European Journal of Human Genetics (2017) 25, 130–136
© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved 1018-4813/17
www.nature.com/ejhg

ARTICLE

Joint association analysis of a binary and a quantitative trait in family samples

Shuai Wang*,[1], James B Meigs[2,3] and Josée Dupuis[1,4]

In recent years, improved genotyping and sequencing technologies have enabled the discovery of new loci associated with various diseases or traits. For instance, by testing the association with each single-nucleotide variant (SNV) separately, genome-wide association studies (GWAS) have achieved tremendous success in identifying SNVs associated with specific traits. However, little is known about the common genetic basis of multiple traits owing to lack of efficient methods. With the use of extended quasi-likelihood, a Wald test has been proposed to perform a bivariate analysis of a continuous and a binary trait in unrelated samples. However, owing to its low computational efficiency, it has not been implemented in real applications to large-scale genetic studies. In this paper, we propose an efficient bivariate robust score test for two traits, one continuous and one binary, based on extended generalized estimating equations. Our approach is applicable to both family-based and unrelated study designs and can be extended to test the association of multiple traits. Our simulation studies demonstrate the type-I error rate of our approach is well controlled in all minor allele frequency (MAF) scenarios, with MAF ranging from 1 to 30%, and the method is more powerful in certain MAF scenarios than univariate testing with correction for multiple testing. Because of the computational advantage of score tests, our approach is readily applicable to GWAS or sequencing studies. Finally, we present a real application to uncover genetic variants associated with body mass index and type-2 diabetes in the Framingham Heart Study.

European Journal of Human Genetics (2017) **25,** 130–136; doi:10.1038/ejhg.2016.134; published online 26 October 2016

INTRODUCTION

In recent years, univariate association test has been implemented as the predominant statistical method in genetic epidemiology and has yielded fruitful results in many applications. For example, univariate association tests have led to tremendous success in the discovery of disease susceptibility loci when applied to genome-wide association studies (GWAS) for various diseases. However, for the genetic association testing of multiple and often correlated traits, univariate association testing combined with multiple testing correction has usually been implemented owing to the ease of computation. Other variations include MultiPhen[1] and Yang's combination of univariate association tests.[2] However, none of these approaches are as powerful or efficient as a joint multivariate test with each trait treated as a dependent variable in discovering genetic loci associated with all traits under study.[1,3,4]

For example, in the case of two continuous traits assumed to be normally distributed, a joint test can be derived as a simple extension of a univariate normal test. However, if one of the two traits is a discrete trait, for example, a binary trait, deriving such a test becomes challenging, and it further complicates in family samples. One reason is that there is no exact closed form of the likelihood function for a binary trait in family samples. Although applications of linear mixed effects models (LMM) have been frequently used to analyze binary traits in GWAS, researchers have demonstrated that, in the presence of relatedness, LMM results in incorrect type-I error rate owing to the violation of homoscedasticity assumption.[5]

Quasi-likelihood-based approaches, such as generalized estimating equations (GEE), have been proposed to address the question of correlated data.[6–8] GEE has been frequently used to analyze correlated data in univariate association tests such as application to GWAS in families.[9,10] For instance, Wang *et al.*[11] applied GEE to test gene-based and single-nucleotide variant (SNV) association with a single binary trait in family data, assuming that the working correlation matrix is a function of the relationship matrix. When treating the correlation parameters as nuisance parameters, the estimators of GEE have been shown to lack asymptotic efficiency,[12] a common weakness of typical GEE approaches. An improved version of GEE was proposed by Zhao and Prentice,[7,8] in which regression parameters and correlation parameters are estimated simultaneously based on pseudo maximum-likelihood approach. However, the improved efficiency comes at the cost of having to specify a correct covariance structure, and the third and fourth moments are necessary for the estimation.[8,12]

Using principles from the extended quasi-likelihood,[13,14] Hall and Severini[12] established the theory of extended generalized estimating equations (EGEE). Instead of treating correlation parameters as nuisance parameters, EGEE estimates them jointly with the regression parameters and does not require correct specification of a working correlation matrix and therefore only requires up to the second order of moments. Hence, EGEE has been proven to be more powerful, more asymptotically efficient and more computer efficient than GEE while retaining many of its good properties.[12]

Based on the idea of EGEE, Liu *et al.*[15] developed an approach specifically for bivariate genetic analysis. They proposed a joint Wald

[1]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; [2]General Medicine Division, Massachusetts General Hospital, Boston, MA, USA; [3]Department of Medicine, Harvard Medical School, Boston, MA, USA; [4]National Heart, Lung, and Blood Institute (NHLBI) Framingham Heart Study, Framingham, MA, USA
*Correspondence: Dr S Wang, Department of Biostatistics, Boston University School of Public Health, 801 Mass Avenue, 3rd floor, Boston, MA 02118, USA. Tel: +1 434 466 4619; E-mail: tutuwang@bu.edu

Received 1 April 2016; revised 5 July 2016; accepted 6 September 2016; published online 26 October 2016

test to evaluate the association between a SNV and the two traits. The joint Wald test asymptotically follows a chi-squared distribution with two degrees of freedom. However, applications to large-scale genetic studies such as GWAS leads to large computational burden because the parameters have to be estimated first before constructing the test statistic each time a SNV is evaluated for association. Another limitation of EGEE application by Liu *et al.*[15] is that it is only intended for unrelated subjects and hence is not applicable to family data. However, there has been an increasing need for methods suitable for family-based study designs because of the presence of related individuals in many existing cohorts, such as the Framingham Heart Study (FHS) and the Family Heart Study. These family-based studies have enabled the discovery of clinical and genetic risk factors influencing cardiovascular and related diseases' risk and have made great contributions to our current understanding of several complex diseases.

In this paper, we construct a model to accommodate familial correlation, and we propose an efficient robust score test to jointly evaluate the association between a SNV and two traits, one continuous and one binary trait. Moreover, our approach has wider applicability: it can also be applied to test the association with two binary traits or a single binary trait. Our simulation studies demonstrate that the type-I error of our approach is well controlled under all minor allele frequency (MAF) scenarios down to 1% MAF. It is also shown that the score test is more powerful in certain scenarios than the univariate testing corrected for multiple testing. Finally, we present a real application to the FHS by analyzing body mass index (BMI) and type-2 diabetes (T2D) as the two traits of interest and report multiple SNV associations in or near genes with prior implication with one or both of these traits. We also report SNVs in genes that have yet to be implicated in the genetics of these traits and hence represent possible new loci. For implementation of source code, please see http://sites.bu.edu/fhspl/publications/bivaregee/.

