USAT: A Unified Score-based Association Test for Multiple Phenotype-Genotype Analysis

Debashree Ray[1], **James S. Pankow**[2], and **Saonli Basu**[1]

[1]Division of Biostatistics, School of Public Health, University of Minnesota, USA

[2]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, USA

Abstract

Genome-wide Association Studies (GWASs) for complex diseases often collect data on multiple correlated endo-phenotypes. Multivariate analysis of these correlated phenotypes can improve the power to detect genetic variants. Multivariate analysis of variance (MANOVA) can perform such association analysis at a GWAS level, but the behavior of MANOVA under different trait models has not been carefully investigated. In this paper, we show that MANOVA is generally very powerful for detecting association but there are situations, such as when a genetic variant is associated with all the traits, where MANOVA may not have any detection power. In these situations, marginal model based methods, however, perform much better than multivariate methods. We investigate the behavior of MANOVA, both theoretically and using simulations, and derive the conditions where MANOVA loses power. Based on our findings, we propose a unified score-based test statistic USAT that can perform better than MANOVA in such situations and nearly as well as MANOVA elsewhere. Our proposed test reports an approximate asymptotic p-value for association and is computationally very efficient to implement at a GWAS level. We have studied through extensive simulations the performance of USAT, MANOVA and other existing approaches and demonstrated the advantage of using the USAT approach to detect association between a genetic variant and multivariate phenotypes. We applied USAT to data from three correlated traits collected on 5, 816 Caucasian individuals from the Atherosclerosis Risk in Communities (ARIC, The ARIC Investigators [1989]) Study and detected some interesting associations.

Keywords

GWAS; MANOVA; Multiple correlated phenotypes; Multivariate analysis; Score-based test; SSU

Corresponding Author: Saonli Basu, Division of Biostatistics, School of Public Health, University of Minnesota, saonli@umn.edu, Phone: 612-624-2135, Fax: 612-626-0660.

1 Introduction

In the study of a complex disease, data on several correlated endo-phenotypes are often collected to get a better understanding of the disease. For example, in the study of thrombosis, the intermediate correlated phenotypes such as Factor VII, VIII, IX, XI, XII, and von Willebrand factor influence greatly the risk of developing thrombosis [Souto et al., 2000; Germain et al., 2011]. An epidemiological study on type 2 diabetes (T2D) typically collects data on a number of risk factors and diabetes-related quantitative traits. The standard approach to analyze these phenotypes is to perform single-trait analyses separately and report the findings for individual trait.

van der Sluis et al. [2013] demonstrated several alternative models which would benefit from a joint analysis. Blair et al. [2013] illustrated the comorbidity between Mendelian disorders and different complex disorders, which indicates that there may be common genetic variants affecting several of these complex traits. Recently, many articles advocating joint analysis over univariate analysis of multiple correlated traits [Ferreira and Purcell, 2009; Zhang et al., 2009; Korte et al., 2012; O'Reilly et al., 2012; Stephens, 2013; Aschard et al., 2014; Galesloot et al., 2014; Zhou and Stephens, 2014; Ried et al., 2014, and references therein] have been published that illustrate the benefits of jointly analyzing these correlated traits to improve the power of detection of genetic variants. Moreover this joint analysis could reveal some pleiotropic genes involved in the biological development of the disease.

Few approaches have been developed to perform association analysis with multivariate traits at a GWAS level. O'Reilly et al. [2012] proposed MultiPhen to detect association between multivariate traits and a single nucleotide polymorphism (SNP) with unrelated individuals. MultiPhen uses ordinal regression to regress a SNP on a collection of phenotypes and tests whether all regression parameters corresponding to the phenotypes in the model are significantly different from zero. It can accommodate both binary and continuous traits but may suffer from lack of power when a SNP is associated with all the highly correlated traits. van der Sluis et al. [2013] proposed Trait-based Association Test (TATES) for testing association between multiple traits and multiple SNPs using extended Simes' procedure on the p-values derived from univariate trait and single SNP association analysis. Even when the phenotypes are strongly correlated, TATES gives appropriate type I error for varying minor allele frequency (m.a.f.). It may have low power when a SNP affects only a few of the strongly correlated traits. Maity et al. [2012] proposed a kernel machine method for unrelated individuals for joint analysis of multimarker effects on multiple traits. Kernel machine is a powerful dimension-reduction tool that can accommodate linear/non-linear effects of multiple SNPs. Their test for association between multiple SNPs and the phenotypes is equivalent to testing the variance components in a multivariate linear mixed model (mvLMM). Implementation of this approach requires parametric bootstrapping to estimate the distribution of the test statistic and could be computationally intensive at a GWAS level. Korte et al. [2012]; Zhou and Stephens [2012] implemented mvLMM for GWAS. Zhou and Stephens [2014] explored efficient algorithms for mvLMM in a GWAS setting.

Recently, data reduction methods such as principal component analysis (PCA) and canonical correlation analysis (CCA) are being explored to perform multivariate association analysis [Tang and Ferreira, 2012; Basu et al., 2013; Aschard et al., 2014]. The advantage of using CCA to perform gene-based tests on multivariate phenotypes has been elaborately discussed in Tang and Ferreira [2012]; Basu et al. [2013]. Previously, Ferreira and Purcell [2009] proposed a multivariate test of association based on CCA to simultaneously test the association between a single SNP and multiple phenotypes. Their CCA approach is equivalent to multivariate analysis of variance (MANOVA) or more generally the Wilk's lambda test in multivariate multiple linear regression (MMLR) approach [Muller and Peterson, 1984]. Basu et al. [2013] extended the MANOVA to family data. Both O'Reilly et al. [2012] and van der Sluis et al. [2013] found significantly high power for MANOVA when a subset of traits was associated with the causal variant or gene. One major advantage of MANOVA is that it can easily be extended to incorporate multiple phenotypes as well as multiple SNPs (such as a gene). Moreover other covariates can easily be incorporated in the model.

In this paper, we explore the performance of MANOVA to detect multi-trait association under various alternative trait models. Our simulation studies consider a single marker to investigate the properties of MANOVA. Further, we theoretically justify the behavior of MANOVA and provide a geometrical explanation as well. We demonstrate that MANOVA may lose significant power when the genetic marker is associated with all the traits and any test that does not consider the within trait correlation can have more power in such a situation. In such a scenario, methods based on marginal models have better power than MANOVA. Stephens [2013], too, emphasizes that "multivariate association analyses are often most advantageous when not all phenotypes are associated with the genetic variant being tested!" Utilizing these findings, we propose a novel unified score-based association test (USAT), which considers an optimal weighted combination of the multivariate score test (MANOVA) and a score-based test from marginal models. USAT maintains good power under various alternative trait models and performs significantly better than MANOVA when all the traits are associated.

This paper evolves as follows. Section 2 describes some popular existing methods for doing association analysis using multiple phenotypes. More specifically, section 2.1 describes the univariate methods that completely ignore trait correlations, section 2.2 describes a method that accounts for the within trait correlation only through the distribution of the test statistic while section 2.3 describes a multivariate method that directly incorporates the trait correlation structure. Section 2.4 theoretically and geometrically justifies some aspects of the behavior of MANOVA, for K traits and a single SNP, in situations that commonly arise in such genetic studies. Section 2.5 introduces our unified approach USAT for association analysis using multiple traits and a single marker for unrelated individuals. Section 3 illustrates a comparison of different existing approaches and USAT using simulated data and a real dataset. Section 4 concludes this article with a short summary and discussion.

2 Methods

Consider K correlated traits $Y_1, Y_2, \ldots, Y_K$ in n unrelated individuals. Let Y_k be the $n \times 1$ vector of k-th trait and Y be the $n \times K$ matrix of traits for all individuals. Consider a GWAS setting with data on a large number $p \ (\gg n)$ of genetic variants. We are interested in testing the association of a single SNP with the K correlated traits. For a given SNP, let X_i be the number of copies of minor alleles (0, 1 or 2) for i-th individual and X be the $n \times 1$ vector of genotypes for all samples. Without loss of generality, it is assumed that the phenotype matrix Y and the genotype vector X are centered but not standardized.

Due to the correlatedness of the traits, a standard approach would be to consider an MMLR model for the association test of K traits and the SNP:

$$Y_{n \times K} = X_{n \times 1} \beta'_{1 \times K} + \mathscr{E}_{n \times K} \quad (1)$$

where $\beta = (\beta_1, \ldots, \beta_K)$ is the vector of fixed unknown genetic effects corresponding to the K correlated traits, and $\mathscr{E}$ is the matrix of random errors. For testing that the SNP is not associated with any of the K traits, the null hypothesis of interest is $H_0 : \beta = 0$.

In the MMLR model (1), each row of $\mathscr{E}$ is i.i.d. with mean $0_{K \times 1}$ and variance $\Sigma_{K \times K}$. In particular, $\mathscr{E}$ may be assumed to be an $n \times K$ normal data matrix from $N_K(0, \Sigma)$, where Σ is a positive definite (p.d.) matrix representing residual covariance among the traits. The likelihood ratio test (LRT) of H_0 based on the MMLR model with matrix normal errors is equivalent to MANOVA [Muller and Peterson, 1984; Yang and Wang, 2012]. One may consider a further partition of $\mathscr{E}$ to arrive at mvLMM:

$$Y_{n \times K} = X_{n \times 1} \beta'_{1 \times K} + W_{n \times K} + \varepsilon_{n \times K}$$

where W is a matrix of random effects representing heritable component of the phenotypes, and ε is the matrix of errors characterizing random variation arising from unmeasured sources. In recent times, mvLMMs have been recognized as powerful tools for testing H_0. mvLMM can not only control population structure, sample relatedness and other confounding factors, but can also account for dependence among multiple traits. Association tests based on mvLMM can be computationally challenging and many efficient algorithms have been developed to this end [Yang et al., 2011; Korte et al., 2012; Zhou and Stephens, 2014].

Apart from multivariate models, one may use marginal/univariate models for such an association test. Although marginal modeling effectively assumes the traits to be uncorrelated, approaches based on marginal models are often computationally faster and easier to implement. The marginal model for testing association of a SNP with k-th trait is given by

$$Y_k = \beta_{M,k} X + e_k, \ e_k \sim N(0, \sigma^2 I_n), \ k = 1, 2, \ldots, K \quad (2)$$

$\beta_{M,k}$ is the k-th genetic effect in the marginal model. For the k-th marginal model, our null hypothesis is $H_{0,k} : \beta_{M,k} = 0$. In order to carry out the simultaneous test H_0, one still needs to devise an approach to combine the results from the marginal tests $H_{0,k}$, $k = 1, 2, \ldots, K$.

