

Supplementary Document for “LPM: a latent probit model to characterize relationship among complex traits using summary statistics from multiple GWASs and functional annotations”

Jingsi Ming¹, Tao Wang^{2,3} and Can Yang^{1*}

¹Department of Mathematics, The Hong Kong University of Science and Technology, Hong Kong,

²Department of Bioinformatics and Biostatistics, Shanghai Jiao Tong University, Shanghai, China

³MoE Key Lab of Artificial Intelligence, Shanghai Jiao Tong University, Shanghai, China

1 Calculation of conditional correlation

In LPM, we assume that $\epsilon_j \sim N(\mathbf{0}, \mathbf{R})$ where $\mathbf{R} \in \mathbb{R}^{K \times K}$ is the correlation matrix among K traits. Suppose we are interested in the relationship among the first k traits conditional on the other traits, we can partition the traits into two disjoint subsets: part a (the first k traits) and part b (other traits). Then we can have the following partitions

$$\epsilon_j = \begin{pmatrix} \epsilon_{ja} \\ \epsilon_{jb} \end{pmatrix}, \mathbf{R} = \begin{pmatrix} \mathbf{R}_{aa} & \mathbf{R}_{ab} \\ \mathbf{R}_{ba} & \mathbf{R}_{bb} \end{pmatrix}.$$

According to the property of Gaussian distribution, we can have

$$\epsilon_{jb} | \epsilon_{ja} \sim N(\mathbf{R}_{ba} \mathbf{R}_{aa}^{-1} \epsilon_{ja}, \mathbf{R}_{bb} - \mathbf{R}_{ba} \mathbf{R}_{aa}^{-1} \mathbf{R}_{ab}).$$

Then the covariance matrix for the first k traits conditional on other traits is given by $\mathbf{R}_{aa} - \mathbf{R}_{ab} \mathbf{R}_{bb}^{-1} \mathbf{R}_{ba}$.

For example, we assume that the relationship among smoking, alcohol drinking and lung cancer can be modeled with the correlation matrix

$$\mathbf{R} = \begin{pmatrix} 1 & 0.8 & 0.9 \\ 0.8 & 1 & 0.72 \\ 0.9 & 0.72 & 1 \end{pmatrix}.$$

In this example, smoking is highly correlated with alcohol drinking ($\rho = 0.8$) and lung cancer ($\rho = 0.9$) which causes alcohol drinking and lung cancer seem to be also correlated ($\rho = 0.72$). Based on the procedure above, we can calculate the covariance matrix for alcohol drinking and lung

*Correspondence should be addressed to Can Yang (macyang@ust.hk)

cancer conditional on smoking. In this case

$$\mathbf{R}_{aa} = 1, \mathbf{R}_{bb} = \begin{pmatrix} 1 & 0.72 \\ 0.72 & 1 \end{pmatrix}, \mathbf{R}_{ba} = \begin{pmatrix} 0.8 \\ 0.9 \end{pmatrix}, \mathbf{R}_{ab} = \begin{pmatrix} 0.8 & 0.9 \end{pmatrix}.$$

The conditional covariance matrix can be calculated as follows

$$\mathbf{R}_{bb} - \mathbf{R}_{ba}\mathbf{R}_{aa}^{-1}\mathbf{R}_{ab} = \begin{pmatrix} 1 & 0.72 \\ 0.72 & 1 \end{pmatrix} - \begin{pmatrix} 0.8 \\ 0.9 \end{pmatrix} \begin{pmatrix} 0.8 & 0.9 \end{pmatrix} = \begin{pmatrix} 0.36 & 0 \\ 0 & 0.19 \end{pmatrix},$$

indicating that alcohol drinking and lung cancer are independent ($\rho = 0$) conditional on smoking.

As a result, after we obtain the estimate $\hat{\mathbf{R}}$ in LPM, we can also evaluate the conditional independence for traits of interest.

2 The detailed derivation of the PX-EM algorithm

For two GWAS, the original bivariate LPM (bLPM) is

$$\tilde{P}_{jk} \sim \begin{cases} U[0, 1], & \tilde{\eta}_{jk} = 0, \\ Beta(\tilde{\alpha}_k, 1), & \tilde{\eta}_{jk} = 1, \end{cases}$$

$$\tilde{\eta}_{jk} = \begin{cases} 1, & \text{if } \tilde{Z}_{jk} > 0, \\ 0, & \text{if } \tilde{Z}_{jk} \leq 0, \end{cases}$$

$$\tilde{\mathbf{Z}}_j = \tilde{\beta} \mathbf{X}_j + \tilde{\epsilon}_j,$$

$$\tilde{\epsilon}_j \sim N(\mathbf{0}, \tilde{\mathbf{R}}),$$

where $\tilde{\mathbf{Z}} \in \mathbb{R}^{M \times 2}$ is the latent variable in probit model, $\mathbf{X} \in \mathbb{R}^{M \times (D+1)}$ is the design matrix of functional annotations, comprised of an intercept and D annotations, $\tilde{\beta} \in \mathbb{R}^{2 \times (D+1)}$ is a matrix of the coefficients. For the j -th SNP, $\tilde{\mathbf{Z}}_j$, \mathbf{X}_j and $\tilde{\epsilon}_j$ are all vectors containing the j -th row of $\tilde{\mathbf{Z}}$, \mathbf{X} and $\tilde{\epsilon}$, respectively. $\tilde{\mathbf{R}} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ is the correlation matrix measuring the pleiotropy of the phenotypes. When the pleiotropy effect exists, ρ is expected to differ from 0.

The expanded model is

$$\tilde{P}_{jk} \sim \begin{cases} U[0, 1], & \tilde{\eta}_{jk} = 0, \\ Beta(\tilde{\alpha}_k, 1), & \tilde{\eta}_{jk} = 1, \end{cases}$$

$$\tilde{\eta}_{jk} = \begin{cases} 1, & \text{if } \tilde{Z}_{jk} > 0, \\ 0, & \text{if } \tilde{Z}_{jk} \leq 0, \end{cases}$$

$$\tilde{\mathbf{Z}}_j = \gamma \mathbf{X}_j + \tilde{\epsilon}_j,$$

$$\tilde{\epsilon}_j \sim N(\mathbf{0}, \Sigma),$$

where $\gamma = \mathbf{D}\tilde{\beta}$, $\Sigma = \mathbf{D}\tilde{\mathbf{R}}\mathbf{D} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ and $\mathbf{D} = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}$ is the auxiliary parameter whose value is fixed at \mathbf{I}_2 in the original model.

Let $\theta = \{\tilde{\alpha}, \tilde{\beta}, \tilde{\mathbf{R}}\}$ and $\Theta = \{\tilde{\alpha}, \gamma, \Sigma\}$ be the collection of model parameters in the original model and the expanded model, respectively.

The complete-data log-likelihood for the expanded model is

$$\begin{aligned}
& \log \Pr(\tilde{\mathbf{P}}, \tilde{\boldsymbol{\eta}}, \tilde{\mathbf{Z}} | \mathbf{X}; \boldsymbol{\Theta}) \\
&= \log \Pr(\tilde{\mathbf{P}} | \tilde{\boldsymbol{\eta}}; \tilde{\boldsymbol{\alpha}}) + \log \Pr(\tilde{\mathbf{Z}} | \mathbf{X}; \boldsymbol{\gamma}, \boldsymbol{\Sigma}) \\
&= \sum_{j=1}^M \sum_{k=1}^2 \tilde{\eta}_{jk} (\log \tilde{\alpha}_k + (\tilde{\alpha}_k - 1) \log \tilde{P}_{jk}) \\
&\quad + \sum_{j=1}^M \left[-\frac{1}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} (\tilde{\mathbf{Z}}_j - \boldsymbol{\gamma}\mathbf{X}_j)^T \boldsymbol{\Sigma}^{-1} (\tilde{\mathbf{Z}}_j - \boldsymbol{\gamma}\mathbf{X}_j) \right].
\end{aligned}$$

The Q function is evaluated as follows

$$\begin{aligned}
Q &= E_{\tilde{\boldsymbol{\eta}}, \tilde{\mathbf{Z}}} \log \Pr(\tilde{\mathbf{P}}, \tilde{\boldsymbol{\eta}}, \tilde{\mathbf{Z}} | \mathbf{X}; \boldsymbol{\Theta}) \\
&= \sum_{j=1}^M \sum_{k=1}^2 E[\tilde{\eta}_{jk}] (\log \tilde{\alpha}_k + (\tilde{\alpha}_k - 1) \log \tilde{P}_{jk}) \\
&\quad + \sum_{j=1}^M \left[-\frac{1}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2} \text{trace}(E[\tilde{\mathbf{Z}}_j \tilde{\mathbf{Z}}_j^T] \boldsymbol{\Sigma}^{-1}) + E[\tilde{\mathbf{Z}}_j^T] \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma} \mathbf{X}_j - \frac{1}{2} \mathbf{X}_j^T \boldsymbol{\gamma}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\gamma} \mathbf{X}_j \right],
\end{aligned}$$

where the expectation is calculated based on the current $\boldsymbol{\theta}$ in the original model.

PX-E step

To obtain the expectation of $\tilde{\boldsymbol{\eta}}$, we calculate the posterior of $\tilde{\boldsymbol{\eta}}$ in the original model:

$$\begin{aligned}
\pi_{j11} &= \Pr(\tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1 | \tilde{\mathbf{P}}, \mathbf{X}; \boldsymbol{\theta}) = \frac{\Phi_{j11} \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1}}{\Phi_j}, \\
\pi_{j10} &= \Pr(\tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 0 | \tilde{\mathbf{P}}, \mathbf{X}; \boldsymbol{\theta}) = \frac{\Phi_{j10} \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1}}{\Phi_j}, \\
\pi_{j01} &= \Pr(\tilde{\eta}_{j1} = 0, \tilde{\eta}_{j2} = 1 | \tilde{\mathbf{P}}, \mathbf{X}; \boldsymbol{\theta}) = \frac{\Phi_{j01} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1}}{\Phi_j}, \\
\pi_{j00} &= \Pr(\tilde{\eta}_{j1} = 0, \tilde{\eta}_{j2} = 0 | \tilde{\mathbf{P}}, \mathbf{X}; \boldsymbol{\theta}) = \frac{\Phi_{j00}}{\Phi_j},
\end{aligned}$$

where

$$\begin{aligned}
\Phi_{j11} &= \Phi_2 \left(\tilde{\beta}_1^T \mathbf{X}_j, \tilde{\beta}_2^T \mathbf{X}_j, \rho \right), \\
\Phi_{j10} &= \Phi_2 \left(\tilde{\beta}_1^T \mathbf{X}_j, -\tilde{\beta}_2^T \mathbf{X}_j, -\rho \right) = -\Phi_{j11} + \Phi \left(\tilde{\beta}_1^T \mathbf{X}_j \right), \\
\Phi_{j01} &= \Phi_2 \left(-\tilde{\beta}_1^T \mathbf{X}_j, \tilde{\beta}_2^T \mathbf{X}_j, -\rho \right) = -\Phi_{j11} + \Phi \left(\tilde{\beta}_2^T \mathbf{X}_j \right), \\
\Phi_{j00} &= \Phi_2 \left(-\tilde{\beta}_1^T \mathbf{X}_j, -\tilde{\beta}_2^T \mathbf{X}_j, \rho \right) = 1 + \Phi_{j11} - \Phi \left(\tilde{\beta}_1^T \mathbf{X}_j \right) - \Phi \left(\tilde{\beta}_2^T \mathbf{X}_j \right), \\
\Phi_j &= \Phi_{j11} \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j10} \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} + \Phi_{j01} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j00},
\end{aligned}$$

and $\Phi(\cdot)$ is the univariate standard normal cumulative distribution function, $\Phi_2\left(\tilde{\beta}_1^T \mathbf{X}_j, \tilde{\beta}_2^T \mathbf{X}_j, \rho\right) = \int_{-\infty}^{\tilde{\beta}_1^T \mathbf{X}_j} \int_{-\infty}^{\tilde{\beta}_2^T \mathbf{X}_j} \phi(x_1, x_2, \rho) dx_1 dx_2$,

$$\phi(x_1, x_2, \rho) = \frac{\exp\left[-\frac{x_1^2+x_2^2-2\rho x_1 x_2}{2(1-\rho^2)}\right]}{2\pi\sqrt{1-\rho^2}}.$$

Therefore, we have

$$E[\tilde{\eta}_{j1}] = \frac{\Phi_{j11}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j10}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}}{\Phi_j},$$

$$E[\tilde{\eta}_{j2}] = \frac{\Phi_{j11}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j01}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1}}{\Phi_j}.$$

The logarithm of the likelihood can be evaluated by

$$L = \sum_{j=1}^M \log \Phi_j.$$

The posterior of $\tilde{\mathbf{Z}}_j$ given $\tilde{\eta}_j$ is a truncated bivariate normal distribution. Using the properties of truncated bivariate normal distribution, we can get

$$E(\tilde{Z}_{j1}) = \tilde{\beta}_1^T \mathbf{X}_j + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) E(\tilde{\epsilon}_{j1} | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta})$$

$$+ \frac{\phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) + \frac{\rho \phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),$$

$$E(\tilde{Z}_{j2}) = \tilde{\beta}_2^T \mathbf{X}_j + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) E(\tilde{\epsilon}_{j2} | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta})$$

$$+ \frac{\rho \phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) + \frac{\phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),$$

$$E(\tilde{Z}_{j1}^2) = \tilde{\beta}_1^T \mathbf{X}_j (2E(\tilde{Z}_{j1}) - \tilde{\beta}_1^T \mathbf{X}_j) + 1$$

$$+ \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) (E(\tilde{\epsilon}_{j1}^2 | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) - 1)$$

$$- \frac{\tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) - \frac{\rho^2 \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),$$

$$\begin{aligned}
E(\tilde{Z}_{j2}^2) = & \tilde{\beta}_2^T \mathbf{X}_j \left(2E(\tilde{Z}_{j2}) - \tilde{\beta}_2^T \mathbf{X}_j \right) + 1 \\
& + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) (E(\tilde{\epsilon}_{j2}^2 | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) - 1) \\
& - \frac{\rho^2 \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) - \frac{\tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{Z}_{j1} \tilde{Z}_{j2}) = & -\tilde{\beta}_1^T \mathbf{X}_j \tilde{\beta}_2^T \mathbf{X}_j + \tilde{\beta}_1^T \mathbf{X}_j E(\tilde{Z}_{j2}) + \tilde{\beta}_2^T \mathbf{X}_j E(\tilde{Z}_{j1}) + \rho \\
& + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) (E(\tilde{\epsilon}_{j1} \tilde{\epsilon}_{j2} | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) - \rho) \\
& - \frac{\rho \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) - \frac{\rho \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1).
\end{aligned}$$

where

$$\begin{aligned}
E(\tilde{\epsilon}_{j1} | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) = & \frac{1}{\Phi_{j11}} \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} \rho \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j2} | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) = & \frac{1}{\Phi_{j11}} \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} \rho \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j2}^2 | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) = & 1 - \frac{1}{\Phi_{j11}} \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c) \\
& - \frac{1}{\Phi_{j11}} \rho^2 \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} c^{-1} \rho \phi(\tilde{\beta}_2^T \mathbf{X}_j) \phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j2}^2 | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) = & 1 - \frac{1}{\Phi_{j11}} \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c) \\
& - \frac{1}{\Phi_{j11}} \rho^2 \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} c^{-1} \rho \phi(\tilde{\beta}_1^T \mathbf{X}_j) \phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j1}\tilde{\epsilon}_{j2}|\tilde{\eta}_{j1}=1,\tilde{\eta}_{j2}=1,\mathbf{X}_j;\boldsymbol{\theta}) = & \rho - \frac{1}{\Phi_{j11}}\rho\tilde{\beta}_1^T\mathbf{X}_j\phi\left(\tilde{\beta}_1^T\mathbf{X}_j\right)\Phi\left(\left(\tilde{\beta}_2^T\mathbf{X}_j-\rho\tilde{\beta}_1^T\mathbf{X}_j\right)c\right) \\
& - \frac{1}{\Phi_{j11}}\rho\tilde{\beta}_2^T\mathbf{X}_j\phi\left(\tilde{\beta}_2^T\mathbf{X}_j\right)\Phi\left(\left(\tilde{\beta}_1^T\mathbf{X}_j-\rho\tilde{\beta}_2^T\mathbf{X}_j\right)c\right) \\
& + \frac{1}{\Phi_{j11}}c^{-1}\phi\left(\tilde{\beta}_1^T\mathbf{X}_j\right)\phi\left(\left(\tilde{\beta}_2^T\mathbf{X}_j-\rho\tilde{\beta}_1^T\mathbf{X}_j\right)c\right),
\end{aligned}$$

$$c = \frac{1}{\sqrt{1-\rho^2}},$$

and $\phi(\cdot)$ is the standard univariate normal density.

PX-M step

We set the derivative of the Q function with respect to the parameters in Θ be zero

$$\begin{aligned}
\frac{\partial Q}{\partial \tilde{\alpha}_k} &= \sum_{j=1}^M E[\tilde{\eta}_{jk}] \left(\frac{1}{\tilde{\alpha}_k} + \log \tilde{P}_{jk} \right) = 0, \\
\frac{\partial Q}{\partial \gamma} &= \sum_{j=1}^M [\mathbf{X}_j E[\tilde{\mathbf{Z}}_j^T] \Sigma^{-1} - \mathbf{X}_j \mathbf{X}_j^T \gamma^T \Sigma^{-1}] = 0, \\
\frac{\partial Q}{\partial \Sigma^{-1}} &= \sum_{j=1}^M \left[\frac{1}{2} \Sigma - \frac{1}{2} E[\tilde{\mathbf{Z}}_j \tilde{\mathbf{Z}}_j^T] + \gamma \mathbf{X}_j E[\tilde{\mathbf{Z}}_j^T] - \frac{1}{2} \gamma \mathbf{X}_j \mathbf{X}_j^T \gamma^T \right] = 0.
\end{aligned}$$

Then we can obtain the updating equations

$$\begin{aligned}
\tilde{\alpha}_k &= -\frac{\sum_{j=1}^M E[\tilde{\eta}_{jk}]}{\sum_{j=1}^M E[\tilde{\eta}_{jk}] \log \tilde{P}_{jk}}, \\
\gamma &= \left(\sum_{j=1}^M E[\tilde{\mathbf{Z}}_j] \mathbf{X}_j^T \right) \left(\sum_{j=1}^M \mathbf{X}_j \mathbf{X}_j^T \right)^{-1} = E[\tilde{\mathbf{Z}}^T] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}, \\
\Sigma &= \frac{1}{M} \left[\left(\sum_{j=1}^M E[\tilde{\mathbf{Z}}_j \tilde{\mathbf{Z}}_j^T] \right) - \gamma \mathbf{X}^T E[\tilde{\mathbf{Z}}] \right].
\end{aligned}$$

We reduce to the original parameters using the reduction function:

$$\begin{aligned}
\tilde{\beta} &= \mathbf{D}^{-1} \gamma, \\
\tilde{\mathbf{R}} &= \mathbf{D}^{-1} \Sigma \mathbf{D}^{-1}.
\end{aligned}$$

Implementation

- Initialize $\tilde{\alpha}$, $\tilde{\beta}$ using warm starts and $\rho = 0$.

- E-step: For $j = 1, \dots, M$ and $k = 1, 2$, calculate

$$E[\tilde{\eta}_{j1}] = \frac{\Phi_{j11}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j10}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}}{\Phi_j},$$

$$E[\tilde{\eta}_{j2}] = \frac{\Phi_{j11}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j01}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1}}{\Phi_j},$$

where

$$\begin{aligned}\Phi_{j11} &= \Phi_2(\tilde{\beta}_1^T \mathbf{X}_j, \tilde{\beta}_2^T \mathbf{X}_j, \rho), \\ \Phi_{j10} &= \Phi_2(\tilde{\beta}_1^T \mathbf{X}_j, -\tilde{\beta}_2^T \mathbf{X}_j, -\rho) = -\Phi_{j11} + \Phi(\tilde{\beta}_1^T \mathbf{X}_j), \\ \Phi_{j01} &= \Phi_2(-\tilde{\beta}_1^T \mathbf{X}_j, \tilde{\beta}_2^T \mathbf{X}_j, -\rho) = -\Phi_{j11} + \Phi(\tilde{\beta}_2^T \mathbf{X}_j), \\ \Phi_{j00} &= \Phi_2(-\tilde{\beta}_1^T \mathbf{X}_j, -\tilde{\beta}_2^T \mathbf{X}_j, \rho) = 1 + \Phi_{j11} - \Phi(\tilde{\beta}_1^T \mathbf{X}_j) - \Phi(\tilde{\beta}_2^T \mathbf{X}_j), \\ \Phi_j &= \Phi_{j11}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j10}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} + \Phi_{j01}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j00}.\end{aligned}$$

And

$$\begin{aligned}E(\tilde{Z}_{j1}) &= \tilde{\beta}_1^T \mathbf{X}_j + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) E(\tilde{\epsilon}_{j1} | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) \\ &\quad + \frac{\phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) + \frac{\rho\phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),\end{aligned}$$

$$\begin{aligned}E(\tilde{Z}_{j2}) &= \tilde{\beta}_2^T \mathbf{X}_j + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) E(\tilde{\epsilon}_{j2} | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) \\ &\quad + \frac{\rho\phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) + \frac{\phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),\end{aligned}$$

$$\begin{aligned}E(\tilde{Z}_{j1}^2) &= \tilde{\beta}_1^T \mathbf{X}_j (2E(\tilde{Z}_{j1}) - \tilde{\beta}_1^T \mathbf{X}_j) + 1 \\ &\quad + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) (E(\tilde{\epsilon}_{j1}^2 | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) - 1) \\ &\quad - \frac{\tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) - \frac{\rho^2 \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),\end{aligned}$$

$$\begin{aligned}E(\tilde{Z}_{j2}^2) &= \tilde{\beta}_2^T \mathbf{X}_j (2E(\tilde{Z}_{j2}) - \tilde{\beta}_2^T \mathbf{X}_j) + 1 \\ &\quad + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) (E(\tilde{\epsilon}_{j2}^2 | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) - 1) \\ &\quad - \frac{\rho^2 \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) - \frac{\tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),\end{aligned}$$

$$\begin{aligned}
E(\tilde{Z}_{j1}\tilde{Z}_{j2}) = & -\tilde{\beta}_1^T \mathbf{X}_j \tilde{\beta}_2^T \mathbf{X}_j + \tilde{\beta}_1^T \mathbf{X}_j E(\tilde{Z}_{j2}) + \tilde{\beta}_2^T \mathbf{X}_j E(\tilde{Z}_{j1}) + \rho \\
& + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) (E(\tilde{\epsilon}_{j1}\tilde{\epsilon}_{j2}|\tilde{\eta}_{j1}=1, \tilde{\eta}_{j2}=1, \mathbf{X}_j; \boldsymbol{\theta}) - \rho) \\
& - \frac{\rho \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) - \frac{\rho \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),
\end{aligned}$$

where

$$\begin{aligned}
E(\tilde{\epsilon}_{j1}|\tilde{\eta}_{j1}=1, \tilde{\eta}_{j2}=1, \mathbf{X}_j; \boldsymbol{\theta}) = & \frac{1}{\Phi_{j11}} \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} \rho \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j2}|\tilde{\eta}_{j1}=1, \tilde{\eta}_{j2}=1, \mathbf{X}_j; \boldsymbol{\theta}) = & \frac{1}{\Phi_{j11}} \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} \rho \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j1}^2|\tilde{\eta}_{j1}=1, \tilde{\eta}_{j2}=1, \mathbf{X}_j; \boldsymbol{\theta}) = & 1 - \frac{1}{\Phi_{j11}} \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c) \\
& - \frac{1}{\Phi_{j11}} \rho^2 \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} c^{-1} \rho \phi(\tilde{\beta}_2^T \mathbf{X}_j) \phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j2}^2|\tilde{\eta}_{j1}=1, \tilde{\eta}_{j2}=1, \mathbf{X}_j; \boldsymbol{\theta}) = & 1 - \frac{1}{\Phi_{j11}} \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c) \\
& - \frac{1}{\Phi_{j11}} \rho^2 \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} c^{-1} \rho \phi(\tilde{\beta}_1^T \mathbf{X}_j) \phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j1}\tilde{\epsilon}_{j2}|\tilde{\eta}_{j1}=1, \tilde{\eta}_{j2}=1, \mathbf{X}_j; \boldsymbol{\theta}) = & \rho - \frac{1}{\Phi_{j11}} \rho \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c) \\
& - \frac{1}{\Phi_{j11}} \rho \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} c^{-1} \phi(\tilde{\beta}_1^T \mathbf{X}_j) \phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c),
\end{aligned}$$

$$c = \frac{1}{\sqrt{1-\rho^2}}.$$

Evaluate L

$$L = \sum_{j=1}^M \log \Phi_j.$$

- M-step: For $k = 1, 2$, update

$$\begin{aligned}\tilde{\alpha}_k &= -\frac{\sum_{j=1}^M E[\tilde{\eta}_{jk}]}{\sum_{j=1}^M E[\tilde{\eta}_{jk}] \log \tilde{P}_{jk}}, \\ \boldsymbol{\gamma} &= E[\tilde{\mathbf{Z}}^T] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}, \\ \boldsymbol{\Sigma} &= \frac{1}{M} \left[\left(\sum_{j=1}^M E[\tilde{\mathbf{Z}}_j \tilde{\mathbf{Z}}_j^T] \right) - \boldsymbol{\gamma} \mathbf{X}^T E[\tilde{\mathbf{Z}}] \right].\end{aligned}$$

Then we reduce to the original parameters using the reduction function:

$$\begin{aligned}\tilde{\boldsymbol{\beta}} &= \mathbf{D}^{-1} \boldsymbol{\gamma}, \\ \tilde{\mathbf{R}} &= \mathbf{D}^{-1} \boldsymbol{\Sigma} \mathbf{D}^{-1}.\end{aligned}$$

- Check the convergence of L .

3 The details of the proposed algorithm

If the correlation coefficient ρ is zero, we can analyze the traits independently, which provides warm starts for generating our three-stage algorithm for bLPM. In the first stage, we set all the coefficients in $\tilde{\beta}$ (except the intercept term) and the correlation coefficient ρ to be zero, and run an EM algorithm to obtain the estimates for $\tilde{\alpha}$ and $\tilde{\beta}_0$. Then in the second stage, we use the estimated parameters as the starting point to obtain $\tilde{\alpha}$ and $\tilde{\beta}$ using a PX-EM algorithm. Finally, in the third stage, we run the above PX-EM algorithm, using initial parameters those obtained in the second stage, and update $\tilde{\alpha}$, $\tilde{\beta}$ and ρ simultaneously until convergence. Since our algorithm is based on the framework of EM and PX-EM, the likelihood is guaranteed to increase at each iteration.

Stage 1:

Suppose all the coefficients in $\tilde{\beta}$ (except the intercept term) and the correlation coefficient ρ are zero, then the model becomes

$$\tilde{P}_{jk} \sim \begin{cases} U[0, 1], & \tilde{\eta}_{jk} = 0, \\ Beta(\tilde{\alpha}_k, 1), & \tilde{\eta}_{jk} = 1, \end{cases}$$

$$\pi_{1k} = \Pr(\tilde{\eta}_{jk} = 1),$$

where $\pi_{1k} = 1 - \Phi(-\tilde{\beta}_{k0})$, $\tilde{\beta}_{k0}$ is the intercept term in the probit model.

We can use EM algorithm to estimate the parameters $\tilde{\alpha}_k$ and π_{1k} .

The complete-data log-likelihood is

$$\begin{aligned} & \log \Pr(\tilde{\mathbf{P}}, \tilde{\eta} | \tilde{\alpha}, \boldsymbol{\pi}_1) \\ &= \log \Pr(\tilde{\mathbf{P}} | \tilde{\eta}; \tilde{\alpha}) + \log \Pr(\tilde{\eta} | \boldsymbol{\pi}_1) \\ &= \sum_{j=1}^M \sum_{k=1}^2 [\tilde{\eta}_{jk} (\log \tilde{\alpha}_k + (\tilde{\alpha}_k - 1) \log \tilde{P}_{jk} + \pi_{1k}) + (1 - \tilde{\eta}_{jk}) \log (1 - \pi_{1k})]. \end{aligned}$$

In the E step, we compute the posterior

$$\tilde{\eta}_{jk} = \Pr(\tilde{\eta}_{jk} = 1 | \tilde{\mathbf{P}}, \tilde{\alpha}, \boldsymbol{\pi}_1) = \frac{\pi_{1k} \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1}}{\pi_{1k} \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} + 1 - \pi_{1k}},$$

and obtain the Q function

$$Q = \sum_{j=1}^M \sum_{k=1}^2 [\tilde{\eta}_{jk} (\log \tilde{\alpha}_k + (\tilde{\alpha}_k - 1) \log \tilde{P}_{jk} + \pi_{1k}) + (1 - \tilde{\eta}_{jk}) \log (1 - \pi_{1k})].$$

The incomplete log-likelihood can be evaluated as

$$L = \sum_{j=1}^M \sum_{k=1}^2 \log (\pi_{1k} \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} + 1 - \pi_{1k}).$$

In the M step, we update $\tilde{\alpha}_k$ and π_{1k} by maximizing the Q function. We have

$$\begin{aligned}\tilde{\alpha}_k &= -\frac{\sum_{j=1}^M \tilde{\eta}_{jk}}{\sum_{j=1}^M \tilde{\eta}_{jk} \log \tilde{P}_{jk}}, \\ \pi_{1k} &= \frac{1}{M} \sum_{j=1}^M \tilde{\eta}_{jk}.\end{aligned}$$

Implementation

Input: $\tilde{\mathbf{P}}$. Initialize: $\tilde{\boldsymbol{\alpha}} = c(0.1, 0.1)$, $\boldsymbol{\pi}_1 = c(0.1, 0.1)$. Output: $\tilde{\boldsymbol{\alpha}}$, $\boldsymbol{\pi}_1$.

- Initialize $\tilde{\boldsymbol{\alpha}} = c(0.1, 0.1)$, $\boldsymbol{\pi}_1 = c(0.1, 0.1)$.
- E-step: For $j = 1, \dots, M$ and $k = 1, 2$, calculate

$$\tilde{\eta}_{jk} = \frac{\pi_{1k} \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1}}{\pi_{1k} \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} + 1 - \pi_{1k}}.$$

Evaluate L

$$L = \sum_{j=1}^M \sum_{k=1}^2 \log \left(\pi_{1k} \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} + 1 - \pi_{1k} \right).$$

- M-step: For $k = 1, 2$, update

$$\begin{aligned}\tilde{\alpha}_k &= -\frac{\sum_{j=1}^M \tilde{\eta}_{jk}}{\sum_{j=1}^M \tilde{\eta}_{jk} \log \tilde{P}_{jk}}, \\ \pi_{1k} &= \frac{1}{M} \sum_{j=1}^M \tilde{\eta}_{jk}.\end{aligned}$$

- Check the convergence of L .

Stage 2:

Suppose the correlation coefficient ρ is zero, then the model becomes

$$\tilde{P}_{jk} \sim \begin{cases} U[0, 1], & \tilde{\eta}_{jk} = 0, \\ Beta(\tilde{\alpha}_k, 1), & \tilde{\eta}_{jk} = 1, \end{cases}$$

$$\tilde{\eta}_{jk} = \begin{cases} 1, & \text{if } \tilde{Z}_{jk} > 0, \\ 0, & \text{if } \tilde{Z}_{jk} \leq 0, \end{cases}$$

$$\tilde{Z}_{jk} = \tilde{\beta}_k^T \mathbf{X}_j + \tilde{\epsilon}_{jk},$$

$$\tilde{\epsilon}_{jk} \sim N(0, 1),$$

where $\tilde{\mathbf{Z}} \in \mathbb{R}^{M \times 2}$ is the latent variable in probit model, $\mathbf{X} \in \mathbb{R}^{M \times (D+1)}$ is the design matrix of functional annotations, comprised of an intercept and D annotations, $\tilde{\beta} \in \mathbb{R}^{2 \times (D+1)}$ is a matrix of the coefficients, $\tilde{\beta}_k$ is a vector of the k -th row of $\tilde{\beta}$. For the j -th SNP, \mathbf{X}_j is a vector containing the j -th row of \mathbf{X} .

The expanded model is

$$\tilde{P}_{jk} \sim \begin{cases} U[0, 1], & \tilde{\eta}_{jk} = 0, \\ Beta(\tilde{\alpha}_k, 1), & \tilde{\eta}_{jk} = 1, \end{cases}$$

$$\tilde{\eta}_{jk} = \begin{cases} 1, & \text{if } \tilde{Z}_{jk} > 0, \\ 0, & \text{if } \tilde{Z}_{jk} \leq 0, \end{cases}$$

$$\tilde{Z}_{jk} = \gamma_k^T \mathbf{X}_j + \tilde{\epsilon}_{jk},$$

$$\tilde{\epsilon}_{jk} \sim N(0, \sigma_k^2),$$

where $\gamma_k = \sigma_k \tilde{\beta}_k$ and σ_k is the auxiliary parameter whose value is fixed at 1 in the original model.

Let $\boldsymbol{\theta} = \{\tilde{\alpha}, \tilde{\beta}\}$ and $\boldsymbol{\Theta} = \{\tilde{\alpha}, \gamma, \sigma\}$ be the collection of model parameters in the original model and the expanded model, respectively.

For the expanded model, the complete-data log-likelihood is

$$\begin{aligned} & \log \Pr(\tilde{\mathbf{P}}, \tilde{\eta}, \tilde{\mathbf{Z}} | \mathbf{X}; \boldsymbol{\Theta}) \\ &= \log \Pr(\tilde{\mathbf{P}} | \tilde{\eta}; \tilde{\alpha}) + \log \Pr(\tilde{\mathbf{Z}} | \mathbf{X}; \gamma, \Sigma) \\ &= \sum_{j=1}^M \sum_{k=1}^2 \tilde{\eta}_{jk} (\log \tilde{\alpha}_k + (\tilde{\alpha}_k - 1) \log \tilde{P}_{jk}) \\ &+ \sum_{j=1}^M \sum_{k=1}^2 \left[-\frac{1}{2} \log (2\pi) - \frac{1}{2} \log \sigma_k^2 - \frac{1}{2\sigma_k^2} (\tilde{Z}_{jk} - \gamma_k^T \mathbf{X}_j)^2 \right]. \end{aligned}$$

The Q function is evaluated as follows

$$\begin{aligned}
Q &= E_{\tilde{\eta}, \tilde{\mathbf{Z}}} \log \Pr(\tilde{\mathbf{P}}, \tilde{\eta}, \tilde{\mathbf{Z}} | \mathbf{X}; \boldsymbol{\Theta}) \\
&= \sum_{j=1}^M \sum_{k=1}^2 E[\tilde{\eta}_{jk}] (\log \tilde{\alpha}_k + (\tilde{\alpha}_k - 1) \log \tilde{P}_{jk}) \\
&\quad + \sum_{j=1}^M \sum_{k=1}^2 \left[-\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \sigma_k^2 - \frac{1}{2\sigma_k^2} E[\tilde{Z}_{jk}^2] + \frac{1}{\sigma_k^2} E[\tilde{Z}_{jk}] \boldsymbol{\gamma}_k^T \mathbf{X}_j - \frac{1}{2\sigma_k^2} (\boldsymbol{\gamma}_k^T \mathbf{X}_j)^2 \right],
\end{aligned}$$

where the expectation is calculated based on the current $\boldsymbol{\theta}$ in the original model.

