A statistical critique of normalization methods in basic science research

By
Alese Halvorson

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biostatistics

December 14, 2019

Nashville, Tennessee

Approved:

Tatsuki Koyama, Ph.D.

Holly M. Scott Algood, Ph.D.

ACKNOWLEDGMENTS

To the members of my thesis committee – I am incredibly indebted to Dr. Tatsuki Koyama for his teaching and mentorship through this thesis project and the entirety of my time at Vanderbilt University. Without his discovery of the motivating examples for this work, this project would not have been possible. I am so appreciative of Dr. Holly Algood for her subject-matter expertise and patience in explaining foreign scientific concepts to us statisticians. A special thanks also to Dr. Robert Greevy for giving the go-ahead for this project and for his thoughtful advice along the way.

To the Vanderbilt Department of Biostatistics – I would like to thank Dr. Jeffrey Blume and Dr. Greevy for their fearless leadership and advisement as Graduate Studies Directors during my time in the program; they provided me with such rewarding opportunities and were undying in their support of what was in my best interest as a student. I owe a great deal of thanks to the advisors of my Research Assistantships which provided financial support throughout my time in the program. I am also incredibly grateful for the dedication of the department faculty members I had the pleasure of working with – they provided us a truly exemplary education. I thoroughly appreciated the comradery and collaborative spirit of my fellow classmates, we made it through this program together and they were rooting me on every step of the way. It has been an honor to be a part of the Vanderbilt Biostatistics community.

To my family – I wouldn't be where I am today without the endless support and encouragement of my parents and sister throughout my life. I especially appreciate them being a sounding board whenever I need to talk through life decisions, both big and small.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

Basic science experiments often aim to compare normalized knockout and wild-type gene expression in the hopes of concluding the effectiveness of a given treatment. Statistically, this typically takes the form of an evaluation of the p-value from a two-sample Student's t-test for difference in means of relative quantities. What has previously gone unnoticed by researchers, however, is the underlying assumptions made when using this statistical test and their potential downstream effects. Typically, when test assumptions are not met, statistical inference will be compromised. In basic science research, there is a tendency to utilize group-specific control measures for normalization and a need to keep the sample size of experiments low.

Fold change is a relative measure which describes the extent to which a treatment quantity is larger than an averaged corresponding control quantity. The intention of this calculation is to scale measures such that varying treatment groups become more comparable to one another; however, the analytic methods historically used for these data have unintended statistical consequences. We aim to explore and identify patterns of error rate inflation in these methods.

Relative fold change-like measures using Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) data are calculated in a different way. The comparative C_T method ($\Delta\Delta C_T$) is an analytic methodology where relative quantities of gene expression are calculated by a series of normalizations between target and housekeeping gene expression measures, as well as treatment groups. These relative measures allow for the comparison of knockout and wild-type groups. Again, statistical issues arise when using this common method. For reference, a "RT-PCR relative" search of PubMed.gov yields 13,937 relevant publications in the last 10 years [1].

In this paper, we discuss common normalization methods used in basic science research and shed light on their pitfalls, ultimately making recommendations of statistically sound and valid alternative methods. We first discuss the motivation behind this work and examine previously published research in Chapter 2. Then for each normalization method, we lay out scenarios in which these derivations might take place and make recommendations for alternative methods in Chapters 3 and 4, respectively.

CHAPTER 2

BACKGROUND

2.1 FOLD CHANGE

Basic science experiments aim to evaluate an effect of treatment and often have separate control measures for each of the treatment groups [2]. In an attempt to make treatment measures comparable across groups, a relative measure is calculated as the fold change increase in treatment values in relation to its corresponding control [3]. As evidenced by a review of literature, the standard fold change measure is utilized quite frequently. Note that a search of "fold change" on PubMed.gov yields 13,695 articles published in the last 10 years [1]. There are several ways in which fold change measures are compared to one another. In one case, fold change measures are calculated relative to group-specific controls and then a t-test is performed to evaluate whether knockout mice have a significantly different treatment effect within groups. We view this to be an unnecessarily intensive course of analysis, as there are other simpler methods available, but it is a statistically valid test. An example of this setting is shown in Figure 2.1 where each of the ODC and E2F2 groups have their own wild-type which is used as the basis for fold change quantification, thus the reason both groups' wild-type messenger RNA levels are 1 [4]. The statistical tests are then performed between *Bves*$^{-/-}$ mice and wild-type mice within the ODC and E2F2 groups separately.

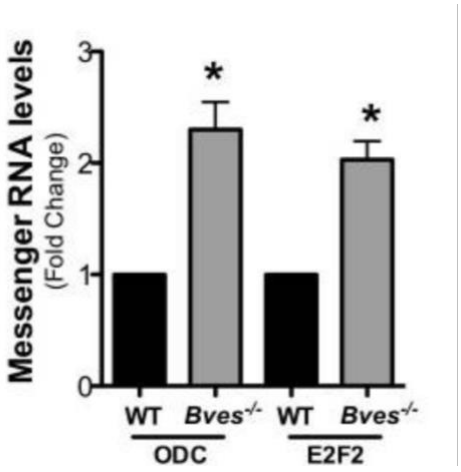

Figure 2.1: c-Myc signaling is dysregulated in *Bves*$^{-/-}$ mice in inflammatory carcinogenesis. (Parang et al., 2017) qPCR for Odc and E2f2 in enteroid cultures Student's t test, *p<0.05.

As another example, see Figure 2.2 where relative expressions are calculated based on controls for each of the PHF8, FBXO7, NCOA3, and TFAP2c treatment groups and subsequently compared within groups [5].

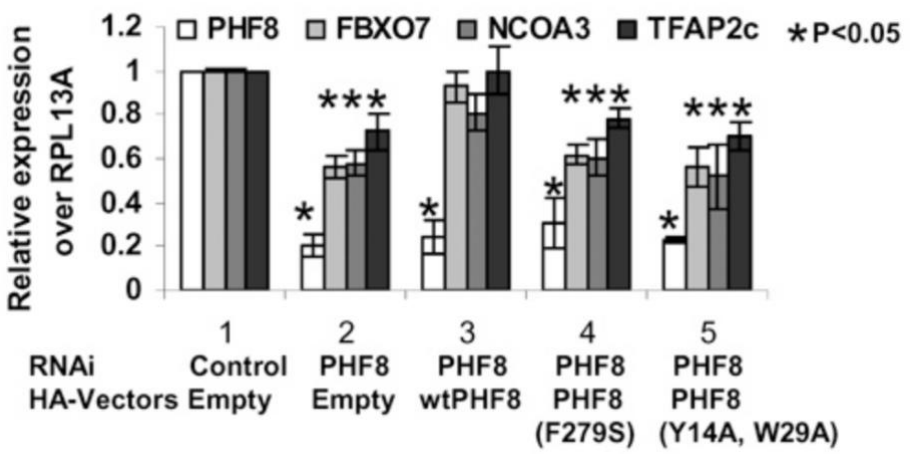

Figure 2.2: Depletion of PHF8 increases H3K9me1, H4K20me1, L3MBTL1 at TSS-bound PHF8 target genes. (Qi et al., 2010) HeLa stable cell lines were established that co-express control or PHF8 shRNA and/or indicated HA-tagged PHF8 constructs. mRNA expression of the three selected genes and endogenous PHF8 was measured by RT-Real time PCR. RPL13A was used as an internal control. P-values were obtained by t-test by comparing the data from column 2, 3, 4 and 5 with those of the corresponding controls (column 1).

In another case, fold changes are calculated in the same manner, but the t-test is performed to evaluate whether knockout mice have a significantly different treatment effect across groups. As we will explore further in Chapter 3, this analysis setting comes with statistical consequences that have previously not been acknowledged by researchers using it. An example of this setting is shown in Figure 2.3. The problematic step is in the comparison of Vector to SIRT1 to shRNA-SIRT1 as we suspect these fold change quantities were calculated relative to separate control measures based on the “Relative Luciferase Unit” label [6].

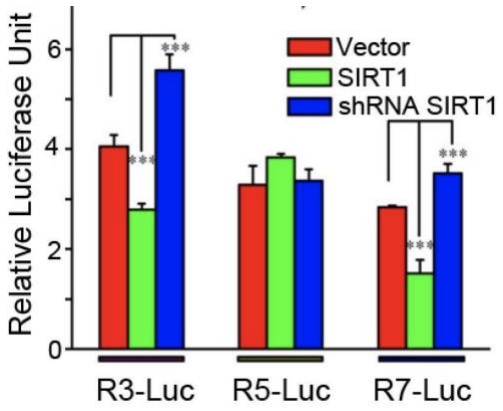

Figure 2.3: SIRT1 regulates CREB through miR-134. (Gao et al., 2010) Reporter constructs containing R3, R5, or R7 regions upstream of a minimal promoter in a luciferase reporter were co-transfected with SIRT1, SIRT1 shRNA, or empty vector.

Throughout the literature, there are more examples falling in the harmless case of fold change analysis, but the harmful case is still present and published in a variety of high-impact journals. It remains useful to examine the assumptions made by researchers using this type of parametric test and how the violation of these assumptions can affect the validity of the conclusions made.

2.2 RT-PCR

The process of RT-PCR involves the amplification of ribonucleic acid (RNA) targets in order to measure gene expression (Figure 2.4) [7]. In this process, RNA is isolated from cells and a portion of the RNA is transcribed into complementary DNA (cDNA) by reverse transcriptase. The cDNA may then be diluted to fall within the range of the machine's capability. Calibrator and experimental samples are then set up in a 96-well plate for analysis [8]. A calibrator, or control, sample of cDNA from untreated cells or tissues is used as the basis for comparison in the relative gene expression method. For the purposes of our case study and simulation, the housekeeping gene is glyceraldehyde 3-phosphate dehydrogenase (*Gapdh*) and the target gene of interest is *IL23a* encoding the interleukin-23 subunit alpha protein [9]. The final step in the RT-PCR process is to amplify the diluted target and housekeeping genes through PCR and analyze using scientific instruments. The output of RT-PCR is a count of the number of cycles needed for the fluorescence generated within a reaction to cross a pre-specified threshold (C_T). The C_T values follow a logarithmic distribution and are used directly in the comparative C_T method of

analysis in which data are normalized to *Gapdh* and standardized to the average of a control group. Further explanation of this analytic method as well as an evaluation of its statistical validity by way of simulation study are described in Chapter 4.

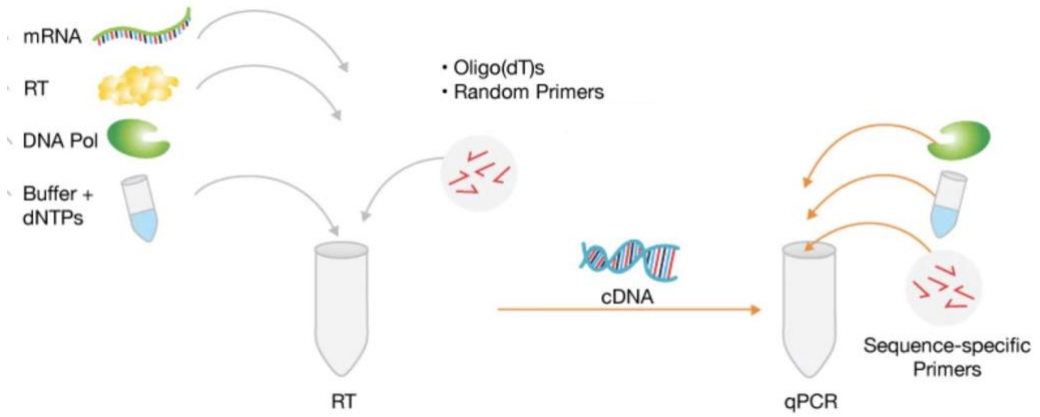

Figure 2.4: Two-step reverse transcription PCR process. (ThermoFisher)

Previous papers have shown evaluations of RT-PCR data using plots of relative quantities as in Figure 2.5 [10,11]. In this particular study, the researchers aimed to determine if IL-21-deficient mice express higher expression levels in Th17-inducing cytokines. To do so, they measured innate cytokines which were known to have potential impacts on Th17 cell expression in the PPs of wild-type and knockout mice. As mentioned in the Figure 2.2 caption, these relative measures normalized cycle threshold values to both *Gapdh* and uninfected wild-type mice. They then performed a Student's t-test and reported significant p-values within each cytokine group. Based on these results, the researchers concluded that *H. pylori*-infected IL-21$^{-/-}$ mice expressed significantly higher levels of *Il1b* and *Il23a* than wild-type mice, but no significant differences in the remaining cytokine transcript levels was seen.

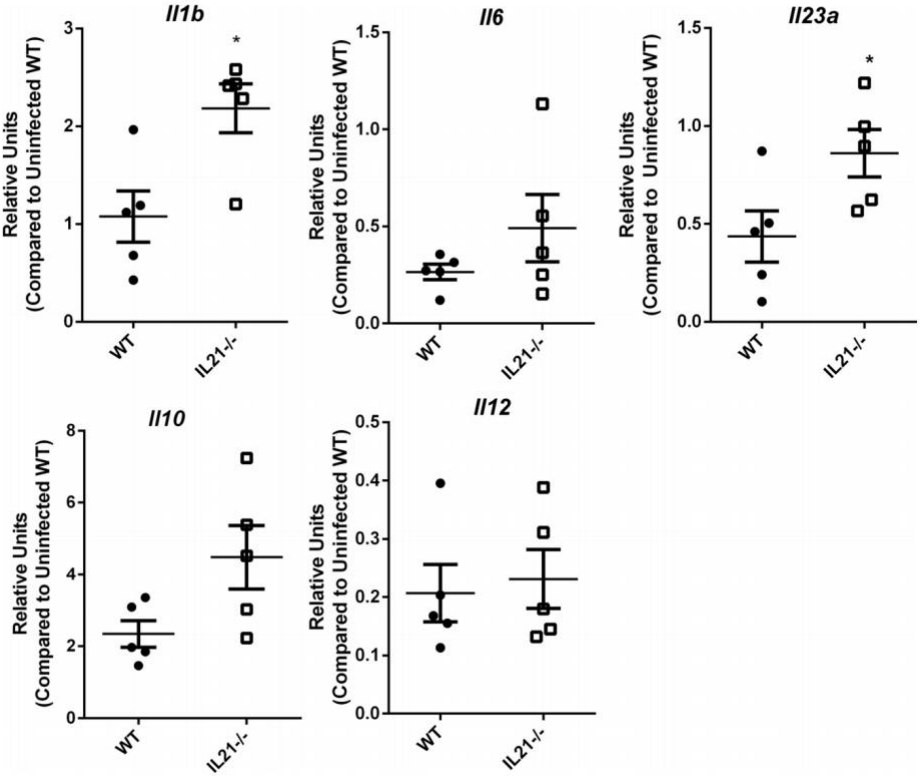

Figure 2.5: Proinflammatory gene expression in the PPs of *H. pylori*-infected mice. (Yasmin et al., 2019) Real time RT-PCR was used to measure *Il1b*, *Il6*, *Il10*, *Il12*, and *Il23* expression levels in PPs from *H. pylori*-infected mice at 1 month post-infection. Relative units were calculated using *Gapdh* as the endogenous control and tissue from uninfected WT mice as the calibrator sample. Five mice per genotype were measured in these assays. An unpaired Student's t-test was performed to test for statistical significance. *, $P < 0.05$.

