Automated Analysis of Multidimensional Flow Cytometry Data Improves Diagnostic Accuracy Between Mantle Cell Lymphoma and Small Lymphocytic Lymphoma

Habil Zare, PhD,1,2* Ali Bashashati, PhD,3* Robert Kridel, MD,4,5 Nima Aghaeepour, MSc,1 Gholamreza Haffari, PhD,3,6 Joseph M. Connors, MD,4,7 Randy D. Gascoyne, MD,4,5 Arvind Gupta, PhD,2 Ryan R. Brinkman, PhD,1,8* and Andrew P. Weng, MD, PhD1,5*

Key Words: Automated data analysis; Flow cytometry; Diagnostic algorithm; Mantle cell lymphoma; Small lymphocytic lymphoma; Immunopathology; Bioinformatics

Abstract

Mantle cell lymphoma (MCL) and small lymphocytic lymphoma (SLL) exhibit similar but distinct immunophenotypic profiles. Many cases can be diagnosed readily by flow cytometry (FCM) alone; however, ambiguous cases are frequently encountered and necessitate additional studies, including immunohistochemical staining for cyclin D1 and fluorescence in situ hybridization for IgH-CCND1 rearrangement. To determine if greater diagnostic accuracy could be achieved from FCM data alone, we developed an unbiased, machine-based algorithm to identify features that best distinguish between the 2 diseases. By applying conventional diagnostic criteria to the flow cytometry data, we were able to assign 28 of 44 (64%) MCL and 48 of 70 (69%) SLL cases correctly. In contrast, we were able to assign all 44 (100%) MCL and 68 of 70 (97%) SLL cases correctly using a novel set of criteria, as identified by our automated approach. The most discriminating feature was the CD20/CD23 mean fluorescence intensity ratio, and we found unexpectedly that inclusion of FMC7 expression in the diagnostic algorithm actually reduced its accuracy. This study demonstrates that computational methods can be used on existing clinical FCM data to improve diagnostic accuracy and suggests similar computational approaches could be used to identify novel prognostic markers and perhaps subdivide existing or define new diagnostic entities.
Classification of cases with borderline CD20/CD23 ratios.

Several groups have attempted to address this challenge by closer analysis of FCM data, but most resulting diagnostic algorithms compromise sensitivity for specificity or vice versa. For example, the approach suggested by Morice et al was reported to have 82% sensitivity to CLL/SLL and 56% sensitivity to MCL for 175 studied cases that were CD5+. Another example is the Matutes score that can be computed based on monoclonal light chain immunoglobulin, CD5, FMC7, CD23, and CD22. This approach relies on a subjective assessment of positive vs negative and moderate/strong vs weak staining for each marker, however, and thus is highly sensitive to interobserver variation. Additional markers such as CD54 and CD200 have improved on SLL/MCL discrimination, but their routine use at present is not widespread.

It is widely recognized that data analysis is by far one of the most challenging and time-consuming aspects of FCM experiments and is a primary source of variation in clinical tests. Investigators have traditionally relied on intuition rather than standardized statistical inference in the analysis of FCM data. Our hypothesis is that the accuracy of diagnosis can be significantly improved by using the information that already exists in FCM data but is missed by traditional data analysis approaches. The goal of this study was to find more sensitive and specific FCM features to reduce diagnostic errors, time and effort required for data analysis, and unnecessary use of ancillary tests. Our approach was to use an unbiased algorithm to analyze retrospectively multidimensional FCM data to identify the most informative features. We report that the CD20/CD23 ratio is the single most powerful FCM feature for discriminating between SLL and MCL and can improve diagnostic accuracy over conventional approaches involving binary (ie, positive vs negative) decision criteria applied to each marker individually. In addition, surface immunoglobulin light chain (sIg) intensity and CD11c expression are useful in classifying cases with borderline CD20/CD23 ratios. Unexpectedly, we observed that while FMC7 expression generally correlates with MCL cases overall, it can confound accurate classification of cases with borderline CD20/CD23 ratios.

