# The supplementary document of "IGESS: A Statistical Approach to Integrating Individual-Level Genotype Data and Summary Statistics in Genome-Wide Association Studies"

Mingwei Dai [1,2], Jingsi Ming [2], Mingxuan Cai [2], Jin Liu [3],
Can Yang [2,*], Xiang Wan [2,*], and Zongben Xu [1,*]
[1]School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China
[2]Department of Mathematics, Hong Kong Baptist University, Hong Kong
[3]Centre of Quantitative Medicine, Duke-NUS Medical School, Singapore
[4]Department of Computer Science, Hong Kong Baptist University, Hong Kong

## 1 Variational Expectation-Maximization Algorithm

### 1.1 E-Step

The joint probability in the main text could be rewritten as, let $\theta = \{\sigma_\beta^2, \sigma_e^2, \pi, \{\alpha_k\}_{k=1}^K\}$ be the collection of model parameters.

$$\begin{aligned}
&\Pr(\mathbf{y}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, \mathbf{P}|\mathbf{X}; \boldsymbol{\theta}) \\
=&\Pr(\mathbf{y}|\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, \mathbf{X}; \boldsymbol{\theta})\Pr(\tilde{\boldsymbol{\beta}}; \boldsymbol{\theta})\Pr(\boldsymbol{\gamma}|\boldsymbol{\theta})\Pr(\mathbf{P}|\boldsymbol{\gamma}; \boldsymbol{\theta}) \\
=&N(\mathbf{y}|\sum_j \mathbf{x}_j\tilde{\beta}_j\gamma_j, \sigma_e^2\mathrm{I})\prod_{j=1}^M N(\tilde{\beta}_j|0, \sigma_\beta^2)\pi^{\gamma_j}(1-\pi)^{1-\gamma_j}(\prod_{k=1}^K \alpha_k p_{jk}^{\alpha_k - 1})^{\gamma_j}
\end{aligned} \tag{S1}$$

The logarithm of the marginal likelihood is

$$\begin{aligned}
\log\Pr(\mathbf{y}, \mathbf{P}|\mathbf{X}; \boldsymbol{\theta}) &= \log\sum_{\boldsymbol{\gamma}}\int_{\tilde{\boldsymbol{\beta}}}\Pr(\mathbf{y}, \mathbf{P}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}|\mathbf{X}, \theta)d\tilde{\beta} \\
&\geq \sum_{\boldsymbol{\gamma}}\int_{\tilde{\boldsymbol{\beta}}} q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma})\log\frac{\Pr(\mathbf{y}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, \mathbf{P}|\mathbf{X}, \boldsymbol{\theta})}{q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma})}d\tilde{\beta} \\
&= \mathbb{E}_q[\log\Pr(\mathbf{y}, \mathbf{P}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}|\mathbf{X}; \theta) - \log q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma})] \\
&:= \mathcal{L}(q)
\end{aligned} \tag{S2}$$

where $L(q)$ is the lower bound implied by the Jensen's inequality and the equality holds if and only if $q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma})$ is the true posterior $\Pr(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}|\mathbf{y}, \mathbf{P}, \mathbf{X}; \boldsymbol{\theta})$. Instead of working with the marginal likelihood,

*To whom correspondence should be addressed.

we iteratively maximizing $\mathcal{L}(q)$. As it is stated in the main text, we employ the following variational distribution to make it feasible to evaluate the lower bound,

$$q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}) = \prod_{j=1}^{M} q_j(\tilde{\boldsymbol{\beta}}_j, \boldsymbol{\gamma}_j). \tag{S3}$$

According to the nice property of factorized distributions in variational inference, we can obtain the best approximation as

$$\log q_j(\tilde{\beta}_j, \gamma_j) = \mathbb{E}_{i \neq j}[\log \Pr(\mathbf{y}, \mathrm{P}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma} | \mathbf{X}; \boldsymbol{\theta})] + \text{Const}, \tag{S4}$$

where the expectation is taken with respect to all of the other factors $\{q_i(\tilde{\beta}_i, \gamma_i)\}$ for $i \neq j$

The logarithm of the joint probability function is

$$\begin{aligned}
\log \Pr(\mathbf{y}, \mathbf{P}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma} | \mathbf{X}; \boldsymbol{\theta}) = & -\frac{N}{2}\log(2\pi\sigma_e^2) - \frac{\mathbf{y}^T\mathbf{y}}{2\sigma_e^2} \\
& + \frac{\sum_{j=1}^{M} \gamma_j \tilde{\beta}_j \mathbf{x}_j^T \mathbf{y}}{\sigma_e^2} - \frac{1}{2\sigma_e^2}\sum_{j=1}^{M}\left((\gamma_j\tilde{\beta}_j)^2 \mathbf{x}_j^T\mathbf{x}_j\right) \\
& - \frac{1}{2\sigma_e^2}\left(\sum_{j=1}^{M}\sum_{j'\neq j}^{M} \gamma_j\tilde{\beta}_j\gamma_{j'}\tilde{\beta}_{j'}\mathbf{x}_k\mathbf{x}_k\right) \\
& - \frac{M}{2}\log(2\pi\sigma_\beta^2) - \frac{1}{2\sigma_\beta^2}\sum_{j=1}^{M}\tilde{\beta}_j^2 \\
& + \log\pi \sum_j \gamma_j + \log(1-\pi)\sum_j(1-\gamma_j) \\
& + \sum_j \gamma_j \sum_k \log(\alpha_k p_{jk}^{\alpha_k - 1})
\end{aligned} \tag{S5}$$

Before proceeding, we should keep several things in our mind. First, $q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma})$ is the variational approximation to the posterior $\Pr(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma} | \mathbf{y}, \mathbf{X}; \boldsymbol{\theta})$. Second, we assumed $q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}) = \prod_{j=1}^{M} q(\tilde{\beta}_j, \gamma_j)$. Third, $q(\tilde{\beta}_j, \gamma_j) = q(\tilde{\beta}_j | \gamma_j) q(\gamma_j)$.

To take the expectation in (S4), we rearrange (S5) into the terms with and without index

$$\begin{aligned}
\log \Pr(\mathbf{y}, \mathbf{P}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma} | \mathbf{X}; \boldsymbol{\theta}) = & -\frac{N}{2}\log(2\pi\sigma_e^2) - \frac{\mathbf{y}^T\mathbf{y}}{2\sigma_e^2} \\
& + \frac{\gamma_j\tilde{\beta}_j\mathbf{x}_j^T\mathbf{y}}{\sigma_e^2} + \frac{\sum_{k\neq j}\gamma_k\tilde{\beta}_k\mathbf{x}_j^T\mathbf{y}}{\sigma_e^2} \\
& - \frac{1}{2\sigma_e^2}\left((\gamma_j\tilde{\beta}_j)^2\mathbf{x}_j^T\mathbf{x}_j\right) - \frac{1}{2\sigma_e^2}\sum_{k\neq j}\left((\gamma_k\tilde{\beta}_k)^2\mathbf{x}_k^T\mathbf{x}_k\right) \\
& - \frac{1}{\sigma_e^2}\left(\sum_{k\neq j}^{M}\gamma_j\tilde{\beta}_j\gamma_k\tilde{\beta}_k\mathbf{x}_j^T\mathbf{x}_k\right) - \frac{1}{2\sigma_e^2}\left(\sum_{k\neq j}\sum_{k'\neq j}\gamma_k\tilde{\beta}_k\gamma_{k'}\tilde{\beta}_{k'}\mathbf{x}_k^T\mathbf{x}_{k'}\right) \\
& - \frac{M}{2}\log(2\pi\sigma_\beta^2) - \frac{1}{2\sigma_\beta^2}\tilde{\beta}_j^2 - \frac{1}{2\sigma_\beta^2}\sum_{k\neq j}\tilde{\beta}_k^2 \\
& + \log\pi\sum_j\gamma_j + \log(1-\pi)\sum_j(1-\gamma_j) \\
& + \sum_j\gamma_j\sum_k\log(\alpha_k p_{jk}^{\alpha_k - 1})
\end{aligned} \tag{S6}$$

Now we can take expectation of $\log \Pr(\mathbf{y}, \mathrm{P}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma} | \mathbf{X}; \boldsymbol{\theta})$ under the distribution $q(\tilde{\beta}_{-j}, \gamma_{-j})$. When $\gamma_j = 1$, we have

$$\log q(\tilde{\beta}_j|\gamma_j = 1) = \left(-\frac{1}{2\sigma_e^2}\mathbf{x}_j^T\mathbf{x}_j - \frac{1}{2\sigma_\beta^2}\right)\tilde{\beta}_j^2 + \frac{\mathbf{x}_j^T\mathbf{y} - \sum_{k\neq j}^{M}\mathbb{E}_k[\gamma_k\tilde{\beta}_k]\mathbf{x}_j^T\mathbf{x}_k}{\sigma_e^2}\tilde{\beta}_j + \text{Const} \tag{S7}$$

Because $\log q(\tilde{\beta}_j|\gamma_j = 1)$ is a quadratic form, we know $\beta_j$ with be Gaussian $N(\mu_j, s_j^2)$,we could easily get

$$\begin{aligned}
s_j^2 &= \frac{\sigma_e^2}{\mathbf{x}_j^T\mathbf{x}_j + \frac{\sigma_e^2}{\sigma_\beta^2}} \\
\mu_j &= \frac{\mathbf{x}_j^T\mathbf{y} - \sum_{k\neq j}\mathbb{E}_k[\gamma_k\beta_k]\mathbf{x}_j^T\mathbf{x}_k}{\mathbf{x}_j^T\mathbf{x}_j + \frac{\sigma_e^2}{\sigma_\beta^2}}
\end{aligned} \tag{S8}$$

Similarly, when $\gamma_j = 0$, we have

$$\log q(\tilde{\beta}_j|\gamma_j = 0) = -\frac{1}{2\sigma_\beta^2}\tilde{\beta}_j^2 + \text{Const} \tag{S9}$$

Thus we know $q(\tilde{\beta}_j|\gamma_j = 0) = N(\tilde{\beta}_j|0, \sigma_\beta^2)$. This is a very good property as it says that the posterior distribution of $\tilde{\beta}_j$ will be the same as its prior if this variable is irrelevant ( $\gamma_j = 0$ ). Note that $\gamma_j$ is a binary variable and then denote $\pi_j = q(\gamma_j = 1)$. Therefore we have

$$q(\tilde{\beta}_j, \gamma_j) = (\pi_j N(\mu_j, s_j^2))^{\gamma_j}((1-\pi_j)N(0, \sigma_\beta^2))^{1-\gamma_j} \tag{S10}$$

Now we evaluate the variational lower bound $L(q)$ (S2).

$$\mathcal{L}(q) = \mathbb{E}_q[\log \Pr(\mathbf{y}, \mathbf{P}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma} | \mathbf{X}; \boldsymbol{\theta})] - \mathbb{E}_q[\log q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma})] \tag{S11}$$

where

$$\begin{aligned}
\mathbb{E}_q[\log \Pr(\mathbf{y}, \mathbf{P}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma} | \mathbf{X}; \theta)] = & -\frac{N}{2}\log(2\pi\sigma_e^2) - \frac{\mathbf{y}^T\mathbf{y}}{2\sigma_e^2} \\
& + \frac{\sum_{j=1}^{M} \mathbb{E}[\gamma_j \tilde{\beta}_j] \mathbf{x}_j^T \mathbf{y}}{\sigma_e^2} \\
& - \frac{1}{2\sigma_e^2} \sum_{j=1}^{M} \left( \mathbb{E}[(\gamma_j \tilde{\beta}_j)^2] \mathbf{x}_j^T \mathbf{x}_j \right) \\
& - \frac{1}{2\sigma_e^2} \left( \sum_{j=1}^{M} \sum_{j' \neq j}^{M} \mathbb{E}[\gamma_j \tilde{\beta}_j] \mathbb{E}[\gamma_{j'} \tilde{\beta}_{j'}] \mathbf{x}_j^T \mathbf{x}_{j'} \right) \\
& - \frac{M}{2}\log(2\pi\sigma_\beta^2) - \frac{1}{2\sigma_\beta^2} \sum_{j=1}^{M} \mathbb{E}[\tilde{\beta}_j^2] \\
& + \log \pi \sum_j \mathbb{E}[\gamma_j] + \log(1-\pi) \sum_j \mathbb{E}[1-\gamma_j] \\
& + \sum_j \gamma_j \sum_k \log(\alpha_k p_{jk}^{\alpha_k - 1})
\end{aligned} \tag{S12}$$

and

$$\begin{aligned}
-\mathbb{E}_q[\log q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma})] = & \frac{M}{2}\log\sigma_\beta^2 + \sum_j \frac{1}{2}\pi_j(\log s_j^2 - \log\sigma_\beta^2) \\
& - \sum_j (\pi_j \log \pi_j + (1-\pi_j)\log(1-\pi_j))
\end{aligned} \tag{S13}$$

Now we substitute $\mathbb{E}[\gamma_j \tilde{\beta}_j] = \pi_j \mu_j$, $\mathbb{E}[(\gamma_j \tilde{\beta}_j)^2] = \pi_j(s_j^2 + \mu_j^2)$, $\mathbb{E}[\tilde{\beta}_j^2] = \pi_j(s_j^2 + \mu_j^2) + (1-\pi_j)\sigma_\beta^2$, $\mathbb{E}[\gamma_j] = \pi_j$ and $\mathbb{E}[1-\gamma_j] = 1 - \pi_j$

We rearrange the lower bound

$$\begin{aligned}\mathcal{L}(q) &= \mathbb{E}_q[\log \Pr(\mathbf{y}, \mathbf{P}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}|\mathbf{X}, \theta)] - \mathbb{E}_q[\log q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma})] \\ &= -\frac{N}{2}\log(2\pi\sigma_e^2) - \frac{\|\mathbf{y} - \sum_j \pi_j \mu_j \mathbf{x}_j\|^2}{2\sigma_e^2} - \frac{1}{2\sigma_e^2}\sum_{j=1}^{M}[\pi_j(s_j^2 + \mu_j^2) - (\pi_j\mu_j)^2]\mathbf{x}_j^T\mathbf{x}_j \\ &\quad - \frac{M}{2}\log(2\pi) - \frac{1}{2\sigma_\beta^2}\sum_{j=1}^{M}[\pi_j(\mu_j^2 + s_j^2) + (1-\pi_j)\sigma_\beta^2] \\ &\quad + \sum_j \pi_j \log(\frac{\pi}{\pi_j}) + \sum_j (1-\pi_j)\log(\frac{1-\pi}{1-\pi_j}) + \sum_{j=1}^{M}\pi_j \sum_k \log(\alpha_k p_{jk}^{\alpha_k - 1}) \\ &\quad + \sum_j \frac{1}{2}\pi_j(\log s_j^2 - \log\sigma_\beta^2)\end{aligned} \tag{S14}$$

To get $\pi_j$, we set $\frac{\partial \mathcal{L}(q)}{\partial \pi_j} = 0$,yielding

$$\pi_j = \frac{1}{1+\exp(-w_j)}, \text{where } w_j = \log\frac{\pi}{1-\pi} + \frac{1}{2}\log\frac{s_j^2}{\sigma_\beta^2} + \frac{\mu_j^2}{2s_j^2} + \sum_k \log(\alpha_k p_{jk}^{\alpha_k - 1}) \tag{S15}$$

## 1.2 M-Step

We will update the model parameters $\boldsymbol{\theta} = \{\sigma_\beta^2, \sigma_e^2, \pi, \{\alpha_k\}_{k=1}^K\}$ sequentially by maximizing the lower bound $\mathcal{L}(q)$.

