

Methodology article

Open Access

## Model based analysis of real-time PCR data from DNA binding dye protocols

Mariano J Alvarez\*<sup>1,2,4</sup>, Guillermo J Vila-Ortiz<sup>1</sup>, Mariano C Salibe<sup>1</sup>,  
Osvaldo L Podhajcer<sup>†2</sup> and Fernando J Pitossi<sup>†3</sup>

Address: <sup>1</sup>Gentron Research Unit, Arenales 1457 – 2° Piso, Buenos Aires C1061AAO, Argentina, <sup>2</sup>Gene Therapy Laboratory, Leloir Institute, CONICET, University of Buenos Aires, Patricias Argentinas 435, Buenos Aires C1405BWE, Argentina, <sup>3</sup>Neuroimmunomodulation and Gene Therapy Laboratory, Leloir Institute, CONICET, University of Buenos Aires, Patricias Argentinas 435, Buenos Aires C1405BWE, Argentina and <sup>4</sup>Joint Centers for Systems Biology, Columbia University, 1130 St Nicholas Avenue, New York, NY 10032, USA

Email: Mariano J Alvarez\* - malvarez@c2b2.columbia.edu; Guillermo J Vila-Ortiz - gvila@leloir.org.ar;

Mariano C Salibe - msalibe@leloir.org.ar; Osvaldo L Podhajcer - opodhajcer@leloir.org.ar; Fernando J Pitossi - fpitossi@leloir.org.ar

\* Corresponding author †Equal contributors

Published: 9 March 2007

Received: 5 January 2007

BMC Bioinformatics 2007, 8:85 doi:10.1186/1471-2105-8-85

Accepted: 9 March 2007

This article is available from: <http://www.biomedcentral.com/1471-2105/8/85>

© 2007 Alvarez et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Reverse transcription followed by real-time PCR is widely used for quantification of specific mRNA, and with the use of double-stranded DNA binding dyes it is becoming a standard for microarray data validation. Despite the kinetic information generated by real-time PCR, most popular analysis methods assume constant amplification efficiency among samples, introducing strong biases when amplification efficiencies are not the same.

**Results:** We present here a new mathematical model based on the classic exponential description of the PCR, but modeling amplification efficiency as a sigmoidal function of the product yield. The model was validated with experimental results and used for the development of a new method for real-time PCR data analysis. This model based method for real-time PCR data analysis showed the best accuracy and precision compared with previous methods when used for quantification of *in-silico* generated and experimental real-time PCR results. Moreover, the method is suitable for the analyses of samples with similar or dissimilar amplification efficiency.

**Conclusion:** The presented method showed the best accuracy and precision. Moreover, it does not depend on calibration curves, making it ideal for fully automated high-throughput applications.

### Background

The reverse transcription polymerase chain reaction (RT-PCR) is the most sensitive method for the detection of specific mRNAs [1]. However, due to the exponential nature of the PCR amplification process, small differences in amplification efficiency among samples may lead to very different product yields, making RT-PCR unsuitable for quantitative purposes. The use of exogenously added standard sequences as internal competitors has overcome

partially this issue [2-4], but the setting up of the system must be done for each target sequence taking into account multiple error sources [5], making competitive PCR unsuitable for high-throughput applications. The recent introduction of fluorescence techniques and instruments able to quantify the DNA content in each cycle had led in the last years to the development of real-time PCR [6-8]. Because the high sensitivity of fluorescent product detection, real-time PCR does not rely on end-point analyses.

Moreover, cycle-by-cycle data generated by real-time PCR provides information about the kinetics of the amplification process, overcoming the limitations of classical RT-PCR. Recently, the introduction of double-stranded DNA specific dyes [9] allowed the quantification of multiple targets without the need of specific fluorescent probes, making real-time PCR a popular method for microarray data validation [10].

Most currently used real-time PCR data analysis methods are based on determining the threshold cycle (*CT*), which is the cycle number at which a fixed amount of product is formed [11]. The comparative *CT* method, a broadly used semi-quantitative one, is based on the exponential description of the PCR process assuming constant amplification efficiency equal to 1 [12]. However, in our applications of SYBR-Green I real-time PCR, we have found amplification efficiency always lower than 1. Moreover, it has been shown that a difference as small as 4% in PCR efficiency could be translated to 400% error in comparative *CT* method based quantifications [13]. Thus, reliable quantitative real-time PCR depends on good estimations of PCR efficiency. Several methods have been proposed for amplification efficiency estimation. One of them uses a dilution curve to estimate the amplification efficiency of each target sequence [14], while the others estimate the amplification efficiency from single reaction data [13,15-17]. Here we introduce a new mathematical model based on the classic exponential description of the PCR, in which amplification efficiency was modelled as a sigmoid function of the product yield. The model was validated with experimental data and it was used for the development of a new method for real-time PCR data analysis. This model based method for real-time PCR data analysis (MoBPA) estimates PCR amplification efficiency from single sample reaction data, eliminating the need of calibration curves, which is a drawback for high-throughput implementations [18]. Moreover, MoBPA showed the highest accuracy and precision compared with previous methods when used for quantification of samples amplified with similar or dissimilar efficiency.

## Results and discussion

### The model

According to its discrete nature, the PCR process can be expressed by the difference equation,

$$T_{n+1} = T_n \cdot (1 + E_n); \quad E_n \in (0,1) \quad (1)$$

where  $T_n$  is the PCR product yield at cycle  $n$ , and  $E$  is the amplification efficiency of the reaction. In a previous work, we have modelled the amplification efficiency as a linear function of the product yield (model 1; Fig. 1A) [5]. However, a recent kinetic description of the real-time PCR showed a sigmoid relationship between the effective

amplification efficiency and the product yield, either whether primer, nucleotides or DNA polymerase become limiting [19]. This prompted us to evaluate two additional empirical models describing the amplification efficiency as a function of PCR product yield: a three parameters, and a two parameters sigmoid models (models 2 and 3; Fig. 1B and 1C). Both models imply that the amplification efficiency change dynamically also during the exponential phase, which is in agreement with the work of Liu, et.al [15]. However, our models differ from the sigmoid description of the PCR reaction, because we model the amplification efficiency as a function of the product yield, instead of describing it as a function of the cycle number. The models were fitted to the same real-time PCR dataset comprising 1723 PCR reactions, with initial target amount spanning 9 orders of magnitude and *CT* values that range from 10 to 35.9. Despite the fact that the three models explained more than 95% of the data variance, only models 2 and 3 showed a random distribution of residuals (Fig. 1). To identify the best way of describing the amplification efficiency, we compared the models using a corrected form of Akaike's Information Criterion (AIC) [20]. Model 3 AIC value was lower than model 1 and 2 ( $p < 10^{-15}$  and  $p < 10^{-4}$ , Wilcoxon paired test, respectively). Thus, the following two parameters sigmoid expression for  $E$  was used through this work,

$$E_n = \left( 1 + e^{\left( \frac{T_n - T_m}{b} \right)} \right)^{-1} \quad (2)$$

