Supplement to "A Minimax Optimal Ridge-Type Set Test for Global Hypothesis with Applications in Whole Genome Sequencing Association Studies"

Yaowu Liu, Zilin Li and Xihong Lin *

In Section 1, we provide the derivation of $Q(\tau)$, and the proofs of Theorems 1 and 2, Corollaries 1 and 2, as well as other technical lemmas. In Section 2, we describe an algorithm to compute τ_c^* using the bisection search method. In Section 3, we provide the complete table for the type I error simulation results with all the sample sizes. In Section 4, we present additional simulation results and the genomic landscapes of significant sliding windows in ARIC WGS data analysis.

1 Proof of main results

1.1 Derivation of $Q(\tau)$

In the marginal model of $\mathbf{S}$, we test $H_0 : \mathbf{S} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$ against $H_a : \mathbf{S} \sim N_p(\mathbf{0}, \mathbf{\Sigma} + \tau\mathbf{\Sigma}^2)$. The likelihood ratio test statistic is equivalent to $\mathbf{S}^T\mathbf{\Sigma}^{-1}\mathbf{S} - \mathbf{S}^T[\mathbf{\Sigma} + \tau\mathbf{\Sigma}^2]^{-1}\mathbf{S}$. Using the eigenvalue decomposition of $\mathbf{\Sigma}$, we have $\mathbf{\Sigma} = \mathbf{U} \cdot \text{diag}\{\lambda_i\} \cdot \mathbf{U}^T$, where $\text{diag}\{\lambda_i\}$ is a diagonal matrix whose elements are the eigenvalues and $\mathbf{U}$ is a matrix of eigenvectors. Hence,

$$\begin{aligned} & \mathbf{S}^T\mathbf{\Sigma}^{-1}\mathbf{S} - \mathbf{S}^T[\mathbf{\Sigma} + \tau\mathbf{\Sigma}^2]^{-1}\mathbf{S} \\ = \; & \mathbf{S}^T\mathbf{U} \cdot \text{diag}\{\frac{1}{\lambda_i}\} \cdot \mathbf{U}^T\mathbf{S} - \mathbf{S}^T\mathbf{U} \cdot \text{diag}\{\frac{1}{\lambda_i(1+\lambda_i)}\} \cdot \mathbf{U}^T\mathbf{S} \\ = \; & \mathbf{S}^T\mathbf{U} \cdot \text{diag}\{\frac{\tau}{1+\tau\lambda_i}\} \cdot \mathbf{U}^T\mathbf{S} = \tau\mathbf{S}^T\left(\mathbf{I} + \tau\mathbf{\Sigma}\right)^{-1}\mathbf{S} = \tau Q(\tau). \end{aligned}$$

1.2 Proof of Theorem 1

(a) Let $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_p)$ be the eigenvalues of $\mathbf{\Sigma}$ and $\hat{\boldsymbol{\lambda}}_n = (\hat{\lambda}_{1n}, \cdots, \hat{\lambda}_{pn})$ be the eigenvalues of $\hat{\mathbf{\Sigma}}_n$. Because $\hat{\mathbf{\Sigma}}_n \xrightarrow{p} \mathbf{\Sigma}$, then $\hat{\boldsymbol{\lambda}}_n \xrightarrow{p} \boldsymbol{\lambda}$. Since τ is a fixed constant, it follows from the Slutsky's theorem

*Yaowu Liu is Associate Professor at the School of Statistics at the Southwestern University of Finance and Economics (*yaowuliu615@gmail.com*). Zilin Li is a postdoctoral fellow in the Department of Biostatistics at Harvard T.H. Chan School of Public Health(*li@hsph.harvard.edu*). Xihong Lin is Professor of Biostatistics at Harvard T.H. Chan School of Public Health and Professor of Statistics at Harvard University (*xlin@hsph.harvard.edu*).

that $Q_n(\tau) \xrightarrow{d} \sum_{i=1}^{p} \frac{\lambda_i}{1+\lambda_i \tau} \chi^2_{1,i}$. It is easy to see that the quantile $q_\alpha(\tau, \boldsymbol{\lambda})$ is continuous w.r.t. $\boldsymbol{\lambda}$. Hence $q_\alpha(\tau, \hat{\boldsymbol{\lambda}}_n) \xrightarrow{p} q_\alpha(\tau, \boldsymbol{\lambda})$. We complete the proof of Theorem 1(a).

(b) We first consider the case where $\boldsymbol{\Sigma}$ is not an identity matrix up to some constant, i.e., the eigenvalues of $\boldsymbol{\Sigma}$ are not all the same. By Lemma 3 that will be introduced in Section 1.5, $\tau_c^*(\boldsymbol{\lambda})$ is continuous in $\mathcal{D}_{\boldsymbol{\lambda}}(c_1, c_2, \delta)$ for any $0 < c_1 < c_2 < +\infty$ and $\delta > 0$, where $\mathcal{D}_{\boldsymbol{\lambda}}(c_1, c_2, \delta)$ is the domain defined in (1). Since $\hat{\boldsymbol{\lambda}}_n \xrightarrow{p} \boldsymbol{\lambda}$, for sufficiently large n, we can find some c_1, c_2 and δ such that $P\{\hat{\boldsymbol{\lambda}}_n \notin \mathcal{D}_{\boldsymbol{\lambda}}(c_1, c_2, \delta)\}$ is sufficiently small. This along with the continuity of $\tau_c^*(\cdot)$ gives that $\tau_c^*(\hat{\boldsymbol{\lambda}}_n) \xrightarrow{p} \tau_c^*(\boldsymbol{\lambda})$. The rest of proof is similar to that for the part (a) of Theorem 1.

Next, we consider the case where $\boldsymbol{\Sigma}$ is an identity matrix. Since $\hat{\lambda}_{in} = 1 + o_p(1)$ for any $1 \le i \le p$, we have

$$\frac{\hat{\lambda}_{in}}{1 + \hat{\lambda}_{in} \tau_c^*(\hat{\boldsymbol{\lambda}}_n)} = \frac{1}{1/\hat{\lambda}_{in} + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)} = \frac{1}{1 + o_p(1) + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)} = (1 + o_p(1)) \cdot \frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)},$$

where $o_p(1)$ denotes a random variable converges to 0 in probability. Let χ^2_p denote the chi-squared distribution with p degrees of freedom and $q_\alpha(\chi^2_p)$ be its upper α-quantile. Because

$$\sum_{i=1}^{p} \frac{\hat{\lambda}_{in}}{1 + \hat{\lambda}_{in} \tau_c^*(\hat{\boldsymbol{\lambda}}_n)} \chi^2_{1,i} = \frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)} \chi^2_p + o_p\left(\frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)}\right),$$

then the critical value

$$q_\alpha(\tau_c^*(\hat{\boldsymbol{\lambda}}_n)) = \frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)} q_\alpha(\chi^2_p) + o_p\left(\frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)}\right).$$

Further, the test statistic

$$Q_n(\tau_c^*(\hat{\boldsymbol{\lambda}}_n)) = \frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)} [\chi^2_p + o_p(1)] + o_p\left(\frac{\chi^2_p + o_p(1)}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)}\right) = \frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)} \chi^2_p + o_p\left(\frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)}\right).$$

Then, the size of the test is

$$\begin{aligned} & P\left[Q_n(\tau_c^*(\hat{\boldsymbol{\lambda}}_n)) \ge q_\alpha(\tau_c^*(\hat{\boldsymbol{\lambda}}_n))\right] \\ = \; & P\left[\frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)} \chi^2_p + o_p\left(\frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)}\right) \ge \frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)} q_\alpha(\chi^2_p) + o_p\left(\frac{1}{1 + \tau_c^*(\hat{\boldsymbol{\lambda}}_n)}\right)\right] \\ = \; & P\left[\chi^2_p \ge q_\alpha(\chi^2_p) + o_p(1)\right] \to \alpha. \end{aligned}$$

Therefore, the test has an asymptotic α-level.

1.3 Proof of Theorem 2

Note that $H(\hat{\boldsymbol{\theta}}_n) = \hat{\boldsymbol{\beta}}_n \xrightarrow{d} N(\boldsymbol{\beta}, \boldsymbol{\Omega}(\boldsymbol{\theta}))$. We first consider the case where $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is not an identity matrix up to some constant. Since $\boldsymbol{\Omega}(\boldsymbol{\theta})$ does not depend on $\boldsymbol{\beta}$, in the limiting multivariate normal model, the test that $\tau_c^*(\boldsymbol{\lambda})$ corresponds to is minimax optimal w.r.t. the risk we defined in the main text. Analogous to the proof of Theorem 1, we can show that for $\tau_c^*(\hat{\boldsymbol{\lambda}}_n)$ and any $\tau \geq 0$, $Q_n^W(\tau_c^*(\hat{\boldsymbol{\lambda}}_n))$ and $Q_n^W(\tau)$ converge in distribution to their corresponding test statistics in the limiting model, respectively. Hence, $Q_n^W(\tau_c^*(\hat{\boldsymbol{\lambda}}_n))$ is asymptotically minimax optimal. When $\boldsymbol{\Omega}(\boldsymbol{\theta})$ is an identity matrix, the risk for $Q_n^W(\tau_c^*(\hat{\boldsymbol{\lambda}}_n))$ goes to to 0 asymptotically and therefore $Q_n^W(\tau_c^*(\hat{\boldsymbol{\lambda}}_n))$ is optimal. In fact, for any $\tau > 0$, the risk for $Q_n^W(\tau)$ also goes to 0 and the tests based on $Q_n^W(\tau_c^*(\hat{\boldsymbol{\lambda}}_n))$ and $Q_n^W(\tau)$ are asymptotically equivalent.

1.4 Proof of Corollaries 1 and 2

Note that $\hat{\boldsymbol{\beta}}_n$ and $\mathbf{S}_n(\boldsymbol{\theta})$ converge to a multivariate normal distribution and the Fisher's information matrix is continuous w.r.t. $\boldsymbol{\theta}$ under regularity conditions. Corollaries 1 and 2 directly follow from Theorem 1.

1.5 Technical Lemmas

The goal of this section is to show that τ_c^*, as a function of the eigenvalues, is continuous for any fixed $0 < \alpha < 1$. Since τ_c^* involves max and argmin, we first introduce two lemmas about the continuity of functions that involve max and argmin, respectively.

