The role of PTHrP in chondrocyte differentiation.

Hoogendam, J.

Citation

Hoogendam, J. (2006, December 6). *The role of PTHrP in chondrocyte differentiation*. Ponsen & Looijen b.v., Wageningen. Retrieved from https://hdl.handle.net/1887/5422

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/5422

Note: To cite this publication please use the final published version (if applicable).

Confidence bands and p-values for qPCR data using the double delta model (DDM)

6

Jakomijn Hoogendam[1], Geertje van der Horst[2], Jan M. Wit[1], Marcel Karperien[1,2], Paul H.C. Eilers[3]

[1]Department of Paediatrics, Leiden University Medical Center, Leiden, The Netherlands
[2]Department of Endocrinology and Metabolic Diseases, Leiden University Medical Center, Leiden, The Netherlands
[3]Department of Medical Statistics, Leiden University Medical Center, Leiden, The Netherlands

Submitted for publication

Abstract

Background

Several mathematical models exist to calculate data derived from qPCR analysis. However, a model providing computations for expression ratios and p-values in experiments in which all samples are run in triplicate in a single qPCR experiment, well organized in a single spreadsheet, is not available.

Results

We have developed the double delta model (DDM) to calculate p-values and confidence bands in single qPCR experiments in which reference and target samples are run in triplicate. In this model, which is derived from the $2^{-\Delta\Delta Ct}$ method, the variances of the gene of interest as well as the variances of the reference are taken into account.

Conclusions

This model is particularly useful when working with many samples, because the calculations of the ratios and the calculations of the p-values are well organized in a single spreadsheet. Because the DDM does not include corrections for qPCR efficiency, it can only be used when the amplification efficiencies of the target and reference gene are close to 1. The DDM may also be applied to other data sets, in which experimental values are correlated with reference values.

Background

The quantitative Polymerase Chain Reaction (qPCR) is a powerful method to quantify gene expression (for review see[1]). Two different methods to analyze qPCR data are mainly used. The standard curve method is performed in circumstances where it is necessary to determine the absolute transcript copy number[2]. This method is also called absolute quantification and is usually performed by relating the PCR signal to a standard curve. In most situations, however, relative quantification will suffice. This is for example the case in experimental designs in which one would like to examine the effect of a drug or treatment on the expression of a given gene. Many mathematical models exist to calculate the mean normalized gene expression[3-6]. In these methods, the expression of the gene of interest is determined in control and treated samples and related to the expression of a reference gene, which is measured in the same samples. This reference gene is often a house-keeping gene, of which the expression is presumed not to be affected by the treatment.

An appropriate method for calculating expression ratios in data derived from a single qPCR experiment, in which the samples of the gene of interest and the reference gene are run in triplicate, is the $2^{-\Delta\Delta Ct}$ method. The derivations and assumptions have been described elsewhere[5]. Since different efficiencies will generate errors when using this method, the amplification efficiencies of the target and reference gene must be close to 1. This is often the case for amplicons designed to be less than 150 bp and for which the PCR conditions are properly optimized. However, this method does not include calculations of p-values and confidence bands.

A suitable model for calculating expression ratios and p-values in a single qPCR experiment, in which all samples are run in triplicate, is the relative expression software tool (REST). This tool is based on the correction for PCR efficiencies and the mean Ct-value deviation

between sample and control group[6]. In addition, REST also includes statistical analysis by using the pair-wise fixed reallocation randomization test. However, a disadvantage of the REST tool is the requirement of a new spreadsheet for the calculations of each sample in experiments in which the expression of 1 gene is determined in many different samples. This is a lot of work and could be confusing. Examples of such experiments are time series, dose range series or series using different stimuli.
In this letter we propose a novel statistical model, the double delta model (DDM). For calculations of expression ratios, p-values and confidence intervals, using qPCR data derived from a single qPCR experiment in which all samples are run in triplicate. This model is based on the $2^{-\Delta\Delta Ct}$ method[5]. Like the $2^{-\Delta\Delta Ct}$ method, the variances of the Ct-values of the gene of interest as well as the variances of the Ct-values of the reference gene are taken into account. This model does not include corrections for amplification efficiencies, so we want to emphasize that this model can only be used when these values are close to 1. This model is especially useful for calculating differences in gene expression in experiments with many samples. The computations of the ratios and p-values can be calculated in a single spreadsheet, making it well organized and easy to produce graphs.