2.3 STATISTICAL EVALUATION

In both fold change and RT-PCR normalization methods, we observe that in previous research, the differences between relative quantities have been evaluated by way of a p-value from a two-sample t-test. The use of this statistical test involves a set of assumptions, including: the data follow the normal probability distribution, the variances of the two populations are equal, and the two samples are independent. If these assumptions hold true, then the null and alternative hypotheses in (2.1) may be used to evaluate whether the relative quantity in the knockout group is different from that of the wild-type group and the inferences from the test will be valid.

$$H_0: \mu_{KO} = \mu_{WT}$$
$$H_A: \mu_{KO} \neq \mu_{WT} \qquad (2.1)$$

CHAPTER 3

FOLD CHANGE NORMALIZATION

Fold change normalization techniques are commonly used by researchers performing studies on mice. In this setting, there are two genotype groups: wild-type and knockout, as well as two treatment groups: control and treatment. The goal of these experiments is to determine the impact of a treatment by comparing the treatment values in wild-type and knockout groups, while simultaneously accounting for the difference in control values from both genotypes. A popular method used by scientists attempting to answer this research question is by calculating a fold change ratio which places treatment values relative to the average of their corresponding control values. In this chapter, we examine the Type I error rates of a variety of simulation studies in an attempt to identify patterns of error rate inflation.

We know based on the formula of the t-statistic (Equation 3.7) and the underlying assumptions of the t-test itself, that there are a variety of factors which may impact Type I error. In the following sections, we explore the effects of changes in theoretical fold change values, fold change magnitude, sample size, variance, and relative quantity derivation on error rate. We also make recommendations as to when the t-test is an appropriate analysis technique versus when alternative methods should be used.

3.1 DEFINITIONS AND DERIVATIONS

Fold changes are calculated as normalized measures within each genotype group. It is a measure in which treatment values are made relative to control values by way of dividing or subtracting. Fold changes calculated by dividing treatment values by a common control group average is by far the most popular. For the purposes of our work, we will simulate our own data, calculate relative fold change measures, and then evaluate analytic methods for statistical validity. As a default case, we will consider data which are generated from a normal distribution, $Y_{Normal} \sim N(\mu, \sigma^2)$. The probability density function, expectation, and variance are known to be the following.

$$f(Y_{Normal}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.1}$$

$$E[Y_{Normal}] = \mu \tag{3.2}$$

$$Var[Y_{Normal}] = \sigma^2 \tag{3.3}$$

In this standard normalization case, the fold change relative ratios are calculated in each genotype group, wild-type and knockout, as the following.

$$FC = \frac{x_{1T}}{\bar{x}_C}, \frac{x_{2T}}{\bar{x}_C}, \frac{x_{3T}}{\bar{x}_C}, \ldots \tag{3.4}$$

It is important to note that the fold change values theoretically follow a Cauchy distribution, $Y_{Cauchy} \sim Cauchy(x_0, \gamma)$ where x_0 represents location and γ represents scale, as the relative quantities involve a ratio of normal distributions. The probability density function of a Cauchy is known, but its expectation and variance are undefined. In order to approximate these moments, we recognize that the fold change distribution will be one of ratios. Because of this, we make use of the equations for approximating the expectation and variance of a ratio of random variables 3.5 and 3.6 which are based on Taylor series expansion theory [12,13]. We will make use of these approximation formulas to describe the distribution of fold change values.

$$E[FC] = E\left[\frac{Y_T}{Y_C}\right] \approx \frac{\mu_T}{\mu_C} \tag{3.5}$$

$$\begin{aligned} Var[FC] &= Var\left[\frac{Y_T}{Y_c}\right] \\ &\approx \frac{1}{\mu_C^2} Var[Y_T] + \frac{\mu_T^2}{\mu_C^4} Var[Y_C] - 2\frac{\mu_T}{\mu_C^3} Cov[Y_T, Y_C] \\ &= \left(\frac{\mu_T}{\mu_C}\right)^2 \left(\frac{Var[Y_T]}{\mu_T^2} + \frac{Var[Y_C]}{\mu_C^2} - 2\frac{Cov[Y_T, Y_C]}{\mu_T \mu_C}\right) \\ &= \left(\frac{\mu_T}{\mu_C}\right)^2 \left(\frac{\sigma_T^2/n_T}{\mu_T^2} + \frac{\sigma_C^2/n_C}{\mu_C^2}\right) \end{aligned} \tag{3.6}$$

Where $Cov[Y_T, Y_C] = 0$ since these values are independent.

After the fold changes are calculated, it is common in the basic science field to evaluate differences between wild-type and knockout groups using Student's t-test. This statistical test has the following test statistic (3.7) and degrees of freedom using a Satterthwaite approximation (3.8) for samples with unequal variance.

$$t = \frac{\overline{FC}_{WT} - \overline{FC}_{KO}}{\sqrt{\frac{sd\overline{FC}_{WT}^2}{n_{WT}} + \frac{sd\overline{FC}_{KO}^2}{n_{KO}}}} \tag{3.7}$$

$$df = \frac{\left(\frac{sd\overline{FC}_{WT}^2}{n_{WT}} + \frac{sd\overline{FC}_{KO}^2}{n_{KO}}\right)^2}{\frac{(sd\overline{FC}_{WT}^2/n_{WT})^2}{n_{WT}-1} + \frac{(sd\overline{FC}_{KO}^2/n_{KO})^2}{n_{KO}-1}} \tag{3.8}$$

This test statistic yields a p-value which is often reported along with a conclusion about treatment effect. What has previously gone unnoticed, however, is the fact that these methods lead to Type I error inflation due to a combination of the dependency induced in the calculation of fold change where values are divided by a common random variable as well as the non-normality of the fold change distribution. In order to avoid this inflation, we recommend performing a linear regression as follows.

$$\begin{aligned} Y = \beta_0 + \beta_1[Genotype = KO] + \beta_2[Group = Trt] \\ + \beta_3[Genotype = KO] \times [Group = Trt] \end{aligned} \tag{3.9}$$

Another improvement which controls Type I error rate would be to weight the linear regression model by $\frac{1}{S^2}$ where S represents the observed standard deviation of the generated data. These models will allow for the proper evaluation of varying treatment effects between the wild-type and knockout groups without making any underlying assumptions, instead using the data itself.

3.2 SIMULATION SETUP

We make use of simulation studies in order to determine whether the conclusions drawn from are statistically valid. In particular, we focus on the Type I error rates observed from simulation. First, we make random draws from the designated data-generating distribution to simulate raw data values for this study based on the user inputs of number of samples, means, and standard deviations for each of the wild-type control, wild-type treatment, knockout control, and knockout treatment groups of mice. As these experiments may be performed on a wide variety of measures, there is no upper limit on the range of possible values, the generated data must only be non-negative.

Next, we derive relative fold change quantities according to equation (3.4) in the standard case or using an alternative definition as will be described in section 3.4.5. We also capture the observed mean and variance of the raw and fold change measures to use later for validation and comparison. We then run both parametric and non-parametric statistical tests between the wild-type and knockout groups of the designated fold change measure. Student's t-test will be used as the parametric test of the difference in means between the wild-type and knockout fold changes under an assumption that the fold changes are normally distributed. The Wilcoxon rank sum test also tests the difference in means between the wild-type and knockout values of the fold changes relative measures without any parametric assumptions. We also run a series of linear regression models, both unweighted and weighted by $\frac{1}{S^2}$, and capture the p-value from the test of the hypothesis that the relationship between treatment group and the outcome of interest is different in wild-type and knockout groups.

The error rate for each analytic method is calculated as the proportion of simulation runs for which the statistical test yielded a p-value less than the designated Type I error rate of 0.05. For this study, we use 10,000 simulation runs. Finally, we observe plots of the average fold changes seen across all simulations to assess distribution shape as well as p-value patterns. The R code corresponding to this simulation may be found in the Appendix section A.1.

3.3 STATISTICAL CONSIDERATIONS

Before examining the results of our different simulation scenarios, we note the statistical test assumptions that may be violated using this method of normalization. The underlying theoretical distribution of fold change values and the potential for correlation between samples depending on the relative measure calculation used are of particular concern. As mentioned in section 3.2, under our assumption that raw values are normally distributed, it follows that the distribution of fold changes will be Cauchy. Thus, the normality assumption of Student's t-test will be violated. We aim in the following sections to assess the impact of this violation.

The Cauchy is known to be a skewed distribution with undefined first and second moments and is also a distribution for which the Central Limit Theorem (CLT) does not apply. Because of this, we expect to see that increasing the sample size does not affect the error rates seen. We see an example of the skewness of the distribution in Figure 3.1 below.

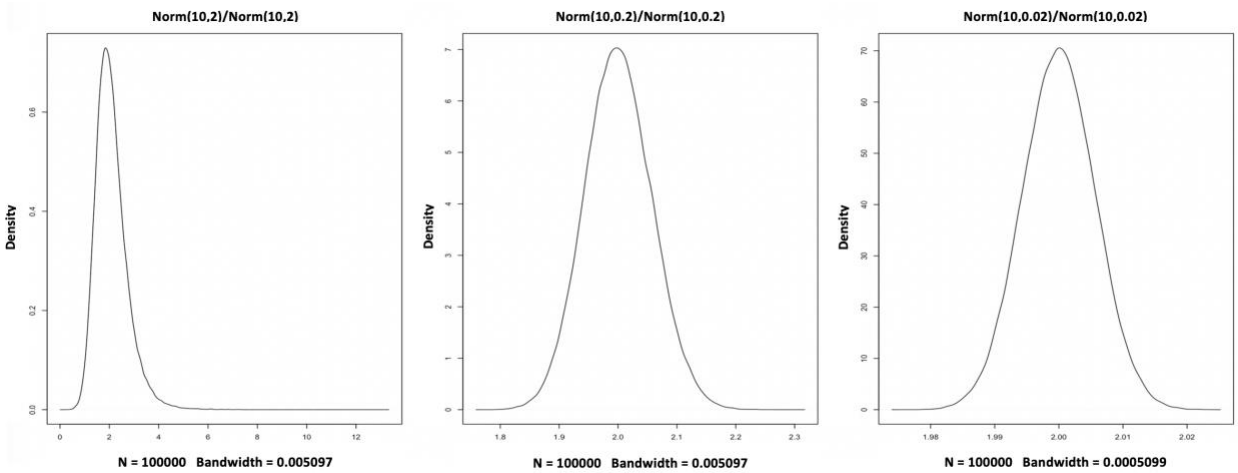

Figure 3.1: Density plots of fold change quantities with various standard deviation settings.

As seen in Figure 3.1, it is possible for the relative quantity distribution to be nearly symmetric, and therefore not in direct violation of the normality assumption of the t-test, when there is very small variability in the treatment and control measures; although this information is mostly useful on a theoretical level, researchers do not have control of this measure in practice.

As displayed in Section 3.4.5 below, there is potential for an induced correlation structure between treatment and control values depending on how the relative quantity is being calculated. This correlation will not be accounted for by the statistical tests used and will violate the independence assumption of both parametric and non-parametric tests when it is present.

3.4 ERROR RATE INFLATION PATTERNS

Let us examine one particular simulation scenario where the theoretical fold change value is 2 ($\mu_c = 200, \mu_T = 400$), there will be 3 control mice and 9 treatment mice, and the treatment standard deviation is double the control ($\sigma_c = 10, \sigma_T = 20$). We will consider this scenario to be the basis for comparison when the simulation settings are changed.

Once the simulation has been run, we observe the Type I error rates for the various analytic methods in Table 3.1. We see Type I error for both parametric and non-parametric statistical tests inflated above the 0.05 level. We also see that the error rate is controlled slightly more using unweighted linear regression and is completely controlled using weighted linear regression. In Figure 3.2, we see that under these simulation settings, the fold change distribution appears to be only slightly skewed. Figure 3.3 indicates that in simulation runs where the average fold change value is more extreme in both genotype groups simultaneously, the p-value is more

likely to be less than 0.05 and contribute to the inflated Type I error. We observe a similar pattern in that when the average fold change denominator is more extreme in both genotype groups simultaneously, the p-value is more likely to be less than 0.05. We do not see this pattern hold when observing the fold change numerator values.

Table 3.1: Type I error rates for the default normalization scenario $n = \{3, 3, 9, 9\}, \mu = \{200, 200, 400, 400\}$ and $\sigma = \{10, 10, 20, 20\}$. Where, $n = \{n_{C,WT}, n_{C,KO}, n_{T,WT}, n_{T,KO}\}, \mu = \{\mu_{C,WT}, \mu_{C,KO}, \mu_{T,WT}, \mu_{T,KO}\}$, and $\sigma = \{\sigma_{C,WT}, \sigma_{C,KO}, \sigma_{T,WT}, \sigma_{T,KO}\}$.

