


Statistical Issues and Group Classification in Plasma MicroRNA Studies With Data Application

Evolutionary Bioinformatics
Volume 16: 1–9
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1176934320913338



Shesh N Rai^{1,2,3,4*} , Chen Qian^{1,2*}, Jianmin Pan¹,
Marion McClain^{3,4,5}, Maurice R Eichenberger⁶,
Craig J McClain^{3,4,5,7} and Susan Galandiuk⁶

¹Biostatistics and Bioinformatics Facility, James Graham Brown Cancer Center, University of Louisville, Louisville, KY, USA. ²Department of Bioinformatics & Biostatistics, University of Louisville, Louisville, KY, USA. ³Alcohol Research Center, University of Louisville, Louisville, KY, USA. ⁴Hepatobiology & Toxicology COBRE Center, University of Louisville, Louisville, KY, USA. ⁵Department of Medicine, University of Louisville, Louisville, KY, USA. ⁶Price Institute of Surgical Research, Hiram C. Polk Jr. M.D. Department of Surgery, School of Medicine, University of Louisville, Louisville, KY, USA. ⁷Gastroenterology, Robley Rex Louisville VA Medical Center, Louisville, KY, USA.

ABSTRACT: The analysis of plasma microRNAs (miRNAs) has been widely used as a method for finding potential biomarkers for human diseases, especially those with a link to cancer. Methods of analyzing plasma miRNA have been thoroughly discussed from sample extraction to data modeling. However, some issues exist within the process that have rarely been talked about. Rice et al. discussed some issues in plasma miRNA studies, such as the lack of standard methodology including the use of different cycle threshold, time to plasma extraction, among others. These issues can lead to inconsistent data, and thus impact the result and assay reproducibility. Other external issues, such as batch effect and operator effect, may also indirectly impact the statistical analysis. Here, we discuss issues in plasma miRNA studies from a statistical point of view. The interaction effect of different ways of calculating fold-change, the choice of housekeeping genes, and methods of normalization are among the issues we discuss, with data demonstrations. *P* values are calculated and compared to determine the effect of those issues on statistical conclusions. Statistical methods such as analysis of variance and analysis of covariance are crucial in the analysis of miRNA but investigators are often confused about them; therefore, a brief explanation of these statistical methods is also included. In addition, 3-group classification is discussed, as it is often challenging, compared with 2-group classification.

KEYWORDS: ANCOVA, ANOVA, batch effect, classification, fold-change, housekeeping genes, normalization, operator effect, plasma miRNA, quantile normalization, varying threshold

RECEIVED: December 23, 2019. **ACCEPTED:** February 18, 2020.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: C Qian was supported by the National Institute of Health grant 5P50 AA024337 (CJM) and the University of Louisville Fellowship. SN Rai was partly supported with Wendell Cherry Chair in Clinical Trial Research Fund and NIH grants 5P20 GM113226 (CJM), 1P20 GM125504, 1P42 ES023716, 2P20 GM103492.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Shesh N Rai, Biostatistics and Bioinformatics Facility, James Graham Brown Cancer Center, University of Louisville, Louisville, KY 40202, USA. Email: shesh.raai@louisville.edu

Introduction

MicroRNAs (abbreviated miRNAs) are a group of noncoding RNA molecules of approximately 22 nucleotides. Research on miRNAs, especially from plasma, has become a hot research topic, as miRNAs are known to have links with diseases, such as cancer. However, there is no standard procedure as to how to extract and analyze them. Methods vary depending on laboratories and scientists, which diminishes assay reproducibility.¹ Other external issues, such as varying thresholds, may also lead to different statistical conclusions. Batch effect and operator effect are additional issues that investigators should pay attention to. Those issues will be further discussed in the next section using the first data set as an example.

Similar issues are observed in statistical analysis. For example, the use of different methods for calculating fold-change might result in different results upon which investigators would base their conclusions. The choice of

housekeeping genes is also another problem. It is very important to choose the right housekeeping gene to generate a reliable and reproducible marker. Moreover, the normalization of data is a crucial part of the analysis. Types of normalization include mean normalization, quantile normalization, and delta-Ct normalization. The details and data demonstration will be discussed in the section “Methods and Issues for Data Set 2”, using the second data set. Finally, with the development of artificial intelligence, machine learning is becoming more prevalent in almost every subject. In medical research, predicting patients’ potential risk for disease has been very important but challenging, as early detection could potentially save lives, or risk analysis could help create new therapies. To achieve these goals, the classification of existing samples by different diseases based on given predictor variables can be extremely helpful. Algorithms will work well on future predictions if the accuracy of the model in the sample is high. However, current classification is limited to 2 groups so as to differentiate controls from cases. Some studies, such

*These authors contributed equally to this work.



