Prognostic relevance of CD163 and CD8 combined with EZH2 and gain of chromosome 18 in follicular lymphoma: a study by the Lunenburg Lymphoma Biomarker Consortium

Wendy B.C. Stevens,† Mattias Mendeville,‡ Robert Redd,§ Andrew J. Clear,‖ Reno Bladegroen,*, Maria Calaminici,*, Andreas Rosenwald,*, Eva Hoster,*, Wolfgang Hiddemann,*, Philippe Gaulard,*, Luc Xerri,*, Gilles Salles,*, Wolfram Klapper,† Michael Pfundschuh,† Andrew Jack,‡ Randy D. Gascoyne,‡ Yasodha Natkunam,‡ Ranjana Advani,‡ Eva Kimby,‖ Birgitta Sander,† Laurie H. Sehn,*, Anton Hagenbeek,‖ John Raemaekers,† John Gribben,*, Marie José Kersten,‖ Bauke Ylstra,‖ Edie Weller*, and Daphne de Jong‡

†Department of Hematology, Radboudumc, Nijmegen, the Netherlands; ‡Department of Pathology, VU University Medical Center, Amsterdam, the Netherlands; §Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA, USA; †Department of Hematology, Barts Cancer Institute, University of London, UK; *Institute of Pathology, Comprehensive Cancer Center Mainfranken, University of Würzburg, Germany; ‖Department of Medicine III, University Hospital Grosshadern, Munich, Germany; ††Department of Pathology and Inserm U955, Hôpital Henri Mondor, University Paris-Est Créteil, France; ‡‡De部ement de Biopathologie, Institut Paoli-Calmettes, Marseille, France; ‡§Service d'Hématologie, Hospices Civils de Lyon & Université Claude Bernard Lyon-1, UMR CNRS 5239, France; †¶Institute of Pathology, University of Schleswig-Holstein, Kiel, Germany; †‖Centre for Haematology, St. Bartholomew's Hospital, London, UK; †‡Department of Pathology, Stanford University School of Medicine, CA, USA; ‡‖Department of Haematology, Stanford University School of Medicine, CA, USA; †¶¶Department of Medicine, Division of Hematology, Karolinska Institute, Stockholm, Sweden; ‡§§Department of Laboratory Medicine, Division of Pathology, Karolinska Institute and Karolinska University Hospital, Stockholm, Sweden and †¶¶¶Department of Hematology, Academic Medical Center, Amsterdam, the Netherlands

*WBSC and MM contributed equally to this work

ABSTRACT

In follicular lymphoma, studies addressing the prognostic value of microenvironment-related immunohistochemical markers and tumor cell-related genetic markers have yielded conflicting results, precluding implementation in practice. Therefore, the Lunenburg Lymphoma Biomarker Consortium performed a validation study evaluating published markers. To maximize sensitivity, an end of spectrum design was applied for 122 uniformly immunochemotherapy-treated follicular lymphoma patients retrieved from international trials and registries. The criteria were: early failure, progression or lymphoma-related death <2 years versus long remission, response duration of >5 years. Immunohistochemical staining for T cells and macrophages was performed on tissue microarrays from initial biopsies and scored with a validated computer-assisted protocol. Shallow whole-genome and deep targeted sequencing was performed on the same samples. The 96/122 cases with complete molecular and immunohistochemical data were included in the analysis. EZH2 wild-type (P=0.006), gain of chromosome 18 (P=0.002), low percentages of CD8+ cells (P=0.011) and CD163+ areas (P=0.038) were associated with early failure. No significant differences in other markers were observed, hereby refuting previous claims of their prognostic significance. Using an optimized study design, this Lunenburg Lymphoma Biomarker Consortium study substantiates wild-type EZH2 status, gain of chromosome 18, low percentages of CD8+ cells and CD163+ area as predictors of early failure to immunochemotherapy in follicular lymphoma treated with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP [-like]), while refuting the prognostic impact of various other markers.
Introduction

The disease course of follicular lymphoma (FL) is characterized by multiple relapses with variable remission durations, which tend to get shorter after each line of treatment. Approximately 15% of patients die within the first few years, largely due to histological transformation or refractory disease. In contrast, the majority of patients show prolonged survival without relapse and a substantial number of patients never require treatment. Currently, clinical factors captured in the Follicular Lymphoma International Prognostic Index (FLIPI) and the adjusted FLIPI are the primary prognostic tools utilized to predict disease progression. FLIPI is based on easily available clinical data designed to offer an accurate yet simple prognostic index. However, the biological behavior of FL is likely determined by a more complex interaction between tumor genetics, microenvironment and patient characteristics.