Implementation

Here, we present a short derivation of the required formulas and show how to implement them as an Excel spreadsheet (table 1). Let *ypcr* be a probe concentration, where *p* indicates the type of probe (reference or not), *c* indicates the condition, and *r* indicates a repetition. Let $X_{pcr} = log_2 Y_{pcr}$. We assume that *x* is approximately normally distributed. This is more realistic than assuming a normal distribution for the concentrations, which are inherently positive and can vary over an enormous range. We also assume that the normal distribution has the same variance for all combinations of *p* and *c*.
The mean of all X_{pcr} for any combination of *p* and *c* will be indicated by

$$mpc = \sum_{r} X_{pcr}/n_{pc},$$

where *npc* is the number of observations. The DDM computes the contrast

$$d = (m_{22}-m_{21})-(m_{12}-m_{11}) = m_{22} - m_{21} - m_{12} + m_{11} .$$

For explanation of symbols, see result section and table 1. The variance in each group can be estimated as

$$vp = \sum_{r} (X_{pcr} - m_{pc})^2 / n_{pc} - 1 .$$

These can be combined to estimate the pooled variance as

$$v = \sum_{p} \sum_{c} v_{pc} / DF ,$$

where

$$DF = \sum_p \sum_c n_{pc} - 1$$

Let se_{pc} indicate the standard error of m_{pc}. It is given by $sepc = \sqrt{v/n_{pc}}$. In other words: the variance of m_{pc} is v/n_{pc}. From elementary mathematical statistics it shows that the variance of a sum or difference of independent variables is equal to the sum of the variances. We thus find:

$$var(d) = v(1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22}),$$

and

$$se(d) = \sqrt{v(1/n11 + 1/n12 + 1/n21 + 1/n22)}$$

Under the assumption of normality, $d/\sqrt{se(d)}$ has a *t*-distribution with degrees of freedom (DF). This allows the computation of confidence bands and p-values. The confidence limits (CL) are computed as: $d \pm \tilde{t}(0.05;DF)*se(d)$, where $\tilde{t}(0.05;DF)$ is the critical value of the *t*-distribution with *DF* degrees of freedom at the 0.05 confidence level.

1	A[1] Group	B[2] Stim.	C[3] Type	D[4] Symbol	E Data1	F Data2	G Data3	H Data4	I Formula
2	ref.	-	PCR 1	x111	21,7	21,7	21,7	21,7	
3	ref.	-	PCR 2	x112					
4	ref.	-	PCR 3	x113	22,1	22,1	22,1	22,1	
5	ref.	+	PCR 1	x121	21,7	21,7	20,9	20,9	
6	ref.	+	PCR 2	x122		21,4	21,3	21	
7	ref.	+	PCR 3	x123	22,1	21,9	21,4	21,5	
8	GOI	-	PCR 1	x211	28,6	28,6	28,6	28,6	
9	GOI	-	PCR 2	x212	28	28	28	28	
10	GOI	-	PCR 3	x213	28,8	28,8	28,8	28,8	
11	GOI	+	PCR 1	x221	28,6	26,6	25,5	27,4	
12	GOI	+	PCR 2	x222	28	26,7	26	27,2	
13	GOI	+	PCR 3	x223	28,8	26,4	26	27,2	
14									
15	ref.	-	Mean	m11	21,90	21,90	21,90	21,90	=AVERAGE(E$2:E$4)
16	ref.	+	Mean	m12	21,90	21,67	21,20	21,13	=AVERAGE(E$5:E$7)
17	GOI	-	Mean	m21	28,47	28,47	28,47	28,47	=AVERAGE(E$8:E$10)
18	GOI	+	Mean	m22	28,47	26,57	25,83	27,27	=AVERAGE(E$11:E$13)
19	ref.	-	Var	v11	0,080	0,080	0,080	0,080	=VAR(E$2:E$4)
20	ref.	+	Var	v12	0,080	0,063	0,070	0,103	=VAR(E$5:E$7)
21	GOI	-	Var	v21	0,173	0,173	0,173	0,173	=VAR(E$8:E$10)
22	GOI	+	Var	v22	0,173	0,023	0,083	0,013	=VAR(E$11:E$13)
23	ref.	-	Count	n11	2	2	2	2	=COUNTIF(E$2:E$4, ">-9e99")
24	ref.	+	Count	n12	3	3	3	3	=COUNTIF(E$5:E$7, ">-9e99")
25	GOI	-	Count	n21	3	3	3	3	=COUNTIF(E$8:E$10, ">-9e99")
26	GOI	+	Count	n22	3	3	3	3	=COUNTIF(E$11:E$13, ">-9e99")
27									
28	Pooled	Pooled	SSQ	s	0,853	0,600	0,733	0,660	=(E23-1)*E19+(E24-1)*E20 +(E25-1)*E21+(E26-1)*E22
29	Pooled	Pooled	DF	DF	6	7	7	7	=SUM(E23:E26)-4
30	Pooled	Pooled	Var	v	0,142	0,086	0,105	0,094	=E28/E29
31									
32	Contrast		Mean	d	0,00	1,76	1,93	0,43	=(E17-E15)-(E18-E16)
33	Contrast		SE	se(d)	0,49	0,36	0,40	0,38	=SQRT(E30*(1/E23+1/E24+1/E25+1/E26))
34	Contrast		t	t	0,00	4,65	4,88	1,15	=E32/E33
35	Contrast		p-value	P	1,00	0,002	0,002	0,287	=TDIST(ABS(E34),E29,2)
36	Contrast		t crit	t(0,05;DF)	2,45	2,36	2,36	2,36	=TINV(0,05,E29)
37	Contrast		CL-		-1,19	0,82	1,00	-0,46	=E32-E36*E33
38	Contrast		CL+		1,19	2,51	2,87	1,32	=E32+E36*E33

Table 1: Example of the double delta model (DDM).
[1]) ref. = reference and GOI = gene of interest. [2]) Stim. = stimulation. [3]) var = variance. SSQ = sum of squares. DF = degrees of freedom. SE = standard error. CL = confidence limit. [4]) Symbols used in the mathematical equations and corresponding with the name in column C. xabc = Ct-value of group (a), stimulation (b), and type (c). mab/vab/nab = mean/variance/counts of group (a) and stimulation (b).

