Electronic Journal of Statistics
Vol. 4 (2010) 1527–1546
ISSN: 1935-7524
DOI: 10.1214/10-EJS592

Detecting column dependence when rows are correlated and estimating the strength of the row correlation

Omkar Muralidharan*

390 Serra Mall
Stanford, CA 94305
e-mail: omkar@stanford.edu

Abstract: Microarray experiments often yield a normal data matrix X whose rows correspond to genes and columns to samples. We commonly calculate test statistics $Z = Xw$, where Z_i is a test statistic for the ith gene, and apply false discovery rate (FDR) controlling methods to find interesting genes. For example, Z could measure the difference in expression levels between treatment and control groups and we could seek differentially expressed genes. The empirical cdf of Z is important for FDR methods, since its mean and variance determine the bias and variance of FDR estimates. Efron (2009b) has shown that if the columns of X are independent, the variance of the empirical cdf of Z only depends on the mean-squared row correlation.

Microarray data, however, frequently shows signs of column dependence. In this paper, we show that Efron's result still holds under column dependence, and give a conservative (upwardly biased) estimator for the mean-squared row correlation. We show Fisher's transformation for sample correlations is still normalizing and variance stabilizing under column dependence, and use it to construct a permutation-invariant test of column independence. Finally, we argue that estimating the mean-squared row correlation under column dependence is impossible in general. Code to perform our test is available in the R package "colcor," available on CRAN.

Keywords and phrases: Fisher transformation, sample correlation, column dependence, root mean squared correlation, matrix normal.

Received October 2010.

1. Introduction

Microarray experiments often yield an $N \times n$ normal data matrix X whose rows represent genes and columns represent samples. Usually, N is much larger than n. One way to find "interesting" genes is to form a vector $Z = Xw$, where Z_i is a test statistic for gene i, and apply false discovery rate methods to find significant Z_is. For example, if the rows of X have unit variance and $w = (1, \dots, 1)/\sqrt{n}$, Z consists of the rescaled mean expression levels (that is, the one-sample t-statistics, but with known variance in the denominator). We could search for significantly large Z_i to find overexpressed genes.

*Supported by an NSF VIGRE Fellowship

The empirical distribution of the entries of Z, $\hat{F}$, is an important quantity for false discovery rate methods. Its mean and covariance determine the bias and variance of false discovery rate estimates. Efron (2009b) has approximated the mean and covariance of $\hat{F}$ when the columns of X are independent. But many have noted that microarray data sets show signs of column dependence, possibly caused by preparation and lab effects (Chen et al., 2004; Piper et al., 2002). Efron (2009a) developed a permutation test that finds dependence in standard datasets (Efron, 2009a; Cosgrove et al., 2010). In this paper, we consider how to use Efron's approximations under column dependence and propose a permutation-invariant test for column dependence.

Suppose the columns of X are iid $\mathcal{N}(0, \Sigma)$, with $\Sigma_{ii} = 1$, and let $\rho_{ii'} = \Sigma_{ii'}/\sqrt{\Sigma_{ii}\Sigma_{i'i'}}$ be the correlation between X_{ij} and $X_{i'j}$. Assuming the columns are independent, Efron (2009b) approximates the mean and covariance of $\hat{F}$ using only the mean and mean-squared correlations,

$$\alpha_1 = \frac{1}{\binom{N}{2}} \sum_{i<i'} \rho_{ii'}$$

$$\alpha_2 = \frac{1}{\binom{N}{2}} \sum_{i<i'} \rho_{ii'}^2,$$

not the full correlation structure Σ. Efron shows that these quantities can be estimated well, in contrast to Σ, which is very hard to estimate. If X is not centered and scaled, then we also need the mean of X and the variances Σ_{ii}.

Efron's approximations let us calculate the bias and variance of FDR estimates. Suppose we want to find rows with positive mean, and we reject the null of zero mean for all $Z_i \geq t$. The standard estimator of the false discovery rate for this rejection rule is

$$F\hat{D}R(t) = \frac{\hat{\pi}_0 (1 - \Phi(t))}{1 - \hat{F}(t)},$$

where $\hat{\pi}_0$ is the estimated fraction of Z_is with zero mean. Since $\hat{\pi}_0$ is usually close to 1 and $\Phi(t)$ is known, the bias and variance of $\hat{F}$ determine the bias and variance of $F\hat{D}R$. Efron's approximations give us a better picture of the behavior of $F\hat{D}R$ by approximating the mean and variance of $\hat{F}$ under row correlation.

Most microarray analyses assume independent columns, but column dependence can cause serious problems. For example, it can cause some of the over- or underdispersion commonly seen in microarray data. Suppose X is a standardized matrix of expression levels and we want to find genes (rows) that are significantly over- or underexpressed. A standard approach is to calculate a one-sample t-statistic Z_i for each row, then use an FDR procedure to find Z_is that are significantly far from 0. If the rows of X have unit variance, we can instead use the scaled row means by taking $w = (1, \ldots, 1)/\sqrt{n}$ and $Z = Xw$. Each Z_i is $\mathcal{N}(0, 1)$ under the null - if the columns of X are independent. If the columns

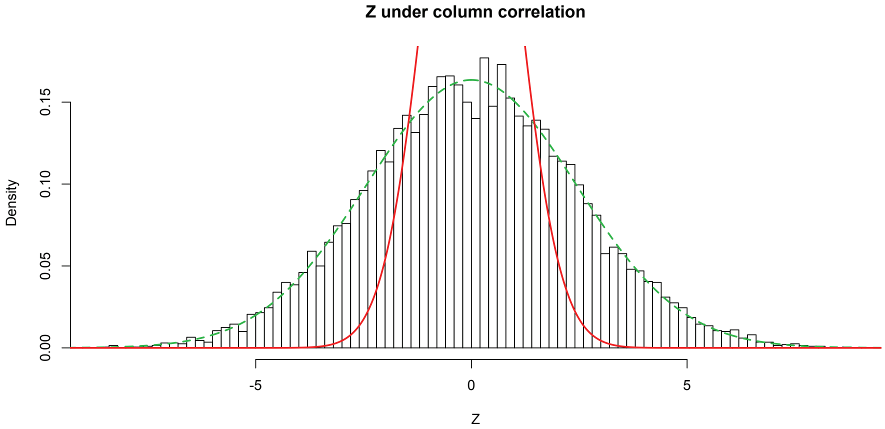

Fig 1. *Histogram of Z_i under column correlation. We took $N = 10,000$, $n = 100$, $\Sigma = I$ and generated X with zero mean and common column correlation 0.05, then calculated the scaled row means $Z_i = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_{ij}$. The Z_i are highly overdispersed; the solid curve is the $\mathcal{N}(0,1)$ null. Lemma 2 shows that the Z_i are $\mathcal{N}\left(0, 2.44^2\right)$, which agrees with the histogram (dashed curve).*

are correlated the Z_i can be substantially over- or underdispersed, as Figure 1 illustrates. Assuming column independence and using the $\mathcal{N}(0,1)$ null can give very misleading results when columns are correlated.

In Section 2 we show that Efron's approximations still work under column dependence, so they can be used to assess the variability of FDR estimates even when columns are correlated. Estimating α_1 and α_2, however, becomes more difficult. We find the uniformly minimum variance unbiased (UMVU) estimators of α_1 and α_2 under column independence. The estimator of α_1 remains unbiased under column correlation, while the α_2 estimator is upwardly biased. The bias is small unless the columns are quite strongly correlated. Our estimators are conservative: using them in Efron's approximations when columns are correlated gives an upwardly biased estimate of the variance of $\hat{F}$.