Lemma 1. *Let $f(x, \mathbf{z})$ be a uniformly continuous function w.r.t. $(x, \mathbf{z}) \in \mathcal{D}_x \times \mathcal{D}_{\mathbf{z}}$, where $x \in \mathcal{D}_x \subseteq \mathbb{R}, \mathbf{z} \in \mathcal{D}_{\mathbf{z}} \subseteq \mathbb{R}^k$ for some $k \geq 1$. For any $\mathbf{z}_0 \in \mathcal{D}_{\mathbf{z}}$, there exists at least one $x_0^{max} \in \mathcal{D}_x$, such that $\max_{x \in \mathcal{D}_x} f(x, \mathbf{z}_0) = f(x_0^{max}, \mathbf{z}_0)$. Then $\max_{x \in \mathcal{D}_x} f(x, \mathbf{z})$ is uniformly continuous w.r.t. $\mathbf{z} \in \mathcal{D}_{\mathbf{z}}$.*

Proof. It suffices to show that $\forall \epsilon > 0, \exists \delta > 0, \forall ||\mathbf{z}_1 - \mathbf{z}_2|| < \delta, |\max_{x \in \mathcal{D}_x} f(x, \mathbf{z}_1) - \max_{x \in \mathcal{D}_x} f(x, \mathbf{z}_2)| < \epsilon$. Let $x_1^{\max}$ (or $x_2^{\max}$) denote a point in $\mathcal{D}_x$ that achieves the maximum of $f(x, \mathbf{z}_1)$ (or $f(x, \mathbf{z}_2)$). We write

$$I = \max_{x \in \mathcal{D}_x} f(x, \mathbf{z}_1) - \max_{x \in \mathcal{D}_x} f(x, \mathbf{z}_2) = f(x_1^{\max}, \mathbf{z}_1) - f(x_2^{\max}, \mathbf{z}_2).$$

First, we have

$$I = [f(x_1^{\max}, \mathbf{z}_1) - f(x_1^{\max}, \mathbf{z}_2)] + [f(x_1^{\max}, \mathbf{z}_2) - f(x_2^{\max}, \mathbf{z}_2)] = I_1 + I_2.$$

If follows from the uniform continuity of $f(x, \mathbf{z})$ that $I_1 < \epsilon$. Since $x_2^{\max}$ is the maximum point, $I_2 \leq 0$. Hence, we obtain $I < \epsilon$. Similarly,

$$I = [f(x_1^{\max}, \mathbf{z}_1) - f(x_2^{\max}, \mathbf{z}_1)] + [f(x_2^{\max}, \mathbf{z}_1) - f(x_2^{\max}, \mathbf{z}_2)] > -\epsilon.$$

□

Lemma 2. *Let $h(y, \mathbf{z})$ be a uniformly continuous function w.r.t. $(y, \mathbf{z}) \in \mathcal{D}_y \times \mathcal{D}_{\mathbf{z}}$, where $y \in \mathcal{D}_y \subseteq \mathbb{R}, \mathbf{z} \in \mathcal{D}_{\mathbf{z}} \subseteq \mathbb{R}^k$ for some $k \geq 1$. For any $\mathbf{z}_0 \in \mathcal{D}_{\mathbf{z}}$, there exists a point $y_0^{min} \in \mathcal{D}_y$, such that $h(y, \mathbf{z}_0)$ is strictly increasing for $y \geq y_0^{min}$ and strictly decreasing for $y \leq y_0^{min}$. Then $\operatorname{argmin}_{y \in \mathcal{D}_y} h(y, \mathbf{z})$ is continuous w.r.t. $\mathbf{z} \in \mathcal{D}_{\mathbf{z}}$.*

Proof. If $\operatorname{argmin}_{y \in \mathcal{D}_y} h(y, \mathbf{z})$ is not continuous for any $\mathbf{z} \in \mathcal{D}_{\mathbf{z}}$, there exist a $\mathbf{z}_0 \in \mathcal{D}_{\mathbf{z}}$, some $\epsilon > 0$ and a sequence of $\mathbf{z}_n \to z_0$, for any n, $|y_n^{\min} - y_0^{\min}| > \epsilon$, where $y_n^{\min} = \operatorname{argmin}_{y \in \mathcal{D}_y} h(y, \mathbf{z}_n)$ and $y_0^{\min} = \operatorname{argmin}_{y \in \mathcal{D}_y} h(y, \mathbf{z}_0)$. Let

$$\begin{aligned} I &= h(y_n^{\min}, \mathbf{z}_n) - h(y_0^{\min}, \mathbf{z}_0) \\ &= [h(y_n^{\min}, \mathbf{z}_n) - h(y_n^{\min}, \mathbf{z}_0)] + [h(y_n^{\min}, \mathbf{z}_0) - h(y_0^{\min}, \mathbf{z}_0)] = I_1 + I_2. \end{aligned}$$

It follows from Lemma 1 that $I \to 0$ as $n \to +\infty$. By the uniform continuity, we have also $I_1 \to 0$. This implies that $I_2 \to 0$. However, since $h(y, \mathbf{z}_0)$ is strictly increasing for $y \geq y_0^{\min}$ and decreasing for $y \leq y_0^{\min}$, then $I_2 \geq \max\{h(y_0^{\min} + \epsilon, \mathbf{z}_0), h(y_0^{\min} - \epsilon, \mathbf{z}_0)\} - h(y_0, \mathbf{z}_0) > 0$ for any n. This is in contradiction to $I_2 \to 0$. □

Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_p)$ be the vector of eigenvalues. Our goal is to show that $\tau_c^*(\boldsymbol{\lambda})$ is continuous for any fixed $0 < \alpha < 1$. To investigate the continuity, we first need to define the domain of $\boldsymbol{\lambda}$. It can be seen that our proposed test does not depend on the scale of $\lambda_i's$. Hence, the domain could be $\sum_{i=1}^p \lambda_i = c$ for any constant $c > 0$. However, for some technical reasons, we would like the domain to be a bounded and closed set in $\mathbb{R}^p$. Hence, we expand the domain to be $c_1 \leq \sum_{i=1}^p \lambda_i \leq c_2$. In addition, if $\lambda_1 = \lambda_2 = \cdots = \lambda_p$, the risk function $R_\alpha(\tau_a, \tau_c) = 0$ for any $\tau_a, \tau_c \geq 0$ and therefor τ_c^* is not unique. To avoid this situation, we also need to set the domain away from the hyperplane $\lambda_1 = \lambda_2 = \cdots = \lambda_p$. Formally, we define the domain of $\boldsymbol{\lambda}$ as

$$\mathcal{D}_{\boldsymbol{\lambda}}(c_1, c_2, \delta) = \{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0 : c_1 \leq \sum_{i=1}^p \lambda_i \leq c_2, \max_{1 \leq i \leq p-1} (\lambda_i - \lambda_{i-1}) \geq \delta\}, \tag{1}$$

where constants $0 < c_1 < c_2 < +\infty$ and $\delta > 0$.

Lemma 3. *For any* $0 < \alpha < 1$, $\tau_c^*(\boldsymbol{\lambda})$ *is continuous in* $\mathcal{D}_{\boldsymbol{\lambda}}(c_1, c_2, \delta)$ *for any* $0 < c_1 < c_2 < +\infty$ *and* $\delta > 0$.

Proof. We rewrite $\mathcal{D}_{\boldsymbol{\lambda}}(c_1, c_2, \delta)$ as $\mathcal{D}_{\boldsymbol{\lambda}}$ to simplify the exposition, and rewrite $\Psi_\alpha(\tau_a, \tau_c)$ as $\Psi_\alpha(\tau_a, \tau_c, \boldsymbol{\lambda})$ to make the dependency on $\boldsymbol{\lambda}$ explicit. It is easy to see that $\Psi_\alpha(\tau_a, \tau_c, \boldsymbol{\lambda})$ is continuous. Note that the domain of τ_c or τ_a is $[0, +\infty)$. If τ_c and τ_a is defined on $[0, C_0]$ for any $C_0 > 0$, then the uniform continuity of $\Psi_\alpha(\tau_a, \tau_c, \boldsymbol{\lambda})$ follows from the fact that a function that is continuous in a bounded and closed set is also uniformly continuous. Because $\mathcal{D}_{\boldsymbol{\lambda}}$ is bounded from above and below and $0 < \alpha < 1$ is fixed, we can find a $C_0 > 0$, for any $\tau_a, \tau_c > C_0$ and $\boldsymbol{\lambda} \in \mathcal{D}_{\boldsymbol{\lambda}}$, such that $1 - \Psi_\alpha(\tau_a, \tau_c, \boldsymbol{\lambda}) > \epsilon$ for some sufficiently small $\epsilon > 0$. Statistically speaking, when the signal strength τ_a is sufficiently large, any test in the class $\{Q(\tau_c)\}$ would have a power close to 1. Therefore, $\Psi_\alpha(\tau_a, \tau_c, \boldsymbol{\lambda})$ is uniformly continuous in $[0, +\infty) \times [0, +\infty) \times \mathcal{D}_{\boldsymbol{\lambda}}$.

Since the risk function $R_\alpha(\tau_a, \tau_c, \boldsymbol{\lambda})$ is the difference between two power functions, it is also uniformly continuous. Then it follows from Lemma 1 that $\max_{\tau_a \geq 0} R_\alpha(\tau_a, \tau_c, \boldsymbol{\lambda})$ is uniformly continuous. In the rest of the proof, we write τ_a as x to make the notation more clear. Since $\boldsymbol{\lambda} \in \mathcal{D}_{\boldsymbol{\lambda}}$ is bounded away from the hyperplane $\lambda_1 = \lambda_2 = \cdots = \lambda_p$, for any $\boldsymbol{\lambda}_0 \in \mathcal{D}_{\boldsymbol{\lambda}}$ and $x_0 > 0$, $R_\alpha(x_0, \tau_c, \boldsymbol{\lambda}_0)$ is strictly increasing when $\tau_c > x_0$ and strictly decreasing when $\tau_c < x_0$. This implies that the maximum risk on the left-hand side $\max_{0 \leq x \leq \tau_c} R_\alpha(x, \tau_c, \boldsymbol{\lambda}_0)$ is strictly increasing w.r.t. τ_c and the maximum risk on the right-hand side $\max_{x \geq \tau_c} R_\alpha(x, \tau_c, \boldsymbol{\lambda}_0)$ is strictly decreasing w.r.t. τ_c. Hence, the maximum risk $\max_{x \geq 0} R_\alpha(x, \tau_c, \boldsymbol{\lambda}_0)$ is strictly decreasing when $\tau_c \leq \tau_c^*(\boldsymbol{\lambda}_0, \alpha_0)$ and strictly increasing when $\tau_c \geq \tau_c^*(\boldsymbol{\lambda}_0, \alpha_0)$. By Lemma 2, $\tau_c^*(\boldsymbol{\lambda})$ is continuous in $\mathcal{D}_{\boldsymbol{\lambda}}$. □

2 Algorithm

In this section, we describe an algorithm to compute τ_c^* using the bisection search method.

