Review

Critical appraisal of quantitative PCR results in colorectal cancer research: Can we rely on published qPCR results?

J.R. Dijkstra\textsuperscript{a,}\*, L.C. van Kempen\textsuperscript{b}, I.D. Nagtegaal\textsuperscript{a}, S.A. Bustin\textsuperscript{c}

\textsuperscript{a}Department of Pathology, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands
\textsuperscript{b}Lady Davis Institute for Medical Research, Jewish General Hospital, LDI-E447, 3755 Cote Ste-Catherine, Montreal, QC, Canada H3T 1E2
\textsuperscript{c}Postgraduate Medical Institute, Faculty of Health, Social Care and Education, Anglia Ruskin University, Bishops Hall Lane, Chelmsford, Essex CM1 1SQ, United Kingdom

\textbf{ARTICLE INFO}

Article history:
Received 13 October 2013
Received in revised form
15 December 2013
Accepted 26 December 2013
Available online 2 January 2014

Keywords:
qPCR
Real time PCR
Colorectal cancer
MIQE

\textbf{ABSTRACT}

The use of real-time quantitative polymerase chain reaction (qPCR) in cancer research has become ubiquitous. The relative simplicity of qPCR experiments, which deliver fast and cost-effective results, means that each year an increasing number of papers utilizing this technique are being published. But how reliable are the published results? Since the validity of gene expression data is greatly dependent on appropriate normalisation to compensate for sample-to-sample and run-to-run variation, we have evaluated the adequacy of normalisation procedures in qPCR-based experiments. Consequently, we assessed all colorectal cancer publications that made use of qPCR from 2006 until August 2013 for the number of reference genes used and whether they had been validated. Using even these minimal evaluation criteria, the validity of only three percent (6/179) of the publications can be adequately assessed. We describe common errors, and conclude that the current state of reporting on qPCR in colorectal cancer research is disquieting. Extrapolated to the study of cancer in general, it is clear that the majority of studies using qPCR cannot be reliably assessed and that at best, the results of these studies may or may not be valid and at worst, pervasive incorrect normalisation is resulting in the wholesale publication of incorrect conclusions. This survey demonstrates that the existence of guidelines, such as MIQE, is necessary but not sufficient to address this problem and suggests that the scientific community should examine its responsibility and be aware of the implications of these findings for current and future research.

\copyright\ 2014 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

1. Introduction

Since its introduction in 1992 (Higuchi et al., 1992), the real-time quantitative polymerase chain reaction (qPCR) has become a commonly used technique for detecting and measuring RNA expression in cancer research. Despite the introduction of next-generation sequencing (NGS), qPCR remains an essential technique for confirmation of NGS findings. The perceived
simplicity of qPCR experiments, which can deliver fast and cost-effective results, has resulted in an increasing number of publications that utilize this technique. But how reliable are the published results? PCR is comprised of several critical parameters that must be evaluated carefully and optimized to obtain meaningful and reproducible results (Deroux et al., 2010; Tichopad et al., 2009). In 2009 the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines were introduced to facilitate critical assessment of those parameters (Bustin et al., 2009). The need for these guidelines is emphasized by a recently described example of conflicting results in publications reporting inadequately transparent experimental detail (Bustin et al., 2013). Endogenous control genes (or reference genes) are one of the crucial parameters incorporated in the MIQE guidelines and their use is currently the most accurate method for correcting variability associated with template input and RT efficiency (Bustin et al., 2009). An important prerequisite for a reference gene is that its expression should remain as stable as possible. Furthermore, gene expression is not only highly tissue specific but also dependent on the experimental setting (Caradec et al., 2010; Radonic et al., 2004), which suggests that it is highly unlikely that universal reference genes exist (Vandesompele et al., 2002). Indeed, most well-known reference genes such as GAPDH or β-actin are not stably expressed (Bustin, 2000; Caradec et al., 2010; Greer et al., 2010; Suzuki et al., 2000; Thellin et al., 1999; Warrington et al., 2000). Nevertheless, there are numerous studies that continue to use these, and other genes, without proper validation. Another consideration when choosing a reference gene is dependent on the method of analysis. The ΔΔCq method is currently the most commonly used method for studies that report changes in the expression of genes of interest relative to a reference gene (Dijkstra et al., 2012; Schmittgen and Livak, 2008). In common with all other relative quantification methods, this approach assumes two basic conditions: (i) the efficiency of individual assays must be consistent from one run to another and (ii) the effect of any variations on Cq value must be equivalent for reference genes and genes of interest (Schmittgen and Livak, 2008). It is therefore essential that the efficiencies of the assays for all targets are known and comparable and, should this not be the case that a correction factor is applied. These assumptions do not apply to the reporting of quantitative relative to target-specific standard curves; nevertheless an inefficient assay is likely to be nonrobust, perform poorly and hence will result in increased variability of the results (Tichopad et al., 2004, 2003).