Several gene expression profiling studies have underlined the major influence of characteristics of the non-malignant tumor microenvironment on prognosis in FL. However, evaluation by immunohistochemistry (IHC) of T-cell and macrophage populations for their prognostic translational impact have produced valuable yet conflicting results. Contradictory results are likely caused by multiple factors, including variable patient selection, heterogeneity of treatments across cohorts, insufficient statistical power due to underrepresentation of poor outcome patients, technical issues in IHC staining and inter-observer variability in scoring. Previous IHC-based validation studies of microenvironment cell populations by the Lunenburg Lymphoma Biomarker Consortium (LLBC), in both diffuse large B-cell Lymphoma (DLBCL) and FL, have highlighted the poor reproducibility of manual scoring even by experienced pathologists, and advocated computer-assisted scoring as being the more reliable technique.

Recently, the mutational spectrum of FL tumor cells and their prognostic value have been reported. Most notably, Pastore et al. demonstrated the value of combining mutation status with clinical FLIPI and performance status to improve upon the prognostic value of FLIPI alone (48-55%) to 64-72% for M7-FLIPI, with a comparable negative prognostic value. This method may also be valuable for very high-risk FL cohorts, as reported by the same group. The objective of the study herein by the LLBC consortium was to critically assess, in a homogeneously treated patient cohort, whether the previously implicated microenvironmental and molecular markers of the tumor have clinically relevant prognostic value. We hypothesized that the microenvironmental and molecular markers would be most prominent when we compared tissue samples of patients with an extremely poor prognosis (i.e., progression or death within 2 years, a well-established criterion for poor prognosis in FL) with those with a very favorable prognosis (a response to first-line treatment lasting >5 years). The LLBC gene panel incorporated the molecular markers which were frequently mutated, and were at that time published, as well as markers that were rarely mutated such as FAS and MYD88, since with our study design we hypothesized the ability to also reveal the prognostic value of rare mutations.

To avoid interlaboratory technical variations and interpretation, all assays were uniformly processed in a single laboratory and IHC was scored using computer-assisted technology based on a previous LLBC FL validation study, and molecular analysis was performed using established next-generation sequencing (NGS) procedures for mutation and copy number analyses.

Methods

Patient selection for an end of spectrum design

Tumor specimens of patients with FL, histologic grades 1, 2 and 3A were retrieved from the randomized Lymphoma Study Association (LYSA) FL2000 study, the German Low-Grade Lymphoma Study Group (GLSG) GLSG2000 study, and the St Bartholomew’s Hospital Registry, London, UK. The patients required treatment and the inclusion criteria in the trials were comparable. All patients were treated with R-CHOP or R-CHOP-like regimens, with or without interferon-α (IFN) maintenance for 2 years. Patients were selected upon the following criteria for an end of spectrum design: (1) early failure (EF), defined as no remission or progression, or lymphoma-related death within 2 years after start of first-line treatment, or (2) long remission (LR), defined as a complete or partial remission lasting >5 years after starting first-line treatment. Patients that fell in between these criteria were not included in this study. The availability of complete clinical information at diagnosis, follow-up data until relapse, progression or at least 5 years post-treatment if the patient was still in remission, and availability of formalin-fixed paraffin-embedded (FFPE) diagnostic biopsy samples were prerequisites for inclusion. Detailed clinical information on demographic parameters, staging procedures, treatment regimens and outcome were collected by the involved data centers (LYSA, GLSG and St Bartholomew’s Hospital).

Microenvironment analysis on tissue microarrays using immunohistochemistry and automated image analysis scoring

Tissue microarrays (TMAs) were constructed centrally according to LLBC validated protocols at the Department of Pathology, Würzburg, Germany, from the biopsy part identified by the pathologist (AR), using duplicate cores of 1 mm diameter. Three µm slides were stained for CD3, CD4, CD8, CD68, CD163, FOXP3, PD1, and p53 (Online Supplementary Table S1) according to standard procedures at the Bartholomew’s Pathology Research Laboratory, London, UK. A computerized system with automated scanning microscope and computerized image analysis (Ariol SL-8, Leica Microsystems, Wetzlar, Germany) was used for scoring as described in the LLBC validation study. Macrophages and all T-cell populations were scored for the whole core, and in the perifollicular and interfollicular areas separately, as described by Wahlin et al. Color and shape class defined positive and negative nucleated cells (T-cell classes), or positive and negative areas (macrophage classes). Cores with less than 50% scorable core surface (non-representative areas or tissue artifacts) were excluded and average scores of duplicates were used when available.

