Open Access Medical Statistics

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Statistical analysis of repeated microRNA high-throughput data with application to human heart failure: a review of methodology

This article was published in the following Dove Press journal:
Open Access Medical Statistics
12 April 2012
Number of times this article has been viewed

Shesh N Rai[1]
Herman E Ray[2]
Xiaobin Yuan[1]
Jianmin Pan[1]
Tariq Hamid[3,4]
Sumanth D Prabhu[3,4]

[1]Biostatistics Shared Facility, JG Brown Cancer Center and Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, USA; [2]Department of Mathematics and Statistics, Kennesaw State University, Kennesaw, GA, USA; [3]Division of Cardiovascular Medicine, University of Louisville, Louisville, KY, USA; [4]Division of Cardiovascular Disease, University of Alabama – Birmingham and Birmingham VAMC, Birmingham, AL, USA

Abstract: Complex experimental designs present unique challenges in the analysis of microRNA (miRNA) cycle to threshold (*Ct*) values. In this paper, we discuss various statistical techniques and their application in an analysis performed at the JG Brown Cancer Center. We consider data quality evaluation, data normalization, and statistical hypothesis procedures in the context of maintaining patients prior to heart transplantation. The research involved repeated sampling over time, and the intra-subject correlation created by the repeated sampling should be incorporated into the analysis resulting in additional significant miRNAs. The statistical techniques leveraged to analyze miRNA *Ct* values resulting from qPCR should incorporate key features of the experimental design. When an experiment collects multiple samples from the same individuals over time this may cause issues with the commonly used methodologies – these issues are discussed.

Keywords: miRNA, repeated measurements, normalization, hypothesis testing

Introduction

Several studies have examined the role of microRNAs (miRNAs) in various diseases such as cancer[1] and heart disease.[2] miRNAs are short, noncoding RNA molecules that affect gene expression. The clinical understanding of the role of miRNAs in disease is growing very quickly. Several techniques including microarray analysis and TaqMan polymerase chain reaction (PCR) from Applied Biosystems may be used to obtain the expression levels of the miRNAs.[3] The reproducibility of experiments performed with Taq-Man PCR has been found to be high.[4] There are also several different normalization techniques that can be employed to remove systematic differences between samples that do not represent true biologic differences.[5] The *Ct* value represents the cycle number at which the fluorescent signal of the reporter dye crosses a threshold value.[6] The threshold is placed such that the PCR is in the exponential phase.

Typically, a hypothesis-testing procedure is applied once the *Ct* values are normalized. Student's *t*-test is a popular procedure for comparing the mean of the normalized *Ct* values between the two groups.[7–9] Resulting *P* values need to be adjusted to control the Type I error rate using an appropriate method such as the Benjamini–Hochberg[10] method.

Experiments in this field are becoming more complex as they are designed to examine the relationships between the disease process, treatments, and the expression of miRNAs. They often involve repeated measurements on the same subjects over time and require specialized statistical techniques to handle the additional correlation. Montenegro et al[11] developed an experiment that examines the expression of miRNAs

Correspondence: Shesh N Rai
Clinical and Translational Research Building, Room 211, 505 South Hancock Street, Louisville, KY 40202, USA
Tel +1 502 852 4030
Fax +1 502 852 7979
Email shesh.rai@louisville.edu

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/OAMS.S27907

at different gestational ages. The authors used a generalized estimating equation (GEE)[12] model:

$$g(E[Y_{ijk}|x_{ij}]) \tag{1}$$

with an exchangeable correlation structure where Y_{ijk} is the kth *Ct* value for the ith subject and the *j*th gestational age. The x_{ij} is the *j*th covariate for the *i*th subject. The model included the obstetric condition and gestational age. The GEE model is in the class of semi-parametric models since it does not require full specification of the likelihood to calculate the parameter estimates. The model is easily applied in the repeated sampling situation created by the qPCR experiment.

The writing of this paper was motivated by an analysis of miRNA *Ct* values performed at the James Graham Brown Cancer Center. The experiment was designed to perform an exploratory analysis of changes in the cardiac expression of miRNAs in patients with end-stage heart failure (HF) undergoing placement of a left ventricular assist device (LVAD) and subsequent heart transplantation. The experimental design presented some unique challenges in the analysis of the data that require a description of the experiment for full appreciation. The remaining sections will describe the experiment, compare various analysis techniques, discuss the results, and provide conclusions.

Motivating example

The experiment that inspired the current paper was designed to analyze miRNA expression profiles in patients with advanced HF undergoing surgical implantation of an LVAD (a mechanical pump designed to assist in blood flow from the weakened heart) as a bridge to heart transplantation, ie, to maintain the patient until the heart could be replaced.

The initial assessment of miRNA expression levels was an exploratory analysis of 384 unique miRNAs. The selected miRNAs were selected based on resources and what was known at the time of experimentation about their features in heart functioning. Each subject (that is, each patient under study) had a sample of the left ventricle removed at the time of LVAD implantation (IMP), a sample taken of the left ventricle at the time of heart transplant and LVAD explantation (ELV), and a sample taken of the right ventricle at the time of LVAD explant (ERV). Therefore, each subject receiving an LVAD in the study contributed three samples at two different times.

