Available online at www.sciencedirect.com
SCIENCE DIRECT
Computational Biology and Chemistry
Computational Biology and Chemistry 30 (2006) 21–26
www.elsevier.com/locate/compbiolchem

Statistical models in assessing fold change of gene expression in real-time RT-PCR experiments

Wenjiang J. Fu [a,*], Jianbo Hu [b], Thomas Spencer [b], Raymond Carroll [c], Guoyao Wu [b]

[a] *Department of Epidemiology, Michigan State University, East Lansing, MI 48824, USA*
[b] *Department of Animal Science, Texas A & M University, College Station, TX 77843, USA*
[c] *Department of Statistics, Texas A & M University, College Station, TX 77843, USA*

Received 3 June 2005; received in revised form 5 October 2005; accepted 5 October 2005

Abstract

Real-time RT-PCR has been frequently used in quantitative research in molecular biology and bioinformatics. It provides remarkably useful technology to assess expression of genes. Although mathematical models for gene amplification process have been studied, statistical models and methods for data analysis in real-time RT-PCR have received little attention. In this paper, we briefly introduce current mathematical models, and study statistical models for real-time RT-PCR data. We propose a generalized estimation equations (GEE) model that properly reflects the structure of repeated data in RT-PCR experiments for both cross-sectional and longitudinal data. The GEE model takes the correlation between observations within the same subjects into consideration, and prevents from producing false positives or false negatives. We further demonstrate with a set of actual real-time RT-PCR data that different statistical models yield different estimations of fold change and confidence interval. The SAS program for data analysis using the GEE model is provided to facilitate easy computation for non-statistical professionals.

Keywords: Correlation; Differential expression; Longitudinal studies; Repeated samples

1. Introduction

Real-time reverse-transcript (RT) polymerase chain reaction (PCR) has gained increasing popularity in quantitative research of molecular biology and bioinformatics (Walker, 2002). It has now become a standard technique for assessing expression of genes that potentially have impact on disease pathology or health prognosis. In RT-PCR experiments, total RNA is extracted from tissues or cells in response to different treatments (e.g. various levels of irradiation, dietary nutrient intake or therapeutic agents), or different physiological and pathological conditions (e.g. age, pregnancy, lactation, and tumorigenesis), and is then reverse transcribed and amplified through DNA polymerase. In the process of RT-PCR analysis, specific primers are used to amplify target genes through numerous cycles, where each cycle consists of three steps, denaturation, annealing, and extension. During the amplification process, one copy of the target gene is doubled and then quadrupled, and is amplified exponentially to 2^n copies after n cycles.

In real-time RT-PCR experiments, fluorescence techniques, such as Taqman, SYBR green and Beacons, are used to detect the amplification of target genes and to assess their expression levels. A fluorescence threshold (ΔR_n) is usually predetermined for an experiment. A cycle threshold (C_T), the value of cycle at which the fluorescence achieves the predetermined threshold (ΔR_n), can then be determined for each target gene through monitoring the fluorescence in individual PCR wells or tubes, where target genes are amplified in separate wells or tubes. The C_T values of target genes are recorded and output with the computer system connected to the PCR. Since the cycle threshold and the logarithm of the input expression level of one specific gene form an inverse linear relationship, a unique line is determined for each target gene as shown in Fig. 1. Different genes may have different slopes. Such a linear relationship can be used to determine the level of absolute gene expression in the input RNA sample if the slope is known or can be calculated (Pfaffl, 2001). However, the slope also depends on experimental factors which may vary from one experiment to another. In an RT-PCR experiment, usually the C_T value corresponding to the threshold (ΔR_n) is recorded for each target gene, which is not enough to determine the slope and thus the level of absolute

* Corresponding author

1476-9271/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiolchem.2005.10.005