Using the Ariol software algorithm, CD3, CD4, CD8, PD1, FOXP3 and p53 positive nucleated cells were scored as the percentage of all nucleated cells. For CD68 and CD163 the positive area versus the whole area was scored to accommodate the large size and long cytoplasmic extensions of macrophages as an optimal determination of cell numbers. The typical perifollicular infiltration pattern of FOXP3 was scored manually by 3 observers, and was considered positive if at least 2 out of the 3 pathologists (DJf, AR and MC) identified and counted at least one rim of...
densely packed positive cells at the periphery of a follicle.14 All histopathological assessments were carried out without insight into the patient's clinical data and outcome.

**Gene mutation and copy number analysis using NGS**

DNA was extracted from FFPE cores with a QiAamp DNA FFPE Tissue Kit (QiAGEN, Hilden, Germany) and quantified using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Carlsbad, CA, USA). 250ng DNA from each patient sample was sheared on a Covaris S2 (Covaris Inc., Woburn, MA, USA), with settings adjusted to DNA from FFPE tissue.14 NGS libraries were prepared using KAPA Library Preparation kits (KAPA Biosystems, Inc., Wilmington, MA, USA). In short, uniquely 8-bp indexed adapters (Roche NimbleGen, Madison, WI, USA) were ligated to the FFPE-extracted DNA, followed by size selection of fragments in the range of 150 to 400bp.

One aliquot of this library was subjected to shallow whole-genome sequencing (WGS) for genome-wide copy number analysis,15 and another aliquot was subjected to hybrid capture target enrichment (Roche NimbleGen, Madison, WI, USA) for mutation analysis. Eight libraries were equimolarly pooled per capture. The hybrid capture panel covered 122 exons (~50,000 base pairs) of 11 frequently mutated genes (KMT2D, CREBBP, MEF2B, EZH2, EP300, BCL2, FAS, TNNFRF14, CARD11, TNNFAP3 and MYD88) in the FL-LLBC-NGS target enrichment panel (Follicular Lymphoma Biomarker Consortium NGS-panel, Online Supplementary Table S2). All libraries were sequenced on a HiSeq 2000 (Illumina, San Diego, CA, USA); 50bp single-end for shallow WGS and 125bp paired-end for mutation analysis. All sequence lanes were multiplexed with up to 24 barcoded sample libraries. Shallow WGS data was analyzed using the Bioconductor package QDNAseq (v1.5.1).16 For gene mutation analysis, variant calling was performed by VarScan 2 (v2.3.7)15 using very strict criteria, which excluded any synonymous mutations, any intronic mutations with predicted low impact and all germline variants reported in the Single Nucleotide Polymorphism database (dbSNP build142). For prognostic analysis, only non-silent mutations (missense, nonsense, in-frame or frame-shift insertions and deletions) were included.

For detailed laboratory data analysis procedures see the Online Supplementary Methods section.

**Ethical Committee statement**

The study and protocols to obtain human archival tissues and patient data were approved by the local ethical committee of the VU University Medical Center, Amsterdam (FWA00017598) for all patient data were approved by the local ethical committee of the VU University Medical Center, Amsterdam (FWA00017598) for all collaborating centers and comply with the Code for Proper Secondary Use of Human Tissue in The Netherlands.

**Statistics**

Univariate and multivariate analyses were used to evaluate the distribution of the biomarkers for the 2 cohorts. For microenvironment analysis, biomarkers were included in the analysis using the scoring categories as defined above, and the average of the biomarker score from 2 cores was used in the analysis. For mutation analysis, the genes were included in the analysis as mutated or wild-type. To correct for multiple comparisons, the Benjamini-Hochberg method was used, and P-values of less than 0.05 were considered significant. Patient characteristics were summarized with descriptive statistics. The Fisher’s exact test was used to test for association between pairs of categorical variables, and univariate and multivariable logistic regression was used for the binary outcome of cohorts. Odds ratios (OR) and 95% confidence intervals (CI) were reported. The Wilcoxon rank-sum test was used to assess a location shift in the distribution of continuous variables between 2 groups. In a secondary analysis, optimal cut points for 8 continuous markers were determined using recursive partitioning models for a binary outcome. To evaluate agreement among the 9 pathologists for the FOXP3 patterns the free-marginal k statistics of Brennan and Prediger are reported with a bootstrap confidence CI.16,37 Analyses were performed using SAS Software version 9.3 (SAS Institute Inc., Cary, NC, USA; 2005) and R version 3.3.0 (R Core Team, Vienna, Austria; 2016).38

