Prognostic significance of immunohistochemical biomarkers in diffuse large B-cell lymphoma: a study from the Lunenburg Lymphoma Biomarker Consortium

Gilles Salles,1 Daphne de Jong,2 Wanling Xie,3 Andreas Rosenwald,4 Mukesh Chhanabhai,5 Philippe Gaulard,6 Wolfram Klappler,7 Maria Calaminici,8 Birgitta Sander,9 Christoph Thorns,10 Elias Campo,11 Thierry Molina,12 Abigail Lee,8 Michael Pfundshoch,13 Sandra Horning,14 Andrew Lister,8 Laurie H. Sehn,5 John Raemaekers,15 Anton Hagenbeck,16 Randy D. Gascoyne,5 and Edie Weller3

1Hospices Civils de Lyon and Université Claude Bernard Lyon-1, Lyon, France; 2Netherlands Cancer Institute, Amsterdam, The Netherlands; 3Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA; 4Institute of Pathology, University of Würzburg, Würzburg, Germany; 5Department of Pathology and Medical Oncology, BC Cancer Agency, University of British Columbia, Vancouver, BC; 6Department of Pathology, Hôpital Henri Mondor, Créteil, France; 7Department of Pathology, Hematopathology Section, University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany; 8Cancer Research-UK Medical Oncology Unit, St Bartholomew’s Hospital, London, United Kingdom; 9Karolinska Institutet, Stockholm, Sweden; 10Department of Pathology, Universitätsklinikum Schleswig-Holstein, Campus Luebeck, Luebeck, Germany; 11Hospital Clinic, University of Barcelona, Barcelona, Spain; 12Université Paris-Descartes; Assistance Publique des Hôpitaux de Paris, Hôtel-Dieu, Paris, France; 13Department of Internal Medicine I, University of Saarland, Homburg, Germany; 14Department of Medicine, Stanford University Medical Center, Palo Alto, CA; 15Department of Hematology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands; and 16Department of Haematology, Academic Medical Center, Amsterdam, The Netherlands

The Lunenburg Lymphoma Biomarker Consortium (LLBC) evaluated the prognostic value of IHC biomarkers in a large series of patients with diffuse large B-cell lymphoma (DLBCL). Clinical data and tumor samples were retrieved from 12 studies from Europe and North America, with patients treated before or after the rituximab era. Using tissue microarrays from 1514 patients, IHC for BCL2, BCL6, CD5, CD10, MUM1, Ki67, and HLA-DR was performed and scored according to previously validated protocols. Optimal cut points predicting overall survival of patients treated in the rituximab era could only be determined for CD5 (P = .003) and Ki67 (P = .02), whereas such cut points for BCL2, BCL6, HLA-DR, and MUM1 could only be defined in patients not receiving rituximab. A prognostic model for patients treated in the rituximab era identified 4 risk groups using BCL2, Ki67, and International Prognostic Index (IPI) with improved discrimination of low-risk patients. Newly recognized correlations between specific biomarkers and IPI highlight the importance of carefully controlling for clinical and biologic factors in prognostic models. These data demonstrate that the IPI remains the best available index in patients with DLBCL treated with rituximab and chemotherapy. (Blood. 2011;117(26):7070-7078)

Introduction

The prediction of outcome in patients with lymphoma is an important challenge facing clinicians. After the description of various clinical prognostic factors, decisive progress was made > 15 years ago with the establishment of the International Prognostic Index (IPI) for diffuse large B-cell lymphoma (DLBCL), the result of a large, international cooperative effort.1 However, it was clearly assumed that the variables used to build this index were surrogates of DLBCL biology. This heterogeneity was already appreciated through various morphologic, immunologic, and cytogenetic features, resulting in numerous studies attempting to assess the prognostic value of single biomarkers.2 More recently, gene-expression profiling data revealed a greater level of molecular complexity in DLBCL, with the identification of specific gene-expression signatures such as germinal-center B-cell (GCB) or activated B-cell (ABC) signatures, which are associated with distinct genetic alterations and significantly different survival rates.3,5 Given the difficulties in applying these molecular approaches in daily practice, an attempt was made to transfer this molecular classification into an IHC score.6 However, the prognostic value of this algorithm has provided very inconsistent results.7,13 These discrepant data may be related to patient selection criteria and the limited sample sizes of some series, but also to variability in technical aspects such as antibodies, different scoring methods, criteria, and cut points. A validation study assessing the reproducibility of IHC marker scoring conducted by our consortium established that the staining and scoring methodologies should be normalized to obtain a reasonable level of confidence in interpreting these markers.14,15 Furthermore, it has been shown that the predictive value of markers may be dependent on treatment, especially with the use of anti-CD20 antibodies.16,17 Given these challenges, the Lunenburg Lymphoma Biomarker Consortium (LLBC) decided to launch a study based on a large series of patients with DLBCL uniformly treated with rituximab plus cyclophosphamide, hydroxydaunorubicin, oncovin, and prednisone/prednisolone (R-CHOP) within clinical studies. In this effort, controlled marker assessment as developed in the previous “validation study”14 was applied to determining the prognostic value of individual IHC markers or to explore a possible combined clinico-biologic index. Two additional series of CHOP-treated patients were also studied to provide further

