Published in final edited form as:

Biometrics. 2019 December ; 75(4): 1076–1085. doi:10.1111/biom.13076.

Pleiotropy informed adaptive association test of multiple traits using GWAS summary data

Maria Masotti, Bin Guo, Baolin Wu*

Division of Biostatistics, School of Public Health, University of Minnesota

Summary:

Genetic variants associated with disease outcomes can be used to develop personalized treatment. To reach this *precision medicine* goal, hundreds of large-scale genome-wide association studies (GWAS) have been conducted in the past decade to search for promising genetic variants associated with various traits. They have successfully identified tens of thousands of diseases related variants. However in total these identified variants explain only part of the variation for most complex traits. There remain many genetic variants with small effect sizes to be discovered, which calls for the development of (1) GWAS with more samples and more comprehensively genotyped variants, e.g., the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program is planning to conduct whole genome sequencing (WGS) on over 100,000 individuals; and (2) novel and more powerful statistical analysis methods. The current dominating GWAS analysis approach is the “single trait” association test, despite the fact that many GWAS are conducted in *deeply-phenotyped cohorts* including many correlated and well-characterized outcomes, which can help improve the power to detect novel variants if properly analyzed, as suggested by increasing evidence that pleiotropy, where a genetic variant affects multiple traits, is the norm in genome-phenome associations. We aim to develop pleiotropy informed powerful association test methods across multiple traits for GWAS. Since it is generally very hard to access individual-level GWAS phenotype and genotype data for those existing GWAS, due to privacy concerns and various logistical considerations, we develop rigorous statistical methods for pleiotropy informed adaptive multi-trait association test methods that need only summary association statistics publicly available from most GWAS. We first develop a pleiotropy test, which has powerful performance for truly pleiotropic variants but is sensitive to the pleiotropy assumption. We then develop a pleiotropy informed adaptive test that has robust and powerful performance under various genetic models. We develop accurate and efficient numerical algorithms to compute the analytical p-value for the proposed adaptive test without the need of resampling or permutation. We illustrate the performance of proposed methods through application to joint association test of GWAS meta-analysis summary data for several glycemic traits. Our proposed adaptive test identified several novel loci missed by individual trait based GWAS meta-analysis. All the proposed methods are implemented in a publicly available R package.

* baolin@umn.edu.

Supporting Information

Data on glycemic traits have been contributed by MAGIC investigators and have been downloaded from www.magicinvestigators.org. An R package implementing the proposed methods is publicly available at <http://www.github.com/baolinwu/MTAR>. Supporting information referenced in Sections 3, 4 and 5 and sample scripts to install and use the R package are available with this paper at the *Biometrics* website on Wiley Online Library.

Keywords

Adaptive association test; Glycemic traits; GWAS; Meta-analysis; Pleiotropy

1. Introduction

In the past decades many significant variants have been identified in genome-wide association studies (GWAS; Visscher et al., 2017). Despite these discoveries, there are still substantial missing heritability that may yet be explained by common genetic variants with smaller effect sizes or rare and low frequency variants (Manolio et al., 2009). This has led to the development of GWAS with increasingly large sample sizes. We note that many existing GWAS are conducted in large cohort studies with well-characterized phenotypes including many correlated phenotypes, but current GWAS are primarily based on the paradigm of testing single outcome. An increasingly acknowledged useful approach is jointly testing multiple correlated phenotypes, which can significantly improve the power to detect novel genetic variants. Especially for those pleiotropic variants, which have effects on multiple phenotypes, properly analyzing multiple phenotypes jointly can further improve power to detect pleiotropic variants.