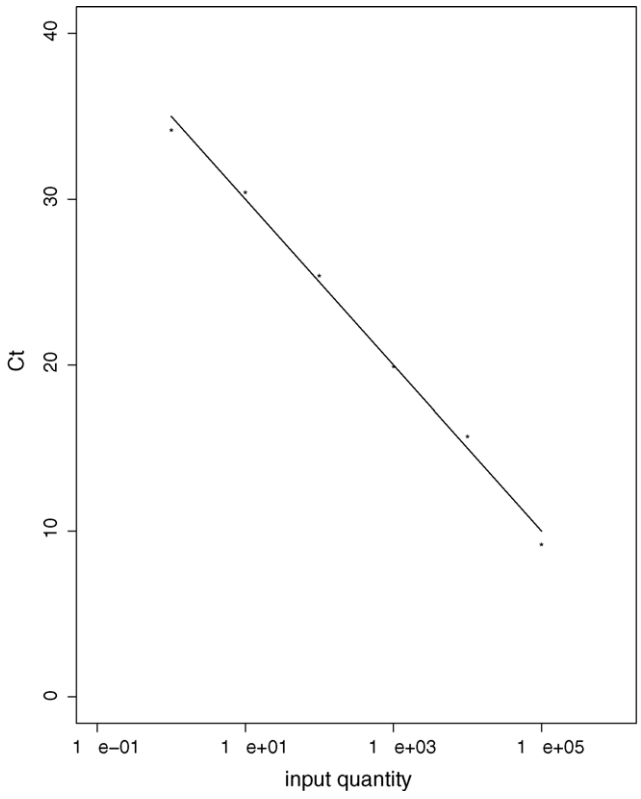

Fig. 1. The inverse linear relationship between cycle threshold (C_T) and the logarithm of input quantity of target gene demonstrated with hypothetical data. The horizontal axis is plotted in logarithm scale.

gene expression in the input RNA sample. Fortunately, the relative expression of target genes in terms of the expression ratio can be determined on the basis of C_T values under certain assumptions, and the differentiation of gene expressions and the fold change between control and treatment groups can be further determined. To validate the C_T readings of RT-PCR, an endogenous reference gene is usually used in the experiment to adjust for any potential unaccounted variation or bias.

Since RT-PCR is used frequently in quantitative research to assess fold changes of gene expression, statistical methods of computing such fold changes play a critical role to ensure the validity of results. Very often biologists extract repeated samples from the same tissue to ensure protocol compliance or experimental consistency. However, statistical models or methods that fail to reflect such a repeated sample structure result in misleading conclusions and may potentially produce either false positives or false negatives. In this article, we briefly discuss current mathematical models used in RT-PCR studies. We then propose a generalized estimating equations (GEE) model to analyze RT-PCR data, and demonstrate the practical usefulness of the GEE method using a real data set collected from an actual RT-PCR experiment. We further provide comparison of the GEE model with a few other models to compute fold changes of genes. Finally we conclude that the GEE model properly reflects the repeated sample structure either in cross-sectional or longitudinal studies and yields correct results.

2. Materials and methods

2.1. RT-PCR data

We recently conducted a nutritional study involving dietary supplementation of arginine [1.51% arginine–HCl or 2.55% alanine (isonitrogenous control) in drinking water] to 9-week-old male Zucker diabetic fatty (ZDF) rats (Fu et al., 2005). At the end of a 10-week supplementation period, various tissues were isolated from control and arginine-treated 19-week-old ZDF rats. Statistical analysis indicated that abdominal and epididymal adipose tissues were 44.5% and 24.7% lower ($P < 0.01$) in arginine-treated ZDF rats when compared with control ZDF rats. The weights of all non-fat tissues (including skeletal muscle, liver, heart, kidneys and brain) did not differ ($P > 0.05$) between control and arginine-supplemented ZDF rats. In search for genes that are differentially expressed and potentially responsible for the difference in abdominal fat mass, a microarray study was conducted to identify the differentially expressed genes in the two groups of ZDF rats, followed by an RT-PCR experiment to confirm the findings of the microarray study. For the RT-PCR study, total RNA samples were extracted from six study rats (three arginine-treated rats and three control rats), using TRIzol reagent (Life Technologies, Gaithersburg, MD). To achieve accurate reading, two repeated samples from each rat tissue were subjected to RT-PCR analysis, using the SYBR Green method and the Amp 5700 Sequence Detection System (Applied Biosystems, Foster City, CA). Table 1 lists the C_T readings of the samples by target genes. Carnitine palmitoyltransferase-2 (CPT-2) was used as an endogenous reference gene and its fold change is thus set at one.

