Supplementary Information: Detecting genetic association through shortest paths in a bi-directed graph

Masao Ueki[*], Yoshinori Kawasaki[†], Gen Tamiya[‡],

and for Alzheimer's Disease Neuroimaging Initiative[§]

November 25, 2016

---

[*] uekimrsd@nifty.com. Biostatistics Center, Kurume University, 67 Asahi-Machi, Kurume, Fukuoka 830-0011, Japan.

[†] kawasaki@ism.ac.jp. The Institute of Statistical Mathematics, The Graduate University for Advanced Studies, 10-3 Midori-Cho, Tachikawa, Tokyo 190-8562, Japan

[‡] gtamiya@genetix-h.com. Tohoku Medical Megabank Organization, Tohoku University, 2-1 Seiryo-Machi, Aoba-Ku, Sendai 980-8573, Japan.

[§] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`adni.loni.usc.edu`). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`

1 Proof of Proposition 1

For notational simplicity, we write $x = x_j$ and $x_k = x_{j_k}$. The vertexes 0, j, and j_k correspond to μ, x, and x_k, respectively. We use induction for the proof. For $l = 1$, if we consider a vector x_1 between y and x,

$$\mu^T Q_{x_1} x = \mu^T x - \mu^T P_{x_1} x = \mu^T x - \frac{(\mu^T x_1)(x^T x_1)}{||x_1||^2},$$

which is zero because $\mu^T x = (\mu^T x_1)(x^T x_1) = 0$, and hence the statement is true. Next, assume that the statement of Proposition 1 is true for any l vectors $x_1, \ldots, x_l$ placed between y and x, where $l > 1$. Then, for any $A \subset \{j_1, \ldots, j_l, j_{l+1}\}$ such that $|A| \in \{0, 1, \ldots, l\}$, $\mu^T Q_{X_A} x = 0$. Therefore the proposition is proved if $\mu^T Q_{(x_1,\ldots,x_l,x_{l+1})} x = 0$ holds. First, if $x_{l+1} \subset \text{span}(x_1, \ldots, x_l)$, then $\mu^T Q_{(x_1,\ldots,x_l,x_{l+1})} x = \mu^T Q_{(x_1,\ldots,x_l)} x = 0$. Next, we consider the case where $x_{l+1} \not\subset \text{span}(x_1, \ldots, x_l)$, and hence, $Q_{(x_1,\ldots,x_l)} x_{l+1} \neq 0$. For such x_{l+1}, adopting the decomposition $P_{(x_1,\ldots,x_l,x_{l+1})} = P_{(x_1,\ldots,x_l)} + P_{Q_{(x_1,\ldots,x_l)} x_{l+1}}$ (Lemma 3, Takane and Yanai 1999) and the assumption of induction, we have

$$\begin{aligned} \mu^T Q_{(x_1,\ldots,x_l,x_{l+1})} x &= -\mu^T P_{Q_{(x_1,\ldots,x_l)} x_{l+1}} x \\ &= -\frac{(\mu^T Q_{(x_1,\ldots,x_l)} x_{l+1})(x^T Q_{(x_1,\ldots,x_l)} x_{l+1})}{||Q_{(x_1,\ldots,x_l)} x_{l+1}||^2}. \end{aligned}$$

If we assume that the left-hand side is nonzero, then both $\mu^T Q_{(x_1,\ldots,x_l)} x_{l+1} \neq 0$ and $x^T Q_{(x_1,\ldots,x_l)} x_{l+1} \neq 0$ must be true. Reversing the assumption of induction, there is then a path between vertexes 0 and j_{l+1} directly (i.e. adjacent) or with inner vertexes being a subset of $\{j_1, \ldots, j_l\}$, and there is a path between vertexes j and j_{l+1} directly or with inner vertexes being a subset of $\{j_1, \ldots, j_l\}$. This means that there should be a path between vertexes j and 0 with inner vertexes comprising a subset of $\{j_1, \ldots, j_l, j_{l+1}\}$, which contradicts the assumption of induction that there is no path between vertexes j and 0 with inner vertexes of any subsets from $\{j_1, \ldots, j_l, j_{l+1}\}$. Therefore, $\mu^T Q_{(x_1,\ldots,x_l,x_{l+1})} x = 0$.

2 Proof of Proposition 2

We use induction for the proof, where we continue to use the simplified notation as in the proof of Proposition 1. For $l = 1$, namely, the length of the shortest path being 1, μ and x are not adjacent — $\mu^T x = 0$. If the shortest path between μ and x has an inner vertex x_1,

$$\mu^T Q_{x_1} x = -\frac{(\mu^T x_1)(x^T x_1)}{||x_1||^2}.$$

The right-hand side is nonzero due to the assumptions $x^T x_1 \neq 0$ and $\mu^T x_1 \neq 0$, proving the claim that $\mu^T Q_{x_1} x \neq 0$.

Next, we assume that the statement of Proposition 2 is true for $l > 1$. Suppose then that there exists at least one path between vertexes 0 and j, and that the shortest path among them is the path with inner vertexes

$j_1 \leftrightarrow \cdots \leftrightarrow j_l \leftrightarrow j_{l+1}$. The length is $l+1$. Then, because there are no paths between vertexes 0 and j whose length is less than $l+1$, Proposition 1 implies that, for any $B \subset \{j_1, \ldots, j_l, j_{l+1}\}$ with $|B| \leq l$, $\mu^T Q_{X_B} x = 0$.

Because there is no path between vertexes j and 0 with inner vertexes $\{j_1, \ldots, j_l\}$, it follows from decomposition of the projection (Takane and Yanai 1999) used in the proof of Proposition 1 that

$$\mu^T Q_{(x_1,\ldots,x_l,x_{l+1})} x = -\frac{(\mu^T Q_{(x_1,\ldots,x_l)} x_{l+1})(x^T Q_{(x_1,\ldots,x_l)} x_{l+1})}{||Q_{(x_1,\ldots,x_l)} x_{l+1}||^2}. \tag{1}$$

The shortest path between vertexes 0 and j with inner vertexes $j_1 \leftrightarrow \cdots \leftrightarrow j_l \leftrightarrow j_{l+1}$ can be decomposed into two shortest paths: the path between vertexes 0 and j_{l+1}, with inner vertexes $j_1 \leftrightarrow \cdots \leftrightarrow j_l$, and the path between j_{l+1} and j, with length zero (i.e. adjacent). Applying the assumption of induction to the former path, of which the length is l, we have $\mu^T Q_{(x_1,\ldots,x_l)} x_{l+1} \neq 0$.

Next, if there exists $a \in \{1, \ldots, l\}$ such that $x^T x_a \neq 0$, it implies the existence of a path between vertexes 0 and j with inner vertexes $j_1 \leftrightarrow \cdots \leftrightarrow j_{a-1} \leftrightarrow j_a$. Its length is $a < l+1$. This contradicts the assumption of induction because the path between vertexes 0 and j with inner vertexes $j_1 \leftrightarrow \cdots \leftrightarrow j_l \leftrightarrow j_{l+1}$, whose length is $l+1$, is assumed to be the shortest. Consequently, $x^T x_a = 0$ must hold for $a = 1, \ldots, l$, or in vector form $X_l^T x = 0$, where $X_l = (x_1, \ldots, x_l)$. Then, the residual $Q_{X_l} x = x - P_{X_l} x = x$, leading to $x^T Q_{X_l} x_{l+1} = x^T x_{l+1}$. This is nonzero as there is a path between j and j_{l+1} with length zero (i.e. adjacent). Therefore, the numerator of (1) is nonzero.

