Selection of Superior Reference Genes’ Combination for Quantitative Real-time PCR in B-cell Lymphomas

Daria Potashnikova1,*, Aleena Gladkikh1,*, and Ivan A. Vorobjev1,2

1 Department of Cell Biology and Histology, Faculty of Biology, M.V. Lomonosov Moscow State University, Moscow, and 2 A.N. Belozersky Institute of Physico-Chemical Biology, M.V. Lomonosov Moscow State University, Moscow, Russia

Abstract. Normalization of real-time quantitative PCR data to appropriate reference genes is crucial to accurately interpret results. Many genes commonly used as reference standards do not perform as expected, depending on cell type and experimental design. In our previous work, we addressed the issue of suitable reference genes for lymphoid tissue and successfully applied the normalization factor-based approach to discriminate lymphoid malignancies according to their cyclin D1 mRNA level. Here, we addressed the problem of reference gene selection and sufficient number on an enlarged sample set with seven candidate genes. The experimental set included 165 samples of spleens, lymph nodes, and peripheral blood mononuclear cells from patients with different types of non-Hodgkin lymphomas along with non-neoplastic lymphoid specimens. For the first time, we compared all major stability ranking algorithms of Visual Basic for Applications (VBA) applets geNorm, BestKeeper, and NormFinder and tested candidate reference genes on a large and heterogeneous set of fresh clinical lymphoid samples. We concluded that a normalization-based approach using three reference genes (YWHAZ, UBC and ACTB) allows for robust reduction of the variance in real-time PCR results and that the further addition of reference genes does not improve data normalization. This creates a clinically applicable tool for PCR-based lymphoma diagnostics.

Keywords: real-time PCR, lymphoid tissue, non-Hodgkin lymphomas, reference genes, gene expression.

Abbreviations: PCR: polymerase chain reaction; NHL: Non-Hodgkin lymphoma; B-CLL: B-cell chronic lymphocytic leukemia; MCL: mantle cell lymphoma; PBMC: peripheral blood mononuclear cells; YWHAZ: tyrosine 3-monooxygenase tryptophan 5-monooxygenase activation protein, zeta polypeptide; UBC: ubiquitin C; HPRT1: hypoxanthine phosphoribosyl-transferase 1; GAPDH: glyceraldehyde 3-phosphate dehydrogenase; RPL13A: ribosomal protein L13a; B2M: β2-microglobulin; ACTB: β-actin; PRKG1: protein kinase cGMP-dependent, type I; GUSB: β-glucuronidase; TBP: TATA box binding protein; HPLC: high performance liquid chromatography; SD: standard deviation; SEM: standard error of mean; NF: normalization factor.

Introduction

Gene expression analysis is very important in bioscience and medicine. To assess the expression of specific mRNAs, researchers use a variety of techniques, including Northern blotting, nuclease protection assays, microarrays, and real-time PCR. Real-time PCR focuses on the expression-level assessment of several genes in large numbers of biological samples and has enjoyed great popularity over the last two decades due to its accuracy, sensitivity, and reproducibility [1,2,3]. Although this method has proved valuable in both fundamental and clinical research, there are nevertheless concerning matters, and getting accurate results depends on how well they are addressed in the experiment. With increasingly user-friendly technologies and advanced machinery, software, and reagents, the bottleneck appears to be optimal experimental design and result interpretation. To be considered are the amounts of starting material, enzymatic and primer efficiencies, and differences in tissue and cell types and their transcriptional activity. Ignoring these factors may yield inaccurate data [4], and the more scientists study gene expression, the more rigorous standards for analysis are developed to ensure the high quality of published results [5,6]. However, the major issue that determines the method’s accuracy is correct data normalization.

*These authors contributed equally to this study. Address correspondence to Daria Potashnikova, M.Sc., Department of Cell Biology and Histology, School of Biology, M.V. Lomonosov Moscow State University, 1/12 Leninskie Gory, 119991, Moscow, Russia; phone: +7 910 442 2728; fax: +7495 939 2084; e-mail: dpotashnikova@gmail.com

0091-7370/15/0100-064. © 2015 by the Association of Clinical Scientists, Inc.
There are several approaches to real-time PCR data normalization [7,8]. To date, the most common method for minimizing non-biological variances has been the use of internal control genes to normalize the mRNA fraction. Previous studies proved that normalization to a single control gene was inappropriate for human tissue samples [9,10]. The strategy based on using several internal control genes combined into a normalization factor was applied to a wide range of different experimental setups and increased the reliability of clinical and experimental real-time PCRs. Panels of reference genes were tested for many cell and nosology types; for example, in cancer research this was done for colon cancer [11,12].