To get $\sigma_\beta^2$, we set $\frac{\partial \mathcal{L}(q)}{\partial \sigma_\beta^2} = 0$

$$\frac{\partial \mathcal{L}(q)}{\partial \sigma_\beta^2} = \frac{\partial \mathbb{E}_q[\log \Pr(\mathbf{y}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, \mathbf{P}|\mathbf{X}, \boldsymbol{\theta})]}{\partial \sigma_\beta^2} = 0, \tag{S16}$$

yielding

$$\sigma_\beta^2 = \frac{\sum_j \pi_j(\mu_j^2 + s_j^2)}{\sum_j \pi_j} \tag{S17}$$

To get $\sigma_e^2$, we set $\frac{\partial \mathcal{L}(q)}{\partial \sigma_e^2} = 0$

$$\begin{aligned}\frac{\partial \mathcal{L}(q)}{\partial \sigma_e^2} = \frac{\partial \mathbb{E}_q[\log \Pr(\mathbf{y}, \tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}, \mathbf{P}|\mathbf{X}, \boldsymbol{\theta})]}{\partial \sigma_e^2} &= -\frac{N}{2}\frac{1}{\sigma_e^2} + \frac{\mathbf{y}^T\mathbf{y}}{2\sigma_e^4} \\ &\quad - \frac{\sum_j \mathbb{E}[\gamma_j\tilde{\beta}_j]\mathbf{x}_j^T\mathbf{y}}{\sigma_e^4} \\ &\quad + \frac{1}{2\sigma_e^4}\sum_{j=1}^{M}(\mathbb{E}[(\gamma_j\tilde{\beta}_j)^2]\mathbf{x}_j^T\mathbf{x}_j) \\ &\quad + \frac{1}{2\sigma_e^4}\left(\sum_{j=1}^{M}\sum_{j'\neq j}^{M}\mathbb{E}[\gamma_j\tilde{\beta}_j]\mathbb{E}[\gamma_{j'}\tilde{\beta}_{j'}]\mathbf{x}_j^T\mathbf{x}_{j'}\right) = 0,\end{aligned} \tag{S18}$$

yielding

$$\begin{aligned}\sigma_e^2 &= \frac{1}{N}\left(\mathbf{y}^T\mathbf{y} - 2\sum_j \mathbb{E}[\gamma_j\tilde{\beta}_j]\mathbf{x}_j^T\mathbf{y} + \sum_{j=1}^{M}(\mathbb{E}[(\gamma_j\tilde{\beta}_j)^2]\mathbf{x}_j^T\mathbf{x}_j) + \sum_{j=1}^{M}\sum_{j'\neq j}^{M}\mathbb{E}[\gamma_j\tilde{\beta}_j]\mathbb{E}[\gamma_{j'}\tilde{\beta}_{j'}]\mathbf{x}_j^T\mathbf{x}_{j'}\right) \\ &= \frac{1}{N}\left(\mathbf{y}^T\mathbf{y} - 2\sum_j \mathbb{E}[\gamma_j\tilde{\beta}_j]\mathbf{x}_j^T\mathbf{y} + \sum_{j=1}^{M}\sum_{j'=1}^{M}\mathbb{E}[\gamma_j\tilde{\beta}_j]\mathbb{E}[\gamma_{j'}\tilde{\beta}_{j'}]\mathbf{x}_j^T\mathbf{x}_{j'} + \sum_{j=1}^{M}(\mathbb{E}[(\gamma_j\tilde{\beta}_j)^2]\mathbf{x}_j^T\mathbf{x}_j) - \sum_{j=1}^{M}(\mathbb{E}[\gamma_j\tilde{\beta}_j])^2\mathbf{x}_j^T\mathbf{x}_j\right) \\ &= \frac{1}{N}\left(\|\mathbf{y} - \sum_j \pi_j\mu_j\mathbf{x}_j\|^2 + \sum_{j=1}^{M}(\pi_j(s_j^2+\mu_j^2) - (\pi_j\mu_j)^2)\mathbf{x}_j^T\mathbf{x}_j\right)\end{aligned} \tag{S19}$$

To get $\pi$, we set $\frac{\partial\mathcal{L}(q)}{\partial\pi} = 0$, yielding

$$\pi = \frac{1}{M}\sum_j \pi_j \tag{S20}$$

The parameter of the beta distribution, $\alpha_k$, could be obtained by maximizing $\mathcal{L}(q)$

$$\alpha_k = \frac{\sum_{j=1}^{M}\pi_j}{\sum_{j=1}^{M}\pi_j(-\log p_{jk})} \tag{S21}$$

### 1.3 Interpretation of variational EM algorithm

The above derivation assumes that the posterior $q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma})$ can be factorized as $q(\tilde{\boldsymbol{\beta}}, \boldsymbol{\gamma}) = \prod_{j=1}^{M} q_j(\tilde{\beta}_j, \gamma_j)$, which is known as "mean-field approximation". Without this approximation, we may use Monte-Carlo expectation-maximization (MCEM) algorithm which is a golden standard for statistical inference. It can be shown that, the M-step remains the same and the E-step can done via Gibbs sampling:

$$q_j(\tilde{\beta}_j|\gamma_j, \tilde{\boldsymbol{\beta}}_{-j}, \boldsymbol{\gamma}_{-j}; \boldsymbol{\theta}) = \begin{cases} \mathcal{N}(\tilde{\beta}_j|\mu_j, s_j^2) & \text{If } \gamma_j = 1, \\ \mathcal{N}(\tilde{\beta}_j|0, \sigma_\beta^2) & \text{If } \gamma_j = 0, \end{cases} \tag{S22}$$

where

$$\begin{aligned} s_j^2 &= \frac{\sigma_e^2}{\mathbf{x}_j^T\mathbf{x}_j + \frac{\sigma_e^2}{\sigma_\beta^2}}, \\ \mu_j &= \frac{\mathbf{x}_j^T\mathbf{y} - \sum_{i\neq j}[\gamma_i\tilde{\beta}_i]\mathbf{x}_j^T\mathbf{x}_i}{\mathbf{x}_j^T\mathbf{x}_j + \frac{\sigma_e^2}{\sigma_\beta^2}}, \end{aligned} \tag{S23}$$

and the probability for $\gamma_j = 1$ given $\{\tilde{\boldsymbol{\beta}}_{-j}, \boldsymbol{\gamma}_{-j}; \boldsymbol{\theta}\}$ is given as

$$\pi_j = \frac{1}{1+\exp(-u_j)}, \text{ where } u_j = \log\frac{\pi}{1-\pi} + \frac{1}{2}\log\frac{s_j^2}{\sigma_\beta^2} + \frac{\mu_j^2}{2s_j^2} + \sum_{k=1}^{K}\log(\alpha_k p_{jk}^{\alpha_k - 1}). \tag{S24}$$

As we can see, the only difference between our variational EM and MCEM is that the sample drawn from posterior distribution $[\gamma_i\tilde{\beta}_i]$ in equation (S23) is replaced by its expectation $\mathbb{E}[\gamma_i\tilde{\beta}_i]$. Such a replacement is known to have a consequence: variational inference may often underestimate the variance of the posterior density [1].

Despite the undesired property, this assumption brings a great computational advantage:

- It enables us to derive an efficient variational EM algorithm, with guaranteed convergence, for statistical inference, as we demonstrated in our paper.
- It also gives us an interpretable variational approximation to the true posterior. For example, the resulting approximated posterior of $\tilde{\beta}_j$ remains the same as its prior, i.e., $\tilde{\beta}_j \sim \mathcal{N}(\tilde{\beta}_j|0, \sigma_\beta^2)$ if SNP $j$ is irrelevant to the phenotype ($\gamma_j = 0$), which is expected and partially justifies our assumption.

# 2 Algorithms

## 2.1 Basic Algorithm Steps

Now we describe an algorithm:

- Initialize $\{\pi_j, \mu_j\}_{j=1}^M, \sigma_\beta^2, \sigma_e^2, \{\alpha_k\}_{k=1}^K$. Let $\tilde{\mathbf{y}} = \sum_j \pi_j\mu_j\mathbf{x}_j$.
- E − Step: For $j = 1, \ldots, M$, first obtain

$$\tilde{\mathbf{y}}_j = \tilde{\mathbf{y}} - \pi_j\mu_j\mathbf{x}_j, \tag{S25}$$

  and then update $\mu_j, s_j^2, \pi_j$ and $\tilde{\mathbf{y}}$ as follows

$$s_j^2 = \frac{\sigma_e^2}{\mathbf{x}_j^T\mathbf{x}_j + \sigma_e^2/\sigma_\beta^2}, \tag{S26}$$

$$\mu_j = \frac{\mathbf{x}_j^T(\mathbf{y} - \tilde{\mathbf{y}}_j)}{\mathbf{x}_j^T\mathbf{x}_j + \sigma_e^2/\sigma_\beta^2}, \tag{S27}$$

$$\pi_j = \frac{1}{1+\exp(-w_j)}, \text{where } w_j = \log\frac{\pi}{1-\pi} + \frac{1}{2}\log\frac{s_j^2}{\sigma_\beta^2} + \frac{\mu_j^2}{2s_j^2} + \sum_k \log(\alpha_k p_{jk}^{\alpha_k - 1}), \tag{S28}$$

$$\tilde{\mathbf{y}} = \tilde{\mathbf{y}}_j + \pi_j\mu_j\mathbf{x}_j. \tag{S29}$$

- M − Step

$$\begin{aligned}
\sigma_e^2 &= \left(\|\mathbf{y} - \tilde{\mathbf{y}}\|^2 + \sum_{j=1}^M(\pi_j(s_j^2 + \mu_j^2) - (\pi_j\mu_j)^2)\mathbf{x}_j^T\mathbf{x}_j\right)/N, \\
\sigma_\beta^2 &= \frac{\sum_j \pi_j(\mu_j^2 + s_j^2)}{\sum_j \pi_j}, \\
\pi &= \frac{1}{M}\sum_j \alpha_j.
\end{aligned} \tag{S30}$$

### 2.2 Efficiency Analysis and Improvements

It is noticed that the main burden of calculations lie in the E-Step. Three key calculations are $\tilde{\mathbf{y}}_j = \tilde{\mathbf{y}} - \pi_j\mu_j\mathbf{x}_j$, $\mathbf{x}_j^T\tilde{\mathbf{y}}_j$ in the $\mu_j$ update and $\tilde{\mathbf{y}} = \tilde{\mathbf{y}}_j + \pi_j\mu_j\mathbf{x}_j$, the time required is roughly $3NP$ for each iteration and it needs to create new space for the vector subtraction, which is also time consuming.

Substituting Eq. (S25) into Eq. (S27) leads to

$$\mu_j = \frac{\mathbf{x}_j^T(\mathbf{y} - (\tilde{\mathbf{y}} - \pi_j\mu_j\mathbf{x}_j))}{\mathbf{x}_j^T\mathbf{x}_j + \sigma_e^2/\sigma_\beta^2} = \frac{\mathbf{x}_j^T\mathbf{y} + \pi_j\mu_j\mathbf{x}_j^T\mathbf{x}_j - \mathbf{x}_j^T\tilde{\mathbf{y}}}{\mathbf{x}_j^T\mathbf{x}_j + \sigma_e^2/\sigma_\beta^2}, \tag{S31}$$

and substituting Eq. (S25) into Eq. (S29), denote $\pi_j\mu_j$ in Eq. (S25) as $\pi_j^0\mu_j^0$,

$$\tilde{\mathbf{y}} = \tilde{\mathbf{y}} + (\pi_j\mu_j - \pi_j^0\mu_j^0)\mathbf{x}_j, \tag{S32}$$

The time required after the above transformations decreased to roughly $2NP$ and avoid allocating $2NP$ new space for each iteration, the efficiency of algorithm is improved by nearly four times.

