How to analyze qPCR data

PREFACE

This document introduces import and statistical analysis qPCR data in Qlucore Omics Explorer (QOE). It is assumed that the user is familiar with the basic functionality of QOE. In addition to the statistical tests that are highlighted below other functionality such as plots and visualizations are also relevant for qPCR data analysis.

Contents

1. QPCR DATA

Polymerase chain reaction, or PCR, uses repeated cycles of heating and cooling to make many copies of a specific region of DNA, like a gene. Real-time PCR (also called quantitative PCR or shorter qPCR) can be used to detect the amount of a gene that is present in a sample.

The CT (cycle threshold) value of a given gene represents the PCR amplification cycle at which the gene is detected above a certain threshold. It is a relative measure of the concentration of target in the PCR reaction. Higher CT values correspond to lower gene expression levels.

Kommenterad [CI1]: @Jan Nilsson Visst är det CT värden vi ivll att användaren skall importera. Borde inte det skrivas?

QOE is used for data visualization, exploration and analysis of many different types of data, and you can also use it to analyze qPCR data.

2. DATA PREPARATION

qPCR data can be imported both as already normalized values (typically called the delta-CT values, ΔCT) or un-normalized data (CT values). If your data is not normalized, then you can normalize the data using Housekeeping normalization in QOE. Then you import the CT values.

The Wizard can import tab separated .txt files, or comma separated .csv files. If your data is in a Microsoft .xlsx or .xls file you need to open the data and save it in a txt or csv format.

3. IMPORT QPCR DATA FILE

Start QOE, select the file menu and then "Open with Wizard…" . The Wizard and steps are described in detail in the reference manual and in several videos that can be viewed at qlucore.com (Tip: Search for "Wizard" on https://qlucore.com/videos).

The data in the .csv file may be organized as vectors (columns) or as a matrix, the Wizard can handle both.

After import, you save the imported data as a .gedata file.

4. IMPORT VARIABLE LIST WITH HOUSEKEEPING VARIABLES

To be able to do normalization of the qPCR data a list of housekeeping variables must be imported as a list. Such a list can be constructed in Microsoft Excel and saved as a .csv file. You can then import it to QOE, and the variable list will be available as a list in the Variables tab.

A list can be constructed with each housekeeping gene (they must also exist in the qPCR dataset) on individual rows in the file. Below an example where GAPDH, ACTB and 18S rRNA are the three housekeeping variables.

	A
1	GAPDH
2	ACTB
3	18S rRNA

***Figure**: Example of a tab separated txt file with 3 housekeeping variables, one per row*

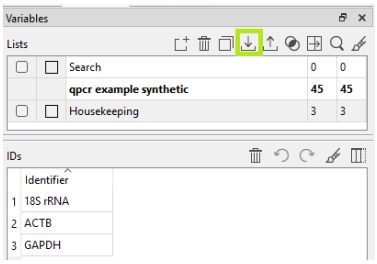

***Figure**: Import of a variable list in the Variables tab.*

5. THE METHODS TAB

After completing the import through the Wizard, you can select to show your data in a table view. This can give you a feeling for the data to be analyzed. To do so, select the Table plot in the Methods tab.

Then switch to the Data tab to prepare your data before analysis begins.

6. THE DATA TAB

6.1. OVERVIEW

The Data tab allows you to make operations on your imported data.

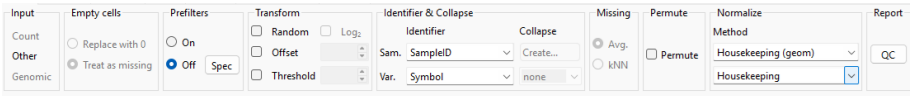

***Figure**: The Data tab*

6.2. EMPTY CELLS OR CELLS WITH MISSING VALUES

Your data may contain cells that are empty. Such a cell may in your data be represented as empty (no value), or as a "NA" or "NULL" value. Only in those cases where you have missing values may you select which method to use for handling missing values.

Such empty cells are in the Table plot indicated with a number in *Italic*. The number is a result of the by default active missing value reconstruction.

An empty cell may indicate that the equipment could not detect anything for the variable for a certain sample and the resulting cell was registered as empty.

If there are empty cells in your data, you can in this step select "Treat as missing". At later stage, under "Missing" in the Data tab, you can select which missing value reconstruction you want to use.

6.3. PREFILTERS

Prefilters allow you to take various actions on your data. You may not need to do any pre-filtering, but it can be useful in some situations. You can select to remove variables that are

missing, and to remove variables with too high or too low values.

6.4. **TRANSFORM**

As a part of the process of normalizing the data, the data is often log transformed. You can do this under Transform.

Log transformation shall only be done if the data has not been log transformed before import.

Your qPCR data may contain cells that have zero values (0). Zero values cannot be log transformed, so before log transformation, the values need to be larger than 0. There are two ways to achieve that:

- Offset. Here you add a value (like 0.01 or 1) to each and every cell in the dataset. This will lift the dataset values and make log transformation possible. The value you add must be high enough to lift every cell value above 0.

- Threshold. Here is set the minimum value of every cell to a fixed value (like 0.01 or 1). Only those cells with a value lower than the Threshold value are affected.

The advantage with the Offset approach compared with the Threshold approach is that distances between values are kept after transformation.