In the PX-E step, we compute the posterior of $\tilde{\eta}_{jk}$ as follows

$$E[\tilde{\eta}_{jk}] = \Pr(\tilde{\eta}_{jk} = 1 | \tilde{\mathbf{P}}, \mathbf{X}; \boldsymbol{\theta}) = \frac{\Phi(\tilde{\beta}_k^T \mathbf{X}_j) \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1}}{\Phi(\tilde{\beta}_k^T \mathbf{X}_j) \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} + 1 - \Phi(\tilde{\beta}_k^T \mathbf{X}_j)}.$$

The posterior of \tilde{Z}_{jk} given $\tilde{\eta}_{jk}$ is a truncated bivariate normal distribution. Using the properties of truncated bivariate normal distribution, we can get

$$\begin{aligned}
E(\tilde{Z}_{jk} | \tilde{\mathbf{P}}, \mathbf{X}; \boldsymbol{\theta}) &= \tilde{\beta}_k^T \mathbf{X}_j + E[\tilde{\eta}_{jk}] \frac{\phi(\tilde{\beta}_k^T \mathbf{X}_j)}{\Phi(\tilde{\beta}_k^T \mathbf{X}_j)} - (1 - E[\tilde{\eta}_{jk}]) \frac{\phi(\tilde{\beta}_k^T \mathbf{X}_j)}{1 - \Phi(\tilde{\beta}_k^T \mathbf{X}_j)} \\
&= \tilde{\beta}_k^T \mathbf{X}_j + \frac{\phi(\tilde{\beta}_k^T \mathbf{X}_j) (\tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} - 1)}{\Phi(\tilde{\beta}_k^T \mathbf{X}_j) \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} + 1 - \Phi(\tilde{\beta}_k^T \mathbf{X}_j)} \\
&= \tilde{\beta}_k^T \mathbf{X}_j + v_{jk}, \\
E(\tilde{Z}_{jk}^2 | \tilde{\mathbf{P}}, \mathbf{X}; \boldsymbol{\theta}) &= \tilde{\beta}_k^T \mathbf{X}_j E(\tilde{Z}_{jk} | \tilde{\mathbf{P}}, \mathbf{X}; \boldsymbol{\theta}) + 1.
\end{aligned}$$

The incomplete log-likelihood can be evaluated as

$$L = \sum_{j=1}^M \sum_{k=1}^2 \log \left(\Phi(\tilde{\beta}_k^T \mathbf{X}_j) \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} + 1 - \Phi(\tilde{\beta}_k^T \mathbf{X}_j) \right).$$

In the PX-M step, we set the derivative of the Q function with respect to the parameters in $\boldsymbol{\Theta}$ be zero and

$$\frac{\partial Q}{\partial \tilde{\alpha}_k} = \sum_{j=1}^M E[\tilde{\eta}_{jk}] \left(\frac{1}{\tilde{\alpha}_k} + \log \tilde{P}_{jk} \right) = 0,$$

$$\frac{\partial Q}{\partial \boldsymbol{\gamma}_k} = \sum_{j=1}^M \left[\frac{1}{\sigma_k^2} E[\tilde{Z}_{jk}] \mathbf{X}_j^T - \frac{1}{\sigma_k^2} \boldsymbol{\gamma}_k^T \mathbf{X}_j \mathbf{X}_j^T \right] = 0,$$

$$\frac{\partial Q}{\partial \sigma_k^2} = \sum_{i=1}^M \left[-\frac{1}{2\sigma_k^2} + \frac{1}{2\sigma_k^4} E[\tilde{Z}_{jk}^2] - \frac{1}{\sigma_k^4} E[\tilde{Z}_{jk}] \boldsymbol{\gamma}_k^T \mathbf{X}_j + \frac{1}{2\sigma_k^4} (\boldsymbol{\gamma}_k^T \mathbf{X}_j)^2 \right] = 0.$$

Then we can obtain the updating equations

$$\tilde{\alpha}_k = -\frac{\sum_{j=1}^M E[\tilde{\eta}_{jk}]}{\sum_{j=1}^M E[\tilde{\eta}_{jk}] \log \tilde{P}_{jk}},$$

$$\boldsymbol{\gamma} = E[\tilde{\mathbf{Z}}^T] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1},$$

$$\sigma_k^2 = \frac{1}{M} \sum_{j=1}^M \left[E[\tilde{Z}_{jk}^2] - 2E[\tilde{Z}_{jk}] \boldsymbol{\gamma}_k^T \mathbf{X}_j + (\boldsymbol{\gamma}_k^T \mathbf{X}_j)^2 \right].$$

We reduce to the original parameters using the reduction function:

$$\tilde{\beta}_k = \frac{\boldsymbol{\gamma}_k}{\sigma_k}.$$

Implementation

Input: $\tilde{\mathbf{P}}$, \mathbf{X} , $\tilde{\boldsymbol{\alpha}}$, $\tilde{\beta}_{k0} = -qnorm(\pi_{1k})$. Output: $\tilde{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\beta}}$.

- Initialize $\tilde{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\beta}} = \begin{pmatrix} \tilde{\beta}_{10} & 0 & \cdots & 0 \\ \tilde{\beta}_{20} & 0 & \cdots & 0 \end{pmatrix}$.
- E-step: For $j = 1, \dots, M$ and $k = 1, 2$, calculate

$$E[\tilde{\eta}_{jk}] = \frac{\Phi(\tilde{\boldsymbol{\beta}}_k^T \mathbf{X}_j) \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1}}{\Phi(\tilde{\boldsymbol{\beta}}_k^T \mathbf{X}_j) \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} + 1 - \Phi(\tilde{\boldsymbol{\beta}}_k^T \mathbf{X}_j)}.$$

$$\begin{aligned} E(\tilde{Z}_{jk} | \tilde{\mathbf{P}}, \mathbf{X}; \boldsymbol{\theta}) &= \tilde{\boldsymbol{\beta}}_k^T \mathbf{X}_j + \frac{\phi(\tilde{\boldsymbol{\beta}}_k^T \mathbf{X}_j) (\tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} - 1)}{\Phi(\tilde{\boldsymbol{\beta}}_k^T \mathbf{X}_j) \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} + 1 - \Phi(\tilde{\boldsymbol{\beta}}_k^T \mathbf{X}_j)} \\ &= \tilde{\boldsymbol{\beta}}_k^T \mathbf{X}_j + v_{jk}, \\ E(\tilde{Z}_{jk}^2 | \tilde{\mathbf{P}}, \mathbf{X}; \boldsymbol{\theta}) &= \tilde{\boldsymbol{\beta}}_k^T \mathbf{X}_j E(\tilde{Z}_{jk} | \tilde{\mathbf{P}}, \mathbf{X}; \boldsymbol{\theta}) + 1. \end{aligned}$$

Evaluate L

$$L = \sum_{j=1}^M \sum_{k=1}^2 \log \left(\Phi(\tilde{\boldsymbol{\beta}}_k^T \mathbf{X}_j) \tilde{\alpha}_k \tilde{P}_{jk}^{\tilde{\alpha}_k - 1} + 1 - \Phi(\tilde{\boldsymbol{\beta}}_k^T \mathbf{X}_j) \right).$$

- M-step: For $k = 1, 2$, update

$$\tilde{\alpha}_k = -\frac{\sum_{j=1}^M E[\tilde{\eta}_{jk}]}{\sum_{j=1}^M E[\tilde{\eta}_{jk}] \log \tilde{P}_{jk}},$$

$$\boldsymbol{\gamma} = E[\tilde{\mathbf{Z}}^T] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1},$$

$$\sigma_k^2 = \frac{1}{M} \sum_{j=1}^M \left[E[\tilde{Z}_{jk}^2] - 2E[\tilde{Z}_{jk}] \boldsymbol{\gamma}_k^T \mathbf{X}_j + (\boldsymbol{\gamma}_k^T \mathbf{X}_j)^2 \right].$$

Then we reduce to the original parameters using the reduction function:

$$\tilde{\beta}_k = \frac{\gamma_k}{\sigma_k}.$$

- Check the convergence of L .

Stage 3:

Implementation

Input: $\tilde{\mathbf{P}}$, \mathbf{X} , $\tilde{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\beta}}$. Initialize: $\rho = 0$. Output: $\tilde{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\beta}}$, ρ .

- Initialize $\tilde{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\beta}}$ using warm starts and $\rho = 0$.
- E-step: For $j = 1, \dots, M$ and $k = 1, 2$, calculate

$$E[\tilde{\eta}_{j1}] = \frac{\Phi_{j11}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j10}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}}{\Phi_j},$$

$$E[\tilde{\eta}_{j2}] = \frac{\Phi_{j11}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j01}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1}}{\Phi_j},$$

where

$$\begin{aligned}\Phi_{j11} &= \Phi_2\left(\tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j, \tilde{\boldsymbol{\beta}}_2^T \mathbf{X}_j, \rho\right), \\ \Phi_{j10} &= \Phi_2\left(\tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j, -\tilde{\boldsymbol{\beta}}_2^T \mathbf{X}_j, -\rho\right) = -\Phi_{j11} + \Phi\left(\tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j\right), \\ \Phi_{j01} &= \Phi_2\left(-\tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j, \tilde{\boldsymbol{\beta}}_2^T \mathbf{X}_j, -\rho\right) = -\Phi_{j11} + \Phi\left(\tilde{\boldsymbol{\beta}}_2^T \mathbf{X}_j\right), \\ \Phi_{j00} &= \Phi_2\left(-\tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j, -\tilde{\boldsymbol{\beta}}_2^T \mathbf{X}_j, \rho\right) = 1 + \Phi_{j11} - \Phi\left(\tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j\right) - \Phi\left(\tilde{\boldsymbol{\beta}}_2^T \mathbf{X}_j\right), \\ \Phi_j &= \Phi_{j11}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j10}\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} + \Phi_{j01}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + \Phi_{j00}.\end{aligned}$$

And

$$\begin{aligned}E(\tilde{Z}_{j1}) &= \tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) E(\tilde{\epsilon}_{j1}|\tilde{\eta}_{j1}=1, \tilde{\eta}_{j2}=1, \mathbf{X}_j; \boldsymbol{\theta}) \\ &\quad + \frac{\phi(\tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) + \frac{\rho\phi(\tilde{\boldsymbol{\beta}}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),\end{aligned}$$

$$\begin{aligned}E(\tilde{Z}_{j2}) &= \tilde{\boldsymbol{\beta}}_2^T \mathbf{X}_j + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) E(\tilde{\epsilon}_{j2}|\tilde{\eta}_{j1}=1, \tilde{\eta}_{j2}=1, \mathbf{X}_j; \boldsymbol{\theta}) \\ &\quad + \frac{\rho\phi(\tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) + \frac{\phi(\tilde{\boldsymbol{\beta}}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),\end{aligned}$$

$$\begin{aligned}E(\tilde{Z}_{j1}^2) &= \tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j (2E(\tilde{Z}_{j1}) - \tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j) + 1 \\ &\quad + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1}\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) (E(\tilde{\epsilon}_{j1}^2|\tilde{\eta}_{j1}=1, \tilde{\eta}_{j2}=1, \mathbf{X}_j; \boldsymbol{\theta}) - 1) \\ &\quad - \frac{\tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j \phi(\tilde{\boldsymbol{\beta}}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1\tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) - \frac{\rho^2 \tilde{\boldsymbol{\beta}}_2^T \mathbf{X}_j \phi(\tilde{\boldsymbol{\beta}}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2\tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),\end{aligned}$$

$$\begin{aligned}
E(\tilde{Z}_{j2}^2) = & \tilde{\beta}_2^T \mathbf{X}_j \left(2E(\tilde{Z}_{j2}) - \tilde{\beta}_2^T \mathbf{X}_j \right) + 1 \\
& + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) (E(\tilde{\epsilon}_{j2}^2 | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) - 1) \\
& - \frac{\rho^2 \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) - \frac{\tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{Z}_{j1} \tilde{Z}_{j2}) = & -\tilde{\beta}_1^T \mathbf{X}_j \tilde{\beta}_2^T \mathbf{X}_j + \tilde{\beta}_1^T \mathbf{X}_j E(\tilde{Z}_{j2}) + \tilde{\beta}_2^T \mathbf{X}_j E(\tilde{Z}_{j1}) + \rho \\
& + \frac{\Phi_{j11}}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - \tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - \tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} + 1) (E(\tilde{\epsilon}_{j1} \tilde{\epsilon}_{j2} | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) - \rho) \\
& - \frac{\rho \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_1 \tilde{P}_{j1}^{\tilde{\alpha}_1-1} - 1) - \frac{\rho \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j)}{\Phi_j} (\tilde{\alpha}_2 \tilde{P}_{j2}^{\tilde{\alpha}_2-1} - 1),
\end{aligned}$$

where

$$\begin{aligned}
E(\tilde{\epsilon}_{j1} | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) = & \frac{1}{\Phi_{j11}} \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} \rho \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j2} | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) = & \frac{1}{\Phi_{j11}} \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} \rho \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j1}^2 | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) = & 1 - \frac{1}{\Phi_{j11}} \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c) \\
& - \frac{1}{\Phi_{j11}} \rho^2 \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} c^{-1} \rho \phi(\tilde{\beta}_2^T \mathbf{X}_j) \phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j2}^2 | \tilde{\eta}_{j1} = 1, \tilde{\eta}_{j2} = 1, \mathbf{X}_j; \boldsymbol{\theta}) = & 1 - \frac{1}{\Phi_{j11}} \tilde{\beta}_2^T \mathbf{X}_j \phi(\tilde{\beta}_2^T \mathbf{X}_j) \Phi((\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j) c) \\
& - \frac{1}{\Phi_{j11}} \rho^2 \tilde{\beta}_1^T \mathbf{X}_j \phi(\tilde{\beta}_1^T \mathbf{X}_j) \Phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c) \\
& + \frac{1}{\Phi_{j11}} c^{-1} \rho \phi(\tilde{\beta}_1^T \mathbf{X}_j) \phi((\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j) c),
\end{aligned}$$

$$\begin{aligned}
E(\tilde{\epsilon}_{j1}\tilde{\epsilon}_{j2}|\tilde{\eta}_{j1}=1, \tilde{\eta}_{j2}=1, \mathbf{X}_j; \boldsymbol{\theta}) = & \rho - \frac{1}{\Phi_{j11}} \rho \tilde{\beta}_1^T \mathbf{X}_j \phi\left(\tilde{\beta}_1^T \mathbf{X}_j\right) \Phi\left(\left(\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j\right) c\right) \\
& - \frac{1}{\Phi_{j11}} \rho \tilde{\beta}_2^T \mathbf{X}_j \phi\left(\tilde{\beta}_2^T \mathbf{X}_j\right) \Phi\left(\left(\tilde{\beta}_1^T \mathbf{X}_j - \rho \tilde{\beta}_2^T \mathbf{X}_j\right) c\right) \\
& + \frac{1}{\Phi_{j11}} c^{-1} \phi\left(\tilde{\beta}_1^T \mathbf{X}_j\right) \phi\left(\left(\tilde{\beta}_2^T \mathbf{X}_j - \rho \tilde{\beta}_1^T \mathbf{X}_j\right) c\right), \\
c = & \frac{1}{\sqrt{1-\rho^2}}.
\end{aligned}$$

Evaluate L

$$L = \sum_{j=1}^M \log \Phi_j.$$

- M-step: For $k = 1, 2$, update

$$\begin{aligned}
\tilde{\alpha}_k = & - \frac{\sum_{j=1}^M E[\tilde{\eta}_{jk}]}{\sum_{j=1}^M E[\tilde{\eta}_{jk}] \log \tilde{P}_{jk}}, \\
\boldsymbol{\gamma} = & E[\tilde{\mathbf{Z}}^T] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}, \\
\boldsymbol{\Sigma} = & \frac{1}{M} \left[\left(\sum_{j=1}^M E[\tilde{\mathbf{Z}}_j \tilde{\mathbf{Z}}_j^T] \right) - \boldsymbol{\gamma} \mathbf{X}^T E[\tilde{\mathbf{Z}}] \right].
\end{aligned}$$

Then we reduce to the original parameters using the reduction function:

$$\begin{aligned}
\tilde{\boldsymbol{\beta}} &= \mathbf{D}^{-1} \boldsymbol{\gamma}, \\
\tilde{\mathbf{R}} &= \mathbf{D}^{-1} \boldsymbol{\Sigma} \mathbf{D}^{-1}.
\end{aligned}$$

- Check the convergence of L .

4 Identification of risk SNPs

If we consider the traits separately, the association mapping of the j -th SNP on the k -th trait can be inferred from

$$\Pr(\eta_{jk} = 1 | P_{jk}, \mathbf{X}) = \frac{\Phi\left(\hat{\beta}_k^T \mathbf{X}_j\right) \hat{\alpha}_k P_{jk}^{\hat{\alpha}_k - 1}}{\Phi\left(\hat{\beta}_k^T \mathbf{X}_j\right) \hat{\alpha}_k P_{jk}^{\hat{\alpha}_k - 1} + 1 - \Phi\left(\hat{\beta}_k^T \mathbf{X}_j\right)}.$$

In this case, the relationship among traits is ignored and only the current GWAS data is used.

If another GWAS data set is integrated, risk SNPs for both the k -th trait and the k' -th trait can be inferred from

$$\Pr(\eta_{jk} = 1, \eta_{jk'} = 1 | P_{jk}, P_{jk'}, \mathbf{X}) = \frac{\Phi_{j11} \hat{\alpha}_k P_{jk}^{\hat{\alpha}_k - 1} \hat{\alpha}_{k'} P_{jk'}^{\hat{\alpha}_{k'} - 1}}{\Phi_j},$$

where

$$\begin{aligned}\Phi_{j11} &= \Phi_2\left(\hat{\beta}_k^T \mathbf{X}_j, \hat{\beta}_{k'}^T \mathbf{X}_j, \hat{\rho}_{kk'}\right), \\ \Phi_{j10} &= \Phi_2\left(\hat{\beta}_k^T \mathbf{X}_j, -\hat{\beta}_{k'}^T \mathbf{X}_j, -\hat{\rho}_{kk'}\right) = -\Phi_{j11} + \Phi\left(\hat{\beta}_k^T \mathbf{X}_j\right), \\ \Phi_{j01} &= \Phi_2\left(-\hat{\beta}_k^T \mathbf{X}_j, \hat{\beta}_{k'}^T \mathbf{X}_j, -\hat{\rho}_{kk'}\right) = -\Phi_{j11} + \Phi\left(\hat{\beta}_{k'}^T \mathbf{X}_j\right), \\ \Phi_{j00} &= \Phi_2\left(-\hat{\beta}_k^T \mathbf{X}_j, -\hat{\beta}_{k'}^T \mathbf{X}_j, \hat{\rho}_{kk'}\right) = 1 + \Phi_{j11} - \Phi\left(\hat{\beta}_k^T \mathbf{X}_j\right) - \Phi\left(\hat{\beta}_{k'}^T \mathbf{X}_j\right), \\ \Phi_j &= \Phi_{j11} \hat{\alpha}_k P_{jk}^{\hat{\alpha}_k - 1} \hat{\alpha}_{k'} P_{jk'}^{\hat{\alpha}_{k'} - 1} + \Phi_{j10} \hat{\alpha}_k P_{jk}^{\hat{\alpha}_k - 1} + \Phi_{j01} \hat{\alpha}_{k'} P_{jk'}^{\hat{\alpha}_{k'} - 1} + \Phi_{j00}.\end{aligned}$$

In addition, we can calculate the marginal posterior that the j -th SNP is associated with the k -th trait

$$\Pr(\eta_{jk} = 1 | P_{jk}, P_{jk'}, \mathbf{X}) = \Pr(\eta_{jk} = 1, \eta_{jk'} = 1 | P_{jk}, P_{jk'}, \mathbf{X}) + \Pr(\eta_{jk} = 1, \eta_{jk'} = 0 | P_{jk}, P_{jk'}, \mathbf{X}),$$

where

$$\Pr(\eta_{jk} = 1, \eta_{jk'} = 0 | P_{jk}, P_{jk'}, \mathbf{X}) = \frac{\Phi_{j10} \hat{\alpha}_k P_{jk}^{\hat{\alpha}_k - 1}}{\Phi_j}.$$

Similarly, we can consider more than two traits, e.g., three traits, and obtain

$$\Pr(\eta_{jk} = 1, \eta_{jk'} = 1, \eta_{jk''} = 1 | P_{jk}, P_{jk'}, P_{jk''}, \mathbf{X}),$$

$$\Pr(\eta_{jk} = 1, \eta_{jk'} = 1 | P_{jk}, P_{jk'}, P_{jk''}, \mathbf{X}),$$

and

$$\Pr(\eta_{jk} = 1 | P_{jk}, P_{jk'}, P_{jk''}, \mathbf{X}),$$

where

$$\Pr(\eta_{jk} = 1, \eta_{jk'} = 1, \eta_{jk''} = 1 | P_{jk}, P_{jk'}, P_{jk''}, \mathbf{X}) = \frac{\Phi_{j111} \hat{\alpha}_k P_{jk}^{\hat{\alpha}_k - 1} \hat{\alpha}_{k'} P_{jk'}^{\hat{\alpha}_{k'} - 1} \hat{\alpha}_{k''} P_{jk''}^{\hat{\alpha}_{k''} - 1}}{\Phi_j},$$

$$\begin{aligned}\Phi_{jlmn} &= \Phi_3(\boldsymbol{\mu}_{jkk'k''}, \mathbf{R}_{kk'k''}), \text{ for } l, m, n \in \{1, 0\}, \\ \Phi_j &= \sum_{l=0}^1 \sum_{m=0}^1 \sum_{n=0}^1 \Phi_{jlmn} \left(\hat{\alpha}_k P_{jk}^{\hat{\alpha}_k - 1} \right)^l \left(\hat{\alpha}_{k'} P_{jk'}^{\hat{\alpha}_{k'} - 1} \right)^m \left(\hat{\alpha}_{k''} P_{jk''}^{\hat{\alpha}_{k''} - 1} \right)^n,\end{aligned}$$

$$\begin{aligned}\boldsymbol{\mu}_{jkk'k''} &= \begin{pmatrix} (2l-1) \hat{\beta}_k^T \mathbf{X}_j \\ (2m-1) \hat{\beta}_{k'}^T \mathbf{X}_j \\ (2n-1) \hat{\beta}_{k''}^T \mathbf{X}_j \end{pmatrix}, \\ \mathbf{R}_{kk'k''} &= \begin{pmatrix} 1 & (2l-1)(2m-1)\hat{\rho}_{kk'} & (2l-1)(2n-1)\hat{\rho}_{kk''} \\ (2l-1)(2m-1)\hat{\rho}_{kk'} & 1 & (2m-1)(2n-1)\hat{\rho}_{k'k''} \\ (2l-1)(2n-1)\hat{\rho}_{kk''} & (2m-1)(2n-1)\hat{\rho}_{k'k''} & 1 \end{pmatrix},\end{aligned}$$

and $\Phi_3(\boldsymbol{\mu}, \Sigma) = \int_{-\infty}^{\mu_1} \int_{-\infty}^{\mu_2} \int_{-\infty}^{\mu_3} \phi_3(\mathbf{x}, \Sigma) d\mathbf{x}$,

$$\phi_3(\mathbf{x}, \Sigma) = \frac{\exp\left[-\frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{x}\right]}{\sqrt{(2\pi)^3 |\Sigma|}}.$$

Accordingly, we can calculate the local false discovery rate of the j -th SNP for the k -th trait when considering one or more traits simultaneously, i.e.,

$$\begin{aligned}fdr_1(P_{jk}) &= 1 - \Pr(\eta_{jk} = 1 | P_{jk}, \mathbf{X}), \\ fdr_1(P_{jk}, P_{jk'}) &= 1 - \Pr(\eta_{jk} = 1 | P_{jk}, P_{jk'}, \mathbf{X}), \\ fdr_1(P_{jk}, P_{jk'}, P_{jk''}) &= 1 - \Pr(\eta_{jk} = 1 | P_{jk}, P_{jk'}, P_{jk''}, \mathbf{X}),\end{aligned}$$

and for both the k -th and k' -th trait

$$\begin{aligned}fdr_2(P_{jk}, P_{jk'}) &= 1 - \Pr(\eta_{jk} = 1, \eta_{jk'} = 1 | P_{jk}, P_{jk'}, \mathbf{X}), \\ fdr_2(P_{jk}, P_{jk'}, P_{jk''}) &= 1 - \Pr(\eta_{jk} = 1, \eta_{jk'} = 1 | P_{jk}, P_{jk'}, P_{jk''}, \mathbf{X}),\end{aligned}$$

and for three traits

$$fdr_3(P_{jk}, P_{jk'}, P_{jk''}) = 1 - \Pr(\eta_{jk} = 1, \eta_{jk'} = 1, \eta_{jk''} = 1 | P_{jk}, P_{jk'}, P_{jk''}, \mathbf{X}).$$

We use the following approach to control the global false discovery rate (FDR). We sort SNPs by fdr from the smallest to the largest and regard the j -th re-ordered SNP as a risk SNP for each

of the three cases if

$$FDR_{(j)} = \frac{\sum_{i=1}^j fdr_{(i)}}{j} \leq \tau,$$

where $fdr_{(i)}$ is the i -th ordered fdr , $FDR_{(j)}$ is the corresponding global FDR, and τ is the threshold of global FDR.

5 Estimation of standard error of $\hat{\beta}$

To estimate the standard error of $\hat{\beta}_{kd}$, we consider the single trait case. The log-likelihood for the k -th trait is

$$L = \sum_{j=1}^M \log \left(\Phi \left(\beta_k^T \mathbf{X}_j \right) \alpha_k P_{jk}^{\alpha_k-1} + 1 - \Phi \left(\beta_k^T \mathbf{X}_j \right) \right).$$

The information matrix of log-likelihood can be computed by

$$\mathcal{I}(\hat{\theta}_k) = -H|_{\hat{\theta}_k} = -\nabla^2 L(\theta_k)|_{\hat{\theta}_k},$$

where

$$\nabla^2 L(\theta_k) = \begin{pmatrix} \frac{\partial^2 L}{\partial \alpha^2} & \frac{\partial^2 L}{\partial \alpha \partial \beta} \\ \frac{\partial^2 L}{\partial \beta^T \partial \alpha} & \frac{\partial^2 L}{\partial \beta^T \partial \beta} \end{pmatrix},$$

$$\begin{aligned} \frac{\partial L}{\partial \alpha} &= \sum_{j=1}^M \frac{\Phi \left(\beta^T \mathbf{X}_j \right) P_j^{\alpha-1} (1 + \alpha \log P_j)}{\Phi \left(\beta^T \mathbf{X}_j \right) \alpha P_j^{\alpha-1} + 1 - \Phi \left(\beta^T \mathbf{X}_j \right)}, \\ \frac{\partial L}{\partial \beta} &= \sum_{j=1}^M \frac{(\alpha P_j^{\alpha-1} - 1) \phi \left(\beta^T \mathbf{X}_j \right) \mathbf{X}_j^T}{\Phi \left(\beta^T \mathbf{X}_j \right) \alpha P_j^{\alpha-1} + 1 - \Phi \left(\beta^T \mathbf{X}_j \right)}, \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L}{\partial \alpha^2} &= - \sum_{j=1}^M \left(\frac{\Phi(\beta^T \mathbf{X}_j) P_j^{\alpha-1} (1 + \alpha \log P_j)}{\Phi(\beta^T \mathbf{X}_j) \alpha P_j^{\alpha-1} + 1 - \Phi(\beta^T \mathbf{X}_j)} \right)^2 \\
&\quad + \sum_{j=1}^M \frac{\Phi(\beta^T \mathbf{X}_j) P_j^{\alpha-1} (2 + \alpha \log P_j) \log P_j}{\Phi(\beta^T \mathbf{X}_j) \alpha P_j^{\alpha-1} + 1 - \Phi(\beta^T \mathbf{X}_j)}, \\
\frac{\partial^2 L}{\partial \alpha \partial \beta} &= - \sum_{j=1}^M \frac{\Phi(\beta^T \mathbf{X}_j) P_j^{\alpha-1} (1 + \alpha \log P_j) (\alpha P_j^{\alpha-1} - 1) \phi(\beta^T \mathbf{X}_j) \mathbf{X}_j^T}{\left(\Phi(\beta^T \mathbf{X}_j) \alpha P_j^{\alpha-1} + 1 - \Phi(\beta^T \mathbf{X}_j) \right)^2} \\
&\quad + \sum_{j=1}^M \frac{P_j^{\alpha-1} (1 + \alpha \log P_j) \phi(\beta^T \mathbf{X}_j) \mathbf{X}_j^T}{\Phi(\beta^T \mathbf{X}_j) \alpha P_j^{\alpha-1} + 1 - \Phi(\beta^T \mathbf{X}_j)}, \\
\frac{\partial^2 L}{\partial \beta^T \partial \beta} &= - \sum_{j=1}^M \frac{\left((\alpha P_j^{\alpha-1} - 1) \phi(\beta^T \mathbf{X}_j) \right)^2 \mathbf{X}_j \mathbf{X}_j^T}{\left(\Phi(\beta^T \mathbf{X}_j) \alpha P_j^{\alpha-1} + 1 - \Phi(\beta^T \mathbf{X}_j) \right)^2} \\
&\quad + \sum_{j=1}^M \frac{-(\alpha P_j^{\alpha-1} - 1) \phi(\beta^T \mathbf{X}_j) \beta^T \mathbf{X}_j \mathbf{X}_j \mathbf{X}_j^T}{\Phi(\beta^T \mathbf{X}_j) \alpha P_j^{\alpha-1} + 1 - \Phi(\beta^T \mathbf{X}_j)}.
\end{aligned}$$

Then the inverse of the observed information matrix is an estimator of the asymptotic covariance matrix $\text{Var}(\hat{\theta}_k) = [\mathcal{I}(\hat{\theta}_k)]^{-1}$.

6 More simulation results

6.1 The relationship graph

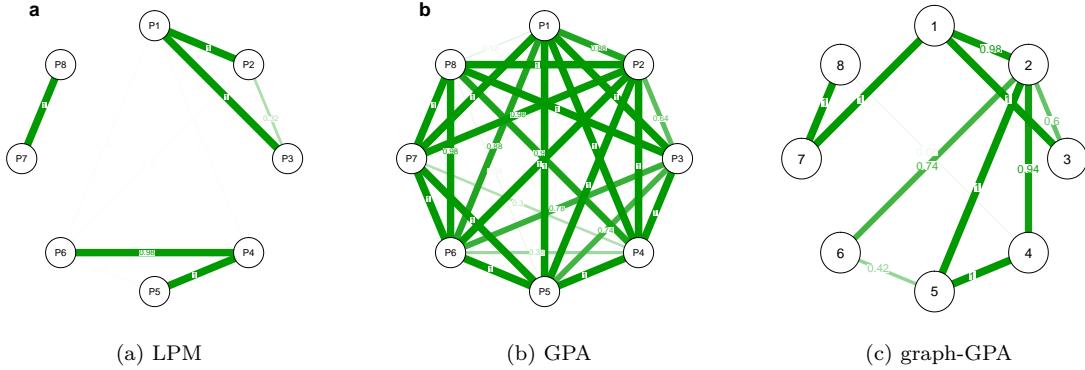


Figure S1: Relationship test graphs of (a) LPM, (b) GPA, and (c) graph-GPA. The numbers on the edges and the widths of the edges indicate the type I error or power of the relationship test for the connected traits. For LPM and GPA, we controlled family-wise error rate at 0.05. The results are summarized from 50 replications.

6.2 The relationship graph for the case when functional annotations have no role

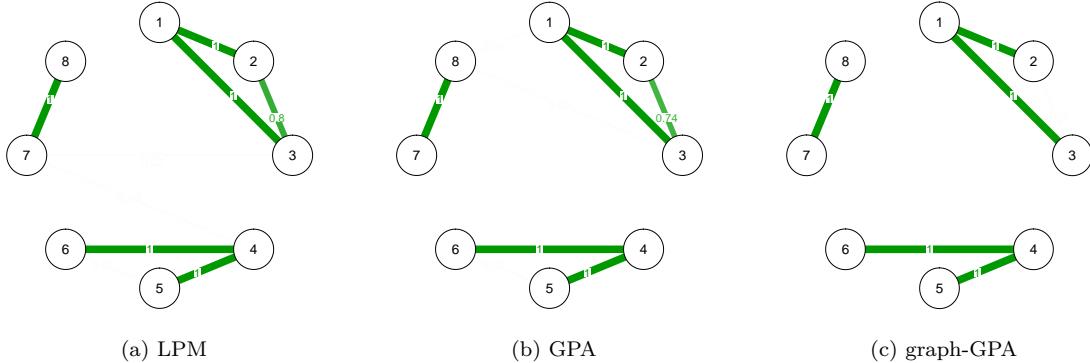


Figure S2: Relationship test graphs of (a) LPM, (b) GPA, and (c) graph-GPA when all functional annotations have no role. The numbers on the edges and the widths of the edges indicate the type I error or power of the relationship test for the connected traits. For LPM and GPA, we controlled family-wise error rate at 0.05. The results are summarized from 50 replications.

6.3 Performance in characterizing the correlations among traits with a dense correlation graph

We conducted simulations to investigate the performance of LPM, GPA (Chung et al., 2014) and graph-GPA (Chung et al., 2017) with a dense correlation graph among eight traits (denoted as P1, P2, ..., P8) which is depicted in Figure S3a. Here P1 - P6 (group 1) were all inter-connected and P7, P8 (group 2) were connected. We didn't assume all traits were correlated because it is also of great interest to see whether the type I error rate can be well controlled. All the other simulation settings and procedures were the same with the simulation of eight traits in the main text.

As shown in Figure S3b, LPM can provide accurate estimates of the correlations. The performance of LPM, GPA and graph-GPA for the relationship test is shown in Figure S4. The type I error rates can be evaluated from the pairs between groups and powers can be evaluated from the pairs within group. For LPM, the type I error rates are almost 0 and the powers are almost 1 except for one pairs (P2 and P6) where the correlation between them is relatively small ($\rho = 0.21$) and the signal strength is relatively weak ($\alpha_2 = 0.35$, $\alpha_6 = 0.55$). For GPA and graph-GPA, the type I error rate for some pairs of traits are not controlled because the relationship they measured does not adjust the effect of functional annotations.

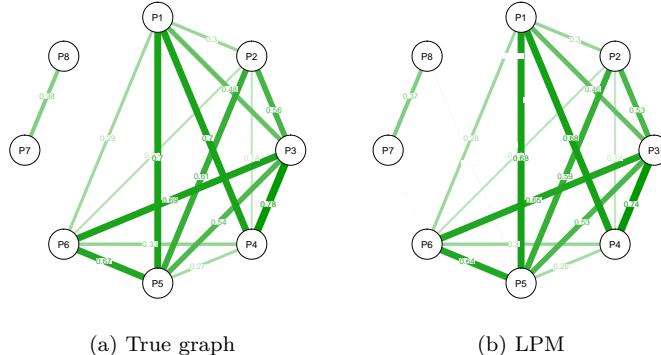


Figure S3: Correlation graphs. (a) true graph. (b) average estimated correlation graph using LPM. The numbers on the edges and the widths of the edges indicate the correlation between the connected traits. The results are summarized from 50 replications.

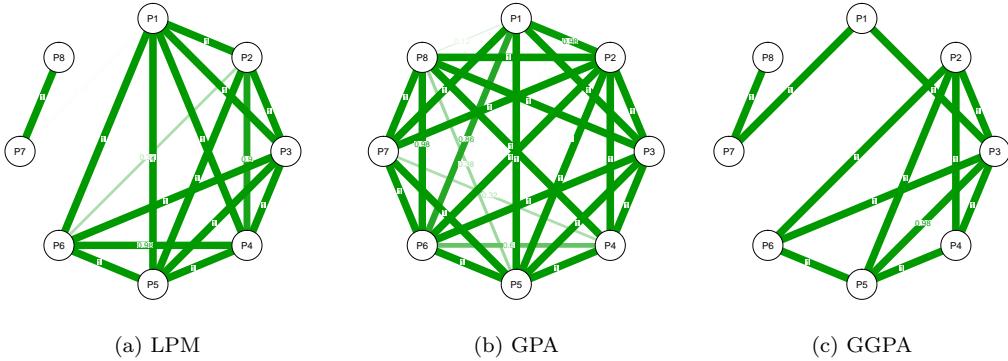


Figure S4: Relationship test graphs of (a) LPM, (b) GPA, and (c) graph-GPA. The numbers on the edges and the widths of the edges indicate the type I error or power of the relationship test for the connected traits. For LPM and GPA, we controlled family-wise error rate at 0.05. The results are summarized from 50 replications.

We also simulated the case when all functional annotations have no role. The results are shown in Figure S5 and Figure S6. The relationship test graphs of LPM and GPA are similar. For graph-GPA, some connections are not detected as graph-GPA represents a conditional independent structure.

To summarize, no matter the correlation graph assumed is sparse or dense, LPM performs quite well and the differences between GPA, graph-GPA and LPM are similar.

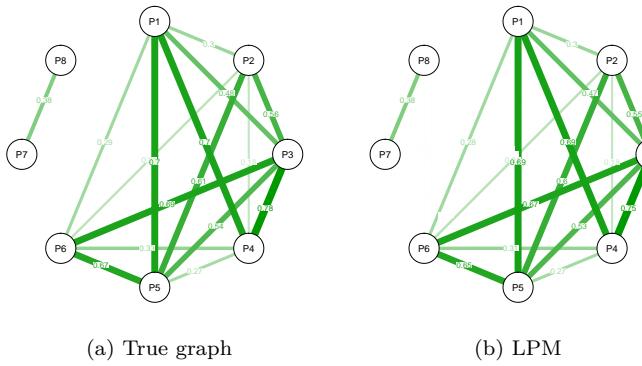


Figure S5: Correlation graphs when all functional annotations have no role. (a) true graph. (b) average estimated correlation graph using LPM. The numbers on the edges and the widths of the edges indicate the correlation between the connected traits. The results are summarized from 50 replications.

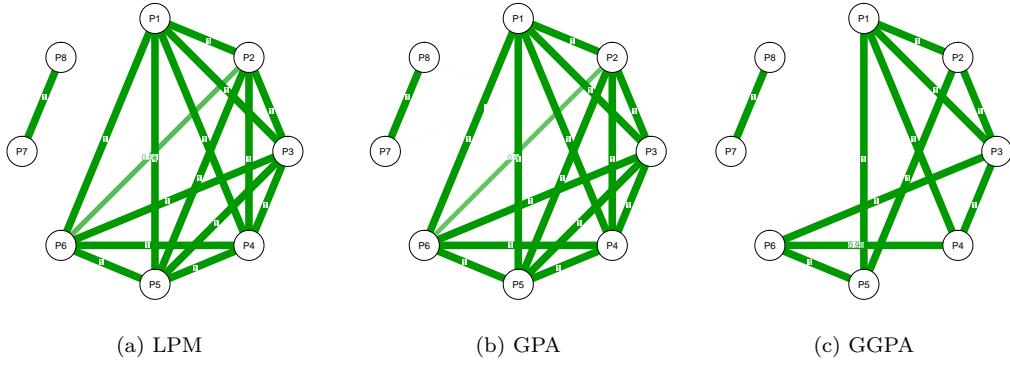


Figure S6: Relationship test graphs of (a) LPM, (b) GPA, and (c) graph-GPA when all functional annotations have no role. The numbers on the edges and the widths of the edges indicate the type I error or power of the relationship test for the connected traits. For LPM and GPA, we controlled family-wise error rate at 0.05. The results are summarized from 50 replications.