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

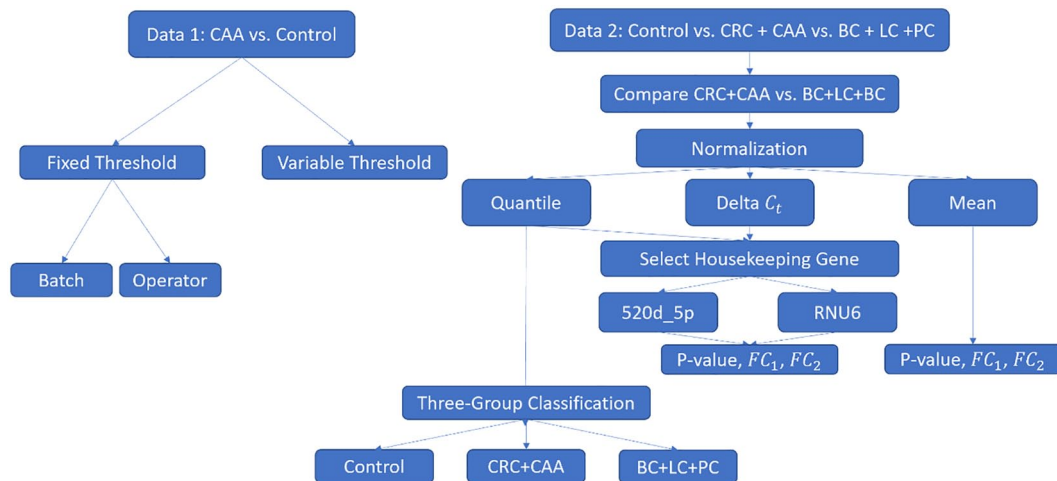


Figure 1. Flowchart of the data processing. BC indicates breast cancer; CAA, colorectal advanced adenoma; CRC, colorectal cancer; FC, fold-change; LC, lung cancer; PC, pancreatic cancer.

Table 1. List of statistical issues in plasma miRNA analysis.

Varying threshold
Batch effect
Operator effect
Normalization methods
Choice of housekeeping genes
Methods of fold-change
Sample size issue
Group classification

Abbreviation: miRNA, microRNA.

as the data we are using in this article, require the classification of 3 or more groups. Especially in miRNA studies where each miRNA carries certain information that might relate to the disease, multigroup classification becomes necessary. A flowchart of the methods detailed in this article is given in Figure 1. A list of issues that are discussed in this article is given in Table 1.

A summary of the experimentation is reproduced from Carter et al.² Total RNA quantity and the purity of each sample were determined using a Nanodrop 2000 spectrophotometer (ThermoFisher Scientific, Middlesex, MA). For each sample, 384 miRNAs were screened to identify dysregulated miRNA expression within each group as compared with controls (TaqMan Low Density Array human miRNA card A; Life Technologies, Carlsbad, CA). Quantitative real-time polymerase chain reaction (qRT-PCR) was performed using a ViiA 7 Real-Time PCR System (ThermoFisher Scientific) with fixed and varying thresholds on selected miRNA from screening phase. Data sets 1 and 2 are based on results from qRT-PCR studies.

Methods and Issues for Data Set 1: Comparing Colorectal Advanced Adenoma Versus Control

Data information and hypothesis

The study was approved by the Institutional Review Board at the University of Louisville. Informed consent was obtained from all participating patients. The first data set consisted of patients from a control group ($n = 16$) and a colorectal advanced adenoma (CAA) group ($n = 16$). CAA was determined with adenomatous polyps greater than 0.6 cm in maximal diameter. Samples from each subject were run by 2 operators, and each operator had 2 batches. The goal of this demonstration is to show the impact of a varying threshold, batch effect, and operator effect. P values were calculated for group comparisons for 11 different miRNAs within samples.

Method

Varying threshold. The use of a fixed threshold and a variable threshold were compared in this data demonstration. A fixed threshold might not intersect the linear phase, as different miRNAs on the same plate may carry different linear phases. Therefore, the variable threshold was the default setting. However, Rice et al¹ suggested that the use of a fixed threshold at 0.03 was preferable to the variable threshold when the missing values were less than 10%. Details of statistical methods in terms of calculating fixed and variable thresholds can be found in that article. In our demonstration, we chose a fixed threshold at 0.03 to compare with a variable threshold. The results can be found in Table 2.

