

**Chebyshev's Inequality for Non-parametric Testing with Small  $N$  and  $\alpha$  in  
Microarray Research**

T. Mark Beasley<sup>1</sup>, Grier P. Page<sup>1</sup>, Jaap P.L. Brand<sup>1</sup>, Gary L. Gadbury<sup>4</sup>,  
John D. Mountz<sup>3</sup>, David B. Allison<sup>1,2</sup>

<sup>1</sup>Department of Biostatistics, Section on Statistical Genetics and

<sup>2</sup>Clinical Nutrition Research Center

<sup>3</sup>Department of Medicine, Division of Clinical Immunology and Rheumatology

University of Alabama at Birmingham

Birmingham, Alabama

<sup>4</sup>Department of Mathematics and Statistics

University of Missouri at Rolla

Rolla, Missouri

This research was supported in part by NIH grants R01DK56366, P30DK56336, P01AG11915, R01AG018922, P20CA093753, R01AG011653, U24DK058776 and R01ES09912; NSF grant 0090286; and a grant from the UAB Health Services Foundation.

## Abstract

**Motivation:** Microarrays are a powerful new technology that allow for the measurement of the expression of thousands of genes simultaneously. Often such measurements are obtained on samples of cases from two or more populations that differ with respect to some characteristic and investigators wish to test whether gene expression levels differ across the populations for each gene. Additionally, because of the relatively high costs of this type of research, the sample sizes tend to be quite small. Yet, if investigators test each gene separately without any correction for multiple testing, then the Type I error rate may be unacceptably high. If investigators apply a correction for multiple testing, a very small p-value will be required to declare significance. Finally, parametric statistical tests may not be valid with such small sample sizes and alpha levels and conventional non-parametric procedures may have zero power because of the limited number of permutations that can be obtained in small sample sizes. Therefore, a procedure is needed that is both non-parametric but capable of yielding p-values on the interval  $[0, 1]$ . Here we use a modified Chebyshev's inequality to develop a test having these properties.

**Results:** We evaluate its properties via simulation and show that it both holds the Type I error rate below nominal levels in almost all conditions and can yield p-values denoting significance even with very small sample sizes and stringent corrections for multiple testing.

**Availability:** We illustrate its use by application to a dataset involving changes in gene expression in response to the application of TNF- $\alpha$  in normal rheumatoid arthritis synovial fibroblasts (RASf) cells compared with RASf cell where Ad-I $\kappa$ B-DN has blocked the action of TNF- $\alpha$ .

**Contact:** [mbeasley@uab.edu](mailto:mbeasley@uab.edu) or [dallison@uab.edu](mailto:dallison@uab.edu)

**Supplementary Information:** Information concerning these procedures are available at:

<http://www.soph.uab.edu/Statgenetics/Research/chebby.htm>

## Introduction

Microarrays are a powerful new technology that allow for the measurement of the expression of thousands of genes simultaneously. Often such measurements are obtained on samples of cases from two or more populations that differ with respect to some characteristic and investigators wish to test whether gene expression levels differ across the populations for each gene. This presents a challenging multiple testing issue. As Gibson (2002) stated, "... it is not clear how to assess the appropriate level of significance for microarrays in which thousands of comparisons are performed..." (p. 20). Additionally, because of the relatively high costs of this type of research, the sample sizes tend to be quite small. For example, Kayo et al. (2001) used only  $n=3$  monkeys per group when studying aging and caloric restriction. Corominola *et al.* (2001) used only  $n=4$  humans per group when studying obesity and diabetes in humans.

This large number of variables ( $k$ ) to small number of total cases ( $N$ ) ratio renders multivariate solutions difficult and in some cases intractable. Thus, the researchers often conduct multiple univariate tests. Some novel methods involve combining the results of these multiple tests in second level analyses (Allison *et al.*, 2002). However, if investigators test each gene separately without any correction for multiple testing, then the experiment-wise type 1 error rate will be unacceptably high (Tusher, *et al.*, 2001). Therefore, with thousands of variables to test, microarray researchers may wish to apply a correction for multiple testing (Allison and Coffey, 2002; Wu, 2001).

Using standard adjustments for multiple tests yields extremely small alphas, which reduces statistical power. For example, with  $\alpha=.05$  and  $k=1000$  variables (genes) the Bonferroni adjustment results in  $\alpha_{\text{BON}}=\alpha/k=.00005$ ; whereas, the Dunn-Šidák (Dunn, 1961; Šidák, 1967) correction yields a slightly larger value,  $\alpha_{\text{DS}}=(1-(1-\alpha)^{1/k})=.00005129$ . These adjustments are

based on the multiple tests being independent; however, it is likely that there is some correlation (i.e., dependence) among several tests conducted on data from the same study. Allison and Beasley (1998) proposed a Monte Carlo based procedure for adjusting alpha when multiple dependent tests are conducted. But, with the small sample sizes available to microarray researchers, it may not be possible to obtain stable and accurate correlation coefficients necessary for this procedure. Nevertheless, a large number of tests may call for a very small per-test alpha level, regardless of the adjustment procedure.

This problem of a large  $k/N$  ratio is even more pernicious because it presents several other complications. First, estimates of skew and kurtosis are also unstable with small samples; therefore, applied researchers cannot test whether data are normally distributed with any reasonable degree of power (Esteban *et al.*, 2001). Also, due to small sample sizes, parametric statistical tests of the differences between the mean levels of gene expression for each of the genes will be more sensitive to assumed distributional forms of the expression data (i.e., normality), and therefore, resulting p-values may not be accurate when there are departures from normality. Moreover, by applying an alpha adjustment due to the large number of tests, the resulting alpha will be extremely small. It is well documented that even when parametric tests are robust at, for example, the  $\alpha=.05$  level under violations of normality, they are often far less robust when very small alpha levels are used (Bradley, 1968; Hotelling, 1961). Thus, researchers cannot rely on robust asymptotic properties of parametric tests with such small sample sizes and small alphas.

Although a nonparametric test might be sought, conventional non-parametric tests also have severe limitations in this context. Consider the study by Lee *et al.* (2000), which used three mice per group. With  $n$  as small as 3 per group, conventional nonparametric tests for comparing

two groups (e.g., the Mann-Whitney U test; Di Bucchianico, 1999) cannot possibly yield two-tailed p-values  $< .10$ . This is because conventional non-parametric tests are based on the number of group combinations of ranks and there are a limited number of unique combinations that exist for finite data sets. Thus, in many situations, it is impossible to obtain “ $p < .05$ ,” let alone the far smaller p-values required if the significance level is corrected for multiple testing. Furthermore, in terms of robustness to heterogeneity of variance, rank-based tests can be nearly as sensitive as parametric procedures with small, unbalanced samples (Brunner and Langer, 2000; Zimmerman, 1996). When considering heterogeneity of variance in the present context, we are more concerned about robustness in terms of Type II error rate (the complement of power) than we are about type I error rate for reasons articulated below.

