Protocol for qPCR analysis that corrects for cDNA amplification efficiency

This protocol presents a variation on the $2^{-\Delta\Delta Ct}$ technique for qPCR analysis. Our approach requires the inclusion of a standard curve on each qPCR plate, and like the $2^{-\Delta\Delta Ct}$ technique, is dependent on the stability of housekeeping gene expression. However, unlike the $2^{-\Delta\Delta Ct}$ technique, our approach corrects for imperfect cDNA amplification efficiency and allows for the use of multiple housekeeping genes. Collectively, this approach enhances analytical accuracy and thereby reduces the type I and II statistical errors in the generated data.

Publisher’s note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.
Protocol

Protocol for qPCR analysis that corrects for cDNA amplification efficiency

Mads V. Damgaard1,2,* and Jonas T. Treebak1,3,*

1Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
2Technical contact
3Lead contact
*Correspondence: damgaard@sund.ku.dk (M.V.D.), jttreebak@sund.ku.dk (J.T.T.)
https://doi.org/10.1016/j.xpro.2022.101515

SUMMARY

This protocol presents a variation on the 2^ΔΔCt technique for qPCR analysis. Our approach requires the inclusion of a standard curve on each qPCR plate, and like the 2^ΔΔCt technique, is dependent on the stability of housekeeping gene expression. However, unlike the 2^ΔΔCt technique, our approach corrects for imperfect cDNA amplification efficiency and allows for the use of multiple housekeeping genes. Collectively, this approach enhances analytical accuracy and thereby reduces the type I and II statistical errors in the generated data.

BEFORE YOU BEGIN

This protocol illustrates how to normalize qPCR data more accurately than the commonly used 2^ΔΔCt method (Livak and Schmittgen, 2001) and includes the option of utilizing multiple housekeeping genes. Herein we use qPCR data from mouse quadriceps muscle, and the examples are based on the housekeeping genes 18s and Gapdh as well as the gene of interest Nmrk2. However, since this method is merely a more robust way to analyze qPCR data, it is applicable to any species, tissues, and target genes. For an example of the use of this protocol, please refer to Damgaard et al. (2022).

To utilize the protocol and improve the accuracy in qPCR data analysis, a series of standards has to be included in each qPCR run, for genes of interest as well as for housekeeping genes. This allows the calculation of an experimental amplification factor, which can be used to circumvent two linked assumptions that are integral to the regular 2^ΔΔCt method and thus improve accuracy in reporting. The first assumption is that the amplification factor of a qPCR is 2, or in other words that the target cDNA is exactly doubled with every cycle, which is the theoretically perfect result. The second assumption is that the amplification factor is the same in all qPCR runs, irrespective of primer choice, temperature choice etc. Both assumptions are very rarely satisfied. If the experimental amplification factor of the housekeeping gene transcript is different from that of the gene of interest and this is not accounted for, it can disguise or exaggerate differences between test groups, depending on the direction of these differences.

Institutional permissions

The samples used in these examples were obtained in accordance with the European directive 2010/63/EU of the European Parliament and of the Council for the protection of animals used for scientific purposes. Ethical approval was given by The Danish Animal Experiments Inspectorate (#2019-15-0201-01630). Readers are reminded to always acquire similar permission from relevant institutions before performing animal experiments.
Prepare standard series

© Timing: ~30 min

1. Create a mix of cDNA from all qPCR-ready samples in the current study by taking for example 5–20 μL from each sample and adding to a single tube. Call this tube “Standard 1” and assign it a value of 1 arbitrary unit (AU) for calculation purposes.

   Note: Standard 1 contains the average cDNA concentration of all the experimental samples.

2. From Standard 1 make a 2-fold serial dilution in nuclease-free water all the way through Standard 6, as in Table 1.

   Note: If extremely large variance in target genes is expected, i.e., a CT difference of more than 10 or roughly a 1000-fold difference between samples, it is encouraged to increase the number of standards or make a 3-fold serial dilution instead.