Each of the three biologic samples that the subject contributed is referred to as a plate, this being the specific set of miRNAs contributed by a subject from a specific point in time and location in the heart. All participants signed an informed consent form for the use of the tissue and the study was approved by the University of Louisville's Institutional Review Board (IRB, IRB# 101.04 JH). There are also archived control samples which represent hearts not experiencing failure.

The experiment was intended to be an exploratory analysis and there are a limited number of wells. Therefore, in order to maximize the number of the miRNAs to be included in the analysis, there are no technical replicates. Each of the 384 wells contains a unique miRNA except for the endogenous controls which may be repeated a few times. There are challenges created by the time required to collect the data as well as the multiple time points in the trial. The specific challenges will be discussed in the following section.

Statistical methodologies

The experimental design presents two unique challenges to analysis. First, in order to maximize the number of unique miRNAs that could be included in the experiment, technical replicates were not included. This implies that normal data quality techniques are not available, and a different approach is required. Second, the repeated sampling of miRNAs from the same subjects over time presents a challenge as typical normalization techniques are not designed to preserve naturally occurring correlation structures. However, there are statistical models that can be employed that include the correlation structure.

In this paper, we discuss the most commonly used methodologies or those methodologies with readily available software. We apply them to the data discussed in the motivating example.

Data quality

The quality of the data had to be assessed once this was ready for analysis. Usually, technical replicates are used to assess the quality of *Ct* values, and technical replicates can be used to determine if information is truly missing or missing at random ('missing' was defined as a *Ct* value greater than 35, even though the software can detect values up to 40). If the values are missing at random then an imputation algorithm can be utilized, while truly missing values should be unaltered.

A different approach was required, however, as the experiment was constructed to include as many unique miRNAs as possible, excluding technical replicates. First, the number of plates with values larger than 35 for each miRNA was calculated. All of the plates, regardless of time during treatment, were simultaneously included in the analysis. miRNAs that were missing across a large number of plates were then excluded from the study. This had the benefit of reducing the number of miRNAs included in the hypothesis testing.

Although using a fixed threshold value is suggested and routinely used, this method is subject to selection bias. An alternative is to use a varying threshold for each plate, in which case *Ct* values and varying threshold values must be combined in a parametric model (this is considered in another manuscript).

Normalization

The second issue encountered during the analysis of the *Ct* values was appropriate normalization. Normalization is required to remove unwanted technical variation from the sample.[13] Many of the normalization techniques are developed from the analysis of microarray datasets and may not be completely applicable. The number of measurements is much smaller in miRNA data, and the majority of miRNA are either not expressed or are expressed at very low levels.[5]

In this analysis, the delta-*Ct* method,[14] the mean normalization,[13] quantile normalization,[15] and rank invariant normalization were considered.[16] These normalization techniques are commonly used and the relevant software for this is readily available. The coefficient of variation associated with the raw data is included as a reference against which to evaluate the normalization techniques.

Let N be the total number of subjects included in the study after filtering and $i = 1, \ldots, N$ be the individual patient number. Let $j = 1, 2, 3$ be the repeated sample number for each subject. In the motivating example, $j = 1$ corresponds with the IMP biologic sample, $j = 2$ corresponds with the ELV biologic sample, and $j = 3$ corresponds with the biologic ERV sample. Then $M = 3N$ is the total number of plates included in the experiment where $m = 1, \ldots, M$. We will also let K be the unique count of miRNAs included in the analysis after filtering where k indicates the kth miRNA for $k = 1, 2, \ldots, K$. Then Ct_{ijk} represents the *Ct* value from the ith subject, jth sample, and kth miRNA. For simplicity, the calculations that are plate-specific will be discussed in terms of the subscript m where $m = 1, ..., M$. Note that $M = 3N$ which represents the total number of plates to be analyzed. The calculations that are sample- and person-specific will include all three subscripts i, j, and k.

Delta-*Ct*

The delta-*Ct* method subtracts the mean of the endogenous controls from the remaining *Ct* values. Two endogenous controls were selected for the analysis RNU24 and RNU48. The algebraic equation representing this is:

$$\Delta Ct_{mk} = Ct_{mk} - Ct_{\mathrm{e}} \qquad (2)$$

where Ct_e is the average of the *Ct* values from the endogenous controls and Ct_{mk} is the individual values for all the other miRNAs in the sample. The delta-*Ct* is a popular method of normalization due to the natural biologic motivation, an explanation of which is contained in the Appendix.