There are increasing evidences suggesting that pleiotropy is the norm in genome-phenome associations (Solovieff et al., 2013; Pickrell et al., 2016; Visscher and Yang, 2016). However, many existing methods for testing multi-trait association do not explicitly account for pleiotropy information (e.g., Ferreira and Purcell, 2009; Liu et al., 2009; Yang et al., 2010; Rasmussen-Torvik et al., 2010; Lee et al., 2012; O'Reilly et al., 2012; Maity et al., 2012; van der Sluis et al., 2013; He et al., 2013; Schifano et al., 2013; Tang and Ferreira, 2012; Seoane et al., 2014). Furthermore, generally it is difficult to access the individual-level GWAS phenotype and genotype data due to privacy concerns and various logistical considerations. Nonetheless for most published GWAS, the association test summary statistics are publicly available. They include the minor allele frequency (MAF), the estimated regression parameters and their standard errors, and significance p-values for each analyzed SNP. Therefore, it would be useful to develop new multi-trait association test methods that just depend on these GWAS summary data (Pasaniuc and Price, 2017). One such class of pleiotropy informed testing approach is based on the false discovery rate (Andreassen et al., 2013a,b; Chung et al., 2014). They typically assume *multiple but disjoint GWAS data*, and jointly model the distribution of individual GWAS summary statistics (e.g., z-scores or p-values). Due to their nature of combinatorial modeling (Chung et al., 2014), it is generally difficult to extend them to simultaneously model more than two GWAS data. Assuming *disjoint GWAS summary statistics* essentially amounts to assuming independence and thus largely simplifies the complexity of modeling.

In this work, we aim to develop efficient GWAS summary data based multi-trait test methods in the same vein as the works of Zhu et al. (2015); Ray and Boehnke (2018). Specifically we develop powerful and robust pleiotropy informed adaptive multi-trait association test methods using multiple GWAS summary statistics from a same cohort of samples. We want to note that the pleiotropy effects discussed in the paper refer to the

pleiotropy association of a variant with multiple traits. The multi-trait association test methods mainly focus on testing the null hypothesis that none of the traits are associated with the variant, and represent a first step toward dissecting the genetic architecture of complex traits. After the methods have identified a significant joint association, more efforts are needed to further dissect the causal structure of multi-trait association.

The rest of the paper is organized as following. We introduce the proposed pleiotropy informed adaptive multi-trait association test in Section 2. Efficient numerical algorithm is developed to quickly and accurately compute analytical p-values without the need of permutation or resampling. And hence the proposed methods are scalable to millions of SNPs. Section 3 is devoted to simulation studies and an application to glycemic GWAS meta-analysis summary data is discussed in Section 4. We end the paper with a discussion in Section 5. All technical derivations are delegated to the appendix.

2. Statistical methods

To conduct association test using the GWAS summary statistics for multiple traits with properly controlled type I errors, we first need to estimate the trait correlation matrix. Stephens (2013) and Zhu et al. (2015) noted that the GWAS summary Z-statistics (standardized regression parameter estimates) across different traits share the same correlation as the traits. Hence a simple and commonly used approach is to estimating the correlation matrix using the sample correlation matrix of summary Z-statistics, which however may perform unsatisfactorily: (i) many SNPs are non-null exhibiting some effects due to the polygenic/omnigenic nature of most traits (Boyle et al., 2017); and (ii) significant genetic covariances have been observed among many different traits (Bulik-Sullivan et al., 2015). Here we adopt the recently proposed LD score regression approach of Bulik-Sullivan et al. (2015) to more accurately estimate the trait correlation matrix using the GWAS summary statistics. The key insight is that the non-null contribution of genetic covariance to the summary statistics correlation can be quantified using the LD scores pre-computed from outside reference population sample (e.g., the 1000 Genomes Project; Abecasis et al., 2012).

Specifically consider GWAS summary statistics for K traits measured on the same cohort of individuals. Denote the trait correlation matrix as $\Sigma = (r_{kj})$. For simplicity of notation, consider a single SNP and denote its association test summary Z-statistics across K traits as $(z_1, \dots, z_K)$. Bulik-Sullivan et al. (2015) showed that under the polygenic model for most traits, $E(z_k z_j) = r_{kj} + g_{kj} \ell$ where ℓ is the SNP's LD score, which essentially sums its total LD with other SNPs, and can be *pre-computed* using the public panel of reference population. The intercept is the trait correlation r_{kj} and the LD score coefficient g_{kj} quantifies the genetic covariance/similarity between traits. The trait correlation matrix Σ was estimated by regressing $z_k z_j$ on the LD scores ℓ over genome-wide SNPs (Bulik-Sullivan et al., 2015).