Results

We want to illustrate this model with an example in which the effect of treatment with a drug is examined on the expression of a particular gene at several time points. RNA is isolated from untreated (time point of stimulation) and treated specimens and prepared for qPCR analysis. Samples are run in triplicate in a single qPCR experiment and Ct-values are calculated for the gene of interest and the reference gene (see column E, F, G, and H table 1). Using the DDM we show that the drug has a significant effect on gene expression after 1 (p = 0.002) and after 2 hours (p = 0.002) (table 1 and fig. 1).

The computations can conveniently be organized as an Excel spreadsheet, as shown in table 1. The numbers in one data set are contained in one column (here columns E, F, G, and H). The upper rows (2– 13) hold the individual measurements (Ct-values). Samples (groups) are indicated in column A. Stimulation, where appropriate, is indicated in column B. Column C gives a name to each number in columns E, F, G, and H, while column D shows the correspondence with symbols in the mathematical equations. Column I shows the Excel formulas. This is done here for illustration purposes; in practical use this column (and perhaps also the fourth, containing the symbols) could probably be deleted. The middle rows (15–26) contain the calculations of mean, variance and sample size for each of the four groups. The lower block (rows 28–38) contains the pooled results and the final statistics, labelled as "contrast". The present spreadsheet assumes a maximum group size of 3. It is not difficult to add extra rows for larger sizes. Smaller group sizes or missing data are no problem: one simply leaves spreadsheet cells empty; the calculations are organized in such a way that they automatically handle missing data. In addition, it is not difficult to add extra columns for additional data sets. Figure 1 shows the graph with the mean expression of a gene of interest (GOI) at 0 (data1), 1 (data2), 2 (data3) and 4 (data4) hours after stimulation, compared to the mean expression at time point 0, described in table 1.

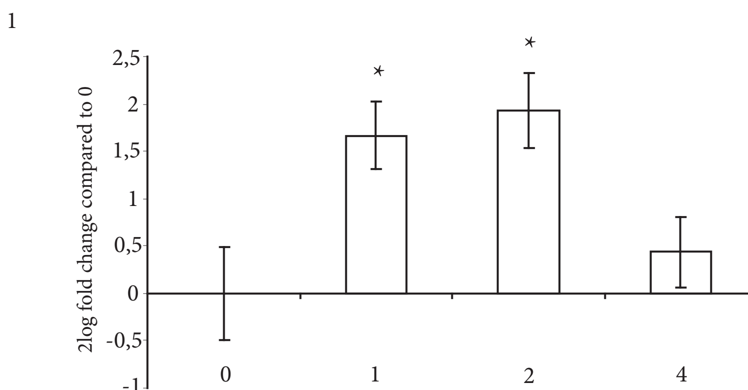

Figure 1. Expression of gene of interest.
The mean expression of the gene of interest (GOI) is shown as 2log fold change compared to 0, time point of stimulation (data1, table 1) at 1 (data2, table 1), 2 (data3, table 1) and 4 (data4, table 1) hours after stimulation. *Significant $p<0.05$.

Conclusions

Several mathematical methods exist to calculate data derived from qPCR analysis[3-6]. However, a model providing calculations for expression ratios and p-values in experiments in which all samples are run in triplicate in a single qPCR experiment, well organized in 1 spreadsheet, is not available. Therefore, we have developed a novel statistical model, the double delta model (DDM) to calculate p-values and confidence bands using data derived from a single qPCR experiment. The DDM is particularly useful when working with many samples. In addition, the DDM is also applicable to other data sets, in which experimental values are correlated with reference values. Examples are experiments, in which enzymatic activity is corrected for cell number or DNA concentration and transient transfection assays, in which luciferase activity is corrected for the transfection efficiency.

Availability

The double delta model is freely available on request from the corresponding author.

Authors' contributions

PE developed the algorithms for the DDM. GH, JW and MK provided biological insight and actively participated in discussion of the project. JH led the project and wrote the paper. All authors read and approved the final manuscript.

Reference list

1. Wong ML, Medrano JF 2005 Real-time PCR for mRNA quantitation. Biotechniques 39:75-85

2. Rutledge RG, Cote C 2003 Mathematics of quantitative kinetic PCR and the application of standard curves. Nucleic Acids Res 31:e93

3. 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

4. Simon P 2003 Q-Gene: processing quantitative real-time RT-PCR data. Bioinformatics 19:1439-1440

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. Pfaffl MW, Horgan GW, Dempfle L 2002 Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res 30:e36