Supplements for “A method for analyzing multiple continuous phenotypes in rare variant association studies allowing for flexible correlations in variant effects”

by

Jianping Sun, Karim Oualkacha, Vincenzo Forgetta, Hou-Feng Zheng, J. Brent Richards, Antonio Ciampi, Celia M.T. Greenwood, and the UK10K Consortium.

Contents

S.1 Deriving $S_\rho$ and its distribution 2
S.2 Search range for correlation parameter $\rho$ 2
S.3 Approximate $S_\rho$ as the sum of two asymptotically independent random variables 4
S.4 Well-replicated known bone genes in UK10K analysis 6
S.5 Computation time and memory usage 6
S.6 Misspecified $\Sigma_\beta$ 8
S.7 Density of the grid values for $\rho$ 9
S.8 Antagonistic pleiotropy 9
S.9 References 10
S.10 Tables 11
S.11 Figures 14
S.1 Deriving $S_\rho$ and its distribution

For a fixed $\rho$, define $V \triangleq Var(Y) = G^*\Sigma_\rho G^{*T} + \Sigma_\epsilon$, and the log likelihood function for model (2) is $\log f(Y) \propto -\frac{1}{2} \log |V| - \frac{1}{2}(Y - X^*\alpha)^T V^{-1}(Y - X^*\alpha)$. Under the null hypothesis, $H_0 : \tau^2 = 0$, we have the score for $\tau^2$ is $U_{\tau^2} = \frac{\partial}{\partial \tau^2} \log f(Y) \big|_{\tau^2 = 0} = -\frac{1}{2} tr \left( \Sigma^{-1}_e G^* [R \otimes W] G^{*T} \right) + \frac{1}{2}(Y - X^*\alpha)^T \left( \Sigma^{-1}_e G^* [R \otimes W] G^{*T} \Sigma^{-1}_e \right)(Y - X^*\alpha)$. Since the first term in $U_{\tau^2}$ does not depend on $Y$, we can drop the first term and ignore the factor of $\frac{1}{2}$ in the second term to obtain our score statistic,

$$S_\rho = (Y - X^*\hat{\alpha})^T \left( \widehat{\Sigma}_e^{-1} G^* [R \otimes W] G^{*T} \widehat{\Sigma}_e^{-1} \right)(Y - X^*\hat{\alpha}),$$

where $\hat{\alpha} = (X^{*T} \widehat{\Sigma}_e^{-1} X^*)^{-1} X^{*T} \widehat{\Sigma}_e^{-1} Y$ and $\widehat{\Sigma}_e$ are estimated under the null model $Y = X^*\alpha + \epsilon$.

Let $P = I - X^*(X^{*T} \widehat{\Sigma}_e^{-1} X^*)^{-1} X^{*T} \widehat{\Sigma}_e^{-1}$, we have $Y - X^*\hat{\alpha} = PY$ and $PX^* = 0$. Hence, $S_\rho$ can be rewritten as $S_\rho = [P(Y - X^*\alpha)]^T \left( \widehat{\Sigma}_e^{-1} G^* [R \otimes W] G^{*T} \widehat{\Sigma}_e^{-1} \right) [P(Y - X^*\alpha)]$. Further denote $P_0 = \widehat{\Sigma}_e^{-1} P$, we obtain $S_\rho = \tilde{Y}^T \left[ \widehat{\Sigma}_e^{1/2} P_0^T G^* (R \otimes W) G^{*T} P_0 \widehat{\Sigma}_e^{1/2} \right] \tilde{Y}$, where $\tilde{Y} = \widehat{\Sigma}_e^{-1/2}(Y - X^*\alpha)$ asymptotically has a $N(0, I_{KN})$ distribution under the null model. Thus, $S_\rho$ has a mixture of chi-square null distribution, $\sum_{s=1}^{l'} \lambda_s \chi^2_{1,s}$, when $R$ is a symmetric matrix, where $\chi^2_{1,s}$ are independent chi-square random variables with one degree of freedom. The parameters, $\{\lambda_s\}$, are eigenvalues of the matrix, $\widehat{\Sigma}_e^{1/2} P_0 G^* (R \otimes W) G^{*T} P_0 \widehat{\Sigma}_e^{1/2}$.