2.1 Pleiotropy informed adaptive association test of multiple traits

Consider a SNP with association test summary Z-statistics $Z = (z_1, \dots, z_K)^T$. For simplicity of notation, denote the estimated correlation matrix as Σ . Our proposed multi-trait test is based on the statistic

$$Q_\rho = (1 - \rho)Z^T \Sigma^{-1} Z + \rho \frac{(\mathbf{1}_K^T \Sigma^{-1} Z)^2}{\mathbf{1}_K^T \Sigma^{-1} \mathbf{1}_K},$$

where $\mathbf{1}_K$ is a column vector of K ones and ρ parameterizes the degree of pleiotropy (similarity among variant effects). We then define the adaptive test as the minimum p-value $T = \min_\rho P(Q_\rho)$, where $P(Q_\rho)$ calculates the significance p-value for Q_ρ .

The first component of Q_ρ statistic, $Q = Z^T \Sigma^{-1} Z$, is an omnibus K -DF chi-square test, which works well for detecting general deviations from the null; the second component, $C = (\mathbf{1}_K^T \Sigma^{-1} Z)^2 / (\mathbf{1}_K^T \Sigma^{-1} \mathbf{1}_K)$, can be derived as an 1-DF score test under the assumption of pleiotropy: C performs the best for pleiotropy variants with similar effects across traits (see Appendix for technical details). To obtain optimal test power, we consider the adaptive test based on their weighted average, $Q_\rho = (1 - \rho)Q + \rho C$, and use the minimum p-value as the test statistic.

In the following, we first offer some intuitive justifications to the proposed tests and then show that the statistic Q_ρ is essentially a variance component test statistic accounting for pleiotropy.

2.2 Association model accounting for pleiotropy

Assume that we know the effect sizes up to some constant, $\gamma \propto \Gamma$, where Γ is a given vector. Under this assumption, we can then derive a score statistic, $Z^T \Sigma^{-1} \Gamma$, to test $H_0: \gamma = 0$. Under the extreme scenario of pleiotropy association with constant effect sizes, i.e., $\Gamma = \mathbf{1}_K$, it leads to the test C . Without any assumption on the effect sizes, intuitively we can estimate Γ with Z , leading to the omnibus test Q . Q_ρ tries to account for the unknown pleiotropy effects by taking a weighted average of these two tests with a parameter ρ . By adaptively estimating ρ from the data, we can achieve robust and powerful performance.

Assume that $z_j \sim \mathcal{N}(\delta_j, 1)$, where δ_j denotes the effect size for the j -trait. We can model the effect sizes $\gamma = (\delta_1, \dots, \delta_K)^T$ with a zero mean multivariate normal distribution with some covariance. Generally due to the correlation among traits, their effect sizes also tend to be correlated. We could assume the de-correlated effects $\Sigma^{-1/2} \gamma$ vary independently. So we can model $\text{Cor}(\Sigma^{-1/2} \gamma) = \mathbf{I}_K$, and hence $\text{Cor}(\gamma) = \Sigma$. Here $\mathbf{I}_K$ denotes the K -th order identity matrix. One extreme scenario of the pleiotropy association is that we have constant effect sizes, $\delta_j = \text{const}$. We note that this constant effect size assumption can be equivalently described as $\text{Cor}(\delta_j, \delta_k) = 1, \forall j, k$, i.e., $\text{Cor}(\gamma) = \mathbf{J}_K$, where $\mathbf{J}_K$ is a $K \times K$ matrix with all elements equal to one. To allow more flexible pattern of effect sizes, we can model $\text{Cor}(\gamma) = (1 - \theta)\Sigma + \theta\mathbf{J}_K$, where the weight θ parameterizes the correlation (degree of similarity; pleiotropy) among variant effects across traits.

