Clinical quantitation of diagnostic and predictive gene expression levels in follicular and diffuse large B-cell lymphoma by RT-PCR gene expression profiling

Ebrahim Sakhinia,1 Caroline Glennie,1 Judith A. Hoyland,1 Lia P. Menasce,2 Gerard Brady,3 Crispin Miller,4 John A. Radford,5 and Richard J. Byers1

1Division of Regenerative Medicine, School of Medicine, Faculty of Medical and Human Sciences, The University of Manchester, Manchester, United Kingdom; 2Department of Histopathology, Christie Hospital National Health Service (NHS) Trust, Manchester, United Kingdom; 3Epistem Ltd, Manchester, United Kingdom; 4Paterson Institute for Cancer Research, Manchester, United Kingdom; 5Cancer Research United Kingdom Department of Medical Oncology, The University of Manchester and Christie Hospital NHS Trust, Manchester, United Kingdom

Recent microarray gene expression profiling studies have identified gene signatures predictive of outcome, so-called “indicator” genes, for diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL). However, measurement of these genes in routine practice remains difficult. We applied real-time polymerase chain reaction (PCR) to polyA cDNAs prepared from 106 archived human frozen lymph nodes (63 of FL, 25 of DLBCL, 10 reactive lymph nodes, and cases with paired samples of FL [4] and subsequent DLBCL [4]). Reverse transcription and polyA reverse transcriptase (RT)–PCR was performed, and resultant cDNA was probed by real-time PCR for 36 candidate indicator genes, selected from microarray studies. Nine genes showed statistically significant different expression between FL and DLBCL, including cyclin B, COL3A1, NPM3, H731, PRKCB1, OVGL, ZFPC150, HLA-DQ-a, and XPB. Of these, cyclin B, NPM3, and COL3A1 were higher in DLBCL. Six genes showed statistically significant higher expression in the neo-plastic nodes compared with reactive nodes, namely PRKCB1, BCL-6, EAR2, ZFX, cyclin B, YY1. High levels of YY1 were associated with a shorter survival interval in both FL and DLBCL. The method is simple, sensitive, and robust, facilitating routine use and may be used as a platform for clinical measurement of prognostic gene signatures. (Blood. 2007; 109:3922-3928)

© 2007 by The American Society of Hematology

Introduction

Microarray gene expression profiling has identified gene signatures, or “indicator” genes, predictive of outcome in many cancer types, including diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL).1-4 There is now an urgent need to translate these signatures to clinical use. However, gene microarrays rely on relatively large amounts of fresh starting tissue obviating measurement of indicator genes in routine practice,1 and there is a need for development of another, simple, robust, relatively inexpensive and sensitive method for their translation to clinical use. To test the use of indicator genes as a diagnostic tool, specifically in lymphoma, we have developed a simple, practical polyA polymerase chain reaction (PCR)–based method for analysis of indicator profiles in DLBCL and FL, using very small tissue samples.

PolyA PCR coordinately amplifies cDNA copies of all polyadenylated mRNAs, thereby generating a PCR product (polyA cDNA) whose composition reflects the relative abundance of all expressed genes in the starting sample.5-7 PolyA PCR enables global mRNA amplification from picogram amounts of RNA and has been routinely used to analyze expression in small samples, including single cells.5 The polyA cDNA pool generated is also indefinitely renewable and as such represents a “molecular block.”5,10 The polyA cDNA can then be assayed for the expression of particular genes, either by hybridization with cDNA microarrays5 or by real-time PCR. Real-time PCR measurement, however, enables more precise quantitation of the expression levels of specific indicator genes. As such it is better suited for measurement in the clinical arena, while also focusing on diagnostically relevant indicator genes. This approach thereby enables gene signatures to be detected within very small amounts of starting material.11

In this study, polyA reverse transcriptase (RT)–PCR was applied to RNA extracted from archived human frozen lymph nodes, and the resultant cDNA was analyzed by TaqMan real-time PCR for 36 indicator genes (abstracted from Husson et al,2 Rosenwald et al,3 Shipp et al4) to find prognostic and diagnostic genes within this set of microarray-identified candidate genes.