The bisection search method in Algorithm 1 is computationally more efficient than the grid search method. Suppose that K (e.g., $K = 200$) points are selected for τ_a and τ_c, respectively. To find the minimax solution, grid search needs to evaluate the risk function at $K \times K$ points, which is computationally intensive. In contrast, due to the monotonicity of the risk function w.r.t. τ_c described in Section 2.6, we do not need K points for τ_c. Instead, we use the bisection search (w.r.t. τ_c) that typically takes several iterations to converge within a given error tolerance, say the number of iterations $s = 5$. Then Algorithm 1 only requires to evaluate the risk function at $K \times s$ points.

Algorithm 1 Bisection search to compute τ_c^*

input: eigenvalues $\lambda_1, \cdots, \lambda_p$; significance level α; error bound Δ
initialize: choose a grid $0 < \tau_1 < \tau_2 < \cdots < \tau_K$, calculate $\Psi_\alpha(\tau_k, \tau_k)$ for $k = 1, \cdots, K$.
$\tau_c^* \leftarrow \text{argmin}_{\{\tau_k; 1 \le k \le K\}} |\Psi_\alpha(\tau_k, \tau_k) - 0.5|$; $\tau_L^* \leftarrow \tau_1$; $\tau_U^* \leftarrow \tau_K$;
$R_U \leftarrow \max_{\tau_k > \tau_c^*} R(\tau_k, \tau_c^*)$; $R_L \leftarrow \max_{\tau_k < \tau_c^*} R(\tau_k, \tau_c^*)$;
while $|R_U - R_L| > \Delta$ **do**
  if $R_U > R_L$ **then**
    $\tau_L^* \leftarrow \tau_c^*$; $\tau_c^* \leftarrow (\tau_c^* + \tau_U^*)/2$;
  else
    $\tau_U^* \leftarrow \tau_c^*$; $\tau_c^* \leftarrow (\tau_c^* + \tau_L^*)/2$;
  end if
  $R_U \leftarrow \max_{\tau_k > \tau_c^*} R(\tau_k, \tau_c^*)$; $R_L \leftarrow \max_{\tau_k < \tau_c^*} R(\tau_k, \tau_c^*)$;
end while

The next question is how to choose K equally spaced points since the range of τ_a and τ_c is $[0, \infty]$. We need to set a finite maximum value for τ_a. Let τ_{max} denote the maximum value. We choose τ_{max} such that when $\tau_a > \tau_{max}$, the power of SKAT is greater than $1 - \Delta$, where Δ is the error bound in Algorithm 1. Finding such a τ_{max} is easy and fast. Due to the monotonicity of the risk function w.r.t. τ_c again and the fact that SKAT is $Q(0)$, the power of any test in $\{Q(\tau) : \tau \ge 0\}$ is also greater than $1 - \Delta$ when $\tau_a > \tau_{max}$. This means that when the signal strength τ_a is greater than τ_{max}, the risk of any test in $\{Q(\tau) : \tau \ge 0\}$ is less than the error bound Δ. Therefore we do not need to consider the values of τ_a beyond τ_{max}. Then we can choose K points in $[0, \tau_{max}]$.

3 Supplementary Tables

Table 1: Type I error of MORST computed over 10^6 replications. Here n is the sample size and p is the number of predictors. The columns corresponds to exchangeable correlation, AR(1) correlation and the correlation of sequenced genotypes, respectively.

n	p	α	Continuous outcomes			Binary outcomes		
			Exchangeable	AR(1)	Sequencing	Exchangeable	AR(1)	Sequencing
2000	100	0.05	$4.57 \cdot 10^{-2}$	$4.88 \cdot 10^{-2}$	$4.95 \cdot 10^{-2}$	$4.68 \cdot 10^{-2}$	$4.97 \cdot 10^{-2}$	$5.01 \cdot 10^{-2}$
		$1 \cdot 10^{-2}$	$8.41 \cdot 10^{-3}$	$9.58 \cdot 10^{-3}$	$9.75 \cdot 10^{-3}$	$8.81 \cdot 10^{-3}$	$9.86 \cdot 10^{-3}$	$1.00 \cdot 10^{-2}$
		$1 \cdot 10^{-3}$	$7.97 \cdot 10^{-4}$	$8.91 \cdot 10^{-4}$	$9.61 \cdot 10^{-4}$	$8.64 \cdot 10^{-4}$	$9.61 \cdot 10^{-4}$	$9.58 \cdot 10^{-4}$
		$1 \cdot 10^{-4}$	$7.90 \cdot 10^{-5}$	$8.50 \cdot 10^{-5}$	$9.10 \cdot 10^{-5}$	$8.60 \cdot 10^{-5}$	$1.02 \cdot 10^{-4}$	$1.03 \cdot 10^{-4}$
	200	0.05	$4.30 \cdot 10^{-2}$	$4.83 \cdot 10^{-2}$	$4.94 \cdot 10^{-2}$	$4.40 \cdot 10^{-2}$	$4.86 \cdot 10^{-2}$	$4.99 \cdot 10^{-2}$
		$1 \cdot 10^{-2}$	$7.68 \cdot 10^{-3}$	$9.27 \cdot 10^{-3}$	$9.75 \cdot 10^{-3}$	$7.91 \cdot 10^{-3}$	$9.36 \cdot 10^{-3}$	$9.91 \cdot 10^{-3}$
		$1 \cdot 10^{-3}$	$6.79 \cdot 10^{-4}$	$8.63 \cdot 10^{-4}$	$9.56 \cdot 10^{-4}$	$6.89 \cdot 10^{-4}$	$8.68 \cdot 10^{-4}$	$9.35 \cdot 10^{-4}$
		$1 \cdot 10^{-4}$	$6.60 \cdot 10^{-5}$	$8.50 \cdot 10^{-5}$	$9.80 \cdot 10^{-5}$	$4.80 \cdot 10^{-5}$	$9.50 \cdot 10^{-5}$	$1.02 \cdot 10^{-4}$
5000	100	0.05	$4.84 \cdot 10^{-2}$	$4.93 \cdot 10^{-2}$	$4.99 \cdot 10^{-2}$	$4.89 \cdot 10^{-2}$	$4.99 \cdot 10^{-2}$	$4.99 \cdot 10^{-2}$
		$1 \cdot 10^{-2}$	$9.37 \cdot 10^{-3}$	$9.78 \cdot 10^{-3}$	$1.02 \cdot 10^{-2}$	$9.48 \cdot 10^{-3}$	$9.92 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$
		$1 \cdot 10^{-3}$	$9.43 \cdot 10^{-4}$	$9.62 \cdot 10^{-4}$	$9.63 \cdot 10^{-4}$	$8.71 \cdot 10^{-4}$	$9.79 \cdot 10^{-4}$	$9.43 \cdot 10^{-4}$
		$1 \cdot 10^{-4}$	$9.60 \cdot 10^{-5}$	$9.20 \cdot 10^{-5}$	$8.80 \cdot 10^{-5}$	$9.20 \cdot 10^{-5}$	$1.06 \cdot 10^{-4}$	$7.80 \cdot 10^{-5}$
	200	0.05	$4.73 \cdot 10^{-2}$	$4.93 \cdot 10^{-2}$	$4.96 \cdot 10^{-2}$	$4.80 \cdot 10^{-2}$	$4.92 \cdot 10^{-2}$	$4.98 \cdot 10^{-2}$
		$1 \cdot 10^{-2}$	$9.02 \cdot 10^{-3}$	$9.71 \cdot 10^{-3}$	$9.96 \cdot 10^{-3}$	$9.21 \cdot 10^{-3}$	$9.80 \cdot 10^{-3}$	$9.94 \cdot 10^{-3}$
		$1 \cdot 10^{-3}$	$8.40 \cdot 10^{-4}$	$9.00 \cdot 10^{-4}$	$1.02 \cdot 10^{-3}$	$8.77 \cdot 10^{-4}$	$8.91 \cdot 10^{-4}$	$9.89 \cdot 10^{-4}$
		$1 \cdot 10^{-4}$	$8.20 \cdot 10^{-5}$	$7.60 \cdot 10^{-5}$	$9.20 \cdot 10^{-5}$	$8.10 \cdot 10^{-5}$	$8.60 \cdot 10^{-5}$	$9.30 \cdot 10^{-5}$
10000	100	0.05	$4.90 \cdot 10^{-2}$	$5.02 \cdot 10^{-2}$	$4.94 \cdot 10^{-2}$	$4.94 \cdot 10^{-2}$	$4.98 \cdot 10^{-2}$	$5.02 \cdot 10^{-2}$
		$1 \cdot 10^{-2}$	$9.65 \cdot 10^{-3}$	$9.99 \cdot 10^{-3}$	$9.89 \cdot 10^{-3}$	$9.75 \cdot 10^{-3}$	$1.00 \cdot 10^{-2}$	$9.96 \cdot 10^{-3}$
		$1 \cdot 10^{-3}$	$9.60 \cdot 10^{-4}$	$9.81 \cdot 10^{-4}$	$1.01 \cdot 10^{-3}$	$9.96 \cdot 10^{-4}$	$9.85 \cdot 10^{-4}$	$9.28 \cdot 10^{-4}$
		$1 \cdot 10^{-4}$	$1.01 \cdot 10^{-4}$	$1.07 \cdot 10^{-4}$	$9.20 \cdot 10^{-5}$	$1.02 \cdot 10^{-4}$	$1.06 \cdot 10^{-4}$	$1.00 \cdot 10^{-4}$
	200	0.05	$4.87 \cdot 10^{-2}$	$4.97 \cdot 10^{-2}$	$5.01 \cdot 10^{-2}$	$4.89 \cdot 10^{-2}$	$4.96 \cdot 10^{-2}$	$5.00 \cdot 10^{-2}$
		$1 \cdot 10^{-2}$	$9.46 \cdot 10^{-3}$	$9.87 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$	$9.60 \cdot 10^{-3}$	$9.65 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$
		$1 \cdot 10^{-3}$	$9.56 \cdot 10^{-4}$	$9.96 \cdot 10^{-4}$	$9.84 \cdot 10^{-4}$	$9.17 \cdot 10^{-4}$	$9.56 \cdot 10^{-4}$	$9.59 \cdot 10^{-4}$
		$1 \cdot 10^{-4}$	$9.60 \cdot 10^{-5}$	$1.02 \cdot 10^{-4}$	$9.50 \cdot 10^{-5}$	$1.07 \cdot 10^{-4}$	$9.60 \cdot 10^{-5}$	$1.05 \cdot 10^{-4}$
20000	100	0.05	$4.95 \cdot 10^{-2}$	$5.00 \cdot 10^{-2}$	$5.00 \cdot 10^{-2}$	$5.00 \cdot 10^{-2}$	$5.00 \cdot 10^{-2}$	$5.00 \cdot 10^{-2}$
		$1 \cdot 10^{-2}$	$9.75 \cdot 10^{-3}$	$9.90 \cdot 10^{-3}$	$1.00 \cdot 10^{-2}$	$9.95 \cdot 10^{-3}$	$9.97 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$
		$1 \cdot 10^{-3}$	$9.87 \cdot 10^{-4}$	$9.86 \cdot 10^{-4}$	$9.94 \cdot 10^{-4}$	$1.03 \cdot 10^{-3}$	$9.65 \cdot 10^{-4}$	$9.58 \cdot 10^{-4}$
		$1 \cdot 10^{-4}$	$9.90 \cdot 10^{-5}$	$1.01 \cdot 10^{-4}$	$1.00 \cdot 10^{-4}$	$8.90 \cdot 10^{-5}$	$9.80 \cdot 10^{-5}$	$9.90 \cdot 10^{-5}$
	200	0.05	$4.94 \cdot 10^{-2}$	$4.99 \cdot 10^{-2}$	$5.00 \cdot 10^{-2}$	$4.95 \cdot 10^{-2}$	$5.01 \cdot 10^{-2}$	$5.01 \cdot 10^{-2}$
		$1 \cdot 10^{-2}$	$9.73 \cdot 10^{-3}$	$9.84 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$	$9.84 \cdot 10^{-3}$	$1.01 \cdot 10^{-2}$	$1.00 \cdot 10^{-2}$
		$1 \cdot 10^{-3}$	$9.37 \cdot 10^{-4}$	$9.77 \cdot 10^{-4}$	$9.87 \cdot 10^{-4}$	$1.04 \cdot 10^{-3}$	$1.01 \cdot 10^{-3}$	$1.02 \cdot 10^{-3}$
		$1 \cdot 10^{-4}$	$9.50 \cdot 10^{-5}$	$9.00 \cdot 10^{-5}$	$9.60 \cdot 10^{-5}$	$8.89 \cdot 10^{-5}$	$1.17 \cdot 10^{-4}$	$1.07 \cdot 10^{-4}$