Broadly, the different statistical approaches for testing our global null hypothesis of no association can be classified into three categories: (1) tests that completely ignore the within trait correlation; (2) tests that incorporate within trait correlation only in deriving the distribution of the test statistic; and (3) tests that incorporate the within trait correlation directly in deriving the test statistic. We compare through extensive simulation studies these three broad approaches and discuss their advantages and shortcomings under various alternative trait models.

2.1 Combination Tests that completely ignore within trait correlation

This category of tests considers separate regression models for the K traits (i.e., K univariate analyses), thereby treating the traits as uncorrelated. Let p_k be the p-value for testing $H_{0,k}$ based on the k-th marginal model in (2). This class of tests proposes several approaches of combining the p-values $p_1, \ldots, p_K$ for testing our global null hypothesis $H_0 = \cap_{k=1}^{K} H_{0,k}$.

2.1.1 Fisher's Test—Fisher's method [Fisher, 1925] involves combining the logarithmic transformation of the p-values $p_1, \ldots, p_K$. The test statistic is $-2\log_e \sum_{k=1}^{K} p_k$, which under H_0 and the assumption of independent tests, has a χ^2_{2K} distribution. In the presence of strong correlation among traits, inflated type-I error is observed ('anti-conservative').

2.1.2 minP Test—The minP test statistic is based on the minimum of adjusted p-values, where adjustment is usually done by Bonferroni's method to take care of multiple-testing issue. It is given by $p_{\min} = \min_{k=1}^{K} \{K p_k\}$. Under H_0 and the assumption of independence among the phenotypes, $p_{\min}$ is distributed as the minimum of independent $U(0, 1)$ variables. In the presence of correlation structure, this test can be conservative. To take care of this conservativeness, van der Sluis et al. [2013] proposed TATES which combines p-values from univariate analyses while correcting for the relatedness among the phenotypes.

2.2 Test that incorporates trait correlation only through distribution

This category of tests does not explicitly consider the trait correlation in the test statistics. The correlation is taken into account in finding the true null distribution of the test statistic due to which the statistic maintains proper type I error. A notable test in this category is the Sum of Squared Score (SSU) test as outlined by Yang and Wang [2012], an extension of the SSU test for association of multiple SNPs with a single trait proposed by Pan [2009].

2.2.1 SSU Test—SSU is a score-based test where the score vector is derived from the marginal normal models in equation (2). Under the global null H_0, the $K \times 1$ vector of marginal scores is given by

$$U_M = \frac{1}{\hat{\sigma}_0^2} Y' X$$

where $\hat{\sigma}_0^2 = \frac{1}{K(n-1)} \sum_{i=1}^{n} \sum_{k=1}^{K} Y_{ik}^2$ is the maximum likelihood estimate (MLE) of σ^2 in equation (2) under the null. The SSU test statistic is $T_S = U'_M U_M$, which has an approximate asymptotic scaled and shifted chi-squared distribution $a\chi_d^2 + b$ [Zhang, 2005] under H_0. The distributional parameters are determined as

$$a = \frac{\sum c_k^3}{\sum c_k^2}, \quad b = \sum c_k - \frac{(\sum c_k^2)^2}{\sum c_k^3}, \quad d = \frac{(\sum c_k^2)^3}{(\sum c_k^3)^2} \quad (3)$$

where $\{c_k\}_{k=1}^{K}$ are the ordered eigenvalues of $\mathrm{Cov}(U_M) = X' X Y' Y / (n\hat{\sigma}_0^4)$.

An important aspect of the SSU test is that the test statistic does not incorporate the trait covariance structure. Notice that, according to equation (3), $\mathrm{Cov}(U_M)$ contains information on within trait correlations and is used in deriving the distribution of the statistic. If U be the score vector from MMLR model (1) under H_0, a test statistic of the form $U'U$ will not be an SSU type test since the within trait covariance matrix is incorporated in U.

2.3 Multivariate Test that incorporates within trait correlation directly in the test statistic

This class of tests explicitly incorporates the within trait correlation structure in the test statistics as well as in finding their distributions.

2.3.1 MANOVA—Consider the MMLR model in equation (1). Assume each row of $\mathscr{E}$ to be i.i.d. $N_K(\mathbf{0}, \mathbf{\Sigma})$. The log-likelihood for the data matrix Y is given by

$$l(\boldsymbol{\beta}, \textstyle\sum) = -\frac{1}{2} n \log |2\pi \textstyle\sum| - \frac{1}{2} \mathrm{tr} \left\{ \textstyle\sum^{-1} (Y - X\boldsymbol{\beta}')' (Y - X\boldsymbol{\beta}') \right\} \quad (4)$$

For testing H_0, the LRT is equivalent to the MANOVA test statistic (Wilk's Lambda), which is the ratio of generalized variances $|E|/|H + E|$. Here H is the hypothesis sum of squares and cross product (SSCP) matrix and E is the error SSCP matrix. The explicit forms of these SSCP matrices in terms of phenotype and genotype data are $H = \hat{\boldsymbol{\beta}}(X'X)\hat{\boldsymbol{\beta}}'$ and $E = Y'Y - \hat{\boldsymbol{\beta}}(X'X)\hat{\boldsymbol{\beta}}'$, where $\hat{\boldsymbol{\beta}} = Y'X(X'X)^{-1}$ is the MLE of $\boldsymbol{\beta}$. Thus, H is calculated as the covariance matrix of the fitted values, and E is calculated as the covariance matrix of the residuals of the model. Under H_0, $-2\log\Lambda = -n\log(|E|/|H + E|)$ has an approximate asymptotic χ_K^2 distribution.

Another such multivariate approach is MultiPhen where the genotype is modeled as ordinal using a proportional odds regression model. O'Reilly et al. [2012] empirically showed that for a single SNP, MultiPhen's performance is similar to MANOVA. For testing the global null of no association, whether one treats Y as a multivariate response and X as a predictor

(as done in MANOVA) or treats X as a univariate response and Y as multiple predictors (as done in MultiPhen), both approaches lead to same p-values [Anderson, 1984; Stephens, 2013].

2.4 MANOVA and its behavior

A major challenge in multivariate disease-related trait analysis is the lack of a test that is uniformly most powerful under different patterns/levels of association and different within trait correlation structures. The association tests which do not consider within trait correlation at all are either 'conservative' or 'anti-conservative'. Our simulation studies with compound symmetry (CS) correlation structure show that MANOVA generally has better performance but loses significant power when within trait correlation is high and is in the same direction as all the genetic effects. For a moderate number of traits, MANOVA may fail to detect pleiotropy (phenomenon where a single genetic variant affects all the traits) even at low within trait correlations (refer sections 3.1, 3.2).

The following theorems provide conditions under which MANOVA loses power when a SNP is associated with all K correlated traits. We assume a CS residual correlation structure. Theorem proofs are provided in Appendix S1.

Theorem 1: Consider the MMLR model $Y_{n \times K} = X_{n \times 1} \beta'_{1 \times K} + \mathscr{E}_{n \times K}$ with $vec(\mathscr{E}) \sim N_{nK}(\mathbf{0}, I_n \otimes \Sigma)$, $\Sigma = \sigma^2 ((1 - \rho)I_K + \rho \mathbf{1}\mathbf{1}')$, $\sigma^2 > 0$, $\rho (> 0)$ is the within trait correlation such that Σ is a p.d. matrix, and $\beta = (\beta_1, \ldots, \beta_K)$ is the vector of genetic effects. Assume that the genetic effects of the associated traits are equal in size and in positive direction. Consider two scenarios of association: 'partial association' (when the SNP is associated with u $(< K)$ traits), and 'complete association' (when all K traits are associated). For testing $H_0 : \beta = \mathbf{0}$, the power of MANOVA under partial association will be asymptotically more than that

under complete association if $\frac{u}{K} > \frac{1-\rho}{1+(K-u-1)\rho} = \frac{2nd\ eigenvalue\ of\ \sum_{K-u}}{1st\ eigenvalue\ of\ \sum_{K-u}}$. Here Σ_{K-u} is the CS residual covariance matrix of the $K - u$ truly unassociated traits.

For $K = 2$ traits, Theorem 1 can be generalized further to encompass genetic effects in opposite direction, and negative within trait correlation.

Theorem 2: Consider the MMLR model in Theorem 1 with $K = 2$ traits. The genetic effects of the associated traits may or may not be equal in size or in same direction. The within trait correlation ρ may or may not be positive. For testing $H_0 : \beta_1 = \beta_2 = 0$, the power of MANOVA when only one trait is associated is asymptotically more than when both traits are associated if $0 < \beta_2 < 2\rho\beta_1$ or $0 > \beta_2 > 2\rho\beta_1$.

Corollary 1: In particular, let us assume that the genetics effects of the associated traits are equal in size. That is, $|\beta_1| = |\beta_2|$ when the SNP is associated with both the correlated traits. Asymptotically, the power of MANOVA under $H_{a1} : \beta_1 > 0, \beta_2 = 0$ will exceed the power of MANOVA under

    i. $H_{a2,1} : \beta_1 = \beta_2 > 0$ *when* $\rho > 1/2$;

ii. $H_{a2,2} : \beta_1 = -\beta_2 > 0$ *when* $\rho < -1/2$.