6.4 Type I error rate and power of LPM for the relationship test between two traits

In this simulation, we set the signal strength of the traits to be the same, i.e., $\alpha_1 = \alpha_2 = \alpha$. We varied α in $\{0.2, 0.4, 0.6\}$ and r in $\{0.25, 1, 4\}$ to obtain the type I error rate, and varied ρ in $\{0, 0.05, 0.1, 0.15, 0.2, 0.25\}$ to obtain the power of LPM for the relationship test between the traits.

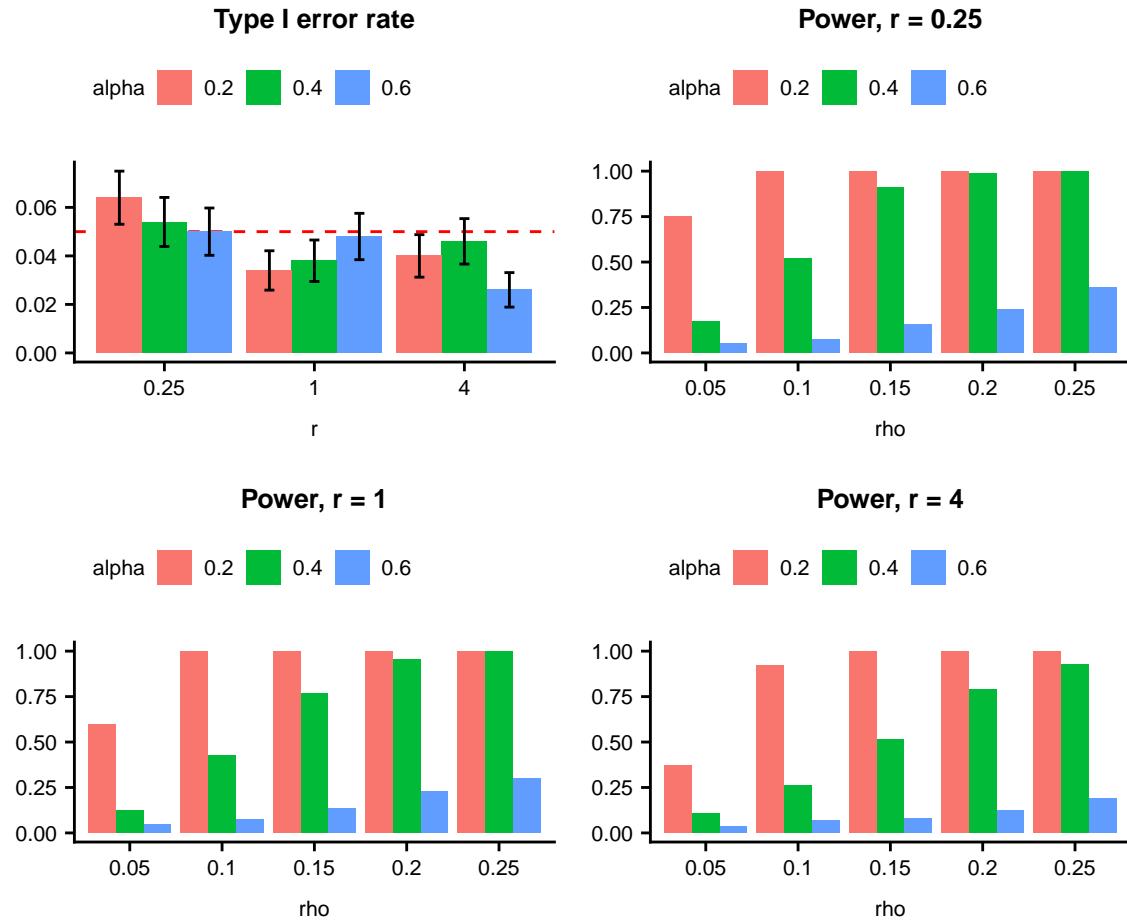


Figure S7: Type I error rate and power of LPM for the relationship test between two traits. The bars represent one standard error. We evaluate type I error rate and power at 0.05 significance level. The results are summarized from 500 replications.

6.5 The performance for the identification of risk SNPs of one specific trait

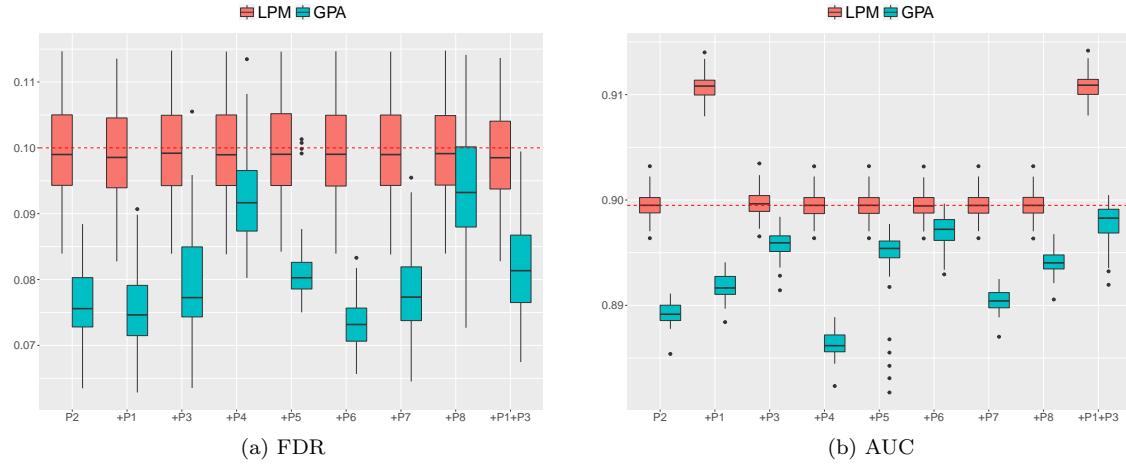


Figure S8: FDR and AUC of LPM and GPA for identification of risk SNPs for P2. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

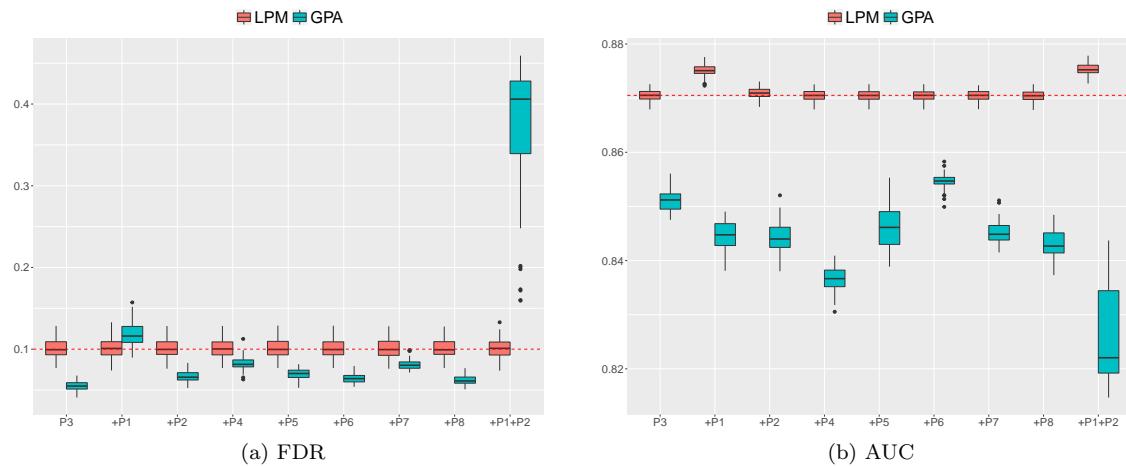


Figure S9: FDR and AUC of LPM and GPA for identification of risk SNPs for P3. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

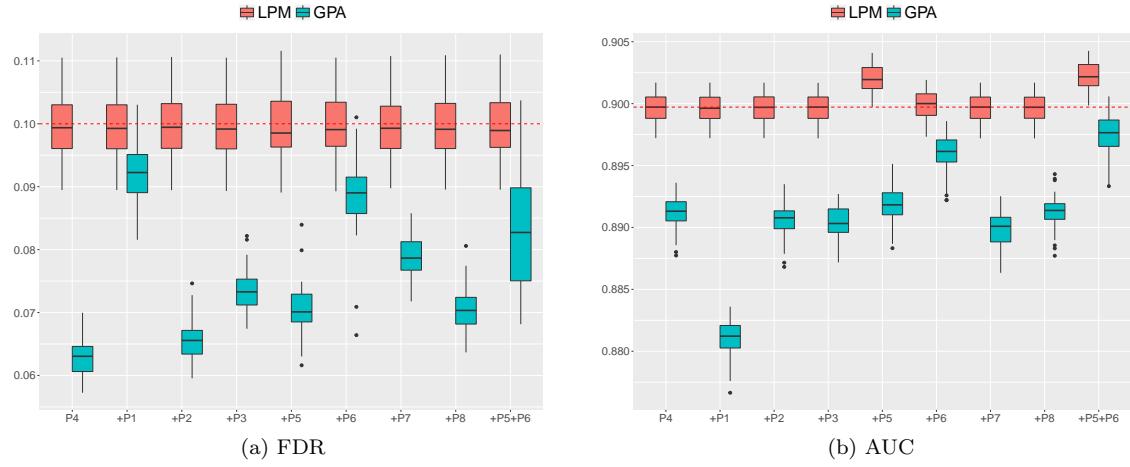


Figure S10: FDR and AUC of LPM and GPA for identification of risk SNPs for P4. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

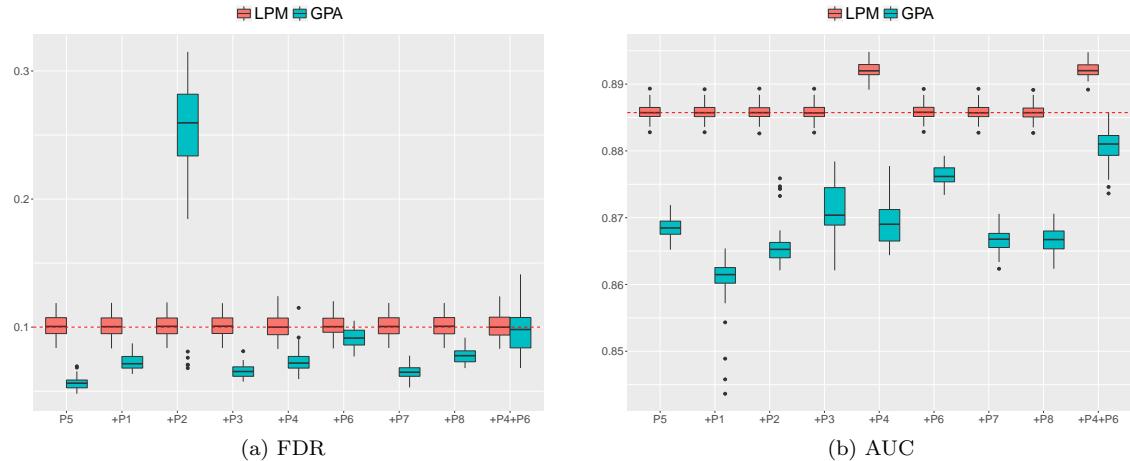


Figure S11: FDR and AUC of LPM and GPA for identification of risk SNPs for P5. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

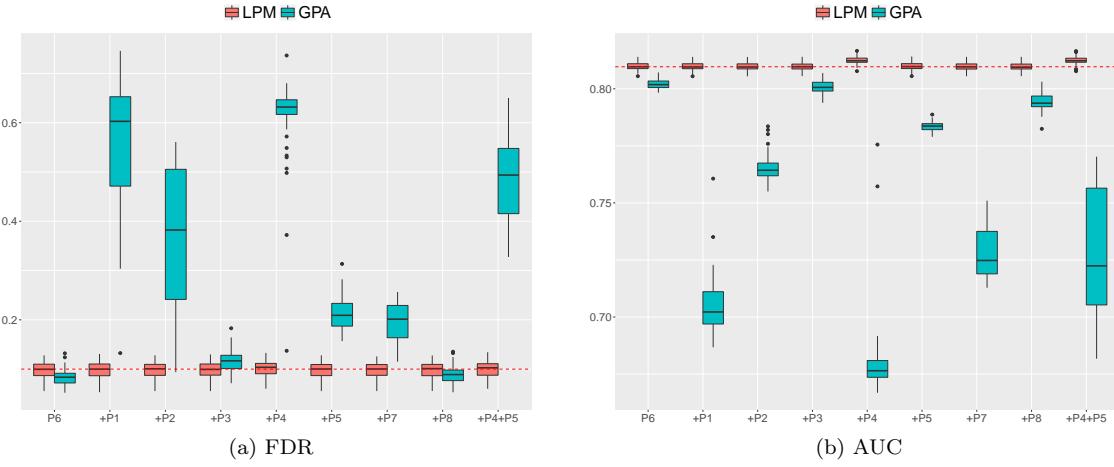


Figure S12: FDR and AUC of LPM and GPA for identification of risk SNPs for P6. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

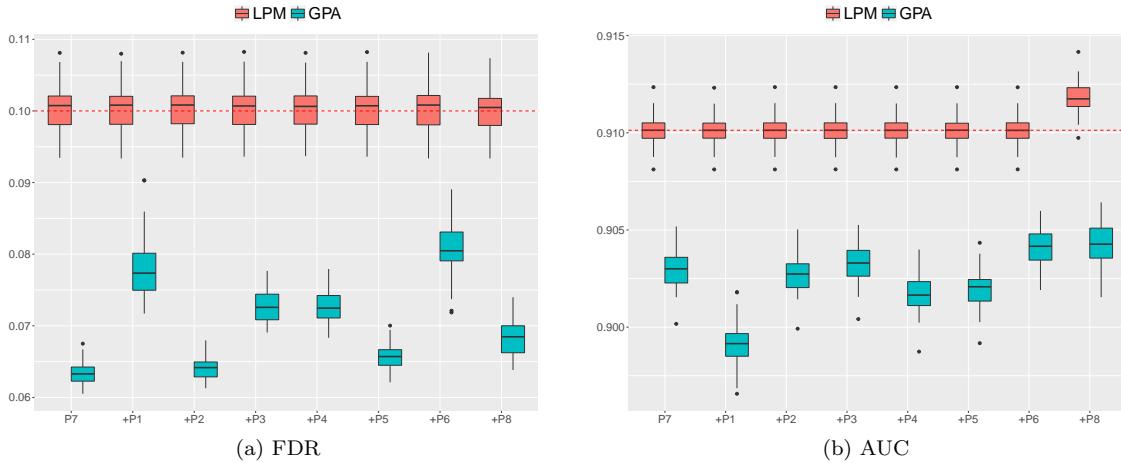


Figure S13: FDR and AUC of LPM and GPA for identification of risk SNPs for P7. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

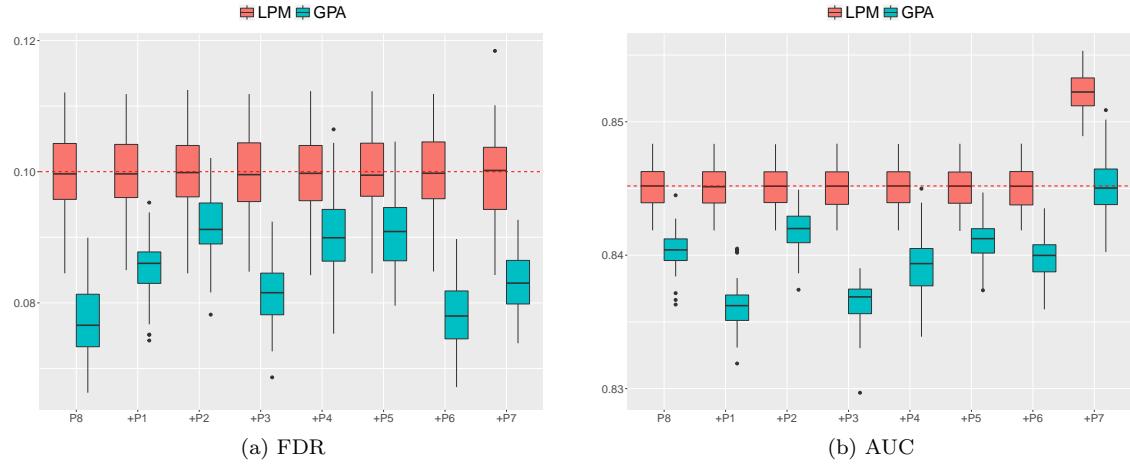


Figure S14: FDR and AUC of LPM and GPA for identification of risk SNPs for P8. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

6.6 The performance for the identification of risk SNPs of one specific trait when correlated trait is integrated

To provide a better illustration for the performance of LPM for the identification of risk SNPs of one specific trait when correlated trait is integrated, we conducted simulations which contain only two traits. In this simulation, we suppose the first trait is the target trait with $\alpha_1 = 0.2$. We varied α_2 at $\{0.2, 0.4, 0.6\}$ and ρ at $\{0, 0.2, 0.4, 0.6\}$ to obtain the FDR and AUC of LPM for the identification of risk SNPs of the target trait. The results are shown in Figure S15.

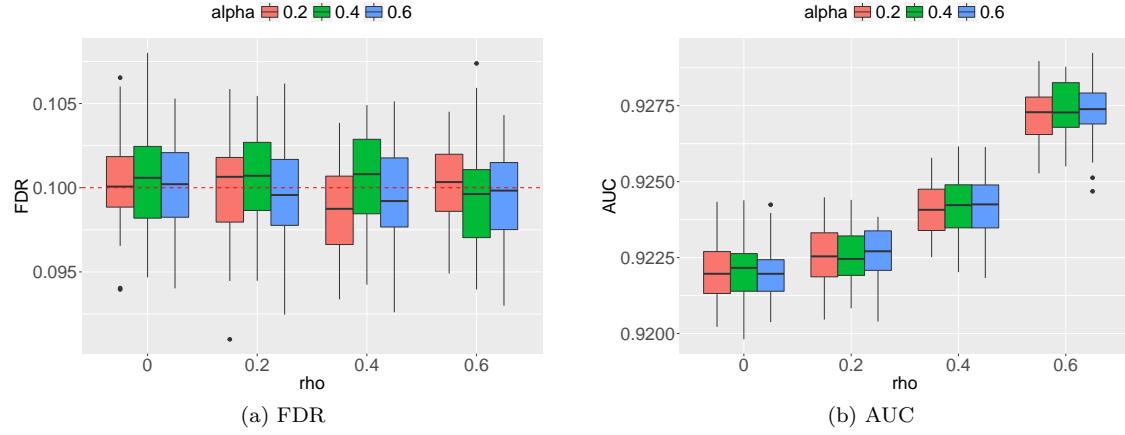


Figure S15: FDR and AUC of LPM for identification of risk SNPs for the target trait when correlated trait is integrated. We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

6.7 Comparison with RiVIERA (Li and Kellis, 2016)

We conducted simulations which contain only two traits. In this simulation, we suppose that the signal strength of the traits were the same, i.e., $\alpha_1 = \alpha_2 = \alpha = 0.2$. We set $\rho = 0.6$. For the analysis of RiVIERA, we set the size of loci to be 1,000 and specify the HMC step size, step number and burn-in proportion to be 1e-3, 100 per MCMC iteration and 20%, respectively. We evaluated the proportion of risk variants which were identified by each method as a function of increasing number of top variants selected. The results are shown in Figure S16. Figure S17 shows the ROC curves of LPM and RiVIERA for the identification of risk SNPs. The receiver operating characteristic curve (AUC) of LPM is larger than RiVIERA indicating that LPM has a better performance.

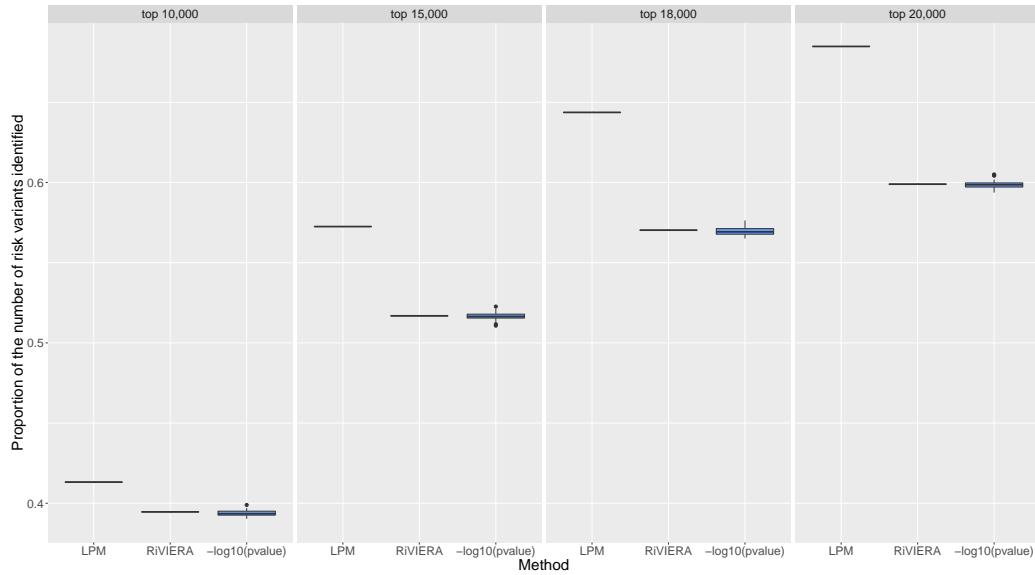


Figure S16: Proportion of the number of risk variants identified by each method with an increasing number of top variants selected. The results are summarized from 50 replications.

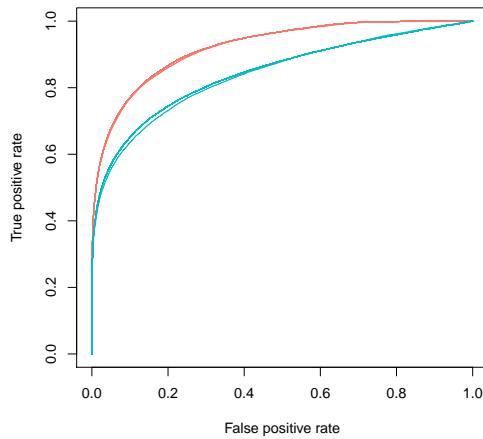


Figure S17: The ROC curve of LPM (red) and RiVIERA (blue) for the identification of risk SNPs. The results are summarized from 50 replications.

6.8 The performance for the identification of risk SNPs of two specific traits

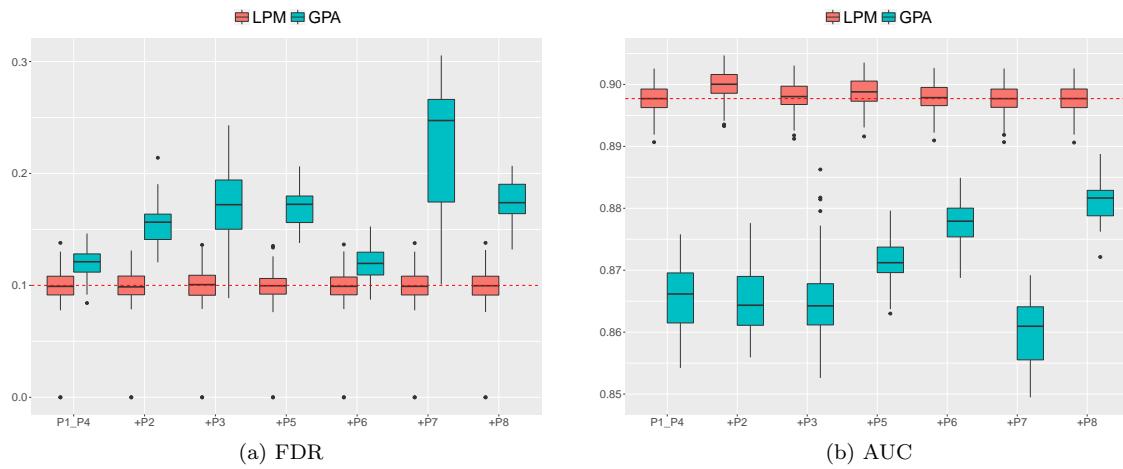


Figure S18: FDR and AUC of LPM and GPA for identification of risk SNPs for both P1 and P4. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in joint analysis of P1 and P4 using LPM as a reference line. The results are summarized from 50 replications.

6.9 The performance for the identification of risk SNPs of three specific traits

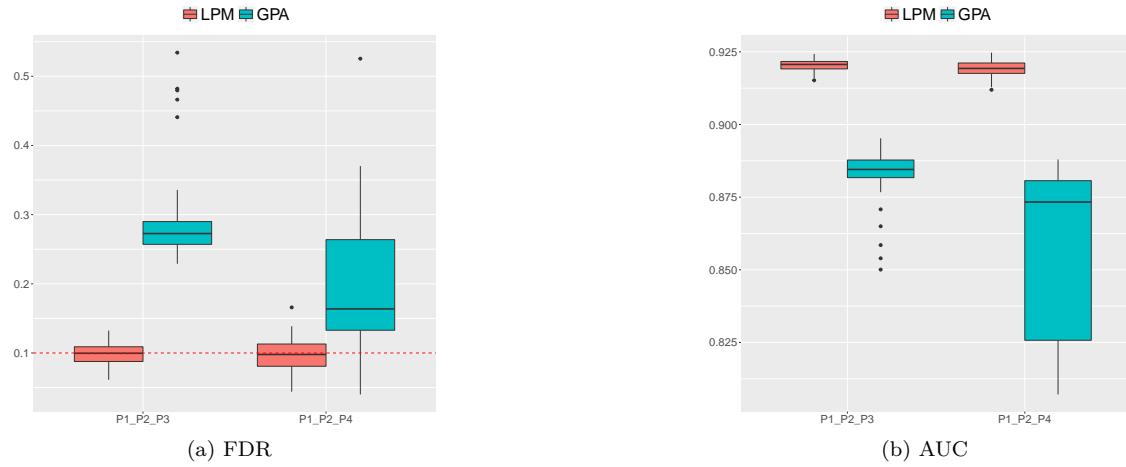


Figure S19: FDR and AUC of LPM and GPA for identification of risk SNPs for three specific traits. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

6.10 The performance for the identification of risk SNPs of one specific trait when there is no functional annotation

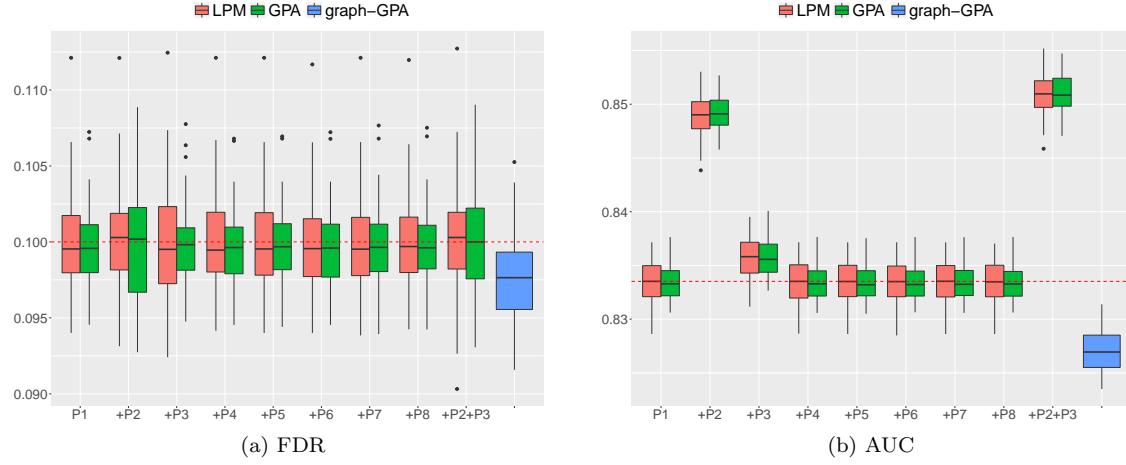


Figure S20: FDR and AUC of LPM, GPA and graph-GPA for identification of risk SNPs for P1. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

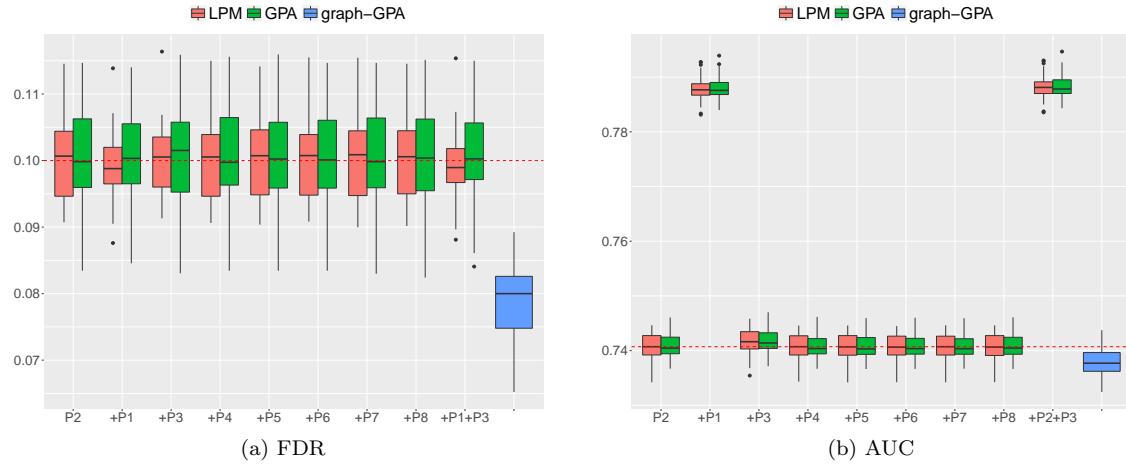


Figure S21: FDR and AUC of LPM, GPA and graph-GPA for identification of risk SNPs for P2. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

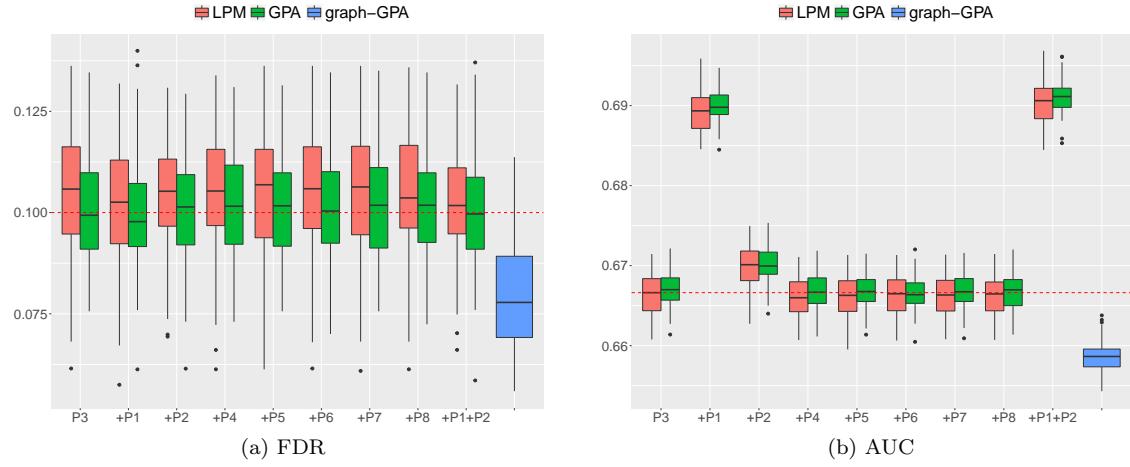


Figure S22: FDR and AUC of LPM, GPA and graph-GPA for identification of risk SNPs for P3. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

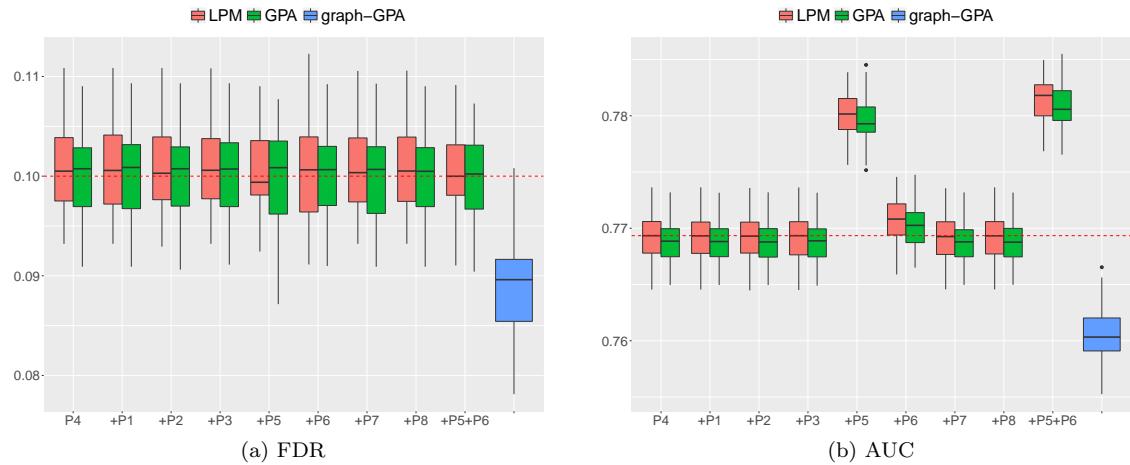


Figure S23: FDR and AUC of LPM, GPA and graph-GPA for identification of risk SNPs for P4. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

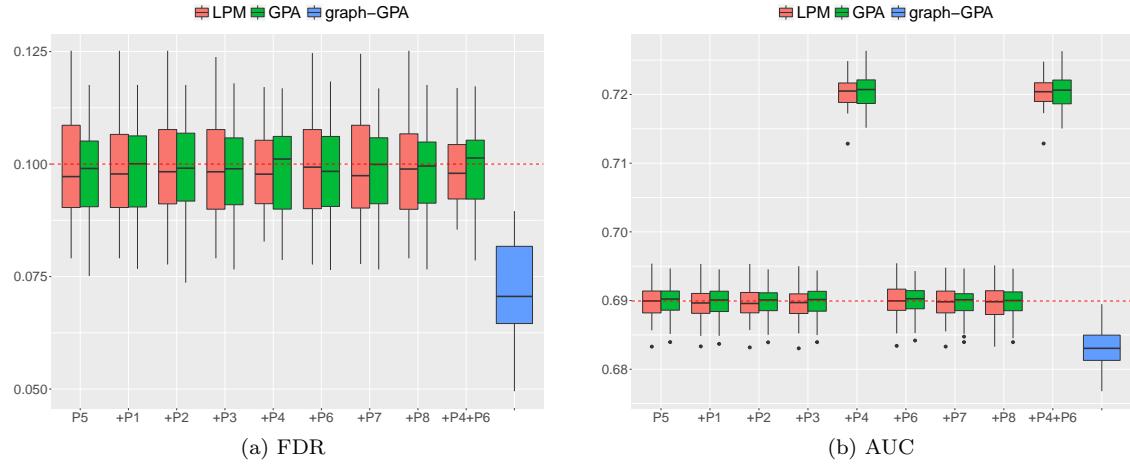


Figure S24: FDR and AUC of LPM, GPA and graph-GPA for identification of risk SNPs for P5. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

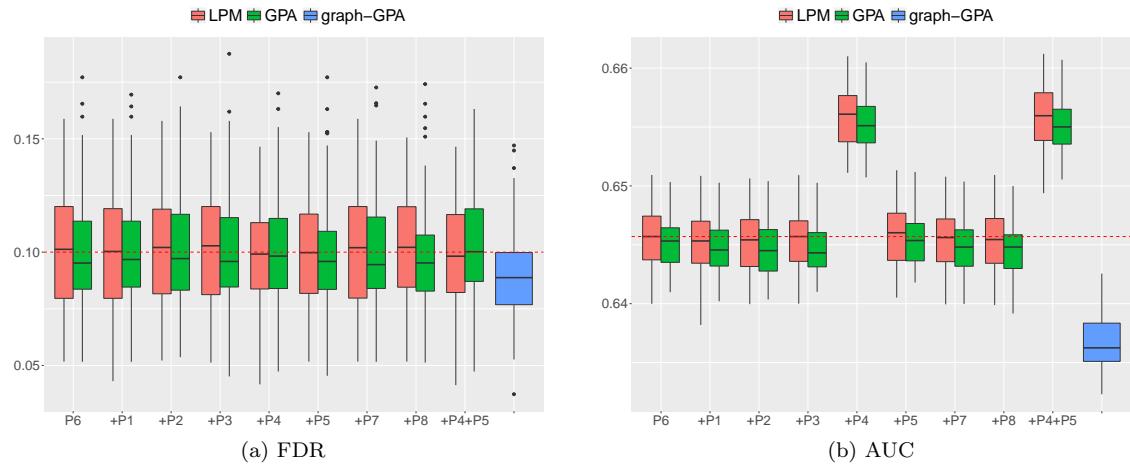


Figure S25: FDR and AUC of LPM, GPA and graph-GPA for identification of risk SNPs for P6. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

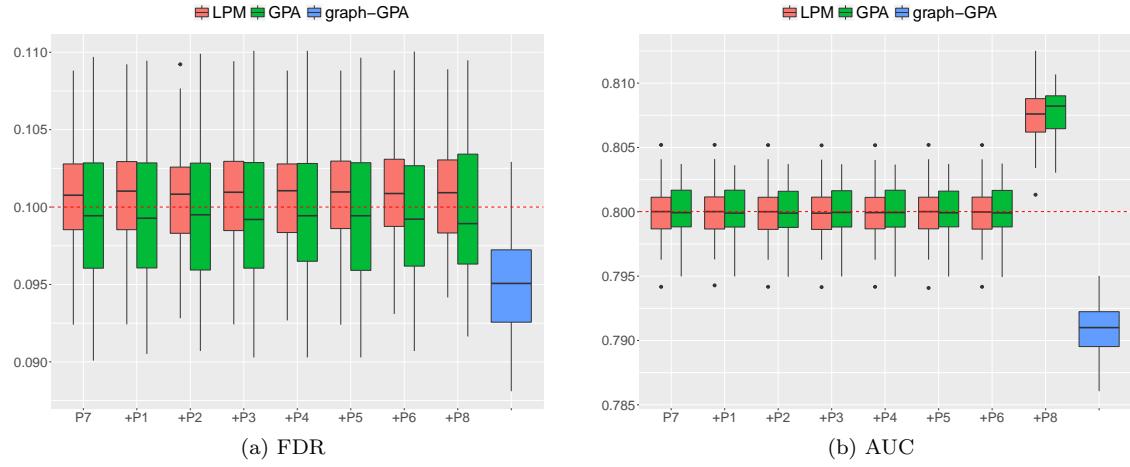


Figure S26: FDR and AUC of LPM, GPA and graph-GPA for identification of risk SNPs for P7. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

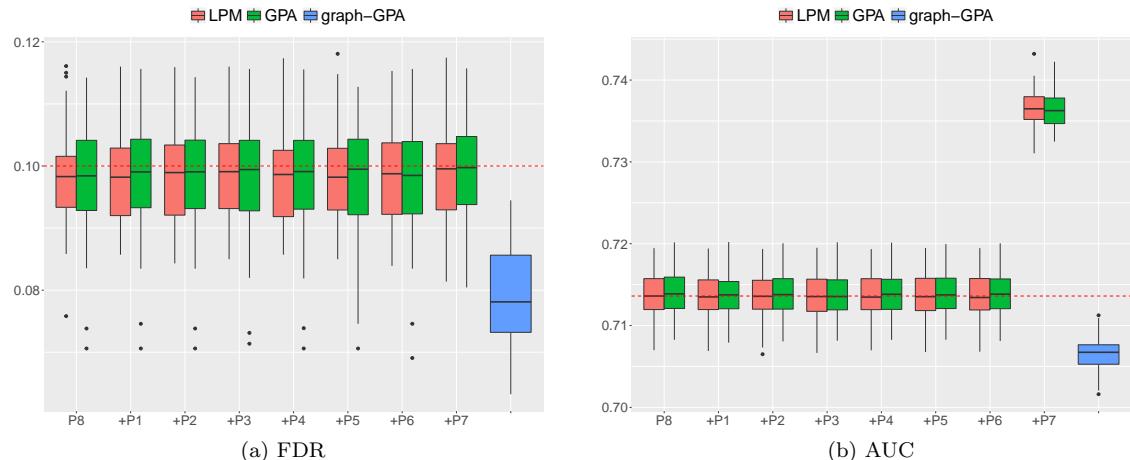


Figure S27: FDR and AUC of LPM, GPA and graph-GPA for identification of risk SNPs for P8. The text in the x-axis indicated the GWAS data sets used in analysis. We controlled global FDR at 0.1 to evaluate empirical FDR. The red horizontal line in (b) was set at the median AUC in separate analysis using LPM as a reference line. The results are summarized from 50 replications.

