BCL6)+(-0.1878\times FN1)+(0.0346\times CCND2)+(0.1888\times SCYA3)+(0.5527\times BCL2)$. For example, the negative weighting value assigned to *LMO2* indicates that higher expression correlates with longer survival. The positive value for *CCND2* indicates that higher expression correlates with shorter survival.

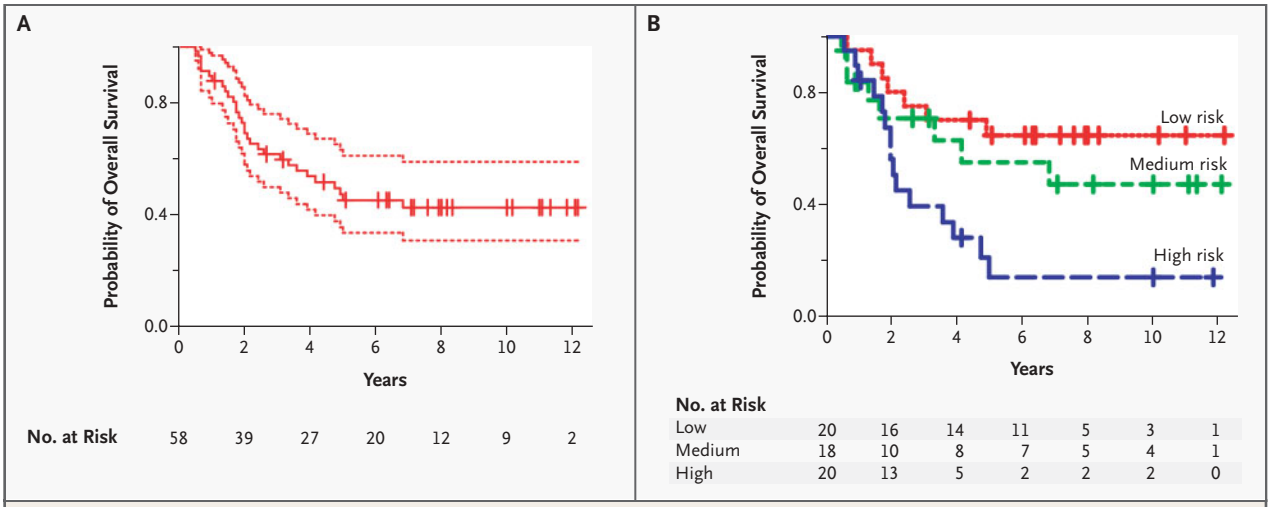


**Figure 2. Development of the Six-Gene Model.**

Panel A shows Kaplan–Meier estimates of overall survival in the 66 patients with diffuse large-B-cell lymphoma, analyzed by quantitative reverse-transcriptase polymerase chain reaction with TaqMan probe-based assays. The dotted lines represent 95 percent confidence intervals. Panel B shows Kaplan–Meier curves for overall survival in the three groups (at low, medium, and high risk of death) as defined by a prediction model based on the weighted expression of six genes (*LMO2, BCL6, FN1, CCND2, SCYA3,* and *BCL2*). According to log-likelihood estimates, P=0.001 for the model based on a continuous variable, and P=0.02 for the model based on the three discrete groups shown in the figure.

We ranked the patients with known IPI scores according to their mortality-predictor scores and divided them into three groups according to whether they had a low, medium, or high risk of death (low risk, lower than 0.063; medium risk, from −0.063 to <0.093; and high risk, 0.093 or higher). Table 2 shows the clinical characteristics of the patients according to these risk groups. The rates of overall survival at five years in the low-risk, medium-risk, and high-risk groups were 65 percent, 49 percent, and 15 percent, respectively (P=0.004). The mean survival times were 8.7 years (95 percent confidence interval, 4.9 to not reached), 7.1 years (95 percent confidence interval, 3.3 to not reached), and 3.8 years (95 percent confidence interval, 1.8 to 5.0), respectively (Fig. 2).