In the current paper we provide an assessment of the reliability of published qPCR studies in all colorectal cancer-associated publications that used qPCR between 2006 and August 2013. The evaluation is based on an analysis of reference genes, since their rigorous selection and validation implies both an understanding of the technique and a willingness of the researcher to perform a well-designed experiment.

2. Method

A PubMed search was performed to retrieve all English-language publications using the keywords “colorectal cancer” and “real-time PCR”/“qPCR” in the period between 01/01/2006 and 01/08/2013. Only full text publications, in which qPCR was used to assess the diagnostic, prognostic or predictive value of mRNA or miRNA expression, were considered for this survey. All publications were screened for: 1. Journal name, 2. Year of publication, 3. Impact Factor (obtained from Web of Science: http://isiknowledge.com/wos), 4. Type of original sample (e.g. FFPE, frozen, blood), 5. qPCR starting material (RNA/ miRNA), 6. Name and accession number of reference gene(s), 7. Citation of PCR efficiencies (claimed comparable efficiency and performed standard curves were accepted), 8. Use of efficiency-dependent method of analysis (e.g. $\Delta\Delta$Cq, $\Delta$Cq, Ratio), 9. Total number of reference genes, 10. Validity testing of reference gene(s), 11. Citation of MIQE, 12. Availability of online supplemental qPCR related data. Screening was performed by the four authors and to assure concordance in screening each author screened ten publications initially screened by another author.

Results were classified as follows:

- If they abided by the individual MIQE criteria being assessed they were scored as a “1” if they did not as a “0”.
- The number of reference genes used to normalize the data were recorded and scored “0” if no or only a single reference gene was used. If multiple reference genes were used they were scored as “1”.
- A lack of clarity or complete absence of information in some publications made it impossible to score individual parameters. Hence a third category, “other”, was included which comprises results that could not be scored in category ’0’ or ’1’.

Comparisons of the overall results in relation to the year of publication or impact factor are depicted by changes in compliance to the MIQE criteria.

3. Results

The initial PubMed based search to retrieve all publications on “colorectal cancer” using “real-time PCR”/”qPCR” in the period between 01/01/2006 and 01/08/2013 identified 378 publications. In total, 199 publications were excluded from the analysis because qPCR was not used to assess the diagnostic, prognostic or predictive value of the expression ($n = 87$), or the methods were DNA based ($n = 87$). Eighteen papers were excluded because they did not fulfill the eligibility criteria (i.e., cell line experiments, the aim of the publication being finding general reference genes, knock-down checks, etc) and we were unable to obtain the full text from seven papers, even after contacting the corresponding author. This resulted in 179 publications meeting our criteria and therefore being suitable for reviewing (Figure 1).

The individual results for every publication are provided in Supplemental data 1. There were fewer miRNA-based qPCR experiments than those targeting mRNA based (resp. 28 and 151, Figure 1). Since all data on both miRNA and RNA were comparable (data not shown), except for the used reference genes (Table 1), the groups were analysed together. In general (Figure 2), almost all studies used a similar strategy in performing the experiment and reporting the experimental
setting. Commercial assays were used in 32% of the published experiments. In 92% of the publications, only a single reference gene was used. Validation of the reference genes and online supplemental data were not reported in 87% of the publications. Crucially, 91% used an efficiency-dependent method of analysis despite the fact that only 18% reported the PCR efficiencies. Finally, the MIQE guidelines were cited in only 1% of publications.

Figure 3 displays the results in relation to the year of publication. Besides a normal fluctuation over the years there is no clear trend visible for any parameter, except possibly for the use of the efficiency-dependent method of analysis. When the data are stratified into two groups using either 2009 (the implementation of MIQE) or 2012 (when MIQE might be expected to have started to have an impact) as a threshold value it becomes clearer that none of the critical parameters show any improvement.