Advances in molecular biology, genetics, and immunology have resulted in extensive changes in NHL classification in the last decades, but the molecular mechanisms of lymphomagenesis, including gene expression variation, remain poorly understood. For instance, cyclin D1 is one of the important markers in lymphoma differential diagnostics. Its overexpression in certain cases can be determined immunohistochemically (Figure 1). However, immunohistochemistry has a very limited dynamic detection range that restricts its use and calls for other more universal detection methods. Hence the need for real-time PCR that has both a high sensitivity and a large dynamic detection range.

Non-Hodgkin lymphoma (NHL) is an extremely heterogeneous group of malignancies that display distinct behavioral, prognostic, and epidemiological characteristics. There have been marked increases in the incidence of NHL in many parts of the world, and in 2008 it accounted for 2.8% of all cancers (355,900 cases and 191,400 deaths) [13].

Different lymphoma studies suggested different housekeeping genes as PCR endogenous controls, including GAPDH (glyceraldehyde 3-phosphate dehydrogenase), B2M (β2-microglobulin), ACTB (β-actin), ribosomal genes, PRKG1 (protein kinase cGMP-dependent, type I), TBP (TATA box binding protein), and others [14-17]. Traditionally,
the stability estimation was limited to standard deviation assessment for each gene in individual data sets. So far, very few assays have been performed with the use of several genes as internal controls.

The diversity of biopsy specimens is characteristic of NHL studies. The most common types of material are peripheral blood mononuclear cells (PBMC) and lymph node and spleen biopsies. Combined with the diversity of nosological forms that comprise the NHL group, the wide variety of biopsy specimens makes correct normalization for quantitative PCR experiments a complicated task. Normalization to total RNA levels in transcriptionally different tissues (such as lymph node and peripheral blood) is not an option [7]. Internal control genes therefore have to be carefully chosen. Several approaches to reference gene choice [18-20] have evolved over the last decade and helped to improve data analysis in many experimental cases [21-25]. Despite apparent differences, all of these algorithms aim to produce a numeric measure of gene expression stability by directly or indirectly comparing their expression in the sample set. Each algorithm thus produces a ranking of genes in which the best-ranked genes are the most stable ones and the cut-offs for sufficient gene expression stability are later determined experimentally. The aim of our study was to test and compare these approaches on a large, biologically diverse set of lymphoid samples to identify several of the most stable genes and to determine the number of genes required for sufficient normalization. We examined seven putative candidate genes that have previously shown high stability results [19] and demonstrated that the three best genes (*YWHAZ* (tyrosine 3-monooxygenase tryptophan 5-monooxygenase activation protein, zeta polypeptide), *UBC* (ubiquitin C), and *ACTB* (beta-actin)) are sufficient, while the addition of more reference genes degrades the normalization factor.

### Materials and Methods

**Patients.** The tissue material included samples of spleens, lymph nodes, and peripheral blood mononuclear cells (48 spleen biopsies, 23 lymph node biopsies, and 94 PBMC) from 165 patients undergoing diagnostic procedures in the “GeneTechnology” Diagnostic Centre (Moscow, Russia). The diagnoses were established by flow cytometry and immunohistochemistry. None of the patients received treatment prior to material withdrawal. Of the 165 patients, 33 were diagnosed with mantle cell lymphoma, and 97 had B-CLL. The translocation data t(11;14)(q13;q32) were available for 12 of 33 MCL patients. The remaining cases consisted of 17 CD5-negative B-cell lymphomas (2 follicular lymphomas, 4 large β-cell lymphomas, 11 marginal zone lymphomas) and 18 reactive cases (non-malignant proliferation). Analysis was performed according to specific research ethics protocols and local guidelines for secondary use of anonymized waste material.

