

Kinetics quality assessment for relative quantification by real-time PCR

Tzachi Bar and Anders Muszta
Chalmers University of Technology, Gothenburg, Sweden

BioTechniques 39:333-340 (September 2005)

For proper relative quantification by real-time PCR, compared samples should have similar PCR efficiencies. To test this prerequisite, we developed two quality tests: (i) adjustment of a test for kinetic outlier detection (KOD) to relative quantification; and (ii) comparison of the efficiency variance of test samples with the efficiency variance of samples with highly reproducible quantification. The tests were applied on relative quantification of two genes in 30 sets of 5 replicate samples (same treatment, different animals). Ten low-quality sets and 28 outliers were identified. The low-quality sets showed higher coefficient of variation (CV)% of DNA quantities in replicate experiments than high-quality sets (63% versus 26%; $P = 0.001$) and contained a higher proportion of outlying quantities (35% versus 5.9%; $P = 0.001$) when individual samples were detected by adjusted KOD. Outlier detection with adjusted KOD reduced the false detection of outliers by 2/3 compared with the previous, nonadjusted version of KOD (20% versus 5.9%; $P = 0.001$). We conclude that the presented tests can be used to assign technical reasons to outlying observations.

INTRODUCTION

Like other quantitative techniques, quantification of DNA by real-time PCR requires determining the origin of outlying results. While biological outliers are interesting, outliers due to technical malfunction should be eliminated. Although intense attention has been paid to quality control in the steps preceding PCR, entailing mainly the use of reference genes (1,2), quality assessment of real-time PCR itself has been left virtually untreated, despite the prevalence of PCR inhibition that might substantially impair the accuracy of quantification (3–6).

Quantification by real-time PCR is sensitive to subtle differences in PCR efficiency between samples. Thus, most quantification methods require that the compared samples have similar PCR efficiency (ABI PRISM 7700 Sequence Detection System, User Bulletin #2. Applied Biosystems, Foster City, CA, USA) (7,8). For samples that have a cycle threshold (C_t) of less than 35, indicating reproducible quantification, as suggested by Soong et al. (9), a possible quality assessment for real-time PCR should focus on the strict requirement for similar PCR

efficiencies among compared reactions of the same gene.

In our earlier work, we described kinetic outlier detection (KOD) as a tool for identifying samples with outlying efficiencies based on their sample-specific efficiency (kinetics). KOD detects outliers by comparing the efficiency of a test sample to the mean efficiency of standard curve samples (10), where the efficiency SD of standard curve is used as a criterion for high-quality quantification. That is, according to KOD, an outlier is defined as a test sample whose sample-specific efficiency differs from the mean efficiency of the standard curve by more than 2 SD of the efficiency of standard curve.

The KOD method is effective for absolute quantification by real-time PCR, but it was not designed for relative quantification procedures. There are important methodological differences between absolute and relative quantification. While in absolute quantification, a test sample is compared with standard curve samples (2,11), in relative quantification, a test sample is compared with other test samples (12,13). Therefore, in relative quantification, a test sample should

have efficiency similar to the other test samples. As long as all test samples have similar efficiencies, differences in efficiency between test samples and standard curve samples (if used) will affect the significance, but not the trend of the end result, as can be deduced from Pfaffl's work (13). Therefore, the aim of the study described here was to develop and test statistical tools for verification of efficiency similarity between test samples in relative quantification.

To assess whether test samples indeed have similar efficiencies, we estimated sample-specific efficiencies using a modified version of the method published by Liu and Saint (14) and Tichopad et al. (15) (see the supplementary material available online at www.BioTechniques.com for details). We then applied two methods to assess the comparability and quality of test samples: (i) adjustment of KOD to relative quantification (hereafter called "adjusted KOD"), where the efficiency of a test sample is compared with the mean of the efficiencies of other test samples, and (ii) a "variance test," which is a comparison of test samples' efficiency variance ($S_{E_test}^2$) to the efficiency variance of highly reproducible quantifications ($S_{E_rep}^2$).

