J Mol Med
DOI 10.1007/s00109-006-0097-6

REVIEW

Quantitative real-time RT-PCR data analysis: current concepts and the novel "gene expression's C_T difference" formula

Jan H. Schefe · Kerstin E. Lehmann · Ivo R. Buschmann · Thomas Unger · Heiko Funke-Kaiser

Received: 17 May 2006 / Accepted: 20 July 2006

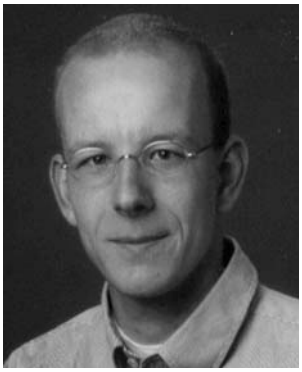
Abstract For quantification of gene-specific mRNA, quantitative real-time RT-PCR has become one of the most frequently used methods over the last few years. This article focuses on the issue of real-time PCR data analysis and its mathematical background, offering a general concept for efficient, fast and precise data analysis superior to the commonly used comparative C_T ($\Delta\Delta C_T$) and the standard curve method, as it considers individual amplification efficiencies for every PCR. This concept is based on a novel formula for the calculation of relative gene expression ratios, termed GED (Gene Expression's C_T Difference) formula. Prerequisites for this formula, such as real-time PCR kinetics, the concept of PCR efficiency and its determination, are discussed. Additionally, this article offers some technical considerations and information on statistical analysis of real-time PCR data.

Keywords Quantification · Expression ratio · PCR efficiency · TaqMan · SYBR Green

J. H. Schefe · T. Unger · H. Funke-Kaiser (✉)
Center for Cardiovascular Research (CCR)/
Institute of Pharmacology and Toxicology,
Charité-Universitätsmedizin Berlin,
Hessische Strasse 3-4,
10115 Berlin, Germany
e-mail: heiko.funke-kaiser@charite.de

K. E. Lehmann · I. R. Buschmann
Center for Cardiovascular Research (CCR)
Department of Cardiology, Experimental and Clinical
Arteriogenesis, Charité-Universitätsmedizin Berlin,
Berlin, Germany

JAN H. SCHEFE studies medicine at the Charité-Universitätsmedizin Berlin. He is supported by scholarships of the Studienstiftung des deutschen Volkes and the Deutsches Institut für Bluthochdruckforschung (DIB). He is currently working in the Center for Cardiovascular Research (CCR)/ Institute of Pharmacology at the Charité-Universitätsmedizin Berlin. His research interests include signal transduction mechanisms and gene regulation.



HEIKO FUNKE-KAISER studied medicine at the Universities of Düsseldorf and Berlin, Germany. He received his MD degree at the Institute of Clinical Pharmacology and Toxicology of the Freie Universität Berlin. After a specialization in Pharmacology and Toxicology, he currently works in the Center for Cardiovascular Research (CCR)/ Institute of Pharmacology at the Charité-Universitätsmedizin Berlin. His research focus is on eukaryotic gene regulation.

Introduction

The quantification of gene-specific mRNA expression is one of the major issues in life science and basic clinical research. Since its establishment in the 1990s [1–4], quantitative real-time RT-PCR (qPCR) has been widely used, allowing fast, accurate and sensitive mRNA quanti-

fication with a high throughput of samples. Unfortunately, the issue of real-time PCR data analysis is often underestimated by researchers.

Two approaches of data analysis—the comparative C_T (cycle threshold) method (also known as $\Delta\Delta C_T$ method [5]) and the standard curve method [6]—are commonly used, but both suffer from major limitations, as discussed later in this article.

In the past years several authors have published approaches for enhanced qPCR data analysis [7–10]. Based on these approaches, we present a novel general strategy, which enables researchers to perform their qPCR data analysis in an efficient, fast and precise manner.

After introducing the background of qPCR kinetics and the concept of PCR efficiency and its determination, the calculation of relative gene expression ratios considering individual amplification efficiencies for every PCR—according to our novel GED (Gene Expression's C_T Difference) formula—is derived and explained. Furthermore, our article is arranged to include several figures containing further supporting information on, e.g. technical issues and statistical analysis of qPCR data.

Basic principle

Before introduction of the qPCR technique, conventional RT-PCR [11, 12] was widely used for gene-specific mRNA quantification. This method of end-point PCR product analysis is not able to determine the initial quantity of template molecules for a gene-specific PCR, because at the end of amplification cycles the amount of amplicon depends not only on the input amount but also technical variations occurring during the reaction [13]. This can be compensated for by the competitive PCR approach based on an internal, co-amplified standard, which represents a mutated form of the amplicon [14, 15]. Due to its labor-intensity, this method has lost significance in the last years.

qPCR offers the opportunity to observe the amplification kinetics of a PCR in "real time" via accumulation and measurement of specific fluorescence signals with each cycle [1, 3, 4].

Generally, a PCR consists of four different kinetic stages [16, 17], which are basically similar to the well-known bacterial growth kinetics as observed by Monod in 1949 [18]. First, there is a lag phase where exponential amplification is already ongoing within the PCR tube, but no fluorescence signal above the background level is measurable. Secondly, in the logarithmic (*log*) phase the exponential growth of PCR product—ideally, there is a doubling of PCR product every cycle—is measureable as fluorescent signal. Thirdly, in the retardation phase, accumulation of PCR inhibiting factors and loss of enzyme and substrates for the PCR decelerates the reaction. Fourthly, the PCR reaches a steady state in the stationary phase, and no more amplicons are produced [16, 19]. For real-time PCR data analysis, the second phase—the *log* phase, with its measurable exponential growth conditions—is crucial.