Operator effect and batch effect. Rice et al¹ also mentioned issues such as intra- and inter-operator variabilities. The problem could occur when a different operator handles the experiment. They may or may not closely follow the exact

procedure, and therefore, could create different data that could lead to distinct conclusions.

Batch effect is a technical issue that also could impact statistical analysis. Sources of this issue include, but not limited to, the person who handles the experiment, the machine used to extract the sample, and location or the environment of the experiment. There were different ways to adjust batch effects in the analysis as proposed by Johnson et al³ and Guo et al.⁴ It is one of the issues that investigators should pay attention to. In

this article, analysis of variance (ANOVA) was used, and the goal was to test whether a batch effect exists in our data set.

Results and issues

There is some variation in terms of the threshold. For example, miR523 was not significantly different between CAA and control group when using a variable threshold ($P=.0718$), but became significantly different when using a fixed threshold at 0.03 ($P=.0443$) (Table 2).

When the threshold was fixed at 0.03 along with group adjustment, there was no batch effect in the data set (Table 3). However, it did not mean that the batch effect was negligible. Every data set is different, and it is always important to pay additional attention. On the other hand, there was an operator effect in miR21 ($P<.0001$) after adjusting for the group. Other miRNAs such as miR523 ($P=.0549$) and miR520-5p ($P=.0707$) were close to being significantly different. It is certainly an issue to focus on.

When the batch effect and operator effect were considered, group significance (for comparing CAA versus control) also changed. By comparing Table 2 with Table 3, percent change indicated that for most miRNAs, the P value for group significance had increased (Table 3).

Methods and Issues for Data Set 2: Comparing CRC + CAA Versus BC + LC + PC

Data information and hypothesis

This study was approved by the Institutional Review Board at the University of Louisville. Informed consent was obtained from every participating patient. The study population consisted of patients from the University of Louisville colon and

Table 2. Comparison between fixed threshold and variable threshold.

MicroRNAs	FIXED THRESHOLD (0.03)	VARIABLE THRESHOLD
	P VALUE	P VALUE
miR374	.0001	<.0001
miR142-3p	<.0001	<.0001
miR523	.0443	.0718
miR374-5p	<.0001	<.0001
miR376c	.3540	.2097
miR27a	.0016	.0005
miR520d-5p	.7852	.4820
miR122	.5186	.6140
miR485-3p	.2369	.5905
miR21	.0052	.0067
miR218	.0044	.0012

Significance level = 0.05.

Table 3. Comparison between batch effect and operator effect.

MicroRNAs	FIXED THRESHOLD = 0.03			
	P VALUE (BATCH)	P VALUE (OPERATOR)	P VALUE (GROUP)	% CHANGE IN P VALUE (GROUP)
miR374	.9930	.5532	.0002	10.8
miR142-3p	.6838	.9329	<.0001	22.4
miR523	.5541	.0549	.0426	-3.79
miR374-5p	.7785	.1278	<.0001	-23.4
miR376c	.4627	.2647	.4848	36.9
miR27a	.1894	.5129	.0016	2.3
miR520d-5p	.2651	.0707	.9788	24.7
miR122	.6168	.7082	.4998	-3.6
miR485-3p	.7991	.7192	.2403	1.4
miR21	.9481	<.0001	.0026	-50.8
miR218	.7525	.4029	.0054	24.6

Significance level = 0.05.

rectal surgery practice and samples from patients housed in the University of Louisville Surgical Biorepository. In this data set, there are 6 groups in total: control group, breast cancer (BC), lung cancer (LC), pancreatic cancer (PC), colorectal cancer (CRC), and CAA. Each group had 20 samples for a total of 120 samples. In the analysis, we combined BC, LC, and PC as one group, and combined CRC and CAA as another group.

The goal is to find whether there is any difference between colorectal neoplasia (CRC and CAA) and other cancers (BC, LC, and PC) in terms of selected miRNAs with different approaches.

