Supplementary Materials for "Joint Analysis of SNP and Gene Expression Data in Genetic Association Studies of Complex Diseases" by Yen-Tsung Huang, Tyler J. VanderWeele and Xihong Lin

1 Causal mediation model

1.1 Definitions

We first define the counterfactual notation. Let $Y(\mathbf{s}, g)$ be the potential outcome that would have been observed if SNPs ($\mathbf{S}$) and gene expression (G) had been set to $\mathbf{s}$ and g, respectively, and $G(\mathbf{s})$ be the potential outcome of the gene expression had the SNPs ($\mathbf{S}$) been set to $\mathbf{s}$. The notation without parenthesis (e.g., $\mathbf{S}$, G) denotes observed values. Note that $Y(\mathbf{s}, g)$ and $G(s)$ may or may not be observed. They are equivalent to observed values when their determinants are set to the observed values: $Y_i(\mathbf{s} = \mathbf{S}_i, g = G_i) = Y_i$, $G_i(\mathbf{s} = \mathbf{S}_i) = G_i$, known as the assumption of consistency.

Similar to the definitions by VanderWeele and Vansteelandt (2010), we define the direct, indirect and total effects as follows. The *direct effect* of SNPs is the effect of the SNPs on the disease outcome that is not through gene expression, whereas the *indirect effect* is the effect of the SNPs on the disease outcome that is through the gene expression. With these counterfactual notations, we can define the *direct effect* (DE), the *indirect effect* (IE) and the *total effect* (TE) of the SNPs, respectively, on the log odds ratio (OR) scale as:

$$\begin{aligned}
\log[\mathrm{OR}^{\mathrm{DE}}_{\mathbf{s}_1,\mathbf{s}_0|\mathbf{x}}(\mathbf{s}_0)] &= \mathrm{logit}[P(Y_i(\mathbf{s}_1, G_i(\mathbf{s}_0)) = 1|\mathbf{X}_i = \mathbf{x})] - \mathrm{logit}[P(Y_i(\mathbf{s}_0, G_i(\mathbf{s}_0)) = 1|\mathbf{X}_i = \mathbf{x})] \\
\log[\mathrm{OR}^{\mathrm{IE}}_{\mathbf{s}_1,\mathbf{s}_0|\mathbf{x}}(\mathbf{s}_1)] &= \mathrm{logit}[P(Y_i(\mathbf{s}_1, G_i(\mathbf{s}_1)) = 1|\mathbf{X}_i = \mathbf{x})] - \mathrm{logit}[P(Y_i(\mathbf{s}_1, G_i(\mathbf{s}_0)) = 1|\mathbf{X}_i = \mathbf{x})] \\
\log[\mathrm{OR}^{\mathrm{TE}}_{\mathbf{s}_1,\mathbf{s}_0|\mathbf{x}}] &= \log[\mathrm{OR}^{\mathrm{DE}}_{\mathbf{s}_1,\mathbf{s}_0|\mathbf{x}}(\mathbf{s}_0)] + \log[\mathrm{OR}^{\mathrm{IE}}_{\mathbf{s}_1,\mathbf{s}_0|\mathbf{x}}(\mathbf{s}_1)] \\
&= \mathrm{logit}[P(Y_i(\mathbf{s}_1, G_i(\mathbf{s}_1)) = 1|\mathbf{X}_i = \mathbf{x})] - \mathrm{logit}[P(Y_i(\mathbf{s}_0, G_i(\mathbf{s}_0)) = 1|\mathbf{X}_i = \mathbf{x})].
\end{aligned}$$

Note that the two counterfactual conditions we compare in the TE is what the outcome would have been, had SNPs been $\mathbf{s}_1$ versus $\mathbf{s}_0$. All three effects are expressed by conditioning on the covariates ($\mathbf{X}$), which are the measured potential confounding factors.