The theoretical 95% acceptance regions of SSU and MANOVA for $K = 2$ correlated traits in Figure 1 provide a geometrical explanation of the above theorems. The acceptance region of SSU is drawn using the marginal scores $U_{M,1}$ and $U_{M,2}$. MANOVA's acceptance region is drawn using the 2 components Z_1 and Z_2 of vector $\boldsymbol{Z}$ since MANOVA is asymptotically equivalent to the test $\boldsymbol{Z'I(0)Z}$. Here $\boldsymbol{Z}$ is an $N(\boldsymbol{0}, \boldsymbol{I(0)}^{-1})$ variable and $\boldsymbol{I(0)}$ is Fisher Information matrix under $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$. Details of this equivalence and the acceptance region plots are provided in Appendix S2. For SSU, a high true value of $\beta_1(\beta_2)$ will be reflected by a high value of $U_{M,1}(U_{M,2})$. In Figure 1, observe that the SSU acceptance regions are almost circular in shape irrespective of correlation ρ. With increase in ρ, the shape of the acceptance region remains same. Only the size increases a little which causes slight loss in power to reject H_0. For MANOVA, a high true value of $\beta_1(\beta_2)$ will be reflected by a high value of $Z_1(Z_2)$. When $\rho \to 1$, notice that the acceptance region for MANOVA becomes elongated along the direction of $\boldsymbol{1}$ vector in Figure 1. Recall that for a CS correlation matrix, the eigen vector corresponding to the largest eigenvalue (for $\rho > 0$) is along the direction of $\boldsymbol{1}$ vector. When the true genetic effect sizes are equal and in the same direction, the corresponding components of $\boldsymbol{Z}$ are equal as well and they will lie on vector $\boldsymbol{1}$. This suggests that the Z's (and hence the non-zero genetic effects) need to be really large to cross the MANOVA acceptance region boundary for high ρ. The black box in Figure 1 represents such a situation, and it arises when the SNP is associated with both the correlated traits. This fail-to-reject situation will prevail even when the genetic effects are unequal but similar in magnitude. In genetic association studies, we may not expect equal effect sizes but we can expect them to be very close since each effect size is very small. On the other hand, if the effect sizes are very different, the $\boldsymbol{Z}$ vector will lie in some direction significantly away from the major axis $\boldsymbol{1}$ of the acceptance region. The closer it gets towards the minor axis, the greater is the chance for MANOVA statistic to fall outside the boundary and reject the null. The gray triangle in Figure 1 represents a situation where MANOVA's power to reject H_0 is higher when ρ is higher. This is the situation when only one of the two traits is associated. Furthermore, Figure 1 shows that MANOVA's loss in power will not be observed (irrespective of the strength and direction of within trait correlation) in studies where the effect sizes are reasonably large. This was observed in our simulation study with large genetic effects (simulation results not provided). It is also to be noted that if all the traits are associated but not all are correlated, MANOVA is not expected to lose power (refer section 3.4).

2.5 An alternative test: A Unified Score-based Association Test (USAT)

Our proposed test is motivated by the geometrical findings in section 2.4. As mentioned earlier, SSU test statistic does not explicitly incorporate within trait correlation and hence its acceptance region is not much affected when we increase the degree of dependency among the traits. On the other hand, MANOVA suffers from lack of power when the correlation is high and the genetic effect sizes are similar in magnitude and in same direction as the correlation. One, of course, does not know the true size and direction of the genetic effects and hence one would not know which association test to use. In such a scenario, one can see the clear advantage of combining MANOVA and SSU. We decided to choose the weight

optimally from the data. We call our test unified score-based association test (USAT). The USAT test statistic is not exactly the best weighted combination of MANOVA and SSU. It is the minimum of the p-values of the different weighted combinations. Lee et al. [2012] proposed a similar test statistic based on minimum p-value in the context of rare variants in sequencing association studies.

Let T_M be the MANOVA test statistic based on Wilk's lambda. From Bartlett's approximation, $T_M \overset{a}{\sim} \chi^2_K$. On the other hand, the SSU test statistic, denoted as T_S, has an approximate $a\chi^2_d + b$ distribution, where the parameters a and b and the degrees of freedom d are estimated from the data using equation (3). Consider the weighted statistic $T_\omega = \omega T_M + (1 - \omega)T_S$, where $\omega \in [0, 1]$ is the weight. Both MANOVA and SSU are special cases of the class of statistics T_ω. Under the null hypothesis of no association, for a given ω, T_ω is approximately a linear combination of chi-squared distributions. For a given ω, the p-value p_ω of the test statistic T_ω can be calculated using Liu et al. [2009] algorithm for chi-square approximation of non-negative quadratic forms. It is worth noting that the calculation of p_ω does not require independence assumption of the two test statistics (refer Appendix S3).

Apriori the optimal weight ω is not known. We propose our optimal unified test USAT as

$$T_{USAT} = \min_{0 \leq \omega \leq 1} p_\omega$$

For practical purposes, a grid of 11 ω values were considered: $\{\omega_1 = 0, \omega_2 = 0.1, \ldots, \omega_{10} = 0.9, \omega_{11} = 1\}$. A finer grid of more ω values did not change the USAT power curve much.

To find the p-value of our USAT test statistic, we need the null distribution of USAT. One option is to calculate the empirical p-value by considering several permuted datasets or by generating several datasets under the null (as done for Figure 4). Finding empirical p-values is computationally intensive and is not suitable when USAT is applied on a GWAS scale with large number of traits. We propose an approximate p-value calculation using a one-dimensional numerical integration. Observe that the p-value of statistic T_{USAT} is

$$p_{USAT} = P(T_{USAT} \leq t_{USAT}) = 1 - P(T_{USAT} \geq t_{USAT})$$
$$= 1 - P(T_{\omega_1} < q_{\min}(\omega_1), \ldots, T_{\omega_{11}} < q_{\min}(\omega_{11})) \approx 1 - \int F_{T_S}(\delta_\omega(x)|x) f_{T_M}(x) dx$$

where t_{USAT} is the observed value of USAT test statistic for a given dataset, $q_{\min}(\omega_b)$ is the $(1 - t_{USAT})$-th percentile of the distribution of T_{ω_b} for a given $\omega = \omega_b$, $F_{T_S}(.)$ is the cdf of SSU test statistic T_S, $\delta_\omega(x) = \min_{\omega \in \{\omega_1, \ldots, \omega_{11}\}} \frac{q_{\min}(\omega) - \omega x}{1 - \omega}$ and $f_{T_M}(.)$ is the pdf of MANOVA test statistic T_M. Mathematical details are provided in Appendix S3.

3 Results

We compared the performances of different methods mentioned in sections 2.1, 2.2, 2.3. We investigated their type I errors and powers by simulating data on unrelated subjects under a

variety of trait models. In Simulation 1 (section 3.1), we considered $K = 2$ correlated traits with genetic effects in different directions and correlation ρ varying between −1 and 1. For Simulation 2 (section 3.2), we considered $K = 5, 10, 20$ traits with genetic effects in the same direction as the positive correlation. CS correlation structure was considered. As part of Simulation 2, we also compared the performance of USAT against MANOVA and SSU. In Simulation 3 (section 3.3), we used data from Simulation 2 and investigated the type I error of USAT using the p-value approximation method described in section 2.5. In Simulation 4 (section 3.4), we used the same set-up as Simulation 2 to investigate the behaviors of existing methods under correlation structures other than CS.

For our simulation studies, we first simulated X taking values 0, 1, 2 with probabilities $(1 − f)^2, 2f(1 − f), f^2$ respectively. $f = 0.2$ was the m.a.f. of the the single SNP. The two alleles at the SNP were sampled independently to ensure Hardy-Weinberg equilibrium. Conditional on X, we simulated Y for a fixed K using the simulation model $Y = \beta_0 \mathbf{1} + \beta X + \varepsilon$, where the vectors Y, $\mathbf{1}$, β, ε are K-dimensional. We took $\beta_0 = 1$ and simulated ε from $N_K(\mathbf{0}, \sigma^2 R(\rho))$, where $R(\rho)$ is a CS correlation matrix. The genetic effect size for a given associated trait was chosen such that $h\%$ $(0 < h < 1)$ of the total phenotypic variance was explained by the SNP. For a given trait, the choice of h and the minor allele frequency of the SNP determine the effect size for the SNP. One may refer Section 2.4 of Basu et al. [2013] to find details on how the SNP effect sizes were found. The specific choices of h, β, σ^2 and ρ for each simulation are given in their respective sections (3.1 and 3.2). Before applying any method on the simulated datasets, we centered both Y and X for each dataset. We are interested in testing $H_0 : \beta = \mathbf{0}$. All the association tests except MultiPhen were coded by us in R 3.0.1 [R Development Core Team, 2014]. For MultiPhen, we used 'Joint Model' output (p-value) from the R package MultiPhen 2.0.0.

3.1 Simulation 1: $K = 2$ traits

We first studied the performances of different association tests by considering only 2 correlated traits so that the genetic effects and the pairwise correlation can have different directions. We considered genetic effects β such that 0.2% of the total variance of an associated trait was explained by the SNP. The total variance of an associated trait was taken to be 10. This ensured that the variance due to SNP was 0.02 while the residual variance was $\sigma^2 = 9.98$. For an unassociated trait, the variance explained by SNP was 0 and hence its residual variance was same as the total variance. We considered 3 possible levels of association: no trait was associated ($\beta_1 = 0 = \beta_2$), only the first trait was associated ($\beta_1 = 0.25, \beta_2 = 0$) and both the traits were associated ($\beta_1 = 0.25 = \beta_2$). We also considered genetic effects in opposite directions ($\beta_1 = 0.25, \beta_2 = −0.25$).

First, type I error comparison was done for the 6 existing methods. Estimation of type I error generally requires a large number of replicates (N) and a large sample size (n) in order to generate a reliable estimate at a stringent level of significance α. The type I error is estimated as the proportion of null datasets in which the p-value α. We considered significance levels of $\alpha = 0.01$ and 0.05. At $\alpha = 0.01$, one expects to find one 'non-null' dataset out of every 100 null datasets. Hence, for this purpose, we simulated $N = 10, 000$ null datasets with $n = 4, 000$ independent individuals. Table 1 shows the type I errors (along

with $100(1 - \alpha)$% confidence intervals of the estimates) for each of the methods for 4 values of ρ: −0.8, −0.2, 0.2, 0.8. For high magnitude of correlation ρ, notice that Fisher's method has inflated type I error while minP is conservative. Unlike minP, TATES is not conservative since it corrects for the relatedness among the traits. SSU maintains proper type I error since the distribution of the test statistic incorporates the within trait correlation structure. As expected, MANOVA and MultiPhen maintain correct type I error.