Mean normalization

The mean normalization subtracts the average of plate m's *Ct* values from all *Ct* values contained on plate m. The mathematical representation is:

$$\Delta Ct^{m}_{mk} = Ct_{mk} - \frac{\sum_{k=1}^{K} Ct_{mk}}{K}, k = 1, \ldots, K \text{ and } m = 1, \ldots, M \qquad (3)$$

where Ct_{mk} is the kth miRNA from the mth plate and $M = 3N$. The method is similar, in essence, to the delta-*Ct* method but relies on an average of all *Ct* values to perform the normalization.

Quantile normalization

The quantile normalization forces the distribution of *Ct* values to be the same across all the plates. The method takes the largest value and replaces it with the mean of the largest values, and then repeats for each subsequent data point. Let:

$$q_k^* = \frac{1}{\sqrt{M}} \sum_{m=1}^{M} q_{mk}\, d = \left(\frac{1}{M} \sum_{m=1}^{M} q_{mk}, \ldots, \frac{1}{M} \sum_{m=1}^{M} q_{mk} \right) \qquad (4)$$

where

$$d = \left(\frac{1}{\sqrt{M}}, \ldots, \frac{1}{\sqrt{M}} \right) \qquad (5)$$

and q_k is the kth row of ordered *Ct* values. The *Ct* values are ordered for each plate independent of the other plates. The quantile normalization methodology is commonly used in the analysis of microarray expression values but the technique assumes that the distribution of the expression values is the same.

Rank invariant normalization

The rank invariant method attempts to determine miRNAs which have a rank that does not change across the plates. The rank invariant miRNAs are then used to create a smooth curve applied to the entire sample. The rank invariant normalization is completed in two steps. First, the *k*th miRNA is considered to be rank invariant if the absolute value of the change in the relative rank (*r*) of the miRNA in the m_i th plate and the m_j th plate is less than 0.05 or:

$$\frac{\left|r_{m_i k} - r_{m_j k}\right|}{r_{m_j k}} < 0.05. \tag{6}$$

A smooth line is then fitted through the set of rank invariant genes that is applied to all miRNAs. The rank invariant method is another technique resulting from the analysis of microarray expression values.

Coefficient of variation

The cumulative distribution of the coefficient of variation is used to compare the various normalization techniques. We calculated the coefficient of variation for each miRNA over all the plates. Let *K* be the total number of miRNAs then the coefficient of variation is:

$$CV_k = \frac{\text{sd}(Ct)}{\text{mean}(Ct)},\ k = 1, 2, \ldots, K \tag{7}$$

Next we created a cumulative distribution of the coefficient of variation as:

$$\tilde{F} = \frac{1}{K}\sum_{k=1}^{K} I\{CV_k \le t\} \tag{8}$$

Hypothesis testing

The final issue to consider during the analysis was the appropriate hypothesis testing procedure. A *t*-test, Mann–Whitney *U* test, and a testing procedure proposed by Pounds and Rai[18] were considered, as well as a model based on the GEE approach.

The experiment collected data on the same individuals at two different time points. The first sample was taken when the LVAD was implanted but only the left ventricle could be sampled. The second sample was taken when the heart transplant was performed and both the left and right ventricles were sampled.

GEE model

A GEE model was constructed to consider the repeated sampling. The Y_{ijk} is the repeated *Ct* values taken at each of the time points described for each individual. The x_{ij} represent the covariates including the three groups, that is, the sample taken at the time of LVAD implant (IMP), the left ventricle sample at the time of explant (ELV), and the right ventricle sample at the time of explant (ERV). Contrasts were constructed to analyze the difference between the *Ct* values taken at the time of implant and the left ventricle sample at the time of explant, as well as the difference in *Ct* values between the left and right ventricles at the time of the explant. The analysis currently assumes an exchangeable correlation structure that accounts for the correlation between the samples for each subject. The Gaussian distribution with the identity link function was selected given the relatively normal distribution of the normalized *Ct*.

t-test

If the intra-subject correlation is ignored then a paired *t*-test can be employed to compare the average expression values from two of the three samples. The mathematical notation is:

$$t_d = \frac{\sqrt{N}\bar{d}_k}{s_d} \tag{9}$$

where

$$\bar{d}_k = \frac{\sum_{n=1}^{N}(Ct_{i1k} - Ct_{i2k})}{N} \tag{10}$$

and Ct_{i1k} is the *Ct* value for the *k*th miRNA from the first sample for the *i*th subject. The value s_d is the appropriate estimate of the standard error for the paired difference. Under the null hypothesis of no difference, the test statistic t_d follows a *t*-distribution with *N*–1 degrees of freedom.

Mann–Whitney *U* test

The Mann–Whitney *U* test is a non-parameter test that evaluates the population medians based on two samples. The procedure also ignores intra-subject correlation and does not require the assumption that sampling statistics follows the normal distribution. The Mann–Whitney *U* test statistic is:

$$U = N^2 + \frac{N(N+1)}{2} + R_1 \tag{11}$$

where R_1 is the sum of the ranks, based on the entire combined sample, associated with just the first sample.

In both the *t*-test and the Mann–Whitney *U* test, as well as the GEE model, there are total *K* test statistics and corresponding hypothesis tests. A multiplicity adjustment should be applied in order to control the total Type I error rate or the false discovery rate.