The PCR kinetics in this stage can be described with the following exponential equation describing the *log* phase:

$$R_n = R_0 \cdot (1 + E)^n \quad (1)$$

with R_n and R_0 being the amount of fluorescence signal (in arbitrary units and proportional to the amount of DNA amplicons) after 0 (R_0) or n (R_n) cycles, respectively (see Fig. 1). The efficiency E of the reaction ($0 \leq E \leq 1$) is defined in the following section.

C_T values and efficiency

After correct setup, run and technical quality control of a qPCR (see Fig. 2), two parameters for data analysis must be determined for each well: the C_T value and the PCR efficiency E.

The C_T value is defined, according to the so-called fit point method [20], as a fractional number of cycles, where the PCR kinetic curve (see Fig. 1) reaches a user- or program-defined threshold amount of fluorescence. This intersection point must be set in the exponential phase of the curve, i.e. above the background level and before reaching the retardation phase. This can be visualized in a half-logarithmic (*y*-axis *log*-scaled) plot of the kinetic curve, in which the exponential phase corresponds to the linear part of this graph (see Fig. 1).

The PCR efficiency E is a major issue in qPCR data analysis. In a perfect situation, we would always achieve perfect PCR amplification. A PCR with a perfect setup would theoretically double the gene-specific amplicons from cycle to cycle, which would be equivalent to E=1=100%. In fact, E, empirically determined, is generally between 0.65 and 0.9 (65 and 90%) [21] and therefore a hardly reproducible parameter for each PCR, even though using identical reaction setups with identical templates. Several factors, such as phenol, ethanol, haemoglobin, heparin and even the reverse transcriptase, are known to inhibit PCR efficiency [22, 23].

Given identical initial copy numbers (R_0), the PCR efficiency E defines the C_T value and the resulting gene expression ratios, as these are calculated on the basis of the C_T values [8]. We may illustrate this problem as follows:

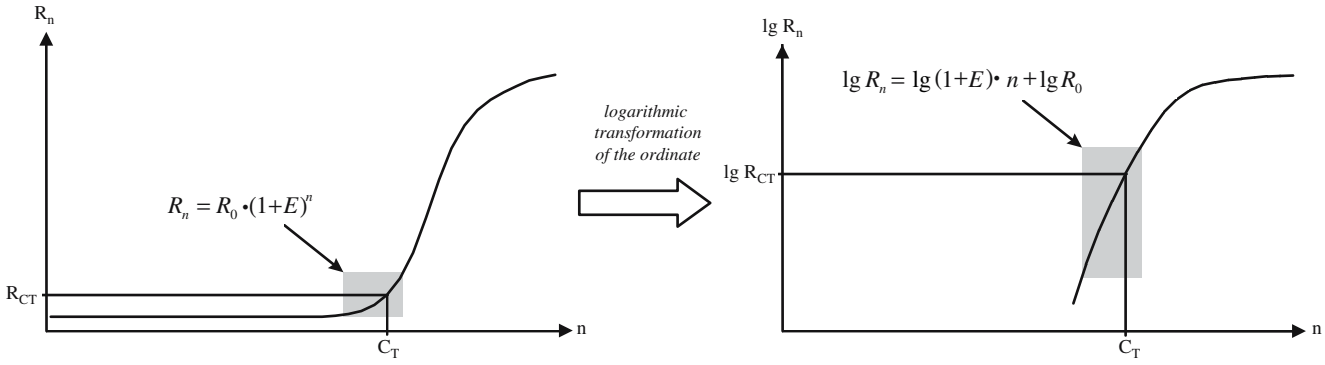


Fig. 1 qPCR amplification plot. Both *graphs* show a typical PCR kinetic curve with R_n plotted against n before and after logarithmic transformation of R_n; the y axis is linear (*left*) and *lg* scaled (*right*). The *gray squares* indicate the exponential regions, where the threshold for C_T value determination must be set. Within these regions of the curve the PCR kinetics obey the indicated formulas. n number of cycles, R_n fluorescence amount after n cycles, R_{CT} fluorescence amount after C_T cycles (identical to fluorescence threshold), R_0 initial fluorescence amount, E PCR efficiency

Fig. 2 Practical considerations—the DOs and DON'Ts

Ensure good primer and probe design. The use of primer and probe design software is highly recommended (e. g., *Primer3* [http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi] and *JaMBW* [http://www.bioinformatics.org/JaMBW/]).

- **RNA preparation and cDNA synthesis are crucial for gene quantification.** All samples should be treated according to identical protocols and – in the optimal case – in parallel. A DNase treatment must be included [32]. Check the quality and concentration of RNA by conventional electrophoresis and spectrophotometry or with devices such as Agilent's Bioanalyzer.
- **Include *NTC*s and *NAC*s in your qPCR runs.** *NTC*s are no-template controls where water instead of cDNA is included in the reaction mixture; this checks for purity of the PCR components. *NAC*s are no-amplicon controls, also known as reverse transcriptase-minus (RT-) controls, where a mock reverse transcription (RT) reaction with all components except the reverse transcriptase serves as template; this checks for genomic contamination and purity of the RT and PCR components.
- **Use an appropriate normalization control.** The use of an invariant endogenous control (housekeeping gene) in the assay can correct for minor sample-to-sample variations in cDNA quantity [31]. There are various options, common choices are *18S rRNA*, *GAPDH, cyclophilin* or *β-Actin*. Unregulated expression of these genes under the chosen experimental conditions is a major problem in this context, and invariance must be ensured [32; http://normalisation.gene-quantification.info]. In many cases the use of a normalization index (2 or more endogenous control genes) may be useful [33].
- **Pay attention to the technical analysis of your qPCR run.** In case of a *SYBR Green I* qPCR assay, a melting (dissociation) curve analysis should be performed yielding only one sharp peak in the first derivative plot. Ensure accurate baseline and threshold setting.