Type I Error Rate			
T-test on FC	Wilcox test on FC	Unweighted linear regression	Linear regression weighted by $\frac{1}{S^2}$
0.30	0.28	0.011	0.048

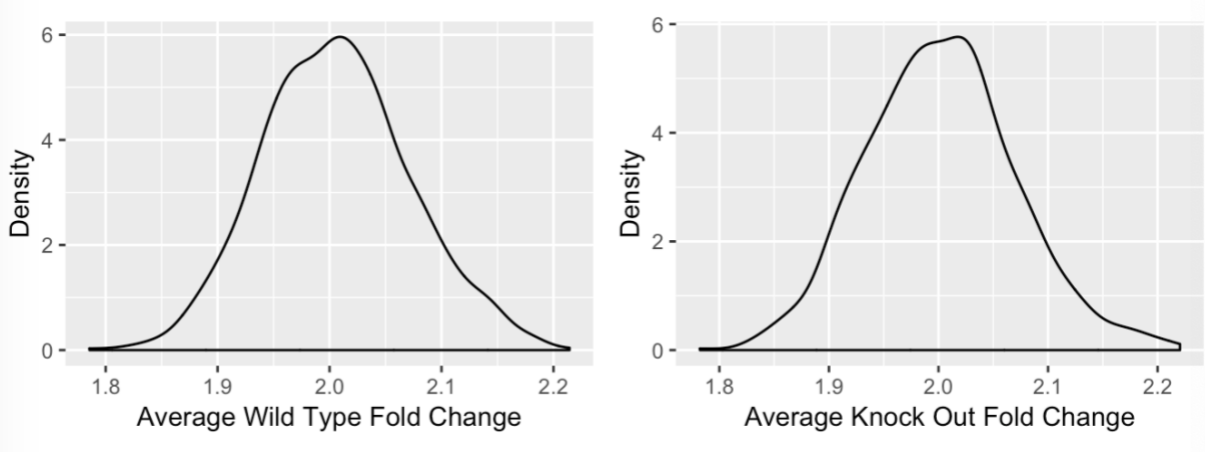

Figure 3.2: Density of average fold change values across all simulations for each genotype group.

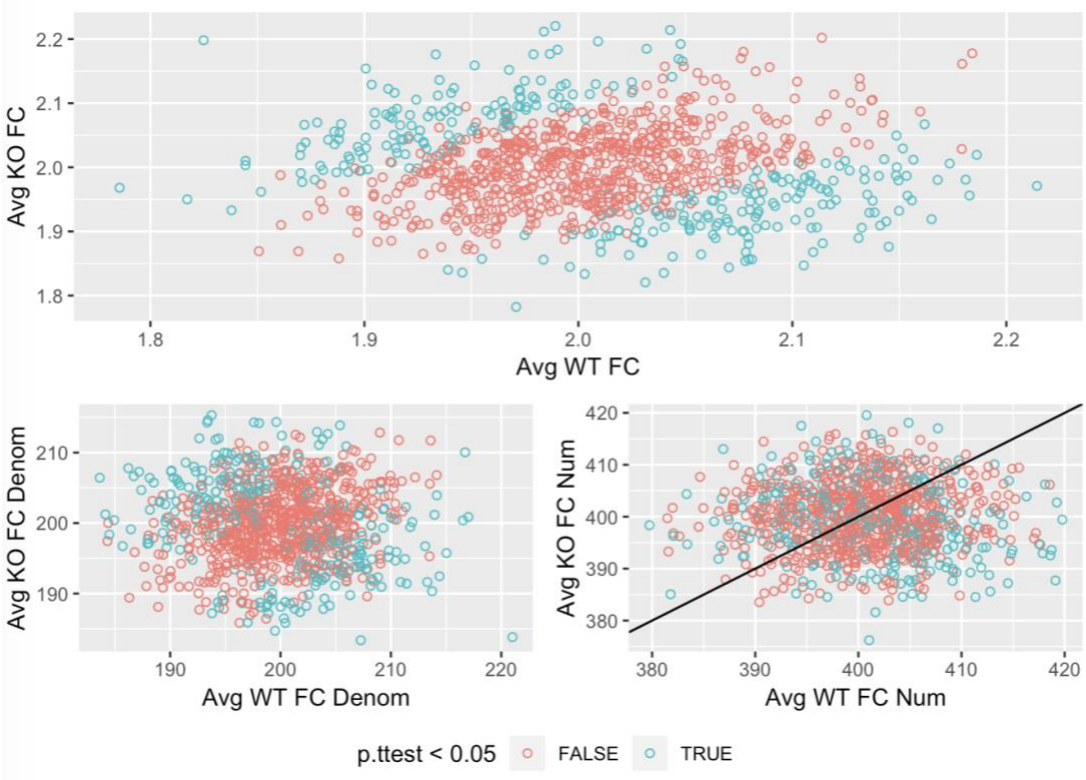

Figure 3.3: Density of average fold change values across all simulations for each genotype group.

Based on equation (3.6), we expect the theoretical variance for both the wild-type and knockout genotypes to equal 0.0044 as shown in equation (3.10).

$$Var[FC] \approx \left(\frac{400}{200}\right)^2 \left(\frac{20^2/9}{400^2} + \frac{10^2/3}{200^2}\right) = 0.00\overline{4} \qquad (3.10)$$

After simulation, we observe the variance of the wild-type fold changes to be 0.0043 and the variance of the knockout fold changes to be 0.0047. This serves as evidence that the Taylor series approximation formula is adequate for calculating the variance of the fold change distribution based on scenario parameters without the need for simulation.

3.4.1 Changing Theoretical Fold Change

We first observe the changes in error rate seen when the theoretical fold change value increases. Recall that according to equation (3.5), the theoretical fold change value is equivalent to $\frac{\mu_T}{\mu_C}$ when data are generated using the normal distribution. The results of fold change values of 2, 4, 10, and 200 are listed in Table 3.2 below. We notice that the t-test Type I error rate is inflated in each case and increases with increased theoretical fold change when the sample size

and variability parameters are all held constant. Also observe that the non-parametric Wilcox test yielded only marginally smaller error rates. The value of the theoretical fold change value did little to change the error rates seen for both linear regressions. However, the weighted linear regression is controlling the error rate better than the unweighted counterpart.

Table 3.2: Type I error rates for normalization scenarios with changing theoretical fold change $n = \{n_{C,WT}, n_{C,KO}, n_{T,WT}, n_{T,KO}\}, \mu = \{\mu_{C,WT}, \mu_{C,KO}, \mu_{T,WT}, \mu_{T,KO}\}$, and $\sigma = \{\sigma_{C,WT}, \sigma_{C,KO}, \sigma_{T,WT}, \sigma_{T,KO}\}$.

$n = \{3, 3, 9, 9\}$ $\sigma = \{10, 10, 20, 20\}$	Type I Error Rate			
	T-test on FC	Wilcox test on FC	Unweighted linear regression	Linear regression weighted by $\frac{1}{S^2}$
$FC = 2$ $(\mu_C = 200, \mu_T = 400)$	0.30	0.28	0.011	0.048
$FC = 4$ $(\mu_C = 100, \mu_T = 400)$	0.56	0.53	0.016	0.052
$FC = 10$ $(\mu_C = 40, \mu_T = 400)$	0.81	0.79	0.007	0.047
$FC = 200$ $(\mu_C = 40, \mu_T = 8000)$	0.99	0.99	0.009	0.054

3.4.2 Scaling Fold Change Magnitude

Next, we observe the effect that a higher magnitude of mean parameters in both the control and treatment groups has on error rate. In both cases shown in Table 3.3 below, the theoretical fold change value is $\frac{\mu_T}{\mu_C} = \frac{400}{200} = \frac{4000}{2000} = 2$ even though the magnitude of mean parameters is 10 times larger. In this case, the t-test Type I error stays constant regardless of the scale of the fold-change parameters. We again see that the weighted linear regression analysis option is controlling the error rate appropriately.

Table 3.3: Type I error rates for normalization scenarios with scaled fold change magnitudes $n = \{n_{C,WT}, n_{C,KO}, n_{T,WT}, n_{T,KO}\}, \mu = \{\mu_{C,WT}, \mu_{C,KO}, \mu_{T,WT}, \mu_{T,KO}\}$, and $\sigma = \{\sigma_{C,WT}, \sigma_{C,KO}, \sigma_{T,WT}, \sigma_{T,KO}\}$.

$n = \{3, 3, 9, 9\}$ $\sigma = \{10, 10, 20, 20\}$	Type I Error Rate			
	T-test on FC	Wilcox test on FC	Unweighted linear regression	Linear regression weighted by $\frac{1}{S^2}$
$FC = 2$ $(\mu_C = 200, \mu_T = 400)$	0.30	0.28	0.011	0.048
$FC = 2$ $(\mu_C = 2000, \mu_T = 4000)$	0.31	0.28	0.011	0.059

3.4.3 Scaling Sample Size

We can also observe what happens to Type I error rates among the various analytic methods when the sample size is increased. Notice in Table 3.4 that the theoretical fold change and variability parameters remain constant as the sample size gets 100 times larger. We observe that this change does not impact the error rates observed, so increasing the number of samples does not deflate Type I error.

Table 3.4: Type I error rates for normalization scenarios with scaled sample sizes $n = \{n_{C,WT}, n_{C,KO}, n_{T,WT}, n_{T,KO}\}, \mu = \{\mu_{C,WT}, \mu_{C,KO}, \mu_{T,WT}, \mu_{T,KO}\}$, and $\sigma = \{\sigma_{C,WT}, \sigma_{C,KO}, \sigma_{T,WT}, \sigma_{T,KO}\}$.

$FC = 2\ (\mu_C = 200, \mu_T = 400)$ $\sigma = \{10, 10, 20, 20\}$	Type I Error Rate			
	T-test on FC	Wilcox test on FC	Unweighted linear regression	Linear regression weighted by $\frac{1}{S^2}$
$n = \{3, 3, 9, 9\}$	0.30	0.28	0.011	0.048
$n = \{300, 300, 900, 900\}$	0.33	0.33	0.007	0.049

3.4.4 Altering Standard Deviation Ratio

It is also possible to alter the ratios of variability parameters $\frac{\sigma_T}{\sigma_C}$ between the wild-type and knockout phenotype groups. As shown in Table 3.5, within each ratio grouping, the Type I error rate remains the same as long as the ratio between $\frac{\sigma_T}{\sigma_C}$ stays constant, regardless of scale. Although, increasing the knockout group's standard deviation ratio yielded a deflation of t-test Type I error. In the case where the sigma ratios are equal, we also see that it is possible for the unweighted linear regression to outperform the weighted version and be closest to the 0.05 level.

Table 3.5: Type I error rates for normalization scenarios with various sigma ratios $n = \{n_{C,WT}, n_{C,KO}, n_{T,WT}, n_{T,KO}\}, \mu = \{\mu_{C,WT}, \mu_{C,KO}, \mu_{T,WT}, \mu_{T,KO}\}$, and $\sigma = \{\sigma_{C,WT}, \sigma_{C,KO}, \sigma_{T,WT}, \sigma_{T,KO}\}$.

	$n = \{3, 3, 9, 9\}$ $FC = 2\ (\mu_C = 200, \mu_T = 400)$	Type I Error Rate			
		T-test on FC	Wilcox test on FC	Unweighted linear regression	Linear regression weighted by $\frac{1}{S^2}$
WT ratio=1 KO ratio=1	$\sigma = \{10, 10, 10, 10\}$	0.56	0.52	0.054	0.069
	$\sigma = \{10, 20, 10, 20\}$	0.52	0.51	0.058	0.083

WT ratio=1 KO ratio=2	$\sigma = \{10, 20, 10, 40\}$	0.36	0.31	0.015	0.059
	$\sigma = \{20, 40, 20, 80\}$	0.32	0.32	0.018	0.070

3.4.5 Various Relative Quantity Normalization Scenarios

As previously mentioned, the standard method used in calculating relative fold change measure involves dividing treatment data by a common averaged control value. In this section, we observe the t-test Type I error rates yielded when the relative quantity calculation is changed. We are particularly interested in a calculation of a relative ratio without a common control, and the calculation of relative differences between common and pairwise controls.

3.4.5.1 Ratio relative to pairwise controls

In order to assess whether the use of a common control in calculating relative measures is the sole cause of error rate inflation, we explore an alternate relative quantity definition. In the case where the relative quantity involves a ratio of treatment data relative to pairwise controls, the fold changes in each genotype group are calculated as the following.

$$FC = \frac{x_{1T}}{x_{1C}}, \frac{x_{2T}}{x_{2C}}, \frac{x_{3T}}{x_{3C}}, \ldots \qquad (3.11)$$

Using this alternative method to calculate fold change yields a t-test Type I error of approximately 7% as compared to 40% for a standard fold change case. We likely see this slightly inflated Type I error rate since the fold change measures still follow a non-normal distribution. Although, since the fold changes are calculated using pairwise controls in this case, the values within treatment group are no longer correlated.

3.4.5.2 Difference relative to average control

We are also curious if the error rate inflation is still seen when the relative quantity is derived using a difference instead of a ratio. In this case, we are no longer working with the Cauchy distribution; instead, the fold change distribution should remain normally distributed. We will first explore the difference between the treatment data and a common control. In the case where the relative quantity involves a difference between treatment data relative to the

average of its corresponding control, the relative differences (RD) in each genotype group are calculated as in equation (3.11).

$$RD = (x_{1T} - \bar{x}_C), (x_{2T} - \bar{x}_C), (x_{3T} - \bar{x}_C), \ldots \quad (3.12)$$

Using this derivation yields a t-test Type I error rate of approximately 15% as compared to 40% for a standard fold change case. Based on the error rates seen in previous settings, we were under the impression that Cauchy distribution was causing most of the error rate inflation, although even without dividing, we encounter the problem. We hypothesize that the induced dependency due to a common control inflates the Type I error rate in this case. Although, the inflation is not as extreme since the relative difference distribution is normal instead of Cauchy.

3.4.5.3 Difference relative to pairwise controls

Finally, we again consider a relative quantity calculation involving a difference, but no longer with a common control. Instead, subtracting paired, and thus no longer independent data. In the case where the relative quantity involves a difference between treatment data relative to pairwise controls, the relative differences in each genotype group are calculated as the following.

$$RD = (x_{1T} - x_{1C}), (x_{2T} - x_{2C}), (x_{3T} - x_{3C}), \ldots \quad (3.13)$$

This final derivation method yields a t-test Type I error rate of approximately 5% as compared to 40% for a standard fold change case. Since these relative differences are normally distributed and independent from one another, the Type I error rate is controlled.

3.5 RECOMMENDATIONS

As shown in the various scenarios in section 3.4, we can manipulate enough settings within the simulation to yield a t-test Type I error ranging anywhere from 0.05 to 0.99. It is also important to note that nearly all cases had inflated error. Through examining these various scenarios, it is clear that there is not a single root explanation as to why the t-test error rate is being inflated in these settings. Instead, it appears that the Cauchy distribution, measures relative to a common control, theoretical fold change values, and variability relationships between treatment groups all work in tandem to invalidate the use of the t-test in these settings. A general pattern we noticed when raw data are generated from a normal distribution is that the magnitude of parameters when comparing wild-type to knockout does not matter, it is the relationship of treatment and control parameters that makes a difference.