2.2. Mathematical models for RT-PCR

Mathematical models for the amplification of DNA sequences in RT-PCR are based on the following considerations. DNA sequences are amplified in RT-PCR through DNA polymerase. During the amplification process, a copy of target gene doubles in one cycle, and then quadruples in the next cycle. Therefore, the amplification is in the power of 2, namely exponential amplification. An equation describes this process is:

$$Y_n = Y_0 2^n,$$

where Y_0 is the initial expression level of a target gene, and Y_n is the expression level of the gene after n cycles. Because the amplification in an experiment is subject to variations in experimental conditions and may not be 100% efficient and the amplification process may not end with an exact number of cycles, the above equation may be written as

$$Y_t = Y_0(1 + e)^t,$$

where t is the duration of the amplification process in continuous number of cycles, and e is the efficiency, which usually ranges between 0 (completely inefficient) and 1 (fully efficient). The efficiency e may depend on many experimental factors, such as primer concentration, buffer solutions, the fluorescence detection system, and other unknown factors, which may vary with

Table 1
C_T readings of genes from RT-PCR analysis of ZDF rat abdominal adipose tissue

Treatment[a]	Rat no	Sample	CPT-2	AMPK	Calcin	HO-3	ODC	NOS-I	PGC-1α
Ala	1	1	18.67	22.37	17.62	27.45	17.26	31.72	24.15
Ala	1	2	18.62	21.98	17.49	27.39	16.95	32.96	24.19
Ala	2	1	18.12	20.14	17.17	25.77	17.36	–[b]	22.52
Ala	2	2	17.78	20.26	17.35	25.90	17.31	–[b]	22.35
Ala	3	1	18.02	21.24	17.12	26.23	17.33	29.56	24.24
Ala	3	2	18.32	21.60	17.36	26.28	17.42	28.86	24.36
Arg	4	1	19.36	22.04	18.02	24.16	17.99	31.42	21.56
Arg	4	2	19.36	21.16	18.17	23.77	17.54	31.18	21.36
Arg	5	1	18.59	21.12	18.12	24.03	18.11	30.55	21.31
Arg	5	2	18.82	21.40	18.13	23.89	17.94	31.13	21.34
Arg	6	1	18.95	21.23	17.38	23.32	17.77	28.57	21.43
Arg	6	2	18.76	21.24	17.56	23.14	17.74	28.33	21.51

CPT-2, carnitine palmitoyltrrasferase; AMPK, AMP-activated protein kinase; Calcin, calcineurin; HO-3, heme oxygenase-3; ODC, ornithine decarboxylase; NOS-I, nitric oxide synthase-I; PGC-1α: PPARγ coactivator-1α.

[a] Arg, arginine treatment group; Ala, alanine control group.
[b] Missing data.

experiments. Thus, it is important to have an endogenous gene to serve as an internal reference for ensuring the validity of RT-PCR results. Presumably, the less the endogenous gene varies with the study treatment, the better the experimental outcome.