First of all, we convert Eq. (1) into a form in which the C_T value may be calculated based on known fluorescence amounts R_0 and R_{CT} (R_{CT} is the fluorescence amount at the C_T value) and the PCR efficiency E.

$$\begin{aligned} R_n = R_0 \cdot (1+E)^n \Rightarrow R_{C_T} &= R_0 \cdot (1+E)^{C_T} \\ \Leftrightarrow \lg R_{C_T} &= \lg R_0 + C_T \cdot \lg(1+E) \\ \Leftrightarrow C_T &= \frac{\lg R_{C_T} - \lg R_0}{\lg(1+E)} \end{aligned} \quad (2)$$

Now we compare two cDNAs (I and II) containing the same initial amount of template molecules (R_0=10), but with different Es [E(cDNA$_{\mathrm{I}}$)=0.85 and E(cDNA$_{\mathrm{II}}$)=0.70] reflecting improper—due to variations—PCR conditions, such as suboptimal template quality and inherent and unavoidable minor variations in pipetting. We set our threshold to $R_{CT} = 5 \times 10^7$ and calculate the C_T value according to Eq. (2):

$$C_T = \frac{\lg R_{C_T} - \lg R_0}{\lg(1+E)}$$

$$C_T(cDNA_1) = \frac{\lg(5 \cdot 10^7) - \lg 10}{\lg(1+0.85)} = 25.07$$

$$C_T(cDNA_2) = \frac{\lg(5 \cdot 10^7) - \lg 10}{\lg(1+0.70)} = 29.07$$

In this example, an efficiency difference of 0.15 results in C_T value difference of $\Delta C_T = C_T(\mathrm{cDNA_1}) - C_T(\mathrm{cDNA_2}) = -4$. According to the $\Delta\Delta C_T$ method, perfect PCR conditions with an efficiency of E=1 are always assumed, i.e. a difference of one C_T value between two samples corresponds to a doubling of the amount of PCR product in the sample with the lower C_T value. By ignoring the real efficiencies of the PCR within the data analysis, we would calculate the relative gene expression in cDNA$_1$ 16 times higher than in cDNA$_2$ (ΔC_T=29.07−25.07=4 corresponds to $2^{-(-4)}$=16-fold expression according to the $\Delta\Delta C_T$ method) with these C_T values.

This example demonstrates that ignoring different PCR efficiencies may result in calculation of significant mRNA expression differences comparing two cDNAs with identical initial template amount.

Thus, we see that E strongly influences the technically determined C_T value. Due to the exponential nature of qPCR kinetics, small variations in C_T values have large effects on calculated gene expression ratios.

Therefore, a major prerequisite for valid qPCR data analysis is the determination of E in every PCR run on every sample and gene due to two reasons: 1) if we assume E=1=100%, gene expression is over- or underestimated (see Fig. 4, which is introduced later) in the case of $E \neq 1$ (e.g. E(cDNA1)=E(cDNA2)=0.8), which is the most common case [19]; 2) if we assume one identical E for every single PCR (i.e. well), we neglect inter-well differences [i.e. E(cDNA1) ≠ E(cDNA2)] leading to wrong determination of gene expression, as seen above.

The $\Delta\Delta C_T$ method and the standard curve method do not fulfil both criteria, as they always assume E=1 or determine an averaged E, respectively. Because of this, these methods are only applicable under certain conditions.

There are different approaches possible to determine the efficiency E of the PCR. The prominent method is an external standard curve with serial dilution series of a template (e.g. cDNA or plasmid DNA) measured in separate wells [6]. However, this conservative method has numerous drawbacks.

In practice, a standard curve often only covers two orders of magnitude, whereas a standard curve for adequate determination of E has to cover at least 3–5 orders of magnitude [24] for valid anchoring of the trendline. In this case, the dilutions are very susceptible to pipetting errors leading to misdetermination of E, with a tendency to overestimate E [8]. Moreover, screening for inter-well variations concerning E is not possible. More practical disadvantages of the standard curve approach are the high consumption of reagents (and cDNA) and the occupancy of many wells during each PCR run. Therefore, approaches to determine E directly from the PCR kinetic curve are preferable, and a number of calculation methods for E have been published so far [7–9, 25, 26; http://www.efficiency.gene-quantification.info].

The linear regression [9] method has proved to work stably and accurately and it performs best in our hands. This method is based on the calculation of the *log* of the amount of fluorescence at each cycle, resulting in a linear graph in the exponential phase of the PCR kinetic curve (see above).

In this linear range, a trendline including 3–6 data points, with the highest possible slope and with the highest correlation coefficient (also known as Pearson correlation coefficient R^2) must be defined [9] (see Fig. 3).

