

eQTL epistasis: detecting epistatic effects and
inferring hierarchical relationships of genes in
biological pathways

Mingon Kang, Chunling Zhang, Hyung-Wook Chun, Chris Ding,
Chunyu Liu, and Jean Gao

Proposition 2.1. Let $\gamma_i = y_i - \beta_1 x_{i1} - \beta_2 x_{i2}$. Given $\beta_1, \beta_2, \beta_{12}$, and r_2 , the optimal genotypic effect size r_1 is given by,

$$r_1 = \operatorname{argmin}_{s_i \in \mathcal{S}} F(\mathbf{X}, \mathbf{y}; \boldsymbol{\beta}, s_i, r_2), \quad (1)$$

where a set \mathcal{S} is defined as,

$$\begin{aligned} \mathcal{S} &= \left\{ \{0, s_1, \dots, s_k\} \mid d_h s^h + d_{h-1} s^{h-1} + \dots + d_1 s^1 + d_0 = 0, \right. \\ d_I &= \begin{cases} \sum_{i=1}^n -\gamma_i \beta_{12} r_2^{\varphi(x_{i2})} \varphi(x_{i1}) & \text{if } \varphi(x_{i1}) - 1 = I \\ \sum_{i=1}^n \beta_{12}^2 r_2^{2\varphi(x_{i2})} \varphi(x_{i1}) & \text{if } 2\varphi(x_{i1}) - 1 = I \\ 0 & \text{if } \varphi(x_{i1}) = 0, \end{cases} \\ &\left. s \geq 0, \quad s \in \mathfrak{R}, \quad 0 \leq I \leq h, \quad 1 \leq k \leq h \right\}. \end{aligned} \quad (2)$$

r_2 is given by,

$$r_2 = \operatorname{argmin}_{t_j \in \mathcal{T}} F(\mathbf{X}, \mathbf{y}; \boldsymbol{\beta}, r_1, t_j), \quad (3)$$

where a set \mathcal{T} is defined as,

$$\begin{aligned} \mathcal{T} &= \left\{ \{0, t_1, \dots, t_k\} \mid d_h t^h + d_{h-1} t^{h-1} + \dots + d_1 t^1 + d_0 = 0, \right. \\ d_I &= \begin{cases} \sum_{i=1}^n -\gamma_i \beta_{12} r_1^{\varphi(x_{i1})} \varphi(x_{i2}) & \text{if } \varphi(x_{i2}) - 1 = I \\ \sum_{i=1}^n \beta_{12}^2 r_1^{2\varphi(x_{i1})} \varphi(x_{i2}) & \text{if } 2\varphi(x_{i2}) - 1 = I \\ 0 & \text{if } \varphi(x_{i2}) = 0, \end{cases} \\ &\left. t \geq 0, \quad t \in \mathfrak{R}, \quad 0 \leq I \leq h, \quad 1 \leq k \leq h \right\}. \end{aligned} \quad (4)$$

Proof. By the Karush-Juhn-Tucker (KKT) complementarity slackness condition for the constraint, $r_1 \geq 0$, we consider the equation for the optimum of r_1 ,

$$\frac{\partial F(\mathbf{X}, \mathbf{y}; \boldsymbol{\beta}, \mathbf{r})}{\partial r_1} r_1 = 0. \quad (5)$$

The optimum exists when $\frac{\partial F(\mathbf{X}, \mathbf{y}; \boldsymbol{\beta}, \mathbf{r})}{\partial r_1} = 0$ or $r_1 = 0$. The formal $\frac{\partial F(\mathbf{X}, \mathbf{y}; \boldsymbol{\beta}, \mathbf{r})}{\partial r_1}$ is,

$$\begin{aligned} &= \sum_{i=1}^n (\gamma_i - \beta_{12} r_1^{\varphi(x_{i1})} r_2^{\varphi(x_{i2})}) (-\beta_{12} r_2^{\varphi(x_{i2})} \varphi(x_{i1}) r_1^{\varphi(x_{i1})-1}) \\ &= \sum_{i=1}^n -\gamma_i \beta_{12} r_2^{\varphi(x_{i2})} \varphi(x_{i1}) r_1^{\varphi(x_{i1})-1} + \\ &\quad \sum_{i=1}^n \beta_{12}^2 r_2^{2\varphi(x_{i2})} \varphi(x_{i1}) r_1^{2\varphi(x_{i1})-1} = 0. \end{aligned} \quad (6)$$