METHODS

We first state the assumptions and define the model equations for one continuous and one binary trait in family samples. We assume that there are N independent families ($i=1,\ldots, N$) with a total sample size of n, and the family size (n_i) depends on the family index (i). The model is composed of two simultaneous equations written as:

$$\begin{aligned}&\boldsymbol{Y}_{\mathrm{c}} = \boldsymbol{X}_{\mathrm{c}}\boldsymbol{\beta}_{\mathrm{c}} + \mathbf{G}\beta_{\mathrm{cG}} + \boldsymbol{b} + \boldsymbol{\epsilon}\\ &\mathrm{g}(E[\boldsymbol{Y}_{\mathrm{b}}]) = \boldsymbol{X}_{\mathrm{b}}\boldsymbol{\beta}_{\mathrm{b}} + \mathbf{G}\beta_{\mathrm{bG}}\end{aligned}$$

where the continuous trait $\boldsymbol{Y}_{\mathrm{c}}$ and the binary trait $\boldsymbol{Y}_{\mathrm{b}}$ are $n\times 1$ vectors; $\boldsymbol{X}_{\mathrm{c}}$ is the design matrix for the continuous trait-specific covariates, including an intercept, with a dimension of $n\times p_{\mathrm{c}}$; $\boldsymbol{\beta}_{\mathrm{c}}$ is a $p_{\mathrm{c}}\times 1$ coefficient vector for the intercept and the $(p_{\mathrm{c}}-1)$ covariates; $\boldsymbol{X}_{\mathrm{b}}$ is the design matrix for the binary trait-specific covariates, including the intercept, with a dimension of $n\times p_{\mathrm{b}}$; $\boldsymbol{\beta}_{\mathrm{b}}$ is a $p_{\mathrm{b}}\times 1$ coefficient vector for the intercept and the $(p_{\mathrm{b}}-1)$ covariates; $\mathbf{G}$ is an $n\times 1$ genotype vector for the SNV; β_{cG} and β_{bG} are the corresponding SNV coefficients for the continuous and the binary traits, respectively; and $\mathbf{b}$ is the random intercept following a normal distribution of $N\left(\mathbf{0}, \sigma_{\mathrm{a}}^2\boldsymbol{\Phi}\right)$ with the relationship matrix $\boldsymbol{\Phi}$ being twice the kinship matrix. The vector $\boldsymbol{\varepsilon}$ is a random error term assumed to follow a normal distribution of $N\left(\mathbf{0}, \sigma_{\mathrm{e}}^2\boldsymbol{I}\right)$ where $\boldsymbol{I}$ is the $n\times n$ identity matrix.

We account for within-family correlation by defining the overall variance matrix of the two traits in family blocks as $\boldsymbol{V} = \begin{bmatrix}\boldsymbol{V}_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \boldsymbol{V}_N\end{bmatrix}$ where $\boldsymbol{V_i}$ ($i=1, \ldots, N$) is the variance matrix of the two traits for the ith family with a dimension of $2n_i\times 2n_i$. The within-family covariance matrix has a form $\begin{pmatrix}\boldsymbol{Var}(\boldsymbol{Y}_{\mathbf{c}i}) & \boldsymbol{cov}(\boldsymbol{Y}_{\mathbf{c}i}, \boldsymbol{Y}_{\mathbf{b}i})\\ \boldsymbol{cov}(\boldsymbol{Y}_{\mathbf{c}i}, \boldsymbol{Y}_{\mathbf{b}i})^T & \boldsymbol{Var}(\boldsymbol{Y}_{\mathbf{b}i})\end{pmatrix}$ where $\boldsymbol{Var}(\boldsymbol{Y}_{\mathbf{c}i})$ is the covariance matrix of the continuous trait, $\boldsymbol{cov}(\boldsymbol{Y}_{\mathbf{c}i},\boldsymbol{Y}_{\mathbf{b}i})$ is the covariance matrix between the continuous and the binary trait and $\boldsymbol{Var}(\boldsymbol{Y}_{\mathbf{b}i})$ is the covariance matrix of the binary trait. Because the variance matrix is crucial to the parameter estimation, we further define the individual components of the variance matrix explicitly as follows:

For the ith family, the covariance matrix of the continuous trait is expressed as

$$Var(Y_{\mathrm{c}i}) = \sigma_{\mathrm{a}}^2\boldsymbol{\Phi_i} + \sigma_{\mathrm{e}}^2\boldsymbol{I_i};$$

The covariance matrix of the binary trait $Var(Y_{\mathrm{b}i})$ and the covariance matrix between the continuous and the binary trait $cov(Y_{\mathrm{c}i}, Y_{\mathrm{b}i})$ have the following forms:

$$\begin{aligned}Var(Y_{\mathrm{b}i}) &= \begin{pmatrix}\sqrt{Var(Y_{\mathrm{b}i1})} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sqrt{Var(Y_{\mathrm{b}in_i})}\end{pmatrix}\boldsymbol{R_i}\begin{pmatrix}\sqrt{Var(Y_{\mathrm{b}i1})} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sqrt{Var(Y_{\mathrm{b}in_i})}\end{pmatrix}\\ cov(Y_{\mathrm{c}i}, Y_{\mathrm{b}i}) &= \begin{pmatrix}\sqrt{Var(Y_{\mathrm{c}i1})} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sqrt{Var(Y_{\mathrm{c}in_i})}\end{pmatrix}\boldsymbol{R}_{\mathrm{bc}\boldsymbol{i}}\begin{pmatrix}\sqrt{Var(Y_{\mathrm{b}i1})} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sqrt{Var(Y_{\mathrm{b}in_i})}\end{pmatrix}\end{aligned}$$

where $\boldsymbol{\Phi_i}$ ($i=1, \ldots, N$) is the ith family relationship matrix with a dimension of $n_i\times n_i$ and $\boldsymbol{I_i}$ is the $n_i\times n_i$ identity matrix. We use the same working correlation matrix $\boldsymbol{R_i}=\boldsymbol{\Phi_i}\phi$ (ϕ is an unknown parameter) as in Wang *et al.*[11] with the diagonal elements fixed to 1. The elements of $\boldsymbol{R}_{\mathbf{bc}\boldsymbol{i}}$ ($-1\leq r\leq 1$ is an unknown parameter) are defined as follows ($\forall 1\leq j,\ j'\leq n_i$):

$$\mathrm{r}_{jj'} = \begin{cases} r^{\frac{1}{(\Phi_i)_{jj'}}} & (\Phi_i)_{jj'} \neq 0\\ 0 & (\Phi_i)_{jj'} = 0\end{cases}$$

Where $(\Phi_i)_{jj'}$ is the jj'th element of the relationship matrix $\boldsymbol{\Phi_i}$.

Then, based on the EGEE score equations,[12] $\sum_{i=1}^N \boldsymbol{U_i}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^N\begin{pmatrix}\boldsymbol{D_i}' & \mathbf{0}\\ \mathbf{0} & \boldsymbol{F_i}'\end{pmatrix}\begin{pmatrix}\boldsymbol{V_i}^{-1} & \mathbf{0}\\ \mathbf{0} & \boldsymbol{I}\end{pmatrix}\begin{pmatrix}\boldsymbol{y_i} - \boldsymbol{\mu_i}\\ \boldsymbol{s_i} - \boldsymbol{\sigma_i}\end{pmatrix} = \mathbf{0}$, the Fisher's scoring algorithm is implemented iteratively to update the regression parameters $\boldsymbol{\beta}=(\beta_{\mathrm{c}}, \beta_{\mathrm{c}G}, \beta_{\mathrm{b}}, \beta_{\mathrm{b}G})^T$ and the correlation parameters $\boldsymbol{\alpha} = \left(\sigma_a^2, \sigma_{\mathrm{e}}^2, \phi, r\right)^T$ until some convergence criterion is met.[12,15] The $(m+1)$th iteration equations are:

$$\begin{pmatrix}\boldsymbol{\beta}^{(m+1)}\\ \boldsymbol{\alpha}^{(m+1)}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\beta}^{(m)}\\ \boldsymbol{\alpha}^{(m)}\end{pmatrix} + \boldsymbol{U}^*\left(\boldsymbol{\beta}^{(m)}, \boldsymbol{\alpha}^{(m)}\right)^{-1}\sum_{i=1}^N \boldsymbol{U_i}\left(\boldsymbol{\beta}^{(m)}, \boldsymbol{\alpha}^{(m)}\right)$$

where $\quad \boldsymbol{U}^*\left(\boldsymbol{\beta}^{(m)}, \boldsymbol{\alpha}^{(m)}\right) = -E\left[\mathbf{D}\sum_{i=1}^N U_i\left(\boldsymbol{\beta}^{(m)}, \boldsymbol{\alpha}^{(m)}\right)\right] = \sum_{i=1}^N\begin{pmatrix}\boldsymbol{D_i}'\boldsymbol{V_i}^{-1}\boldsymbol{D_i} & \mathbf{0}\\ \boldsymbol{F_i}'\frac{\partial\boldsymbol{\sigma_i}}{\partial\boldsymbol{\beta}'} & \boldsymbol{F_i}'\frac{\partial\boldsymbol{\sigma_i}}{\partial\boldsymbol{\alpha}'}\end{pmatrix}$[15,16];

$\mathbf{D}f$ denotes the Jacobian of f; $\boldsymbol{D_i}=\frac{\partial\boldsymbol{\mu_i}}{\partial\boldsymbol{\beta}}$ is the stacked matrix with a size of $2n_i\times(p_{\mathrm{c}}+p_{\mathrm{b}}+2)$; $\boldsymbol{F_i} = \frac{\partial\boldsymbol{vec}\left(\boldsymbol{V_i}^{-1}\right)}{\partial\boldsymbol{\alpha}^T}$; and $\boldsymbol{\sigma_i}$ is the vectorized $\boldsymbol{V_i}$. We are estimating both regression parameters $\boldsymbol{\beta}$ and the correlation parameters $\boldsymbol{\alpha}$ simultaneously, while in Wang's method for a single binary trait,[11] the estimates of regression parameters are first updated based on the scoring equations for $\boldsymbol{\beta}$ only, and the correlation parameter ϕ is then updated based on the formula of Pearson residuals.[17] The convergence of Wang's method is solely based on $\boldsymbol{\beta}$. However, the convergence of our novel approach is based on the Euclidean distance between iterations for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$.

Note that when the approach is applied to unrelated samples, it is equivalent to specifying $\boldsymbol{\Phi}_{\mathbf{i}}=\boldsymbol{I}$, $\phi=1$, $\sigma_{\mathrm{a}}^2 = 0$, reducing the score equations above to the form proposed by Liu *et al.*[15]

Robust score test

Breslow[18] developed a score test for overdispersed Poisson regression and other quasi-likelihood models in 1990, and then Guo *et al.*[19] demonstrated its advantage over the sandwich estimator. Following the same rationale, we derive a robust score test to evaluate the null hypothesis of no association between the genotypes and the two traits. Equivalently, we are testing $H_0 : \beta_{cG} = \beta_{bG} = 0$. Note this could be easily extended to analyze two binary traits or a single binary trait.

Let $\boldsymbol{U}^{(\mathbf{1})} = \begin{pmatrix}\boldsymbol{U}_{\boldsymbol{\beta}_{\mathrm{c}}}\\ \boldsymbol{U}_{\boldsymbol{\beta}_{\mathrm{b}}}\end{pmatrix}$ denote the vector of score function with respect to $\boldsymbol{\theta}^{(\mathbf{1})} = \left(\boldsymbol{\beta}_{\boldsymbol{c}}^T, \boldsymbol{\beta}_{\boldsymbol{b}}^T\right)^T$, $\boldsymbol{U}^{(\mathbf{2})} = \begin{pmatrix}U_{\beta_{cG}}\\ U_{\beta_{bG}}\end{pmatrix}$ denote the vector of score function with respect to $\boldsymbol{\theta}^{(2)} = (\beta_{cG}, \beta_{bG})^T$ and let $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\alpha}}_0$ denote the parameter estimates under H_0. We propose the following score test statistic:

$$S = \left(\mathbf{A}\left(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\alpha}}_0\right)\mathbf{U}\left(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\alpha}}_0\right)\right)^T \left\{\mathbf{A}\left(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\alpha}}_0\right) \sum_{i=1}^{N} \left[\boldsymbol{U}_i\left(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\alpha}}_0\right) U_i^T\left(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\alpha}}_0\right)\right] \mathbf{A}^T\left(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\alpha}}_0\right)\right\}^{-1} \mathbf{A}\left(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\alpha}}_0\right)\mathbf{U}\left(\hat{\boldsymbol{\theta}}_0, \hat{\boldsymbol{\alpha}}_0\right).$$

where $\boldsymbol{A} = \left(-\boldsymbol{U}^*_{\theta^{(2)}\theta^{(1)}}\left(\boldsymbol{U}^*_{\theta^{(1)}\theta^{(1)}}\right)^{-1}, \quad \boldsymbol{I}\right)$; $\boldsymbol{U} = \begin{pmatrix}\boldsymbol{U}^{(1)}\\ \boldsymbol{U}^{(2)}\end{pmatrix} = \sum_{i=1}^{N} \boldsymbol{U}_i = \sum_{i=1}^{N}\begin{pmatrix}\boldsymbol{U}_i^{(1)}\\ \boldsymbol{U}_i^{(2)}\end{pmatrix}$; $\boldsymbol{U}^*$ is as previously defined; and $\boldsymbol{I}$ is the 2×2 identity matrix. (see Appendix for derivation details). The proposed test statistic asymptotically follows a χ^2_2 (termed as 'BivarEGEE'). When the covariance structure is correctly specified,[18] that is, $E\left[\boldsymbol{U}\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}\right)\boldsymbol{U}^T\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}\right)\right] = -E\left[\frac{\partial \boldsymbol{U}\left(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}\right)}{\partial \boldsymbol{\theta}^T}\right]$, the variance formula of $\boldsymbol{U}^{(2)}$ will reduce to $\boldsymbol{U}^*_{22} - \boldsymbol{U}^*_{21}\boldsymbol{U}^{*-1}_{11}\boldsymbol{U}^*_{12}$ (the subscript 1 and 2 corresponds to $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$, respectively). The test statistic with this restriction is termed as 'BivarEGEE$_R$'.

Simulations

We conduct simulation studies to evaluate the validity of our approach to test the association between SNVs with different MAF and two traits. We also compare the power of our approach to a univariate approach to determine under which circumstances it is more powerful.