Assume the predetermined threshold for the fluorescence level is Y_T. For the target gene x under treatment, we have the following equation at the threshold C_T:

$$Y_{T,x} = Y_{0,x}(1 + e_x)^{C_{T,x}}. \quad (1)$$

Similarly for the endogenous reference gene R:

$$Y_{T,R} = Y_{0,R}(1 + e_R)^{C_{T,R}}. \quad (2)$$

In the same way, we have for the target gene x under control:

$$Y^*_{T,x} = Y^*_{0,x}(1 + e_x)^{C^*_{T,x}}, \quad (3)$$

and for the endogenous reference gene R under control:

$$Y^*_{T,R} = Y^*_{0,R}(1 + e_R)^{C^*_{T,R}}. \quad (4)$$

The efficiencies e_x and e_R remain the same for both treatment and control groups. Taking the ratio of Eqs. (1) to (3) and the ratio of (2) to (4) yields:

$$Z_{T,x} = Z_{0,x}(1 + e_x)^{C_{T,x} - C^*_{T,x}} \quad (5)$$

and

$$Z_{T,R} = Z_{0,R}(1 + e_R)^{C_{T,R} - C^*_{T,R}}, \quad (6)$$

where $Z_{T,x}$ and $Z_{0,x}$ are the ratios of $Y_{T,x}$ to $Y^*_{T,x}$ and $Y_{0,x}$ to $Y^*_{0,x}$, respectively. Similarly, $Z_{T,R}$ and $Z_{0,R}$ are the ratios of $Y_{T,R}$ to $Y^*_{T,R}$ and $Y_{0,R}$ to $Y^*_{0,R}$, respectively. The model by Pfaffl (2001) can thus be derived with Eq. (5) and (6) by setting $E_{target} = 1 + e_x$ and $E_{ref} = 1 + e_R$.

If only the C_T values of the target and reference genes at the predetermined threshold are recorded and no other information is available, the efficiencies e_x and e_R cannot be determined and thus assumptions are needed in order to calculate the fold changes of genes. Under the assumption of equal efficiency for both target and reference genes, i.e. $e_x = e_R = e$, we have the model:

$$Z_0 = K(1 + e)^{-\Delta C_T} \quad (7)$$

with $\Delta C_T = C_{T,x} - C^*_{T,x}$ and $\Delta C_T = C_{T,R} - C^*_{T,R}$ for target and reference genes, respectively. K is a constant and remains invariant with respect to genes, but may vary from one experiment to another. Thus, if full efficiency is assumed with $e = 1$ as in the User Bulletin # 2 for ABI Prism 7700 Sequence Detection System (Applied Biosystems Inc., Foster City, CA), the fold change of the target gene expression between treatment and control is calculated with:

$$2^{-\Delta\Delta C_T}, \quad (8)$$

where $\Delta\Delta C_T = (C_{T,x} - C^*_{T,x}) - (C_{T,R} - C^*_{T,R}) = (C_{T,x} - C_{T,R}) - (C^*_{T,x} - C^*_{T,R})$. The term $\Delta\Delta C_T$ measures the relative change of expression of gene x from treatment to control compared to the reference gene R.

2.3. Statistical models and methods

Although calculation of the relative change $\Delta\Delta C_T$ and the fold change in Eq. (8) is straightforward for many target genes, estimation of these quantities and standard errors depends on statistical models and may vary largely. So far, we are not aware of any statistical models or methods published in peer-reviewed journals for the estimation of the change ($\Delta\Delta C_T$) and its standard error. The method for calculating the standard error of the fold change provided in the User Bulletin #2 (Applied Biosystems Inc.) is based on an incorrect formula on the coefficient of variation for a ratio of two random variables (page 34 of the User Bulletin, Applied Biosystems Inc., 1997). We compare three simple statistical models for RT-PCR data, the model averaging repeated samples, the independent sample model and the GEE model, with the above data set on gene expression in abdominal adipose tissue of ZDF rats. We demonstrate that for given readings of C_T values, different statistical models lead to different estimations, and improper statistical models that fail to reflect the structure of repeated data may potentially lead to false positives or false negatives. We conclude that the GEE

model reflects the repeated sample structure either in cross-sectional or longitudinal RT-PCR data and thus leads to correct estimation.