Ten miRNAs were selected in the screening phase and evaluated as they were among the most significantly dysregulated miRNAs as noted in Carter et al.²

Method

Normalization. Three normalization methods were used and compared, including delta-Ct normalization, mean normalization, and quantile normalization. Normalization is required in miRNA studies, as unwanted technical variation exists and must be removed. Detailed formulas are not shown in this article, as we are only focusing on issues. However, readers can refer to Rai et al⁵ for a detailed explanation of those 3 methods. As we want to incorporate housekeeping gene information in the comparison, we have discussed the effect of the quantile and delta-Ct normalizations in the next section as a housekeeping gene does not play a role in the mean normalization. The mean normalization will be discussed in the discussion section.

Specifically, for quantile normalization, there are 2 approaches. In miRNA studies, there is always more than one group. One approach is to do the scaling separately for each group, and then combine them. The other way is to combine all groups and do the scaling. The earlier method is usually preferred as it maintains more information from the original data, but it depends on the specific data set. The comparison will also be made between 2 quantile normalization approaches in the discussion section.

Housekeeping genes. Housekeeping genes play an extremely important role when generating reliable and reproducible markers. Two housekeeping genes (520d-5p and RNU6) were compared, as both were recommended by Rice et al.⁶ We will conduct a comparison under the delta-Ct normalization and quantile normalization (Table 4).

Fold-change. The reporting value of fold-change has frequently been used by investigators. However, different methods of calculating fold-change values might lead to different results. To calculate fold-change, we first introduced the cycle threshold (Ct) value. This has been discussed in Rice et al.¹ The cycle threshold we are using in the formula is the value after normalization. Two ways of calculating fold-change will be

demonstrated in this article. The modification was made based on Pfaffl.⁷ The formulas are listed below

$$FC_1 = 2^{\left(\frac{1}{n_1} \sum_{i=1}^{n_1} \Delta Ct_{miR_i}^{(1)} - \frac{1}{n_2} \sum_{i=1}^{n_2} \Delta Ct_{miR_i}^{(2)}\right)},$$

$$FC_2 = \frac{\frac{1}{n_1} \sum_{i=1}^{n_1} 2^{\Delta Ct_{miR_i}^{(1)}}}{\frac{1}{n_2} \sum_{i=1}^{n_2} 2^{\Delta Ct_{miR_i}^{(2)}}},$$

where $\Delta Ct_{miR_i}^{(j)}$ and $TH_{miR_i}^{(j)}$ are the ΔCt and the threshold values for miR_i in group j ($1 = \text{Control}$, $2 = \text{Case}$), HKG is the housekeeping gene. $\Delta Ct_{miR_i}^{(j)}$ is defined as follows:

$$\Delta Ct_{miR_i}^{(j)} = Ct_{miR_i}^{(j)} - Ct_{HKG}^{(j)} \quad \text{for fixed threshold.}$$

$$\Delta Ct_{miR_i}^{(j)} = Ct_{miR_i}^{(j)} - Ct_{HKG}^{(j)} - \frac{\left(\log\left(TH_{miR_i}^{(j)}\right) - \log\left(TH_{HKG}^{(j)}\right)\right)}{\log(2)}$$

for variable threshold

Fold-changes are calculated under different housekeeping genes and every normalization method (Table 4).

Results and issues

The selection of housekeeping genes does make an impact on the conclusion. In this data set, under delta-Ct normalization, miR122 ($P = .0525$) was not quite significant under housekeeping gene miR520-5p, but was significant under housekeeping gene RNU6 ($P = .0115$). On the other hand, miR372 showed the opposite results. Two fold-change methods also produced slightly different results. Similarly, under the quantile normalization, there was one that produced the opposite result when using different housekeeping genes (Table 4). Under both normalization methods, significant miRNAs were quite matched, given certain housekeeping genes. Differences were only found in miR372 and miR122.

Statistical Methods in miRNA Studies: ANOVA and Analysis of Covariance

In plasma miRNA analysis, the analysis of variance (ANOVA) is one of the most common statistical methods for finding significant differences, especially when there are more than 2 groups. Detailed derivation methodology will be omitted here as it is a common procedure. To begin, the sample variance can be calculated using the following formula

$$MS_{\text{total}} = \frac{1}{n-1} \sum_{i=1}^n (\Delta Ct_{miR_i} - \overline{\Delta Ct})^2 \triangleq \frac{SS_{\text{total}}}{N-1},$$

where the latter term is called the total sum of squares. The total sum of squares (SS_{total}) divided by the degree of freedom results in the total mean square (MS_{total}). In addition, the total sum of squares can be separated into 2 parts, the sum of squares for treatment and the sum of squares for errors, and similarly

Table 4. Interaction effect between normalizations, fold-changes, and housekeeping genes.