Efron (2009a) shows that the strong row correlations typically seen in microarray data (Qiu et al., 2005; Owen, 2005) make detecting column dependence tricky. He gave a permutation-based test for column dependence, but desired a test that does not depend on the ordering of the columns.

In Section 3, we show that Fisher's transformation can explain why column dependence is hard to detect and why it makes estimating α_2 difficult. Let $z(\cdot) = \tanh^{-1}(\cdot)$ be Fisher's transformation, and $\hat{\rho}_{ii'}$ the sample row correlations. We show that when the columns are dependent, $z(\hat{\rho}_{ii'})$ is still approximately normal with mean $z(\rho_{ii'})$ and constant variance, but the variance depends on the strength of the column dependence. This explains why it is difficult to estimate α_2 under column dependence: α_2 is approximately the variance of $z(\rho_{ii'})$, so estimating it is like separating out the variance of $z(\hat{\rho}_{ii'})$ into the variance of $z(\rho_{ii'})$ and the unknown noise variance. Transposing the argument shows why row dependence makes column dependence hard to detect.

Finally, in Section 4, we use Fisher's transformation to give a permutation-invariant test for column correlation. Our test formalizes an FDR-based heuristic that Efron uses. The test is particularly useful since its p-values are independent of those produced by Efron's permutation test. Our test has good power in simulations, especially when most columns are weakly correlated but some are highly correlated. It has little power when the column correlations are all of similar size.

2. A conservative estimator

2.1. Setup

We use the matrix normal distribution to model correlated columns. Our model is $X \sim \mathcal{N}(0, \Sigma \otimes \Delta)$, meaning that $cov\left(X_{ij}, X_{i'j'}\right) = \Sigma_{ii'}\Delta_{jj'}$. This notation for the matrix normal is nonstandard, but is used by Efron (2009a), whose results we will need. In this model, Σ controls the row covariance and Δ controls the column covariance.

We usually assume X is double standardized, so its rows and columns have zero mean and unit variance. Nearly all matrices can be double standardized by successive row and column standardization. Some pathological matrices cannot be double standardized, but Olshen and Rajaratnam (2010) show that these matrices are a set of Lebesgue measure zero in $\mathbb{R}^{N\times n}$. Accordingly, we will assume throughout that X is double standardized.

In our model, double standardization lets us take $\Sigma_{ii} = \Delta_{jj} = 1$, and forces each row and column of Σ and Δ to sum to zero. Double standardizing a normal matrix, however, does not yield a normal matrix. Although centering preserves normality, scaling does not. Modeling X as normal before double standardization is more realistic, but makes many computations intractable. We will usually approximate X as normal *after* double standardization to make computations easier.

This approximation generally works well, though rigorously quantifying the error of approximation is difficult. Lemma 1 approximates the kurtosis introduced by the first row scaling. Since the double standardization procedure usually converges quickly (Olshen and Rajaratnam, 2010), Lemma 1 gives us some idea of the non-normality introduced by double standardizing a normal matrix.

Lemma 1. *Suppose $X \sim \mathcal{N}(0, \Sigma \otimes \Delta)$ and assume $\Delta_{jj} = 1$. Let Y be the row standardized matrix, so $Y_{ij} = \frac{X_{ij}}{\sqrt{\frac{1}{n}\sum_j X_{ij}^2}}$. Then the skewness of Y_{ij} is zero and the kurtosis is approximately $-\frac{4}{n}\sum_{j'}\Delta_{jj'}^2 + 2\frac{\|\Delta\|_F^2}{n^2}$, where $\|A\|_F^2 = tr(A'A)$ is the Frobenius norm.*

Lemma 1 says that row standardization produces no skew and small kurtosis if column correlations are small. If Δ is close to the identity, row standardization introduces a small negative kurtosis, but depending on the correlation

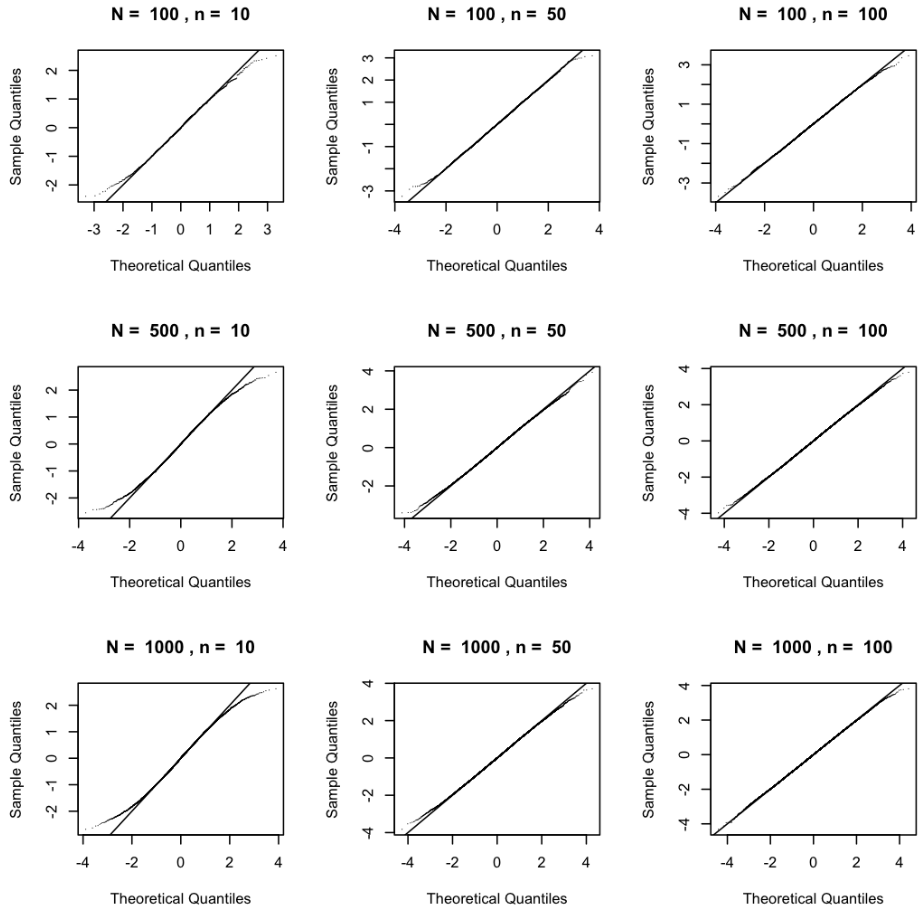

Fig 2. *Normal quantile-quantile plots of double standardized normal matrix entries for various* (N, n). *We generated* $X \sim \mathcal{N}(0, \Sigma \otimes \Delta)$, *then double standardized. The off-diagonal entries of* Σ, Δ *were all* 0.1 *to simulate moderate row and column correlation.*

structure, column correlation can introduce positive kurtosis as well. Transposing Lemma 1 shows that column standardization would introduce a kurtosis of $-\frac{4}{N}\sum_{i'}\Sigma_{ii'} + 2\frac{\|\Sigma\|_F^2}{N^2}$, if it were done on a normal matrix. Thus if row standardization approximately preserves normality, the following column standardization approximately preserves normality, as long as the row correlations are not too extreme. Subsequent row and column standardizations have much less of an effect, since after the first row and column standardizations, the matrix is approximately centered and scaled.

The lemma suggests that approximating X as normal after double standardization is reasonable, provided n, N are large and the row and column correlations are not too extreme. Figure 2 shows the distribution of the entries of a double-standardized X with weak row and column correlation for different values of n and N. The entries look fairly normal for $N \geq 100$ and $n \geq 50$.