**Results**

**Patient selection and immunohistochemical biomarker assessment**

A total of 122 patients fulfilled the selection criteria and had biopsy material available that met the input requirements for both IHC and gene mutation analysis (EF, n=49 and LR, n=73). In 105 cases a complete set of IHC markers was available (Online Supplementary Table S3). For NGS analysis, DNA of sufficient quality and with sufficient NG read depth could be retrieved for 111 cases, resulting in the complete molecular data for gene mutation analysis and copy number profiles.

For 96 of the 122 cases (79%) both IHC and molecular data could be generated that met our quality criteria and was included for downstream analysis. Table 1 shows the clinical characteristics of the 96 patients (Online Supplementary Table S4 provides the characteristics for all 122 patients, with marginal statistical differences for prognostic subgroup representation). The FLIPI high-risk category was overrepresented in the EF cohort (P=0.009), underpinning the validity of the selection criteria for this end of spectrum design.

**Impact of microenvironment T-cell and macrophage cell populations by cohort**

The distribution of IHC markers per end of spectrum prognostic subgroup analyzed in the whole core is shown in Figure 1 and the Online Supplementary Table S5 (Online Supplementary Table S6 shows the distribution of IHC markers of the 105 cases). Statistically significant differences were found in the median percentage of CD8+ nucleated cells (median EF vs. LR is 7.9% vs. 8.6%, P=0.011) and CD163+ macrophages area (median EF vs. LR is 3.6% vs. 5.2%, P=0.038). In logistic regression analyses, the estimated OR for CD8+ cells are 3.9 (95% CI: [1.5-12.1], P=0.01) for a decrease of 10% CD8+ cells. For the CD163+ area the odds are 2.0 (95% CI: [1.1-4.4], P=0.04) for a decrease of 10% CD163+ area (Online Supplementary Table S7). Adjusting for other IHC markers without the FLIPI score in a multivariable model, only the CD8+ T cells retained significance (OR=4.5, 95% CI: [1.1-21.2], P=0.04), however, with the FLIPI score included they no longer reached significance. No significant differences were found for the other markers (CD3, CD4, CD68, CD163, PD1, FOXP3 and p53) between the two cohorts (Online Supplementary Table S7).

Since CD163 was binary scored as positive or negative, a high overall score for the 2 prognostic subgroups might be caused by a higher cell density and/or by larger individual cells. Visual assessment by two pathologists (DdJ, MC) was performed and confirmed that the higher percentage of positive areas was due to higher cell density. A secondary analysis using recursive partitioning was performed to evaluate markers that could separate the 2 prognostic sub-
groups and to determine the optimal cut points for the identified markers for the whole core. The optimal cut points were 12.6% and 6.3% for CD8+ and CD163+, respectively. The percentage of patients with low levels of both CD8+ T cells and CD163+ area (defined as lower than the optimal cut points) was 79% (95% CI 64 – 91%) vs. 39% (95% CI 16 - 52%) for EF (n=39) vs. LR (n=57) (P<0.001). Similarly, results were obtained if the upper quartiles were used instead of the optimal cut points (12.2% (n=39) vs. 8.4% (n=57) for EF [79% (95% CI: 64 – 91%)] vs. LR [46% (95% CI: 82 – 59%)], P=0.001).

Impact of spatial distribution and perifollicular pattern of FOXP3 by prognostic subgroup

The spatial distribution of T-cell populations and macrophages has been claimed to be of greater influence on prognosis than the overall numbers of infiltrating cells and therefore intrafollicular and interfollicular populations were assessed separately.14,17,19,22,39 However, in this series, we could not validate this claim for most T-cell or macrophage classes. Except in the univariate analysis of the interfollicular population, differences in CD8+ cells and CD163+ area were significant; CD8+ cells with an OR 3.6 (95% CI: [1.2-6.4], P=0.03) and CD163+ area 1.9 (95% CI: [1.1-3.6], P=0.03). Only CD8+ interfollicular populations have a minor influence in multivariate analysis excluding the FLIPI score (OR 3.7, 95% CI [1.2-15.5], P<0.01) (Online Supplementary Tables S8 and S9).