Data are plotted as lg R_n on the ordinate and n on the abscissa, and the slope of the trendline can then be determined from the equation:

$$\begin{aligned} R_n &= R_0 \cdot (1+E)^n \\ \lg R_n &= \lg(1+E) \cdot n + \lg R_0 \end{aligned}$$

Therefore, the slope s equals lg (1+E) according to the standard straight line formula (y=mx+b, where m is the

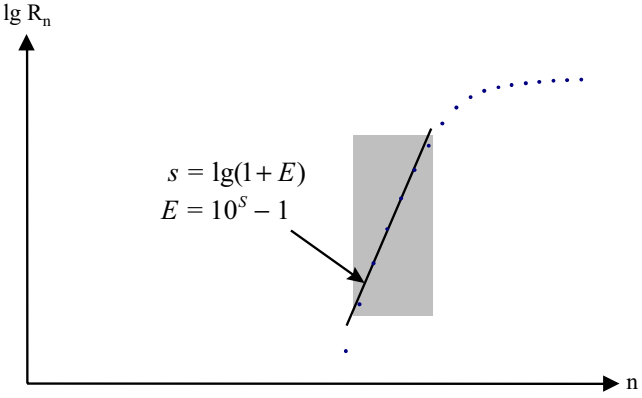


Fig. 3 PCR efficiency determination. In a plot of the PCR kinetic curve with a half-logarithmic scale, the PCR efficiency E can be calculated from the slope of a trendline in the exponential phase of the curve following the indicated formula. n number of cycles, R_n fluorescence amount after n cycles, s slope of the trendline, E PCR efficiency

slope of the straight line), which enables us to calculate E from the slope of the straight line:

$$\begin{aligned} & s = lg\,(1+E) \\ \Leftrightarrow\ & 10^s = 1+E \\ \Leftrightarrow\ & E = 10^s - 1 \end{aligned} \qquad (3)$$

Thus far, the manufacturers of real-time PCR instruments have not included algorithms for determining PCR efficiency in their data analysis software packages. Nevertheless, the above calculation can be easily performed using, e.g. the publicly available programs LinRegPCR [9] or DART-PCR [8].

Unfortunately, PCR efficiency determination is generally applied on few data points and is therefore prone to errors in measurement. This drawback may partially be compensated by highly sensitive assays utilizing gene-specific reverse transcription in combination with hydrolysis probes (e.g. TaqMan) or hairpin probes (e.g. Molecular Beacons, Scorpions) which can expand the linear range of data acquisition [10]. Besides these technical aspects, novel algorithms (such as the sigmoidal and the logistic models [21, 25, 27]) have been developed in recent years to include more data points in the data analysis, increasing its precision. These algorithms work user-independent, but are not publicly available as software packages yet (for review of efficiency estimation methods, see Wong et al. [10]).

In essence, there are two ways to work with efficiencies determined for each well: First, E may be considered separately for each well. Secondly, a mean value of all Es is calculated and applied to all wells.

It has been shown that data analysis with a mean value of E shows smaller variations concerning the calculated initial template amount R_0 than an "each-well-separately" approach [8].

An important prerequisite for the use of a mean value of E is the standard deviation (SD) of the determined Es. As discussed above, even small variations in E may result in miscalculations of gene expression ratios. Nevertheless, the determination of E is limited by the number of cycles over which linear regression is applicable. Therefore, the accuracy of E determination is also limited. Thus, small ranges of E have to be accepted, but outliers must be detected, if we would like to calculate our relative expression ratios (*rER*; see below) with averaged Es.

Bar et al. investigated the distribution of Es determined from PCR kinetic curves [28]. They showed that the determined Es are normally distributed. They calculated a maximum SD of E [SD(E)] of 0.02 for a large training set and various genes. For kinetic outlier detection (KOD), a 95% confidence interval, which equals a mean value of E±1.96 SD (i.e. mean value of E±0.039; SD=0.02) was applied and only Es within this interval were included in data analysis (the residuals were defined as outliers). In this way, Bar et al. proved that *rER* calculation with averaged Es works well under these conditions. If the Es in our qPCRs fit these requirements (especially, if SD(E)≤0.02), we can, after elimination of outliers, calculate *rER* with averaged Es.

Another option has been proposed by Peirson et al. [8]. These authors also assumed a normal distribution of the determined Es and performed analysis of variance (ANOVA) to exclude outliers and to check for affiliation to a common (homogenous) distribution of the determined Es before calculating with averaged Es. Peirson et al. also provide a free tool for this process, named DART-PCR. In case of non-homogenous distribution of E in the compared samples, the data analysis must be performed with the individually calculated Es for all wells. Generally, we try to avoid data analysis under these conditions, as the calculation of E is an estimation—not resulting in a "true" value—and is therefore prone to small errors that may cause large variations in *rER*.

Determination of expression ratios

After introduction and discussion of C_T values and efficiency E, we can now use both parameters to calculate gene-specific relative expression ratios *rERs*.

A gene-specific *rER* is defined as the expression level of a gene-of-interest (*GOI*) in one sample of interest (*SOI*, normally a cDNA template) vs a reference sample (e.g.