We are interested in the empirical cdf $\hat{F}$ of $Z = Xw$, taking without loss of generality $\|w\| = 1$. When the columns are independent, $Z \sim \mathcal{N}(0, \Sigma)$. It is easy to calculate the mean and covariance of Z for general Δ.

Lemma 2. *Suppose $X \sim \mathcal{N}(0, \Sigma \otimes \Delta)$ before double standardization. Then Z has mean 0 and covariance $(w'\Delta w)\,\Sigma$. If we approximate $X \sim \mathcal{N}(0, \Sigma \otimes \Delta)$ after double standardization, then $Z \sim \mathcal{N}(0, (w'\Delta w)\,\Sigma)$.*

Allen and Tibshirani (2010) find the variance of each Z_i when Z is a vector of two-sample t-statistics; Lemma 2 generalizes their result. It has two important implications. First, as Allen and Tibshirani (2010) note, $w'\Delta w$ changes the variance of each Z_i. Column dependence can thus cause the over- or underdispersion often seen in microarray data.

Second, the approximate normality of Z lets us use Efron's approximations for the mean and variance of $\hat{F}$. Efron's approximations estimate the variance of Z, so they can account for the factor $w'\Delta w$; it is easy to estimate since it is common to all the Z_i. The lemma shows that column dependence does not affect the correlation structure of Z, so to use Efron's approximations, we still need the same mean and mean-squared correlations α_1 and α_2 as we would if the columns were independent.

Estimating α_1 and α_2 becomes more difficult when the columns are dependent. As Efron (2009a) notes, double standardization forces the row and column sample correlations to have the same mean and variance. For example, the mean-squared row correlation is $\frac{1}{n^2N^2} tr\left((XX')\,(XX')\right) = \frac{1}{n^2N^2} tr\left((X'X)\,(X'X)\right)$, which is just the mean-squared column correlation. This makes it hard to estimate α_2 and to test for column correlation. Row correlation can create the appearance of column correlation, and column correlation can inflate estimates of row correlation.

2.2. The independence-UMVU estimator

We can, however, find conservative estimators of α_1 and α_2, in the sense that under column correlation, the estimators are unbiased and positively biased, respectively. We begin by calculating their uniformly minimum variance unbiased (UMVU) estimators assuming the columns are independent. This result assumes that $X \sim \mathcal{N}(0, \Sigma \otimes I)$ and is *not* double standardized. This lets us apply a result of Olkin and Pratt (1958) to find the UMVU estimators easily.

Lemma 3. *Suppose $X \sim \mathcal{N}(0, \Sigma \otimes I)$ is not double standardized. Let $\hat{\rho}_{ii'}$ be the sample row correlations. Let $F(z; a, b, c) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b+k)\Gamma(c)}{\Gamma(a)\Gamma(b)\Gamma(c+k)} \frac{z^k}{k!}$ be the Gaussian (2,1) hypergeometric function, and let $f_1(r) = rF(1 - r^2; \frac{1}{2}, \frac{1}{2}, \frac{n-1}{2})$, $f_2(r) = 1 - \frac{n-2}{n-1}(1 - r^2)F(1 - r^2; 1, 1, \frac{n+1}{2})$.*

Then the UMVU estimators of α_1 and α_2 are $\tilde{\alpha}_1 = \frac{1}{\binom{N}{2}} \sum_{i<i'} f_1(\hat{\rho}_{ij})$ and $\tilde{\alpha}_2 = \frac{1}{\binom{N}{2}} \sum_{i<i'} f_2(\hat{\rho}_{ij})$. These are approximated to $O(n^{-2})$ by

$$\begin{aligned} \hat{\alpha}_1 &= \frac{1}{\binom{N}{2}} \frac{\Gamma(\frac{n-1}{2})\Gamma(\frac{n-3}{2})}{\Gamma(\frac{n-2}{2})^2} \sum_{i<j} \hat{\rho}_{ii'} \\ \hat{\alpha}_2 &= -\frac{1}{n-3} + \frac{n-1}{n-3}\left(\frac{1}{\binom{N}{2}} \sum_{i<i'} \hat{\rho}^2_{ii'} \right) \end{aligned}$$

and the approximations are particularly accurate if most $\hat{\rho}_{ii'}$ are small.

Lemma 3 gives the independence-UMVU estimators of α_1 and α_2, and more useful linear approximations (accuracy bounds for the linear approximations are given in the Appendix). These approximations let us avoid forming the $N \times N$ sample row correlation matrix $\hat{\Sigma}$, since $\sum_{i<i'} \hat{\rho}^2_{ii'} = \frac{1}{2}(\|\hat{\Sigma}\|^2_F - N)$, and we can compute the Frobenius norm using $\|XX'\|^2_F = \|X'X\|^2_F$.

We now calculate these estimators' bias under column dependence. To do this, we approximate X as $\mathcal{N}(0, \Sigma \otimes \Delta)$ after double standardization. Centering makes each row and column of Σ and Δ sum to zero. This fixes $\alpha_1 = -\frac{1}{N-1}$. The α_1 estimators, however, are defined more generally, and the bias of the α_1 estimators under column correlation may be of interest when α_1 is not fixed. We thus assume that $\Sigma_{ii} = \Delta_{jj} = 1$, but do not require Σ and Δ to be centered. Our basic tool is the following lemma, proved by Efron (2009a).

Lemma 4 (Efron)**.** *Suppose $X \sim \mathcal{N}(0, \Sigma \otimes \Delta)$ with $\Sigma_{ii} = \Delta_{jj} = 1$. Let $\hat{\sigma}_{ii'}$ be the sample covariance between columns i and i' in the double standardized model, and let $\|\Delta\|^2_F = tr(\Delta^2)$ be the Frobenius norm. Then*

$$\begin{aligned} E\left(\hat{\sigma}_{ii'}\right) &= \Sigma_{ii'} \\ cov\left(\hat{\sigma}_{ii'}, \hat{\sigma}_{kk'}\right) &= \frac{\|\Delta\|^2_F}{n^2}\left(\Sigma_{ik}\Sigma_{i'k'} + \Sigma_{ik'}\Sigma_{i'k}\right) \end{aligned}$$

Lemma 4 says that the sample row covariances are unbiased and have a Wishart covariance structure, but column dependence reduces the degrees of freedom from n to $\frac{n^2}{\|\Delta\|^2_F}$. An analogous formula holds for column covariances.

We now apply Lemma 4 to approximate the bias of the independence-UMVU estimators and their linear approximations. Lemma 5 gives a simple and accurate approximation; messier but more precise formulas are in the Appendix.

Lemma 5. *Suppose $X \sim \mathcal{N}(0, \Sigma \otimes \Delta)$ with $\Sigma_{ii} = \Delta_{jj} = 1$. Let δ be the mean-squared correlation of the columns, so $\delta = \frac{1}{\binom{n}{2}} \sum \Delta^2_{jj'}$. Then $E(\hat{\alpha}_2) = \alpha_2 + \delta(\alpha_2 + 1) + O(n^{-2})$ and $E(\tilde{\alpha}_2) = \alpha_2 + \delta(\alpha_2 + 1) + O(n^{-2})$. In particular, if $\alpha_2\delta$ is small,*

$$E\left(\hat{\alpha}_2\right) \approx E\left(\tilde{\alpha}_2\right) \approx \alpha_2 + \delta$$

In addition, $\hat{\alpha}_1$ is nearly unbiased: $E(\hat{\alpha}_1) = \alpha_1 + O(n^{-1})$ and $E(\tilde{\alpha}_1) = \alpha_1 + O(n^{-1})$.