## 3 More results in the simulation experiments

In this section, we present more simulation results, which contain two main parts for quantitative trait studies and case-control studies. As described in the main text, we generated individual-level data $\{\mathbf{X}, \mathbf{y}\}$ and $\{\mathbf{X}^k, \mathbf{y}^{(k)}\}_{k=1,\ldots,K}$. For the individual-level data set, the genotype matrices $\{\mathbf{X}, \mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(K)}\}$ were first simulated from normal distribution, where autoregressive correlation $\rho^{|j-j'|}$ was set to mimic the linkage disequilibrium between variants $j$ and $j'$. Next, their entries are discretized to genotype codes $\{0, 1, 2\}$ according to the Hardy-Weinberg principle based on the minor allele frequencies drawn from $\mathcal{U}[0.05, 0.5]$. Based on $\mathbf{X}^{(k)}$ and $\mathbf{y}^{(k)}$, we were able to get $z$-values and $p$-values by applying univariate linear regression to $\{\mathbf{x}_j^{(k)}, \mathbf{y}^{(k)}\}$, where $\mathbf{x}_j^{(k)}$ corresponds to the genotypes of the $j$-th variant in the $k$-th study.

For IGESS, both the individual-level data $\{\mathbf{X}, \mathbf{y}\}$ and the $p$-value matrix $\mathbf{P}$ were used, where $\{\mathbf{X}^k, \mathbf{y}^k\}$ were pretended to be unavailable. For BVSR, only the individual-level data set $\{\mathbf{X}, \mathbf{y}\}$ was used and its performance could serve as a reference. For CPASSOC, an $M \times (1 + K)$ matrix $\mathbf{Z}$ comprised of the $z$-values from $\{\mathbf{X}, \mathbf{y}\}$ and the $z$-values from $K$ studies were used as its input. Only individual-level data set $\{\mathbf{X}, \mathbf{y}\}$ was used for BVSR and Lasso, whose performance served as reference.

If not explicitly specified, the number of samples and the number of variants were set to be $N = 2,000$ and $M = 10,000$ in the simulation. The heritability was pre-specified at $\{0.3, 0.5, 0.8\}$ with respect to 500 nonzero entries. The number of summary statistics data sets $K$ was set to be $1, 2$ and 6. The autoregressive correlation $\rho$ varies in $\{0.0, 0.3, 0.6\}$ .

### 3.1 Quantitative trait studies

#### 3.1.1 The performance of risk variant identification

Figure S1 shows the results for comparison of risk variant identification measured by AUC, with respect to the methods of IGESS, BVSR, CPASSOC, Lasso and p-values-based ranking. The FDR

of IGESS, BVSR and CPASSOC were evaluated with the nominal FDR controlled at 0.1 and the results are shown in Figure S2. The FDR of IGESS is well controlled in most cases except for the setting $\rho = 0.6$ and $h = 0.8$ (strong correlation and very high heritability). We also see that FDR inflation of CPASSOC is much severer than IGESS. In real data analysis, we are often interested in the performance of different methods with small false positive rate (FPR), e.g., with FPR $<$ 0.1. Here we show the ROC curves of IGESS, BVSR, CPASSOC in Figure S3 for comparison. All results are summarized from 50 replications.

In the main text, we have stated that CPASSOC outperforms IGESS in terms of AUC when heritability is very small (e.g., $h = 0.3$). A closer examination reveals that both methods have nearly zero power with the nominal FDR controlled at 0.1 (see the top right panel of Figure S4). This implies that the slightly better AUC of CPASSOC is due to ranking results of risk variants with a larger false positive rate which is not of interest in practice. Next, we increased the sample sizes $N_k$ of GWAS data $\mathbf{X}^{(k)}, \mathbf{y}^{(k)}$ to simulate summary statistics. Specifically, with heritability fixed at $h = 0.1$, we set $N_k$ to be $k * 2000$ ($k = 1, 2, \ldots, 5$) rather than an identical value of 2000. The lower panel of Figure S5 shows the performance comparison of IGESS and CPASSOC. Due to the larger sample size, the power of both methods is no longer near zero despite small heritability. In this case, IGESS also outperforms CPASSOC.

To have more comprehensive results, we set the sample sizes $\{N_k\}_{k=1,\ldots,5}$ of individual-level datasets $\{\mathbf{X}^{(k)}, \mathbf{y}^{(k)}\}_{k=1,\ldots,5}$ to be $\{1000, 2000, 3000, 4000, 5000\}$, respectively. We evaluated the performance in terms of AUC and Power between IGESS and CPASSOC and the results are summarized in Figure S6. The results show that the performance of IGESS is better than CPASSOC in terms of AUC and Power.

At last, we investigated the robustness of modeling $p$-values from the non-null group using the beta distribution $\mathcal{B}(\alpha, 1)$. To justify the usage of the beta distribution more comprehensively, we further conducted simulation studies as follows. In stead of obtaining $p$-values from individual-level data, we directly simulated $z$-values and then converted them to $p$-values. Here $z$-values from the null group follow a standard normal distribution and $z$-values from the non-null group follow an alternative distribution. We considered six density functions as given in Table S1. Clearly, the $p$-values converted from the $z$-values will not be a mixture of uniform and Beta distributions.

| Scenario | Alternative distribution for $z$-score |
|---|---|
| spiky | $0.4N(0, 0.25^2) + 0.2N(0, 0.5^2) + 0.2N(0, 1^2) + 0.2N(0, 2^2)$ |
| near normal | $2/3N(0, 1^2) + 1/3N(0, 2^2)$ |
| flattop | $1/7[N(-1.5, .5^2) + N(-1, .5^2) + N(-0.5, .5^2) +$ $N(0, .5^2) + N(0.5, .5^2) + N(1.0, .5^2) + N(1.5, .5^2)]$ |
| skew | $1/4N(-2, 2^2) + 1/4N(-1, 1.5^2) + 1/3N(0, 1^2) + 1/6N(1, 1^2)$ |
| big-normal | $N(0, 4^2)$ |
| bimodal | $0.5N(-2, 1^2) + 0.5N(2, 1^2)$ |

Table S1: Six density functions for $z$-values from the non-null group.

We evaluated the FDR control of IGESS and CPASSOC in this setting. The results shown in Figure S7 indicates that (a) both IGESS and CPASSOC are robust to different alternative distribution of $z$-values from the non-null group. (b) IGESS performs more stably than CPASSOC (small variance in the FDR control can be seen from the boxplots).

### 3.1.2 The risk prediction accuracy

Figure S8 shows the results for comparison of risk prediction measured by correlation between the observed phenotype values and the predicted values, with respect to the methods of IGESS, BVSR, Lasso. All results are summarized based on 50 replications.

### 3.1.3 Forward stepwise strategy vs. backward stepwise strategy

We use prediction accuracy (measured by AUC) by cross-validation as the criterion to select summary statistics from relevant study. In forward stepwise selection, IGESS tries to add one summary-statistic data at a time and picks the summary-statistic data set which maximizes prediction accuracy. If prediction accuracy gets worse when a summary-statistic data is incorporated, IGESS will automatically exclude this study in next steps. If we take a backward approach, we remove the study which leads to the largest increase of the prediction accuracy. According to the simulations, the performance of these two strategies are comparable and the forward-stepwise strategy is a little better (Figure S9). We recommend the forward stepwise strategy because it also has the computational advantage.

## 3.2 Case-Control Studies

### 3.2.1 The performance of risk variant identification

Figure S10 shows the results for comparison of risk variant identification measured by AUC, with respect to the methods of IGESS, BVSR, CPASSOC, Lasso and p-values-based ranking in case-control studies. Figure S11 summarizes the FDR of IGESS, BVSR and CPASSOC evaluated with the nominal FDR controlled at 0.1. All results are summarized based on 50 replications.

### 3.2.2 The prediction accuracy

Risk prediction accuracy comparison results of IGESS, BVSR, Lasso in Case-Control studies are shown in Figure S12. The classification accuracy is measured by AUC for the observed phenotype labels and the predicted values (using the independent test data). All results are summarized based on 50 replications.

### 3.2.3 The performance evaluation in presence of irrelevant studies

Figure S13 and Figure S14 show the performance evaluation in presence of irrelevant studies. They evaluate the performance with respect to risk variant identification measured by AUC, and classification accuracy measured by AUC for the observed phenotype labels and the predicted values (independent test data), respectively. $\{\mathbf{X}, \mathbf{y}\}$ in the $x$-axis indicates the performance of individual-level data only, $u$ corresponds to the performance of integrating the individual-level data with $p$-values from a study simulated using parameter $u_k = \Pr(\gamma_j^{(k)} = 1 | \gamma_j = 1) \in \{0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 1\}$ and $s$ indicates the stepwise performance achieved at the $s$-step.

# 4 More results in real data analysis

## 4.1 Analysis of Crohn's disease

Kang et al (2010) [3] argued that the application of a linear model in case-control studies for risk variant identification could be feasible by connecting it to the Armitage trend test. Recently, Chen et al. (2016) [2] showed that type I error control of linear model in case-control studies might fail in presence of population structure, where they used the real data of Asthma GWAS in Hispanic/Latino population as an example. Furthermore, they proposed generalized linear mixed model association test (GMMAT) to address this issue.

In the main article, we claimed that linear models (Gaussian noise case) often provide satisfactory results if the population structure is not very complex, e.g., European population considered in [3]. To provide evidence of our claim, we used GMMAT and GEMMA (linear mixed model association test) to analyze Crohn's diseases and the results are shown in Figure S15. We did not see clear difference of the results between GEMMA and GMMAT.

We also compared logistic models with linear models using the GWAS data of Crohn's disease. Specifically, we considered $L_1$-regularized logistic regression, $L_1$-regularized linear regression, variational Bayesian logistic regression and variational Bayesian linear regression. Their accuracies measured by AUC are all evaluated via cross-validation. The results for ten repetitions are summarized in Figure S16, they indicates that there is no clear prediction advantage of logistic regression models over linear models in the Crohn's disease analysis.

## 4.2 Analysis of Rheumatoid Arthritis

For the second real data example, we considered the WTCCC data of Rheumatoid Arthritis (RA). We applied the same quality control strategy to preprocess the RA data set (with details provided in the main text) and finally we had 4,944 individuals with 304,011 SNPs. Beside the WTCCC individual-level data, the summary statistics of RA from five GWAS in [4] are publicly available from `http://www.broadinstitute.org/ftp/pub/rheumatoid_arthritis/Stahl_etal_2010NG`: BRASS (483 cases, 1,449 controls), Canada (589 cases, 1,472 controls), EIRA (1,173 cases, 1,089 controls), NARAC1 (867 cases, 1,041 controls), NARAC2 (902 cases, 4,510 controls).

We applied IGESS to integrate individual-level RA data and summary-level RA data. The results are given in Figure S17. Again, we can see that the prediction accuracy has been improved 1% with the help of summary-level data. Here the best prediction is obtained when only one summary-level data is incorporated in IGESS.

# References


[1] David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *arXiv preprint arXiv:1601.00670*, 2016.

[2] Han Chen, Chaolong Wang, Matthew P Conomos, Adrienne M Stilp, Zilin Li, Tamar Sofer, Adam A Szpiro, Wei Chen, John M Brehm, Juan C Celedón, et al. Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *The American Journal of Human Genetics*, 98(4):653–666, 2016.

[3] Hyun Min Kang, Jae Hoon Sul, Susan K Service, Noah A Zaitlen, Sit-yee Kong, Nelson B Freimer, Chiara Sabatti, Eleazar Eskin, et al. Variance component model to account for sample structure in genome-wide association studies. *Nature genetics*, 42(4):348–354, 2010.

[4] Eli A Stahl, Soumya Raychaudhuri, Elaine F Remmers, Gang Xie, Stephen Eyre, Brian P Thomson, Yonghong Li, Fina AS Kurreeman, Alexandra Zhernakova, Anne Hinks, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature genetics*, 42(6):508–514, 2010.

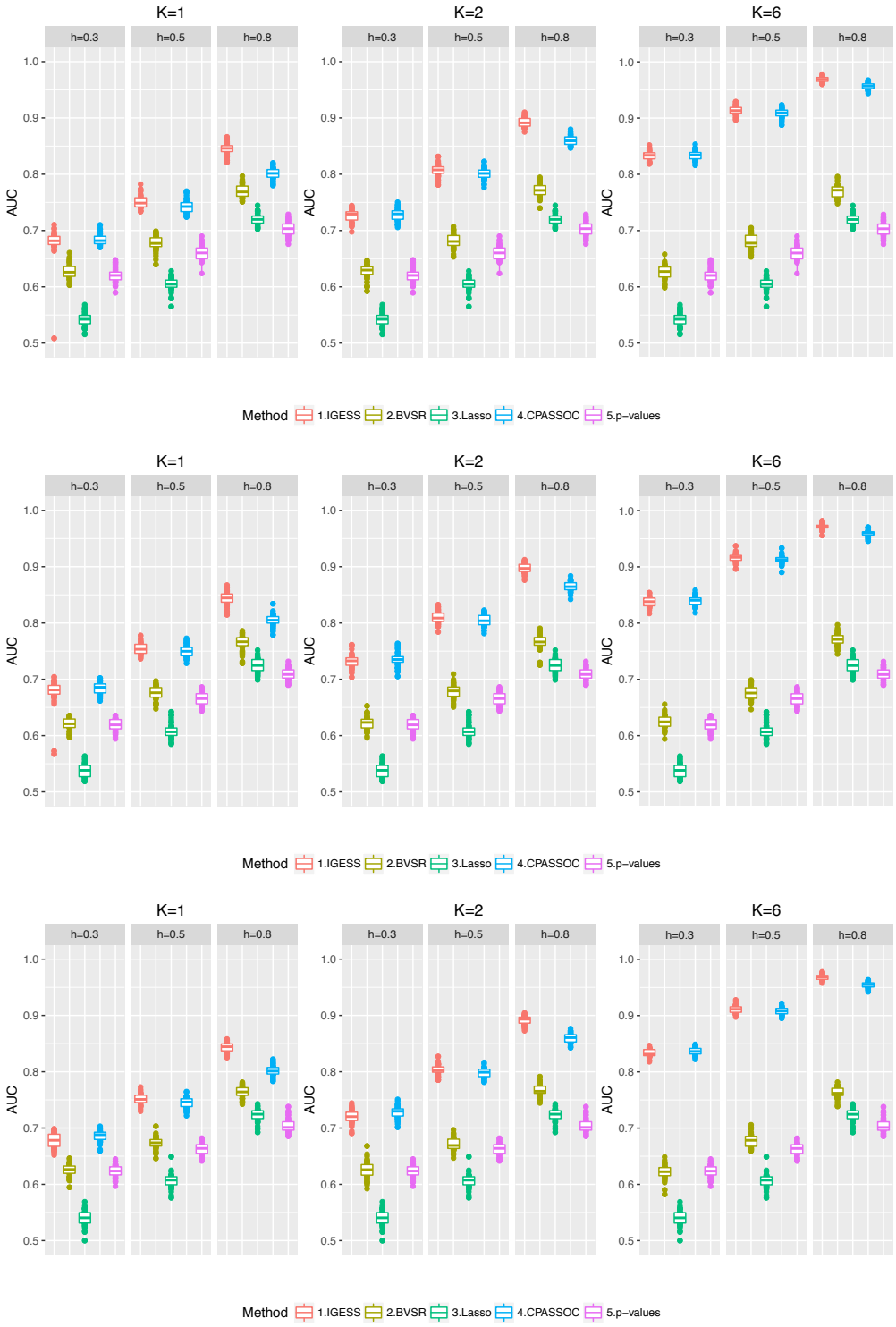


Figure S1: Performance of risk variant identification measured by AUC. Upper panel: autoregressive correlation $\rho = 0$. Middel panel: autoregressive correlation $\rho = 0.3$. Lower panel: autoregressive correlation $\rho = 0.6$.

