Let d_1 & d_2 be the expection of the numerator of the t statistic for genes 1 & 2, resp.
Let $Z_1 + d_1$ and $Z_2 + d_2$ be the numerators of the t-statistics with $E(Z_1)=0$, $E(Z_2)=0$.
The goal is to find $r = Cor\left(\frac{Z_1+d_1}{S_1}, \frac{Z_2+d_2}{S_2}\right)$

Assuming $S_1 \perp\!\!\!\perp Z_1$ and $S_2 \perp\!\!\!\perp Z_2$, it can be shown that

$$r = \frac{d_1 d_2 \, Cov\left(\frac{1}{S_1}, \frac{1}{S_2}\right) + E\left(\frac{Z_1 Z_2}{S_1 S_2}\right)}{\sqrt{\left[1 + d_1^2 \, var\left(\frac{1}{S_1}\right)\right]\left[1 + d_2^2 \, var\left(\frac{1}{S_2}\right)\right]}}$$

For the rest of this document, our goal will be to find ① $E\left(\frac{Z_1 Z_2}{S_1 S_2}\right)$ ② $var\left(\frac{1}{S_1}\right)$ and $var\left(\frac{1}{S_2}\right)$

③ $cov\left(\frac{1}{S_1}, \frac{1}{S_2}\right)$.

Assumptions: a) Normality (praitcally this may not be much stronger that $S_1 \perp\!\!\!\perp Z_1$ & $S_2 \perp\!\!\!\perp Z_2$):

b) The covariance matrix for the two genes is

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}, \quad \sigma_{12} = \sigma_{21} \quad \rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

c) For notational simplicity, we here consider a one-sample problem: $\frac{Z_i + d_i}{S_i}$ is the one-sample t-statistic, where

Z_i i.i.d is the sample mean, $S_1 = \hat{\sigma}_1 / \sqrt{n}$, where n is the sample size.

to be discussed later: Note: with a two-sample problem (or some other design) the derivation will still hold b/c the only thing needs to change is the d.f., which will cancel out as $n \to \infty$ $\square$

Theorem 1

$$n^{3/2} \begin{pmatrix} S_1^2 - \sigma_1^2/n \\ S_2^2 - \sigma_2^2/n \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2\sigma_1^4 & 2\sigma_{12}^2 \\ 2\sigma_{12}^2 & 2\sigma_2^4 \end{pmatrix} \right)$$

This follows from the following Lemma

Lemma 1

$$\sqrt{n} \begin{pmatrix} \hat{\sigma}_1^2 - \sigma_1^2 \\ \hat{\sigma}_2^2 - \sigma_2^2 \end{pmatrix} \xrightarrow{D} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2\sigma_1^4 & 2\sigma_{12}^2 \\ 2\sigma_{12}^2 & 2\sigma_2^4 \end{pmatrix} \right)$$

Pf (brief): Normality follows from properties of MLE & asymptotic unbiasedness.

The asymptotic ~~var~~ covariance can be derived by looking at the information matrix. Alternatively, one can rely on the following fact:

If $X_i = (X_1^i, \cdots, X_p^i) \overset{i.i.d}{\sim} MVN(\mu, \Sigma_X)$, $i = 1 \cdots n$

then $\underbrace{\sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T}_{p \times p \text{ matrix}} \sim \underbrace{W_p(\Sigma, n-1)}_{\text{Wishart distribution with } n-1 \text{ degrees of freedom.}}$

It is helpful to keep in mind that the MGF of $W_p(\Sigma, n-1)$ is given by $m(\Theta) = \det(I - 2\Theta\Sigma)^{-\frac{n-1}{2}}$ $\square$

Theorem 1 combined by Delta method yields

Theorem 2 $var(\frac{1}{S_1}) \to \frac{1}{2\sigma_1^2}$ $var(\frac{1}{S_2}) \to \frac{1}{2\sigma_2^2}$

$Cov(\frac{1}{S_1}, \frac{1}{S_2}) \to \frac{\rho^2}{2\sigma_1\sigma_2}$

$E(\frac{1}{nS_1S_2}) \to \frac{1}{\sigma_1\sigma_2}$

It follows that $E\left(\frac{Z_1Z_2}{S_1S_2}\right) = E(nZ_1Z_2)E\left(\frac{1}{nS_1S_2}\right)$

$\to \left(n \cdot \frac{\rho\sigma_1\sigma_2}{n}\right)\left(\frac{1}{\sigma_1\sigma_2}\right) = \rho.$

Therefore, $r = \frac{\rho + \frac{1}{2}\rho^2 R_1R_2}{\sqrt{(1+\frac{1}{2}R_1^2)(1+\frac{1}{2}R_2^2)}}$ where $R_1 = \frac{d_1}{\sigma_1}$, $R_2 = \frac{d_2}{\sigma_2}$

Two-sample t-test

With a two-sample analysis, the derivation needs to be changed only with regard the following aspects;

(1) One degree of freedom will be lost for $\hat{\sigma}_1$ & $\hat{\sigma}_2$ (and hence S_1 & S_2). But this won't affect the conclusion as $n \to \infty$.

(2) d_i (defined to be E(numerator of t statistic)) needs to be replaced by $\Delta_i/2$, where Δ_i is the DE effect of gene i.

It can be shown (details omitted) that

$$r = \frac{\rho + \frac{1}{8}\rho^2 R_1 R_2}{\sqrt{(1+\frac{1}{8}R_1^2)(1+\frac{1}{8}R_2^2)}}, \quad R_i = \frac{\Delta_i}{\sigma_i}, \quad i = 1, 2$$