Material and methods

Clinical samples

One hundred eleven archived human frozen lymph nodes (LNs), (with at least 5 years follow-up) with a diagnosis of follicular lymphoma or diffuse large B-cell lymphoma were obtained, with informed consent, from the archives of the Christie Hospital NHS Trust, Manchester, United Kingdom. This work was approved by the Central Manchester Multicentre Research Ethical Committee. The samples were selected from the archive on a...
sequential chronologic basis, taking all samples for which consent was given, in the archive from 1995 backward to 1989. These were initially assessed for suitability on the basis of biopsy size, amount of necrosis, and confidence of diagnosis. All cases unlikely to have sufficient material for analysis, with more than 10% necrosis or with uncertain or complex/mixed diagnoses, were excluded. Following this initial selection 111 cases were available for study, and paraffin-embedded sections from these cases were reviewed by 2 pathologists (R.J.B. and L.P.M.). In 5 of the cases the diagnosis was revised to mantle-cell lymphoma and excluded from the study. The remaining 106 cases were classified into the following groups (n refers to numbers of frozen samples): group 1, follicular lymphoma (WHO grade 1 or 2) (frozen tissue available) [DLBCL(F) post-FL(F)]; group 2, diffuse large B-cell lymphoma arising de novo without any evidence of preexisting follicular lymphoma (DLBCL) (n = 25); group 3, paired frozen samples of pure follicular lymphoma (WHO grade 1 or 2) (frozen tissue available) and subsequent transformed diffuse large B-cell lymphoma (frozen tissue available) [FL(F) pre-t-DLBCL(F)] (n = 4); these were matched pairs with samples in group 4; group 4, paired frozen samples of diffuse large B-cell lymphoma (frozen tissue available) for which there had been a previous biopsy showing pure follicular lymphoma (WHO grade 1 or 2) (frozen tissue available) [DLBCL(F) post-FL(F)] (n = 4); these were matched pairs with samples in group 3; group 5, reactive lymph node (RLN) (n = 10).

The groups and clinical data are summarized in Table 1.

### Extraction of RNA and global amplification of polyadenylated mRNAs (PolyA RT-PCR)

The lymph nodes were homogenized using a Mixer Mill MM 300 (Qiagen, Crawley, West Sussex, United Kingdom). Total RNA was then extracted by an RNeasy mini kit (Qiagen), as recommended by the manufacturer; DNase II (Qiagen) was used to remove the contaminating genomic DNA, and PolyA RT-PCR was carried out.

### Specific RT-PCR

Taqman PCR primers and probes were designed for 36 indicator genes and 4 housekeeping genes using Primer Express Software (Applied Biosystems, Foster City, CA) and are listed in Table S1 (available on the Blood website; see the Supplemental Table link at the top of the online article). All PCR primer pairs were designed for mRNA sequence within 300 base pair (bp) of the 3’ end of each indicator gene (Table S1) and were tested in PCR reactions carried out in 25 μL containing 1 ng PolyA cDNA, 0.33 μM of each oligonucleotide, 0.5 U Ex-Taq polymerase (TaKaRa), and 0.25 μM dNTPs in the buffer supplied by the manufacturer. PCR was performed using the following thermal cycle: 5 minutes at 94°C, followed by 30 cycles of 30 seconds at 94°C, 30 seconds at 56°C, and 1 minute at 72°C.

### Taqman real-time quantitative PCR

For each gene, Taqman PCR was applied to 1 ng PolyA cDNA from each sample and to 10 μL serially diluted human genomic standards, using a Taqman Gold kit. All samples were analyzed using an ABI Prism 7700 sequence detection system (Foster City, CA). The copy number of each gene was determined by reference, after normalization to Mhouse (below), of the real-time PCR expression level to human genomic DNA standards.