Fig. 4 Derivation of the GED formula

We start our derivation with the basic PCR kinetics equation and set the values for the gene-of-interest *GOI* and the housekeeping gene *HKG* (with respect to R_{C_T}, R_0, and C_T) for one sample.

$$R_{C_T} = R_0 \cdot (1+E)^{C_T}$$
$$\Rightarrow R_{C_T}(GOI) = R_0(GOI) \cdot (1+E(GOI))^{C_T(GOI)} \qquad (I)$$
$$\Rightarrow R_{C_T}(HKG) = R_0(HKG) \cdot (1+E(HKG))^{C_T(HKG)} \qquad (II)$$

$$\frac{(I)}{(II)} \rightarrow K = \frac{R_{C_T}(GOI)}{R_{C_T}(HKG)} = \frac{R_0(GOI) \cdot (1+E(GOI))^{C_T(GOI)}}{R_0(HKG) \cdot (1+E(HKG))^{C_T(HKG)}}$$
$$R_{\text{norm}} = \frac{R_0(GOI)}{R_0(HKG)} \Rightarrow K = R_{\text{norm}} \cdot \frac{(1+E(GOI))^{C_T(GOI)}}{(1+E(HKG))^{C_T(HKG)}} \Leftrightarrow R_{\text{norm}} = K \cdot \frac{(1+E(HKG))^{C_T(HKG)}}{(1+E(GOI))^{C_T(GOI)}}$$

R_{norm} is the relative quantity of the *GOI* relative to the *HKG*. Now we set the R_{norm} of our sample-of-interest *SOI* relative to the R_{norm} of a reference sample *ref*.

$$\frac{R_{\text{norm}}(SOI)}{R_{\text{norm}}(ref)} = \frac{K \cdot \frac{(1+E(HKG;SOI))^{C_T(HKG;SOI)}}{(1+E(GOI;SOI))^{C_T(GOI;SOI)}}}{K \cdot \frac{(1+E(HKG;ref))^{C_T(HKG;ref)}}{(1+E(GOI;ref))^{C_T(GOI;ref)}}}$$
$$\Leftrightarrow \frac{R_{\text{norm}}(SOI)}{R_{\text{norm}}(ref)} = \frac{(1+E(HKG;SOI))^{C_T(HKG;SOI)}}{(1+E(GOI;SOI))^{C_T(GOI;SOI)}} \cdot \frac{(1+E(GOI;ref))^{C_T(GOI;ref)}}{(1+E(HKG;ref))^{C_T(HKG;ref)}}$$

$$\Leftrightarrow \frac{R_{\text{norm}}(SOI)}{R_{\text{norm}}(ref)} = \frac{(1+E(HKG;SOI))^{C_T(HKG;SOI)}}{(1+E(GOI;SOI))^{C_T(GOI;SOI)}} \cdot \frac{(1+E(HKG;ref))^{-C_T(HKG;ref)}}{(1+E(GOI;ref))^{-C_T(GOI;ref)}} \qquad (5)$$

Equation 3 offers the opportunity to insert different efficiencies *E* for all genes and samples. If we calculate with averaged *E*s for the *GOI* and *HKG*, the equation can be further simplified.

$$\frac{R_{\text{norm}}(SOI)}{R_{\text{norm}}(ref)} = \frac{(1+E(HKG))^{C_T(HKG;SOI)-C_T(HKG;ref)}}{(1+E(GOI))^{C_T(GOI;SOI)-C_T(GOI;ref)}}$$

$$\Rightarrow \Delta C_T(gene) = C_T(gene;SOI) - C_T(gene;ref)$$

$$rER = \frac{R_{\text{norm}}(SOI)}{R_{\text{norm}}(ref)} = \frac{(1+E(HKG))^{\Delta C_T(HKG)}}{(1+E(GOI))^{\Delta C_T(GOI)}} = \frac{(1+E(GOI))^{-\Delta C_T(GOI)}}{(1+E(HKG))^{-\Delta C_T(HKG)}} \qquad (4)$$

R_0 and R_{C_T} are initial and threshold fluorescence, respectively. *E(gene; sample)* is the PCR efficiency of one gene in one sample. C_T is the fractional cycle number where the PCR kinetic curve reaches R_{C_T}. *K* is the quotient of the threshold fluorescences R_{C_T}*(GOI)* and R_{C_T}*(HKG)* and is, therefore, identical for all samples on the same plate/data analysis.

Equations 4 and 5 can be transformed into the well-known $\Delta\Delta C_T$ formula with *E(GOI; sample)* = *E(HKG; sample)* = 1. It should be mentioned that this transformation is not possible if *E(GOI; sample)* = *E(HKG; sample)* $\neq 1$, indicating that the $\Delta\Delta C_T$ method is only applicable in the case of $E = 1$ for all samples and genes.

cDNA from healthy tissue, unstimulated cells, etc.). We call this process relative quantification, as usually we have no information concerning the initial copy number of the gene of interest (i.e. its concentration) within this reference sample. Nevertheless, an absolute quantification would be possible if we knew the concentration of interest in the reference sample. With the exception of special cases, in which a precise determination of initial copy number is essential (e.g. virology, clinical chemistry), the application of an external standard curve will still remain the "gold standard" [29, 30].

Normally, the measurement of the *GOI* is performed relative to at least one internal standard (sometimes also called endogenous control): an "unregulated" housekeeping gene *HKG* (the issue of *HKG* selection and *HKG* indices is addressed in Fig. 1).

We looked for a PCR data analysis formula that is not dependent on standard curves (for economic reasons) and integrates the calculated PCR efficiencies (for increased accuracy). For this reason, we deduced our formula from the basic PCR kinetics formula (Eq. 1) retaining E for both PCRs (for complete derivation, see Fig. 4) to determine the relative expression ratio *rER* of a gene-of-interest *GOI* (relative to a housekeeping gene *HKG*) in a sample-of-interest *SOI* (relative to a reference sample *ref*).

Formula for the use with averaged efficiencies E for each gene:

$$rER = \frac{R_{norm}(SOI)}{R_{norm}(ref)} = \frac{(1+E(GOI))^{-\Delta C_T(GOI)}}{(1+E(HKG))^{-\Delta C_T(HKG)}} \quad (4)$$

with $\Delta C_T(gene) = C_T(gene;SOI) - C_T(gene;ref)$.