2.4. Generalized estimating equations model

Although RT-PCR data are usually obtained from cross-sectional studies (i.e. experimental tissues obtained at the same time point), longitudinal studies using microarrays and other technologies become more and more popular (Zhu and Hero, 2005). It is known that data collected from the same subjects either repeatedly in cross-sectional experiments or in longitudinal experiments are correlated. Statistical models that fail to address such a structure of repeated data yield incorrect conclusion. Liang and Zeger (1986) proposed generalized estimating equations (GEE) model to incorporate correlation structure into the model by specifying a working correlation structure between observations of the same subjects. One advantage of this approach is that although the working correlation structure may be specified in many different ways, the parameter estimations obtained from such a model are consistent. That means the GEE model results in correct estimation of the parameters even with incorrect specification of correlation structure. We now briefly introduce the GEE model for correlated observations. More details can be found in the original paper by Liang and Zeger (1986).

Assume that a longitudinal study has K subjects. Each subject $k = 1, \ldots, K$ has n_k observations Y_{ki} with $i = 1, \ldots, n_k$ with corresponding covariate representing demographic and clinical conditions x_{kj}, $j = 1, \ldots, p$. If the investigators are mainly interested in how the response variable Y_{ki} varies with the covariates, Liang and Zeger (1986) proposed the following GEE model based on marginal distribution $f(y_{ki}) = \exp[\{y_{ki}\theta_{ki} - b(\theta_{ki}) + a(Y_{ki})\}\phi]$.

$$\sum_{k=1}^{K} D_k V_k^{-1} S_k = 0, \tag{9}$$

where $V_k = A_k^{1/2} R(\alpha) A_k^{1/2}/\phi$ is the working covariance matrix, $D_k = \mathrm{d}\{b_k'(\theta)\}/\mathrm{d}\beta = A_k \Delta_k X_k$, $\Delta_k = \mathrm{diag}(\mathrm{d}\theta_{ki}/\mathrm{d}\eta_{ki})$, and $S_k = y_k - b_k'(\theta)$. The GEE estimator $\hat{\beta}$ is consistent, i.e. the estimation converges to true parameter value as sample size K increases to sufficiently large or infinity.

The GEE model was developed initially for longitudinal studies where multiple observations along time were taken from each subject (Liang and Zeger, 1986). The GEE model can also be applied to clustered data where correlated observations may not be ordered in time or other factors, such as survey data within household. The GEE model can be run with the SAS GENMOD procedure as shown in the Appendix A.

2.5. Comparison of statistical models and computational methods

We fit three statistical models to the ZDF rat RT-PCR data in Table 1, and compare the results.

2.5.1. Model 1: simple model averaging repeated sample readings

Since two repeated samples were drawn from the same animal fat tissue, averaging the C_T readings of the two repeated samples achieved accuracy in C_T values for each rat tissue. Further analysis was thus based on the averaged readings. Although such an approach improved accuracy for the reading of each animal tissue, information is lost by averaging out the raw data within each animal, resulting in a smaller sample size and larger variability. The larger variability led to larger standard errors and larger confidence intervals of the change in expression levels. Thus, this method is not preferred unless the experiment is not carefully designed and has largely unequal numbers of repeated samples for different animals.

2.5.2. Model 2: ANOVA model assuming independent samples

Although it is known that repeated samples from the same rat tissue are correlated, such correlation may often be overlooked in practical data analysis by non-statistical professionals. A model that treats all samples independent and ignores such a correlation is expected to yield different results. This model regarded all six samples in each treatment group as independent samples, and the sample size was thus incorrectly considered to be 12 for the two treatment groups although the two repeated samples from the same animal tissue were highly correlated. The effect of the endogenous reference gene CPT-2 was adjusted by subtracting the C_T readings of CPT-2 from those of the target genes. Note that the adjustment of the reference gene CPT-2 by incorporating its C_T reading as a model covariate does not comply with the mathematical models (5) and (6) because this adjustment almost surely yields a parameter estimate for the effect of CPT-2 not equal to one, which makes the model adjustment invalid.