Next, we compared the powers of the methods. Compared to type I error analysis, we reduced the number of datasets to reduce computation time while ensuring reasonably good estimates of the power for different methods. $N = 500$ datasets with $n = 4,000$ unrelated individuals were simulated for different levels of association. 8 different values of correlation ρ were used: −0.8, −0.6, −0.4, −0.2, 0.2, 0.4, 0.6, 0.8. Since the methods do not have comparable type I errors (as seen in Table 1), we plotted empirical power curves for comparison. The empirical power at 5% error level was calculated in the following way. For each of Fisher's method, MANOVA and SSU, the 95-th quantile of the empirical distribution of the test statistic was determined based on the $N = 500$ test statistics obtained from N null datasets. Empirical power for these methods was calculated as the proportion of test statistics that exceeded the 95-th quantile. For each of minP and TATES, the 5-th quantile of the empirical distribution of the test statistic was determined using the $N = 500$ test statistics under null. Empirical power was, then, calculated as the proportion of test statistics that could not exceed the 5-th quantile. The empirical power of MultiPhen was determined using p-values in a way similar to empirical power calculation of minP and TATES. From Figure 2, we observe that, irrespective of the value of ρ, the tests that do not consider within trait correlation have increase in power with increase in the number of associated traits. They seem to have similar performance when both traits are associated. On the other hand, MANOVA and MultiPhen have similar performance and are usually the most powerful approaches for detecting association. But, both experience power loss when $\rho > 0.5$ and both traits have same direction of association. For traits with genetic effects in opposite directions, similar behavior of MANOVA was observed: MANOVA's power drops when $\rho < -0.5$ (refer Figure S8 in Appendix S10). These empirical observations on MANOVA are consistent with Corollary 1 of Theorem 2. No such power loss is observed for multiple univariate model based approaches. In particular, SSU maintains correct type I error and does not experience power loss like MANOVA. This observation on SSU is consistent with our geometrical insight from Figure 1.

3.2 Simulation 2: K = 5, 10, 20 traits

To further study the performance of different tests with increase in the number of correlated traits, we simulated three sets of data where the first set had $K = 5$, second had $K = 10$ and the third had $K = 20$ correlated traits. Since this experiment involves a large number of comparisons for varying K and ρ, we reduced the sample size further to save computation time. We considered $N = 500$ simulated datasets for each scenario with $n = 400$ unrelated individuals. When we plotted estimated power as a function of N for five different simulation runs (results not provided), no appreciable between-simulation variability was observed at $N = 500$. This ensures that reasonably good estimates of power can be obtained for our choice of N and n. For this simulation study, we considered only non-negative

genetic effects, and positive correlation $\rho = 0.2, 0.4, 0.6$ between each pair of traits. The total variance of a trait was fixed at 10. Following Ferreira and Purcell [2009], $\boldsymbol{\beta}$ was chosen such that 0.5% of the total variance of an associated trait was explained by the single SNP. We considered 6 possible levels of association: 0%, 20%, 40%, 60%, 80% or 100% of the traits were associated with the SNP. Empirical power curves are presented for comparison.

From Figure 3, we again observe how MANOVA suffers from power loss at 'complete association' when the within trait correlation is high. This power loss increases with increase in total number of correlated traits. At 'complete association' (where MANOVA loses power), the power difference between MANOVA and other methods (such as SSU) increases with increase in number of correlated traits and decrease in correlation ρ. At a given 'partial association', MANOVA is seen to dominate over other methods. Here, the difference in powers of MANOVA and any other method increases with increase in number of traits as well as the correlation. MANOVA's performance in this experiment is consistent with the asymptotic result in Theorem 1.

Next we studied the performance of our approach USAT compared to MANOVA and SSU. Since all three approaches can maintain proper type I error (type I error performance of USAT is established in section 3.3), asymptotic power curves are plotted for comparison. For each method, the asymptotic power is estimated as the proportion of datasets that give p-value ≤ 0.05. For the calculation of asymptotic power of USAT, we used the p-value approximation method described in section 2.5. In Figure 4, observe that USAT has better power than MANOVA whenever it suffers from power loss due to same direction of residual correlation and equal-sized genetic effects. In such situations, SSU performs significantly better than MANOVA, and USAT follows the SSU power curve closely. In other situations where MANOVA is seen to be most powerful among existing methods, USAT tends to have power close to MANOVA. USAT maximizes power by adaptively using the data to combine the MANOVA and the SSU approach.

3.3 Simulation 3: p-value approximation for USAT

In this section, we applied our approximate p-value approach for finding USAT p-values to study its impact on type I error. We generated $N = 100,000$ independent datasets (as in section 3.2) with $n = 10,000$ unrelated individuals under H_0. Since we aim to consider stringent error levels, a large sample size as well as a large number of datasets have been chosen. The type I error was estimated by the proportion of datasets in which the asymptotic approximate p-value of USAT test statistic was $\leq 10^{-4}$, $\leq 10^{-3}$, and $\leq 10^{-2}$. Table 2 gives the estimated type I error rates (along with $100(1 - \alpha)\%$ confidence intervals of the estimates) for USAT using p-value approximation. The estimated values of type I error for different values of K and ρ were very close to the true error level α in most scenarios. In general, USAT correctly maintains type I error at moderately low error levels (Table 2) but can exhibit slight inflatedness for more stringent levels (Table S1).

3.4 Simulation 4: Other correlation structures

We first considered an independent structure. Apart from the residual correlation matrix $\boldsymbol{R}(\rho)$, the data simulation was exactly same as in Simulation 2 (section 3.2). The figures and

detailed explanations can be found in Appendix S4. When all the traits are independent (i.e., $R(\rho) = I_K$), MANOVA does not suffer from power loss at any level of association. Empirical power curves (Figure S5) showed that the performances of all the methods described in sections 2.1, 2.2, 2.3, except minP and TATES, were similar. As expected, the powers steadily increased with increase in number of associated traits. Next we considered a correlation structure where the first 80% of the traits had pairwise correlation ρ while the rest were independent. Empirical power curves (Figure S6) showed that MANOVA suffered power loss when only the correlated traits were associated. Performance of MANOVA improved when the SNP was associated with some of the uncorrelated traits. This simulation study showed us that MANOVA may not experience power loss even when all the traits are associated if some of them are uncorrelated. Appendix S5 provides a theoretical support for this observation. The third type of non-CS correlation structure that we considered was AR1(ρ) (Figure S7). MANOVA's power loss was mainly observed for small K and strong ρ. With increase in K and decrease in ρ, MANOVA did not experience power loss even at 'complete association'. The strength of AR1(ρ) correlation becomes negligible at or near 'complete association' when ρ is small and K is moderately large. For all these trait models, the power curves of multiple univariate model based approaches rose with increase in number of associated traits (irrespective of strength or direction of residual correlation). All these observations on MANOVA for various correlation structures were expected based on our geometrical insight from Figure 1 (section 2.4).

3.5 Real Data Analysis

The ARIC study is an ongoing prospective study designed to investigate the etiology and natural history of atherosclerosis and its clinical manifestations, and to measure variation in cardiovascular risk factors, medical care and disease by race, gender, place and time [The ARIC Investigators, 1989]. ARIC has collected measures on many T2D-related traits at 4 separate visits over a 9-year period. For our analysis, we focused on the Caucasian participants and the following 3 T2D related quantitative traits measured at visit 4 (1996 – 98): fasting glucose; 2-hour glucose from an oral glucose tolerance test; fasting insulin. The pairwise correlations among these 3 traits were within (0.2, 0.35). As in most studies of T2D [Dupuis et al., 2010; Manning et al., 2012; Scott et al., 2012, for example], BMI was adjusted as a covariate. Individuals with diagnosed or treated diabetes at visit 4 and individuals with missing traits were excluded, leaving 5, 816 in our analytic sample. More details on the phenotypes and the choice of covariates can be found in Appendix S6.

The ARIC cohort has been genotyped using the Affymetrix Genome-Wide SNP Array 6.0. Genotyping was completed at the Broad Institute of MIT and Harvard in three batches; the Birdseed algorithm was used for genotype calling. Imputation was performed using Mach 1.0 86 and HapMap release 21 (Build 35). SNPs with a call rate < 90%, m.a.f. < 1%, or deviation from Hardy-Weinberg equilibrium ($p < 10^{-6}$) were excluded for imputation. There was a total of 2.5 million genotyped or imputed SNPs. Apart from USAT and MANOVA, we also performed separate univariate analyses to emphasize the importance of joint analysis over univariate ones. Before implementing any of these approaches, we centered both phenotype and genotype data. SNPs with m.a.f. < 5% were excluded. All statistical models were adjusted for Age, Sex and BMI.

Figure 5 shows the manhattan plots of negative log-transformed p-values for the single trait single SNP analyses for chromosomes 1–22. The red horizontal line (at 7.3) in each plot indicates the log-transformed GWAS significance p-value 5×10^{-8}. There were 53 significant SNPs for fasting glucose, and none for the other two traits. On the other hand, there were 96 and 77 signals for MANOVA and USAT respectively that reached this stringent threshold (refer Figure 6, and Table S3 in Appendix S8). Most of these signals mapped near the genes GCKR, ABCB11, C2orf16, CCDC121, ZNF512, FAM148A, C2CD4A, which are already known to be associated with diabetes related traits [Yamauchi et al., 2010; Kraja et al., 2011, for example]. It is worth noting that these detected SNPs are in high linkage disequilibrium (LD). Among the SNPs reported in Table S3, MANOVA and USAT respectively detected 44 and 27 SNPs that none of the univariate analyses could detect. Most notable genes that the univariate analyses completely missed are GCKR (on chr 2) and FAM148A (on chr 15).

Since most of the detected SNPs in Figure 6 are in high LD, Table 3 reports only the important SNPs after removing the ones in high LD. In a group of highly correlated SNPs (i.e., SNPs with estimated absolute pairwise correlation coefficient > 0.8 with another SNP), we kept one SNP as a representative. The choice of representative SNP was based on previous reports of association. The correlation coefficients (as measures of LD) were obtained from PLINK [Purcell et al., 2007] using the command plink --file mydata --r. All of the signals in Table 3 have been previously reported. The minor allele T of rs1260326 (gene GCKR of chr 2) is known to be associated with T2D and hypertriglyceridemia. Risk allele A of rs13022873 (gene ZNF512 of chr 2) was found to be significantly associated with waist circumference (a T2D related trait highly correlated with BMI) and triglycerides [Kraja et al., 2011]. rs13431652 (gene G6PC2 of chr 2) was reported to be a potentially causative SNP linking G6PC2 to increased fasting plasma glucose levels and elevated promoter activity [Bouatia-Naji et al., 2010]. The rs1402837 T allele (gene G6PC2 of chr 2) is known to be associated with blood sugar levels (glycated hemoglobin levels). McCaffery et al. [2013] reported that SNPs in ABCB11 (like rs484066) of chr 2 are associated with weight loss and regain. Meta-analysis of several GWAS found rs17271305 (gene VPS13C of chr 15) to be associated with glucose levels 2 hours after an oral glucose challenge [Saxena et al., 2010]. The diabetogenic A allele of rs7172432 (gene VPS13C/C2CD4A/C2CD4B of chr 15) significantly impairs glucose-stimulated insulin response in non-diabetics [Grarup et al., 2011].