Formula for the use with individual efficiencies E for each gene and also each well:

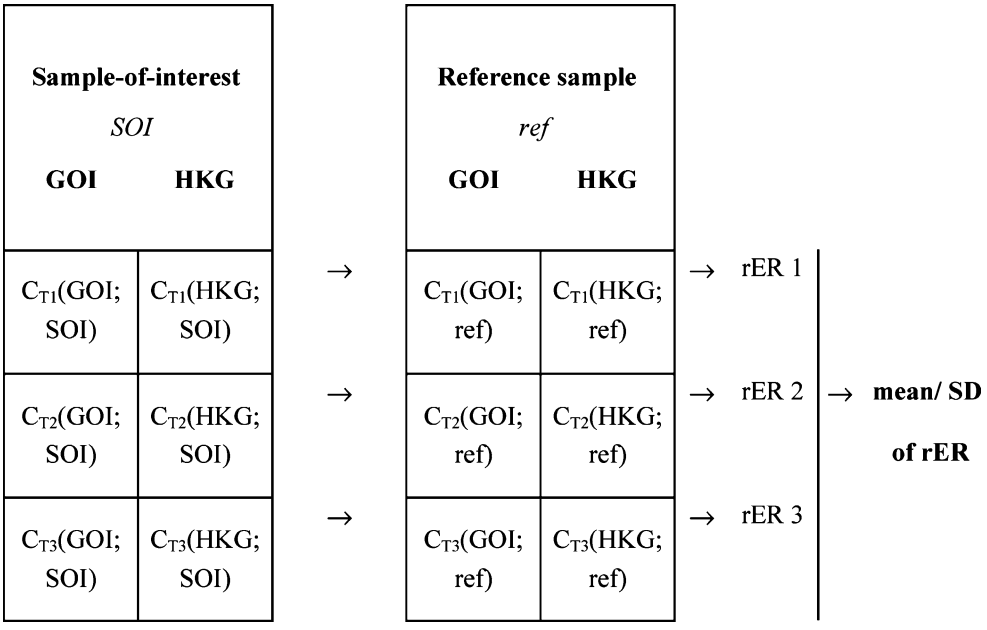
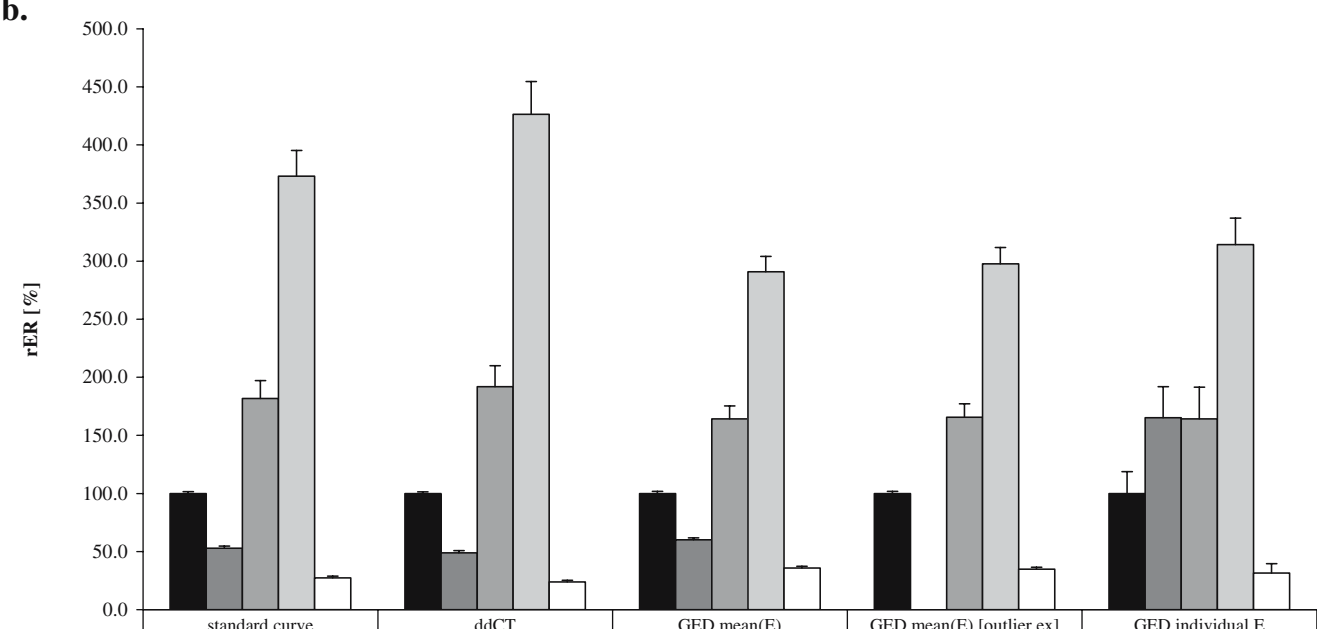