From www.bloodjournal.org by guest on June 9, 2017. For personal use only.
Methods

Patients

Prospective clinical studies organized by European and American collaborative groups or population-based reference centers were selected for this study based on (1) inclusion of patients with DLBCL treated with standard anthracycline-based therapy, with at least 6 cycles of CHOP or CHOP-like chemotherapy with or without rituximab; (2) availability of complete clinical data at study registration and an updated follow-up of at least 3-years; and (3) the possibility of retrieving paraffin blocks from the diagnostic biopsy samples. Six of these studies—GELA 98-5 and 05-1 (NCT00135499), ECOG4494,22 BCMA,23 MINT,21 and HOVON-46 (NCT00287171)—performed between February 1998 and August 2005, included patients treated with R-CHOP in 2-armed studies with a control group treated with CHOP only or patients from registry data selected within 18 months of when rituximab was introduced (BCCA only). These cohorts are designated as rituximab-CHOP (r-CHOP) and control-CHOP (c-CHOP), respectively. The other studies (EORTC-20901, HOVON-25/26,26 GLHSG-B1/2/23,24 and UK-Bart’s & Manchester) included patients treated before the rituximab era, between August 1989 and February 2004; these patients are designated as earlier CHOP (e-CHOP) patients.

This study was approved by the ethical committees of all collaborating trial organizations and centers to comply with the Declaration of Helsinki. Data obtained from each corresponding group were electronically transferred to the LLBC statisticians and checked for data completeness and consistency with previous original publications of the clinical studies.

Sample collection, processing, IHC, and marker assessment

Tissue microarrays (TMAs) were constructed by each pathology reference center for all available samples using 0.6-mm core samples in triplicate, as described previously.14 Based on the results of the validation study, a policy of “one marker/one laboratory” was followed, implying that sections of all available TMAs for the complete series were stained for individual markers in one selected laboratory that demonstrated the most reliable results in the validation study. To avoid any potential bias,_string sections were distributed between the laboratories, and each sample was counted with 5% of the scores used for this study. The median value was taken. If 2 pathologists agreed and 1 could not score a sample, these data on the patients from that TMA group were used to impute a value for this patient. With the imputed value, the final score for this patient was determined using the same rule. When 2 or 3 pathologists were unable to score a sample, this sample was considered as not scored. The adjusted algorithm to distinguish germinal center (GCB) versus nongerminal center (non-GCB) samples evaluated in this paper defines BCL6-positive as any staining and uses 25% as the cut point for MUM1. Although the original cut points used for this index were chosen arbitrarily and differ from the present one,25 based on our former validation study, agreement was significantly higher when the BCL6 (and CD10) descriptive categories of no staining, weak/variable weak staining, and strong/variable strong staining were used instead of percentages of patients scored. Therefore, this BCL6 scoring was implemented in this study.

Statistical methods

Univariate and multivariate analyses were used to assess the prognostic significance of biomarkers and the IPI for overall survival (OS), defined as time from treatment initiation to death or date of last follow-up. Correlations between biomarker and IPI were evaluated using the χ² test adjusting for multiple comparisons using the Bonferroni method.

Because of the high percentage of cases with incomplete scores for all 8 markers (48%), classification and regression trees for survival outcomes (using the “part” function in R with 10-fold cross-validation) were implemented to determine prognostic models for the r-CHOP cohort using the IPI groups as well as the biomarkers. In this analysis each biomarker was included as originally scored. Models based on an optimal cut point identified from a univariate analysis were also evaluated. In the univariate analysis, the optimal cut point for each IHC biomarker was selected as the one with the maximum log-rank statistic to predict patient OS. Regression tree results are similar, with biomarkers included using the categories and the cut points.

The impact of the biomarkers and IPI on OS was also evaluated among the c-CHOP series, which included a broader cross-section of age and IPI groups. Given the large number of patients in the c-CHOP series, a test validation (2/3 to 1/3) design stratified by study group was implemented. Prognostic models were evaluated and applied to the e-CHOP and r-CHOP series of data.