Finally, the denominator of (1), $||Q_{X_l} x_{l+1}||^2$, is no greater than $||x_{l+1}||^2$, which is never infinite. As a consequence, $\mu^T Q_{(x_1,\ldots,x_l,x_{l+1})} x \neq 0$, completing the proof.

3 Conservative family-wise error rate control

Let $M = \{1, \ldots, p\}$. We denote the sequence of indexes $j_1 \leftrightarrow \cdots \leftrightarrow j_L$ by $(j_1, \ldots, j_L)$. For $s = (j_1, \ldots, j_L)$, define $|s| = L$. Further, let $t(s) = j_L$, which extracts the last index, and let $s_- = (j_1, \ldots, j_{L-1})$, i.e. the sequence where the last index is removed from s. Let

$$
\begin{aligned}
A(j) = \{(j_1, \ldots, j_L) :& \\
& \text{the path between } j_L \text{ and } j,\ y \leftrightarrow j \leftrightarrow j_1 \leftrightarrow \cdots \leftrightarrow j_L, \\
& \text{is the shortest}, j_l \in M \setminus \{j\}\}.
\end{aligned}
$$

For $k \in M$ and $e \subset M$, let $T_{k|e}$ denote a test statistic for testing the null hypothesis $\beta_k = 0$ (β_k is the regression coefficient of X_k in the multiple regression model $\theta = \beta_0 + X_k\beta_k + X_e\beta_e$), in which θ is a conditional mean transformed by a link function in a generalized linear model with log-likelihood function $\ell(\theta)$. Denote the first and second derivatives of $\ell(\theta)$ with respect to θ by $\dot{\ell}(\theta)$ and $\ddot{\ell}(\theta)$, respectively. For n subjects, the log-likelihood function is given by $\sum_{i=1}^n \ell(\theta_i)$, where θ_i is the regression model θ for subject i as mentioned above. If $e = \emptyset$, $T_{k|e}$ reduces to the test statistic for marginal association, which is denoted by T_k. Under the null hypothesis of no association of any SNPs with phenotype, it can be shown that $T_{k|e} \sim N(0,1)$ for any $l \in e$. Thus, the p-value is represented by $1 - F_{\chi_1^2}(T_{k|e}^2)$, where $F_{\chi_1^2}$ is the cumulative distribution function of a 1 df χ^2-distribution. Furthermore, $T_{k|e}$ is asymptotically independent of T_l. To see this, let $X = (X_1, X_2)$, let $\widetilde{\beta}_1$ be the maximum likelihood estimator in the model $X_1\beta_1$, and let $\widehat{\beta}_2$ be the maximum likelihood estimator of β_2 in the model $X_1\beta_1 + X_2\beta_2$. If both models are correct, it follows from standard asymptotic theory for generalized linear models (Dobson 2002) that

$$\begin{aligned}\widetilde{\beta}_1 - \beta_1 &\approx (X_1^T W X_1)^{-1} X_1^T \dot{\ell} \\ \widehat{\beta}_2 - \beta_2 &\approx E_2^T (X^T W X)^{-1} X^T \dot{\ell},\end{aligned}$$

where β_1 and β_2 are the true regression coefficients, $E_2^T = (O_{p_2 \times p_1}, I_{p_2})$ is a $p \times p_2$ matrix, $O_{p_2 \times p_1}$ is the zero matrix whose size is shown by its subscripts, and $\dot{\ell} = (\dot{\ell}(\theta_1), \ldots, \dot{\ell}(\theta_n))^T$ and $W = -\mathrm{diag}\{\ddot{\ell}(\theta_1), \ldots, \ddot{\ell}(\theta_n)\}$ are evaluated at the true parameter. Here, the notation '$\approx$' means equality ignoring terms of order $o_p(n^{-1/2})$. Under the null model, both models are correct and $\beta_1 = 0$,

$\beta_2 = 0$, and $E(\dot{\ell}\dot{\ell}^T) = W$. Consequently,

$$\begin{aligned}
\operatorname{cov}(\widehat{\beta}_2, \widetilde{\beta}_1) &\approx E_2^T (X^T W X)^{-1} X^T W X_1 (X_1^T W X_1)^{-1} \\
&= (O_{p_2 \times p_1}, I_{p_2})(X^T W X)^{-1} X^T W X \begin{pmatrix} (X_1^T W X_1)^{-1} \\ O_{p_2 \times p_1} \end{pmatrix} \\
&= O_{p_2 \times p_1}.
\end{aligned}$$

Hence, $\widetilde{\beta}_1$ and $\widehat{\beta}_2$ are asymptotically independent, and so are the two Wald tests based on these two regression coefficients.

Letting $u = F_{\chi_1^2}^{-1}(1 - \alpha_1)$ and $v_k = F_{\chi_1^2}^{-1}\{1 - \alpha_2/|A(k)|\}$, an upper bound on the family-wise error rate is

$$\begin{aligned}
&P(\exists k \in M, \exists e \in A(k) : T_k^2 > u, T_{t(e)|\{k,e_-\}}^2 > v_k) \\
&\quad \leq \sum_{k \in M} \sum_{e \in A(k)} P(T_k^2 > u, T_{t(e)|\{k,e_-\}}^2 > v_k) \\
&\quad = \sum_{k \in M} \sum_{e \in A(k)} P(T_k^2 > u) P(T_{t(e)|\{k,e_-\}}^2 > v_k) \\
&\quad = \sum_{k \in M} \sum_{e \in A(k)} P\{1 - F_{\chi_1^2}(T_k^2) < \alpha_1\} P[\{1 - F_{\chi_1^2}(T_{t(e)|\{k,e_-\}}^2)\}|A(k)| < \alpha_2] \\
&\quad = \sum_{k \in M} \sum_{e \in A(k)} \alpha_1 \frac{\alpha_2}{|A(k)|} \\
&\quad = p\alpha_1\alpha_2.
\end{aligned}$$

The equality in the third line is due to the asymptotic independence between T_k and $T_{t(e)|\{k,e_-\}}$. Letting α be the nominal family-wise error rate (such as

5%), we can control the family-wise error rate at α by taking $\alpha_2 = \alpha/(\alpha_1 p)$ for any α_1. The expected number of indexes satisfying $T_k^2 > u$ is $p\alpha_1$ if all SNPs have no effect. For example, if we take $\alpha_1 = 10^{-3}$ when $p = 10^6$, the expected number is 1,000, which is the expected cardinality of H in step 2 in Section 2.2 in the main text.

4 Marginal association testing under a bivariate model

Consider two SNPs, SNP$_1$ (A/a) and SNP$_2$ (B/b), whose joint distribution of allele frequencies is as in the following table.