4 Supplementary Figures

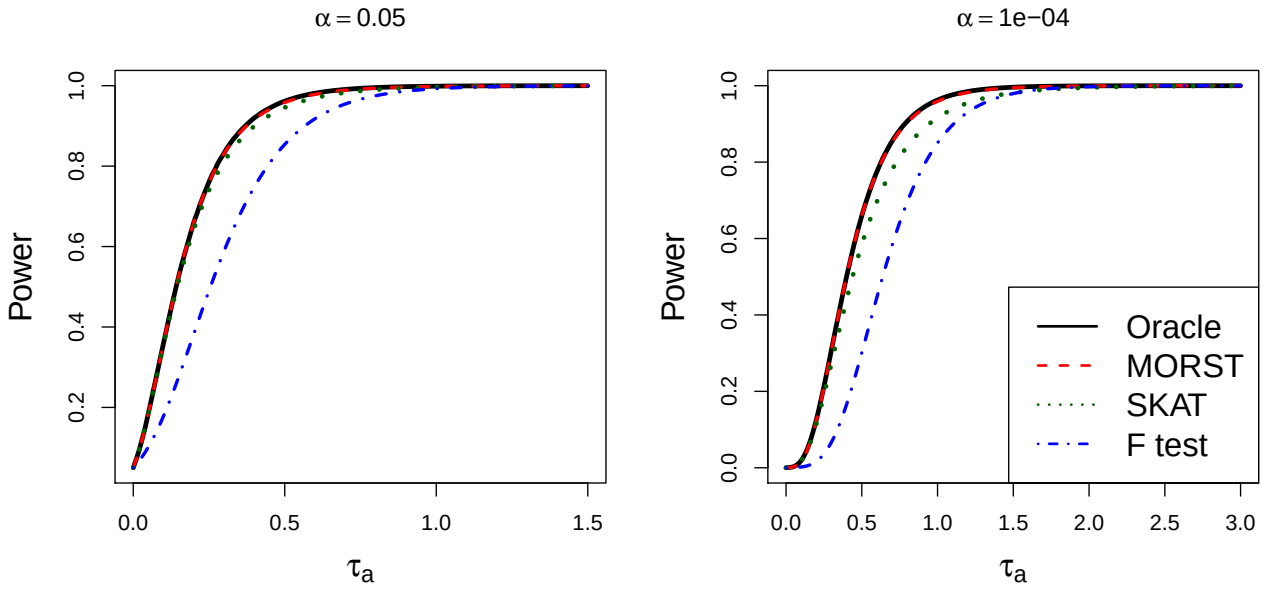

Figure 1: An example of the power curves (i.e., $\Psi_\alpha(\tau_a, \tau_c)$) of the oracle, MORST, SKAT and F tests under $\alpha = 0.05, 10^{-4}$. Here, $p = 100$ and $\mathbf{\Sigma} = (\sigma_{kl})$ with $\sigma_{kl} = 0.8^{|k-l|}$ for any $1 \leq k, l \leq p$.

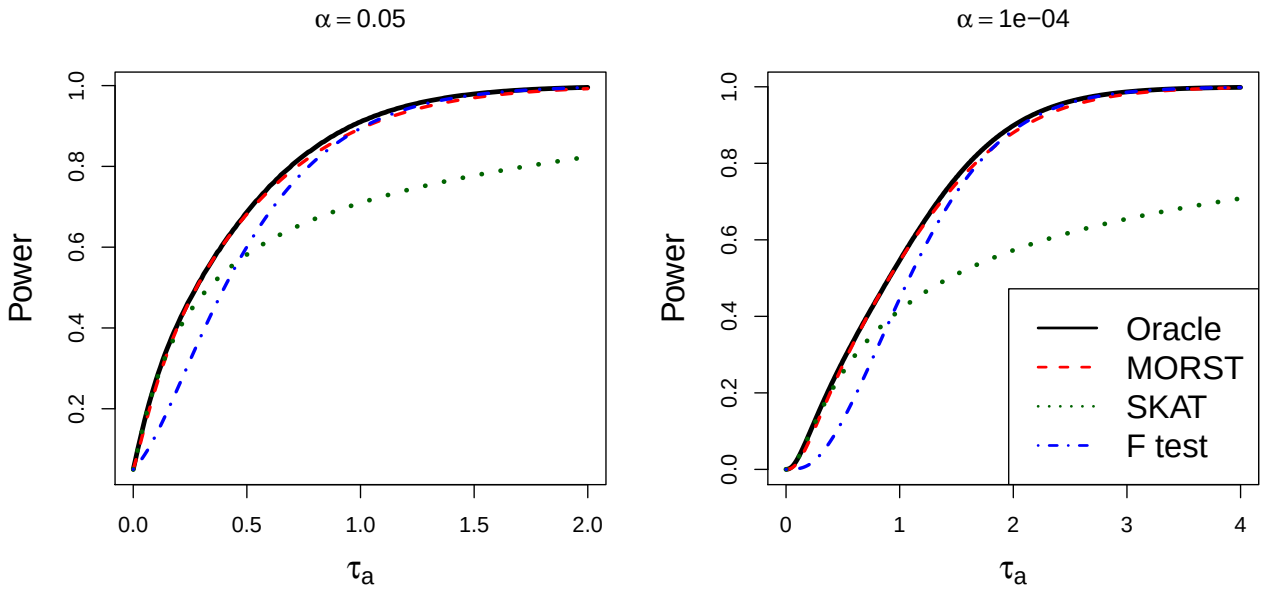

Figure 2: An example of the power curves (i.e., $\Psi_\alpha(\tau_a, \tau_c)$) of the oracle, MORST, SKAT and F tests under $\alpha = 0.05, 10^{-4}$. Here, $p = 50$ and $\mathbf{\Sigma} = (\sigma_{ij})$ with $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.4$ for any $1 \le i < j \le p$.