1.2 Assumptions

To identify DE and IE using the observed data requires four assumptions, which are a multivariate extension of those used in VanderWeele and Vansteelandt (2010). We use $A \coprod B|C$ to denote that A is independent of B conditioning on C. The four assumptions are: after controlling for the covariates ($\mathbf{X}$), (1) $Y(\mathbf{s}, g) \coprod \mathbf{S}|\mathbf{X}$: no unmeasured confounding for the effect of SNPs ($\mathbf{S}$) on the outcome (Y); (2) $Y(\mathbf{s}, g) \coprod G|\mathbf{S}, \mathbf{X}$: no unmeasured confounding for the effect of gene expression (G) on the outcome (Y) after controlling for SNPs ($\mathbf{S}$); (3) $G(\mathbf{s}) \coprod \mathbf{S}|\mathbf{X}$: no unmeasured confounding for the effect of SNPs ($\mathbf{S}$) on gene expression (G); (4) $Y(\mathbf{s}, g) \coprod G(\mathbf{s}^*)|\mathbf{X}$: there is no downstream effect of SNPs ($\mathbf{S}$) that can confound gene expression-outcome relation. We also make the rare disease assumption to approximate logit by log and vice versa.

The event that SNPs regulate gene expression occurs within a cell, it is plausible to assume that no phenotypic covariates exert undue influence on SNP-expression relationship (assumption (3)) and that it is unlikely that downstream factors of SNPs should confound expression-disease relation (assumption (4)). Based on the same reason, the confounders we collect and adjust for to ensure the causal intepretation for SNP-disease association (assumption (1)) should be very similar to those for expression-disease association (assumption (2)). However, if we model a multigenetic disease with a partial list of genetic factors, the remaining causal genetic factors may violate the above four assumptions. Thus, in addition to controlling for all possible confounding covariates, we either assume the complex genetic architecture does not violate our assumptions or all causal genetic factors must be analyzed simultaneously in the model. Later, we will show that these assumptions can be relaxed substantially in developing tests for the TE.

1.3 Direct, indirect and total effects

With the above four assumptions and models (2.1) and (2.2), it can be shown:

$$
\begin{aligned}
&\text{logit}[P(Y_i(\mathbf{s}_a, G_i(\mathbf{s}_b)) = 1|\mathbf{X}_i = \mathbf{x})] \\
&\approx \log[\int P(Y_i(\mathbf{s}_a, g) = 1|\mathbf{X}_i = \mathbf{x}, G_i(\mathbf{s}_b) = g)P(G_i(\mathbf{s}_b) = g|\mathbf{X}_i = \mathbf{x})dg] \\
&= \log[\int P(Y_i(\mathbf{s}_a, g) = 1|\mathbf{X}_i = \mathbf{x}, \mathbf{S}_i = \mathbf{s}_a, G_i = g)P(G_i(\mathbf{s}_b) = g|\mathbf{X}_i = \mathbf{x})dg] \quad \text{(by assmptions 1, 2 \& 4)} \\
&= \log[\int P(Y_i = 1|\mathbf{X}_i = \mathbf{x}, \mathbf{S}_i = \mathbf{s}_a, G_i = g)P(G_i(\mathbf{s}_b) = g|\mathbf{X}_i = \mathbf{x})dg] \quad \text{(by consistency)} \\
&= \log[\int P(Y_i = 1|\mathbf{X}_i = \mathbf{x}, \mathbf{S}_i = \mathbf{s}_a, G_i = g)P(G_i(\mathbf{s}_b) = g|\mathbf{X}_i = \mathbf{x}, \mathbf{S}_i = \mathbf{s}_b)dg] \quad \text{(by assmption 3)} \\
&\approx \log[\int \exp(\mathbf{x}^T\boldsymbol{\alpha} + \mathbf{s}_a^T\boldsymbol{\beta}_S + g\beta_G + \mathbf{s}_a^T g\boldsymbol{\gamma})P(G_i = g|\mathbf{X}_i = \mathbf{x}, \mathbf{S}_i = \mathbf{s}_b)dg] \quad \text{(by consistency)} \\
&= \mathbf{x}^T\boldsymbol{\alpha} + \mathbf{s}_a^T\boldsymbol{\beta}_S + (\beta_G + \mathbf{s}_a^T\boldsymbol{\gamma})(\mathbf{x}^T\boldsymbol{\phi} + \mathbf{s}_b^T\boldsymbol{\delta}) + \frac{1}{2}(\beta_G + \mathbf{s}_a^T\boldsymbol{\gamma})^2\sigma_G^2. \qquad \text{(A. 1)}
\end{aligned}
$$