Bootstrap techniques are often suggested as an alternative (Kerr and Churchill, 2001) because they need not assume normality or homogeneity of variance (Good, 1999) and are therefore less restrictive. If one chooses the bootstrap as an alternative method to “nonparametrically” produce the distribution of p-values, however, a similar complication arises when resampling from very few cases. Specifically, the maximum number of different bootstrap samples for a two-group design with sample sizes of  $n_1$  and  $n_2$  is:

$$W_{\max} = [(2n_1-1)!/(n_1!(n_1-1)!)] [(2n_2-1)!/(n_2!(n_2-1)!)]$$

(Efron and Tibshirani, 1993). If sample sizes are very small (e.g.,  $n < 5$ ), p-values will be affected by the discreteness of the bootstrapped distribution and there will be a limited number of “distinct” p-values. Table 1 shows the minimum two-tailed p-values of the nonparametric Mann-Witney U test and bootstrap technique for several two-group balanced designs with small sample sizes.

Nevertheless, even with  $n=3$  per group, at times a group difference is so large that

common sense seems to suggest that the observed sample difference is significant regardless of whether the data are normally distributed, even though a non-parametric or bootstrap p-value is not significant. Consider a case where the expression of a tightly regulated gene is completely ablated by an intervention (e.g., knocking out a gene in a model organism) and an array experiment yields values of [10000.1; 10100.2; 10200.4] in the 1<sup>st</sup> group and [1.1; 2.5; 0.8] in the 2<sup>nd</sup> group. Real-life examples of huge differences in gene expression are readily found in the literature. One example is the effect of knocking out the interleukin-6 (IL-6) gene which causes roughly a 35-fold change in expression of IGFBP-1 gene (Li *et al.*, 2001). Such results sometimes lead biologists state that they do not need statisticians to tell them a difference is ‘real’ and if a statistician says the difference is not significant, something must be wrong with the statistician or his or her methods. We acknowledge the sensibleness of this position, but believe that most scientists wish to say more than “its *obviously* real.” So, we seek a method that is non-parametric and yet theoretically capable of yielding p-values that are continuous on the interval  $0 < p \leq 1$  with *any*  $N > 4$  cases (i.e., a balanced two-group experiment with  $[n_1 = n_2] > 2$ ).

### Test Procedures

We use Chebyshev’s classic inequality or variations thereof (DasGupta, 2000; Saw, *et al.*, 1984) to construct tests we will call the “*Chebby Checkers*.” We recognize that these methods are likely to be very conservative and have very low power, but we only expect them to be used in those situations where investigators have very small samples and due to multiple variables (genes) must test at very small alphas in order to make confident and objective statements that their largest effects are significant. Chebyshev’s inequality states that the probability of a random variable ( $\tau$ ) exceeding any real value  $T > 0$  is:

$$\mathbb{P}\left(\left|\frac{\tau - \mu_\tau}{\sigma_\tau}\right| \geq T\right) \leq \frac{1}{T^2}, \quad (1)$$

where  $\mu_\tau$  and  $\sigma_\tau$  are the mean and standard deviation of  $\tau$ , respectively. For unimodal, symmetrically distributed random variables, Gauss (1823) showed that Chebyshev's original inequality can be tightened by multiplying the right hand side by (4/9) (see Mallows, 1956). DasGupta (2000) proved that for a normally distributed random variable this bound can be tightened further by multiplying the right hand side by (1/3). Furthermore, he showed that this improvement holds for a larger family of distributions.

Thus, the following inequality can be used to compute two-tailed p-values for hypothesis tests when the assumptions of the usual test statistic (e.g., normality) are potentially violated but not testable:

$$\mathbb{P}\left(\left|\frac{\tau - \mu_\tau}{\sigma_\tau}\right| \geq T\right) \leq \frac{1}{3T^2}. \quad (2)$$

Saw *et al.* (1984) proposed a variant of Chebyshev's inequality for sample data. That is, the population mean and variance are not known but are replaced with sample estimates. Saw *et al.* show that for a sample of fixed size ( $N$ ) the upper bound for the Chebyshev's inequality approaches:

$$\mathbb{P}\left(\left|\frac{\tau - \mu_\tau}{\sigma_\tau}\right| \geq T\right) \leq \frac{1}{(N+1)T^2}. \quad (3)$$

as the random variable ( $\tau$ ) becomes large. We propose combining Chebyshev's inequality in (1) and the Saw *et al.* (1984) bound in (3) and defining:

$$\mathbb{P}\left(\left|\frac{\tau - \mu_\tau}{\sigma_\tau}\right| \geq T\right) \leq \frac{1}{(N+1)T^2}. \quad (4)$$

DasGupta (2000) showed that the p-value for a given value of a  $t$ -statistic can be “reproduced” by using the following bound:

$$P\left(\left|\frac{\tau - \mu_\tau}{\sigma_\tau}\right| \geq T\right) \leq \frac{T(v + T^2)}{v(T^2 - 1)} f_v(T), \text{ for } T > 1, \quad (5)$$

where  $f_v(T)$  is the pdf for Student’s  $t$ -distribution with  $v$  degrees-of-freedom ( $df$ ), and the quantity  $|\tau - \mu_\tau / \sigma_\tau|$  is defined as the  $t$ -statistic (see 6 below). Because this method provides accurate estimates of the actual p-values (DasGupta, 2000), employing this bound directly would lead to the same problems as simply using the  $t$ -test. Namely, the p-values may not be accurate when assumptions are violated, especially for small and unequal samples.

Thus, for testing differences between two group means (e.g., Corominola *et al.*, 2001; Kayo *et al.*, 2001; Lee *et al.*, 1999), we suggest the independent samples  $t$ -test:

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{S_{(\bar{Y}_1 - \bar{Y}_2)}} \quad (6)$$

as the random variable ( $\tau$ ) for these procedures. If the gene expression levels are continuously distributed, as they should be when dealing with clonal or inbred strains (Goddard, 2001), the distribution of  $\tau$  will be unimodal and will be symmetric under the null hypothesis of no mean difference between the two groups,  $\mu_\tau = 0$ , if the group sample sizes are equal. Under normality, the  $t$ -statistic follows Student’s  $t$  distribution with  $v=(N-2)$   $df$  and has a standard deviation of  $\sigma_\tau = [v/(v-2)]^{1/2}$ . If the dependent variable ( $Y$ ) is negatively kurtotic, however, then the conventional  $t$ -test will be conservative (Box and Watson, 1962). By contrast, if  $Y$  has positive kurtosis, then the conventional  $t$ -test will be too liberal (for some values of alpha), but the sampling variance of  $s^2$  will increase at a faster rate than the sampling variance of  $\bar{Y}$  which further implies that under

positive kurtosis,  $\sigma_\tau < [v/(v-2)]^{1/2}$ . Furthermore, if  $Y$  is skewed, the  $t$ -test can be very liberal, especially with unequal sample sizes (Wilcox, 1993, 1997).

