Analysis of RT-qPCR Data

James R. Perkins[1], Nor Izayu Abdul Rahman[2] and Matthias Kohl[2,*]

[1]*Institute of Structural and Molecular Biology, Division of Biosciences, University College London, Gower Street, London, WC1E 6BT, UK*

[2]*Department of Mechanical and Process Engineering, Furtwangen University, Jakob-Kienzle-Str. 17, D-78054 VS-Schwenningen, Germany*

Abstract: We give a brief overview of the necessary steps in the analysis of real-time reverse transcription quantitative polymerase chain reaction (RT-qPCR) data. We cover determination of amplification efficiency, background correction, normalization, quality control, and statistical analysis.

Keywords: RT-qPCR, DNA, miRNA, Cq value, amplification efficiency, background correction, normalisation, reference gene, quality control, statistical analysis.

INTRODUCTION

Real-time reverse transcription quantitative polymerase chain reaction (RT-qPCR) is perhaps the most commonly used technology for measuring gene expression [1]. It combines reverse transcription of the RNA species of interest to its DNA complement, followed by quantification using real-time qPCR. There is a wide range of different platforms available, allowing the user to measure expression for a small number of genes (such as Quiagen Rotor-Gene Q) to medium and high-throughput machines, such as Roche LightCycler and Applied Biosciences Taqman Gene Signature arrays. The expression level of a given gene (or other biological molecule such as an miRNA) is estimated as a function of the number of cycles necessary for the detection of a given DNA sequence complementary to the target sequence in a sample above background noise. This number can be called Ct (cycle threshold), Cp (crossing point), or TOP (take-off point) value. However these are all manufacturer-specific terms and refer to the same thing. In the MIQE guidelines it is thus recommended to use Cq (quantification cycle) value only for the sake of consistency and clarity [2].

DETERMINATION OF AMPLIFICATION EFFICIENCY

An important parameter for calculating the Cq values is the PCR amplification efficiency, defined as the increase in the amount of the amplified target sequence that takes place in one PCR cycle. It can be expressed in two formulas for determining the number of DNA products after n cycles

$$N_n = N_o E^n; 1 < E < 2$$

or

$$N_n = N_o (1 + \tilde{E})^n; 0 < \tilde{E} < 1$$

where N_0 is the amount of primer at the onset of the reaction and N_n is the amount of reaction products after n cycles. The second formula is normally used when efficiency is expressed as a percentage; see [3] for more details.

If small differences between the target and reference genes are not taken into account this may lead to large errors for the calculated expression ratios and hence a clear over- or underestimation of changes. Therefore it is recommended to assess the exact efficiencies of target and reference genes [4].

There are four main methods for determining the efficiency; see [3, 4] for more details. First, it can be calculated from the slopes of so-called calibration or standard curves derived from a dilution series. This tends to overestimate efficiencies and frequently returns practically impossible values larger than E = 2 (resp. $\tilde{E}$ = 100%). The second method is based on the fluorescence increase in the third linear phase in the logarithmic plot of the fluorescence values. Again the efficiency is calculated as the slope of a fitted linear regression model. This method often underestimates the true efficiencies and typically yields values between E = 1.35 (resp. $\tilde{E}$ = 35%) and E = 1.60 (resp. $\tilde{E}$ = 60%). Thirdly, one can determine the efficiency using the data from the second (real) exponential phase and fit a polynomial curve to the data. With this method one

*Address correspondence to this author at the Department of Mechanical and Process Engineering, Furtwangen University, Jakob-Kienzle-Str. 17, D-78054 VS-Schwenningen, Germany; Tel: +49 (0) 7720 307-4746;
Fax: +49 (0) 7720 307-4727; E-mail: Matthias.Kohl@hs-furtwangen.de

E-ISSN: 1929-6029/12

typically obtains realistic efficiencies between E = 1.75 (resp. $\tilde{E}$ = 75%) and E = 1.90 (resp. $\tilde{E}$ = 90%). Finally, one can also use all data points and fit a sigmoidal or logistic curve to the data. This method is easy to perform and yields good estimators for the maximum curve slope. However, the derived slope parameters do not represent the true (unknown) PCR efficiency. Refer to [5] for more details.

It is important to use the same method throughout the whole experiment and also if one wants to compare results from different experiments. It is still an open question which method is the best one in terms of reliability and reproducibility.

BACKGROUND CORRECTION AND NORMALISATION

Before calculating Cq values one should perform a background correction. Ideally the background is constant over the cycles but often it is rather noisy showing drifts upwards or downwards. Hence, there are various approaches for background correction ranging from subtracting a constant to fitting nonlinear models [3, 4].

The normalisation of Cq values is an essential step in RT-qPCR analysis, necessary to make different samples comparable to each other. By far the most common way to do this is through the use of one or several reference (or housekeeping) genes. Under the assumption that the reference genes are constitutively expressed and do not change in expression between the different cells and/or conditions investigated by the experiment, the user can subtract the Cq values of the reference genes from the other genes in a sample. This will control for a number of different experimental factors that might differ between samples, including (sample specific) amplification efficiency, amount of starting sample and RNA integrity. Two frequently chosen reference genes are glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and β2-microglobulin (B2M). However many widely used reference genes do actually change in expression under certain conditions and the existence of a universal reference gene for all samples is debatable. Some commonly used reference genes such as GAPDH have been shown to vary in expression between different tissues and conditions [6]. It is therefore essential to measure the expression of a number of reference genes in order to validate them for use under the conditions of interest. Following this the best performing reference genes should be combined to calculate a normalisation factor [7]. Several procedures exist to determine the best reference genes for a given experiment; the number of reference genes that should be used depends on the nature of the experimental data [8]. The most frequently used ones are probably geNorm [7] and NormFinder [9], both included in the freely available Bioconductor package NormqPCR [10]. In addition, the MIQE guidelines state that reference gene abundance should show strong correlation with total mRNA amounts and that this correlation should be reported [2].