Overall we can model the variation of effects with $\text{Cov}(\gamma) = \tau^2\{(1 - \theta)\Sigma + \theta\mathbf{J}_K\}$. Therefore marginally $Z \sim \mathcal{N}(0, \Sigma + \tau^2\{(1 - \theta)\Sigma + \theta\mathbf{J}_K\})$, and the multi-trait association test now reduces

to testing $H_0 : \tau^2 = 0$. We can then show that the variance component test statistic is (see Appendix for technical details)

$$Z^T \Sigma^{-1} \{(1 - \theta)\Sigma + \theta J_K\} \Sigma^{-1} Z = (1 - \theta) Z^T \Sigma^{-1} Z + \theta (\mathbf{1}_K^T \Sigma^{-1} Z)^2,$$

which is proportional to Q_ρ when setting $\rho = \theta (\mathbf{1}_K^T \Sigma^{-1} \mathbf{1}_K) / \{1 - \theta + \theta (\mathbf{1}_K^T \Sigma^{-1} \mathbf{1}_K)\}$. Given a data, by choosing the minimum p-value over ρ 's, we can adapt to the unknown pleiotropy.

We can derive the *exact formula* and an efficient algorithm to compute the *analytical p-value* of the proposed pleiotropy adaptive test (see Appendix for technical details).

In the next two sections, we conduct simulation studies and applications to illustrate the performance of the proposed methods.

3. Simulation study

We first conducted a set of simulations to investigate the impact of estimating trait correlation matrix using the GWAS summary data. The complete results are available at the Supporting Information Section S1. Our preliminary simulations suggested that the LD score regression approach could lead to more accurate estimates and better control of type I errors than the sample correlation matrix.

Next we investigate the performance of test methods under various simulation scenarios. For simplicity of computation, we assume known trait correlation matrix and directly simulate summary statistics. In Supporting Information Section S3.1, we have done simulations to check that this approach leads to the same conclusions regarding the relative performance of different methods as the full simulation of estimating trait correlation matrix based on the LD score regression approach.

To investigate the type I errors, we considered $K = 5, 10$ traits and set the trait correlation matrix Σ as an exchangeable matrix with pairwise correlation r (ranging from complete independence $r = 0$ to strong dependence $r = 0.8$). For each setting, we simulate 10^{10} null multivariate normal random vectors from $\mathcal{N}(0, \Sigma)$ to estimate the type I errors at the $\alpha = 10^{-6}, 10^{-7}, 5 \times 10^{-8}$ significance levels. Overall all proposed methods have well controlled type I errors under all settings. Table 1 summarizes the empirical type I errors for the proposed pleiotropy based test C (denoted as PT), the omnibus chi-square test Q (denoted as OT), and the adaptive test T (denoted as AT) with 0.4 pairwise trait correlation. We provide the complete results at the Supporting Information Section S2.

Next we evaluate the power under 5×10^{-8} significance level based on simulating multivariate normal random vectors from $\mathcal{N}(\Delta, \Sigma)$, where Δ denotes the marginal variant effect sizes across traits. We include the minimum individual trait test p-value (denoted as minP) in the comparison. The analytical p-value of minP can be quickly computed based on the multivariate normal distribution function (Genz et al., 2018; Botev, 2017). In addition, we also include the following tests in the comparison: the sum test (ST), which directly sums the summary Z-statistics; and the sum of squares test (SQT), which is defined as the sum of

squared summary Z-statistics. Both ST and SQT can be treated as some variant of the metaanalysis test. The metaUSAT approach of Ray and Boehnke (2018) is based on adaptively weighting the SQT and OT. We found significantly inflated type I errors associated with metaUSAT in our simulation studies (see Supporting Information Section S2.1). So we do not include its results in the main paper. Instead, we study a slightly revised implementation of metaUSAT with valid type I error control and include the comparison results in the Supporting Information (Section S2 and S3).

For the power comparison, we have considered (1) varying the number of traits associated with the variant; (2) simulating the effect sizes from various distributions; (3) various pairwise trait correlation patterns: ranging from complete independence $r = 0$ to strong dependence $r = 0.8$; and a mix of positive and negative correlations. We have obtained similar conclusions across all simulation scenarios and provided the complete results at the Supporting Information Section S2.