The Cox proportional hazards (PH) regression model was used to estimate the relative risk for each of the prognostic groups. A global measure of fit, the Bayesian information criterion (BIC) was assessed, as well as a measure of discrimination, the area under the receiver operating characteristic curve for survival outcomes (the c-index).25 Low values of BIC indicate better fit and high values of c-index indicate better discrimination. A multiple imputation procedure (n = 5 datasets) was also used to evaluate the effect of missing scores. Similar results were obtained applying recursive partitioning and Cox PH regression to the multiple imputation datasets (results not shown). Analyses were performed using SAS Version 9 software and the R package Version 2.8.22.

Results

Patient population and biomarker assessment

Clinical data were collected for 2451 patients with DLBCL from 12 studies (r-CHOP, 674 patients; c-CHOP, 620 patients; and e-CHOP, 1157 patients). Material for TMA was available for 1514 of these patients with 347 of 674 (52%) from r-CHOP, 289 of 620 (47%) from c-CHOP, and 878 of 1157 (76%) from e-CHOP. Individual IPI factors and IPI index were assessed for each cohort (not shown) and among those with available TMAs (Table 1). In these subsets, complete data for IPI were available for all c-CHOP and r-CHOP patients and for 845 of 878 (96%) e-CHOP patients. As a result of the inclusion criteria of the recent randomized R-CHOP versus CHOP clinical trials, a significantly higher percentage of patients were older than 60 years of age in the r-CHOP (n = 231, 67%) and c-CHOP (n = 211, 73%) cohorts compared with the e-CHOP cohort (n = 411, 47%; P < .001). Compared with those without TMA submission, patients with available TMA material were significantly younger (P < .001, 56% vs 69% > 60 years of age), had less-advanced disease (59% vs 72% stage
Table 1. Distribution of IPI and survival estimates in the 3 groups among the 1514 of 2451 patients for which blocks were available for construction of a TMA

<table>
<thead>
<tr>
<th>Analysis group</th>
<th>All patients</th>
<th>Age &gt; 60 y</th>
<th>No. of patients (%)</th>
<th>4-year OS (95% CI)</th>
<th>IPI category low</th>
<th>IPI category low/intermediate</th>
<th>IPI category high/intermediate</th>
<th>IPI category high</th>
</tr>
</thead>
<tbody>
<tr>
<td>r-CHOP (n = 347)</td>
<td>69% (64.74%)</td>
<td>231 (67%)</td>
<td>94 (27%)</td>
<td>85% (75, 91%)</td>
<td>80 (23%)</td>
<td>80% (69, 87%)</td>
<td>87 (25%)</td>
<td>68% (57, 77%)</td>
</tr>
<tr>
<td>c-CHOP (n = 289)</td>
<td>55% (49, 61%)</td>
<td>211 (73%)</td>
<td>79 (27%)</td>
<td>64% (73, 91%)</td>
<td>63 (22%)</td>
<td>63% (50, 74%)</td>
<td>79 (27%)</td>
<td>48% (37, 58%)</td>
</tr>
<tr>
<td>e-CHOP (n = 878)</td>
<td>59% (55, 62%)</td>
<td>411 (47%)</td>
<td>428 (50%)</td>
<td>76% (72, 80%)</td>
<td>208 (25%)</td>
<td>49% (42, 56%)</td>
<td>142 (17%)</td>
<td>40% (31, 48%)</td>
</tr>
</tbody>
</table>

Median follow-up was 4.4, 5.2, and 5.8 years among the r-CHOP, c-CHOP, and e-CHOP patients, respectively.

*Thirty-three e-CHOP patients (4%) did not have IPI data.

III/IV), had fewer extranodal sites of involvement (20% vs 41% with >1 extranodal site), and were less likely to have elevated lactate dehydrogenase levels (47% vs 59% above normal value), resulting in a higher percentage of low-risk patients (41% vs 20%, P < .001). As expected, OS was significantly longer (P < .001) in the r-CHOP patients compared with the c-CHOP patients. In addition, highly significant differences in OS were detected by IPI risk groups (P < .0001 for all 3 cohorts; Table 1 and Figure 1).

Of the 1514 patients, 793 (52%) had all 8 immunohistochemical markers scored, 320 (21%) patients had 7 of the 8 markers scored, and 137 (9%) had 6 of 8 markers scored. Markers could not be scored in individual cases for various reasons; for example, not representative or missing cores in the immunostained TMA section, inadequate staining results, or absence of an internal control staining precluding scoring as absence of staining in tumor cells. The proportion of patients whose TMA could be scored by each marker was as follows: 1306 (86%) for BCL2, 985 (65%) for BCL6, 1344 (89%) for CD10, 1366 (90%) for CD5, 1377 (91%) for HLA-DR, 1138 (75%) for Ki67, and 1249 (82%) for MUM1 (supplemental Table 1, available on the Blood Web site; see the Supplemental Materials link at the top of the online article). A detailed distribution of biomarkers scores in the r-CHOP cohort is provided in Figure 2.