MATERIALS AND METHODS

Experimental Design

A natural means of studying the effect of efficiency differences between samples on quantification would be to compare the amounts of a quantified gene. However, in practice, the steps preceding PCR (i.e., nucleic acid extraction, purification, and reverse transcription) also influence the quantification. The common procedure to deal with this problem is to normalize the expression of target genes to that of one or more stably expressed reference genes in the same sample (1,16). Therefore, to study the effect of unequal efficiencies on quantification, we applied this internal reference concept. Instead of studying absolute amounts, the effect of dissimilar efficiencies was studied from the ratios of two genes

Protocol 1. Adjusted KOD to Relative Quantification

Step 1: Select a source of variance as described below (i.e., internal or external) and calculate $S_{E_rep}^2$.

Step 2: Estimate the mean efficiency, \bar{E} of the experimentally relevant samples to which test samples will be compared. In relative quantification, these are the other n test samples, whereas in absolute quantification, these are the n samples of the standard curve.

Step 3: If external source of variance is used, use Equation 1 to evaluate the probability that the efficiency of a test sample, E_{test} , originates from the same distribution as the mean efficiency found in Step 2. Note that, unlike in our previous publication (10), this equation takes into account the number of samples used to estimate \bar{E} .

$$P = 2 \left[1 - \Phi \left(\frac{|\bar{E} - E_{test}|}{\sqrt{S_{E_rep}^2 \left(1 + \frac{1}{n}\right)}} \right) \right] \quad [\text{Eq. 1}]$$

Where Φ is the cumulative distribution function for the standard normal distribution.
KOD: kinetic outlier detection.

Protocol 2. A Variance Test for Relative Quantification

Step 1: Group all the test samples to be compared and estimate their variance ($S_{E_test}^2$).

Step 2: Select a source of variance as described below (i.e. internal or external) and calculate $S_{E_rep}^2$.

Step 3: If internal source of variance was selected, use an F-test to test the similarity of $S_{E_test}^2$ and $S_{E_rep}^2$. If external source of variance was selected, use Chi-Square test (Equation 2) to test if $S_{E_test}^2 \leq S_{E_rep}^2$.

$$\chi^2 = \frac{(n-1) * S_{E_test}^2}{S_{E_rep}^2} \quad [\text{Eq. 2}]$$

Here n is the number of test samples used to estimate $S_{E_test}^2$. The calculated statistic χ^2 obtained from Equation 2 is tested against a table value with $n - 1$ degrees of freedom and significance level α of 0.05 or less.

(*Hprt* and *Ppia*), where the efficiency of one gene (*Ppia*) is similar in all compared samples, and the efficiency of the other gene (*Hprt*) varies.

Ppia and *Hprt* expression was quantified in 75 cDNA samples from 5 different brain parts of 15 rats (Figure 1). *Hprt* expression was quantified in each cDNA sample, first by an inexperienced investigator and then by an expert. *Ppia* expression was quantified by the expert only. This gave a total of 225 measurements of gene expression level (150 performed by the expert and 75 by the inexperienced investigator). Efficiencies were estimated from every amplification curve using the method described in the supplementary material.

High similarity of efficiencies in all replicate sets of one gene (*Ppia*) and a range of similarities in the other gene (*Hprt*) reduced the number of occurrences in which inhibition in one gene in a sample might be compensated by similar inhibition in the other gene in the same sample. This simplified the demonstration of the effect of efficiency dissimilarity on quantification.

One out of the 225 test reactions was excluded because no signal was obtained. Two samples with anomalously low expression of *Ppia* were excluded from the study despite the fact

that they were not identified as outliers by the examination of the estimated efficiency. It could be that these two low quantities resulted from temporal inhibition of *Taq* DNA polymerase at early cycles, inhibition that probably ended before the detectable phase and therefore could not be identified by kinetics quality assessment tests.

Mathematical Models for Quality Assessment Tests

Both the adjusted KOD and the variance test use the efficiency variance of highly reproducible quantifications: $S_{E_rep}^2$. The source of these quantifications can be a standard curve with high correlation between the C_t values and $\text{Log}(\text{DNA amount})$, typically 0.99 in 10-fold dilution series, or with a mean square error (the mean distance of data points from their regression) of less than 0.1. An alternative source for reproducible quantifications can be a set of PCR replicates with a spread of C_t values similar to the recommended spread defined by the manufacturer [e.g., LightCycler® Technical Note No. LC 10/update 2003 (Roche Applied Science, Indianapolis, IN, USA) and iCycler™ Technical Note 2620 (Bio-Rad Laboratories, Hercules, CA, USA)]. If the high-quality quantifica-

tions were assayed together with the test samples, we refer to this source of efficiency variance as “internal.” After several experiments, when several $S_{E_rep}^2$ are available, an alternative source for $S_{E_rep}^2$ could be the average $S_{E_rep}^2$ from these experiments. We consider this source of variance “external.”