Assumption adequacy averaging

The concept of assumption adequacy averaging (AAA) was proposed by Pounds and Rai[18] as a technique for developing more robust methods that incorporate assessments of assumption adequacy into the analysis. The technique utilized empirical Bayesian principles described by Efron et al,[19] as well as Pounds and Morris,[20] to develop a method that averages the results from different testing procedures with weights determined by tests of assumption adequacy. The method combines results from the classical *t*-test and rank-sum tests with weights determined by the Shapiro–Wilk's test to assess the normality assumption.

Sample size justification

Many of these studies are designed on an ad hoc basis – unlike in clinical trials, experiments are not usually planned. However, post-analysis justification of sample size is essential. In high-throughput data analyses, where the number of hypotheses rapidly becomes large, one of the primary objectives is to have a high probability of declaring a hypothesis (such as a specific miRNA) to be significant (differentially expressed) if they are truly significant (truly expressed), while keeping the probability of making false declarations low. There are two approaches to controlling error rates: false discovery rate (FDR) and family-wise error rate (FWER). Following Benjamini and Hochberg (1995),[9] the FDR is the expected value of the proportion of the non-prognostic genes (in our case miRNAs) among the discovered genes (in our case miRNAs). We will use FDR approach for sample size justification.

Three repeat measurements are included in the motivating example, which allowed two pairwise comparisons and a comparison of the overall effect. The pairwise comparison was based on a paired *t*-test. It was determined that a two-sided test was preferable as information was not available regarding whether miRNA were downregulated or upregulated.

Adjusted significance level

Following Chow et al (2008),[21] the adjusted significance level is given as:

$$\propto^{*} = \frac{r_1 f}{m_0(1-f)} \qquad (12)$$

In the above expression r_1 is the desired number of the alternative hypothesis (# of miRNAs to be discovered) to be declared significant at f false discovery rate from m total hypotheses (total # of miRNAs), with m_1 potentially alternative hypotheses (potentially significant # of miRNAs) and m_0 ($= m - m_1$) null hypotheses (not significantly expressed miRNAs). Once the level of significance is determined it is straightforward to determine design parameters (power, effect size, or sample size).

Demonstration to motivating example

Based on previous experience, it was expected that $m_1 = 40$ (around 10% of the 384 miRNAs) of which approximately $r_1 = 10$ miRNAs were expected to be identified. The resulting adjusted significance levels were 0.0015 and 0.0032 at FDR = 5% and 10%, respectively. Using a one-sided paired *t*-test, with n = 9 for 80% power, and significance level of 0.0015 and 0.0032, effect sizes of 1.69 SD (standard deviation) units and 1.52 SD, respectively, could be detected. Assuming equal variances in the repeat measures, unit fold (ratio of means), a quantity most commonly used in the collaborative research, could be identified at 2.69 fold for upregulated miRNA or 0.37 fold for downregulated miRNA at an FDR level of 5%.

Results

The methods described above were applied to the *Ct* values resulting from the motivating example. The first item to consider is the data quality resulting from the qPCR experiment. One component of the analysis considered a comparison between the expression values from the left ventricle at the time of explant (ELV) to the expression values from the right ventricle at the time of explant (ERV). There were nine subjects with the PCR performed on the same platform. miRNAs with missing values on 13 or more of the 18 plates (or more than 72.2%) were excluded from further analysis.

A similar approach was used to reduce the number of *Ct* values greater than 35 in an analysis that considered the three samples (IMP, ELV, and ERV) in one model. miRNAs were excluded from future analysis if the *Ct* values were missing from 19 or more of the 27 plates (or more than 70.4%). After filtering, each individual contributed the same miRNAs at each of the three time points resulting in a balanced repeated design. Table 1 reports the percentage of the plates with *Ct* values larger than 35 for the different comparisons before and after filtering. In each case, the percentage of reasonable *Ct* values less than 35 increased.

Table 1 Effect of filtering on percentage of *Ct* values deemed undetermined

Experiment	Data	Before filtering	After filtering
IMP vs ELV	*Ct* values ≤ 35	60.5%	88.2%
	Ct values > 35	39.5%	11.8%
ELV vs ERV	*Ct* values ≤ 35	61.1%	90.2%
	Ct values > 35	35.9%	9.8%
IMP vs ELV vs ERV	*Ct* values ≤ 35	69.5%	90.0%
	Ct values > 35	30.5%	10.0%

Abbreviations: Ct, cycle to threshold; HF, end-stage heart failure; IMP, the sample taken from the left ventricle at time of implant; ERV, the sample taken from the right ventricle at time of explant; ELV, the sample taken from the left ventricle at the time of the explant.