Assessing the optimal cut point that predicted patient outcome for each biomarker

In a univariate analysis, we selected optimal cut points for OS based on the maximum log-rank statistic (Table 2). Among the r-CHOP patients, only CD5 (75% vs > 75%) and Ki67 (75% vs > 75%) could discriminate patient outcome (P < .05). A marginal difference (P = .09) was detected for BCL2 (75% vs > 75%) but not for BCL6 (Figure 3). Among the e-CHOP patients, which represent a broader range of age groups, the optimal cut points could be determined for BCL2 (75% vs > 75%), BCL6 (no staining vs any staining), CD5 (no staining vs any staining), HLA-DR (positive vs negative), and MUM1 (75% vs > 75%; Table 2). Several markers identified in the e-CHOP cohort retained their significance in the c-CHOP cohort (BCL2, BCL6, and marginally for MUM1). The optimal Ki67 cut point of 75% vs > 75% determined in the r-CHOP cohort was not found to be predictive in univariate analysis for the e-CHOP cohort (or for the c-CHOP cohort). No difference in OS was found for CD10 (positive vs negative; Table 2). Similar cut points were found using data obtained with imputation for missing scores (not shown).

The impact of combined markers that might define GCB versus non-GCB patients on OS was evaluated using an adjusted GCB/non-GCB algorithm (where BCL6 was dichotomized as no-staining/staining) and an optimized LLBC algorithm that uses optimal cut points determined from the univariate analysis (GCB/non-GCB LLBC with MUM1 dichotomized as ≤ 75% vs > 75% and BCL6 dichotomized as no-staining/staining). The adjusted GCB/non-GCB algorithm was not of prognostic value for c-CHOP or r-CHOP patients (hazard ratio [HR] = 1.3; 95% confidence interval [CI] 0.8-1.9 and HR = 1.1; CI 0.6-1.8, respectively, P > .25), but was of prognostic for the e-CHOP
cohort (HR = 1.4; CI 1.1-1.7, P = .01; Figure 4). The GCB/non-GCB LLBC also lacked prognostic significance in the r-CHOP cohort (HR = 1.4; CI 0.8-2.6, P = .23 for r-CHOP), whereas it appeared to have a higher relative risk in the 2 CHOP patient cohorts (HR = 1.7; CI 1.3-2.2, P < .0001 for e-CHOP; HR = 1.8; CI 1.1-2.8, P = .01 for c-CHOP).

Pairwise correlation for prognostic biomarkers and the IPI categories

The large number of cases with available IHC data allowed assessment of the associations between the expression of each dichotomized biomarker and IPI category distribution using all 3 patient cohorts. As shown in Table 3, the percentage of patients expressing certain biomarkers differed between IPI categories. Patients with tumor cells that overexpressed BCL2 (> 75%) presented significantly more frequently with a higher IPI (P < .0001), whereas those that were either CD10 or HLA-DR positive had a significantly lower IPI at diagnosis (P = .008 and P = .04, respectively).

Interestingly, pairs of markers also showed strong correlations. As expected, CD10 and BCL6 were correlated and CD10 and MUM1 inversely correlated. High BCL2 expression was also more frequently observed in patients who highly expressed MUM1. Expression of CD5 was also correlated with the high expression of BCL2 as well as lack of CD10. Finally, cases with high expression of MUM1 were found to be HLA-DR negative and had high expression of Ki67. (P < .002 for all; Table 4). These findings highlight the coordinated expression of biomarkers associated with the same impact on patient outcome and possibly a common biologic background.

Developing a prognostic model using biomarkers

We developed a model using the biomarkers as originally scored on all r-CHOP patients with at least one marker scored (n = 342 of 347). The low-risk patients were separated by BCL2 and the low-intermediate/ high-intermediate patients were separated by Ki67 (Figure 5A). This model distinguished 4 groups (groups 1–4) of approximately equal size with improved model evaluation measures relative to the IPI (BIC = 1145 vs 1153; c-index = 0.69 vs 0.67; Table 5 and Figure 5B). The most favorable group comprised patients with a low IPI and BCL2 ≤ 75% and had an expected 4-year OS of 94%. Group 2 (81%