Table 1: Allele frequency in the general population

SNP$_1$ / SNP$_2$	B	b	Total
A	P_{11}	P_{12}	$P_{1\cdot}$
a	P_{21}	P_{22}	$P_{2\cdot}$
Total	$P_{\cdot 1}$	$P_{\cdot 2}$	1

Now we assume that a quantitative phenotype is generated from the following bivariate linear regression model, M_{true} say,

$$y = \beta_0 1 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I),$$

where $x_1 = I(\text{aA}) + 2I(\text{aa})$ and $x_2 = I(\text{bB}) + 2I(\text{bb})$.

The effect of x_1 can be tested under the true model M_{true}; call this Method I. On the other hand, we also consider the situation where the effect of x_1 is

Table 2: Conditional expectation of y given two SNP genotypes for true model M_{true}

SNP_1 / SNP_2	BB	bB	bb
AA	β_0	$\beta_0 + \beta_2$	$\beta_0 + 2\beta_2$
aA	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2$	$\beta_0 + \beta_1 + 2\beta_2$
aa	$\beta_0 + 2\beta_1$	$\beta_0 + 2\beta_1 + \beta_2$	$\beta_0 + 2\beta_1 + 2\beta_2$

tested by a marginal linear regression model, M_1 (i.e. x_2 is omitted),

$$y = \beta_{1,0}1 + \beta_{1,1}x_1 + \epsilon,$$

which is exactly the standard marginal association test in SNP-GWAS; call this Method II.

We compare the statistical power of Methods I and II. The tests we consider are based on the Wald test, in which the corresponding test statistics $\widehat{t}_{1,I}$ and $\widehat{t}_{1,II}$ are given in (4) and (5), respectively. Let $\rho_{12} = \mathrm{cor}(x_1, x_2)$ and $v_j = \mathrm{var}(x_j)$ for $j = 1, 2$. In the above setting, $v_1 = 2P_{1\cdot}P_{2\cdot}$ and $v_2 = 2P_{\cdot 1}P_{\cdot 2}$. Further, assuming Hardy–Weinberg equilibrium, $\rho_{12} = r$, where r is the Pearson correlation between alleles at SNP_1 and SNP_2 (Wellek and Ziegler 2009). Let the scaled predictor be $x'_j = x_1/\sqrt{v_j}$, giving the corresponding regression coefficient $\beta'_j = \beta_j\sqrt{v_j}$. We assume that $\sigma^2 \gg |\beta'_j|$ for typical GWAS data. Then,

$$E\widehat{t}_{1,I} \approx \frac{\sqrt{n}}{\sigma}\sqrt{1 - \rho_{12}^2}\beta'_1 \tag{2}$$

$$E\widehat{t}_{1,II} \approx \frac{\sqrt{n}}{\sigma}(\beta'_1 + \beta'_2\rho_{12}). \tag{3}$$

The derivation is given in the following subsection. The above equations imply that Methods I and II have equivalent power when $\beta_2 = 0$ or when $\rho_{12} = 0$, because then $\widehat{t}_{1,I} \approx \widehat{t}_{1,II}$. The condition $\beta_2 = 0$ implies that x_2 has no effect on phenotype. The condition $\rho_{12} = 0$ implies that x_1 and x_2 are uncorrelated, or not in LD. Hence, even though the underlying model is the bivariate model M_{true}, the marginal association test under model M_1 works when $\beta_2 = 0$ or $\rho_{12} = 0$. Next, we consider the case where $\beta_2 = 0$ and $\rho_{12} \neq 0$, which is more common in SNP-GWAS. Then, we have from (2) and (3) that $E\widehat{t}_{1,I} \approx \frac{\sqrt{n}}{\sigma}\sqrt{1-\rho_{12}^2}\beta_1'$ and $E\widehat{t}_{1,II} \approx \frac{\sqrt{n}}{\sigma}\beta_1'$. Thus, $\widehat{t}_{1,I}^2 < \widehat{t}_{1,II}^2$ due to $\rho_{12} \neq 0$, implying that Method II outperforms Method I. This is also supported by our power simulation scenario 1.

If $\beta_1' \approx -\beta_2'\rho_{12}$, then no marginal effect of x_1 appears, because $E\widehat{t}_{1,II} \approx 0$. Given that $\beta_1' \approx -\beta_2'\rho_{12}$, $\beta_1' \neq 0$ implies that both $\beta_2' \neq 0$ and $\rho_{12} \neq 0$. In this case, Method I outperforms Method II if $\rho_{12}^2 < 1$, because $E\widehat{t}_{1,I}$ is approximately nonzero. The above argument implies that, for the effect of SNP$_1$ ($\beta_1 \neq 0$) to be hidden by SNP$_2$, the effect of SNP$_2$ (β_2) is nonzero and the two SNPs are in LD ($\rho_{12} \neq 0$). Although x_1 has no marginal effect, x_2 has a nonzero marginal effect because

$$\beta_2' + \beta_1'\rho_{12} \approx \beta_2'(1-\rho_{12}^2) \neq 0.$$

Finally, we describe particular situations where Method I outperforms Method II. First, suppose that $b = \beta_1' = \beta_2' > 0$. Then, $\widehat{t}_{1,I}^2 > \widehat{t}_{1,II}^2$ if

$(1-\rho_{12}^2)/(1+\rho_{12}) > 1$, or

$$\rho_{12} < 0.$$

Hence, when x_1 and x_2 have the same effect, the negative correlation leads to higher power of Method I compared to that of Method II. Next, suppose that $b = \beta_1' = -\beta_2' > 0$. Then, $\widehat{t}_{1,I}^2 > \widehat{t}_{1,II}^2$ if $(1-\rho_{12}^2)/(1-\rho_{12}) > 1$, or

$$\rho_{12} > 0.$$

Hence, when x_1 and x_2 have opposite effects, the positive correlation leads to higher power of Method I compared to that of Method II.

4.1 Derivation of (2) and (3)

Let $\widetilde{x}_1 = Q_1 x_1$ and $\widetilde{x}_2 = Q_1 x_2$. The least squares estimator of β_1 and its variance estimator under M_{true} are given by

$$\widehat{\beta}_{1,1} = (Q_{\widetilde{x}_2}\widetilde{x}_1)^T y / ||Q_{\widetilde{x}_2}\widetilde{x}_1||^2 \quad \text{and} \quad \widehat{\sigma}_I^2 / ||Q_{\widetilde{x}_2}\widetilde{x}_1||^2,$$

where $\widehat{\sigma}_I^2 = ||Q_{(\widetilde{x}_1,\widetilde{x}_2)} y||^2/n$, and the t statistic is given by

$$\widehat{t}_{1,I} = \widehat{\sigma}_I^{-1} (Q_{\widetilde{x}_2}\widetilde{x}_1)^T y / ||Q_{\widetilde{x}_2}\widetilde{x}_1||. \tag{4}$$

On the other hand, the least squares estimator of $\beta_{1,1}$ and its variance estimator under M_1 are given by

$$\widehat{\beta}_{1,1} = \widetilde{x}_1^T y / ||\widetilde{x}_1||^2 \quad \text{and} \quad \widehat{\sigma}_{II}^2 / ||\widetilde{x}_1||^2,$$

where $\widehat{\sigma}_{II}^2 = ||Q_{\widetilde{x}_1} y||^2 / n$, and the t statistic is given by

$$\widehat{t}_{1,II} = \widehat{\sigma}_{II}^{-1} \widetilde{x}_1^T y / ||\widetilde{x}_1||. \tag{5}$$

Both $\widehat{t}_{1,I}$ and $\widehat{t}_{1,II}$ are designed to be asymptotically $N(0,1)$ under the null hypothesis. Therefore, for a given nominal level α, the effect of x_1 is detected if

$$\widehat{t}_{1,I}^2 > q$$

with Method I, and if

$$\widehat{t}_{1,II}^2 > q$$

with Method II. Here q is the $(1-\alpha)$th quantile of the χ_1^2 distribution. Consequently, statistical power is determined by the noncentrality parameters $E\widehat{t}_{1,I}$ and $E\widehat{t}_{1,II}$ for Methods I and II, respectively.

