HGAL is a novel interleukin-4–inducible gene that strongly predicts survival in diffuse large B-cell lymphoma

Izidore S. Lossos, Ash A. Alizadeh, Ranjani Rajapaksa, Robert Tibshirani, and Ronald Levy

Introduction

Diffuse large B-cell lymphomas (DLBCLs) constitute 30% to 40% of adult non-Hodgkin lymphoma (NHL).1 There is a consensus that DLBCL represents a diverse group of neoplasms with heterogeneous genetic abnormalities, clinical features, treatment responses, and prognoses.2 Previous attempts to subclassify these neoplasms on morphologic grounds have been hampered by irreproducibility, thus leading to their categorization as one single group in the Revised European American Lymphoma (REAL) classification.2,3 In view of this heterogeneity, only 50% of DLBCL patients are cured with standard chemotherapy. Therefore, the establishment of prognostic models based on pretreatment characteristics of patients or tumors is of paramount importance to guide the choices of treatment intensities. The International Prognostic Indicator (IPI), based on clinical characteristics at diagnosis, has been constructed and successfully used to define prognostic subgroups in DLBCL.4 However, the differences in clinical features and in treatment responses of DLBCL are probably caused by the marked genetic and molecular heterogeneity underlying disease aggressiveness and tumor progression.

Examination of gene expression profiles in DLBCL tumors and application of a pattern recognition algorithm termed hierarchical clustering identified 2 molecularly distinct forms of the disease: germinal center B-cell–like DLBCL, characterized by the expression of genes normally expressed in germinal center B cells, and activated B-cell–like DLBCL, characterized by the expression of genes normally induced during in vitro activation of B cells.5 Patients with these 2 forms of DLBCL were found to have very different prognoses: those with germinal center B-cell–like DLBCL had a significantly better overall survival than those with activated B-cell–like DLBCL. However, the relative prognostic contribution of the individual genes defining these 2 DLBCL subgroups could not be assessed by this method. This aim can be achieved by using clinical data to supervise the discovery of genes with expression patterns that correlate with outcome, as was recently reported.6,7 This supervised approach may allow identification of genes that play a role in determining prognosis and pathophysiology, including the discovery of previously unknown genes of major clinical relevance among the multiple expressed sequence tags (ESTs) present on the arrays.

By conducting a search for genes predicting DLBCL outcome, we have identified and cloned a novel gene, termed HGAL (human germinal center–associated lymphoma). This gene is mainly expressed in germinal center (GC) B cells and is stimulated specifically by the lymphokine interleukin-4 (IL-4).

Patients, materials, and methods

Cell lines, normal tissues, and tumor specimens

Ten non-Hodgkin lymphoma (NHL) cell lines (Raji, Jurkat, Daudi, Ramos, SU-DHL6, OCI-Ly3, OCI-Ly8, OCI-Ly10, HFi, and RC-K8) and 2 myeloid cell lines (HL60 and K562) were selected for this study. All cell lines except OCI-Ly3 and OCI-Ly10 were grown in RPMI 1640 medium (Fisher Scientific, Santa Clara, CA), supplemented with 10% fetal calf serum (FCS), 2 mM L-glutamine (Gibco BRL, Grand Island, NY), and penicillin/streptomycin (Gibco BRL). The OCI-Ly3 and OCI-Ly10 cell lines were grown in Iscove modified Dulbecco medium (IMDM; Fisher Scientific) supplemented with 20% fresh human plasma and 50 μM mercaptoethanol.

Biopsy specimens from patients with primary untreated DLBCL (54 patients), follicle-center lymphoma (FCL) (21 patients), chronic lymphocytic leukemia (CLL) (16 patients), mantle cell lymphoma (MCL) (4 patients), nodal marginal zone lymphoma (MZL) (3 patients), and T-cell lymphoblastic lymphoma (T-LL) (2 patients), classified according to the Revised European-American Lymphoma Classification,2 were used in this study. The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked “advertisement” in accordance with 18 U.S.C. section 1734.
study. DLBCL tissues were chosen randomly from a collection of specimens obtained during the course of diagnostic procedures between 1983 and 1993. All specimens were distinct from the tumor samples used in our previous analysis of gene expression profiles in DLBCL. Tumor tissues were stored either as fresh-frozen biopsy specimens embedded in TissueTek Optimal Cutting Temperature (OCT) compound 4583 (Miles, Elkhart, IN) and preserved at ~80°C or as frozen viable cell suspensions in 10% dimethyl sulfoxide in liquid nitrogen. DLBCL specimens were chosen based on diagnosis of primary DLBCL; availability of the tissue specimen obtained at diagnosis, before the initiation of therapy; treatment with an anthracycline-containing chemotherapy; and follow-up at Stanford University Hospital and availability of the outcome data. In all DLBCL patients chosen for the study, disease dissemination was evaluated before treatment by physical examination, bone marrow biopsy, and computed tomography of the chest, abdomen, and pelvis. All patients were staged according to the Ann Arbor system. IPI scores could be determined from the records of 49 of these patients. All DLBCL tumors had the histologic appearance of centroblastic-large-cell lymphomas demonstrating diffuse patterns of involvement without evidence of residual reactive follicles.

