EJHG Open

ARTICLE

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

Jianping Sun[1,2], Karim Oualkacha[3], Vincenzo Forgetta[1,2,4,5], Hou-Feng Zheng[6,7], J Brent Richards[1,2,4,5,8], Antonio Ciampi[1] and Celia MT Greenwood*[,1,2,4,9] UK10K Consortium

For region-based sequencing data, power to detect genetic associations can be improved through analysis of multiple related phenotypes. With this motivation, we propose a novel test to detect association simultaneously between a set of rare variants, such as those obtained by sequencing in a small genomic region, and multiple continuous phenotypes. We allow arbitrary correlations among the phenotypes and build on a linear mixed model by assuming the effects of the variants follow a multivariate normal distribution with a zero mean and a specific covariance matrix structure. In order to account for the unknown correlation parameter in the covariance matrix of the variant effects, a data-adaptive variance component test based on score-type statistics is derived. As our approach can calculate the P-value analytically, the proposed test procedure is computationally efficient. Broad simulations and an application to the UK10K project show that our proposed multivariate test is generally more powerful than univariate tests, especially when there are pleiotropic effects or highly correlated phenotypes.

INTRODUCTION

Advances in next-generation sequencing technologies have led to identification of millions of rare variants (with minor allele frequencies (MAFs) < 0.05) that can be tested in genetic association studies. There is a growing literature on methods for analysis of association for a set of rare variants: for example, the burden test,[1,2] sequence kernel association test (SKAT),[3] optimal SKAT (SKAT-O),[4] mixed effects score test,[5] and many others.[6] Most of the existing rare-variant association tests can be thought of as univariate tests. That is, they focus on the association between a set of rare variants and a single phenotype.

However, multiple measures of related traits, that is, a multivariate phenotype, are often available and are of interest. Some phenotypes such as blood pressure, which is measured by both systolic and diastolic pressure, are intrinsically multivariate. In other cases, there may be several correlated measures of a trait of interest; however, the one most directly influenced by genetic variation is unknown. For example, excess weight can be measured by body mass index, waist circumference or skin-fold thickness; by bringing in serum lipids as well, the composite phenotype will have a different emphasis.

Many patterns of genetic architectures could link variants at a locus to a set of phenotypes:[7] pleiotropy describes one genetic variant influencing multiple traits;[8] endophenotypes (measures assumed to be proximal to genetic effects) can be highly multidimensional; locus heterogeneity can refer to a single trait caused by variants at different chromosomal loci,[9] or alternatively several variants at a locus may influence a disease, whereas symptoms (visible phenotypes) vary from person to person (Figure 1). Thus, multivariate methods may help in rare-variant association studies, as they may add ability to investigate the architectures discussed above, and increase power to detect disease-associated loci.[10]

For associations between a single genetic variant and multiple traits, a recent review paper[11] summarizes several proposed methods. However, these single-variant-based multivariate methods usually have limited power to detect rare-variant effects. Although there are a few approaches that have been developed to test associations between multivariate traits and multiple variants in a genetic region, several of these[12,13] employ the idea of combining the test statistics obtained from single-variant-based multivariate tests or multiple-variant-based univariate tests. Notably, Maity *et al*[14] recently proposed a kernel machine regression to test multiple phenotypes and a set of markers simultaneously, and the method of Maity *et al* showed power improvements over single phenotype tests when the correlations are strong. However, this method can be computationally inefficient and may not have optimal power to detect pleiotropic associations, as there is no explicit consideration of correlations between variant effects. Given the rapid increase in large-scale sequencing studies, there is a need for a powerful yet easily implemented statistical method to simultaneously detect associations between multiple phenotypes and rare variants in a region.

In this study, we develop a novel region-based multivariate test, MURAT (Multivariate Rare-Variant Association Test), for identifying

[1]Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada; [2]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, QC, Canada; [3]Department of Mathematics, Université du Québec À Montréal, Montreal, QC, Canada; [4]Department of Human Genetics, McGill University, Montreal, QC, Canada; [5]Department of Medicine, Jewish General Hospital, Montreal, QC, Canada; [6]Institute of Aging Research, School of Medicine, Hangzhou Normal University, Hangzhou, Zhejiang, China; [7]Affiliated Hospital of Hangzhou Normal University, Hangzhou, Zhejiang, China; [8]Department of Twin Research, King's College London, London, UK; [9]Department of Oncology, McGill University, Montreal, QC, Canada
*Correspondence: Dr CMT Greenwood, Division of Clinical Epidemiology, Jewish General Hospital, 3755 Cote Ste Catherine, H414, Montreal, Quebec, Canada H3T 1E2. Tel: +1 514 340 8222x8397; Fax: +1 514 340 7564; E-mail: celia.greenwood@mcgill.ca
Received 27 March 2015; revised 22 December 2015; accepted 30 December 2015; published online 10 February 2016
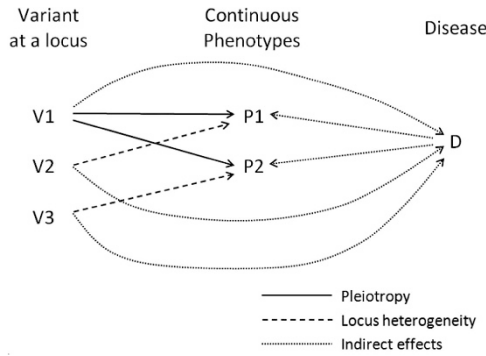
Figure 1 Possible patterns of genetic architecture that could link variants at a locus to a set of continuous phenotypes. The solid line represents pleiotropy, where a single variant influences multiple phenotypes. The dashed line describes the locus heterogeneity, where different variants in the same gene or region each influence a different phenotype. The dotted line represents a situation where correlation among phenotypes arises indirectly due to the variants' effects on disease; the phenotypes could be symptoms, endophenotypes, or continuous manifestations of disease.

rare-variant associations when multiple correlated continuous phenotypes are observed. By assuming the variant effects to be random and correlated, we propose a linear mixed model that leads to a data-adaptive variance component test based on a score-type statistic. This choice accounts for unknown correlations among variant effects, while testing for association. We show that there are two special cases of our model depending on the assumed correlation structure, one where it reduces to SKAT and another where it becomes a multivariate fixed effects model. In addition, we show that our analytic P-value derivation only depends on a one-dimensional numerical integration, and hence that this method is computationally efficient and can be considered for analysis of genome-wide sequencing data.