We now assess behavior of the above t statistics as $n \to \infty$. For large n,

$$E\{(Q_{\widetilde{x}_2} \widetilde{x}_1)^T y\}/n = \beta_1 ||Q_{\widetilde{x}_2} \widetilde{x}_1||^2 / n,$$

and

$$\begin{aligned}||Q_{\widetilde{x}_2}\widetilde{x}_1||^2/n &\approx ||\widetilde{x}_1||^2/n - (\widetilde{x}_1^T\widetilde{x}_2/n)^2/(||\widetilde{x}_2||^2/n)\\ &\approx v_1 - \mathrm{cov}(x_1,x_2)^2/v_2\\ &= v_1(1-\rho_{12}^2).\end{aligned}$$

Also, $\widehat{\sigma}_I^2 \approx \sigma^2$. Therefore, under the true model M_{true},

$$E\widehat{t_{1,I}} \approx \frac{\beta_1||Q_{\widetilde{x}_2}\widetilde{x}_1||}{\sigma} \approx \frac{\sqrt{nv_1(1-\rho_{12}^2)}\beta_1}{\sigma}. \tag{6}$$

On the other hand, for large n,

$$\begin{aligned}E(\widetilde{x}_1^T y)/n &= \widetilde{x}_1^T(\beta_1 x_1 + \beta_2 x_2)/n\\ &= (\beta_1||\widetilde{x}_1||^2 + \beta_2\widetilde{x}_1^T\widetilde{x}_2)/n\\ &\approx \beta_1 v_1 + \beta_2\mathrm{cov}(x_1,x_2)\\ &= \sqrt{v_1}(\sqrt{v_1}\beta_1 + \beta_2\sqrt{v_2}\rho_{12}).\end{aligned}$$

Also,

$$\begin{aligned}E(\widehat{\sigma}_{II}^2) &= \sigma^2 + \beta_2^2||Q_{\widetilde{x}_1}\widetilde{x}_2||^2/n\\ &\approx \sigma^2 + \beta_2^2 v_2(1-\rho_{12}^2).\end{aligned}$$

Thus,

$$E\widehat{t}_{1,II} \approx \frac{\sqrt{n}(\sqrt{v_1}\beta_1 + \beta_2\sqrt{v_2}\rho_{12})}{\sqrt{\sigma^2 + \beta_2^2 v_2(1-\rho_{12}^2)}}. \tag{7}$$

5 Additional tables

		Quantitative				Binary			
$a_{\min}$	α_1	MePL	MaPL	MeNP	MaNP	MePL	MaPL	MeNP	MaNP
0.1	0.01	4	9	369	10341	4	9	364	12051
0.1	0.001	4	7	354	2757	4	7	354	3082
0.1	0.0001	3	4	309	544	3	5	309	646
0.1	0.00001	1	1	84	93	1	1	76	85
0.3	0.01	6	12	344	27020	6	12	304	24464
0.3	0.001	5	9	434	7643	5	9	289	4396
0.3	0.0001	4	5	253	730	4	5	210	511
0.3	0.00001	1	1	42	62	1	1	160	199
0.5	0.01	4	8	45	1498	4	9	47	2293
0.5	0.001	4	6	48	489	4	6	46	544
0.5	0.0001	3	4	33	68	3	4	46	95
0.5	0.00001	1	1	11	13	1	1	15	26
0.7	0.01	4	7	12	441	4	7	12	455
0.7	0.001	3	5	11	117	3	5	12	134
0.7	0.0001	3	3	10	37	3	3	8	23
0.7	0.00001	0	0	1	2	0	0	2	6

Table 3: Summary of the shortest paths in all bi-directed graphs in our proposed method from quantitative- and binary-phenotype type I error simulations with 16 combinations of $(a_{\min}, \alpha_1) \in \{0.1, 0.3, 0.5, 0.7\} \times \{0.01, 0.001, 0.0001, 0.00001\}$. MePL, mean shortest path length; MaPL, maximum shortest path length; MeNP, mean number of shortest paths; MaNP, maximum number of shortest paths.

Scenario		1		2		3		4		5		6	
$a_{\min}$	α_1	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5
0.1	0.01	3	3	3	3	4	3	3	3	3	3	4	4
0.1	0.001	3	3	3	3	3	3	3	3	3	3	2	3
0.1	0.0001	2	3	2	3	1	3	2	3	1	3	1	2
0.1	0.00001	1	3	1	3	1	2	1	3	1	2	0	1
0.3	0.01	5	5	5	5	5	5	5	5	5	5	5	6
0.3	0.001	4	5	4	5	4	5	4	5	4	5	3	5
0.3	0.0001	3	5	3	5	2	4	3	5	2	4	1	3
0.3	0.00001	2	4	1	4	1	3	2	5	1	3	0	2
0.5	0.01	4	4	4	4	4	4	4	4	3	4	3	4
0.5	0.001	3	4	3	4	3	4	3	4	3	3	2	3
0.5	0.0001	2	4	2	4	1	3	2	4	1	3	1	2
0.5	0.00001	1	4	1	3	1	3	2	3	0	2	0	1
0.7	0.01	3	3	3	3	3	3	3	3	3	3	3	3
0.7	0.001	2	3	3	3	2	3	3	3	2	3	2	3
0.7	0.0001	2	3	2	3	1	3	2	3	1	3	0	1
0.7	0.00001	1	3	1	3	1	2	1	3	0	1	0	1

Table 4: Mean shortest path length in all bi-directed graphs in our proposed method from quantitative power simulations with 16 combinations of $(a_{\min}, \alpha_1) \in \{0.1, 0.3, 0.5, 0.7\} \times \{0.01, 0.001, 0.0001, 0.00001\}$ under six scenarios. The two columns in each scenario indicate results for $m = 1$ and $m = 1.5$, respectively.