Germinal center (GC) B cells were purified from 3 human tonsils, as previously described, and pooled. Peripheral blood mononuclear cells from healthy donors were isolated by Ficoll-Hypaque density centrifugation (Amersham Pharmacia Biotech, Piscataway, NJ). B cells were enriched to more than 90% by human B-cell enrichment cocktail (StemCell Technologies, Vancouver, BC, Canada), as determined by fluorescence-activated cell sorter (FACS) analysis. Enriched B cells were plated at 5 × 10^5 cells per well in 6-well plates (Corning Glassworks, Corning, NY) in IMDG (Gibco BRL) supplemented with 2% fetal calf serum (FCS), 0.5% bovine serum albumin (Sigma, St Louis, MO), 50 μg/mL human transferrin (Sigma), 5 μg/mL bovine insulin (Sigma), and 15 μg/mL gentamicin (Cellgro-Mediatac, Herndon, VA) at 37°C in 5% CO2. The cells were stimulated with 100 U/mL IL-4 (R&D Systems, Minneapolis, MN) and goat F(ab')2 antihuman μ antibodies (Biosource International, Camarillo, CA) for 6 and 24 hours. For CD40 activation, purified B cells were stimulated by culture on irradiated (55 Gy) CD40L-expressing mouse L cells (a gift of Yong-Jun Liu, DNAX, Palo Alto, CA).

RNA isolation and reverse transcription reaction

Total cellular RNA was isolated from 5 × 10^5 to 1.0 × 10^6 cells using the RNeasy kit (Qiagen, Valencia, CA) according to the manufacturer’s instructions. Total RNA isolated from the cells was quantified using spectrophotometric OD260 measurements. RNA (2 μg) was reverse transcribed using the High-Capacity cDNA Archive kit (Applied Biosystems, Foster City, CA) according to the manufacturer’s protocol with a minor modification that consisted of the addition of RNase inhibitor (Applied Biosystems) at a final concentration of 1 U/μL. Samples were incubated at 25°C for 10 minutes and 37°C for 120 minutes. For the determination of HGAL mRNA expression in normal tissues, we used Multiple Tissue cDNA (Clontech, Palo Alto, CA).

cDNA cloning and sequencing

Clone 814622 (GE: 2210537) was used for the design of 5′ SMART rapid amplification of cDNA ends (RACE) polymerase chain reaction (PCR) primer, 5′-GCCAAAGAGGGTAGTGGGATTACG-3′. We performed 5′ SMART RACE cDNA amplification using Advantage 2 polymerase mix according to the manufacturer’s protocol (Clontech). PCR amplicon was cloned into a TA-PCR cloning vector (Invitrogen, Carlsbad, CA). After the transformation of competent Escherichia coli (1 Shot INV F; Invitrogen) and plating on selective agar (50 μg/mL kanamycin, 40 μL of 40 μg/mL X-gal), 10 white colonies were picked for plasmid purification using QIAprep kit (Qiagen, Valencia, CA). DNA sequencing was performed on a 373 automatic DNA sequencer (Applied Biosystems) using ABI Prism Big Dye Terminator Kit (Perkin Elmer, Foster City, CA) as recommended by the manufacturer. HGAL full-length cloning was completed with 3′ SMART RACE cDNA amplification. The cDNA sequence was confirmed by reverse transcription (RT)–PCR spanning the coding region of the gene.

Northern blot analysis

Total RNA from normal spleen (Ambion, Austin, TX) and DLBCL cell lines were size-fractionated in 1% agarose–glyoxal–dimethyl sulfoxide gels and transferred to Hybond N+ positively charged nylon membranes (Amersham Pharmacia Biotech, Piscataway, NJ) according to standard protocols. The membranes were hybridized in ULTRAhyb (Ambion) buffer with entire coding region HGAL cDNA probe, labeled with [α-32P]dATP (Amersham Pharmacia Biotech, Piscataway, NJ) during asymmetric PCR.