Importantly, these variants of Chebyshev's inequality are likely to hold for *any* distribution with finite 1<sup>st</sup> and 2<sup>nd</sup> moments. Furthermore, it is not necessary for the first two moments to be known or estimated for Chebyshev's inequality to hold; they only need to exist (DasGupta, 2000). This is particularly useful because the mean and variance of a distribution of  $t$ -statistics (6), although they may be close to their expected values, are typically unknown when underlying assumptions (e.g., normality) are severely violated. Thus, by defining  $\tau$  as the  $t$ -test (6) and  $\sigma_\tau$  as  $[v/(v-2)]^{1/2}$ , we can be reasonably certain that  $\tau$  has finite 1<sup>st</sup> and 2<sup>nd</sup> moments, even if  $Y$  is non-normal. That is, when the data are non-normal the first two moments of the distribution of  $t$ -statistics may not follow Student's  $t$ -distribution exactly, but they most assuredly exist. Therefore with small sample sizes, one can employ the  $t$ -statistic as a random variable ( $\tau$ ) and apply the DasGupta (2000) bound for p-values (5) and the modification of the Saw *et al.* (1984) version of Chebyshev's inequality (4).

It should be noted that by defining  $\tau$  as the  $t$ -statistic (6) and  $\sigma_\tau$  as  $[v/(v-2)]^{1/2}$ , the p-value from the inequality in (2),  $P_{(2)}$ , is always less than the two-tailed p-value for a computed  $t$ -statistic (6),  $P_{(t)}$ . Likewise, the p-value resulting from the inequality in (5),  $P_{(5)}$ , is always smaller than  $P_{(t)}$ . However, as the  $t$ -statistic (6) becomes large,  $P_{(5)}$  becomes virtually identical to  $P_{(t)}$ , and thus, the use of  $P_{(5)}$  could potentially suffer from the same problems of using the  $t$ -test when conducting significance tests at small alphas. By contrast, for small values of the  $t$ -statistic, p-values resulting from the inequality in (4),  $P_{(4)}$ , are smaller than  $P_{(t)}$ . However, there is a "crossover point" where,  $P_{(4)} > P_{(t)}$ . Thus, we propose using:

$$P_{(7)} = \max \{P_{(4)} P_{(t)}\} . \quad (7)$$

Thus, for smaller values of  $t$  (6),  $P_{(7)} = P_{(t)}$ , and for larger values of  $t$ ,  $P_{(7)} = P_{(4)}$ . Thus, for significance testing at larger alphas (e.g.,  $\alpha = .05$  or  $.01$ ), where the  $t$ -test tends to be more robust,  $P_{(t)}$  tends to be used. Likewise, at smaller alphas,  $P_{(4)}$  tends to be used. To illustrate, Table 2 shows these p-values for several values of  $t$  (6) and  $N$ .

### Simulation Study of Type I Error Rate

#### Method

To evaluate the performance of these testing procedures, we conducted a simulation study. We used a two-group design with small, unbalanced sample sizes of  $N = 6$  and  $8$ . Data for the two groups were sampled separately (i.e., independently) from the normal and several non-normal distributions with homogeneous variances. Therefore, the conditions of our simulations assume that the error terms for each group were IID $[0, \sigma^2]$ . Consistent with the challenge faced by microarray researchers, we suppose a situation in which the researcher will test differences in group means across  $k=1,000$  genes and thus to control for inflation of the Type I error rate will use a very small alpha of  $\alpha_{\text{BON}}=.0005$ . Because alpha was set at such a small value we conducted 5 million replications in order for the expected number of rejections to be reasonably large (i.e., 250 rejections) under the null hypothesis.

Using SAS/IML, we simulated normally distributed data ( $Y$ ) using the RANNOR function (SAS Institute, 2001). To create the non-normal distributions, we used a system of non-linear transformations from the generalized lambda distribution (Karian and Dudewicz, 2000):

$$Y = \lambda_1 + \frac{U^{\lambda_3} - (1-U)^{\lambda_4}}{\lambda_2}, \quad (8)$$

where  $U$  is a random deviate from the uniform distribution generated with the SAS RANUNI function. The resultant variable,  $Y$ , has a mean of 0 and unit variance with the skew and kurtosis specified in Table 3. To simulate an extremely non-normal distribution, we generated a lognormal distribution by:  $Y = (\exp^X)$ , where  $X$  is a random deviate from the unit normal distribution generated with the RANNOR function. The resultant variable,  $Y$ , has a mean of  $e^{5\sigma^2} = 1.65$  and a variance of  $e^{\sigma^2}(e^{\sigma^2} - 1) = 4.67$  with a skew of  $(e^{\sigma^2} + 2)\sqrt{(e^{\sigma^2} - 1)} = 6.18$  and kurtosis of  $(e^{\sigma^2})^4 + 2(e^{\sigma^2})^3 + 3(e^{\sigma^2})^2 - 3 = 113.94$ . The  $Y$  values were then standardized to have a mean 0 and unit variance. The lognormal distribution was chosen because this has been shown in past research to be especially troublesome for parametric statistics (Cressie and Whitford, 1986; Wilcox, 1997).

Relative Bias was calculated by dividing the empirical Type I error rate by the nominal alpha. If the Type I error rate is controlled then relative bias should be 1. Using Bradley's (1978) stringent criterion we considered any relative bias higher than 1.1 to be an inflation of the Type I error rate.

## Results

Table 4 shows the relative bias for the five test procedures under normality assumptions (i.e.,  $NID[0, \sigma^2]$ ) and under five non-normal  $IID[0, \sigma^2]$  conditions. As can be seen, the parametric  $t$ -test (6) inflated the Type I error rates with a mesokurtic (i.e., a distribution with kurtosis equal to the normal), but skewed distribution (GLD-A) and with a light-tailed distribution (GLD-C). It performed even worse for the extremely skewed and heavy-tailed Lognormal distribution. Four of the Chebby Checker (CC) procedures held the Type I error rates well below the nominal alpha and in fact rarely committed a false positive. The CC procedure based on (5) also controlled the

Type I error rate, except for the light-tailed distribution (GLD-C) with  $N=8$  and the Lognormal distribution with unequal samples of  $n_1=2$  and  $n_2=4$ . Table 5 shows that the Type I rate inflation that occurred at  $\alpha=.0005$  did not occur to any great extent at  $\alpha=.05$ . This shows the importance of conducting such simulation studies with the small alpha levels that may eventually be used.

### Simulation Study of Statistical Power

To investigate the statistical power of these methods, a simulation study similar to the first simulation study was conducted. A constant ( $\delta$ ) was added to the dependent variable ( $Y$ ) for the first group in order to create standardized group mean differences (i.e., effect sizes). Effect size constants of  $\delta = 2, 5, 10,$  and  $20$  were used. Because rejection rates were expected to be larger than  $\alpha = .0005$ , one million replications were conducted.

Tables 6, 7, 8 and 9 show the empirical power estimates for effect sizes of  $\delta = 2, 5, 10,$  and  $20$ , respectively. Conditions in which Type I error rates exceeded the nominal alpha with a relative bias greater than 1.1 are not reported. As would be expected the  $CC$  tests had low statistical power, even with an extremely large effect size of  $\delta = 10$ . Of these tests,  $CC(5)$  was the most powerful; however, this procedure inflated the Type I error rate in a few conditions. Furthermore, for several conditions in which  $CC(5)$  did not inflate the Type I error rate, the more powerful  $t$ -test (6) did not either, thus making the  $t$ -test preferable. Of the remaining tests  $CC(4)$  was slightly more powerful than either  $CC(1)$  or  $CC(2)$ .