We recommend researchers instead use analysis of variance with weights $\frac{1}{S^2}$ in order to control the Type I error of their analysis. Although, with the inevitably small sample sizes seen in this field of research, this recommended method of analysis will suffer from low power. We also saw good performance of relative measures calculated using differences instead of ratios. We next examine normalization schemes using RT-PCR data which rely more heavily on relative differences to see if the error rate inflation issues persist.

CHAPTER 4

COMPARATIVE C_T NORMALIZATION METHOD USING RT-PCR DATA

Scientists often make use of RT-PCR data in order to compare gene expression levels between knockout and wild-type groups, while simultaneously accounting for the differences between treated and untreated samples. The current analysis procedures used with RT-PCR data present another normalization technique involving quantities relative to a common control. In contrast to the fold change setting, there are derived relative quantities in RT-PCR analysis which involve only differences of these shared controls. Although, another commonly used relative quantity does involve a ratio calculation with respect to a shared control as part of the normalization process, echoing the fold change method. We will examine in the following sections if the same Type I error inflation issues persist when using the t-test in this similar setting.

There are several analytic methods that are programmed into the RT-PCR instruments used by scientific research labs, such as Applied Biosystems®, which aid researchers in determining this effect. Researchers make use of the Relative Standard Curve method for analyzing RT-PCR data when testing small numbers of target genes and samples. This method makes no assumption about the PCR efficiencies, instead measuring relative quantities directly. Due to the fact that this method is able to interpolate unknown sample quantities by use of a standard curve, discrete changes in expression may be measured and little validation is needed. Alternatively, when testing a large number of target genes and samples or when validating results from previous experiments, the Comparative C_T Method ($\Delta\Delta C_T$) is commonly used. This is also regularly referred to as the delta-delta method. The $\Delta\Delta C_T$ method comes with the caveat that in order for the relative calculation to be valid, the efficiency of the target and reference amplification must be relatively equivalent. This measure of amplification efficiency is typically found through previously performed validation experiments. Generally, both methods require specific and proper plate set-up in order for the software to make the appropriate derivations.

4.1 DEFINITIONS AND DERIVATIONS

We focus on the calculation and use of a relative delta-delta measure for comparison, as it is a popular analytic method in which the process of normalization should be evaluated for statistical validity. The following arithmetic formula is a normalized, relative quantity (RQ) which represents the amount target gene which is normalized to a housekeeping gene and relative to a calibrator sample [14].

$$RQ = 2^{-\Delta\Delta C_T} \tag{4.1}$$

Theoretically, this formula used is based on the general equation which calculates the exponential amplification of PCR (Equation 4.2) [15].

$$X_n = X_0 \times (1 + E_X)^n \tag{4.2}$$

where:

X_n = Number of target molecules at cycle n
X_0 = Initial number of target molecules
E_X = Efficiency of target amplification
n = Number of cycles

In our simulation of RT-PCR data, we are interested in quantifying the number of molecules in both the target (e.g., *IL23a*) and housekeeping (e.g., *Gapdh*) reference groups at the cycle number at which the fluorescence threshold is crossed; $X_{T,IL23a}$ and $X_{T,Gapdh}$, respectively. Using equation (4.2) as a basis, the number of cycles may be replaced with the cycle number at which the amount of amplified target reaches a fixed threshold C_T and we know that the threshold number of molecules will be an integer constant. Equation (4.3) describes the exponential amplification of the *IL23a* target reaction.

$$X_{T,IL23a} = X_{0,IL23a} \times (1 + E_{IL23a})^{C_{T,IL23a}} \tag{4.3}$$

Similarly, equation 4.4 describes the same quantity for the *Gapdh* housekeeping reaction.

$$X_{T,Gapdh} = X_{0,Gapdh} \times (1 + E_{Gapdh})^{C_{T,Gapdh}} \tag{4.4}$$

In order to normalize, we relate the target molecules to the reference molecules by taking a ratio.

$$\frac{X_{T,IL23a}}{X_{T,Gapdh}} = \frac{X_{0,IL23a} \times (1 + E_{IL23a})^{C_{T,IL23a}}}{X_{0,Gapdh} \times (1 + E_{Gapdh})^{C_{T,Gapdh}}} \tag{4.5}$$

The $IL23a_T$ and $Gapdh_T$ threshold values are dependent on several factors related to scientific variation and the machine itself and therefore will not necessarily equal one another, though will equal a constant K. Under the assumption that the efficiencies of the target and reference reactions are equal $E_{IL23a} = E_{Gapdh} = E$, the equation may be rewritten as in equation (4.6).

$$K = \frac{X_{0,IL23a}}{X_{0,Gapdh}} \times (1+E)^{C_{T,IL23a} - C_{T,Gapdh}} \tag{4.6}$$

Next, $\frac{X_{0,IL23a}}{X_{0,Gapdh}}$ is defined as the normalized amount of initial target molecules X_N and the difference in threshold cycles $C_{T,IL23a} - C_{T,Gapdh}$ for target and reference molecules is defined as ΔC_T. By applying these new definitions and rearranging, we yield Equation (4.7).

$$X_N = K \times (1+E)^{-\Delta C_T} \tag{4.7}$$

Finally, we again normalize by taking a ratio such that the treated samples T are relative to their untreated counterparts UT.

$$\frac{X_{N,T}}{X_{N,UT}} = \frac{K \times (1+E)^{-\Delta C_{T,T}}}{K \times (1+E)^{-\Delta C_{T,UT}}} = (1+E)^{-\Delta\Delta C_T} \tag{4.8}$$

Notice that the delta-delta relative quantity $\Delta\Delta C_T$ is equal to $\Delta C_{T,T} - \Delta C_{T,UT}$. Under the assumption set by Applied Biosystems® that the efficiency is nearly equal to one, the point estimate of the relative quantity becomes $RQ = 2^{-\Delta\Delta C_T}$. Statistical tests are then performed on RQ_{WT} and RQ_{KO} to test the following hypotheses:

$$H_0: \mu_{RQ_{KO}} = \mu_{RQ_{WT}}$$

$$H_A: \mu_{RQ_{KO}} \neq \mu_{RQ_{WT}}$$

In terms of quantifying the variation of the above relative quantity, we know that at the first normalization step of comparing *IL23a* to *Gapdh* genes, the standard error is calculated as in equation (4.9) where n is the number of wells for a given sample.

$$\Delta C_T SE = \sqrt{\frac{C_T SD^2_{target}}{n} + \frac{C_T SD^2_{GAPDH}}{n}} \tag{4.9}$$

The quantification of variance for the second normalization step for comparing treated to untreated samples is less clear. According to the Applied Biosystems® Guide to Performing Relative Quantitation of Gene Expression Using RT-PCR, the subtraction of $\Delta C_{T,UT}$ is considered to be the subtraction of an "arbitrary constant" meaning that it has no associated

variability [14]. It follows from this logic that the standard deviation of the $\Delta\Delta C_T$ value is the same as the standard deviation of the treated quantities $\Delta C_{T,T}$. This assumption, however, ignores a non-negligible amount of variability in the normalized untreated quantities and calls into question the RT-PCR instrument's automatic calculation of a confidence interval surrounding the $2^{-\Delta\Delta C_T}$ relative quantity.

4.2 SIMULATION SETUP

In order to assess the validity of the statistical conclusions being made in this field of work, we examine error rates from simulation. We first make random draws from a normal distribution to simulate cycle threshold C_T values for the *IL23a* target and *Gapdh* housekeeping genes used for this study. We assume that the distribution of C_T values for the *Gapdh* housekeeping genes are the same for both wild-type and knockout groups and follow,

$$C_{T\,T,Gapdh} \sim Normal(\mu_{T,Gapdh}, \sigma_{T,Gapdh})$$

$$C_{T\,UT,Gapdh} \sim Normal(\mu_{UT,Gapdh}, \sigma_{UT,Gapdh})$$

for treated and untreated samples, respectively. Similarly, we assume that the *IL23a* target genes for both wild-type and knockout groups follow,

$$C_{T\,T,IL23a} \sim Normal(\mu_{T,IL23a}, \sigma_{T,IL23a})$$

$$C_{T\,UT,IL23a} \sim Normal(\mu_{UT,IL23a}, \sigma_{UT,IL23a})$$

for treated and untreated samples, respectively. For these raw measures, it is helpful to note that a smaller C_T value is indicative of larger cytokine expression since it crossed the fluorescence threshold at an earlier cycle, meaning there was more of that cytokine. Generally speaking, there do not tend to be C_T values above 35 cycles.

Next, we derive relative quantities according to the equations in section 4.1 based on the corresponding normalization scenario detailed in section 4.4. We then run both parametric and non-parametric statistical tests between the wild-type and knockout groups of various relative measures. Student's t-test will be used as the parametric test of the difference in means between the wild-type and knockout quantities of the $\Delta\Delta C_T$ and $2^{-\Delta\Delta C_T}$ relative measures under an assumption that the relative measures are normally distributed. The Wilcoxon rank sum test evaluates the difference in central tendency between the wild-type and knockout values of the $\Delta\Delta C_T$ and $2^{-\Delta\Delta C_T}$ relative measures with no parametric assumptions.

Finally, the error rate for each test in the normalization scenario is calculated as the proportion of simulation runs for which the statistical test yielded a p-value less than the designated Type I error rate of 0.05. For the purposes of this study, we use 10,000 simulation runs. We are particularly interested in knowing whether the Type I error for any of the tests is: below the 0.05 level, indicating a loss of statistical power; at the 0.05 level, demonstrating statistical validity; or inflated. The R code corresponding to this simulation may be found in the Appendix section A.2.

4.3 STATISTICAL CONSIDERATIONS

Before examining any specific normalization scenarios, we can evaluate potential threats to statistical validity based on the test assumptions that could be violated. The underlying theoretical distributions of the relative quantities being tested and the correlation between samples are of particular concern. Under our assumption that raw C_T values are normally distributed, we can follow through the derivation formulas as in section 4.1 and observe that the calculation of $\Delta\Delta C_T$ involves a series of addition, subtraction, and scaling of known normal distributions as shown in the expansion in (4.10). Thus, $\Delta\Delta C_T$ values will follow a normal distribution and not violate the normality assumption of Student's t-test.

$$\begin{aligned}\Delta\Delta C_T &= \Delta C_{T\,Treated} - \Delta C_{T\,Untreated} \\ &= (C_{T\,T,IL23a} - C_{T\,T,Gapdh}) - (C_{T\,UT,IL23a} - C_{T\,UT,Gapdh}) \\ &= \left(\frac{C_{T1\,T,IL23a} + C_{T2\,T,IL23a}}{2} - \frac{C_{T1\,T,Gapdh} + C_{T2\,T,Gapdh}}{2}\right) \\ &\quad - \left(\frac{C_{T1\,UT,IL23a} + C_{T2\,UT,IL23a}}{2} - \frac{C_{T1\,UT,Gapdh} + C_{T2\,UT,Gapdh}}{2}\right)\end{aligned} \tag{4.10}$$

The relative quantity $2^{-\Delta\Delta C_T}$, however, does not follow a theoretically normal distribution. As described above, we know the $-\Delta\Delta C_T$ quantity is normal, but 2 raised to the power of a normal is not a known distribution. The only recognized relationship is that $-log_2(2^{-\Delta\Delta C_T})$ is normally distributed. As seen in Figure 4.1, it is possible for the relative quantity distribution to be nearly symmetric, and therefore not in direct violation of the normality assumption of the t-test, when the standard deviation of the $\Delta\Delta C_T$ value is very low; although the variance of the normalized ΔC_T values is not something within the control of researchers when performing experiments. What Figure 4.1 also highlights is the potential for relative quantity

outliers when the $\Delta\Delta C_T$ values have a moderate amount of variability. This combination of skewness and wide variability calls into question the use of a parametric statistical test for evaluating research questions of the difference between group averages.

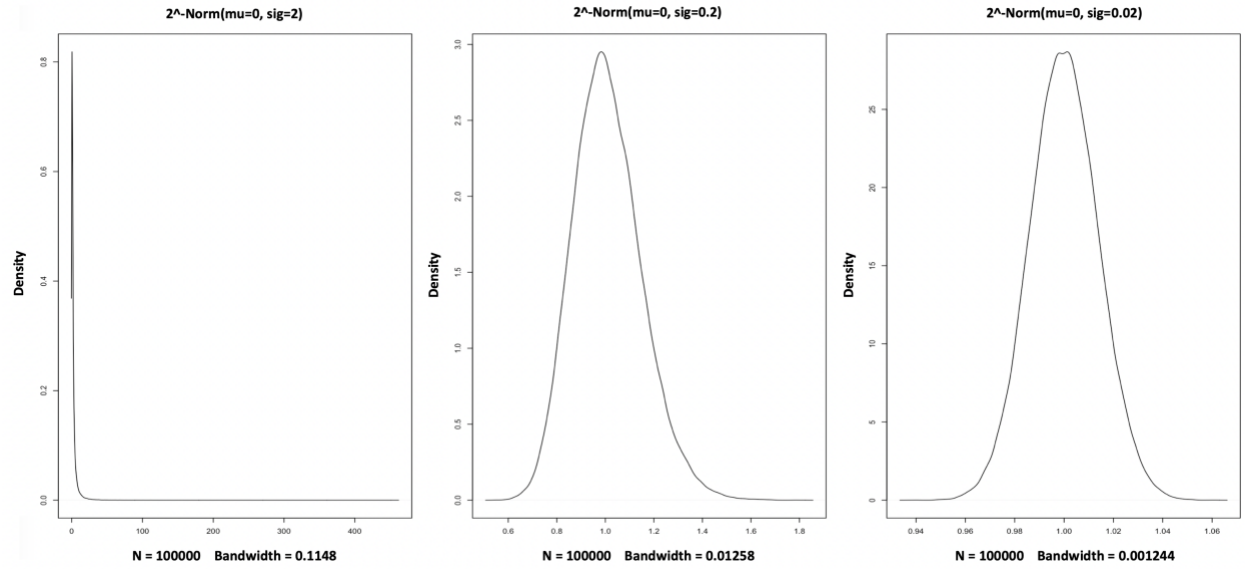

Figure 4.1: Density plots of $2^{-\Delta\Delta C_T}$ relative quantities with various standard deviation settings for the normally distributed $\Delta\Delta C_T$ values.