S.2 Search range for correlation parameter $\rho$

In our proposed MURAT, we search over a grid of values for $\rho \in [0, 1]$ to obtain a minimum $p$ value as the overall test statistic. Mathematically, the range for this grid could be $[-1, 1]$, since $\rho$ is a correlation parameter. We have chosen to search over the range $[0, 1]$ for two reasons. Firstly, intuitively, $\rho$ describes a pleiotropic effect. Without loss of generality, we can assume two traits are positively correlated, otherwise, one could switch the sign on one of the traits prior to analysis with MURAT so they would be positively correlated. Under this assumption, variants that are beneficial for one phenotype and detrimental to another (antagonistic pleiotropy) are uncommon. Also, when
$K \geq 3$, the assumption of a common $\rho$ across variant effects requires that $\rho > 0$.

In addition, we can show that when $\rho = 0$ or when $\rho = 1$, the test statistic for a fixed $\rho$, i.e. $S_\rho$, will degenerate to two test statistics commonly used for multivariate tests.

To see these limiting cases, suppose there are two correlated traits. The underlying model for MURAT with respect to individual $i$ can be written as,

$$
Y_i = \left( \begin{array}{c} y_{i1} \\ y_{i2} \end{array} \right) = \left( \begin{array}{c} X_i^T \alpha_1 \\ X_i^T \alpha_2 \end{array} \right) + \left( \begin{array}{c} g_{i1} \beta_{11} + \cdots + g_{ip} \beta_{1p} \\ g_{i1} \beta_{21} + \cdots + g_{ip} \beta_{2p} \end{array} \right) + \left( \begin{array}{c} \epsilon_{i1} \\ \epsilon_{i2} \end{array} \right) \text{ for } i = 1, \ldots, N. \quad (S1)
$$

If we assume weight matrix $W = I_p$, the test statistic when $\rho$ is fixed is

$$
S_\rho = (Y - X^*\widehat{\alpha})^T (\widehat{\Sigma}_e^{-1} G^* [R \otimes I_p] G^{*T} \widehat{\Sigma}_e^{-1}) (Y - X^*\widehat{\alpha})
$$

$$
= (1 - \rho) (Y - X^*\widehat{\alpha})^T (\widehat{\Sigma}_e^{-1} G^* [I_2 \otimes I_p] G^{*T} \widehat{\Sigma}_e^{-1}) (Y - X^*\widehat{\alpha})
$$

$$
+ \rho (Y - X^*\widehat{\alpha})^T (\widehat{\Sigma}_e^{-1} G^* [(1_2 1_2^T) \otimes I_p] G^{*T} \widehat{\Sigma}_e^{-1}) (Y - X^*\widehat{\alpha}),
$$

where $\rho = Corr(\beta_{1j}, \beta_{2j})$ for $j = 1, \ldots, p$.

Unlike in MURAT, if we assume $\beta_k = (\beta_{k1}, \ldots, \beta_{kp})^T \sim N(0, \tau^2 I_p)$ for $k = 1, 2$, and $\beta_1$ and $\beta_2$ are independent, i.e. $\rho = 0$, then the above model (S1) degenerates to the underlying model for Maity’s method with a linear kernel, and it is also a bivariate SKAT model by simply assuming $\epsilon_{i1}$ and $\epsilon_{i2}$ are correlated. Under this situation, the score test statistic for testing $H_0 : \beta_1 = \beta_2 = 0$ will be $(Y - X^*\widehat{\alpha})^T (\widehat{\Sigma}_e^{-1} G^* [I_2 \otimes I_p] G^{*T} \widehat{\Sigma}_e^{-1}) (Y - X^*\widehat{\alpha})$. Hence, when $\rho = 0$, $S_\rho$ is the same as the test statistic for Maity’s method with linear kernel or multivariate SKAT.