6.11 Performance of LPM for the hypothesis testing of annotation enrichment

To provide a better illustration for the performance of LPM for the hypothesis testing of annotation enrichment, we conducted simulations which contain only two traits. In this simulation, we suppose the signal strength of the traits are the same, i.e., $\alpha_1 = \alpha_2 = \alpha$, the number of annotations to be $D = 1$ and no correlation $\rho = 0$. We varied α at $\{0.2, 0.4, 0.6\}$ to obtain the type I error rate and varied and the coefficient of the annotation β at $\{-0.4, -0.3, -0.2, -0.1, 0.1, 0.2, 0.3, 0.4\}$ to obtain the power of LPM for the enrichment test of the annotation. The results are shown in Figure S28.

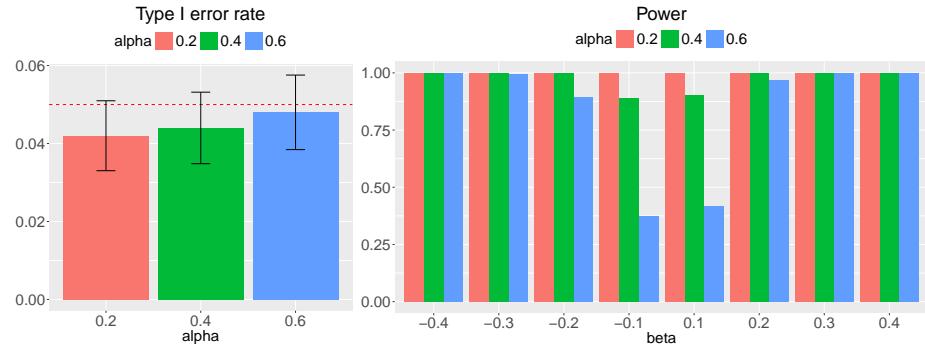


Figure S28: Type I error rate and power of LPM for the annotation enrichment test. The bars represent one standard error. We evaluate type I error rate and power at 0.05 significance level. The results are summarized from 500 replications.

When functional annotations only correspond to a small proportion of genome, we considered the two traits case with same signal strength ($\alpha_1 = \alpha_2 = 0.2$). We assume the number of annotations to be $D = 1$ and no trait-correlation $\rho = 0$. We set the coefficient of the annotation β be zero to obtain the type I error rate and varied β at $\{-0.3, -0.2, -0.1, 0.1, 0.2, 0.3\}$ to evaluate the power of LPM for annotation enrichment test with the functional proportion (the proportion SNPs are annotated) varied at $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.2\}$. The results are shown in Figure S29 and Figure S30.

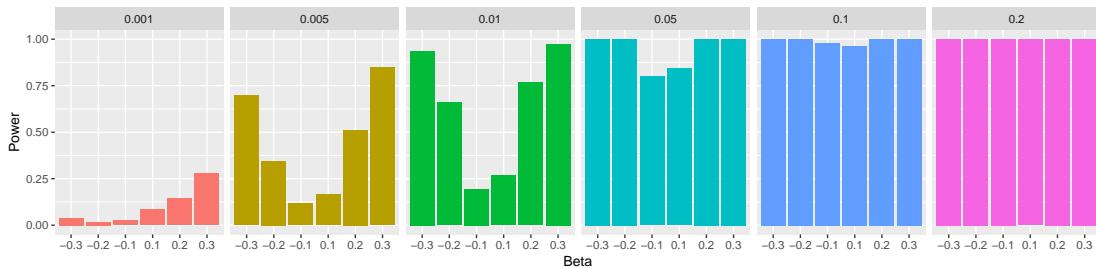


Figure S29: Power of enrichment test for annotations with different sparsity. We evaluate power at 0.05 significance level. The results are summarized from 500 replications.

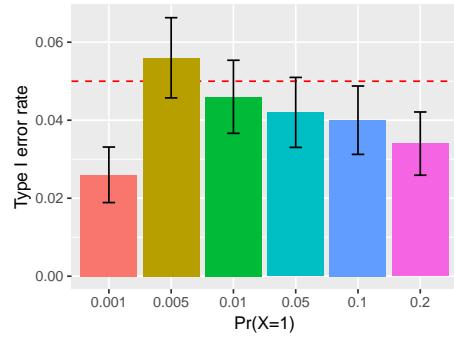


Figure S30: Type I error rate of enrichment test for annotations with different sparsity. The bars represent one standard error. We evaluate type I error rate at 0.05 significance level. The results are summarized from 500 replications.

6.12 Parameter estimation using LPM

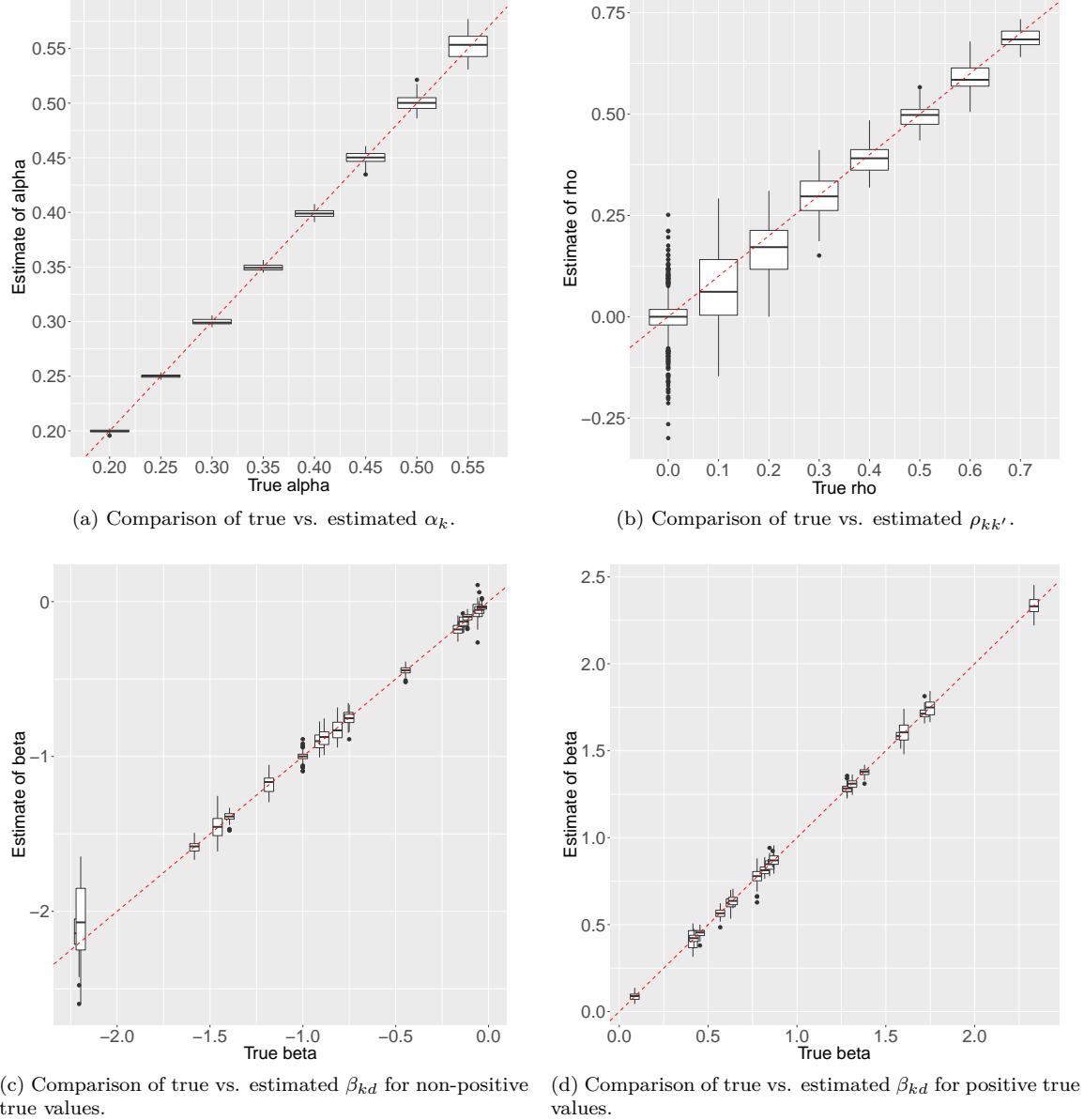


Figure S31: Parameter estimation performance of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and \mathbf{R} . We separated the comparison graph of β_{kd} for better visualization in (c) and (d). The red dashed lines indicate the diagonal. The results are summarized from 50 replications.

6.13 Parameter estimation using bLPM

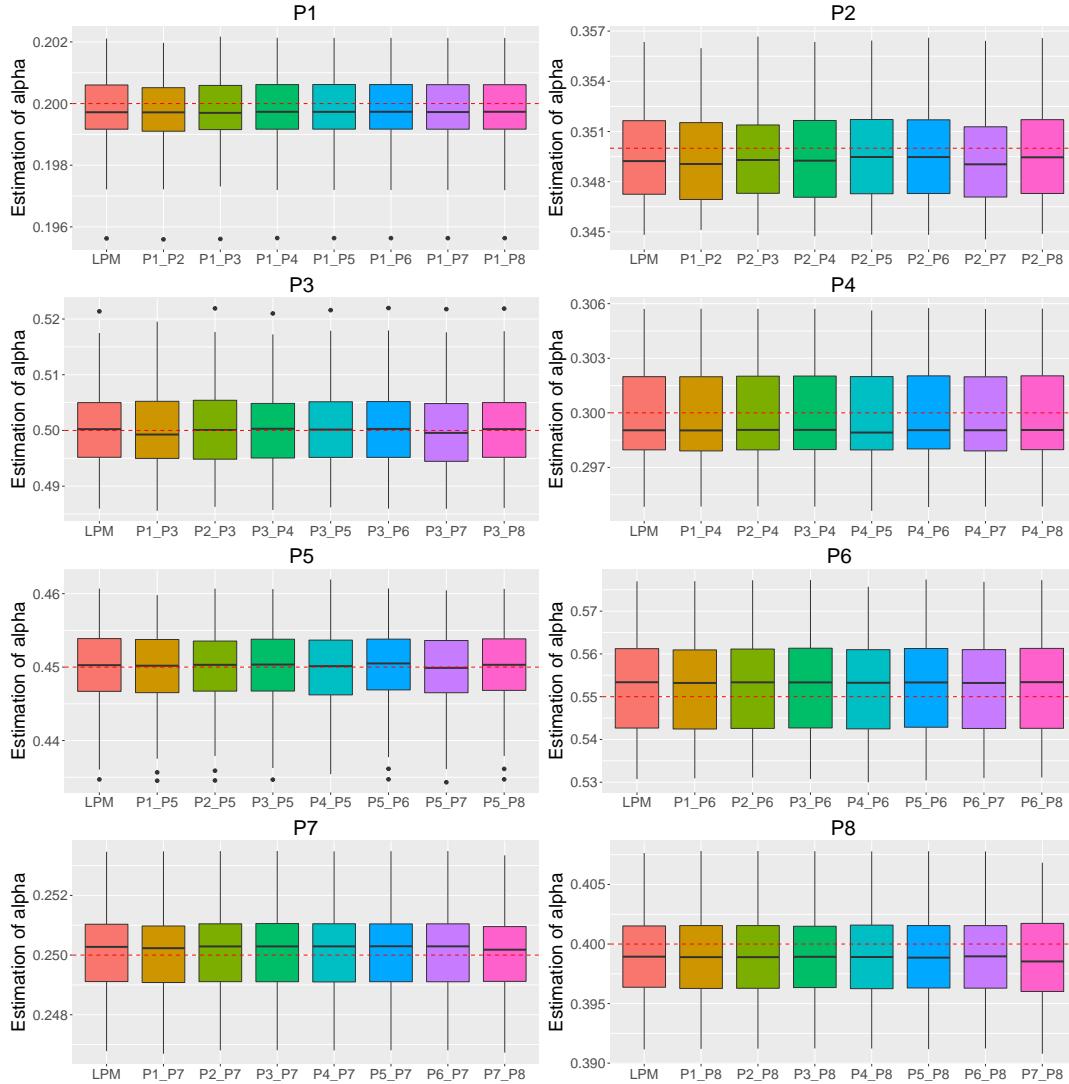


Figure S32: Comparison of the parameter estimation of α using LPM and $\tilde{\alpha}$ using bLPM. The red dashed lines indicate the true values. The results are summarized from 50 replications.

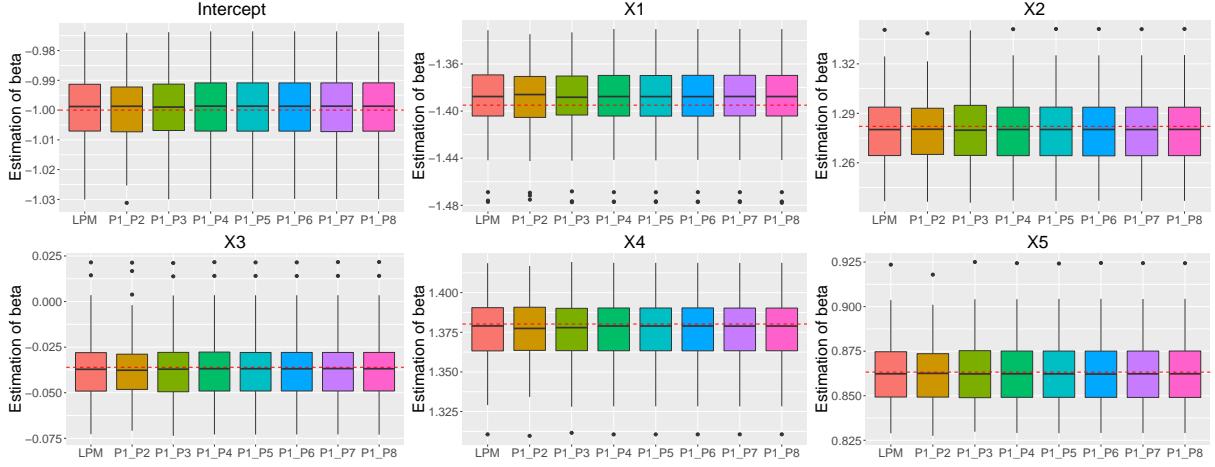


Figure S33: Comparison of the parameter estimation of β using LPM and $\tilde{\beta}$ using bLPM for P1. The red dashed lines indicate the true values. The results are summarized from 50 replications.

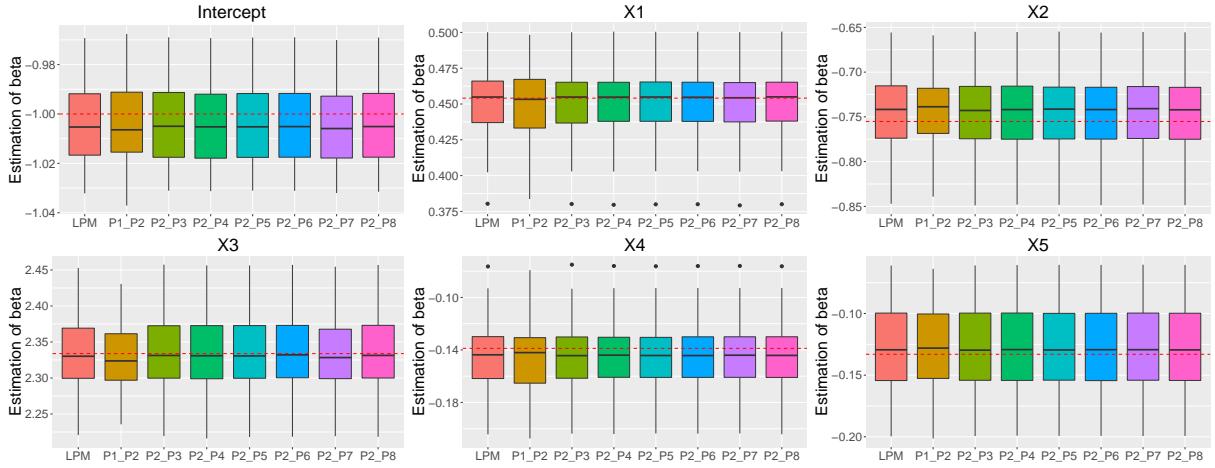


Figure S34: Comparison of the parameter estimation of β using LPM and $\tilde{\beta}$ using bLPM for P2. The red dashed lines indicate the true values. The results are summarized from 50 replications.

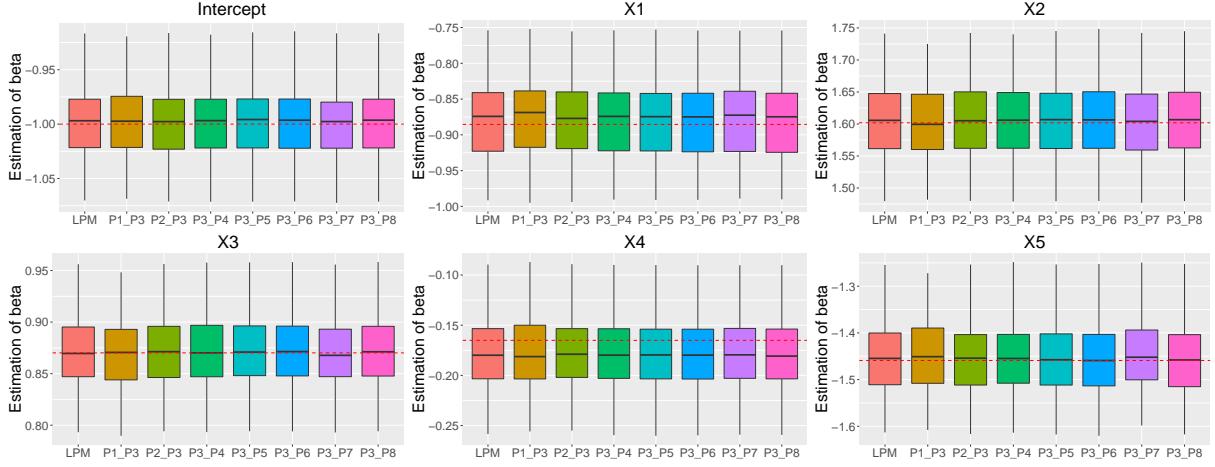


Figure S35: Comparison of the parameter estimation of β using LPM and $\tilde{\beta}$ using bLPM for P3. The red dashed lines indicate the true values. The results are summarized from 50 replications.

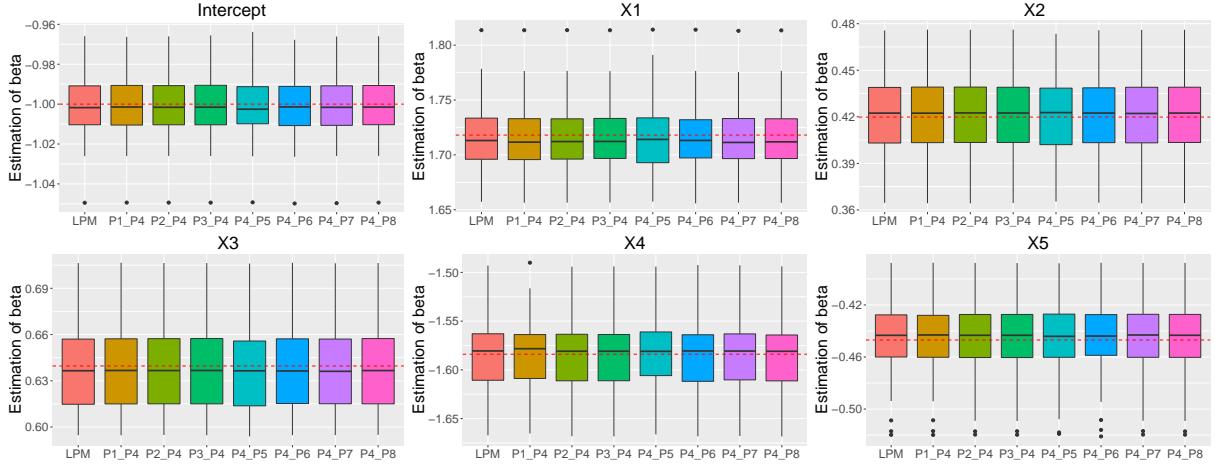


Figure S36: Comparison of the parameter estimation of β using LPM and $\tilde{\beta}$ using bLPM for P4. The red dashed lines indicate the true values. The results are summarized from 50 replications.

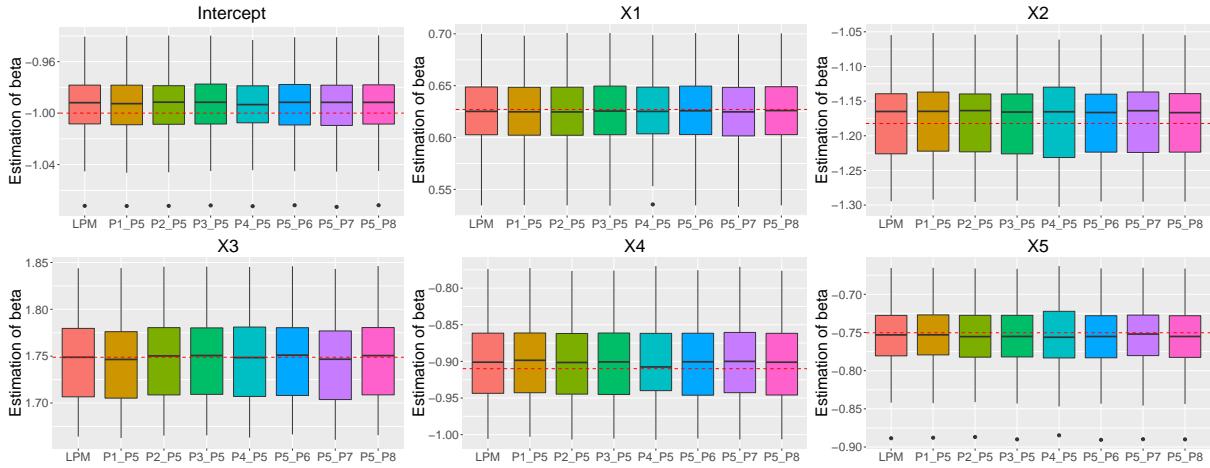


Figure S37: Comparison of the parameter estimation of β using LPM and $\tilde{\beta}$ using bLPM for P5. The red dashed lines indicate the true values. The results are summarized from 50 replications.

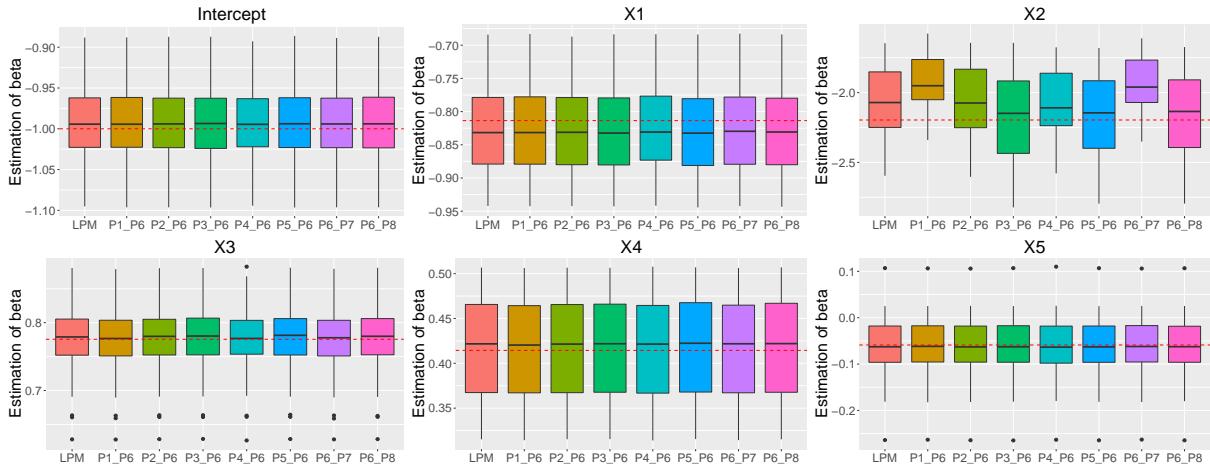


Figure S38: Comparison of the parameter estimation of β using LPM and $\tilde{\beta}$ using bLPM for P6. The red dashed lines indicate the true values. The results are summarized from 50 replications.

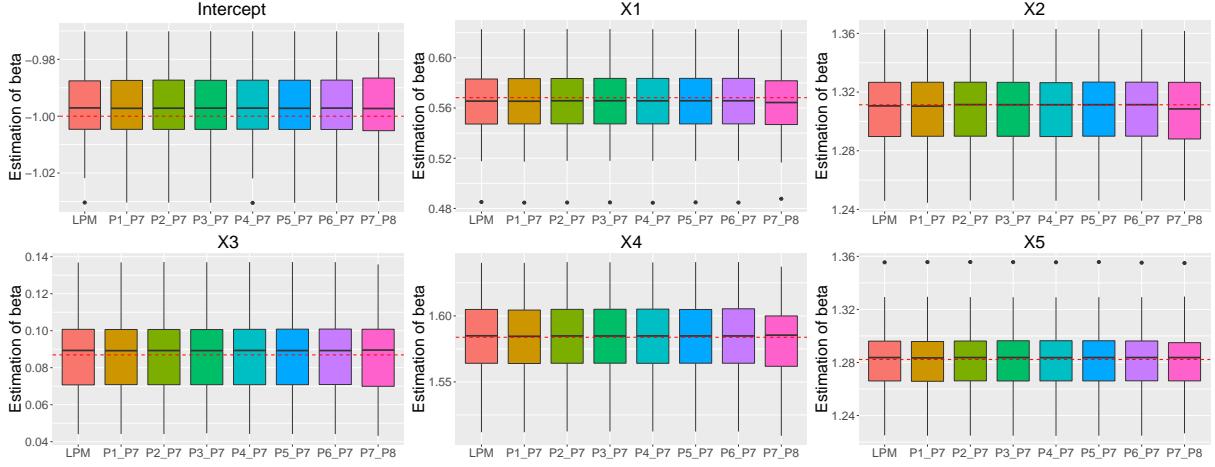


Figure S39: Comparison of the parameter estimation of β using LPM and $\tilde{\beta}$ using bLPM for P7. The red dashed lines indicate the true values. The results are summarized from 50 replications.

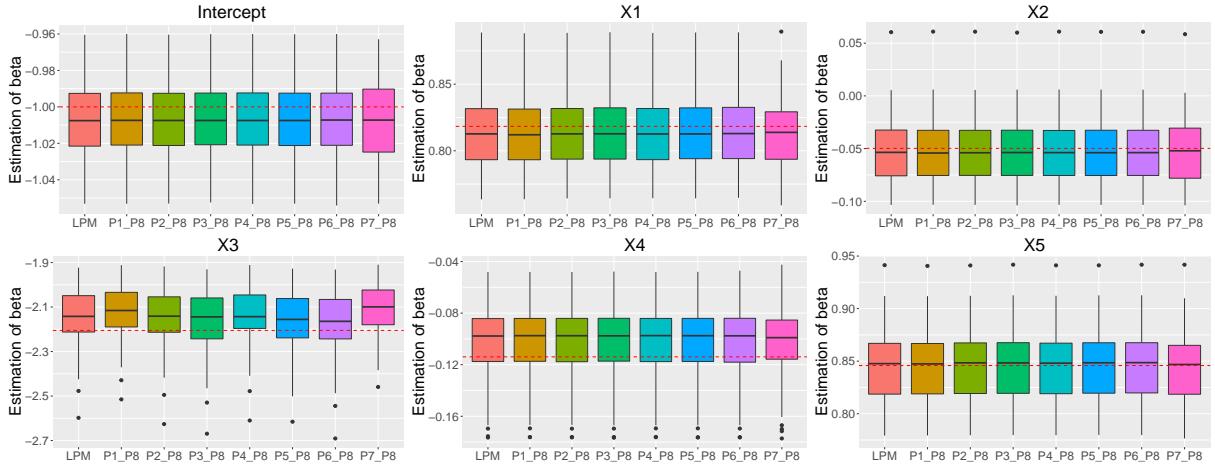


Figure S40: Comparison of the parameter estimation of β using LPM and $\tilde{\beta}$ using bLPM for P8. The red dashed lines indicate the true values. The results are summarized from 50 replications.

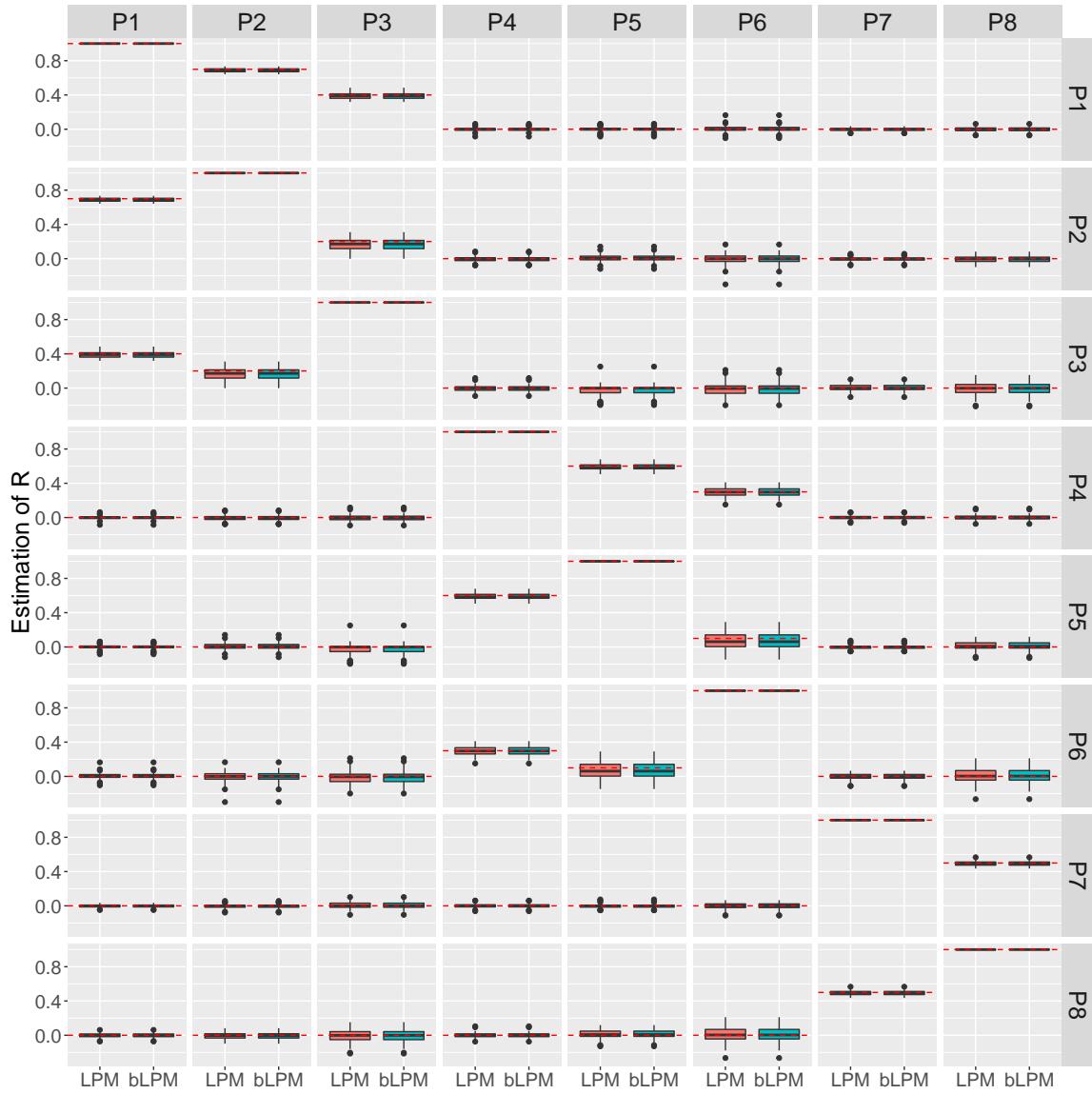


Figure S41: Comparison of the parameter estimation of \mathbf{R} using LPM and $\tilde{\mathbf{R}}$ using bLPM. The red dashed lines indicate the true values. The results are summarized from 50 replications.

6.14 Simulations when P -values are obtained from individual-level data

We generated the P -values for two GWASs from individual-level data instead of directly simulating from the generative model. We assumed that one GWAS was based on $N_1 = 10,000$ individuals and the other was based on $N_2 \in \{3000, 5000, 10,000\}$ individuals. We set the number of SNPs and functional annotations to be $M = 20,000$ and $D = 5$ respectively. The procedure to generate \mathbf{X} , $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ was the same with that in the main text with $\rho = 0$. The entries in the first column of $\boldsymbol{\beta}$ (i.e., the intercept terms) were set to be -3 . Following the Hardy-Weinberg principle, the genotype matrix \mathbf{G} is generated with minor allele frequency from $U[0.05, 0.5]$. The effect sizes β_{SNP} were simulated from $N(0, 1)$ if the corresponding entries in $\boldsymbol{\eta}$ were 1 and were set to be 0 otherwise. The noise level σ_e^2 was specified to control heritability $h^2 = \frac{var(\mathbf{G}\boldsymbol{\beta}_{SNP})}{var(\mathbf{G}\boldsymbol{\beta}_{SNP}) + \sigma_e^2}$ at given levels (0.3, 0.5 or 0.8). The phenotype data \mathbf{y} was generated based on $\mathbf{y} = \mathbf{G}\boldsymbol{\beta}_{SNP} + \mathbf{e}$, where $e_i \sim N(0, \sigma_e^2)$ for $i = 1, \dots, N$. Finally, we calculated P -values for each SNP using univariate linear regression.

Figure S42 shows the estimates of α_2 using LPM with different sample sizes and heritability. When the heritability is fixed, larger sample size corresponds to smaller value of α , indicating stronger signal strength of GWAS. As the heritability increases, the corresponding estimates of α_2 also become smaller, given fixed sample size. Hence, we can use α to indicate the signal strength of GWAS in our paper.

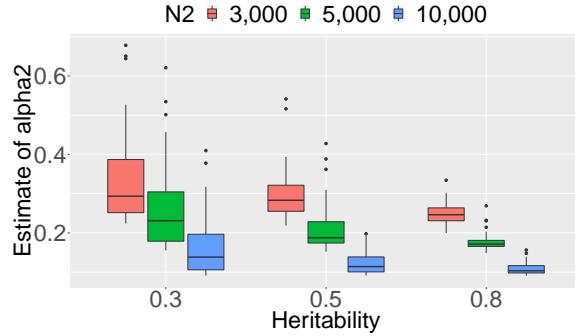


Figure S42: The estimates of parameter α_2 using LPM from individual-level data with different sample sizes and heritability. The results are summarized from 50 replications.