This is a polynomial function of r_1 . The polynomial function forms,

$$d_h r_1^h + d_{h-1} r_1^{h-1} + \dots + d_1 r_1 + d_0 = 0, \quad (7)$$

where the coefficient of each degree of the function is denoted as,

$$d_I = \begin{cases} \sum_{i=1}^n -\gamma_i \beta_{12} r_2^{\varphi(x_{i2})} \varphi(x_{i1}) & \text{if } \varphi(x_{i1}) - 1 = I \\ \sum_{i=1}^n \beta_{12}^2 r_2^{2\varphi(x_{i2})} \varphi(x_{i1}) & \text{if } 2\varphi(x_{i1}) - 1 = I \\ 0 & \text{if } \varphi(x_{i1}) = 0. \end{cases} \quad (8)$$

Let $\mathbf{S} = \{s_1, \dots, s_k\}$ be roots of the polynomial function, where $1 \leq k \leq h$ and $s_i \in \mathfrak{R}$. Among these h roots of Eq. (7), only those which satisfy $s_i \geq 0$ are admissible. We assume that there are k admissible roots. The roots of the polynomial functions can be easily estimated by computing the eigenvalues of the companion matrix. According to KKT condition of Eq. (5), $r_1 = 0$ is an admissible solution, i.e., $\mathbf{S} = \{0, s_1, \dots, s_k\}$. Therefore, the element of \mathbf{S} , $s_i \in \mathbf{S}$, that minimizes the score function of Eq. (2) is considered as the optimal solution,

$$r_1 = \underset{s_i \in \mathbf{S}}{\operatorname{argmin}} F(\mathbf{X}, \mathbf{y}; \boldsymbol{\beta}, s_i, r_2). \quad (9)$$

The optimal solution of r_2 is derived by the same way. Thus, we finish the proof. \square

Proposition 2.2. *Given r_1 and r_2 , the optimal parameters, β_{12} , β_1 , and β_2 , are given by,*

$$\beta_{12} = \sum_{i=1}^n \frac{\gamma_i r_1^{\varphi(x_{i1})} r_2^{\varphi(x_{i2})}}{r_1^{2\varphi(x_{i1})} r_2^{2\varphi(x_{i2})} + \frac{\lambda}{w_i}} \quad (10)$$

$$\beta_1 = \sum_{i=1}^n \frac{y_i x_{i1} - \beta_2 x_{i1} x_{i2} - \beta_{12} x_{i1} r_1^{\varphi(x_{i1})} r_2^{\varphi(x_{i2})}}{x_{i1}^2 + \frac{\lambda}{w_i}} \quad (11)$$

$$\beta_2 = \sum_{i=1}^n \frac{y_i x_{i2} - \beta_2 x_{i1} x_{i2} - \beta_{12} x_{i2} r_1^{\varphi(x_{i1})} r_2^{\varphi(x_{i2})}}{x_{i2}^2 + \frac{\lambda}{w_i}}. \quad (12)$$

Proof. The derivatives of F with respect to β_{12} , β_1 , and β_2 are

$$\begin{aligned}\frac{\partial F}{\partial \beta_1} &= \sum_{i=1}^n -\frac{1}{w_i} (\gamma_i - \beta_{12} r_1^{\varphi(x_{i1})} r_2^{\varphi(x_{i2})}) x_{i1} + \lambda \beta_1 = 0 \\ \frac{\partial F}{\partial \beta_2} &= \sum_{i=1}^n -\frac{1}{w_i} (\gamma_i - \beta_{12} r_1^{\varphi(x_{i1})} r_2^{\varphi(x_{i2})}) x_{i2} + \lambda \beta_2 = 0 \\ \frac{\partial F}{\partial \beta_{12}} &= \sum_{i=1}^n -\frac{1}{w_i} (\gamma_i - \beta_{12} r_1^{\varphi(x_{i1})} r_2^{\varphi(x_{i2})}) r_1^{\varphi(x_{i1})} r_2^{\varphi(x_{i2})} \\ &\quad + \lambda \beta_{12} = 0.\end{aligned}$$