### Normalization

The expression levels of 4 housekeeping genes (IF2-beta, GAP, human ribosomal protein S9 mRNA, and β-actin) were measured by RT-PCR in each sample. Copy numbers obtained for the mean (Mhouse) of the 4 housekeeping genes (IF2-b, GAP, RbS9, and β-actin) in each sample were divided by the highest Mhouse in all samples, resulting in a normalization correction factor.

### Statistical analysis

The data were not normally distributed; therefore, nonparametric tests were used. Statistical analysis of the expression levels of the 36 genes in the 5 groups was performed using the Mann-Whitney test and Kruskal-Wallis with a P value at least .05 for statistical significance. All the tests were performed using SPSS software v13 (Woking, Surrey, United Kingdom).

### False discovery analysis

Because multiple statistical tests were performed, false discovery rate analysis was performed to identify genes significant after accounting for multiple testing. Specifically, the sīggens package in Bioconductor was used to estimate false discovery rate (FDR) by applying the SAM method on the log2 expression levels for all the genes fall across a large range, the log2 scale was used to plot all values to aid visual comparison. The ranking of the expression levels of indicator genes between samples was presented as the mean rank statistical difference.

### Results

PolyA cDNA was generated from mRNA extracted from all the 106 archived human frozen lymph nodes. The copy number for each gene in each sample was reproducible over at least duplicate tests, demonstrating reliability of the method. Specific PCR of these polyA cDNAs was positive for each of the selected indicator genes, and bands of appropriate sizes were present in all samples, demonstrating presence of the target transcript in the polyA cDNA (data not shown). There was no statistically significant difference in the value of Mhouse for the different diagnostic groups (data not shown). Following normalization of the data, the following comparisons were made.
Comparison of expression among FL (group 1), DLBCL (group 2), and RLN (group 5)

Ten genes, namely PRKCB1, OVGL, ZFPC150, BCL6, XBP, EAR.2, ZFX, NPM3, cyclin B, and YY1, showed statistically different expression between reactive, FL, and DLBCL samples, of which PRKCB1 ($\text{P} \leq .004$), OVGL ($\text{P} \leq .025$), ZFPC150 ($\text{P} \leq .031$), BCL6 ($\text{P} \leq .001$), XBP ($\text{P} \leq .048$), EAR2 ($\text{P} \leq .032$), and ZFX ($\text{P} \leq .004$) were up-regulated in FL compared with RLN samples, whereas NPM3 ($\text{P} \leq .052$) and YY1 ($\text{P} \leq .031$) were up-regulated in FL and DLBCL compared with RLN samples, and cyclin B ($\text{P} \leq .012$) was up-regulated in DLBCL compared with FL and RLN samples (Figure 1). Comparison between both FL and DLBCL ($n = 88$) with RLN ($n = 10$) samples identified 6 genes showing statistically significant difference, specifically PRKCB1 ($\text{P} \leq .042$), BCL6 ($\text{P} \leq .001$), EAR2 ($\text{P} \leq .041$), ZFX ($\text{P} \leq .007$), cyclin B ($\text{P} \leq .045$), and YY1 ($\text{P} \leq .020$) (ACTA and HSF were close to significance at $\text{P} \leq .053$), each of which showed higher expression in the neoplastic nodes (Figure 2). With a false discovery rate (FDR) of 5.5%, 3 of these genes (BCL6, YY1, ZFX) were differentially expressed; at 8.0%, 6 of them (PRKCB1, BCL6, XPB, ZFX, cyclin B, YY1) were differentially expressed; and at 11%, 9 of them (PRKCB1, OVGL, ZFPC150, BCL6, XPB, ZFX, NMP3, cyclin B, YY1) were differentially expressed.