**Cell isolation.** Blood mononuclear cells were isolated by Ficoll-Histopaque (Sigma-Aldrich) gradient centrifugation according to standard protocol. Isolated cells were washed twice in RPMI-1640 medium and immediately frozen in liquid nitrogen. Single-cell suspensions were prepared from lymph node and spleen specimens (approximately 20 mg of tissue) in Medi-machine (BD Biosciences) and were filtered through a 40-micron filter. Suspensions were also frozen in liquid nitrogen. All samples were stored at −70°C for a week before RNA extraction.

**RNA extraction and reverse transcription.** RNA was extracted from frozen cell samples using the RNeasy Mini Kit (Qiagen) according to the manufacturer’s protocol. The sample was then incubated with DNase

### Table 1. Primer sequences for genes under study.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>YWHAZ</em></td>
<td>ACTTTTGGTACATTGTGGCTTCAA</td>
<td>CCGCCAGGACAAAACCAGTAT</td>
</tr>
<tr>
<td><em>UBC</em></td>
<td>ATTTGGGTCGCACTTCTTGG</td>
<td>TGCCATTGACATTCTGATGTT</td>
</tr>
<tr>
<td><em>HPRT1</em></td>
<td>TGAACCTGGCAGAATAAGC</td>
<td>GGTCTTTTCACGAGCAGAGCT</td>
</tr>
<tr>
<td><em>GAPDH</em></td>
<td>TGCAACCCAACCTGGCTTACG</td>
<td>GCGATGACTGTGGTCTGATGAG</td>
</tr>
<tr>
<td><em>ACTB</em></td>
<td>CTGGAACGGTGAAAGTGACA</td>
<td>AAGGAGCTTCTGTAACATGCC</td>
</tr>
<tr>
<td><em>B2M</em></td>
<td>TGCTGTCTCCATGTTGTATCT</td>
<td>TCTCTGCTCCCCACCTCTGA</td>
</tr>
<tr>
<td><em>RPL13A</em></td>
<td>CCTGGAGGAGAAGAGAAGAGA</td>
<td>TTGAGGACCTCTGTGTATTTGTCAA</td>
</tr>
</tbody>
</table>
Superior reference genes for RTqPCR in B-cell lymphomas

The primer sequences were used as proposed by Vandesompele et al with minor modifications (Table 1) [19]. To our knowledge all genes have independent cellular functions and were assumed not to be co-regulated. The amplicon sizes ranged between 86 and 140 bp for all analyzed genes (Table 2). The products were verified by electrophoresis in TBP 1.5% agarose gel with ethidium bromide. All primers were synthesized and HPLC-purified by Syntol Ltd. (Russia).

A cDNA sample was used in x1, x10, x100, and x1000 dilutions to create a linear regression plot for each gene. The PCR efficiency for each individual gene was calculated as $10^{1/slope}$.

Data analysis and statistics. The Ct values were determined for real-time PCR curves by setting the threshold at 5 SD for each run. Relative cDNA quantity was calculated as $Q = E^{C(t)\text{min}-C(t)n}$, where $E$–PCR efficiency, $C(t)n$–averaged triplicate $C(t)$ value for each patient’s sample, $C(t)\text{min}$–minimal average $C(t)$ value for the gene in the experimental set.

Results were analyzed using free VBA applets: geNorm (Center of medical Genetics, Ghent University hospital, geNorm version 3.5, 2002), NormFinder (Molecular Diagnostic Laboratory, Department of clinical Biochemistry, Aarhus University Hospital, Denmark, 2004), and BestKeeper (FML-Weihenstephan, Centre of Life and Food Science, Technical University of Munich, Germany, 2003). While geNorm and NormFinder used relative quantities, BestKeeper required cycle raw data.

Statistical analysis was performed using GraphPad® Prism (GraphPad Software, version 5, San Diego, CA, USA). The normality test was performed using Kolmogorov-Smirnov and D’Augostino-Pearson criteria.
Results

Initial screening of transcription profiles of the candidate reference genes showed that all were expressed in the tissues studied. The PCR products for each gene were tested by agarose electrophoresis – a single band of expected product size was always observed. In real-time PCR experiments gene-specific amplification was confirmed by a single peak in the melting curves. For each sample, the standard deviation from the mean Ct in triplicate was always less than 0.25. The Ct ranges, means, SD, and SEM for each gene are presented in Table 3.