Real-time polymerase chain reaction measurement of HGAL mRNA expression

HGAL mRNA expression was measured using the TaqMan technology on an ABI Prism 7900HT Sequence Detector System (Applied Biosystems). Probe and primers were designed using Primer Express software (Applied Biosystems) and were chosen to hybridize to sequences at the junction between 2 exons to avoid amplification of genomic DNA, as follows: forward primer, 5′-CCCCAAGAGGTAATGAAAAAGGTT-3′ (900 nM); reverse primer, 5′-GGGTATGGCAGTTCTCCTGAATG-3′ (900 nM); probe, 5′-CCATCAGGAGCAATGTT-3′ (250 μM), labeled with 6-carboxy-fluorescein phosphoramidite (FAM) at the 5′ end and with nonfluorescent quencher (NFQ) at the 3′ end. PCR reactions were prepared in a final volume of 20 μL, with final concentrations of 1× TaqMan Universal PCR Master Mix (Applied Biosystems) and cDNA equivalent to 20 ng input RNA. Reaction mixtures were assembled at 4°C, followed by PCR consisting of 50°C UNG initiation for 2 minutes, AmpliTaq Gold activation at 95°C for 10 minutes, followed by 95°C for 15 seconds and 60°C for 1 minute, for 40 cycles. Each PCR run included the 5 points of the calibration curve for HGAL and GAPDH (5-fold diluted human RNA), a no-template control, the calibrator Raji cDNA, and the patient’s samples, all in triplicate. Threshold cycle (Ct) was chosen at 10 times the standard deviation of the baseline fluorescence signal of the first few PCR cycles.

Because the quality of the RNA (ie, extent of RNA degradation) and consequently the amount of cDNA added to each reaction was difficult to assess, we also quantified the level of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as the endogenous RNA control using a commercially available kit (PE Applied Biosystems), and each sample was normalized on the basis of its GAPDH content, as was previously reported. For each experimental sample, the amount of the target HGAL and the endogenous reference (GAPDH) were determined from the calibration curves obtained by serial dilution of control human RNA (Applied Biosystems). HGAL amount was then divided by the endogenous reference (GAPDH) amount to obtain a normalized value. To allow the relative expression of HGAL gene to be compared across all the tested samples, HGAL/GAPDH expression ratios were also normalized to the HGAL/GAPDH values concomitantly measured in Raji cells (calibrator), as suggested in ABI 7700 User Bulletin 2 (Applied Biosystems).

Search for HGAL gene mutations

To determine whether the HGAL gene is somatically mutated, high-molecular-weight DNA was extracted from 5.0 × 10^5 cells by a commercially available kit (QIamp Tissue Kit; Qiagen, Valencia, CA) from 19 NHL (8 DLBCL and 11 FCL) specimens, 2 DLBCL cell lines and normal T lymphocytes from 2 tumor specimens, enriched by CD3 microbeads (Miltenyi Biotec, Auburn, CA). Approximately 1000 bp from the transcription initiation site, including the first exon and the 5′ portion of the first intron, were amplified and directly sequenced using 2 pairs of primers: Mut-1 forward-5′-AGCACAAGGCG- AAGAAGGAGGTT-3′; Mut-1 reverse-5′-CTGAAAAGGGTTGAGTTTGAA- TGC-3′; Mut-2 forward-5′-GTCAAAATCACCACCTCTTCTTACG-3′; and Mut-2 reverse-5′-TCTGTTGTCCACTCTCAGTTGAC-3′. PCR was performed in a final volume of 50 μL containing 0.5 μM of each primer, 20 μM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl2, 200 μM each dNTP, and 2.5 U Taq DNA polymerase (Gibco BRL). The PCR conditions were 1 cycle at 94°C for 5 minutes, 56°C (Mut-1 primers) or 54.5°C (Mut-2 primers) for 1 minute, and 72°C for 3 minutes, 30 cycles at
94°C for 30 seconds, 56°C (Mut-1 primers) or 54.5°C (Mut-2 primers) for 30 seconds, and 72°C for 30 seconds, and 1 cycle at 72°C for 7 minutes. PCR products were analyzed by 2% agarose gel electrophoresis and stained with ethidium bromide. Bands of appropriate size were excised from the gels purified by adsorption to a silica matrix (QIAquick columns; Qiagen). Amplicons were directly sequenced in both directions using the same primers.