Note the expression $Y(\mathbf{s}_a, G(\mathbf{s}_b))$ requires a joint manipulation of both SNPs $\mathbf{S}$ and gene expression G, respectively, to be $\mathbf{s}_a$ and $G(\mathbf{s}_b)$, which is another potential outcome with SNPs assigned to be $\mathbf{s}_b$ that may or may not be the same as $\mathbf{s}_a$. It can be intepreted as the following hypothetical intervention steps: 1) intervene and set SNPs to $\mathbf{s}_b$; 2) observe the potential expression $G(\mathbf{s}_b)$; 3) return to the pre-intervention state; 4) intervene to set gene expression to $G(\mathbf{s}_b)$ and SNPs to $\mathbf{s}_a$; 5) observe $Y(\mathbf{s}_a, G(\mathbf{s}_b))$. Robins and Richardson (2010) discuss an example where such an intervention may be possible if a suitable technology is available.

With the above definitions of DE, IE and TE as well as the result in (A. 1), we can derive the expression of the *direct effect*, the *indirect effect* and the *total effect* of the SNPs, respectively, on the log odds ratio scale in terms of the regression coefficients in models (2.1) and (2.2):

$$\begin{aligned}
\log[\mathrm{OR}^{\mathrm{DE}}_{\mathbf{s}_1,\mathbf{s}_0|\mathbf{x}}(\mathbf{s}_0)] \approx & (\mathbf{s}_1-\mathbf{s}_0)^T[\boldsymbol{\beta}_S + \boldsymbol{\gamma}(\mathbf{x}^T\boldsymbol{\phi} + \mathbf{s}_0^T\boldsymbol{\delta} + \beta_G\sigma_G^2)] + \frac{1}{2}\sigma_G^2(\mathbf{s}_1+\mathbf{s}_0)^T\boldsymbol{\gamma}(\mathbf{s}_1-\mathbf{s}_0)^T\boldsymbol{\gamma} \\
\log[\mathrm{OR}^{\mathrm{IE}}_{\mathbf{s}_1,\mathbf{s}_0|\mathbf{x}}(\mathbf{s}_1)] \approx & (\mathbf{s}_1-\mathbf{s}_0)^T\boldsymbol{\delta}(\beta_G + \mathbf{s}_1^T\boldsymbol{\gamma}) \\
\log[\mathrm{OR}^{\mathrm{TE}}_{\mathbf{s}_1,\mathbf{s}_0|\mathbf{x}}] \approx & (\mathbf{s}_1-\mathbf{s}_0)^T[\boldsymbol{\beta_S} + \beta_G\boldsymbol{\delta} + \boldsymbol{\gamma}(\mathbf{x}^T\boldsymbol{\phi} + \mathbf{s}_0^T\boldsymbol{\delta} + \beta_G\sigma_G^2) + \boldsymbol{\delta}\mathbf{s}_1^T\boldsymbol{\gamma}] \\
& + \frac{1}{2}\sigma_G^2(\mathbf{s}_1+\mathbf{s}_0)^T\boldsymbol{\gamma}(\mathbf{s}_1-\mathbf{s}_0)^T\boldsymbol{\gamma}
\end{aligned}$$