The first test we examined was adjusted KOD. The steps describing the test are listed in Protocol 1. A small P value (e.g., 0.05 or lower) indicates low probability that the efficiency of the test sample originated from the same distribution as the mean efficiency of experimentally relevant samples: that is, the tested sample is an outlier. If an internal source of variance was used, Φ should be replaced with the cumulative distribution function for the student t distribution with $n - 1$ degrees of freedom.

The second test we examined was the variance test, which is described in the steps listed in Protocol 2. The χ^2 value calculated from the equation in Protocol 2 is tested against a table value with $n - 1$ degrees of freedom and a significance level α of 0.05 or less. If the calculated χ^2 is larger than the table value, the hypothesis that $S_{E_test}^2 \leq S_{E_rep}^2$ is rejected, and the set is considered low quality.

RESULTS AND DISCUSSION

The data set used in this study consists of relative quantifications of *Hprt* and *Ppia* levels from 75 cDNA samples arranged in sets of 5 replicate animals each (Figure 1). To reduce the effect of steps preceding the PCR on quantification, the ratio of *Hprt*/*Ppia* quantities was examined, as is often done in normalization of gene expression by a reference gene (17,18).

Figure 2 shows the spread of *Hprt*/*Ppia* quantities (cv%) versus the spread of efficiencies of *Hprt* $S_{E_test}^2$ in each set of 5 replicate animals. High correlation was obtained ($R = 0.81, P < 0.0001$) between the spread of quantities and spread of efficiencies. The strong correlation between efficiencies spread and quantities spread is the rationale for the variance test, which alerts the user of a too-high efficiency spread that might be associated with the high quantities spread.

In the next step, we implemented the variance test on the sets mentioned above, with external source of variance equal to the mean $S_{E_rep}^2$ of 21 standard

curves of 7 genes. This was possible after low correlation was found between mean efficiencies of the standard curves and their variances ($R = 0.17, P = 0.45$; Supplementary Figure S4). Ten low-quality sets, with a significantly higher spread of efficiencies compared with highly reproducible quantifications, were identified (in Figure 2, these are the sets to the right of the vertical line). The spread of quantities (cv% of *Hprt*/*Ppia*) in low-quality sets was significantly higher as compared with the spread of quantities in high-quality sets; cv% = 63% versus 26% when using an external source of variance (Wilcoxon rank-sum, $P = 0.0001$) or cv% = 55% versus 27% when using an internal source of variance (Wilcoxon rank-sum, $P = 0.001$; data not shown).

Outliers were identified by two different versions of KOD: (i) adjusted KOD for relative quantification (i.e., the efficiency of a test sample was compared with the mean efficiency of the other test samples); and (ii) the original KOD (10), a version for absolute quantification (i.e., the efficiency of a test sample was

compared with the efficiency of the standard curve samples). Both versions of KOD were tested with external and internal sources of variance.

Table 1 summarizes the results of outlier detection and variance testing. It lists the proportion of outliers identified by adjusted KOD and standard KOD in high- and low-quality sets found by internal and external variance tests. When using adjusted KOD, the proportion of outliers was higher in the low-quality sets compared with the high-quality sets (z test, $P = 0.0001$). False outlier detection, the proportion of outliers in high-quality sets, was lower in adjusted KOD than in standard KOD (z test, $P = 0.001$ for internal source of variance and $P = 0.02$ for external source of variance).

All low-quality sets and most of the outliers were found in the inexperienced worker data. However, the data from both workers showed a similar expression pattern—stable expression of *Hprt*/*Ppia* at the different drug doses and brain parts (Kruskal-Wallis test, $P = 0.17$ for the inexperienced worker and $P = 0.47$ for the experienced worker).