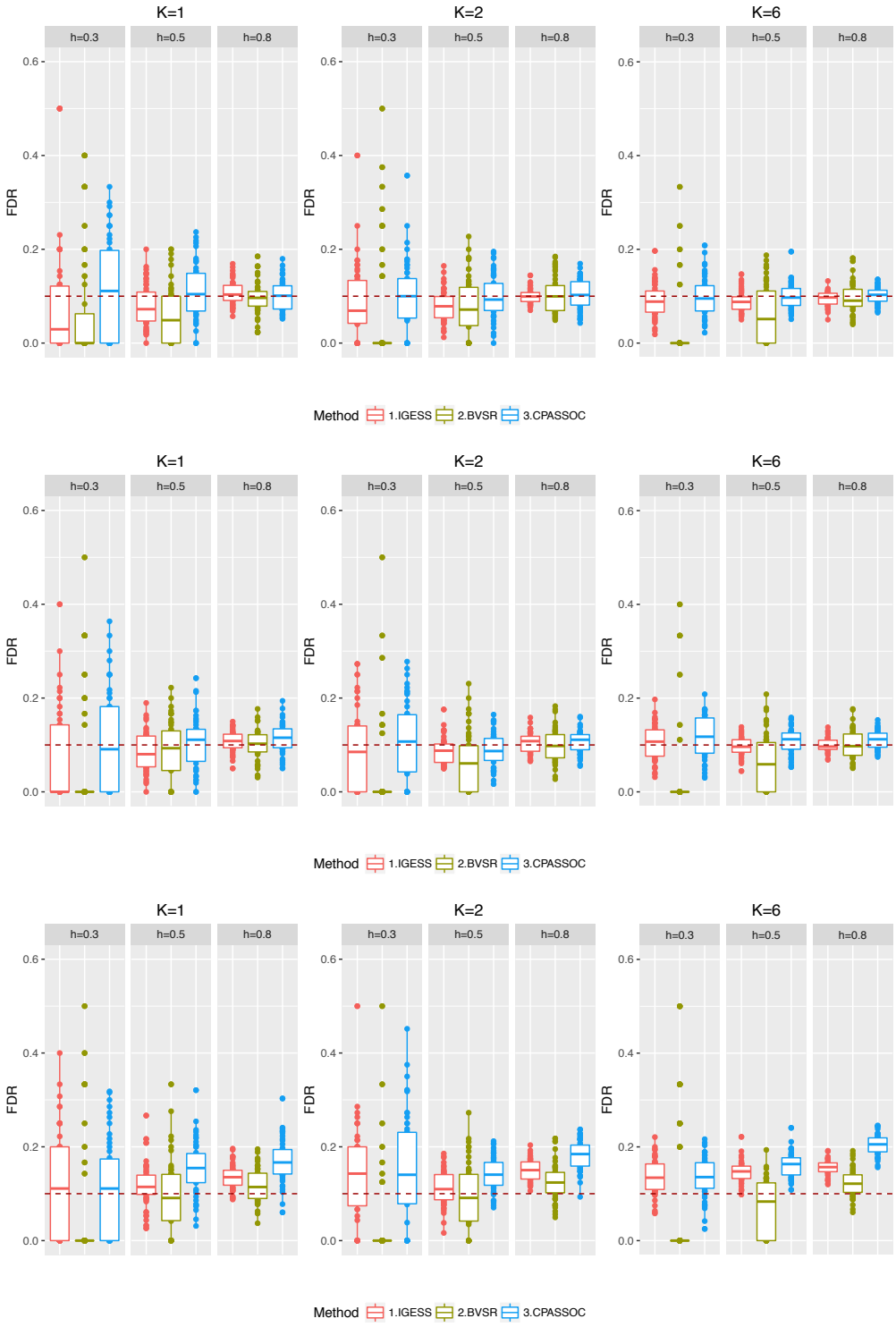


Figure S2: The FDR of IGESS, BVSR and CPASSOC are evaluated with the nominal FDR controlled at 0.1. Upper panel: autoregressive correlation $\rho = 0$. Middel panel: autoregressive correlation $\rho = 0.3$. Lower panel: autoregressive correlation $\rho = 0.6$.

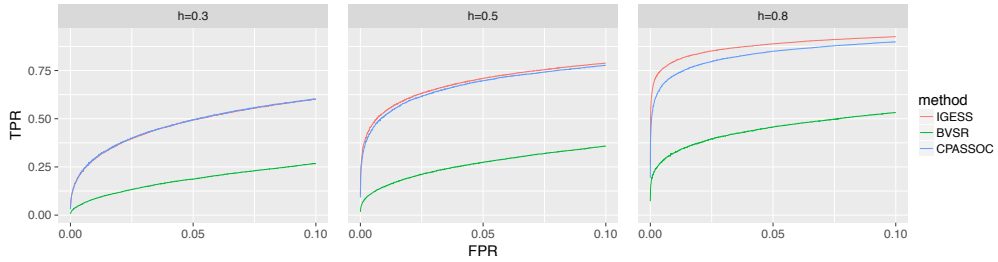


Figure S3: The ROC Curves of IGESS, BVSR and CPASSOC with FPR $< 0.1$.

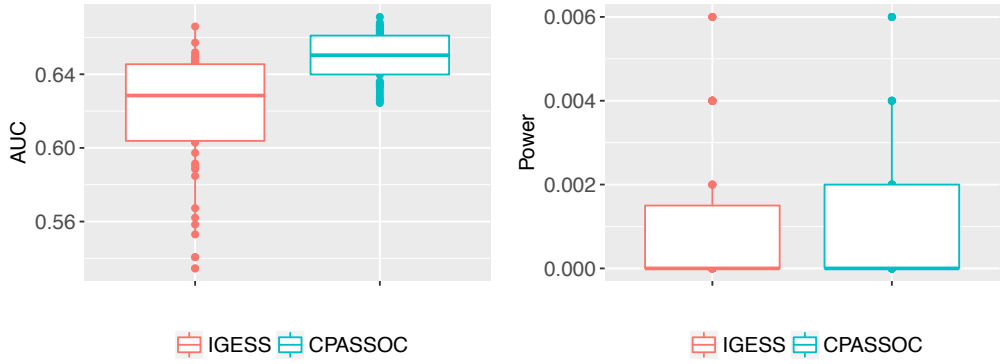


Figure S4: Performance comparison with $h = 0.1$ and $N_k = 2000$ for $k = 1, 2, \ldots, 5$. Left panel: Performance of risk variant identification measured by AUC. Right Panel: Performance of risk variant identification measured by power.

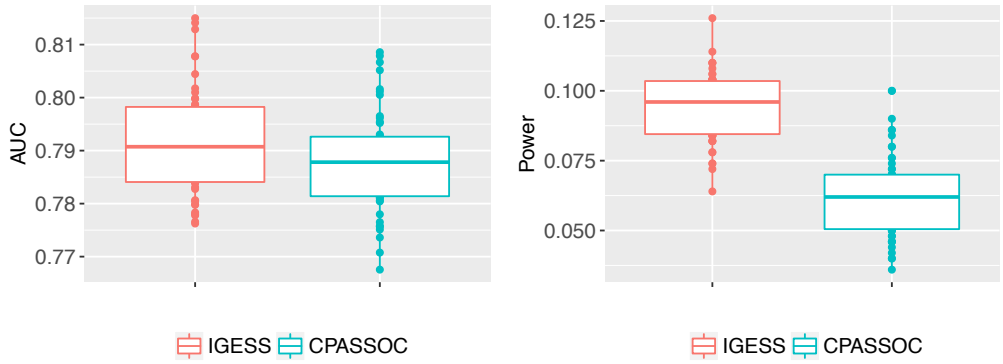


Figure S5: Performance comparison when $h = 0.1$ and $N = \{2000, 4000, 6000, 8000, 10000\}$ for GWAS. Left panel: Performance of risk variant identification measured by AUC. Right Panel: Performance of risk variant identification measured by power.

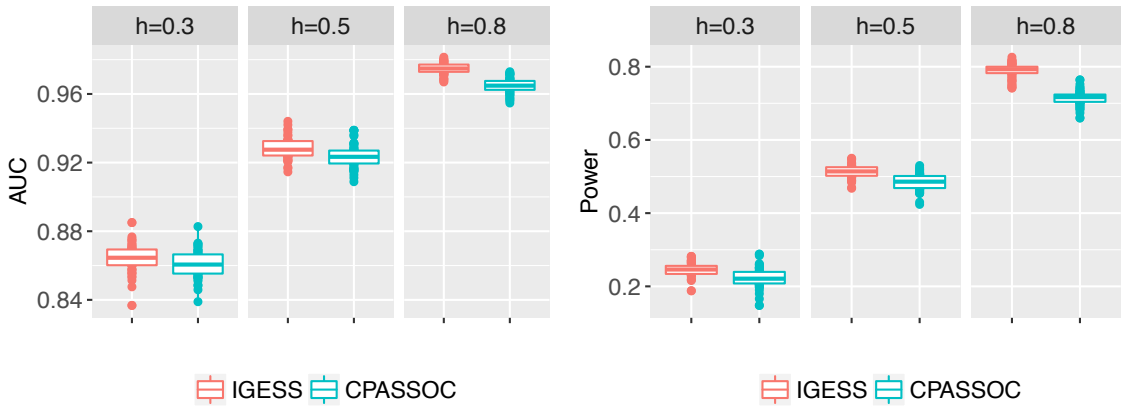


Figure S6: Left panel: Performance of risk variant identification measured by AUC. Right Panel: Performance of risk variant identification measured by Power.

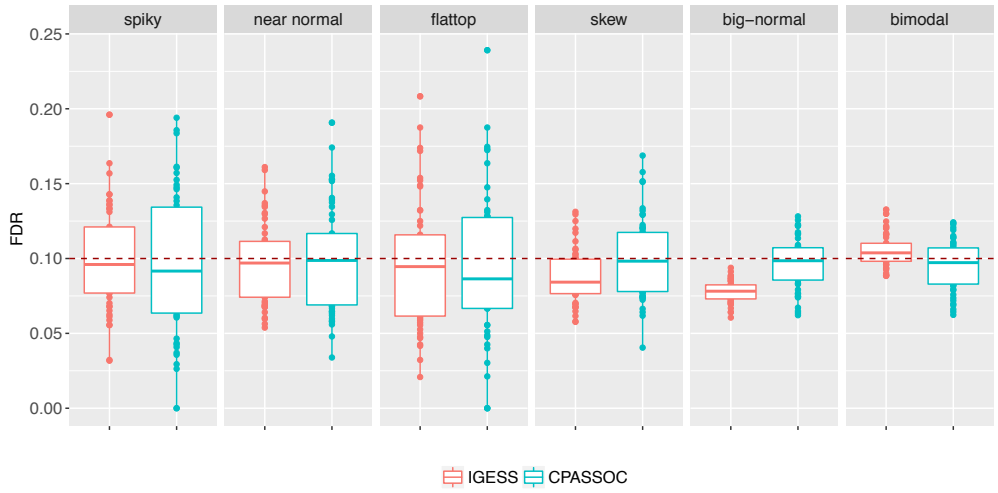


Figure S7: The FDR control evaluated for different different alternative distribution of $z$-values from the non-null group. Here the nominal FDR was controlled at 0.1.