Statistical analysis
To identify genes potentially predicting DLBCL outcome, we have applied the significance analysis of microarrays (SAM) method9 to previously published DLBCL gene expression profiling data.5 The SAM method ranks genes according to their Cox proportional hazards regression score. A significant gene list is formed by cutting off the ranked list at a given threshold. The threshold is chosen by assessment of the false discovery rate (FDR), estimated by random permutation of the survival times. We chose an FDR of 15%. Comparison of clinical characteristics between DLBCL patients with high and low HGAL expression was performed by Student ttest for age and by Fisher exact test for all the other variables using GraphPad Prism version 2.0 software (San Diego, CA). Overall survival (OS) time of patients with DLBCL was calculated from the date of diagnosis until the date of death or last follow-up examination. Disease-free survival (DFS) was measured as the interval between the date of complete remission after induction treatment and the date of relapse, death, or last follow-up evaluation. Survival curves were estimated using the product-limit method of Kaplan-Meier and were compared using the log-rank test. Multivariate regression analysis according to the Cox proportional hazards regression model10 with OS as the dependent variable was used to adjust the effect of HGAL expression for IPI and BCL6 expression. Two-tailed P < .05 was considered significant.

Results
Identification and cloning of the HGAL gene
We performed an analysis of previously reported data on the gene expression in DLBCL5 supervised by survival information. For this purpose, we used the SAM method.9 This analysis identified 234 genes whose expression correlated strongly with clinical outcome of DLBCL patients (data not shown). The gene whose expression best predicted DLBCL OS was an EST—UniGene cluster 49614 (clone 814622; GI: 2210537). It had nucleotide sequence homology to the previously reported mouse GC-specific gene M17.11 In this data set, DLBCL patients with mRNA expression of clone 814622 greater than the median expression of all analyzed patients exhibited significantly longer OS (P = .008) than patients with low expression (Figure 1). By hierarchical clustering, this gene was contained within the GC gene cluster. On the microarray, its expression was high in GC lymphocytes, intermediate in memory B cells, and relatively low in peripheral blood B cells.5 In tumors, its expression was high in FCL tumors, low in CLL, and heterogeneous in DLBCL specimens.5 These observations suggested that this gene is up-regulated at specific stages of B-cell differentiation, especially in GC lymphocytes. A full-length cDNA was cloned from sorted GC lymphocytes and from the Ramos cell line and was termed HGAL (GenBank accession number AF521911) (Figure 2A). The full-length sequence of HGAL cDNA perfectly matched another EST on the array (clone 1339726; GI: 2883522), which also was among the top genes that predicted DLBCL outcome and had an expression pattern similar to that of clone 814622.5 PCR amplification of the HGAL mRNA by 3′ RACE revealed the presence of 2 amplicons—a major amplicon of 1659 bp and a minor amplicon of approximately 3500 bp, the latter containing the

Figure 1. Kaplan-Meier curves of OS in DLBCL patients. Kaplan-Meier curves of OS in 42 patients with DLBCL according to high (greater than the median expression of all analyzed patients, §) and low (lower than the median expression of all analyzed patients, ¶) HGAL mRNA expression as determined in previous microarray analysis of DLBCL.5

Figure 2. Analysis of HGAL sequence. (A) Nucleotide sequence of HGAL cDNA and putative amino acid sequence of HGAL protein. (B) Comparison of ITAM motif in HGAL protein to the ITAM motifs of M17 and representative receptor proteins.
sequence of the shorter amplicon and differing in the length of the 3′ untranslated sequence. Northern blot analysis of normal spleen and several NHL cell lines confirmed the PCR findings and demonstrated the presence of 2 RNA transcripts of approximately 1.7 kb (major transcript) and 3.5 kb (Figure 3).