The four different normalization techniques were applied to the miRNA values after filtering. The cumulative distribution associated with the various normalization techniques is depicted in Figure 1. Based on the reduction of the coefficient of variation, the quantile normalization performed the best (this is shown by the quantile normalization value reaching 1 most quickly). The other methods introduce more variation than is seen in the raw data. The endogenous controls were carefully selected for the delta-*Ct* method based on a review of literature including analysis conducted by Applied Biosystems.[17] The stability of the controls was also assessed, and the average of the triplicates was used as the Ct_e.

The effects of the normalization techniques are depicted in Figures 2 and 3. Figure 2 displays the density estimates of the raw *Ct* values for each plate. We can see that the distributions appear to follow a normal distribution with a small deviation near $Ct = 40$. This deviation is caused by the number of miRNAs with *Ct* values of 40, which implies that a cycle to threshold value was not determined. It also supports the choice of the quantile normalization technique since all of the curves are very similar in shape, location, and scale. Figure 3 displays the density estimates based on the quantile normalized *Ct* values. There are still 27 curves (or one for each plate) but the normalization technique forces each distribution to be nearly identical. The results appear to be extremely normal but a small deviation from the normal curve still appears near the *Ct* value 40.

The final issue to consider during the analysis was the appropriate hypothesis testing procedure. A *t*-test, Mann–Whitney *U* test, and a testing procedure proposed by Pounds and Rai[18] were considered, as well as a model based on the GEE approach.

The distribution of the adjusted *P* values is depicted in Figures 4 and 5. The false discovery rate described by Benjamini and Hochberg[9] was used. An adjusted *P* value smaller than 5% was considered to be statistically significant. Figure 4 displays the distribution of the adjusted *P* values resulting from the contrast comparing the average *Ct* values from the left ventricle at the time of LVAD implantation (IMP) with the average *Ct* values from the left ventricle at the time of heart transplant (ELV). There are many miRNAs with significantly different expression values. The GEE model was also used to compare the expression values between the ELV and ERV but there are not as many significantly expressed miRNAs. The results are displayed in Figure 5.

Figure 6 contains the histogram of FDR adjusted *P* values resulting from a comparison between the IMP and ELV based on the paired *t*-test. We can see there are fewer significantly different miRNAs than in the same comparison based on the GEE model. Although the *t*-test is paired, it does not have the ability to incorporate the additional correlation. The distribution of *P* values associated with the comparison between the ELV and ERV time periods based on the paired *t*-test is similar to Figure 6.

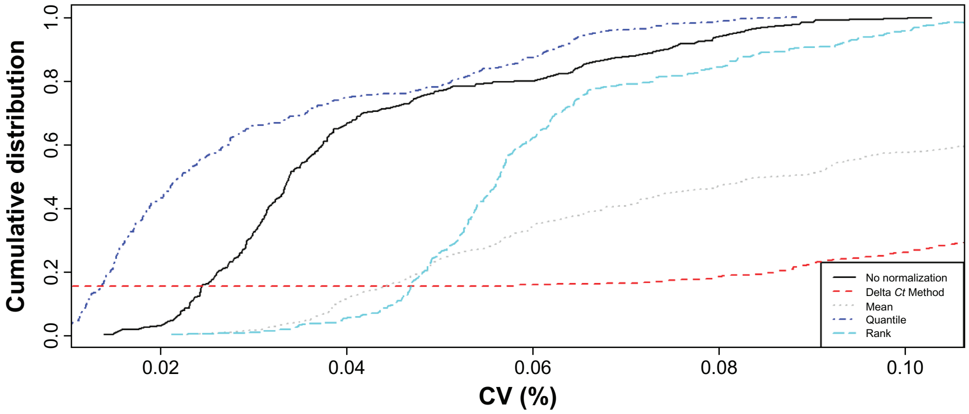

Figure 1 Cumulative distribution of the coefficient of variation.
Abbreviation: CV, coefficient of variation.

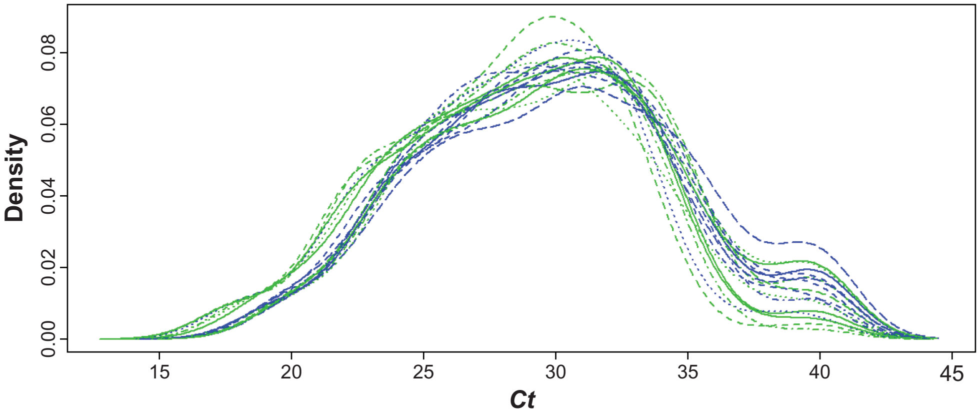

Figure 2 Density estimate of *Ct* values for each plate over all miRNAs (no normalization).
Abbreviation: *Ct*, cycle to threshold.