The perifollicular pattern of FOXP3+ T cells was scored manually by 3 pathologists (DdJ, AR, BS). Agreement

Table 1. Clinical characteristics of patients with all immunohistochemical and molecular markers available.

<table>
<thead>
<tr>
<th></th>
<th>Total n = 96</th>
<th>Early failure n = 39</th>
<th>Long remission n = 57</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Barts</td>
<td>8 (8%)</td>
<td>6 (15%)</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>GLSG</td>
<td>80 (83%)</td>
<td>29 (74%)</td>
<td>51 (89%)</td>
</tr>
<tr>
<td>LYSA</td>
<td>8 (8%)</td>
<td>4 (10%)</td>
<td>4 (7%)</td>
</tr>
<tr>
<td>Age at diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (range) &lt; 60</td>
<td>58 (27-75)</td>
<td>61 (27-75)</td>
<td>58 (32-69)</td>
</tr>
<tr>
<td>Age at diagnosis 40-59</td>
<td>50 (52%)</td>
<td>18 (46%)</td>
<td>32 (56%)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>47 (49%)</td>
<td>17 (44%)</td>
<td>30 (53%)</td>
</tr>
<tr>
<td>Male</td>
<td>49 (51%)</td>
<td>22 (56%)</td>
<td>27 (47%)</td>
</tr>
<tr>
<td>Grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 1, 2</td>
<td>75 (78%)</td>
<td>28 (72%)</td>
<td>47 (82%)</td>
</tr>
<tr>
<td>Grade 3a</td>
<td>6 (6%)</td>
<td>3 (8%)</td>
<td>3 (5%)</td>
</tr>
<tr>
<td>Missing</td>
<td>15 (16%)</td>
<td>8 (21%)</td>
<td>7 (12%)</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I-II</td>
<td>2 (2%)</td>
<td>1 (3%)</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>Stage III-IV</td>
<td>94 (97%)</td>
<td>38 (97%)</td>
<td>55 (96%)</td>
</tr>
<tr>
<td>Missing</td>
<td>1 (1%)</td>
<td>0</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>B-symptoms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absent</td>
<td>59 (61%)</td>
<td>23 (59%)</td>
<td>36 (63%)</td>
</tr>
<tr>
<td>Present</td>
<td>35 (36%)</td>
<td>16 (41%)</td>
<td>19 (33%)</td>
</tr>
<tr>
<td>Missing</td>
<td>2 (2%)</td>
<td>0</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>ECOG PS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>32 (33%)</td>
<td>13 (33%)</td>
<td>19 (33%)</td>
</tr>
<tr>
<td>1</td>
<td>58 (60%)</td>
<td>21 (54%)</td>
<td>37 (65%)</td>
</tr>
<tr>
<td>2</td>
<td>2 (2%)</td>
<td>2 (5%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>3</td>
<td>1 (1%)</td>
<td>1 (3%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Missing</td>
<td>3 (2%)</td>
<td>2 (5%)</td>
<td>1 (2%)</td>
</tr>
<tr>
<td>FLIPI risk categories</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>10 (10%)</td>
<td>1 (3%)</td>
<td>9 (16%)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>35 (36%)</td>
<td>11 (28%)</td>
<td>24 (42%)</td>
</tr>
<tr>
<td>High</td>
<td>46 (48%)</td>
<td>26 (67%)</td>
<td>20 (36%)</td>
</tr>
<tr>
<td>Missing</td>
<td>5 (5%)</td>
<td>1 (3%)</td>
<td>4 (7%)</td>
</tr>
<tr>
<td>First-line therapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-CHOP</td>
<td>87 (91%)</td>
<td>34 (87%)</td>
<td>53 (93%)</td>
</tr>
<tr>
<td>R-CHOP-I</td>
<td>9 (9%)</td>
<td>5 (13%)</td>
<td>4 (7%)</td>
</tr>
</tbody>
</table>

Barts: Bartholomew’s Hospital Registry London, UK; GLSG: German Low-grade Lymphoma Study Group; LYSA: the Lymphoma Study Association; ECOG: Eastern Cooperative Oncology Group; PS: performance score; FLIPI: follicular lymphoma international prognostic index; R-CHOP: rituximab, cyclophosphamide, adriamycin, vincristine and prednisone; R-CHOP-I: rituximab, cyclophosphamide, adriamycin, etoposide, prednisolone and interferon-α2a.
between pathologists reached levels similar to the validation study (Brennan-Prediger estimate of 0.78 and 0.83 for cores 1 and 2 in the study herein versus 0.85 for the validation study). The perifollicular pattern did not differ significantly ($P=0.46$) between the 2 prognostic subgroups, 10/39 (26%) in the EF subgroup and 11/57 (19%) in the LR subgroup (Online Supplementary Table S10).