MicroRNAs	DELTA-CT NORMALIZATION						QUANTILE NORMALIZATION					
	HKG=miR520-5P			HKG=RNU6			HKG=miR520-5P			HKG=RNU6		
	P VALUE	FC1	FC2	P VALUE	FC1	FC2	P VALUE	FC1	FC2	P VALUE	FC1	FC2
miR192	.9377	0.9763	0.1758	.2516	0.7494	0.3513	.3342	0.8350	0.3916	.7616	0.9232	0.5166
miR29c	.0301	0.4801	0.1246	< .0001	0.3612	0.2693	< .0001	0.3848	0.1947	.0003	0.4014	0.2212
miR21	.0125	0.4185	0.1895	< .0001	0.3292	0.2242	< .0001	0.3535	0.1774	.0004	0.3943	0.2043
miR19a	.0488	0.5574	0.196	.0002	0.4177	0.2241	< .0001	0.3841	0.1693	.0006	0.4061	0.1889
miR150	.3785	1.2773	0.3885	.9134	0.9793	0.9782	.4749	0.8639	0.6431	.7932	0.9342	0.9943
miR374	.0012	0.2648	0.0989	< .0001	0.2045	0.1254	< .0001	0.2528	0.1968	.0001	0.2794	0.1723
miR193a-5p	.4168	1.2684	0.3363	.9311	0.9796	0.6821	.6508	0.8921	0.0397	.6539	1.1342	1.0682
miR346	.8836	1.0265	0.7007	.2965	0.7321	0.5721	.9617	1.0091	0.4971	.7547	0.8985	0.5378
miR372	.0003	2.1137	2.4506	.1037	1.7181	1.7517	.2731	1.2768	0.8368	.0301	1.8760	1.0264
miR122	.0525	0.4163	0.0563	.0115	0.3289	0.043	< .0001	0.2106	0.0051	.0009	0.2451	0.0503

Abbreviations: FC, fold-change; HKG, housekeeping genes.
Significance level = 0.05.

Table 5. Sample size required at power of 90% at different FDR level.

FDR	ADJUSTED α	SAMPLE SIZE REQUIRED	
		ONE-SIDED	TWO-SIDED
1%	0.00033	23	25
5%	0.0018	18	20
10%	0.0036	16	18

Abbreviation: FDR, false discovery rate.

for the degrees of freedom. The F -test for 1-way ANOVA can be obtained by the formula

$$F = \frac{MS_{\text{treatment}}}{MS_{\text{error}}} = \frac{SS_{\text{treatment}} / (M - 1)}{SS_{\text{error}} / (N - M)},$$

where M is the number of treatments and N is the overall sample size.

On the other hand, the analysis of covariance (ANCOVA) is not used as often as ANOVA. It is used when there are additional variables available that could be incorporated into the model. In our data set, there is no other covariate that we need to include in the model, but in other data sets, covariates such as age, can be added to the model by using ANCOVA. The model would then produce a better estimate because the error variance is reduced.

Sample Size Justification

Jung⁸ introduced a sample size calculation method for a specified number of true rejections while controlling the false discovery rate (FDR) at the desired level with a closed-form formula if the projected effect sizes are equal among differentially expressed genes, otherwise, requiring a numerical method to solve an equation. In addition, rank-based procedures can also be used, in which case α can be adjusted using α divided by the number of groups. We consider a hypothetical example. Suppose there are 2 types of samples (cases and controls) and the potential number of types of miRNAs is 384, of which about 30% are potential candidates to differ between cases and controls. Of that 30%, our focus is to identify at least 10 (2.5% of all 384). The adjusted α values for each level of FDR are listed in Table 5. The outcome measure is based on $\Delta Ct_{\text{miR}_i}^{(j)}$ for i th miRNA from the j th group and compare the mean difference of ΔCt_{miR_i} in 2 groups for each miRNA. To detect at least 1.5 SD (standard deviation units) at a power of 90% for a 1-sided hypothesis, we need 23, 18, and 16 samples of each type. If a 2-sided hypothesis is used, the corresponding sample sizes will be 25, 20, and 18 (Table 5). If we consider hypothesis-generating pre-clinical studies that do not adjust alpha, we will be able to detect reasonably small effect sizes (1.0 SD) with 12 samples in each group.

The classification requires a different approach for sample size justification. If 10 miRNAs are used to build a classifier

Table 6. Sample size required for classification at $\alpha=0.05$.