Materials and Methods

Patient Samples

We identified 114 lymph node biopsy specimens with final pathologic diagnosis of MCL (44 cases [38.6%]) or SLL (70 cases [61.4%]) and for which FCM analysis was performed at the British Columbia (BC) Cancer Agency, Vancouver, Canada, between 1997 and 2010. Patient characteristics for the SLL cases were as follows: 19 females (27%), 51 males (73%), and a median age of 65 years. Patient characteristics for the MCL cases were as follows: 14 females (32%), 30 males (68%), and a median age of 65.5 years.

Pathologic Classification

Final pathologic diagnoses for all cases were determined by expert staff hematopathologists at the BC Cancer Agency after integration of findings from biopsy histology, immunohistochemical and/or FISH analysis where indicated, FCM, and clinical history. All cases with MCL diagnoses were confirmed positive for cyclin D1 by immunohistochemical analysis, and 9 of 10 with available FISH results were positive for IGH-CCND1 translocation. Cyclin D1 immunohistochemical analysis and/or IGH-CCND1 FISH analysis were also performed on 35 (50%) and 6 (9%) of 70 cases of SLL, respectively; all were negative.

FCM Data Acquisition


Photomultiplier voltage settings were substantially modified twice during the 3-color data acquisition period and once during the 8-color data acquisition period, as reported. Because changes in photomultiplier voltage settings of the instrument alter the mean fluorescent intensity (MFI) of the corresponding markers and can mask biologic information, we segregated the data into 5 distinct periods such that photomultiplier voltage settings were essentially constant within each period. The 5 periods included 23, 21, and 44 cases during 3-color era and 19 and 7 cases during 8-color era.

Definitional Criteria for Positive and Negative Marker Expression

According to World Health Organization immunophenotypic classification, typical MCL is CD5+, CD23−, and FMC7+, whereas typical SLL is CD5+, CD23+, FMC7−. In routine clinical practice, the distinction between positive and negative expression for a given marker is often based on absolute threshold values as defined by an internal negative control population in the same staining tube or parallel analysis of unstained patient cells, patient cells stained with isotype control antibodies, or staining of cells from a “normal” control sample. While this approach gives clear results when cell
populations of interest exhibit uniform and bright expression of a given marker, it is much less informative when expression is variable or dim. To compare our automated approach with conventional practice, we applied the \(t\) test as a statistical measure to define positive vs negative marker expression rather than subjective interpretation. Internal control populations were used to define “negative” expression for each marker. For example, CD19– cells were defined as the negative control population for CD11c, CD23, and FMC7 markers, whereas normal CD19+/CD5– B cells served as negative controls for CD5 expression on malignant B cells.

Computational Method for Data Processing

We previously developed the SamSPECTRAL method (publicly available from Bioconductor [http://www.bioconductor.org/packages/devel/bioc/html/SamSPECTRAL.html]) to cluster individual flow data points and, thus, define cell populations automatically. As a preprocessing step to cluster individual flow data points and, thus, define cell populations automatically, we first apply a method similar to the Bolasso algorithm, which is a novel feature selection technique similar to the Bolasso algorithm, to identify FCM features that were most useful in discriminating between MCL and SLL diagnoses.

Setting of Thresholds

For each feature under study, we used the standard R density method, which uses a gaussian kernel, to estimate the densities of MCL vs SLL cases in each of the 5 time frames. We then computed the Bayes error for 1,000 points uniformly distributed in the range of values for each feature and selected the optimal threshold that minimized the Bayes error (ie, the probability of misclassification between MCL and SLL was minimized).

Online Supplementary Material

Supplementary materials for this report can be found at http://www.cs.ubc.ca/~zare/MCL_SLL.html that include a figure showing 4 examples of typical and atypical SLL and MCL cases, and a table of comparative discriminative values based on all 114 study cases. Also, a computational package is provided to reproduce the data analyses performed in this study.