Lemma 5 shows that our α estimators are conservative if the columns are in fact correlated. Efron's approximations for the variance of $\hat{F}$ increase linearly as α_1 and α_2 increase. The lemma thus guarantees that even under column correlation, using our estimators in Efron's approximations gives conservative, upwardly biased, estimates of the variance of $\hat{F}$.

Lemma 5 also tells us that our α_2 estimators' bias is just the mean-squared correlation of the columns. The bias is small unless the column correlations are comparable in strength to the row correlations. In the microarray setting, this would mean that the correlations between arrays are comparable to the correlations between genes, which would be a very bad experimental situation.

For the rest of this paper, we assume that X has been centered, so α_1 is known and only α_2 is left to estimate. To simplify notation, we drop the subscript and denote α_2 by α. This gives us different notation from Efron (2009b), who uses α for the *root*-mean-squared correlation.

3. Transforming correlations

Estimating δ, the mean-squared column correlation, would let us assess the extent of column correlation and improve our estimate of α. We will now see why this is difficult – α and δ are hard to separate. Fisher's transformation reveals that estimating α and δ is like trying to estimate the variances of two independent random variables based on observing their sum.

More generally, Fisher's transformation simplifies the study of row and column correlations by making the sample correlation distribution easier to understand and manipulate. The distribution of a sample row correlation depends on the true row and column correlations in a complicated way. Fisher's transformation simplifies this dependence by normalizing the sample row correlation and stabilizing its variance. The transformed sample row correlation is approximately normal with mean depending on the true row correlation and variance depending on the mean-squared column correlation. The same holds for sample column correlations as well, and we will use this to test for column correlation.

3.1. Fisher's transformation

We first show that Fisher's transformation for correlations still works when columns are dependent. We assume that X is normal after double standardization, so $X \sim \mathcal{N}(0, \Sigma \otimes \Delta)$ with $\Sigma_{ii} = \Delta_{jj} = 1$, and Σ, Δ have zero row and column sums.

Theorem 1. *Let $z(r) = \tanh^{-1} r$. Then*

$$z(\hat{\rho}_{ii'}) \dot{\sim} \mathcal{N}\left(z(\rho_{ii'}), \frac{\|\Delta\|_F^2}{n^2}\right).$$

An analogous formula holds for the transformed column correlations. Theorem 1 follows from the next two lemmas, which both use the delta method. Lemma 6 shows that the transformation is still variance stabilizing.

Lemma 6. *The delta method approximation for the mean and variance of $z(\hat{\rho}_{ii'})$ is $z(\hat{\rho}_{ii'}) \dot{\sim} \left(z(\rho_{ii'}), \frac{\|\Delta\|_F^2}{n^2}\right)$.*

Lemma 7 shows Fisher's transformation is still approximately normalizing.

Lemma 7. *The skewness and kurtosis of $\hat{\rho}_{ii'}$ are approximately*

$$\begin{aligned} skew\left(\hat{\rho}_{ii'}\right) &\approx \frac{6\Sigma_{ii'} + 2\Sigma_{ii'}^3}{\left(1-\Sigma_{ii'}^2\right)^3} \frac{tr\left(\Delta^3\right)}{\|\Delta\|_F^3} \\ kurt\left(\hat{\rho}_{ii'}\right) &\approx \frac{6 + 36\Sigma_{ii'}^2 + 6\Sigma_{ii'}^4}{\left(1-\Sigma_{ii'}^2\right)^4} \frac{tr\left(\Delta^4\right)}{\|\Delta\|_F^4}. \end{aligned}$$

These are also the delta method approximations to the skewness and kurtosis of $z\left(\hat{\rho}_{ii'}\right)$.

Lemma 7 shows that the skewness and kurtosis of the sample correlations is usually small. Since the off-diagonal elements of Δ^k decay rapidly, the $\frac{tr(\Delta^k)}{\|\Delta\|_F^k}$ terms behave roughly like $n^{1-\frac{k}{2}}$. Switching Σ and Δ gives analogous formulas for the sample column correlations.

Lemmas 6 and 7 together show that Fisher's transformation is still approximately normalizing and variance stabilizing when the columns are correlated. The variance of the transformed row correlations, though, depends on the mean-squared column correlation, and vice versa. The transformation is not perfect. It works very well in the center. If columns are highly correlated, however, the transformed row correlations can have lighter or heavier tails, and vice versa.

3.2. Interpretation

Theorem 1 relates estimating α and δ to a more familiar problem. If the columns were independent, the transformed row correlations $z(\hat{\rho}_{ii'})$ would have known, equal variance. Since $z(\rho) \approx \rho$ for small ρ, α is roughly the variance of the $z(\rho_{ii'})$. We could thus estimate α by measuring the variance of the $z(\hat{\rho}_{ii'})$ and subtracting the known noise variance.

When the columns are dependent, however, the variance of the $z(\hat{\rho}_{ii'})$ is equal but unknown. This makes it impossible to separate the variance of the $z(\hat{\rho}_{ii'})$ into row correlation (the variance of $z(\rho_{ii'})$) and column correlation (noise variance) components.

Estimating α and δ this way is like trying to estimate the variances of two independent random variables after only observing their sum. It is thus impossible without some strong assumptions on Σ or Δ. This explains why estimating α under column dependence is so difficult: any estimator must either rely on the failure of Theorem 1 or use some information not captured by treating the sample correlations as one large vector.

This view also explains the conservatism of our α estimators. Our estimators were designed for independence; they essentially take the variance of the $\hat{\rho}_{ii'}$ and adjust it using the known noise variance. Dependence increases the noise variance by roughly δ, since $\|\Delta\|_F^2 = n + n(n-1)\delta$. This widens the $\hat{\rho}_{ii'}$ distribution and biases our estimates upward. Lemma 5 says that the bias of our α estimators is roughly the increase in noise variance.

4. Testing for column correlation

Testing for column dependence is easier than estimating it. Efron (2009a) gave a permutation test of column dependence, but desired a test that does not depend on the ordering of the columns. In this section, we use our results on Fisher's transformation to give such a test.

Even if the columns are independent, standardization introduces small correlations between them. Our null hypothesis is thus not $\Delta = I$, but $\Delta = \frac{n}{n-1}(I - \frac{1}{n}\mathbf{1}\mathbf{1}')$, where $\mathbf{1}$ is an n-vector of ones. Our previous results make it easy to approximate the null distribution of the off-diagonal correlations, assuming X is normal after double standardization.

Lemma 8. *Under the null hypothesis* $\Delta = \frac{n}{n-1}(I - \frac{1}{n}\mathbf{1}\mathbf{1}')$, *if* $X \sim \mathcal{N}(0, \Sigma \otimes \Delta)$, $z(\hat{\Delta}_{jj'}) \dot{\sim} \mathcal{N}(z(-\frac{1}{n-1}), \frac{\|\Sigma\|_F^2}{N^2})$. *Also,* $cor(z(\hat{\Delta}_{jj'}), z(\hat{\Delta}_{ll'})) = \mathcal{O}(n^{-1})$ *for* $(j, j') \neq (l, l')$.

The transformed correlations all have the same mean and variance, and are themselves nearly uncorrelated. Testing for column dependence is thus approximately the same as observing $\binom{n}{2}$ independent normal random variables of the same unknown variance, and testing if they all have the same mean. If we center and scale the $z(\hat{\Delta}_{jj'})$, this reduces to testing whether a collection of $\binom{n}{2}$ independent $\mathcal{N}(0,1)$ random variables all have zero mean, or if some have nonzero mean.

Figure 3 illustrates this idea. They show the centered and scaled column correlations for two data sets considered by Efron (2009a). There are many more large correlations than we would expect under the $\mathcal{N}(0,1)$ null. This suggests both data sets have some column dependence, and indeed, our test strongly rejects the null in both cases. Efron (2009a) used a similar idea to explore the extent of column correlation. Our test builds on Efron's idea by using Fisher's transformation to justify an approximately normal null, and a different test for nonzero means.