### Discussion

It should be reiterated that we simulated data under  $\text{IID}[0, \sigma^2]$  assumptions. It is well known that the Type I error rate of the parametric  $t$ -test (6) is affected by between-group

differences in variance (e.g., Boneau, 1960; Scheffé, 1959) and skew (Cressie and Whitford, 1986; Wilcox, 1993), especially for unequal sample sizes. Preliminary simulation work has shown that even the most conservative Chebby Checker,  $CC(1)$ , can inflate Type I error rates with heterogeneous variances or with different distributions for the two groups. Satterthwaite (1949) adjustments could possibly remedy the heteroscedasticity problem (Algina et al., 1994), but there are no known adjustments for differences in distributional shapes. Thus, the IID assumptions are critical to valid Type I error rates for any of these procedures. Thus, it may be asked why we did not use conditions that violate the IID assumptions.

First, the IID assumptions are difficult to evaluate with small sample sizes; thus, for practical purposes a researcher would never know if they hold. More importantly, we contend that if a test statistic becomes large enough to become a “significant result” when the IID assumptions are not met, even though population means are identical, then it is still a valuable result to microarray researchers (see Cliff, 1993). That is, differences in variance or distributional shape are results not merely nuisances. Therefore, the procedures and IID conditions we simulated seem reasonable in that a microarray researcher would not want to contend there are group differences when in fact the populations are identical (in terms of location, spread and shape), but the rejection is due to the sensitivity of the  $t$ -test to non-normality. However, if the IID assumptions do not hold, it should be viewed as a result that can be detected with some degree of statistical power. Typically, departures from IID can be tested with non-parametric methods (Mann-Whitney U, Cliff’s  $d$ ), but as stated previously they suffer from a limited range of  $p$ -values for small samples (see Table 1).

The methods we have proposed are certainly not the sole methods that should be applied to microarray data. We acknowledge that these are methods with very low power for all but the

largest effects. Nevertheless, in many cases, those are exactly the effects observed and of interest. As Richmond and Somerville (2000) stated, “in ‘marker discovery’ experiments, the goal is to discover a limited number of highly specific marker genes for a cell type, a developmental stage or an environmental treatment. In such experiments, the researcher is often interested only in genes that show a dramatic and selective induction or repression of expression.” (p. 108).

In the majority of this paper, we refer to testing for differences between two groups, but the method generalizes to many other testing situations. For example, with multiple group comparisons, one could use the ANOVA  $F$ -statistic or some post-hoc test statistic (e.g., Tukey’s HSD, Fisher’s LSD) as the random variable ( $\tau$ ) and their known expected values and standard deviations as  $\mu_\tau$  and  $\sigma_\tau$ , respectively.

### Computational Example

Rheumatoid arthritis synovial fibroblasts (RASf) are abnormal cells found in the synovium around joints in individuals with rheumatoid arthritis (RA) and are associated with inflammation of the joints (Mountz et al., 2001). TNF- $\alpha$ , a cytokine (a class of secreted proteins that can stimulate cells to grow or differentiate), is a key pro-inflammatory molecule that contributes to the initiation and perpetuation of RA (Mountz and Zhang, 2001) and is a known cause RASf cells to grow and divide and thus increasing the progression of RA (Miyazawa et al., 1998). Drugs that inhibit TNF- $\alpha$  are effective treatment for RA. TNF- $\alpha$  binds to RASf and causes the phosphorylation of the protein inhibitor of NF- $\kappa$ B (I- $\kappa$ B) and nuclear translocation of the transcription factor nuclear factor- $\kappa$ B (NF- $\kappa$ B). NF- $\kappa$ B then causes many genes involved in inflammation, cell growth, and cell division to then be turned on. The I- $\kappa$ B gene may be mutated in such a way that TNF- $\alpha$  can no longer cause its phosphorylation and blocks the translocation of NF- $\kappa$ B. This mutation is called a dominant negative mutation and it has been cloned into an adenovirus (Ad-I $\kappa$ B-DN), which enables high-efficiency transduction of RASf and results in high levels of expression of I $\kappa$ B-DN

Suppose a microarray experiment examining the changes in gene expression in response to the application of TNF- $\alpha$  in normal RASf cells compared with RASf cell where Ad-I $\kappa$ B-DN has blocked the action of TNF- $\alpha$ . In this experiment, primary RASf cells are isolated from synovial fibroblasts from a single human and grown in-vitro to retain their original in-vivo dependence on TNF- $\alpha$ . Six samples are taken from these primary passage RASf. The Ad-I $\kappa$ B-DN construct is added to three of the samples. A control construct is added to the other three. After 15 hours TNF- $\alpha$  is added to all 6 samples. After 3 more hours the RNA is extracted from

the 6 samples. The RNA is then labeled and run the Affymetrix, Inc. (Santa Clara, Ca) U95Av2 microarray. Now suppose that the 2 groups of  $n = 3$  samples are compared for differential expression across  $k = 800$  suspected genes. With  $\alpha=.05$  and  $k=800$  variables (genes) the Bonferroni adjustment results in  $\alpha_{\text{BON}}=\alpha/k=.0000625$ . Table 10 shows actual data from such as experiment for 5 extreme cases. Means, standard deviations,  $t$ -statistics (6),  $P_{(t)}$ , and p-values from the Chebby Checker ( $CC$ ) procedures are also displayed.

As can be seen, the  $t$ -test (6) is statistically significant for all 5 of these genes ( $P_{(t)} < .0000625$ ). However, one may question whether these data were sampled from a normal distribution. Without knowledge of the distributions from which these data were sampled, one may also question whether these results are valid or whether they are false positives (Type I errors) due to the sensitivity of the  $t$ -test to departure from normality. These results also show that the  $CC(5)$  procedure also yields statistically significant results for these 5 genes. The  $CC(7)$  method yields statistically significant p-values for the 408-at and 35922\_at genes Even the less powerful  $CC(2)$  procedure yields a statistically significant p-value for the 408-at gene. This demonstrates that the Chebby Checker methods can be used in order to make confident and objective statements that the large effects are statistically significant.

## References

- Algina, J., Oshima, T.C. and Lin, W.Y. (1995) Type I error rates for Welch's test and James's second-order test under nonnormality and inequality of variance when there are two groups. *J. Educ. Behav. Stat.*, **19**, 275-291.
- Allison, D.B. and Beasley, T.M. (1998) Method and computer program for controlling the family-wise alpha rate in gene association studies involving multiple phenotypes. *Genet. Epidemiol.*, **15**, 87-101.
- Allison, D.B. and Coffey C.S. (2002) Two-stage testing in microarray analysis: What is gained? *J. Gerontol. A Biol. Sci. Med. Sci.*, **57**, B189-B192.
- Allison, D.B., Gadbury, G.L., Heo, M., Fernández, J.R., Lee, C-K., Prolla, T.A. and Weindruch, R. (2002) A mixture model approach for the analysis of microarray gene expression data. *Comp. Stat. Data Anal.*, **39**, 1-20
- Boneau, C.A. (1960) The effects of violations of assumptions underlying the t test. *Psychol. Bull.*, **57**, 49-64.
- Box G.E.P. and Watson, G.S. (1962) Robustness to non-normality of regression tests. *Biometrika*, **49**, 93-106
- Bradley, J.V. (1968) *Distribution-free statistical tests*. Prentice-Hall, Englewood Cliffs, NJ.
- Bradley, J.V. (1978) Robustness? *Brit. J. Math. Stat. Psychol.*, **31**, 144-152.
- Brunner. E. and Munzel, U. (2000) The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation. *Biometrical J.*, **42**, 17-25.
- Cliff, N. (1993) Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychol. Bull.*, **114**, 494-509