When considering the non-parametric Wilcoxon rank sum test, it is important to note that due to the nature of the underlying ranking involved, the test will not be able to detect small group differences at a 0.05 level when the sample size is less than 4 [16].

As shown below in Section 4.4, depending on the normalization scheme used for deriving relative quantities, there is potential for an induced correlation structure that is not being accounted for in the use of statistical tests. The group-level correlation between samples is in direct violation of the independence assumption of both parametric and non-parametric tests.

Finally, recall that equation (4.8), only yields the $2^{-\Delta\Delta C_T}$ relative measure due to the assumption that the amplification efficiency is equal to one. Previous work has quantified the impact of this efficiency assumption in terms of error percent [17].

$$ErrorPercent = \left(100 \times \frac{2^{C_T}}{(1+E)^{C_T}}\right) - 100, \qquad (4.11)$$

meaning, that if the efficiency is actually 0.9 instead of 1 in 25 cycles, for example, we will have a very high percentage of error (Equation 4.12), calling into question the practicality of this assumption.

$$ErrorPercent = \left(100 \times \frac{2^{25}}{(1+0.9)^{25}}\right) - 100 = 261\% \qquad (4.12)$$

4.4 NORMALIZATION SCENARIOS

There are several plate setups used by various scientific labs which dictate how the relative quantities used for comparison are derived. Within a single target gene, the calculation of ΔC_T remains the same, regardless of scenario type. The calculation of $\Delta\Delta C_T$, however, depends on the setup of the plate and the research question being asked as described in sections (4.4.1-4) below.

4.4.1 Single Treated relative to Single Untreated

Suppose, for the purposes of explanation, that we have a single set of treated samples and a single set of untreated samples for both the wild-type and knockout groups. In this setting, we assume that $C_{T\ Gapdh} \sim Normal(20, 2)$ and $C_{T\ Il23a} \sim Normal(30, 2)$ regardless of whether the sample was treated or untreated. Table 4.1 shows an example of simulated data from one run of the simulation under this normalization scenario.

Table 4.1: Single *IL23a* target gene subset of simulated Applied Biosystems® instrument data when normalization is performed on a single treated relative to a single untreated sample.

	Plate	Well	Sample.Name	Target.Name	CT	CT.Mean	CT.Sd	Delta.CT.Mean	Delta.CT.SE	DelDel.CT	RQ
Untreated WT1	1	B1	WT_Th17	Gapdh	17.06	18.20	1.61				
	1	B2	WT_Th17	Gapdh	19.33						
	1	B9	WT_Th17	Il23a	26.26	28.31	2.91	10.12	2.35	0	1
	1	B10	WT_Th17	Il23a	30.37						
Treated WT1	1	C1	WT_Th17+Il17a	Gapdh	23.32	21.83	2.12				
	1	C2	WT_Th17+Il17a	Gapdh	20.33						
	1	C9	WT_Th17+Il17a	Il23a	30.06	30.98	1.29	9.15	1.75	-0.97	1.95
	1	C10	WT_Th17+Il17a	Il23a	31.89						
Untreated KO1	1	F1	IL17RA-/-_Th17	Gapdh	18.10	18.86	1.07				
	1	F2	IL17RA-/-_Th17	Gapdh	19.62						
	1	F9	IL17RA-/-_Th17	Il23a	28.94	30.10	1.65	11.24	1.39	0	1
	1	F10	IL17RA-/-_Th17	Il23a	31.27						
Treated KO1	1	H1	IL17RA-/-_Th17+Il17a	Gapdh	19.01	17.91	1.56				
	1	H2	IL17RA-/-_Th17+Il17a	Gapdh	16.80						
	1	H9	IL17RA-/-_Th17+Il17a	Il23a	30.45	29.15	1.83	11.25	1.70	0.01	0.99
	1	H10	IL17RA-/-_Th17+Il17a	Il23a	27.86						

Let's examine the steps for deriving the treated wild-type relative quantity of 1.95. First, recall that the C_T values are randomly generated from the designated normal distributions and note that

there are two wells assigned for each cytokine. We calculate the $C_{T\,Mean}$ values as the average of the two wells for each target or housekeeping gene within each wild-type group.

$$\begin{array}{rll} \text{Column:} & \textbf{CT.Mean} \quad \textbf{CT} & \textbf{CT.Mean} \quad \textbf{CT} \\ \text{Untreated WT:} & 18.20 = \dfrac{17.06 + 19.33}{2}; & 28.31 = \dfrac{26.26 + 30.37}{2} \\ \text{Treated WT:} & 21.83 = \dfrac{23.32 + 20.33}{2}; & 30.98 = \dfrac{30.06 + 31.89}{2} \end{array} \tag{4.11}$$

Next, the ΔC_T values (are calculated by subtracting the housekeeping $C_{T\,Mean}$ from the target $C_{T\,Mean}$ within each treated status wild-type group.

$$\begin{array}{rr} \text{Column:} & \textbf{Delta.CT.Mean} \quad \textbf{CT.Mean} \\ \text{Untreated WT:} & 10.12 = 28.31 - 18.20 \\ \text{Treated WT:} & 9.15 = 30.98 - 21.83 \end{array} \tag{4.12}$$

The standard errors of the ΔC_T values are calculated according to equation (4.9).

$$\begin{array}{rr} \text{Column:} & \textbf{Delta.CT.SE} \quad \textbf{CT.sd} \\ \text{Untreated WT:} & 2.35 = \sqrt{\dfrac{2.91^2}{2} + \dfrac{1.61^2}{2}} \\ \text{Treated WT:} & 1.75 = \sqrt{\dfrac{1.29^2}{2} + \dfrac{2.12^2}{2}} \end{array} \tag{4.13}$$

The $\Delta\Delta C_T$ values are derived as the difference between the treated and untreated ΔC_T values of the wild-type group.

$$\begin{array}{rr} \text{Column:} & \textbf{DelDel.CT} \quad \textbf{Delta.CT.Mean} \\ \text{Untreated WT:} & 0 = 10.12 - 10.12 \\ \text{Treated WT:} & -0.97 = 9.15 - 10.12 \end{array} \tag{4.14}$$

Finally, the relative quantity is calculated as $2^{-\Delta\Delta C_T}$.

$$\begin{array}{rr} \text{Column:} & \textbf{RQ} \quad \textbf{DelDel.CT} \\ \text{Untreated WT:} & 1 = 2^{-0} \\ \text{Treated WT:} & 1.95 = 2^{-(-0.97)} \end{array} \tag{4.15}$$

A similar exercise may be performed to yield the relative quantities within the knockout group. Note that in this scenario, statistical tests may not be performed as there is only one wild-type and one knockout value to be compared. We instead view this example as an illustration of relative measure derivation.

4.4.2 Multiple Treated relative to Single Untreated

In this scenario, the $C_{T\,Mean}$, ΔC_T, $\Delta C_T SE$, and relative quantity $2^{-\Delta\Delta C_T}$ values are calculated in the same way as in the previous normalization setting (equations 4.11, 4.12, 4.13, 4.15). What deviates is the derivations of $\Delta\Delta C_T$ values. Table 4.2 shows an example of simulated data that may be seen in this normalization scenario.

Table 4.2: Single *IL23a* target gene subset of simulated Applied Biosystems® instrument data when normalization is performed on multiple treated samples all relative to the same single untreated sample.

	Plate	Well	Sample.Name	Target.Name	CT	CT.Mean	CT.Sd	Delta.CT.Mean	Delta.CT.SE	DelDel.CT	RQ
Untreated WT 1	1	A1	WT_Th17	Gapdh	17.06	18.20	1.61				
	1	A2	WT_Th17	Gapdh	19.33						
	1	A9	WT_Th17	Il23a	26.26	28.31	2.91	10.12	2.35	0	1
	1	A10	WT_Th17	Il23a	30.37						
Treated WT 1	1	B1	WT_Th17+Il17a	Gapdh	23.32	21.83	2.12				
	1	B2	WT_Th17+Il17a	Gapdh	20.33						
	1	B9	WT_Th17+Il17a	Il23a	30.06	30.98	1.29	9.15	1.75	-0.97	1.95
	1	B10	WT_Th17+Il17a	Il23a	31.89						
Treated WT 2	1	C1	WT_Th17+Il17a	Gapdh	22.35	21.28	1.51				
	1	C2	WT_Th17+Il17a	Gapdh	20.21						
	1	C9	WT_Th17+Il17a	Il23a	29.55	30.64	1.54	9.36	1.53	-0.76	1.70
	1	C10	WT_Th17+Il17a	Il23a	31.73						
Treated WT 3	1	D1	WT_Th17+Il17a	Gapdh	20.17	19.90	0.38				
	1	D2	WT_Th17+Il17a	Gapdh	19.63						
	1	D9	WT_Th17+Il17a	Il23a	24.45	24.78	0.47	4.88	0.43	-5.23	37.60
	1	D10	WT_Th17+Il17a	Il23a	25.12						
Untreated KO 1	1	E1	IL17RA-/-_Th17	Gapdh	18.10	18.86	1.07				
	1	E2	IL17RA-/-_Th17	Gapdh	19.62						
	1	E9	IL17RA-/-_Th17	Il23a	28.94	30.10	1.65	11.24	1.39	0	1
	1	E10	IL17RA-/-_Th17	Il23a	31.27						
Treated KO 1	1	F1	IL17RA-/-_Th17+Il17a	Gapdh	19.01	17.91	1.56				
	1	F2	IL17RA-/-_Th17+Il17a	Gapdh	16.80						
	1	F9	IL17RA-/-_Th17+Il17a	Il23a	30.45	29.15	1.83	11.25	1.70	0.01	0.99
	1	F10	IL17RA-/-_Th17+Il17a	Il23a	27.86						
Treated KO 2	1	G1	IL17RA-/-_Th17+Il17a	Gapdh	20.96	22.56	2.26				
	1	G2	IL17RA-/-_Th17+Il17a	Gapdh	24.16						
	1	G9	IL17RA-/-_Th17+Il17a	Il23a	28.57	28.93	0.51	6.36	1.64	-4.88	29.36
	1	G10	IL17RA-/-_Th17+Il17a	Il23a	29.29						
Treated KO 3	1	H1	IL17RA-/-_Th17+Il17a	Gapdh	19.86	19.60	0.37				
	1	H2	IL17RA-/-_Th17+Il17a	Gapdh	19.35						
	1	H9	IL17RA-/-_Th17+Il17a	Il23a	31.61	33.15	2.19	13.55	1.57	2.31	0.20
	1	H10	IL17RA-/-_Th17+Il17a	Il23a	34.70						

In this case, within the wild-type group, each treated ΔC_T value is relative to the same untreated ΔC_T value of 10.12 in calculating $\Delta\Delta C_T$ values.

Column:	**DelDel.CT**	**Delta.CT.Mean**
Untreated WT:	$0 = 10.12 - 10.12$	
Treated WT 1:	$-0.97 = 9.15 - 10.12$	
Treated WT 2:	$-0.76 = 9.36 - 10.12$	
Treated WT 3:	$-5.23 = 4.88 - 10.12$	

(4.16)

We run several simulation settings, altering the number of treated samples between 2 and 5 and the extremes of 9 and 90 as well as the standard deviations of the C_T values of 2 and 0.2. The error rates output from the simulation may be examined for patterns of statistical validity.

Table 4.3: Type I error rates for the normalization scenario where relative measures are calculated for multiple treated samples all relative to the same single untreated sample where $\mu = \{\mu_{UT,Gapdh}, \mu_{T,Gapdh}, \mu_{UT,IL23a}, \mu_{UT,IL23a}\}$ and $\sigma = \{\sigma_{UT,Gapdh}, \sigma_{T,Gapdh}, \sigma_{UT,IL23a}, \sigma_{UT,IL23a}\}$.

		Type I Error Rate			
		T-test on $\Delta\Delta C_T$	Wilcox test on $\Delta\Delta C_T$	T-test on $2^{-\Delta\Delta C_T}$	Wilcox test on $2^{-\Delta\Delta C_T}$
μ ={20, 20, 30, 30}, σ ={2, 2, 2, 2}	1 Untreated, 2 Treated	0.061	0	0.037	0
	1 Untreated, 3 Treated	0.186	0	0.038	0
	1 Untreated, 4 Treated	0.30	0.23	0.054	0.23
	1 Untreated, 5 Treated	0.36	0.31	0.083	0.31
	1 Untreated, 9 Treated	0.50	0.47	0.22	0.47
	1 Untreated, 90 Treated	0.84	0.83	0.74	0.83
μ ={20, 20, 30, 30}, σ ={0.2,0.2,0.2,0.2}	1 Untreated, 2 Treated	0.06	0	0.043	0
	1 Untreated, 3 Treated	0.19	0	0.063	0
	1 Untreated, 4 Treated	0.29	0.23	0.095	0.23
	1 Untreated, 5 Treated	0.36	0.31	0.14	0.31
	1 Untreated, 9 Treated	0.51	0.48	0.34	0.48
	1 Untreated, 90 Treated	0.83	0.83	0.79	0.83

As shown in Table 4.3, under this normalization scenario, a 10-fold decrease in the standard deviation of C_T values did little to alter the error rates observed. This is a surprising pattern; based on the results from the fold change simulation, we expected that the significantly smaller

variance would have very different error rates. Again, in contrast to the fold change setting, we see that an increase of sample size leads to a greater inflation of Type I error. This is likely seen since there is more correlation present that when there are more treated samples and this dependency is not accounted for in statistical testing. We also observe that for this particular scenario, the t-test on $2^{-\Delta\Delta C_T}$ is statistically valid and only slightly underpowered when no more than 3 treated samples are normalized to a single untreated sample for each group, as is a common sample size for these experiments.

4.4.3 Multiple Treated relative to Multiple Pairwise Untreated

This scenario is also known as the triplicate case, where single treated and untreated samples are placed on 3 different plates and the relative quantities are paired accordingly. Again, the only deviation in the derivation equations from section 4.4.1 is in the calculation of $\Delta\Delta C_T$ values. Table 4.4 shows an example of simulated data seen in this normalization scenario.

Table 4.4: Single *IL23a* target gene subset of simulated Applied Biosystems® instrument data when normalization is performed on multiple treated samples all relative to multiple paired untreated samples across 3 plates.