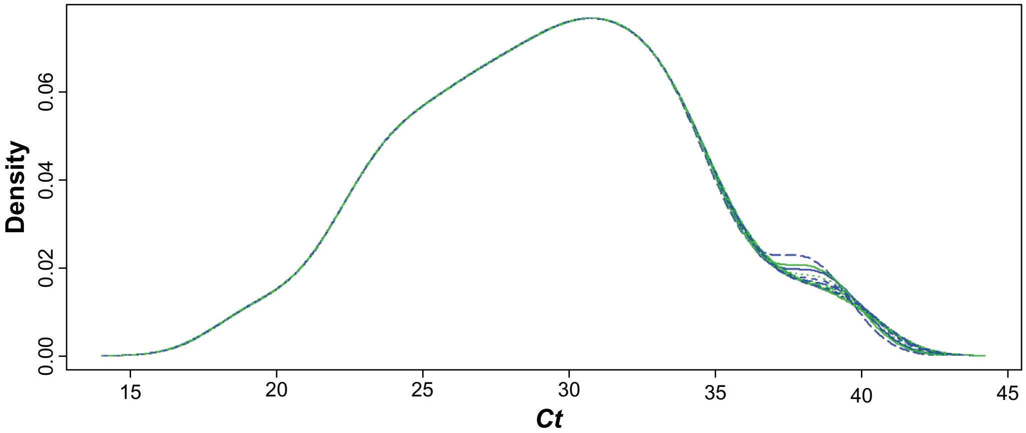

Figure 3 Density estimate of *Ct* values for each plate over all miRNAs (quantile normalization).
Abbreviation: *Ct*, cycle to threshold.

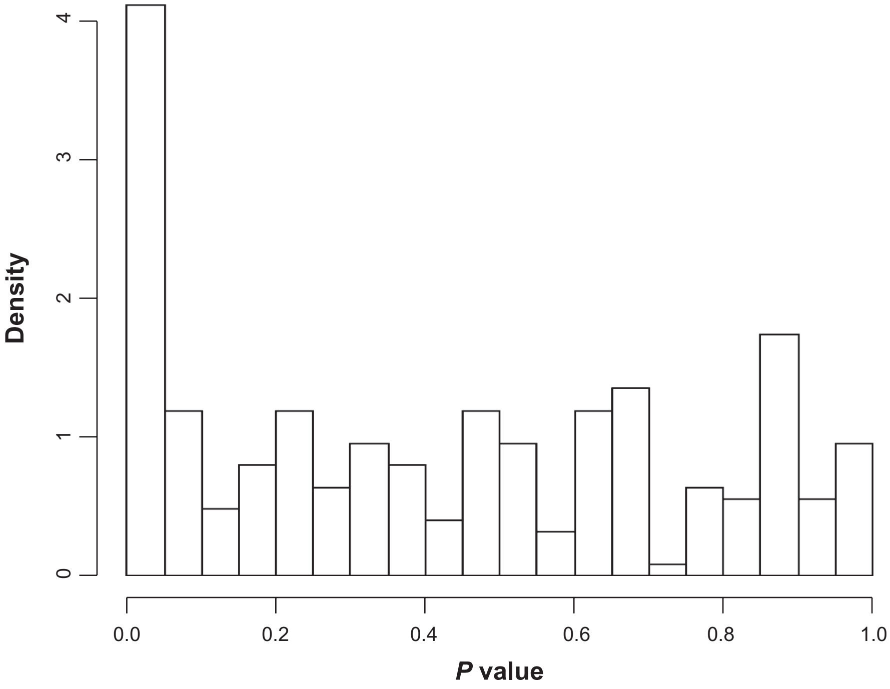

Figure 4 Histogram of *P* values of comparison between IMP and ELV based on GEE.
Abbreviations: GEE, generalized estimating equations; IMP, the sample taken from the left ventricle at time of implant; ELV, the sample taken from the left ventricle at the time of the explant.