Table 2. Optimal cut points of IHC markers used to predict OS†

<table>
<thead>
<tr>
<th>Marker</th>
<th>Optimal cut point selected†</th>
<th>r-CHOP HR (95% CI)‡</th>
<th>Log-rank P</th>
<th>e-CHOP HR (95% CI)‡</th>
<th>Log-rank P</th>
<th>c-CHOP HR (95% CI)‡</th>
<th>Log-rank P</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCL2</td>
<td>≤ 75% vs &gt; 75%</td>
<td>1.4 (0.9, 2.2)</td>
<td>.09</td>
<td>1.7 (1.4, 2.1)</td>
<td>&lt; .0001</td>
<td>1.6 (1.1, 2.3)</td>
<td>.02</td>
</tr>
<tr>
<td>BCL6</td>
<td>No staining vs staining</td>
<td>1.1 (0.5, 2.7)</td>
<td>.76</td>
<td>1.7 (1.3, 2.3)</td>
<td>&lt; .0003</td>
<td>2.9 (1.7, 5.0)</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>CD5</td>
<td>≤ 75% vs &gt; 75%</td>
<td>2.4 (1.3, 4.3)</td>
<td>.003</td>
<td>1.6 (1.2, 2.2)</td>
<td>.004</td>
<td>1.0 (0.6, 1.9)</td>
<td>.9</td>
</tr>
<tr>
<td>CD10</td>
<td>Negative vs positive</td>
<td>1.1 (0.8, 1.7)</td>
<td>.56</td>
<td>1.2 (0.9, 1.5)</td>
<td>.17</td>
<td>1.4 (1.0, 2.0)</td>
<td>.06</td>
</tr>
<tr>
<td>HLA-DR</td>
<td>Negative vs positive</td>
<td>1.3 (0.7, 2.4)</td>
<td>.36</td>
<td>1.6 (1.2, 2.1)</td>
<td>.001</td>
<td>1.3 (0.8, 1.9)</td>
<td>.33</td>
</tr>
<tr>
<td>Ki67</td>
<td>≤ 75% vs &gt; 75%</td>
<td>1.8 (1.1, 2.9)</td>
<td>.02</td>
<td>0.8 (0.7, 1.0)</td>
<td>.11</td>
<td>1.0 (0.6, 1.5)</td>
<td>.86</td>
</tr>
<tr>
<td>MUM1</td>
<td>≤ 75% vs &gt; 75%</td>
<td>1.4 (0.8, 2.3)</td>
<td>.22</td>
<td>1.3 (1.0, 1.7)</td>
<td>.03</td>
<td>1.5 (1.0, 2.4)</td>
<td>.06</td>
</tr>
</tbody>
</table>

†The number of the 1514 patients scored for BCL2, BCL6, CD5, CD10, HLA-DR, Ki67, MUM1, are 1306, 985, 1366, 1344, 1377, 1138, 1249, respectively
‡When optimal cut points were not determined using the r-CHOP series, the e-CHOP cut points are presented. Results obtained from applying the cut points to all cohorts are shown.
§Reference groups: ≤ 75% staining for BCL2, CD5, MUM1, and Ki67; any staining for BCL6; positive for CD10 and HLA-DR.
4-year OS) included low-IPI patients with BCL2 > 75% and low-intermediate/ high-intermediate IPI patients with Ki67 positivity in ≤ 75% of cells. Group 3 (62% 4-year OS) comprised low-intermediate/ high-intermediate patients with Ki67 > 75%, and group 4 comprised patients with a high IPI (4-year OS 45%; Figure 5C). When this model was applied to c-CHOP patients, the prognosis for group 2 and 3 appeared similar (Table 5).

The impact of the biomarkers and IPI on OS was also evaluated among the e-CHOP series, which included a broader cross-section of ages and IPI groups. This analysis also resulted in the discrimination of low-IPI patients according to BCL2 and HLA-DR expression, with 4 separate risk groups identified (not shown). When evaluated in the r-CHOP cohort, this model, which was optimized for the e-CHOP series, did not result in substantial improvement of risk group definitions. Similar results were obtained by applying

---

**Figure 3.** OS by BCL2, BCL6, CD5, and Ki67 categories for the 347 r-CHOP patients. Categories are shown as originally scored. The number of r-CHOP patients scored for BCL2, BCL6, CD5, and Ki67 are 526, 212, 333, and 239, respectively.