Recently, microarray derived “data driven” methods, such as quantile normalisation, have been proposed for normalising high-throughput RT-qPCR experimental data [11]. These methods are applicable when it is expected that the majority of genes are unchanging between experimental conditions.

QUALITY CONTROL

RT-qPCR is relatively easy to perform; however, it is also easy for things to go wrong. It is important to check the quality of the data produced in each stage of the experiment: for the determination of amplification efficiency, calculation of Cq values, and after normalization. Technical replicates should be used to ensure reliable results and the experimental design should be such that technical variation can be separated from biological variation, particularly when separate runs are being performed. Sample extraction and primer design are also essential steps in an RT-qPCR experiment, problems with either could lead to misleading results. Full details on the procedures that should be followed for the different experimental steps are given in [2].

STATISTICAL ANALYSIS

The statistical analysis depends on the experimental design and the biological question being asked. Several kinds of uni- and multivariate statistical methods may be appropriate [12]. In any case the use of biological replicates is essential in order to draw meaningful and reproducible conclusions. Adjustment of the p-values to account for multiple testing, e.g. by false discovery rate (FDR), should be performed. This reduces detection of spurious results, i.e. genes that do not change in expression between conditions being deemed as changing. A further improvement of the reproducibility and reliability of the results can be expected by combining adjusted p-values with additional criteria such as fold change.

ACKNOWLEDGMENTS

We would like to thank two anonymous referees for valuable comments on the manuscript.

APPENDIX OF SYMBOLS

RT	=	Reverse transcription
PCR	=	Polymerase chain reaction
qPCR	=	Real-time quantitative PCR
RT-qPCR	=	Reverse transcription qPCR
DNA	=	Deoxyribonucleic acid
RNA	=	Ribonucleic acid
miRNA	=	microRNA, small RNA molecule (ca. 22 nucleotides)
Cq	=	Quantification cycle
Ct	=	Cycle threshold
Cp	=	Crossing point
TOP	=	Take-off point
MIQE	=	Minimum information for publication of quantitative real-time PCR experiments
GAPDH	=	Glyceraldehyde-3-phosphate dehydrogenase
B2M	=	β2-microglobulin
FDR	=	False discovery rate

REFERENCES

[1] VanGuilder H.D., Vrana K.E., and Freeman W.M. Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques 2008 Apr; 44(5): 619-626. http://dx.doi.org/10.2144/000112776

[2] Bustin S.A., Benes V., Garson J.A., Hellemans J., Huggett J., Kubista M., Mueller R., Nolan T., Pfaffl M.W., Shipley G.L., Vandesompele J., and Wittwer C.T. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin Chem. 2009 Apr; 55(4): 611-22. http://dx.doi.org/10.1373/clinchem.2008.112797

[3] Rebrikov D.V. and Trofimov D. Yu. Real-Time PCR: A Review of Approaches to Data Analysis. Applied Biochemistry and Microbiology 2006; 42(5): 455-463. http://dx.doi.org/10.1134/S0003683806050024

[4] Pfaffl M. W. Chapter 3 - Quantification strategies in real-time PCR. In: Bustin S. A., editor. A-Z of quantitative PCR. International University Line (IUL) 2004.

[5] Spiess A.-N., Feig C., and Ritz C. Highly accurate sigmoidal fitting of real-time PCR data by introducing a parameter for asymmetry. BMC Bioinformatics 2008 Apr; 9:221. http://dx.doi.org/10.1186/1471-2105-9-221

[6] Radonić A., Thulke S., Mackay I.M., Landt O., Siegert W., and Nitsche A. Guideline to reference gene selection for quantitative real-time PCR. Biochem Biophys Res Commun. 2004 Jan 23;313(4): 856-62. http://dx.doi.org/10.1016/j.bbrc.2003.11.177

[7] Vandesompele J., Preter K.D., Pattyn F., Poppe B., Roy N.V., Paepe A.D., and Speleman F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 2002 Jun; 3(7):RESEARCH0034.

[8] Vandesompele J., Kubista M., and Pfaffl M.W. Reference gene validation software for improved normalization. In: Logan J., Edwards K., and Saunders N., editors. Real-Time PCR: Current Technologies and Applications. Caister Academic Press 2009.

[9] Andersen C.L., Jensen JL, Ørntoft TF. 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. 2004 Aug 1;64(15):5245-50. http://dx.doi.org/10.1158/0008-5472.CAN-04-0496

[10] Perkins J.R., Dawes J.M., McMahon S.B., Bennett D.L., Orengo C., and Kohl M. ReadqPCR and NormqPCR: R packages for the reading, quality checking and normalisation of RT-qPCR quantification cycle (Cq) data. BMC Genomics 2012 Jul; 13:296. http://dx.doi.org/10.1186/1471-2164-13-296

[11] Dvinge H., Bertone P. HTqPCR: high-throughput analysis and visualization of quantitative real-time PCR data in R. Bioinformatics. 2009 Dec; 25(24): 3325-6. http://dx.doi.org/10.1093/bioinformatics/btp578

[12] Pfaffl M.W., Vandesompele J., and Kubista M. Data analysis software. Logan J., Edwards K., and Saunders N., editors. Real-Time PCR: Current Technologies and Applications. Caister Academic Press 2009.

Received on 21-11-2012 Accepted on 05-12-2012 Published on 21-12-2012

http://dx.doi.org/10.6000/1929-6029.2012.01.02.10