Characterization of the *3': 5'* ratio for reliable determination of RNA quality

Jose V. Die[1*], Ángeles Obrero[2], Clara I. González-Verdejo[2], Belén Román[2]

[1] Mejora Genética Vegetal, Instituto de Agricultura Sostenible, (IAS-CSIC), 14080 Córdoba, Spain
[2] Biotecnología y Mejora, IFAPA 'Alameda del Obispo', 14080 Córdoba, Spain

* Corresponding author.

* E-mail address : jdieramon@ias.csic.es

* Tel.: +343574992241

* Fax: +34957499252

Abstract

Determination of RNA quality is a critical first step in obtaining meaningful gene expression data. The PCR-based 3': 5' assay is an RNA quality assessment tool. This assay is a simple, fast and low-cost method of selecting samples for further analysis. However, the practical application is unexploited primarily because of the absence of a experimental threshold. We show that, by anchoring the 5'assay a specific distance from the 3' end of the sequence and by spacing the 3' at a distance of a number of nucleotides, a cutoff determines whether a sample is suitable for downstream quantification studies.

Keywords : 3':5' ratio; qPCR; RNA quality

Nowadays the reverse transcription quantitative PCR (RT-qPCR) is the most rapid, sensitive, accurate and precise method available for gene expression analysis [1] The technique requires a combination of various steps with several conditions that have a direct impact on the conclusions. Therefore, minimizing the method's variability and maximizing its reproducibility by quality-assessing every component of the qPCR workflow and adhering to common guidelines for the analysis and accurate interpretation of the data is essential [2-4]. Prior studies have reported that the quality of the template may be the most important determinant of the reproducibility and biological relevance of qPCR results [5]. This notion is supported by the well-known impact that the use of degraded RNA has on the interpretation of data [6-7]. Despite these obvious implications, RNA quality check is deficient or absent from a substantial number of published papers [8]. Therefore this is an issue that requires more attention than it has received to date.

Methods for assessing the quality of total RNA have evolved from highly sample-consuming, low-throughput, qualitative gel-based techniques to an easily automated, high-throughput, quantitative technology. Questions have recently been raised regarding the ability of traditional methods based on the 28S/18S ratio or optical density measurements to measure RNA integrity [9-10]. Currently, lab-on-chip technology is the most efficient system for ensuring RNA quality. This technique offers accuracy, reproducibility and high sensitivity [11]. Although its applications is expected to increase in the future, a major drawback of this technology is that mainly assess the ribosomal RNA profile, providing total RNA quality, but is not focused on the quality of the messenger RNA molecules. Moreover, cost of automated electrophoresis stations, renders the method unfeasible for laboratories with resource constraints.

In the absence of an alternative, the 3':5' assay has been proposed in order to assess mRNA integrity [4]. It is based on the evidence that cDNA yield from sequences near the 5' end of partially degraded mRNAs is significantly lower than those from sequences near the 3' end provided that reverse transcription proceeds from the polyA-tail [12], i.e. poor RNA quality adversely affects the synthesis of first-strand cDNA resulting in the under representation of the 5' moiety of the transcript. The assay has been adopted by microarray users for many years [13], although its practical utility remains to be determined as several different cutoff criteria for the 3':5' ratio have been proposed. Furthermore, the term *3':5' assay* is not used unambiguously by researchers. These inconsistencies have hampered the systematic application as quality assessment tool.

Since the availability of a practical and reliable cutoff prior to RT and qPCR assays has the potential to save substantial costs in wasted reagents and technical time, we investigated which 3':5' ratio can determine the suitability of RNA samples for downstream applications in qPCR-based studies. In this work, we considered this value to depend on the nature of the samples (i.e. fresh frozen), the target's abundance, location between two primer sets and distance from the 3' end. Our hypothesis was that employing specific target design enables the identification of samples with sufficient RNA quality to be suitable for downstream applications.

We designed seven sets of specific primer pairs at various positions on the ubiquitin cDNA sequence (UBQ; DFCI Medicago Gene Index, TC112803) to amplify products of 60 to 110 bp, with an optimal primer melting temperature (Tm) of 60ºC and GC contents between 35% and 65%. To maximize PCR amplification efficiencies, control