NUMBER OF miRNAs	POWER	SAMPLE SIZE REQUIRED PER GROUP
10	90	120
5-8	80	96
5-8	68	72

Abbreviation: miRNA, microRNAs.

and classify samples into 2 groups, we need approximately 120 samples in each group (considering 10 covariates in a predictive model). With this sample size, we can detect sensitivity/specificity from 75% to 90% at $\alpha=0.05$ and power $\geq 90\%$. With the reduced number of significant miRNAs (5-8) used for classification, we need approximately 96 samples in each group to maintain a power of 80% at $\alpha=0.05$. Using the same design parameters with only 72 samples in the group will have only 68% power (Table 6). If attrition is expected, these sample sizes will need to be increased to address the attrition. If the studies are designed to establish a hypothesis or confirmatory point of view (to maintain high rigor and reproducibility), then a lower FDR/alpha ($\leq 5\%$) and a higher power ($\geq 90\%$) are recommended.

Three-Group Classification on Cancer Patients: Control vs CRC + CAA vs BC + LC + PC

We will demonstrate a 3-group classification using the second data set. Quantile normalization with the preferred approach will be used, as it was determined to be the best one in Rai et al.⁵ There are 3 groups for classification: control group (CT), CRC and CAA (CN), and BC, LC, PC (BLP). Two methods will be used and compared: multinomial logistic regression for a parametric model and random forest for a nonparametric model. R is used for computation.

Issues exist in classification as well. In the multinomial logistic regression model, we picked miRNAs that were significant in a previous analysis, which yields an accuracy rate of 0.7837. However, if we include all miRNAs, the accuracy rate would barely pass 0.5. In our random forest model, accuracy reached nearly 1. In both models, certain miRNAs had relatively high importance and dominated the outcome (miR372 and miR19a). While no relation between input and output variables is expected, using a nonparametric model is always preferred. Picking the right model is certainly important. In addition, the same data set is used for model building and classification. The sample size is relatively small. It is better to use different relatively large data sets for training and testing.

Discussion

Issues involved in plasma miRNA analysis are discussed in the article. However, many other problems exist and that should

Table 7. Comparison between 2 quantile normalization methods.

MicroRNAs	QUANTILE NORMALIZATION			QUANTILE NORMALIZATION		
	APPROACH 1			APPROACH 2		
	P VALUE	FC1	FC2	P VALUE	FC1	FC2
miR192	.4079	0.8880	0.9904	.0010	1.6047	1.6992
miR29c	<.0001	0.3805	0.3813	.0086	0.7718	0.7869
miR21	<.0001	0.3687	0.3425	.0044	0.8310	0.7633
miR19a	<.0001	0.3864	0.2960	.0098	0.7875	0.5904
miR150	.4707	0.8844	0.7971	.0006	1.8415	1.6381
miR374	<.0001	0.2592	0.1898	.0052	0.5377	0.3501
miR193a-5p	.8408	1.0438	1.4808	.0411	1.5312	1.5241
miR346	.8474	0.9666	0.8776	.0208	1.4904	1.1032
miR372	<.0001	1.6193	1.6126	.5117	1.0806	1.0344
miR122	<.0001	0.2278	0.0646	.0027	0.4137	0.0805

Abbreviation: FC, fold-change.
Significance level = 0.05.

also be adjusted. Future research could focus on comparing different machine learning models when doing classification. Also, more than 3 groups of classification should be tested when sufficient data are available. In our data set, there are many diseases with limited sample size and miRNAs are also limited. There are possibly additional miRNAs that could connect to the diseases. External issues that lead to inconsistent data should be addressed by introducing a standardized procedure. Further research is needed to determine the best method for the experiment.

Different quantile normalization procedures produced contradictory results (Table 7). Here, we used the first approach as discussed in the “Normalization” section. Please note in this comparison we did not incorporate housekeeping gene information, as our focus is solely on the issues that might occur between 2 quantile normalization approaches. By comparing 2 different approaches, miRNAs such as miR192 ($P=.001$), miR150 ($P=.0006$), miR193a-5p ($P=.0411$), and miR346 ($P=.0208$) were significantly different between groups when using the second approach, but were not in the first approach. On the other hand, miR372 ($P=.5117$) was not significantly different in the second approach but was in the first approach. Also, the values of fold-changes vary in the 2 approaches as well. While the first approach is always recommended, researchers should try both approaches to see if there is any difference in output. It depends on the specific data, and as for our data, the second approach yields a better result. This issue is certainly important and could lead to different conclusions. It is recommended that researchers include both outputs in the supplement file.