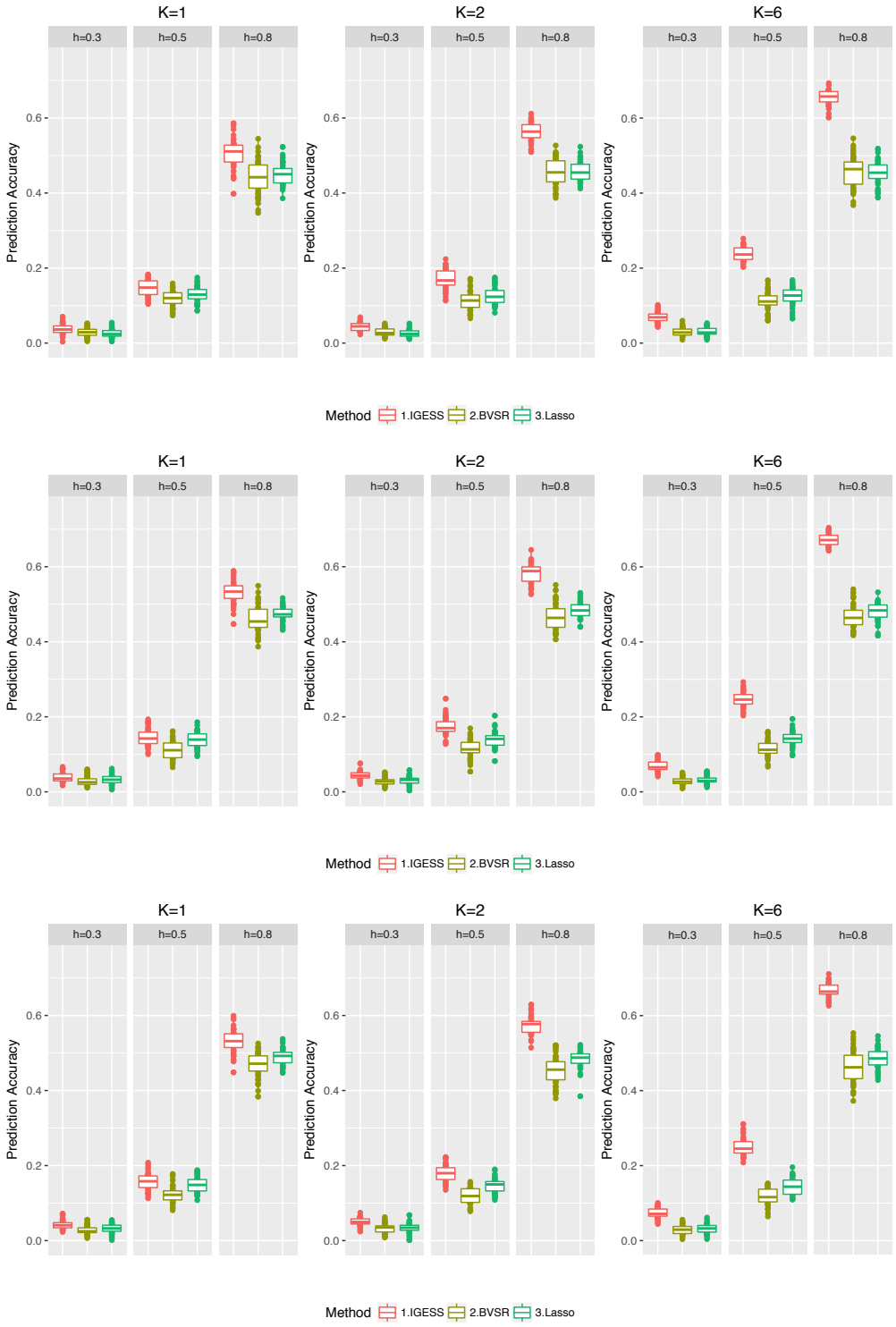


Figure S8: Risk prediction comparison in the quantitative trait studies. Upper panel: autoregressive correlation $\rho = 0$. Middel panel: autoregressive correlation $\rho = 0.3$. Lower panel: autoregressive correlation $\rho = 0.6$.

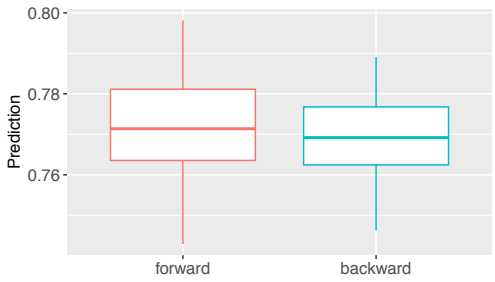


Figure S9: Comparison of forward-stepwise strategy and backward-stepwise strategy with respect to prediction accuracy measured by AUC.


Figure S10: Performance of risk variant identification measured by AUC in case-control studies. Upper panel: autoregressive correlation $\rho = 0$. Middel panel: autoregressive correlation $\rho = 0.3$. Lower panel: autoregressive correlation $\rho = 0.6$.

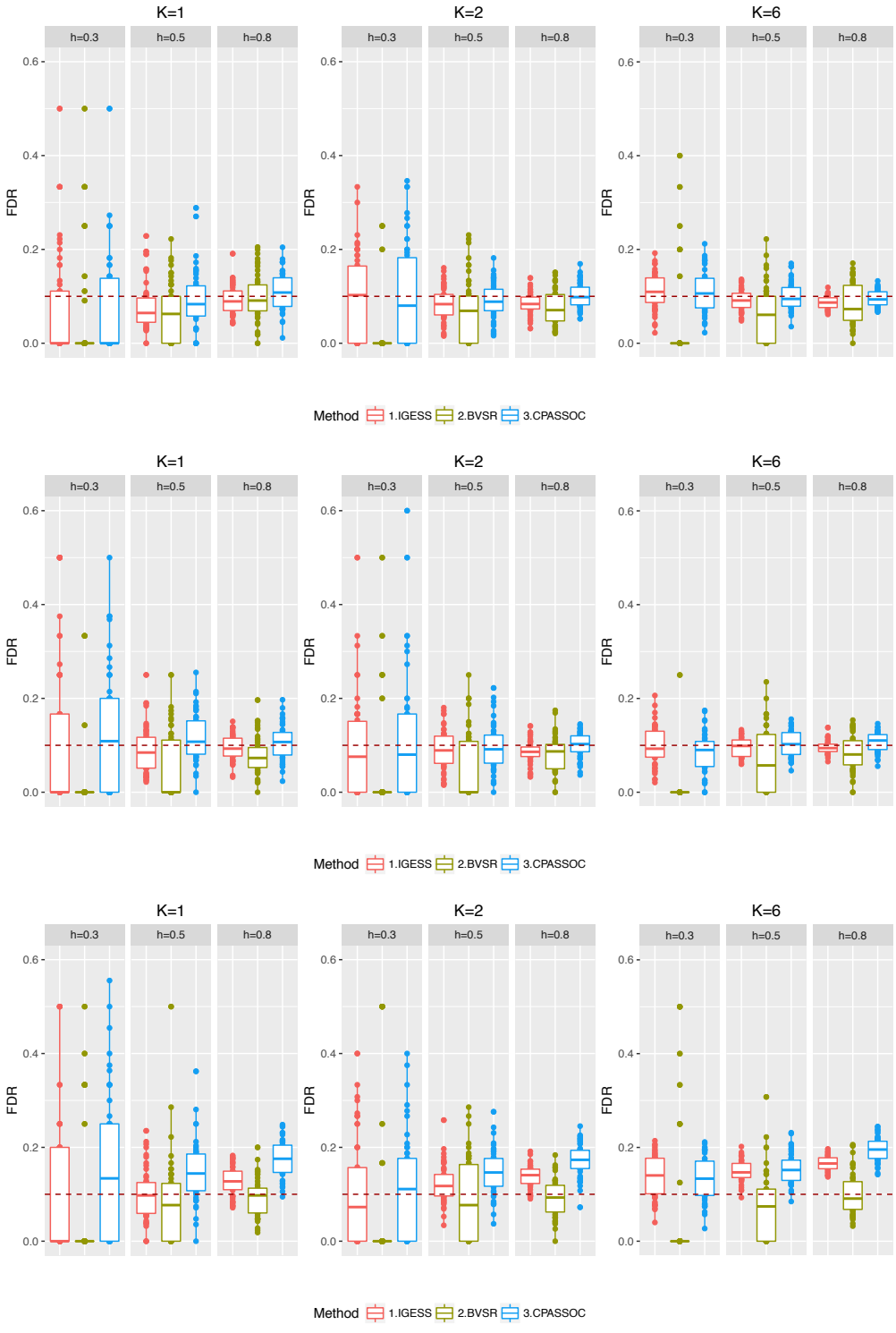


Figure S11: The FDR of IGESS, BVSR and CPASSOC are evaluated with the nominal FDR controlled at 0.1. Upper panel: autoregressive correlation $\rho = 0$. Middel panel: autoregressive correlation $\rho = 0.3$. Lower panel: autoregressive correlation $\rho = 0.6$.

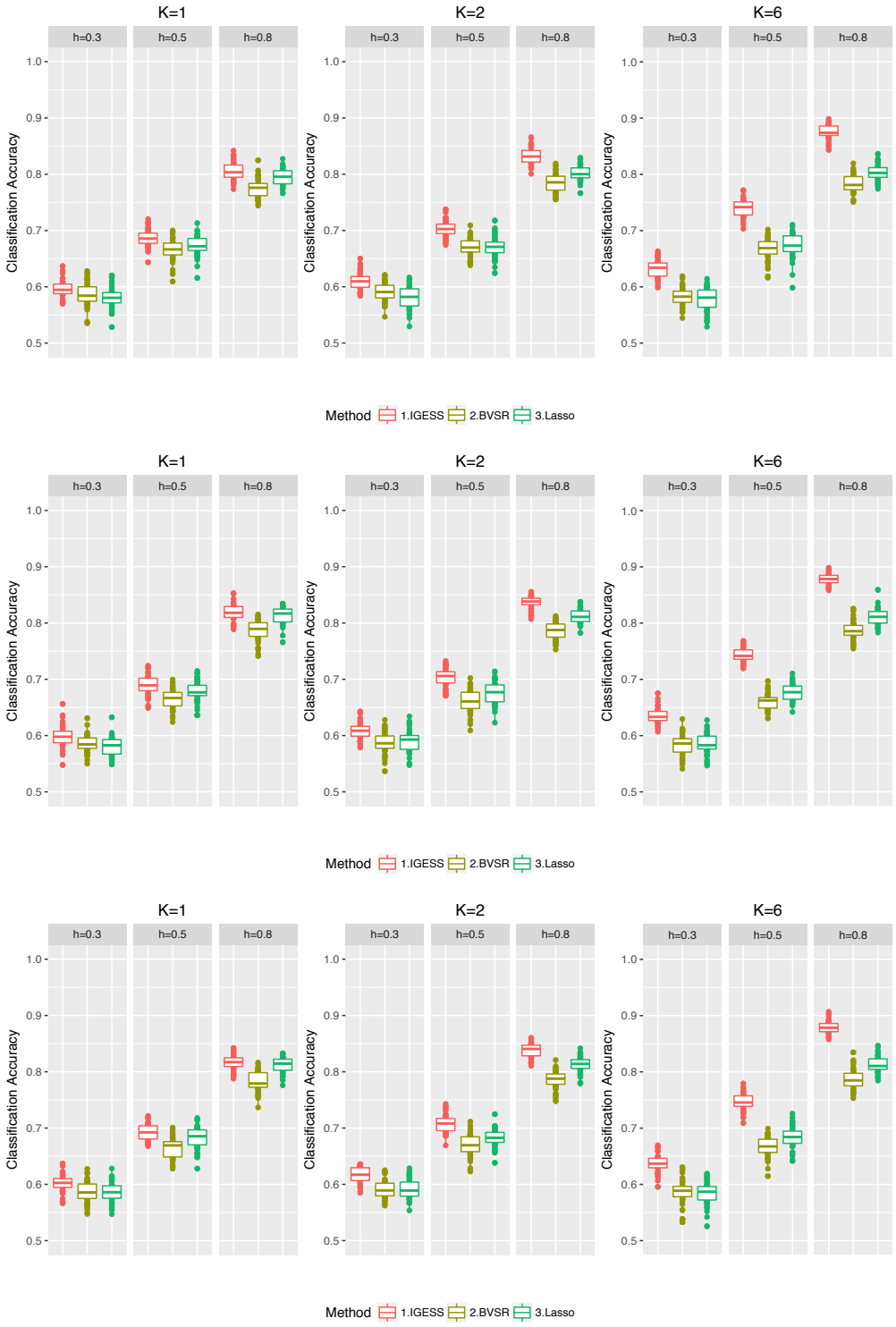


Figure S12: Risk prediction comparison in the case-control studies. Upper panel: autoregressive correlation $\rho = 0$. Middel panel: autoregressive correlation $\rho = 0.3$. Lower panel: autoregressive correlation $\rho = 0.6$.

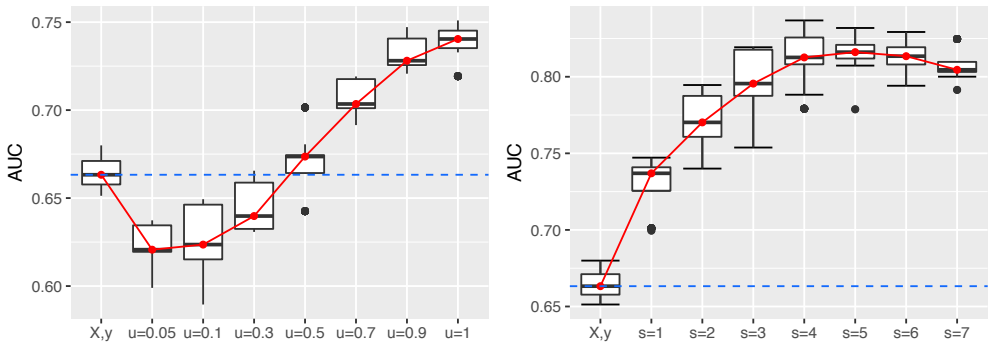


Figure S13: Performance of risk variant identification measured by AUC. Left Panel: Performance of risk variant identification at the first forward step. Right panel: Performance of risk variant identification in the entire forward selection process.