On the other hand, when $\rho = 1$, without loss of generality, we can assume $\beta_{1j} = \beta_{2j} = \beta_j$. By further assuming $\beta_1, \ldots, \beta_p$ are fixed, the model (S1) is equivalent to a multivariate fixed effect model, where we assume the same effect of each variant across the multiple traits. The score statistic for testing $H_0 : \beta_1 = \cdots = \beta_p = 0$ under this fixed effect model is $U_\beta \propto [G^* (1_2 \otimes I_p)]^T \widehat{\Sigma}_e^{-1} (Y - X^*\widehat{\alpha})$, and the square of $U_\beta$, i.e. $U_\beta^T U_\beta$, is $(Y - X^*\widehat{\alpha})^T \widehat{\Sigma}_e^{-1} G^* [(1_2 1_2^T) \otimes I_p] G^{*T} \widehat{\Sigma}_e^{-1} (Y - X^*\widehat{\alpha})$, which equals $S_\rho$ when $\rho = 1$. Further investigation also shows that if there
are no covariates in this multivariate fixed effect model, then \( U_\beta^T U_\beta \) is equivalent to the Lawley-Hotelling statistic used in the hypothesis testing in multivariate multiple regression\(^1\).

The above derivation can be easily generalized to any number of \( K \) correlated traits. Therefore, by searching a grid of \( \rho \) between 0 and 1, our proposed MURAT covers score tests between Maity’s method (or multivariate SKAT) and multivariate fixed effect model.

**S.3 Approximate \( S_\rho \) as the sum of two asymptotically independent random variables**

Denote \( X = (X_1, \ldots, X_N)^T \) and \( G = (G_1, \ldots, G_N)^T \), we have \( S_\rho = \tilde{Y}^T \left\{ \left[ (I - P_x)(GWG^T)(I - P_x) \right] \otimes \left[ \hat{\Sigma}^{-1/2} R \hat{\Sigma}^{-1/2} \right] \right\} \tilde{Y} \), where \( P_x = X(X^T X)^{-1} X^T \) represents the projection matrix of \( X \).

For simplicity, we define the following notations: \( Z = \hat{\Sigma}^{-1/2} Z = Z \cdot 1_K / K \), \( M = \tilde{z}(\tilde{z}^T \tilde{z})^{-1} \tilde{z}^T \), and \( Q = (I_n - P_x)(GWG^T)(I_n - P_x) \). By plugging in \( R = (1 - \rho)I_K + \rho 1_K 1_K^T \), we have

\[
S_\rho = (1 - \rho) \tilde{Y}^T \left\{ Q \otimes [(I - M)ZZ(I - M)] \right\} \tilde{Y} + 2(1 - \rho) \tilde{Y}^T \left\{ Q \otimes [(I - M)ZZM] \right\} \tilde{Y} + \tau(\rho) \tilde{Y}^T \left\{ Q \otimes [\tilde{z}\tilde{z}^T] \right\} \tilde{Y},
\]

where \( \tau(\rho) = \frac{1 - \rho}{(\tilde{z}^T \tilde{z})} \sum_{j=1}^K (\tilde{z}^T Z_j)^2 + K^2 \rho \) is a constant only depending on \( \rho \), and \( Z_j \) is the jth column of matrix \( Z \).

Define random variables \( \xi = \tilde{Y}^T \left\{ Q \otimes [(I - M)ZZ(I - M)] \right\} \tilde{Y}, \quad \zeta = 2 \tilde{Y}^T \left\{ Q \otimes [(I - M)ZZM] \right\} \tilde{Y} \), and \( \eta = \tilde{Y}^T \left\{ Q \otimes [\tilde{z}\tilde{z}^T] \right\} \tilde{Y} \). Since \( \tilde{Y} \) asymptotically has a \( N(0, I_{K_N}) \) distribution under the null model, we have the following asymptotical distributions: \( \xi \sim \sum_{r=1}^8 \lambda_r \chi^2_{1,r} \), where \( \lambda_r \)'s are the eigenvalues of \( Q \otimes [(I - M)ZZ(I - M)] \); \( \eta \sim \sum_{i=1}^J \lambda'_i \chi^2_{1,i} \), where \( \lambda'_i \)'s are the eigenvalues of \( Q \otimes [\tilde{z}\tilde{z}^T] \). In addition, because \( (I - M)\tilde{z}\tilde{z}^T = 0 \), we have \( \xi \) and \( \eta \) are asymptotically independent under the null model by Craig’s Theorem.