In Table 3, we notice one SNP (rs7172432) that USAT missed at the stringent significance level of 5×10^{-8}. One also notices that difference in the p-values of USAT and MANOVA for this SNP is negligible. If one takes a closer look at the manhattan plots of Figure 6, one will find that certain SNPs are prominently visible for USAT but not for MANOVA (even though none could reach genome wide significance). The most noticeable regions (details in Appendix S9) are in chromosomes 1, 5, 12 and 18. Regions in chr 18, for example, have only been identified by USAT, and some recent studies have reported findings on chr 18 for T2D [Morris et al., 2012; Yuan et al., 2014]. For most of these interesting findings, USAT gave more significant p-value than MANOVA by borrowing almost the entire information from SSU.

4 Discussion

In the study of a complex disease, several correlated traits are often measured as risk factors for the disease. There may be genetic variants affecting several of these traits. Analyzing multiple disease-related traits could potentially increase power to detect association of genetic variants with such a disease. The elucidation of genetic risk factors of such diseases will help us in better understanding and developing therapeutics against them. In this paper, we have studied some of the common univariate and multivariate approaches for analyzing association between multiple phenotypes and a genetic variant. Our simulation results showed that no single method perform uniformly better than the others under the simulation scenarios we considered. Multivariate methods like MANOVA and MultiPhen usually had higher power than the univariate tests only in situations where a few of the correlated traits were associated. Univariate model based methods in sections 2.1 & 2.2 outperformed multivariate methods when all the correlated traits were associated and the genetic effects as well as the residual correlations were in the same direction. Under the assumption of a CS residual correlation structure, we not only established theoretical conditions for MANOVA's loss of power but also provided geometrical explanation for the observed dominance of univariate methods over MANOVA (refer section 2.4). Although we have not established similar theoretical conditions for other correlation structures, we have seen similar behavior of MANOVA in our simulation studies.

We also proposed a novel weighted approach USAT, which maximizes power by adaptively using the data to optimally combine MANOVA and the SSU test. Approximate USAT p-values can be computed using a very fast one-dimensional numerical integration, which makes implementation on GWAS data easy. It also allows for adjustment of other covariates (details in Appendix S7). As shown by our simulation studies, USAT maintains correct type I error in most scenarios (refer Table 2). It may be noted that the approximation involved in the p-value calculation can be slightly inflated for stringent error levels (refer Appendix S3). The magnitude of such inflatedness at stringent error levels not only depends on the number of traits and the correlation parameter but also on the correlation structure. To give an idea about the effect of approximation in real GWAS for stringent thresholds, the empirical USAT p-values along with approximate USAT p-values are provided for a few randomly chosen SNPs in the ARIC Study (refer Table S2). Unlike MANOVA, USAT is powerful in detecting pleiotropy under the simulation models we considered. For a large number of traits with weak correlations or for a few traits with strong correlations, USAT has clear advantage over MANOVA (refer Figure 4). For traits with AR1 covariance structure, USAT is more powerful in most situations (refer Appendix S11). Although the reported simulation studies considered multivariate normal traits, USAT performed well for multivariate t distributed traits (non-normal continuous traits) and common genetic variants. The conclusions from the ARIC data analysis are twofold. First, it emphasizes the importance of joint analysis of correlated phenotypes over multiple univariate analyses. Secondly, USAT can give significant findings by giving more weight to MANOVA for certain SNPs while giving more weight to SSU for the rest. For any real data, the underlying association scenario (which may vary from one SNP to another) is not known apriori and a specific

approach of association test (univariate vs multivariate modeling) may not be powerful enough. USAT overcomes this challenge and has proven to be a sensible choice.

Finally, the simulation scenarios we considered are not exhaustive. Under the scenarios we considered, the relative behavior of these two tests did not vary much with change in m.a.f. (refer Appendix S12), or with increase in the number of correlated traits. Our simulation studies also assumed no missing data and no trait outliers. USAT requires complete phenotype data. In presence of missing traits, one may consider imputation before performing association analysis. van der Sluis et al. [2013] showed that 10% missing-completely-at-random data caused quite a drop in power for MANOVA when only 1 trait was associated. O'Reilly et al. [2012] showed that in the presence of outliers in the phenotype distribution, MANOVA and the standard univariate approach were substantially inflated for low m.a.f. We simulated data for an additive model only and did not consider any non-additive genetic model and/or interactions. In future, we intend to study how power of our USAT test would be affected in such situations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Anderson, T. An introduction to multivariate statistical analysis. New York: John Wiley & Son; 1984.

Aschard H, Vilhjálmsson B, Greliche N, Morange PE, Trégouët DA, Kraft P. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. Am J Hum Genet. 2014; 94:662–676. [PubMed: 24746957]

Basu S, Zhang Y, Ray D, Miller M, Iacono W, McGue M. A rapid gene-based genome-wide association test with multivariate traits. Hum Hered. 2013; 76(2):53–63. [PubMed: 24247328]

Blair D, Lyttle C, Mortensen J, et al. A nondegenerate code of deleterious variants in mendelian loci contributes to complex disease risk. Cell. 2013; 155:70–80. [PubMed: 24074861]

Bouatia-Naji N, Bonnefond A, Baerenwald DA, et al. Genetic and functional assessment of the role of the rs13431652-A and rs573225-A alleles in the G6PC2 promoter that are strongly associated with elevated fasting glucose levels. Diabetes. 2010; 59(10):2662–71. [PubMed: 20622168]

Dupuis J, Langenberg C, Prokopenko I, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nat Genet. 2010; 42(2):105–116. [PubMed: 20081858]

Ferreira M, Purcell S. A multivariate test of association. Bioinformatics. 2009; 25:132–133. [PubMed: 19019849]

Fisher, R. Statistical Methods for Research Workers. Oliver and Boyd; Edinburgh: 1925.

Galesloot T, van Steen K, Kiemeney L, Janss L, Vermeulen S. A comparison of multi-variate genome-wide association methods. PLoS One. 2014; 9(4):e95923. [PubMed: 24763738]

Germain M, Saut N, Greliche N, Dina C, et al. Genetics of venous thrombosis: Insights from a new genome wide association study. PLoS One. 2011; 6:9.

Grarup N, Overvad M, Sparsø T, Witte D, et al. The diabetogenic VPS13C/C2CD4A/C2CD4B rs7172432 variant impairs glucose-stimulated insulin response in 5,722 non-diabetic Danish individuals. Diabetologia. 2011; 54(4):789–94. [PubMed: 21249489]

Korte A, Vilhjálmsson B, Segura V, Platt A, Long Q, Nordborg M. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet. 2012; 44:1066–1071. [PubMed: 22902788]

Kraja A, Vaidya D, Pankow J, Goodarzi M, et al. A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium. Diabetes. 2011; 60(4):1329–39. [PubMed: 21386085]

Lee S, Wu M, Lin X. Optimal tests for rare variant effects in sequencing association studies. Biostatistics. 2012; 13:762–775. [PubMed: 22699862]

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.

Maity A, Sullivan P, Tzeng J. Multivariate phenotype association analysis by marker-set kernel machine regression. Genet Epidemiol. 2012; 36(7):686–695. [PubMed: 22899176]

Manning A, Hivert MF, Scott R, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet. 2012; 44(6):659–669. [PubMed: 22581228]

McCaffery J, Papandonatos G, Huggins G, et al. Human cardiovascular disease IBC chip-wide association with weight loss and weight regain in the look AHEAD trial. Hum Hered. 2013; 75(2–4):160–74. [PubMed: 24081232]

Morris A, Voight B, Teslovich T, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet. 2012; 44:981–990. [PubMed: 22885922]

Muller K, Peterson B. Practical methods for computing power in testing the multivariate general linear hypothesis. Comput Stat Data Anal. 1984; 2(2):143–158.

O'Reilly P, Hoggart C, Pomyen Y, Calboli C, et al. MultiPhen: Joint model of multiple phenotypes can increase discovery in gwas. PLoS One. 2012; 7(5):e34861. [PubMed: 22567092]

Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. Genet Epidemiol. 2009; 33:497–507. [PubMed: 19170135]

Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. Sep; 2007 81(3):559–575.10.1086/519795 [PubMed: 17701901]

R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2014. http://www.R-project.org/

Ried J, Shin S-Y, Krumsiek J, et al. Novel genetic associations with serum level metabolites identified by phenotype set enrichment analyses. Hum Mol Genet. 2014

Saxena R, Hivert M, Langenberg C, et al. Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat Genet. 2010; 42(2):142–8. [PubMed: 20081857]

Scott R, Lagou V, Welch R, et al. Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat Genet. 2012; 44(9):991–1005. [PubMed: 22885924]

Souto J, Almasy L, Borrell M, Blanco-Vaca F, Mateo J. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: The GAIT study. genetic analysis of idiopathic thrombophilia. Am J Hum Genet. 2000; 67(6):1452–1459. [PubMed: 11038326]

Stephens M. A unified framework for association analysis with multiple related phenotypes. PLoS One. 2013; 8(7):e65245.10.1371/journal.pone.0065245 [PubMed: 23861737]

Tang C, Ferreira M. A gene-based test of association using canonical correlation analysis. Bioinformatics. 2012; 28(6):845–850. [PubMed: 22296789]

The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol. 1989; 129(4):687–702. [PubMed: 2646917]

van der Sluis S, Posthuma D, Dolan C. TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. PLoS Genet. 2013; 9(1):e1003235. [PubMed: 23359524]

Yamauchi T, Hara K, Maeda S, Yasuda K, et al. A genome-wide association study in the japanese population identifies susceptibility loci for type 2 diabetes at UBE2E2 and C2CD4A-C2CD4B. Nat Genet. 2010; 42(10):864–868. [PubMed: 20818381]

Yang J, Lee H, Goddard M, Visscher P. GCTA: A tool for Genome-wide Complex Trait Analysis. Am J Hum Genet. 2011; 88(1):76–82. [PubMed: 21167468]

Yang Q, Wang Y. Review article: Methods for analyzing multivariate phenotypes in genetic association studies. J Probab Stat. 2012; 2012:13.10.1155/2012/652569

Yuan W, Xia Y, Bell C, et al. An integrated epigenomic analysis for type 2 diabetes susceptibility loci in monozygotic twins. Nat Commun. 2014; 5:5719. [PubMed: 25502755]

Zhang JT. Approximate and asymptotic distributions of chi-squared-type mixtures with applications. J Am Stat Assoc. 2005; 100:273285.