METHODS
Notation and model

Suppose for the ith individual, $i=1,\ldots,N$, we observe K correlated continuous phenotypes, $\mathbf{Y}_i=(y_{i1},y_{i2},\ldots,y_{iK})^T$, m covariates, $\mathbf{X}_i=(x_{i1},x_{i2},\ldots,x_{im})^T$, and genotype data for a set of SNPs that contains p variants $\mathbf{G}_i=(g_{i1},g_{i2},\ldots,g_{ip})^T$, where g_{ij} for $j=1,\ldots,p$ are coded as 0, 1, or 2, representing the number of minor alleles that individual i carries for the jth variant. It is noteworthy that these p variants can be both common and rare, although we focus on rare variants in this study. We further assume that $(\mathbf{Y}_i,\mathbf{X}_i,\mathbf{G}_i)$, $i=1,\ldots,N$, are independent and identically distributed, and our goal is to test the association between $\mathbf{G}_i$, a set of (rare) variants, and the multiple traits $\mathbf{Y}_i$

Assuming that the relationship between $\mathbf{Y}_i$, $\mathbf{X}_i$, and $\mathbf{G}_i$ can be modeled as

$$\mathbf{Y}_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iK} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_i^T\boldsymbol{\alpha}_1 \\ \vdots \\ \mathbf{X}_i^T\boldsymbol{\alpha}_K \end{pmatrix} + \begin{pmatrix} \mathbf{G}_i^T\boldsymbol{\beta}_1 \\ \vdots \\ \mathbf{G}_i^T\boldsymbol{\beta}_K \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iK} \end{pmatrix} \text{ for } i=1,\ldots,N, \quad (1)$$

where for the kth phenotype, $k=1,\ldots,K$, $\boldsymbol{\alpha}_k=(\alpha_{k1},\ldots,\alpha_{km})^T$, and $\boldsymbol{\beta}_k=(\beta_{k1},\ldots,\beta_{kp})^T$ are regression coefficients for $\mathbf{X}_i$ and $\mathbf{G}_i$, respectively. In addition, we also assume $\boldsymbol{\alpha}_k$ is a vector of fixed effects and $\boldsymbol{\beta}_k\sim N(\mathbf{0},\tau^2W)$ is a vector of random effects, where τ^2 is an unknown variance component and $W=diag\{w_1,\ldots,w_p\}$ is a weight matrix such that w_j represents the weight for the jth variant. The weight matrix is predetermined and usually a decreasing function of observed MAFs for the corresponding variants. Finally, we suppose the model error term, $\boldsymbol{\varepsilon}_i=(\varepsilon_{i1},\varepsilon_{i2},\ldots,\varepsilon_{iK})^T$, is multivariate normal with a mean vector of zeros and a variance covariance matrix Σ. As there are

correlations among y_{ik}'s for $k=1,\ldots,K$, we assume Σ is an arbitrary symmetric and positive definite matrix.

Moreover, if we denote $\mathbf{Y}_{KN\times 1}=(\mathbf{Y}_1^T,\ldots,\mathbf{Y}_N^T)^T$, $X^*_{KN\times Km}=(I_K\otimes\mathbf{X}_1,\cdots,I_K\otimes\mathbf{X}_N)^T$, and $G^*_{KN\times Kp}=(I_K\otimes\mathbf{G}_1,\cdots,I_K\otimes\mathbf{G}_N)^T$, where $\otimes$ is the Kronecker product, then the above equation (1) can be written in matrix form as

$$\mathbf{Y} = X^*\boldsymbol{\alpha} + G^*\boldsymbol{\beta} + \boldsymbol{\epsilon}. \quad (2)$$

Here we assume $\boldsymbol{\epsilon}_{KN\times 1}=(\boldsymbol{\varepsilon}_1^T,\ldots,\boldsymbol{\varepsilon}_N^T)^T\sim N(\mathbf{0},\Sigma_e)$ with $\Sigma_e=I_N\otimes\Sigma$; $\boldsymbol{\alpha}=(\boldsymbol{\alpha}_1^T,\ldots,\boldsymbol{\alpha}_K^T)^T$ is a $Km\times 1$ vector denoting all fixed covariate effects, and $\boldsymbol{\beta}=(\boldsymbol{\beta}_1^T,\ldots,\boldsymbol{\beta}_K^T)^T$ is a $Kp\times 1$ vector representing all random variant effects, assumed to follow a multivariate normal distribution with variance covariance matrix Σ_β.

The assumed correlation structures in Σ_e and Σ_β are crucial to this model. To capture trait pleiotropy effects, we assume the correlation between $\boldsymbol{\beta}_k$ and $\boldsymbol{\beta}_{k'}$ is $Corr(\boldsymbol{\beta}_k,\boldsymbol{\beta}_{k'})=\rho I_p$ for $k\neq k'$ and $k,k'\in\{1,\ldots,K\}$, where ρ is unknown. That is, we suppose there is a common correlation for the effects of the same variant on different phenotypes, $Corr(\beta_{kj},\beta_{k'j})=\rho$ for $j=1,\ldots,p$. This assumption also captures the situation where the phenotypes are symptoms of a latent trait. In addition, we assume that effects of different variants are uncorrelated, so that $Corr(\beta_{kj},\beta_{k'j'})=0$ and $Corr(\beta_{kj},\beta_{kj'})=0$ for $j\neq j'$. Therefore, we have $\boldsymbol{\beta}\sim N(\mathbf{0},\Sigma_\beta)$, where $\Sigma_\beta=\tau^2R\otimes W$ is a $Kp\times Kp$ symmetric and positive definite matrix, $R=(1-\rho)I_K+\rho\mathbf{1}_K\mathbf{1}_K^T$, and $\mathbf{1}_K$ is a K-dimensional vector of ones. For example, when $K=2$, we have $\Sigma_\beta=\tau^2\begin{pmatrix} W & \rho W \\ \rho W & W \end{pmatrix}$. Consequently, the variance covariance matrix of $\mathbf{Y}$ is

$$Var(\mathbf{Y}) = G^*\Sigma_\beta G^{*T} + \Sigma_e. \quad (3)$$

The proposed model (2) is a generalization of the underlying model used for the SKAT test, where variant effects are assumed to be random, independent, and to follow a distribution with mean 0 and variance τ^2. In particular, model (2) reduces to SKAT when $K=1$ (one observed phenotype). Although a straightforward multivariate generalization of SKAT could be achieved by implementing SKAT models for each $(y_{ik}, \mathbf{G}_i)$ pair, $k=1,\ldots,K$, and then combining these K models allowing correlations among $\{y_{ik}\}$'s, our proposed MURAT model is richer, because it allows the effects of the same variant on different phenotypes to be correlated as shown in equation (3). On the contrary, if we simply generalized SKAT as described above, the corresponding covariance matrix would be $Var(\mathbf{Y})=\tau^2G^*[I_K\otimes W]G^{*T}+\Sigma_e$, which is a special case of equation (3) when $\rho=0$, under which, in fact model (2) is also equivalent to Maity's test with a linear kernel. Thus, through the correlation parameter ρ, MURAT has the ability to explore pleiotropic effects.