Alignment of the sequenced cDNA to the GenBank database and BLAST search of the human genome identified a perfectly aligning genomic DNA sequence (AC024964 and NT_022484). The gene was located on chromosome 3q13. Comparison of the genomic sequence to the cDNA sequence revealed that HGAL spans 11 kb. Recognition of a Kozak sequence and search for the longest open reading frame (ORF) led to the identification of a putative ORF extending from exon 1 to exon 6 and encoding a 178-amino acid (aa) protein, with 51% identity and 62% similarity to the mouse GC-expressed M17 protein. The HGAL gene product had a hydrophilic profile with no predicted transmembrane domain, and it lacked a nuclear localization sequence. The C-terminal portion of HGAL protein is proline rich. Similar to mouse M17 protein, HGAL contains a modified immunoreceptor tyrosine-based activation motif (D/EX2D/EX;YX2LYX2LY or termed ITAM) (Figure 2B). ITAM fragments are usually found within the cytoplasmic domains of transmembrane receptors, and they play a role in signal transduction in B and T lymphocytes. HGAL and M17 are the only 2 known nonreceptor proteins that contain such a motif. However, the spacing between the 2 YXXL regions is greater than 7 aa. A potential contact site to SH2 domains was identified at aa positions 102 to 105 (YENV). The 3′ untranslated sequence of HGAL contains several cryptic polyadenylation sites and ATTTA motifs capable of mediating rapid degradation of mRNAs.

Analysis of HGAL expression in normal tissues, NHL cell lines, and tumor specimens

Expression of HGAL was further evaluated by real-time RT-PCR among normal tissues. The highest HGAL mRNA expression was observed in GC cells, thus confirming our previous array observations, followed by thymus and spleen (Figure 4A). All nonhematopoietic and nonlymphopoietic organs, except lung, expressed only trace amounts of HGAL. Examination of HGAL mRNA expression in malignant cell lines revealed variable expression in B-cell NHL cell lines, minimal expression in Jurkat T cells, and no expression in nonlymphoid HL60 and K562 cell lines (Figure 4B). We subsequently analyzed HGAL expression in a spectrum of NHL tumor specimens (Figure 4C) that were distinct from specimens analyzed by DNA microarrays. HGAL expression was high in all FCL and low in all CLL, all T-LL, and most MCL specimens. A single MCL specimen that demonstrated high HGAL expression contained multiple normal GC. HGAL expression in DLBCL was highly variable. In some cases it was not expressed at all, whereas in other cases its expression levels were similar to those of FCL specimens. These observed patterns of expression in normal and malignant tissues confirmed our previous array observations (Figure 1B) demonstrating that HGAL is highly expressed in GC lymphocytes and GC-derived tumors. High expression in thymus and relatively low expression in peripheral blood lymphocytes (PBLs) suggested that some immature T lymphocytes might also express high HGAL mRNA levels. However, the limited number of analyzed immature T cell tumors did not exhibit high HGAL expression.

Analysis of HGAL mutations

Given that in normal lymphocyte subsets and in NHL HGAL exhibited the highest expression in GC lymphocytes and GC-derived tumors, both of which exhibit somatic mutational activity affecting Ig, BCL6, and other genes, we evaluated whether the HGAL gene might also be the target of somatic mutations. In the Ig and BCL6 genes, mutations are mainly observed in 1000 to 2000 bp 3′ to the transcription initiation site. Consequently, we sequenced a DNA region of approximately 1000 bp starting at and downstream from the transcription initiation site of HGAL in 19 NHL specimens and 2 cell lines. No mutations from the germline sequence were found in any of these specimens. Four polymorphic variants, also detected in T cells isolated from the 2 lymphoma specimens, were identified at the following positions, numbered according to

Figure 3. Northern blot analysis of HGAL transcript size. Total RNA from spleen (lane 1), Raji cell line (lane 2) and HF1 cell line (lane 3) was transferred to filters and hybridized with 32P-labeled coding region HGAL cDNA probe. M indicates size marker.

Figure 4. HGAL mRNA expression by real-time RT-PCR. (A) HGAL mRNA expression of normal tissues. (B) HGAL mRNA expression in cell lines. (C) HGAL mRNA expression in FCL (n = 21; ■), DLBCL (n = 54; ●), CLL (n = 16; □), MZL (n = 3; ▲), MCL (n = 4; ○), T-LL (n = 2; △), and sorted and pooled GC cells from 3 normal tonsils (○). HGAL mRNA expression was measured by real-time quantitative PCR. HGAL mRNA expression is a relative expression of HGAL normalized to GAPDH in relation to the HGAL/GAPDH ratio of the Raji cell line (see “Patients, materials, and methods”).
sequence AC024964: in the untranslated portion of exon 1 at position 160869 C/T (allele prevalence 0.90 and 0.10), in intron 1 at positions 160976 T/C (allele prevalence 0.81 and 0.19), 161049 G/A (allele prevalence 0.90 and 0.10), and 161640 T/C (allele prevalence 0.95 and 0.05).