for secondary structure of single-stranded DNA and/or at the sites of primer binding was conducted using MFOLD v3.2 software [14]. Using Trizol reagent (Invitrogen, CA, USA), total RNA was extracted from roots and leaves (n=3) obtained from pools of five plants of the model legume *Medicago truncatula* to minimize sample-to-sample variation. RNA was quantified using a Nanodrop Spectrophotometer ND-1000 (Thermo Scientific, MA, USA). Prior to cDNA synthesis, DNase treatment was performed using Turbo DNAseI (Ambion, TX, USA) according to manufacturer's protocol. The absence of contaminating genomic DNA was verified by qPCR. RNA was artificially degraded by heat exposure. Subsequently, samples were subjected to microfluidic electrophoresis analyses performed on the Experion system (Bio-Rad, CA, USA). Each sample (n=9) was classified as (i) high quality (RQI≥9; 0 min at 70ºC), (ii) optimal quality (RQI:7-9; 10 min at 70ºC) or, (iii) partly degraded (RQI:6-7; 30 min at 70ºC). Total RNA electropherograms were in accordance with the degradation exposure indicating a progressive reduction in size of the 18S and 25S peaks and an elevation of the base-line, resulting in a decrease of the RQI, as depicted in Fig. 1. RT was performed on 2 µg total RNA using oligo(dT_{15}) and Superscript III (Invitrogen) following the manufacturer's instructions. Gene expression analysis was performed on 1.5 µl of a 1:4 dilution of cDNA, 7.5 µl of 2xSYBR Green master mix (Promega, WI, USA), and 200 nM of each primer pair in a total volume of 15 µl. PCRs were performed on a Mx3000P Real-time PCR System (Stratagene, CA, USA). Reactions were performed in duplicate on each of the 9 samples across the seven targets. Variations in Cq among technical replicates were assessed (max. variation <0.30) and the mean Cq of replicates was employed in subsequent calculations. The experimental data, primer sequences and target assays are outlined in the Supplementary Content. Formation of specific predicted PCR products was confirmed by melting curve analysis. Relative quantity (RQ) of template was

calculated using the *comparative* C_T *method* [15]. Next, expression data were normalized using a normalization factor consisting of the geometric mean of the expression levels determined by the two assays located closer to the 3' end. This provides the normalized RQ (NRQ) of the specific assay for each biological replicate (Fig. 1D-F). A log transformation of NRQ data was subsequently applied and analysis of variance was employed to compare quality groups using the transformed NRQ. Results exhibiting $P<0.05$ were considered significant, and the least significant difference (LSD) was calculated at 5% level to compare expression levels obtained from the various groups.

When designing a primer pair for use in qPCR, screening the 3' region of a cDNA may be a useful approach, as it reduces the risk of comparing samples with differing degrees of RNA degradation, which predominantly starts at the 5' end. Accordingly, higher RNA quality is associated with an elevated signal from the 5' primer set. Conversely, lower RNA quality is associated with downregulated levels of 5' transcripts and relatively upregulated levels of 3' transcripts. This gene-specific primer design strategy, however, is not possible in many cases in which limited genetic information is available. We did not observe statistically significant differences in expression levels between intact and degraded RNA samples from two target templates designed within the ~800nt sequence closest to the 3' end of the transcript. Nevertheless, significant differences above this threshold were evident between maximum- and minimum-quality groups (Supp. Content). To determine a practical cutoff 3':5' ratio value to identify suitable cDNAs for further analysis, we first defined the 3':5' ratio as the fold change difference in expression values between the two assays. Each amplicon may represent a 3' or 5' assay, as that designation is merely a spatial attribute. The 3' target assay is

designed to amplify an amplicon near the 3’ end, and the 5’ target assay targets the 5’ end of the mRNA sequence. Next, we decided to anchor a 5’ assay at 1526 nt from the end of the sequence, and investigated whether the 3’ targets expression levels were significantly different from the 5’assay, employing the mean expression level of various exons across the UBQ gene that were located 700, 900 and 1100 bp distance from the 5’ assay. Statistical analysis indicated that the 3’:5’ ratio of lowest-quality group differed when the 3’ assay was designed targeting sequences 700 bp ($P<0.01$), 900 bp ($P<0.001$) and 1070 bp downstream ($P<0.001$). Finally, on the basis of the observed 5’ mean expression level (setting at 1), we employed the mean 3’ relative expression value in optimal-quality samples and the 95.45% confidence interval (+95.45% CI, upper interval) as upper limit for scoring a suitable cDNA derived from intact RNA, and mean value of the partly degraded samples and the 95.45% confidence interval (-95.45% CI, lower interval) as lower limit for discarding a cDNA (Table 1).

Some conclusions from the aforementioned : (1) to characterize successfully the extent of RNA degradation in the starting material, 5’ amplicon of the 3’:5’ ratio assay has to be designed in a sufficient distance to the 3’-end of the sequence. Otherwise, measured gene expression from starting materials of equal quality, a critical condition for obtaining meaningful gene expression data, will not be ensured. This has a second meaning suggesting a ‘security zone’ close to the 3’ end that may be utilized, through careful primer design, to enable the quantification of mRNA levels. As mentioned above, this method’s relevance is limited by the availability of information in public databases regarding gene structures; (2) as determined from the fixed distance from the 5’ amplicon to the 3’ end sequence, there are various ratio thresholds, depending on the targets’ spatial separation, that can discriminate between various integrity levels of the

RNA samples. Results indicate that 3':5' ratio values to 2.73, provide full assurance that the input material is of high quality. Moreover, values over the limit of 4.43 for the same distance between amplicons, may characterize cDNAs as unreliable for gene expression studies. The inclusion of samples within this safety margin does not demonstrate that the material is unsuitable for further qPCR assays, but does not guarantee that the sample is in a maximum-quality criteria compliant way.