Results

Identifying Discriminative FCM Data Features

We applied our computational approach to the problem of SLL/MCL discrimination in 2 steps. First, we used the SamSPECTRAL clustering algorithm to identify cell populations; then we used our FeaLect method to identify FCM data features of these populations that differentiated best between SLL and MCL. Given that the 114 cases included in this study spanned 5 distinct periods during which cytometer platform and/or voltage settings had changed significantly, we first analyzed data from the third period because it contained the largest number of cases (44/114 [38.5%]) and defined this set of cases as the “training” set.

When all 15 available markers were analyzed for their discriminative value, our unbiased, automated approach revealed that only a subset of markers was useful. Consistent with expectations from clinical practice, these included CD19, CD20, CD5, CD23, FMC7, CD11c, and slg light chain. It was apparent from the data that the features selected by FeaLect were coming from B-cell populations. Although this observation may seem somewhat trivial given that SLL and MCL are B-cell lymphomas, the unbiased nature of the feature selection approach provided strong reassurance that there were indeed no subtle T-cell phenotypes that could contribute to SLL/MCL discrimination.

To compare the performance of our selected features with conventional practice, we objectively identified B cell–containing populations by applying 2 consecutive \(t\) tests. The first \(t\) test was applied to each cluster with the null hypothesis that the CD19 MFI of the cluster was greater than the CD19 MFI of all live cells. Clusters that passed this test with \(P\) value of .001 were designated CD19+. As the rest of the events included CD19dim+ and CD19– cells (in non-CD19+ clusters), we applied another \(t\) test to identify dim from negative clusters, this time using the null hypothesis that the cluster expressed CD19 at higher levels than the rest of the non-CD19+ events. Clusters that passed this test with \(P\) value of .001 were designated CD19dim+. All remaining clusters were designated as CD19–. CD19+ and CD19dim+ clusters were included, and CD19– clusters were excluded from subsequent analyses. We then examined MFI values for each marker individually and all pairwise combinations of MFI ratios for their ability to discriminate between MCL and SLL.
We observed CD23, FMC7, CD20, immunoglobulin light chain, and CD5 (typically regarded as most useful in MCL vs SLL diagnosis\textsuperscript{10,13-24,48-51}) to give variable results when considered individually. We found that when a subset of these was considered as pairwise ratios (ie, CD20/CD23, FMC7/CD23, and CD20/CD11c), these features showed considerably improved ability in discriminating between MCL and SLL.
Developing a Diagnostic Predictor

We next sought to combine these 3 discriminative ratios (CD20/CD23, FMC7/CD23, and CD20/CD11c) into a composite diagnostic predictor, which we have termed the “combined ratio score,” or CRS. The CRS is derived for each case by counting the number of ratios that are above an empirically defined threshold (see “Materials and Methods”). For example, if all 3 ratios are above their corresponding thresholds, the CRS will have a value of 3, which represents the strongest prediction for MCL. Conversely, if all ratios are below their corresponding thresholds, the CRS will have a value of 0, which represents the strongest prediction for SLL.

Since the CRS was initially defined using the 44-sample training set, we tested its performance on an independent, or “validation” set of 70 samples comprising the remaining cases identified in this study. The performance of the CRS on this 70-sample test set is summarized in Table 1, with comparison with each of the component pairwise ratios and with selected individual markers. Although the CRS achieved 100% sensitivity, 92% specificity, and 96% accuracy in

![Discriminative value of individual markers typically used for diagnosis of mantle cell lymphoma (MCL) vs small lymphocytic lymphoma (SLL).](image)