**HGAL expression is induced by IL-4**

To determine how HGAL expression might be involved in physiologic responses, peripheral blood B lymphocytes were stimulated for 6 and 24 hours with anti-μ antibodies, IL-4, or cells expressing human CD40 ligand (t-CD40L). Following stimulation, total RNA was extracted and HGAL mRNA expression was assessed by real-time RT-PCR. As shown in Figure 5, anti-μ antibodies (0.5 μg/mL) or stimulation with CD40L did not affect HGAL mRNA expression, even though these stimuli induced the expression of activation markers (CD69 and CD71), as expected (data not shown). Remarkably, IL-4 stimulated HGAL mRNA expression by 4- to 5-fold, an effect that persisted for 24 hours (Figure 5). No synergistic effect in the induction of HGAL mRNA expression was observed on costimulation with IL-4 and anti-μ antibodies. These results suggest that HGAL is specifically induced in the response to IL-4.

**HGAL mRNA expression is a predictor of DLBCL survival**

Array data (Figure 1A) suggested that high HGAL mRNA expression predicts improved OS of DLBCL. To confirm this observation, we analyzed the correlation between HGAL mRNA expression and DLBCL survival in an independent group of DLBCL patients by real-time RT-PCR. For this analysis we chose a value of 0.73 for HGAL expression based on HGAL mRNA expression in GC lymphocytes from 3 normal tonsil specimens. With a median follow-up of 49 months (range, 6-171 months), the OS rate was significantly higher in the DLBCL patients with high HGAL gene expression than in the DLBCL patients with low HGAL gene expression (median OS, 67 and 33 months, respectively; P = .01) (Figure 6A). The patient with the longest OS of 171 months in the group with high HGAL gene expression died of a cause unrelated to lymphoma (Figure 6A). To further assess the strength of HGAL in predicting survival, we analyzed its effect on survival when HGAL expression was considered as a continuous variable; exact expression values were converted to ranks to avoid giving undue influence to outlying points. Again, higher HGAL expression correlated with longer OS (P = .01). Clinical characteristics at presentation of patients with high (0.73 or higher) and low (lower than 0.73) HGAL gene expression are presented in Table 1. Patients with low HGAL gene expression tended to have higher levels of lactate dehydrogenase (LDH), but no difference in the distribution of other components of IPI was observed. In addition, no differences in complete remission rates were observed in DLBCL groups with high (0.73 or higher) and low (lower than 0.73) HGAL gene expression. Median DFS in DLBCL with low HGAL expression was 20 months, but it was not reached in DLBCL with high HGAL expression. Differences in DLBCL DFS curves approached but did not reach statistical significance (P = .08; data not shown), probably because of the relatively small number of analyzed patients. Patients with DLBCL tumors expressing high HGAL had significantly better failure-free survival than patients with low HGAL expression (P = .01). We next analyzed whether HGAL expression could add to the OS prognostic value of the IPI. Because of the small number of patients in each IPI score subgroup, we combined patients with low or low-intermediate (low clinical risk) and high-intermediate or high (high clinical risk) scores. Considering patients with low clinical risk separately, as judged by the IPI, patients in the high (0.73 or higher) HGAL gene expression group
had a distinctly better OS than patients with low (lower than 0.73) HGAL gene expression ($P = .04$) (Figure 6B). Interestingly, only a few (3 of 13) patients with high clinical risk, as judged by IPI, had high HGAL gene expression, thus preventing meaningful statistical analysis of the possible predictive effect of HGAL gene expression on OS in this group of patients. We next assessed the strength of HGAL in predicting survival in comparison to the IPI. IPI alone did not reach statistical significance in predicting OS in our group of patients, probably because of the small sample size in each IPI category. In multivariate Cox regression analyses that included IPI scores or IPI individual components and HGAL, with OS as the dependent variable, only HGAL mRNA expression was an independent predictor of OS in DLBCL patients (Table 2, $P = .02$). HGAL expression still predicted DLBCL OS as assessed by a loglikelihood ratio test. This analysis confirmed that HGAL gene expression might accurately stratify patients into good and bad prognostic groups, even in a small group of patients in which IPI does not predict the outcome.