To evaluate the performance of LPM in the trait-relationship test, we calculated the type I error rate of LPM for the trait-relationship test ($H_0 : \rho = 0$). The results are shown in Figure S43. We also varied ρ at $\{0, 0.2, 0.4, 0.6, 0.8\}$ and evaluated the empirical FDR for identification of risk SNPs. The results are shown in Figure S44. The type I error rate of the trait-relationship test and the empirical FDR of risk SNP identification are well controlled at the nominal level.

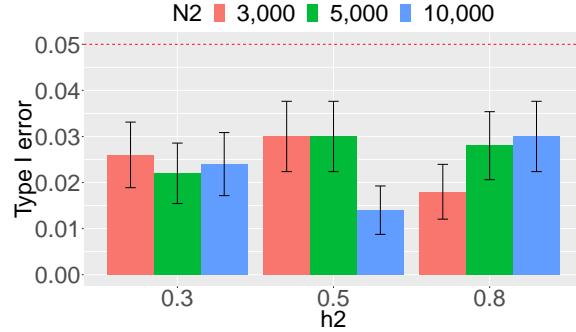


Figure S43: Type I error rate of LPM for the relationship test between the two traits when P -values are simulated from individual-level data with different sample sizes and heritability. The bars represent one standard error. We evaluated type I error rate at 0.05 significance level. The results are summarized from 500 replications.

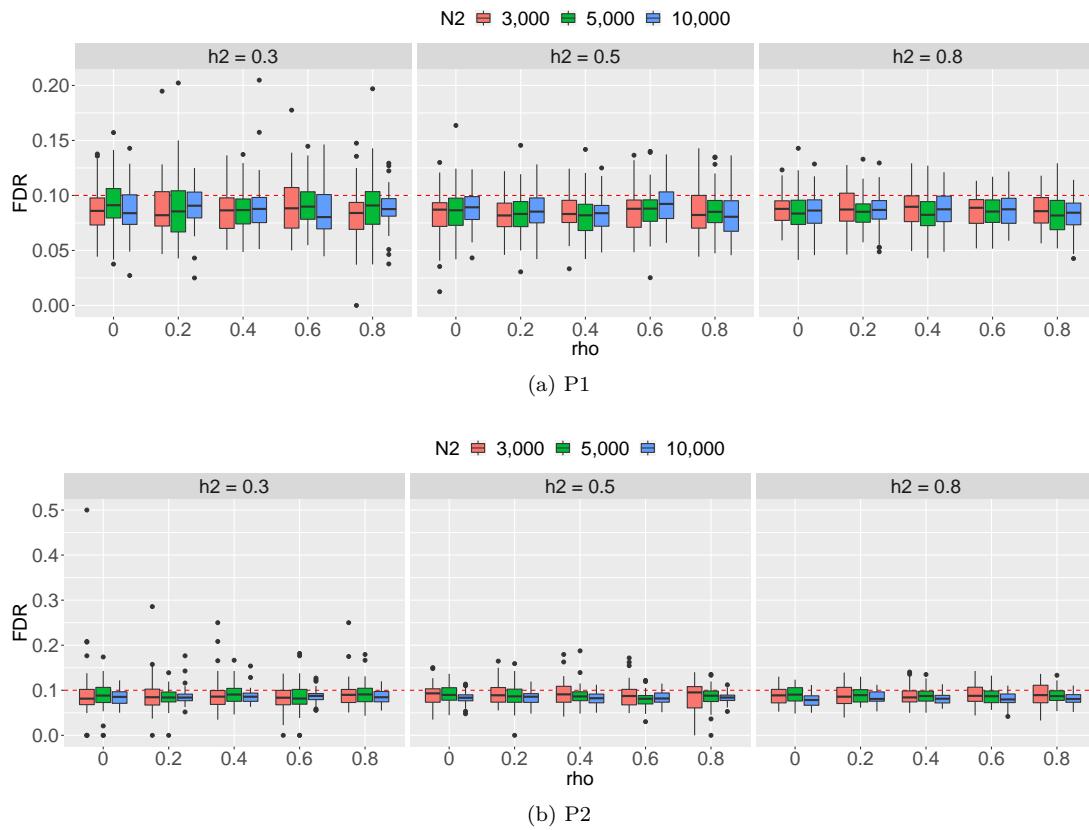


Figure S44: FDR of LPM for identification of risk SNPs for (a) P1 and (b) P2 when P -values are simulated from individual-level data with different sample sizes and heritability. We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

6.15 Simulations if P -values are not from beta distribution

We simulate the two GWASs case when P -values are not from beta distribution. We generated z -values for SNPs in the null group from the standard normal distribution and z -values for SNPs in the non-null group from the distributions in Table S1. Then we converted z -scores to obtain P -values. In these simulations, P -values do not follow the mixture of uniform and beta distribution. We set $M = 100,000$ and $D = 5$. The procedure to generate \mathbf{X} , $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ is the same with that in the main text. As shown in Figure S45-S46, the type I error rate of LPM for the relationship test between the two traits under these distributions and the empirical FDR for identification of risk SNPs are still controlled at the nominal level.

Scenario	Distribution
big-normal	$N(0, 4^2)$
near normal	$\frac{2}{3}N(0, 1^2) + \frac{1}{3}N(0, 2^2)$
skew	$\frac{1}{4}N(-2, 2^2) + \frac{1}{4}N(-1, 1.5^2) + \frac{1}{3}N(0, 1^2) + \frac{1}{6}N(1, 1^2)$
spiky	$0.4N(0, 0.25^2) + 0.2N(0, 0.5^2) + 0.2N(0, 1^2) + 0.2N(0, 2^2)$

Table S1: Alternative distributions for z -scores.

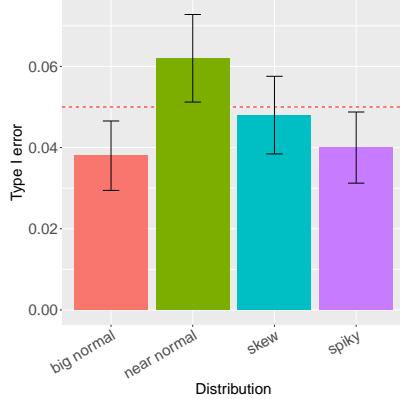
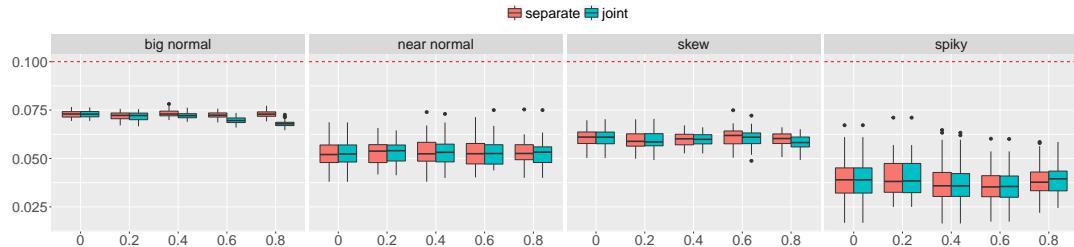
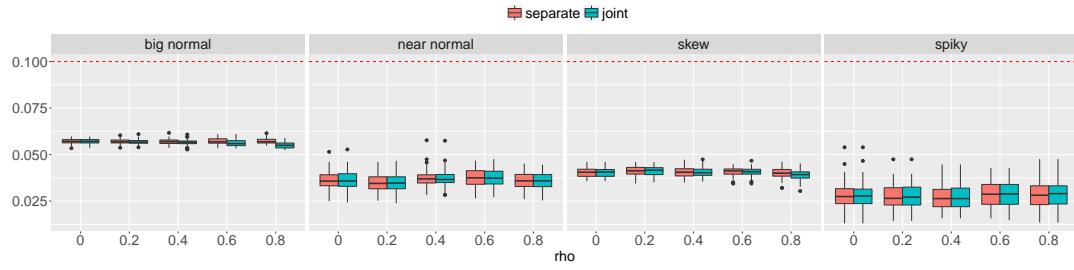


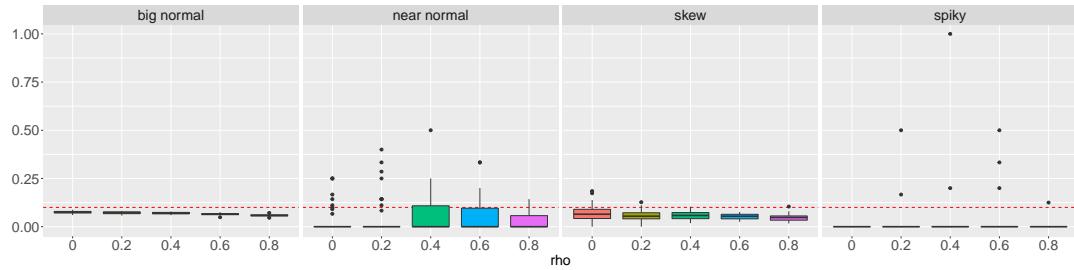
Figure S45: Type I error rate of LPM for the relationship test between the two traits under different distributions. The bars represent one standard error. We evaluate type I error rate at 0.05 significance level. The results are summarized from 500 replications.



(a) P1



(b) P2



(c) P1 and P2

Figure S46: FDR of LPM for identification of risk SNPs for (a) P1, (b) P2, (c) P1 and P2 under different distributions. We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

6.16 Performance of LPM when the proportion of risk SNPs is extremely small

We conducted simulations to investigate the performance of LPM when the proportion of risk SNPs π_1 was extremely small. To better control π_1 , we considered the case of two traits with same signal strength ($\alpha_1 = \alpha_2 = \alpha$) and no annotations. In this simulation, we varied π_1 in $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.15, 0.2\}$ and then transformed π_1 to β_0 which is the intercept in LPM using $\beta_0 = -\Phi^{-1}(1 - \pi_1)$ so that we can use the same procedure to simulate data. We evaluated the accuracy of the estimation of π_1 and ρ using LPM with α varied in $\{0.2, 0.4, 0.6\}$ and ρ varied in $\{0, 0.2, 0.4, 0.6\}$. The results are shown in Figure S47 and Figure S48, respectively.

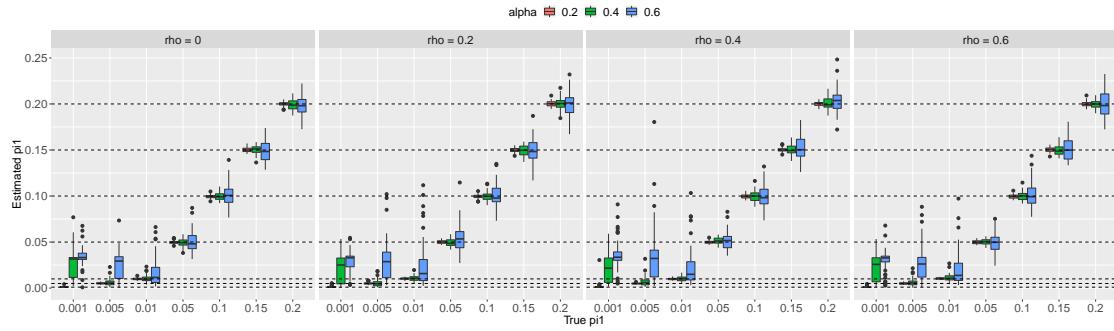


Figure S47: Parameter estimation of π_1 using LPM. The true values are indicated by dotted lines. The results are summarized from 50 replications.

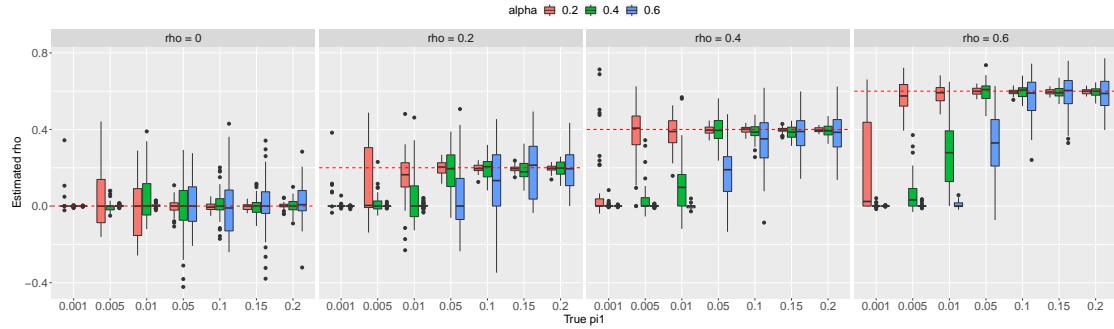


Figure S48: Parameter estimation of ρ using LPM. The true values are indicated by dotted red lines. The results are summarized from 50 replications.

6.17 Simulations to evaluate the LD effects on LPM

we conducted simulations using real genotype and real annotations. Here we used nine genic category annotations (upstream, downstream, exonic, intergenic, intronic, ncRNA exonic, ncRNA intronic, UTR'3 and UTR'5) provided by ANNOVAR (Wang et al., 2010) and the observed genotype data (1,500 individuals from the 1958 British Birth Cohort (58C) and 1,500 individuals from the UK Blood Service (UKBS)) from WTCCC (The Wellcome Trust Case Control Consortium, 2007). For simplicity, we only considered 17,937 SNPs in chromosome 1 which passed quality control and were overlapped with genic category annotations. Then we partitioned the samples into two groups of equal size to mimic the two GWASs case. We considered the case that risk SNPs were nearly uniformly distributed along the chromosome and there existed one risk SNP every 1000 SNPs. So we simulated 17 risk SNPs for each GWAS and these SNPs were assumed to explain 10% phenotypic variance. For the first GWAS, we generated the entries in β from $N(0, 1)$ and then transformed them to control the relative signal strength between annotated part and un-annotated part $r = 0.25$. The risk SNPs for the second GWAS were generated in three cases: (i) no pleiotropy, (ii) 50% shared risk SNPs and (iii) 100% shared risk SNPs with the first GWAS. We used GCTA (Yang et al., 2011) to simulate phenotypes and used PLINK (Purcell et al., 2007) to get P -values of SNPs for these two GWASs. We applied LPM and detected risk SNPs for the first GWAS. The results of empirical FDR for the first trait based on different region sizes are shown in Figure S49. The distance of 100Kb means the region within 100Kb to the true causal SNPs. The results show that for all the cases with distance threshold larger than 300Kb, the empirical FDRs were well controlled indicating that LPM can provide a satisfactory FDR control in identification of a local genomic region of risk SNPs.

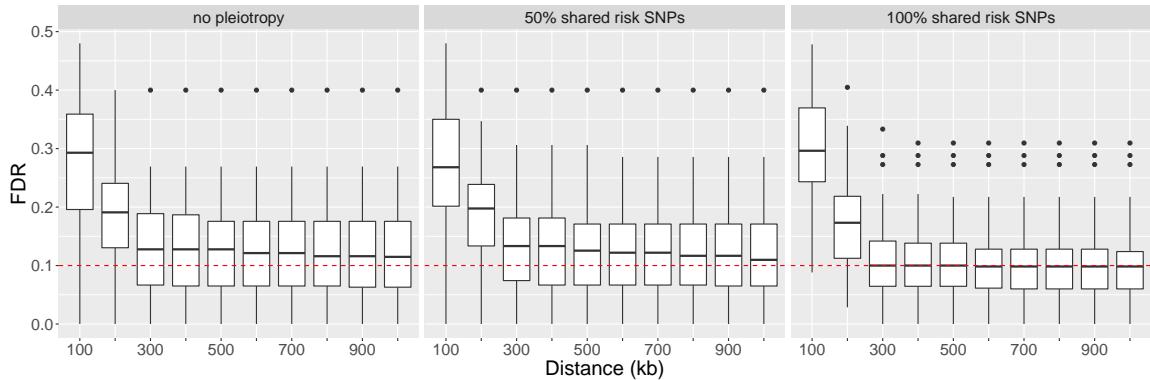


Figure S49: FDR of LPM for identification of risk SNPs with different distance thresholds. We controlled global FDR at 0.1 to evaluate empirical FDR. The results are summarized from 50 replications.

7 Information of 44 GWASs

Age at Menopause	Day et al., 2015, Nature Genetics. download
Alzheimer	Lambert et al., 2013, Nature Genetics. download
Anorexia Nervosa	Duncan et al., 2017, American Journal of Psychiatry. download (ED)
Atopic Dermatitis	Paternooster et al., 2015, Nature Genetics. download
Attention Deficit Hyperactivity Disorder	Demontis et al., 2019, Nature Genetics. download (ADHD-Full GWAS)
Autism Spectrum Disorder	Grove et al., 2019, Nature Genetics. download (iPSYCH-PGC ASD)
Bipolar Disorder	Sklar et al., 2011, Nature Genetics. download (BIP)
BMI	Speliotes et al., 2010, Nature Genetics. download
Celiac Disease	Dubois et al., 2010, Nature Genetics. download
Coronary Artery Disease	Schunkert et al., 2011, Nature Genetics. download
Crohn's Disease	Jostins et al., 2012, Nature. download
Depressive Symptoms	Okbay et al., 2016, Nature Genetics. download
Ever Smoked	Furberg et al., 2010, Nature Genetics. download
Fasting Glucose	Manning et al., 2012, Nature Genetics. download
Fasting Insulin	Manning et al., 2012, Nature Genetics. download
Haemoglobin	Pickrell, 2014, The American Journal of Human Genetics. download
Height	Wood et al., 2014, Nature Genetics. download
High-density Lipoprotein	Willer et al., 2013, Nature Genetics. download
HIV	McLaren et al., 2013, PLoS Pathogens. download
Inflammatory Bowel Disease	Jostins et al., 2012, Nature. download
Low-density Lipoprotein	Willer et al., 2013, Nature Genetics. download
Major Depressive Disorder	Wray et al., 2018, Nature Genetics. download (MDD2 2018)
Mean Cell Haemoglobin	Pickrell, 2014, The American Journal of Human Genetics. download
Mean Cell Haemoglobin Concentration	Pickrell, 2014, The American Journal of Human Genetics. download
Mean Cell Volume	Pickrell, 2014, The American Journal of Human Genetics. download
Multiple Sclerosis	Sawcer et al., 2011, Nature. download
Neuroticism	Okbay et al., 2016, Nature Genetics. download
Packed Cell Volume	Pickrell, 2014, The American Journal of Human Genetics. download
Primary Biliary Cirrhosis	Cordell et al., 2015, Nature Communications. download
Pubertal Growth	Cousminer et al., 2013, Human Molecular Genetics. download
Red Blood Cell Count	Pickrell, 2014, The American Journal of Human Genetics. download
Rheumatoid Arthritis	Okada et al., 2014, Nature. download
Schizophrenia1	Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013, The Lancet. download (SCZ subset)
Schizophrenia2	Ripke and oherd, 2011, Nature Genetics. download(SCZ1)
Schizophrenia3	Ripke et al., 2013, Nature Genetics. download (Sweden+SCZ1)
Schizophrenia4	Ripke et al., 2014, Nature. download (SCZ2)
Systemic Lupus Erythematosus	Bentham et al., 2015, Nature Genetics. download
Total Cholesterol	Willer et al., 2013, Nature Genetics. download
Triglycerides	Willer et al., 2013, Nature Genetics. download
Type 1 Diabetes	Bradfield et al., 2011, PLoS Genetics. download
Type 2 Diabetes	Morris et al., 2012, Nature Genetics. download
Ulcerative Colitis	Jostins et al., 2012, Nature. download
Years of Education1	Rietveld et al., 2013, Science. download
Years of Education2	Okbay et al., 2016, Nature. download

Table S2: The source of the 44 GWASs.

Phenotypes	Sample Size	$\hat{\alpha}$
Age at Menopause	69,360	0.24
Alzheimer	Case (17,008), Control (37,154)	0.46
Anorexia Nervosa	Case (3,495), Control (10,982)	0.76
Atopic Dermatitis	38,176	0.42
Attention Deficit Hyperactivity Disorder	Case (20,183), Control (35,191)	0.58
Autism Spectrum Disorder	Case (18,381), Control (27,969)	0.64
Bipolar Disorder	16,731	0.65
BMI	122,033	0.30
Celiac Disease	Case (4,533), Control (10,750)	0.53
Coronary Artery Disease	76,016	0.50
Crohn's Disease	Case (5,956), Control (14,927)	0.26
Depressive Symptoms	161,460	0.67
Ever Smoked	74,035	0.74
Fasting Glucose	58,074	0.22
Fasting Insulin	51,750	0.37
Haemoglobin	50,709	0.49
Height	242,435	0.21
High-density Lipoprotein	96,902	0.09
HIV	Case (6,334), Control (7,247)	0.75
Inflammatory Bowel Disease	Case (12,882), Control (21,770)	0.34
Low-density Lipoprotein	92,046	0.09
Major Depressive Disorder	52,438	0.62
Mean Cell Haemoglobin	43,223	0.29
Mean Cell Haemoglobin Concentration	46,713	0.56
Mean Cell Volume	48,225	0.32
Multiple Sclerosis	Case (9,772), Control (17,376)	0.33
Neuroticism	170,911	0.49
Packed Cell Volume	44,728	0.53
Primary Biliary Cirrhosis	Case (2,764), Control (10,475)	0.31
Pubertal Growth	10,791	0.67
Red Blood Cell Count	44,966	0.43
Rheumatoid Arthritis	37,681	0.38
Schizophrenia1	Case (9,379), Control (7,736)	0.66
Schizophrenia2	Case (9,394), Control (12,462)	0.61
Schizophrenia3	Case (13,833), Control (18,310)	0.55
Schizophrenia4	Case (36,989), Control (113,075)	0.41
Systemic Lupus Erythematosus	14,267	0.46
Total Cholesterol	97,142	0.10
Triglycerides	93,282	0.08
Type 1 Diabetes	Case (9,934), Control (16,956)	0.46
Type 2 Diabetes	63,038	0.50
Ulcerative Colitis	Case (6,968), Control (20,464)	0.41
Years of Education1	126,559	0.60
Years of Education2	328,917	0.42

Table S3: Sample size and estimates of α for the 44 GWASs.

8 More real data results

8.1 Number of significant correlations among 44 traits

Significance level for Bonferroni correction	0.001	0.01	0.05
Number of significant correlations	763	785	797

Table S4: Number of significant correlations among 44 traits at different significance level.

8.2 Genetic correlation among a subset of GWASs estimated by cross-trait LD Score regression

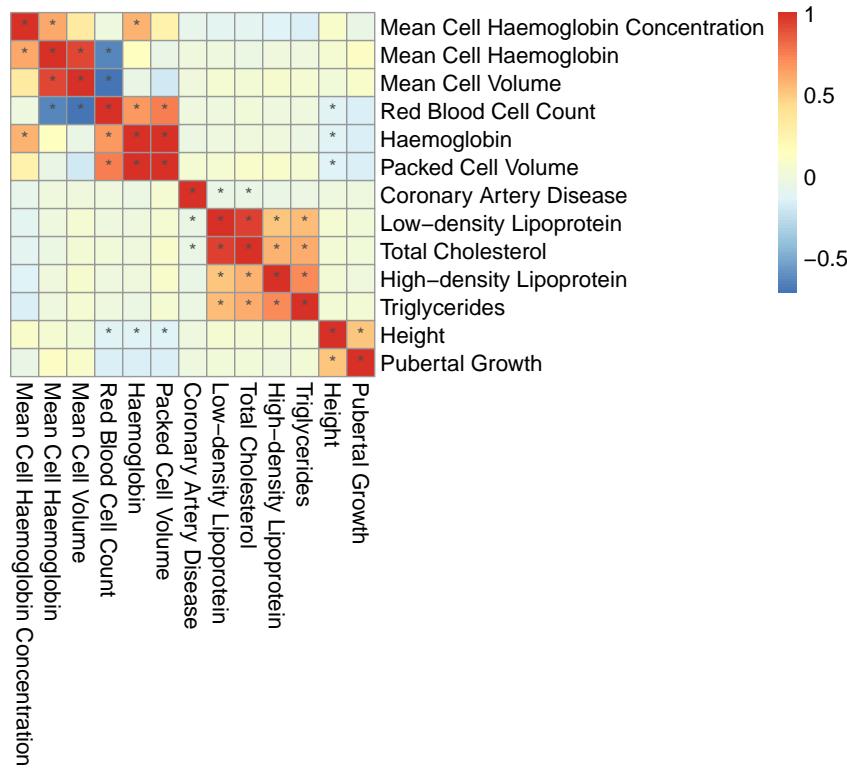


Figure S50: Genetic correlations estimated by cross-trait LD Score regression. The symbol "*" means the P -value is significant after Bonferroni correction at level 0.05.

8.3 Relationship among a subset of GWASs inferred by graph-GPA

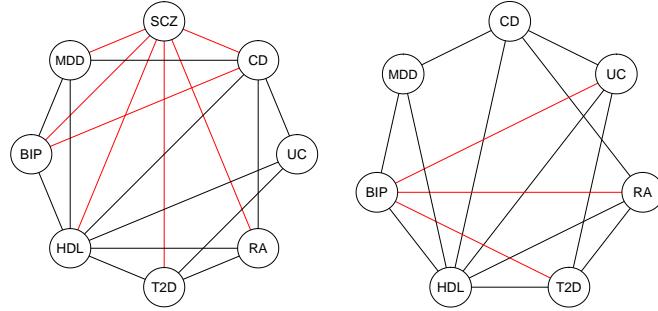


Figure S51: Relationship inferred by graph-GPA when different numbers of GWASs are considered. The red edges denote the changes in the relationship among traits. SCZ denotes Schizophrenia4.

8.4 Estimation of pairwise correlations for 44 GWASs with or without functional annotations using LPM

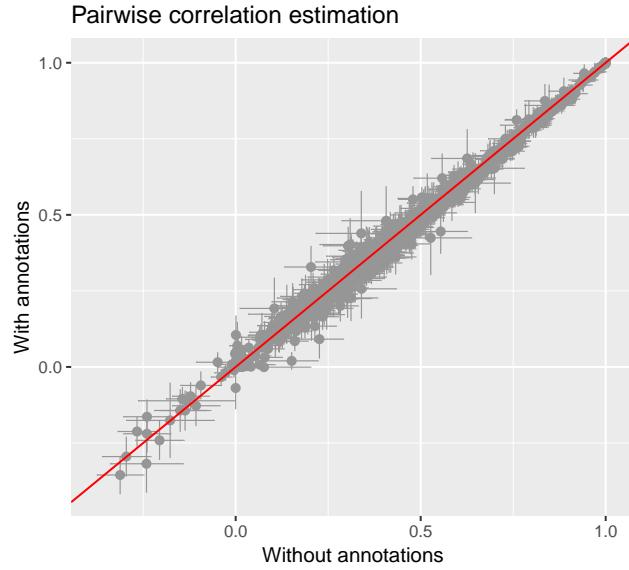


Figure S52: The estimates of pairwise correlations for 44 GWASs with or without functional annotations using LPM. The bars represent the standard errors. The red dashed lines indicate the diagonal.

8.5 Top correlated traits for each of the 44 GWASs

	top1	top 2
Age at Menopause	Fasting Insulin	BMI
Alzheimer	Total Cholesterol	BMI
Anorexia Nervosa	Major Depressive Disorder	Years of Education2
Atopic Dermatitis	Ulcerative Colitis	Multiple Sclerosis
Attention Deficit Hyperactivity Disorder	Major Depressive Disorder	Autism Spectrum Disorder
Autism Spectrum Disorder	Attention Deficit Hyperactivity Disorder	Major Depressive Disorder
Bipolar Disorder	Schizophrenia2	Schizophrenia1
BMI	Years of Education2	Major Depressive Disorder
Celiac Disease	Multiple Sclerosis	Primary Biliary Cirrhosis
Coronary Artery Disease	Triglycerides	Low-density Lipoprotein
Crohn's Disease	Inflammatory Bowel Disease	Ulcerative Colitis
Depressive Symptoms	Neuroticism	Major Depressive Disorder
Ever Smoked	Years of Education2	Major Depressive Disorder
Fasting Glucose	Type 2 Diabetes	Fasting Insulin
Fasting Insulin	Type 2 Diabetes	Triglycerides
Haemoglobin	Packed Cell Volume	Red Blood Cell Count
Height	Pubertal Growth	High-density Lipoprotein
High-density Lipoprotein	Triglycerides	Total Cholesterol
HIV	Crohn's Disease	Attention Deficit Hyperactivity Disorder
Inflammatory Bowel Disease	Ulcerative Colitis	Crohn's Disease
Low-density Lipoprotein	Total Cholesterol	Triglycerides
Major Depressive Disorder	Attention Deficit Hyperactivity Disorder	Depressive Symptoms
Mean Cell Haemoglobin	Mean Cell Volume	Red Blood Cell Count
Mean Cell Haemoglobin Concentration	Red Blood Cell Count	Haemoglobin
Mean Cell Volume	Mean Cell Haemoglobin	Red Blood Cell Count
Multiple Sclerosis	Primary Biliary Cirrhosis	Type 1 Diabetes
Neuroticism	Depressive Symptoms	Major Depressive Disorder
Packed Cell Volume	Haemoglobin	Red Blood Cell Count
Primary Biliary Cirrhosis	Multiple Sclerosis	Inflammatory Bowel Disease
Pubertal Growth	Height	BMI
Red Blood Cell Count	Haemoglobin	Packed Cell Volume
Rheumatoid Arthritis	Primary Biliary Cirrhosis	Multiple Sclerosis
Schizophrenia1	Schizophrenia2	Schizophrenia3
Schizophrenia2	Schizophrenia1	Schizophrenia3
Schizophrenia3	Schizophrenia2	Schizophrenia4
Schizophrenia4	Schizophrenia3	Schizophrenia1
Systemic Lupus Erythematosus	Primary Biliary Cirrhosis	Multiple Sclerosis
Total Cholesterol	Low-density Lipoprotein	Triglycerides
Triglycerides	High-density Lipoprotein	Total Cholesterol
Type 1 Diabetes	Multiple Sclerosis	Primary Biliary Cirrhosis
Type 2 Diabetes	Fasting Insulin	Fasting Glucose
Ulcerative Colitis	Inflammatory Bowel Disease	Crohn's Disease
Years of Education1	Education2	Attention Deficit Hyperactivity Disorder
Years of Education2	Education1	Attention Deficit Hyperactivity Disorder

Table S5: The top 3 correlated traits for the 44 GWASs.

8.6 Results for association mapping

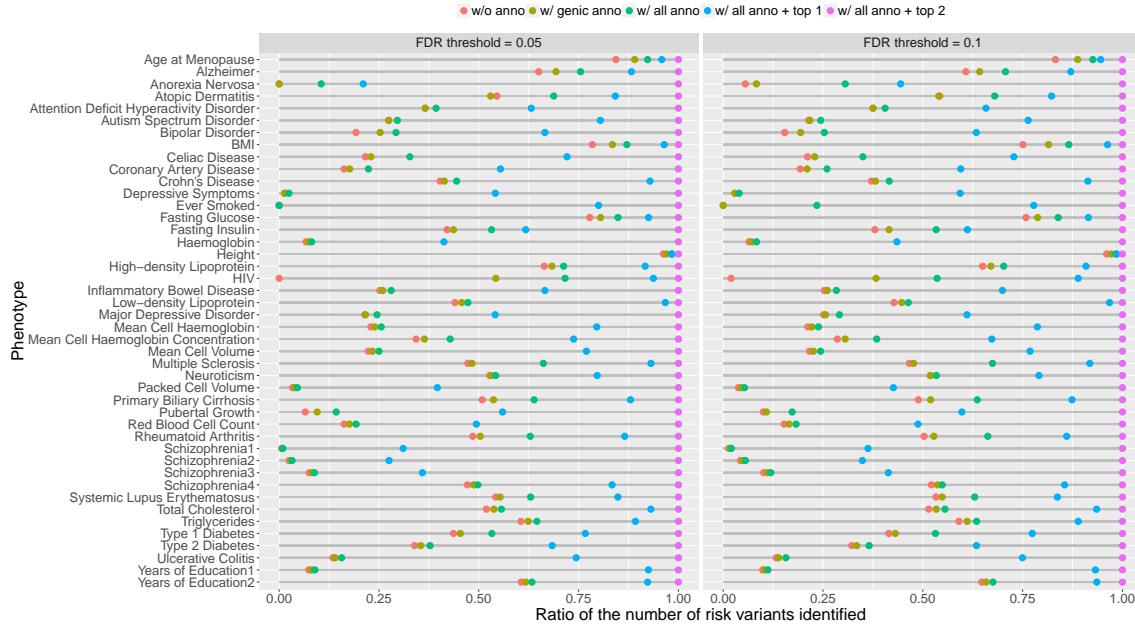


Figure S53: The numbers of variants identified to be associated with each of the 44 traits using LPM by five different analysis approaches: (i) separate analysis without annotation, (ii) separate analysis with genic category annotations, (iii) separate analysis with all annotations, (iv) joint analysis of the top 1 correlated trait with all annotations and (v) joint analysis of the top 2 correlated traits with all annotations. We controlled global FDR at 0.1. For visualization purpose, these numbers are normalized by dividing the corresponding number of variants identified by the fifth approach (joint analysis of the top 2 correlated traits with all annotations).

8.7 Manhattan plot of LDL and TC

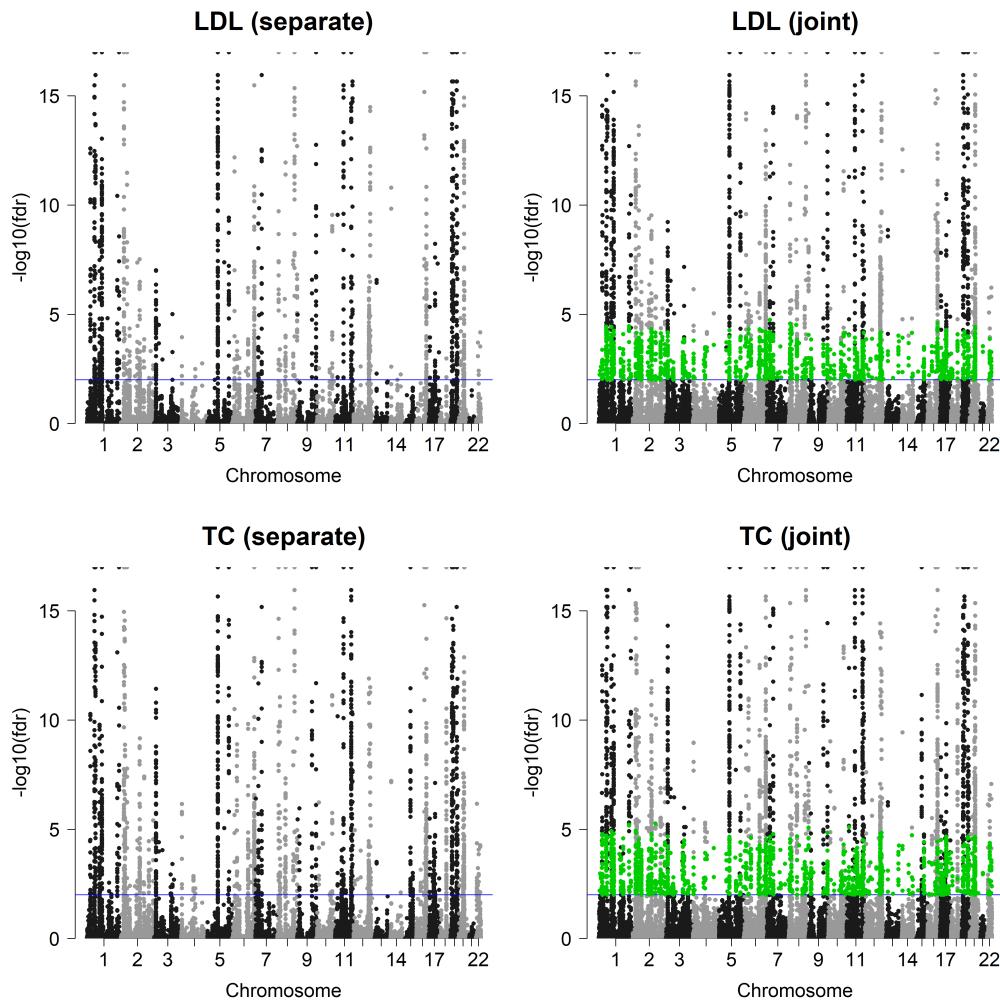


Figure S54: Manhattan plots of LDL and TC. Top left: separate analysis for LDL. Top right: joint analysis of LDL and TC for LDL. Bottom left: separate analysis for TC. Bottom right: joint analysis of LDL and TC for TC. The red lines indicate local $fdr = 0.01$. The green points denote the additional SNPs identified in joint analysis with $fdr \leq 0.01$. For $fdrs$ smaller than 10^{-17} , we threshold them at 10^{-17} .