To test the validity of this model, we applied it to published microarray gene-expression data from Shipp et al.[5] (Fig. 3A and 3B) and from Rosenwald et al.[6] (Fig. 3C and 3D). These tests confirmed the ability of the model to predict survival. In the smaller cohort study of patients with diffuse large-B-cell lymphoma, reported by Shipp et al.,[5] the overall survival in the medium-risk group was similar to that in the high-risk group. However, the medium-risk group did have an intermediate risk in the larger cohort of patients that Rosenwald et al. analyzed.[6]

We next investigated whether our model could add prognostic value beyond that of the IPI. Among patients in our sample who were at high risk for death according to the IPI, the six-gene model could further subdivide the patients into those likely to have longer survival and those likely to have shorter survival, in a manner similar to what we observed in the entire group of 66 patients (P=0.006) (data not shown). But of our 66 patients, too few were in the IPI group with the lowest risk of death for our findings to achieve statistical significance. We therefore tested the added value of the six-gene model by analyzing the larger data set reported by Rosenwald et al.[6] (Fig. 4). We used that study group's three subdivisions, which were based on the IPI (low, medium, and high). In some of these groups, the number of patients was small. But in each stratum of the IPI, we could identify a group with an especially low probability of survival (Fig. 4, blue lines). Thus, by identifying the patients who had either medium-risk or high-risk scores on the IPI along with a high-risk expression profile (the bottom, or highest-risk, group), it was possible to identify the group of approximately 30 percent of all patients with diffuse large-B-cell lymphoma who had especially short survival.

Recently, Barrans et al.[9] reported that immunohistochemical analysis of antibodies to germinal-center markers and Bcl-2, in combination with the