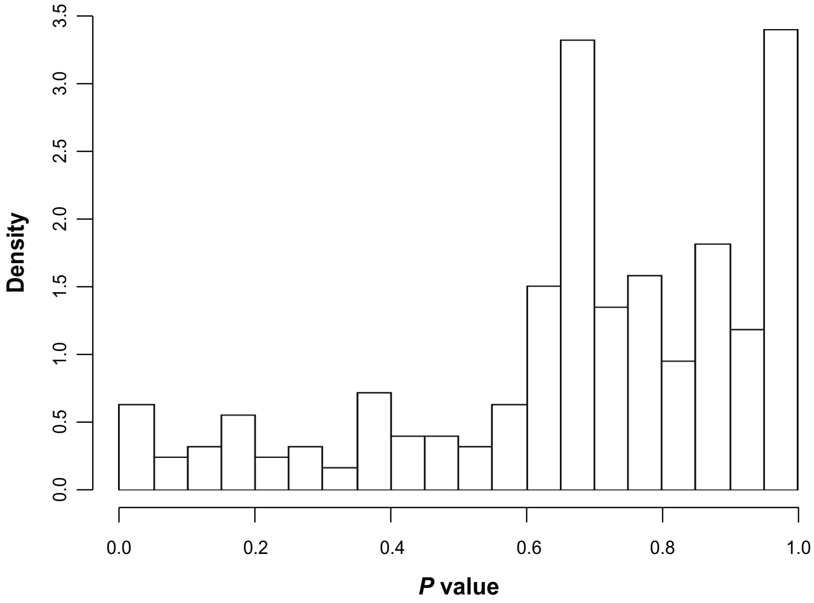

Figure 5 Histogram of *P* values of comparison between ELV and ERV based on GEE.
Abbreviations: GEE, generalized estimating equations; ERV, the sample taken from the right ventricle at time of explant; ELV, the sample taken from the left ventricle at the time of the explant.

The results from the analysis utilizing the AAA methodology, as well as the Mann–Whitney *U* test results, are not included here since the distribution of adjusted *P* values is similar to the previously discussed *t*-test.

Discussion

In reviewing the current methodologies, there are statistical models that incorporate the intra-subject correlation created by repeated measurements on the same individuals. The GEE model is a popular method that incorporates the additional correlation. In the exploratory analysis, it is apparent the GEE model also results in a greater number of significantly expressed miRNAs than the paired *t*-test. Although the statistical model incorporates the additional correlation, what effect does the normalization technique have on the naturally occurring correlation structure?

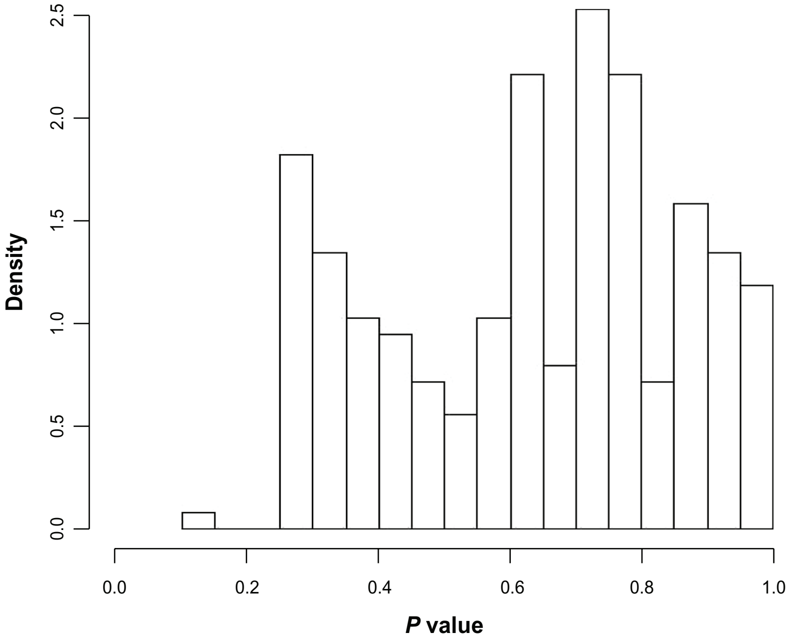

Figure 6 Histogram of *P* values of comparison between IMP and ELV based on *t*-test.

Should the effect be of concern to the analyst? The delta-*Ct* and the mean normalization techniques shift the mean of the expression value of each plate thus preserving the original correlation structures. The quantile normalization is not a simple shift of the center but actually changes the distribution of the *Ct* values, resulting in a different correlation structure than the one which naturally occurs. The methodology also reduces the variance based on the coefficient of variation. Based on the availability of software and the cumulative distribution of the coefficient of variation, it appears that the quantile normalization technique is the best choice. What effect does the normalization technique have on the GEE model and the resulting significantly expressed genes? How does the effect compare to the shift of center normalization procedures that do not reduce the variation in the *Ct* values? Should the analysis be performed on the raw, unnormalized data?

The topics discussed require a method to simulate the correlated *Ct* values. Once the initial problem is solved, then one must evaluate the various combinations of normalization procedures with hypothesis testing procedures to determine the impact on the results.

Conclusions

The motivating example brings to the fore many important questions facing an analyst of miRNA *Ct* values. The increased accuracy and reproducibility of the qPCR methods imply that more researchers are turning to the technology. Experimental designs are becoming more advanced and now include repeated biologic sampling from the same individuals over time. In general, it is important that researchers consider the complexities of the experimental design in order to select analytical techniques that will allow a full understanding of the results.

The analysis presented is not unique, but rather the presentation of an analysis from data preparation, through normalization, and hypothesis testing is unique. The analysis emphasizes the importance of considering the experimental design. The theoretical formulation of many popular methods is also now contained in one place.