	Plate	Well	Sample.Name	Target.Name	CT	CT.Mean	CT.Sd	Delta.CT.Mean	Delta.CT.SE	DelDel.CT	RQ
Untreated WT1	1	B1	WT_Th17	Gapdh	17.06	18.20	1.61				
	1	B2	WT_Th17	Gapdh	19.33						
	1	B9	WT_Th17	Il23a	26.26	28.31	2.91	10.12	2.35	0	1
	1	B10	WT_Th17	Il23a	30.37						
Treated WT1	1	C1	WT_Th17+Il17a	Gapdh	23.32	21.83	2.12				
	1	C2	WT_Th17+Il17a	Gapdh	20.33						
	1	C9	WT_Th17+Il17a	Il23a	30.06	30.98	1.29	9.15	1.75	-0.97	1.95
	1	C10	WT_Th17+Il17a	Il23a	31.89						
Untreated WT2	2	B1	WT_Th17	Gapdh	20.93	19.31	2.29				
	2	B2	WT_Th17	Gapdh	17.69						
	2	B9	WT_Th17	Il23a	30.46	30.12	0.48	10.81	1.65	0	1
	2	B10	WT_Th17	Il23a	29.78						
Treated WT2	2	C1	WT_Th17+Il17a	Gapdh	22.35	21.28	1.51				
	2	C2	WT_Th17+Il17a	Gapdh	20.21						
	2	C9	WT_Th17+Il17a	Il23a	29.55	30.64	1.54	9.36	1.53	-1.46	2.74
	2	C10	WT_Th17+Il17a	Il23a	31.73						
Untreated WT3	3	B1	WT_Th17	Gapdh	20.87	19.22	2.33				
	3	B2	WT_Th17	Gapdh	17.57						
	3	B9	WT_Th17	Il23a	27.94	30.85	4.10	11.62	3.34	0	1
	3	B10	WT_Th17	Il23a	33.75						
Treated WT3	3	C1	WT_Th17+Il17a	Gapdh	20.17	19.90	0.38				
	3	C2	WT_Th17+Il17a	Gapdh	19.63						
	3	C9	WT_Th17+Il17a	Il23a	24.45	24.78	0.47	4.88	0.43	-6.74	106.83
	3	C10	WT_Th17+Il17a	Il23a	25.12						
Untreated KO1	1	F1	IL17RA-/-_Th17	Gapdh	18.10	18.86	1.07				
	1	F2	IL17RA-/-_Th17	Gapdh	19.62						
	1	F9	IL17RA-/-_Th17	Il23a	28.94	30.10	1.65	11.24	1.39	0	1
	1	F10	IL17RA-/-_Th17	Il23a	31.27						
Treated KO1	1	H1	IL17RA-/-_Th17+Il17a	Gapdh	19.01	17.91	1.56				
	1	H2	IL17RA-/-_Th17+Il17a	Gapdh	16.80						
	1	H9	IL17RA-/-_Th17+Il17a	Il23a	30.45	29.15	1.83	11.25	1.70	0.01	0.99
	1	H10	IL17RA-/-_Th17+Il17a	Il23a	27.86						
Untreated KO2	2	F1	IL17RA-/-_Th17	Gapdh	19.59	20.14	0.78				
	2	F2	IL17RA-/-_Th17	Gapdh	20.69						
	2	F9	IL17RA-/-_Th17	Il23a	29.90	27.99	2.69	7.86	1.98	0	1
	2	F10	IL17RA-/-_Th17	Il23a	26.09						
Treated KO2	2	H1	IL17RA-/-_Th17+Il17a	Gapdh	20.96	22.56	2.26				
	2	H2	IL17RA-/-_Th17+Il17a	Gapdh	24.16						
	2	H9	IL17RA-/-_Th17+Il17a	Il23a	28.57	28.93	0.51	6.36	1.64	-1.49	2.82
	2	H10	IL17RA-/-_Th17+Il17a	Il23a	29.29						
Untreated KO3	3	F1	IL17RA-/-_Th17	Gapdh	19.10	20.51	1.99				
	3	F2	IL17RA-/-_Th17	Gapdh	21.92						
	3	F9	IL17RA-/-_Th17	Il23a	33.00	32.62	0.54	12.11	1.46	0	1
	3	F10	IL17RA-/-_Th17	Il23a	32.24						
Treated KO3	3	H1	IL17RA-/-_Th17+Il17a	Gapdh	19.86	19.60	0.37				
	3	H2	IL17RA-/-_Th17+Il17a	Gapdh	19.35						
	3	H9	IL17RA-/-_Th17+Il17a	Il23a	31.61	33.15	2.19	13.55	1.57	1.44	0.37
	3	H10	IL17RA-/-_Th17+Il17a	Il23a	34.70						

In this case, within the wild-type group, each treated ΔC_T value is relative to its paired untreated ΔC_T value when calculating $\Delta\Delta C_T$.

Column:	**DelDel.CT**	**Delta.CT.Mean**	
Treated WT 1:	$-0.97 = 9.15 - 10.12$		
Treated WT 2:	$-1.46 = 9.36 - 10.81$		*(4.17)*
Treated WT 3:	$-6.74 = 4.88 - 11.62$		

We run several simulation settings, altering the number of treated and untreated samples from 3 to 9 as well as the standard deviations of the C_T values of 2 and 0.2. The error rates output from the simulation may be examined for patterns of statistical validity.

Table 4.5: Type I error rates for the normalization scenario where relative measures are calculated for multiple treated samples relative to paired untreated samples where $\mu = \{\mu_{UT,Gapdh}, \mu_{T,Gapdh}, \mu_{UT,IL23a}, \mu_{UT,IL23a}\}$ and $\sigma = \{\sigma_{UT,Gapdh}, \sigma_{T,Gapdh}, \sigma_{UT,IL23a}, \sigma_{UT,IL23a}\}$.

		Type I Error Rate			
		T-test on $\Delta\Delta C_T$	Wilcox test on $\Delta\Delta C_T$	T-test on $2^{-\Delta\Delta C_T}$	Wilcox test on $2^{-\Delta\Delta C_T}$
μ ={20, 20, 30, 30}, σ ={2, 2, 2, 2}	3 Untreated, 3 Treated	0.035	0	0.0085	0
	9 Untreated, 9 Treated	0.048	0.041	0.011	0.041
μ ={20, 20, 30, 30}, σ ={0.2,0.2,0.2,0.2}	3 Untreated, 3 Treated	0.034	0	0.014	0
	9 Untreated, 9 Treated	0.048	0.040	0.018	0.040

As shown in Table 4.5, under this normalization scenario, a 10-fold decrease in the standard deviation of C_T values did relatively little to change the observed error rates. Based on the results from the fold change simulation, this is another surprising pattern. We see that increasing sample size increases Type I error marginally, but not above the 0.05 level. In this case, performing a t-test on the $2^{-\Delta\Delta C_T}$ relative quantity yields a highly conservative Type I error due to the large spread of potential relative quantity values as displayed conceptually in Figure 4.1. Of particular interest in this scenario is the fact that performing a t-test on the $\Delta\Delta C_T$ quantities instead yields a less conservative test. This observation is likely due to the fact that the distribution of $\Delta\Delta C_T$ quantities is theoretically normal, and thus not in violation of the t-test's assumption of normality.

4.4.4 Multiple Treated relative to Single Averaged Untreated

This scenario is an expansion of the triplicate case in section 4.4.3, where single treated and untreated samples are placed on 3 different plates, but the relative quantities are calculated relative to the average of the 3 untreated samples. Again, the only deviation in the derivation equations from section 4.4.1 is in the calculation of $\Delta\Delta C_T$ values. Table 4.6 shows an example of simulated data seen in this normalization scenario.

Table 4.6: Single *IL23a* target gene subset of simulated Applied Biosystems® instrument data when normalization is performed on multiple treated samples all relative to a single averaged untreated sample across 3 plates.

	Plate	Well	Sample.Name	Target.Name	CT	CT.Mean	CT.Sd	Delta.CT.Mean	Delta.CT.SE	Avg.Delta.CT.Mean	DelDel.CT	RQ
	1	B1	WT_Th17	Gapdh	17.06	18.20	1.61					
Untreated WT1	1	B2	WT_Th17	Gapdh	19.33							
	1	B9	WT_Th17	Il23a	26.26	28.31	2.91	10.12	2.35	10.85	0	1
	1	B10	WT_Th17	Il23a	30.37							
	2	B1	WT_Th17	Gapdh	20.93	19.31	2.29					
Untreated WT2	2	B2	WT_Th17	Gapdh	17.69							
	2	B9	WT_Th17	Il23a	30.46	30.12	0.48	10.81	1.65	10.85	0	1
	2	B10	WT_Th17	Il23a	29.78							
	3	B1	WT_Th17	Gapdh	20.87	19.22	2.33					
Untreated WT3	3	B2	WT_Th17	Gapdh	17.57							
	3	B9	WT_Th17	Il23a	27.94	30.85	4.10	11.62	3.34	10.85	0	1
	3	B10	WT_Th17	Il23a	33.75							
	1	C1	WT_Th17+Il17a	Gapdh	23.32	21.83	2.12					
Treated WT1	1	C2	WT_Th17+Il17a	Gapdh	20.33							
	1	C9	WT_Th17+Il17a	Il23a	30.06	30.98	1.29	9.15	1.75		-1.70	3.25
	1	C10	WT_Th17+Il17a	Il23a	31.89							
	2	C1	WT_Th17+Il17a	Gapdh	22.35	21.28	1.51					
Treated WT2	2	C2	WT_Th17+Il17a	Gapdh	20.21							
	2	C9	WT_Th17+Il17a	Il23a	29.55	30.64	1.54	9.36	1.53		-1.49	2.82
	2	C10	WT_Th17+Il17a	Il23a	31.73							
	3	C1	WT_Th17+Il17a	Gapdh	20.17	19.90	0.38					
Treated WT3	3	C2	WT_Th17+Il17a	Gapdh	19.63							
	3	C9	WT_Th17+Il17a	Il23a	24.45	24.78	0.47	4.88	0.43		-5.97	62.51
	3	C10	WT_Th17+Il17a	Il23a	25.12							
	1	F1	IL17RA-/-_Th17	Gapdh	18.10	18.86	1.07					
Untreated KO1	1	F2	IL17RA-/-_Th17	Gapdh	19.62							
	1	F9	IL17RA-/-_Th17	Il23a	28.94	30.10	1.65	11.24	1.39	10.40	0	1
	1	F10	IL17RA-/-_Th17	Il23a	31.27							
	2	F1	IL17RA-/-_Th17	Gapdh	19.59	20.14	0.78					
Untreated KO2	2	F2	IL17RA-/-_Th17	Gapdh	20.69							
	2	F9	IL17RA-/-_Th17	Il23a	29.90	27.99	2.69	7.86	1.98	10.40	0	1
	2	F10	IL17RA-/-_Th17	Il23a	26.09							
	3	F1	IL17RA-/-_Th17	Gapdh	19.10	20.51	1.99					
Untreated KO3	3	F2	IL17RA-/-_Th17	Gapdh	21.92							
	3	F9	IL17RA-/-_Th17	Il23a	33.00	32.62	0.54	12.11	1.46	10.40	0	1
	3	F10	IL17RA-/-_Th17	Il23a	32.24							
	1	H1	IL17RA-/-_Th17+Il17a	Gapdh	19.01	17.91	1.56					
Treated KO1	1	H2	IL17RA-/-_Th17+Il17a	Gapdh	16.80							
	1	H9	IL17RA-/-_Th17+Il17a	Il23a	30.45	29.15	1.83	11.25	1.70		0.85	0.56
	1	H10	IL17RA-/-_Th17+Il17a	Il23a	27.86							
	2	H1	IL17RA-/-_Th17+Il17a	Gapdh	20.96	22.56	2.26					
Treated KO2	2	H2	IL17RA-/-_Th17+Il17a	Gapdh	24.16							
	2	H9	IL17RA-/-_Th17+Il17a	Il23a	28.57	28.93	0.51	6.36	1.64		-4.04	16.45
	2	H10	IL17RA-/-_Th17+Il17a	Il23a	29.29							
	3	H1	IL17RA-/-_Th17+Il17a	Gapdh	19.86	19.60	0.37					
Treated KO3	3	H2	IL17RA-/-_Th17+Il17a	Gapdh	19.35							
	3	H9	IL17RA-/-_Th17+Il17a	Il23a	31.61	33.15	2.19	13.55	1.57		3.15	0.11
	3	H10	IL17RA-/-_Th17+Il17a	Il23a	34.70							

In this case, within the wild-type group, each treated ΔC_T value is relative to a single average of untreated ΔC_T values when calculating $\Delta\Delta C_T$.

Column:	**DelDel.CT**	**Delta.CT.Mean**	
Treated WT 1:	-1.70	$= 9.15 - 10.85$	
Treated WT 2:	-1.49	$= 9.36 - 10.85$	*(4.18)*
Treated WT 3:	-5.97	$= 4.88 - 10.85$	

We run several simulation settings, altering the number of untreated samples to be averaged and treated samples from 3 to 9 as well as the standard deviations of the C_T values of 2 and 0.2. The error rates output from the simulation may be examined for patterns of statistical validity.

Table 4.7: Type I error rates for the normalization scenario where relative measures are calculated for multiple treated samples relative to a single untreated sample which was derived via averaging where $\mu = \{\mu_{UT,Gapdh}, \mu_{T,Gapdh}, \mu_{UT,IL23a}, \mu_{UT,IL23a}\}$ and $\sigma = \{\sigma_{UT,Gapdh}, \sigma_{T,Gapdh}, \sigma_{UT,IL23a}, \sigma_{UT,IL23a}\}$.

		Type I Error Rate			
		T-test on $\Delta\Delta C_T$	Wilcox test on $\Delta\Delta C_T$	T-test on $2^{-\Delta\Delta C_T}$	Wilcox test on $2^{-\Delta\Delta C_T}$
μ ={20, 20, 30, 30}, σ ={2, 2, 2, 2}	3 Untreated averaged, 3 Treated	0.086	0	0.026	0
	9 Untreated averaged, 9 Treated	0.12	0.098	0.040	0.098
μ ={20, 20, 30, 30}, σ ={0.2,0.2,0.2,0.2}	3 Untreated averaged, 3 Treated	0.085	0	0.037	0
	9 Untreated averaged, 9 Treated	0.077	0.065	0.043	0.065

As shown in Table 4.7, under this normalization scenario, decreasing the standard deviation of C_T values by 10-fold did help deflate the Type I error rate when 9 untreated samples were averaged, and we saw little change when 3 untreated samples were averaged. We see that in this case, an increased sample size only marginally increases Type I error regardless of variability parameters. Again, in this case, performing a t-test on the $2^{-\Delta\Delta C_T}$ relative quantity yields an error rate below the 0.05 level. Although, even the error rates from the non-parametric tests are inflated. We were surprised in this scenario not to see more parallels to the standard fold change

setting. With the common averaged untreated sample used in the $2^{-\Delta\Delta C_T}$ in these cases, the setup is similar to a fold change ratio with common averaged control. However, the inflation we see in the RT-PCR case is not as great.

