Introduction to data analysis in R

Quantifying mRNA using the pcr package

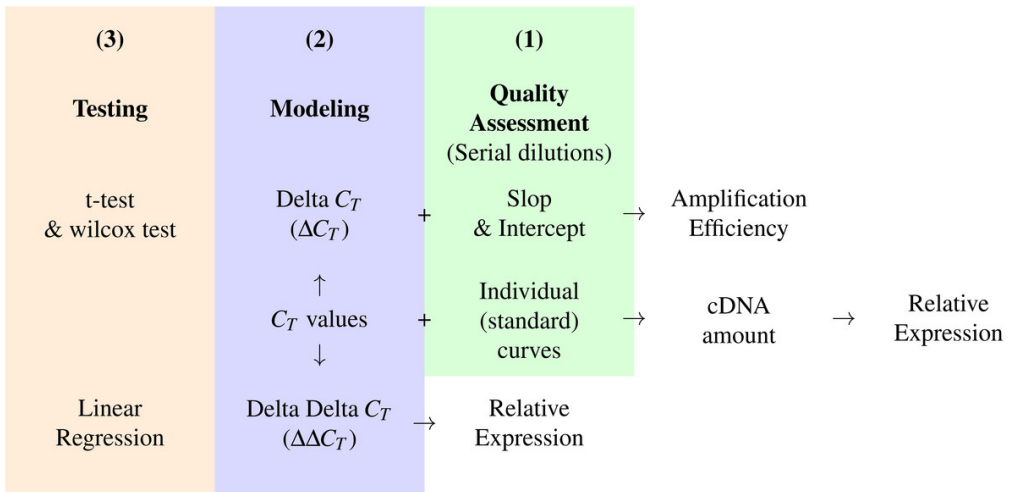
Mahmoud Ahmed

April 2022

Quantitative PCR (qPCR)

Quantitative real-time PCR is an important technique in medical and bio-medical applications.

The pcr package provides a unified interface for quality assessing, analyzing and testing qPCR data for statistical significance.



We will be focusing only on parts 2 and 3 (Modeling and testing.)

Double delta $C_T\Delta\Delta C_T$ model

The comparative C_T methods makes three assumptions:

1. cDNA templates have similar amplification efficiency.
2. Amplification efficiency is near perfect.
3. The expression difference between two genes or two samples can be captured by subtracting one (gene or sample of interest) from another (reference).

This means that at a certain threshold during the linear portion of the PCR reaction, the amount of the gene of the interest and the control double each cycle.

Double delta $C_T \Delta\Delta C_T$ model

The $\Delta\Delta C_T$ is given by:

$$\Delta\Delta C_T = \Delta C_{T,q} - \Delta C_{T,cb}$$

And the relative expression by:

$$2^{-\Delta\Delta C_T}$$

where $\Delta C_{T,q}$ is the difference in the C_T of a gene of interest and a reference gene in a group of interest. $\Delta C_{T,cb}$ is the the difference in the C_T of a gene of interest and a reference gene in a reference group

And the error term is given by:

$$s = \sqrt{s_1^2 + s_2^2}$$

where s_1 and s_2 are the standard deviation of a gene of interest and a reference gene.

The dataset

We will be using the ct1 dataset. It contains values of C-Myc (MYC) and GABDH in 6 brain and 6 kidney tissues samples.

```
ct1 <- read.csv('data/ct1.csv')
head(ct1, n = 3)

##   c_myc GAPDH
## 1 30.72 23.70
## 2 30.34 23.56
## 3 30.58 23.47
```

We need to build a group variable corresponding to rows/samples.

```
# create a group variable
group_var <- rep(c('brain', 'kidney'), each = 6)
head(group_var, n = 3)

## [1] "brain" "brain" "brain"
```

Modeling the relative expression

The goal is to estimate the relative expression of the gene these cells.