2.5.3. Model 3: GEE model accounting for correlation between repeated samples

To account for the correlation between repeated samples from the same rat tissue, we fitted a generalized estimating equations (GEE) model to the RT-PCR data. The GEE model with an exchangeable correlation structure was assumed and the identity link function for a normal random variable was specified. The exchangeable correlation structure specified that the correlation between any two distinct repeated samples from the same rat tissue remains the same regardless of the animals. This special correlation structure is particularly useful for cross-sectional studies with repeated samples because samples from the same animals are not ordered. The GEE model was fitted with the SAS procedure PROC GENMOD (The SAS Institute, 2000), which yielded treatment effect for the C_T change. Since the GEE model accounted for the correlation between repeated samples, its estimated variability was between those of Models 1 and 2.

3. Results and discussion

Table 2 displays the treatment effect ($\Delta\Delta C_T$) and its standard errors estimated for different target genes by Models 1–3. Model 1 yielded the same values in estimation of $\Delta\Delta C_T$ as Model 3.

Table 2
Estimation (S.E.) of treatment effect ($\Delta\Delta C_T$) between arginine and alanine supplementations

Gene	Model 1	Model 2	Model 3	Fold[a]	Correlation[b]
AMP-activated protein kinase	−0.62 (0.399)	−0.87 (0.309)*	−0.62 (0.336)	1.54	0.67
Calcineurin	−0.17 (0.276)	−0.20 (0.455)	−0.17 (0.226)	1.13	0.72
Heme oxygenase-3	−3.50 (0.377)*	−3.71 (0.218)*	−3.50 (0.308)*	11.3	0.89
Nitric oxide synthase-1	−1.14 (1.445)	−1.45 (0.534)*	−1.14 (1.151)	2.20	1.00
Ornithine decarboxylase	−0.14 (0.387)	−0.31 (0.221)	−0.14 (0.316)	1.10	0.90
PPARγ coactivator-1α	−2.94 (0.510)*	−3.52 (0.136)*	−2.94 (0.417)*	7.67	1.00

[a] Fold change calculated based on Model 3, which incorporates the correlation between samples from the same tissues.
[b] Correlation coefficient between repeated samples estimated by the GEE model.
* Statistically significant with $p < 0.05$.

However, Model 1 averaged the repeated samples from each rat tissue, resulted in loss of information, and thus yielded slightly larger variability with larger standard errors compared to Model 3. If the numbers of animals and repeated samples were moderate or large, this loss of information may be large enough to have led to false negatives. Namely, genes that are truly differentially expressed may be claimed as being statistically nonsignificant due to larger standard errors than true values. Although Model 1 yielded correct estimation of the change in gene expression levels, it is somewhat too conservative in identifying statistical significance.

On the contrary, Model 2 yielded very different effect estimation, because it treated all samples as independent and ignored the correlation between the samples from the same rat tissue. As shown in Table 2, Model 2 yielded statistically significant changes in AMPK and NOS expression in addition to HO-3 and PPARγ coactivator-1α (PGC-1α), while Models 1 and 3 identified statistically significant changes only in HO-3 and PGC-1α. Evidently Model 2 yielded incorrect estimation, and may potentially have produced false positives. Another drawback of Model 2 is that if an experiment is not carefully designed and has largely different numbers of repeated samples for different animals, Model 2 will weigh more on the animals with a large number of repeated samples and will thus yield largely biased estimation.

Compared with Models 1 and 2, Model 3 yielded accurate estimation for the change in gene expression levels as Model 1. It also took into account the correlation between repeated samples from the same rat tissue and thus accurately estimated the variability and the standard errors. The SAS program for Model 3 also yielded an estimate of the correlation coefficient between repeated samples (listed in Table 2). These large correlation coefficients between 0.67 and 1 indicate that a proper correlation structure must be incorporated in the model. Failure in doing so will result in biased estimation as in Model 2.