Proposed test statistics

In model (2), the null hypothesis of no association between G^* and $\mathbf{Y}$ is H_0: $\boldsymbol{\beta}=0$. This is equivalent to $H_0:\tau^2=0$. We propose to develop a score test, because it does not require an estimate of τ^2. For a fixed ρ, the variance component score statistic for τ^2 is

$$S_\rho = (\mathbf{Y}-X^*\hat{\alpha})^T\left(\hat{\sum}_e^{-1}G^*[R\otimes W]G^{*T}\hat{\sum}_e^{-1}\right)(\mathbf{Y}-X^*\hat{\alpha}), \quad (4)$$

where $\hat{\alpha}$ and $\hat{\sum}_e^{-1}$ are estimated under the null by regressing $\mathbf{Y}$ on only the covariates X^*. The subscript ρ indicates that this score statistic is a function of ρ. As we assume the K phenotypes are correlated, $\hat{\alpha}$ and $\hat{\sum}_e^{-1}$ can be efficiently estimated using the generalized least-squares method, achieved by using the function *gls()* in the R package *nlme*.

For fixed ρ, it can be shown that under the null hypothesis, S_ρ follows a mixture of χ^2-distributions[15–17] when R is a symmetric matrix (Supplementary S.1). Several methods exist to approximate this mixture χ^2-distribution. For example, both the Davies method[18] based on the inverse characteristic function and the Liu method[19] based on moment matching have been shown to work well.

However, the common correlation ρ is usually unknown. To construct a test statistic for unknown ρ, as suggested by Lee *et al*,[4] we apply a data adaptive approach to select an optimal ρ, to maximize test power. We first calculate S_ρ and its P-value for a grid of values of ρ over [0,1] and then use the smallest P-value as an overall test statistic for testing $H_0:\tau^2=0$. Specifically, for a grid $0=\rho_1<\rho_2<\cdots<\rho_b=1$, we obtain S_{ρ_v} and the corresponding P-value, p_{ρ_v}, for

$v = 1,...,b$. Then, the overall test statistic is defined to be

$$T = \min\left\{p_{\rho_1}, ..., p_{\rho_b}\right\}. \tag{5}$$

We have shown that when $\rho = 0$, S_ρ is the same as the test statistic used in Maity's method with linear kernel or the one obtained from multivariate SKAT by simply assuming correlated model errors. In addition, we can also show that when $\rho = 1$, the proposed S_ρ is equivalent to the square of a score statistic obtained from a multivariate fixed effect model, where we assume the same effect of each variant across the multiple traits (Supplementary S.2). Hence, by searching $0 \leq \rho \leq 1$, MURAT covers all score tests between Maity's method (or multivariate SKAT) and multivariate multiple regression.

Null distribution of the test statistic T

For an observed T, say T_{obs}, let $q(\rho)$ be the $(1 - T_{\text{obs}})^{\text{th}}$ percentile of the null distribution of S_ρ, such that $p_\rho > T_{\text{obs}}$ is equivalent to $S_\rho < q(\rho)$. Then, by definition, the P-value, p_T, for overall test statistic T is defined to be

$$p_T = P\{T < T_{\text{obs}}\} = 1 - P\{S_{\rho_1} < q(\rho_1), S_{\rho_2} < q(\rho_2), \cdots, S_{\rho_b} < q(\rho_b)\}.$$

We show in the Supplementary S.3 that we can approximate S_ρ as a weighted sum of two asymptotically independent random variables, say κ and η, which do not depend on ρ, and both follow a mixture of χ^2-distributions. Consequently, p_T can be calculated by using one-dimensional numerical integration,

$$p_T = 1 - \int_x F_\kappa\left(\min_{v=1,...,b}\left\{\frac{1}{1-\rho_v}[q(\rho_v) - \tau(\rho_v)x]\right\}\right) \cdot f_\eta(x)dx, \tag{6}$$

where $\tau(\rho)$ is a constant, $F_\kappa(\cdot)$ is the distribution function of κ, and $f_\eta(\cdot)$ is the density function of η. To improve tail probability estimates, we used the moment matching method with mean, variance, and kurtosis instead of the first three moments,[4] to approximate $F_\kappa(\cdot)$. Furthermore, to improve accuracy in numerical integration, we used a convolution of independent random variables to calculate $f_\eta(\cdot)$ (see Supplementary S.3).

RESULTS

Simulation studies

We conducted a variety of simulation studies to evaluate the performance of MURAT. In the simulations, we considered K ($K = 2$ or 3) correlated continuous traits and compared type I error and power between MURAT, Maity's test, and SKAT, which tests the multiple traits separately. The genotypes we used for the simulations were taken from sequencing data for a sample of 1000 individuals from the UK10K project (http://www.uk10k.org/). A set of 10 rare variants were randomly selected from a single gene, *BET1L*, in order to preserve the linkage disequilibrium pattern. The MAFs for these 10 variants range from 0.001 to 0.043 in the sample.

For each individual, we generated K correlated continuous phenotypes following the model (1), where two covariates were generated from *Normal*(0,1) and *Bernoulli*(0.5), respectively, both with coefficient 0.5; a common intercept of 0.1 was also included and $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{iK})^T$ was generated from a multivariate normal distribution, $N(\mathbf{0}, \Sigma)$, with $\Sigma = (1 - \rho_e)I_K + \rho_e \mathbf{1}_K \mathbf{1}_K^T$. In simulations, we used four different levels of the residual correlation, $\rho_e = 0$, 0.1, 0.4, and 0.7. Different values for the genetic effects, β_{kj}, where $k = 1,...,K$ and $j = 1,...,10$, were considered, in order to evaluate type I error and power. To avoid potential zero denominators in equation (6), we searched over a grid of values for ρ from 0 to 0.99, with step size 0.1 in MURAT. For simplicity, we used the identity weight matrix, that is, $W = I$, through all the simulation studies. A total of 10 000 simulated data sets were generated for each scenario. Finally, to eliminate the effects caused by different phenotype scales, we first standardized all phenotypes within each simulation before implementing our multivariate test.