Comparison of expression between FL (group 1) and DLBCL (group 2)

Nine genes showed statistically significant difference in expression between FL ($n = 63$) and DLBCL ($n = 25$) in frozen samples, of which cyclin B ($\text{P} \leq .025$), COL3A1 ($\text{P} \leq .025$), and NPM3 ($\text{P} \leq .024$) were up-regulated in FL, and H731 ($\text{P} \leq .02$), PRKCB1 ($\text{P} \leq .01$), OVGL ($\text{P} \leq .019$), ZFPC150 ($\text{P} \leq .030$), HLA-DQ-a ($\text{P} \leq .041$), and XBP ($\text{P} \leq .024$) in DLBCL (Figure 3). With an FDR of 6.0%, 6 of these genes (H731, PRKCB1, OVGL, ZFPC150, HLA-DQ-a, XBP) were found to be differentially expressed, the same 6 genes were found at 8.6%, and 9 genes (cyclin B, COL3A1, NPM3, H731, PRKCB1, OVGL, ZFPC150, HLA-DQ-a, and XBP) were found at 11.8%.

Comparison of expression between (1) matched samples of FL before (group 3) and after (group 4) transformation to DLBCL and (2) FL before transformation (group3) and FL without subsequent transformation (group 1)

Only one gene, PKCG ($\text{P} \leq .029$), showed statistically significant different expression between matched cases of FL before transformation (group 3) and after transformation to DLBCL (group 4) (Figure 4A). It was not possible to choose an appropriate value for $\Delta$ to yield reasonable FDRs, probably a consequence of the small number of replicates in the comparison. Conversely, 2 genes showed statistically significant different expression in cases of FL that either did (group 3) or did not (group 1) go on to transform, namely ACTA ($\text{P} \leq .046$) and HLA-DQ-a ($\text{P} \leq .029$) (Figure 4B); at an FDR of 4.5%, one gene (HLA-DQ) was identified.

Comparison of expression between cases of FL and DLBCL dead or alive after follow-up

YY1 showed statistically significant ($\text{P} \leq .047$) up-regulation in those patients with FL dead ($n = 47$) compared with those alive.
Several genes, namely, HSP40 (P = .007), MINOR (P = .020), H731 (P = .010), and YY1 (P = .003) showed statistically significant difference in expression between patients with DLBCL who were dead (n = 22) and alive (n = 3) after follow-up (Figure 5B).

Survival analysis for all cases of FL and DLBCL

For both FL and DLBCL, the data for each gene was grouped into 4 quartiles for use in Kaplan-Meir survival analysis, including all cases, both dead and alive (Figure 6A). In FL, 3 genes showed statistically significant association with survival, namely XPB, TNF, and YY1, and specifically high levels of YY1 were associated with shorter survival interval (Figure 6B). In DLBCL, 3 genes showed statistically significant association with survival, namely YY1, TNF, and HLA-DQ-a. High levels of YY1 and low levels of HLA-DQ-a and TNF were associated with shorter survival interval (Figure 6C).

Discussion

Recent gene expression profiling has identified gene signatures predictive of outcome, so-called indicator genes, for diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL). However, measurement of indicator genes in routine practice remains difficult. We have demonstrated utility of real-time PCR measurement of indicator genes in globally amplified polyA cDNA as a practical method for their clinical analysis. PolyA PCR enables global mRNA amplification from picogram amounts of RNA, and the polyA cDNA pool generated is indefinitely renewable, representing a molecular block. Real-time PCR measurement of the expression levels of specific indicator genes then allows gene signatures to be detected in the polyA cDNA. In this project we applied real-time PCR for 36 indicator genes to polyA cDNAs prepared from 106 archived human frozen lymph nodes; specifically, 63 cases of FL, 25 of DLBCL, 10 RLNs, and 4 cases with paired samples of FL and subsequent DLBCL were analyzed.

Nine genes showed statistically significant different expression between FL and DLBCL, including cyclin B, COL3A1, NPM3,
Figure 4. Expression levels (log₂) of genes with statically significant difference. Differences were (A) between samples of FL before (group 3, FL pre-t-DLBCL) and after (group 4, DLBCL post-FL) transformation and (B) between FL before transformation (group 3, FL pre-t-DLBCL) and FL without subsequent transformation (group 1, FL). For each box, plots are shown in top panel and mean rank statistic in bottom panel.