Noticing the fact that if a random vector \( x \) follows a multivariate normal distribution with mean
µ and covariance matrix Σ, then for a symmetric matrix Λ, we have \( E(x^T Λ x) = \text{tr}(Λ \Sigma) + µ^T Λ µ \) and \( \text{Var}(x^T Λ x) = 2 \text{tr}(Λ Σ Λ Σ) + 4µ^T Λ Σ Λ µ). \) In addition, for two symmetric matrices, \( Λ_1 \) and \( Λ_2 \), we also have \( \text{Cov}(x^T Λ_1 x, x^T Λ_2 x) = 2 \text{tr}(Λ_1 Σ Λ_2 Σ) + 4µ^T Λ_1 Σ Λ_2 µ). \) In addition, since \( x^T Λ x \) and \( x^T Λ^T x \) are both scalars, we have \( x^T Λ x = x^T Λ^T x \), no matter whether \( Λ \) is symmetric or not. Hence, the above conclusions still hold when \( Λ \) is not symmetric, because we can have an equivalent quadratic form replacing \( Λ \) by \( \tilde{Λ} = \frac{1}{2}(Λ + Λ^T) \), which is symmetric. Therefore, we obtain the following conclusions: \( E(ζ) = 0, \text{Var}(ζ) = 4 \text{tr}[(GWG^T)Q] \cdot \text{tr}[ZMZZ(I − M)Z] \) asymptotically, and \( ζ \) is asymptotically uncorrelated with \( ξ \) and \( η \), i.e. \( \text{Cov}(ζ, ξ) = \text{Cov}(ζ, η) = 0 \).

Define \( κ = ξ + ζ \), and we obtain \( S_ρ = (1 − ρ)κ + τ(ρ)η \). According to the above discussion, \( ξ \) is asymptotically independent from \( η \); \( ζ \) is asymptotically uncorrelated with both \( η \) and \( ζ \); and \( E(ζ) = 0 \). Hence, the correlation between \( κ = ξ + ζ \) and \( η \) is zero. Therefore, following a similar argument in the SKAT-O paper, we can approximate \( S_ρ \) as a weighted sum of two asymptotically independent random variables that do not depend on \( ρ \).

Consequently, we have

\[
\begin{align*}
    p_T &= 1 - P\left\{ S_{ρ_1} < q(ρ_1), S_{ρ_2} < q(ρ_2), \cdots, S_{ρ_b} < q(ρ_b) \right\} \\
         &= 1 - P\left( κ < \min_{v=1,\ldots,b} \left\{ \frac{1}{1−ρ_v}[q(ρ_v) − τ(ρ_v)η] \right\} \right) \\
         &= 1 - E\left[ P\left( κ < \min_{v=1,\ldots,b} \left\{ \frac{1}{1−ρ_v}[q(ρ_v) − τ(ρ_v)η] \right\} \middle| η \right) \right] \\
         &= 1 - \int_x F_κ\left( \min_{v=1,\ldots,b} \left\{ \frac{1}{1−ρ_v}[q(ρ_v) − τ(ρ_v)x] \right\} \right) \cdot f_η(x)dx,
\end{align*}
\]

where \( F_κ(·) \) is the distribution function of \( κ \) and \( f_η(·) \) is the density function of \( η \).

Since \( κ \) and \( η \) are both distributed as mixtures of chi-square distributions, \( F_κ(·) \) can be approximated by using the moment matching method. Following Lee et al., we improve tail probability approximation for \( F_κ(·) \) by matching mean, variance and kurtosis instead of the first three moments. Although we considered moment matching also for estimation of \( f_η(·) \), the approximated function can be discontinuous hence leading to accuracy problems in numerical integration. There-
fore, instead, we use convolution of independent $\chi^2_1$ random variables to calculate $f_{\eta}(\cdot)$, and $p_T$ can be calculated by using one-dimensional numerical integration.