Now you can select to log transform your data. To do so you select the checkbox "Log_2". You can also, as an alternative, select to do the normalization automatically when you select housekeeping normalization based on "geometric mean", which log transforms the data.

6.5. **IDENTIFIER & COLLAPSE**

Your data may have variables that are not unique. In this step you make your variables unique by selecting a method to calculate the collapsed value (using max, min, average, median or sum) of the variables that are not unique.

You can also change the variable identifier here, in which case a collapse may be a logical step thereafter.

Here you can also do a collapse on the sample side. As an example, if you have replicates from the same patient, you may change the sample identifier from "Sample" to "Patient". You can then collapse on the sample side (max, min, average or median). If you select to collapse on the sample side, you will get a new dataset.

6.6. **MISSING**

Here you can select which missing value re-construction method to use, when missing values are imputed. You can (depending on data set size) select between an average calculation or kNN.

6.7. **NORMALIZE**

qPCR data is often normalized using housekeeping normalization. If your data has not been log transformed in a previous step (you have not log transformed the data before import and you have not used the log transform), select the "Housekeeping (geom)" version of the housekeeping normalization option, otherwise just "Housekeeping".

To select “Housekeeping (geom)” or “Housekeeping” you are urged to specify the variable list with housekeeping variables that you would like to use for normalization.

6.8. EXAMINE THE TABLE VIEW

Now when you feel you are ready, have a look in Table plot, where you see the actual values.

6.9. SAVE THE DATASET

Remember to save the dataset after your final changes, you do that by selecting “Save as” under “File” in the top menu of the Program. If you use the same file name as before you will overwrite the originally imported dataset.

7. ADJUSTMENT OF SAMPLES

Go to the Samples tab to examine your samples and sample annotations.

The dataset may have control samples included in the same run as your own samples. These control samples can be excluded from the analysis. You can then uncheck these samples in the Samples tab.

8. ADD MORE ANNOTATIONS TO YOUR SAMPLES

You may have additional information about your samples, information that is not available in data you have imported.

You can import such samples annotation to your dataset. You can create the information in Microsoft Excel and save the information in a tab separated txt file or a csv file and then import the sample annotations into QOE.

To add the new annotation information, go to “File” in the top menu, select “Import” and then “Sample Annotation via Wizard …”.

9. ADJUSTMENT OF VARIABLES

Sometimes you do not want to have the housekeeping variables active in the dataset. You can then omit these variables from the analysis. In order to do so you can use “difference” in the “Set Operations” to create a new variable list which does not include the housekeeping variables.

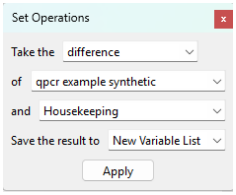

***Figure**: Set operations to create a new variable list*

You can then activate the list to deactivate the Housekeeping variables. Below the name has been updated to “qPCR without Housekeeping”.

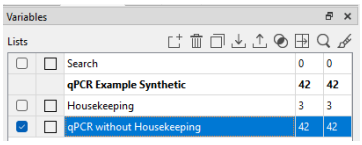

Figure: *Activate the list of variables excluding the housekeeping genes*

10. FINAL DATASET TO BE ANALYSED

Now you should have a final dataset ready for analysis. If you want to save the dataset without the samples and variables you have removed, you can do so. Then go to "Save as" under "File" in the top menu of the Program.

Select a new name for the dataset and select to save only the active samples and variables.

11. DATA OVERVIEW IN PLOTS

A qPCR data set has a limited number of variables, and a heatmap often gives a good overview of the data. You can change the plot to a heatmap in the Methods Tab. You can also select how you want to order the samples, perhaps based on a Sample annotation that you intend to use for statistical analysis. You can also order the variables, one idea is to order them using hierarchical clustering, to better see correlations between variables.

You can also get a feeling for the individual variables in for instance Bar plots and Violin plots.

12. STATISTICAL TESTS

12.1. TEST SELECTION

In QOE you have several two group tests, multi group tests and regression tests to select in between. Below a few examples.

12.2. A TWO GROUP TEST

A two-group comparison can be used to test the significance of the difference between two conditions. QOE supports several two group tests: t-test, Welch test and Mann Whitney U Test (Wilcoxon Rank Sum Test).

The standard t-test assumes that the input values are normally distributed and the variance between the groups is comparable. The Welch test is an alternative when one group has many more samples than the other.

The Mann Whitney U Test (Wilcoxon Rank Sum Test) is used when data is not normally distributed.

If you would like to eliminate the influence from covariates, you can do so for t-tests and Welch tests. You can select the sample annotations that you would like to eliminate on.

12.3. A REGRESSION TEST

If you have an annotation with numerical or ranked values, you can select to do a regression test.

In the case you have an annotation with numerical values, you can use a linear regression test or a quadratic regression test. If you have an annotation with ordered, but not numerical values, you can instead select to run a rank regression test.

13. ACKNOWLEDGEMENTS

Excel is a trademark of Microsoft Corporation.

14. DISCLAIMER

The contents of this document are subject to revision without notice due to continuous progress in methodology, design, and manufacturing.

Qlucore shall have no liability for any error or damages of any kind resulting from the use of this document.

Qlucore Omics Explorer is only intended for research purposes.