Zhang L, Pei Y, Li J, Papasian C, Deng H. Univariate/multivariate genome-wide association scans using data from families and unrelated samples. PLoS One. 2009; 4:e6502. [PubMed: 19652719]

Zhou X, Stephens M. Genome-wide efficient mixed model analysis for association studies. Nat Genet. 2012; 44(7):821–824. [PubMed: 22706312]

Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods. 2014; 11:407–409. [PubMed: 24531419]
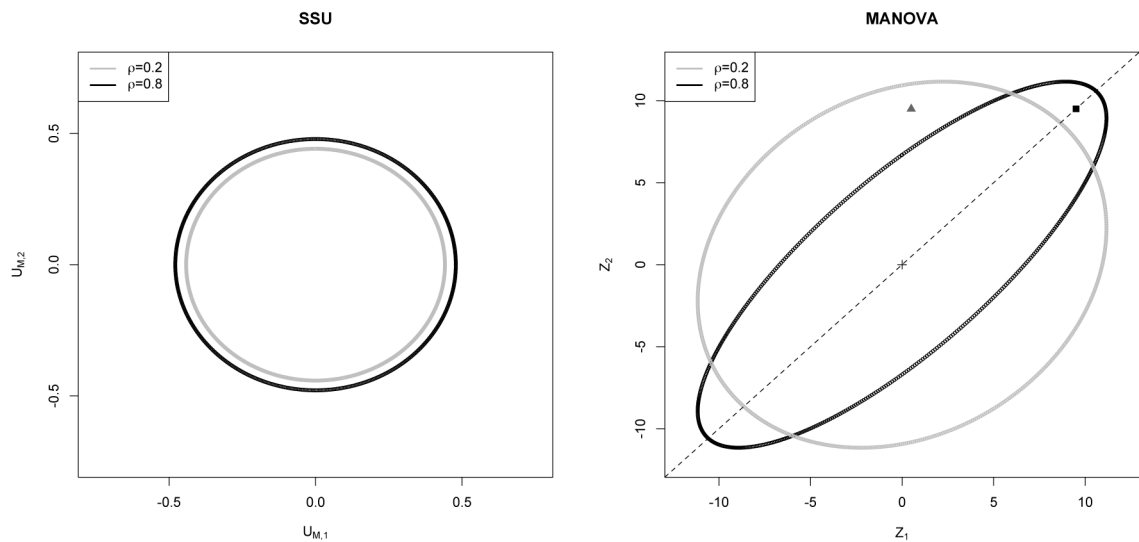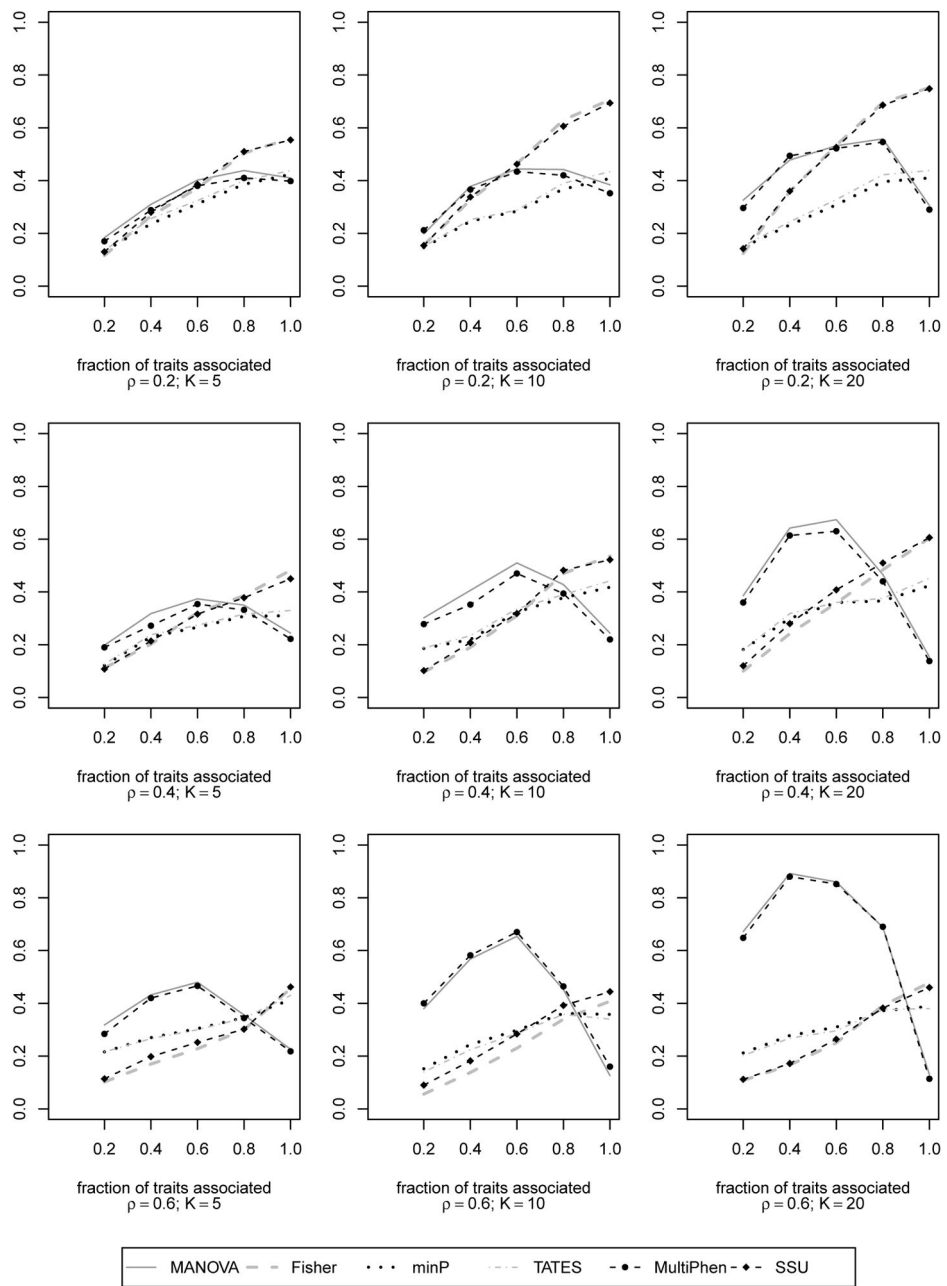
Figure 1.
Comparison of theoretical 95% acceptance regions of SSU test and of MANOVA for $K = 2$ traits and $\rho = 0.2, 0.8$ (Compound Symmetry correlation). The area within the gray (black) ellipse is the acceptance region when $\rho = 0.2$ (0.8). Details of these plots are provided in Appendix S2. SSU's acceptance region is drawn using the marginal scores $U_{M,1}$ and $U_{M,2}$. MANOVA's acceptance region is drawn based on the test $\mathbf{Z}'\mathbf{I(0)Z}$, where $\mathbf{I(0)}$ is Fisher Information matrix under $H_0 : \boldsymbol{\beta} = \mathbf{0}$. The black dotted line is the $\mathbf{1}$ vector and coincides with the major axes of the ellipses. When $\beta_1 = \beta_2$, we expect $Z_1 = Z_2$. The solid black square represents a situation where $\beta_1 = \beta_2 \gg 0$. When β_1 & β_2 are significantly apart, we expect the same for Z_1 & Z_2, and such a situation in represented by the gray triangle.
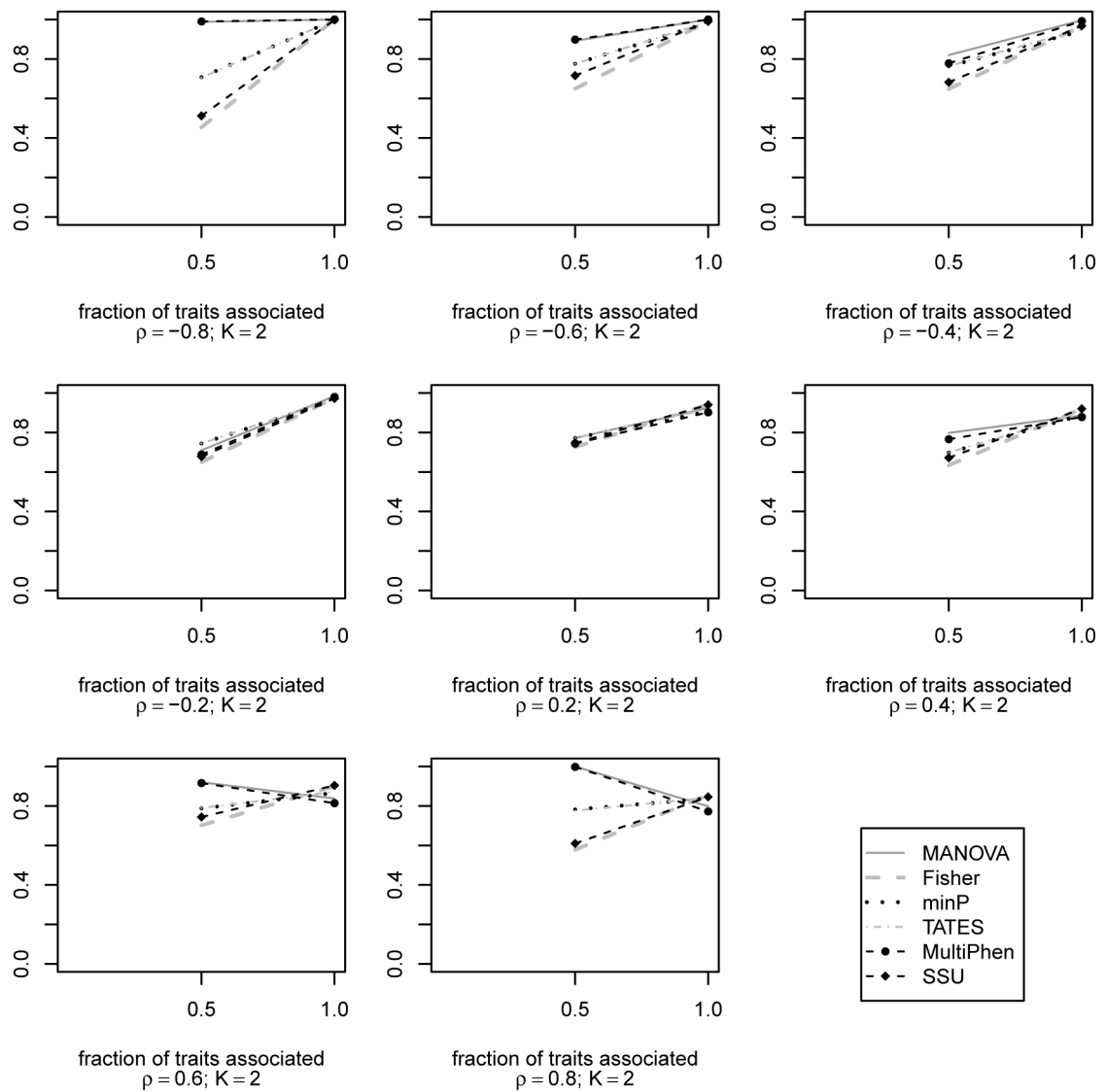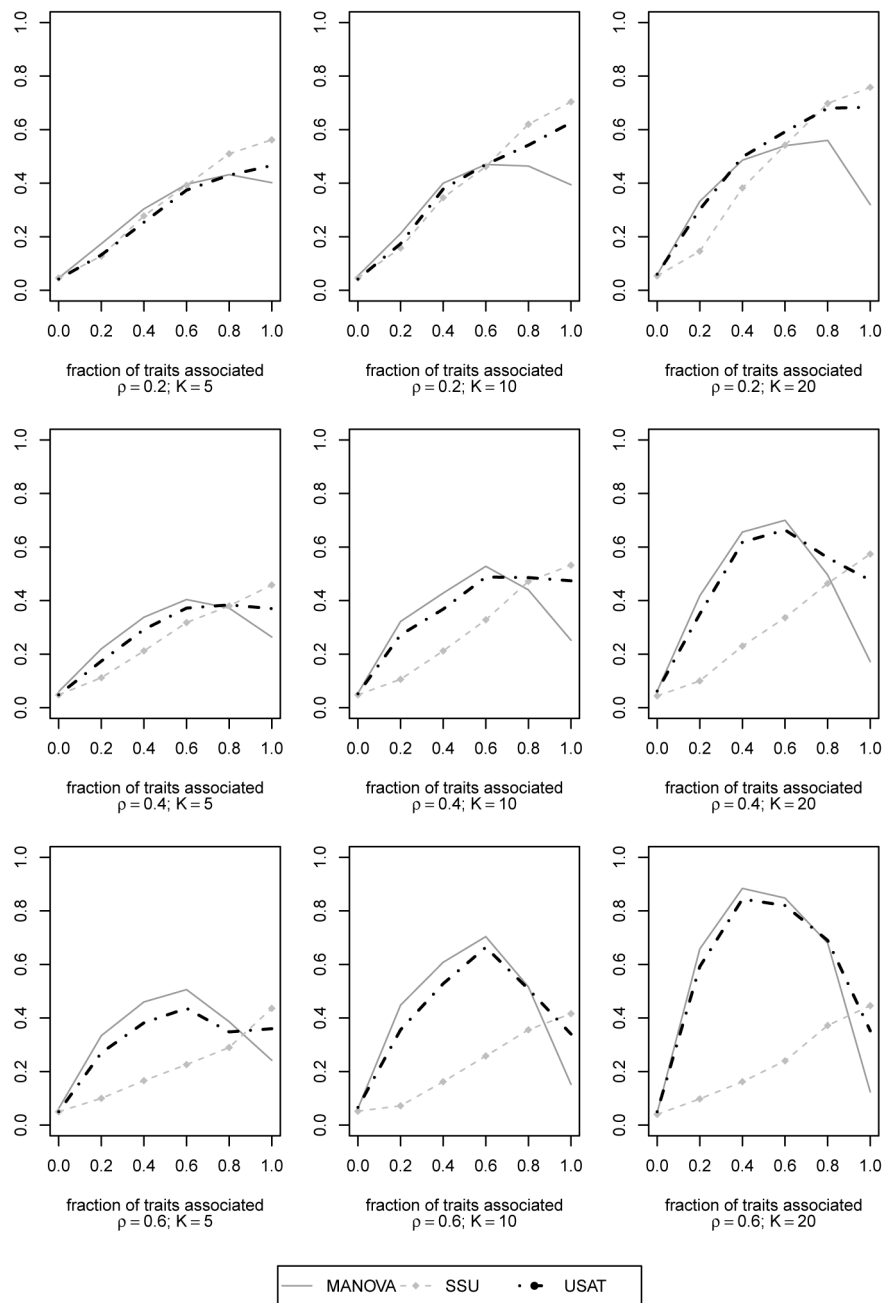
Figure 2.
Empirical power curves of the different existing association tests for $K = 2$ traits and different within trait correlation values $\rho = -0.8, -0.6, -0.4, -0.2, 0.2, \ldots, 0.8$ based on $N = 500$ datasets with $n = 4,000$ unrelated subjects. Same direction and same size genetic effect used when both traits are associated (i.e., datasets are generated from an alternative model $H_{a2,1} : \beta_1 = \beta_2 > 0$). Effect size of 0.25 (proportion of variance explained is 0.2%) is used for the associated traits. The power is plotted along y-axis while the fraction of traits associated with the genetic variant is plotted along x-axis.