Scenario		1		2		3		4		5		6	
$a_{\min}$	α_1	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5
0.1	0.01	6	6	6	6	6	6	6	6	6	6	6	6
0.1	0.001	4	5	4	5	3	5	4	5	4	5	3	4
0.1	0.0001	2	4	2	4	2	4	3	4	2	4	1	3
0.1	0.00001	2	4	1	4	1	3	2	4	1	3	0	2
0.3	0.01	8	8	8	8	8	8	8	9	8	8	8	8
0.3	0.001	6	7	6	7	5	7	6	7	5	7	4	7
0.3	0.0001	4	7	3	6	2	5	5	7	2	5	2	4
0.3	0.00001	2	6	2	5	1	4	3	7	1	4	1	2
0.5	0.01	6	6	6	6	6	6	6	6	5	6	5	5
0.5	0.001	4	5	4	5	4	5	4	6	3	5	3	4
0.5	0.0001	3	5	3	5	2	4	3	5	1	4	1	3
0.5	0.00001	2	5	1	4	1	3	2	5	1	3	0	2
0.7	0.01	4	5	4	4	4	4	4	5	4	5	4	4
0.7	0.001	3	4	3	4	3	4	3	4	3	4	2	3
0.7	0.0001	2	4	2	4	1	3	2	4	1	3	0	2
0.7	0.00001	1	4	1	3	1	3	1	4	0	2	0	1

Table 5: Maximum shortest path length in all bi-directed graphs in our proposed method from quantitative power simulations with 16 combinations of $(a_{\min}, \alpha_1) \in \{0.1, 0.3, 0.5, 0.7\} \times \{0.01, 0.001, 0.0001, 0.00001\}$ under six scenarios. The two columns in each scenario indicate results for $m = 1$ and $m = 1.5$, respectively.

Scenario		1		2		3		4		5		6	
$a_{\min}$	α_1	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5
0.1	0.01	668	652	723	702	745	732	670	667	591	633	705	604
0.1	0.001	568	685	628	768	647	780	591	652	539	582	413	591
0.1	0.0001	424	720	464	761	437	756	431	663	301	551	229	419
0.1	0.00001	288	713	310	723	273	698	305	672	192	456	53	310
0.3	0.01	1391	1326	1408	1214	1340	1349	1485	1378	1148	1076	944	957
0.3	0.001	1096	1275	1349	1422	1221	1567	1084	1367	626	922	913	1077
0.3	0.0001	914	1410	907	1647	636	1659	884	1410	239	644	253	566
0.3	0.00001	698	1418	459	1694	310	1510	697	1443	70	405	127	472
0.5	0.01	58	58	57	62	51	63	50	55	33	43	24	27
0.5	0.001	63	77	67	80	49	72	46	62	26	35	15	27
0.5	0.0001	46	83	50	79	34	72	39	63	15	28	7	23
0.5	0.00001	48	92	29	77	21	69	34	60	3	19	2	14
0.7	0.01	6	5	5	5	5	6	5	5	7	10	4	4
0.7	0.001	5	6	6	6	4	5	4	5	5	9	2	3
0.7	0.0001	4	7	4	6	3	6	4	5	3	6	0	2
0.7	0.00001	4	8	3	7	2	6	3	5	1	4	0	1

Table 6: Mean number of shortest paths in all bi-directed graphs in our proposed method from quantitative power simulations with 16 combinations of $(a_{\min}, \alpha_1) \in \{0.1, 0.3, 0.5, 0.7\} \times \{0.01, 0.001, 0.0001, 0.00001\}$ under six scenarios. The two columns in each scenario indicate results for $m = 1$ and $m = 1.5$, respectively.

Scenario		1		2		3		4		5		6	
$a_{\min}$	α_1	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5
0.1	0.01	1938	2284	2545	2848	2919	3199	2661	3167	2188	2550	2463	2532
0.1	0.001	875	1186	1122	1761	1122	1835	1004	1425	981	1523	624	1106
0.1	0.0001	600	1137	700	1318	610	1373	658	1193	442	1176	275	625
0.1	0.00001	378	1052	433	1151	396	1094	383	1073	275	792	58	387
0.3	0.01	10773	16026	13154	14446	12913	14325	15026	18962	11583	13034	5821	7019
0.3	0.001	5084	6793	8551	11061	7926	9807	4970	10540	1509	8420	1802	2742
0.3	0.0001	3403	6319	3128	9517	1191	8797	3423	6138	385	1825	495	1066
0.3	0.00001	2542	5700	829	8706	556	7799	1324	5427	105	850	233	730
0.5	0.01	452	537	368	565	282	453	321	447	201	336	106	158
0.5	0.001	272	472	244	406	167	300	191	353	71	174	29	79
0.5	0.0001	136	433	114	318	79	247	110	286	37	113	11	50
0.5	0.00001	91	413	66	252	49	174	74	226	5	53	3	20
0.7	0.01	30	33	31	38	28	38	30	40	42	65	22	29
0.7	0.001	15	25	17	26	13	21	15	25	18	37	4	11
0.7	0.0001	11	23	9	20	7	18	10	20	7	24	1	4
0.7	0.00001	8	20	6	16	4	14	6	18	3	13	0	1

Table 7: Maximum number of shortest paths in all bi-directed graphs in our proposed method from quantitative power simulations with 16 combinations of $(a_{\min}, \alpha_1) \in \{0.1, 0.3, 0.5, 0.7\} \times \{0.01, 0.001, 0.0001, 0.00001\}$ under six scenarios. The two columns in each scenario indicate results for $m = 1$ and $m = 1.5$, respectively.

Scenario		1		2		3		4		5		6	
$a_{\min}$	α_1	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5
0.1	0.01	3	3	3	4	4	4	3	4	4	4	4	4
0.1	0.001	3	3	3	3	3	3	3	3	3	4	3	4
0.1	0.0001	2	3	2	3	2	3	3	3	2	4	2	3
0.1	0.00001	1	3	1	3	1	4	2	3	1	3	1	3
0.3	0.01	5	6	5	5	5	5	5	5	5	5	5	5
0.3	0.001	5	5	5	5	5	5	5	5	4	4	4	5
0.3	0.0001	4	5	3	5	4	5	4	5	2	4	2	5
0.3	0.00001	2	5	2	5	2	5	3	5	1	3	1	5
0.5	0.01	4	4	4	4	4	4	4	4	3	3	4	4
0.5	0.001	3	4	4	4	3	4	4	4	3	3	3	4
0.5	0.0001	3	4	2	4	2	4	3	4	1	3	2	4
0.5	0.00001	2	4	1	4	1	4	2	4	0	2	1	4
0.7	0.01	3	3	3	3	3	4	3	3	4	4	3	3
0.7	0.001	3	3	3	3	3	3	3	3	3	4	2	3
0.7	0.0001	2	3	2	3	2	3	2	3	1	4	0	3
0.7	0.00001	1	3	1	3	1	3	1	3	0	3	0	2

Table 8: Mean shortest path length in all bi-directed graphs in our proposed method from binary power simulations with 16 combinations of $(a_{\min}, \alpha_1) \in \{0.1, 0.3, 0.5, 0.7\} \times \{0.01, 0.001, 0.0001, 0.00001\}$ under six scenarios. The two columns in each scenario indicate results for $m = 1$ and $m = 1.5$, respectively.