The results in relation to the impact factor of the journal are depicted in Figure 4, separated into two categories, those with impact factors of <5 and ≥5. There were no differences in any of the reported parameters, except for the more extensive deposition of supplemental data in higher impact factor journals.

Based on the assumption that users are expecting the efficiencies of commercial assays to be tested by the producing company, it might be expected that researchers would be more inclined to assess the efficiency of the PCR for non-commercial primers. However, the PCR efficiency is only tested for 21% of the non-commercial assays, which is just 5% higher than for the commercial assays.

4. Discussion

Although qPCR is often described as a “gold standard” for gene expression studies and has been used in thousands of published papers, there are serious questions over its reliability, reproducibility and the validity of conclusions based on this technique (Bustin, 2010). The increasing focus on molecular biomarkers has led to the publication of numerous papers using qPCR to identify and attempt to validate a wide range of mRNA and miRNA markers for diagnostic, prognostic or predictive purposes in human cancers. But how sound are these studies? Here we provide an assessment of the reliability of published qPCR studies in relevant colorectal cancer-associated publications that used qPCR between 2006 and August 2013. The disturbing finding of this survey is that the vast majority of publications provide insufficient information to allow an assessment of the reliability of these qPCR experiments. Considering that the survey is limited to reference genes, the results are likely to be an underestimation of the reporting quality of qPCR experiments.

A striking and commonly found limitation of qPCR experiments is the number of reference genes that are used to normalize qPCR data. It has become well-established that the use of more than one reference gene increases the accuracy of the measurement compared to the use of a single reference gene, especially when the aim is to show relatively small fold-changes in RNA levels (Meyer et al., 2010; Vandesompele et al., 2002). Furthermore, for large-scale studies it has been established that the use of the mean expression value is the best normalization strategy (Mestdagh et al., 2009). However, 92% of the assessed publications use only a single reference gene, and only 6% use more than two. Furthermore, only 23 (13%) out of the 179 publications report whether the chosen reference gene was validated. Based on our finding that 70% of the publications on RNA based qPCR use just one of three reference genes (ACTB/GAPDH/18S, Table 1), it appears to be highly unlikely that these have been validated. Several publications base the choice of the reference gene either on other (non-comparable) studies (Mazzoccoli et al., 2011) or on results from publications.

When the data are stratified into two groups using either 2009 (the implementation of MIQE) or 2012 (when MIQE might be expected to have started to have an impact) as a threshold value it becomes clear that none of the critical parameters show any improvement.

The results in relation to the impact factor of the journal are depicted in Figure 4, separated into two categories, those with impact factors of <5 and ≥5. There were no differences in any of the reported parameters, except for the more extensive deposition of supplemental data in higher impact factor journals.

Based on the assumption that users are expecting the efficiencies of commercial assays to be tested by the producing company, it might be expected that researchers would be more inclined to assess the efficiency of the PCR for non-commercial primers. However, the PCR efficiency is only tested for 21% of the non-commercial assays, which is just 5% higher than for the commercial assays.
identifying general reference genes (Andersen et al., 2011; Herath et al., 2009; Mazzoccoli et al., 2011; Rubie et al., 2011, 2008). The use of either approach is ill-advised for several reasons, a. the valid use of a reference gene is shown to be tissue- and even experiment-specific, b. the efficiencies/validations reported may not be repeated in a second study, c. the use of the exact same reagents, instruments and protocols is not always achievable, and d. if the referenced publication is not published in an open access journal, claims may not be easily corroborated (Caradec et al., 2010; Radonic et al., 2004). A comprehensive disclosure of the method used does not guarantee a high quality experiment. One of the studies did describe the validation of a reference gene (Koga et al., 2008) and even though it was unstably expressed and therefore unsuitable as a reference gene, it was still used for this purpose. Furthermore, even though it has been established that the best way to normalize miRNA expression is to use other miRNAs (Vandesompele et al., 2002), 52% and 16% of all publications on miRNA expression use a small nuclear RNA or an mRNA, respectively, as a (single) reference gene.