- Corominola H., Conner L.J., Beavers L.S., Gadski R.A., Johnson D., Caro, J.F. and Rafaeloff-Phail, R. (2001) Identification of novel genes differentially expressed in omental fat of obese subjects and obese type 2 diabetic patients. *Diabetes*, **50**, 2822-2830.
- Cressie, N.A. and Whitford, H.J. (1986) How to use the two sample *t*-test. *Biometrical J.*, **28**, 131-148.
- DasGupta, A. (2000) Best constants in Chebychev inequalities with various applications. *Metrika*, **51**, 185-200.
- Di Bucchianico, A. (1999). Combinatorics, computer algebra and the Wilcoxon-Mann-Whitney test. *J. Stat. Plan. Inference*, **79**, 349-364.
- Dunn, O.J. (1961) Multiple comparison among means. *J. Amer. Stat. Assoc.*, **56**, 52-64.
- Efron, B. and Tibshirani, R.J. (1993) *An Introduction to the Bootstrap*. Chapman and Hall, New York
- Esteban, M.D., Castellanos, M.E., Morales, D. and Vajda, I. (2001) Monte Carlo comparison of four normality tests using different entropy estimates. *Commun. Stat. - Simul, Comp.*, **30**, 761-785.
- Gauss, C.F. (1823) *Theoria Combinationis Observatoinum Erroribus Minimus Obnoxiae*, Pars Prior. *Commentationes Societatis regiae Scientiarum Gottingensis Recentiores*, 5.
- Gibson, G. (2002) Microarrays in ecology and evolution: a preview. *Mol. Ecol.*, **11**, 17-24.
- Goddard M.E. (2001) The validity of genetic models underlying quantitative traits. *Livestock Prod. Sci.*, **72**, 117-127.
- Good, P.I. (2000) *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses* (2<sup>nd</sup> ed.). Springer-Verlag, London..

- Hotelling, H. (1961) The behavior of some standard statistical tests under non-standard conditions. In Neyman, J. (ed), *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1. University of California Press, Berkeley CA, pp. 319-359.
- Karian, Z.A. and Dudewicz, E.J. (2000) *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods*. New York, CRC Press.
- Kayo, T., Allison, D.B., Weindruch, R. and Prolla, T.A. (2001) Influences of aging and caloric restriction on the transcriptional profile of skeletal muscle from rhesus monkeys. *PNAS*, **98**, 5093-5098.
- Kerr, M.K. and Churchill, G.A. (2001) Statistical design and the analysis of gene expression microarray data. *Genet. Res.*, **77**, 123-128.
- Lee, C.K., Weindruch, R. and Prolla, T.A. (2000) Gene-expression profile of the ageing brain in mice. *Nat. Genet.*, **25**, 294-297.
- Li, W., Liang, X., Leu, J.I., Kovalovich, K., Ciliberto, G. and Traub, R. (2001) Global changes in interleukin-6-dependent gene expression patterns in mouse livers after partial hepatectomy. *Hepatology*, **6**, 1377-1386.
- Mallows, C.L. (1956) Generalizations of Tchebycheff's inequalities. *J. Royal Stat. Soc. – Ser. B*, **18**, 139-176.
- Miyazawa, K., Mori, A., Yamamoto, K. and Okudaira, H. (1998) Transcriptional roles of CCAAT/enhancer binding protein-beta, nuclear factor-kappaB, and C-promoter binding factor 1 in interleukin (IL)-1beta-induced IL-6 synthesis by human rheumatoid fibroblast-like synoviocytes. *J. Biol. Chem.*, **273**, 7620-7627.

- Mountz, J.D., Hsu, H.C., Matsuki, Y. and Zhang, H.G. (2001) Apoptosis and rheumatoid arthritis: past, present, and future directions. *Curr. Rheumatol. Rep.*, **3**(1), 70-78.
- Mountz, J.D. and Zhang, H.G. (2001) Regulation of apoptosis of synovial fibroblasts. *Curr. Dir. Autoimmun.*, **3**, 216-239.
- Richmond, T. and Somerville, S. (2000) Chasing the dream: plant EST microarrays. *Curr. Opin. Plant Biol.*, **3**, 108-116.
- Satterthwaite, F.E. (1946) An approximate distribution of estimates of variance components. *Biometrics Bull.*, **2**, 110-114.
- Saw, J.G., Yang, M.C.K. and Mo, T.C. (1984) Chebyshev inequality with estimated mean and variance. *Amer. Statistician*, **38**, 130- 132 .
- Scheffé, H. (1959) *The Analysis of Variance*. Wiley, New York.
- Šidák, Z. (1967) Rectangular confidence regions for the means of multivariate normal distributions. *J. Amer. Stat. Assoc.*, **62**, 626-633.
- Tusher, V.G., Tibshirani, R, and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *PNAS*, **98**, 5116-5121.
- Wilcox, R.R. (1993) Robustness in ANOVA. In E. Edwards, E. (ed), *Applied Analysis of Variance in the Behavioral Sciences*, New York, Marcel Dekker, pp. 345-374.
- Wilcox, R.R. (1997) *Introduction to Robust Estimation and Hypothesis Testing*. Academic Press, San Diego, CA.
- Wu, T.D. (2001) Analysing gene expression data from DNA microarrays to identify candidate genes. *J. Pathol.*, **195**, 53-65.
- Zimmerman, D.W. (1996) A note on homogeneity of variance of scores and ranks. *J. Exper. Educ.*, **64**, 351-362.

**Table 1.** Minimum Possible Two-Tailed p-values for the Mann-Witney U and Bootstrap Procedures.

$n$	Mann-Witney U	Bootstrap
3	.10	.02
4	.0285714	.0016327
5	.0079365	.0001260
6	.0021645	.0000094
7	.0005828	.0000007

**Table 2.** Calculated two-tailed p-values for the proposed methods.

$\tau = t(6)$	$N = 6$						$N = 8$					
	$P_{(t)}$	$P_{(1)}$	$P_{(2)}$	$P_{(5)}$	$P_{(4)}$	$P_{(7)}$	$P_{(t)}$	$P_{(1)}$	$P_{(2)}$	$P_{(5)}$	$P_{(4)}$	$P_{(7)}$
2.0	.11612	.50000	.16667	.57735	.07143	.11612	.09243	.37500	.12500	.29910	.04167	.09243
2.5	.06677	.32000	.10667	.26245	.04571	.06677	.04653	.24000	.08000	.13202	.02667	.04653
3.0	.03994	.22222	.07407	.14674	.03175	.03994	.02401	.16667	.05556	.06629	.01852	.02401
3.5	.02490	.16327	.05442	.08993	.02332	.02490	.01283	.12245	.04082	.03563	.01361	.01361
4.0	.01613	.12500	.04167	.05832	.01786	.01786	.00712	.09375	.03125	.02011	.01042	.01042
4.5	.01082	.09877	.03292	.03941	.01411	.01411	.00410	.07407	.02469	.01182	.00823	.00823
5.0	.00749	.08000	.02667	.02752	.01143	.01143	.00245	.06000	.02000	.00719	.00667	.00667
5.5	.00533	.06612	.02204	.01975	.00945	.00945	.00151	.04959	.01653	.00452	.00551	.00551
6.0	.00388	.05556	.01852	.01451	.00794	.00794	.00096	.04167	.01389	.00292	.00463	.00463
7.0	.00219	.04082	.01361	.00831	.00583	.00583	.00042	.03061	.01020	.00131	.00340	.00340
8.0	.00132	.03125	.01042	.00507	.00446	.00446	.00020	.02344	.00781	.00064	.00260	.00260
$p < .00005$	18.53	200.01	115.48	26.23	75.59	75.59	10.26	173.21	100.00	12.61	57.74	57.74