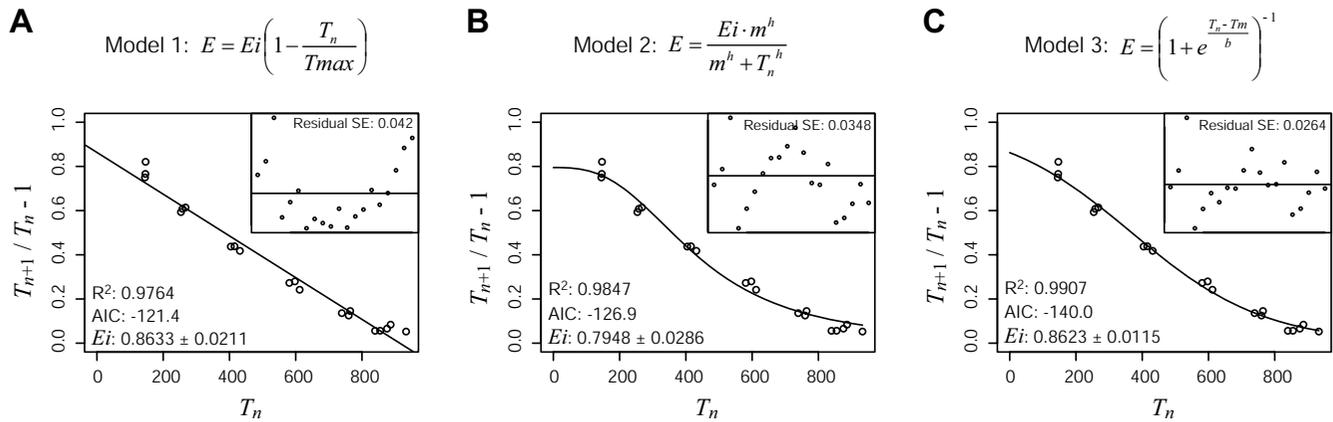
where  $T_m$  and  $b$  are parameters to be fitted by non-linear regression. There is no relationship between these parameters and kinetic parameters such as  $F_{max}$  and  $k$  [18], since our model is defined as a function of the product yield instead of the PCR cycle number.

The intrinsic amplification efficiency  $E_i$  (i.e. the putative amplification efficiency for a product (or template) amount equal zero)[5], is useful for  $T_0$  estimation (see below), and can be obtained from Eq. (2) as,

$$E_i = \left( 1 + e^{\left( \frac{T_m}{b} \right)} \right)^{-1} \quad (3)$$

### Model based estimation of the initial template amount ( $T_0$ )

In dsDNA binding dye protocols, real-time thermocyclers generate fluorescence intensity data. For most applications in which the dsDNA binding dye is in great excess, it can be assumed that the fluorescence intensity is proportional to the amount of double stranded DNA (product



**Figure 1**  
**Models for PCR amplification efficiency.** The effective amplification efficiency for each PCR cycle was calculated as  $T_{n+1} / T_n - 1$ , where  $T_n$  and  $T_{n+1}$  were the PCR product yield at cycles  $n$  and  $n+1$  respectively. Data points are the effective amplification efficiency vs. PCR product yield from a representative PCR reaction performed in triplicate. Lines are the fit of models 1 (A), 2 (B) and 3 (C) to the experimental data. Inserts are the residuals for each fit. The determination coefficient ( $R^2$ ), corrected Akaike's Information Criterion (AIC) and the best fit value for  $Ei \pm$  asymptotic standard error are shown in the graphs.  $Ei$  for model 3 was calculated from Eq. (3). Model 1 was fitted by linear regression, while models 2 and 3 were fitted by non-linear regression.

yield) [18]. However, it has been shown that this is not always the case [21]. Our model parameters are fitted with fluorescence intensity data, thus the product yield ( $T_n$ ) will be expressed in arbitrary fluorescent units, as the initial template amount ( $T_0$ ). Since the fluorescent dye will be in great excess when compared to the initial template amount, we can assume that  $T_0$  is proportional to the initial template amount, and so the semiquantitative comparison between different samples using  $T_0$  is valid.

The simplest way to estimate  $T_0$  is assuming that amplification efficiency at cycle  $CT$  ( $E_{CT}$ ) is very similar to  $Ei$ . This assumption is usually correct because  $CT$  is defined at small amounts of PCR product. Then, solving Eq. (1) for  $n = CT$  and assuming constant amplification efficiency we obtain:

$$T_{CT} = T_0 (1 + E)^{CT} \quad (4)$$

from which,

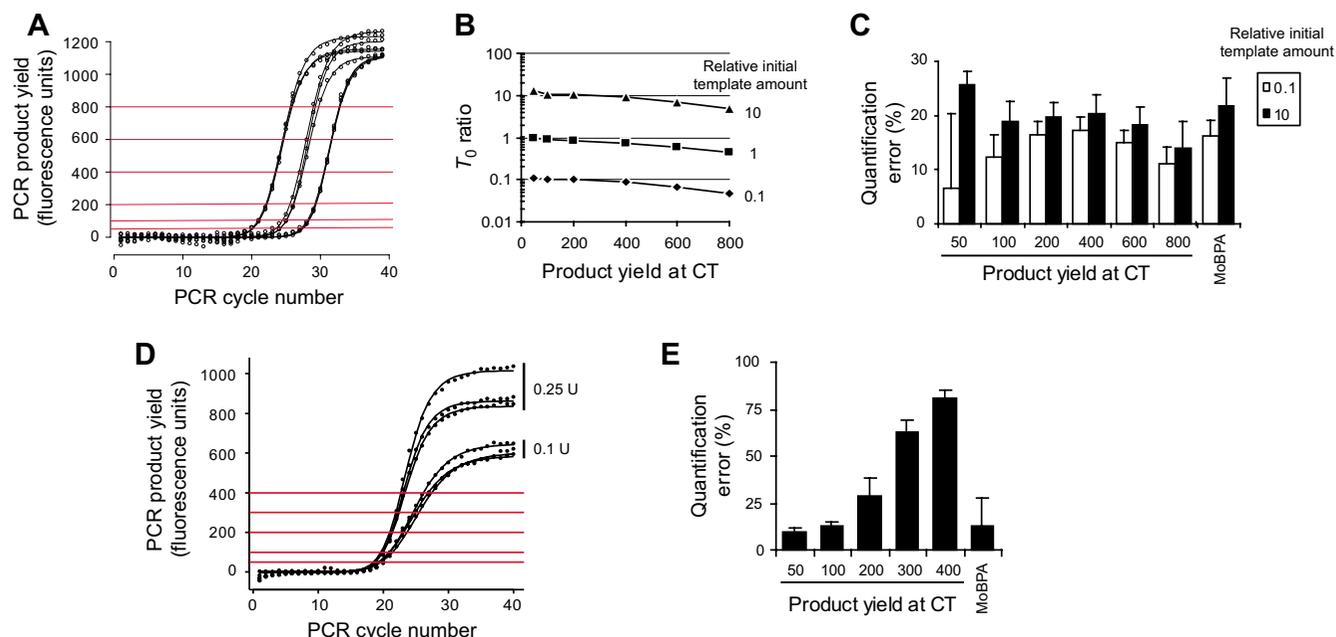
$$T_0 = T_{CT} (1 + E)^{-CT} \quad (5)$$

$Ei$  can be estimated from Eq. (3), and the efficiency parameters  $b$  and  $Tm$  can be obtained fitting Eq. (2) to real-time PCR data by non-linear regression (see methods for details). Then,  $T_0$  is calculated assuming  $E = Ei$ . Since Eq. (5) is a power function of the amplification efficiency, small errors in the estimation of  $Ei$  can be translated to strong errors in  $T_0$  estimation. These errors can be minimized using  $CT$  values as small as possibly, and estimat-

ing the amplification parameters with pooled data from replicates to improve  $Ei$  estimation accuracy. Replicates must be done by splitting a master-mix containing all components of the PCR to minimize the variability introduced by the operator.

To test this method, we amplified three serial dilutions of cDNA from mouse midbrain with  $\beta$ -actin specific primers. Pooled data from triplicate experiments were used for  $Ei$  estimation, which were 0.835, 0.852 and 0.856 for dilutions 10, 1 and 0.1, respectively. Estimations of  $T_0$  from this data were accurate and precise for  $CT$  values calculated from small PCR product yields, but led to different absolute quantification results for higher  $CT$  values (Fig. 2A and 2B). However, since amplification efficiencies are similar, the relative error of quantifications was independent of the product yield at which the  $CT$  was calculated (Fig. 2C). To test the performance of the  $CT$  method when amplification efficiencies are not the same, we amplified the same amount of mouse midbrain cDNA with  $\beta_2$ -microglobulin specific primers, but using different amounts of Taq DNA polymerase (Fig. 2D). Amplification efficiencies estimated with our approach from triplicate results were 0.855 and 0.915 for 0.1 and 0.25 units of Taq DNA polymerase, respectively. Then, an increase of only 7% in amplification efficiency led to errors in the quantifications that were directly related to the PCR product yield at which the  $CT$  was calculated (Fig. 2E).

We formulated an alternative more robust method for  $T_0$  estimation that does not rely on product threshold deter-



**Figure 2**  
**Effect of CT on the initial template amount estimation.** (A) Product yield vs. cycle number for the amplification of three serial dilutions (0.1, 1 and 10) of cDNA from mouse midbrain with  $\beta$ -actin specific primers performed in triplicate. Horizontal lines show the values for the product yield at which CT was calculated. (B) Ratios between each  $T_0$  determination and  $T_0$  estimated for dilution 1 at the smallest CT value.  $T_0$  was calculated assuming constant amplification efficiency and using different amounts of PCR product for the estimation of CT values. Data points are the mean for triplicates. (C) Relative error for the quantification of data presented in A. Bars are the relative error of quantification as percentage (mean  $\pm$  SEM for triplicates). (D) Real-time PCR amplification of the same mouse midbrain cDNA sample with  $\beta_2$  microglobulin specific primers using 0.1 and 0.25 units of Taq DNA polymerase. Horizontal lines show the values for the product yield at which CT was calculated. (E) Relative error for the quantification of data presented in D. Bars are the relative error of quantification as percentage (mean  $\pm$  SEM for triplicates).

minations. This alternative method is based on the fit of Eq. (1) and (2) to real-time PCR data by non-linear regression to obtain the best-fit estimators for the parameters  $b$ ,  $T_m$  and  $T_0$ . This approach does not assume constant amplification efficiency and estimates  $T_0$  from all the data points that fall into the exponential-linear growth phase instead of using only the product yield at cycle CT. Analysis of experimental results showed that  $T_0$  estimations were precise and accurate (Table 1). However, fitting the model to experimental data showed a strong correlation between  $T_0$  and both model 3 parameters  $b$  and  $T_m$ . This means that experimental data do not define the model unambiguously, leading to large uncertainty in the estimation of  $T_0$  (see standard error of  $T_0$  in Table 1). To overcome this problem, we propose a two-step procedure called Model-Based method for real-time PCR data Analysis (MoBPA). In the first step of the MoBPA procedure, we obtain good estimators for  $b$  and  $T_m$  parameters by fitting model 3 (Eq. 2) to experimental data. In the second step, we replace  $b$  and  $T_m$  in Eq. 2 with the values estimated in step 1, and then fit Eq. 1 to experimental data to

obtain a good estimator for  $T_0$ .  $T_0$  estimations by MoBPA were as precise and accurate as using the CT method for small PCR product yield-derived CT values (Fig. 2C and 2E), and the asymptotic standard errors were approximately 10 fold smaller (Table 1).