The stable expression of these two genes is known from other studies, and often these two genes are used as housekeeping genes (1). The similarity in expression pattern and difference between the two workers only in the spread of the quantities in the replicate animals could possibly be explained by the pattern of efficiency spread. Despite the large sample size ($n_1 = 74, n_2 = 73$), no significant difference was found between the mean efficiencies of the data generated by the two workers ($\bar{E}_1 = 0.887, \bar{E}_2 = 0.874$, two-tailed Student's *t*-test, unequal variances, $P = 0.06$); this is contrary to the significant difference in the spread of efficiencies

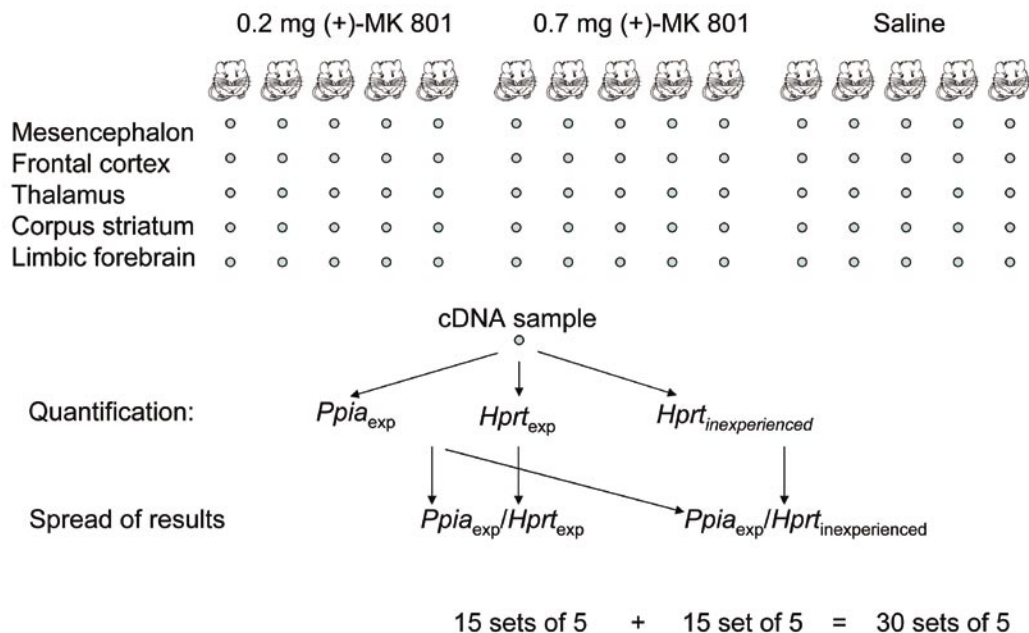


Figure 1. Experimental design. Five rats were treated with 0.2 mg (+)-MK 801 hydrogen maleate per kilogram body weight, 5 rats were treated with 0.7 mg of the same compound per kilogram body weight, and 5 control rats were given saline and total RNA was extracted from 5 different parts of the brain (10). After reverse transcription, *Hprt* was quantified twice, once by an inexperienced worker and again by an expert investigator, and *Ppia* was quantified by an expert investigator (for details about RNA extraction, reverse transcription, and PCR primers and conditions, please see the supplementary materials available online). The ratio *Hprt*/*Ppia* was calculated twice, once for each *Hprt* quantification, giving a total of 30 sets of 5 replicates each.

Table 1. Percentage of Outliers Detected by Adjusted and Standard KOD for Relative Quantification in High- and Low-Quality Sets

	Outliers in High-Quality Samples (%)	Outliers in Low-Quality Samples (%)
Internal $S_{E_rep}^2$		
Adjusted KOD	5.9 (10/170)	35 (18/52)
KOD	20 (34/170)	23 (12/52)
External $S_{E_rep}^2$		
Adjusted KOD	4.9 (9/183)	38 (15/39)
KOD	16 (29/183)	43 (17/39)

The actual numbers of outliers in each sample are given in parentheses. KOD, kinetic outlier detection.

(inexperienced $S_{E_test}^2 = 2.19 \times 10^{-3}$, experienced $S_{E_test}^2 = 6.68 \times 10^{-4}$, F-test, $P < 0.0001$).

The exponential nature of PCR makes it vulnerable to differences in efficiencies of compared samples. Therefore, most quantification methods require similar PCR efficiencies among compared samples. Although reference genes may properly compensate for PCR inhibition in a target gene, this might not be the situation in all cases because target and reference PCR systems may have different PCR efficiencies and their expression levels might differ substantially, resulting in a different quantitative effect of the inhibitor.