Donoho and Jin (2004) introduced the "higher criticism" procedure to test whether many independent random variables are all $\mathcal{N}(0,1)$, or if some are normal with nonzero mean. The idea behind higher criticism is simple. For a fixed rejection threshold t, we can test the $\mathcal{N}(0,1)$ null by seeing if significantly many variables fall outside $[-t, t]$. Donoho and Jin (2004) maximize the resulting t-statistic over a range of t.

The higher criticism statistic takes the following form for our data:

$$\Lambda = \max_{t \in [a,b]} \frac{\left[\hat{G}(-t) + 1 - \hat{G}(1-t)\right] - \left[\Phi(-t) + 1 - \Phi(1-t)\right]}{\left[\frac{1}{n}\left(2\Phi(-t)\right)\left(1 - 2\Phi(-t)\right)\right]^{\frac{1}{2}}},$$

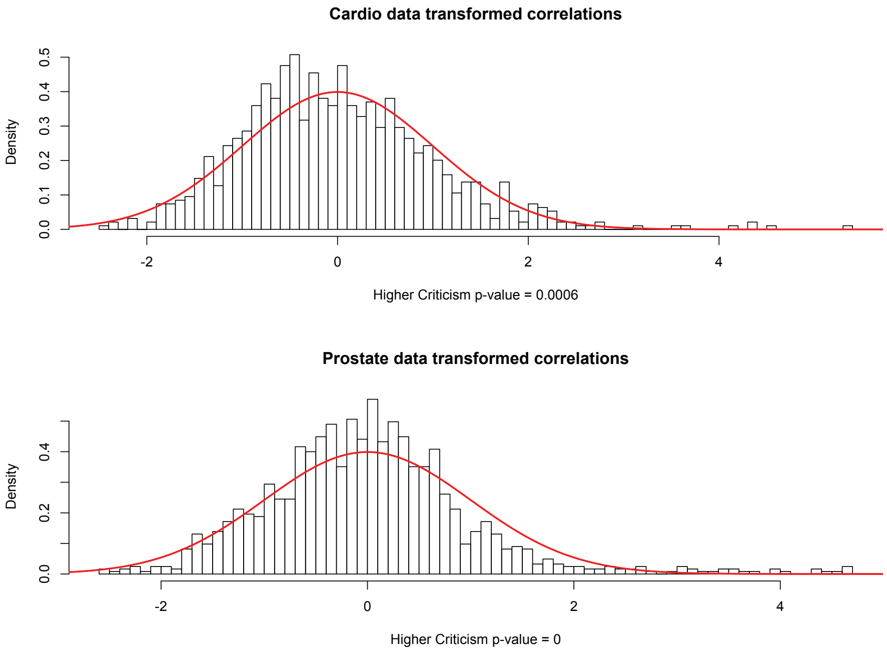

Fig 3. *Centered, scaled, transformed column correlations for the "Cardio Data" ($N = 20246$ genes on $n = 44$ healthy controls) and the "Prostate Data" of Singh et al. (2002) ($N = 6033$ genes on $n = 50$ healthy controls). The red line gives the $\mathcal{N}(0,1)$ distribution we expect under column independence, which our test rejects for both data sets.*

where $\hat{G}$ is the empirical cdf of the centered and scaled $z(\hat{\Delta}_{jj'})$. Based on the recommended interval in Donoho and Jin (2004), we choose the maximizing interval $[a,b]$ to be $[\Phi^{-1}(\frac{1}{4}), \Phi^{-1}(\frac{1}{n(n-1)})]$. The null distribution of Λ can be calculated by repeatedly simulating $\binom{n}{2}$ $\mathcal{N}(0,1)$ random variables, centering and scaling them, and calculating Λ for each simulated collection.

The transformed column correlations can have slightly heavy tails, especially under moderate row correlation. Figure 4 shows a normal quantile-quantile plot of the centered and scaled $z(\hat{\Delta}_{jj'})$ under the null and moderate row correlation. The figure shows that the correlations are quite normal in the center, but somewhat heavy-tailed. This deviation from normality can lead to spuriously large values of Λ. Interestingly, the problem is worst for moderate row correlation. When the row correlation is low, Lemma 7 says that the transformed correlations have low kurtosis. When the row correlation is high, the normal after double standardization approximation begins to break down, and the entries of X are light-tailed, as Lemma 1 predicts. Figure 5 shows normal quantile-quantile plots of the centered and scaled $z(\hat{\Delta}_{jj'})$, under the null and low or high row correlation. The correlations are more normal than they are for moderate correlation.

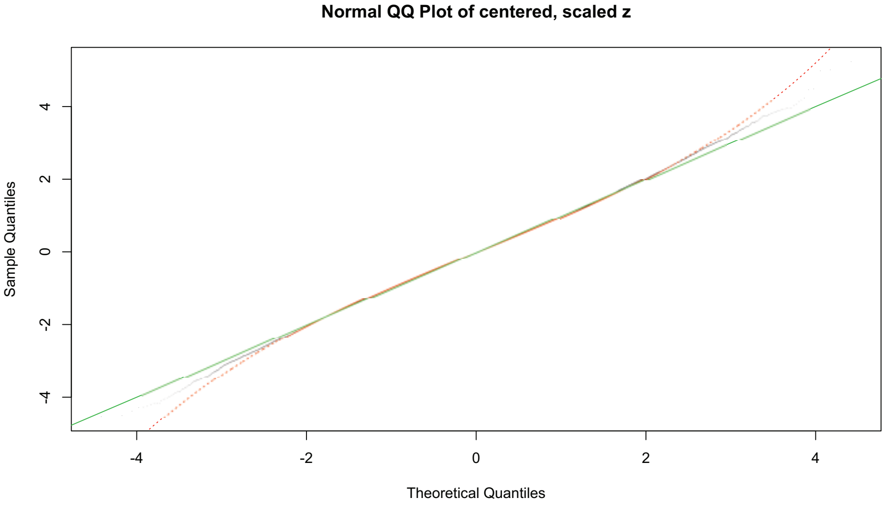

FIG 4. *Normal quantile-quantile plot of $z(\hat{\Delta}_{jj'})$ under null, with row correlations generated from the block setup, $\alpha^{\frac{1}{2}} = 0.083$. The dashed line shows scaled t-quantiles with 14 degrees of freedom, chosen to match the kurtosis of the $z(\hat{\Delta}_{jj'})$. The t distribution is too heavy in the far tail.*

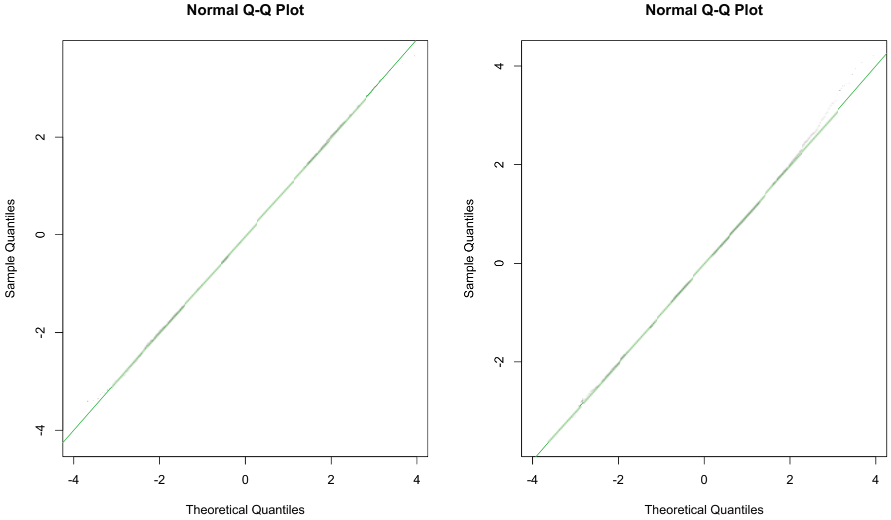

FIG 5. *Normal quantile-quantile plots of $z(\hat{\Delta}_{jj'})$ under the null, with row correlations generated from the block setup, $\alpha^{\frac{1}{2}} = 0.001$ (left) and $\alpha^{\frac{1}{2}} = 0.174$ (right).*