4.5 RECOMMENDATIONS

When using RT-PCR data, the recommendations provided for researchers rely heavily on the method their lab uses to set up plates for normalization. We summarize the methods studied and our corresponding statistical recommendations in Table 4.8 below.

Table 4.8: Summary table of relative quantification cases and corresponding statistical recommendations based on simulation study.

Case		Plate Setup	Type I Error Patterns	Statistical Recommendation
A	1 Untreated, Multiple Treated	Single	Analysis on $\Delta\Delta C_T$ values: • Type I error rate is above the above 0.05 and inflates as number of treated samples grows regardless of standard deviation for both parametric and non-parametric tests Analysis on $2^{-\Delta\Delta C_T}$ values: • Type I error rate inflates as number of treated samples grows regardless of standard deviation for both parametric and non-parametric tests, although the error rate is below 0.05 when there are fewer than 4 infected samples	Run a Student's test on $2^{-\Delta\Delta C_T}$ values as long as there are exactly 3 treated samples.
B	Multiple Untreated, Multiple Treated (Pairwise)	Multiple	Analysis on $\Delta\Delta C_T$ values: • Type I error rate is slightly conservative and increases toward 0.05 with higher numbers of samples regardless of standard deviation for both parametric and non-parametric tests Analysis on $2^{-\Delta\Delta C_T}$ values: • Type I error rate is highly conservative regardless of standard deviation for parametric tests and slightly less conservative for non-parametric tests	Run a statistical test (either parametric or non-parametric) on $\Delta\Delta C_T$ values, and the more replicates of each sample, the better. This alternative method makes no assumptions about the efficiencies of the target and housekeeping genes.

C	1 Untreated (via averaging), Multiple Treated	Multiple	Analysis on $\Delta\Delta C_T$ values: • Type I error rate is inflated above 0.05 with higher numbers of samples regardless of standard deviation for both parametric and non-parametric tests Analysis on $2^{-\Delta\Delta C_T}$ values: • Type I error rate is slightly conservative for parametric tests and slightly inflated for non-parametric tests regardless of standard deviation	Run a statistical test (either parametric or non-parametric) on $2^{-\Delta\Delta C_T}$ values, and the more replicates of each sample, the better.

In Case A where the relative quantities were calculated using multiple treated relative to a single untreated sample, we observed that the t-test maintains an uninflated Type I error rate in the case where there are exactly 3 replicates of treated samples. In Case B where triplicate untreated controls exist, we make a recommendation to change analytic methods and perform statistical testing on the $\Delta\Delta C_T$ quantity instead. The value of this quantity may not be as interpretable as the historically used RQ, but it results in a more powerful test. Performing statistical tests on the $2^{-\Delta\Delta C_T}$ RQ yields highly conservative Type I error rate which usually decreases the power of the test. The patterns seen in Case C where multiple treated relative to single averaged untreated were less clear cut. Although, based on conversations with labs using RT-PCR data, it does not appear that this normalization method is used at the present time.

CHAPTER 5

DISCUSSION AND CONCLUSION

Throughout this work, we detailed various normalization scenarios seen in fold change and RT-PCR relative quantity data and ultimately made recommendations to researchers based on the patterns observed. In order to thoroughly examine each of these data types, we created functions to simulate fold change and RT-PCR relative quantity derivations. Statistical testing, Type I error rate evaluation, and graphical outputs were also observed. Based on the results from simulations run including the changes of many parameters, we were able to identify patterns in error rate inflation. Specifically, we highlighted situations in which the parametric t-test can be appropriately used, and also shed light on alternative analytic methods that answer the same research question while preserving statistical validity. We made progress in ruling out single parameter relationships as sole issues leading to error rate inflation in both fold change and RT-PCR data, although weren't able to provide a concise, full picture of which combinations of factors are to blame. In creating recommendations for researchers, we are sensitive to the fact that experiment parameters are far easier to control in a computer simulation setting than in a science lab. We also recognize the preference of researchers to utilize relative quantities for comparison purposes and the need to keep the sample size of experiments low. Hopefully, the alternative methods we suggest are seen as feasible.

This study opens the door to future work in the field of normalization analysis for basic science research. It would be useful in the future to examine other measures of statistical validity outside of Type I error alone. Additionally, more emphasis could be placed on the standard error calculations and confidence intervals surrounding various relative quantity measures in future work. We began the process of examining patterns seen under a variety of simulation settings in the hopes of identifying the causes of t-test Type I error inflation when using normalized quantities, but there are more scenarios that could be tried and more input from basic science researchers that could be gained.

REFERENCES

[1] National Center for Biotechnology Information. (n.d.). *Search results.* Retrieved from PubMed.gov: ncbi.nlm.nih.gov/pubmed

[2] Koyama, T., & Zhao, A. (2009). Failed "Normalization": Dividing by the Control Average. Nashville, TN, USA: Cancer Biostatistics Center, Vanderbilt-Ingram Cancer Center.

[3] Schagat, T., Paguio, A., & Kopish, K. (2007). Normalizing Genetic Reporter Assays: Approaches and Considerations for Increasing Consistency and Statistical Significance. *Cell Notes*(17), 9-12.

[4] Parang, B., Kaz, A. M., Barrett, C. W., Short, S. P., Ning, W., Keating, C. E., . . . Williams, C. S. (2017, May). BVES regulates c-Myc stability via PP2A and suppresses colitis-induced tumorigenesis. *Gut, 66*(5), 852-862.

[5] Qi, H. H., Sarkissian, M., Hu, G. Q., Wang, Z., Bhattacharjee, A., Gordon, D. B., . . . Shi, Y. (2010, July 22). Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. *Nature, 466*(7305), 503-507.

[6] Gao, J., Wang, W., Mao, Y., Gräff, J., Guan, J., Pan, L., … Tsai, L. (2010, July 11). A novel pathway regulates memory and plasticity via SIRT1 and miR-134. *Nature, 466*(7310), 1105-1109

[7] ThermoFisher Scientific. (n.d.). *Basic Principles of RT-qPCR.* Retrieved from https://www.thermofisher.com/us/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/spotlight-articles/basic-principles-rt-qpcr.html

[8] Kozera, B., & Rapacz, M. (2013). Reference genes in real-time PCR. *Journal of Applied Genetics, 54*(4), 391-406.

[9] Barber, R. D., Harmer, D. W., Coleman, R. A., & Clark, B. J. (2005, May 11). GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. *Physiological Genomics, 21*(3), 389-95.

[10] Yasmin, S., Dixon, B. R., Olivares-Villagómez, D., & Algood, H. M. (2019, October 18). Interleukin-21 (IL-21) Downregulates Dendritic Cell Cytokine Responses to Helicobacter pylori and Modulates T Lymphocyte IL-17A Expression in Peyer's Patches during Infection. *Infection and Immunity, 87*(11).

[11] Dale, B. L., Pandey, A. K., Chen, Y., Smart, C. D., Laroumanie, F., Ao, M., . . . Madhur, M. S. (2019, June 6). Critical role of IL-21 and T follicular helper cells in hypertension and vascular dysfunction. *JCI Insight, 4*(11).

[12] Seltman, H. (n.d.). *Approximations for Mean and Variance of a Ratio.* Retrieved from Carnegie Mellon University: http://www.stat.cmu.edu/~hseltman/files/ratio.pdf

[13] Casella, G., & Berger, R. L. (2002). *Statistical Inference.* Pacific Grove, CA, USA: Duxbury.

[14] Applied Biosystems. (2008). *Guide to Performing Relative Quantitation of Gene Expression Using Real-Time Quantitative PCR.*

[15] Applied Biosystems. (2001). *User Bulletin #2 Subject: Relative Quantitation of Gene Expression.*

[16] Ospina, L. (2013, October 31). *Wilcoxon-Mann-Whitney test and a small sample size*. Retrieved from Oxford Protein Informatics Group: https://www.blopig.com/blog/2013/10/wilcoxon-mann-whitney-test-and-a-small-sample-size/

[17] Ayers, D., Kis, C., & Splittgerber, R. (n.d.). Efficiency Rules in the Analysis of Quantitative Real Time PCR. Nashville, TN, USA: Division of Cancer Biostatistics.