Fig. 5 Statistical analysis. We suggest performing the measurement of a single cDNA template in—at least—technical triplicate. In this case, three different C_T values (C_{T1}, C_{T2}, C_{T3}) are obtained for each gene (*GOI* and *HKG*) and each sample. One of the measured samples is defined as the reference sample *ref*. To compare two samples we calculate three different relative expression ratios *rER*s (with averaged or individual *E*s) by clustering the corresponding C_T values as indicated in the table. If we assume that the calculated *rER*s for *one* sample-of-interest *SOI* are part of a normal distribution (as the technically determined C_T values and *E*s are), we can also calculate the mean value and the standard derivation of these *rER*s. Additionally, we may perform statistical testing for differences between *two SOI*s by applying an unpaired, two-tailed *t*test or an analysis of variation (ANOVA). For statistical analysis of two sample groups (two groups consisting of more than one *SOI*), the assumption of normal distribution within these groups is problematic, i.e. one cannot assume that, for example, the expression of a certain gene in glioblastomas of ten different individuals is normally distributed. Here, a non-parametric statistical analysis as the Mann–Whitney–Wilcoxon rank sum test should be applied. This procedure should also be applied for statistical analysis of biological replicates

a.

sample	C_T(RER)	E(RER)	C_T(18S)	E(18S)
adrenal	24.43	0.67	6.02	0.85
	24.52	0.66	6.15	0.84
	24.3	0.69	5.92	0.86
heart	25.63	0.58	6.27	0.73
	25.73	0.59	6.26	0.75
	25.71	0.56	6.29	0.7
kidney	23.61	0.68	6.23	0.84
	23.83	0.68	6.22	0.86
	23.56	0.67	6.2	0.87
HeLa-S3	22.32	0.67	6.13	0.83
	22.52	0.67	6.15	0.87
	22.46	0.68	6.13	0.86
HEK293	26.61	0.7	6.22	0.88
	26.52	0.68	6.09	0.85
	26.71	0.66	6.16	0.87
NTC	37.67		34.12	
	38.58		35.69	
	37.23		34.65	

b.



	standard curve	ddCT	GED mean(E)	GED mean(E) [outlier ex]	GED individual E
adrenal	100.0	100.0	100.0	100.0	100.0
heart	52.9	49.0	60.1		165.3
kidney	181.8	192.0	164.2	165.6	164.1
HeLa-S3	373.1	426.3	290.7	297.7	314.2
HEK293	27.5	23.8	35.8	34.9	31.5

◀ **Fig. 6** GED method vs standard curve and $\Delta\Delta C_T$ method. qPCR analysis of human renin/prorenin receptor (RER) mRNA expression was performed on several different human tissues and cell lines applying a SYBR Green I reaction mix run on an ABI Prism 7000 Sequence Detection System (Applied Biosystems, Weiterstadt, Germany). RNA of human heart and tissue of human kidney and adrenal were kindly provided by the group of Patricia Ruiz (Charité-Universitätsmedizin Berlin) and the Department of Urology (Charité-Universitätsmedizin Berlin), respectively. 18S rRNA (18S) expression was used for normalization. The individual PCR efficiencies were determined using LinRegPCR [9]. According to the standard curve method the PCR efficiency was determined with 0.87 and 0.95 for RER and 18S rRNA, respectively. **a** Raw data (C_T values and individual PCR efficiencies E) of this PCR run. PCR efficiency on heart cDNA was significantly lower compared to the other samples as detected with ANOVA, and was therefore defined as outlier (*grey underlay*). **b** Comparison of different qPCR data analysis methods. The relative expression ratio (rER) of adrenal tissue was set to 100%. Data analyses based on the standard curve and $\Delta\Delta C_T$ method are shown as indicated. Additionally, we analysed the same data set using our novel GED formula. We calculated each rER with averaged E for all wells [GED mean(E)], averaged E for all wells after exclusion of the outliers [GED mean(E) (outlier ex)], and individual E for all wells separately [GED individual E], respectively. This example demonstrates that usage of the $\Delta\Delta C_T$ method results in an overestimation of the gene expression differences especially compared to the different GED approaches (e.g., 426.3% vs 290.7%; $p<0.01$). Furthermore, this example indicates that the detection of efficiency outliers is crucial. Neglecting individual E in *rER* calculation—by using the standard curve or the $\Delta\Delta C_T$ method—can even cause the change of a relative overexpression to a relative underexpression (165.3% vs 49.0% or 52.9%, respectively; both $p<0.01$)

$$rER = \frac{R_{norm}(SOI)}{R_{norm}(ref)} = \frac{(1+E(HKG;SOI))^{C_T(HKG;SOI)}}{(1+E(GOI;SOI))^{C_T(GOI;SOI)}} \cdot \frac{(1+E(HKG;ref))^{-C_T(HKG;ref)}}{(1+E(GOI;ref))^{-C_T(GOI;ref)}} \quad (5)$$

$R_{\mathrm{norm}}(SOI)$ and $R_{\mathrm{norm}}(ref)$ are the initial fluorescence amounts (proportional to the initial DNA template amounts) in the sample-of-interest *SOI* and the reference sample *ref*, respectively, of the gene-of-interest *GOI* normalised to the housekeeping gene *HKG*. *rER* is the relative expression ratio (for statistical analysis of *rER*s, see Fig. 5). The PCR efficiency E, either applied as mean value [E(*gene*)] or separately for each well [E(*gene*; *sample*); sample is *SOI* or *ref*] (see discussion above). ΔC_T(*gene*) is the C_T value difference of the *SOI* and the *ref* for the *gene* (*GOI* or *HKG*) [ΔC_T(*gene*)=C_T(*gene*; *SOI*)-C_T(*gene*; *ref*)].

In fact, ΔC_T(*gene*) is nothing more than the gene expression's C_T difference between two samples. Therefore, we named our real-time RT-PCR data analysis approach the GED (Gene Expression's C_T Difference) method.