Scenario		1		2		3		4		5		6	
$a_{\min}$	α_1	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5
0.1	0.01	6	6	6	6	6	6	6	6	6	6	6	6
0.1	0.001	4	5	4	5	5	6	4	5	4	5	3	5
0.1	0.0001	3	4	3	5	3	5	3	4	2	5	2	4
0.1	0.00001	2	4	2	4	2	5	2	4	1	4	1	4
0.3	0.01	8	9	8	8	8	8	8	8	7	8	8	8
0.3	0.001	6	8	6	7	6	7	7	8	5	7	5	7
0.3	0.0001	5	8	4	6	4	6	5	7	2	6	3	7
0.3	0.00001	3	7	2	6	2	6	4	7	1	5	1	6
0.5	0.01	6	6	6	6	5	6	6	6	5	6	5	6
0.5	0.001	5	6	5	6	4	6	5	6	3	5	4	6
0.5	0.0001	4	6	3	5	3	6	4	6	2	4	2	5
0.5	0.00001	2	5	2	5	1	5	3	5	1	3	1	5
0.7	0.01	4	4	4	5	4	6	4	5	5	6	4	5
0.7	0.001	3	4	4	5	4	5	3	4	4	5	2	4
0.7	0.0001	2	4	2	5	2	4	2	4	1	5	0	3
0.7	0.00001	2	4	1	4	1	4	1	3	0	4	0	2

Table 9: Maximum shortest path length in all bi-directed graphs in our proposed method from binary power simulations with 16 combinations of $(a_{\min}, \alpha_1) \in \{0.1, 0.3, 0.5, 0.7\} \times \{0.01, 0.001, 0.0001, 0.00001\}$ under six scenarios. The two columns in each scenario indicate results for $m = 1$ and $m = 1.5$, respectively.

Scenario		1		2		3		4		5		6	
$a_{\min}$	α_1	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5
0.1	0.01	707	667	918	848	980	828	717	763	873	812	767	778
0.1	0.001	642	702	831	933	1322	935	649	763	861	990	543	811
0.1	0.0001	515	732	594	973	1094	1113	520	741	494	1124	375	801
0.1	0.00001	363	732	366	976	647	1302	372	704	306	1082	132	786
0.3	0.01	1677	2275	1610	1419	1578	1381	1247	1284	1300	811	1035	814
0.3	0.001	2204	1980	2383	1805	2803	1609	1476	1518	480	609	750	741
0.3	0.0001	2144	2117	2203	2115	2946	2025	1413	1655	157	438	398	857
0.3	0.00001	1655	1933	1918	2886	1525	2477	825	1671	56	312	203	874
0.5	0.01	50	56	59	58	40	79	46	41	30	36	24	27
0.5	0.001	60	69	61	74	47	104	59	53	24	33	24	31
0.5	0.0001	51	84	47	79	39	102	50	62	8	29	16	36
0.5	0.00001	39	86	30	88	19	86	48	64	3	21	7	35
0.7	0.01	6	6	9	7	8	7	6	5	12	9	5	6
0.7	0.001	5	5	6	6	7	6	4	4	9	9	2	3
0.7	0.0001	5	6	5	7	7	6	4	4	2	8	1	1
0.7	0.00001	4	7	4	9	4	7	3	3	1	6	0	1

Table 10: Mean number of shortest paths in all bi-directed graphs in our proposed method from binary power simulations with 16 combinations of $(a_{\min}, \alpha_1) \in \{0.1, 0.3, 0.5, 0.7\} \times \{0.01, 0.001, 0.0001, 0.00001\}$ under six scenarios. The two columns in each scenario indicate results for $m = 1$ and $m = 1.5$, respectively.

Scenario		1		2		3		4		5		6	
$a_{\min}$	α_1	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5	1	1.5
0.1	0.01	2399	3234	3122	4411	4641	4637	2991	3743	4044	4359	3216	4587
0.1	0.001	984	1254	1574	3404	2818	3356	1382	2141	1741	3956	769	2816
0.1	0.0001	690	1131	903	2537	1664	3267	711	1552	675	3241	442	1645
0.1	0.00001	456	1066	496	1984	859	3209	456	1219	334	2587	156	1171
0.3	0.01	11189	36501	13347	14343	13241	16601	9531	13540	8523	8740	7330	7978
0.3	0.001	7205	13790	8102	12944	9171	12247	6201	12024	1081	2908	2404	2247
0.3	0.0001	4758	10673	4823	11421	5688	12220	4525	10216	251	1476	579	1734
0.3	0.00001	3001	7010	3390	10525	2748	11380	2169	9152	70	830	295	1455
0.5	0.01	335	485	407	399	301	687	409	329	144	312	114	181
0.5	0.001	256	434	199	357	157	624	290	303	61	176	57	112
0.5	0.0001	170	405	123	304	84	463	185	278	14	107	28	92
0.5	0.00001	113	359	59	258	35	321	113	243	4	59	13	77
0.7	0.01	34	49	48	56	50	54	38	38	63	70	28	43
0.7	0.001	18	29	20	41	21	40	16	25	27	48	7	15
0.7	0.0001	12	25	14	37	13	35	10	21	5	33	1	5
0.7	0.00001	8	20	9	33	6	33	5	19	1	23	0	2

Table 11: Maximum number of shortest paths in all bi-directed graphs in our proposed method from binary power simulations with 16 combinations of $(a_{\min}, \alpha_1) \in \{0.1, 0.3, 0.5, 0.7\} \times \{0.01, 0.001, 0.0001, 0.00001\}$ under six scenarios. The two columns in each scenario indicate results for $m = 1$ and $m = 1.5$, respectively.

Table 12: Significant shortest paths at $a_{\min} = 0.1$ and $\alpha_1 = 0.01$. P_{marg}, marginal association p-value. P_{mult}, multiple regression p-value applied to each path. SNPs in each path are ordered so that the first SNP is adjacent to the phenotype and the last SNP is tested for association.