Figure 5. Expression levels (log₂) of genes with statically significant difference. Difference was between cases of (A) FL and (B) DLBCL dead or alive after follow-up. For each for each box, plots are shown in top panel and mean rank statistic in bottom panel.
obscure; it was overexpressed in cases of DLBCL refractory to treatment. The data for each gene were grouped into 4 quartiles for use in Kaplan-Meir survival analysis, and these are indicated in the survival curves as 1 = first quartile, 2 = second quartile, 3 = third quartile, and 4 = fourth quartile; first quartile used for lower end of gene expression for each gene.

H731, PRKCB1, OVGL, ZFPC150, HLA-DQ-a, and XPB. Of these, cyclin B, a cell-cycle gene, NPM3, a nucleolar phosphoprotein, and COL3A1 were higher in DLBCL. Interestingly, both cyclin B and NPM3 have been associated with mitogenesis in tumors in other studies, concurrent with up-regulation in DLBCL, which is a more aggressive tumor, with a higher cell growth fraction than FL. Interestingly, Okuda et al. demonstrated that NPM1, a nucleolar phosphoprotein, is a target of another cell-cycle gene, CDK2/cyclin E, and the same association may explain the up-regulation of both cyclin B and NPM3 in DLBCL. The up-regulation of Co3A1 in DLBCL is harder to explain, although this has also been reported for advanced compared with local ovarian carcinoma, although this probably reflects tumor-associated fibroplasia, which may not be relevant for lymphogenesis. Of the genes down-regulated in DLBCL, of particular interest was the reduction in expression of XPB, a DNA helicase showing lower expression in more aggressive splenic marginal zone lymphoma. Of the remaining genes, HLA-DQ-a and OVGL, both down-regulated in DLBCL, were also down-regulated in cases of FL with foci of transformation, indicating a possible role in the process of transformation. Mutations in the HLA class II genes leading to loss of expression of HLA-DQ have been reported in DLBCL, whereas the importance of immune cells in outcome of FL has also been recognised, supporting such a role. OVGL has been associated with ovarian cancer, but its role in lymphoma is obscure; it was overexpressed in cases of DLBCL refractory to treatment.

Six genes showed statistically significant different expression between the neoplastic (FL and DLBCL) and RLNs. Of these 2, PRKCB1 and ZFX, also distinguished between FL and DLBCL. ZFX is a zinc finger binding protein present in prostate adenocarcinoma, consistent with our finding of up-regulation in the neoplastic nodes. Of greater interest, however, is the finding of up-regulation of cyclin B and BCL-6 in the neoplastic nodes, cell-cycle and antiapoptotic genes that have been associated with malignancy in many tissues, and particularly in lymphoma in previous studies. Ear2 encodes a protein with homology to a steroid receptor and binds to the TGACCT direct-repeat motif, but its role in lymphogenesis is unclear. YY1 (ying-yang 1) is a zinc finger protein reported to positively regulate IL-4 gene expression in lymphocytes. It has also been associated with prostate cancer, but interestingly in this study was also up-regulated in cases of FL dead after follow-up. YY1 inhibits FAS-induced apoptosis and its down-regulation increased sensitivity of an NHL B-cell line to rituximab, whereas the same group also shows that the regulation of Fas resistance by NF-κB is mediated via YY1 expression and activity. In support of this possible role, YY1 was also up-regulated in those cases of DLBCL dead after follow-up, although the numbers of patients alive with DLBCL after follow-up were low.

Although distinction between FL and DLBCL can already be made by conventional morphologic and immunohistochemical means, prediction of outcome of FL is less easy, although recent microarray studies have identified an immune signature associated with survival. This signature has not yet been validated in further studies, although YY1 was associated with death in this study. FL often leads to death following transformation to DLBCL, and a signature predictive of this at initial presentation would be clinically useful because it could direct more aggressive treatment or inform closer follow-up. In this study we identified several genes that may fulfill this role. Specifically, ACTA and HLA-DQ-a were significantly up-regulated in cases of FL that subsequently transformed to DLBCL. However, the number of cases of FL with subsequent transformation was low, and these results therefore require validation in a prospective study. Considering the process of transformation in the matched cases of FL with subsequent DLBCL, PKCG, which encodes for protein kinase C-γ, was up-regulated on transformation. It has not been associated with cancer, although PKC-β overexpression is associated with decreased survival in DLBCL.