One may notice that in Table 4, some significant miRNAs do not match under different normalization methods. This leads to the questions of why they were different, and which one a researcher should pick. To solve the mystery, we believe that a normality test is necessary. We ran normality tests on each subgroup under different housekeeping genes for both normalization methods. The results can be found in Table 8. Under delta-Ct normalization, a few of those genes appear to be nonnormally distributed but overall satisfy normality conditions. Under quantile normalization, the majority of our genes are not normally distributed. Thus, it is always important to check normality first. If data are not distributed very well normally, then researchers should pay close attention as to which normalization method they choose. When reporting results, researchers should also place other methods in the supplement file for discussion.

In addition, if data are not normally distributed, then a 2-sample t test might not be the most appropriate method to test significance. Instead, one should consider using Wilcoxon rank-sum test as it is more suitable as discussed by Pounds and Rai.⁹ Results can be found in Table 9. In our data, variation occurs in miR372 when using quantile normalization. The 2-sample t test and the Wilcoxon rank-sum test produced different conclusions using both housekeeping genes.

Previously, we mentioned that the choice of housekeeping gene does not have any impact when using mean normalization. The reason is that when calculating mean normalization with a housekeeping gene, we first take the difference between each miRNA and the housekeeping gene, then we take the

Table 8. *P* value from normality test.

MicroRNAs	DELTA-CT NORMALIZATION				QUANTILE NORMALIZATION			
	HKG = miR520-5P		HKG = RNU6		HKG = miR520-5P		HKG = RNU6	
	CRC + CAA	BC + LC + PC	CRC + CAA	BC + LC + PC	CRC + CAA	BC + LC + PC	CRC + CAA	BC + LC + PC
miR192	.2357	.9302	.9789	.2407	<.0001	<.0001	.5066	.0221
miR29c	.7084	.4263	.0052	.0003	<.0001	<.0001	.0895	.0005
miR21	.6869	.0623	.0785	.0001	<.0001	<.0001	.0767	<.0001
miR19a	.4524	.6339	.5364	.0085	<.0001	<.0001	.2869	.0014
miR150	.9878	.5045	.0839	.4572	.1634	.0045	.0231	.0748
miR374	.9261	.0506	.0546	.0001	.0005	<.0001	.0162	<.0001
miR193a-5p	.6837	.1162	.4115	.9972	<.0001	<.0001	.0010	<.0001
miR346	.4568	.8454	.3524	.2321	.0056	<.0001	.0001	.0004
miR372	.1057	.0047	.2187	.6227	<.0001	<.0001	.0001	<.0001
miR122	.0897	.1458	.1309	.6836	.0448	<.0001	.0101	.0020

Abbreviations: BC, breast cancer; CAA, colorectal advanced adenoma; CRC, colorectal cancer; HKG, housekeeping genes; LC, lung cancer; PC, pancreatic cancer. Significance level = 0.05.

Table 9. Comparison of *P* value between 2-sample *t* test and Wilcoxon rank-sum test.

MicroRNAs	DELTA-CT NORMALIZATION				QUANTILE NORMALIZATION			
	HKG = miR520-5P		HKG = RNU6		HKG = miR520-5P		HKG = RNU6	
	<i>P</i> VALUE	WILCOXON <i>P</i> VALUE	<i>P</i> VALUE	WILCOXON <i>P</i> VALUE	<i>P</i> VALUE	WILCOXON <i>P</i> VALUE	<i>P</i> VALUE	WILCOXON <i>P</i> VALUE
miR192	.9377	.7935	.2516	.2506	.3342	.7789	.7616	.3693
miR29c	.0301	.0289	<.0001	.0005	<.0001	<.0001	.0003	.0007
miR21	.0125	.0082	<.0001	.0002	<.0001	<.0001	.0004	.0003
miR19a	.0488	.0490	.0002	.0015	<.0001	<.0001	.0006	.0009
miR150	.3785	.2656	.9134	.5197	.4749	.7894	.7932	.6294
miR374	.0012	.0014	<.0001	<.0001	<.0001	<.0001	.0001	.0003
miR193a-5p	.4168	.8023	.9311	.7545	.6508	.9152	.6539	.2019
miR346	.8836	.7202	.2965	.2124	.9617	.4269	.7547	.1838
miR372	.0003	.0002	.1037	.1998	.2731	<.0001	.0301	.3617
miR122	.0525	.1916	.0115	.0338	<.0001	.0002	.0009	.0074

Abbreviation: HKG, housekeeping genes. Significance level = 0.05.

difference between each miRNA with the mean. Therefore, the result stays the same. Similar procedures were carried out, and results can be found in Tables 10 and 11. The normality test also showed that some genes were not normally distributed. However, gene significance matched for both the 2-sample *t* test and Wilcoxon rank-sum test.