Unless otherwise specified, to correct for multiple testing across phenotypes, all SKAT results are based on adjusted P-values, defined as K times the minimum of the univariate-based P-values obtained via SKAT.

Type I error rates

To evaluate type I error, we generated phenotypes with $\beta_{kj} = 0$ for $k = 1,...,K$ where $K = 2$ or 3, and $j = 1,...,10$. We assessed type I error first at moderate significance levels of $\alpha = 0.05$, $\alpha = 0.01$, and $\alpha = 0.001$, and the results are summarized in Table 1. All three methods control type I error well under these moderate significance levels, except that SKAT tends to have slightly inflated type I error when $\alpha = 0.001$ and $\rho_e > 0$.

However, in genome-wide studies, such moderate significance levels are usually of secondary interest. As there are around 20,000 genes in the exome, a Bonferroni-adjusted threshold for testing would be $0.05/20\,000 = 2.5 \times 10^{-6}$. Therefore, we also examined the type I error of MURAT and SKAT for smaller α levels of 10^{-4}, 10^{-5}, and 2.5×10^{-6} by performing 10^7 simulations. Owing to the computational burden, these extra simulations only examined $\rho_e = 0$ and 0.4. In addition, as Maity's test is much slower than the other two methods (Supplementary S.5), we excluded it from this comparison.

Table 2 shows that both MURAT and SKAT have inflated type I error when α is very small, and that SKAT performs better than MURAT in this situation. However, the inflation for MURAT is not severe. MURAT's type I error inflation could be due to the approximations required to calculate the significance level analytically, such as matching the first three moments, or the numerical integration in equation (6). Permutation analysis can be used at key regions to obtain more accurate P-values if necessary. To perform permutation, one needs to preserve the dependence structure between the phenotypes and between the genotypes as well. Specifically, for one individual, all his/her phenotypes should be viewed as a single phenotype group or block, and similarly all genotypes in the region being studied can be considered as a single genotype block. Permuting the phenotype blocks with respect to the genotype blocks across individuals leads to data sets that retain the necessary within-block correlations, yet are generated under the null hypothesis of no association between the blocks.

Power

To assess power of the proposed multivariate test, we randomly selected one, two, or five rare variants, to be causal. That is, let $\{v_1, v_2,..., v_u\}$ be u selected causal variants, where $u = 1$, 2, or 5. We set variant effects $\beta_{1j} = \cdots = \beta_{Kj} = 0$ for $j \notin \{v_1, v_2,..., v_u\}$ in the simulations, where $K = 2$ or 3. The nonzero variant effects, β_{kj}'s were generated from $N(\mathbf{0}, \Sigma_\beta)$ with $\Sigma_\beta = \tau^2 R \otimes I_u$ and $R = (1-\rho)I_K + \rho \mathbf{1}_K \mathbf{1}_K^T$ for $k = 1,...,K$ and $j' \in \{v_1, v_2,..., v_u\}$. In the simulations, different values for τ^2 were chosen under different scenarios so that the resulting power was high enough to facilitate comparisons. We evaluated power at significance level $\alpha = 0.05$ for each scenario.

Scenario 1. We first considered the scenario where u causal variants were associated with both $K = 2$ phenotypes. In this setting, we selected $\tau^2 = 3$ and considered six different correlations among the causal variant effects: $\rho = 0$, 0.1, 0.3, 0.5, 0.7, and 0.9, in order to explore pleiotropic effects on power.

The top and middle panels in Figure 2 compare MURAT with SKAT for $u = 2$ causal variants associated with both phenotypes and show that MURAT is overall more powerful than SKAT under all possible combinations of ρ and ρ_e. For small residual correlation ρ_e, the pleiotropic correlation ρ has little effect on MURAT's power; however, for large ρ_e, the power of MURAT decreases as ρ increases.

Table 1 Type I error comparison among MURAT, SKAT, and Maity's method at moderate significance levels, $\alpha = 0.05$, 0.01, and 0.001

	K = 2					
	$\rho_e = 0$			$\rho_e = 0.1$		
α	MURAT	Maity	SKAT	MURAT	Maity	SKAT
5×10^{-2}	5.17×10^{-2}	5.17×10^{-2}	4.76×10^{-2}	5.12×10^{-2}	5.10×10^{-2}	4.81×10^{-2}
1×10^{-2}	0.90×10^{-2}	0.94×10^{-2}	0.94×10^{-2}	0.97×10^{-2}	0.99×10^{-2}	0.95×10^{-2}
1×10^{-3}	1.01×10^{-3}	1.30×10^{-3}	0.90×10^{-3}	1.10×10^{-3}	1.10×10^{-3}	1.20×10^{-3}
	$\rho_e = 0.4$			$\rho_e = 0.7$		
5×10^{-2}	4.99×10^{-2}	4.80×10^{-2}	4.78×10^{-2}	5.01×10^{-2}	4.78×10^{-2}	4.41×10^{-2}
1×10^{-2}	1.02×10^{-2}	0.91×10^{-2}	0.98×10^{-2}	0.98×10^{-2}	0.93×10^{-2}	0.97×10^{-2}
1×10^{-3}	1.02×10^{-3}	0.70×10^{-3}	1.40×10^{-3}	1.10×10^{-3}	0.80×10^{-3}	1.30×10^{-3}
	K = 3					
	$\rho_e = 0$			$\rho_e = 0.1$		
α	MURAT	Maity	SKAT	MURAT	Maity	SKAT
5×10^{-2}	5.08×10^{-2}	4.83×10^{-2}	4.79×10^{-2}	5.12×10^{-2}	4.62×10^{-2}	4.91×10^{-2}
1×10^{-2}	1.03×10^{-2}	0.90×10^{-2}	1.02×10^{-2}	1.06×10^{-2}	0.98×10^{-2}	0.91×10^{-2}
1×10^{-3}	1.00×10^{-3}	0.90×10^{-3}	0.60×10^{-3}	1.30×10^{-3}	0.80×10^{-3}	1.20×10^{-3}
	$\rho_e = 0.4$			$\rho_e = 0.7$		
5×10^{-2}	5.06×10^{-2}	4.69×10^{-2}	4.65×10^{-2}	5.08×10^{-2}	4.67×10^{-2}	4.05×10^{-2}
1×10^{-2}	1.02×10^{-2}	0.93×10^{-2}	0.93×10^{-2}	1.04×10^{-2}	0.99×10^{-2}	0.86×10^{-2}
1×10^{-3}	1.20×10^{-3}	0.90×10^{-3}	1.40×10^{-3}	1.10×10^{-3}	1.00×10^{-3}	1.40×10^{-3}

Abbreviations: MURAT, Multivariate Rare-Variant Association Test; SKAT, sequence kernel association test.
The simulations are based on 10 000 simulated data sets. The results for SKAT are based on adjusted P-values, which are defined as K times the minimum of univariate-based P-values obtained via SKAT.