Run the qPCRs

© Timing: 1–3 h per qPCR

3. Perform qPCRs for both housekeeping genes and genes of interest according to the instructions from the provider of polymerase. Tables 2 and 3 show a master mix and qPCR conditions example.

   Note: Since a series of standards is included with every qPCR, it is not necessary to run housekeeping genes and genes of interest on the same plate or even the same qPCR machine.

   Note: Take into account the melting temperature of the specific primer sets when deciding the annealing temperature.

   Note: Usually 40 cycles are allowed, although in practice CT values above 35 are often associated with increased variance between replicates.

---

### Table 1. Serial dilution example

| Standard # | Concentration [AU] |
|------------|--------------------|
| Standard 1 | 1                  |
| Standard 2 | 0.5                |
| Standard 3 | 0.25               |
| Standard 4 | 0.125              |
| Standard 5 | 0.0625             |
| Standard 6 | 0.03125            |

---

### Table 2. qPCR reaction master mix

| Reagent                  | Volume (pr. Sample) |
|--------------------------|---------------------|
| Brilliant III Ultra-fast SYBR | 5 μL                |
| Forward primer (10 μM)   | 0.3 μL              |
| Reverse primer (10 μM)   | 0.3 μL              |
| Nuclease-free water      | 2.4 μL              |
| cDNA                     | 2 μL                |
| Total                    | 10 μL               |
4. Quality control of each individual qPCR run can be performed in the following ways:
   a. Include technical replicates – triplicates are recommended if possible.

   **Note:** This is especially important if the target gene has low expression, as errors in early amplification cycles are more impactful the lower the initial cDNA concentration is (See problem 1).

   b. Include a melting curve to ensure that only the target gene has been amplified (See problem 2).

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| Critical commercial assays | | |
| RNeasy kit | QIAGEN | 74104 |
| iScript cDNA synthesis kit | BIO-RAD | 1708890 |
| Brilliant III Ultra-fast SYBR qPCR Master Mix | Agilent | 600883 |
| Deposited data | | |
| Supplemental Excel sheets - “Example qPCR” & “Proof of concept” | This paper | |
| | | Mendeley Data: [https://doi.org/10.17632/ytbnp523n9.1](https://doi.org/10.17632/ytbnp523n9.1) |
| Experimental models: Organisms/strains | | |
| 19-week-old male C57BL/6NTac mice | Taconic | B6-M |
| Oligonucleotides | | |
| Primer 18s Forward | 10.1016/j.isci.2022.103863 AGTCCCTGCCCTTTGTACACA |
| Primer 18s Reverse | 10.1016/j.isci.2022.103863 GATCCGAGGGCCCTCACAAAC |
| Primer Gapdh Forward | 10.1016/j.isci.2022.103863 GCACAGTCAAGGCCGAGAAT |
| Primer Gapdh Reverse | 10.1016/j.isci.2022.103863 GCCCTTCCATGGTGTTGAA |
| Primer Nmrk2 Forward | This paper GGGGTGGAAAGGCTGCTATTTA |
| Primer Nmrk2 Reverse | This paper CGTCCAGTCAGAACATCTCA |
| Software and algorithms | | |
| Excel | Microsoft | v2016 |

### MATERIALS AND EQUIPMENT

Microsoft Excel was used for calculations and plotting, but any similar program could be used depending on preference. Similarly, our choices for RNA isolation, cDNA synthesis and qPCR kits are listed in the key resources table, but the calculation method outlined below is independent of kit preference.