**Figure 1** Discriminative value of individual markers typically used for diagnosis of mantle cell lymphoma (MCL) vs small lymphocytic lymphoma (SLL). “Typical” MCL cases were defined as having a CD5+, CD23−, and FMC7+ B-cell signature, whereas typical SLL cases were defined as having a CD5+, CD23+, and FMC7− B-cell signature. Any cases without a typical MCL or SLL immunophenotype were designated atypical. All MCL cases were confirmed as cyclin D1+ by immunohistochemical analysis. Cases left or right of gray lines are confirmed SLL/MCL cases, accordingly, and the green cutoffs show the Bayes decision boundary (see “Materials and Methods”). The blue and pink curves are estimated densities for SLL and MCL, accordingly. For clarity, only cases from the third time frame (14 MCLs shown in pink, 30 SLLs shown in blue) are depicted.
**Figure 2** Most discriminative fluorescence ratios for mantle cell lymphoma (MCL) vs small lymphocytic lymphoma (SLL). Symbols are as depicted in Figure 1.
In contrast, sIg light chain intensity seemed to improve on the discriminative value of the CD20/CD23 ratio in that most cases with borderline ratios would have been assigned correctly if light chain intensity were considered (ie, dim light chains favoring SLL). The incremental value of considering CD11c expression with the CD20/CD23 ratio seemed variable in that some cases would have been assigned correctly, but others would not. 

In diagnosing MCL in this 70-sample validation set, we noted rather unexpectedly that the CRS performed less reliably than did the best pairwise feature, the CD20/CD23 ratio, alone. In examining further the underlying cause for poorer performance of the CRS, it became apparent that FMC7 was the primary confounding variable in that it was expressed at relatively high levels (and scored as positive by t test comparison) in confirmed SLL cases with borderline high CD20/CD23 ratios. In contrast, sIg light chain intensity seemed to improve on the discriminative value of the CD20/CD23 ratio in that most cases with borderline ratios would have been assigned correctly if light chain intensity were considered (ie, dim light chains favoring SLL). The incremental value of considering CD11c expression with the CD20/CD23 ratio seemed variable in that some cases would have been assigned correctly, but others would not.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Discriminative Values on Test Set*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TP</td>
</tr>
<tr>
<td>Novel ratios</td>
<td></td>
</tr>
<tr>
<td>CD20/CD23</td>
<td>30</td>
</tr>
<tr>
<td>FMC7/CD23</td>
<td>30</td>
</tr>
<tr>
<td>CD20/CD11c</td>
<td>24</td>
</tr>
<tr>
<td>Combined ratio score</td>
<td>30</td>
</tr>
<tr>
<td>Common markers</td>
<td></td>
</tr>
<tr>
<td>CD23 intensity</td>
<td>30</td>
</tr>
<tr>
<td>FMC7 intensity</td>
<td>24</td>
</tr>
<tr>
<td>CD20 intensity</td>
<td>24</td>
</tr>
<tr>
<td>Light chain</td>
<td>18</td>
</tr>
<tr>
<td>CD5 intensity</td>
<td>19</td>
</tr>
</tbody>
</table>