Kaplan-Meir survival analysis identified a small number of genes associated with survival interval in both FL and DLBCL. Of these high levels of YY1 were associated with a shorter survival interval in both FL and DLBCL. This is of particular interest given the reported role of YY1 in producing resistance to rituximab resulting from down-regulation of FAS-induced apoptosis. This was reported for a cell culture study and has not been shown in clinical studies, although the result in this project suggests a possible detrimental role for YY1 clinically and that it may act as a predictor of rituximab response. Up-regulation of HLA-DQ-a was associated with longer survival interval, in agreement with its role as a positive predictor of survival in the work of Rosenwald et al. TNF showed differential association with survival interval in FL and DLBCL in which high levels were associated with short and long survival intervals respectively; why this is so is unclear, although it may reflect fundamentally different responses of the immune system to either indolent (FL) or aggressive (DLBCL) disease. Comparison of the gene expression level in patients with either FL or DLBCL alive or dead at the end of the study interval confirmed the importance of YY1 as a predictor of outcome. Interestingly, however, the other genes discriminatory between patients dead and alive at the end of the study period differed from those predictive of length of survival by Kaplan-Meir survival, although the genes in each analysis were statistically significant. This indicates a bimodal survival model in which patients with a
lower survival chance express certain genes, namely H731, MINOR, and HSP40, at higher levels than those with a higher survival chance, because the genes expressed at higher levels in those dead of disease do not match to those predictive of length of survival, with resultant divergence of survival lines in Kaplan-Meir analysis into 2 broad groups, one with long survival and one with relatively short survival. This suggests that FL comprises 2, at least, groups, one in which there is relatively short survival and another of long-term survivors; this matches clinical experience.

This project (1) demonstrates the possibility of using polyA PCR for global amplification of clinical samples and real-time PCR to measure diagnostically informative gene expression profiles in the resultant polyA cDNA molecular block and (2) identifies novel prognostic markers for FL and DLBCL. The method is simple, sensitive, and robust, allowing translation into routine clinical use. Although lymphoma represents a relatively small group of cancer patients, the generic nature of microarray gene profiles for cancer subtypes will facilitate simple extension of the method to other patient groups.

Acknowledgment

This work was supported by the United Kingdom Department of Health, New and Emerging Applications of Technology (NEAT) (grant C020) and was carried out in the Manchester Molecular Diagnostic Centre, Manchester Royal Infirmary, Central Manchester and Manchester Children’s University Hospitals, National Health Service (NHS) Trust.

Authorship

Contribution: E.S. performed technical work and data analysis and wrote the manuscript; C.G. performed technical work and sample tracking; J.A.H. wrote the manuscript and provided scientific support; L.P.M. provided expert histopathology review; G.B. provided scientific support; C.M. provided expert statistical analysis; J.A.R. provided the clinical lead; R.J.B. provided overall supervision, designed experimental plan, analyzed data, and wrote the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: Richard Byers, Rm 1.540, Division of Laboratory and Regenerative Medicine, School of Medicine, Faculty of Medical and Human Sciences, The University of Manchester, Stopford Bldg, Oxford Rd, Manchester, M13 9PT United Kingdom; e-mail: r.byers@manchester.ac.uk.

References

Clinical quantitation of diagnostic and predictive gene expression levels in follicular and diffuse large B-cell lymphoma by RT-PCR gene expression profiling

Ebrahim Sakhinia, Caroline Glennie, Judith A. Hoyland, Lia P. Menasce, Gerard Brady, Crispin Miller, John A. Radford and Richard J. Byers