We have recently reported that BCL6 expression is a good predictor of OS in DLBCL patients.8 Consequently, we wanted to compare the prognostic value of HGAL, BCL6, or their combination on DLBCL outcome (Figure 7). High expression of each gene predicted better DLBCL survival with a similar statistical power (derivation DLBCL group, $P = .007$ and $P = .008$; validation DLBCL group, $P = .01$ and $P = .01$, for BCL6 and HGAL, respectively). DLBCL patients with high expression of both genes had a better OS than DLBCL patients with low expression of both genes ($P = .0006$; Figure 7). Patients whose tumors exhibited high expression of at least one of these genes also had better OS than patients in whom expression of both genes was low ($P = .015$; Figure 7). There was a trend for better OS in patients whose tumors exhibited high expression of both genes than those whose tumors had high expression of only 1 of these 2 genes, but the trend did not reach statistical significance. Multivariate Cox regression analysis that included the individual components of the IPI and HGAL and BCL6 with OS as the dependent variable demonstrated that BCL6 expression was the best independent OS predictor ($P = .02$), followed by HGAL expression ($P = .08$).

### Discussion

Recent application of cDNA microarray gene expression profiling to DLBCL specimens has significantly advanced our understanding of this heterogeneous disease. Definition of gene expression signatures characteristic of GC and of activated normal lymphocytes resulted in recognition of 2 DLBCL subtypes, based on their cell of origin: GC B-cell–like DLBCL and activated B-cell–like DLBCL.3 These 2 subtypes of DLBCL exhibited different survival outcomes, though the study was not designed to find gene expression patterns that predict survival. Further evidence for meaningful biologic differences between these DLBCL subgroups was provided by analysis of immunoglobulin gene mutations18 and BCL-2 gene translocations.22 Moreover, array analysis identified signaling pathways that are abnormally active in activated B-cell–like DLBCL but not in GC B-cell–like DLBCL. Members of these pathways have been suggested as attractive targets for therapeutic interventions.24 The subdivision of DLBCL into GC B-cell–like and activated B-cell–like types has recently been confirmed in a large and independent collection of DLBCL patients,7 and a third group has also been identified.

In the present work we demonstrate yet another potential of gene expression profiling—identification and cloning of physiologically important, previously unknown, genes among the numerous ESTs present on the arrays. We have cloned and characterized a GC-expressed gene, HGAL. HGAL is a highly evolutionary conserved gene with marked similarity to the mouse GC gene, M17.11 Its high expression in GC lymphocytes, spleen, and GC-derived tumors in humans and mice, compared with other tissues, suggests that it has a GC-specific function. In contrast to mice, in which M17 is not expressed in thymus,11 human HGAL is also highly expressed in thymus and at low levels in PBLs, thus suggesting transient expression of this gene during T-cell ontogeny. Analysis of the putative structure of HGAL suggests that it may function as an adaptor protein with a potential regulatory role. Most notable is the presence of an ITAM motif, which contains 2 YXXL substrates for tyrosine phosphorylation that can mediate interactions with SH2 domain–bearing molecules. Usually ITAM motifs are found in membrane-associated receptors,12 HGAL, as far as we know, is the first human cytoplasmic protein that contains such a motif. This motif is also found in the mouse M17 protein.11 Up-regulation of HGAL expression after IL-4 stimulation suggests that HGAL may mediate some IL-4 effects. In vitro studies have indicated that IL-4 is a growth factor promoting GC B-cell differentiation into memory B cells,25 whereas in vivo studies have demonstrated that the absence of IL-4 enhances the GC reaction in secondary immune responses.26 HGAL expression in memory B cells, though at lower levels than in GC lymphocytes, provides indirect support for the involvement of HGAL in the differentiation from GC to memory B cells.

Our study also shows that HGAL has prognostic significance in DLBCL tumors. Remarkably, in an independent group of DLBCL patients, gene expression profiling recently demonstrated that EST corresponding to HGAL mRNA was among the 16 genes predicting DLBCL outcome.3 The survival curves of our 2 DLBCL patient groups, assessed by cDNA microarrays in the first group and by real-time PCR in the second group, were similar, with 60% and 20% of DLBCL patients with high and low HGAL expression, respectively, surviving for more than 50 months after diagnosis. Interestingly, during the initial 24 months following diagnosis, the

![Figure 7. Kaplan-Meier curves of OS in DLBCL patients based on BCL6 and HGAL expression. HGAL 0.73 or higher and BCL6 greater than 1.3, n = 21 (A); HGAL less than 0.73 and BCL6 no more than 1.3, n = 14 (B). HGAL 0.73 or higher and BCL6 no more than 1.3 or HGAL less than 0.73 and BCL6 greater than 1.3, n = 19 (C).](image)