Type-I error

We compare the type-I error rate of our approach to the minimum *P*-value obtained from the univariate association testing for each trait with Bonferroni correction for multiple testing of two traits ('minP'). We simulate the traits under the null hypothesis that there is no genetic association with any of the two traits, that is, $H_0: \beta_{cG} = \beta_{bG} = 0$. We simulate 8 SNV scenarios with MAF ranging from 0.01 to 0.3. For each SNV and trait scenario, we simulate 50 000 replicates and calculate the proportion of simulations reaching the significance threshold of 0.001. In each replicate, we simulate a total of 1000 independent nuclear families with 2 parents and the number of children randomly determined from a discrete uniform distribution ranging from 1 to 4, so that family size ranges from 3 to 6 members. Within each family, we simulate the genotypes of the parents under Hardy–Weinberg equilibrium, and the children's genotypes using random allele dropping. We also simulate two covariates: age and sex. Given a family, the sex of the offspring is randomly assigned and we simulate age in the following way: we first simulate the age of the youngest adult offspring from a continuous uniform distribution ranging from 30 to 50, additional offspring's ages are set to be within 5 years of the first one with at least a 1-year gap so that the possibility of them being twins is excluded. The mother is assumed to be 20–45 years older than all her offspring, and the father's age is set to be within 5-year of the mother's age and he must be at least 20 years older than his oldest offspring. We then simulate two continuous traits influenced by age and sex only, based on the following two equations, so that age and sex explains around 4.5 and 5.4% of the total variance of y_1 *versus* 11 and 0.9% of y_2:

$$y_1 = 0.025\text{age} + 0.5\text{sex} + \varepsilon_1;$$
$$y_2 = 0.04\text{age} + 0.2\text{sex} + \varepsilon_2;$$

where $\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\end{pmatrix} \sim N(0, \boldsymbol{\Sigma_a} \otimes \boldsymbol{\Phi} + \boldsymbol{\Sigma_e} \otimes \mathbf{I})$, the additive covariance matrix is $\boldsymbol{\Sigma_a} = \begin{pmatrix}0.5 & 0.25\\ 0.25 & 0.5\end{pmatrix}$ and the environmental covariance matrix is $\boldsymbol{\Sigma_e} = \begin{pmatrix}0.5 & 0.25\\ 0.25 & 0.5\end{pmatrix}$.

We transform y_2 to a binary variable using a threshold model with a disease prevalence of 30%, assuming a disease with a high prevalence such as obesity or hypertension in older adults. Based on the same trait and covariates data set, in each replicate, we compute the 'minP' as follows: we conduct univariate association testing on y_1 and the transformed binary version of y_2, select the smaller *P*-value, and then multiply it by a factor of 2 (Bonferroni's correction). In both approaches, the type-I error rate is defined as the proportion of replicates with *P*-value$<$0.001.

Power simulation

We compare the power of our approach to the minimum *P*-value obtained from univariate tests (minP) under the same scenarios (Table 1) and the same family structure as above. In addition to the effects of sex and age, we include an additively coded genetic variant to the model, so that the traits are simulated under the alternative hypothesis that there is an association between the genotypes and each of the two traits:

$$y_1 = 0.025\text{age} + 0.5\text{sex} + m\sqrt{\frac{1\%}{2\text{MAF}(1-\text{MAF})}}G + \varepsilon_1;$$

$$y_2 = 0.04\text{age} + 0.2\text{sex} + \sqrt{\frac{1\%}{2\text{MAF}(1-\text{MAF})}}G + \varepsilon_2;$$

where m is used to model the relative strength of association and takes values of -0.5, -0.1, 0.1 and 0.5 under different scenarios; and ε_1 and ε_2 follow the same normal distribution as for the type-I error simulations. We adjust the correlation parameter ρ (=0.2, 0.5 or 0.8) in the additive covariance matrix $\boldsymbol{\Sigma_a} = \begin{pmatrix}0.5 & 0.5\rho\\ 0.5\rho & 0.5\end{pmatrix}$ to reflect different correlation magnitude between the two traits. We set $\boldsymbol{\Sigma_e}$ equal to $\boldsymbol{\Sigma_a}$, except in the last two scenarios (the bottom row in Figure 1), where the covariance term in $\boldsymbol{\Sigma_e}$ is set to be negative.

For each scenario, we simulate 1000 replicates and then compute the power as the proportion of simulations reaching the significance threshold of 0.0001, a threshold that gives a good range of power for the methods compared.

Framingham Heart Study

One important motivation for developing the model and proposing the score test statistic is to provide a computationally efficient approach applicable to large-scale genetic studies such as GWAS, exome sequencing or whole genome sequencing (WGS) studies. In the application section, we perform a genome-wide association of BMI and T2D in the FHS, to better understand the common genetic basis of these two traits.

The FHS was initiated in 1948 and is a longitudinal study consisting of three generations of cohorts: the Original cohort, the Offspring cohort and the third generation (Gen 3) cohort, totaling 14, 428 participants. Some participants were recruited from the same household, and hence are related. Over the years, research efforts in FHS have been rewarded with fruitful results in identifying risk factors of cardiovascular-related traits such as blood pressure and cholesterol levels, as well as glycemic and other metabolic traits.

Obesity is an important risk factor in the development of T2D.[20,21] By applying our approach to BMI, a continuous variable, and T2D, a binary variable, on a genome-wide scale, we hope to better understand their common genetic basis. In our analyses, both traits are adjusted for age and sex.

We analyze the association between these two traits and genotypes from the Framingham SNP Health Association Resource (SHARe) project sponsored by the National Heart, Lung and Blood Institute (NHLBI). Genotypes from Affymetrix 500K genotyping arrays (Affymetrix, Santa Clara, CA, USA), supplemented by the Affymetrix MIPS array, were available on 8481 participants after exclusion for low call rate ($<$97%), heterozygosity rate outside of 5 SDs from the mean or excess Mendelian errors ($>$1000). Additional SNVs were imputed with the software MACH (Markov Chain-based haplotyper) using the HapMap 2 reference haplotypes.[22]

Table 1 Type-I error simulation results

MAF	*BivarEGEE (SD)*	*BivarEGEE$_R$(SD)*	*minP (SD)*
0.01	0.0006 (1.1×10^{-4})	0.0006 (1.1×10^{-4})	0.0009 (1.3×10^{-4})
0.02	0.0006 (1.1×10^{-4})	0.0007 (1.2×10^{-4})	0.0009 (1.3×10^{-4})
0.03	0.0007 (1.2×10^{-4})	0.0007 (1.2×10^{-4})	0.0009 (1.3×10^{-4})
0.04	0.0009 (1.3×10^{-4})	0.0009 (1.3×10^{-4})	0.0009 (1.3×10^{-4})
0.05	0.001 (1.4×10^{-4})	0.010 (1.4×10^{-4})	0.0009 (1.3×10^{-4})
0.1	0.001 (1.4×10^{-4})	0.0011 (1.5×10^{-4})	0.0008 (1.3×10^{-4})
0.2	0.001 (1.4×10^{-4})	0.001 (1.4×10^{-4})	0.001 (1.4×10^{-4})
0.3	0.001 (1.4×10^{-4})	0.001 (1.4×10^{-4})	0.0009 (1.3×10^{-4})