**Note:**  $\tau = t(6)$ ;  $\mu_\tau = 0$ ;  $\sigma_\tau = [v/(v-2)]^{1/2}$ ;  $v=(N-2)$ , and thus,  $T = t/[(N-2)/(N-4)]^{1/2}$ . The last row ( $p < .00005$ ) shows the  $t$ -statistic (6) necessary for each method to reject the null hypothesis at  $\alpha = .00005$ .

**Table 3.** Skew, Kurtosis, and Lambda coefficients for the generation of the Generalized Lambda Distributions (GLD).

Distribution	$\lambda_1$	$\lambda_2$	$\lambda_3$	$\lambda_4$	Skew ( $\gamma^3$ )	Kurtosis ( $\gamma^4$ )
GLD - A	-1.2251	0.1995	0.00685	0.3356	0.8	3.0
GLD - B	0	-0.4595	-0.17740	-0.1774	0.0	14.7
GLD - C	0	0.5773	1.00000	1.0000	0.0	1.8
GLD - D	-0.2937	-0.3658	-0.10470	-0.1833	2.0	21.9
Normal					0.0	3.0
Lognormal					6.2	113.9

**Note:** Normal and Lognormal distributions were generated using SAS RANNOR function.

Table 4. Relative Bias for  $\alpha = 0.00005$  based on 5,000,000 replications.

Distribution	Test	$n_1=3$ $n_2=3$	$n_1=2$ $n_2=4$	$n_1=4$ $n_2=4$	$n_1=3$ $n_2=5$
Normal $\gamma^3 = 0$ $\gamma^4 = 3.0$	$t(6)$	0.964	1.004	0.928	0.940
	$CC(1)$	0.000	0.000	0.000	0.000
	$CC(2)$	0.000	0.000	0.000	0.000
	$CC(4)^7$	0.004	0.004	0.000	0.000
	$CC(5)$	0.224	0.244	0.308	0.240
GLD - A $\gamma^3 = 0.8$ $\gamma^4 = 3.0$	$t(6)$	2.124 *	2.552 *	2.224 *	2.832 *
	$CC(1)$	0.000	0.000	0.000	0.000
	$CC(2)$	0.000	0.000	0.000	0.000
	$CC(4)^7$	0.004	0.012	0.000	0.000
	$CC(5)$	0.556	0.756	0.720	1.024
GLD - B $\gamma^3 = 0$ $\gamma^4 = 14.7$	$t(6)$	0.568	0.880	0.540	0.544
	$CC(1)$	0.000	0.000	0.000	0.000
	$CC(2)$	0.000	0.000	0.000	0.000
	$CC(4)^7$	0.000	0.000	0.000	0.000
	$CC(5)$	0.128	0.236	0.132	0.152
GLD - C $\gamma^3 = 0$ $\gamma^4 = 1.8$	$t(6)$	3.640 *	3.124 *	4.624 *	4.300 *
	$CC(1)$	0.000	0.000	0.000	0.000
	$CC(2)$	0.000	0.000	0.000	0.000
	$CC(4)^7$	0.016	0.020	0.000	0.000
	$CC(5)$	0.976	0.912	1.796 *	1.580 *
GLD - D $\gamma^3 = 2.0$ $\gamma^4 = 21.9$	$t(6)$	0.600	1.064	0.500	0.692
	$CC(1)$	0.000	0.000	0.000	0.000
	$CC(2)$	0.004	0.000	0.000	0.000
	$CC(4)^7$	0.004	0.004	0.000	0.000
	$CC(5)$	0.116	0.236	0.136	0.252
Lognormal $\gamma^3 = 6$ $\gamma^4 = 100$	$t(6)$	1.404 *	5.300 *	0.944	2.696 *
	$CC(1)$	0.000	0.000	0.000	0.000
	$CC(2)$	0.000	0.000	0.000	0.000
	$CC(4)^7$	0.000	0.012	0.000	0.000
	$CC(5)$	0.388	1.508 *	0.288	0.972

Table 5. Relative Bias for  $\alpha = 0.05$  based on 5,000,000 replications.

Distribution	Test	$n_1=3$ $n_2=3$	$n_1=2$ $n_2=4$	$n_1=4$ $n_2=4$	$n_1=3$ $n_2=5$
Normal $\gamma^3 = 0$ $\gamma^4 = 3.0$	$t(6)$	0.998964	0.998336	0.998696	0.997432
	$CC(1)$	0.063324	0.063860	0.030408	0.031028
	$CC(2)$	0.435928	0.434168	0.389708	0.390232
	$CC(4)^7$	0.998964	0.998336	0.998696	0.997432
	$CC(5)$	0.276200	0.276000	0.361484	0.362048
GLD - A $\gamma^3 = 0.8$ $\gamma^4 = 3.0$	$t(6)$	1.022748	1.034504	0.974360	0.975356
	$CC(1)$	0.091324	0.102608	0.044104	0.048376
	$CC(2)$	0.489004	0.513460	0.400964	0.410632
	$CC(4)^7$	1.022748	1.034504	0.974360	0.975356
	$CC(5)$	0.328804	0.351668	0.374604	0.384248
GLD - B $\gamma^3 = 0$ $\gamma^4 = 14.7$	$t(6)$	0.788436	0.872180	0.808344	0.845452
	$CC(1)$	0.041440	0.053612	0.017180	0.019548
	$CC(2)$	0.313712	0.367832	0.275252	0.296748
	$CC(4)^7$	0.788436	0.872180	0.808344	0.845452
	$CC(5)$	0.191468	0.231552	0.252864	0.273720
GLD - C $\gamma^3 = 0$ $\gamma^4 = 1.8$	$t(6)$	1.213360	1.117312	1.122276	1.093584
	$CC(1)$	0.136524	0.114896	0.072296	0.065404
	$CC(2)$	0.626492	0.557924	0.515480	0.488040
	$CC(4)^7$	1.213360	1.117312	1.122276	1.093584
	$CC(5)$	0.438512	0.382508	0.485656	0.459276
GLD - D $\gamma^3 = 2.0$ $\gamma^4 = 21.9$	$t(6)$	0.798420	0.889220	0.812072	0.849284
	$CC(1)$	0.044600	0.060020	0.018900	0.022412
	$CC(2)$	0.324784	0.385412	0.281020	0.304480
	$CC(4)^7$	0.798420	0.889220	0.812072	0.849284
	$CC(5)$	0.200412	0.247012	0.258192	0.281352
Lognormal $\gamma^3 = 6$ $\gamma^4 = 100$	$t(6)$	0.673532	0.862252	0.614032	0.689568
	$CC(1)$	0.058584	0.118964	0.020712	0.034876
	$CC(2)$	0.308380	0.458080	0.215088	0.270964
	$CC(4)^7$	0.673532	0.862252	0.614032	0.689568
	$CC(5)$	0.206172	0.329880	0.199064	0.253184

Table 6. Power for  $\alpha = 0.00005$  based on 1,000,000 replications  $\delta = 2$ .