### STEP-BY-STEP METHOD DETAILS

**Calculate the experimental amplification factor for each qPCR**

- **Timing:** Initial setup - 1 h, subsequent use - 5 min
### Protocol

#### Table A

| Standard # | $C_T(18s)$ | Mean $C_T$ | Concentration |
|------------|-------------|------------|---------------|
| 2          | 16.28       | 16.35      | 16.43         | 16.35         | 1            |
| 3          | 17.48       | 17.30      | 17.61         | 17.46         | 0.5          |
| 4          | 18.25       | 18.41      | 18.15         | 18.27         | 0.25         |
| 5          | 19.16       | 19.14      | 19.08         | 19.13         | 0.125        |
| 6          | 20.22       | 20.15      | 20.29         | 20.22         | 0.0625       |
| 7          | 21.24       | 21.38      | 21.09         | 21.24         | 0.03125      |

#### Table B

| Standard # | $C_T(18s)$ | Mean $C_T$ | Concentration |
|------------|-------------|------------|---------------|
| 2          | 16.28       | 16.35      | 16.43         | 16.35         | 1            |
| 3          | 17.48       | 17.30      | 17.61         | 17.46         | 0.5          |
| 4          | 18.25       | 18.41      | 18.15         | 18.27         | 0.25         |
| 5          | 19.16       | 19.14      | 19.08         | 19.13         | 0.125        |
| 6          | 20.22       | 20.15      | 20.29         | 20.22         | 0.0625       |
| 7          | 21.24       | 21.38      | 21.09         | 21.24         | 0.03125      |

**Cells in Table A:**

- **G2**: =RSQ(E2:E7;LOG(F2:F7))
- **H2**: $R^2 = 0.9977$

#### Table C

| Standard # | $C_T(18s)$ | Mean $C_T$ | Concentration |
|------------|-------------|------------|---------------|
| 2          | 16.28       | 16.35      | 16.43         | 16.35         | 1            |
| 3          | 17.48       | 17.30      | 17.61         | 17.46         | 0.5          |
| 4          | 18.25       | 18.41      | 18.15         | 18.27         | 0.25         |
| 5          | 19.16       | 19.14      | 19.08         | 19.13         | 0.125        |
| 6          | 20.22       | 20.15      | 20.29         | 20.22         | 0.0625       |
| 7          | 21.24       | 21.38      | 21.09         | 21.24         | 0.03125      |

**Cells in Table C:**

- **G2**: =SLOPE(E2:E7;LOG(F2:F7,2))
- **H2**: $R^2 = 0.9977$
- **I2**: $slope = -0.9579$

#### Table D

| Concentration | log2(Concentration) | Mean $C_T$ |
|---------------|---------------------|------------|
| 1             | 0                   | 16.35450   |
| 0.5           | -1                  | 17.4623619 |
| 0.25          | -2                  | 18.272091  |
| 0.125         | -3                  | 19.1263069 |
| 0.0625        | -4                  | 20.2182892 |
| 0.03125       | -5                  | 21.2355209 |

![Graph showing linear regression](image.png)

**Graph Details:**
- Line equation: $y = -0.9579x + 16.383$
- $R^2 = 0.9977$

#### Table E

| Standard # | $C_T(18s)$ | Mean $C_T$ | Concentration |
|------------|-------------|------------|---------------|
| 2          | 16.28       | 16.35      | 16.43         | 16.35         | 1            |
| 3          | 17.48       | 17.30      | 17.61         | 17.46         | 0.5          |
| 4          | 18.25       | 18.41      | 18.15         | 18.27         | 0.25         |
| 5          | 19.16       | 19.14      | 19.08         | 19.13         | 0.125        |
| 6          | 20.22       | 20.15      | 20.29         | 20.22         | 0.0625       |
| 7          | 21.24       | 21.38      | 21.09         | 21.24         | 0.03125      |

**Cells in Table E:**

- **G2**: =2^-1/15
- **H2**: $R^2 = 0.9977$
- **I2**: $slope = -0.9579$
- **H3**: $slope = -0.9579$
- **I3**: $Amp. F = 2.0618$
The purpose of this step is to calculate the individual amplification factors for each qPCR and to ensure that the calculated amplification factors are accurate. This is necessary to correct for any differences in amplification efficiency.

1. Fill out a table as shown in Figure 1A based on the results from the standard series.

2. Calculate R² of a linear correlation between the mean CT values, and the log-transformed standard series concentrations (Figure 1B) – any log base will work for this purpose.