Here we report the empirical power estimated over 10^6 simulations for $K = 5, 10$ with a medium correlation at Table 2 and 3. We have considered five scenarios of simulating the effect sizes. The first three scenarios, where we independently simulate each component of from the uniform distribution $U(1, 4)$, $U(0, 4)$ and the normal distribution $\mathcal{N}(2, 1)$, have similar effect sizes and generally favor the performance of the PT. The PT performs the best under the first scenario with very similar effect sizes across traits. But it has largely reduced power for the last two scenarios, which simulate the effect sizes from $U(-4, 4)$ and $\mathcal{N}(0, 3)$ leading to a mix of positive and negative effect sizes. In contrast, both the OT and AT have more robust and consistent performance. Generally the AT performs the best when the PT has decent power. When we have more traits associated with the variant (from $K = 5$ to $K = 10$), we also see the potential of more power gain for the PT and AT with similar effect sizes. Overall the first three alternative methods (minP, ST and SQT) perform less favorably compared to the three proposed methods, which explicitly account for the correlation among traits.

In addition, we have conducted another set of simulations for three traits with their correlation structure set based on the glycemic traits data analyzed in the following section. Overall we have observed the same conclusions regarding the relative performance of different methods. We provide the complete results at the Supporting Information Section S1 and S3.

4. Analysis of glycemic GWAS summary results

We analyze the GWAS meta-analysis results for fasting glucose (FG), fasting insulin (FI), and indices of β -cell function (HOMA-B) based on 46,186 non-diabetic European samples conducted by the international MAGIC consortium (Dupuis et al., 2010). With the summary

statistics, we first obtain the trait correlation matrix estimate,
$$\begin{bmatrix} 1 & 0.2646 & -0.2317 \\ 0.2646 & 1 & 0.7451 \\ -0.2317 & 0.7451 & 1 \end{bmatrix},$$

based on the LD score regression approach (Bulik-Sullivan et al., 2015). We apply the omnibus chi-square test Q , 1-DF pleiotropy test C , and the pleiotropy adaptive test T based

on the minimum p-value of Q and C . We also apply the minimum individual trait test p-values (minP) as a benchmark.

Figure 1 shows the venn diagram for the total number of identified significant loci by the proposed joint association test methods at the 5×10^{-8} genome-wide significance level. The omnibus chi-square test Q identified 15 significantly associated loci, and the individual trait test based minP only identified 11 of them. The pleiotropy test C identified 3 of the 15 significant loci, and 1 additional significant locus not identified by Q . The adaptive test T identified all 15 significant loci identified by Q , and the additional significant locus identified by C . Overall the adaptive test T performs the best. The supporting information provided the detailed information for the list of 16 significantly associated loci.

Table 4 lists the four additional significant loci identified by joint tests but missed by individual trait based tests. Two of them are common variants and the other two are less frequent. We note the MAGIC consortium performed a followup GWAS meta-analysis of fasting glucose in up to 58,074 non-diabetic European samples (Manning et al., 2012), and a followup replication study using a Metabochip with around 133,010 non-diabetic participants (Scott et al., 2012). Both studies have confirmed that the rs340874 SNP in *PROX1* gene is associated with fasting glycemia. The identified novel SNP, rs7116432, has clearly pleiotropic signals: the three individual Z-scores are very similar, and none are close to genome-wide significance individually. Both the PT C and AT T report genome-wide significant p-values, 2.7×10^{-8} and 4.99×10^{-8} respectively. The OT Q reports a non-significant p-value, 4.5×10^{-7} . The rs7116432 SNP is in the *CD44* gene, which has been confirmed associated with hyperglycemia, insulin resistance and type 2 diabetes (Kodama et al., 2012; Liu et al., 2015; Kodama et al., 2015). More research is needed to further study the identified novel variant rs7116432 and its role in glycemia control.

In our numerical studies, we have searched over $\rho \in \{0, 0.1, 0.2, 0.3, \dots, 0.8, 0.9, 1\}$ for the smallest p-value in the AT. For each SNP, we record $\hat{\rho}$ that leads to the smallest p-value. Figure 2 shows the empirical distribution of $\hat{\rho}$ across all SNPs. We have divided them into three groups based on the AT p-value ($P < 10^{-6}$, $10^{-6} < P < 10^{-3}$, $P > 10^{-3}$ respectively). Note that $\rho = 1$ corresponds to the PT (assuming pleiotropy association of similar trait effects) and $\rho = 0$ corresponds to the OT. Overall we can see the trend that for those SNPs with less significant AT p-values, there are increasingly more proportion of SNPs showing pleiotropy effects. Specifically for those less significant SNPs (AT $P > 10^{-3}$), majority of them show some degrees of pleiotropy effects. This is consistent with the increasingly recognized observation that pleiotropy is the norm in genome-phenome associations (Visscher and Yang, 2016). It also suggests the potential of the proposed methods to discover more novel variants for future studies of polygenic traits with larger sample sizes.