**Frequency distribution of chromosomal copy numbers and gene mutations in FL**

Shallow WGS resulted in high quality genome-wide copy number aberration (CNA) plots for all cases. The most common aberrations, detected in at least 10% of patients, included complete or partial gains of chromosomes 1q, 2, 7, 8, 12 and 18 and losses of 1p, 6q and 10q.

---

**Figure 1. Boxplots per immunohistochemical marker.** For CD4, CD8, CD3, FOXP3, PD1 and P53 they show the percentage of positive nucleated cells of all nucleated cells, and for CD163 and CD68 they show the percentage of positive cell area of the total cell area. Early failure n=39, long remission n=57.
The landscape of CNAs showed notable enrichment of FL-related genes, including focal losses of \(\text{TNFRSF14} (1\text{p}36.32)\), \(\text{TNFAIP3} (6\text{p}23.3)\) and \(\text{FAS} (10\text{q}23.31)\) and focal gains that harbor FL-related oncogenes, like \(\text{BCL11A}\) and \(\text{REL} (2\text{p}16.1)\).

Genes included in the FL-LLBC-NGS target enrichment panel showed non-synonymous mutations in at least 2 FL cases (Figure 3 and Online Supplementary Table S12). \(\text{BCL2}\), a known target for aberrant somatic hypermutation (aSHM) in FL, was most frequently mutated (88/96, 92%) with 0 to 78 mutations per case. Chromatin modifying genes \(\text{KMT2D} (71\%)\) and \(\text{CREBBP} (67\%)\) were mutated with high frequency, and epigenetic modifiers \(\text{EZH2} (18\%), \text{MEF2B} (10\%), \text{EP300} (7\%)\) at lower rates with non-silent mutations in 1-4 of these chromatin modifying genes in 90% of FL patients, consistent with the critical role of epigenetic deregulation in the majority of FL. No patterns of co-occurrence or mutual exclusivity were observed. Non-silent mutations were found in \(\text{TNFRSF14} (50\%), \text{TNFAIP3} (7\%), \text{CARD11} (8\%), \text{FAS} (2\%)\) and \(\text{MYD88} (2\%)\) (Table 2, Online Supplementary Table S13 shows the distribution of mutations of the 111 cases).

**Impact of chromosomal copy numbers and gene mutations by prognostic subgroup**

The distribution of CNAs and gene mutations by prognostic subgroup is shown in Figures 2 and 3. Statistically significant differences after multiple testing correction were only found for gain of chromosomal region \(18\text{p}11.32-q21.33\) (gain \(\text{EF vs. LR} 49\%\) vs. \(12\%, \text{P}<0.001)\), with an estimated OR in logistic regression analysis of 0.15 (95\% CI: [0.05-0.44]) and for \(\text{EZH2}\) mutation status (unmutated \(\text{EF vs. LR} 90\%\) vs. \(72\%, \text{P}<0.001)\) (Table 2). The odds ratio for unmutated \(\text{EZH2}\) was 14.53 (95\% CI: [2.06-635.92]). No significant impact was found for various markers, previously implicated as (borderline) prognostic such as \(\text{CREBBP}\) (OR 0.55, 95\% CI [0.20-1.45]).
In a multivariable model combining the 4 markers, CD8, CD163, EZH2, chromosome 18, and FLIPI mutation, that are statistically significant in the univariate analysis, only the gain of chromosome 18 (OR 0.27 (95% CI: [0.09,0.79], \(P=0.019\)) and EZH2 (OR 13.76 (95% CI: [2.35,264.94], \(P=0.017\)) retained significance. After incorporating the FLIPI score into the model, gain of chromosome 18 (\(P=0.018\)) and EZH2 (\(P=0.036\)) (Table 3) retained significance.