   **Note:** CT values can be viewed as a log₂-transformed measure of the target gene amount, and thus should correlate linearly with the log-transformed version of the standard series concentrations.

   **Note:** The resulting R²-value should exceed 0.99 to ensure that the amplification factor calculated in step 4 is as accurate as possible. If this is not the case, refer to problem 3.

3. Calculate the slope between the mean CT values and the log₂-transformed standard curve concentrations (Figure 1C).

   **Note:** Log base 2 is used here to fit the 2-fold serial dilution. If a 3-fold or 10-fold dilution scheme is chosen, use log base 3 or 10, respectively.

   **Note:** It is recommended to proceed to analysis only if the slope value lies between -0.9 and -1.1, as the expected slope of a perfect qPCR run has a value of -1. If this is not the case, refer to problem 4.

   **Optional:** Data can be plotted as shown in Figure 1D. Transform the standard curve concentrations using the correct log base (as in step 3) and create a scatter plot with a trendline. Display Equation and R-squared value on chart to read the slope and R² directly from the resulting figure. R² and slope values should be as in steps 2 and 3, respectively.

4. Calculate the experimental amplification factor of the qPCR run from the slope by raising 2 to the power of -1/Slope (Figure 1E).

   **Note:** In the calculation above, 2 is the factor used for the serial dilution (adjust this according to dilution scheme as in step 3) and -1 is the expected slope of a perfect qPCR run.

   **Note:** The resulting experimental amplification factor is the factor with which the cDNA was amplified per cycle in each specific qPCR run.

**Calculate the relative expression of the gene(s) of interest**

* Timing: Initial setup - 1 h, subsequent use - 5 min

The purpose of this step is to use the individual amplification factors to accurately calculate the relative expression of the gene(s) of interest.
Figure 2. Examples for calculating the relative expression of the gene(s) of interest

(A) Example of calculating the linear form expression value of a gene (18s).

(B) Example of calculating a normalization factor based on multiple housekeeping genes (18s and Gapdh), by taking the geometric mean of the linear form expression values of each gene.
5. Calculate the linear form expression values for each sample by raising the associated experimental amplification factor to the power of the negative of each sample CT value (Figure 2A).

**Note:** The resulting values are a linear (rather than logarithmic) representation of the amount of replicable cDNA of the target type. If the dataset is from a housekeeping gene, and the housekeeping gene expression is truly stable, the resulting linear form expression values will correlate well with the (unknown) amount of replicable cDNA in each sample. Thus, they can be used as normalization factors.

6. (Recommended) Use multiple housekeeping genes to calculate an even more accurate set of normalization factors.
   a. Perform steps 1–5 for all housekeeping genes
   b. Ensure that the linear form expression values of the housekeeping genes correlate well (see problem 5).
   c. Find the adjusted normalization factors by calculating the geometric means of the linear form expression values from each housekeeping gene (Figure 2B).

**Note:** The geometric mean is used (rather than regular mean) because the data was exponentially transformed. The result is a more robust mean, similar to the regular mean between CT values, except that the experimental amplification factor of each housekeeping gene transcript has been accounted for.

**Note:** Statistically, a normalization factor based on multiple housekeeping genes should correlate even more accurately with the amount of replicable cDNA in each sample than a single housekeeping gene.

7. Calculate the relative expression values for the gene(s) of interest.
   a. Perform steps 1–5 for the gene(s) of interest.
   b. For each sample, divide the linear form expression value by the normalization factor found in steps 5 or 6 (Figures 2C and 2D).

**Note:** Irrespective of the chosen type of normalization factor, the result will be the amount of the gene of interest relative to a reference that correlates with the amount of replicable cDNA in each sample. If a single housekeeping gene is used, the resulting values can be reported as gene of interest/housekeeping gene (here Nmrk2/18s), whereas if multiple housekeeping genes are used it should be reported as “relative expression” or “arbitrary units”.

**Optional:** Adjust the relative expression values so the control group has a mean of 1 arbitrary unit. To do so, divide the relative expression values found in step 7, with the average relative expression value for all samples in the control group (Figure 2E).