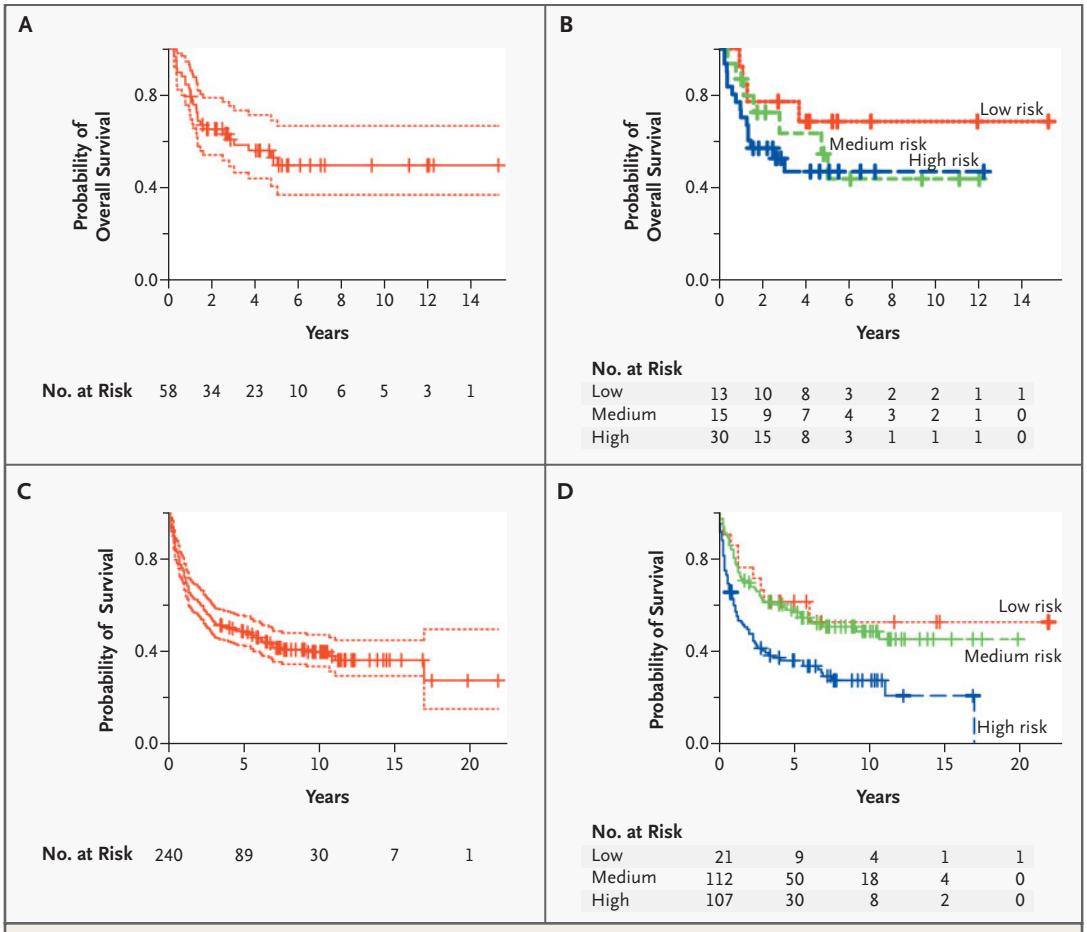


**Figure 3. Validation of the Performance of the Six-Gene Model with the Use of Data from Oligonucleotide Microarrays (Panels A and B) and cDNA Microarrays (Panels C and D).**

Panel A shows Kaplan–Meier estimates of overall survival in all 58 patients with diffuse large-B-cell lymphoma reported by Shipp et al.,[5] and Panel B Kaplan–Meier estimates of overall survival in the 58 patients after subdivision into three groups (at low, medium, and high risk of death) on the basis of the six-gene model for prediction. The dotted lines represent 95 percent confidence intervals. According to log-likelihood estimates, P=0.02 for the model as a continuous variable, and P=0.31 for the model as a class. Similar analyses of the data on the 240 patients with diffuse large-B-cell lymphoma reported by Rosenwald et al.[6] are shown in Panels C and D. P<0.001 for the model based on a continuous variable and for the model based on the three discrete groups shown in the figure.

IPI, could improve risk stratification among patients with diffuse large-B-cell lymphoma. Colomo et al.[29] could not confirm this result but did demonstrate the predictive power of expression of the Bcl-2 protein. Comparison of the predictive power of our model of the expression of six genes with that of a gene-expression model based on only *BCL6* (a germinal-center marker) and *BCL2* showed that the six-gene model predicted overall survival better among patients in our cohort and in the cohorts studied by Shipp et al.[5] and Rosenwald et al.[6] (see the Supplementary Appendix).

## DISCUSSION

As new therapies for lymphoma become available, it will be increasingly important to identify patients who do not benefit from current treatments and who may be candidates for early treatment with these new approaches. The IPI has proved useful for identifying such patients with diffuse large-B-cell lymphoma,[3] but we found that the molecular characteristics of lymphoma can add further predictive power. Simultaneous analysis of the expression of thousands of genes in diffuse large-B-cell lympho-