Table 2 Type I error comparison between MURAT and SKAT at stringent significance levels, $\alpha = 1 \times 10^{-4}$, 1×10^{-5}, and 2.5×10^{-6}

	$\rho_e = 0$		$\rho_e = 0.4$	
α	MURAT	SKAT	MURAT	SKAT
K = 2				
1×10^{-4}	1.68×10^{-4}	1.27×10^{-4}	1.48×10^{-4}	1.24×10^{-4}
1×10^{-5}	2.78×10^{-5}	1.16×10^{-5}	1.29×10^{-5}	1.54×10^{-5}
2.5×10^{-6}	7.60×10^{-6}	3.30×10^{-6}	2.80×10^{-6}	3.30×10^{-6}
K = 3				
1×10^{-4}	1.64×10^{-4}	1.29×10^{-4}	1.35×10^{-4}	1.23×10^{-4}
1×10^{-5}	2.15×10^{-5}	1.25×10^{-5}	1.82×10^{-5}	1.35×10^{-5}
2.5×10^{-6}	9.80×10^{-6}	4.10×10^{-6}	10.20×10^{-6}	3.10×10^{-6}

Abbreviations: MURAT, Multivariate Rare-Variant Association Test; SKAT, sequence kernel association test.
The simulations are based on 10^7 simulated data sets. The results for SKAT are based on adjusted P-values, which are defined as K times the minimum of univariate-based P-values obtained via SKAT.

This is not surprising, because when ρ_e and ρ are both large, the two phenotypes are highly correlated. Hence, in situations of very strong correlation, a multivariate test may gain little power relative to analysis of only a single trait. For all values of ρ_e, the power of SKAT decreases as ρ increases. Similar conclusions can also be drawn when there are only one or five causal variants associated with both phenotypes (Supplementary Figures S1 and S2).

We also compared the power of MURAT and SKAT with that of Maity's test with a linear kernel when $K = 2$ and 3. Owing to the computational time associated with Maity's method, we only evaluated its performance with $u = 5$ causal variants, $\rho_e = 0.4$ and $\rho = 0$, 0.3 or 0.7. The bottom panel of Figure 2 shows that for $K = 2$, when $\rho = 0$, MURAT and Maity's test have similar power, consistent with our expectation, but when $K = 3$ and $\rho = 0$, MURAT seems a little less powerful than Maity's test. As MURAT searches grids of ρ, there could exist slight power differences between MURAT and Maity's method even when the true ρ is zero. It is notable that when ρ is large, MURAT tends to be more powerful than Maity's test for both values of K. For example, when $K = 2$, MURAT is 10% (for $\rho = 0.3$) or 40% (for $\rho = 0.7$) more powerful than Maity's test. Hence, MURAT can exploit pleiotropic effects to gain additional power over a multivariate method that only incorporates correlations among phenotypes. In addition, in this scenario, MURAT gains power as K increases.

Scenario 2. We further conducted a set of simulations in which u causal variants were only associated with the first of two traits. That is, we assumed $\beta_{2j} = 0$ for $j = 1,...,10$ and generated $\beta_{1j'}$, $j' \in \{v_1, v_2,...,v_u\}$, independent and identically distributed from $N(0, \tau^2)$ with $\tau^2 = 5$.