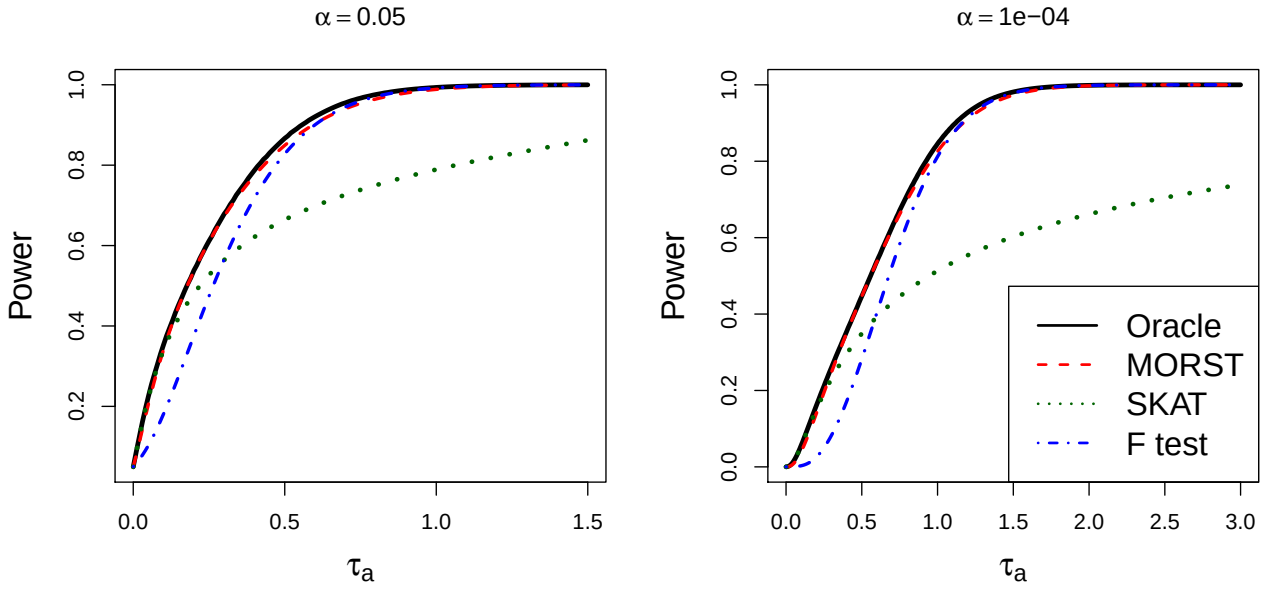

Figure 3: An example of the power curves (i.e., $\Psi_\alpha(\tau_a, \tau_c)$) of the oracle, MORST, SKAT and F tests under $\alpha = 0.05, 10^{-4}$. Here, $p = 100$ and $\mathbf{\Sigma} = (\sigma_{ij})$ with $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.3$ for any $1 \le i < j \le p$.

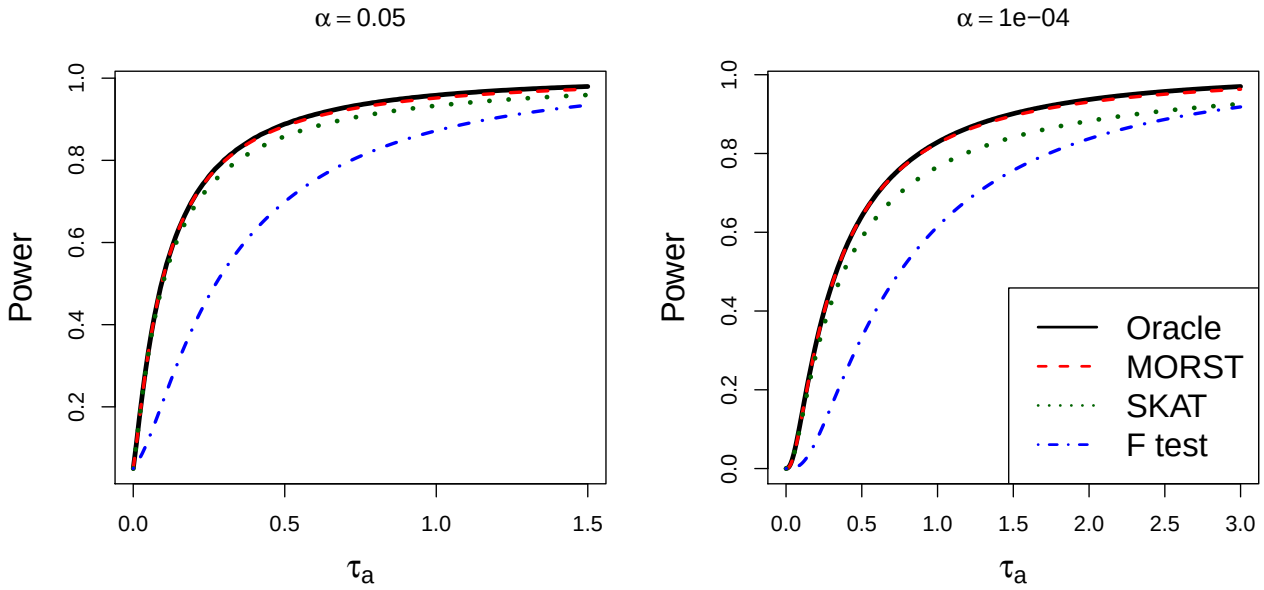

Figure 4: An example of the power curves (i.e., $\Psi_\alpha(\tau_a, \tau_c)$) of the oracle, MORST, SKAT and F tests under $\alpha = 0.05, 10^{-4}$. Here, $\mathbf{\Sigma}$ is the correlation matrix of 50 genetic variants of a continuous region randomly selected from the simulated sequencing data described in Section 4.1 in the main text.

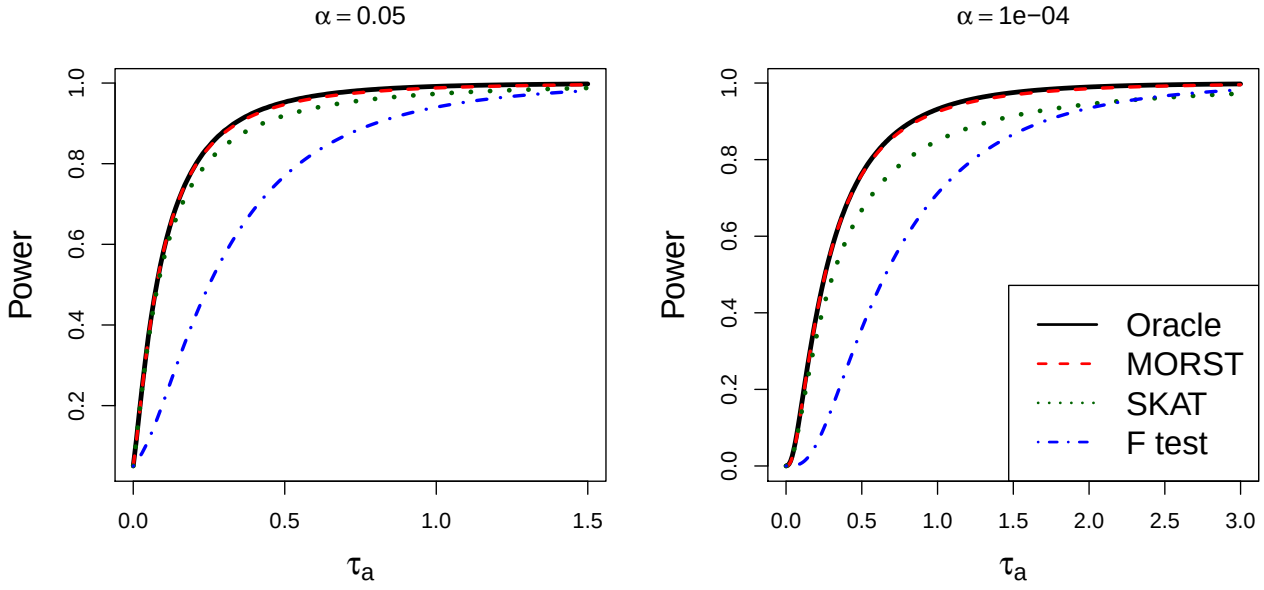

Figure 5: An example of the power curves (i.e., $\Psi_\alpha(\tau_a, \tau_c)$) of the oracle, MORST, SKAT and F tests under $\alpha = 0.05, 10^{-4}$. Here, $\mathbf{\Sigma}$ is the correlation matrix of 100 genetic variants of a continuous region randomly selected from the simulated sequencing data described in Section 4.1 in the main text.

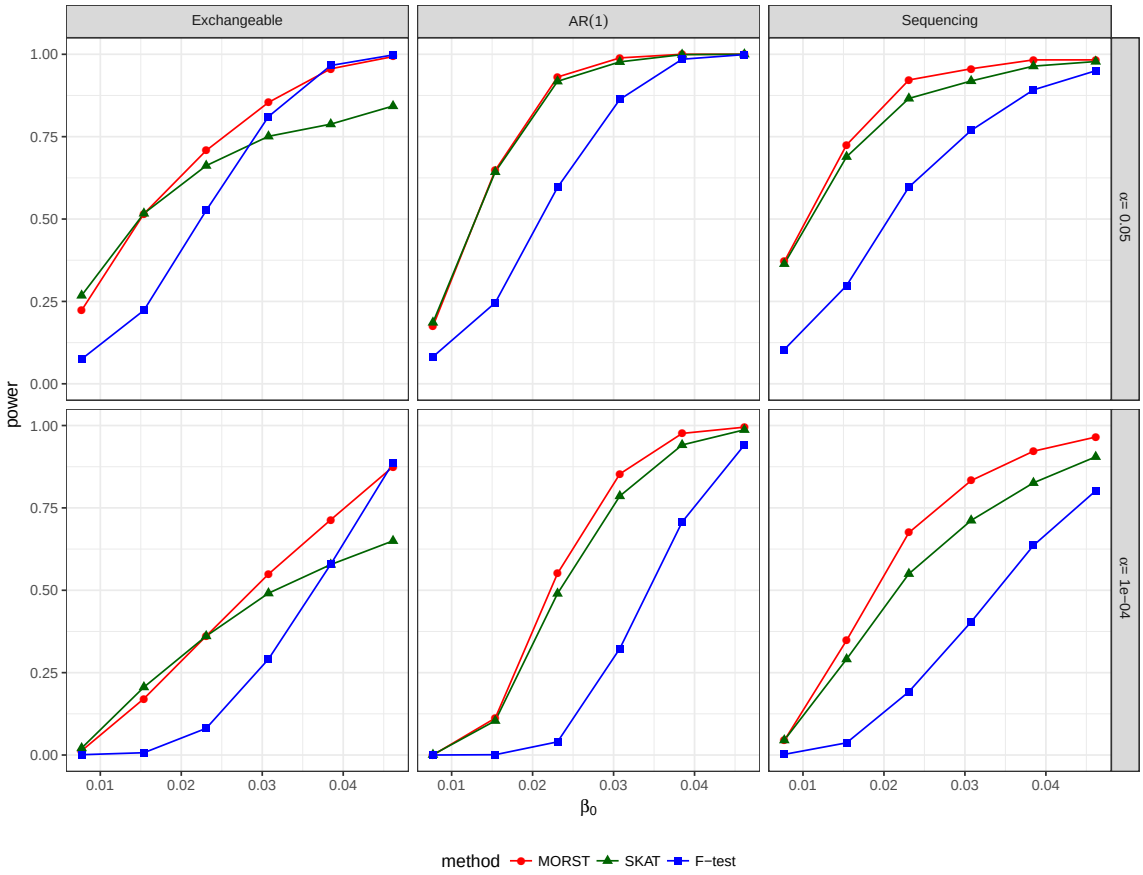

Figure 6: The power of MORST, SKAT and the F test in linear regression with $p = 200$ and 20% non-zero β_i's. The columns from left to right correspond to exchangeable correlation, AR(1) correlation and the correlation of sequenced genotypes, respectively. The rows correspond to $\alpha = 0.05, 0.0001$. The x-axis β_0 is the signal strength (i.e., the common magnitude of non-zero β_i's).