**Figure 4.** OS according to algorithms obtained by assessments of CD10, BCL6, and MUM1 scoring results. Algorithms were used to define an adjusted GCB/non-GCB index (with the MUM1 cut point defined as 25%) or the LLBC GCB/non-GCB index (with the MUM1 cut point defined as 75%) in the r-CHOP (A and D), c-CHOP (B and E), and e-CHOP (C and F), respectively. Among the 1046 patients scored using the adjusted GCB algorithm and with complete IPI data, 558 were CD10⁺ (GCB), 151 were CD10⁺BCL6⁺MUM1⁻ (non-GCB), 242 were CD10⁺BCL6⁺MUM1⁺ (non-GCB), and 95 were CD10⁺BCL6⁺MUM1⁺ (non-GCB). Among the 1046 patients scored using the LLBC GCB algorithm and with complete IPI data, 558 were CD10⁺ (GCB), 270 were CD10⁺BCL6⁺MUM1⁻ (GCB), 123 were CD10⁺BCL6⁺MUM1⁺ (non-GCB), and 95 were CD10⁺BCL6⁺ (non-GCB).
Table 3. Biomarker expression and IPI category distribution in all patients with DLBCL

<table>
<thead>
<tr>
<th>Biomarker expression</th>
<th>No. of patients with biomarker scored and IPI</th>
<th>No. of patients with biomarker expression (%)</th>
<th>Biomarker expression percentage by IPI categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>BCL2 &gt; 75%</td>
<td>1279</td>
<td>630 (49%)</td>
<td>39</td>
</tr>
<tr>
<td>BCL6 staining</td>
<td>965</td>
<td>842 (87%)</td>
<td>91</td>
</tr>
<tr>
<td>CD10 staining</td>
<td>1316</td>
<td>558 (42%)</td>
<td>49</td>
</tr>
<tr>
<td>CD5 &gt; 75%</td>
<td>1339</td>
<td>115 (9%)</td>
<td>7</td>
</tr>
<tr>
<td>HLA-DR staining</td>
<td>1348</td>
<td>1180 (88%)</td>
<td>91</td>
</tr>
<tr>
<td>Ki67 &gt; 75%</td>
<td>1119</td>
<td>627 (56%)</td>
<td>61</td>
</tr>
<tr>
<td>MUM1 &gt; 75%</td>
<td>1225</td>
<td>264 (22%)</td>
<td>20</td>
</tr>
<tr>
<td>GCB/non-GCB§ (adjusted)</td>
<td>1046</td>
<td>709 (68%)</td>
<td>72</td>
</tr>
<tr>
<td>GCB/non-GCB§ (LLBC)</td>
<td>1046</td>
<td>828 (79%)</td>
<td>83</td>
</tr>
</tbody>
</table>

*The number of the 1514 patients scored for BCL2, BCL6, CD5, CD10, HLA-DR, Ki67, MUM1, GCB (adjusted), and GCB (LLBC) are 1306, 985, 1364, 1344, 1377, 1138, 1249, 1067, and 1067, respectively.
†The number of additional patients without IPI for BCL2, BCL6, CD5, CD10, HLA-DR, Ki67, MUM1, GCB (adjusted), and GCB (LLBC) are 27, 20, 18, 27, 29, 19, 24, 21, and 21, respectively.
$\chi^2$ P values are adjusted for 9 multiple comparisons
‡Among the 1046 patients scored using the adjusted GCB algorithm and with complete IPI data, 558 were CD10 (GCB). 151 were CD10 /BCL6/MUM1 (GCB), 242 were CD10 /BCL6/MUM1 non-GCB, and 95 were CD10 /BCL6 (non-GCB). Among the 1046 patients scored using the LLBC GCB algorithm and with complete IPI data, 558 were CD10 (GCB), 270 were CD10 /BCL6/MUM1 (GCB), 123 were CD10 /BCL6/MUM1 (non-GCB), and 95 were CD10 /BCL6 (non-GCB).
NS indicates nonsignificant.

Discussion

In the present study, the LLBC explored the validity of IHC markers as prognostic tools in patients with DLBCL. The availability of 3 large patient cohorts based on clinical studies with optimally documented and homogeneous treatments conducted in various countries resulted in a unique dataset not directly comparable to single institutions or single cooperative groups. This allowed for the evaluation of the impact of covariates and interrelations of biologic and clinical parameters. Although biases remain and low-risk R-CHOP–treated patients are relatively underrepresented, this series constitutes a very reasonable cross-section of all risk and age groups. The approach for staining and scoring IHC markers was standardized and based on previous reproducibility assessments. Although several markers could not be scored in all patients, this evaluation was performed on material collected from multiple prospective clinical trials and involved a panel of hematopathologists, and therefore is representative of clinical investigations. Individual cases stained and scored on whole sections at the time of diagnosis by individual pathologists may more closely reflect clinical practice, but are not practical for the evaluation of large series. Using a step-by-step statistical approach supported by the large number of available cases, several key findings were identified.