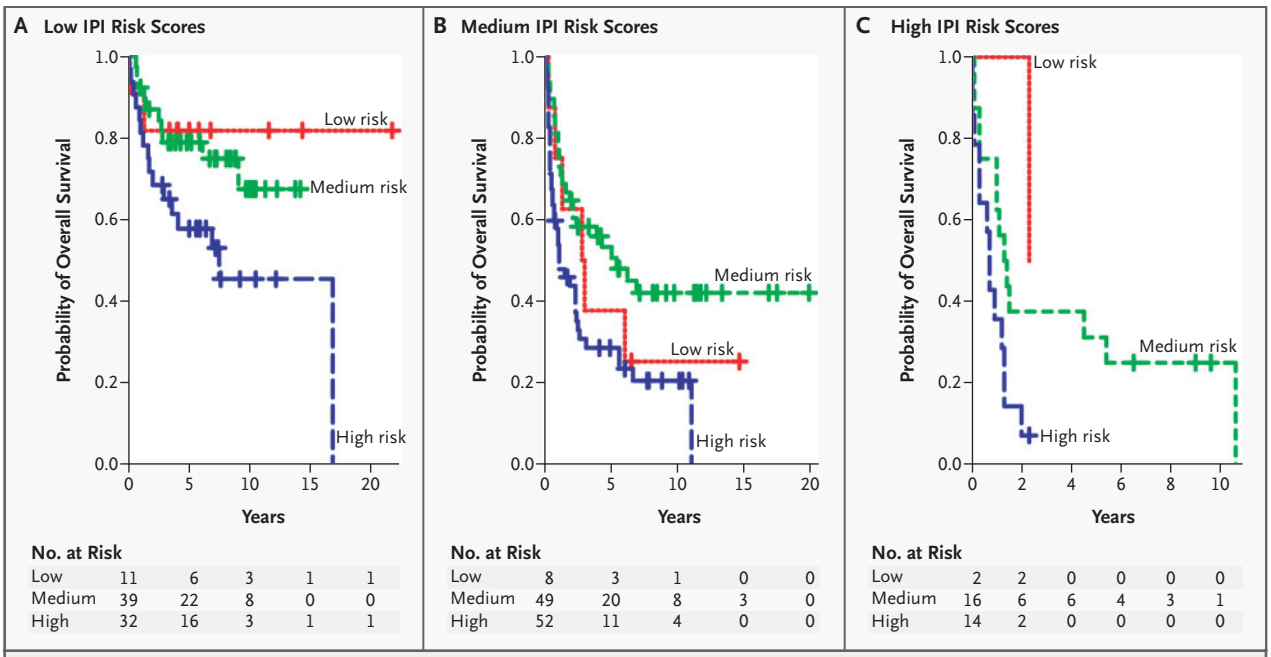


**Figure 4. The Six-Gene Model and the International Prognostic Index.**

The Kaplan–Meier estimates show overall survival for groups of patients with low-risk (Panel A), medium-risk (Panel B), and high-risk (Panel C) scores on the International Prognostic Index, as reported by Rosenwald et al.,[6] after subdivision into three groups (at low, medium, and high risk for death) on the basis of the six-gene model for prediction. According to log-likelihood estimates, P=0.01, P=0.002, and P=0.16 for the model based on a continuous variable applied to the low-risk, medium-risk, and high-risk groups, respectively, and P=0.02, P=0.003, and P=0.01, respectively, for the model based on the three discrete groups shown in the figure.

ma with the use of cDNA and oligonucleotide microarrays by several groups has provided a rich source of data that can be correlated with the clinical outcome.[4-6] Each of these studies has produced lists of genes for use in the stratification of the risk of death among patients with diffuse large-B-cell lymphoma; however, validating such models in other, unrelated groups of patients is essential. In addition, more convenient methods for the measurement of gene expression need to be developed.

The aim of our study was to identify a small group of genes whose expression predicts survival in patients with diffuse large-B-cell lymphoma and can be readily measured. To this end, we evaluated the prognostic significance of 36 genes that were chosen on the basis of previous reports of their prognostic potential and on the basis of our own analysis of the existing microarray data. The results of this evaluation allowed us to design a model consisting of six genes that predicts overall survival in patients with diffuse large-B-cell lymphoma. The model assigned the 66 patients in our series and the 58 and 240 patients with lymphomas analyzed by Shipp et al.[5] and Rosenwald et al.,[6] respectively, to three prognostic groups. Our method stratified all three groups of patients according to risk and was independent of the IPI. Furthermore, our 6-gene model was as robust as the 17-gene model developed by Rosenwald et al.[6] and was also independent of the IPI in its ability to predict the outcome (data not shown). Moreover, our six-gene model could be applied to the data sets derived from the measurement of gene expression by three methods[30-32] without the need for shifting or scaling of the expression data to match their mean and variance levels.[33]

The genes in our model occur in the germinal-center B-cell signature (*LMO2* and *BCL6*), the activated B-cell signature (*BCL2*, *CCND2*, *SCYA3*), and the lymph-node signature (*FN1*).[4,6] However, since many of the other genes in these signatures have no independent predictive power in our model, the model we propose probably refines these signatures by identifying the genes with the highest level of independent prognostic power.

In this study, the expression of *LMO2*, *BCL6*, and *FN1* correlated with prolonged survival. *LMO2*[34] has

an important role in erythropoiesis and angiogenesis[35,36] and is the most frequent site of chromosomal translocation in childhood T-cell acute lymphoblastic leukemia.[34] It is not expressed in normal T lymphocytes,[37] but it is expressed at high levels in germinal-center lymphocytes.[4] *LMO2* has also been implicated in T-cell leukemia, developing after retrovirus-based gene therapy of X-linked severe combined immunodeficiency.[37] The relationship between its ability to cause T-cell leukemia and its correlation with prolonged survival in diffuse large-B-cell lymphoma is unclear.