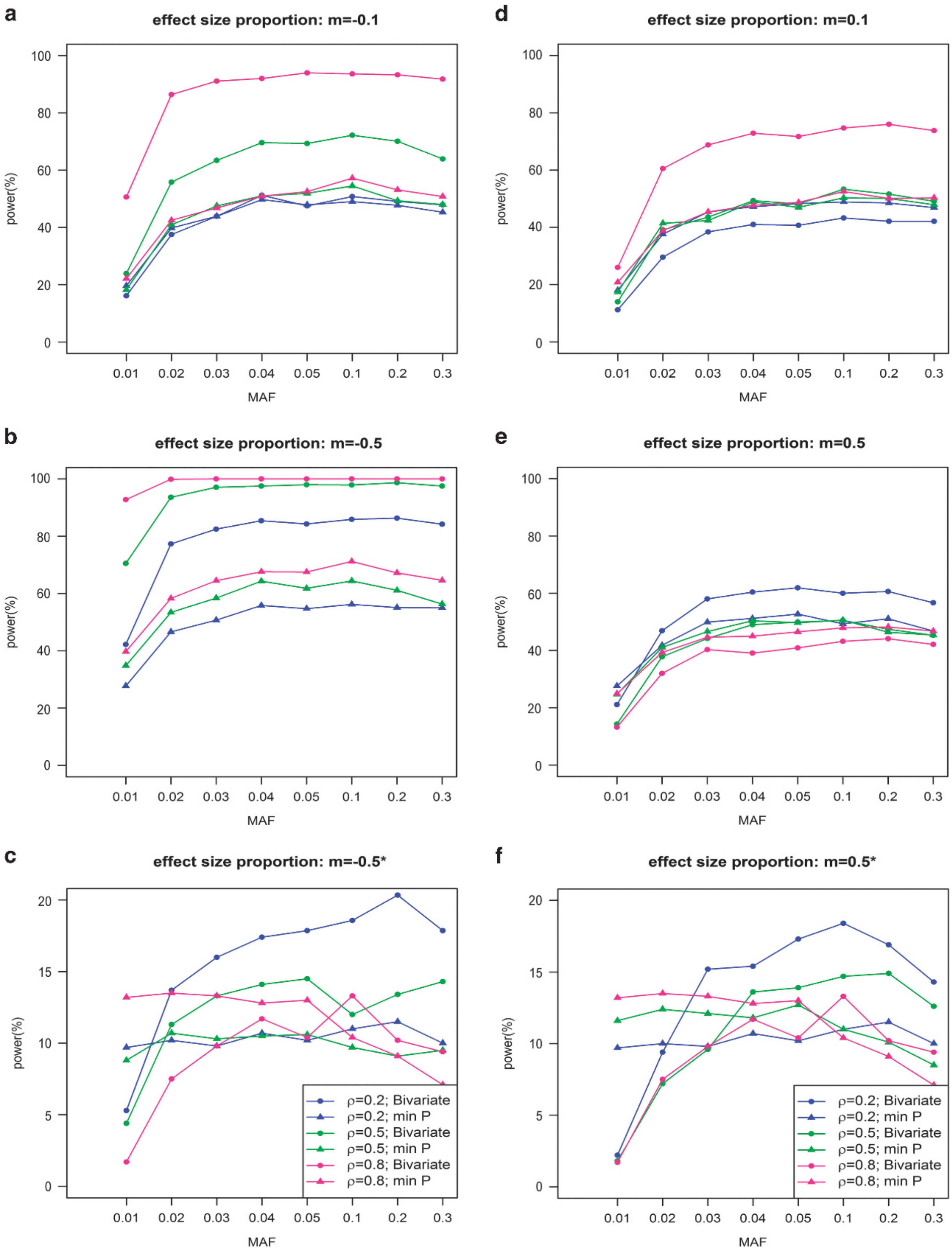

Figure 1 Power (*y* axis) as a function of MAF (*x* axis). Different trait correlation (ρ) values are distinguished by different color lines, and different effect size proportion (m) are presented in each panel: (**a**) $m=-0.1$; (**b**) $m=-0.5$; (**c**) $m=-0.5$ (negative environmental covariance); (**d**) $m=0.1$; (**e**) $m=0.5$; (**f**) $m=-0.5$ (negative environmental covariance).

RESULTS

Type-I error

Simulation results show that the type-I error rate of our proposed approach ('BivarEGEE') is well controlled in all MAF scenarios where MAF ranges from 0.01 to 0.3 (Table 1). We also provide the type-I error rate when the variance structure is assumed to be correctly specified ('BivarEGEE$_R$'). The fact that both approaches yield the same type-I error rate in all MAF scenarios is a good indication that the variance structure is correctly modeled. The type-I error rate of the minP approach is also well controlled at $\alpha=0.001$.

Power simulations

The results of power simulations are presented in Figure 1. The results suggest that when the two untransformed traits have opposite direction of association with the SNV, our proposed approach is consistently more powerful. The highest power gain from BivarEGEE over minP reaches 40%. In the scenarios where both traits have the same direction of association, the power gain differs depending on the relative association strength m and the correlation ρ. For instance, when $m=0.1$, BivarEGEE is more powerful or as powerful as minP when the two untransformed traits are strongly or moderately correlated ($\rho=0.8$ or 0.5), while the power slightly decreases when the two traits have a weak correlation ($\rho=0.2$). When $m=0.5$, BivarEGEE is at least as powerful as minP when the two traits have a weak or moderate correlation, while with increased correlation, the power tends to suffer some small loss. When the covariance term of the environmental covariance matrix $\boldsymbol{\Sigma_e}$ is set to be negative, our approach is consistently more powerful for common variants (MAF > 0.02).

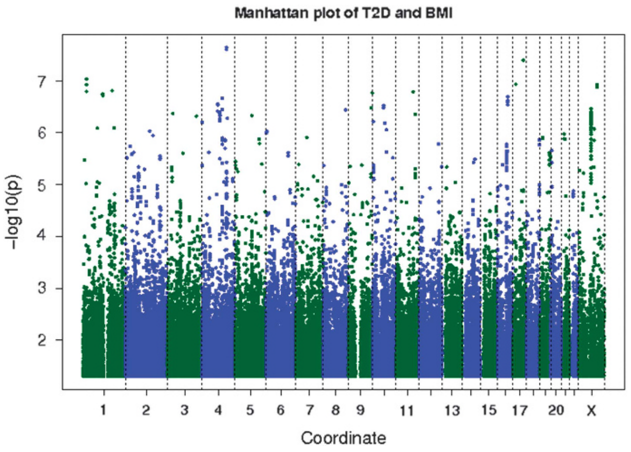

Figure 2 GWAS results for the 23 chromosomes using the FHS SHARe 550 k genotype data. The *y* axis is the −log10-transformed *P*-value, and the *x* axis represents the coordinates of the SNVs on the 23 chromosomes.

Application to the FHS

We apply our approach to study the genome-wide association between genetic variants from the Framingham SHARe and the combination of BMI and T2D status in FHS participants. A total of 7038 genotyped and phenotyped participants in 1185 families are analyzed after participants with missing traits or without genotypes are omitted. Both traits are adjusted for age and sex. We present the genome-wide association results as the minus logarithm base 10 of the *P*-value in Figure 2 and also provide a list of the top 20 SNVs with the smallest *P*-values in Table 2. Three SNVs reach the GWAS significance threshold of 5×10^{-8}, including the top 2 SNVs from chromosome 4, near the height-associated gene *HHIP*.[23] The chromosome 4-associated SNVs are also near *TMEM154*, a T2D-associated gene identified by the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium in 2014.[24]

Among the remaining top 20 SNVs, chromosome 16 SNVs (rs8059849, rs9931529, rs13332434, rs9783765) are near *FTO*, a gene known for its association with both BMI and T2D.[24–29] The SNV rs10894188 (chromosome 11) is near *MTNR1B*, a gene known to be associated with both T2D and obesity-related traits;[26] rs12097783 (chromosome 1) is near previously identified BMI gene *SEC16B*;[29–32] rs11145958 (chromosome 9) is near *GPSM1*, a T2D-associated gene;[33] 5 SNVs on chromosome 1 are near *NOTCH2*[25] and *ADAM30*,[25] two genes known for SNVs associated with T2D; rs17863929 (chromosome 4) is approximately 3 Mb away from *IL2*,[34] a gene known for SNVs in the intron region associated with type-1 diabetes.