Figure 3.
Empirical power curves of the different existing association tests for $K = 5, 10, 20$ traits and different within trait correlation values $\rho = 0.2, \ldots, 0.6$ based on $N = 500$ datasets with $n = 400$ unrelated subjects. Same effect size of 0.395 (proportion of variance explained is 0.5%) is used for all the associated traits. The power is plotted along y-axis while the fraction of traits associated with the genetic variant is plotted along x-axis.

Figure 4.
Asymptotic power curves of the SSU and MANOVA tests along with our novel approach USAT for $CS(\rho)$ within-trait correlation structure. $K = 5, 10, 20$ traits have been simulated at different within trait correlation values $\rho = 0.2, 0.4, 0.6$. For each value of K and ρ, there were $N = 500$ datasets of $n = 400$ unrelated individuals. Same effect size of 0.395 (proportion of variance explained is 0.5%) was used for the traits that are associated. The power is plotted along y-axis while the fraction of traits associated with the genetic variant is plotted along x-axis.

Fasting Glucose
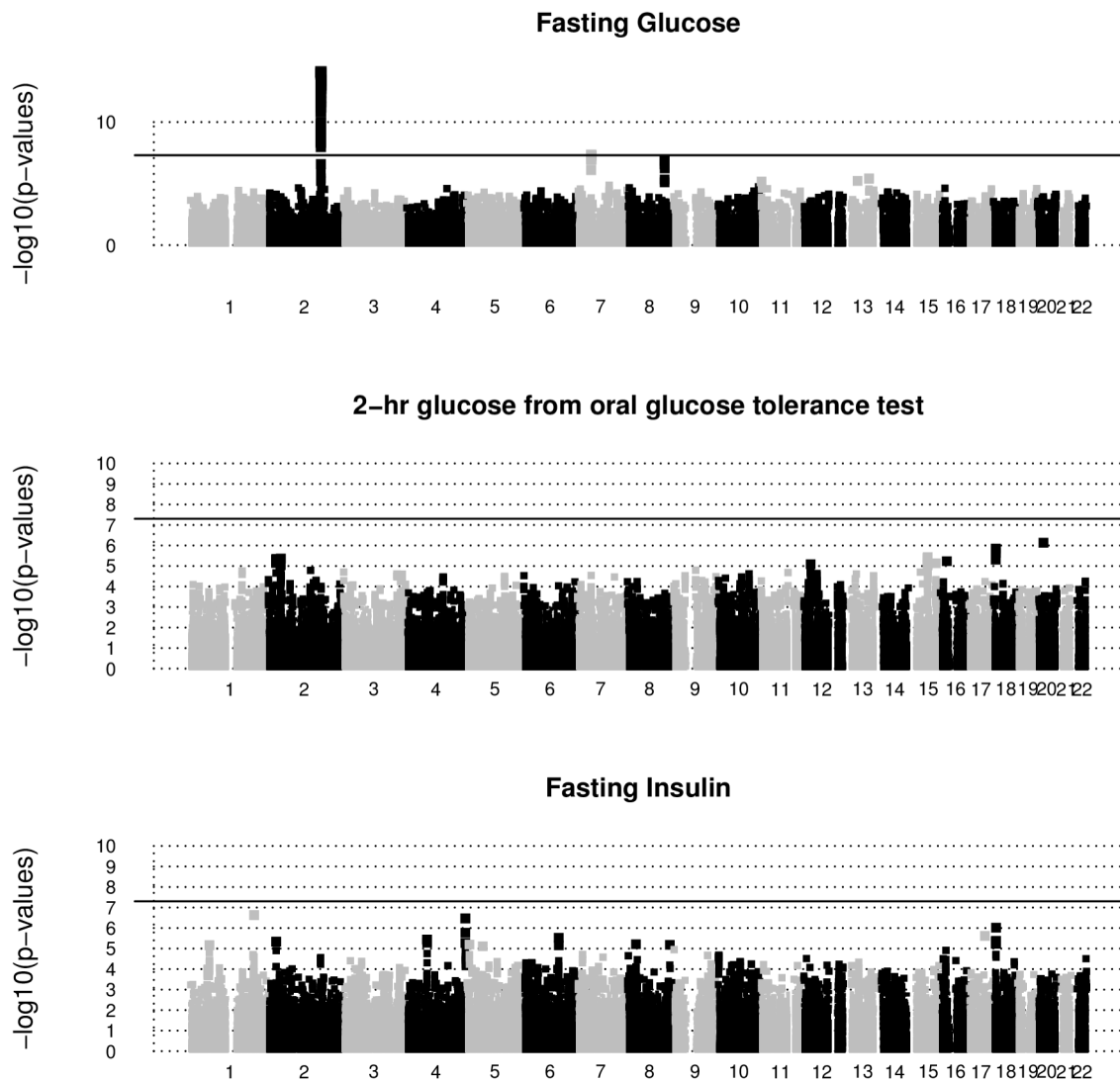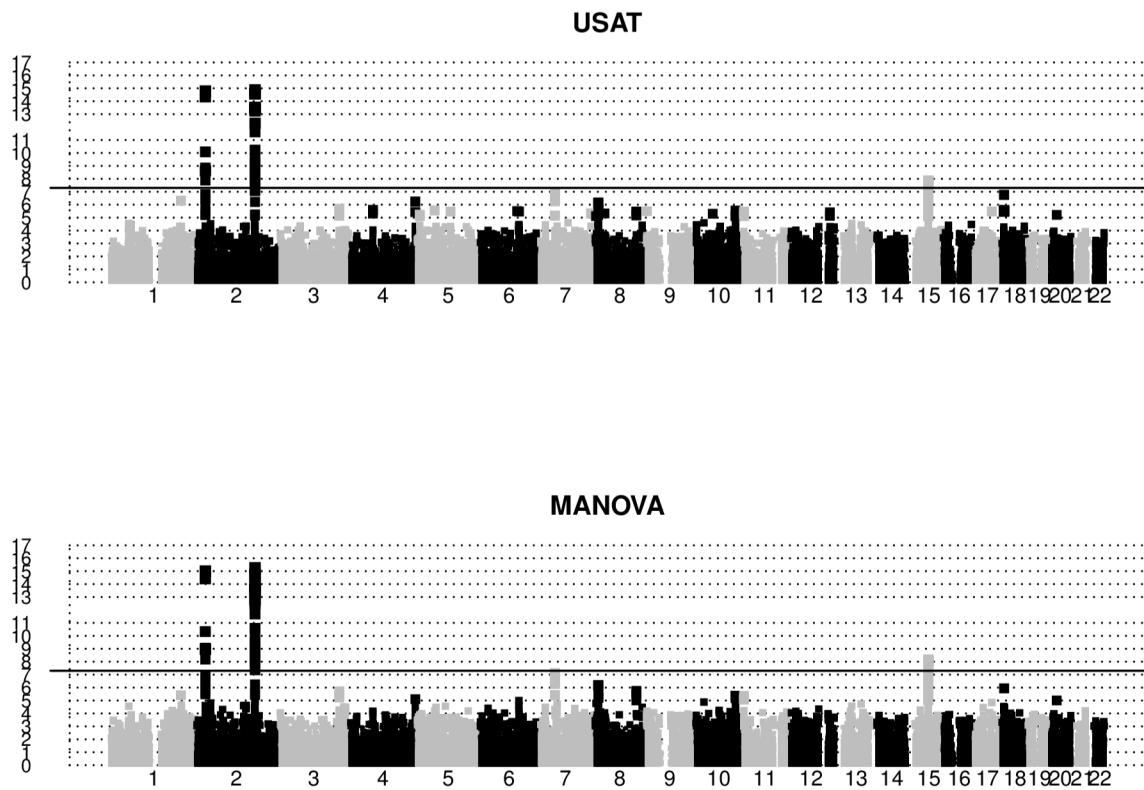