5. Discussion

In the past decade, many significant variants have been identified in GWAS. However there are still significant missing heritability for most traits. Given their polygenic/omnigenic nature, there remain many more genetic variants with small effects to be discovered. This has called for the development of association studies with increased sample sizes and novel

statistical methods with more detection power. Recently there have been increased research interest in integrating multiple correlated traits to boost the power to detect more genetic variants. This is partly owing to the increasingly recognized fact that pleiotropy is the norm in genome-phenome associations. Most existing multi-trait association test methods do not explicitly account for the pleiotropy information. In this work, we develop novel and powerful association test methods that explicitly account for potential pleiotropy association and also have robust performance. For the proposed methods, we develop efficient and accurate methods to quickly compute the analytical p-values without the need of resampling or permutation, and hence they are scalable to genome-wide association test. Various multitrait testing methods have been developed along the line of Stephens (2013) and Zhu et al. (2015), though generally they require computing-intensive Monte Carlo or MCMC sampling to compute p-values (see, e.g., Kim et al., 2015 and Shim et al., 2015). Our proposed methods provide an alternative approach with extreme scalability and robust performance. In the same spirit as many current approaches (Stephens, 2013; Zhu et al., 2015; Shim et al., 2015; Zhu and Stephens, 2017), the proposed methods just need the publicly available GWAS summary statistics without the need to access raw genotype and phenotype data. We expect that these newly developed GWAS summary data based joint testing methods can potentially identify more novel variants and shed more lights into the underlying disease mechanism.

A key component to the multi-trait association test of GWAS summary data has been the estimation of marginal trait correlation matrix. We have conducted more numerical studies to investigate the performance of empirical correlation matrix and the LD score regression approach based estimate (please see Supporting Information Section S1). Our results have suggested that the LD score regression generally provides more accurate estimates and better control of type I errors, though we want to note that an implicit assumption is that the outside reference population samples should come from the same population as the GWAS summary data. The LD score regression approach has been primarily applied to the European samples in the field. It will be worthwhile to study methods that can be adapted to GWAS summary data of more diverse background, e.g., the admixed African samples.

In this paper, we have mainly focused on the test methods that integrate GWAS summary data from all traits and have closed-form formulas to quickly compute significance p-values. It will be interesting to generalize the methods to adaptively integrate informative traits, which poses some challenges for statistical inference due to selection and should take into account the joint effects of all traits. Our preliminary studies have suggested that integrating correlated null traits can also boost the joint association test power, which have also been reported in previous studies (Stephens, 2013; Wu and Pankow, 2015), though no statistical justifications have been explicitly given. In the Supporting Information Section S4, we provide numerical results and also offer some statistical insights and justifications. In summary, the correlated null traits can help to explain some of the test statistics variation in those variant associated traits and thus improve the overall association test power.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

It is wished to be known that the first two authors contributed equally to this work. This research was supported in part by NIH grants GM083345 and CA134848. We acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Minnesota for providing resources that contributed to the research results reported within this paper. We would like to thank the associate editor and reviewers for their constructive comments, which have greatly improved the presentation of the paper.

Appendix Pleiotropy informed adaptive test

Assuming $Z = \mu\Gamma$, where μ is a numeric parameter and Γ is a given vector, we model $Z \sim \mathcal{N}(\mu\Gamma, \Sigma)$. The log likelihood part involving μ is $-(Z - \mu\Gamma)^T \Sigma^{-1} (Z - \mu\Gamma)/2$. We can then readily check that the score test for $H_0: \mu = 0$ is proportional to $U = Z^T \Sigma^{-1} \Gamma$, and $Cov(U) = \Gamma^T \Sigma^{-1} \Gamma$.