This study also reveals that 146/179 publications (82%) do not mention PCR efficiencies of either target or the reference genes, yet an extraordinary 135 of these 146 publications (92%) use a method of analysis that is meaningless unless PCR efficiencies are known. An example of such a method is the ΔΔCq-method, and one of the prerequisites for using this method is comparable assay efficiencies for the reference gene and gene of interest, all other instances require the application of a correction factor (Schmittgen and Livak, 2008). It is therefore essential to provide details of PCR efficiencies when this approach is used. Failure to report PCR efficiencies does not automatically mean these experiments are unreliable, but does make it impossible to assess the validity of the data.

Consequently, the absence of information on assay efficiencies alone renders 75% (135/179) of publications

![Figure 2](image1.png) This stacked bar chart shows the distribution of the results per evaluation criteria. Each bar represents the accordance (1) or non-compliance (0) with the specific criterion. All deviant cases (i.e. not specified, unclear) are represented in the category "other".

![Figure 3](image2.png) Overall results in relation to the year of publication. The depicted line represents the percentage of publications that meet or report information on the specific criterion.
questionable. If the other assessed parameters are included, 173/179 (97%) provide inadequate information. This leaves a mere six (3%) publications reporting sufficient experimental detail to allow a reliable assessment of the qPCR data. Of course, the fact that these six can be assessed for the validity of the reference genes used does not automatically mean they contain reliable results. The results of our analyses are similar to those of a recent study looking at normalisation in the context of the HepaRG cell line, which is widely used as an alternative for primary human hepatocytes (Ceelen et al., 2013). The authors of this study concluded that not one of the 24 reviewed studies used a proper normalization method. They also agree with another more general recent survey of qPCR-based publications that comes to the same conclusion (Bustin et al., 2013). Therefore, proper validation of reference genes might be lacking in the majority of all qPCR experiments in any given setting.

An example of conclusions based on inadequate reporting and possibly inappropriate experimental detail is the report on the expression of CD133 in colorectal cancer patients (Artells et al., 2010). It compares CD133 mRNA levels of normal and cancer samples using 18S rRNA as the single reference gene using a commercial assay, concluding that CD133 mRNA expression is significantly higher in tumour compared with normal tissue. There are a number of issues with these data: first, despite using the deltaDeltaCq normalization method, there is no mention of PCR efficiency of either the gene of interest or reference gene. Second, an analysis of the data shows that the maximum relative expression difference between paired tumour and normal tissue is only 2.5-fold, with 16/53 tissue pairs analysed having lower CD133 mRNA levels. This kind of marginal fold-change requires careful application of multiple validated reference genes and the use of a single, unvalidated one is, at the very least, inadvisable. Third, 18S rRNA genes have been described as a poor choice for use as a reference gene for colorectal cancer, since not only do colorectal cancers contain more ribosomes and rRNA than normal tissue (Tsuji et al., 2002), but the regulation of rRNA synthesis is independent from synthesis of mRNA (Radonic et al., 2004), resulting in an expression pattern that differs from that of mRNA (de Kok et al., 2005). Fourth, it has also been shown to be one of the most variable reference genes in colorectal cancer (Sorby et al., 2010), although ironically an earlier report contradicts this (Tsuji et al., 2002).

Strikingly, there was no correlation between the source of an assay and the reporting of assay efficiencies. One might have assumed that the supplier has already tested the efficiencies of commercial assays; hence there would be less of a need to test their efficiencies compared with non-commercial self-designed assays, which obviously need to be tested. Our results show that this is not so and again poses the question whether researchers are failing to carry out basic quality control analyses because they do not understand the need to do so or because they just cannot be bothered.

The results of the current survey are worrying. If these data are extrapolated to the study of cancer in general, one is forced to conclude that almost all studies that use qPCR cannot be reliably assessed and the results of these studies might or might not be valid. Enormous amounts of money and effort have been put into this kind of research over the years while the practical implications are that new research studies but also extensive research and development efforts by pharmaceutical and/or biotechnology companies are potentially based on inaccurate data. This conclusion agrees with a recent observation that literature data on potential drug targets should be viewed with caution, since most experiments published in the peer-reviewed literature are not reproducible (Prinz et al., 2011). In an area as important as cancer diagnostics there is a necessity to improve and the scientific community should take its responsibility more seriously. The existence of guidelines, such as MIQE, is not sufficient; editors and reviewers should realize its importance for current and future research. Methodological screening of papers should be standard, especially in the current era of seemingly limitless technical possibilities.

Appendix A.
Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.molonc.2013.12.016.

References