**S.4 Well-replicated known bone genes in UK10K analysis**

In our analyses of the UK10K data, we identify three top genes, in which the gene $SLC17A5$ has been associated with free sialic acid storage disease\(^3\),\(^4\), an inherited disorder that primarily influences the nervous system\(^5\).

There are several well-replicated genes known to contain variants that influence bone mineral density\(^6\). Our results did not identify these genes as significant, either because of the limited sample size or due to windows containing too many non-associated variants. Here, in Table S1, we report the observed $p$-values at 15 well-replicated bone-density-associated genes. The results using MURAT have smaller or equivalent $p$-values at 11 of the 15 genes.

**S.5 Computation time and memory usage**

To have a better understanding of computational costs for MURAT, we performed some simulations to compare computational time and memory usage between MURAT, SKAT, and Maity’s method, when the sample size $N$, the numbers of traits, $K$, and variants, $p$, change.

In the simulations, we considered the number of traits to be $K = 2, 3, 5$, the numbers of variants to be $p = 10, 20, 50$, and three different sample sizes, $N = 500, 750, 1000$. The phenotypes were generated from a multivariate normal distribution with mean 1 and covariance matrix $\Sigma = (1 - 0.4)I_K + 0.4 \times 1_K 1_K^T$. The covariates matrix, $X$, was generated such that the first column represents the intercept (1), and elements in the second column were randomly selected from a standard normal distribution. The genotype matrix, $G$, was simply generated based on a Bernoulli distribution with success probability 0.1. All the computation time and memory usage statistics were based on one simulated data set, and measured on a Linux cluster by using one thread on a node equipped with 64 GB RAM and two Intel Xeon CPUs E5645 at 2.40GHz.
Our simulations show that SKAT has the lowest computational cost. Under all combinations of $K$, $p$, and $N$, the computation times for SKAT were within one second and the memory usage ranged from 3.6 to 37 Mb. Since SKAT is much faster and uses less resources than either MURAT or Maity’s method, for the remainder of this section, we only compare the computational costs between MURAT and Maity’s methods.

First, we found that our proposed MURAT uses much less time than Maity’s method under all simulation settings. The computational time for MURAT ranged from 1.74 seconds to 22.14 seconds, while the corresponding time for Maity’s test ranged from 22.17 seconds to 2048 seconds (about 35 minutes). In addition, Figure S3 shows that for both MURAT and Maity’s test, the number of variants, $p$, does not have much effect on the computation time. However, for both methods, the computation time depends significantly on the number of trait $K$ and the sample size $N$. In addition, the $p$-values obtained from Maity’s method, for our comparison here, depend on a resampling process where the default number of resamplings is $10^5$. In order to get comparable $p$ values at more stringent significance levels, more resamplings would be needed for Maity’s test and consequently more computation time.

Since the memory usage is not static, we compared the maximum memory usages for three methods under each combinations of $K$, $p$, and $N$, respectively. Memory usage is affected by all of the three factors, and it increases with the number of traits, variants, and sample size. In general, MURAT needs more memory than Maity’s method. The memory usages for MURAT range from 387Mb to 8321Mb, and for Maity’s method from 274Mb to 5174Mb. This may be because MURAT requires more complex large matrix operation.

In particularly, the runtimes for applying MURAT to UK10K data are summarized in Table S2. These runtimes were measured on the same Linux cluster mentioned above.
S.6 Misspecified $\Sigma_\beta$

In our model, we assumed the variant effects, $\beta$, are correlated random variables and the covariance matrix has a specific structure, $\Sigma_\beta = \tau^2 R \otimes I_p$ with $R = (1 - \rho)I_K + \rho 1_K 1_K^T$. That is, we assumed a common correlation for effects of the same variant on different phenotypes, but no correlation for effects of different variants. Since our test statistic is based on this assumption about the structure of $\Sigma_\beta$, it is of interest to evaluate power of MURAT when the true correlations violate the above assumption.