Shrinking the $z(\hat{\Delta}_{jj'})$ can control the tails and yield better behaved Λs. The obvious approach would be to use the t-distribution that matches the first four cumulants of $z(\hat{\Delta}_{jj'})$. That is, we would base Λ on $\Phi^{-1}(F_t(z))$, where F_t is the scaled t-distribution with appropriate degrees of freedom. However, the correlations are not as heavy tailed as a t-distribution that matches their variance and kurtosis, so the t-transformation overshrinks in the far tail. This hurts power dramatically. One good compromise is a linear approximation to the t-transformation, which shrinks the bulk of the data appropriately but does not overshrink the far tail. Shrinking by a factor of 0.9 approximates the t-transformation with 4 df reasonably well between $[-3, 3]$; this is a conservative choice of degrees of freedom, since it gives the t distribution infinite kurtosis.

The test's power depends on both the row and column correlations, and indirectly on the sample size. The higher α is, the higher the variance of the column correlations, so the lower our power. Increasing N will increase our power if α decreases, but will have little effect if α remains the same. The test is also sensitive to the particular configuration of column correlations. If the nonzero column correlations are all small, they will only inflate the center of the $z(\hat{\Delta}_{jj'})$ histogram; centering and scaling will eliminate this effect, and we will be unable to detect the correlations. We have more power when most column correlations are small but some are large. The larger and more numerous the large correlations are, the more power the test will have. Increasing n will increase power if the fraction of large correlations remains the same, since we will get better estimates of the column correlation distribution.

4.1. Simulations

We investigate the level and power of our test using simulations. We took $N = 1000$, $n = 50$ and generated data with a block row dependence structure. We divided the rows into 5 blocks, with constant correlation within block and no correlation between blocks, then standardized the row correlation matrix. The row correlations ranged from small (independence before standardization) to extreme (correlation of 0.9 within blocks before standardization). We check the level of the test under column independence and its power to detect three column correlation structures. The first corresponds to a "batch effect" - the columns have a block correlation structure like the rows, with 10 blocks. The second corresponds to an adjacent array effect, with $\Delta_{jj'} = \rho^{-|j-j'|}$. The third corresponds to a contamination setting, where most columns are uncorrelated, but a small block of 5 columns has constant correlation. All these descriptions are before standardization, which changes the exact correlations slightly but does not alter their broad structure. We determine the test's power in each of these settings under low, moderate or high row correlation. In every simulation, we first generate $X \sim \mathcal{N}(0, \Sigma \otimes \Delta)$, then double standardize.

Table 1 shows our test maintains its level across the range of row dependence. It is very conservative because of the shrinkage. This effect is most pronounced for weak or strong row dependence, since the correlations are heaviest-tailed for

Table 1
Level of test at different row correlations (simulation described in text). The range of αs correspond to within block row correlation ranging from 0 to 0.9, before standardization. The results are for 1000 replications, so standard errors are all less than 0.0055

$\alpha^{\frac{1}{2}} =$	0.001	0.041	0.083	0.128	0.173	0.222	0.272	0.325	0.380	0.438
Level at nominal 0.1	0.001	0.025	0.028	0.009	0.003	0.002	0	0	0	0
Level at nominal 0.05	0	0.013	0.016	0.005	0.001	0	0	0	0	0
Level at nominal 0.01	0	0	0.003	0.004	0.001	0	0	0	0	0

Table 2
Power to detect batch effect under low, medium, and high row correlation at level 0.05. The within block correlations given are after standardization; about 8.2% of the column correlations had this value, and the rest were small. To interpret the effect size, note that the transformed correlations have standard deviation approximately $\alpha^{\frac{1}{2}}$. Power was calculated over 1000 simulations, so the standard deviation is less than 0.016, and much lower near 0 and 1

Within block cor.	0.074	0.159	0.245	0.334	0.423	0.514	0.605	0.699	0.793	0.890
$\alpha^{\frac{1}{2}} = 0.020$	0.13	1	1	1	1	1	1	1	1	1
$\alpha^{\frac{1}{2}} = 0.128$	0.011	0.011	0.080	0.586	1	1	1	1	1	1
$\alpha^{\frac{1}{2}} = 0.325$	0	0	0	0.001	0.003	0.043	0.224	0.650	0.980	1

moderate row dependence. To interpret the row dependence strength, note that standard microarray datasets have $\hat{\alpha}^{\frac{1}{2}}$ in the range of 0 to 0.25. The prostate data of Singh et al. (2002) and the BRCA data of Hedenfalk et al. (2001) have $\hat{\alpha}^{\frac{1}{2}} = 0$, the leukemia data of Golub et al. (1999) has $\hat{\alpha}^{\frac{1}{2}} = 0.1430$ and the cardio data used by Efron has $\hat{\alpha}^{\frac{1}{2}} = 0.2341$ (all these correlations were estimated on the controls).

Despite its conservatism, our test maintains decent power under low to moderate row correlation. Tables 2, 3 and 4 show the results of the power simulations. As we would expect, the power increases with the column correlation and the number of large column correlations, but decreases with the row correlation. The batch effect is easiest to detect, followed by the adjacent array effect and the contamination effect. Surprisingly, the power of our test can actually fall as the column correlations become stronger. In the adjacent-array scenario under high row correlation, our test's power falls when ρ increases from 0.800 to 0.888. This happens because when ρ becomes very large, many columns become weakly correlated, and these small correlations widen the center of the $z(\hat{\Delta}_{jj'})$ histogram.

Our test seems to be able to detect the kinds of column dependence seen in real data sets. Figure 3 shows that our test strongly rejects on the data sets considered by Efron (2009a). We can gain additional power by combining our test with Efron's permutation test. Since our test is permutation-invariant, the two methods yield independent p-values.

Code to perform our test is available in the R package "colcor," on CRAN.

TABLE 3

Power to detect adjacency effect at level 0.05. The columns have correlation $\rho^{|j-j'|}$, where ρ is given above. To interpret the effect size, note that the transformed correlations have standard deviation approximately $\alpha^{\frac{1}{2}}$. Power was calculated over 1000 simulations, so the standard deviation is less than 0.016, and much lower near 0 and 1

Adjacent column cor.	0.080	0.169	0.259	0.348	0.439	0.529	0.620	0.710	0.800	0.888
$\alpha^{\frac{1}{2}} = 0.020$	0.314	1	1	1	1	1	1	1	1	1
$\alpha^{\frac{1}{2}} = 0.128$	0.009	0.014	0.099	0.770	0.998	1	1	1	1	0.979
$\alpha^{\frac{1}{2}} = 0.325$	0	0	0	0	0	0.010	0.052	0.204	0.392	0.297

TABLE 4

Power to detect contamination effect under low, medium, and high row correlation at level 0.05. The within block correlations given are after standardization; about 0.82% of the column correlations had this value, and the rest were small. To interpret the effect size, note that the transformed correlations have standard deviation approximately $\alpha^{\frac{1}{2}}$. Power was calculated over 1000 simulations, so the standard deviation is less than 0.016, and much lower near 0 and 1

Contaminated block cor.	0.067	0.147	0.230	0.314	0.402	0.492	0.585	0.681	0.780	0.881
$\alpha^{\frac{1}{2}} = 0.020$	0.005	0.912	1	1	1	1	1	1	1	1
$\alpha^{\frac{1}{2}} = 0.128$	0.009	0.007	0.016	0.104	0.447	0.867	0.998	1	1	1
$\alpha^{\frac{1}{2}} = 0.325$	0	0	0	0	0.001	0.006	0.019	0.067	0.300	0.796

Acknowledgements

The author thanks Professor Bradley Efron for many very helpful discussions and suggestions, and the referees, whose comments clarified and improved the paper.

