For a gene i, suppose

$$1. Expression_A = x_i$$

$$2. Expression_B = y_i$$

fold change (FC) between sample AB will be

$$FC_i^{AB} = \frac{x_i}{y_i} = \pi_i^{(1)}$$

$$x_i = \pi_i^{(1)} y_i$$

if true mix ratio is 3:1, and fold change of CD of gene i will be

$$FC_i^{CD} = \frac{3x_i + y_i}{3y_i + x_i} = \frac{3\pi_i^{(1)} y_i + y_i}{3y_i + \pi_i^{(1)} y_i} = \frac{3\pi_i^{(1)} + 1}{3 + \pi_i^{(1)}} = \pi_i^{(2)}$$

But mix ratio might not be exactly 3 and with simply assuming that is exactly 3 will affect the standard error. We have to use regressions to estimate the true mix ratio in samples, since we know $\pi_i^{(1)}$ and $\pi_i^{(2)}$ for every genes from our data, it is just a linear regression. Substituting theoretical mix ratio 3 by true mix ratio α,

$$\frac{\alpha_i \pi_i^{(1)} + 1}{\alpha_i + \pi_i^{(1)}} = \pi_i^{(2)}$$

For all genes,

$$\frac{\alpha \pi^{(1)} + 1}{\alpha + \pi^{(1)}} = \pi^{(2)}$$

$$a(\pi^{(1)} - \pi^{(2)}) + 1 = \pi^{(1)} \pi^{(2)}$$

This is the same format as

$$ax + b + \epsilon = y$$

where $x = \pi^{(1)} - \pi^{(2)}$, $y = \pi^{(1)} \pi^{(2)}$ and ϵ is the error term we are comparing between TGIRT-seq and TruSeq3. The predicted FC^{CD} would be

$$\pi_{predicted}^{(2)} = \frac{a\pi^{(1)} + 1}{a + \pi^{(1)}}$$

and R^2 can be computed as

$$R^2 = 1 - \frac{\sum_i (\pi_{predicted}^{(2)} - \pi_i^{(2)})^2}{\sum_i (\pi_i^{(2)} - \bar{\pi}^{(2)})^2} \bullet = 1 - \frac{\sum_i \epsilon_i^2}{\sum_i (\pi_i^{(2)} - \bar{\pi}^{(2)})^2}$$