Significant path	CHR	SNP	BP	P_{marg}	P_{mult}
1	19	rs440277	50053064	3.8e-03	4.5e-01
1	19	rs387976	50070900	8.8e-02	4.1e-02
1	19	rs157580	50087106	5.2e-05	2.0e-11
1	19	rs8106922	50093506	1.3e-04	1.3e-08
2	19	rs440277	50053064	3.8e-03	1.7e-01
2	19	rs387976	50070900	8.8e-02	4.3e-02
2	19	rs157580	50087106	5.2e-05	1.3e-11
2	19	rs405509	50100676	7.7e-04	1.6e-08
3	19	rs157580	50087106	5.2e-05	1.9e-01
3	19	rs2075650	50087459	1.9e-11	2.7e-08
4	19	rs157580	50087106	5.2e-05	9.2e-11
4	19	rs8106922	50093506	1.3e-04	3.9e-10
5	19	rs157580	50087106	5.2e-05	1.2e-10
5	19	rs405509	50100676	7.7e-04	2.5e-09
6	19	rs157580	50087106	5.2e-05	8.2e-01
6	19	rs429358	50103781	9.5e-23	6.7e-19
7	19	rs2075650	50087459	1.9e-11	5.3e-01
7	19	rs429358	50103781	9.5e-23	6.6e-13
8	19	rs8106922	50093506	1.3e-04	3.9e-10
8	19	rs157580	50087106	5.2e-05	9.2e-11
9	19	rs8106922	50093506	1.3e-04	9.7e-02
9	19	rs2075650	50087459	1.9e-11	8.6e-09
10	19	rs8106922	50093506	1.3e-04	6.2e-01
10	19	rs429358	50103781	9.5e-23	1.4e-19
11	19	rs8106922	50093506	1.3e-04	1.2e-01
11	19	rs405509	50100676	7.7e-04	1.6e-02
11	19	rs439401	50106291	1.3e-05	2.3e-09
12	19	rs405509	50100676	7.7e-04	2.5e-09
12	19	rs157580	50087106	5.2e-05	1.2e-10
13	19	rs405509	50100676	7.7e-04	7.8e-01
13	19	rs2075650	50087459	1.9e-11	6.3e-09
14	19	rs405509	50100676	7.7e-04	5.7e-01
14	19	rs439401	50106291	1.3e-05	8.5e-01
14	19	rs429358	50103781	9.5e-23	3.8e-12
15	19	rs405509	50100676	7.7e-04	1.6e-01
15	19	rs8106922	50093506	1.3e-04	1.6e-01
15	19	rs429358	50103781	9.5e-23	6.9e-20
16	19	rs405509	50100676	7.7e-04	3.8e-01
16	19	rs2075650	50087459	1.9e-11	3.6e-01
16	19	rs429358	50103781	9.5e-23	4.8e-13
17	19	rs405509	50100676	7.7e-04	6.6e-01
17	19	rs157580	50087106	5.2e-05	8.8e-01
17	19	rs429358	50103781	9.5e-23	4.4e-11
18	19	rs405509	50100676	7.7e-04	6.1e-08
18	19	rs439401	50106291	1.3e-05	1.2e-09
19	19	rs439401	50106291	1.3e-05	2.6e-02
19	19	rs2075650	50087459	1.9e-11	2.7e-08
20	19	rs439401	50106291	1.3e-05	5.8e-02
20	19	rs157580	50087106	5.2e-05	1.1e-04
20	19	rs8106922	50093506	1.3e-04	1.1e-09
21	19	rs439401	50106291	1.3e-05	1.2e-09
21	19	rs405509	50100676	7.7e-04	6.1e-08
22	19	rs439401	50106291	1.3e-05	8.4e-01
22	19	rs429358	50103781	9.5e-23	1.5e-18

Table 13: Significant shortest paths at $a_{\min} = 0.1$ and $\alpha_1 = 0.001$. P_{marg}, marginal association p-value. P_{mult}, multiple regression p-value applied to each path. SNPs in each path are ordered so that the first SNP is adjacent to the phenotype and the last SNP is tested for association.

Significant path	CHR	SNP	BP	P_{marg}	P_{mult}
1	5	rs2963337	8988632	1.5e-04	4.0e-08
1	5	rs2963345	8962508	3.2e-03	7.0e-03
1	5	rs6870451	8991885	9.3e-02	8.6e-07
2	5	rs2963337	8988632	1.5e-04	3.8e-08
2	5	rs2963345	8962508	3.2e-03	7.9e-03
2	5	rs11955429	8994715	7.8e-02	9.9e-07
3	19	rs157580	50087106	5.2e-05	1.9e-01
3	19	rs2075650	50087459	1.9e-11	2.7e-08
4	19	rs157580	50087106	5.2e-05	9.2e-11
4	19	rs8106922	50093506	1.3e-04	3.9e-10
5	19	rs157580	50087106	5.2e-05	1.2e-10
5	19	rs405509	50100676	7.7e-04	2.5e-09
6	19	rs157580	50087106	5.2e-05	8.2e-01
6	19	rs429358	50103781	9.5e-23	6.7e-19
7	19	rs2075650	50087459	1.9e-11	5.3e-01
7	19	rs429358	50103781	9.5e-23	6.6e-13
8	19	rs8106922	50093506	1.3e-04	3.9e-10
8	19	rs157580	50087106	5.2e-05	9.2e-11
9	19	rs8106922	50093506	1.3e-04	9.7e-02
9	19	rs2075650	50087459	1.9e-11	8.6e-09
10	19	rs8106922	50093506	1.3e-04	6.2e-01
10	19	rs429358	50103781	9.5e-23	1.4e-19
11	19	rs8106922	50093506	1.3e-04	1.2e-01
11	19	rs405509	50100676	7.7e-04	1.6e-02
11	19	rs439401	50106291	1.3e-05	2.3e-09
12	19	rs405509	50100676	7.7e-04	2.5e-09
12	19	rs157580	50087106	5.2e-05	1.2e-10
13	19	rs405509	50100676	7.7e-04	7.8e-01
13	19	rs2075650	50087459	1.9e-11	6.3e-09
14	19	rs405509	50100676	7.7e-04	5.7e-01
14	19	rs439401	50106291	1.3e-05	8.5e-01
14	19	rs429358	50103781	9.5e-23	3.8e-12
15	19	rs405509	50100676	7.7e-04	1.6e-01
15	19	rs8106922	50093506	1.3e-04	1.6e-01
15	19	rs429358	50103781	9.5e-23	6.9e-20
16	19	rs405509	50100676	7.7e-04	3.8e-01
16	19	rs2075650	50087459	1.9e-11	3.6e-01
16	19	rs429358	50103781	9.5e-23	4.8e-13
17	19	rs405509	50100676	7.7e-04	6.6e-01
17	19	rs157580	50087106	5.2e-05	8.8e-01
17	19	rs429358	50103781	9.5e-23	4.4e-11
18	19	rs405509	50100676	7.7e-04	6.1e-08
18	19	rs439401	50106291	1.3e-05	1.2e-09
19	19	rs439401	50106291	1.3e-05	2.6e-02
19	19	rs2075650	50087459	1.9e-11	2.7e-08
20	19	rs439401	50106291	1.3e-05	5.8e-02
20	19	rs157580	50087106	5.2e-05	1.1e-04
20	19	rs8106922	50093506	1.3e-04	1.1e-09
21	19	rs439401	50106291	1.3e-05	1.2e-09
21	19	rs405509	50100676	7.7e-04	6.1e-08
22	19	rs439401	50106291	1.3e-05	8.4e-01
22	19	rs429358	50103781	9.5e-23	1.5e-18

Table 14: Significant shortest paths at $a_{\min} = 0.1$ and $\alpha_1 = 0.0001$. P_{marg}, marginal association p-value. P_{mult}, multiple regression p-value applied to each path. SNPs in each path are ordered so that the first SNP is adjacent to the phenotype and the last SNP is tested for association.

Significant path	CHR	SNP	BP	P_{marg}	P_{mult}
1	19	rs157580	50087106	5.2e-05	1.9e-01
1	19	rs2075650	50087459	1.9e-11	2.7e-08
2	19	rs157580	50087106	5.2e-05	9.2e-11
2	19	rs8106922	50093506	1.3e-04	3.9e-10
3	19	rs157580	50087106	5.2e-05	1.2e-10
3	19	rs405509	50100676	7.7e-04	2.5e-09
4	19	rs157580	50087106	5.2e-05	8.2e-01
4	19	rs429358	50103781	9.5e-23	6.7e-19
5	19	rs2075650	50087459	1.9e-11	5.3e-01
5	19	rs429358	50103781	9.5e-23	6.6e-13
6	19	rs439401	50106291	1.3e-05	2.6e-02
6	19	rs2075650	50087459	1.9e-11	2.7e-08
7	19	rs439401	50106291	1.3e-05	5.8e-02
7	19	rs157580	50087106	5.2e-05	1.1e-04
7	19	rs8106922	50093506	1.3e-04	1.1e-09
8	19	rs439401	50106291	1.3e-05	1.2e-09
8	19	rs405509	50100676	7.7e-04	6.1e-08
9	19	rs439401	50106291	1.3e-05	8.4e-01
9	19	rs429358	50103781	9.5e-23	1.5e-18

Table 15: Significant shortest paths at $a_{\min} = 0.1$ and $\alpha_1 = 0.00001$. P_{marg}, marginal association p-value. P_{mult}, multiple regression p-value applied to each path. SNPs in each path are ordered so that the first SNP is adjacent to the phenotype and the last SNP is tested for association.