Conclusion

In general, issues in plasma miRNA analysis cannot be neglected, and investigators should put effort into how to correct or minimize the error. We have demonstrated that variations such as batch effect and operator effect could potentially impact the outcome data, and thus lead to wrong statistical

Table 10. P value from normality test for mean normalization.

MicroRNAs	MEAN NORMALIZATION	
	CRC + CAA	BC + LC + PC
miR192	.0145	.0510
miR29c	.5102	.1721
miR21	.7943	.0468
miR19a	.1265	.7713
miR150	.1578	.6555
miR374	.1670	.2009
miR193a-5p	.0012	.6890
miR346	.2017	.5728
miR372	.0040	.0978
miR122	.1909	.7966

Abbreviations: BC, breast cancer; CAA, colorectal advanced adenoma; CRC, colorectal cancer; LC, lung cancer; PC, pancreatic cancer. Significance level = 0.05.

analysis and conclusions. In addition, issues that involve different choices for the experiment, such as varying threshold and housekeeping genes, should also be addressed properly. Investigators should make the right choice when making decisions. In terms of statistical analysis, statisticians should pay attention to the method of normalization and fold-change they are using. As demonstrated, different ways of calculating fold-change and normalization will impact the results. Similarly, when doing classification, researchers should have a good sample size, and pick the most appropriate model depending on the data and experiment.

Author Contributions

SNR, MRE, and SG planned the experiment design. CQ and JP performed the analysis. SNR, CJM, MM, and SG contributed to the discussion and comments. MM verified and proofread the manuscript.

ORCID iD


Shesh N Rai  <https://orcid.org/0000-0002-8377-353X>

Table 11. Mean normalization fold-change and comparison of P value between 2-sample t test and Wilcoxon rank-sum test.

MicroRNAs	MEAN NORMALIZATION			
	P VALUE	FC1	FC2	WILCOXON P VALUE
miR192	.0530	1.2982	1.2687	.1746
miR29c	.0005	0.6296	0.6266	.0015
miR21	.0004	0.5855	0.6065	.0007
miR19a	.0038	0.728	0.6767	.0055
miR150	.0006	1.7068	1.6342	.0031
miR374	<.0001	0.3564	0.2399	.0001
miR193a-5p	.0002	1.7073	1.6767	.0006
miR346	.2896	1.2759	0.5676	.2018
miR372	<.0001	3.121	3.0901	<.0001
miR122	.0807	0.5732	0.2223	.1343

Abbreviation: FC, fold-change. Significance level = 0.05.

REFERENCES

- Rice J, Roberts H, Burton J, et al. Assay reproducibility in clinical studies of plasma miRNA. *PLoS ONE*. 2015;10:e0121948. doi:10.1371/journal.pone.0121948.
- Carter JV, Roberts HL, Pan J, et al. A highly predictive model for diagnosis of colorectal neoplasms using plasma microRNA: improving specificity and sensitivity. *Ann Surg*. 2016;6:575-584.
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics*. 2006;8:118-127.
- Guo Y, Zhao S, Su P, et al. Statistical strategies for microRNAseq batch effect reduction. *Transl Cancer Res*. 2014;3:260-265.
- Rai SN, Ray H, Yuan X, Pan J, Hamid T, Prabhu SD. Statistical analysis of repeated microRNA high-throughput data with application to human heart failure: a review of methodology. *Open Access Med Stat*. 2012;2012:21-31.
- Rice J, Roberts H, Rai SN, Galandiuk S. Housekeeping genes for studies of plasma microRNA: a need for more precise standardization. *Surgery*. 2015;158:1345-1351.
- Pfaffl M. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001;29:e45.
- Jung S-H. Sample size for FDR-control in microarray data analysis. *Bioinformatics*. 2005;21:3097-3104.
- Pounds S, Rai SN. Assumption adequacy averaging as a concept to develop more robust methods for differential gene expression analysis. *Comput Stat Data Anal*. 2009;53:1604-1612.