DISCUSSION

We propose a novel approach to test the association between a genetic variant and two traits, at least one of which is binary, in family samples, based on EGEE. Our approach can handle a range of families,

Table 2 Top 20 SNVs of SHARe GWAS of BMI and T2D

rsID	*dbSNP (GRCh38)*	*P-value*	*Previously associated trait: nearest gene*	*P-value (BMI)*	*P-value (T2D)*
rs17363126	Chr4: g.141595017T>C	2.3×10^{-8}	Height: HHIP T2D: TMEM154	0.09	1.7×10^{-8}
rs17459397	Chr4: g.141541225T>C	2.5×10^{-8}	Height: HHIP T2D: TMEM154	0.07	4.7×10^{-9}
rs8066504	Chr17: g.64334251C>T	4.0×10^{-8}	NA	0.06	1.4×10^{-8}
rs16825415	Chr1: g.21494365G>A	9.3×10^{-8}	NA	0.60	2.3×10^{-8}
rs4545864	Chr17: g.16682354G>C	1.2×10^{-7}	NA	0.21	2.4×10^{-8}
rs16825373	Chr1: g.21469182C>T	1.2×10^{-7}	NA	0.63	2.8×10^{-8}
rs12097783	Chr1: g.171941188C>G	1.6×10^{-7}	BMI: SEC16B	0.53	4.3×10^{-8}
rs2316510	Chr1: g.21442519G>T	1.6×10^{-7}	NA	0.71	3.7×10^{-8}
rs10894188	Chr11: g.99957388G>C	1.6×10^{-7}	Obesity-related traits: MTNR1B T2D: MTNR1B	0.80	3.2×10^{-8}
rs11145958	Chr9: g.136360621G>A	1.7×10^{-7}	T2D: GPSM1	0.48	1.9×10^{-8}
rs12023850	Chr1: g.115262772G>A	1.8×10^{-7}	T2D: NOTCH2 ADAM30	0.10	5.1×10^{-8}
rs12031246	Chr1: g.115262711C>T	1.9×10^{-7}	NA	0.10	5.4×10^{-8}
rs4839429	Chr1: g.115259437C>T	1.9×10^{-7}	NA	0.10	5.3×10^{-8}
rs17033541	Chr1: g.115259165G>A	1.9×10^{-7}	NA	0.10	5.4×10^{-8}
rs17033538	Chr1: g.115259019T>C	1.9×10^{-7}	NA	0.10	5.5×10^{-8}
rs8059849	Chr16: g.60021262A>G	2.0×10^{-7}	T2D and BMI: FTO	0.66	8.4×10^{-8}
rs9931529	Chr16: g.60025332A>C	2.0×10^{-7}	T2D and BMI: FTO	0.65	8.4×10^{-8}
rs13332434	Chr16: g.60041028A>C	2.1×10^{-7}	T2D and BMI: FTO	0.64	8.5×10^{-8}
rs17863929	Chr4: g.117796608C>A	2.2×10^{-7}	T1D: IL2	0.008	3.6×10^{-8}
rs9783765	Chr16: g.60012023A>G	2.4×10^{-7}	T2D and BMI: FTO	0.56	1.3×10^{-7}

Table 3 Computational time (in seconds) for BivarEGEE with different sample sizes and family structures[a]

	1148	*2235*	*4508*	*9027*	*18101*
N(Nfam)[b]	*(250)*	*(500)*	*(1000)*	*(2000)*	*(4000)*
$\text{Time}^{\text{Bivar}}$ (s)	7	14	32	64	168
$\text{Time}^{\text{Cont}}$ (s)	0.6	5	32	190	1470
Time^{bin} (s)	2.4	6	13	31	70

$\text{Time}^{\text{Cont}}$ denotes score test in famskat.[35]
[a]The time includes both estimation of nuisance parameter and computation of test statistic and *P*-value.
[b]N(Nfam): simulated with varying family size (3–6 family members randomly determined) for each family.

including large and complex pedigrees. Using simulation studies, we demonstrate that our approach has well-controlled type-I error rate in all the scenarios evaluated and is more powerful than univariate tests adjusted for multiple testing in certain scenarios.

Our approach is based on extended quasi-likelihood. Fisher's scoring algorithm is implemented for parameter estimation. It is worth noting that we model the covariance matrix of the binary and continuous traits as a function of the kinship matrix. Moreover, we propose to use a conditional correlation matrix to account for the correlation between the two traits, which is novel. All these features lead to a computer-efficient implementation that allows for genome-wide applications. In the simulation studies, our unrestricted approach ('BivarEGEE') has similar type-I error rate as the restricted version ('BivarEGEE$_R$'), so we are confident that the covariance structure is correctly modeled in our approach. However, 'BivarEGEE' is more flexible, because it has no additional restrictions on the covariance structure of the traits. Using a similar framework, our approach can be easily extended to the analysis of two binary traits or a single binary trait, for which R functions and sample codes are also available on the webpage. The approach should readily be extendable to genetic analysis of three or four traits simultaneously. However, extensions to >4 traits might add complexity to the model and implementation.

Although our approach is based on joint estimation and testing, it is computer efficient. Table 3 lists computing time when applied to data with different family structure and sample size, including parameter estimation under the null hypothesis, computing the test statistic and *P*-value on a single node of Intel(R) Xeon(R) CPU E5-2640 0 @ 2.50 GHz Linux machine. As a score test, the parameter estimation is performed only once under the null hypothesis prior to application to a large-scale genetic study, such as GWAS. The computational time for minP is also listed in Table 3. It takes approximately half the time to analyze a single binary trait compared with that to analyze the two traits jointly. The time it takes to analyze a continuous trait using famskat[35] increases exponentially with the sample size. By contrast, it is not computationally affordable to apply the Wald test proposed by Liu *et al*.[15] to a large-scale genetic study, because the parameters always have to be re-estimated each time a new SNV is tested for association.

Bivariate genetic association testing is not new, but it has not been extensively applied, due to various limitations or non-availability of the existing methods and software. In this paper, we develop a bivariate approach, BivarEGEE, and we apply our approach to a real data set and found interesting associations. For instance, we replicate some loci close to relevant genes known to have impact on both traits, such as *FTO* and *MTNR1B*. One novel region (chr1:115,259,019-115,262,711 using GRCh38) on chromosome 1 was among our top findings; however, no prior T2D or BMI associations have been reported in this region. Replication from an independent study using our approach or other multivariate methods is needed to determine whether this finding is spurious or a real replicable association that we have identified using BivarEGEE and would have been undetectable without a powerful bivariate analytic approach. It is worth noting that our approach is not purely driven by the more significantly associated trait. For example, rs1558902 (*FTO*, chromosome 16) is the most significantly associated SNV with BMI ($P=2.6\times10^{-9}$) but is not associated with T2D ($P=0.20$). The overall *P*-value of rs1558902 with both traits ($P=1.7\times10^{-6}$) does not reach the GWAS significance threshold.