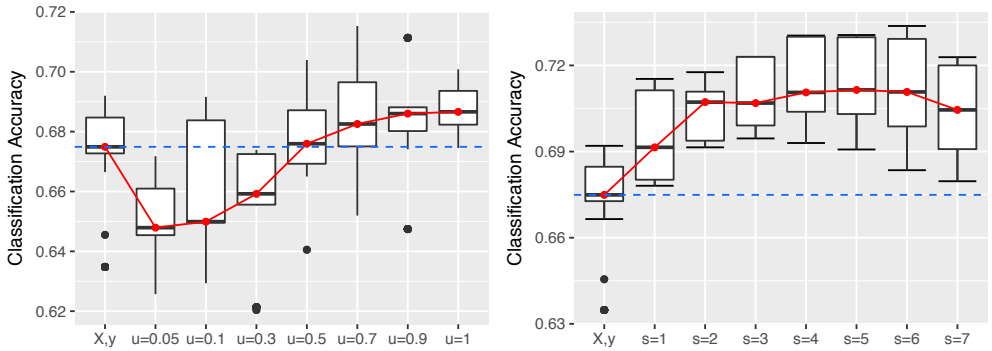


Figure S14: Performance of classification accuracy measured by AUC calculated with the observed phenotype labels and the predicted values (with the independent test data). Left Panel: Performance of classification accuracy at the first forward step. Right panel: Performance of classification accuracy in the entire forward selection process.

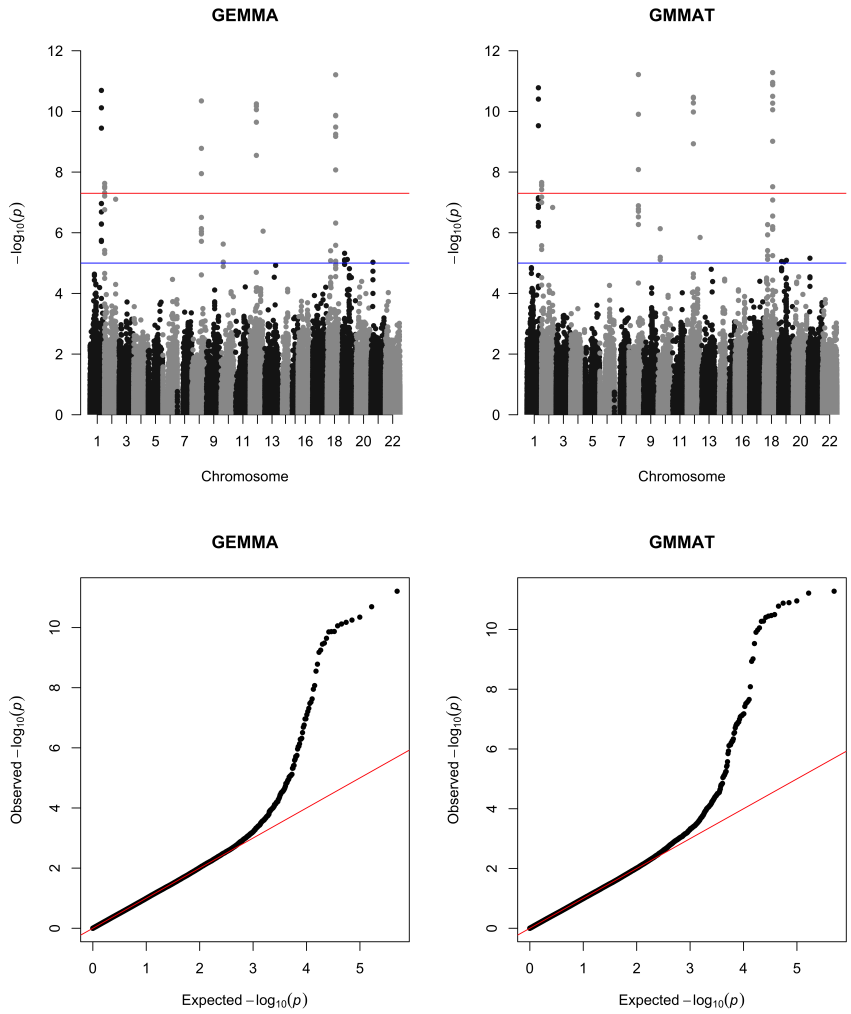


Figure S15: Manhattan plots of Crohn's Diseases and qq-plots based on the analysis results from GEMMA with genomic control factor $\lambda = 1.00106$ (left panel) and GMMAT with genomic control factor $\lambda = 0.9998477$ (right panel).

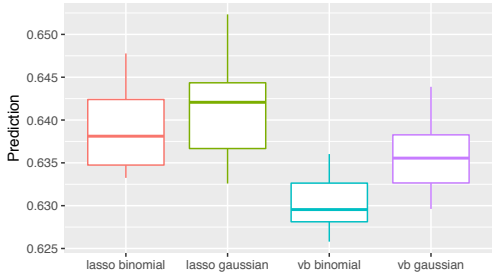


Figure S16: Comparison between logistic models and linear models on the GWAS data of Crohn's disease.

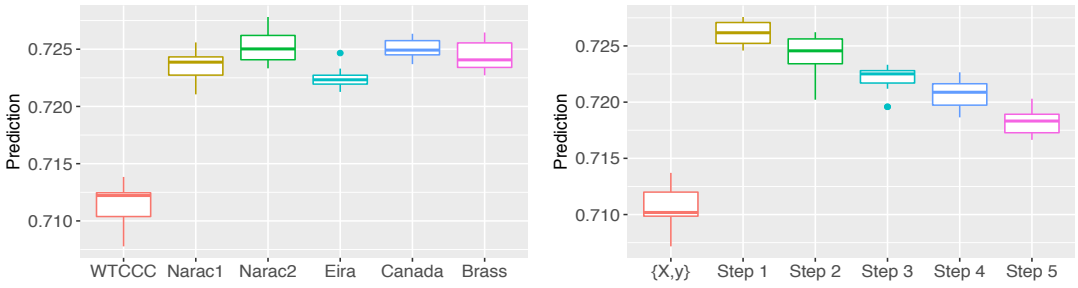


Figure S17: Left Panel: Prediction accuracy measured by AUC by single GWAS with summary-level data. Right Panel: Prediction accuracy measured by AUC during forward stepwise selection.

Table S2:Identified variants using IGESS(WTCCC + Early Onset)

| SNP id | Chr | Pos | Type | Gene |
|---|---|---|---|---|
| rs707472 | chr1 | 7906008 | downstream | PER3 |
| rs7528804 | chr1 | 67593819 | intronic | C1orf141 |
| rs10889656 | chr1 | 67594559 | intronic | C1orf141 |
| rs3762318 | chr1 | 67597119 | intronic | C1orf141 |
| rs4655679 | chr1 | 67599657 | intronic | C1orf141 |
| rs10789224 | chr1 | 67605134 | intergenic | C1orf141(dist=4480),IL23R(dist=27035) |
| rs17375018 | chr1 | 67655147 | intronic | IL23R |
| rs4655689 | chr1 | 67659421 | intronic | IL23R |
| rs11805303 | chr1 | 67675516 | intronic | IL23R |
| rs2201841 | chr1 | 67694202 | intronic | IL23R |
| rs6660226 | chr1 | 67744601 | intergenic | IL23R(dist=18951),IL12RB2(dist=28446) |
| rs12141431 | chr1 | 67747023 | intergenic | IL23R(dist=21373),IL12RB2(dist=26024) |
| rs12119179 | chr1 | 67747415 | intergenic | IL23R(dist=21765),IL12RB2(dist=25632) |
| rs11209039 | chr1 | 67751193 | intergenic | IL23R(dist=25543),IL12RB2(dist=21854) |
| rs6679677 | chr1 | 114303808 | downstream | RSBN1 |
| rs3768566 | chr1 | 155201064 | intergenic | GBAP1(dist=3739),GBA(dist=3175) |
| rs7520184 | chr1 | 155253583 | intronic | HCN3 |
| rs10995238 | chr10 | 64387108 | intronic | ZNF365 |
| rs10761659 | chr10 | 64445564 | intergenic | ZNF365(dist=13793),ADO(dist=118952) |
| rs7095491 | chr10 | 101274058 | intergenic | GOT1(dist=83528),LINC01475(dist=12049) |
| rs7078219 | chr10 | 101274365 | intergenic | GOT1(dist=83835),LINC01475(dist=11742) |
| rs7081330 | chr10 | 101274465 | intergenic | GOT1(dist=83935),LINC01475(dist=11642) |
| rs10883365 | chr10 | 101287764 | ncRNA_exonic | LINC01475 |
| rs10883367 | chr10 | 101287990 | ncRNA_intronic | LINC01475 |
| rs1548962 | chr10 | 101289735 | ncRNA_intronic | LINC01475 |
| rs6584283 | chr10 | 101290301 | ncRNA_intronic | LINC01475 |
| rs10883371 | chr10 | 101292455 | upstream | NKX2-3 |
| rs989979 | chr10 | 101322423 | intergenic | NKX2-3(dist=26143),SLC25A28(dist=47852) |
| rs2033784 | chr15 | 67449660 | intronic | SMAD3 |
| rs7174445 | chr15 | 67451215 | intronic | SMAD3 |
| rs1074631 | chr16 | 28554108 | intergenic | NUPR1(dist=3613),SGF29(dist=11141) |
| rs4788076 | chr16 | 28570005 | intronic | SGF29 |
| rs7193402 | chr16 | 28586127 | intronic | SGF29 |
| rs17707300 | chr16 | 28593347 | intronic | SGF29 |
| rs4788074 | chr16 | 28593597 | intronic | SGF29 |
| rs8062405 | chr16 | 28837906 | intronic | ATXN2L |
| rs12443881 | chr16 | 28841777 | intronic | ATXN2L |
| rs17221417 | chr16 | 50739582 | intronic | NOD2 |
| rs2066843 | chr16 | 50745199 | exonic | NOD2 |
| rs1861759 | chr16 | 50745583 | exonic | NOD2 |
| rs748855 | chr16 | 50751398 | intronic | NOD2 |
| rs8060598 | chr16 | 50781802 | intronic | CYLD |
| rs7342715 | chr16 | 50787483 | intronic | CYLD |
| rs3135503 | chr16 | 50791250 | intronic | CYLD |
| rs589365 | chr18 | 2154783 | intergenic | LINC00470(dist=795153),METTL4(dist=382741) |
| rs672495 | chr18 | 2154927 | intergenic | LINC00470(dist=795297),METTL4(dist=382597) |
| rs2041756 | chr2 | 103049910 | intronic | IL18RAP |
| rs6708413 | chr2 | 103063369 | intronic | IL18RAP |
| rs7559479 | chr2 | 103068787 | UTR3 | IL18RAP(NM_003853:c.*146G>A) |
| rs10210302 | chr2 | 234158839 | intergenic | INPP5D(dist=42290),ATG16L1(dist=1378) |
| rs6752107 | chr2 | 234161448 | intronic | ATG16L1 |
| rs6431654 | chr2 | 234161769 | intronic | ATG16L1 |
| rs6737398 | chr2 | 234170397 | intronic | ATG16L1 |
| rs3828309 | chr2 | 234180410 | intronic | ATG16L1 |
| rs3792106 | chr2 | 234190740 | intronic | ATG16L1 |
| rs2241874 | chr2 | 234247627 | intronic | SAG |
| rs2241873 | chr2 | 234247924 | intronic | SAG |
| rs11679046 | chr2 | 234258101 | intergenic | SAG(dist=2400),DGKD(dist=5052) |
| rs2838517 | chr21 | 45613825 | intergenic | C21orf33(dist=48220),ICOSLG(dist=29053) |
| rs762421 | chr21 | 45615561 | intergenic | C21orf33(dist=49956),ICOSLG(dist=27317) |
| rs762422 | chr21 | 45615638 | intergenic | C21orf33(dist=50033),ICOSLG(dist=27240) |
| rs2838520 | chr21 | 45615896 | intergenic | C21orf33(dist=50291),ICOSLG(dist=26982) |
| rs2838521 | chr21 | 45615917 | intergenic | C21orf33(dist=50312),ICOSLG(dist=26961) |
| rs4410472 | chr3 | 49329090 | intronic | USP4 |
| rs6784820 | chr3 | 49450864 | intronic | TCTA |
| rs6997 | chr3 | 49453834 | UTR3 | TCTA(NM_022171:c.*1539C>T) |
| rs1464567 | chr3 | 49459252 | intronic | AMT |
| rs3870338 | chr3 | 49557051 | intronic | DAG1 |
| rs1801143 | chr3 | 49570200 | exonic | DAG1 |
| rs1050088 | chr3 | 49570882 | UTR3 | DAG1 |
| rs9827708 | chr3 | 49649989 | intronic | BSN |
| rs11919311 | chr3 | 49656789 | intronic | BSN |
| rs9858542 | chr3 | 49701983 | exonic | BSN |
| rs4855881 | chr3 | 49715446 | intronic | APEH |
| rs2271961 | chr3 | 49878113 | intronic | TRAIP |
| rs2352974 | chr3 | 49890613 | intronic | TRAIP |
| rs2240327 | chr3 | 50113034 | intronic | RBM6 |
| rs10512734 | chr5 | 40393605 | intergenic | LINC00603(dist=340179),PTGER4(dist=286427) |
| rs16869934 | chr5 | 40397352 | intergenic | LINC00603(dist=343926),PTGER4(dist=282680) |
| rs17234657 | chr5 | 40401509 | intergenic | LINC00603(dist=348083),PTGER4(dist=278523) |
| rs10213846 | chr5 | 40442869 | intergenic | LINC00603(dist=389443),PTGER4(dist=237163) |
| rs11957215 | chr5 | 40445681 | intergenic | LINC00603(dist=392255),PTGER4(dist=234351) |
| rs4957297 | chr5 | 40455074 | intergenic | LINC00603(dist=401648),PTGER4(dist=224958) |
| rs4957300 | chr5 | 40463739 | intergenic | LINC00603(dist=410313),PTGER4(dist=216293) |
| rs6871834 | chr5 | 40480187 | intergenic | LINC00603(dist=426761),PTGER4(dist=199845) |
| rs6882351 | chr5 | 40481654 | intergenic | LINC00603(dist=428228),PTGER4(dist=198378) |
| rs1505992 | chr5 | 40498577 | intergenic | LINC00603(dist=445151),PTGER4(dist=181455) |
| rs1553576 | chr5 | 40509655 | intergenic | LINC00603(dist=456229),PTGER4(dist=170377) |
| rs1553577 | chr5 | 40510007 | intergenic | LINC00603(dist=456581),PTGER4(dist=170025) |
| rs6896604 | chr5 | 40516017 | intergenic | LINC00603(dist=462591),PTGER4(dist=164015) |
| rs6866402 | chr5 | 40517331 | intergenic | LINC00603(dist=463905),PTGER4(dist=162701) |
| rs1876143 | chr5 | 40521648 | intergenic | LINC00603(dist=468222),PTGER4(dist=158384) |
| rs7718309 | chr5 | 40528899 | intergenic | LINC00603(dist=475473),PTGER4(dist=151133) |
| rs11750156 | chr5 | 40561358 | intergenic | LINC00603(dist=507932),PTGER4(dist=118674) |
| rs10055946 | chr5 | 40570075 | intergenic | LINC00603(dist=516649),PTGER4(dist=109957) |
| rs4434422 | chr5 | 40600917 | intergenic | LINC00603(dist=547491),PTGER4(dist=79115) |
| rs2135330 | chr5 | 40602209 | intergenic | LINC00603(dist=548783),PTGER4(dist=77823) |
| rs10473203 | chr5 | 40606294 | intergenic | LINC00603(dist=552868),PTGER4(dist=73738) |
| rs13181692 | chr5 | 40607998 | intergenic | LINC00603(dist=554572),PTGER4(dist=72034) |
| rs2549794 | chr5 | 96244549 | intronic | ERAP2 |
| rs27307 | chr5 | 96338505 | intronic | LNPEP |
| rs27290 | chr5 | 96350088 | intronic | LNPEP |
| rs27300 | chr5 | 96363407 | intronic | LNPEP |
| rs2136188 | chr5 | 131577514 | intergenic | P4HA2(dist=13958),PDLIM4(dist=15837) |
| rs6890009 | chr5 | 131580033 | intergenic | P4HA2(dist=16477),PDLIM4(dist=13318) |
| rs6871350 | chr5 | 131580220 | intergenic | P4HA2(dist=16664),PDLIM4(dist=13131) |
| rs2285673 | chr5 | 131755969 | ncRNA_intronic | C5orf56 |
| rs11744116 | chr5 | 131779760 | ncRNA_intronic | C5orf56 |
| rs4540166 | chr5 | 131779857 | ncRNA_intronic | C5orf56 |
| rs10077785 | chr5 | 131801158 | ncRNA_intronic | C5orf56 |
| rs6861600 | chr5 | 158819615 | intergenic | LOC285626(dist=29773),LOC285627(dist=55949) |
| rs17056763 | chr5 | 158880527 | ncRNA_exonic | LOC285627 |
| rs11738617 | chr5 | 158881276 | ncRNA_intronic | LOC285627 |
| rs1799964 | chr6 | 31542308 | downstream | LTA |
| rs1052248 | chr6 | 31556581 | UTR3 | LST1 |
| rs9272346 | chr6 | 32604372 | upstream | HLA-DQA1 |
| rs2533146 | chr7 | 153463391 | intergenic | LINC01287(dist=354072),DPP6(dist=120791) |
| rs921720 | chr8 | 126534671 | intergenic | TRIB1(dist=84024),LINC00861(dist=400096) |
| rs6478108 | chr9 | 117558703 | intronic | TNFSF15 |
| rs4263839 | chr9 | 117566440 | intronic | TNFSF15 |