**Discussion**

This LLBC study is, as far as we know, the first to comprehensively explore the combined prognostic impact of microenvironment T-cell and macrophage infiltration and
tumor genetics of FL patients with extremely poor outcome (EF) versus those with a prolonged remission (LR). We show poor outcome to be characterized by a lower number of CD8+ T cells, a smaller CD163+ macrophage area (indicative of fewer macrophages), wild-type EZH2 and a copy number gain of chromosome 18. These observations, in part, confirm previous studies. The gain of chromosome 18, despite its statistically strong prognostic value, has not been reported previously. Equally important, cellular densities of various other cell populations, such as PD1+ T follicular helper (TFH) cells and FOXP3+ T regulatory (Treg) cells, previously claimed to predict clinical outcome, were not confirmed in our study. This study was specifically designed to verify the impact of previously published IHC and molecular markers in FL in the rituximab-chemotherapy era. Therefore, we implemented a dedicated study design to reduce noise such that all cases were retrieved from clinical trials and registries which guaranteed complete clinical information at presentation and detailed treatment information. This allowed us to make a homogeneously treated patient selection, with subgroups at the extreme ends of the prognostic spectrum, the EF and the LR subgroups. By balancing inclusion of the rare EF subgroup, we maximized the sensitivity to observe clinically relevant differences in the microenvironment and mutations, while allowing an overall relatively small patient cohort. To reduce inter-observer variability, the validated quantitative computerized IHC scoring method of the TMA was implemented, which has previously been shown to be more reproducible than any semi-quantitative method.

For microenvironment populations, our results regarding the CD163+ area validate those of Kridel et al., showing that a higher CD163+ pixel count or CD163+ area are independent predictors of prolonged progression-free survival (PFS) in patients treated with R-CHOP (P=0.011-0.050), while the CD68+ macrophages population did not have a significant impact by pixel count or area. Previous reports provided conflicting data for CD68+ macrophages, suggesting correlation with either adverse or favorable outcome. Differences in scoring methods and variations in treatment characteristics may explain these discrepancies.

This LLBC study also showed that a lower percentage of CD8+ cytotoxic T cells in the whole TMA core and inter-follicular areas was associated with EF after treatment with rituximab-chemotherapy. This confirms findings in several series using computerized scoring or flow cytometry of cell suspensions, both in the rituximab era and pre-rituximab era. However, inconsistent results were obtained in studies using manual scoring, corroborating the need for a reliable scoring method of CD8+ T cells. None of the other T-cell markers, including CD8, CD4, PD1 and FOXP3 demonstrated a significant prognostic impact. Debates mostly focus on PD1+ cells, FOXP3+ cells and perifollicular patterns of FOXP3 positive T cells, with conflicting results being published. However, the positive prognostic value of PD1+ cells was seen in studies in which only some of the patients received rituximab, whereas in studies in which the majority of patients were treated with rituximab-chemotherapy, either no effect or a negative effect was reported.}

Frequencies of mutations and distribution of alterations in hotspots and functional domains of the tested genes are largely in line with those previously reported. Almost all FLs are reported to have mutations in 1 or more histone-modifying gene, such as KMT2D (reported as 76-89%), CREBBP (reported as 33-68%), MEF2B (reported as 15%), EP300 (reported as 9-15%) and EZH2 (reported as 7-26%). Histone-modifying genes exert their function largely indirectly via co-activation of various transcription factors, and as such their specific role in B-cell oncogenesis and immunological functions is difficult to predict. Gene expression analysis has suggested that mutated CREBBP may downregulate major histocompatibility complex (MHC) class II genes, resulting in impaired T-cell activation and possibly lower T-cell levels in tumor samples. This hypothesis could not be supported in the study here-in, which showed no association of CREBBP mutation status and total numbers of T cells or specific subsets, apart from an association of CREBBP mutated status with higher numbers of PD1 positive T cells. It should be noted,
however, that the differences, albeit statistically significant, take place in a very narrow dynamic range. KMT2D, which was mutated in 70% of both EF and LR cases, showed a strong correlation to both CD8 and CD163 with high levels of both markers in wild-type cases. KMT2D is known to decrease apoptosis and increase B-cell proliferation both directly and indirectly in germinal center B cells. Conditional mouse models indicate a role in plasma cell and germinal center differentiation. Regulatory alterations, impacting on immunological interactions with T cells or macrophages have not been described, however. In line with this, it is less remarkable that the mutation status of TNFRSF14, a protein that is directly involved in immune response regulation, does correlate with CD4+ and CD8+ T-cell infiltration in tumor samples.