1.4 Assumptions required for testing the total effect (TE)

We now study the assumption for the null hypothesis of no total SNP effects and show that the assumptions mentioned above can be relaxed substantially in developing tests for the total effect of the SNPs on the disease, which still exploits expression data. To perform hypothesis testing for no total effect of the SNPs $\mathbf{S}$ on Y in (4.1), we can substantially weaken the four unmeasured confounding assumptions required for simultaneously estimating both direct and indirect genetic effects described in the above discussion, and only need a single assumption that no unmeasured confounding for the effect of SNPs $\mathbf{S}$ on the outcome Y after controlling for the covariates $\mathbf{X}$, i.e., $Y(\mathbf{s}) \coprod \mathbf{S}|\mathbf{X}$. This is because, under the assumption $Y(\mathbf{s}) \coprod \mathbf{S}|\mathbf{X}$, we can show that $\log[P(Y_i(\mathbf{s})=1|\mathbf{X}_i=\mathbf{x})] = \log[P(Y_i=1|\mathbf{X}_i=\mathbf{x},\mathbf{S}_i=\mathbf{s})]$. It follows that

$$\begin{aligned}
&\mathrm{logit}[P(Y_i(\mathbf{s})=1|\mathbf{X}_i=\mathbf{x})] \approx \log[P(Y_i=1|\mathbf{X}_i=\mathbf{x},\mathbf{S}_i=\mathbf{s})] \\
&\quad = \log[\int P(Y_i=1|\mathbf{X}_i=\mathbf{x},\mathbf{S}_i=\mathbf{s},G_i=g)P(G_i=g|\mathbf{X}_i=\mathbf{x},\mathbf{S}_i=\mathbf{s})dg] \\
&\quad \approx \mathbf{x}^T\boldsymbol{\alpha} + \mathbf{s}^T\boldsymbol{\beta}_S + (\beta_G + \mathbf{s}^T\boldsymbol{\gamma})(\mathbf{x}^T\boldsymbol{\phi} + \mathbf{s}^T\boldsymbol{\delta}) + \frac{1}{2}(\beta_G + \mathbf{s}^T\boldsymbol{\gamma})^2\sigma_G^2.
\end{aligned}$$

Simple calculations by setting $\mathbf{s}$ to $\mathbf{s}_0$ and $\mathbf{s}_1$ show the total SNP effect (TE) that is given in the above. Note $\mathrm{logit}[P(Y_i(\mathbf{s})=1|\mathbf{X}_i=\mathbf{x})] = \mathrm{logit}[P(Y_i(\mathbf{s},G_i(\mathbf{s}))=1|\mathbf{X}_i=\mathbf{x})]$, where the counterfactual $Y(\mathbf{s}) = Y(\mathbf{s},G(\mathbf{s}))$ corresponds to simply setting the SNPs to level $\mathbf{s}$. Hence

estimation of total SNP effect only requires the no unmeasured confounding assumption $Y_i(\mathbf{s}) \coprod \mathbf{S}_i|\mathbf{X}_i$. Note that the above derivation also depends on the linearity of the outcome model as specified in (2.1) in the main text and the normality of the gene expression model.

In contrast, estimation of both direct and indirect effects individually requires estimating $\text{logit}[P(Y_i(\mathbf{s}_a, G_i(\mathbf{s}_b)) = 1|\mathbf{X}_i = \mathbf{x})]$. For $Y_i(\mathbf{s}_a, G_i(\mathbf{s}_b))$, the SNPs are set to one level, $\mathbf{s}_a$, while the gene expression is set to the level it would have been had the SNPs been set to a different level, $\mathbf{s}_b$. Because $\mathbf{s}_a$ and $\mathbf{s}_b$ can be different, the assumptions needed for identification of this counterfactual are much stronger, and the four no unmeasured confounding assumptions as given previously are needed. The three additional unmeasured confounding assumptions allow decomposing the total effect into direct and indirect effects, but are not required for testing for the total genetic effects.