**Table 2. Prognostic factors in multivariate analysis**

<table>
<thead>
<tr>
<th>Factor</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, 60 y or younger vs older than 60 y</td>
<td>.06</td>
</tr>
<tr>
<td>Ann Arbor stage, I-II vs III-IV</td>
<td>.61</td>
</tr>
<tr>
<td>Performance status, 0-1 vs 2-4</td>
<td>.86</td>
</tr>
<tr>
<td>LDH level, normal vs above normal</td>
<td>.37</td>
</tr>
<tr>
<td>Extramedial sites, 0-1 vs more than 1</td>
<td>.99</td>
</tr>
<tr>
<td>HGAL, 0.73 or higher vs less than 0.73</td>
<td>.02</td>
</tr>
</tbody>
</table>
survival of DLBCL patients with high and low HGAL mRNA expression was similar. Conversely, after this initial period, patients with low HGAL mRNA expression died more often than did patients with high HGAL expression. Whether the better survival in the high HGAL expression group could be attributed to a specific function of the gene or simply to the fact that it identifies tumors that originate from GC lymphocytes is unknown. Our previous observations indicate that not every GC marker has prognostic significance. Improved survival of DLBCL patients with tumors expressing high BCL6 mRNA, but not necessarily of those expressing high levels of CD10, suggests a gene-specific effect or an unrecognized stage of GC lymphocyte ontogeny, from which the tumors originate. Regarding the first possibility of a gene-specific effect on tumor survival, it is notable that HGAL (as demonstrated in this study) and BCL2 are IL-4–inducible genes. IL-4 is a growth and differentiation factor for normal B cells. Distinct effects were reported in malignant B cells. It has been demonstrated that IL-4 induced in vitro inhibition of growth in 60% of lymphoma were reported in malignant B cells. It has been demonstrated that growth and differentiation factor for normal B cells. Distinct effects were reported in malignant B cells. It has been demonstrated that IL-4 induced in vitro inhibition of growth in 60% of lymphoma specimens. Other studies indicate that IL-4 provides growth-inhibitory signals to NHL cells, activated through their surface immunoglobulin receptors. Recently, an inverse association between tumor levels of IL-4 and lymphoma proliferation have been noted. Consequently, it is possible that high mRNA expressions of HGAL and BCL6 are markers of tumors in which IL-4 exhibits antilymphoma effects. This group of DLBCL tumors has better clinical outcomes. It is also possible that HGAL and BCL6 are markers of distinct lymphocyte differentiation stages from which a subset of DLBCL tumors arise. Simultaneous evaluation of the prognostic significance of expression of HGAL and BCL6 genes allowed patients to be subclassified into 3 groups (Figure 7): one group with high expression of both genes, one group with low expression of both genes, and one group highly expressing only one of these genes. These groups appear to differ in OS. Interestingly, all of these groups have similar initial decreases in survival, indicating that other factors are more important in determining early death. After this initial period, there was a plateau in OS of patients whose tumors expressed high levels of both genes in contrast to continuous deaths observed in patients highly expressing only one of these genes. Because the number of patients in each group was small, further studies are needed to establish whether patients with high expression of both genes have a better or similar OS than patients whose tumors exhibit high expression of only one of these genes. Interestingly, in contrast to heterogeneous patterns of DLBCL, all FL tumors express high levels of HGAL and BCL6, thus suggesting their uniform origin from a single B-cell counterpart. With the advance of microarray technology, better understanding of DLBCL pathophysiology, and identification of expression signatures that predict tumor outcome, we will be able to improve the management of, and perhaps find new targets for the therapy for, lymphoma. Simple assays, such as RT-PCR or immunoperoxidase staining, will eventually be developed based on the most useful predictive markers. BCL6 and HGAL must now be added to BCL2,31-34 survivin,35 and short lists of other genes identified by microarray studies6,7 as candidates for routine clinical application. In summary, we have cloned and characterized a new gene, HGAL, which is highly expressed in GC lymphocytes and GC-derived tumors. Moreover, we demonstrate that HGAL mRNA expression is an important predictor of OS in patients with DLBCL. HGAL expression defined DLBCL subgroups with distinct outcomes even in patients at low risk, defined by the currently used clinical index (IPI). Further studies will elucidate the biologic function of this gene and will confirm its prognostic value alone or in combination with other genes.

Acknowledgment

R.L. is an American Cancer Society Clinical Research Professor.

References

4. The exression de...


