Accurate Classification of Diffuse Large B-Cell Lymphoma into Germinal Center and Activated B-Cell Subtypes Using a Nuclease Protection Assay on Formalin-Fixed, Paraffin-Embedded Tissues

Lisa M. Rimsza1, George Wright4, Mark Schwartz5, Wing C. Chan6, Elaine S. Jaffe4, Randy D. Gascoyne8, Elias Campo9, Andreas Rosenwald10, German Ott11, James R. Cook6, Raymond R. Tubbs6, Rita M. Braziel7, Jan Delabie12, Tom P. Miller3, and Louis M. Staudt2

Abstract

Classification of diffuse large b-cell lymphoma (DLBCL) into cell-of-origin (COO) subtypes based on gene expression profiles has well-established prognostic value. These subtypes, termed germinal center B cell (GCB) and activated B cell (ABC) also have different genetic alterations and overexpression of different pathways that may serve as therapeutic targets. Thus, accurate classification is essential for analysis of clinical trial results and planning new trials by using targeted agents. The current standard for COO classification uses gene expression profiling (GEP) of snap frozen tissues, and a Bayesian predictor algorithm. However, this is generally not feasible. In this study, we investigated whether the qNPA technique could be used for accurate classification of COO by using formalin-fixed, paraffin-embedded (FFPE) tissues. We analyzed expression levels of 14 genes in 121 cases of R-CHOP-treated DLBCL that had previously undergone GEP by using the Affymetrix U133 Plus 2.0 microarray and had matching FFPE blocks. Results were evaluated by using the previously published algorithm with a leave-one-out cross-validation approach. These results were compared with COO classification based on frozen tissue GEP profiles. For each case, a probability statistic was generated indicating the likelihood that the classification by using qNPA was accurate. When data were dichotomized into GCB or non-GCB, overall accuracy was 92%. The qNPA technique accurately categorized DLBCL into GCB and ABC subtypes, as defined by GEP. This approach is quantifiable, applicable to FFPE tissues with no technical failures, and has potential for significant impact on DLBCL research and clinical trial development. Clin Cancer Res; 17(11); 3727–32. ©2011 AACR.

Introduction

Gene expression profiling (GEP) can be used to classify diffuse large B-cell lymphoma (DLBCL) into 2 biologically based cell-of-origin (COO) subtypes with marked prognostic value that is retained even with immunochromotherapy. The germinal center B cell (GCB) subtype has overexpression of germinal center-related genes, higher rate of BCL2 translocations, and more frequent gains of 12q12. In contrast, the activated B cell (ABC) subtype has overexpression of the NF-xB pathway, and more frequent trisomy 3, gains of 3q and 18q21-q22, losses of 6q21-q22 (1-3). These biological differences strongly imply that different strategies for therapeutic targeting should be used for each subtype (2, 4-7). At this time, there is a need for methods for accurate classification to interpret retrospective and prospective clinical trial results as well as for design of new trials by using targeted agents. The original standard for COO classification used RNA extraction, GEP of snap frozen tissues, and a Bayesian predictor algorithm. Previously 14 key genes were reduced from a larger data set to make this distinction (8). An immunohistochemistry (IHC) classification scheme (9), based on 3 antibodies, is widely used as a substitute for GEP classification; however, IHC does not completely correlate with GEP classification or outcome (10). More recently, an improved IHC scheme based on 5 antibodies has been proposed with improved specificity (11). Both of these techniques rely on semiquantitative assessment of protein levels rather than quantitative mRNA levels. Furthermore, the frequency of cases with failed or uninterpretable IHC results has not

Authors’ Affiliations: 1Department of Pathology, University of Arizona; 2High ThroughPut Genomics; 3Arizona Cancer Center, Tucson, Arizona; 4Metabolism Branch, National Institutes of Health, Bethesda, Maryland; 5Department of Pathology, University of Nebraska, Omaha, Nebraska; 6Pathology and Laboratory Medicine Institute, Cleveland Clinic, Cleveland, Ohio; 7Oregon Health & Sciences University, Portland, Oregon; 8British Columbia Cancer Agency, Vancouver, British Columbia, Canada; 9Department of Pathology, University of Barcelona, Barcelona, Spain; 10Department of Pathology, University of Wurzburg, Wurzburg; 11Institute of Pathology, Robert-Bosch Hospital and Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, Stuttgart, Germany; and 12Department of Pathology, Norwegian Radium Hospital, Oslo, Norway