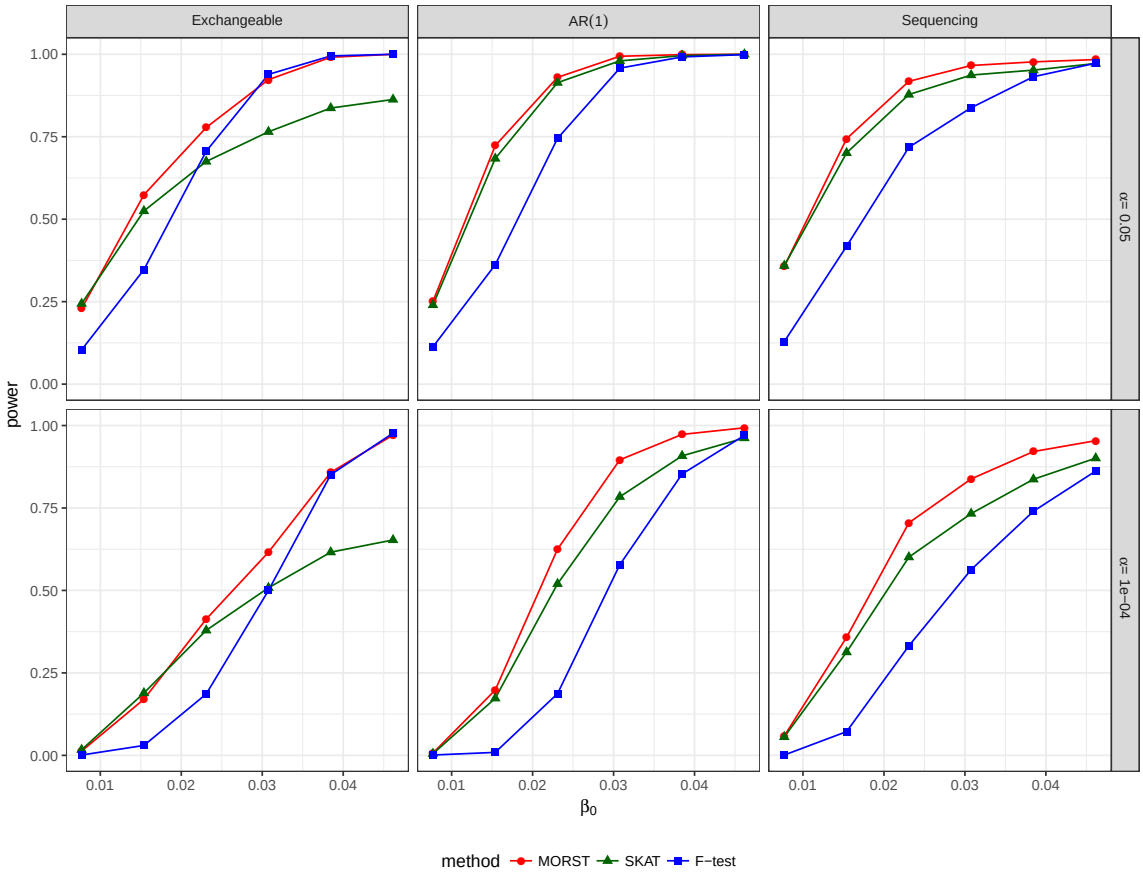

Figure 7: The power of MORST, SKAT and the F test in linear regression with $p = 100$ and 40% non-zero β_i's. The columns from left to right correspond to exchangeable correlation, AR(1) correlation and the correlation of sequenced genotypes, respectively. The rows correspond to $\alpha = 0.05, 0.0001$. The x-axis β_0 is the signal strength (i.e., the common magnitude of non-zero β_i's).

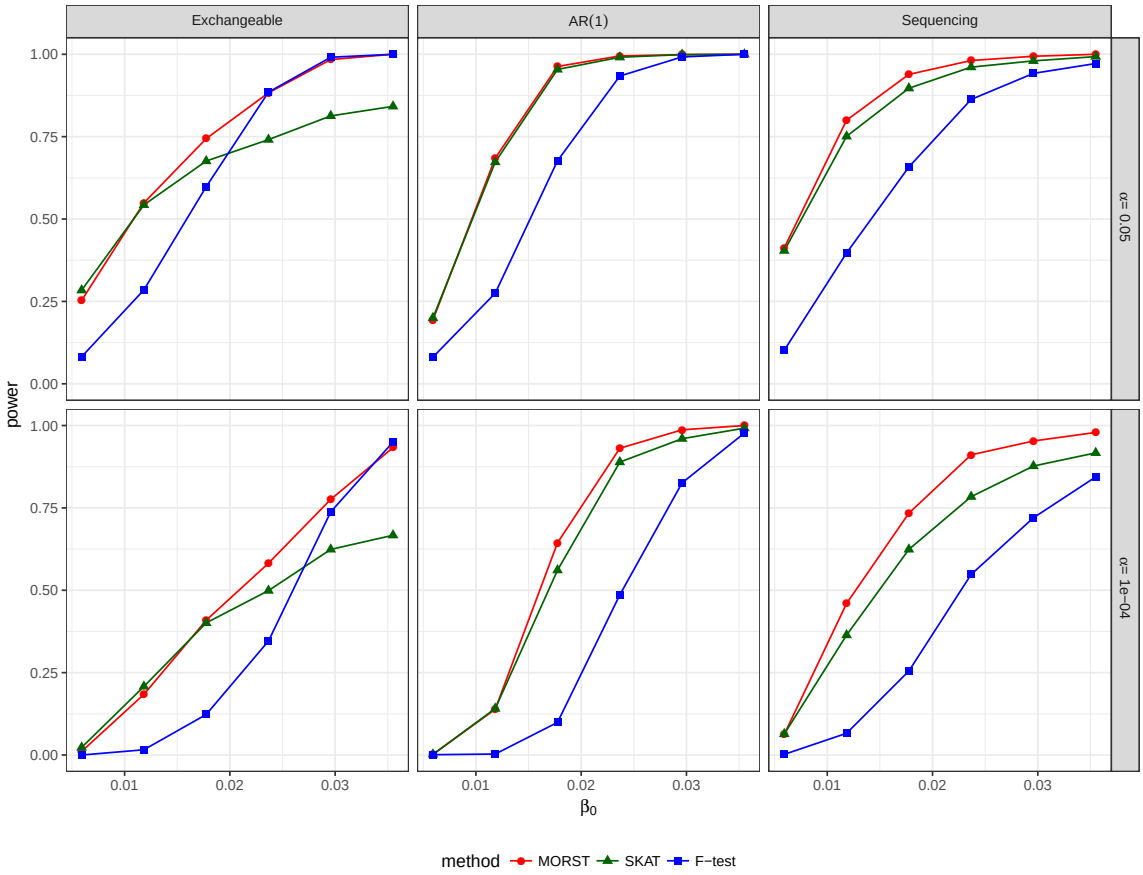

Figure 8: The power of MORST, SKAT and the F test in linear regression with $p = 200$ and 40% non-zero β_i's. The columns from left to right correspond to exchangeable correlation, AR(1) correlation and the correlation of sequenced genotypes, respectively. The rows correspond to $\alpha = 0.05, 0.0001$. The x-axis β_0 is the signal strength (i.e., the common magnitude of non-zero β_i's).