Appendix A: Proofs

A.1. Proof of Lemma 1

Y_{ij} is centered, approximately unit variance, and symmetric, so we only need to approximate the fourth moment. Since X is normal and $\Delta_{jj} = 1$, it is easy to show that

$$\begin{aligned}
E\left(\frac{1}{n}\sum X_{ij}^2\right) &= \Sigma_{ii} \\
E\left(\left(\frac{1}{n}\sum X_{ij}^2\right)^2\right) &= \Sigma_{ii}^2\left(1 + 2\frac{\|\Delta\|_F^2}{n^2}\right) \\
E\left(X_{ij}^2\right) &= \Sigma_{ii} \\
E\left(X_{ij}^4\right) &= 3\Sigma_{ii}^2 \\
E\left(X_{ij}^2\left(\frac{1}{n}\sum X_{ij}^2\right)\right) &= \Sigma_{ii}^2\left(1 + \frac{2}{n}\sum_{j'}\Delta_{jj'}^2\right)
\end{aligned}$$

Using the Taylor series approximation

$$E\left(\frac{x}{y}\right) \approx \frac{E\left(x\right)}{E\left(y\right)}+\begin{pmatrix} 1/E\left(y\right) \\ -E\left(x\right)/E\left(y\right)^2 \end{pmatrix}' \begin{pmatrix} var\left(x\right) & cov\left(x,y\right) \\ cov\left(x,y\right) & var\left(y\right) \end{pmatrix} \begin{pmatrix} 1/E\left(y\right) \\ -E\left(x\right)/E\left(y\right)^2 \end{pmatrix},$$

we have

$$\begin{aligned} E\left(Y_{ij}^4\right) &\approx 1+\frac{1}{\Sigma_{ii}^2}\begin{pmatrix} 1 \\ -1 \end{pmatrix}' \begin{pmatrix} 2\Sigma_{ii}^2 & \Sigma_{ii}^2\left(\frac{2}{n}\sum_{j'}\Delta_{jj'}^2\right) \\ \Sigma_{ii}^2\left(\frac{2}{n}\sum_{j'}\Delta_{jj'}^2\right) & \Sigma_{ii}^2\left(2\frac{\|\Delta\|_F^2}{n^2}\right) \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ &= 3-\frac{4}{n}\sum_{j'}\Delta_{jj'}^2+2\frac{\|\Delta\|_F^2}{n^2} \end{aligned}$$

which yields the kurtosis approximation $E(Y_{ij}^4)-3 \approx -\frac{4}{n}\sum_{j'}\Delta_{jj'}^2+2\frac{\|\Delta\|_F^2}{n^2}$.

A.2. Proof of Lemma 2

Clearly Z is normal with mean 0. We then have

$$\begin{aligned} cov\left(Z_i, Z_{i'}\right) &= cov\left(\sum_j X_{ij}w_j, \sum_{j'} X_{i'j'}w_{j'}\right) \\ &= \sum_{j,j'} w_j w_{j'}\Delta_{jj'}\Sigma_{ii'} \\ &= \left(w'\Delta w\right)\Sigma_{ii'}. \end{aligned}$$

A.3. Proof of Lemma 3

The argument is the same for $\hat{\alpha}_1$ and $\hat{\alpha}_2$, so we will only prove the result for $\hat{\alpha}_2$. Olkin and Pratt (1958) show that for a bivariate normal with correlation ρ, $f_2(\hat{\rho})$ is UMVU for ρ^2. $(X_{ij}, X_{i'j}), j=1,\ldots,n$ are iid bivariate normal with correlation $\rho_{ii'}$, so $E(f_2(\hat{\rho}_{ii'}))=\rho_{ii'}^2$. Summing this equality proves $\hat{\alpha}_2$ is unbiased for α_2.

Now, $\hat{\rho}_{ii'}$ is complete sufficient for $\rho_{ii'}$, so $f_2(\hat{\rho}_{ii'})$ is still UMVU for $\rho_{ii'}$ in the multivariate normal situation. This means it is uncorrelated with any unbiased estimator of 0. Since $\hat{\alpha}_2$ is a linear combination of these, it is also uncorrelated with any unbiased estimator of 0, and is thus UMVU for α_2.

The approximations come from fixing the hypergeometric functions at their maximum at $r=0$. For α_1, we have

$$|\tilde{\alpha}_1-\hat{\alpha}_1| \leq \left(\frac{1}{\binom{N}{2}}\sum_{i<i'}|\hat{\rho}_{ii'}|\right)\left(\frac{\Gamma(\frac{n-1}{2})\Gamma(\frac{n-3}{2})}{\Gamma(\frac{n-2}{2})^2}-1\right)$$

and bounding $F\left(1-r^2;\frac{1}{2},\frac{1}{2};\frac{n-1}{2}\right)$ above and below using its values at $r=0,1$,

$$\left|\frac{1}{\binom{N}{2}}\sum \hat{\rho}_{ii'}\right| \leq |\tilde{\alpha}_1| \leq |\hat{\alpha}_1|\,.$$

Similarly, since $F(1-r^2;1,1,\frac{n+1}{2})$ is decreasing and concave in r, $F(1-r^2;1,1,\frac{n+1}{2})-F(1,1,\frac{n+1}{2},1)$ is in $[-\frac{2r^2}{n-3},0]$, giving

$$|\tilde{\alpha}_2-\hat{\alpha}_2| \quad \leq \quad \frac{2}{n-3}\frac{n-2}{n-1}\left(\frac{1}{\binom{N}{2}}\sum_{i<i'}\hat{\rho}_{ii'}\left(1-\hat{\rho}^2_{ii'}\right)\right)$$

and

$$-\frac{1}{n-1}+\frac{n-2}{n-1}\left(\frac{1}{\binom{N}{2}}\sum_{i<i'}\hat{\rho}^2_{ii'}\right) \geq \tilde{\alpha}_2 \geq \hat{\alpha}_2. \tag{A.1}$$

A.4. Proof of Lemma 5

Since $\Sigma_{ii}=\Delta_{jj}=1$, we can approximate the correlations by covariances: $\hat{\rho}_{ii'}\approx\hat{\sigma}_{ii'}$. Applying Lemma 4 shows that

$$E\left(\frac{1}{\binom{N}{2}}\sum_{i<i'}\hat{\rho}^2_{ii'}\right) \quad = \quad \left(1+\frac{\|\Delta\|_F^2}{n^2}\right)\alpha_2+\frac{\|\Delta\|_F^2}{n^2}$$

so

$$\begin{aligned} E\left(\hat{\alpha}\right) \quad &= \quad \left(\frac{\|\Delta\|_F^2}{n^2}-\frac{1}{n-3}\right)+\frac{n-1}{n-3}\left(1+\frac{\|\Delta\|_F^2}{n^2}\right)\alpha_2 \\ &= \quad \alpha+\delta\left(\alpha+1\right)+O\left(n^{-2}\right). \end{aligned}$$

Similarly, we can show that the expectation of the upper bound in equation A.1 is $(-\frac{1}{n-1}+\frac{\|\Delta\|_F^2}{n})+\frac{n-2}{n-1}(1+\frac{\|\Delta\|_F^2}{n^2})\alpha=\alpha+\delta(\alpha+1)+O(n^{-2})$, so $E(\tilde{\alpha})=\alpha+\delta(\alpha+1)+O(n^{-2})$. The exact expectations of the bounds give exact bounds for $E(\tilde{\alpha})$. The proof for α_1 is similar.