Corresponding Author: Lisa M. Rimsza, Professor and Associate Chair of Research, Department of Pathology, 1501 N. Campbell Ave, Room 5208A, P.O. Box 245043, Tucson, Arizona 85724, Phone: 520-626-8396; Fax: 520-626-6081. E-mail: lrimsza@email.arizona.edu

doi: 10.1158/1078-0432.CCR-10-2573

©2011 American Association for Cancer Research.
been reported in any of the previous papers. We recently described a quantitative nuclease protection assay (qNPA) (ArrayPlate) for quantification of mRNA in tissues with excellent correlation between frozen and formalin-fixed, paraffin-embedded (FFPE) materials and further extended our pilot series to a large group of DLBCL cases (12, 13). Initially, this assay was developed to measure the mRNA levels of genes previously identified by using GEP or reverse transcriptase PCR as highly prognostic in DLBCL (14–17). In this study, we developed and incorporated additional gene probes to measure genes that were part of the Bayesian predictor algorithm for classifying ABC and GCB subtypes (8). Our hypothesis was that qNPA could be used for accurate classification of COO by using FFPE tissues. The overall goal of the project was to develop a robust, reproducible method to classify DLBCL into COO by using FFPE blocks with the same accuracy as GEP for application in future studies.

Methods

Cases

All work was carried out under an IRB-approved protocol from the University of Arizona, Tucson, Arizona. We retrospectively collected 121 cases of R-CHOP treated, de novo DLBCL from adult patients that had undergone GEP by using the Affymetrix U133 Plus 2.0 microarray and had matching FFPE blocks were analyzed with qNPA in duplicate. Of these, 106 cases were part of the originally published case series by the Lymphoma and Leukemia Molecular Profiling Project (18), with the remaining 15 collected later under the same protocol and so not included in the initial publication. All available cases with successful GEP and which had adequately thick FFPE blocks were included. Five unstained sections of 5 μm thickness were cut from each FFPE block and sent to the University of Arizona, where they were anonymously coded and forwarded to High Throughput Genomics (HTG) for analysis. The HTG personnel were blinded to the COO status of the cases.

Assay design

We expanded the previous gene probe repertoire of the DLBCL-ArrayPlate assay to include the 14 genes most pertinent to COO classification. The genes of interest for this project included CD10, LRMP, CCND2, ITPKB, PIM1, IL16, IRF4, FLT8, BCL6, PTPN1, LM02, CD39, MYBL1, and IGHM (8).

Assay performance

The ArrayPlate assay was carried out as previously described (12, 13). Briefly, following lysis of the tissue sections, 50-mer probes specific for the genes of interest were incubated with the samples, forming specific probe-mRNA duplexes, then unhybridized probes were digested by S1 Nuclease, followed by alkaline hydrolysis to destroy the mRNA in the duplexes. This left intact probes with stoichiometric concentrations proportional to the abundance of specific mRNA in the original tumor sections. Samples were transferred for probe detection to ArrayPlates, which contained 16 unique, covalently-bound, 25-mer “anchor” oligonucleotide spots on the bottom of each well. This universal array was modified to bind 50-mer probes for the genes of interest by exposing the array to a mixture of 50-mer Programing Linker oligonucleotides that contained 25-mer sequences designed to bind 1 of the probes at 1 end, and a 25-mer sequence to bind 1 of the anchor oligonucleotides on the other end. After hybridization, probes from the sample were bound to array elements by the Programing Linker oligonucleotides. A mixture of Detection Linker oligonucleotides was added. The 50-mer Detection Linkers contained a 25-mer sequence that bound sample probe on the end not bound by the Programing Linker probe on 1 end, and a common 25-mer sequence to bind a Detection Probe, bound with horseradish peroxidase, on the other end. Finally, a chemiluminescent peroxidase substrate was added to allow quantification (Lumigen PS-atto, Lumigen, Inc.). Figure 1 shows a schematic of the assay. The signals were recorded simultaneously by imaging the plate from the bottom with a CCD-based Omix Imager (HTG). Images were analyzed by using Vuescript software (HTG). Expression levels were normalized to the housekeeping gene TATA Box Binding Protein [TBP (13, 19)].

Data analysis

By using the predictor previously described, samples were initially subtyped into ABC and GCB categories based on a subset of 187 probes from the U133 Plus array, resulting in 58 ABC samples, 49 GCB samples, and 14 with intermediate gene expression which were left unclassified (18).