8.8 Detailed results for the enrichment test of functional annotations

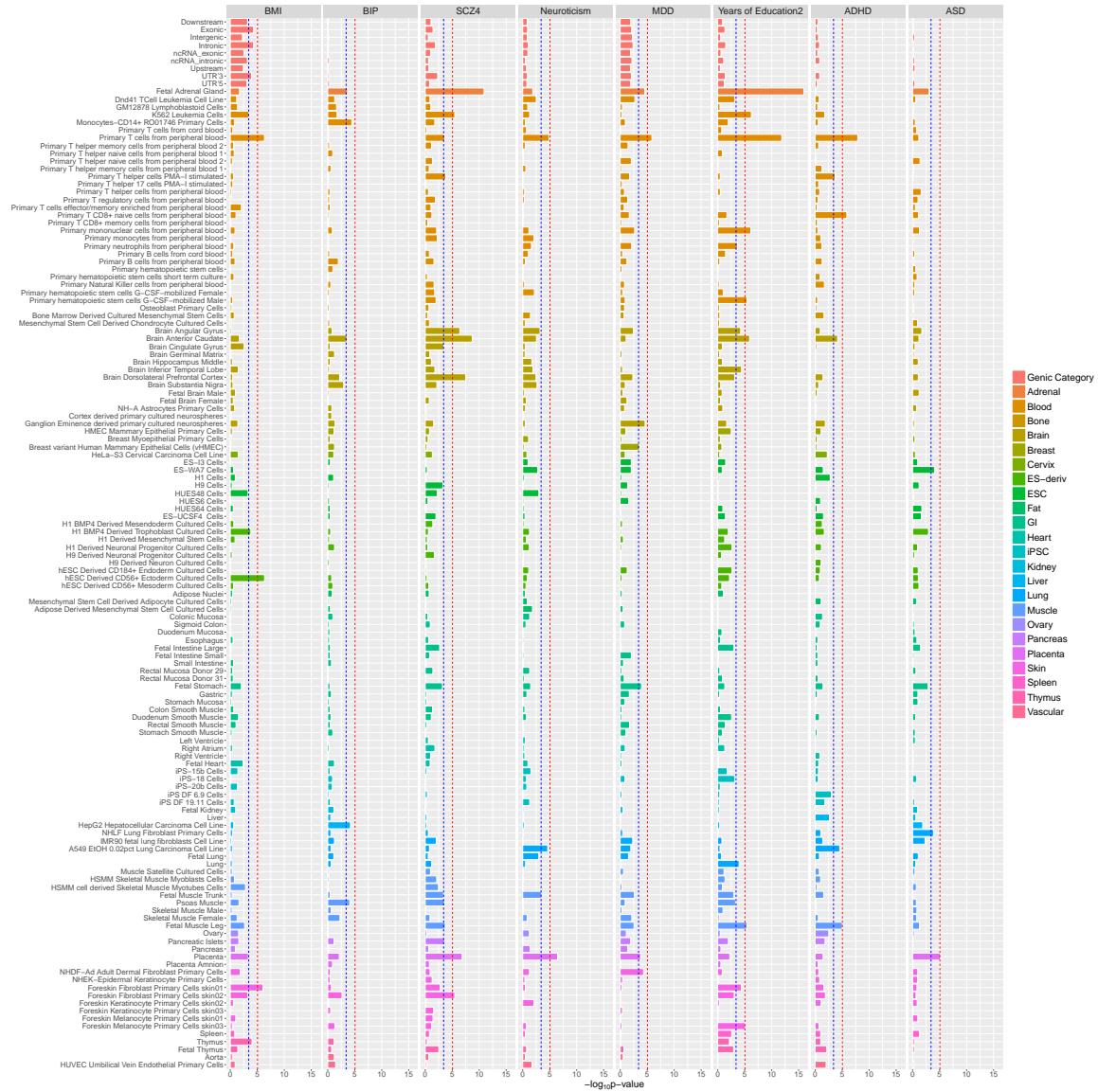


Figure S55: The $-\log_{10}(p\text{-value})$ of the enrichment test for nine genic category annotations and 127 cell-type specific functional annotations for traits in block 1. The red lines represent $-\log_{10}\left(\frac{0.05}{(9+127)\times 44}\right)$. The blue lines represent $-\log_{10}\left(\frac{0.05}{9+127}\right)$.

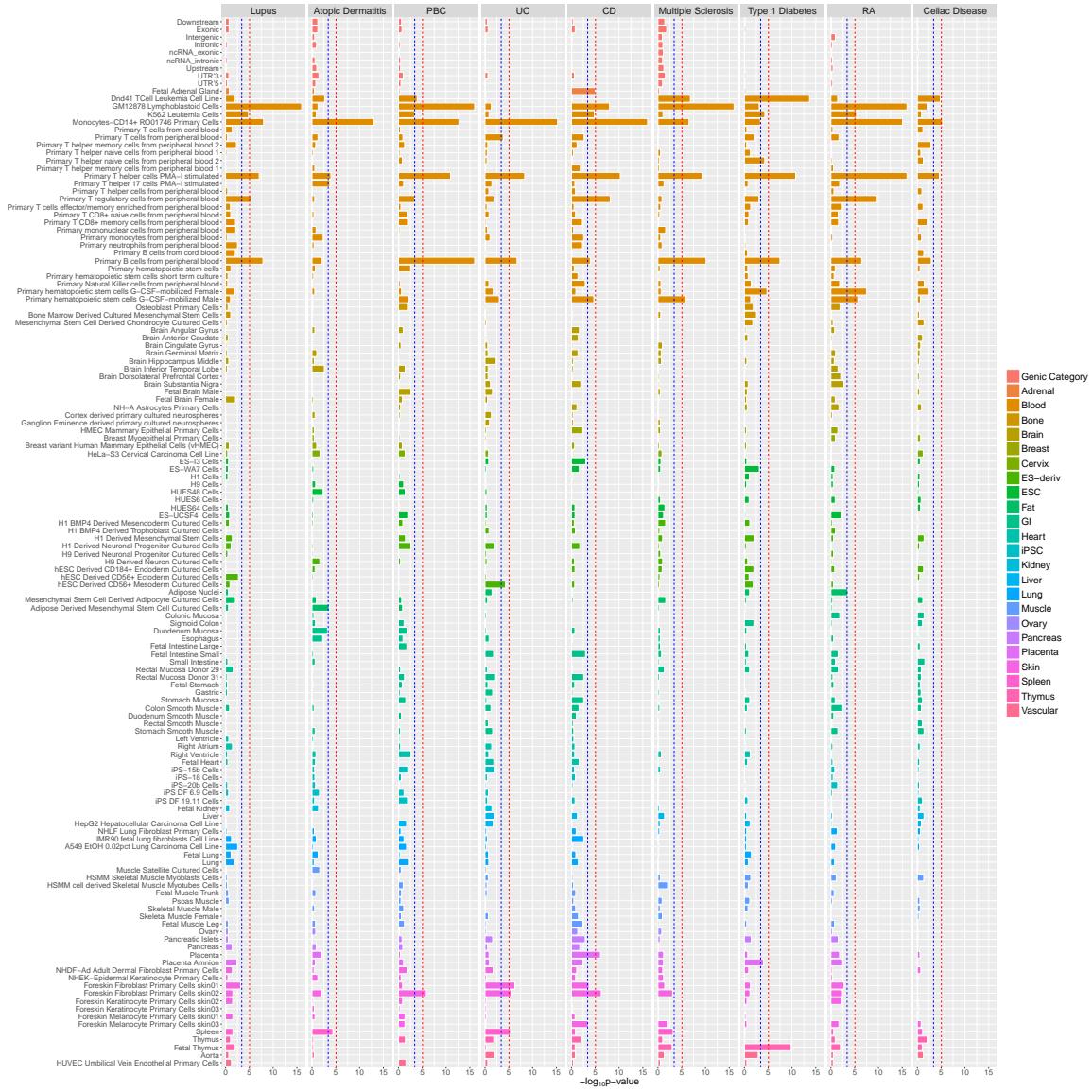


Figure S56: The $-\log_{10}(p\text{-value})$ of the enrichment test for nine genic category annotations and 127 cell-type specific functional annotations for traits in block 2. The red lines represent $-\log_{10}\left(\frac{0.05}{(9+127)\times 44}\right)$. The blue lines represent $-\log_{10}\left(\frac{0.05}{9+127}\right)$.

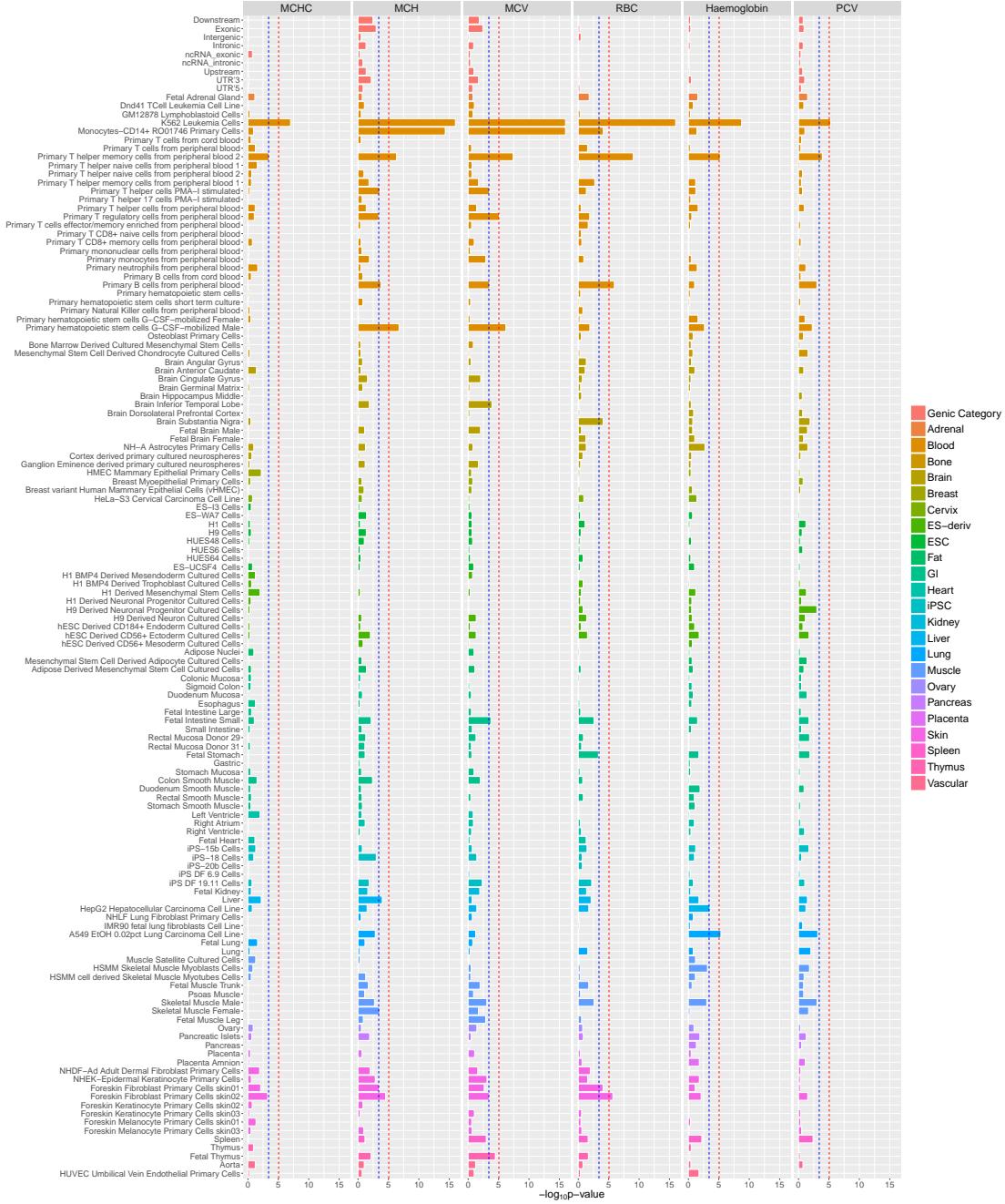


Figure S57: The $-\log_{10}(p\text{-value})$ of the enrichment test for nine genic category annotations and 127 cell-type specific functional annotations for traits in block 3. The red lines represent $-\log_{10}\left(\frac{0.05}{(9+127)\times 44}\right)$. The blue lines represent $-\log_{10}\left(\frac{0.05}{9+127}\right)$.

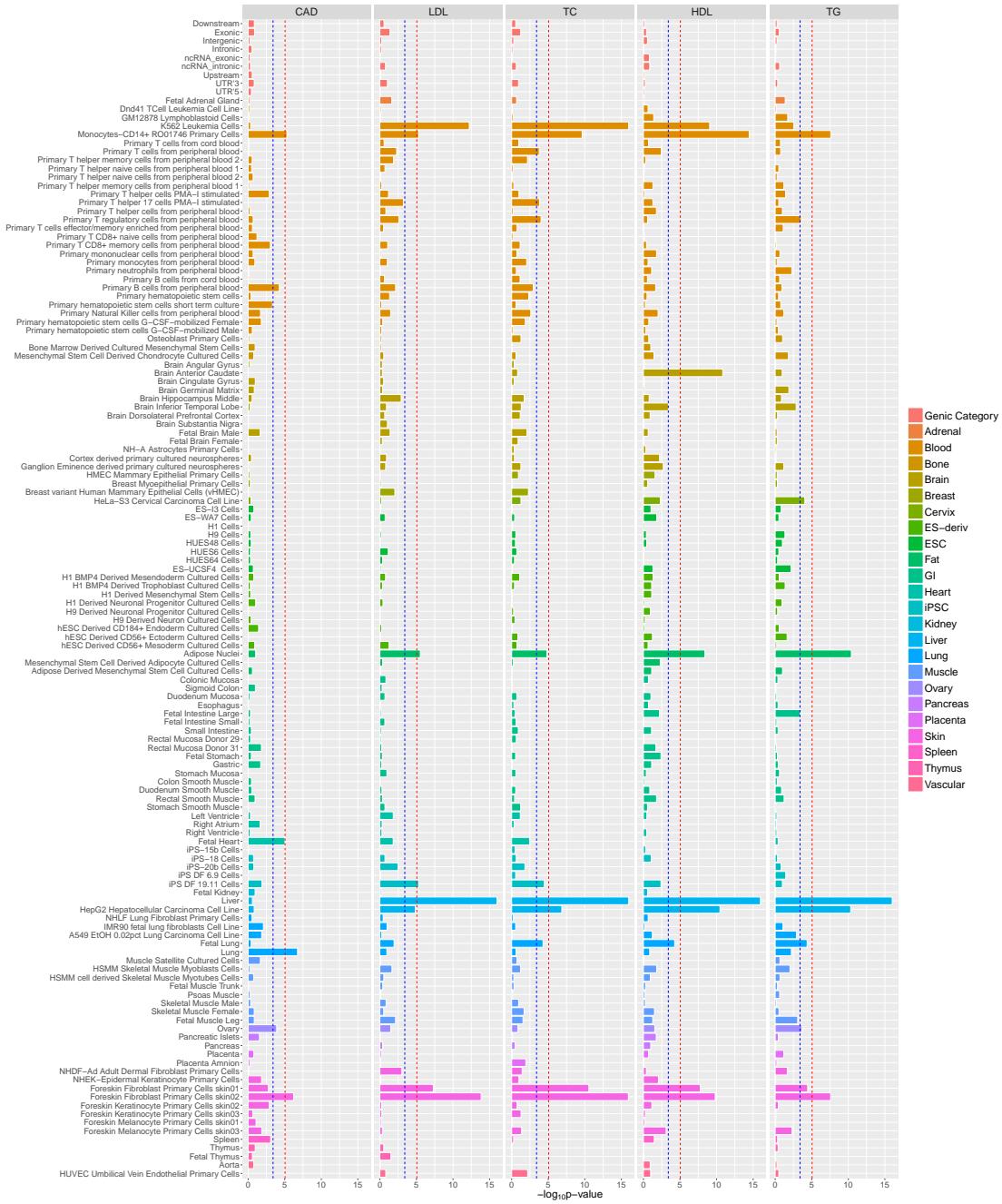


Figure S58: The $-\log_{10}(p\text{-value})$ of the enrichment test for nine genic category annotations and 127 cell-type specific functional annotations for traits in block 4. The red lines represent $-\log_{10}\left(\frac{0.05}{(9+127)*44}\right)$. The blue lines represent $-\log_{10}\left(\frac{0.05}{9+127}\right)$.

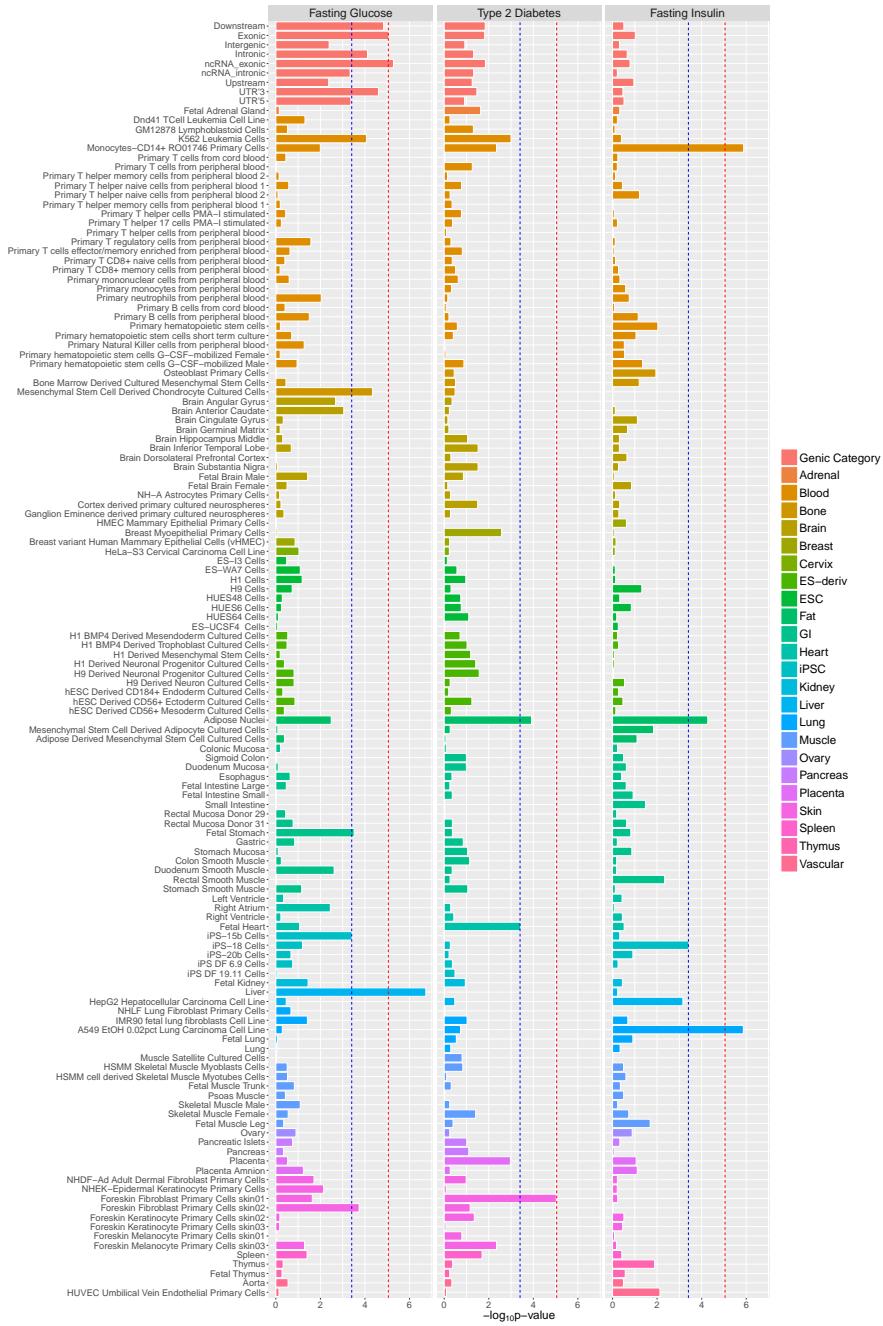


Figure S59: The $-\log_{10}(p\text{-value})$ of the enrichment test for nine genic category annotations and 127 cell-type specific functional annotations for traits in block 5. The red lines represent $-\log_{10}\left(\frac{0.05}{(9+127)\times 44}\right)$. The blue lines represent $-\log_{10}\left(\frac{0.05}{9+127}\right)$.

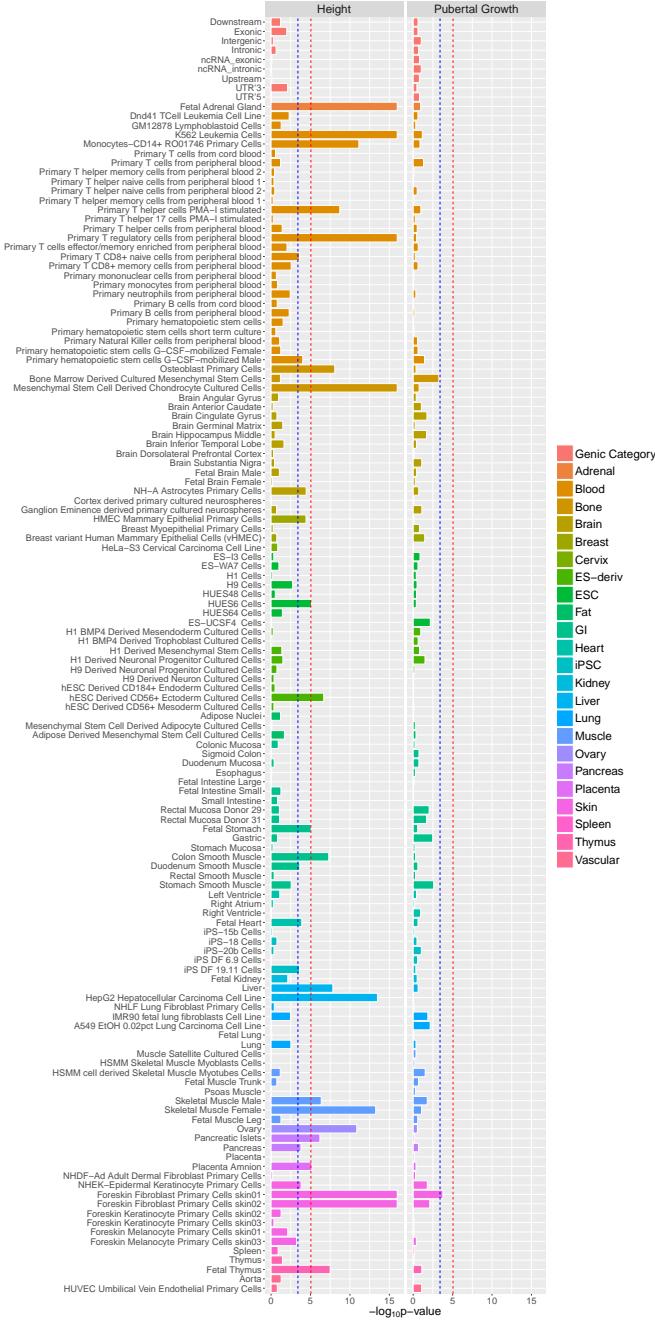


Figure S60: The $-\log_{10}(p\text{-value})$ of the enrichment test for nine genic category annotations and 127 cell-type specific functional annotations for traits in block 6. The red lines represent $-\log_{10}\left(\frac{0.05}{(9+127)\times 44}\right)$. The blue lines represent $-\log_{10}\left(\frac{0.05}{9+127}\right)$.

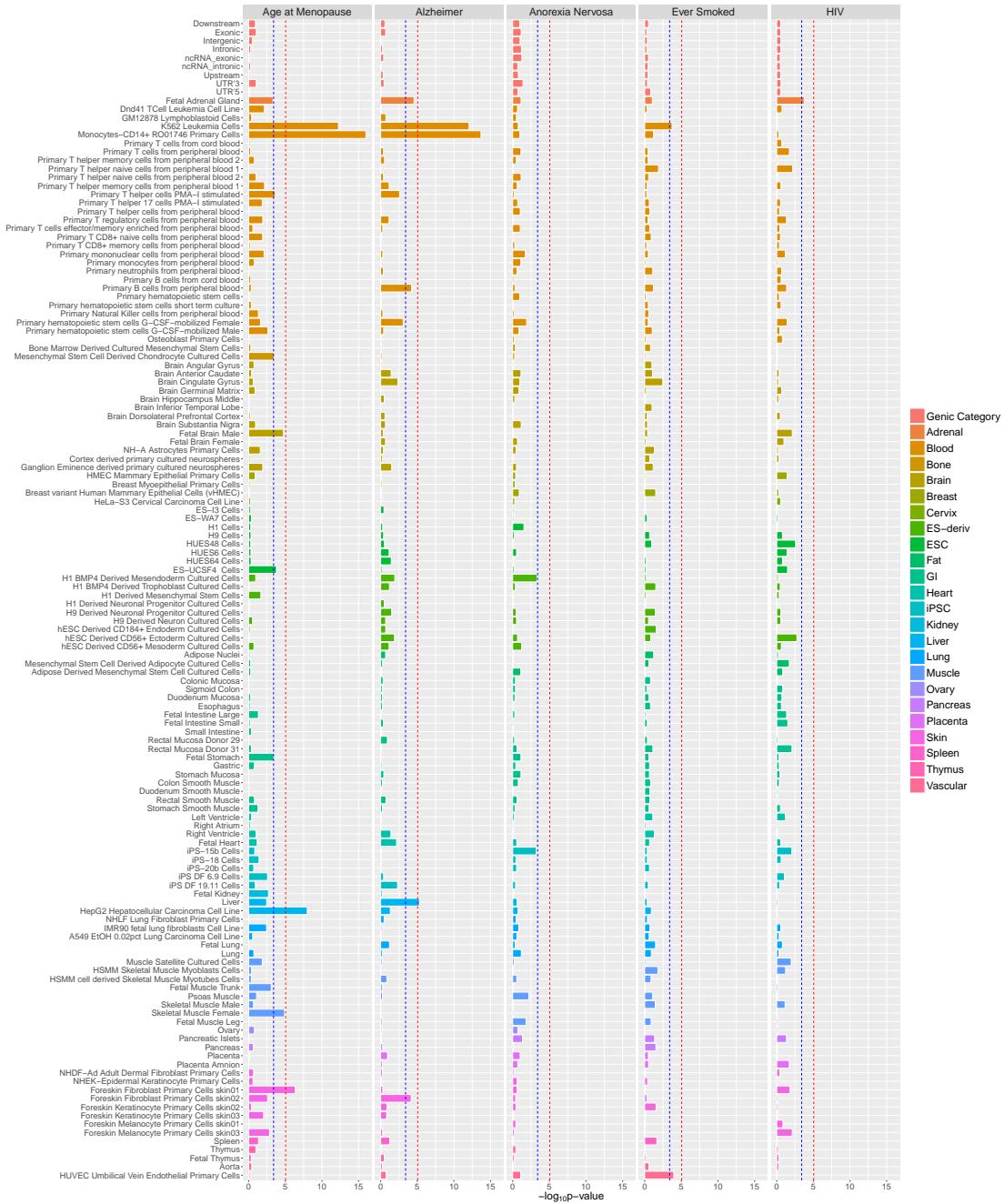


Figure S61: The $-\log_{10}(p\text{-value})$ of the enrichment test for nine genic category annotations and 127 cell-type specific functional annotations for other traits. The red lines represent $-\log_{10}\left(\frac{0.05}{(9+127)\times 44}\right)$. The blue lines represent $-\log_{10}\left(\frac{0.05}{9+127}\right)$.

8.9 Estimated coefficients of 9 genic category annotations

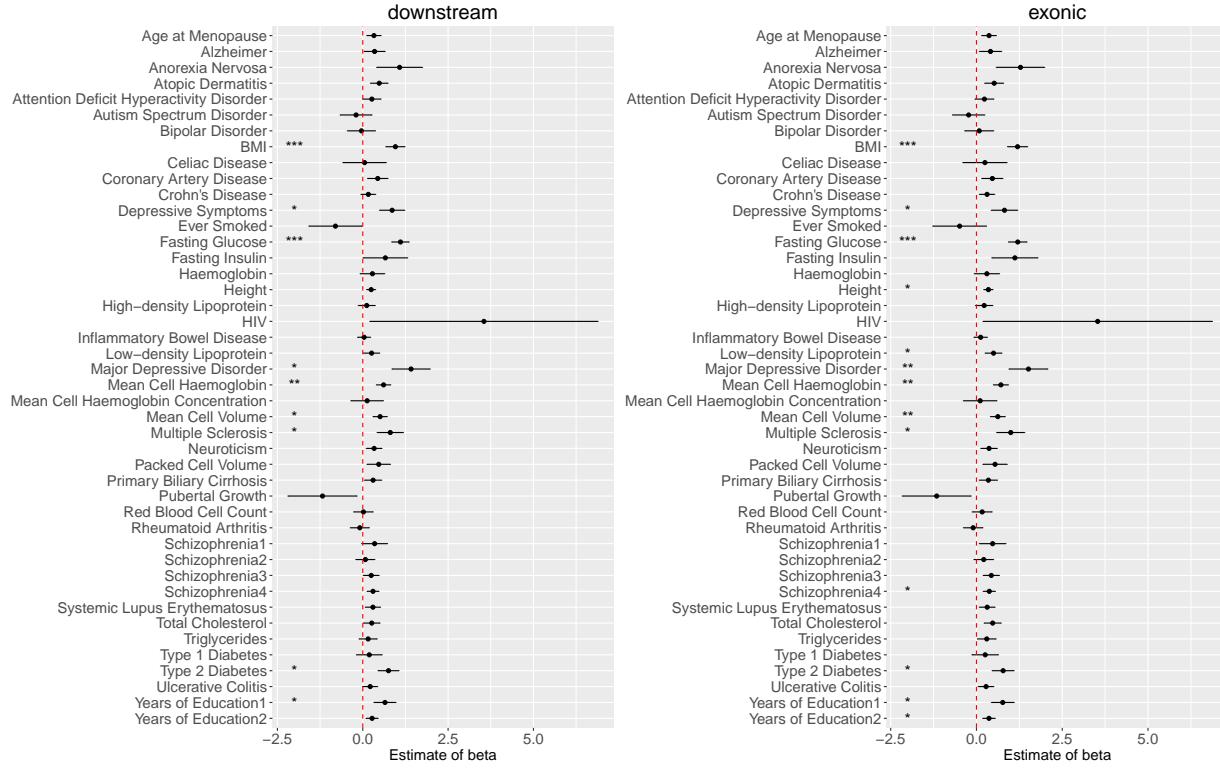


Figure S62: The estimated coefficients of downstream and exonic for 44 GWASs. The bars represent one standard error. The symbol “*”, “**” and “***” means the p value for the test of annotation enrichment is ≤ 0.05 , ≤ 0.01 and ≤ 0.001 respectively.

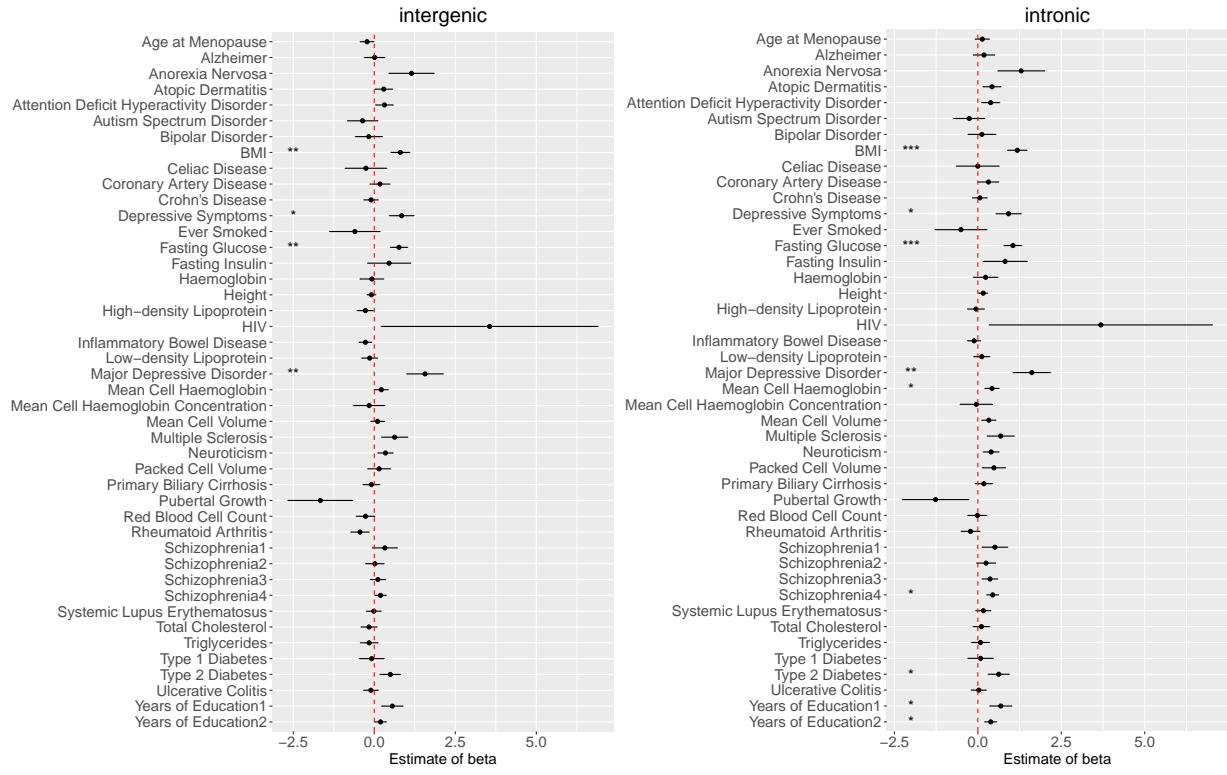


Figure S63: The estimated coefficients of intergenic and intronic for 44 GWASs. The bars represent one standard error. The symbol “*”, “**” and “***” means the p value for the test of annotation enrichment is ≤ 0.05 , ≤ 0.01 and ≤ 0.001 respectively.

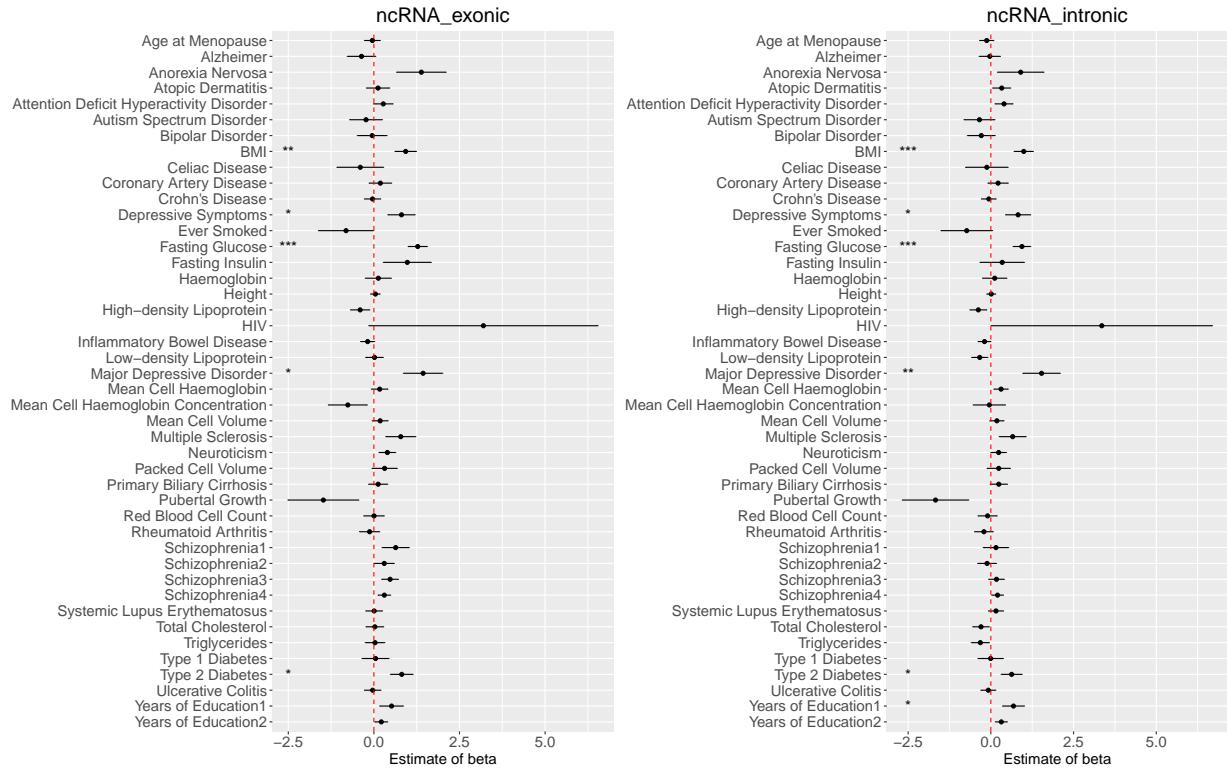


Figure S64: The estimated coefficients of ncRNA_exonic and ncRNA_intronic for 44 GWASs. The bars represent one standard error. The symbol “*”, “**” and “***” means the p value for the test of annotation enrichment is ≤ 0.05 , ≤ 0.01 and ≤ 0.001 respectively.