FN: false-negative; FP: false-positive; MCL: mantle cell lymphoma; SLL: small lymphocytic lymphoma; TN: true-negative; TP: true-positive.
* MCL is considered as positive and SLL as negative. TP is the number of MCL cases diagnosed correctly; TN, the number of SLL cases diagnosed correctly; FP, the number of SLL cases wrongly diagnosed as MCL; FN, the number of MCL cases wrongly diagnosed as SLL; sensitivity to MCL is the portion of MCL cases correctly diagnosed by a feature [TP/(TP + FN)]; specificity is the portion of SLL cases diagnosed correctly [TN/(TN + FP)]; accuracy is the portion of all correctly diagnosed cases. Performance higher than 95% is shown in bold face.
Despite these potential limitations of the CRS, it is worth noting that the superior performance of the CD20/CD23 ratio critically depends on optimized setting of the threshold value between MCL and SLL. In fact, while cases with CD20/CD23 ratios distant from the threshold value may be regarded with confidence, cases with CD20/CD23 ratios close to the borderline are less clear. It is in this context that perhaps the CRS holds value. In fact, when the CRS is applied to the entire data set of 114 cases, 36 (82%) of 44 MCL cases received a score of 3, whereas 67 (96%) of 70 SLL cases received a score of 0 or 1. Only 11 (9.6%) of 114 cases received an ambiguous score of 2 and included the remaining 8 (18%) of 44 MCL cases and 3 (4%) of 70 SLL cases. In comparison, using conventional criteria of FMC7+/CD23– for MCL and FMC7–/CD23+ for SLL, only 76 (66.7%) of 114 cases were correctly assigned, leaving 38 (33.3%) of 114 cases with ambiguous FMC7–/CD23– or FMC7+/CD23+ phenotypes (16 MCL, 22 SLL). While most experienced flow cytometrists would incorporate other features such as CD20 and sIg intensity into their diagnostic assessment (albeit in a highly subjective manner), in practice, these ambiguous FMC7–/CD23– and FMC7+/CD23+ phenotypes often elicit ordering of ancillary immunohistochemical and/or FISH tests. Thus, use of the CRS in clinical practice may help reduce the number of additional tests required to establish the diagnosis.

Sensitivity to Thresholds

A common approach to assess the value and robustness of a diagnostic test is through receiver operating characteristic (ROC) analysis, which is performed by plotting the true-positive rate (or sensitivity) vs false-positive rate (or 1 – specificity) while varying the test threshold from the minimum observed value to the maximum observed value. The area under the ROC curve is a measure of overall test performance, and if a curve for a given test lies above that for an alternative test, the former is considered more robust. We thus performed ROC analysis to compare the performance of the CRS against each of the component pairwise ratios and select individual markers. ROC analysis confirmed that the CD20/CD23 ratio was indeed the most robust FCM data feature in discrimination between MCL and SLL.

Discussion

By using an automated, unbiased approach to examine multidimensional FCM data, we attempted in this study to improve on diagnostic accuracy in distinguishing between 2 immunophenotypically related lymphomas, SLL and MCL. The use of CD19 or CD20, CD5, CD23, and FMC7 as positive vs negative markers with consideration of features such as intensity of CD20 and sIg light chain expression is widespread. However, some cases cannot be confidently diagnosed by applying these conventional approaches to FCM data analysis. Our automated algorithm has identified that the CD20/CD23 ratio is the most robust FCM feature for discriminating SLL from MCL. Unexpectedly, inclusion of additional features such as FMC7 and CD11c expression actually results in a greater likelihood of misdiagnosis, most specifically in SLL cases with borderline high CD20/CD23 ratios. In contrast, consideration of immunoglobulin light chain expression may improve diagnostic accuracy for these
borderline cases. Our findings provide new insight into the relative contribution of each FCM data feature to the overall diagnostic algorithm and, by improving the diagnostic accuracy of FCM data analysis, can potentially reduce the amount of confirmatory ancillary testing required.

Because most clinical FCM laboratories already routinely acquire CD20 and CD23 expression data in their standard diagnostic panels for assessment of lymphoproliferative disorders, calculation of the CD20/CD23 ratio should be easy to implement. The optimal cutoff value for CD20/CD23 ratio to discriminate between SLL and MCL, however, will be sensitive to interlaboratory variables such as staining protocols, choice of antibody clones and fluorochrome conjugates, and instrumentation settings and sensitivity. Therefore, our method currently may require an initial calibration step by each laboratory using its own set of training samples to determine the optimal cutoff.

While some studies have shown that the biologic information of FMC7 can also be captured by CD20 intensity,\textsuperscript{54-56} it remains common practice to consider FMC7 expression and CD20 intensity in evaluating FCM data. It is interesting that analysis of the discriminative value of each marker individually suggested FMC7 is superior to CD20 (Figure 1); however, when taken in the context of the CD20/CD23 ratio, FMC7 actually compromised diagnostic accuracy. This apparent contradiction may be due to the fact that while most MCL/SLL cases tend to be positive/negative for FMC7, respectively, this correlation seems to be less meaningful in the particular context of cases with borderline phenotypes.