Therefore, we performed a set of simulations with misspecified correlations. For clarity, from now on, we call $\rho$, the common correlation for effects of the same variant on different phenotypes, as “between” correlation and denote it by $\rho = \rho_b = \text{Corr}(\beta_{kj}, \beta_{k'j})$ for $j = 1, \ldots, p$, $k \neq k'$, and $k, k' \in \{1, \ldots, K\}$. To simulate a misspecified variant effect correlation matrix, we also assume there exists a “within” correlation for effects of different variants on the same phenotype, and denote this by $\rho_w = \text{Corr}(\beta_{kj}, \beta_{kj'}) \neq 0$ for $k = 1, \ldots, K$, $j \neq j'$ and $j, j' \in \{1, \ldots, p\}$. In the simulation, we considered $\rho_w = 0, 0.1, 0.3$, and $\rho_b = 0, 0.1, 0.3, 0.5, 0.7$. The other simulation settings are similar to the ones used for the Simulation Studies section in the main paper. We randomly chose two causal variants out of 10 variants and set $\rho_e = 0.1$.

Since the null model does not depend on $\Sigma_\beta$, the type I error for MURAT would not change when $\Sigma_\beta$ is misspecified. Hence, based on the conclusion drawn in the main paper, our proposed multivariate test has correct type I error rate under this situation. Powers of MURAT under different $\rho_w$ and $\rho_b$ combinations are listed in Table S3. Note that when $\rho_w = 0$, the covariance matrix for variant effects, $\Sigma_\beta$, does not violate our assumption. Hence the first row in Table S3 indicates the power of our proposed test when $\Sigma_\beta$ is not misspecified. Compared with the first row, the other rows show only very small differences in terms of test power. Hence this kind of misspecification of $\Sigma_\beta$ does not have a substantial impact on our proposed test power.
S.7 Density of the grid values for $\rho$

We performed some additional simulations to investigate whether different densities of grid searches for $\rho$ would affect the performance of MURAT. In the simulations, we considered three different grids: $\rho$ ranging from 0 to 0.99 by 0.15, 0.05, and 0.01. We compared the corresponding type I error rates and powers with the ones obtained in the main manuscript section ”Simulations studies”, where the grid search for $\rho$ ranged from 0 to 0.99 by 0.1.

For simplicity, we assumed two correlated traits, and used the same model as in the ”Simulation studies” section, to generate 10,000 simulated data sets. Then, we applied MURAT to analyze each data set with four different different grid settings. When considering type I error, we generated phenotypic values under the null model with $\rho_e = 0.4$. For power comparison, the simulated traits were generated when five causal variants were associated with both traits, $\rho_e = 0.4$, and $\rho = 0.3$.

Tables S4 and S5 list estimated type I errors and powers for these four different grids, where the step size of the grid search is given in the header. For example, ”grid 0.05” implies a search from 0.05 to 0.95 by 0.05. From these two tables, we can see that the density of grid has little effect on MURAT.

By further investigation, we find that when applying different grids on the same data set, the finally-selected $\rho_v$s, which result the smallest $p$-value, are quite close to each other. Hence, the resulting test statistics from different grid searches are very similar, and consequently, different grids tend to produce similar test results. For example, when comparing grid 0.05 and grid 0.01 for type I error, 4287 out of 10000 simulated data sets return same selected $\rho_v$s, and for a total of 9851 data sets, the distances between two selected $\rho_v$s from different grids are less or equal to 0.04. Similar phenomena can also be found when comparing other grids for type I errors or powers.

S.8 Antagonistic pleiotropy

In the two simulation scenarios in our Simulation Studies section of the main paper, the variant effects $\beta$ were randomly generated from multivariate normal distribution with mean zero. Hence,
there was no restriction on the direction of \( \beta \), apart from that induced by the assumed common correlations. However, there may exist situations where \( \beta \) values have opposing directions for different phenotypes. This concept was described as anatogistic pleiotropy by Williams in 1957\(^7\). Antagonistic pleiotropy refers to the situation where one gene is associated with multiple traits, and some of these traits are beneficial but others are detrimental to the organism. This could happen when a single variant is associated with highly negatively correlated traits. We realize that when there are more than two phenotypes, it may not always be possible to construct a set of phenotypes that are positively correlated. This is a limitation of our method. To investigate the power of MURAT under this special situation, we conducted a third simulation when variants have opposing effects.