Current GWAS often involve meta-analysis of independent studies in a consortium, because meta-analysis can greatly increase sample size and power. In the future, we aim to develop meta-analysis method for the BivarEGEE approach. This will provide a more powerful bivariate approach to study two traits that commonly occur in human physiology and disease and offers a powerful approach to identify novel SNV associations with multiple correlated traits.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the National Heart, Lung and Blood Institute's Framingham Heart Study (Contract No. N01-HC-25195 & HHSN268201500001I) and its contract with Affymetrix, Inc for genotyping services (Contract No. N02-HL-6-4278), and by grants from the National Institute of National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) R01 DK078616, NIDDK K24 DK080140 and American Diabetes Association Mentor-Based Postdoctoral Fellowship Award #7-09-MN-32.

1 O'Reilly PF, Hoggart CJ, Pomyen Y *et al*: MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS One* 2012; **7**: e34861.
2 Yang Q, Wu H, Guo C, Fox CS: Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. *Genet Epidemiol* 2010; **34**: 444–454.
3 Lange C, Van Steen K, Andrew T *et al*: A family-based association test for repeatedly measured quantitative traits adjusting for unknown environmental and/or polygenic effects. *Stat Appl Genet Mol Biol* 2004; **3**: 1–27.
4 Klei L, Luca D, Devlin B, Roeder K: Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genet Epidemiol* 2008; **32**: 9–19.
5 Chen H, Wang C, Conomos MP *et al*: Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *Am J Hum Genet* 2016; **98**: 653–666.
6 Zeger SL, Liang K, Albert PS: Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988; **44**: 1049–1060.
7 Prentice RL, Zhao LP: Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics* 1991; **47**: 825–839.
8 Zhao LP, Prentice RL: Correlated binary regression using a quadratic exponential model. *Biometrika* 1990; **77**: 642–648.
9 Kathiresan S, Manning AK, Demissie S *et al*: A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med Genet* 2007; **8**: 1.
10 Levy D, Larson MG, Benjamin EJ *et al*: Framingham Heart Study 100K project: Genome-wide associations for blood pressure and arterial stiffness. *BMC Med Genet* 2007; **8**: 1.
11 Wang X, Lee S, Zhu X, Redline S, Lin X: GEE-Based SNP set association test for continuous and discrete traits in family-based association studies. *Genet Epidemiol* 2013; **37**: 778–786.
12 Hall DB, Severini TA: Extended generalized estimating equations for clustered data. *J Am Stat Assoc* 1998; **93**: 1365–1375.
13 Nelder JA, Pregibon D: An extended quasi-likelihood function. *Biometrika* 1987; **74**: 221–232.
14 McCullagh P, Nelder JA: *Generalized Linear Models*. CRC Press, 1989; Vol 37.
15 Liu J, Pei Y, Papasian CJ, Deng H: Bivariate association analyses for the mixture of continuous and binary traits with the use of extended generalized estimating equations. *Genet Epidemiol* 2009; **33**: 217–227.
16 Hall DB: On the application of extended quasi-likelihood to the clustered data case. *Can J Stat* 2001; **29**: 77–97.
17 McCullagh P, Nelder JA, McCullagh P: *Generalized Linear Models*. London: Chapman and Hall, 1989; Vol 2.

18 Breslow N: Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *J Am Stat Assoc* 1990; **85**: 565–571.
19 Guo X, Pan W, Connett JE, Hannan PJ, French SA: Small-sample performance of the robust score test and its modifications in generalized estimating equations. *Stat Med* 2005; **24**: 3479–3495.
20 Chan JM, Rimm EB, Colditz GA, Stampfer MJ, Willett WC: Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. *Diabetes Care* 1994; **17**: 961–969.
21 Colditz GA, Willett WC, Stampfer MJ *et al*: Weight as a risk factor for clinical diabetes in women. *Am J Epidemiol* 1990; **132**: 501–513.
22 Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR: MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* 2010; **34**: 816–834.
23 Lango Allen H, Estrada K, Lettre G *et al*: Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010; **467**: 832–838.
24 DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) ConsortiumAsian Genetic Epidemiology Network Type 2 Diabetes (AGEN-T2D) ConsortiumSouth Asian Type 2 Diabetes (SAT2D) Consortium*et al*: Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* 2014; **46**: 234–244.
25 Zeggini E, Scott LJ, Saxena R *et al*: Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008; **40**: 638–645.
26 Voight BF, Scott LJ, Steinthorsdottir V *et al*: Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010; **42**: 579–589.
27 Perry JR, Voight BF, Yengo L *et al*: Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. *PLoS Genet* 2012; **8**: e1002741.
28 Willer CJ, Speliotes EK, Loos RJ *et al*: Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 2009; **41**: 25–34.
29 Berndt SI, Gustafsson S, Magi R *et al*: Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet* 2013; **45**: 501–512.
30 Wen W, Zheng W, Okada Y *et al*: Meta-analysis of genome-wide association studies in east asian-ancestry populations identifies four new loci for body mass index. *Hum Mol Genet* 2014; **23**: 5492–5504.
31 Monda KL, Chen GK, Taylor KC *et al*: A meta-analysis identifies new loci associated with body mass index in individuals of african ancestry. *Nat Genet* 2013; **45**: 690–696.
32 Speliotes EK, Willer CJ, Berndt SI *et al*: Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 2010; **42**: 937–948.
33 Hara K, Fujita H, Johnson TA *et al*: Genome-wide association study identifies three novel loci for type 2 diabetes. *Hum Mol Genet* 2014; **23**: 239–246.
34 Barrett JC, Clayton DG, Concannon P *et al*: Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009; **41**: 703–707.
35 Chen H, Meigs JB, Dupuis J: Sequence kernel association test for quantitative traits in family samples. *Genet Epidemiol* 2013; **37**: 196–204.

APPENDIX

Below we show how the coefficient matrix **A** of the score function is derived.