Significant path	CHR	SNP	BP	P_{marg}	P_{mult}
1	19	rs2075650	50087459	1.9e-11	1.3e-08
1	19	rs157580	50087106	5.2e-05	1.3e-02
1	19	rs12610605	50062678	2.3e-01	4.7e-04
2	19	rs2075650	50087459	1.9e-11	5.3e-01
2	19	rs429358	50103781	9.5e-23	6.6e-13

Table 16: Significant shortest paths at $a_{\min} = 0.3$ and $\alpha_1 = 0.01$, $a_{\min} = 0.3$ and $\alpha_1 = 0.0001$, and $a_{\min} = 0.5$ and $\alpha_1 = 0.01, 0.001, 0.0001, 0.00001$. P_{marg}, marginal association p-value. P_{mult}, multiple regression p-value applied to each path. SNPs in each path are ordered so that the first SNP is adjacent to the phenotype and the last SNP is tested for association.

Significant path	CHR	SNP	BP	P_{marg}	P_{mult}
1	19	rs2075650	50087459	1.9e-11	5.3e-01
1	19	rs429358	50103781	9.5e-23	6.6e-13

Table 17: 6. Significant shortest paths at $a_{\min} = 0.3$ and $\alpha_1 = 0.001$. P_{marg}, marginal association p-value. P_{mult}, multiple regression p-value applied to each path. SNPs in each path are ordered so that the first SNP is adjacent to the phenotype and the last SNP is tested for association.

Significant path	CHR	SNP	BP	P_{marg}	P_{mult}
1	5	rs2963327	8987159	2.0e-04	8.1e-08
1	5	rs2963345	8962508	3.2e-03	7.1e-03
1	5	rs6870451	8991885	9.3e-02	1.4e-06
2	5	rs2963327	8987159	2.0e-04	7.8e-08
2	5	rs2963344	8962428	3.7e-03	1.1e-02
2	5	rs6870451	8991885	9.3e-02	2.0e-06
3	5	rs2963327	8987159	2.0e-04	7.7e-08
3	5	rs2963345	8962508	3.2e-03	8.0e-03
3	5	rs11955429	8994715	7.8e-02	1.6e-06
4	5	rs2963327	8987159	2.0e-04	7.0e-08
4	5	rs2963344	8962428	3.7e-03	1.3e-02
4	5	rs11955429	8994715	7.8e-02	2.3e-06
5	5	rs2963337	8988632	1.5e-04	4.0e-08
5	5	rs2963345	8962508	3.2e-03	7.0e-03
5	5	rs6870451	8991885	9.3e-02	8.6e-07
6	5	rs2963337	8988632	1.5e-04	3.9e-08
6	5	rs2963344	8962428	3.7e-03	1.1e-02
6	5	rs6870451	8991885	9.3e-02	1.3e-06
7	5	rs2963337	8988632	1.5e-04	3.8e-08
7	5	rs2963345	8962508	3.2e-03	7.9e-03
7	5	rs11955429	8994715	7.8e-02	9.9e-07
8	5	rs2963337	8988632	1.5e-04	3.5e-08
8	5	rs2963344	8962428	3.7e-03	1.2e-02
8	5	rs11955429	8994715	7.8e-02	1.4e-06
9	19	rs2075650	50087459	1.9e-11	5.3e-01
9	19	rs429358	50103781	9.5e-23	6.6e-13

Table 18: Results from lasso. SNPs that were included in the optimal model chosen by AIC. P_{marg}, marginal association p-value. P_{puma}, PUMA p-value.

CHR	SNP	BP	P_{marg}	P_{puma}
1	rs4970843	109688714	5.6e-06	8.3e-05
1	rs6427160	167158805	4.3e-06	6.5e-03
1	rs4434815	244153230	1.4e-05	2.1e-06
2	rs4072374	3564846	3.2e-05	1.3e-04
2	rs10929604	9782998	8.1e-06	4.8e-03
2	rs6751810	19110421	7.1e-05	1.2e-01
2	rs2346567	19126150	6.0e-05	6.1e-01
2	rs4953672	42953942	9.0e-06	1.2e-04
2	rs2883782	171142071	1.5e-05	2.9e-02
2	rs2854711	208720112	3.5e-04	9.6e-04
3	rs7641243	11855760	3.3e-05	1.3e-07
3	rs9821223	36858373	6.8e-04	6.4e-05
3	rs10510985	69746561	2.9e-05	1.2e-02
4	rs6534996	135217173	9.5e-07	3.4e-06
4	rs11726692	142139487	1.2e-06	3.4e-02
5	rs12657303	172075156	7.6e-06	2.1e-05
6	rs2305955	21308261	2.9e-05	7.6e-06
6	rs7767652	55114756	2.8e-05	4.8e-03
7	rs799449	44751222	1.1e-04	2.2e-04
7	rs757839	70213204	5.5e-05	9.9e-05
7	rs1006091	122330026	1.6e-04	1.9e-03
8	rs1155490	69262212	7.3e-05	1.5e-04
8	rs769031	70513720	4.7e-05	7.8e-05
8	rs10110185	131548563	8.1e-04	1.1e-05
9	rs2518112	107327881	6.3e-05	1.9e-04
9	rs4978345	110122092	5.9e-05	6.4e-03
10	rs11598825	7668437	1.3e-04	2.4e-04
11	rs10892831	121763566	2.9e-05	3.9e-03
12	rs7137643	113775994	6.3e-05	7.9e-05
13	rs6491232	27328216	1.1e-04	4.7e-03
14	rs2693693	98745939	6.0e-05	2.3e-04
15	rs11072670	75828844	5.9e-04	9.7e-06
15	rs1077541	78885162	7.4e-05	6.0e-04
18	rs8182439	70026851	1.0e-04	8.7e-03
19	rs4807567	4229502	9.5e-05	1.8e-03
19	rs429358	50103781	9.5e-23	2.5e-25
20	rs6130775	43206582	1.9e-04	7.9e-02
22	rs739049	48094617	4.7e-05	4.1e-03

References

[1] Takane Y, Yanai H (1999) On oblique projectors. *Lin Algeb Appl* **289**, 297–310.

[2] Dobson A (2002) *An Introduction to Generalized Linear Models, 2nd Ed.* Chapman & Hall, London.

[3] Wellek S, Ziegler A (2009) A genotype-based approach to assessing the association between single nucleotide polymorphisms. *Hum Hered* **67**, 128–139.