Here we have presented two statistical tools to assess this requirement

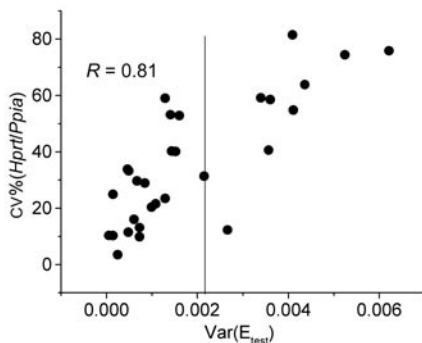


Figure 2. Variance of sample-specific efficiencies as indicator for comparability of samples. The coefficient of variation (cv)% of *Hprt/Ppia* quantities in 30 sets of 5 replicates was plotted versus the variance of their sample-specific efficiencies, $S_{E_test}^2$. The vertical line indicates the critical value for low-quality sets, according to external source of variance equal to the average $S_{E_rep}^2$ of 21 standard curves ($S_{E_rep}^2 = 7.93 \times 10^{-4}$) and Chi-square test at $P = 0.05$. The correlation between cv% and $S_{E_test}^2$, R , is given in the upper left-hand side.

in relative quantification. Both tools are based on two heuristics: (i) that $S_{E_rep}^2$ reflects the technical component of $S_{E_test}^2$, including instrumental and labor-related errors, and (ii) that PCR efficiency and DNA quantity are associated via the basic equation of PCR (19). Then, if test samples have efficiencies that are not outliers according to Equation 1, or $S_{E_test}^2 \leq S_{E_rep}^2$, the samples quantification is expected to be accurate with comparability similar to the comparability of the highly reproducible quantifications used in the tests. A quantitative measure for the comparability of samples quantified by a standard curve, in the form of minimum detectable fold difference in initial copy number with 95% confidence is given in the following equation.

$$\text{Comparability} = (1 + E)^{1.96 * MSE}$$

where E is the efficiency of the standard curve and MSE is the mean square error of the standard curve (as in LightCycler Technical Note No. LC 11/2000). If the PCR replicates were used as highly reproducible quantifications, then MSE is replaced by the replicates $SD(C_i)$.

Adjusted KOD reduces false-positive outlier detection caused by dissimilarity between the efficiencies of the standard curve and the test samples. Although the dissimilarity in efficiency in this study could be a result of two different sources of starting material (purified PCR product versus cDNA), such differences might also appear when inhibitors of *Taq* DNA polymerase are present in the cDNA pool used for constructing the standard curve (3–6). Interestingly, different

inhibitory effects even between replicates were reported earlier in a stringent study of PCR inhibitors by internal control (20), emphasizing the need for kinetics quality assessment for each sample. The frequency of PCR inhibition (15) suggests that adjusted KOD would be relevant to a large number of studies.

Adjusted KOD and the variance test may be combined by first detecting and excluding outliers and then testing the variance. In this sequence, the variance test can alert the user of the presence of subsets of efficiencies, as in bipolar or multipolar distribution, which may have different efficiencies even though they are not outliers. Unfortunately, we could not show subsets of efficiencies in the small size sets in this study ($n \leq 5$). Based on the $S_{E_rep}^2$ used in this study, undetectable differences in efficiencies might amount to 0.06–0.09 with profound quantitative consequences. The variance test may be found useful when parameters known to affect PCR efficiency vary between compared samples [e.g., type of tissue (21) or quality of SYBR® green and dNTPs in and between runs (22,23)].

We conclude that once KOD is adjusted to the method of quantification, it can be used with the variance test for kinetics quality assessment of relative quantification by real-time PCR. However, exclusion of low-quality samples and sets from further analysis should not be automatic because reference genes may properly compensate for the dissimilarity in efficiencies in the target gene. Therefore, until a comprehensive model for kinetics quality assessment for a multiple-gene system is published, we suggest using the variance test and KOD as tools to help the real-time PCR user assign a technical reason to outlying results. That is, samples that are both biological outliers and are of low technical quality should be considered suspect. Once suspect samples are identified, their subsequent processing could include their dilution to reduce inhibitor concentration or repetition of earlier steps in the quantification, possibly with an additional step for the removal of PCR inhibitors.

ACKNOWLEDGMENTS

We thank Mikael Kubista, Jose Manuel Andrade, and Tamar Schirman-Hildesheim for critically reading the manuscript, and Carlsson Research for kindly supplying us with data. This work was financially supported by SWEGENE and the Chalmers Bioscience Effort.