Let $\boldsymbol{U}^{(3)}$ denote the score function vector with respect to the correlation parameter vector α, and $\boldsymbol{U}^{(1)}$, $\boldsymbol{U}^{(2)}$ is as defined in the methods section. We apply the first-order Taylor expansion around α, $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}_{\mathbf{0}}^{(2)}$ to the score function vector $\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}^{(1)} \\ \boldsymbol{U}^{(2)} \\ \boldsymbol{U}^{(3)} \end{pmatrix}$, substitute the estimates from Fisher's algorithm $\widehat{\boldsymbol{\theta}_{\mathbf{0}}^{(1)}}$, $\widehat{\boldsymbol{\alpha}_{\mathbf{0}}}$, $\boldsymbol{\theta}_{\mathbf{0}}^{(2)}$ at H_0 and thus we obtain the following equations:

$$\mathbf{0} = \boldsymbol{U}^{(1)}(\widehat{\boldsymbol{\theta}_{\mathbf{0}}^{(1)}}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \hat{\boldsymbol{\alpha}}_{\mathbf{0}}) \approx \boldsymbol{U}^{(1)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{0}^{(2)}, \alpha) + \frac{\partial \boldsymbol{U}^{(1)}}{\partial \boldsymbol{\theta}^{(1)}}(\widehat{\boldsymbol{\theta}_{\mathbf{0}}^{(1)}} - \boldsymbol{\theta}^{(1)}) + \frac{\partial \boldsymbol{U}^{(1)}}{\partial \boldsymbol{\alpha}}(\widehat{\boldsymbol{\alpha}_{\mathbf{0}}} - \boldsymbol{\alpha})$$

$$\mathbf{0} = \boldsymbol{U}^{(3)}(\widehat{\boldsymbol{\theta}_{\mathbf{0}}^{(1)}}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \hat{\boldsymbol{\alpha}}_{\mathbf{0}}) \approx \boldsymbol{U}^{(3)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{0}^{(2)}, \alpha) + \frac{\partial \boldsymbol{U}^{(3)}}{\partial \boldsymbol{\theta}^{(1)}}(\widehat{\boldsymbol{\theta}_{\mathbf{0}}^{(1)}} - \boldsymbol{\theta}^{(1)}) + \frac{\partial \boldsymbol{U}^{(3)}}{\partial \boldsymbol{\alpha}}(\widehat{\boldsymbol{\alpha}_{\mathbf{0}}} - \boldsymbol{\alpha}).$$

Using the principle of Fisher's scoring algorithm by replacing the second-order derivative with its expectation, we get

$$\begin{aligned}
\begin{pmatrix} \widehat{\boldsymbol{\theta}_{\mathbf{0}}^{(1)}} - \boldsymbol{\theta}^{(1)} \\ \widehat{\boldsymbol{\alpha}_{\mathbf{0}}} - \boldsymbol{\alpha} \end{pmatrix} &\approx \left[-E \begin{pmatrix} \frac{\partial \boldsymbol{U}^{(1)}}{\partial \boldsymbol{\theta}^{(1)}} & \frac{\partial \boldsymbol{U}^{(1)}}{\partial \boldsymbol{\alpha}} \\ \frac{\partial \boldsymbol{U}^{(3)}}{\partial \boldsymbol{\theta}^{(1)}} & \frac{\partial \boldsymbol{U}^{(3)}}{\partial \boldsymbol{\alpha}} \end{pmatrix} \right]^{-1} \begin{pmatrix} \boldsymbol{U}^{(1)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \boldsymbol{\alpha}) \\ \boldsymbol{U}^{(3)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \boldsymbol{\alpha}) \end{pmatrix} \\
&= (\boldsymbol{U}^{*})^{-1} \begin{pmatrix} \boldsymbol{U}^{(1)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{0}^{(2)}, \boldsymbol{\alpha}) \\ \boldsymbol{U}^{(3)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{0}^{(2)}, \boldsymbol{\alpha}) \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{U}_{11}^{*} & \mathbf{0} \\ \boldsymbol{U}_{31}^{*} & \boldsymbol{U}_{33}^{*} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{U}^{(1)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{0}^{(2)}, \boldsymbol{\alpha}) \\ \boldsymbol{U}^{(3)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{0}^{(2)}, \boldsymbol{\alpha}) \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{U}_{11}^{*-\mathbf{1}} & \mathbf{0} \\ -\boldsymbol{U}_{33}^{*-\mathbf{1}} \boldsymbol{U}_{31}^{*} \boldsymbol{U}_{11}^{*-1} & \boldsymbol{U}_{33}^{*-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{U}^{(1)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \boldsymbol{\alpha}) \\ \boldsymbol{U}^{(3)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \boldsymbol{\alpha}) \end{pmatrix}.
\end{aligned}$$

Next we substitute these equations into the equation for the first-order Taylor expansion of $U^{(2)}$ around α, $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}_{\mathbf{0}}^{(2)}$ and replace the first-order derivative by its expectation to obtain the following equation evaluated at:

$$\begin{aligned}
\widehat{\boldsymbol{\theta}_{\mathbf{0}}^{(1)}}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \hat{\boldsymbol{\alpha}}_{\mathbf{0}} : \boldsymbol{U}^{(2)}(\widehat{\boldsymbol{\theta}_{\mathbf{0}}^{(1)}}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \hat{\boldsymbol{\alpha}}_{\mathbf{0}}) &\approx \boldsymbol{U}^{(2)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \boldsymbol{\alpha}) \\
&\quad + E\left[\frac{\partial \boldsymbol{U}^{(2)}}{\partial \boldsymbol{\theta}^{(1)}}\right](\widehat{\boldsymbol{\theta}_{\mathbf{0}}^{(1)}} - \boldsymbol{\theta}^{(1)}) + E\left[\frac{\partial \boldsymbol{U}^{(2)}}{\partial \boldsymbol{\alpha}}\right](\widehat{\boldsymbol{\alpha}_{\mathbf{0}}} - \boldsymbol{\alpha}) \\
&= \boldsymbol{U}^{(2)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \alpha) \\
&\quad + \left(E\left[\frac{\partial \boldsymbol{U}^{(2)}}{\partial \boldsymbol{\theta}^{(1)}}\right] E\left[\frac{\partial \boldsymbol{U}^{(2)}}{\partial \boldsymbol{\alpha}}\right] \right) \begin{pmatrix} \boldsymbol{U}_{11}^{*-1} & \mathbf{0} \\ -\boldsymbol{U}_{33}^{*-\mathbf{1}} \boldsymbol{U}_{31}^{*} \boldsymbol{U}_{11}^{*-\mathbf{1}} & \boldsymbol{U}_{33}^{*-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{U}^{(1)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \boldsymbol{\alpha}) \\ \boldsymbol{U}^{(3)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \boldsymbol{\alpha}) \end{pmatrix} \\
&= \boldsymbol{U}^{(2)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \boldsymbol{\alpha}) \\
&\quad + \begin{pmatrix} -\boldsymbol{U}_{21}^{*} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{U}_{11}^{*-\mathbf{1}} & 0 \\ -\boldsymbol{U}_{33}^{*-\mathbf{1}} \boldsymbol{U}_{31}^{*} \boldsymbol{U}_{11}^{*-\mathbf{1}} & \boldsymbol{U}_{33}^{*-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{U}^{(1)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \boldsymbol{\alpha}) \\ \boldsymbol{U}^{(3)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \boldsymbol{\alpha}) \end{pmatrix} \\
&= \boldsymbol{U}^{(2)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{\mathbf{0}}^{(2)}, \boldsymbol{\alpha}) - \boldsymbol{U}_{21}^{*} \boldsymbol{U}_{11}^{*-\mathbf{1}} \boldsymbol{U}^{(1)}(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}_{0}^{(2)}, \boldsymbol{\alpha}).
\end{aligned}$$

Hence $\mathbf{A} = \left(-\boldsymbol{U}_{\boldsymbol{\theta}^{(2)}\boldsymbol{\theta}^{(1)}}^{*} (\boldsymbol{U}_{\boldsymbol{\theta}^{(1)}\boldsymbol{\theta}^{(1)}}^{*})^{-\mathbf{1}}, \quad \boldsymbol{I} \right)$.