**Note:** An advantage to this optional adjustment is that it may facilitate visualization of the data. A disadvantage is that when investigating multiple genes, it removes the ability to gauge which genes are highly expressed and which are not. Resulting values should be reported as “relative expression” or “arbitrary units”.

---

**Figure 2. Continued**

(C) Example of calculating the relative expression of a gene of interest (Nmrk2) compared to a single housekeeping gene (18s).

(D) Example of calculating the relative expression of a gene of interest (Nmrk2) compared to a normalization factor based on multiple housekeeping genes (18s and Gapdh).

(E) Example of normalizing data such that the mean of the control group gets an arbitrary value of 1.

Abbreviations) Amp. F. = Amplification factor (experimental), Lin. Form = Linear form, GeoMean = Geometric mean, Rel. Exp. = Relative expression, Norm. Exp. = Normalized expression.
EXPECTED OUTCOME

The resulting relative measures of the gene(s) of interest in step 7 are similar to results from the $2^{-\Delta Ct}$ method, and the optional adjustment makes them similar to results from the $2^{-\Delta\Delta Ct}$ method. The difference being that the individual experimental amplification factors are accounted for in both cases to increase accuracy and thus reduce the risk of type I or type II errors. Either of the two measures are relative expression levels that can be statistically tested and reported.

The supplemental excel sheet “Proof of concept” contains 7 sheets with different scenarios and explanatory notes that show the comparison between the $2^{-\Delta Ct}$ method and the adjusted version presented herein. The individual scenarios can also be interacted with to explore further.

LIMITATIONS

It stands to reason that this method, similarly to the original $2^{-\Delta Ct}$ and $2^{-\Delta\Delta Ct}$ methods, is never better than the correlation between the chosen housekeeping gene and the amount of replicable cDNA of the experimental samples. Resultantly, the recommendation from here is to utilize multiple housekeeping genes to minimize the risk that a housekeeping gene unexpectedly affected by the treatment(s) skews the results. Another limitation is that the inclusion of a standard curve with each qPCR as well as (optionally) using multiple housekeeping genes, increases the necessary amount of sample material. This can be a problem if sample material is limited. Some research groups try to circumvent both these issues by measuring and normalizing to total input RNA or cDNA instead of utilizing housekeeping genes. However, that strategy does not account for how much of the sample material is actually replicable (non-fragmented) and thus comes with its own set of challenges.

TROUBLESHOOTING

Problem 1
There is disagreement in $C_T$ values between technical replicates during quality control step a.

Potential solution
Some disagreement between technical replicates is inevitable due to the stochastic nature of the qPCR method. This is one of the reasons technical replicates are encouraged. If technical replicates disagree more than $0.30 C_T$ from the mean ($\sim23\%$ assuming perfect amplification), it is usually caused by low expression of the gene of interest or by imprecise pipetting. If $C_T$ levels are higher than $35$ it may be necessary to use cDNA of a higher concentration to get precise amplification. If $C_T$ levels are generally low, try using larger volumes of cDNA while reducing nuclease-free water, as this will reduce the impact of slight pipetting errors.

Problem 2
The melting curve suggests more than one melting temperature during quality control step b.

Potential solution
If multiple melting temperature peaks are observed, it means either that primers are not specific to one target, or that the target has isoforms in the amplified region. Since it is not possible to distinguish whether it is caused by isoforms or alternative targets in such cases, another set of primers should be used, as the data cannot be meaningfully interpreted.

Problem 3
$R^2$ after step 2 is below 0.99 for one or more standard curve(s).

Potential solution
Consider which standard(s) cause the problem (for example via plotting). If the standard curve is generally imprecise, it is often due to imprecise pipetting, unspecific primers, or wrong temperature during qPCR. However, if it is an issue only in the lowest concentrated standards, it can relate to the
general issues with technical replicability of lowly concentrated cDNA. In this case, if the C_T values of all samples are lower (meaning higher concentration of cDNA) than Standard 5, you can exclude Standard 6 from your assessment, since the samples will still be covered by the standard curve. If this did not fix the issue, use similar reasoning to see whether you can exclude Standard 5 as well (i.e., Do all samples have lower C_T values than Standard 4?). It is not recommendable to reduce below 4 sequential standards.