In particular, we considered two causal variants associated with three traits. We further assumed that the variant effects with respect to trait 1 were all positive, i.e. \( \beta_{11} > 0 \) and \( \beta_{12} > 0 \), and the variant effects with respect to trait 2 were all negative, i.e. \( \beta_{21} < 0 \) and \( \beta_{22} < 0 \). For the effects with respect to the third trait, we considered three different settings: 1. \( \beta_{31} > 0 \) and \( \beta_{32} > 0 \); 2. \( \beta_{31} > 0 \) and \( \beta_{32} < 0 \); and 3. \( \beta_{31} < 0 \) and \( \beta_{32} < 0 \). All the simulations were based on 10000 simulated data sets. In addition, to generate phenotypes, we set \( \rho_e = 0.4 \) and repeatedly drew \( \beta \) from multivariate normal distribution with mean zero and covariance matrix \( R = (1 - 0.3)I_3 + 0.3 \times I_31_3^T \) until we obtained enough \( \beta_s \) that satisfied restrictions.

Table S6 lists the power comparison between MURAT and SKAT under our antagonistic pleiotropy scenarios. For MURAT, it seems there is no clear trend when we change the direction of \( \beta \). In addition, under all three settings, the multivariate test has larger power than the univariate test.

### S.9 References

1. Rencher AC: Methods of multivariate analysis, 2nd Edition. *John Wiley & Sons* 2002.

2. Lee S, Wu MC, Lin X: Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 2012; 13:762-775.
3. Aula N, Salomaki P, Timonen R, Verheijen F, Mancini G, Mansson JE et al: The spectrum of SLC17A5-gene mutations resulting in free sialic acid-storage diseases indicates some genotype-phenotype correlation. *Am. J. Hum. Genet.* 2000; 67:83240.

4. Landau D, Cohen D, Shalev H, Pinsky V, Yerushalmi B, Zeigler M et al: A novel mutation in the SLC17A5 gene causing both severe and mild phenotypes of free sialic acid storage disease in one inbred Bedouin kindred. *Molecular Genetics and Metabolism* 2004; 82:167-72.

5. Suwannarat P: Disorders of free sialic. *Molecular Genetics and Metabolism* 2005; 85:85-7.

6. Richards JB, Zheng HF, Spector TD: Genetics of osteoporosis from genome-wide association studies: advances and challenges. *Nature Reviews Genetics* 2012; 13(8):576-588.

7. Williams G.C.: (1957). Pleiotropy, natural selection, and the evolution of senescence. *Evolution*. 1957; 11: 398411.

**S.10 Tables**
Table S1: MURAT and adjusted SKAT $p$ values for 15 well-replicated bone-density-associated genes obtained in UK10K analyses with 1005 samples.

| Gene | Chr | MURAT | adj. SKAT |
|------|-----|-------|-----------|
| $SPP1$ | 4 | 0.27 | 0.33 |
| $RSP03$ | 6 | 0.68 | 0.88 |
| $ESR1$ | 6 | 0.83 | 0.87 |
| $RUN2$ | 6 | 0.23 | 0.79 |
| $WNT16$ | 7 | 0.11 | 0.14 |
| $OPG$ | 8 | 0.33 | 0.08 |
| $DKK1$ | 10 | 0.43 | 0.55 |
| $LRP5$ | 11 | 0.20 | 0.23 |
| $LRP4$ | 11 | 0.13 | 0.13 |
| $SOX6$ | 11 | 0.75 | 0.71 |
| $PTH1L$ | 12 | 0.40 | 0.38 |
| $SP7$ | 12 | 0.21 | 0.26 |
| $RANKL$ | 13 | 0.60 | 0.53 |
| $SOST$ | 17 | 0.11 | 0.21 |
| $SOX9$ | 17 | 0.60 | 0.81 |