The qNPA data were normalized to the housekeeping gene, TBP as in previous studies (13, 19). The data were log2 transformed. Following a leave-one-out framework,
Bayesian predictive models were generated based on those qNPA assays that were found to be significantly different between ABC and GCB with \( P < 0.001 \) (8). Briefly this was done by, removing a test sample from the data set and predicting its subtype based on a predictor trained on the remaining samples. To do this, a predictor score was calculated, which was equal to the weighted sum of the log transformed normalized qNPA values for the left out sample, with weights equal to the \( t \) statistic for ABC/GCB differential expression for that qNPA probe on the remaining training set. This score was then compared with the distributions of scores within the ABC and GCB subsets in the training set to arrive at a Bayesian probability for the sample to be ABC versus GCB.

The classification results based on qNPA were then compared with the previous COO classification based on frozen tissue GEP profiles (18). Because different cross-validation runs included different weights and reference distributions, the predictor scores for individual cross-validation runs were not directly comparable. To present comparable scores in Figure 2, the scores of an individual cross-validation run included different weights and refer-ences.

**Results**

All 14 genes in all 121 cases were successfully analyzed. Although gene selection was cross-validated, the data were sufficiently uniform that all cross-validated runs on qNPA identified the same 12 probe sets with the most significance (\( P < 0.001 \)). The 2 eliminated genes were \( IgHM \) (\( P = 0.06 \)) and \( PTEN \) (\( P = 0.017 \)). The \( P \)-values for the remaining 12 genes were as follows: \( CD10, P = 6.85 \times 10^{-13}; LRMP, P = 2.28 \times 10^{-15}; CCND2, P = 6.87 \times 10^{-6}; ITPKB, P = 1.6 \times 10^{-5}; PIM1, P = 5.62 \times 10^{-3}; IL16, P = 6.43 \times 10^{-7}; IRF4, P = 3.87 \times 10^{-12}; FLI18, P = 1.23 \times 10^{-11}; BCL6, P = 1.29 \times 10^{-7}; LM02, P = 7.96 \times 10^{-5}; CD39, P = 1.11 \times 10^{-7}; MYBL1, P = 1.11 \times 10^{-14} \) (\( P \)-values presented here were calculated across the entire set of 121 samples). All further analysis was carried out with the 12 most significant genes.

Of the 121 cases, 49 were GCB, 58 were ABC, and 14 were unclassifiable as categorized by GEP in the original publication (18). For each case, a probability statistic was generated indicating the likelihood that the classification by using qNPA was accurate. Of the GCB cases, 36/49 (73%) were classified correctly by qNPA with a confidence cutoff of more than 0.9 and 39/49 (80%) classified correctly with a confidence cutoff of more than 0.8. Of the ABC cases, 46/58 (80%) were classified correctly as ABC by using qNPA with a confidence cutoff of more than 0.9 and 39/49 (80%) classified correctly with a confidence cutoff of more than 0.8. These results are tabulated in the top portion of Table 1. When the data were dichotomized into GCB and non-GCB with a confidence cutoff of above or below 50th percentile of the probability statistic, and unclassifiable cases were excluded, 45/49 cases were correctly classified as GCB and 53 of 58 cases were correctly classified as ABC for an accuracy rate of 92%. These data are summarized in the bottom portion of Table 1. The calculated sensitivity is 92% and specificity is 90% for the GCB subtype.
Observed model prediction differences can result from technology platform differences, different cases used in the training set, or from model formulation (12 genes vs. 187 genes). We therefore undertook a comparison of 3 different models for ABC and GCB distinction by examining the probability scores derived from (i) 12 genes measured with the qNPA platform, model derived by using cross-validation as described above, (ii) the same 12 genes measured with Affymetrix platform retrained via cross-validation on the 121 test samples, and (iii) the 187 genes predictor trained on a separate set of 181 CHOP cases as previously described (18). A summary of these models is shown in Table 2. Two-way graphs comparing the model scores generated by each model plotted against each other are shown in Figure 2. On the left, 2 sets of Affymetrix data are shown, 1 with 12 genes and the other with 187 genes. Inspection reveals good general agreement between the 2 models. Any scatter in the plot was considered to result primarily from the reduced number of genes in the model and possibly the different set of cases used for cross-validation, however, not from the technical platform, which was the same. On the right, is a comparison of the 12 gene model by using qNPA versus the 187 gene model by using Affymetrix plotted against each other, R-scale value of 0.89, P < 0.0001. Of interest, the amount of scatter in the plots is similar indicating that most of the variability in the results was because of the use of 12 (vs. 187) genes rather than a difference in the technical platform. Afy, Affymetrix.