COMPETING INTERESTS STATEMENT

T.B. declares that a provisional patent application including the material published in this article was filed with the U.S. Patent Office. A.M. declares no competing interests.

REFERENCES

- Thellin, O., W. Zorzi, B. Lakaye, B. De Borman, B. Coumans, G. Hennen, T. Grisar, A. Igout, and E. Heinen. 1999. Housekeeping genes as internal standards: use and limits. *J. Biotechnol.* 75:291-295.
- Bustin, S.A. 2000. Absolute quantification of mRNA using real-time reverse transcriptase polymerase chain reaction assays. *J. Mol. Endocrinol.* 25:169-193.
- Wilson, I.G. 1997. Inhibition and facilitation of nucleic acid amplification. *Appl. Environ. Microbiol.* 63:3741-3751.
- Wiedbrauk, D.L., J.C. Werner, and A.M. Drevon. 1995. Inhibition of PCR by aqueous and vitreous fluids. *J. Clin. Microbiol.* 33:2643-2646.
- Rossen, L., P. Norskov, K. Holmstrom, and O.F. Rasmussen. 1992. Inhibition of PCR by components of food samples, microbial diagnostic assays and DNA-extraction solutions. *Int. J. Food Microbiol.* 17:37-45.
- Chandler, D.P., C.A. Wagnon, and H. Bolton, Jr. 1998. Reverse transcriptase (RT) inhibition of PCR at low concentrations of template and its implications for quantitative RT-PCR. *Appl. Environ. Microbiol.* 64:669-677.
- Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29:E45.
- Gentle, A., F. Anastasopoulos, and N.A. McBrien. 2001. High-resolution semi-quantitative real-time PCR without the use of a standard curve. *BioTechniques* 31:502-508.
- Soong, R. and A. Ladanyi. 2003. Improved indicators for assessing the reliability of detection and quantification by kinetic PCR. *Clin. Chem.* 49:973-976.
- Bar, T., A. Ståhlberg, A. Muszta, and M. Kubista. 2003. Kinetic Outlier Detection (KOD) in real-time PCR. *Nucleic Acids Res.* 31:E105.
- Rutledge, R.G. and C. Cote. 2003. Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Res.* 31:E93.
- Livak, K.J. and T.D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25:402-408.
- Pfaffl, M.W. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29:E45.
- Liu, W. and D.A. Saint. 2002. Validation of a quantitative method for real time PCR kinetics. *Biochem. Biophys. Res. Commun.* 294:347-353.
- Tichopad, A., A. Dzidic, and M.W. Pfaffl. 2002. Improving quantitative real-time RT-PCR reproducibility by boosting primer-linked amplification efficiency. *Biotechnol. Lett.* 24:2053-2056.
- Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3:RESEARCH0034.
- Andersen, C.L., J.L. Jensen, and T.F. Orntoft. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64:5245-5250.
- Szabo, A., C.M. Perou, M. Karaca, L. Perreard, J.F. Quackenbush, and P.S. Bernard. 2004. Statistical modeling for selecting housekeeper genes. *Genome Biol.* 5:R59.
- Wiesner, R.J., J.C. Ruegg, and I. Morano. 1992. Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues. *Biochem. Biophys. Res. Commun.* 183:553-559.
- Rosenstraus, M., Z. Wang, S.Y. Chang, D. DeBonville, and J.P. Spadaro. 1998. An internal control for routine diagnostic PCR: design, properties, and effect on clinical performance. *J. Clin. Microbiol.* 36:191-197.
- Tichopad, A., A. Didier, and M.W. Pfaffl. 2004. Inhibition of real-time RT-PCR quantification due to tissue-specific contaminants. *Mol. Cell. Probes* 18:45-50.
- Kubista, M., A. Ståhlberg, and T. Bar. 2001. Light-up probe based real-time Q-PCR., p. 53-58. *In* T.W. Raghavachari R. (Ed.), *Genomics and Proteomics Technologies. Proceedings of SPIE.*
- Karsai, A., S. Muller, S. Platz, and M.T. Hauser. 2002. Evaluation of a homemade SYBR green I reaction mixture for real-time PCR quantification of gene expression. *BioTechniques* 32:790-796.

Received 10 January 2005; accepted 3 March 2005.

Address correspondence to Tzachi Bar, Ygaal Alon 31, K. Tivon 36083, Israel. e-mail: tzachi.bar@gmail.com

To purchase reprints of this article, contact apfeffer@BioTechniques.com