**Comparison of MoBPA with previous methods**

Next, we evaluated the performance of different methods for quantification when amplification efficiency for different samples is not the same. For this, we analysed two different datasets: 1) *in-silico* generated data of real-time PCR runs starting at the same initial template amount but differing in their intrinsic amplification efficiency; and 2) real experimental data obtained by amplifying the same amount of mouse midbrain cDNA with  $\beta_2$ -microglobulin specific primers, but with different amounts of Taq DNA polymerase (Fig. 2D).

Simulation data was generated *in-silico* from eqs. (1) and (2), using parameters that resemble real PCR runs. Results from amplification of mouse midbrain cDNA with  $\beta$ -actin

**Table 1: Estimation of model parameters.**

Dilution	A					B		
	$T_0$	SE	Mean $\pm$ SEM	Correlation		$T_0$	SE	Mean $\pm$ SEM
				$T_0 \times b$	$T_0 \times T_m$			
10	8.52	2.27	8.63 $\pm$	0.963	-0.97	12.4	0.29	12.17 $\pm$
10	7	2	0.98	0.963	-0.977	11.2	0.34	0.52
10	10.4	1.95		0.963	-0.969	12.9	0.36	
1	0.42	0.06		0.967	-0.962	0.981	0.018	
1	1.08	0.56	1 $\pm$ 0.32	0.964	-0.948	0.804	0.024	1 $\pm$ 0.12
1	1.51	0.74		0.965	-0.977	1.21	0.043	
0.1	0.09	0.04	0.073 $\pm$	0.968	-0.965	0.121	0.0032	0.12 $\pm$
0.1	0.065	0.015	0.0087	0.969	-0.974	0.111	0.0026	0.0028
0.1	0.063	0.023		0.968	-0.921	0.117	0.0028	

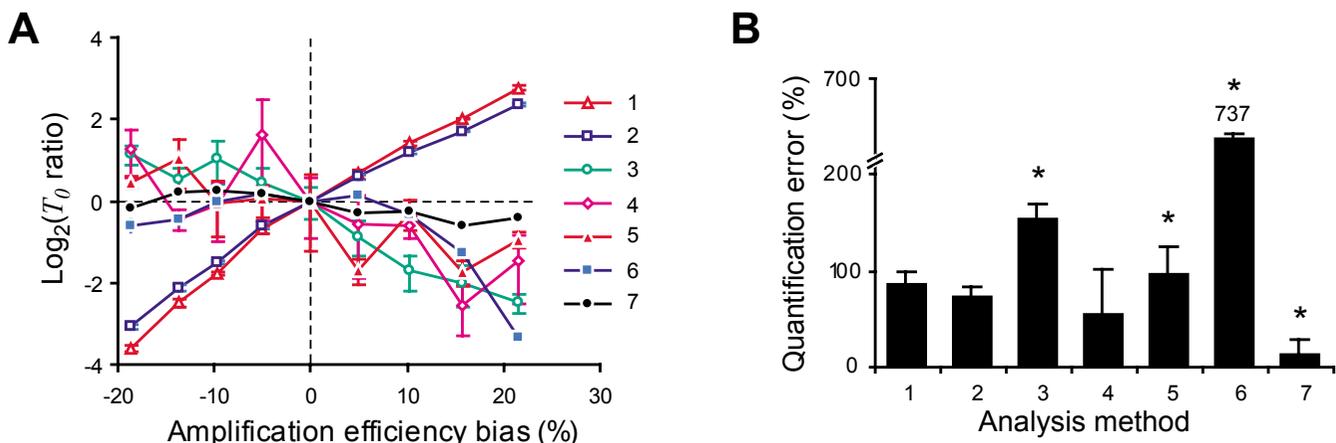
Estimation of  $T_0$ ,  $b$  and  $T_m$  by fitting Eq. (1) and (2) to experimental data (A), and estimation of  $T_0$  by fitting Eq. (1) and (2) to experimental data using  $b$  and  $T_m$  values previously obtained from the fit of Eq. (2) to experimental amplification efficiencies (B). Shown is the dilution of mouse midbrain cDNA used for each real-time PCR run, the best fit values for  $T_0$ , the asymptotic estimation of the standard error (SE), and the  $T_0$  mean value  $\pm$  standard error of the mean (SEM). For (A), the correlation between  $T_0$  and the other parameters estimated by non-linear regression is also shown.

and  $\beta_2$ -microglobulin specific primers were used to estimate plausible values for the simulation parameters. The analysis of these simulation results must be taken with care, because we used the same model for generating and fitting the data, so some overfitting is expected. However, we also evaluated the performance of the different methods on real PCR data. To obtain different amplification efficiency in our PCR runs, we used two different amount of DNA polymerase. Efficiencies estimated with our approach from triplicate results were 0.855 and 0.915 for 0.1 and 0.25 units of Taq DNA polymerase, respectively. The use of more than 0.25 units of DNA polymerase did not led to additional increments in amplification efficiency, while PCR with less than 0.1 units of DNA polymerase showed no product. We also tried to partially reduce the PCR amplification efficiency by adding  $Mg^{2+}$  chelating agents like EDTA, by increasing the amount of dNTPs, by lowering the amount of  $Mg^{2+}$ , and by adding known DNA polymerase inhibitors like phenol. However, we could not find the appropriate conditions to achieve the partial inhibition of DNA polymerase in a reproducible way.

Both *in-silico* and experimental real-time PCR data were analysed with different methods. The simpler and widespread used methods are based on  $CT$  value determinations, assuming constant amplification efficiency during the exponential phase of the PCR and among different samples [11,22]. Then, from Eq. (5) is easy to solve the initial template amount ratio between two samples as:

$$\frac{T_{0A}}{T_{0B}} = (1 + E)^{CT_B - CT_A} \tag{6}$$

where  $T_0$  is the initial template amount,  $E$  is the amplification efficiency, and  $CT_i$  the threshold cycle for each sample. The simplest form of threshold-based methods even assumes amplification efficiency equal 1. Thus, the ratio of initial template amounts between two samples will be  $2^{\Delta CT}$  [12]. Because of the exponential nature of this expression, amplification efficiencies below 1 will lead to unreliable quantifications. Amplification efficiency below 1 is frequent and indeed, most of our real-time PCR reactions showed amplification efficiencies between 0.7 and 0.95. To overcome this limitation, a dilution series based method for amplification efficiency estimation has been developed [14,23]. A curve is constructed by amplification of serial dilutions of one reference sample and plotting the resulting  $CT$  values against the base 10 logarithm of the dilution factor. Assuming constant amplification efficiency between dilutions and over the number of thermocycles required to reach  $CT$ , the amplification efficiency can be obtained from Eq. (4) as  $E = 10^{-1/slope} - 1$ . However, sample contamination with salt, phenol, chloroform, etc. may result in a lower-than-expected PCR efficiency [13,24,25]. Dilution of this sample will also dilute the contaminant decreasing its effect on the PCR reaction, thereby increasing the PCR efficiency with each dilution step. Indeed, we usually observe such dilution effect on PCR amplification efficiency, see for example data presented in Fig. 2A, in which estimated efficiency from the dilution series was 0.853, while estimations of the intrinsic amplification were 0.835, 0.852 and 0.856 for dilutions 100, 10 and 1, respectively. Amplification efficiency estimations from such dilution series will be inaccurate, introducing a bias in the quantification. Moreover, contaminants can quantitatively and qualitatively differ



**Figure 3**  
**Effect of amplification efficiency over the quantifications performed by different methods.** (A) *In-silico* generated PCR data with initial template amount  $T_0 = 0.001$  and different intrinsic amplification efficiencies ( $E_i$ ) ranging from 0.65 to 0.972 analysed by different methods (see below). Data points represent the base 2 logarithm of the ratio between  $T_0$  estimations from each simulated reaction and efficiency 0.8 ones vs. the amplification efficiency bias as mean  $\pm$  SEM of triplicates. (B) Analysis of experimental results by different methods (see below). Bars represent the error of quantifications as mean  $\pm$  SEM for triplicates. Bars marked with (\*) are under-estimations, conversely, the rest of the bars are over-estimations. Method 6 underestimated  $T_0$  by 737%, note that it is out of scale in the graph. Data was analysed with the CT method assuming constant amplification efficiency equal to 1 {1}; with the CT method assuming constant amplification efficiency equal to 0.8 for *in-silico* data, or 0.855 for experimental data {2}; assuming constant amplification efficiency that was estimated from two threshold values {3} [15]; using the assumption-free analysis proposed by Ramakers et.al. {4} [13]; using the standardized determination of PCR efficiency from single reaction proposed by Tichopad et.al. {5} [16]; using the sigmoid model proposed by Liu et. al. {6} [17]; and with our model based real-time PCR analysis method (MoBPA) {7}.

between samples, thus, the assumption of equal amplification efficiencies between samples can lead to strong errors in quantifications.

Analysis of *in-silico* PCR results showed that the bias in quantifications by these methods is directly related to the difference in amplification efficiency between templates. In such a way, a 20% difference in amplification efficiency led to 11 and 7.4 fold bias in quantifications depending on whether efficiency 1 or reference sample efficiency (0.8) was used for calculations (Fig. 3A). Accordingly, the quantification of  $\beta 2$ -microglobulin was 86% and 74% overestimated by the use of efficiency 1 or 0.855 in threshold-based methods, respectively (Fig. 3B). Despite the poor accuracy of these methods,  $T_0$  estimation from replicated samples was very precise (mean CV: 6%, range: 0.75–12%), possibly because these approaches are only affected by errors in CT estimation. Similar precision was obtained with experimental real-time PCR results (Fig. 3B), and it was previously reported by other authors [26].

Four different methods for single reaction amplification efficiency estimation have been proposed. All of them use the kinetic data generated by real-time PCR cyclers, they do not assume equivalent amplification efficiency among samples and have the additional advantage that no dilu-

tions curve is needed. Three of them assume constant amplification efficiency during the exponential phase of the PCR and estimate it from the few data points that fall into this phase [13,15,16]. The simplest one is based on the product yield measured at two thresholds along the exponential phase of the PCR, from which the amplification efficiency is estimated as  $E = (T_1/T_2)^{1/(CT_1-CT_2)} - 1$  [15]. Then, the initial template amount is calculated from Eq. (5). Based on the same model for the PCR reaction, Ramakers et. al. proposed the use of all the points of the background subtracted and logarithm transformed PCR data that fall into a "window-of-linearity", for amplification efficiency estimation by linear regression [13]. Recently, Tichopad et.al. introduced a similar approach based on a new statistical method for the identification of the exponential phase and estimation of amplification efficiency by non-linear regression [16]. Analysis of *in-silico* generated results with different amplification efficiencies showed better accuracy of these three approaches when compared with the threshold-based methods, since they introduced only 2.3 – 4.8 fold bias in quantifications (Fig. 3A). However, since  $T_0$  determinations depend on both efficiency and CT estimations, precision in quantification of replicates was very poor (mean CV: 58%, range: 19–134%). Indeed, quantification of samples that differ less than 10% in amplification efficiency were more accurate

**Table 2: Analysis of real-time PCR results with similar amplification efficiency among samples. Data represent the quantification of dilutions 0.1 and 10 as the mean ± SEM for 12 experiments performed in triplicate.**

Dilution	Analysis method						
	1	2	3	4	5	6	7
0.1	0.084 ± 0.003	0.11 ± 0.0033	0.57 ± 0.30	0.37 ± 0.19	0.75 ± 0.45	0.71 ± 0.63	0.12 ± 0.022
1	1 ± 0.019	1 ± 0.017	1 ± 0.11	1 ± 0.20	1 ± 0.12	1 ± 0.10	1 ± 0.018
10	14 ± 0.680	10 ± 0.35	242 ± 206	83 ± 52	15 ± 5.5	48 ± 1.0	10 ± 1.1

- (1) CT method assuming constant amplification efficiency equal to 1 [12].
- (2) CT method with amplification efficiency estimated from a dilution series [14].
- (3) Amplification efficiency estimated at two product yield thresholds [15].
- (4) Amplification efficiency estimated with LinRegPCR software [13].
- (5) Amplification efficiency estimated with the Tichopad et.al. approach [16].
- (6) Amplification efficiency estimated from the model proposed by Liu et.al. [17].
- (7) Our model based real-time PCR analysis method (MoBPA).

when analysed with the Pfaffl or efficiency 1 threshold-based methods [12,14], most likely due to the poor precision of single reaction based efficiency estimations. These observations were further confirmed with experimental real-time PCR data (Fig. 3B).