2 Derivation of model (4.7) in the main text

When $[Y|\mathbf{S}, G, \mathbf{X}]$ follows the interaction model (2.1) in the main text, by plugging in (2.2) into (2.1) in the main text, the true $[Y|\mathbf{S}, G, \mathbf{X}]$ model can be written as

$$\begin{aligned}\text{logit}[P(Y_i = 1|\mathbf{S}_i, \mathbf{X}_i, \epsilon_i)] &= \mathbf{X}_i^T\boldsymbol{\alpha} + \mathbf{S}_i^T\boldsymbol{\beta}_S + (\mathbf{X}_i\boldsymbol{\phi} + \mathbf{S}_i^T\boldsymbol{\delta} + \epsilon_i)\beta_G + (\mathbf{X}_i\boldsymbol{\phi} + \mathbf{S}_i^T\boldsymbol{\delta} + \epsilon_i)\mathbf{S}_i^T\boldsymbol{\gamma} \\ &= \mathbf{X}_i^T(\boldsymbol{\alpha} + \boldsymbol{\phi}\beta_G) + \mathbf{S}_i^T(\boldsymbol{\beta}_S + \boldsymbol{\delta}\beta_G) \\ &\quad + \mathbf{X}_i^T\boldsymbol{\phi}\mathbf{S}_i^T\boldsymbol{\gamma} + \mathbf{S}_i^T\boldsymbol{\delta}\mathbf{S}_i^T\boldsymbol{\gamma} + (\beta_G + \mathbf{S}_i^T\boldsymbol{\gamma})\epsilon_i.\end{aligned}$$

Integrating out ϵ_i, we have

$$\text{logit}[P(Y_i = 1|\mathbf{S}_i, \mathbf{X}_i, \epsilon_i)] \approx c_i^* \left\{\mathbf{X}_i^T(\boldsymbol{\alpha} + \boldsymbol{\phi}\beta_G) + \mathbf{S}_i^T(\boldsymbol{\beta}_S + \boldsymbol{\delta}\beta_G) + \mathbf{X}_i^T\boldsymbol{\phi}\mathbf{S}_i^T\boldsymbol{\gamma} + \mathbf{S}_i^T\boldsymbol{\delta}\mathbf{S}_i^T\boldsymbol{\gamma}\right\}$$

where $c_i^* = \{1 + 0.35\sigma_G^2(\beta_G + \mathbf{S}_i^T\boldsymbol{\gamma})^2\}^{-1/2}$ (Zeger et al., 1988; Breslow and Clayton, 1993).

3 Asymptotic distribution of Q

First note that $Q = n^{-1}(\mathbf{Y} - \widehat{\boldsymbol{\mu}}_0)^T(a_1\mathbb{S}\mathbb{S}^T + a_2\mathbf{G}\mathbf{G}^T + a_3\mathbb{C}\mathbb{C}^T)(\mathbf{Y} - \widehat{\boldsymbol{\mu}}_0)$, can be expressed in terms of the L2 norm of the scores, $\|n^{-1/2}\sum_{i=1}^n \mathbf{V}_i(Y_i - \widehat{\mu}_i)\|_2^2$. We denote this quantity as $\|\sqrt{n}\widehat{S}_V\|_2^2$, where $S_U(\boldsymbol{\theta}) = \begin{bmatrix} S_X(\boldsymbol{\theta})_{q\times 1} \\ S_V(\boldsymbol{\theta})_{(2p+1)\times 1}\end{bmatrix} = n^{-1}\sum_{i=1}^n \mathbf{U}_i(Y_i - \mu_i)$, $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta_S}^T, \beta_G, \boldsymbol{\gamma}^T)^T$, and $\widehat{S}_V$ is the counterpart of $S_V(\boldsymbol{\theta})$ by plugging in $\boldsymbol{\theta}$ with $\widehat{\boldsymbol{\theta}_0} = (\widehat{\boldsymbol{\alpha}}_0^T, \mathbf{0}_{2p+1}^T)^T$, $\widehat{\boldsymbol{\alpha}}_0$ is the MLE of $\boldsymbol{\alpha}$ under H_0.