The analysis also raises several important research questions regarding determining the most appropriate analysis methods for miRNA *Ct* values obtained from experiments collecting multiple samples on the same individuals over time. The sharing of such information will aid future researchers with the analysis of qPCR *Ct* values using freely available software and methods.

Acknowledgments

We thank Dr D Miller for advancing free use of these methods with future publication of the R routine and manual for public use at http://browncancercenter.louisville.edu/biostats. SN Rai was partly supported with Wendell Cherry Chair in Clinical Trial Research Fund and HE Ray was partly supported by Kennesaw State Univeristy. SD Prabhu was partially supported by V.A. Merit Award, NIH grant HL-99014, UofL CEGIB Director's Award, and the UofL Department of Medicine Bales Fund (S.D.P.). T Hamid was partially supported by an AHA SDG award 0835456N.

Disclosure

The authors report no conflict of interest in this work.

References

1. Ryan BM, Robles AI, Harris CC. Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer*. 2010;10(6):389–402.
2. van Rooij E, Marshall WS, Olson EN. Toward microRNA-Based therapeutics for heart disease: the sense in antisense. *Circ Res*. 2008;103(9):919–928.
3. Ach RA, Wang H, Curry B. Measuring microRNAs: comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods. *BMC Biotechnol*. 2008;8:69.
4. Chen Y, Gelfond JA, McManus LM, Shireman PK. Reproducibility of quantitative RT-PCR array in miRNA expression profiling and comparison with microarray analysis. *BMC Genomics*. 2009;10:407.
5. Pradervand S, Weber J, Thomas J, et al. Impact of normalization on miRNA microarray expression profiling. *RNA*. 2009;15(3):493–501.
6. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc*. 2008;3(6):1101–1108.
7. Yuan JS, Reed A, Chen F, Stewart CN. Statistical analysis of real-time PCR data. *BMC Bioinformatics*. 2006;7:85.
8. Schonrock N, Ke YD, Humphreys D, et al. Neuronal microRNA deregulation in response to Alzheimer's Disease Amyloid-ß. *PLoS ONE*. 2010;5(6):e11070.
9. Melkamu T, Zhang X, Tan J, Zeng Y, Kassie F. Alteration of microRNA expression in vinyl carbamate-induced mouse lung tumors and modulation by the chemopreventive agent indole-3-carbinol. *Carcinogenesis*. 2010;31(2):252–258.
10. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B*. 1995;57(1):289–300.
11. Montenegro D, Romero R, Pineles BL, et al. Differential expression of microRNAs with progression of gestation and inflammation in the human chorioamniotic membranes. *Am J Obstet Gynecol*. 2007;197(3):289. e1–e6.
12. Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics*. 1988;44(4): 1049–1060.
13. Mestdagh P, Van Vlierberghe P, De Weer A, et al. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol*. 2009;10(6):R64.
14. 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;25(4):402–408.
15. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185–193.

16. Tseng GC, Oh MK, Rohlin L, Liao JC, Wong WH. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res*. 2001;29(12):2549–2557.
17. Wong L, Lee K, Russell I, Chen C. *Endogenous Controls for Real-time Quantitation of MiRNA Using TaqMan MicroRNA Assays.* Carlsbad, CA: Applied Biosystems; 2010.
18. Pounds S, Rai SN. Assumption adequacy averaging as a concept for developing more robust methods for differential gene expression analysis. *Comput Stat Data Anal*. 2009;53(5):1604–1612.
19. Efron B, Tibshirani R, Storey JD, Tusher V. Empirical Bayes analysis of a microarray experiment. *J Am Stat Assoc*. 2001;96(456): 1151–1160.
20. Pounds S, Morris SW. Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of P-values. *Bioinformatics*. 2003;19(10):1236–1242.
21. Chow SC, Shao J, Wang H. *Sample Size Calculations in Clinical Research*, *2nd Edition*. Boca Raton, FL: Chapman and Hall/CRC; 2008.

Appendix

Typically, the delta-*Ct* normalization technique is derived through the ration of the target gene efficiency (E_T) raised to the power:

$$\Delta Ct_T = Ct_T - Ct_e \tag{A1}$$

and the reference gene efficiency (ER) raised to the power:

$$\Delta Ct_R = Ct_R - Ct_e \tag{A2}$$

where Ct_T is the *Ct* value for the target group and Ct_R is the *Ct* values for the reference group (or control group). The ratio is:

$$\frac{E_T^{\Delta Ct_T}}{E_T^{\Delta Ct_R}} \tag{A3}$$

If the PCR amplification efficiency achieves the maximum value then both $E_T = E_R = 2$ and the ratio is written as:

$$2^{-\Delta\Delta Ct} \tag{A4}$$

where $-\Delta\Delta Ct = \Delta Ct_T - \Delta Ct_{R.}$ The first step, or:

$$\Delta Ct = Ct - \Delta Ct_e \tag{A5}$$

is a normalization technique using the endogenous controls as the reference.