Recently, Liu et.al. proposed a three parameters sigmoidal function for modelling the whole kinetic process of real-time PCR. Then, the initial template amount is estimated after fitting the sigmoidal model to background subtracted real-time PCR results [17]. Analysis of *in-silico* PCR results showed a good accuracy and moderate precision of this method for most PCR simulations, with a fold bias < 1.4 and a mean CV of 12% (range: 5–21%) (Fig. 3A). However, the model presented systematic deviations from the data that were particularly noted with amplification efficiencies above 0.9, both with experimental and *in-silico* generated data. These deviations had a strong effect on  $T_0$  estimation, leading to unreliable quantifications (Fig. 3B). Similar deviations were recently described in the plateau phase by Rutledge, who showed that elimination of these data points improve the accuracy of this method [18].

Finally, quantification from the same *in-silico* and experimental real-time PCR results with our model based approach showed the highest accuracy and precision. Analysis of *in-silico* data showed that 20% difference in amplification efficiency between samples only led to 0.31 fold bias, with a mean CV of only 2.6% (range: 0.8–4.2%) (Fig. 3A). Moreover, an increase of 7% in experimental real-time PCR amplification efficiency only led to 13% under-estimated quantification (mean CV: 10.5%) (Fig. 3B).

To test the reliability of the different methods in conditions of similar amplification efficiency between samples, we amplified three dilutions of mouse midbrain cDNA with  $\beta$ -actin specific primers. Among threshold-based methods, the use of amplification efficiency estimated for

each experiment from the dilution series [14] produced the best results (Table 2). Most methods for the estimation of amplification efficiency from single reaction results introduced strong deviations in the quantification. Only the approach suggested by Tichopad et.al. [16] showed moderate deviations (Table 2). Similar observation have been reported by Peirson et.al., who showed that estimation of amplification efficiency from single reactions data introduces systemic errors and increases the assay noise [27], suggesting that these methods should be used only when the amplification efficiency between samples is not the same. A statistical method for the detection of samples with dissimilar efficiencies, called kinetic outlier detection (KOD), was recently developed [28]. KOD method can be used to decide whether to use single reaction or global amplification efficiency estimated from dilution series for quantifications. To note, our model based approach gave the most accurate and precise results, which were similar to the ones produced by Pfaffl approach [14] (Table 2), indicating that MoBPA produce reliable quantifications no matter whether samples are amplified with similar or different efficiency.

Our model assumes that the signal is proportional to the amount of product, which is often the case for SYBR-Green I real-time PCR performed with saturating concentrations of dye. In such conditions centrally symmetric amplification curves are expected. However, in TaqMan applications, where the Taq DNA polymerase digests a probe labelled with a fluorescent reporter and quencher dye, the signal diverges from the product resulting in non-symmetric amplification curves (Supplementary Fig. 1A in Additional file 1) [8]. To test whether our approach is also suitable for TaqMan data analysis, we quantified  $\beta$ -actin and hypoxanthine phosphoribosyl transferase (HPRT) in 1/10 dilutions of the same cDNA sample using TaqMan probes. The error of quantifying a 10-fold concentrated sample was only slightly higher using TaqMan when compared to SYBR-Green I (Fig. 2C and Supplementary Fig. 1B in Additional file 1). These results suggest

that MoBPA is robust enough to deal with non-symmetric amplification curves, possibly because it makes use of only the data points that fall into the exponential-lineal growth phases of the PCR.

## Conclusion

Following the emergence of functional genomic methodologies, the development of high-throughput methods for microarray-derived data validation is becoming indispensable. Nevertheless, high-throughput quantification by real-time PCR is difficult to achieve, primarily due to deficiencies of the threshold-based methodologies, which require reliable estimation of amplification efficiencies [27]. In addition to the technical challenge of generating dilution curves for each target sequence, these methods assume similar amplification efficiencies between samples, and this assumption has been reported to be invalid [13,24,25]. Here we introduce a new method for the analysis of real-time PCR results termed: Model Based method for real-time PCR data Analysis (MoBPA). The new method produced the most accurate and precise quantifications when compared to previous methods regardless of whether samples were amplified with similar or different amplification efficiencies. Moreover, no calibration curve is needed for quantifications, since amplification efficiency is estimated directly from real-time PCR data. MoBPA provides many of the fundamental capabilities required for fully automated high-throughput quantification [18], including routine assessment of amplification efficiency within individual samples and no need of PCR-generated standard curves. In addition, quantitative data can be easily derived from arbitrary fluorescence units, simply applying a calibration factor that relates fluorescence to DNA mass [18]. In conclusion, MoBPA combines the well accepted exponential model of the PCR reaction with a sigmoid description of amplification efficiency to obtain a powerful method for real-time PCR data analysis that expand the applicability of real-time PCR to fully automated high-throughput applications.

## Methods

### RNA extraction and reverse transcription

Mice were killed by cervical dislocation and brains removed. Midbrains were dissected, snap frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until RNA extraction. Total RNA was isolated using TRIzol Reagent (Invitrogen, MD), genomic DNA contaminant was removed using DNase I (Ambion, Inc., TX), and mRNA was purified by Micro-Poly(A)Pure kit (Ambion, Inc., TX). First-strand complementary DNA was synthesized at  $42^{\circ}\text{C}$  by priming with oligo-dT<sub>12-18</sub> (Invitrogen, MD) and using SuperScriptII reverse transcriptase according to the protocol provided by the manufacturer (Invitrogen, MD).