The GED formula can easily be implemented into spreadsheet programs such as Microsoft Excel. By using this method in daily practice we observed significant differences comparing the GED method with $\Delta\Delta C_T$ and standard curve method. Especially detection and mathematical consideration of efficiency differences enabled a more accurate calculation of *rER* with the GED formula (see Fig. 6).

Conclusion

The incorrect assumption of a 100% PCR efficiency strongly confounds accurate calculation of gene-specific relative expression ratios—the essential purpose of real-time PCR analysis. Assumption of an identical efficiency for each well may also confound data analysis. The GED formula, presented in this paper, compensates for these major drawbacks. The formula overcomes the requirement for an additional standard curve for every gene on every plate, and is able to consider non-homogenous and suboptimal PCR efficiencies. Therefore, it is more economic than the standard curve method and ensures a higher accuracy of gene expression ratio determination in comparison to the $\Delta\Delta C_T$ and standard curve method.

Acknowledgements The authors would like to thank Dave Hancock and Justin Cross for critically reading the manuscript.

References

1. Higuchi R, Dollinger G, Walsh PS, Griffith R (1992) Simultaneous amplification and detection of specific DNA sequences. Biotechnology (NY) 10:413–417
2. Higuchi R, Fockler C, Dollinger G, Watson R (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology (NY) 11:1026–1030
3. Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. Genome Res 6:986–994
4. Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP (1997) Continuous fluorescence monitoring of rapid cycle DNA amplification. Biotechniques 22:130–138
5. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25:402–408
6. Morrison TB, Weis JJ, Wittwer CT (1998) Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. Biotechniques 24:954–962
7. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45
8. Peirson SN, Butler JN, Foster RG (2003) Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. Nucleic Acids Res 31:e73
9. Ramakers C, Ruijter JM, Deprez RH, Moorman AF (2003) Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 339:62–66
10. Wong ML, Medrano JF (2005) Real-time PCR for mRNA quantitation. Biotechniques 39:75–85

11. Murphy LD, Herzog CE, Rudick JB, Fojo AT, Bates SE (1990) Use of the polymerase chain reaction in the quantitation of mdr-1 gene expression. Biochemistry 29:10351–10356
12. Becker-Andre M, Hahlbrock K (1989) Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). Nucleic Acids Res 17:9437–9446
13. Liu W, Saint DA (2002) Validation of a quantitative method for real time PCR kinetics. Biochem Biophys Res Commun 294:347–353
14. Becker-Andre M, Hahlbrock K (1989) Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). Nucleic Acids Res 17:9437–9446
15. Ding C, Cantor CR (2004) Quantitative analysis of nucleic acids—the last few years of progress. J Biochem Mol Biol 37:1–10
16. Freeman WM, Walker SJ, Vrana KE (1999) Quantitative RT-PCR: pitfalls and potential. Biotechniques 26:112–125
17. Wilhelm J, Pingoud A (2003) Real-time polymerase chain reaction. Chembiochem 4:1120–1128
18. Monod J (1949) The growth of bacterial cultures. Annu Rev Microbiol 3:371–394
19. Pannetier C, Delassus S, Darche S, Saucier C, Kourilsky P (1993) Quantitative titration of nucleic acids by enzymatic amplification reactions run to saturation. Nucleic Acids Res 21:577–583
20. Luu-The V, Paquet N, Calvo E, Cumps J (2005) Improved real-time RT-PCR method for high-throughput measurements using second derivative calculation and double correction. Biotechniques 38:287–293
21. Tichopad A, Dilger M, Schwarz G, Pfaffl MW (2003) Standardized determination of real-time PCR efficiency from a single reaction set-up. Nucleic Acids Res 31:e122
22. Suslov O, Steindler DA (2005) PCR inhibition by reverse transcriptase leads to an overestimation of amplification efficiency. Nucleic Acids Res 33:e181
23. Bustin SA, Nolan T (2004) Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction. J Biomol Tech 15:155–166
24. Livak KJ (2001) User Bulletin #2, ABI PRISM 7700 Sequence detection system. PE applied biosystems. [http://www.docs.appliedbiosystems.com/pebiodocs/04303859.pdf]
25. Liu W, Saint DA (2002) A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. Anal Biochem 302:52–59
26. Tichopad A, Didier A, Pfaffl MW (2004) Inhibition of real-time RT-PCR quantification due to tissue-specific contaminants. Mol Cell Probes 18:45–50
27. Tichopad A, Dzidic A, Pfaffl MW (2002) Improving quantitative real-time RT-PCR reproducibility by boosting primer-linked amplification efficiency. Biotechnol Lett 24:2053–2056
28. Bar T, Stahlberg A, Muszta A, Kubista M (2003) Kinetic Outlier Detection (KOD) in real-time PCR. Nucleic Acids Res 31:e105
29. Mohey R, Jorgensen LB, Moller BK, Black FT, Kjems J, Obel N (2005) Detection and quantification of proviral HIV-1 184 M/V in circulating CD4(+) T cells of patients on HAART with a viremia less than 1,000 copies/ml. J Clin Virol 34:257–267
30. Stahlberg A, Aman P, Ridell B, Mostad P, Kubista M (2003) Quantitative real-time PCR method for detection of B-lymphocyte monoclonality by comparison of kappa and lambda immunoglobulin light chain expression. Clin Chem 49:51–59
31. Ivell R (1998) A question of faith—or the philosophy of RNA controls. J Endocrinol 159:197–200
32. Bustin SA (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. J Mol Endocrinol 29:23–39
33. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:Research 1–11