The *BCL6* gene encodes a transcriptional repressor,[38-40] is normally expressed in B cells and CD4+ T cells within the germinal center, and controls germinal-center formation and T-cell–dependent immune responses.[41-43] It is expressed in non-Hodgkin's lymphomas that originate from germinal-center B cells. *BCL6* expression has been reported to predict survival in patients with diffuse large-B-cell lymphoma,[8] and our findings confirm this observation.

Fibronectin 1 (*FN1*), an extracellular glycoprotein, is a ligand for the integrin family of cell-adhesion receptors that regulate cytoskeletal organization. The expression of *FN1*, by hepatocytes, stromal fibroblasts, and some tumor cells,[44] has been associated with metastasis.[45] *FN1* is a component of the lymph-node signature in diffuse large-B-cell lymphoma.[6] Its expression may reflect the presence of mesenchymal cells in the biopsy specimen or a response of the lymph node to tumor cells, since we found only low levels of *FN1* transcripts in the purified lymphoma cells (data not shown).

In our study, the expression of *BCL2*, *CCND2*, and *SCYA3* correlated with short survival. These three genes are included in the activated B-cell–like signature that we and others have associated with short survival.[4] The Bcl-2 protein is present at low levels in normal germinal-center B cells but at increased levels in some non-Hodgkin's lymphoma cells that have a t(14;18) translocation.[4,46] Bcl-2 prevents apoptosis, and elevated levels of this protein, as detected immunohistologically, serve as an independent marker of a poor prognosis in patients with diffuse large-B-cell lymphoma.[20-23]

*CCND2* encodes a protein belonging to the cyclin family, whose members are characterized by dramatic periodicity in the amount of protein that is present during the cell cycle. *CCND2* controls the progression of the cell cycle from the G1 phase to the S phase by serving as a mediator of S-phase commitment and DNA synthesis.[47] Overexpression of *CCND2* occurs in chronic lymphocytic leukemia and mantle-cell lymphoma.[48]

*SCYA3*, otherwise known as MIP-1-alpha or CCL3, is a CC chemokine that recruits a variety of cells to sites of inflammation.[49] Its function in B-cell lymphomas is unknown, but it is expressed mainly in the activated B-cell–like subgroup of diffuse large-B-cell lymphoma.[4] The promoter regions of the *CCND2* and *SCYA3* genes contain high-affinity binding sites for Bcl-6, and the expression of these two genes is repressed by Bcl-6.[25] This observation underscores the relations among the individual genes that we and others have implicated as determining the prognosis for patients with diffuse large-B-cell lymphoma.

Diffuse large-B-cell lymphoma is heterogeneous and may require a risk-adjusted approach to therapy. For simplicity, one can focus on the high-risk group in our model, because patients in this group have an especially poor prognosis with respect to long-term survival in the three data sets that we examined. Even among patients with a low risk of death according to the IPI, our six-gene model identified patients with a five-year survival rate of only 57 percent. Among patients with a medium or high clinical risk of death (medium or high scores on the IPI), the five-year survival rate in the high-risk group in our model is less than 27 percent. This group, representing approximately one third of all patients with diffuse large-B-cell lymphoma, may require a different therapeutic approach from that used in other patients. The six-gene model and quantitative PCR can most likely be used to identify patients who may benefit from new treatments.

Supported by grants (CA33399 and CA34233) from the Public Health Service, National Institutes of Health.