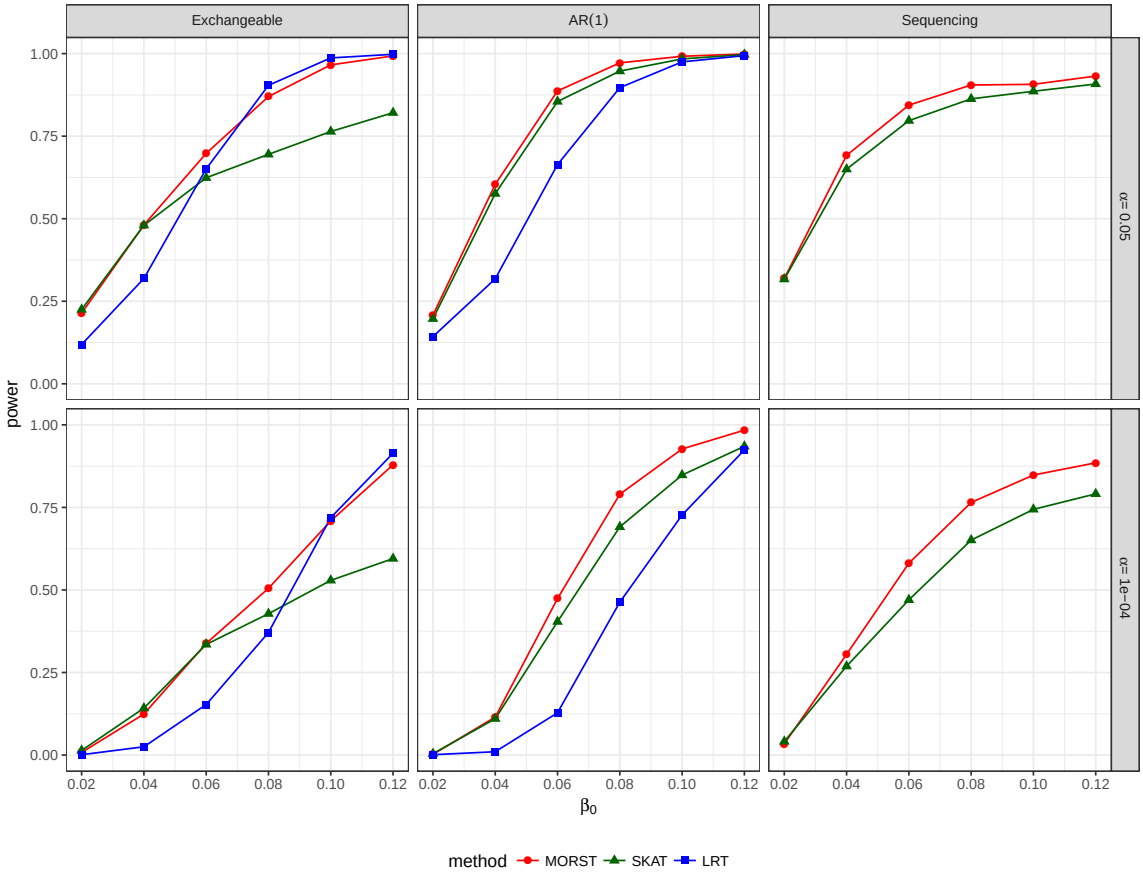

Figure 9: The power of MORST, SKAT and LRT in logistic regression with $p = 100$ and 20% non-zero β_i's. The columns from left to right correspond to exchangeable correlation, AR(1) correlation and the correlation of sequenced genotypes, respectively. The rows correspond to $\alpha = 0.05, 0.0001$. The x-axis β_0 is the signal strength (i.e., the common magnitude of non-zero β_i's). The LRT is not included in the panels on the right because the MLE does not exist for some genetic variants such as singletons.

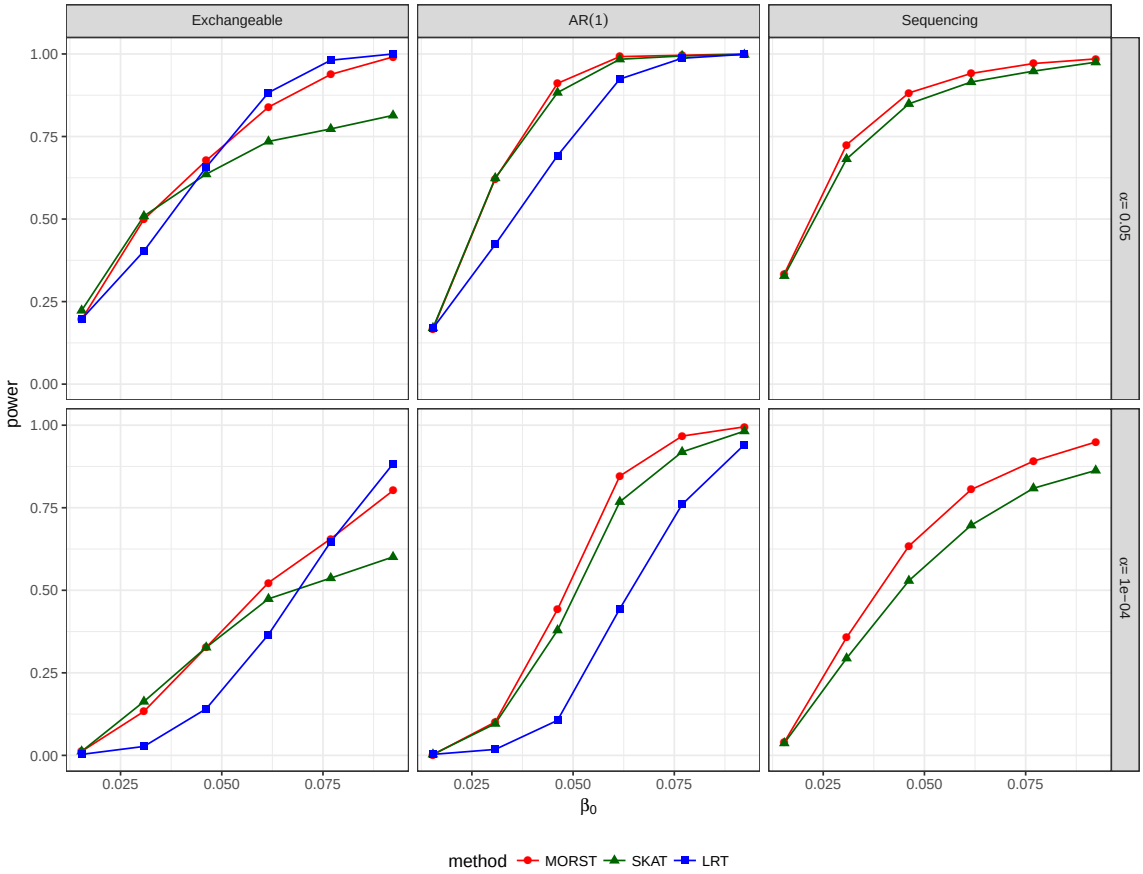

Figure 10: The power of MORST, SKAT and the LRT in logistic regression with $p = 200$ and 20% non-zero β_i's. The columns from left to right correspond to exchangeable correlation, AR(1) correlation and the correlation of sequenced genotypes, respectively. The rows correspond to $\alpha = 0.05, 0.0001$. The x-axis β_0 is the signal strength (i.e., the common magnitude of non-zero β_i's). The LRT is not included in the panels on the right because the MLE does not exist for some genetic variants such as singletons.

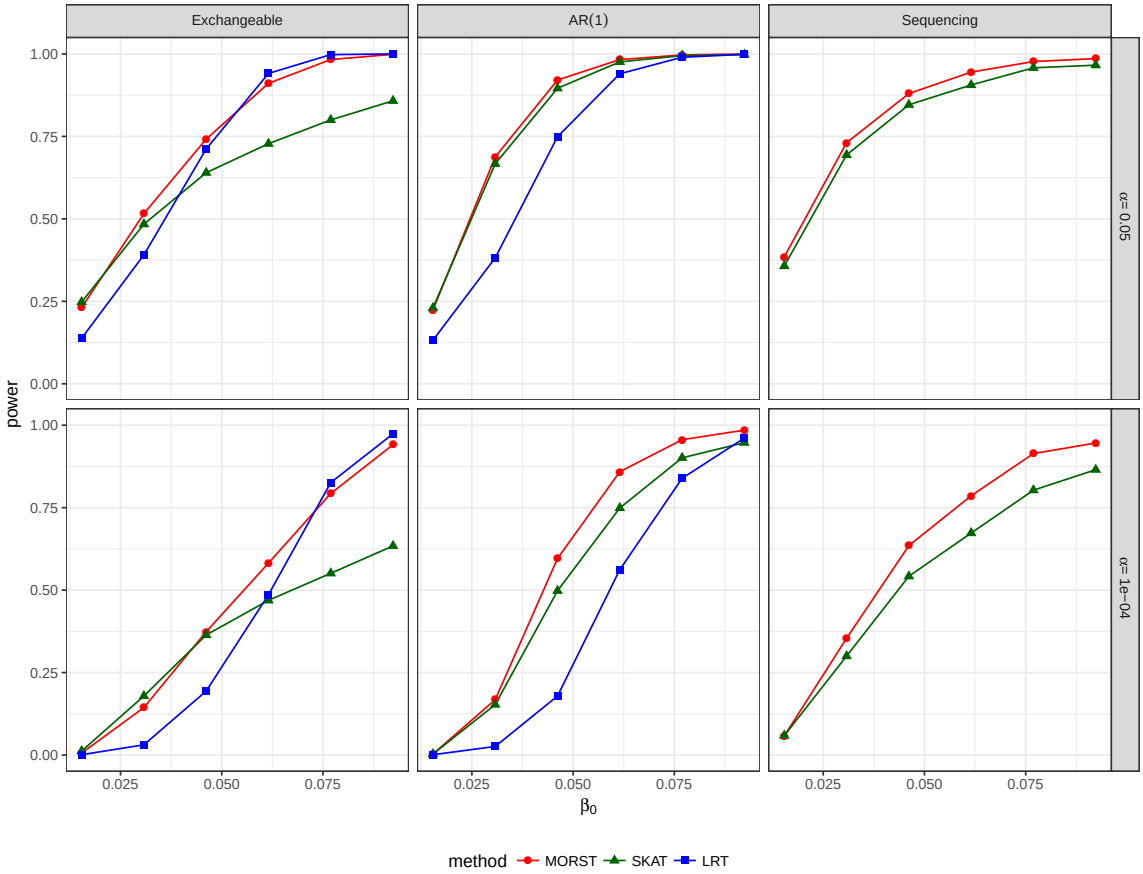

Figure 11: The power of MORST, SKAT and the LRT in logistic regression with $p = 100$ and 40% non-zero β_i's. The columns from left to right correspond to exchangeable correlation, AR(1) correlation and the correlation of sequenced genotypes, respectively. The rows correspond to $\alpha = 0.05, 0.0001$. The x-axis β_0 is the signal strength (i.e., the common magnitude of non-zero β_i's). The LRT is not included in the panels on the right because the MLE does not exist for some genetic variants such as singletons.