Table S3 : Identified variants using IGESS(WTCCC + Early Onset + Cedar2 + NIddkJ + NiddkNJ)

| SNP id | Chr | Pos | Type | Gene |
|---|---|---|---|---|
| rs707472 | chr1 | 7906008 | downstream | PER3 |
| rs7528804 | chr1 | 67593819 | intronic | C1orf141 |
| rs10889656 | chr1 | 67594559 | intronic | C1orf141 |
| rs3762318 | chr1 | 67597119 | intronic | C1orf141 |
| rs4655679 | chr1 | 67599657 | intronic | C1orf141 |
| rs10789224 | chr1 | 67605134 | intergenic | C1orf141(dist=4480),IL23R(dist=27035) |
| rs17375018 | chr1 | 67655147 | intronic | IL23R |
| rs4655689 | chr1 | 67659421 | intronic | IL23R |
| rs11209018 | chr1 | 67667291 | intronic | IL23R |
| rs11805303 | chr1 | 67675516 | intronic | IL23R |
| rs2201841 | chr1 | 67694202 | intronic | IL23R |
| rs6660226 | chr1 | 67744601 | intergenic | IL23R(dist=18951),IL12RB2(dist=28446) |
| rs12141431 | chr1 | 67747023 | intergenic | IL23R(dist=21373),IL12RB2(dist=26024) |
| rs12119179 | chr1 | 67747415 | intergenic | IL23R(dist=21765),IL12RB2(dist=25632) |
| rs11209039 | chr1 | 67751193 | intergenic | IL23R(dist=25543),IL12RB2(dist=21854) |
| rs6679677 | chr1 | 114303808 | downstream | RSBN1 |
| rs3768566 | chr1 | 155201064 | intergenic | GBAP1(dist=3739),GBA(dist=3175) |
| rs11264345 | chr1 | 155213124 | intronic | GBA |
| rs7520184 | chr1 | 155253583 | intronic | HCN3 |
| rs11264359 | chr1 | 155282829 | intronic | FDPS |
| rs5005770 | chr1 | 155345043 | intronic | ASH1L |
| rs1325908 | chr1 | 155413304 | intronic | ASH1L |
| rs11264375 | chr1 | 155424065 | intronic | ASH1L |
| rs475550 | chr1 | 155652081 | intronic | YY1AP1 |
| rs821551 | chr1 | 155688580 | intronic | DAP3 |
| rs822490 | chr1 | 155822971 | intronic | GON4L |
| rs4916197 | chr1 | 172831286 | intergenic | FASLG(dist=195274),TNFSF18(dist=179074) |
| rs12037853 | chr1 | 172852403 | intergenic | FASLG(dist=216391),TNFSF18(dist=157957) |
| rs10489276 | chr1 | 172862939 | intergenic | FASLG(dist=226927),TNFSF18(dist=147421) |
| rs10922215 | chr1 | 197363948 | intronic | CRB1 |
| rs3024505 | chr1 | 206939904 | intergenic | MAPKAPK2(dist=32274),IL10(dist=1044) |
| rs303436 | chr10 | 30731893 | intronic | MAP3K8 |
| rs10995238 | chr10 | 64387108 | intronic | ZNF365 |
| rs7071642 | chr10 | 64414060 | intronic | ZNF365 |
| rs7089612 | chr10 | 64414164 | intronic | ZNF365 |
| rs7915131 | chr10 | 64418656 | intronic | ZNF365 |
| rs729739 | chr10 | 64430302 | UTR3 | ZNF365(NM_199451:c.*241G>A,NM_199452:c.*241G>A) |
| rs10761659 | chr10 | 64445564 | intergenic | ZNF365(dist=13793),ADO(dist=118952) |
| rs224136 | chr10 | 64470675 | intergenic | ZNF365(dist=38904),ADO(dist=93841) |
| rs224147 | chr10 | 64485815 | intergenic | ZNF365(dist=54044),ADO(dist=78701) |
| rs224063 | chr10 | 64503349 | intergenic | ZNF365(dist=71578),ADO(dist=61167) |
| rs224067 | chr10 | 64506971 | intergenic | ZNF365(dist=75200),ADO(dist=57545) |
| rs224090 | chr10 | 64541319 | intergenic | ZNF365(dist=109548),ADO(dist=23197) |
| rs224092 | chr10 | 64541404 | intergenic | ZNF365(dist=109633),ADO(dist=23112) |
| rs224302 | chr10 | 64592805 | intergenic | EGR2(dist=13878),NRBP2(dist=300202) |
| rs1250538 | chr10 | 81037800 | intronic | ZMIZ1 |
| rs1250564 | chr10 | 81047342 | intronic | ZMIZ1 |
| rs7095491 | chr10 | 101274058 | intergenic | GOT1(dist=83528),LINC01475(dist=12049) |
| rs7078219 | chr10 | 101274365 | intergenic | GOT1(dist=83835),LINC01475(dist=11742) |
| rs7081330 | chr10 | 101274465 | intergenic | GOT1(dist=83935),LINC01475(dist=11642) |
| rs10883365 | chr10 | 101287764 | ncRNA_exonic | LINC01475 |
| rs10883367 | chr10 | 101287990 | ncRNA_intronic | LINC01475 |
| rs11190139 | chr10 | 101288099 | ncRNA_intronic | LINC01475 |
| rs1548962 | chr10 | 101289735 | ncRNA_intronic | LINC01475 |
| rs6584283 | chr10 | 101290301 | ncRNA_intronic | LINC01475 |
| rs10883371 | chr10 | 101292455 | upstream | NKX2-3 |
| rs989979 | chr10 | 101322423 | intergenic | NKX2-3(dist=26143),SLC25A28(dist=47852) |
| rs7927894 | chr11 | 76301316 | intergenic | EMSY(dist=37373),LRRC32(dist=67252) |
| rs10784518 | chr12 | 40737115 | intronic | LRRK2 |
| rs1365765 | chr12 | 40741821 | intronic | LRRK2 |
| rs4768233 | chr12 | 40743788 | intronic | LRRK2 |
| rs4768234 | chr12 | 40743830 | intronic | LRRK2 |
| rs744910 | chr15 | 67446785 | intronic | SMAD3 |
| rs12441344 | chr15 | 67447895 | intronic | SMAD3 |
| rs2033784 | chr15 | 67449660 | intronic | SMAD3 |
| rs7174445 | chr15 | 67451215 | intronic | SMAD3 |
| rs1074631 | chr16 | 28554108 | intergenic | NUPR1(dist=3613),SGF29(dist=11141) |
| rs4788076 | chr16 | 28570005 | intronic | SGF29 |
| rs7193402 | chr16 | 28586127 | intronic | SGF29 |
| rs17707300 | chr16 | 28593347 | intronic | SGF29 |
| rs4788074 | chr16 | 28593597 | intronic | SGF29 |
| rs8062405 | chr16 | 28837906 | intronic | ATXN2L |
| rs12443881 | chr16 | 28841777 | intronic | ATXN2L |
| rs4785393 | chr16 | 50259483 | intronic | PAPD5 |
| rs745230 | chr16 | 50670182 | UTR3 | NKD1(NM_033119:c.*2490G>C) |
| rs7199150 | chr16 | 50676514 | intergenic | NKD1(dist=1743),SNX20(dist=23697) |
| rs17221417 | chr16 | 50739582 | intronic | NOD2 |
| rs2066843 | chr16 | 50745199 | exonic | NOD2 |
| rs1861759 | chr16 | 50745583 | exonic | NOD2 |
| rs748855 | chr16 | 50751398 | intronic | NOD2 |
| rs8060598 | chr16 | 50781802 | intronic | CYLD |
| rs3785142 | chr16 | 50787147 | intronic | CYLD |
| rs7342715 | chr16 | 50787483 | intronic | CYLD |
| rs3135503 | chr16 | 50791250 | intronic | CYLD |
| rs2919366 | chr18 | 68056742 | intergenic | LOC101060542(dist=4915),GTSCR1(dist=241013) |
| rs12712135 | chr2 | 102930948 | intronic | IL1RL1 |
| rs13019081 | chr2 | 102950822 | intronic | IL1RL1 |
| rs1362349 | chr2 | 102951972 | intronic | IL1RL1 |
| rs1035127 | chr2 | 103019919 | intergenic | IL18R1(dist=4684),IL18RAP(dist=15331) |
| rs2041756 | chr2 | 103049910 | intronic | IL18RAP |
| rs6708413 | chr2 | 103063369 | intronic | IL18RAP |
| rs7559479 | chr2 | 103068787 | UTR3 | IL18RAP(NM_003853:c.*146G>A) |
| rs759382 | chr2 | 103094213 | intronic | SLC9A4 |
| rs7587251 | chr2 | 198930197 | intronic | PLCL1 |
| rs10210302 | chr2 | 234158839 | intergenic | INPP5D(dist=42290),ATG16L1(dist=1378) |
| rs6752107 | chr2 | 234161448 | intronic | ATG16L1 |
| rs6431654 | chr2 | 234161769 | intronic | ATG16L1 |
| rs6737398 | chr2 | 234170397 | intronic | ATG16L1 |
| rs3828309 | chr2 | 234180410 | intronic | ATG16L1 |
| rs3792106 | chr2 | 234190740 | intronic | ATG16L1 |
| rs4663421 | chr2 | 234201700 | intronic | ATG16L1 |
| rs2241874 | chr2 | 234247627 | intronic | SAG |
| rs2241873 | chr2 | 234247924 | intronic | SAG |
| rs11679046 | chr2 | 234258101 | intergenic | SAG(dist=2400),DGKD(dist=5052) |
| rs1736135 | chr21 | 16805220 | intergenic | NRIP1(dist=368094),USP25(dist=297124) |
| rs991774 | chr21 | 16811110 | intergenic | NRIP1(dist=373984),USP25(dist=291234) |
| rs1736145 | chr21 | 16811996 | intergenic | NRIP1(dist=374870),USP25(dist=290348) |
| rs1736020 | chr21 | 16812552 | intergenic | NRIP1(dist=375426),USP25(dist=289792) |
| rs2838517 | chr21 | 45613825 | intergenic | C21orf33(dist=48220),ICOSLG(dist=29053) |
| rs2838519 | chr21 | 45615023 | intergenic | C21orf33(dist=49418),ICOSLG(dist=27855) |
| rs762421 | chr21 | 45615561 | intergenic | C21orf33(dist=49956),ICOSLG(dist=27317) |
| rs762422 | chr21 | 45615638 | intergenic | C21orf33(dist=50033),ICOSLG(dist=27240) |
| rs2838520 | chr21 | 45615896 | intergenic | C21orf33(dist=50291),ICOSLG(dist=26982) |
| rs2838521 | chr21 | 45615917 | intergenic | C21orf33(dist=50312),ICOSLG(dist=26961) |
| rs181359 | chr22 | 21928641 | intronic | UBE2L3 |
| rs181360 | chr22 | 21928916 | intronic | UBE2L3 |
| rs2283790 | chr22 | 21956653 | intronic | UBE2L3 |
| rs4821116 | chr22 | 21973319 | intronic | UBE2L3 |
| rs11708786 | chr3 | 48752654 | intronic | IP6K2 |
| rs4410472 | chr3 | 49329090 | intronic | USP4 |
| rs9864406 | chr3 | 49352619 | intronic | USP4 |
| rs9863142 | chr3 | 49366741 | intronic | USP4 |