Since the power function 2^t is nonlinear, confidence intervals for the fold change are preferable than standard errors. Confidence intervals can be calculated with lower limit $2^{-\Delta\Delta C_T - Z_{(1-\alpha/2)}\text{S.E.}}$ and upper limit $2^{-\Delta\Delta C_T + Z_{(1-\alpha/2)}\text{S.E.}}$ based on the upper and lower confidence limits $-\Delta\Delta C_T \pm Z_{(1-\alpha/2)}\text{S.E.}$ of the change $\Delta\Delta C_T$, where S.E. is the standard error of the change $-\Delta\Delta C_T$. $Z_{(1-\alpha/2)}$ is the $100 \times (1-\alpha/2)$-th percentile of the standard normal distribution Normal (0, 1), and α is the level for a $(1-\alpha) \times 100\%$ confidence interval. A standard error of the fold change, if preferred, can be calculated with the statistical delta method (Lehmann and Casella, 1998).

In summary, if repeated samples are taken from the same animal tissues, specifying a GEE model with an exchangeable correlation structure as in Model 3 yields accurate estimation of the change in gene expression and its standard error. Even if unequal numbers of repeated samples are collected from different animals, the GEE model still yields robust and accurate estimation and standard error of the change in gene expression.

4. Conclusion

Real-time RT-PCR has been used frequently in quantitative research in biology and bioinformatics to assess differential expression of genes. Although mathematical models for the fold change of genes are straightforward, different statistical models may yield different estimations and confidence intervals for changes in gene expression. It is thus critical to specify a statistical model that properly reflects data structure. We proposed a GEE model to reflect the repeated sample structure in RT-PCR data, and demonstrated that it yielded accurate estimation and prevented from producing false positives or false negatives.

Acknowledgements

This research was supported, in part, by grants #R25-CA90301 (W. Fu) and #5R01-CA57030 (R. Carroll) from NIH/NCI, by Center for Environmental and Rural Health grant #P30-ES09106 from NIEHS, by a grant #1R01-HD38274 (T. Spencer) from NIH, and by a grant #0255878Y (G. Wu) from the American Heart Association. We thank Dr. Qi Ding for valuable discussion.

Appendix A. SAS program for the GEE model

```
*** Assume the data set in Table 1 is named
    PCR;
Data pcrdiff;
Set pcr;
Respvar = ampk - cpt2;
**** cpt2 serves as the endogenous
    **** reference gene;
**** The Respvar can be changed by
    **** replacing ampk to other genes
**** to compute the change of expression
    **** levels of other genes;
```

```
Run;

Proc sort data = pcrdiff;
By treat rat samp;
Run;

Proc GENMOD data = pcrdiff;
Class treat rat;
Model Respvar = treat /dist = normal link
    = identity type1 type3;
Repeated subject = rat /type = exch
    corrw;
Run;

**** The effect estimate of the treat
    **** effect in the SAS GENMOD
**** output is the value of
**** ΔΔC_T;
**** But be cautious with the sign, which
    **** depends on what level
    **** of treatment is set as the reference
    **** in the SAS output;
```

References

Applied Biosystems, 1997. User Bulletin #2: ABI Prism 7700 Sequence Detection System. Applied Biosystems Inc., Foster City, CA, p. 34.

Fu, W.J., Haynes, T.E., Kohli, R., Hu, J., Spencer, T., Shi, W., Carroll, R.J., Meininger, C.J., Wu, G., 2005. Dietary *l*-arginine supplementation reduces fat mass in Zucker diabetic fatty rats. J. Nutr. 135, 714–721.

Lehmann, E.L., Casella, G., 1998. Theory of Point Estimation, 2nd ed. Springer, New York.

Liang, K-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika. 73, 13–22.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucl. Acid Res. 29, 2002–2007.

The SAS Institute, 2000. Statistical Analysis System, Version 8. The SAS Institute Inc., Cary, NC.

Walker, N., 2002. A technique whose time has come. Science. 296, 557–558.

Zhu, D.X., Hero, A.O., 2005. Identify differentially expressed genes from probe level intensities in longitudinal affymetrix microarray experiments. Available at http://www-personal.umich.edu/zhud/DE.pdf.