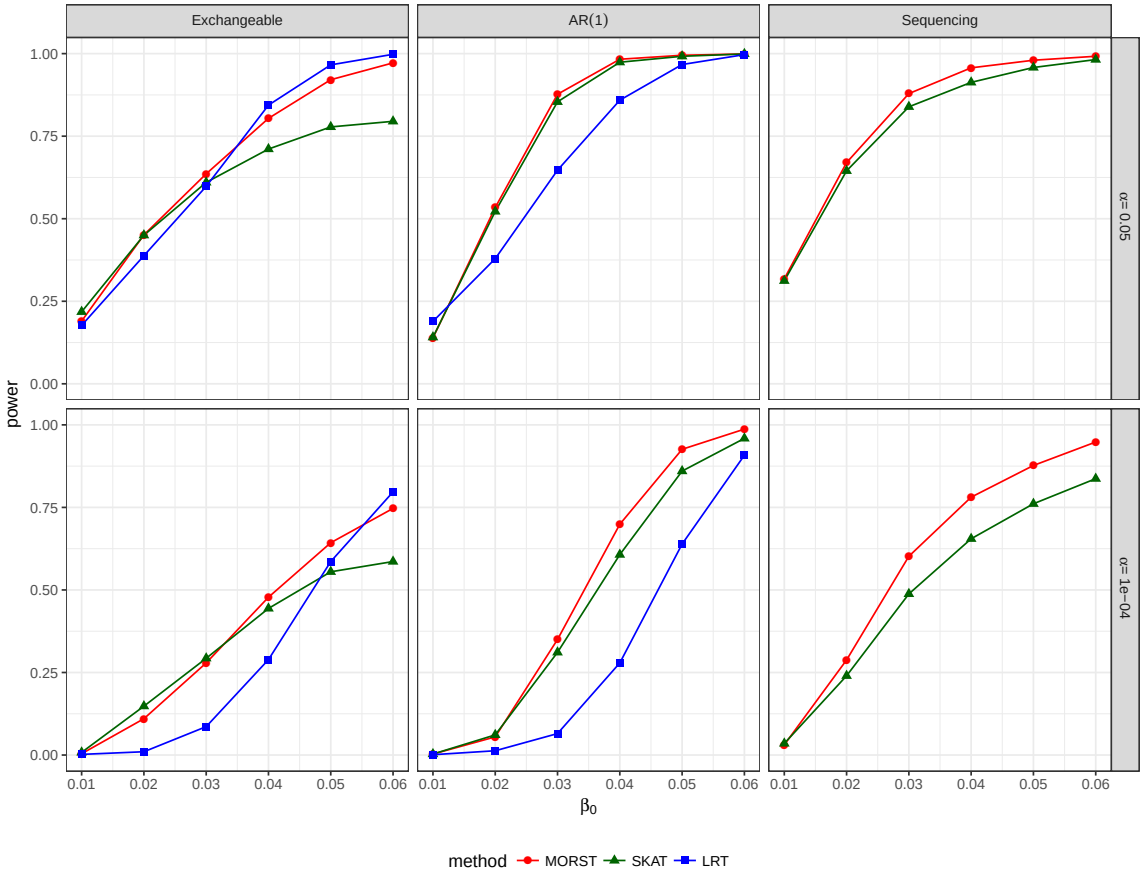

Figure 12: The power of MORST, SKAT and the LRT in logistic regression with $p = 200$ and 40% non-zero β_i's. The columns from left to right correspond to exchangeable correlation, AR(1) correlation and the correlation of sequenced genotypes, respectively. The rows correspond to $\alpha = 0.05, 0.0001$. The x-axis β_0 is the signal strength (i.e., the common magnitude of non-zero β_i's). The LRT is not included in the panels on the right because the MLE does not exist for some genetic variants such as singletons.

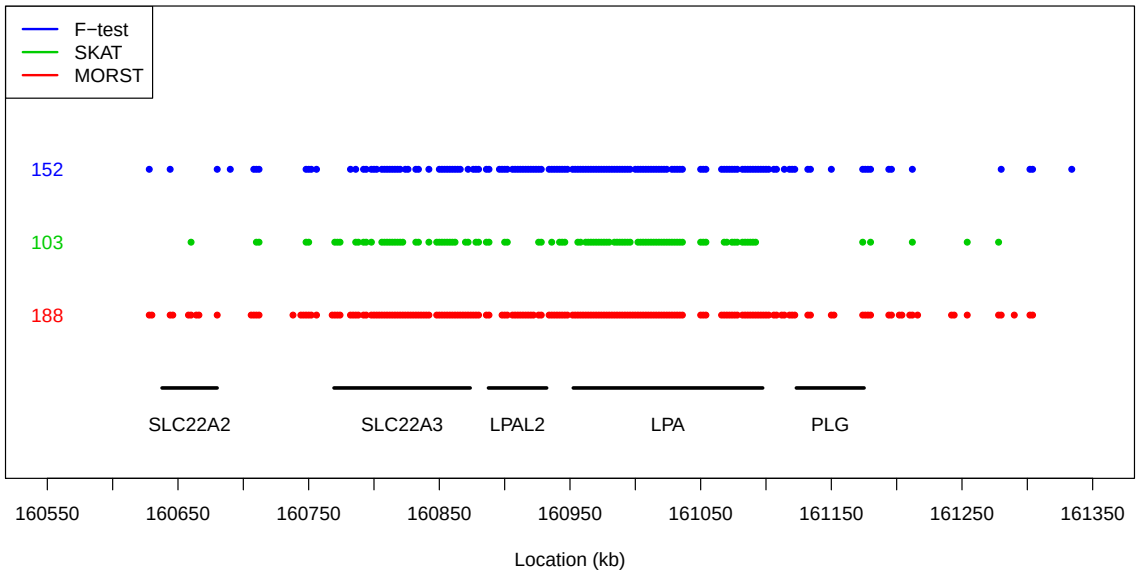

Figure 13: The genomic landscape of significant 4kb sliding windows on chromosome 6 identified by MORST, SKAT and the F test in the analysis of Lipoprotein(a) among AAs. A dot means that the sliding window at this location is significant by the method that the color of the dot represents. The numbers on the left of the plot show the number of significant windows identified by each method.

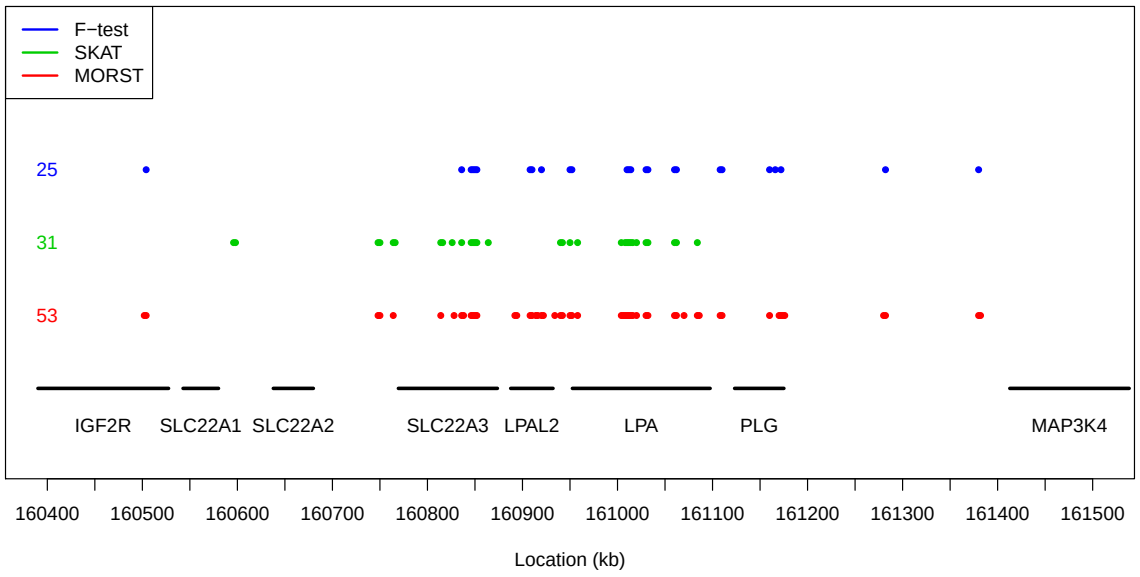

Figure 14: The genomic landscape of significant 4kb sliding windows on chromosome 6 identified by MORST, SKAT and the F test in the analysis of Lipoprotein(a) among EAs. A dot means that the sliding window at this location is significant by the method that the color of the dot represents. The numbers on the left of the plot show the number of significant windows identified by each method.

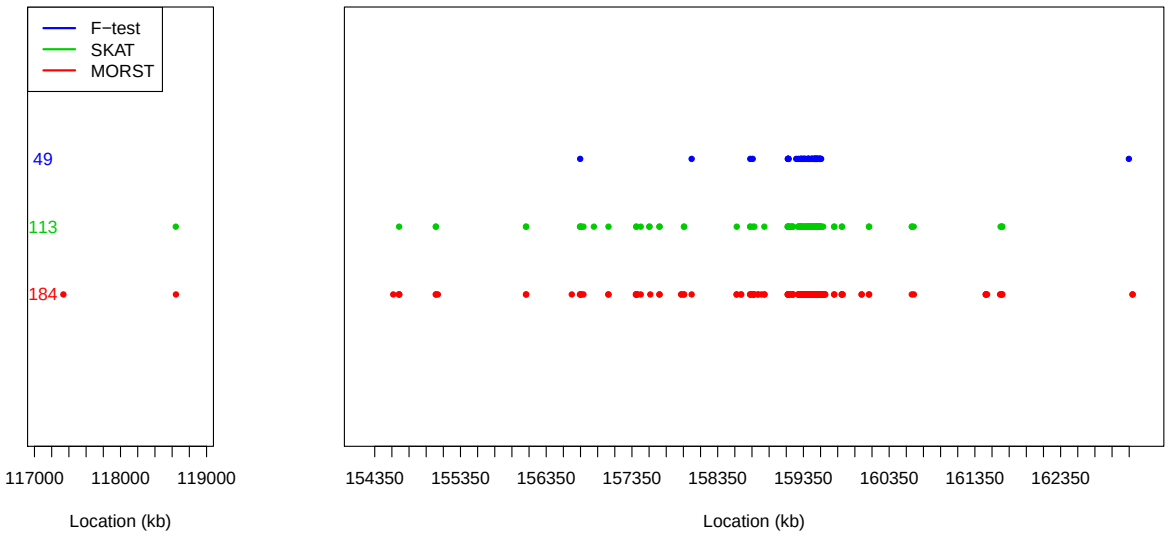

Figure 15: The genomic landscape of significant 4kb sliding windows on chromosome 1 identified by MORST, SKAT and the F test in the analysis of Neutrophil count among AAs. A dot means that the sliding window at this location is significant by the method that the color of the dot represents. The numbers on the left of the plot show the number of significant windows identified by each method.