APPENDIX

A.1 FOLD CHANGE FUNCTION R CODE

```
#Include relevant libraries
library(kableExtra)
library(ggplot2)
library(ggpubr)

#Define Simulation Inputs
#sims: Number of simulation runs
#dist={‘Normal’, ‘Lognormal’, ‘Exponential’}: representing data-generating distribution
#n={nc1,nc2,nt1,nt2}: representing number of control and treatment groups, respectively
#m={mc1,mc2,mt1,mt2}: representing means of control and treatment groups, respectively
#s={sc1,sc2,st1,st2}: representing standard deviations of control and treatment groups,
                      respectively
#r={rc1,rc2,rt1,rt2}: representing exponential parameters for control and treatment groups,
                      respectively

#Run simulation
mousesim <- function(sims, dist, n, m, s, r){
  ## Wild-type is group 1
  ## Knockout is group 2

  #Calculate Variance of Ratio based on parameters for both
  #wild-type and knockout groups
  m <- ifelse(rep(dist=='Exponential', length(n)), 1/r, m)
  s <- ifelse(rep(dist=='Exponential', length(n)), sqrt(1/r^2), s)
  mu.WT1.TRT <- m[3]
  mu.WT1.Cont <- m[1]
  var.WT1.TRT <- s[3]^2/n[3]
  var.WT1.Cont <- s[1]^2/n[1]
  var.ratio.WT1 <- (mu.WT1.TRT/mu.WT1.Cont)^2*(var.WT1.TRT/mu.WT1.TRT^2 +
                                                 var.WT1.Cont/mu.WT1.Cont^2)

  mu.KO2.TRT <- m[4]
  mu.KO2.Cont <- m[2]
  var.KO2.TRT <- s[4]^2/n[4]
  var.KO2.Cont <- s[2]^2/n[2]
  var.ratio.KO2 <- (mu.KO2.TRT/mu.KO2.Cont)^2*(var.KO2.TRT/mu.KO2.TRT^2 +
                                                 var.KO2.Cont/mu.KO2.Cont^2)

  #Initialize output variables
  simdat <- matrix(NA, nrow=sims, ncol=25)
  colnames(simdat) <- c('seed', 'denom1', 'denom2', 'denomc', 'numer1', 'numer2',
                        'fc1', 'fc2', 'fcc1', 'fcc2', 'log2fc1', 'log2fc2',
                        'log2fcdiv1', 'log2fcdiv2', 'sdfc1', 'sdfc2', 'p.ttest',
                        'p.ztest', 'p.ttest.c', 'p.ttest.log2', 'p.ttest.log2div',
                        'p.wilcoxtest', 'p.lmUnweighted', 'p.lmWeightedSigma',
                        'p.lmWeightedSd')
  simdat[,'seed'] <- sample(1:100000, sims)

  #Run simulation
  for(i in seq_len(sims)){
    set.seed(simdat[i,'seed'])

    #Create dataframe with raw and derived quantities
    y <- if (dist=='Normal') {rnorm(sum(n), mean=rep(m,n), sd=rep(s,n))
    } else if (dist=='Lognormal') {rlnorm(sum(n), meanlog=rep(m,n), sdlog=rep(s,n))
    } else {rexp(sum(n), rate=rep(r,n))
    }
```

```
gen <- rep(c('WT1','KO2','WT1','KO2'), n)
trt <- rep(c('Cont','TRT'), c(n[1]+n[2],n[3]+n[4]))
gentrt <- paste(gen,trt, sep='.')
dat <- data.frame(gen=factor(gen, levels=unique(gen)),
                  trt=factor(trt, levels=unique(trt)),
                  gentrt=factor(gentrt, levels=unique(gentrt)),
                  n=rep(n,n),
                  distribution=rep(dist,sum(n)),
                  mu=ifelse(rep(dist=='Normal'|dist=='Lognormal',sum(n)),
                            rep(m,n), rep(1/r,n)),
                  sigma=ifelse(rep(dist=='Normal'|dist=='Lognormal', sum(n)),
                               rep(s,n), rep(1/r^2,n)),
                  lambda=ifelse(rep(dist=='Exponential', sum(n)), rep(r,n),
                                NA), y=y)
dat$xbar <- rep(aggregate(dat$y, list(dat$gentrt), mean)$x, n)
dat$sd <- rep(aggregate(dat$y, list(dat$gentrt), sd)$x, n)
denominators <- c(dat$xbar[dat$gentrt=='WT1.Cont'][1],
                  dat$xbar[dat$gentrt=='KO2.Cont'][1])
dat$denom <- rep(rep(denominators,2), n)
dat$fc <- dat$y/dat$denom
dat$fcnoavg <- dat$y
dat$log2fc <- log2(dat$fc)
dat$log2fcdiv <- log(dat$y)/log(dat$denom)

dat$denomc <- mean(dat$y[dat$trt=='Cont'])
dat$fcc <- dat$y/dat$denomc

#Simulation Outputs
simdat[i,'denom1'] <- denominators[1]
simdat[i,'denom2'] <- denominators[2]
simdat[i,'denomc'] <- mean(dat$y[dat$trt=='Cont'])

simdat[i,'numer1'] <- dat$xbar[dat$gentrt=='WT1.TRT'][1]
simdat[i,'numer2'] <- dat$xbar[dat$gentrt=='KO2.TRT'][1]

simdat[i,'fc1'] <- mean(dat$fc[dat$gentrt=='WT1.TRT'])
simdat[i,'fc2'] <- mean(dat$fc[dat$gentrt=='KO2.TRT'])

simdat[i,'log2fc1'] <- mean(dat$log2fc[dat$gentrt=='WT1.TRT'])
simdat[i,'log2fc2'] <- mean(dat$log2fc[dat$gentrt=='KO2.TRT'])

simdat[i,'log2fcdiv1'] <- mean(dat$log2fcdiv[dat$gentrt=='WT1.TRT'])
simdat[i,'log2fcdiv2'] <- mean(dat$log2fcdiv[dat$gentrt=='KO2.TRT'])

simdat[i,'sdfc1'] <- sd(dat$fc[dat$gentrt=='WT1.TRT'])
simdat[i,'sdfc2'] <- sd(dat$fc[dat$gentrt=='KO2.TRT'])

simdat[i,'fcc1'] <- mean(dat$fcc[dat$gentrt=='WT1.TRT'])
simdat[i,'fcc2'] <- mean(dat$fcc[dat$gentrt=='KO2.TRT'])

#Perform statistical tests between wild-type and knockout groups
#and capture p-values
t.testFC <- t.test(dat$fc[dat$gentrt=='WT1.TRT'],
                   dat$fc[dat$gentrt=='KO2.TRT'])
simdat[i,'p.ttest'] <- t.testFC$p.value

z.testFC <- (mean(dat$fc[dat$gentrt=='WT1.TRT']) -
               mean(dat$fc[dat$gentrt=='KO2.TRT']))/sqrt(var.ratio.WT1/n[3] +
                                                         var.ratio.KO2/n[4])
simdat[i,'p.ztest'] <- 2*pnorm(-abs(z.testFC))
```

```
  t.testFCc <- t.test(dat$fcc[dat$gentrt=='WT1.TRT'],
                      dat$fcc[dat$gentrt=='KO2.TRT'])
  simdat[i,'p.ttest.c'] <- t.testFCc$p.value

  t.testlog2FC <- t.test(dat$log2fc[dat$gentrt=='WT1.TRT'],
                         dat$log2fc[dat$gentrt=='KO2.TRT'])
  simdat[i,'p.ttest.log2'] <- t.testlog2FC$p.value

  t.testlog2FCdiv <- t.test(dat$log2fcdiv[dat$gentrt=='WT1.TRT'],
                            dat$log2fcdiv[dat$gentrt=='KO2.TRT'])
  simdat[i,'p.ttest.log2div'] <- t.testlog2FCdiv$p.value

  t.testFC.wilcoxtest <- wilcox.test(dat$fc[dat$gentrt=='WT1.TRT'],
                                     dat$fc[dat$gentrt=='KO2.TRT'])
  simdat[i,'p.wilcoxtest'] <- t.testFC.wilcoxtest$p.value

  lmUnweighted <- lm(y ~ gen*trt, data=dat)
  simdat[i,'p.lmUnweighted'] <- summary(lmUnweighted)$coefficients[4,4]

  lmWeightedSigma <- lm(y ~ gen*trt, weights=1/sigma^2, data=dat)
  simdat[i,'p.lmWeightedSigma'] <- summary(lmWeightedSigma)$coefficients[4,4]

  lmWeightedSd <- lm(y ~ gen*trt, weights=1/sd^2, data=dat)
  simdat[i,'p.lmWeightedSd'] <- summary(lmWeightedSd)$coefficients[4,4]
}
simdat <- data.frame(simdat)

#Calculate Type I Errors
typei <- data.frame(sum(simdat$p.ttest<0.05)/sims,
                    sum(simdat$p.ztest<0.05)/sims,
                    sum(simdat$p.ttest.c<0.05)/sims,
                    sum(simdat$p.ttest.log2<0.05)/sims,
                    #sum(simdat$p.ttest.log2div<0.05)/sims,
                    sum(simdat$p.wilcoxtest<0.05)/sims,
                    sum(simdat$p.lmUnweighted<0.05)/sims,
                    sum(simdat$p.lmWeightedSigma<0.05)/sims,
                    sum(simdat$p.lmWeightedSd<0.05)/sims)
names(typei) <- c('typei.ttest',
                  'typei.ztest',
                  'typei.ttest.c',
                  'typei.ttest.log2',
                  #'typei.ttest.log2div',
                  'typei.wilcoxtest',
                  'typei.lmUnweighted',
                  'typei.lmWeightedSigma',
                  'typei.lmWeightedSd')

#Observed Variance of Ratio
var.obs.WT1 <- sd(simdat$fc1)^2
var.obs.KO2 <- sd(simdat$fc2)^2
varFC <- data.frame(var.ratio.WT1, var.obs.WT1, var.ratio.KO2, var.obs.KO2)

#Produce descriptive plots
plot.fc1 <- ggplot(simdat, aes(x=fc1)) + geom_density() +
  labs(x="Average Wild Type Fold Change", y="Density")
plot.fc2 <- ggplot(simdat, aes(x=fc2)) + geom_density() +
  labs(x="Average Knock Out Fold Change", y="Density")
plot.fc <- ggplot(simdat, aes(x=fc1, y=fc2, color=p.ttest<0.05)) +
  geom_point(shape=1) +
  xlab("Avg WT FC") + ylab("Avg KO FC") + theme(legend.position="none")
```

```
  plot.denom <- ggplot(simdat, aes(x=denom1, y=denom2, col=p.ttest<0.05)) +
    geom_point(shape=1) +
    xlab("Avg WT FC Denom") + ylab("Avg KO FC Denom")
  plot.numer <- ggplot(simdat, aes(x=numer1, y=numer2, col=p.ttest<0.05)) +
    geom_point(shape=1) + geom_abline() +
    xlab("Avg WT FC Num") + ylab("Avg KO FC Num")
  plots <- ggarrange(ggarrange(plot.fc1, plot.fc2, ncol = 2),
                     plot.fc,
                     ggarrange(plot.denom, plot.numer, ncol = 2,
                               common.legend = TRUE, legend="bottom"), nrow = 3)

  #Output simulation-specific outputs
  list(simdat, typei, plots, varFC)
}

#Example of implementation (Normal distribution)
sim.norm.1a <- mousesim(sims=10000, dist='Normal', n=c(3,3,9,9), m=c(200,200,400,400),
s=c(10,10,20,20))
simdat.norm.1a <- sim.norm.1a[[1]]
sim.norm.1a[[2]] %>% kable() %>% kable_styling()
sim.norm.1a[[4]] %>% kable() %>% kable_styling()
sim.norm.1a[[3]]
```

A.2 RT-PCR FUNCTION R CODE

```
#Include relevant libraries
library(knitr)
library(kableExtra)
library(ggplot2)
library(ggpubr)

#Define Simulation Inputs
#sims: Number of simulation runs
#case: 2 if Single Treated relative to Multiple Untreated
       3 if Multiple Treated relative to Multiple Pairwise Untreated
       4 if Multiple Treated relative to Single Averaged Untreated
#n={n.ui, n.i}: representing number of untreated and treated samples, respectively
#m={m.gapdh.ui,m.gapdh.i,m.il23.ui,m.il23.i}: representing means of housekeeping Gapdh genes
                                              and target IL23a genes, respectively
#s={s.gapdh.ui,s.gapdh.i,s.il23.ui,s.il23.i}: representing standard deviations of housekeeping
                                              Gapdh genes and target IL23a genes, respectively
#Simulation of error rates from Delta-Delta normalization of RT-PCR data
rtpcrsim <- function(sims, case, n, m, s){
  #Initialize output variables
  var.ui <- NULL
  var.deldel.wt <- NULL
  var.deldel.ko <- NULL
  var.rq.wt <- NULL
  var.rq.ko <- NULL
  pval.deldel <- NULL
  pval.w.deldel <- NULL
  pval.rq <- NULL
  pval.w.rq <- NULL

  #Assign number of treated and untreated samples based on function input
  nsamp.ui=n[1]
  nsamp.i=n[2]

  #Run simulation
  for(i in 1:sims){
    #Create dataframe of raw data based on function inputs
    dat <- data.frame("sample" = rep(1:nsamp.i, times=2),
                      "trt" = rep(c('WT', 'KO'), each=nsamp.i),
                      "ui1.gapdh" = ifelse(rep(case==2, each=nsamp.i*2),
                                           rep(rnorm(nsamp.ui*2, m[1], s[1]),
                                               each=nsamp.i),
                                           rep(rnorm(nsamp.i*2, m[1], s[1]))),
                      "ui2.gapdh" = ifelse(rep(case==2, each=nsamp.i*2),
                                           rep(rnorm(nsamp.ui*2, m[1], s[1]),
                                               each=nsamp.i),
                                           rep(rnorm(nsamp.i*2, m[1], s[1]))),
                      "i1.gapdh" = rnorm(nsamp.i*2, m[2], s[2]),
                      "i2.gapdh" = rnorm(nsamp.i*2, m[2], s[2]),
                      "ui1.il23" = ifelse(rep(case==2, each=nsamp.i*2),
                                          rep(rnorm(nsamp.ui*2, m[3], s[3]),
                                              each=nsamp.i),
                                          rep(rnorm(nsamp.ui*2, m[3], s[3]))),
                      "ui2.il23" = ifelse(rep(case==2, each=nsamp.i*2),
                                          rep(rnorm(nsamp.ui*2, m[3], s[3]),
                                              each=nsamp.i),
                                          rep(rnorm(nsamp.ui*2, m[3], s[3]))),
                      "i1.il23" = rnorm(nsamp.i*2, m[4], s[4]),
                      "i2.il23" = rnorm(nsamp.i*2, m[4], s[4]))
```

```
  #Perform derivations to yield relative quantities
  dat$ui.ct.mean.gapdh.raw <- rowMeans(dat[,c('ui1.gapdh', 'ui2.gapdh')])
  dat$ui.ct.mean.gapdh <- ifelse(rep(case==4, each=nsamp.i*2),
                                 rep(c(mean(dat[which(dat$trt=='WT'),
                                                'ui.ct.mean.gapdh.raw']),
                                       mean(dat[which(dat$trt=='KO'),
                                                'ui.ct.mean.gapdh.raw'])),
                                     each=nsamp.i), dat$ui.ct.mean.gapdh.raw)
  dat$i.ct.mean.gapdh <- rowMeans(dat[,c('i1.gapdh', 'i2.gapdh')])
  dat$ui.ct.mean.il23.raw <- rowMeans(dat[,c('ui1.il23', 'ui2.il23')])
  dat$ui.ct.mean.il23 <- ifelse(rep(case==4, each=nsamp.i*2),
                                rep(c(mean(dat[which(dat$trt=='WT'),
                                               'ui.ct.mean.il23.raw']),
                                      mean(dat[which(dat$trt=='KO'),
                                               'ui.ct.mean.il23.raw'])),
                                    each=nsamp.i), rep(dat$ui.ct.mean.il23.raw))
  dat$i.ct.mean.il23 <- rowMeans(dat[,c('i1.il23', 'i2.il23')])
  dat <- transform(dat, ct.sd.gapdh=apply(dat[,c('i1.gapdh', 'i2.gapdh')], 1, sd))
  dat <- transform(dat, ct.sd.il23=apply(dat[,c('i1.il23', 'i2.il23')], 1, sd))
  dat$delta.ct.mean <- dat$i.ct.mean.il23 - dat$i.ct.mean.gapdh
  dat$delta.ui <- dat$ui.ct.mean.il23 - dat$ui.ct.mean.gapdh
  dat$delta.ct.se <- sqrt(dat$ct.sd.il23^2/2 + dat$ct.sd.gapdh^2/2)
  dat$delta.delta.ct <- dat$delta.ct.mean - dat$delta.ui
  dat$rq <- 2^-dat$delta.delta.ct

  #Capture variance of Untreated Deltas that are subtracted from each measure
  var.ui[i] <- var(unique(c(dat[which(dat$trt=='WT'),'delta.ui'],
                            dat[which(dat$trt=='KO'),'delta.ui'])))

  var.deldel.wt[i] <- var(unique(dat[which(dat$trt=='WT'),'delta.delta.ct']))
  var.deldel.ko[i] <- var(unique(dat[which(dat$trt=='KO'),'delta.delta.ct']))
  var.rq.wt[i] <- var(unique(dat[which(dat$trt=='WT'),'rq']))
  var.rq.ko[i] <- var(unique(dat[which(dat$trt=='KO'),'rq']))

  #Test w/ Delta Delta
  ttest.deldel <- t.test(dat[which(dat$trt=='WT'),'delta.delta.ct'],
                         dat[which(dat$trt=='KO'),'delta.delta.ct'])
  wiltest.deldel <- wilcox.test(dat[which(dat$trt=='WT'),'delta.delta.ct'],
                                dat[which(dat$trt=='KO'),'delta.delta.ct'])
  pval.deldel[i] <- ttest.deldel$p.value
  pval.w.deldel[i] <- wiltest.deldel$p.value

  #Test w/ RQ
  ttest.rq <- t.test(dat[which(dat$trt=='WT'),'rq'],
                     dat[which(dat$trt=='KO'),'rq'])
  wiltest.rq <- wilcox.test(dat[which(dat$trt=='WT'),'rq'],
                            dat[which(dat$trt=='KO'),'rq'])
  pval.rq[i] <- ttest.rq$p.value
  pval.w.rq[i] <- wiltest.rq$p.value
}

#Output table of error rates for statistical tests on various relative quantities
tab <- kable(dat) %>%
  kable_styling() %>%
  scroll_box(width = "100%")
tie <- data.frame("Delta-Delta t-test" = sum(pval.deldel<0.05)/sims,
                  "Delta-Delta Non-parametric" = sum(pval.w.deldel<0.05)/sims,
                  "RQ t-test" = sum(pval.rq<0.05)/sims,
                  "RQ Non-parametric" = sum(pval.w.rq<0.05)/sims)
tab.tie <- kable(tie) %>%
  kable_styling()
```

```
  list(tab.tie, tab, var.ui, pval.deldel, pval.w.deldel, pval.rq, pval.w.rq,
       var.deldel.wt, var.deldel.ko, var.rq.wt, var.rq.ko)
}

#Example of implementation (Case 2, 3 Treated relative to 1 Untreated:
c2.n2 <- rtpcrsim(sims=10000, case=2, n=c(1,2), m=c(20,20,30,30), s=c(2,2,2,2))
c2.n2[[1]]
c2.n2[[2]]
par(mfrow=c(1,2))
plot(c2.n2[[3]], c2.n2[[4]], xlab="Var[delta.ui]", ylab="p-value from Delta-Delta t-test")
plot(c2.n2[[3]], c2.n2[[6]], xlab="Var[delta.ui]", ylab="p-value from RQ t-test")
```