A.5. Proof of Lemma 6

Let D be the sample row covariance matrix of the centered, column standardized matrix (so $\Delta_{jj}=1$ but not necessarily Σ_{ii}). Using the argument of Lemma 4, D has the mean and covariance of a $(\frac{\|\Delta\|_F^2}{n^2})^{-1}Wishart(\Delta,\frac{\|\Delta\|_F^2}{n^2})$ random matrix. Now apply the delta method to get the variances of $\hat{\rho}_{ii'}$ as a function of D. The only effect of dependence is to replace $\frac{1}{N}$ with $\frac{\|\Sigma\|_F^2}{N^2}$, so the delta method under independence yields the result.

A.6. Proof of Lemma 7

We first approximate the third and fourth moments of the sample correlations. The only approximation we make is to replace the standard deviations denominator by their expectation, 1, yielding $n\hat{\rho}_{ii'} = \sum_j x_{ij}x_{i'j}$.

Consider the third moment. We have

$$E\left((n\hat{\rho}_{ii'})^3\right) = \sum_{j,j',j''} E\left(x_{ij}x_{i'j}x_{ij'}x_{i'j'}x_{ij''}x_{i'j''}\right).$$

Isserlis' theorem (Isserlis, 1918) for normal moments says that

$$E\left(x_{ij}x_{i'j}x_{ij'}x_{i'j'}x_{ij''}x_{i'j''}\right) = \sum\prod E\left(y_1y_2\right)$$

where $\sum\prod$ means we sum over all ways to partition the x's into pairs, and multiply the expectations for each partition. This representation, with a few computational tricks, lets us find

$$E\left((n\hat{\rho}_{ii'})^3\right) = \Sigma_{ii'}^3 + \left(3\Sigma_{ii'} + 3\Sigma_{ii'}^3\right) ntr\left(\Delta^2\right) + tr\left(\Delta^3\right)\left(6\Sigma_{ii'} + 2\Sigma_{ii'}^3\right)$$

and,

$$\begin{aligned} E\left((n\hat{\rho}_{ii'})^4\right) &= \left(6 + 36\Sigma_{ii'}^2 + 6\Sigma_{ii'}^4\right) tr\left(\Delta^4\right) + \left(3 + 6\Sigma_{ii'}^2 + 3\Sigma_{ii'}^4\right) tr\left(\Delta^2\right)^2 \\ &\quad + \left(24\Sigma_{ii'}^2 + 8\Sigma_{ii'}^4\right) ntr\left(\Delta^3\right) + \left(6\Sigma_{ii'}^2 + 6\Sigma_{ii'}^4\right) n^2 tr\left(\Delta^2\right) + n^4\Sigma_{ii'}^4. \end{aligned}$$

Expanding the skewness and kurtosis in terms of the raw moments and some algebra completes the proof.

References

Genevera Allen and Robert Tibshirani. Inference with transposable data: Modeling the effects of row and column correlations. 2010.

James J. Chen, Robert R. Delongchamp, Chen-An Tsai, Huey-miin Hsueh, Frank Sistare, Karol L. Thompson, Varsha G. Desai, and James C. Fuscoe. Analysis of variance components in gene expression data. *Bioinformatics*, 20(9):1436–1446, 2004. DOI 10.1093/bioinformatics/bth118. URL http://bioinformatics.oxfordjournals.org/content/20/9/1436.abstract.

Elissa Cosgrove, Timothy Gardner, and Eric Kolaczyk. On the choice and number of microarrays for transcriptional regulatory network inference. *BMC Bioinformatics*, 11(1):454, 2010. ISSN 1471-2105. DOI 10.1186/1471-2105-11-454. URL http://www.biomedcentral.com/1471-2105/11/454.

David Donoho and Jiashun Jin. Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, 32(3):962–994, 2004. ISSN 00905364. URL http://www.jstor.org/stable/3448581. MR2065195

Bradley Efron. Are a set of microarrays independent of each other? *Annals of Applied Statistics*, 3, 2009a.

Bradley Efron. Correlated z-values and the accuracy of large-scale statistical estimates. 2009b.

T.R. Golub, D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, J.R. Downing, M.A. Caligiuri, C.D. Bloomfield, and E.S. Lander. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science*, 286(5439):531–537, 1999. DOI 10.1126/science.286.5439.531. URL http://www.sciencemag.org/cgi/content/abstract/286/5439/531.

I. Hedenfalk, D. Duggan, Y.D. Chen, M. Radmacher, M. Bittner, R. Simon, P. Meltzer, B. Gusterson, M. Esteller, O.P. Kallioniemi, B. Wilfond, A. Borg, J. Trent, M. Raffeld, Z. Yakhini, A. Ben-Dor, E. Dougherty, J. Kononen, L. Bubendorf, W. Fehrle, S. Pittaluga, and Gruvberg. Gene-expression profiles in hereditary breast cancer. *New England Journal of Medicine*, 344:539–548, 2001. URL http://dx.doi.org/10.1056/NEJM200102223440801.

L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12(1/2):134–139, 1918. ISSN 00063444. URL http://www.jstor.org/stable/2331932.

Ingram Olkin and John W. Pratt. Unbiased estimation of certain correlation coefficients. *Annals of Mathematical Statistics*, 29(1):201–211, 1958. MR0093854

Richard A. Olshen and Bala Rajaratnam. Successive normalization of rectangular arrays. *The Annals of Statistics*, 38:1638–1664, 2010. MR2662355

Art B. Owen. Variance of the number of false discoveries. *Journal of the Royal Statistical Society, Series B*, 67:411–426, 2005. MR2155346

Matthew D. W. Piper, Pascale Daran-Lapujade, Christoffer Bro, Birgitte Regenberg, Steen Knudsen, Jens Nielsen, and Jack T. Pronk. Reproducibility of oligonucleotide microarray transcriptome analyses. *Journal of Biological Chemistry*, 277(40):37001–37008, 2002. DOI 10.1074/jbc.M204490200. URL http://www.jbc.org/content/277/40/37001.abstract.

Xing Qiu, Andrew Brooks, Lev Klebanov, and Andrei Yakovlev. The effects of normalization on the correlation structure of microarray data. *BMC Bioinformatics*, 6(1):120, 2005. ISSN 1471-2105. DOI 10.1186/1471-2105-6-120. URL http://www.biomedcentral.com/1471-2105/6/120.

Dinesh Singh, Phillip G. Febbo, Kenneth Ross, Donald G. Jackson, Judith Manola, Christine Ladd, Pablo Tamayo, Andrew A. Renshaw, Anthony V. D'Amico, Jerome P. Richie, Eric S. Lander, Massimo Loda, Philip W. Kantoff, Todd R. Golub, and William R. Sellers. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 1(2):203–209, 2002. ISSN 1535-6108. DOI 10.1016/S1535-6108(02)00030-2. URL http://www.sciencedirect.com/science/article/B6WWK-45J85YN-F/2/b0c5e920001196813bd1821d0191f4b9.