The current study was limited to examination of FCM data from lymph node samples to avoid other potential variables from confounding this initial analysis. Further studies are warranted to determine if our approach yields similar diagnostic accuracy in peripheral blood and bone marrow samples. Also, as CD5+ marginal zone lymphoma and CD5+ diffuse large B-cell lymphoma are very rare and our approach requires sufficient numbers of cases to perform the automated clustering/feature discovery and statistical analyses, we were not able to include these lymphoma types in the current study.

Existing high dimensional FCM data may potentially contain valuable biologic information that is hidden from conventional analyses because such approaches rely on bivariate plots and manual gating. Our identification of the CD20/CD23 ratio is an example of revealing novel features that capture inapparent but meaningful diagnostic information that already resides within existing FCM data. While our goal for the current study was to develop and test the automated algorithm for improving diagnostic accuracy, this approach is capable of identifying novel multidimensional features that could provide valuable prognostic information or aid in recognition and definition of novel biologic subtypes that are currently subsumed under a single diagnostic heading. Unsupervised clustering of lymphoma samples facilitated by our unbiased automatic approach will be a focus of future work and can potentially lead to discovery of such novel subtypes.

From the \textsuperscript{1}Terry Fox Laboratory, \textsuperscript{2}Department of Molecular Oncology, and \textsuperscript{3}Center for Lymphoid Cancers, British Columbia Cancer Agency, Vancouver, Canada; \textsuperscript{4}Department of Computing Science, \textsuperscript{5}Department of Pathology and Laboratory Medicine, \textsuperscript{7}Faculty of Medicine, and \textsuperscript{8}Department of Medical Genetics, University of British Columbia, Vancouver, Canada; and \textsuperscript{6}Faculty of Information Technology, Monash University, Victoria, Australia.

**Figure 5** Receiver operating characteristic curves obtained for all 114 cases (44 mantle cell lymphoma [MCL] and 70 small lymphocytic lymphoma [SLL]). The CD20/CD23 ratio (brown) is the most accurate (100%) and most robust (area under the curve = 1) flow cytometric feature.
Supported by funding from Natural Sciences and Engineering Research Council of Canada, Ottawa; Mathematics of Information Technology and Complex Systems, Vancouver, Canada; Canadian Cancer Society grant 700374, Toronto, Canada; National Institutes of Health/National Institute of Biomedical Imaging and Bioengineering grant EB008400, Bethesda, MD; Canadian Institutes of Health Research grant 94132, Ottawa; the Terry Fox Foundation, Chilliwack, Canada; and the Terry Fox Research Institute, Vancouver. The British Columbia Cancer Agency received an unrestricted grant from F. Hoffmann-LaRoche, Basel, Switzerland, that was used to support research on the integration of positron emission tomography scanning into lymphoma management. R.R. Brinkman and A.P. Weng are Michael Smith Foundation for Health Research scholars.

Address reprint requests to Dr Brinkman or Dr Weng: Terry Fox Laboratory, BC Cancer Agency, 675 W 10th Ave, Vancouver, BC, Canada V5Z 1L3.

* H. Zare and A. Bashashati contributed equally as first authors; R.R. Brinkman and A.P. Weng contributed equally as senior authors.

Acknowledgments: We acknowledge the hematopathology staff at British Columbia Cancer Agency (BCCA), D. Banerjee, MD, M. Chhanabhai, MD, M. Hayes, MD, A. Karsan, MD, B. Skinnider, MD, and G. Slack, MD, for expert pathologic diagnoses, and the BCCA Clinical Flow Cytometry Laboratory staff for excellent technical work.

References


© American Society for Clinical Pathology