### Polymerase chain reaction

PCR amplifications were obtained using an Icyler IQ Real-Time PCR Detection System (BioRad, CA). cDNA samples were assayed by triplicate. PCR reactions were performed in a final volume of 25  $\mu\text{l}$  containing 1  $\mu\text{l}$  of cDNA, 2.5  $\mu\text{l}$  of the reaction buffer (200 mM Tris-HCl pH 8.4, 500 mM KCl), 3 mM  $\text{MgCl}_2$ , 0.3 mM of dNTPs mix, 0.2 nM of each primer, 0.3  $\times$  SYBR-Green I (Molecular Probes, OR), 100  $\mu\text{g/ml}$  BSA, 0.25  $\mu\text{l}$  ROX Reference Dye (Invitrogen, MD), 1% glycerol, and 1.25 U of Taq Platinum Polymerase (Invitrogen, MD). The primer sequences used were:  $\beta$ 2-microglobulin, sense: TGA CCG GCT TGT ATG CTA TC and antisense: CAG TGT GAG CCA GGA TAT AG;  $\beta$ -actin, sense: CAA TGT GGC TGA GGA CTT TG and antisense: ACA GAA GCA ATG CTG TCA CC. PCR was performed as follows: one initial cycle of  $94^{\circ}\text{C}$  for 2.5 min, 40 cycles of  $94^{\circ}\text{C}$  for 30 sec,  $58^{\circ}\text{C}$  for 30 sec, and  $72^{\circ}\text{C}$  for 15 sec. TaqMan real-time PCRs were performed as the SYBR-Green I assays, but with no addition of SyBrGreen I and with the following primers and probes:  $\beta$ -actin sense: AGA AAA TCT GGC ACC ACA CC, antisense: CAG AGG CGT ACA GGG ATA GC, and probe: ACC GCG AGA AGA TGA CCC AGA TCA T; HPRT sense: AGA CTG AAG AGC TAT TGT AAT, antisense: CAG CAA GCT TGC GAC CTT GAC, and probe: TGC TTT CCT TGG TCA GGC AGT ATA.

### Analysis of real-time PCR data

ROX base-line corrected real-time PCR results were analysed with different methods as described [12-17] using the R-System v2.2.0 [29] or the LinRegPCR software [13]. For the LinRegPCR software, we used the default fit option, which iteratively searches for lines consisting in 4 to 6 data points with the highest  $R^2$  value. Then we inspected the fit for each PCR curve and manually corrected the windows of linearity when needed.

The implementation of our method comprise the following steps: 1) Identification of ground, exponential, lineal growth and plateau phases, 2) background subtraction, 3) effective amplification efficiency estimation and fitting eq (2) to experimental data, 4) initial template amount calculation.

The ground phase was identified as described [16], using a p-value cut-off of 0.01, while the beginning and end of the lineal growth phase was identified as the second derivative maximum and minimum, respectively, of a four parameters sigmoid function (eq. 7) fitted to the data (Supplementary Figure 2A in Additional file 1).

$$\gamma = \frac{(mi - ma)c^h}{c^h + x^h} \quad (7)$$

Here,  $\gamma$  is the PCR product yield (fluorescence units),  $x$  is the cycle number,  $mi$  and  $ma$  are the signal-offset and sat-

uration value, respectively,  $c$  is the point of inflexion and  $h$  is the exponent parameter. Non-linear least square regressions were performed with a Gauss-Newton algorithm implemented in the `nls` function of the R-system software.

The background level was calculated as the last ground phase data point, which in turn was estimated by a linear regression over the last five data points of this phase [16].

The effective amplification efficiency for each PCR cycle ( $E_n$ ) was solved from Eq. (1) as  $E_n = T_{n+1}/T_n - 1$  [5] and calculated using the background subtracted data. Then, the parameters  $b$  and  $T_m$  were estimated by fitting Eq (2) to the experimental effective amplification efficiency by non-linear regression. We only used data points that fall into the exponential and linear growth phase of the amplification curve, because efficiency calculated at the early ground phase and late plateau phase showed strong dispersion due to experimental background noise or showed no information, respectively (Supplementary Figure 2 in Additional file 1). Finally, the initial template amount  $T_0$  was estimated by fitting our discrete model, shown below in R-code:

```
T <- function(x, t0, b, Tm) {
  output <- NULL
  for (i in 1:length(x)) {
    t00 <- t0
    for (ii in 1:x[i]) t00 <- t00*(1+1/
(1+exp((t00-Tm)/b)))
    output <- c(output, t00)
  }
  return(output)
}
```

to background subtracted experimental data by the `nls` function of R-system, using the values for the parameters  $b$  and  $T_m$  estimated previously.

Our method was implemented in the R-System. The source code and Windows binary of MoBPA package are available for non-commercial research use (see Additional files 2 and 3).

### Quantification error and Akaike's Information Criterion

For calculating the quantification error, we defined a reference sample ( $RS$ ) and a target sample ( $TS$ ), which are derived from the same original sample and differ by a known dilution factor or PCR amplification efficiency. We calculated the relative bias as  $RE = (TS - RS)/RS$ , and the quantification error as  $SQE = RE * 100$  for  $RE \geq 0$ , or  $SQE = -100 * RE / (1 + RE)$  for  $RE < 0$ .

For comparing alternative models we used a corrected form of the Akaike's Information Criterion ( $AIC$ ), defined as:

$$AIC = N \log \left( \frac{SS}{N} \right) + 2K + \frac{2K(K+1)}{N-K-1} \quad (7)$$

where  $N$  is the number of data points,  $K$  is the number of parameters fitted by the regression plus one, and  $SS$  is the sum of the square of the vertical distances of the points from the curve [20].

### In silico generation of PCR data

Simulation data was generated *in silico* from eqs. (1) and (2), using parameters that resemble real PCR runs. Results from amplification of mouse midbrain cDNA with  $\beta$ -actin and  $\beta_2$ -microglobulin specific primers were used to estimate plausible values for the simulation parameters. In such a way, the initial template amount in fluorescent arbitrary units ( $T_0$ ) was set to  $10^{-3}$ , and the product amount at the PCR plateau phase ( $T_{max}$ ) was set to 1200. The amplification parameters  $b$  and  $T_m$  were solve from eq. (2) and (3) as,

$$b = \frac{T_n}{\log \frac{E_n - E_i E_n}{E_i - E_i E_n}}, T_m = b \log \left( \frac{1}{E_i} - 1 \right),$$

Then, for  $T_n = T_{max}$ ,  $E_n$  tends to zero, so we calculated approximated values for  $b$  and  $T_m$  using  $T_n = T_{max} = 1200$  and  $E_n = 0.001$ .