Figure 3 shows that there is usually some power loss for MURAT compared with SKAT when causal variants are only associated with trait 1. However, the power loss diminishes as ρ_e increases and there is a power gain when ρ_e is 0.7. Hence, when traits are strongly correlated, there is a potential benefit of a multivariate test even if only one trait is truly under genetic influence.
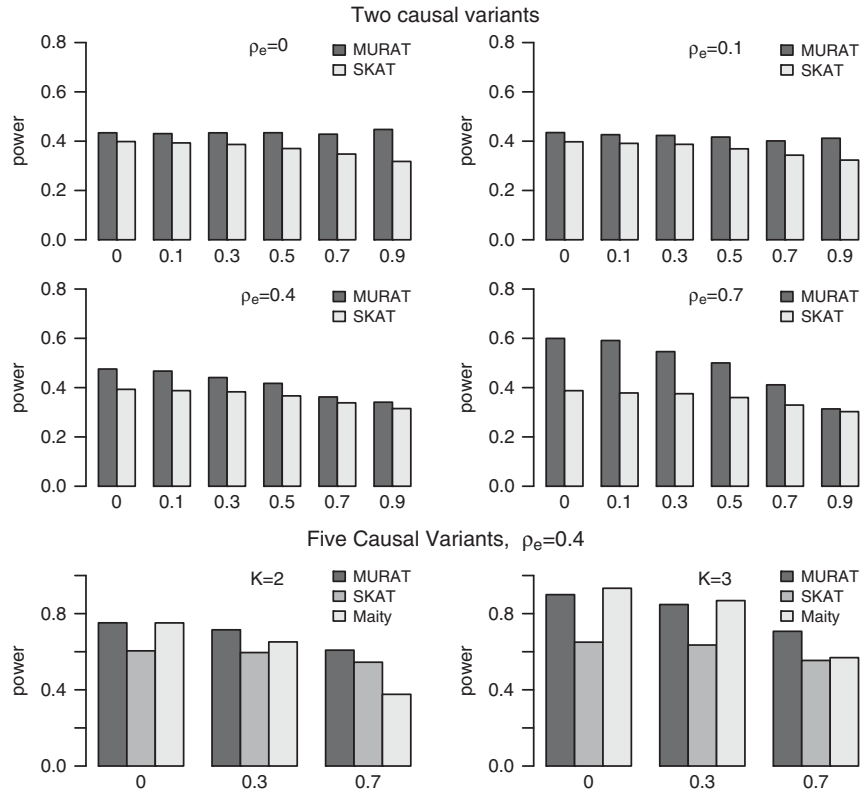
Figure 2 Power comparisons among MURAT, SKAT, and Maity's method under power simulation scenario 1. The top and middle panels show empirical powers of MURAT, which tests associations with two phenotypes simultaneously, versus SKAT, which tests associations with the two phenotypes separately, when two causal variants are associated with both traits. The bottom panel shows empirical powers of MURAT, Maity's test with a linear kernel, and SKAT, when five causal variants are associated with multiple traits. The significance level is 0.05 for all tests and the univariate tests, SKAT, are corrected for multiple comparisons.
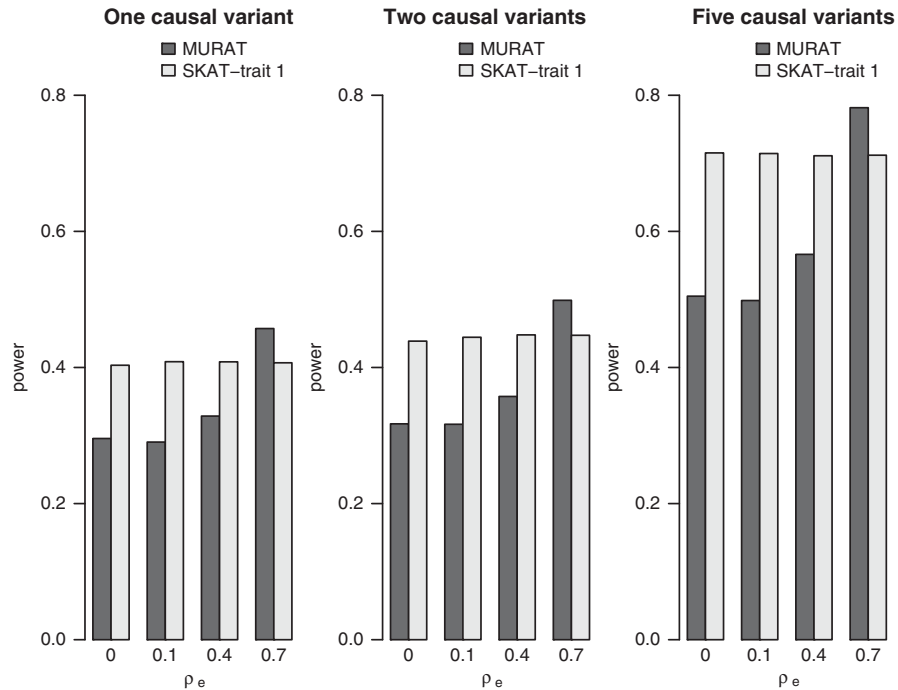
Figure 3 Empirical powers of MURAT versus SKAT for trait 1 at significance level of 0.05. Causal variants are associated with only the first trait. The results for SKAT on trait 1 are not adjusted for multiple testing.

Using scenario 2, we also compared MURAT with Maity's test with a linear kernel when there were five causal variants and $\rho_e = 0.4$. The empirical power for Maity's test (0.64) was higher than for MURAT (0.57) but lower than the power for testing only the first associated trait with SKAT (0.71).

Bone mineral density and exome-sequencing variants

We have analyzed bone mineral density (BMD) phenotypes and tested for genetic associations with sequence-identified variants in data from the TwinsUK participants (www.twinsuk.ac.uk) in the UK10K consortium. The UK10K project (http://www.uk10k.org/) is a whole-genome and exome-sequencing study focused on the genetic architecture of complex diseases. This analysis is intended to demonstrate the performance of our methods and is not definitive; investigations of genetic associations with BMD in larger sample sizes have been undertaken by the GEFOS consortium.[20]

In our analysis, we selected measures of BMD at two positions: lumbar spine (LS) and femoral neck (FN), in 1005 individuals. These are strongly correlated traits with a correlation of 0.685. We performed gene-based analysis on variants identified by exome sequencing on all autosomes, therefore testing association with BMD at 19 123 gene-based sets. All analyses included two covariates, weight and age, as these strongly affect BMD. As in our simulation studies, we separately standardized the 1005 measurements of LS and FN before applying MURAT.

We compared the strategy of simultaneously testing LS and FN for association using MURAT, with separate tests of LS and FN using SKAT. As the analysis included both common and rare variants, we applied SKAT with the weighted linear kernel and set the MURAT weight matrix as $W = diag\{w_1,...,w_p\}$, where $(w_j)^{1/2} = \beta(MAF_j, 1, 25)$ is the weight for the jth variant in a particular gene as suggested by Wu *et al*.[3]

The Bonferroni-corrected significance threshold for 19 123 genes is around $0.05/19\,123 = 2.61 \times 10^{-6}$. As this correction is stringent and the sample size is small, in Table 3 we list all genes that have P-values $< 1 \times 10^{-5}$, from either MURAT or SKAT (SKAT P-values here are not adjusted for multiple comparisons). For genes *SLC17A5* and *C20orf187*, the MURAT P-values are smaller than those obtained from SKAT and the difference is notable at *C20orf187*. In addition, we applied Maity's test to *C20orf187* and obtained a P-value of 8×10^{-6}, supporting the benefit of a multivariate test and the potential benefits of MURAT. Gene *SNX17* showed association with only one of the bone phenotypes; the result with MURAT was less significant than the better of the two univariate tests; similar results were seen in our simulations. However, MURAT was only slightly less significant than the better univariate result.

To adjust single phenotype results for multiple testing, MURAT region-based P-values are compared with adjusted SKAT results in the right panel of Figure 4, clearly showing that many MURAT P-values tend to be smaller than the adjusted SKAT P-values. In addition, we also show the corresponding Q–Q plots for MURAT and adjusted SKAT, respectively, in the left panel of Figure 4. The MURAT Q–Q plot indicates a small amount of inflation from the null distribution. In contrast, SKAT suffers from conservativeness for large P-values (ie, earlier departure from the null distribution) and later inflation that exceeds MURAT. Hence, MURAT has the potential to improve test power over single phenotype tests by reducing adjustments for multiple testing in the tail of the distribution.

Several genes are well-known to contain variants that influence BMD.[21] Our results did not identify these genes as significant, either because of the limited sample size or due to performing region-based analysis of windows containing too many non-associated variants. In Supplementary S.4, we report the region-based P-values at 15 well-replicated bone-density-associated genes; MURAT has smaller or equivalent P-values at 11 of these 15 genes.