## REFERENCES

1. Jaffe ES. Histopathology of the non-Hodgkin's lymphomas and Hodgkin's disease. In: Canellos GP, Lister TA, Sklar JL, eds. The lymphomas. Philadelphia: W.B. Saunders, 1998:77-106.
2. Vose JM. Current approaches to the management of non-Hodgkin's lymphoma. Semin Oncol 1998;25:483-91.
3. 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.
4. 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.
5. Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 2002;8:68-74.
6. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse

large-B-cell lymphoma. N Engl J Med 2002; 346:1937-47.
**7.** Lossos IS, Alizadeh AA, Rajapaksa R, Tibshirani R, Levy R. HGAL is a novel interleukin-4-inducible gene that strongly predicts survival in diffuse large B-cell lymphoma. Blood 2003;101:433-40.
**8.** Lossos IS, Jones KD, Warnke R, et al. The expression of a single gene, BCL-6, strongly predicts survival in patients with diffuse large B-cell lymphoma. Blood 2001; 98:945-51.
**9.** Barrans SL, Carter I, Owen RG, et al. Germinal center phenotype and bcl-2 expression combined with the International Prognostic Index improves patient risk stratification in diffuse large B-cell lymphoma. Blood 2002;99:1136-43.
**10.** Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Natl Acad Sci U S A 2002;99:6562-6.
**11.** Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J Natl Cancer Inst 2003;95: 14-8.
**12.** Harris NL, Jaffe ES, Stein H, et al. A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. Blood 1994;84:1361-92.
**13.** Lossos IS, Czerwinski DK, Wechser MA, Levy R. Optimization of quantitative real-time RT-PCR parameters for the study of lymphoid malignancies. Leukemia 2003; 17:789-95.
**14.** Thorstenson YR, Shen P, Tusher VG, et al. Global analysis of ATM polymorphism reveals significant functional constraint. Am J Hum Genet 2001;69:396-412.
**15.** Krenacs L, Himmelmann AW, Quintanilla-Martinez L, et al. Transcription factor B-cell-specific activator protein (BSAP) is differentially expressed in B cells and in subsets of B-cell lymphomas. Blood 1998;92: 1308-16.
**16.** Moller MB, Kania PW, Ino Y, et al. Frequent disruption of the RB1 pathway in diffuse large B cell lymphoma: prognostic significance of E2F-1 and p16INK4A. Leukemia 2000;14:898-904.
**17.** Drillenburg P, Wielenga VJ, Kramer MH, et al. CD44 expression predicts disease outcome in localized large B cell lymphoma. Leukemia 1999;13:1448-55.
**18.** Ichikawa A, Kinoshita T, Watanabe T, et al. Mutations of the p53 gene as a prognostic factor in aggressive B-cell lymphoma. N Engl J Med 1997;337:529-34.
**19.** Koduru PR, Raju K, Vadmal V, et al. Correlation between mutation in P53, p53 expression, cytogenetics, histologic type, and survival in patients with B-cell non-Hodgkin's lymphoma. Blood 1997;90:4078-91.
**20.** Gascoyne RD, Adomat SA, Krajewski S, et al. Prognostic significance of Bcl-2 protein expression and Bcl-2 gene rearrangement in diffuse aggressive non-Hodgkin's lymphoma. Blood 1997;90:244-51.
**21.** Kramer MH, Hermans J, Parker J, et al. Clinical significance of bcl2 and p53 protein expression in diffuse large B-cell lymphoma: a population-based study. J Clin Oncol 1996;14:2131-8.
**22.** Hermine O, Haioun C, Lepage E, et al. Prognostic significance of bcl-2 protein expression in aggressive non-Hodgkin's lymphoma. Blood 1996;87:265-72.
**23.** Hill ME, MacLennan KA, Cunningham DC, et al. Prognostic significance of Bcl-2 expression and bcl-2 major breakpoint region rearrangement in diffuse large cell non-Hodgkin's lymphoma: a British National Lymphoma Investigation Study. Blood 1996; 88:1046-51.
**24.** Adida C, Haioun C, Gaulard P, et al. Prognostic significance of survivin expression in diffuse large B-cell lymphomas. Blood 2000;96:1921-5.
**25.