Goal

Does local ancestry affect the gene network?

Two nodes and their correlations

Consider two genes k_1, k_2 and the corresponding gene expression vector of y_{k_1} and y_{k_2}, each of length n where n is the number of subjects. We also have two local ancestry vectors x_{k_1} and x_{k_2}. The temporary question of interest is whether the correlation between y_{k_1} and y_{k_2} is driven by either x_{k_1} or x_{k_2}.

Without loss of generality, let's see how much the correlation between the two expression levels changes with respect to the values of x_{k_1}. Each element of the local ancestry vector x takes one of three values : 0, 1, or 2. When we divide the n samples into three groups based on local ancestry each with size n_0, n_1, n_2, we can observe three correlation coefficients: ρ_0, ρ_1, and ρ_2. Using Fisher's transformation, we consider three normally distributed z values for each group : z_0, z_1, and z_2.

$$z_i = \frac{1}{2}ln\left(\frac{1+\rho_i}{1-\rho_i}\right) \sim N\left(\frac{1}{2}ln\left(\frac{1+\tilde{\rho}_i}{1-\tilde{\rho}_i}\right), \frac{1}{\sqrt{n_i-3}}\right)$$

where $\tilde{\rho}_i$ is the underlying true correlation coefficient for the group with local ancestry i.

We assume that z_0, z_1, and z_2 have linear relationship. We formulate the model like below.

$$\begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix} \sim MVN\left(\begin{bmatrix} \mu \\ \mu+\beta \\ \mu+2\beta \end{bmatrix}, \begin{bmatrix} \frac{1}{n_0-3} & 0 & 0 \\ 0 & \frac{1}{n_1-3} & 0 \\ 0 & 0 & \frac{1}{n_2-3} \end{bmatrix}\right)$$

with hypotheses

$$H_0 : \beta = 0$$

$$H_A : \beta \neq 0$$

For notational convenience, we define the following matrix and vectors.

$$\Sigma = \begin{bmatrix} \frac{1}{n_0-3} & 0 & 0 \\ 0 & \frac{1}{n_1-3} & 0 \\ 0 & 0 & \frac{1}{n_2-3} \end{bmatrix} = \begin{bmatrix} \sigma_0^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_2^2 \end{bmatrix}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu \\ \mu \\ \mu \end{bmatrix}, \quad \boldsymbol{\mu_\beta} = \begin{bmatrix} \mu \\ \mu+\beta \\ \mu+2\beta \end{bmatrix}, \quad \boldsymbol{z} = \begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix}$$

Also for notational convenience, we introduce Φ and Λ.

$$\Phi = \sigma_0^2\sigma_1^2 + \sigma_1^2\sigma_2^2 + \sigma_2^2\sigma_0^2$$

$$\Lambda = \sigma_1^2\sigma_2^2 z_0 + \sigma_0^2\sigma_2^2 z_1 + \sigma_0^2\sigma_1^2 z_2$$

Now, we set up the log likelihood of the two cases: when β is 0 and when β is the MLE estimator.

$$\ell_{\beta=0} = -\frac{1}{2}log|2\pi\Sigma| - \frac{1}{2}((\boldsymbol{z}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{z}-\boldsymbol{\mu})) \tag{1}$$

$$\frac{\partial \ell_{\beta=0}}{\partial \mu} = \frac{z_0-\mu}{\sigma_0^2} + \frac{z_1-\mu}{\sigma_1^2} + \frac{z_2-\mu}{\sigma_2^2} = 0 \tag{2}$$

$$\Rightarrow \hat{\mu}_{0,MLE} = \frac{\sigma_1^2\sigma_2^2 z_0 + \sigma_2^2\sigma_0^2 z_1 + \sigma_0^2\sigma_1^2 z_2}{\sigma_0^2\sigma_1^2 + \sigma_1^2\sigma_2^2 + \sigma_2^2 + \sigma_0^2} = \frac{\Lambda}{\Phi} \tag{3}$$

is the MLE estimator for the μ under the null hypothesis.

For the alternative hypothesis,

$$\ell_{\beta\neq 0} = -\frac{1}{2}log|2\pi\Sigma| - \frac{1}{2}((\boldsymbol{z}-\boldsymbol{\mu_\beta})^T\Sigma^{-1}(\boldsymbol{z}-\boldsymbol{\mu_\beta}))$$
$$\frac{\partial \ell_{\beta\neq 0}}{\partial\mu} = \frac{z_0-\mu}{\sigma_0^2} + \frac{z_1-\mu-\beta}{\sigma_1^2} + \frac{z_2-\mu-2\beta}{\sigma_2^2}$$
$$\Rightarrow \hat{\mu}_{A,MLE} = \frac{\sigma_1^2\sigma_2^2 z_0 + \sigma_2^2\sigma_0^2(z_1-\beta) + \sigma_0^2\sigma_1^2(z_2-2\beta)}{\sigma_0^2\sigma_1^2+\sigma_1^2\sigma_2^2+\sigma_2^2\sigma_0^2} = \frac{\Lambda - \sigma_2^2\sigma_0^2\beta - 2\sigma_0^2\sigma_1^2\beta}{\Phi}$$

is the MLE estimator for μ under the alternative hypothesis. Now we plug in the above to get the MLE estimator for β.

$$\frac{\partial \ell_{\beta\neq 0}}{\partial\beta} = \frac{z_1-\mu-\beta}{\sigma_1^2} + \frac{2(z_2-\mu-2\beta)}{\sigma_2^2}$$
$$\Rightarrow \sigma_2^2(z_1-\hat{\mu}_{A,MLE}-\hat{\beta}_{A,MLE}) + 2\sigma_1^2(z_2-\hat{\mu}_{A,MLE}-2\hat{\beta}_{A,MLE}) = 0$$
$$\Rightarrow \frac{\Phi z_1 - \Lambda + (\sigma_0^2\sigma_2^2 + 2\sigma_0^2\sigma_1^2 - \Phi)\hat{\beta}_{A,MLE}}{\Phi\sigma_1^2} + \frac{2(\Phi z_2 - \Lambda + (\sigma_0^2\sigma_2^2 + 2\sigma_0^2\sigma_1^2 - 2\Phi)\hat{\beta}_{A,MLE})}{\Phi\sigma_2^2} = 0$$
$$\Rightarrow \hat{\beta}_{A,MLE} = \frac{(\sigma_2^2+2\sigma_1^2)\Lambda - (\sigma_2^2 z_1 + 2\sigma_1^2 z_2)\Phi}{(\sigma_2^2+2\sigma_1^2)(\sigma_0^2\sigma_2^2+2\sigma_0^2\sigma_1^2-2\Phi)}$$

Now we can perform the likelihood ratio test and see if β is significantly different from 0. Under the null,

$$-2log\left(\frac{N_3(\boldsymbol{z};\hat{\boldsymbol{\mu}},\Sigma)}{N_3(\boldsymbol{z};\hat{\boldsymbol{\mu}}_{\boldsymbol{\beta}},\Sigma)}\right) \sim \chi_1^2$$

and we can compute the corresponding p-value.