Table S2: Computation time (in hours) when using MURAT to perform exome-wide rare variant multivariate analysis of BMD phenotypes, for each of 22 autosomes, in data from the UK10K project.

| Chr | # of Genes | time | Chr | # of Genes | time |
|-----|------------|------|-----|------------|------|
| 1   | 2004       | 5.99 | 12  | 1046       | 3.64 |
| 2   | 1280       | 5.00 | 13  | 324        | 1.57 |
| 3   | 1069       | 4.09 | 14  | 792        | 2.72 |
| 4   | 734        | 2.96 | 15  | 575        | 2.19 |
| 5   | 876        | 3.59 | 16  | 827        | 2.87 |
| 6   | 1031       | 3.57 | 17  | 1165       | 3.48 |
| 7   | 936        | 3.37 | 18  | 280        | 1.18 |
| 8   | 660        | 2.56 | 19  | 1432       | 4.25 |
| 9   | 772        | 2.84 | 20  | 553        | 1.71 |
| 10  | 743        | 2.68 | 21  | 241        | 0.87 |
| 11  | 1296       | 4.10 | 22  | 487        | 1.79 |
Table S3: Power of MURAT when $\Sigma_{\beta}$ is misspecified by having both within and between correlations. The significance level is $\alpha = 0.05$.

| $\rho_w$ | $\rho_b$ | 0.0 | 0.1 | 0.3 | 0.5 | 0.7 |
|----------|----------|-----|-----|-----|-----|-----|
| 0.0      | 0.668    | 0.664| 0.659| 0.641| 0.614|     |
| 0.1      | 0.662    | 0.667| 0.656| 0.644| 0.622|     |
| 0.3      | 0.656    | 0.660| 0.654| 0.647| 0.617|     |

Table S4: Estimated type I errors when applying different grids for $\rho$ under different significance levels. The grid values in the table correspond to the step sizes in the grid search.

| $\alpha$ | grid 0.1 | grid 0.15 | grid 0.05 | grid 0.01 |
|----------|----------|-----------|-----------|-----------|
| 0.05     | 0.0499   | 0.0515    | 0.0516    | 0.0516    |
| 0.01     | 0.0102   | 0.0103    | 0.0104    | 0.0104    |
| 0.001    | 0.0010   | 0.0011    | 0.0011    | 0.0011    |

Table S5: Estimated powers when applying different grids for $\rho$ under different significance levels. The grid values in the table correspond to the step sizes in the grid search.

| $\alpha$ | grid 0.1 | grid 0.15 | grid 0.05 | grid 0.01 |
|----------|----------|-----------|-----------|-----------|
| 0.05     | 0.7225   | 0.7225    | 0.7210    | 0.7200    |
| 0.01     | 0.4062   | 0.4053    | 0.4042    | 0.4026    |
| 0.001    | 0.1336   | 0.1329    | 0.1333    | 0.1325    |

Table S6: Power comparison between MURAT and SKAT under antagonistic pleiotropy. The results for SKAT have been corrected for multiple testing.

|                | MURAT | SKAT |
|----------------|-------|------|
| $\beta_{31} > 0, \beta_{32} > 0$ | 0.55  | 0.36 |
| $\beta_{31} > 0, \beta_{32} < 0$ | 0.57  | 0.35 |
| $\beta_{31} < 0, \beta_{32} < 0$ | 0.54  | 0.35 |
Figure S1: Empirical powers of MURAT which tests associations with two phenotypes simultaneously, versus SKAT, which tests associations with the two phenotypes separately, at significance level $\alpha = 0.05$. The results for SKAT are adjusted for multiple testing. One causal variant is associated with both traits.
Figure S2: Empirical powers of MURAT which tests associations with two phenotypes simultaneously, versus SKAT, which tests associations with the two phenotypes separately, at significance level $\alpha = 0.05$. The results for SKAT are adjusted for multiple testing. Five causal variants are associated with both traits.
Figure S3: Computation time comparison between MURAT and Maity’s method. For MURAT, the computation time is measured in seconds. For Maity’s test, the computation time is measured in minutes.