Table 1. Molecular subclassification using quantitative nuclease protection data

<table>
<thead>
<tr>
<th>Quantitative nuclease protection assay</th>
<th>ABC</th>
<th>GCB</th>
<th>Unclassified</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Affymetrix GEP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABC (N = 58)</td>
<td>46</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>GCB (N = 49)</td>
<td>1</td>
<td>36</td>
<td>12</td>
</tr>
<tr>
<td>Unclassifiable (N = 14)</td>
<td>3</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

Quantitative nuclease protection assay

<table>
<thead>
<tr>
<th>GCB</th>
<th>Non-GCB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Affymetrix GEP</strong></td>
<td></td>
</tr>
<tr>
<td>GCB (N = 49)</td>
<td>45</td>
</tr>
<tr>
<td>ABC (N = 58)</td>
<td>5</td>
</tr>
</tbody>
</table>

*92% sensitivity and 90% specificity for GCB subtype.
Conclusion

The goal of the project was to develop a technique that would be useful for assessment of FFPE blocks with the same accuracy as GEP so that we could use this for interrogation of other collected case series. We reasoned that such a technique would have wide application in DLBCL research and clinical trials. In this study, we show that the qNPA technology can categorize DLBCL into GCB and ABC subtypes with high accuracy as compared with GEP. Importantly, both the laboratory work and classification analysis were carried out blinded to the COO category based on the previous GEP and used cases from the 1 of our previously published papers which was used to define ABC and GCB in the literature. Also, the algorithm previously used for GEP classification was again employed with no modifications. Of note, there were no technical difficulties with any of the pathologic materials although they were collected retrospectively from a variety of institutions and countries with different fixation methods. This could be a significant advantage of the qNPA technique over IHC, which may be sensitive to fixation and antigen retrieval methods. Of note, neither previous IHC nor molecular studies have reported on the rate of unsuccessful or uninterpretable cases.

Other methods for comparing mRNA levels in paraffin-embedded tissues rather than frozen tissues are also under development as recently shown by using the Affymetrix system (20). Although accurate, the Affimmetrix method still requires RNA extraction and performs best with a linear amplification technique to provide a sufficient amount of template. Additionally, the study excluded any cases previously assigned to the Unclassifiable category using complete GEP from previous studies, thus eliminating 11 of 59 samples, resulting in a 48-case series with a call rate of 93.8% using 100 of the most predictive genes (21). In the current study, we used all available samples including those with gene expression profiles that had previously fallen into the Unclassifiable category (15, 18).

As shown in the model comparison, the larger number of genes that can be included in an assay, the more similar the results will be to the original complete GEP model with 187 genes. The assay tested in this study was based on the most statistically powerful 14 genes, of which 12 were finally used. However, additional genes could be readily added to the design that would undoubtedly further improve the accuracy compared with complete GEP.

Limitations of the qNPA assay include the relatively complicated technology compared with IHC and use of a patented array that is not widely implemented at this time. Of interest, a new imaging platform has just been announced that will use a Luminex bead technology called the qBead Assay (Luminex Corporation) rather than the Omix imager used for this project. Advantages include the lack of RNA extraction, use of FFPE tissues, and more quantitative data than IHC. The qNPA method yielded data on all cases with a perfect “intent to categorize” rate, whereas the failure rate of other methods has not been reported. In summary, we report here a method that has the potential for significant impact on future DLBCL research and clinical trial development not only with regard to COO classification but also in the development of other applications of GEP data.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 24, 2010; revised February 4, 2011; accepted February 7, 2011; published OnlineFirst March 1, 2011.

Table 2. Models for comparison

<table>
<thead>
<tr>
<th>Training set</th>
<th>Genes</th>
<th>Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Gold standard”</td>
<td>74 ABC, 76 GCB (8)</td>
<td>187</td>
</tr>
<tr>
<td>qNPA</td>
<td>58 ABC, 49 GCB Cross-validated</td>
<td>12</td>
</tr>
<tr>
<td>Affymetrix equivalent to qNPA</td>
<td>58 ABC, 49 GCB Cross-validated</td>
<td>12</td>
</tr>
</tbody>
</table>

References


Accurate Classification of Diffuse Large B-Cell Lymphoma into Germinal Center and Activated B-Cell Subtypes Using a Nuclease Protection Assay on Formalin-Fixed, Paraffin-Embedded Tissues

Lisa M. Rimsza, George Wright, Mark Schwartz, et al.

Clin Cancer Res 2011;17:3727-3732. Published OnlineFirst March 1, 2011.