Distribution	Test	$n_1=3$ $n_2=3$	$n_1=2$ $n_2=4$	$n_1=4$ $n_2=4$	$n_1=3$ $n_2=5$
Normal $\gamma^3 = 0$ $\gamma^4 = 3.0$	$t(6)$	0.001223	0.001034	0.004810	0.004118
	$CC(1)$	0.000000	0.000000	0.000000	0.000000
	$CC(2)$	0.000000	0.000000	0.000000	0.000000
	$CC(4)^7$	0.000005	0.000004	0.000000	0.000000
	$CC(5)$	0.000314	0.000264	0.001571	0.001339
GLD - A $\gamma^3 = 0.8$ $\gamma^4 = 3.0$	$t(6)$	0.001849 *	0.002177 *	0.006487 *	0.007274 *
	$CC(1)$	0.000000	0.000000	0.000000	0.000000
	$CC(2)$	0.000001	0.000002	0.000000	0.000000
	$CC(4)^7$	0.000007	0.000013	0.000000	0.000002
	$CC(5)$	0.000492	0.000669	0.002376	0.002779
GLD - B $\gamma^3 = 0$ $\gamma^4 = 14.7$	$t(6)$	0.002467	0.002089	0.012032	0.010731
	$CC(1)$	0.000000	0.000000	0.000000	0.000000
	$CC(2)$	0.000002	0.000000	0.000000	0.000000
	$CC(4)^7$	0.000007	0.000007	0.000000	0.000000
	$CC(5)$	0.000644	0.000482	0.004333	0.003746
GLD - C $\gamma^3 = 0$ $\gamma^4 = 1.8$	$t(6)$	0.001504 *	0.001191 *	0.004551 *	0.003936 *
	$CC(1)$	0.000000	0.000000	0.000000	0.000000
	$CC(2)$	0.000001	0.000003	0.000000	0.000000
	$CC(4)^7$	0.000006	0.000003	0.000001	0.000000
	$CC(5)$	0.000439	0.000328	0.001705 *	0.001474 *
GLD - D $\gamma^3 = 2.0$ $\gamma^4 = 21.9$	$t(6)$	0.002780	0.002649	0.013622	0.012882
	$CC(1)$	0.000000	0.000000	0.000000	0.000000
	$CC(2)$	0.000002	0.000001	0.000001	0.000000
	$CC(4)^7$	0.000010	0.000008	0.000002	0.000002
	$CC(5)$	0.000737	0.000706	0.004881	0.004509
Lognormal $\gamma^3 = 6$ $\gamma^4 = 100$	$t(6)$	0.002491 *	0.003530 *	0.008926	0.009598 *
	$CC(1)$	0.000000	0.000000	0.000000	0.000000
	$CC(2)$	0.000001	0.000003	0.000000	0.000000
	$CC(4)^7$	0.000010	0.000014	0.000002	0.000006
	$CC(5)$	0.000690	0.001093 *	0.003653	0.004094

Table 7. Power for  $\alpha = 0.00005$  based on 1,000,000 replications  $\delta = 5$ .

Distribution	Test	$n_1=3$ $n_2=3$	$n_1=2$ $n_2=4$	$n_1=4$ $n_2=4$	$n_1=3$ $n_2=5$
Normal $\gamma^3 = 0$ $\gamma^4 = 3.0$	$t(6)$	0.023454	0.019093	0.186161	0.165556
	$CC(1)$	0.000001	0.000000	0.000000	0.000000
	$CC(2)$	0.000013	0.000012	0.000001	0.000001
	$CC(4)^7$	0.000082	0.000087	0.000021	0.000015
	$CC(5)$	0.006392	0.005238	0.079843	0.069655
GLD - A $\gamma^3 = 0.8$ $\gamma^4 = 3.0$	$t(6)$	0.030457 *	0.028508 *	0.199101 *	0.182952 *
	$CC(1)$	0.000002	0.000002	0.000000	0.000000
	$CC(2)$	0.000039	0.000033	0.000001	0.000002
	$CC(4)^7$	0.000185	0.000168	0.000029	0.000043
	$CC(5)$	0.009121	0.008736	0.092453	0.085545
GLD - B $\gamma^3 = 0$ $\gamma^4 = 14.7$	$t(6)$	0.053648	0.045468	0.333857	0.307165
	$CC(1)$	0.000006	0.000004	0.000000	0.000000
	$CC(2)$	0.000058	0.000038	0.000001	0.000003
	$CC(4)^7$	0.000271	0.000221	0.000075	0.000069
	$CC(5)$	0.016056	0.013192	0.182376	0.164040
GLD - C $\gamma^3 = 0$ $\gamma^4 = 1.8$	$t(6)$	0.017608 *	0.013905 *	0.125957 *	0.109888 *
	$CC(1)$	0.000001	0.000001	0.000000	0.000000
	$CC(2)$	0.000009	0.000010	0.000001	0.000000
	$CC(4)^7$	0.000069	0.000057	0.000015	0.000015
	$CC(5)$	0.004996	0.003823	0.049855 *	0.042814 *
GLD - D $\gamma^3 = 2.0$ $\gamma^4 = 21.9$	$t(6)$	0.059218	0.051257	0.353966	0.329362
	$CC(1)$	0.000006	0.000004	0.000000	0.000000
	$CC(2)$	0.000055	0.000053	0.000004	0.000001
	$CC(4)^7$	0.000303	0.000260	0.000111	0.000101
	$CC(5)$	0.017753	0.015223	0.197642	0.181093
Lognormal $\gamma^3 = 6$ $\gamma^4 = 100$	$t(6)$	0.041527 *	0.039814 *	0.163409	0.154608 *
	$CC(1)$	0.000006	0.000013	0.000000	0.000001
	$CC(2)$	0.000057	0.000067	0.000003	0.000007
	$CC(4)^7$	0.000330	0.000307	0.000122	0.000131
	$CC(5)$	0.014480	0.014212 *	0.093889	0.088761

Table 8. Power for  $\alpha = 0.00005$  based on 1,000,000 replications  $\delta = 10$ .