A simple Taylor series expansion shows

$$\sqrt{n}S_X(\widehat{\boldsymbol{\theta}_0}) = \sqrt{n}S_X(\boldsymbol{\theta}_0) - \sqrt{n}\mathbf{D}_{XX}(\widehat{\boldsymbol{\alpha}_0} - \boldsymbol{\alpha}_0) + o_p(1) \qquad \text{(A. 2)}$$

Using (A. 2), simple calculations show that

$$\begin{aligned}\sqrt{n}\widehat{S}_V &= \sqrt{n}S_V(\widehat{\boldsymbol{\theta}_0}) = \sqrt{n}S_V(\boldsymbol{\theta}_0) - \sqrt{n}\mathbf{D}_{VX}(\widehat{\boldsymbol{\alpha}_0} - \boldsymbol{\alpha}_0) + o_p(1)\\ &= \sqrt{n}S_V(\boldsymbol{\theta}_0) - \sqrt{n}\mathbf{D}_{VX}\mathbf{D}_{XX}^{-1}S_X(\boldsymbol{\theta}_0) + o_p(1)\\ &= \sqrt{n}\mathbf{A}S_U(\boldsymbol{\theta}_0) + o_p(1).\end{aligned}$$

By the central limit theorem, $\sqrt{n}S_U(\theta_0) \xrightarrow{D} \boldsymbol{\epsilon}$ in distribution as $n \to \infty$, where $\boldsymbol{\epsilon}$ follows $N(0, \mathbf{D})$. From the Slutsky theorem and the above results, we then have $\sqrt{n}\widehat{S}_V \xrightarrow{D} \mathbf{A}\boldsymbol{\epsilon}$. Finally, it follows from the continuous mapping theorem, $Q = \|\sqrt{n}\widehat{S}_V\|_2^2 \xrightarrow{D} \|\mathbf{A}\boldsymbol{\epsilon}\|_2^2 = \sum_{l=1}^{2p+1}(\mathbf{A}_l\boldsymbol{\epsilon})^2$.

The expectation and variance of Q can be obtained from the results of the quadratic function. More specifically, we first express $Q = n^{-1}(\mathbf{Y} - \widehat{\boldsymbol{\mu}}_0)^T(a_1\mathbb{S}\mathbb{S}^T + a_2\mathbf{G}\mathbf{G}^T + a_3\mathbb{C}\mathbb{C}^T)(\mathbf{Y} - \widehat{\boldsymbol{\mu}}_0)$ as a quadratic function of $\mathbf{Y}$, $\mathbf{Y}^T\mathbf{B}\mathbf{Y}$, where $\mathbf{B} = n^{-1}(\mathbf{I} - \mathbf{H})(a_1\mathbb{S}\mathbb{S}^T + a_2\mathbf{G}\mathbf{G}^T + a_3\mathbb{C}\mathbb{C}^T)(\mathbf{I} - \mathbf{H})$ and $\mathbf{H} = \mathbf{W}^{1/2}\mathbb{X}(\mathbb{X}^T\mathbf{W}\mathbb{X})\mathbb{X}^T\mathbf{W}^{1/2}$. It follows that $\mathrm{E}(Q) = tr(\mathbf{B}\mathbf{W})$ and $\mathrm{Var}(Q) = 2tr(\mathbf{B}\mathbf{W}\mathbf{B}\mathbf{W})$.

References

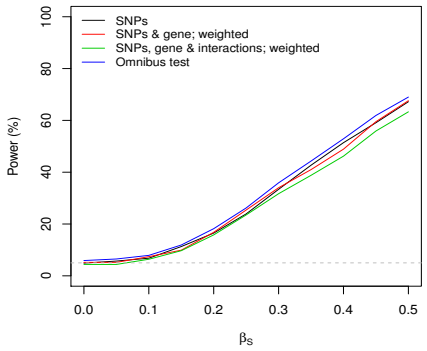
Breslow, N. and Clayton, D. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9-25.