By shallow WGS, copy number profiles of all tumors were studied in both prognostic groups. Of all previously published numerical alterations in FL, only gain of chromosome 18 stood out as an independent prognostic marker. Our results confirm EZH2 as a strong prognostic marker in FL, with wild-type gene status associated with poor disease outcome (EF), but other markers, such as EP300 and TNFRSF14, which have been implicated to have a significant, though minor, impact on prognosis, were not substantiated in our study. This is likely due to selection bias and relative underrepresentation of poor prognosis patients in previous series for which our study design was specifically optimized. This LLBC study design precludes integration of a complete multifactorial prognostic model such as the M7-FLIPI index, however, the prognostic trend of EZH2 as reported by Pastore et al. follows the same direction as in our study, where statistical significance is reached and lack of significance of 4 other markers is confirmed.

Figure 4. Correlation molecular markers and IHC markers. A: CREBBP B: TNFRSF14. Blue bars are mutated, gray bars are wild-type. On the X-axis the number of cases per IHC marker, on the Y-axis the percentages of positive cells or area.
The pertinent question is whether these findings on the prognostic value of microenvironment populations and genomic alterations can be translated to application in daily clinical practice. The mutations in EZH2 are largely clustered in codon Y646 (Online Supplementary Figure S1) and are therefore technically very easily amenable to simple polymerase chain reaction (PCR) techniques, while chromosomal gains can be monitored using fluorescence in situ hybridization (FISH) or NGS.\textsuperscript{36,37} This signifies that EZH2 mutation status and chromosome 18 gain have a high potential for clinical implementation, in contrast to the microenvironment population markers CD8 and CD163. With current techniques, even if optimized, the absolute quantitative differences are too small between these 2 extreme cohorts to become a powerful and clinically useful tool for the scores around the cut point. At best the prediction of the extremes can be used for this purpose.

In conclusion, the literature with regard to IHC prognostic markers in FL has produced highly conflicting results, and concerning mutation analysis, only very limited data are currently available on prognostic impact. By making use of a specialized study design and a homogeneously rituximab-chemotherapy treated group of patients, we can now confirm that lower percentages of CD8+ T cells, CD163+ M2 macrophage areas, EZH2 wild-type status and gain of chromosome 18 in the initial tumor biopsy specimen predict a poor prognosis in FL for this treatment cohort. Of equal importance, in the study herein we could not substantiate the previously reported claims on the prognostic impact of the other most commonly mutated genes, such as TNFRSF14 and EP300, of T-cell populations and classes of macrophages, as well as a perifollicular distribution of FOXP3+ T cells for patients treated with R-CHOP (-like).\textsuperscript{14,16,18,22,24,31,39-41,54} Moreover, the study herein provides further insight into the relationship between gene mutation status and the most relevant microenvironmental populations in FL.

Contributors

The Lunenburg Lymphoma Biomarker Consortium (LLBC) is a collaboration of 10 international lymphoma research groups, each represented by a clinical investigator and one or more hematopathologists and supported by a team of statisticians. Foundation of the LLBC was made possible with a grant from the Van Vlietingen Lymphoma Foundation. EORTC Lymphoma group: Daphne de Jong, John Raemaekers. HOVON Lymphoma group: Daphne de Jong, Marie-José Kersten, Anton Hagenbeek. LLBC: Philippe Gauthier, Gilles Salles, Luc Xerri. British Columbia Cancer Agency: Randy D. Gascoyne, Laurie Senn. ECOG: Randy D. Gascoyne. GLSG: Andreas Rosenwald, Wolfram Klapper, Michael Freindelschuh, Wolfgang Hiddemann, Eva Hoster. NGL: Birgitta Sandue, Eva Kimby. Barst Cancer Institute: Andrew J. Cleary, Maria Calaminici, John Gribben. Leeds Registry: Andrew Jack Stanford. Yasodha Ndtkumon, Ranjana Adani. Dana-Farber Cancer Institute, Boston, USA: Edie Weller, Robert Redd.

Acknowledgments

The authors would like to thank Daoud Sie for his advice and providing genomics and compute infrastructure.

Funding

This study was supported by unrestricted grants from: Genentech/Roche, GlaxoSmithKline, Pfizer Pharma, Teva, Pharmaceuticals/Cephalon and Millennium Pharmaceuticals Inc, Celgene.

References