Distribution	Test	$n_1=3$ $n_2=3$	$n_1=2$ $n_2=4$	$n_1=4$ $n_2=4$	$n_1=3$ $n_2=5$
Normal $\gamma^3 = 0$ $\gamma^4 = 3.0$	$t(6)$	0.220771	0.185707	0.913637	0.891029
	$CC(1)$	0.000030	0.000022	0.000000	0.000003
	$CC(2)$	0.000269	0.000205	0.000031	0.000040
	$CC(4)^7$	0.001433	0.001084	0.000896	0.000767
	$CC(5)$	0.073230	0.059821	0.719817	0.680424
GLD - A $\gamma^3 = 0.8$ $\gamma^4 = 3.0$	$t(6)$	0.236473 *	0.206769 *	0.908810 *	0.880787 *
	$CC(1)$	0.000050	0.000043	0.000004	0.000005
	$CC(2)$	0.000416	0.000395	0.000072	0.000071
	$CC(4)^7$	0.002107	0.001872	0.001655	0.001517
	$CC(5)$	0.087820	0.076629	0.707468	0.663854
GLD - B $\gamma^3 = 0$ $\gamma^4 = 14.7$	$t(6)$	0.358471	0.317768	0.886256	0.871281
	$CC(1)$	0.000087	0.000073	0.000004	0.000006
	$CC(2)$	0.000783	0.000631	0.000195	0.000192
	$CC(4)^7$	0.003918	0.003146	0.004110	0.003622
	$CC(5)$	0.151775	0.129613	0.767953	0.743374
GLD - C $\gamma^3 = 0$ $\gamma^4 = 1.8$	$t(6)$	0.165405 *	0.134663 *	0.964096 *	0.942765 *
	$CC(1)$	0.000020	0.000024	0.000001	0.000003
	$CC(2)$	0.000205	0.000172	0.000022	0.000018
	$CC(4)^7$	0.001055	0.000834	0.000469	0.000379
	$CC(5)$	0.051710	0.041546	0.718957 *	0.662762 *
GLD - D $\gamma^3 = 2.0$ $\gamma^4 = 21.9$	$t(6)$	0.377997	0.338651	0.889098	0.874494
	$CC(1)$	0.000088	0.000076	0.000007	0.000009
	$CC(2)$	0.000838	0.000732	0.000238	0.000192
	$CC(4)^7$	0.004278	0.003678	0.004893	0.004221
	$CC(5)$	0.164341	0.142126	0.779204	0.756100
Lognormal $\gamma^3 = 6$ $\gamma^4 = 100$	$t(6)$	0.212114 *	0.196178 *	0.542337	0.524592 *
	$CC(1)$	0.000118	0.000100	0.000010	0.000014
	$CC(2)$	0.000861	0.000823	0.000239	0.000226
	$CC(4)^7$	0.003994	0.003686	0.003845	0.003509
	$CC(5)$	0.099934	0.092146 *	0.415645	0.398133

Table 9. Power for  $\alpha = 0.00005$  based on 1,000,000 replications  $\delta = 20$ .

Distribution	Test	$n_1=3$ $n_2=3$	$n_1=2$ $n_2=4$	$n_1=4$ $n_2=4$	$n_1=3$ $n_2=5$
Normal $\gamma^3 = 0$ $\gamma^4 = 3.0$	$t(6)$	0.861007	0.813792	1.000000	0.999999
	$CC(1)$	0.000481	0.000349	0.000057	0.000063
	$CC(2)$	0.003899	0.003033	0.001794	0.001660
	$CC(4)^7$	0.019348	0.015549	0.036978	0.031892
	$CC(5)$	0.518978	0.458934	0.999944	0.999864
GLD - A $\gamma^3 = 0.8$ $\gamma^4 = 3.0$	$t(6)$	0.851466 *	0.797048 *	1.000000 *	1.000000 *
	$CC(1)$	0.000718	0.000580	0.000157	0.000145
	$CC(2)$	0.005784	0.004967	0.003271	0.002871
	$CC(4)^7$	0.026296	0.022421	0.047415	0.042064
	$CC(5)$	0.517857	0.460239	0.999998	0.999996
GLD - B $\gamma^3 = 0$ $\gamma^4 = 14.7$	$t(6)$	0.861556	0.832753	0.995299	0.994641
	$CC(1)$	0.001319	0.001075	0.000440	0.000383
	$CC(2)$	0.010728	0.008654	0.008823	0.007513
	$CC(4)^7$	0.048294	0.039478	0.107488	0.095122
	$CC(5)$	0.636445	0.591405	0.987491	0.985490
GLD - C $\gamma^3 = 0$ $\gamma^4 = 1.8$	$t(6)$	0.895919 *	0.835260 *	1.000000 *	1.000000 *
	$CC(1)$	0.000315	0.000246	0.000050	0.000026
	$CC(2)$	0.002767	0.002127	0.000988	0.000811
	$CC(4)^7$	0.013704	0.010542	0.019808	0.016361
	$CC(5)$	0.451868	0.382262	1.000000 *	1.000000 *
GLD - D $\gamma^3 = 2.0$ $\gamma^4 = 21.9$	$t(6)$	0.866753	0.838634	0.993795	0.992641
	$CC(1)$	0.001493	0.001152	0.000509	0.000448
	$CC(2)$	0.011937	0.009818	0.010365	0.008784
	$CC(4)^7$	0.053124	0.044561	0.119827	0.106829
	$CC(5)$	0.655212	0.609306	0.985178	0.982586
Lognormal $\gamma^3 = 6$ $\gamma^4 = 100$	$t(6)$	0.563911 *	0.535049 *	0.868312	0.858626 *
	$CC(1)$	0.001363	0.001267	0.000511	0.000431
	$CC(2)$	0.009973	0.008840	0.007162	0.006408
	$CC(4)^7$	0.037678	0.033684	0.059442	0.054211
	$CC(5)$	0.374103	0.348641 *	0.796664	0.783783

Table 10. Two-tailed p-values for the proposed methods based on the Computational Example.

Gene	Treated	Control	$t(6)$	$P_{(t)}$	$P_{(1)}$	$P_{(2)}$	$P_{(5)}$	$P_{(7)}$
408_at	116.8	1540.7	-76.29	.00000037 *	.00017180	.00005727 *	.00000071 *	.00002454 *
	60.8	1528.7						
	51.3	1506.2						
Mean	76.3	1525.2						
SD	35.4	17.5						
35992_at	302.4	1619.4	-57.22	.00000082 *	.00030542	.00010181	.00000223 *	.00004363 *
	248.5	1615.4						
	229.0	1576.6						
Mean	260.0	1603.8						
SD	38.0	23.6						
38488_at	-0.4	75.9	-47.68	.00000105 *	.00043980	.00014660	.00000461 *	.00006281
	2.1	76.6						
	4.5	77.9						
Mean	2.1	76.8						
SD	2.5	1.1						
37762_at	834.5	390.1	47.68	.00000116 *	.00043982	.00014661	.00000462 *	.00006282
	851.9	407.4						
	844.4	416.4						
Mean	843.6	404.6						
SD	8.7	13.4						
37032_at	510.2	1314.4	-42.92	.00000179 *	.00054297	.00018099	.00000704 *	.00007757
	558.4	1306.6						
	540.8	1278.3						
Mean	536.5	1299.8						
SD	24.4	19.0						

**Note:**  $\tau = t(6)$ ;  $\mu_\tau = 0$ ;  $\sigma_\tau = [v/(v-2)]^{1/2}$ ;  $v=(N-2)$ , and thus,  $T = t/[(N-2)/(N-4)]^{1/2}$ . The Kruskal-Wallis test was  $\chi^2 = 3.857$  with an exact two-tailed  $p = .10$  for all cases.