We also tested our genes for normal distribution of their expression levels in our sample set. The normality test was performed using Kolmogorov-Smirnov and D’Augostino-Pearson criteria. Two genes, YWHAZ and UBC, fit both of the normality criteria (p>0.1, α=0.05), while other genes (HPRT1, GAPDH, ACTB, RPL13A, B2M) fit neither. Figure 2 shows the distribution of each candidate gene in the sample set.

To determine the expression stability of the selected genes we used geNorm, NormFinder, and BestKeeper algorithms. geNorm is based on stepwise elimination of bad candidates according to their stability measure. The best genes are included in the normalization factor by geometric averaging. We performed geNorm analysis on the set of all samples. Next we divided our data set into three groups by tissue type (spleen, peripheral blood, and lymph node) and into four groups by nosologies (CLL, MCL, CD5-negative lymphomas, and reactive lymphoid tissue) and repeated the analysis for them separately. We ranked the selected genes according to their gene stability measure M, calculated as the average pairwise variation of a particular gene with all other control genes; the genes with higher M had greater variations of expression. The stability ranking was unaffected by tissue type (data not shown). Genes were ranked as follows, starting from the most stable pair (Table 4): YWHAZ and UBC, ACTB, HPRT1, GAPDH, RPL13A, B2M. In our total sample set and in nosology groups “CLL” and “reactive lymphoid tissue”, the ranking order of reference genes was also similar. In “MCL”, ACTB and HPRT1 were interchanged, and in “CD5-negative lymphomas”, the ranking was as follows: UBC and ACTB, YWHAZ, GAPDH, HPRT1, RPL13A, B2M. YWHAZ, UBC and ACTB still had the lowest M values, while RPL13A and B2M had the highest M values.

BestKeeper algorithm is based on multiple Pearson correlation analyses and standard deviation assessment for every individual gene. The ranking it produced was similar to the one of geNorm with only two genes (YWHAZ and UBC) fitting below the proposed threshold of SD<1.

We also used the NormFinder application to evaluate our candidate reference genes [18].
Unlike geNorm, NormFinder evaluates every gene independently. The stability rankings produced by the two programs are presented in Table 4. We first tested our genes on the whole sample set. We then divided the set into groups by nosology and tissue type to find the best pair of reference genes based on minimal inter- and intra-group variation. NormFinder ranked YWHAZ as the best solitary gene, with a stability value of 0.045. Less predictably, the best pair of genes as suggested by NormFinder was YWHAZ and B2M, with a stability value of 0.032.

To estimate the number of genes sufficient for normalization, we analyzed pairwise variations (V) defined as SD of log2-transformed expression ratios of any pair of reference genes as proposed in geNorm and performed NF linearization analysis. The lowest V values were V5/6 for “MCL” and “CLL” subgroups and V4/5 for “CD5-negative lymphomas” and “reactive lymphoid tissue” subgroups and the overall sample set (Figure 3). The NF linearization plots (Figure 4) show better linearization for NF3 vs NF4 ($R^2=0.81$), which suggests the addition of a fourth gene is unnecessary for NF.

**Discussion**

As no gene in the real world fits all the criteria for an ideal reference [18,26,27], researchers started to regard the choice of the “right” normalizer as an experiment-dependent matter. Initially, direct analysis of $C_t$ standard deviation for every single gene under study was used as an expression stability measure [28-30]. The majority of works on gene expression in B-cell lymphomas use a single gene as the normalizer [15-17]. Several attempts to choose the best reference gene for lymphoid tissue have been made, but the results have been controversial.
For instance, Lossos et al stated that PRKG1 is quite suitable for both T- and B-cell tumor studies [14], while Green et al named this gene among poor normalizers and recommend GUSB as the best reference gene for snap-frozen and formalin-fixed lymphoid tissue [31].

The best single gene SD-based approach assumes that gene expression levels are normally distributed throughout the sample set [32]. However, the normality of distribution was rarely directly analyzed in experimental studies. Over the last decade, a number of algorithms that do not require reference genes to be normally distributed have been introduced along with the idea that several “best” reference genes should be included in gene expression studies for data normalization [10,11,33]. In our experience, using only one of our best-ranked reference genes (YWHAZ or UBC) allowed us to obtain statistically significant differences in cyclin D1 mRNA expression for mantle cell lymphoma, CD5-negative lymphomas, and reactive lymphoid tissue, but not for B-CLL [34].