The computation time for running MURAT on all 22 autosomes was manageable. On chromosome 1, for example, 2004 genes were analyzed and computational time was ~ 6 h using one processor on a node equipped with 64 GB RAM and two Intel Xeon CPUs E5645 at 2.40 GHz. In contrast, analysis of the 241 genes on chromosome 21 took only 52 min with MURAT. The runtimes for each chromosome are summarized in Supplementary Table S2.

DISCUSSION

We have developed a new method, MURAT, to detect association between multiple correlated continuous phenotypes and a set of rare genetic variants. Our test can be applied to studies where features of a complex disease are measured by multiple correlated traits. Similarly, MURAT may be also beneficial for studies in which multiple questionnaires assess psychological or behavioral phenotypes. Through simulations, we show that MURAT has the potential to increase test power. As for most other rare-variant association tests, weights can be used in MURAT, to prioritize variants either based on MAFs or external annotation information.

By assuming the variant effects to be random yet correlated, MURAT is based on a linear mixed model that incorporates arbitrary correlations among multivariate phenotypic traits. It is a generalization of SKAT when there exist correlations among observed phenotypes. Through the analytic derivation of the distribution of the P-value, we can avoid the need for permutation tests, except for very small significance thresholds. The computational costs for running our proposed test are much less expensive than for Maity's method

Table 3 Top genes selected from association testing of BMD phenotypes with exome-sequencing variants

| | | | *P-values* | | | |
| | | | *SKAT* | | *MURAT* | |
Chr	Gene	No. of SNPs	FN	LS	FN and LS	ρ_v in MURAT
2	*SNX17*	34	1.02×10^{-1}	2.46×10^{-9}	3.05×10^{-8}	0.5
6	*SLC17A5*	42	2.39×10^{-6}	4.43×10^{-2}	6.12×10^{-7}	0.6
20	*C20orf187*	3	1.12×10^{-2}	1.14×10^{-1}	7.73×10^{-6}	0

Abbreviations: BMD, bone mineral density; FN, femoral neck; LS, lumbar spine; MURAT, Multivariate Rare-Variant Association Test; SKAT, sequence kernel association test; SNP, single-nucleotide polymorphism.
Results with P-values $\leq 1 \times 10^{-5}$ are shown. SKAT tests were not corrected for multiple comparisons. The last column reports the value of the correlation parameter ρ_v that gave the minimum P-value in the MURAT test.
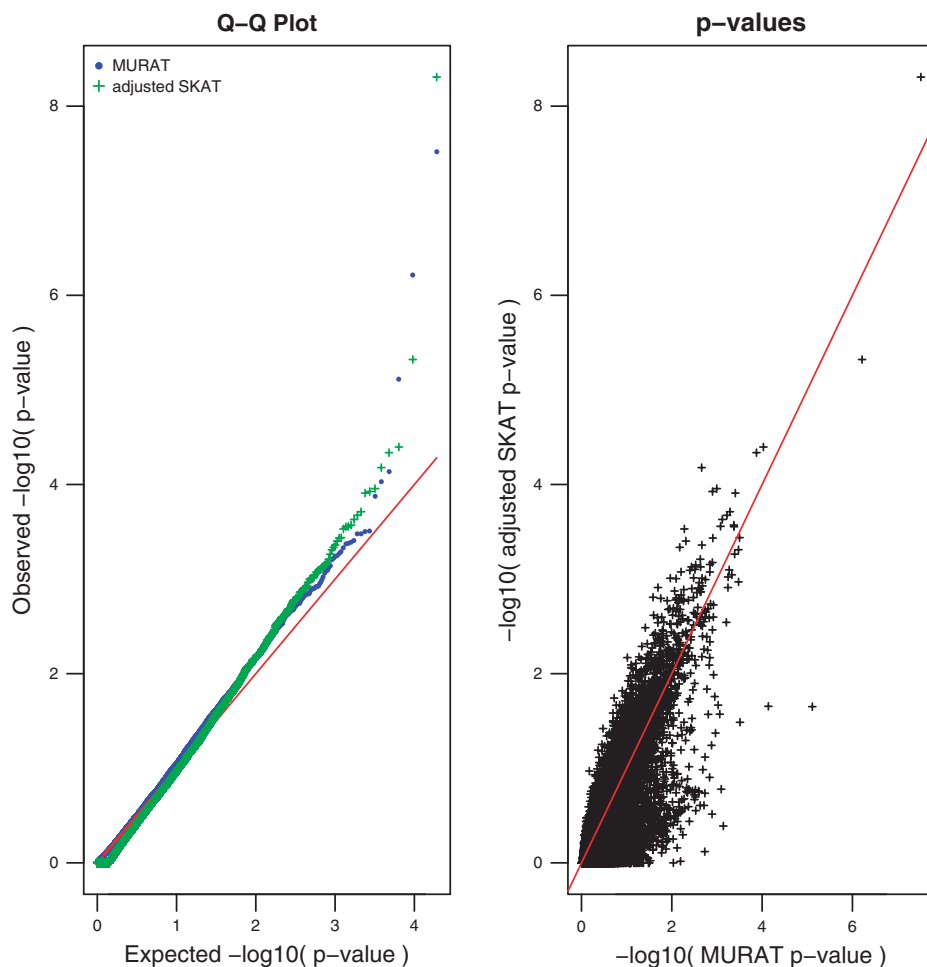
Q–Q Plot

p–values

Figure 4 The left panel shows the *Q–Q* plot for MURAT *P*-values and adjusted SKAT *P*-values on 19 123 genes in UK10K data analysis. The slopes for the MURAT *Q–Q* plot and adjusted SKAT *Q–Q* plot are 1.04 and 1.02, respectively. The right panel shows the comparison of $-\log_{10}$(*P*-values) between MURAT and SKAT tests on each of the 19 123 genes. The SKAT results are corrected for multiple comparisons and the adjusted *P*-values are defined as twice the minimum of the LS- and FN-based *P*-values obtained via SKAT.

(Supplementary S.5). An R package for MURAT can be downloaded from http://greenwoodlab.github.io/software/.