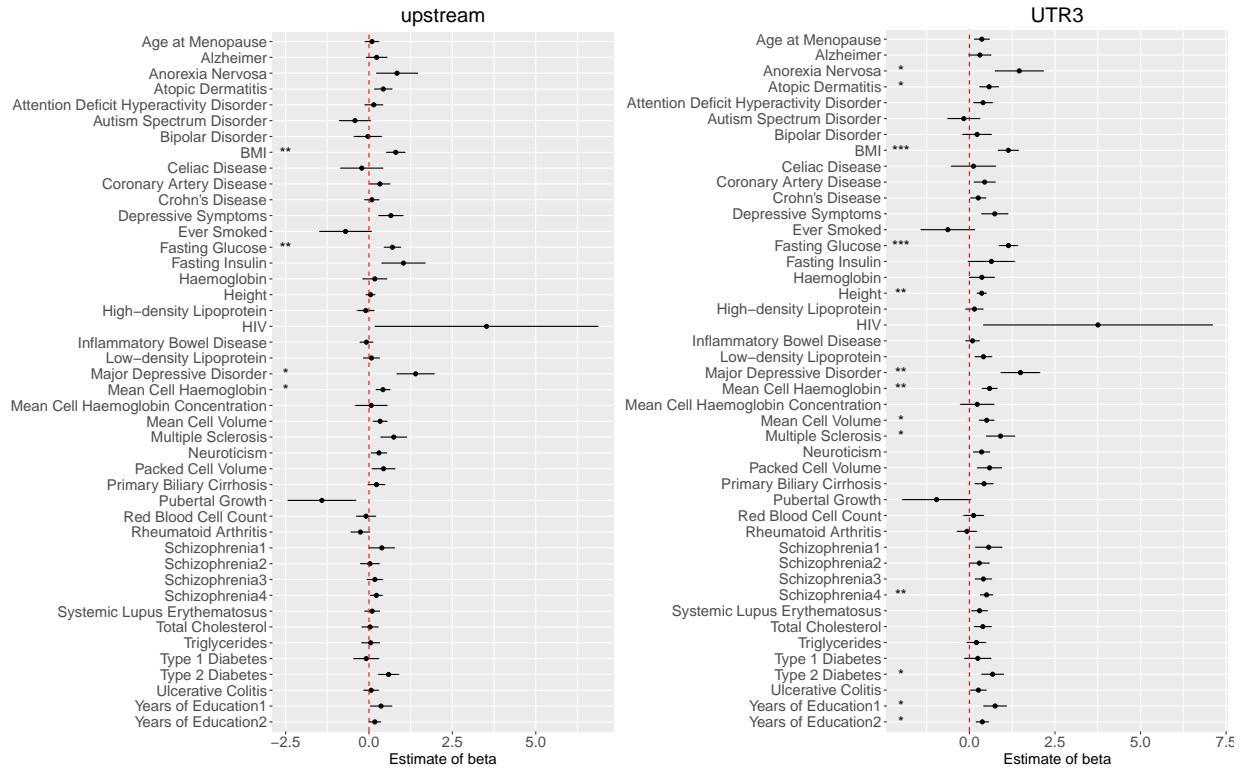


Figure S65: The estimated coefficients of upstream and UTR'3 for 44 GWASs. The bars represent one standard error. The symbol “*”, “**” and “***” means the p value for the test of annotation enrichment is ≤ 0.05 , ≤ 0.01 and ≤ 0.001 respectively.

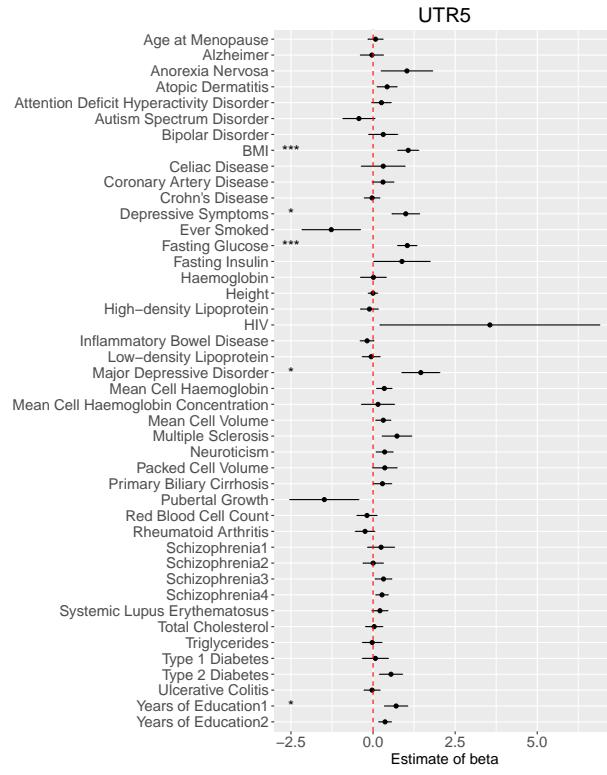


Figure S66: The estimated coefficients of UTR'5 for 44 GWASs. The bars represent one standard error. The symbol “*”, “**” and “***” means the p value for the test of annotation enrichment is ≤ 0.05 , ≤ 0.01 and ≤ 0.001 respectively.

8.10 Estimated coefficients of 127 cell-type specific functional annotations

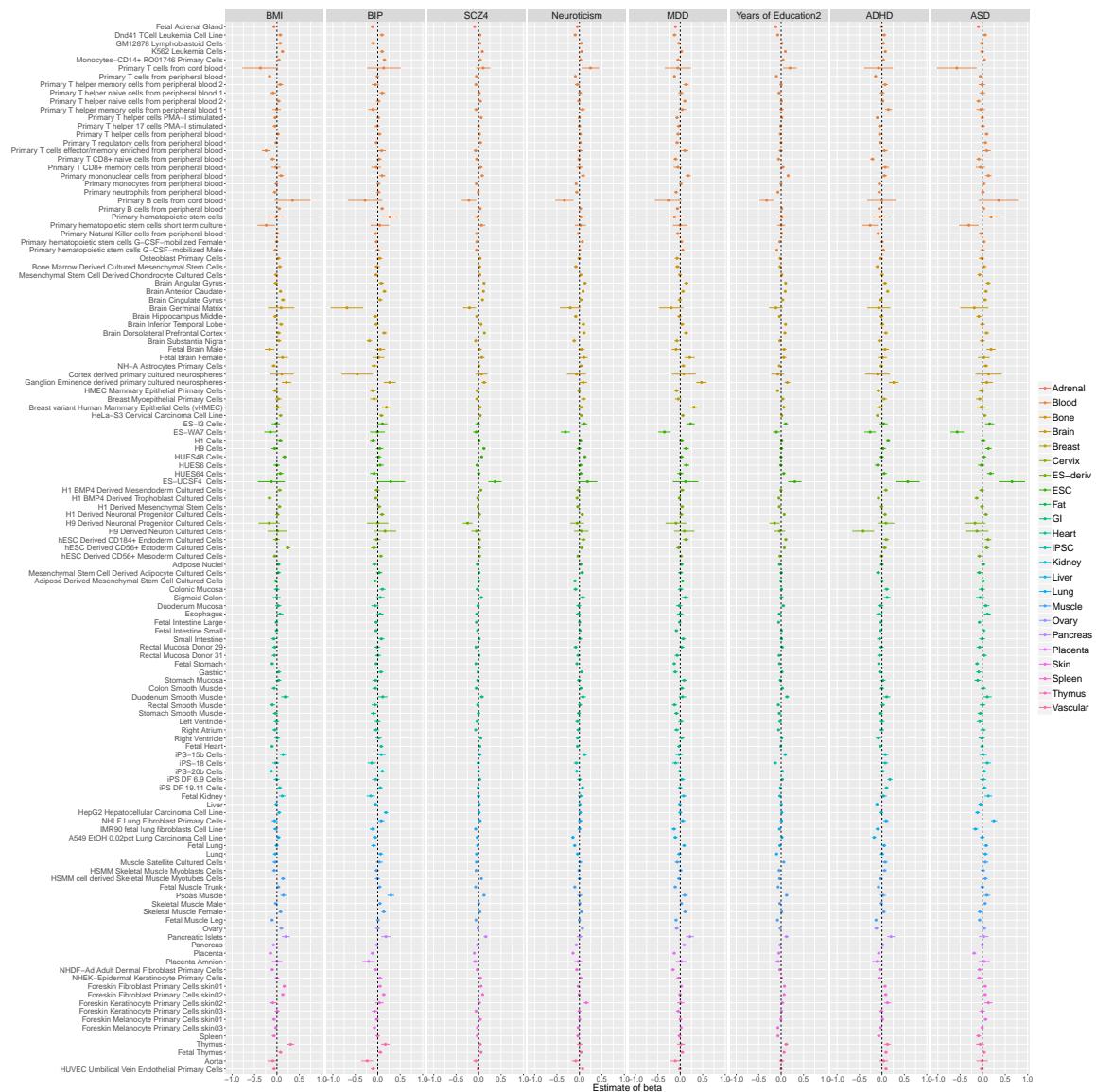


Figure S67: The estimated coefficients of 127 cell-type specific functional annotations for traits in block 1. The bars represent one standard error.

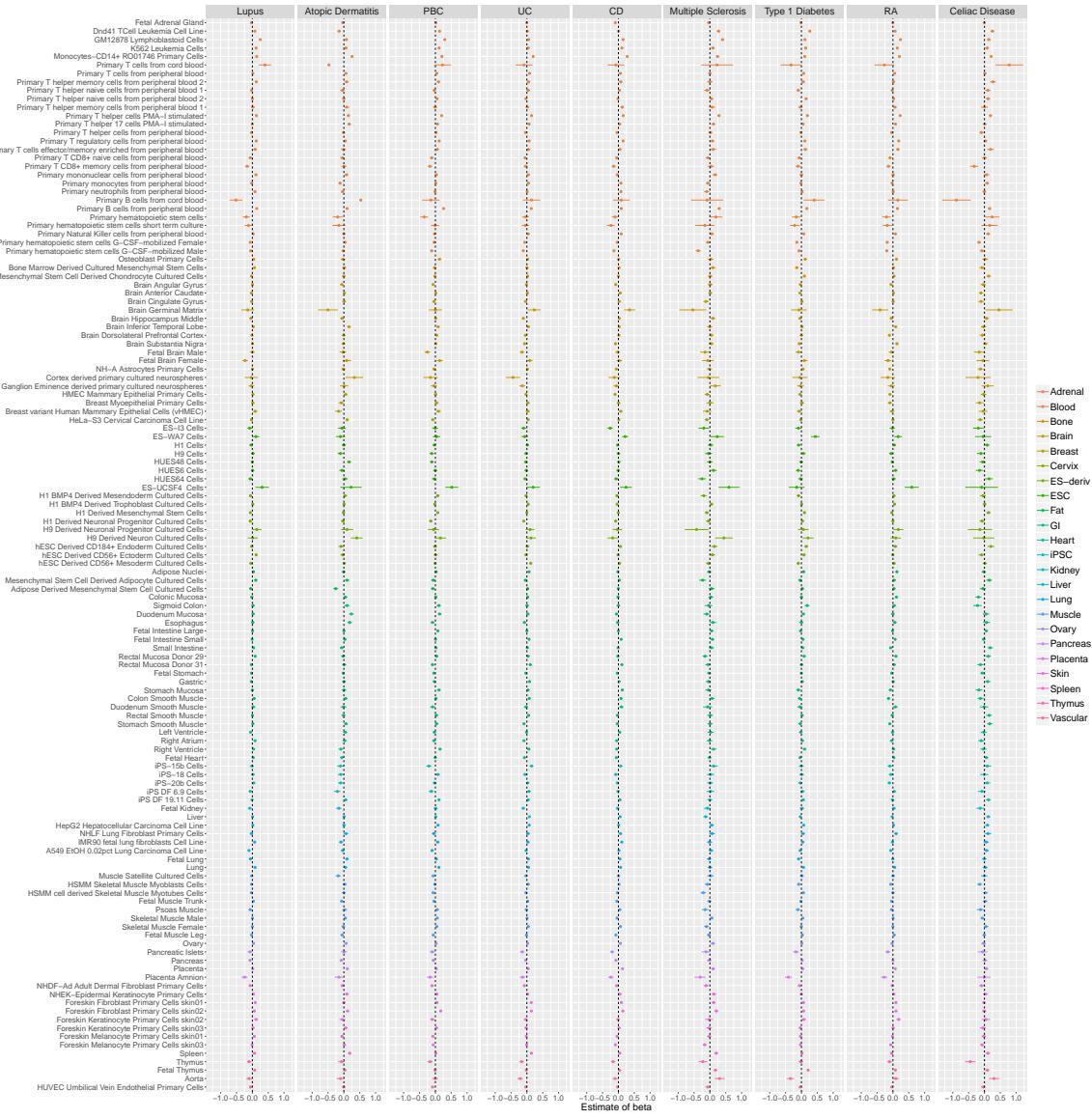


Figure S68: The estimated coefficients of 127 cell-type specific functional annotations for traits in block 2. The bars represent one standard error.

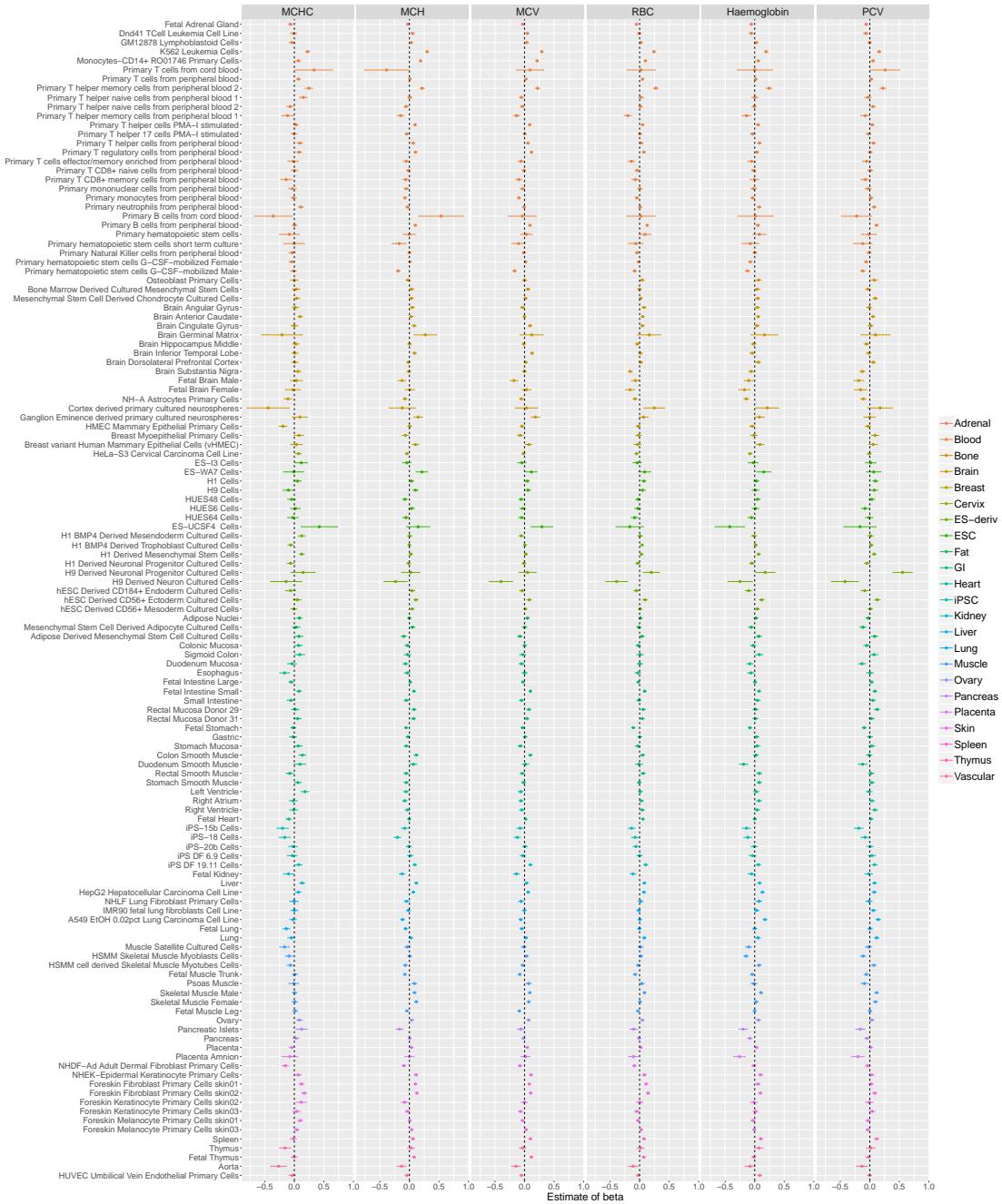


Figure S69: The estimated coefficients of 127 cell-type specific functional annotations for traits in block 3. The bars represent one standard error.

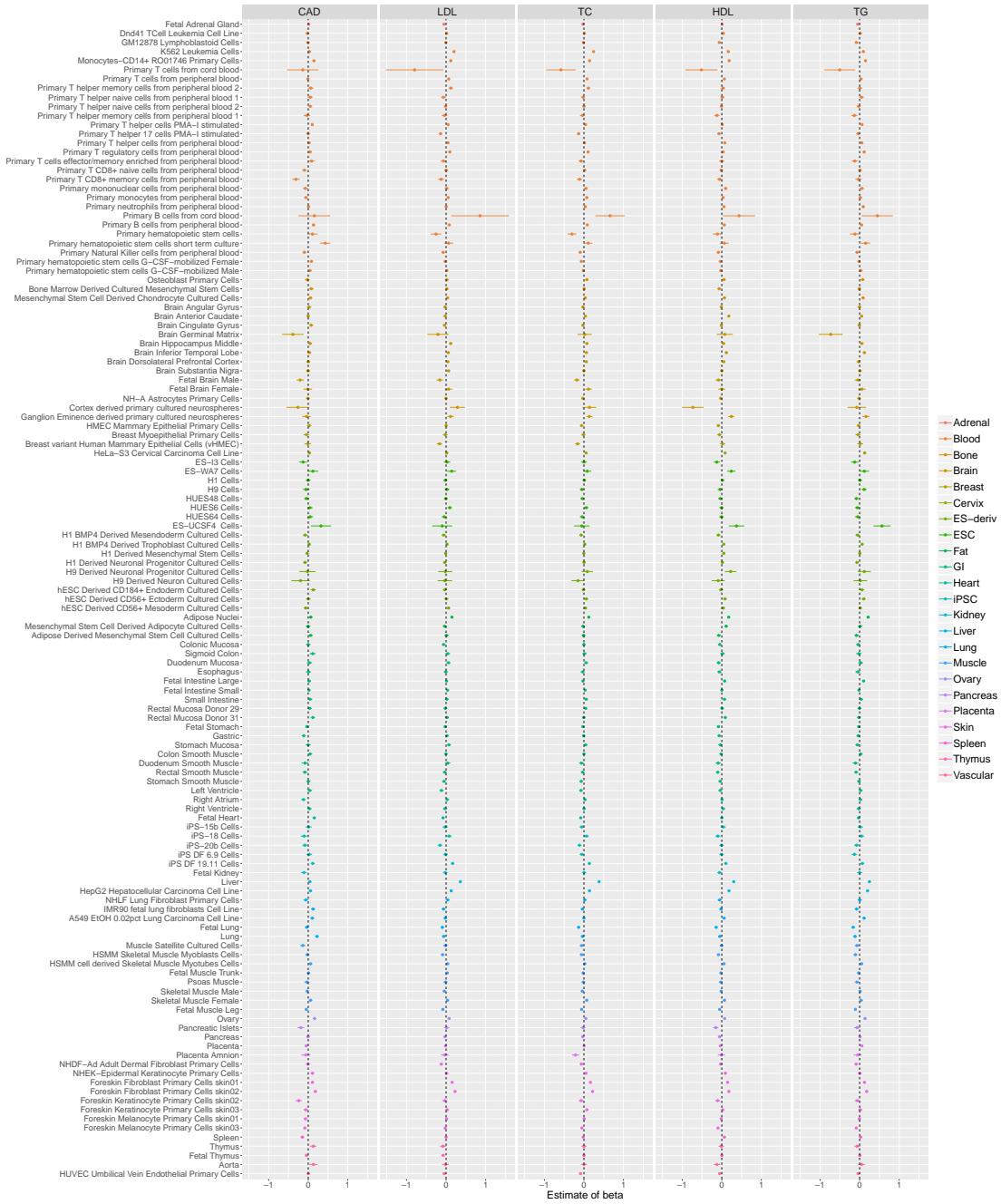


Figure S70: The estimated coefficients of 127 cell-type specific functional annotations for traits in block 4. The bars represent one standard error.

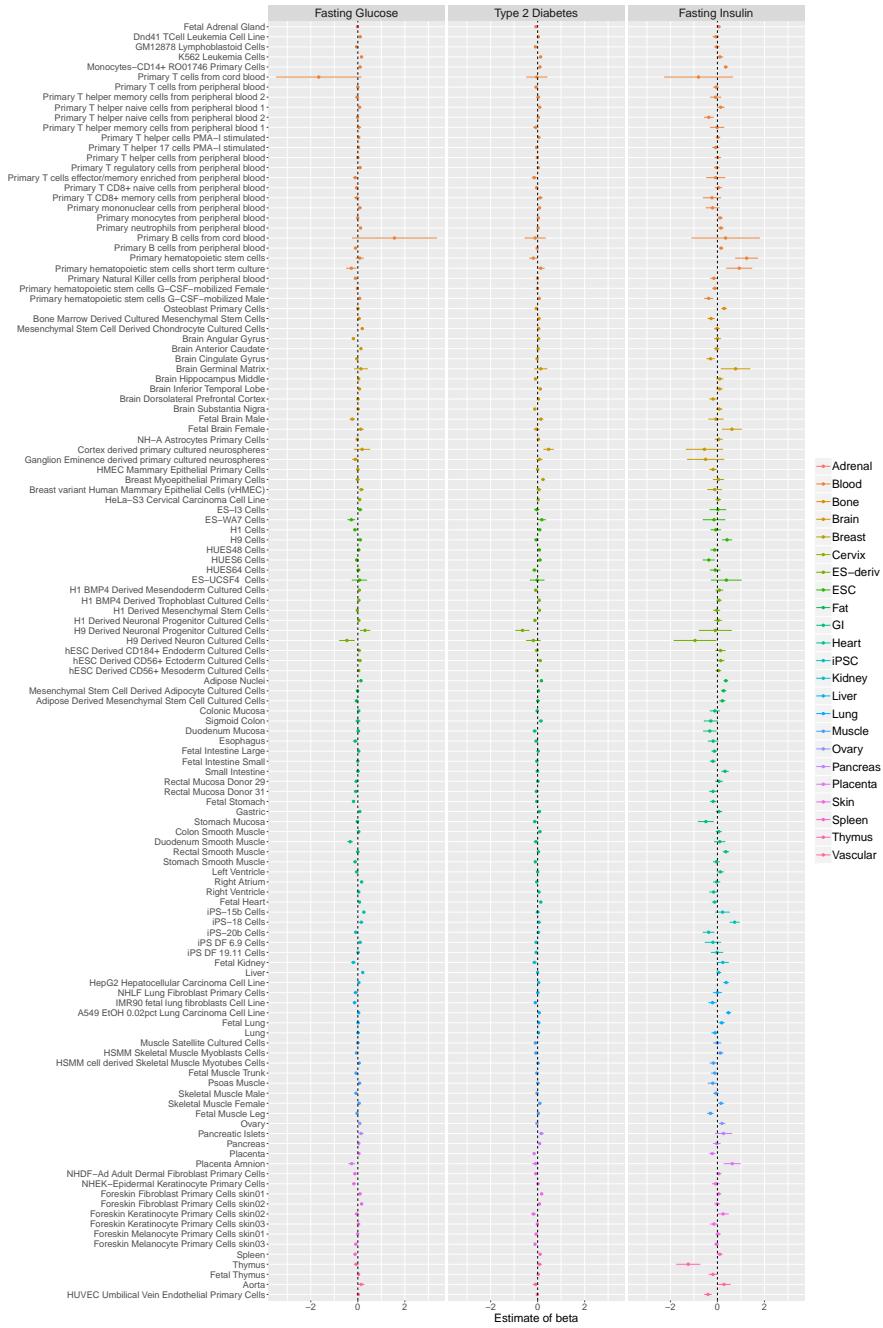


Figure S71: The estimated coefficients of 127 cell-type specific functional annotations for traits in block 5. The bars represent one standard error.

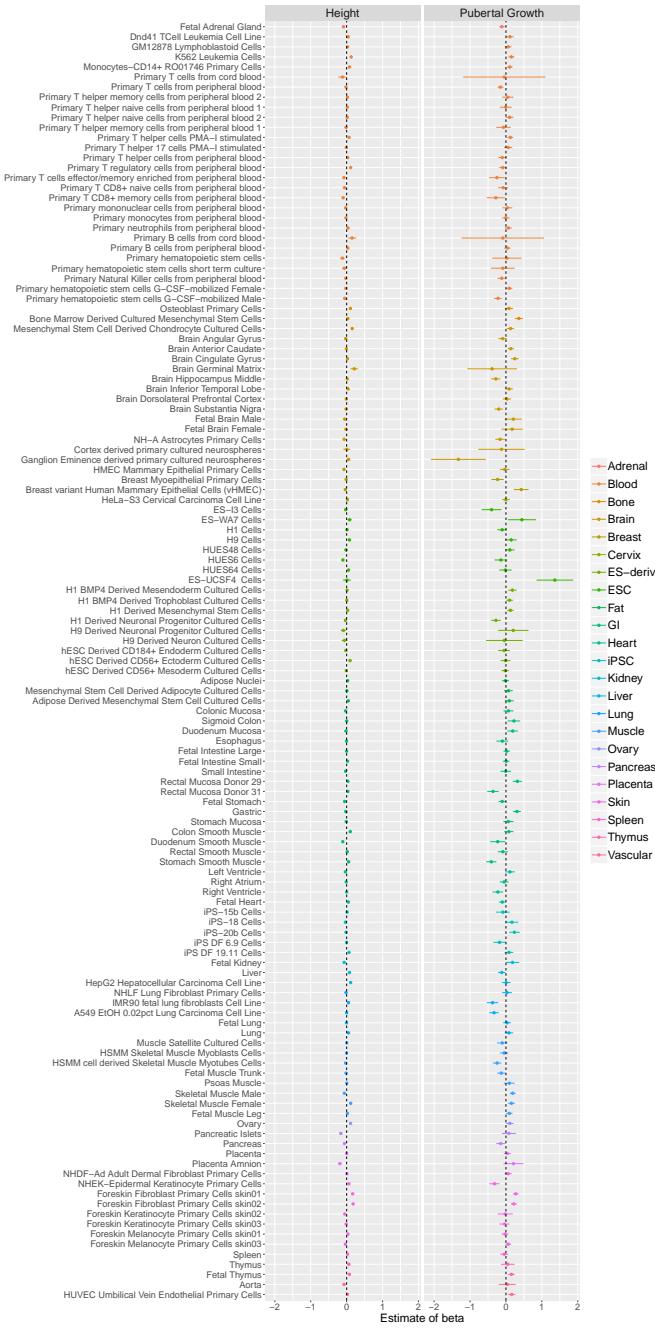


Figure S72: The estimated coefficients of 127 cell-type specific functional annotations for traits in block 6. The bars represent one standard error.

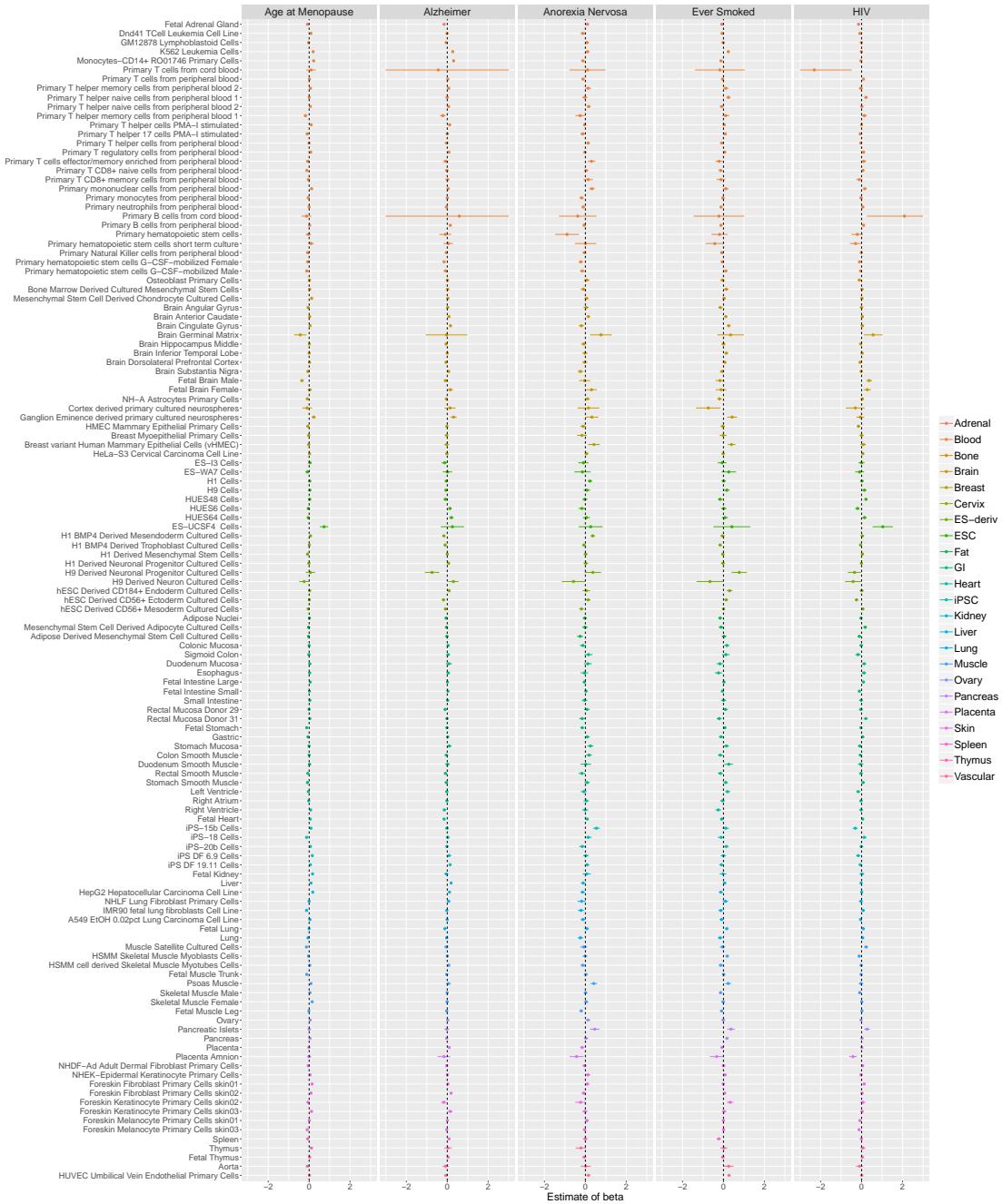


Figure S73: The estimated coefficients of 127 cell-type specific functional annotations for other traits. The bars represent one standard error. The error bars for Primary T cells from cord blood and Primary B cells from cord blood in Alzheimer and HIV are out of the axis limits. We set their error bars within -3 and 3.

8.11 Four Schizophrenia GWASs

Table S6: Sample sizes for four schizophrenia GWASs.

	Schizophrenia1	Schizophrenia2	Schizophrenia3	Schizophrenia4
case	9,379	9,394	13,833	36,989
control	7,736	12,462	18,310	113,075

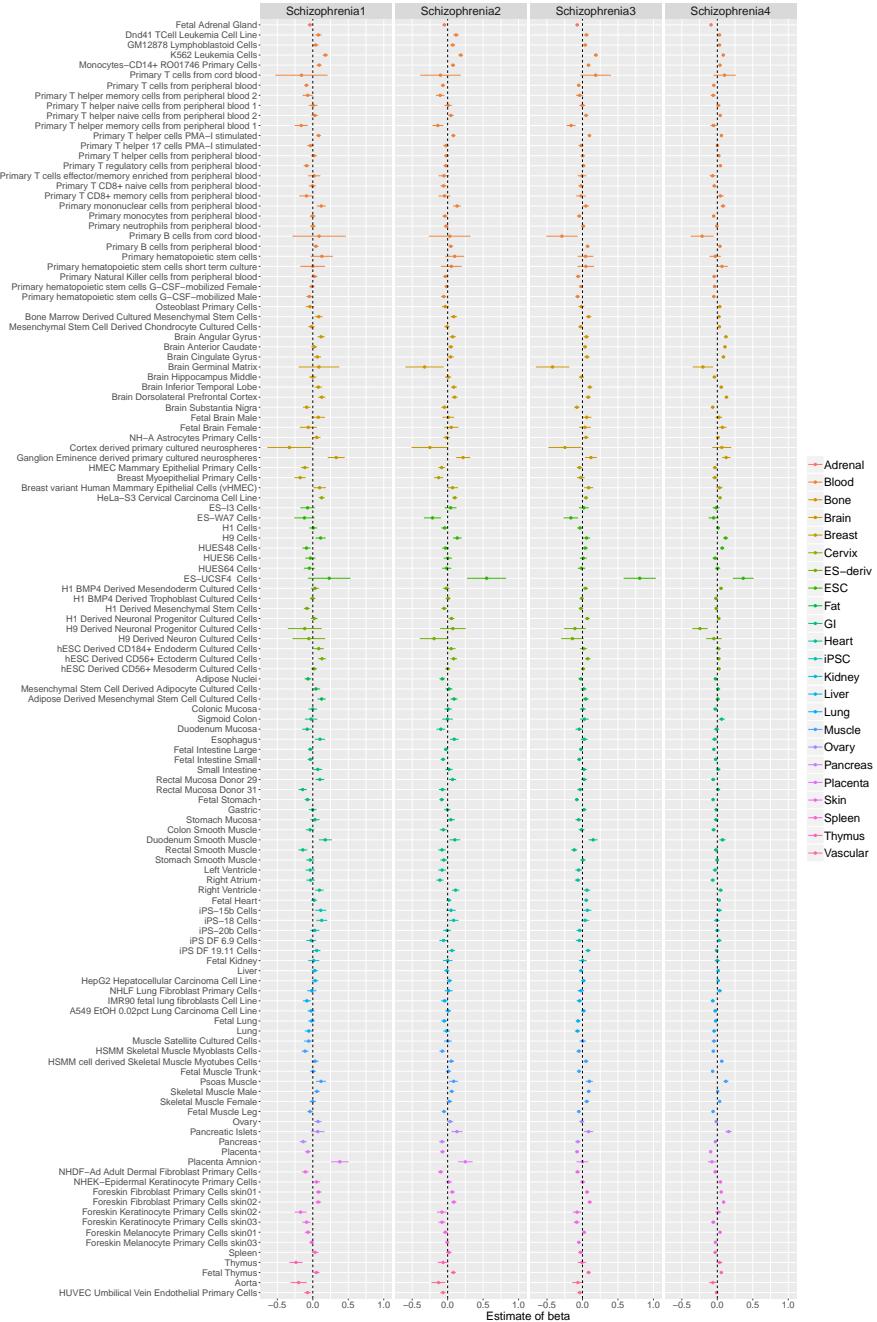


Figure S74: The estimated coefficients of 127 cell-type specific functional annotations for four SCZ GWASs. The bars represent one standard error.

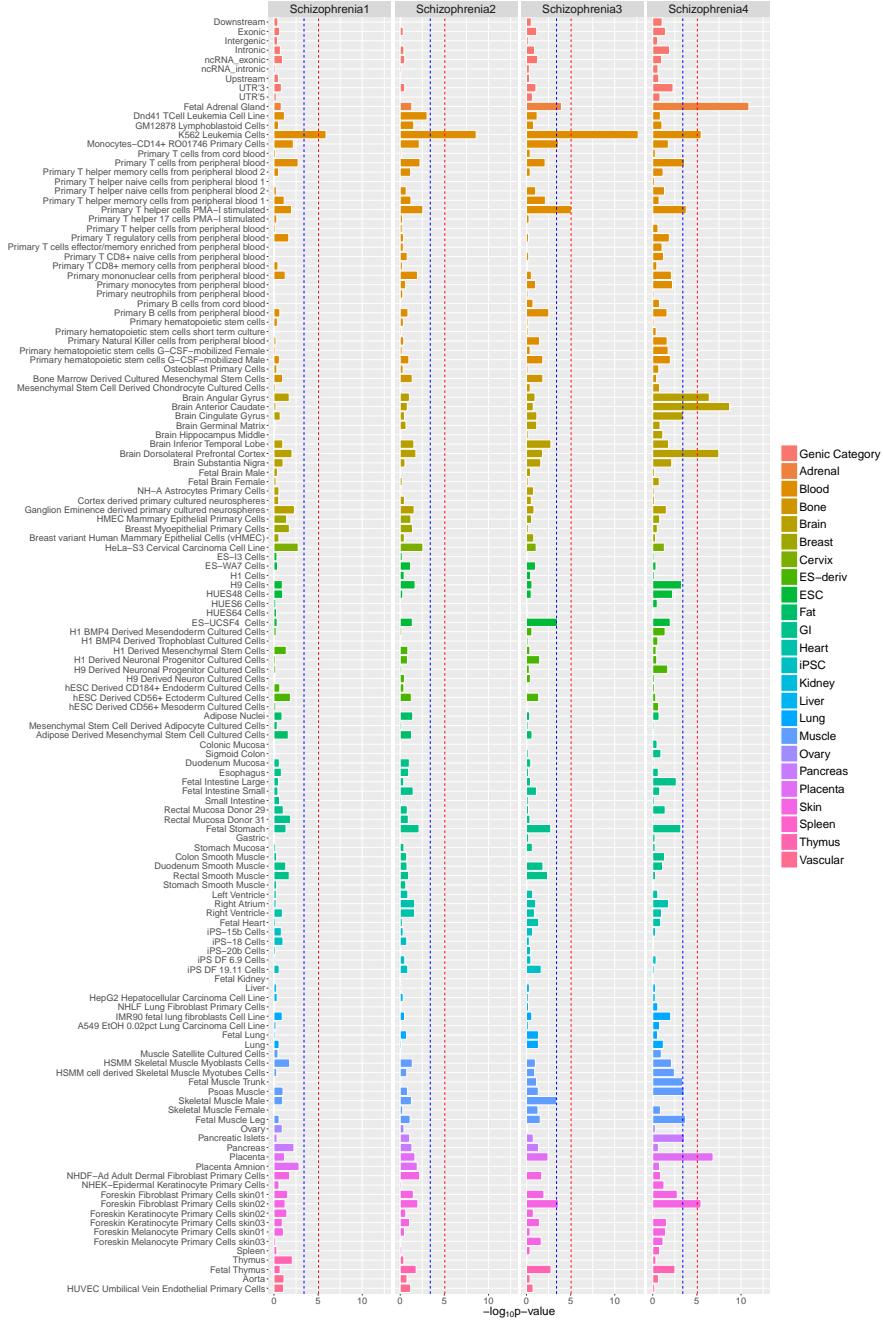


Figure S75: The $-\log_{10}(p\text{-value})$ of the enrichment test for nine genic category annotations and 127 cell-type specific functional annotations for four SCZ GWASs. The red lines represent $-\log_{10}\left(\frac{0.05}{(9+127)\times 44}\right)$. The blue lines represent $-\log_{10}\left(\frac{0.05}{9+127}\right)$.