The solutions can be easily induced from the above equations. \square

Algorithm 1 Generic epistasis model

Require: \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{y} , λ , and $\varphi(x)$

- 1: Initialize the parameters $\beta_1, \beta_2, \beta_{12}, r_1, r_2$
 - 2: **repeat**
 - 3: Estimate the parameters r_1 and r_2 by Proposition. **2.1**.
 - 4: Estimate the parameters β_{12}, β_1 , and β_2 by Proposition. **2.2**.
 - 5: **until** all parameter converges
 - 6: **if** β_{12} is significant **then**
 - 7: Compute the significant test
 - 8: Then, determine the functional hierarchy.
 - 9: **end if**
-

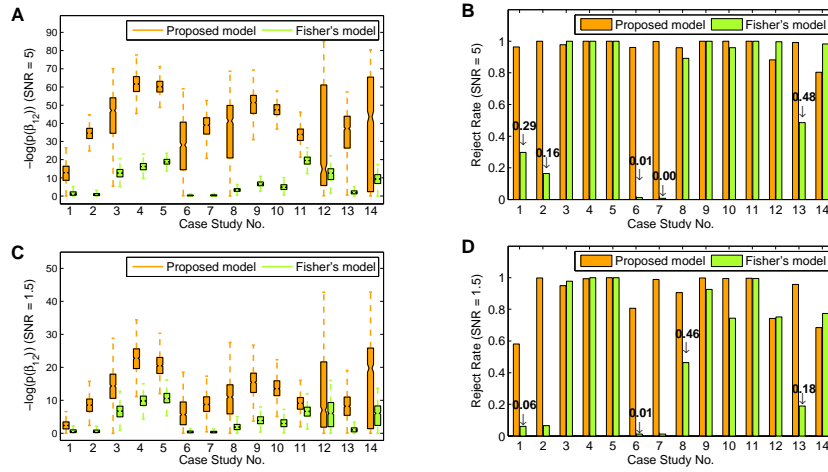


Figure S1: The comparison of the performance between GEM and Fisher's model. Negative logarithmic p-values of the interaction term, $-\log_{10}(p(\beta_{12}))$, (A and B), and reject rates to the null hypothesis (C and D) for SNR = 5 and 1.5, respectively.

Table S1: The signaling pathways and genes associated with schizophrenia.

Signaling pathway	Associated genes
PI3K/AKT signaling growth factors and related	ANK3, AKT1, BDNF, BMP6, CABIN1, EGFR, FYN, GSK3B, IGF1, IMPA2, NRG1, PIK3C3, PIK4CA, PIP5K2A, PDLIM5, RGS4, TCF4, WDFY2
Adhesion/ Junctions	CNTNAP2, JAM3, NCAM1, NRXN1
NMDA and glutamate-related	CIT, DAO, DAOA, DTNBP1, GRIA1, GRID1, GRIN1, GRIN2A, GRIN2B, GRIK4, GRM3, GRM4, GRM7, NOS1, NOSIAP, PCLO, SNAP25, SYN3, SYNJ1
Dopaminergic/ Serotonergic	COMT, DRD2, DRD3, HTR2A, HTR5A, HTR6, NCKAP5, SLC6A3, SLC6A4, SLC18A1, SLC18A2, MAOA, TDO2, TH
GABA	GABRA1, GABRB2, GAD1
Circadian	CLOCK, NPAS2, NR1D1, TIMELESS, PER3
Cytokines	CSF2RB, IL1B, IL1RN, IL6, IL10, TNFA
Oxidative and other stress	ND4, NDUFV2, MTHFR, MTHFD, MTR, NOS3