Only recently have several genes have been compared by geNorm for lymphoid specimens [31,35]. Green et al were the first to apply a specific expression stability estimation method for lymphoid cells, but their specimens included only snap-frozen and formalin-fixed tissues, and the whole cohort (21 pairs) was rather small [31]. Differences in gene expression stability in fresh and formalin-fixed tissue may result from specific patterns of RNA degradation [36,37]. Another study focusing on applicability of reference genes included only highly purified B-CLL cells from peripheral blood and showed that the best result is obtained using four reference genes [35]. However, the authors concluded that 3 genes (B2M, HPRT1, and GUSB) may also be sufficient [35]. As we did not apply sorting methods to our sample set, closely following the methodology of routine clinical investigations, the ranking of reference genes for our data set was different and pairwise variation achieved was higher for all groups.

The approaches to stability estimation and choice of genes suggested by different authors vary [18-20]. We therefore used three free VBA applets (geNorm, BestKeeper and NormFinder) to investigate the expression stability of our candidate reference genes (Table 4). Both geNorm and BestKeeper rank YWHAZ and UBC as the best-choice reference genes. NormFinder ranks YWHAZ as the most stable and includes it in the best reference pair along with B2M, which was considered a poor normalizer by geNorm and BestKeeper. Such a difference in gene selection is probably due to the differences in the algorithm used, as BestKeeper and geNorm allow “bad” gene exclusion while NormFinder uses a model-based approach and does not exclude unstably-expressed genes from further analysis.

Of the three VBA applets, one – geNorm – focuses on estimating the adequate number of reference genes by introducing their pairwise variation (V) as its measure. BestKeeper only assesses genes separately, and NormFinder provides the stability values for single genes and for the best pair, not addressing their sufficiency. Many authors have contemplated the use of more than two genes for accurate normalization [10,38,39]. Our determination of a sufficient number of reference genes included V analysis and the NF linearity test, as the proposed cut-off values for M and V were never reached in our sample set. In accordance with the results of Valcekeine et al [35], the lowest pairwise variation obtained in our total dataset was V4/5 (0.24), slightly lower than V3/4 (0.28) (see Figure 3). However, since both these values are above the proposed cut-off of
0.15, we speculated that such low pairwise variation is unachievable on a highly heterogeneous clinical sample set and used NF linearization analysis to determine the new cut-off value for V. Data linearization shows whether the addition of another gene to the normalization factor significantly impacts it. The highest linear regression coefficient ($R^2=0.81$) of NF plots was obtained on NF$_3$ data plotted against NF$_4$ (Figure 4), indicating that three genes are sufficient for the normalization factor and setting the V cut-off to 0.28. We suggest this value can be used for the comparison of a heterogeneous group of specimens. In our previous dataset, the third-best gene ranked by geNorm was HPRT1, which was very close to ACTB in M values [34]. Therefore, it is not surprising that these two genes switched rankings as the dataset enlarged. Normalizing with a combination of YWHAZ, UBC, and HPRT1 and a combination of YWHAZ, UBC, and ACTB yielded similar results in our current dataset. This shows that while a third gene is necessary for accurate normalization, there may be several candidates that improve accuracy by an equal amount.

It is noteworthy that YWHAZ and UBC were the only genes out of seven candidate genes that passed the normality tests. Their inclusion proves to be the most important for accurate normalization, i.e. for decreasing M value, whereas adding genes that deviate far from normal distribution may bias the results [34]. Since accurate normalization increases the method’s resolution, in the previous study we were able to resolve groups with only a 10-fold difference in gene expression with statistical significance [34] whereas other authors could only discriminate among sample groups with expression differences of 2 and more orders of magnitude [15,40]. Having expanded the set of candidates for reference genes, we conclude that a normalization-based approach using three reference genes (YWHAZ, UBC, and ACTB) allows the robust reduction of variance in real-time PCR results on a large and heterogeneous set of clinical lymphoid samples.

Acknowledgements

The work was supported by RFBR grant #13-04-40189 to I.A.V. We are grateful to “GeneTechnology” Diagnostic Centre for providing human tissue samples (Moscow, Russia), and to Anna V. Tvorogova for providing technical assistance in preparation of this manuscript.

References