2–hr glucose from oral glucose tolerance test

Fasting Insulin

Figure 5.
ARIC Study: Manhattan plots of negative log-transformed p-values of the univariate analyses are plotted against base pair positions for chromosomes 1–22. Only SNPs with m.a.f. > 5% are reported. Age, Sex and BMI were adjusted in the statistical models. The solid black horizontal line in each plot corresponds to the genome-wide significance level 5×10^{-8}. 53 SNPs (all from chr 2) detected as significant for fasting glucose; and none for the other traits. Note that many of these significant SNPs are in high linkage disequilibrium.

USAT

MANOVA

Figure 6.
ARIC Study: Manhattan plots of negative log-transformed p-values of multivariate analyses (USAT and MANOVA) are plotted against base pair positions for chromosomes 1–22. Only SNPs with m.a.f. > 5% are reported. Age, Sex and BMI were adjusted in the statistical models. The solid black horizontal line in each plot corresponds to genome-wide significance level 5×10^{-8}. Note that many of these SNPs are in high linkage disequilibrium.

Author Manuscript   Author Manuscript   Author Manuscript   Author Manuscript

Table 1

Estimated type I errors of the afore mentioned existing association tests for $K = 2$ correlated traits. 4 values of pairwise correlation ρ were considered. The p-values were calculated for 10, 000 null datasets with 4, 000 unrelated individuals. Type I error rate was calculated as the proportion of null datasets with p-value α. The $100(1 − \alpha)\%$ confidence intervals for the estimates are provided in square braces.

α	ρ	Fisher	minP	TATES	SSU	MANOVA	MultiPhen
0.01	−0.8	0.029 [0.024, 0.033]	0.009 [0.006, 0.011]	0.012 [0.009, 0.014]	0.010 [0.008, 0.013]	0.009 [0.007, 0.012]	0.010 [0.008, 0.013]
	−0.2	0.012 [0.009, 0.015]	0.011 [0.008, 0.014]	0.011 [0.009, 0.014]	0.011 [0.008, 0.013]	0.011 [0.008, 0.014]	0.011 [0.008, 0.013]
	0.2	0.011 [0.008, 0.014]	0.010 [0.008, 0.013]	0.010 [0.008, 0.013]	0.010 [0.007, 0.012]	0.011 [0.008, 0.013]	0.010 [0.008, 0.013]
	0.8	0.026 [0.022, 0.030]	0.008 [0.006, 0.011]	0.012 [0.009, 0.015]	0.009 [0.007, 0.012]	0.009 [0.007, 0.012]	0.009 [0.007, 0.012]
0.05	−0.8	0.079 [0.074, 0.084]	0.039 [0.036, 0.044]	0.053 [0.049, 0.057]	0.049 [0.045, 0.054]	0.049 [0.045, 0.054]	0.048 [0.044, 0.053]
	−0.2	0.053 [0.049, 0.057]	0.051 [0.046, 0.055]	0.052 [0.048, 0.056]	0.049 [0.045, 0.054]	0.051 [0.046, 0.055]	0.051 [0.047, 0.055]
	0.2	0.051 [0.047, 0.056]	0.049 [0.045, 0.054]	0.051 [0.047, 0.056]	0.05 [0.046, 0.054]	0.052 [0.048, 0.056]	0.05 [0.046, 0.054]
	0.8	0.079 [0.074, 0.084]	0.039 [0.035, 0.043]	0.055 [0.05, 0.059]	0.047 [0.043, 0.051]	0.052 [0.047, 0.056]	0.051 [0.047, 0.055]

Table 2

Estimated type I errors of the approximate p-value calculation approach for our USAT test. The p-values were calculated for 100, 000 null datasets with 10, 000 unrelated individuals. Type I error rate was calculated as the proportion of datasets that had approximate p-value ≤ α. The $100(1 - α)$% confidence intervals for the estimates are provided in square braces.

K	**5**			**10**			**20**		
ρ	**0.2**	**0.4**	**0.6**	**0.2**	**0.4**	**0.6**	**0.2**	**0.4**	**0.6**
$α = 10^{-4}$	0.00008 [-0.00003, 0.00019]	10^{-4} [-0.00002, 0.00022]	0.00008 [-0.00003, 0.00019]	0.00015 [-0.00000, 0.00030]	10^{-4} [-0.00002, 0.00022]	0.00008 [-0.00003, 0.00019]	0.00015 [-0.00000, 0.00030]	0.00011 [-0.00002, 0.00024]	10^{-4} [-0.00002, 0.00022]
$α = 10^{-3}$	0.00081 [0.00051, 0.00111]	0.00092 [0.00060, 0.00123]	0.00086 [0.00055, 0.00116]	0.00108 [0.00074, 0.00142]	0.00099 [0.00066, 0.00132]	0.00089 [0.00058, 0.00120]	0.00149 [0.00109, 0.00189]	0.00117 [0.00081, 0.00153]	0.00109 [0.00075, 0.00143]
$α = 10^{-2}$	0.0082 [0.0075, 0.0089]	0.0091 [0.0083, 0.0099]	0.0093 [0.0085, 0.0101]	0.0094 [0.0086, 0.0102]	0.0098 [0.0089, 0.0106]	0.0097 [0.0089, 0.0105]	0.0103 [0.0095, 0.0111]	0.0104 [0.0096, 0.0112]	0.0103 [0.0095, 0.0112]

Table 3

List of SNPs that exceed the genome-wide significance level 5×10^{-8} for USAT and MANOVA. SNPs with m.a.f. $< 5\%$ have been screened out. The SNPs listed here are the ones left after LD screening. In a group of highly correlated SNPs (i.e., SNPs with estimated absolute pairwise correlation coefficient > 0.8 with another SNP), one SNP was kept as a representative. p values for the univariate analysis of the individual traits are also provided for these significant SNPs. SNP rs ID in bold is the one detected solely by MANOVA but not by USAT at this stringent genome-wide significance level. ω denotes the weight assigned to MANOVA statistic in the USAT approach. The abbreviations used are FG (Fasting Glucose), 2-hr GL (2-hour glucose from an oral glucose tolerance test), FI (Fasting Insulin).

			MANOVA	USAT		Univariate Analysis p			
chr	SNP	m.a.f.	p	p	ω	FG	2-hr GL	FI
2	*rs1260326*	0.41	3.77×10^{-15}	4.44×10^{-15}	1.00	1.24×10^{-4}	6.26×10^{-6}	1.24×10^{-5}
2	*rs13022873*	0.27	9.94×10^{-10}	2.34×10^{-9}	1.00	1.49×10^{-2}	6.01×10^{-6}	1.07×10^{-2}
2	*rs13431652*	0.29	1.85×10^{-13}	5.48×10^{-13}	1.00	2.24×10^{-12}	9.57×10^{-1}	2.85×10^{-1}
2	*rs1402837*	0.23	4.91×10^{-9}	1.15×10^{-8}	1.00	2.78×10^{-10}	5.18×10^{-2}	8.07×10^{-1}
2	*rs484066*	0.36	2.01×10^{-12}	2.32×10^{-12}	1.00	6.10×10^{-12}	8.54×10^{-1}	4.87×10^{-1}
15	*rs17271305*	0.39	6.87×10^{-9}	1.35×10^{-8}	1.00	6.29×10^{-3}	1.17×10^{-5}	2.86×10^{-1}
15	***rs7172432***	0.41	1.68×10^{-8}	5.98×10^{-8}	1.00	7.98×10^{-4}	3.52×10^{-4}	4.88×10^{-2}