Note that the usefulness of the 3':5' ratio relies on the oligo(dT) priming method for cDNA synthesis because, with this method, the progress of the reverse transcriptase is dependent on the intactness of the mRNA. Moreover, experiments may be conducted in order to determine whether 3':5' ratios are RT enzyme-dependent or using more cell tissue types or targeting the 5' and 3'-end of other reference genes with different transcript abundance. Furthermore, it would be important to confirm these results in formalin fixed paraffin embedded tissues. However, data within the ratios shown in Table 1 can ensure that observed biological differences are not due to poor RNA integrity.

To summarize, the quality criteria proposed in this study should provide a useful guideline for future studies using qPCR to assess mRNA levels. The suggested cutoffs derived from specific distances between amplicons may help determine whether a certain cDNA is suitable for downstream quantification assays. Although we have conducted the experiment in plant tissues, should be applicable for the assessment of gene expression data obtained from other cell populations.

References

[1] H.D. VanGuilder, K.E. Vrana, W.M. Freeman, Twenty-five years of quantitative PCR for gene expression analysis, Biotech. 44 (2008) 619-626.

[2] S. Derveaux, J. Vandesompele, J. Hellemans, How to do successful gene expression analysis using real-time PCR, Methods 50 (2010) 227-230.

[3] S. Bustin, V. Benes, J.A. Garson, J. Hellemans, J. Huggett, M. Kubista et al., The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments, Clin. Chem. 55 (2009) 611-622.

[4] T. Nolan, R.E. Hands, S.A. Bustin, Quantification of mRNA using real-time RT-PCR, Nat. Protocols 1 (2006) 1559:1582.

[5] S.A. Bustin, T. Nolan, Analysis of mRNA expression by real-time PCR, in: J. Logan, K. Edwards, N. Saunders (Eds.) Real-time PCR: current technology and Applications (2009) ISBN : 978-1-904455-39-4, pp. 111-135.

[6] J.Vermeulen, K. De Preter, S. Lefever, J. Nuytens, S. Derveaux, J. Hellemans et al., Measurable impact of RNA quality on gene expression results from quantitative PCR, Nucl. Acids. Res. (2011) doi:10.1093/nar/gkr065

[7] S. Fleige, M.W. Pfaffl, RNA integrity and the effect on the real-time qRT-PCR performance, Mol. Aspects Med. 27 (2006) 126-139.

[8] J. Murphy, S.A. Bustin, Reliability of real-time reverse-transcription PCR in clinical diagnostics: gold standard or substandard?, Expert Rev. Mol. Diagn. 9 (2009) 187-197.

[9] C. Strand, J. Enell, I. Hedenfalk, M. Fernö, RNA quality in frozen breast cancer samples and the influence on gene expression analysis--a comparison of three evaluation methods using microcapillary electrophoresis traces, BMC Mol. Biol. 8 (2007) 38.

[10] O. Mueller, K. Hahnenberger, M. Dittmann, H. Yee, R. Dubrow, R. Nagle, D. Ilsley, A microfluidic system for high-speed reproducible DNA sizing and quantitation, Electrophoresis 21 (2000) 128–134.

[11] I. Riedmaier, M. Bergmaier, M.W. Pfaffl, Comparison of two Available Platforms for Determination of RNA Quality, Biotechnol. Biotechnol. Equip. 24 (2011) 2154-2159.

[12] G.H. Swift, M.J. Peyton, R.J. MacDonald, Assessment of RNA quality by semi-quantitative RT-PCR of multiple regions of a long ubiquitous mRNA, Biotechniques 28 (2000) 524–531.

[13] H. Auer, S. Lyianarachchi, D. Newsom, M.I. Klisovic, G. Marcucci, K. Kornacker, Chipping away at the chip bias: RNA degradation in microarray analysis. Nat. Genet. 35 (2003) 292-293.

[14] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31(2003) 3406–3415.

[15] T.D. Schmittgen, K.J. Livak, Analyzing real-time PCR data by the comparative CT method, Nat. Protocols 3 (2008) 1101-1108.

Internet resources

The Gene Index Project : http://compbio.dfci.harvard.edu/tgi/

Figure Legend

Fig. 1. Virtual gel image and electropherograms from three representative RNA leaves samples showing different degrees of degradation according to the RQI (A) 9.1, (B) 7.2, or (C) 6.0. (D-F) Expression profiles of different amplicons throughout the UBQ gene. Normalized values are rescaled to 543-assay. Mean expression level (+ SEM) from three samples is shown for each quality group. The X axis represents amplicon distance from the 3'end.