Robins, J. and Richardson, T. (2010) Alternative graphical causal models and the identification of direct effects. *working paper no. 100, Center of Statistics and the Social Sciences, University of Washington.*

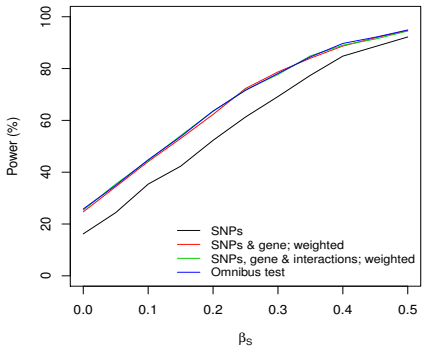
Zeger, S., Liang, K. and Albert P. (1988). Models for Longitudinal Data: A Generalized Estimating Equation Approach. *Biometrics* **44**, 1049-1060.

Table 1: (Supplemental Table 1) Empirical sizes ($\times 10^{-2}$) of the proposed variance component score tests using Davies' approximation and perturbation. The empirical size was calculated at the significance level of 0.05, 5×10^{-3}, 5×10^{-4}, based on 1,000,000 simulations. Q_S: SNP-only analysis; Q_{SG}: joint analyses of SNP and gene expression without interaction; Q_{SGC}: joint analyses of SNP, gene expression and their interaction

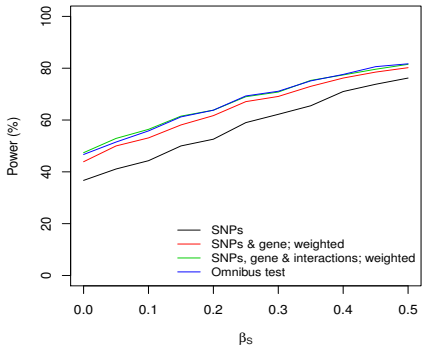
	Davies' method			Perturbation			
Significance level	Q_S	Q_{SG}	Q_{SGC}	Q_S	Q_{SG}	Q_{SGC}	Omnibus
0.05	5.04	5.01	5.02	4.74	4.94	4.78	4.91
5×10^{-3}	0.49	0.47	0.46	0.52	0.58	0.52	0.62
5×10^{-4}	0.048	0.045	0.038	0.08	0.08	0.08	0.18