We assumed a common correlation for the effect of the same variant on different traits, that is, the variance–covariance matrix for β is $\Sigma_{\beta} = \tau^2 R \otimes W$, where $R = (1-\rho)I_K + \rho \mathbf{1}_K \mathbf{1}_K^T$. This form for Σ_{β} is a generalization of the correlation structures used in other popular rare-variant association tests also based on mixed models. For example, when $K=1$, the matrix Σ_{β} reduces to $\tau^2 W$, which is exactly the variance used for SKAT with p variants. When $p=1$ and $K>1$, our test becomes a single-variant-based multivariate test and Σ_{β} reduces to a $K \times K$ matrix $\Sigma_{\beta} = \tau^2 R$. Hence, the matrix Σ_{β} in MURAT becomes the same as that used in SKAT-O, if the identity weight matrix is assumed for a set of K variants, as SKAT-O assumes an exchangeable correlation structure for variant effects. If we further assume $\Sigma_e = I_N \otimes \Sigma = I_N \otimes (\sigma^2 I_K)$ in equation (3), that is, the residuals are uncorrelated, then our proposed model (2) reduces to the one used for the SKAT-O test.

The specific correlation matrix structure that we assumed is of course a simplification. A more realistic assumption might allow an arbitrary correlation matrix without any constraints. However, this assumption leads to many parameters and it becomes impractical to

derive a feasible procedure to estimate significance. Therefore, to compromise between biological reality and statistical feasibility, we assumed there is only one common correlation, ρ, for the effects of the same variant. Simulation shows that MURAT power is affected very little by the misspecified correlation matrix for variant effects (Supplementary S.6).

Through extra simulations, we also find that MURAT is robust to different densities of grid for ρ (Supplementary S.7) and direction of variant effect β, especially under scenarios of antagonistic pleiotropy,[22] where one gene is associated with multiple traits; however, the variants are beneficial for some traits and detrimental for others (Supplementary S.8).

It could be argued that observing K phenotypes for an individual is conceptually similar to observing only one phenotype in a family with K members. Although equation (1) could be adapted to univariate-trait family-based designs, the specific assumptions of our model apply only to the multiple phenotype situation. We assume there is a correlation ($0 \leq \rho \leq 1$) between the effects of a single variant on different phenotypes; in a family situation, it would make more sense to assume a variant always has the same effect on different family members ($\rho=1$). As we have previously demonstrated, our test

simplifies in this situation to a score test for a multivariate fixed effect model, where we assume the same effect of each variant across the multiple traits. In families, controlling for relatedness not linked to the locus of interest is necessary and several linear mixed models[23–25] have been proposed that use the kinship matrix or realized relationship matrix to capture these residual genetic correlations. In contrast, in MURAT, the residual correlations in Σ_e could be due to either genetic or other factors. Covariates capturing cryptic population structure (eg, principal components[26]) can be added as fixed effect covariates, if necessary.

A useful extension of our current work might be to generalize the correlated phenotypes from continuous variables to binary ones, or a mixture of these two types, and to extend the proposed multivariate test to family-based studies. Research along this line will be presented in future work.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Li B, Leal SM: Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet* 2008; **83**: 311–321.
2 Madsen BE, Browning SR: A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* 2009; **5**: e1000384.
3 Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X: Rare variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011; **89**: 82–93.
4 Lee S, Wu MC, Lin X: Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 2012; **13**: 762–775.
5 Sun J, Zheng Y, Hsu L: A unified mixed-effects model for rare-variant association in sequencing studies. *Genet Epidemiol* 2013; **37**: 334–344.
6 Burkett K, Greenwood C: A sequence of methodological changes due to sequencing. *Curr Opin Allergy Clin Immunol* 2013; **13**: 470–477.
7 Shriner D: Moving toward system genetics through multiple trait analysis in genome-wide association studies. *Front Genet* 2012; **3**: 1.
8 Bauman LE, Almasy L, Blangero J, Duggirala R, Sinsheimer JS, Lange K: Fishing for pleiotropic QTLs in a polygenic sea. *Ann Hum Genet* 2005; **69**: 590–611.
9 Turnpenny PD, Ellard S: *Emery's Elements of Medical Genetics*. 14th edn. Elsevier/Churchill Livingstone, 2012.
10 Zhu W, Zhang H: Why do we test multiple traits in genetic association studies? *J Korean Stat Soc* 2009; **38**: 1–10.
11 Yang Q, Wang Y: Methods for analyzing multivariate phenotypes in genetic association studies. *J Probab Stat* 2012; **2012**: 652569.
12 Guo X, Liu Z, Wang X, Zhang H: Genetic association test for multiple traits at gene level. *Genet Epidemiol* 2013; **37**: 122–129.
13 Zhao J, Thalamuthu A: Gene-based multiple trait analysis for exome sequencing data. *BMC Proc* 2011; **5** (Suppl 9): S75.
14 Maity A, Sullivan PF, Tzeng JY: Multivariate phenotype association analysis by marker-set kernel machine regression. *Genet Epidemiol* 2012; **36**: 686–695.
15 Zhang D, Lin X: Hypothesis testing in semiparametric additive mixed models. *Biostatistics* 2003; **4**: 57–74.
16 Liu D, Lin X, Ghosh D: Semiparametric regression of multidimensional genetic pathway data: least-squares kernel machines and linear mixed models. *Biometrics* 2007; **63**: 1079–1088.
17 Liu D, Ghosh D, Lin X: Estimation and testing for the effect of a genetic pathway on a disease outcome using logistic kernel machine regression via logistic mixed models. *BMC Bioinformatics* 2008; **9**: 292.
18 Davies RB: Algorithm AS 155: The distribution of a linear combination of 2 random variables. *Appl Stat* 1980; **29**: 323–333.
19 Liu H, Tang Y, Zhang H: A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Comput Stat Data Anal* 2009; **53**: 853–856.
20 Zheng HF, Forgetta V, Hsu YH *et al*: Whole-genome sequence-based identification of EN1 as a determinant of bone density and fracture. *Nature* 2015; **526**: 112–117.
21 Richards JB, Zheng HF, Spector TD: Genetics of osteoporosis from genome-wide association studies: advances and challenges. *Nat Rev Genet* 2012; **13**: 576–588.
22 Williams GC: Pleiotropy, natural selection, and the evolution of senescence. *Evolution* 1957; **11**: 398–411.
23 Chen H, Meigs JB, Dupuis J: Sequence kernel association test for quantitative traits in family samples. *Genet Epidemiol* 2013; **37**: 196–204.
24 Oualkacha K, Dastani Z, Li R *et al*: Adjusted sequence kernel association test for rare variants controlling for cryptic and family relatedness. *Genet Epidemiol* 2013; **37**: 366–376.
25 Schifano ED, Epstein MP, Bielak LF *et al*: SNP set association analysis for familial data. *Genet Epidemiol* 2012; **36**: 797–810.
26 Kang HM, Sul JH, Service SK *et al*: Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 2010; **42**: 348–354.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)