First we need to load pcr using the command library.

```
# load library
library(pcr)

# get to the help page of pcr
?pcr

# get to the help page of pcr_analyze
?pcr_analyze
```

Note: output not shown

The required inputs

The function pcr_analyze takes as input:

- A data.frame with columns containing the genes and the rows the C_T values from different samples.
- A grouping variable
- The names of the reference gene and group

Modeling the relative expression

We can put all these pieces together in one function call to pcr_analyze.

```
# calculate all values and errors in one step
pcr_analyze(ct1,
            group_var = group_var,
            reference_gene = 'GAPDH',
            reference_group = 'brain')
```

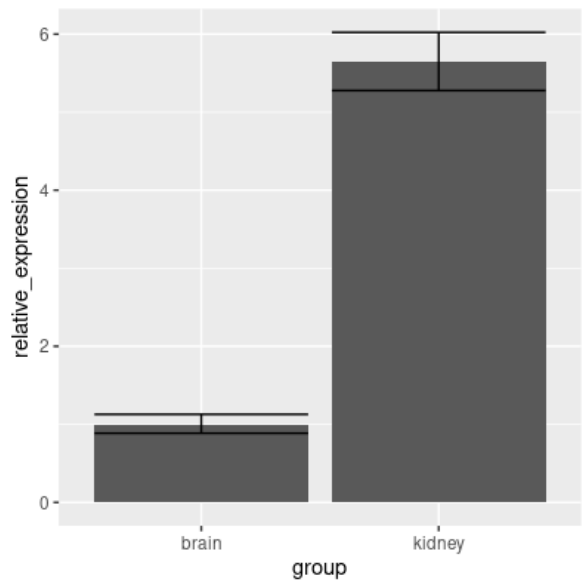
	group	gene	normalized	calibrated	relative_expression	error	lower	upper
1	brain	c_myc	6.860	0.000	1.000000	0.17395402	0.886410	1.128146
2	kidney	c_myc	4.365	-2.495	5.637283	0.09544632	5.276399	6.022850

The output of pcr_analyze is 8 columns (names are self-explainatory!).

Visualizing relative expression

Setting plot = TRUE in pcr_analyze displays the output as a bar graph.

```
# set plot = TRUE in pcr_analyze
pcr_analyze(ct1,
            group_var = group_var,
            reference_gene = 'GAPDH',
            reference_group = 'brain',
            plot = TRUE)
```



Testing the difference in expression

Testing for statistical significance between conditions is important to ensure the validity and replicability of the analysis.

Different statistical methods require different assumptions.

So the choice of which test to use depends on many factors.

1. The number of the conditions/groups
2. The sample and replicate sizes
3. The type of desired comparison

Testing the difference in expression

pcr_test provides a unified interface to different testing methods. We will be using a simple t.test to compare the relative expression of MYC in brain and kidney tissue.

```
pcr_test(ct1,
         group_var = group_var,
         reference_gene = 'GAPDH',
         reference_group = 'brain')
```

```
   group  gene normalized calibrated relative_expression      error    lower    upper
1  brain c_myc      6.860      0.000            1.000000 0.17395402 0.886410 1.128146
2 kidney c_myc      4.365     -2.495            5.637283 0.09544632 5.276399 6.022850
```

References

1. Ahmed M, Kim DR. pcr: an R package for quality assessment, analysis and testing of qPCR data. PeerJ. 2018 Mar 16;6:e4473. doi: 10.7717/peerj.4473.
2. 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 Dec;25(4):402-8. doi: https://pubmed.ncbi.nlm.nih.gov/11846609/.
3. Yuan JS, Reed A, Chen F, Stewart CN Jr. Statistical analysis of real-time PCR data. BMC Bioinformatics. 2006 Feb 22;7:85. doi: 10.1186/1471-2105-7-85.

Summary

What you learned

- Double delta C_T ($\Delta\Delta C_T$) model
- Modeling C_T values using pcr_analyze
- Testing C_T values using pcr_test
- Tests

What's next

- Practice (Link)
- Homework (Link)
- Module 4: Quantifying protein co-localization in fluorescence images using the colocr package (Link)