8.12 More replications results

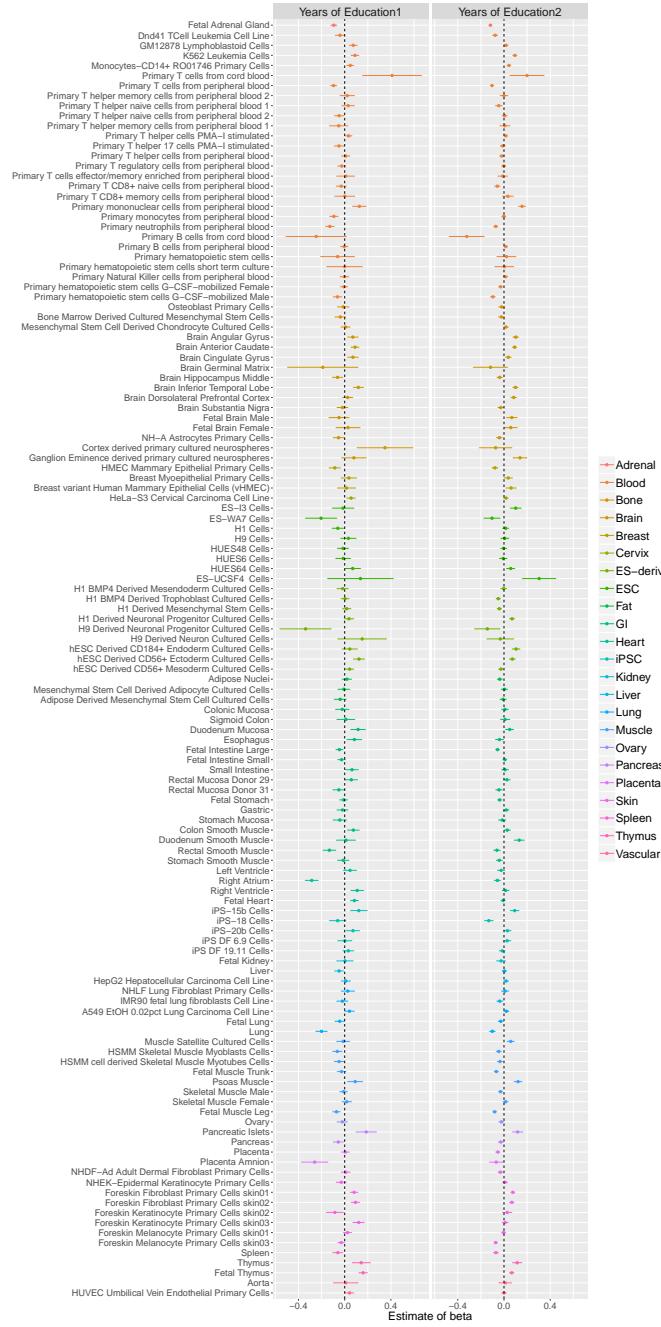


Figure S76: The estimated coefficients of 127 cell-type specific functional annotations for two years of education GWASs. The bars represent one standard error.

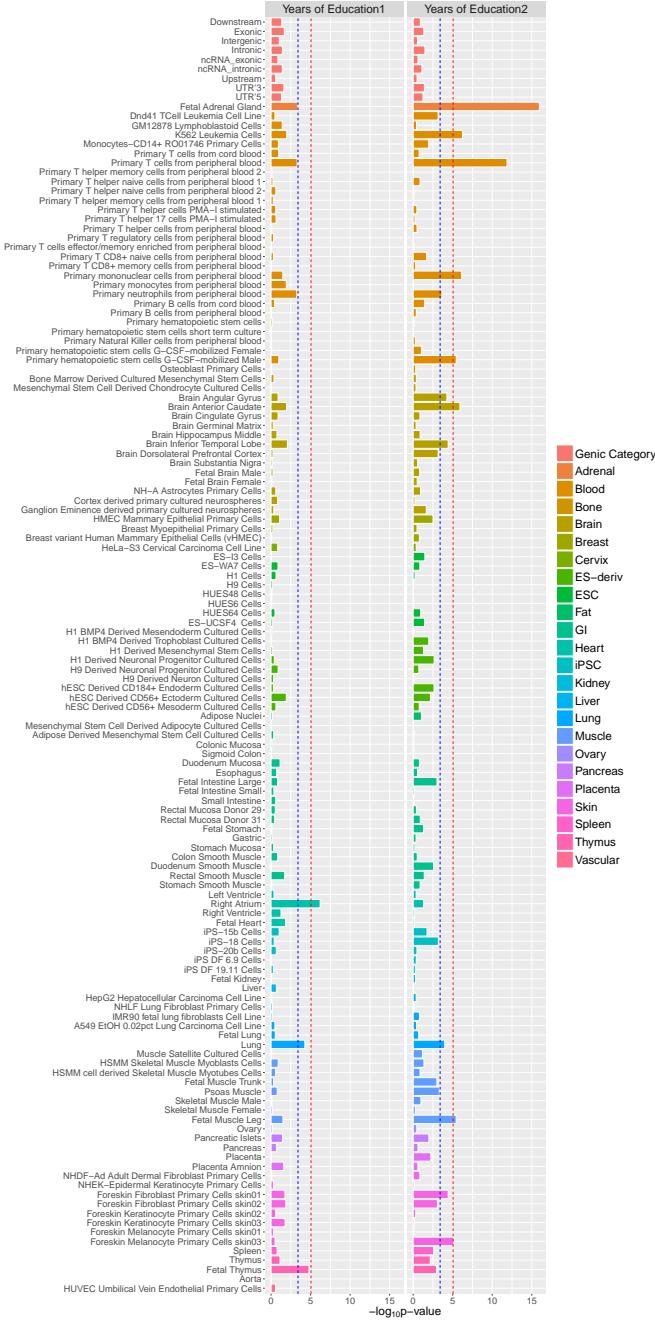


Figure S77: The $-\log_{10}(p\text{-value})$ of the enrichment test for nine genic category annotations and 127 cell-type specific functional annotations for two years of education GWASs. The red lines represent $-\log_{10}\left(\frac{0.05}{(9+127)\times 44}\right)$. The blue lines represent $-\log_{10}\left(\frac{0.05}{9+127}\right)$.

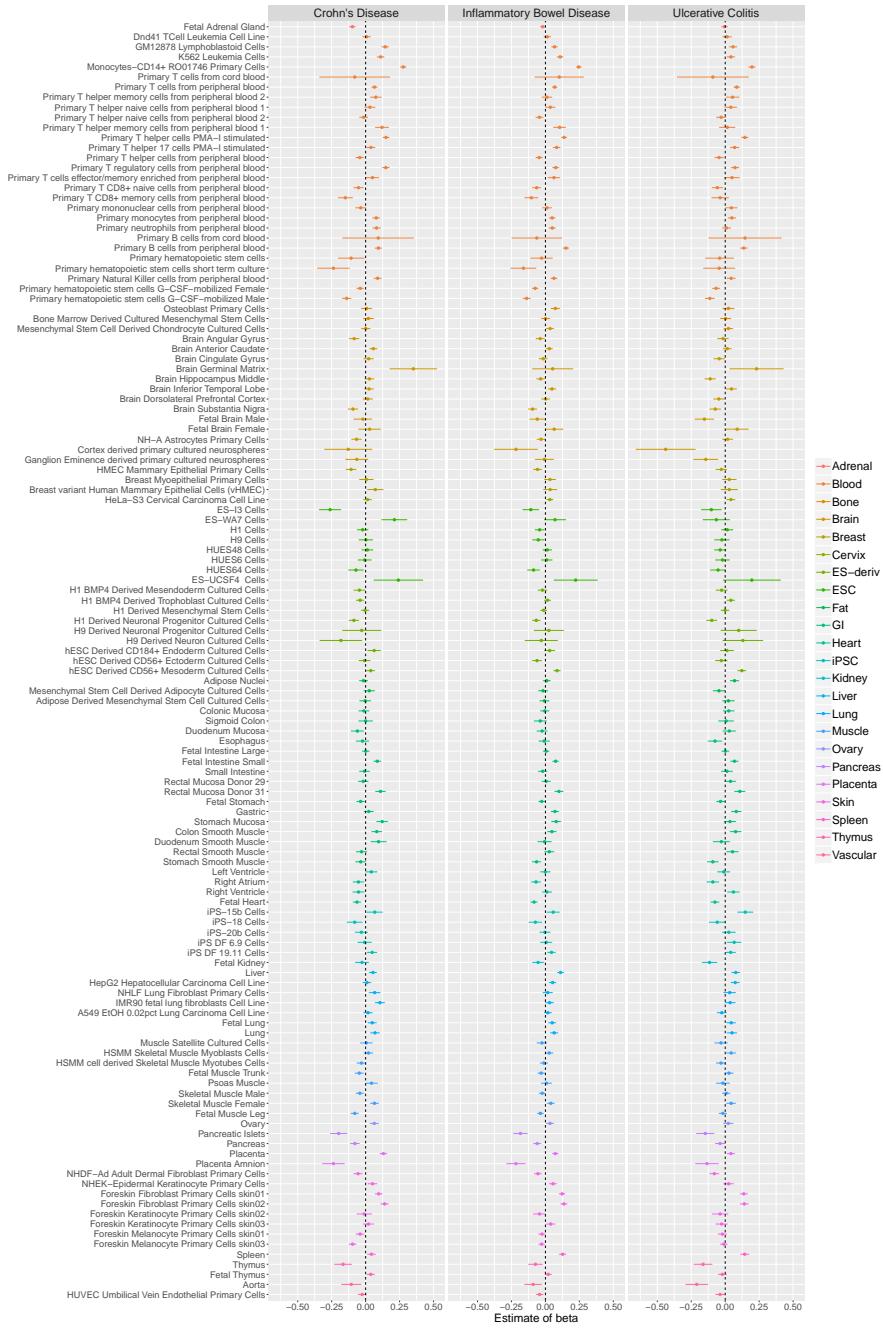


Figure S78: The estimated coefficients of 127 cell-type specific functional annotations for UC, CD and IBD. The bars represent one standard error.

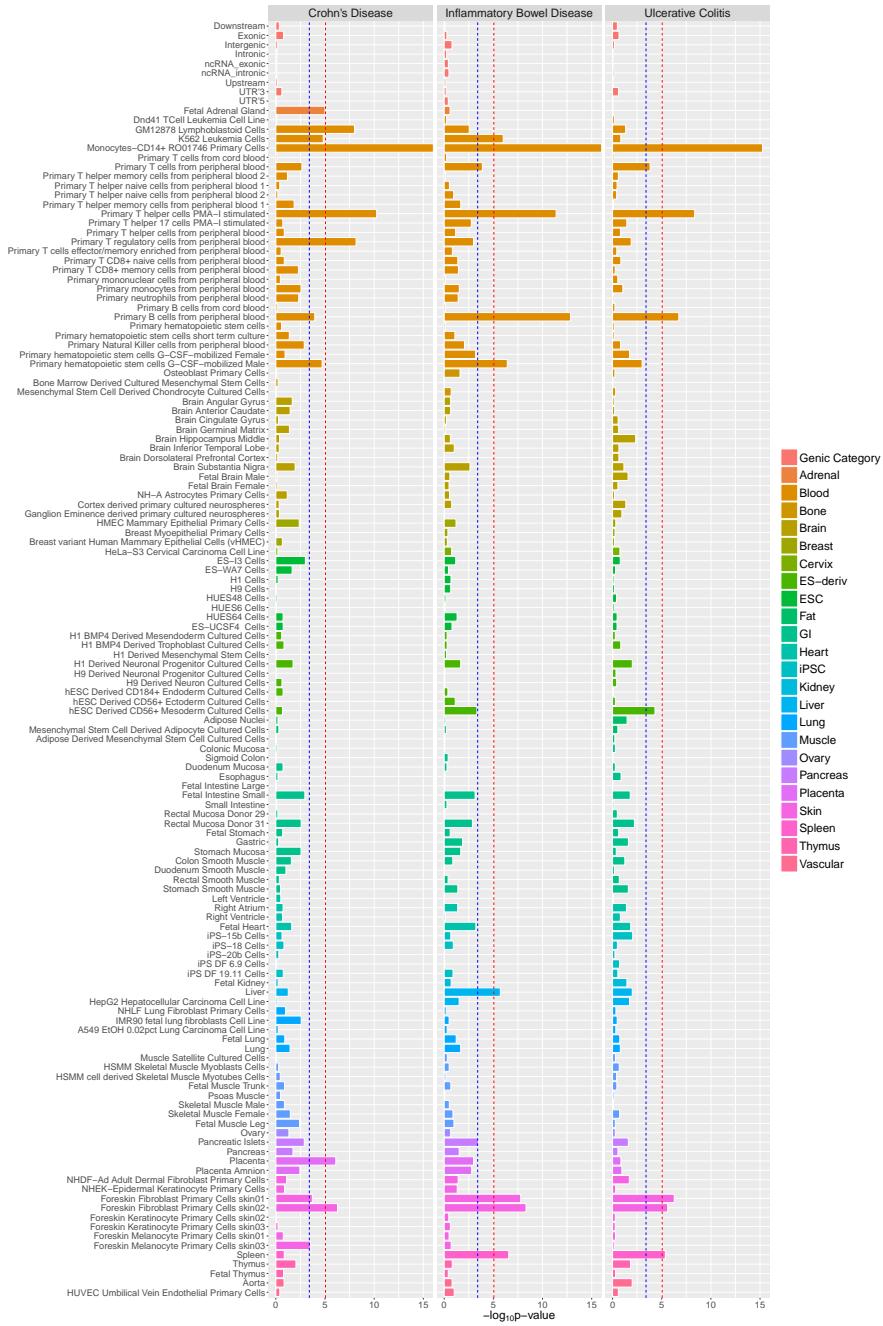


Figure S79: The $-\log_{10}(p\text{-value})$ of the enrichment test for nine genic category annotations and 127 cell-type specific functional annotations for UC, CD and IBD. The red lines represent $-\log_{10}\left(\frac{0.05}{(9+127)\times 44}\right)$. The blue lines represent $-\log_{10}\left(\frac{0.05}{9+127}\right)$.

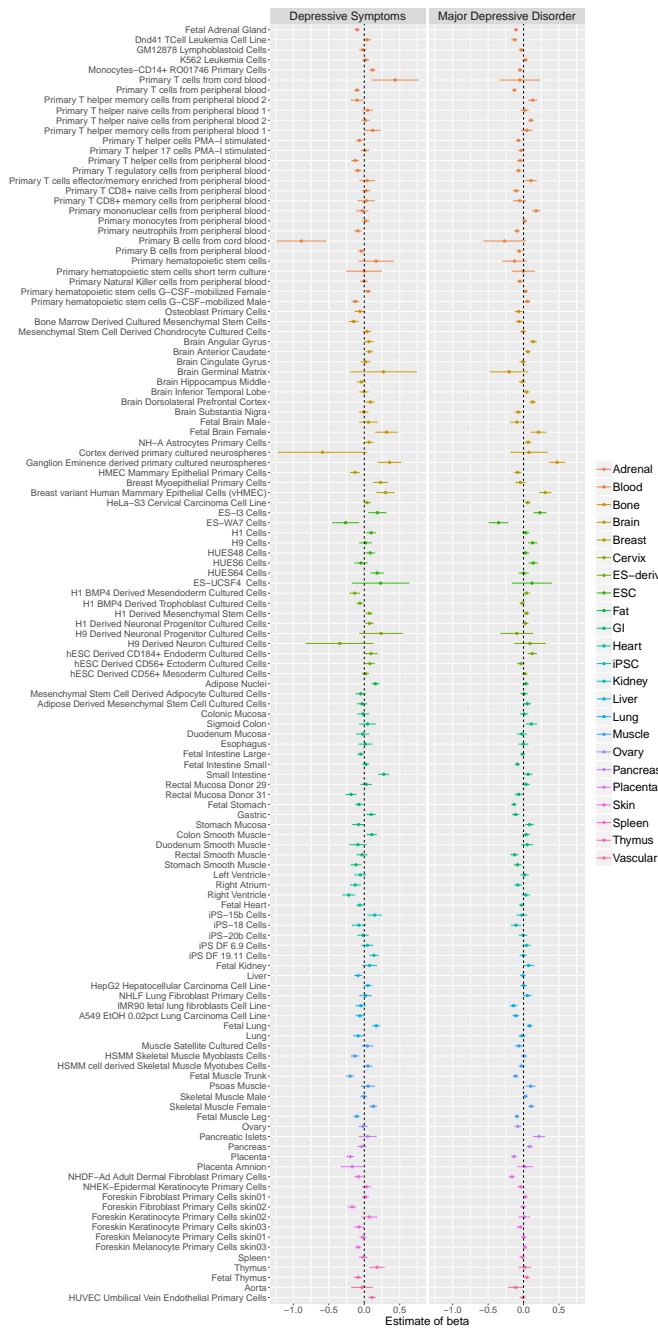


Figure S80: The estimated coefficients of 127 cell-type specific functional annotations for depressive symptoms and MDD. The bars represent one standard error.

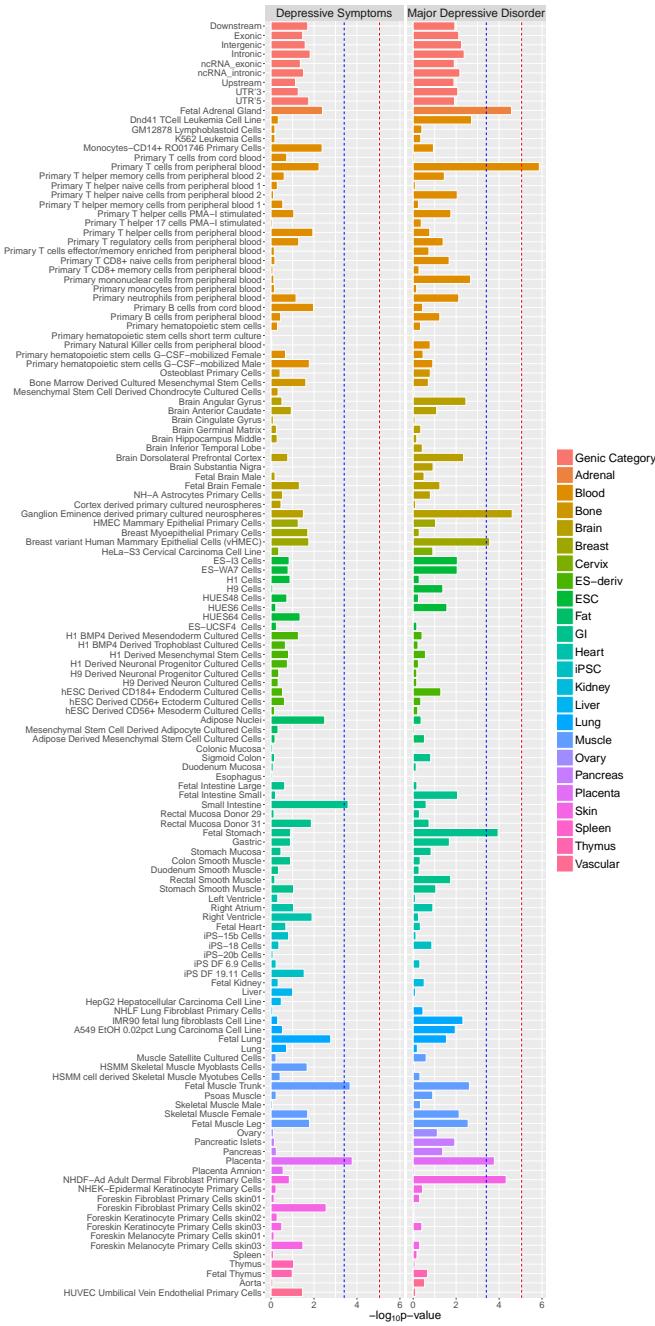


Figure S81: The $-\log_{10}(p\text{-value})$ of the enrichment test for nine genic category annotations and 127 cell-type specific functional annotations for depressive symptoms and MDD. The red lines represent $-\log_{10}\left(\frac{0.05}{(9+127)\times 44}\right)$. The blue lines represent $-\log_{10}\left(\frac{0.05}{9+127}\right)$.

8.13 Comparison of enrichment tests when testing all annotations or one annotation at a time

We chose two traits, depressive symptoms and major depressive disorder to demonstrate. The comparison results of enrichment test are shown in Figure S82. More annotations are detected to be relevant when we test one annotation at a time. If we test all annotations, only the one with the largest significance among the correlated annotations will be identified which is also conclusive.

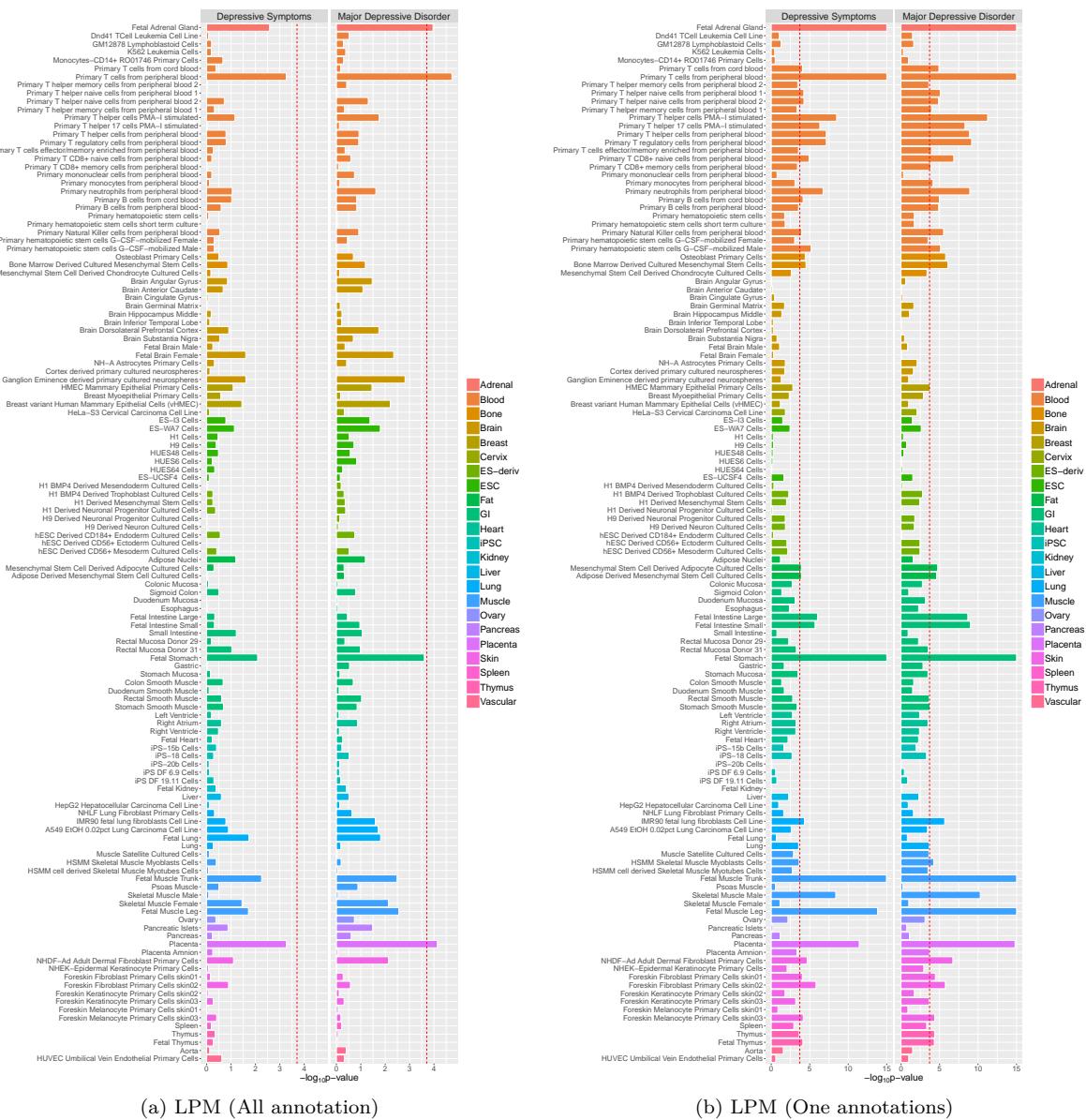


Figure S82: The $-\log_{10}(p\text{-value})$ of the enrichment test for 127 cell-type specific functional annotations for depressive symptoms and major depressive disorder when test (a) one annotation or (b) all annotations at a time. The red lines represent $-\log_{10}\left(\frac{0.05}{127 \times 2}\right)$.

9 Theorem based on composite likelihood approach

We use the same notation as in the main text. For the marginal model for phenotypes k and k' , let $\boldsymbol{\theta}_{kk'} = \{\tilde{\alpha}_{kk'}, \tilde{\beta}_{kk'}, \tilde{\mathbf{R}}_{kk'}\}$ be the collection of parameters, and $l_{kk'}(\boldsymbol{\theta}_{kk'}) = \sum_{j=1}^M l_{jkk'}(\boldsymbol{\theta}_{kk'})$ the corresponding log-likelihood.

Let $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_{kk'}, 1 \leq k < k' \leq K\}$. Consider the pairwise log-likelihood constructed from all possible pairs of phenotypes

$$cl(\boldsymbol{\Theta}) = \sum_{1 \leq k < k' \leq K} l_{kk'}(\boldsymbol{\theta}_{kk'}) = \sum_{1 \leq k < k' \leq K} \sum_{j=1}^M l_{jkk'}(\boldsymbol{\theta}_{kk'}).$$

This function is also known as a composite or pseudo likelihood (Varin et al., 2011). Maximizing it with respect to $\boldsymbol{\Theta}$ is equivalent to our proposed method of fitting all pairwise marginal models.

While the classical composite likelihood approach allows one or more parameters to be present in different pieces of the likelihood, we assume that each marginal likelihood has a separate set of parameters, that is, all elements of $\boldsymbol{\Theta}$ are distinct. Let $\boldsymbol{\Theta}^*$ be the set of all parameters in the joint model. One can check that α_k and β_k in $\boldsymbol{\Theta}^*$ have $K - 1$ counterparts in $\boldsymbol{\Theta}$, while $\rho_{kk'}$ has a single counterpart. Consequently, there exists a matrix \mathbf{A} such that $\boldsymbol{\Theta}^* = \mathbf{A}\boldsymbol{\Theta}$. As argued in the main text, the motivation of this separate parameterization is to ease the computational burden of joint maximization.

Let

$$\mathbf{s}_{jkk'}(\boldsymbol{\theta}_{kk'}) = \partial l_{jkk'}(\boldsymbol{\theta}_{kk'}) / \partial \boldsymbol{\theta}_{kk'}$$

and

$$\mathbf{H}_{jkk'}(\boldsymbol{\theta}_{kk'}) = \partial \mathbf{s}_{jkk'}(\boldsymbol{\theta}_{kk'}) / \partial \boldsymbol{\theta}_{kk'}.$$

Denote by $\hat{\boldsymbol{\Theta}}$ the maximizer of $cl(\boldsymbol{\Theta})$. We have the following theorem.

Theorem 1. *Let $\boldsymbol{\Omega} = -\sum_{j=1}^M E\{\mathbf{H}_{jkk'}(\boldsymbol{\theta}_{kk'})\}/M$ and $\boldsymbol{\Sigma} = \sum_{j=1}^M \text{cov}\{\mathbf{s}_{jkk'}(\boldsymbol{\theta}_{kk'})\}/M$. Then, under mild regularity conditions, $\sqrt{M} \text{vec}(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})$ converges in distribution to a multivariate normal distribution with mean vector zero and variance-covariance matrix $\boldsymbol{\Omega}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Omega}^{-1}$.*

The proof of this theorem, which we omit, follows directly from the general pseudo likelihood theory. See, for example, Geys et al. (1999) and Steffen and Geert (2006). To estimate the variance-covariance matrix, we can drop the expectations and covariances, and replace the unknown parameters by their estimates. This empirical estimator is known as the “sandwich” estimator. Let $\hat{\boldsymbol{\Theta}}^* = \mathbf{A}\hat{\boldsymbol{\Theta}}$.

Corollary 2. *Under mild regularity conditions, $\sqrt{M} \text{vec}(\hat{\boldsymbol{\Theta}}^* - \boldsymbol{\Theta}^*)$ converges in distribution to a multivariate normal distribution with mean vector zero and variance-covariance matrix $\mathbf{A}' \boldsymbol{\Omega}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Omega}^{-1} \mathbf{A}$.*

References

- Bentham, J., D. L. Morris, D. S. Cunningham Graham, C. L. Pinder, P. Tombleson, T. W. Behrens, J. Martín, B. P. Fairfax, J. C. Knight, L. Chen, et al. (2015, 12). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nature Genetics* 47(12), 1457–1464.
- Bradfield, J. P., H.-Q. Qu, K. Wang, H. Zhang, P. M. Sleiman, C. E. Kim, F. D. Mentch, H. Qiu, J. T. Glessner, K. A. Thomas, et al. (2011, 09). A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLOS Genetics* 7(9), 1–8.
- Chung, D., H. J. Kim, and H. Zhao (2017, 02). graph-GPA: A graphical model for prioritizing GWAS results and investigating pleiotropic architecture. *PLOS Computational Biology* 13(2), 1–20.
- Chung, D., C. Yang, C. Li, J. Gelernter, and H. Zhao (2014, 11). GPA: A statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. *PLOS Genetics* 10(11), 1–14.
- Cordell, H. J., Y. Han, G. F. Mells, Y. Li, G. M. Hirschfield, C. S. Greene, G. Xie, B. D. Juran, D. Zhu, D. C. Qian, et al. (2015). International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nature Communications* 6, 8019.
- Cousminer, D. L., D. J. Berry, N. J. Timpson, W. Ang, E. Thiering, E. M. Byrne, H. R. Taal, V. Huikari, J. P. Bradfield, M. Kerkhof, et al. (2013). Genome-wide association and longitudinal analyses reveal genetic loci linking pubertal height growth, pubertal timing and childhood adiposity. *Human Molecular Genetics* 22(13), 2735–2747.
- Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet* 381(9875), 1371–1379.
- Day, F. R., K. S. Ruth, D. J. Thompson, K. L. Lunetta, N. Pervjakova, D. I. Chasman, L. Stolk, H. K. Finucane, P. Sulem, B. Bulik-Sullivan, et al. (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nature Genetics* 47(11), 1294–1303.
- Demontis, D., R. K. Walters, J. Martin, M. Mattheisen, T. D. Als, E. Agerbo, G. Baldursson, R. Belliveau, J. Bybjerg-Grauholt, M. Bækvad-Hansen, et al. (2019, January). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics* 51(1), 63–75.
- Dubois, P. C. A., G. Trynka, L. Franke, K. A. Hunt, J. Romanos, A. Curtotti, A. Zhernakova, G. A. R. Heap, R. Ádány, A. Aromaa, et al. (2010, February). Multiple common variants for celiac disease influencing immune gene expression. *Nature Genetics* 42, 295.
- Duncan, L., Z. Yilmaz, H. Gaspar, R. Walters, J. Goldstein, V. Anttila, B. Bulik-Sullivan, S. Ripke, , L. Thornton, A. Hinney, et al. (2017). Significant locus and metabolic genetic correlations revealed

- in genome-wide association study of anorexia nervosa. *American Journal of Psychiatry* 174(9), 850–858. PMID: 28494655.
- Furberg, H. et al. (2010, April). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature Genetics* 42, 441.
- Geys, H., G. Molenberghs, and L. M. Ryan (1999). Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association* 94(447), 734–745.
- Grove, J., S. Ripke, T. D. Als, M. Mattheisen, R. K. Walters, H. Won, J. Pallesen, E. Agerbo, O. A. Andreassen, R. Anney, et al. (2019, March). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics* 51(3), 431–444.
- Jostins, L., S. Ripke, R. K. Weersma, R. H. Duerr, D. P. McGovern, K. Y. Hui, J. C. Lee, L. Philip Schumm, Y. Sharma, C. A. Anderson, et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491(7422), 119–124.
- Lambert, J.-C., C. A. Ibrahim-Verbaas, D. Harold, A. C. Naj, R. Sims, C. Bellenguez, G. Jun, A. L. DeStefano, J. C. Bis, G. W. Beecham, et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* 45(12), 1452–1458.
- Li, Y. and M. Kellis (2016, 07). Joint bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases. *Nucleic Acids Research* 44(18), e144–e144.
- Manning, A. K., M.-F. Hivert, R. A. Scott, J. L. Grimsby, N. Bouatia-Naji, H. Chen, D. Rybin, C.-T. Liu, L. F. Bielak, I. Prokopenko, et al. (2012, May). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genetics* 44, 659.
- McLaren, P. J., C. Coulonges, S. Ripke, L. van den Berg, S. Buchbinder, M. Carrington, A. Cos sarizza, J. Dalmau, S. G. Deeks, O. Delaneau, et al. (2013, 07). Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. *PLOS Pathogens* 9(7), 1–9.
- Morris, A. P. et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nature Genetics* 44(9), 981–990.
- Okada, Y., D. Wu, G. Trynka, T. Raj, C. Terao, K. Ikari, Y. Kochi, K. Ohmura, A. Suzuki, S. Yoshida, et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506(7488), 376–381.
- Okbay, A., B. M. L. Baselmans, J.-E. De Neve, P. Turley, M. G. Nivard, M. A. Fontana, S. F. W. Meddens, R. K. Linnér, C. A. Rietveld, J. Derringer, et al. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics* 48(6), 624–633.
- Okbay, A., J. P. Beauchamp, M. A. Fontana, J. J. Lee, T. H. Pers, C. A. Rietveld, P. Turley, G.-B. Chen, V. Emilsson, S. F. W. Meddens, et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 533(7604), 539–542.

- Paternoster, L. et al. (2015, October). Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. *Nature Genetics* 47, 1449.
- Pickrell, J. K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *The American Journal of Human Genetics* 94(4), 559–573.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 81(3), 559 – 575.
- Rietveld, C. A., S. E. Medland, J. Derringer, J. Yang, T. Esko, N. W. Martin, H.-J. Westra, K. Shakhsbazov, A. Abdellaoui, A. Agrawal, et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* 340(6139), 1467–1471.
- Ripke, S. et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511(7510), 421–427.
- Ripke, S., C. O'Dushlaine, K. Chambert, J. L. Moran, A. K. Kähler, S. Akterin, S. E. Bergen, A. L. Collins, J. J. Crowley, M. Fromer, et al. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics* 45(10), 1150–1159.
- Ripke, S. and otherd (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics* 43(10), 969–976.
- Sawcer, S. et al. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476(7359), 214–219.
- Schunkert, H., I. R. König, S. Kathiresan, M. P. Reilly, T. L. Assimes, H. Holm, M. Preuss, A. F. R. Stewart, M. Barbalic, C. Gieger, et al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nature Genetics* 43(4), 333–338.
- Sklar, P. et al. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nature Genetics* 43(10), 977–983.
- Speliotes, E. K., C. J. Willer, S. I. Berndt, K. L. Monda, G. Thorleifsson, A. U. Jackson, H. L. Allen, C. M. Lindgren, J. Luan, R. Mägi, et al. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics* 42(11), 937–948.
- Steffen, F. and V. Geert (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* 62(2), 424–431.
- The Wellcome Trust Case Control Consortium (2007, June). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447(7145), 661–678.
- Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica* 21(1), 5–42.
- Wang, K., M. Li, and H. Hakonarson (2010, 07). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* 38(16), e164–e164.

- Willer, C. J. et al. (2013). Discovery and refinement of loci associated with lipid levels. *Nature Genetics* 45(11), 1274–1283.
- Wood, A. R., T. Esko, J. Yang, S. Vedantam, T. H. Pers, S. Gustafsson, A. Y. Chu, K. Estrada, J. Luan, Z. Kutalik, et al. (2014). Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics* 46(11), 1173–1186.
- Wray, N. R., S. Ripke, M. Mattheisen, M. Trzaskowski, E. M. Byrne, A. Abdellaoui, M. J. Adams, E. Agerbo, T. M. Air, T. M. F. Andlauer, et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* 50(5), 668–681.
- Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher (2011). GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* 88(1), 76 – 82.