Note: Approximately half of the samples will have a higher concentration (lower C_T) than Standard 1. This is not an issue as they are less impacted by potential errors in early amplification cycles than the samples with low concentration. Thus, it can be expected that these samples have been amplified similarly to the standard curve – assuming that the quality of the run is otherwise confirmed.

Problem 4
The slope of the standard curve after step 3 is nowhere near the expected value of -1.

Potential solution
Check that the correct log base has been used for the calculation. The log base should correspond to the serial dilution scheme. i.e., for a 3-fold serial dilution scheme, use log base 3.

Since the slope of the standard curve is used to correct later calculations, it is in principle possible to proceed with any slope value as long as R^2 is above 0.99. However, a slope below -1.1 or above -0.9 demonstrates that something in the qPCR run could be improved. If there is confirmation from other runs that the standards are correctly diluted, a different set of primers or a different annealing temperature can be utilized.

Problem 5
The chosen housekeeping genes do not correlate well with each other in the recommended step 6 (R^2<0.7).

Potential solution
The genes that are considered housekeeping genes generally have quite stable expression but can sometimes be affected by treatments (Bustin, 2000). If the chosen housekeeping genes do not correlate well, it signifies that (at least) one of them is not stable and cannot be used for normalization. If only two housekeeping genes were used, the recommendation is to test at least one additional housekeeping gene and assess which of the two original housekeeping genes it correlates best with. Optionally, it can be tested whether the housekeeping gene that does not correlate well with the others is affected by treatment. This is done by performing step 7 as if it were a gene of interest.

RESOURCE AVAILABILITY

Lead contact
Further requests for information should be directed to and will be fulfilled by the lead contact, Jonas T. Treebak (jttreebak@sund.ku.dk).

Materials availability
This study did not generate new unique reagents or materials.

Data and code availability
The Excel sheet used to create exemplary figures (“Example qPCR”) as well as the Excel sheet that compares the regular 2^{ΔΔCT} method to the adjusted version herein (“Proof of concept”) have been deposited to Mendeley Data: https://doi.org/10.17632/ytbnp523n9.1.
ACKNOWLEDGMENTS
This work was supported by a PhD scholarship to M.V.D. from the Danish Diabetes Academy, which is funded by the Novo Nordisk Foundation (NNF17SA0031406). Further financial support was provided by Novo Nordisk Foundation Center for Basic Metabolic Research (CBMR). CBMR is an independent Research Center at the University of Copenhagen, which is partially funded by an unrestricted donation from the Novo Nordisk Foundation (NNF18CC0034900). The graphical abstract was created using assets from BioRender.com.

AUTHOR CONTRIBUTIONS
Conceptualization, M.V.D.; Methodology, M.V.D.; Writing – Original Draft, M.V.D.; Visualization, M.V.D.; Writing – Review & Editing, M.V.D. and J.T.T.; Funding Acquisition, M.V.D. and J.T.T.; Supervision, J.T.T.

DECLARATION OF INTERESTS
The Authors declare no competing interests.

REFERENCES
Bustin, S.A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J. Mol. Endocrinol. 25, 169–193. https://doi.org/10.1677/jme.0.0250169.

Damgaard, M.V., Nielsen, T.S., Basse, A.L., Chubanava, S., Trost, K., Moritz, T., Dellinger, R.W., Larsen, S., and Treebak, J.T. (2022). Intravenous nicotinamide riboside elevates mouse skeletal muscle NAD(+) without impacting respiratory capacity or insulin sensitivity. iScience 25, 103863. https://doi.org/10.1016/j.isci.2022.103863.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2−ΔΔCT method. Methods 25, 402–408. https://doi.org/10.1006/meth.2001.1262.