

Text S2

To understand at which conditions the test of Ueki and Cordell [1] is valid for testing gene-gene interactions, we look at the true parameters $\bar{\xi}^1$ and $\bar{\xi}^0$, where

$$\bar{\xi}^0 = \left\{ \log \left(2 \frac{P_{110}P_{000}}{P_{010}P_{100}} - 1 \right), \log \left(\frac{P_{120}P_{000}}{P_{020}P_{100}} \right), \log \left(\frac{P_{210}P_{000}}{P_{010}P_{200}} \right), \log \left(\frac{P_{220}P_{000}}{P_{020}P_{200}} \right) / 2 \right\}^T$$

and

$$\bar{\xi}^1 = \left\{ \log \left(2 \frac{P_{111}P_{001}}{P_{011}P_{101}} - 1 \right), \log \left(\frac{P_{121}P_{001}}{P_{021}P_{101}} \right), \log \left(\frac{P_{211}P_{001}}{P_{011}P_{201}} \right), \log \left(\frac{P_{221}P_{001}}{P_{021}P_{201}} \right) / 2 \right\}^T.$$

Consider the genotype combination in which SNP1 has the g genotype and SNP 2 has the h genotype. Let P_{gh} be the probability of observing the genotype combination in the population, p_{gh1} be the probability in cases, and p_{gh0} be that in controls. Then

$$\frac{p_{gh1}P_{001}}{p_{0h1}P_{g01}} = \frac{\pi_{gh}\pi_{00}}{\pi_{0h}\pi_{g0}} \frac{P_{gh}P_{00}}{P_{0h}P_{g0}}, \quad \frac{p_{gh0}P_{000}}{p_{0h0}P_{g00}} = \frac{(1-\pi_{gh})(1-\pi_{00})}{(1-\pi_{0h})(1-\pi_{g0})} \frac{P_{gh}P_{00}}{P_{0h}P_{g0}} \quad \text{for } g = 1,2, h = 1,2.$$

For notational convenience, the common allele of a SNP is labeled 0 and the rare allele is labeled 1. Let $\psi_{ij}, i = 0,1, j = 0,1$ denote the haplotype frequencies in the population. For a random mating population, Ueki and Cordell [1] showed that

$$\frac{P_{11}P_{00}}{P_{01}P_{10}} = \frac{e^\delta + 1}{2}, \quad \frac{P_{12}P_{00}}{P_{02}P_{10}} = e^\delta, \quad \frac{P_{21}P_{00}}{P_{01}P_{20}} = e^\delta, \quad \frac{P_{22}P_{00}}{P_{02}P_{20}} = e^{2\delta}.$$

where $\delta = \log([\psi_{11}\psi_{00}]/[\psi_{10}\psi_{01}])$. If we also assume no gene-gene interaction, then

$$\bar{\xi}^0 = \bar{\xi}^1 = \left\{ \log \left(\frac{(1+e^{\mu+\alpha_1})(1+e^{\mu+\beta_1})}{(1+e^{\mu+\alpha_1+\beta_1})(1+e^\mu)} (e^\delta + 1) - 1 \right), \delta + \log \left(\frac{(1+e^{\mu+\alpha_1})(1+e^{\mu+\beta_2})}{(1+e^{\mu+\alpha_1+\beta_2})(1+e^\mu)} \right), \right. \\ \left. \delta + \log \left(\frac{(1+e^{\mu+\alpha_2})(1+e^{\mu+\beta_1})}{(1+e^{\mu+\alpha_2+\beta_1})(1+e^\mu)} \right), \delta + \log \left(\frac{(1+e^{\mu+\alpha_2})(1+e^{\mu+\beta_2})}{(1+e^{\mu+\alpha_2+\beta_2})(1+e^\mu)} \right) / 2 \right\}^T \quad (1)$$

When one or both SNPs are associated with disease, the theoretical optimal weights will be different for cases and for controls. Therefore, in order to ensure that the weighted sums are the same for cases and controls, the four values in (1) must be the same. A necessary condition is

$$\frac{(1 + e^{\mu+\alpha_1})(1 + e^{\mu+\beta_2})}{(1 + e^{\mu+\alpha_1+\beta_2})(1 + e^{\mu})} = \frac{(1 + e^{\mu+\alpha_2})(1 + e^{\mu+\beta_1})}{(1 + e^{\mu+\alpha_2+\beta_1})(1 + e^{\mu})},$$

which implies that

$$(1 + e^{\mu+\alpha_1})(1 + e^{\mu+\beta_2})(1 + e^{\mu+\alpha_2+\beta_1}) = (1 + e^{\mu+\alpha_2})(1 + e^{\mu+\beta_1})(1 + e^{\mu+\alpha_1+\beta_2}).$$

Comparing the coefficients of $e^{3\mu}$, $e^{2\mu}$ and e^{μ} of the left hand side and the right hand side, we have

$$(e^{\alpha_1} - e^{\alpha_2})(1 - e^{\beta_1+\beta_2}) = (e^{\beta_1} - e^{\beta_2})(1 - e^{\alpha_1+\alpha_2}).$$

The condition of α_1 and α_2 such that the equation is true for arbitrary β_1 and β_2 is $\alpha_1 = \alpha_2 = 0$. It is easy to see that this condition is also sufficient, i.e., it leads to

$$\bar{\xi}^0 = \bar{\xi}^1 = \{\delta, \delta, \delta, \delta\}^T.$$

This result indicates that the test of Ueki and Cordell [1] is a valid test for gene-gene interactions when at most one SNP has main effect.

1. Ueki, M. and H.J. Cordell, *Improved Statistics for Genome-Wide Interaction Analysis*. Plos Genetics, 2012. 8(4): p. 141-159.