First, the optimal cut points for single markers that provided the best discrimination for patient outcome were identified on a rational basis and appeared to be different from those arbitrarily used in previous studies. Second, new correlations between individual biomarkers and IPI, as well as between biomarkers themselves, were identified. These were probably underestimated in previous studies and could likely account for some discordant results. The positive correlations between CD10 and BCL6 expression and the negative correlation between CD10 and MUM1 expression was expected, because they represent features consistent with the GCB subtype definition. Further, BCL2 overexpression has been reported in ABC-type DLBCL cases that overexpress MUM1. As described previously, a correlation among CD5, BCL2, and an inverse with CD10 was noted, which may represent a biologically separate, small subclass of CD5+ DLBCL. These interrelations between biomarkers themselves and IPI parameters highlight the importance of carefully controlling for clinical and biologic characteristics when multiple markers are tested in prognostic models.

Table 4. Pairwise frequency of biomarkers

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Score No staining</th>
<th>Staining</th>
<th>No staining</th>
<th>Staining</th>
<th>≤75%</th>
<th>&gt;75%</th>
<th>≤75%</th>
<th>&gt;75%</th>
<th>≤75%</th>
<th>&gt;75%</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCL2</td>
<td>0%-75%</td>
<td>56 (12%)</td>
<td>433 (88%)</td>
<td>378 (59%)</td>
<td>267 (41%)</td>
<td>615 (95%)*</td>
<td>33 (5%)*</td>
<td>75 (12%)</td>
<td>56 (88%)</td>
<td>233 (42%)</td>
</tr>
<tr>
<td></td>
<td>&gt; 75%</td>
<td>58 (13%)</td>
<td>381 (87%)</td>
<td>351 (57%)</td>
<td>265 (43%)</td>
<td>554 (87%)*</td>
<td>81 (13%)*</td>
<td>86 (13%)</td>
<td>53 (87%)</td>
<td>246 (48%)</td>
</tr>
<tr>
<td>BCL6</td>
<td>No staining</td>
<td>96 (82%)*</td>
<td>21 (18%)*</td>
<td>104 (87%)</td>
<td>15 (13%)</td>
<td>14 (11%)</td>
<td>109 (89%)</td>
<td>49 (44%)</td>
<td>62 (56%)</td>
<td>78 (68%)</td>
</tr>
<tr>
<td></td>
<td>Staining</td>
<td>411 (49%)*</td>
<td>422 (51%)*</td>
<td>789 (93%)</td>
<td>56 (7%)</td>
<td>93 (11%)</td>
<td>750 (89%)</td>
<td>331 (42%)</td>
<td>458 (58%)</td>
<td>642 (79%)</td>
</tr>
<tr>
<td>CD10</td>
<td>No staining</td>
<td>669 (89%)</td>
<td>84 (11%)</td>
<td>109 (14%)</td>
<td>643 (86%)</td>
<td>298 (47%)</td>
<td>332 (53%)</td>
<td>503 (72%)</td>
<td>200 (28%)</td>
<td>440 (89%)</td>
</tr>
<tr>
<td></td>
<td>Staining</td>
<td>529 (95%)</td>
<td>29 (5%)</td>
<td>154 (13%)</td>
<td>1057 (87%)</td>
<td>448 (44%)</td>
<td>572 (56%)</td>
<td>880 (79%)</td>
<td>235 (21%)</td>
<td>77 (71%)</td>
</tr>
<tr>
<td>CD5</td>
<td>0%-75%</td>
<td>52 (37%)</td>
<td>88 (63%)</td>
<td>100 (66%)</td>
<td>52 (34%)</td>
<td>40 (43%)</td>
<td>54 (57%)</td>
<td>77 (71%)</td>
<td>31 (29%)</td>
<td>408 (87%)</td>
</tr>
<tr>
<td></td>
<td>&gt; 75%</td>
<td>40 (35%)</td>
<td>50 (65%)</td>
<td>51 (33%)</td>
<td>83 (67%)</td>
<td>442 (45%)</td>
<td>534 (55%)</td>
<td>858 (80%)</td>
<td>209 (20%)</td>
<td>424 (69%)</td>
</tr>
</tbody>
</table>