Table S3 Continued

| SNP id | Chr | Pos | | Gene |
|---|---|---|---|---|
| rs6784820 | chr3 | 49450864 | intronic | TCTA |
| rs6997 | chr3 | 49453834 | UTR3 | TCTA(NM_022171:c.*1539C>T) |
| rs1464567 | chr3 | 49459252 | intronic | AMT |
| rs3870338 | chr3 | 49557051 | intronic | DAG1 |
| rs1801143 | chr3 | 49570200 | exonic | DAG1 |
| rs1050088 | chr3 | 49570882 | UTR3 | DAG1 |
| rs9827708 | chr3 | 49649989 | intronic | BSN |
| rs11919311 | chr3 | 49656789 | intronic | BSN |
| rs9858542 | chr3 | 49701983 | exonic | BSN |
| rs4855881 | chr3 | 49715446 | intronic | APEH |
| rs11130214 | chr3 | 49735746 | intronic | RNF123 |
| rs2291542 | chr3 | 49751585 | exonic | RNF123 |
| rs3749237 | chr3 | 49770032 | intronic | IP6K1 |
| rs2271961 | chr3 | 49878113 | intronic | TRAIP |
| rs2352974 | chr3 | 49890613 | intronic | TRAIP |
| rs2240327 | chr3 | 50113034 | intronic | RBM6 |
| rs10512734 | chr5 | 40393605 | intergenic | LINC00603(dist=340179),PTGER4(dist=286427) |
| rs16869934 | chr5 | 40397352 | intergenic | LINC00603(dist=343926),PTGER4(dist=282680) |
| rs17234657 | chr5 | 40401509 | intergenic | LINC00603(dist=348083),PTGER4(dist=278523) |
| rs10213846 | chr5 | 40442869 | intergenic | LINC00603(dist=389443),PTGER4(dist=237163) |
| rs11957215 | chr5 | 40445681 | intergenic | LINC00603(dist=392255),PTGER4(dist=234351) |
| rs4957297 | chr5 | 40455074 | intergenic | LINC00603(dist=401648),PTGER4(dist=224958) |
| rs4957300 | chr5 | 40463739 | intergenic | LINC00603(dist=410313),PTGER4(dist=216293) |
| rs6871834 | chr5 | 40480187 | intergenic | LINC00603(dist=426761),PTGER4(dist=199845) |
| rs6882351 | chr5 | 40481654 | intergenic | LINC00603(dist=428228),PTGER4(dist=198378) |
| rs1505992 | chr5 | 40498577 | intergenic | LINC00603(dist=445151),PTGER4(dist=181455) |
| rs1553576 | chr5 | 40509655 | intergenic | LINC00603(dist=456229),PTGER4(dist=170377) |
| rs1553577 | chr5 | 40510007 | intergenic | LINC00603(dist=456581),PTGER4(dist=170025) |
| rs6896604 | chr5 | 40516017 | intergenic | LINC00603(dist=462591),PTGER4(dist=164015) |
| rs6866402 | chr5 | 40517331 | intergenic | LINC00603(dist=463905),PTGER4(dist=162701) |
| rs1876143 | chr5 | 40521648 | intergenic | LINC00603(dist=468222),PTGER4(dist=158384) |
| rs10941516 | chr5 | 40522212 | intergenic | LINC00603(dist=468786),PTGER4(dist=157820) |
| rs7718309 | chr5 | 40528899 | intergenic | LINC00603(dist=475473),PTGER4(dist=151133) |
| rs11750156 | chr5 | 40561358 | intergenic | LINC00603(dist=507932),PTGER4(dist=118674) |
| rs10055946 | chr5 | 40570075 | intergenic | LINC00603(dist=516649),PTGER4(dist=109957) |
| rs4434422 | chr5 | 40600917 | intergenic | LINC00603(dist=547491),PTGER4(dist=79115) |
| rs2135330 | chr5 | 40602209 | intergenic | LINC00603(dist=548783),PTGER4(dist=77823) |
| rs10473203 | chr5 | 40606294 | intergenic | LINC00603(dist=552868),PTGER4(dist=73738) |
| rs13181692 | chr5 | 40607998 | intergenic | LINC00603(dist=554572),PTGER4(dist=72034) |
| rs2278019 | chr5 | 96225252 | intronic | ERAP2 |
| rs10434709 | chr5 | 96225774 | intronic | ERAP2 |
| rs2548533 | chr5 | 96238401 | intronic | ERAP2 |
| rs2549794 | chr5 | 96244549 | intronic | ERAP2 |
| rs1056893 | chr5 | 96245439 | exonic | ERAP2 |
| rs2549797 | chr5 | 96245518 | intronic | ERAP2 |
| rs2910787 | chr5 | 96274223 | intronic | LNPEP |
| rs27307 | chr5 | 96338505 | intronic | LNPEP |
| rs27290 | chr5 | 96350088 | intronic | LNPEP |
| rs27300 | chr5 | 96363407 | intronic | LNPEP |
| rs152125 | chr5 | 131427161 | intergenic | CSF2(dist=15298),P4HA2-AS1(dist=93408) |
| rs39897 | chr5 | 131436896 | intergenic | CSF2(dist=25033),P4HA2-AS1(dist=83673) |
| rs715285 | chr5 | 131485383 | intergenic | CSF2(dist=73520),P4HA2-AS1(dist=35186) |
| rs2278398 | chr5 | 131530441 | intronic | P4HA2 |
| rs4361509 | chr5 | 131536753 | intronic | P4HA2 |
| rs3792894 | chr5 | 131547271 | intronic | P4HA2 |
| rs9791170 | chr5 | 131569627 | intergenic | P4HA2(dist=6071),PDLIM4(dist=23724) |
| rs2136188 | chr5 | 131577514 | intergenic | P4HA2(dist=13958),PDLIM4(dist=15837) |
| rs6890009 | chr5 | 131580033 | intergenic | P4HA2(dist=16477),PDLIM4(dist=13318) |
| rs6871350 | chr5 | 131580220 | intergenic | P4HA2(dist=16664),PDLIM4(dist=13131) |
| rs10463891 | chr5 | 131597392 | intronic | PDLIM4 |
| rs2285673 | chr5 | 131755969 | ncRNA_intronic | C5orf56 |
| rs11744116 | chr5 | 131779760 | ncRNA_intronic | C5orf56 |
| rs4540166 | chr5 | 131779857 | ncRNA_intronic | C5orf56 |
| rs10077785 | chr5 | 131801158 | ncRNA_intronic | C5orf56 |
| rs11957134 | chr5 | 150230950 | intergenic | IRGM(dist=2719),ZNF300(dist=43004) |
| rs6893009 | chr5 | 150233304 | intergenic | IRGM(dist=5073),ZNF300(dist=40650) |
| rs4958847 | chr5 | 150239587 | intergenic | IRGM(dist=11356),ZNF300(dist=34367) |
| rs1000113 | chr5 | 150240076 | intergenic | IRGM(dist=11845),ZNF300(dist=33878) |
| rs11747270 | chr5 | 150258867 | intergenic | IRGM(dist=30636),ZNF300(dist=15087) |
| rs10041072 | chr5 | 150259642 | intergenic | IRGM(dist=31411),ZNF300(dist=14312) |
| rs3900064 | chr5 | 150264414 | intergenic | IRGM(dist=36183),ZNF300(dist=9540) |
| rs270661 | chr5 | 158560154 | intergenic | LOC101927740(dist=15668),RNF145(dist=24263) |
| rs1363670 | chr5 | 158784111 | ncRNA_intronic | LOC285626 |
| rs6861600 | chr5 | 158819615 | intergenic | LOC285626(dist=29773),LOC285627(dist=55949) |
| rs17388425 | chr5 | 158824174 | intergenic | LOC285626(dist=34332),LOC285627(dist=51390) |
| rs4921227 | chr5 | 158849837 | intergenic | LOC285626(dist=59995),LOC285627(dist=25727) |
| rs17056763 | chr5 | 158880527 | ncRNA_exonic | LOC285627 |
| rs11738617 | chr5 | 158881276 | ncRNA_intronic | LOC285627 |
| rs6888934 | chr5 | 158931798 | intergenic | LOC285627(dist=38514),LOC101927766(dist=2719 84) |
| rs1799964 | chr6 | 31542308 | downstream | LTA |
| rs1052248 | chr6 | 31556581 | UTR3 | LST1 |
| rs3130484 | chr6 | 31715882 | ncRNA_intronic | MSH5-SAPCD1 |
| rs3131379 | chr6 | 31721033 | ncRNA_intronic | MSH5-SAPCD1 |
| rs1150753 | chr6 | 32059867 | intronic | TNXB |
| rs206015 | chr6 | 32182759 | intronic | NOTCH4 |
| rs3129934 | chr6 | 32336187 | intronic | C6orf10 |
| rs2894254 | chr6 | 32345689 | intergenic | C6orf10(dist=6000),HCG23(dist=12598) |
| rs9450667 | chr6 | 88094386 | intergenic | C6orf163(dist=19205),LINC01590(dist=12456) |
| rs4707364 | chr6 | 88097764 | intergenic | C6orf163(dist=22583),LINC01590(dist=9078) |
| rs9401937 | chr6 | 127388087 | intergenic | MIR588(dist=582228),RSPO3(dist=51961) |
| rs9375486 | chr6 | 127388186 | intergenic | MIR588(dist=582327),RSPO3(dist=51862) |
| rs2800708 | chr6 | 127437617 | intergenic | MIR588(dist=631758),RSPO3(dist=2431) |
| rs1936805 | chr6 | 127452116 | intronic | RSPO3 |
| rs2489623 | chr6 | 127455821 | intronic | RSPO3 |
| rs2503322 | chr6 | 127457260 | intronic | RSPO3 |
| rs9285458 | chr6 | 127463645 | intronic | RSPO3 |
| rs9491700 | chr6 | 127482008 | intronic | RSPO3 |
| rs9491701 | chr6 | 127482207 | intronic | RSPO3 |
| rs6569474 | chr6 | 127493611 | intronic | RSPO3 |
| rs9491706 | chr6 | 127496226 | intronic | RSPO3 |
| rs9491706 | chr6 | 127496226 | intronic | RSPO3 |
| rs9968920 | chr6 | 127529209 | intergenic | RSPO3(dist=8583),RNF146(dist=58618) |
| rs4895494 | chr6 | 137937380 | intergenic | OLIG3(dist=121849),LOC102723649(dist=49403) |
| rs487438 | chr6 | 137947988 | intergenic | OLIG3(dist=132457),LOC102723649(dist=38795) |
| rs1819333 | chr6 | 167373547 | intergenic | RNASET2(dist=3470),MIR3939(dist=37748) |
| rs9366076 | chr6 | 167373708 | intergenic | RNASET2(dist=3631),MIR3939(dist=37587) |
| rs386548 | chr6 | 167385533 | intergenic | RNASET2(dist=15456),MIR3939(dist=25762) |
| rs408918 | chr6 | 167399282 | intergenic | RNASET2(dist=29205),MIR3939(dist=12013) |
| rs933243 | chr6 | 167403873 | intergenic | RNASET2(dist=33796),MIR3939(dist=7422) |
| rs422562 | chr6 | 167406318 | intergenic | RNASET2(dist=36241),MIR3939(dist=4977) |
| rs9457252 | chr6 | 167433925 | intronic | FGFR1OP |
| rs1894603 | chr6 | 167434686 | intronic | FGFR1OP |
| rs7749278 | chr6 | 167435325 | intronic | FGFR1OP |
| rs9295385 | chr6 | 167448181 | intronic | FGFR1OP |
| rs12209395 | chr6 | 167463914 | intergenic | FGFR1OP(dist=8008),CCR6(dist=61381) |
| rs720325 | chr6 | 167467349 | intergenic | FGFR1OP(dist=11443),CCR6(dist=57946) |
| rs1358883 | chr6 | 167467433 | intergenic | FGFR1OP(dist=11527),CCR6(dist=57862) |
| rs4710175 | chr6 | 167467800 | intergenic | FGFR1OP(dist=11894),CCR6(dist=57495) |
| rs12203510 | chr6 | 167473006 | intergenic | FGFR1OP(dist=17100),CCR6(dist=52289) |
| rs6921588 | chr6 | 167494397 | intergenic | FGFR1OP(dist=38491),CCR6(dist=30898) |
| rs2533146 | chr7 | 153463391 | intergenic | LINC01287(dist=354072),DPP6(dist=120791) |
| rs921720 | chr8 | 126534671 | intergenic | TRIB1(dist=84024),LINC00861(dist=400096) |
| rs6478108 | chr9 | 117558703 | intronic | TNFSF15 |
| rs4263839 | chr9 | 117566440 | intronic | TNFSF15 |
| rs10448340 | chr9 | 139320069 | intergenic | PMPCA(dist=1856),INPP5E(dist=2998) |
| rs3812591 | chr9 | 139341612 | intronic | SEC16A |
| rs11145756 | chr9 | 139364585 | intronic | SEC16A |
| rs4379550 | chr9 | 139376426 | intronic | SEC16A |