(a) $\beta_G = 0, \gamma = 0; \delta = 1.0$

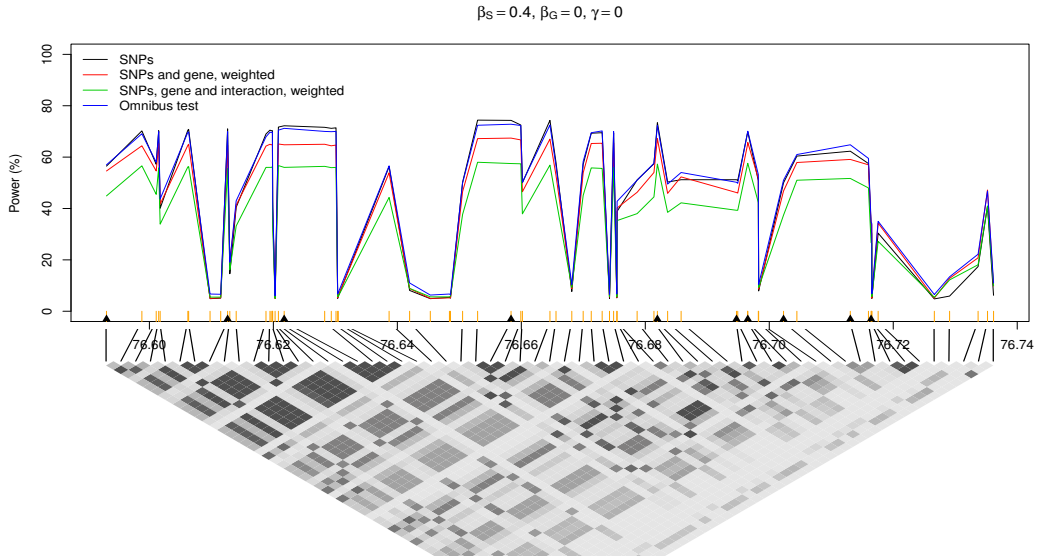


(b) $\beta_G = 0.2, \gamma = 0; \delta = 1.0$

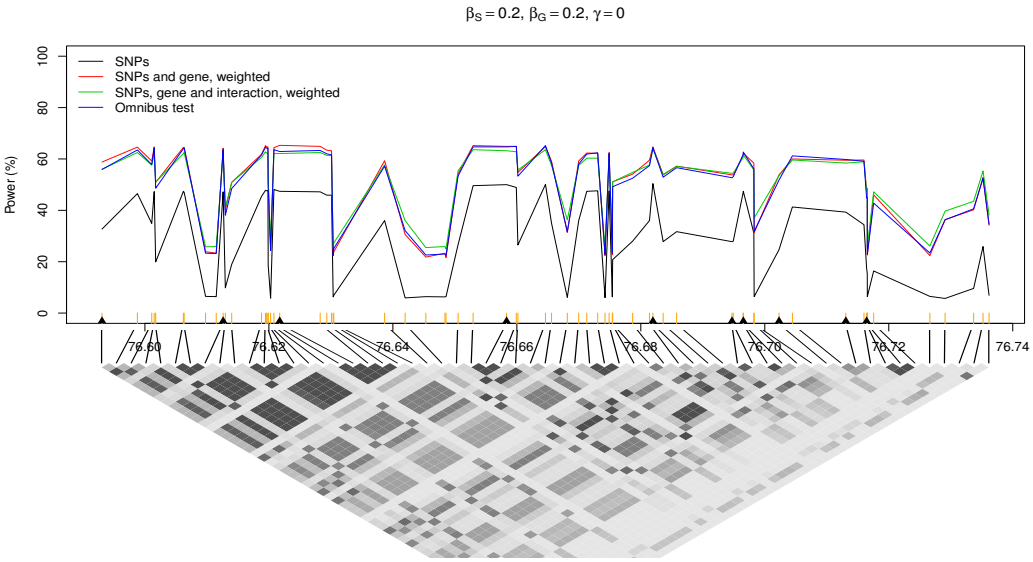


(c) $\beta_G = 0.2, \gamma = 0.2; \delta = 1.0$

Figure 1: (Supplemental Figure 1) Empirical power (%) under the setting with three causal SNPs of the ORMDL3 gene. SNPs are assumed to be eQTL SNPs ($\delta = 1$). Each figure plots the powers of the proposed tests as a function of the main effect of SNP (β_s). The three figures correspond to the three different true models, the model with only SNP effects, the model with only main effects of SNP and gene expression, and the model with SNPs, gene expression and their interaction effects. The dashed line in (a) indicates 5% type I error rate.



(a) $\beta_S = 0.4, \beta_G = 0, \gamma = 0$



(b) $\beta_S = 0.2, \beta_G = 0.2, \gamma = 0$

Figure 2: (Supplemental Figure 2) Simulated power curves at 15q24-15q25.1. The x-axis indicates the physical location (Mb) of the 69 HapMap SNPs at 15q24-15q25.1. The orange vertical bar indicates the relative locations of the causal SNP and the black triangles indicate the ten typed SNPs. Different lines indicate the powers of different tests. The lower panel of each subfigure is the plot for linkage disequilibrium, measured as r^2 ranging from 0 (white) to 1 (black).