*Biomarkers that are significantly (P < .002) correlated. Adjusting for multiple comparisons using the Bonferroni method, $\chi^2$ P values < .05/21 are considered significant.
Among 347 patients treated with R-CHOP, this study demonstrates that only CD5 retained a prognostic significance. This is in contrast to the e-CHOP and c-CHOP cohorts in which BCL2, BCL6, and MUM1 and IHC algorithms related to gene expression in CHOP-treated patients were prognostic for OS. This difference clearly confirms that response to rituximab may differ in distinct biologic groups. It also emphasizes that the impact of prognostic factors may be strongly influenced by the nature of the therapy administered, and that immunotherapy may differ in that way from chemotherapy. The present r-CHOP (and c-CHOP) cohorts are, however, biased toward patients older than 60 years of age with adverse IPI scores, and such findings may need confirmation in larger r-CHOP cohorts. It should be noted that the GCB/non-GCB algorithms used in this study were not specifically designed as optimized surrogates for the biologic gold standard of GCB/ABC subtypes based on gene-expression analysis, but rather represent potential prognostic tools validated using IHC. However, the present data indicate that the GCB/non-GCB IHC-based algorithm based on CD10, BCL6, and MUM1 expression does not provide prognostic information in the rituximab era, in agreement with another large patient series reported recently.32 Overall, these results emphasize that prognostic indicators developed before the use of rituximab need to be thoroughly reassessed before any implementation in the current era of therapy may be considered.

Figure 5. Prognostic model using biomarkers and IPI. (A) Hierarchical tree model. Numbers indicate the number of deaths observed at each level in the population at risk. (B) Number of patients, 4-year OS, and HR for the risk of death (without and with imputation for samples with missing scores) in the r-CHOP cohort. (C) OS for r-CHOP patients according to the IPI and the biomarker and IPI model, respectively. Log-rank P values for both models were < .0001.
Most importantly, this study demonstrates that variations in biologic features are largely overshadowed by the IPI for prognostic impact, and that biomarkers only allow subtle refinements of the IPI. This is demonstrated by the modest HR for death (ranging from 1.3 up to 1.8) conferred for individual biomarkers in the r-CHOP cohort (as well as in larger e-CHOP cohorts). Attempts to build an index using biomarkers alone, the GCB LLBC classification, or the combination of biomarkers with IPI individual factors (instead of IPI categories) did not result in an improved model fit or discrimination (data not shown). Several biomarkers appear to be able to further discriminate subgroups with distinct outcome within IPI categories. BCL2 was found to discriminate the outcome of low- or intermediate-risk patients treated with (and without) rituximab. This may be clinically relevant, because BCL2 overexpression is well known to confer chemotherapy resistance, and therapeutic targeting of this protein is currently under development. The adverse outcome associated with CD5 expression, also reported by others, may reflect the distinct origin of DLBCL for which targeted therapies are investigated.

**Acknowledgments**

The Lunenburg Lymphoma Biomarker Consortium is a collaboration of 9 international lymphoma collaborative groups, each represented by clinical investigators and hematopathologists, supported by a team of data managers and statisticians, and with the active participation of other experts. Those include in particular: R. Gascoyne, M. Chhanabhai, L. Sehn (British Columbia Cancer Center); R. Gascoyne, S. Horning (ECOG); D. De Jong, M. van Glabbeke, H. Kluin-Nelemans, J. Raemaekers (EORTC Lymphoma Group); P. Gaulard, T. Molina, J. Briere, M. Fournier, B. Coiffier, H. Tilly, G. Salles (GELA); C. Thorns, A. Rosenwald, W. Klapper, G. Ott, S.H.-W. Bernd, L. Trumper, N. Schmitz, M. Pfreundschuh (German High Grade Non-Hodgkin Lymphoma Group–DSHNHL); P. Sonneveld, J. Doorduijn, P. Huijgens, L.F. Verdonck, G. van Imhoff, M. Steijger-M. de Meulendijks, M. Testroote, W. van Putten, W. van der Holt, W. Graveland; A. Mulder, D. de Jong, K. Lam, J. van den Tweel (HOVON); B. Sander, E. Kimby (Nordic Lymphoma Study Group); J. Radford (Manchester, United Kingdom); M. Calaminici, A. Lee, A. Norton, A. Clear, A. Lister (St Bartholomew’s Hospital); E. Campo, Independent pathology advisor (Barcelona, Spain); E. Weller, W. Xie, Y. He, B. Giblin (Dana-Farber Cancer Institute); M.-J. Kersten (Amsterdam, The Netherlands).

This study was supported by unrestricted grants from Bayer-Schering-Pharma, Eli-Lilly, Genentech, Millennium, and Roche.

**Authorship**

**References**


Prognostic significance of immunohistochemical biomarkers in diffuse large B-cell lymphoma: a study from the Lunenburg Lymphoma Biomarker Consortium

Gilles Salles, Daphne de Jong, Wanling Xie, Andreas Rosenwald, Mukesh Chhanabhai, Philippe Gaulard, Wolfram Klapper, Maria Calaminici, Birgitta Sander, Christoph Thorns, Elias Campo, Thierry Molina, Abigail Lee, Michael Pfreundschuh, Sandra Horning, Andrew Lister, Laurie H. Sehn, John Raemaekers, Anton Hagenbeek, Randy D. Gascoyne and Edie Weller