Zero mean, normally distributed random noise was added to the *in-silico* results. The background mean ( $Bg$ ) and intra- and inter-replicates standard deviation (*intraSD* and *interSD*) were estimated from 48 triplicate samples to be 625, 17.94 and 19.06 respectively. The mean standard deviation for fluorescence intensity in the early ground phase of each PCR reaction (*mSD*), also estimated from experimental data, was 4.9. We found no correlation between the standard deviation of fluorescence intensity in the early ground phase of each PCR reaction and  $Bg$  values, thus, random noise generated from a normal distribution with mean 0 and standard deviation *mSD* was added to *in-silico* PCR data regardless of the background fluorescence level.

## Authors' contributions

MJA carried out the design of the study, participated in data analysis and drafted the manuscript. GJV carried out the real time PCR. MCS participated in data collection and analysis. OLP and FJP participated in the design of the study and critically revised the manuscript. All authors read and approved the final manuscript.

## Additional material

### Additional File 1

supplementary figures 1 and 2.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-85-S1.pdf>]

### Additional File 2

Source files for the R-system package MoBPA.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-85-S2.gz>]

### Additional File 3

Windows binary files for the R-system package MoBPA.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2105-8-85-S3.zip>]

## Acknowledgements

We thank Dr. Andrea Califano for helpful comments on the manuscript. We specially thank Cynthia Serra and Daniela Celi for technical assistance. We are in debt with María Romina Girotti and Andrea Sabina Llera for facilitating their TaqMan PCR results.

## References

- Rappolee DA, Wang A, Mark D, Werb Z: **Novel method for studying mRNA phenotypes in single or small numbers of cells.** *J Cell Biochem* 1989, **39(1)**:1-11.
- Gilliland G, Perrin S, Blanchard K, Bunn HF: **Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction.** *Proc Natl Acad Sci USA* 1990, **87**:2725-2729.
- Becker-André M, Hahlbrock K: **Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY).** *Nucleic Acids Res* 1989, **17(22)**:9437-9447.
- Wang AM, Doyle MV, Mark DF: **Quantitation of mRNA by the polymerase chain reaction.** *Proc Natl Acad Sci USA* 1989, **86**:9717-9721.
- Alvarez MJ, Depino AM, Podhajcer OL, Pitossi FJ: **Bias in estimations of DNA content by competitive polymerase chain reaction.** *Anal Biochem* 2000, **287**:87-94.
- Ozawa T, Tanaka M, Ikebe S, Ohno K, Kondo T, Mizuno Y: **Quantitative determination of deleted mitochondrial DNA relative to normal DNA in parkinsonian striatum by a kinetic PCR analysis.** *Biochem Biophys Res Commun* 1990, **172(2)**:483-489.
- Higuchi R, Fockler C, Dollinger G, Watson R: **Kinetic PCR analysis: real-time monitoring of DNA amplification reactions.** *Biotechnology* 1993, **11(9)**:1026-1030.
- Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K: **Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization.** *PCR Methods Appl* 1995, **4(6)**:357-362.
- Morrison TB, Weis JJ, Wittwer CT: **Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification.** *Biotechniques* 1998, **24(6)**:954-962.
- Chuaqui RF, Bonner RF, Best CJ, Gillespie JW, Flaig MJ, Hewitt SM, Phillips JL, Krizman DB, Tangrea MA, Ahram M, Linehan WM, Knezevic V, Emmert-Buck MR: **Post-analysis follow-up and validation of microarray experiments.** *Nat Genet* 2002, **32 Suppl**:509-514.
- Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP: **Continuous fluorescence monitoring of rapid cycle DNA amplification.** *Biotechniques* 1997, **22(1)**:130-138.
- Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods* 2001, **25(4)**:402-408.
- Ramakers C, Ruijter JM, Deprez RH, Moorman AF: **Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data.** *Neurosci Lett* 2003, **339(1)**:62-66.
- Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic Acids Res* 2001, **29(9)**:e45.
- Liu W, Saint DA: **A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics.** *Anal Biochem* 2002, **302(1)**:52-59.
- Tichopad A, Dilger M, Schwarz G, Pfaffl MW: **Standardized determination of real-time PCR efficiency from a single reaction set-up.** *Nucleic Acids Res* 2003, **31(20)**:e122.
- Liu W, Saint DA: **Validation of a quantitative method for real time PCR kinetics.** *Biochem Biophys Res Commun* 2002, **294(2)**:347-353.
- Rutledge RG: **Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications.** *Nucleic Acids Res* 2004, **32(22)**:e178.
- Mehra S, Hu WS: **A kinetic model of quantitative real-time polymerase chain reaction.** *Biotechnol Bioeng* 2005, **91(7)**:848-860.
- Motulsky HJ, Christopoulos A: **Fitting models to biological data using linear and nonlinear regression. A practical guide to curve fitting.** San Diego, CA, GraphPad Software Inc (Available at <http://www.graphpad.com/manuals/Prism4/RegressionBook.pdf>); 2003.
- Zipper H, Brunner H, Bernhagen J, Vitzthum F: **Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications.** *Nucleic Acids Res* 2004, **32(12)**:e103.
- Freeman WM, Walker SJ, Vrana KE: **Quantitative RT-PCR: Pitfalls and potential.** *BioTechniques* 1999, **26(1)**:112-125.
- Rutledge RG, Côté C: **Mathematics of quantitative kinetic PCR and the application of standard curves.** *Nucleic Acids Res* 2003, **31(16)**:e93.
- Tichopad A, Didier A, Pfaffl MW: **Inhibition of real-time RT-PCR quantification due to tissue-specific contaminants.** *Molecular and Cellular Probes* 2004, **18(1)**:45-50.
- Stahlberg A, Aman P, Ridell B, Mostad P, Kubista M: **Quantitative real-time PCR method for detection of B-lymphocyte monoclonality by comparison of kappa and lambda immunoglobulin light chain expression.** *Clinical Chemistry* 2003, **49(1)**:51-59.
- Rajeevan MS, Ranamukhaarachchi DG, Vernon SD, Unger ER: **Use of real-time quantitative PCR to validate the results of cDNA array and differential display PCR technologies.** *Methods* 2001, **25(4)**:443-451.
- Peirson SN, Butler JN, Foster RG: **Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis.** *Nucleic Acids Res* 2003, **31(14)**:e73.
- Bar T, Stahlberg A, Muszta A, Kubista M: **Kinetic Outlier Detection (KOD) in real-time PCR.** *Nucleic Acids Res* 2003, **31(17)**:e105.
- The R Project for Statistical Computing [<http://www.r-project.org/>]