** Shaffer AL, Yu X, He Y, Boldrick J, Chan EP, Staudt LM. BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. Immunity 2000;13:199-212.
**26.** Zhang A, Ohshima K, Sato K, et al. Prognostic clinicopathologic factors, including immunologic expression in diffuse large B-cell lymphomas. Pathol Int 1999;49:1043-52.
**27.** Cox DR. Regression models and life-tables. J R Stat Soc [B] 1972;34:187-220.
**28.** Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116-21. [Erratum, Proc Natl Acad Sci U S A 2001;98:10515.]
**29.** Colomo L, Lopez-Guillermo A, Perales M, et al. Clinical impact of the differentiation profile assessed by immunophenotyping in patients with diffuse large B-cell lymphoma. Blood 2003;101:78-84.
**30.** Li J, Pankratz M, Johnson JA. Differential gene expression patterns revealed by oligonucleotide versus long cDNA arrays. Toxicol Sci 2002;69:383-90.
**31.** Yuen T, Wurmbach E, Pfeffer RL, Ebersole BJ, Sealfon SC. Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. Nucleic Acids Res 2002; 30:e48 (Web only). (Accessed April 5, 2004, at http://www3.oup.co.uk/nar/methods.)
**32.** Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS. Analysis of matched mRNA measurements from two different microarray technologies. Bioinformatics 2002; 18:405-12.
**33.** Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci U S A 2003; 100:9991-6.
**34.** Boehm T, Foroni L, Kaneko Y, Perutz MF, Rabbitts TH. The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. Proc Natl Acad Sci U S A 1991;88: 4367-71.
**35.** Warren AJ, Colledge WH, Carlton MB, Evans MJ, Smith AJ, Rabbitts TH. The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. Cell 1994;78:45-57.
**36.** Yamada Y, Pannell R, Forster A, Rabbitts TH. The oncogenic LIM-only transcription factor Lmo2 regulates angiogenesis but not vasculogenesis in mice. Proc Natl Acad Sci U S A 2000;97:320-4.
**37.** Royer-Pokora B, Loos U, Ludwig WD. TTG-2, a new gene encoding a cysteine-rich protein with the LIM motif, is overexpressed in acute T-cell leukaemia with the t(11;14)(p13;q11). Oncogene 1991;6:1887-93.
**38.** Chang CC, Ye BH, Chaganti RS, Dalla-Favera R. BCL-6, a POZ/zinc-finger protein, is a sequence-specific transcriptional repressor. Proc Natl Acad Sci U S A 1996;93: 6947-52.
**39.** Kerckaert JP, Deweindt C, Tilly H, Quief S, Lecocq G, Bastard C. LAZ3, a novel zinc-finger encoding gene, is disrupted by recurring chromosome 3q27 translocations in human lymphomas. Nat Genet 1993;5:66-70.
**40.** Seyfert VL, Allman D, He Y, Staudt LM. Transcriptional repression by the proto-oncogene BCL-6. Oncogene 1996;12:2331-42.
**41.** Cattoretti G, Chang CC, Cechova K, et al. BCL-6 protein is expressed in germinal-center B cells. Blood 1995;86:45-53.
**42.** Dent AL, Hu-Li J, Paul WE, Staudt LM. T helper type 2 inflammatory disease in the absence of interleukin 4 and transcription factor STAT6. Proc Natl Acad Sci U S A 1998; 95:13823-8.
**43.** Ye BH, Cattoretti G, Shen Q, et al. The BCL-6 proto-oncogene controls germinal-centre formation and Th2-type inflammation. Nat Genet 1997;16:161-70.
**44.** Mosher DF. A role for fibronectin in self-repair after ischemic injury. Nat Med 2001;7:290-2.
**45.** Clark EA, Golub TR, Lander ES, Hynes RO. Genomic analysis of metastasis reveals an essential role for RhoC. Nature 2000; 406:532-5. [Erratum, Nature 2001;411:974.]
**46.** Kramer MH, Hermans J, Wijburg E, et al. Clinical relevance of BCL2, BCL6, and MYC rearrangements in diffuse large B-cell lymphoma. Blood 1998;92:3152-62.
**47.** Sherr CJ. Mammalian G1 cyclins and cell cycle progression. Proc Assoc Am Physicians 1995;107:181-6.
**48.** Delmer A, Ajchenbaum-Cymbalista F, Tang R, et al. Overexpression of cyclin D2 in chronic B-cell malignancies. Blood 1995; 85:2870-6.
**49.** Proost P, Wuyts A, van Damme J. The role of chemokines in inflammation. Int J Clin Lab Res 1996;26:211-23.