Under the assumption of common effects, we have $\Gamma = \mathbf{1}_K$ and can derive the score test C . Define $\eta = \Sigma^{-1} \mathbf{1}_K$. We have $C = (\eta^T Z)^2 / (\eta^T \Sigma \eta)$. Under null, $C \sim \chi_1^2$. Define a rank one projection matrix, $H = \Sigma^{1/2} \eta (\Sigma^{1/2} \eta)^T / (\eta^T \Sigma \eta)$. Consider the transformed random variables $Y = \Sigma^{-1/2} Z$. Under null, $Y \sim \mathcal{N}(0, \mathbf{I}_K)$. We have $Q = Y^T Y$ and $C = Y^T H Y$. Under null, $W = Z^T Z - C = Y^T (\mathbf{I}_K - H) Y \sim \chi_{K-1}^2$ and W is independent of C , since $H(\mathbf{I}_K - H) = 0$. We have $Q_\rho = (1 - \rho)W + C$.

Denote the p-value of $Q_\rho = (1 - \rho)Q + \rho C$ as p_ρ . Given a grid of ρ 's: $\rho_1 < \dots < \rho_M = 1$, denote the minimum p-value as $P = \min_{1 \leq i \leq M} p_{\rho_i}$. Denote the observed minimum p-value as T , and let q_i be the $(1 - T)$ -th quantile of distribution of Q_{ρ_i} . We compute the p-value of T as

$$\Pr(P \leq T) = 1 - \Pr(Q_{\rho_i} \leq q_i, i = 1, \dots, M).$$

Denote $v_x = \min_{i < M} \{(q_i - x)/(1 - \rho_i)\}$. We have

$$\Pr(Q_{\rho_i} \leq q_i, \forall i) = \Pr(W \leq v_C; C \leq q_M) = \int_0^{q_M} F_{\chi_{K-1}^2}^{(v_x)} f_{\chi_1^2}^{(x)} dx,$$

where $f_{\chi_1^2}^{(x)}$ denotes the density of χ_1^2 distribution, and $F_{\chi_{K-1}^2}^{(v_x)}$ the χ_{K-1}^2 distribution function.

Note $\int_0^{q_M} f_{\chi_1^2}^{(x)} dx = 1 - T$, hence

$$\Pr(P \leq T) = 1 - \int_0^{q_M} F_{\chi_{K-1}^2}^{(v_x)} f_{\chi_1^2}^{(x)} dx = T + \int_0^{q_M} [1 - F_{\chi_{K-1}^2}^{(v_x)}] f_{\chi_1^2}^{(x)} dx.$$

For a given ρ value, Q_ρ is distributed as the weighted sum of independent χ_1^2 and χ_{K-1}^2 random variables. Their distribution can be accurately computed using the Davies method (Davies, 1980; Wu et al., 2016).

Appendix Variance component test

Given $Z \sim \mathcal{N}(0, \Sigma + \tau^2[(1 - \theta)\Sigma + \theta J_K])$, we want to test $H_0 : \tau^2 = 0$. The log likelihood is

$$L = -\frac{1}{2} \log |\Sigma + \tau^2[(1 - \theta)\Sigma + \theta J_K]| - \frac{1}{2} Z^T \{\Sigma + \tau^2[(1 - \theta)\Sigma + \theta J_K]\}^{-1} Z.$$

We can check that

$$\begin{aligned} \frac{\partial L}{\partial \tau^2} = & -\frac{1}{2} Z^T (\Sigma + \tau^2[(1 - \theta)\Sigma + \theta J_K])^{-1} [(1 - \theta)\Sigma + \theta J_K] (\Sigma + \tau^2[(1 - \theta)\Sigma + \theta J_K])^{-1} Z \\ & - \frac{1}{2} \text{Tr} \left\{ \{\Sigma + \tau^2[(1 - \theta)\Sigma + \theta J_K]\}^{-1} [(1 - \theta)\Sigma + \theta J_K] \right\}, \end{aligned}$$

where $\text{Tr}()$ denotes matrix trace. We use the part involving data Z as the test statistic, leading to the following variance component test statistic at $\tau^2 = 0$

$$Z^T \Sigma^{-1} [(1 - \theta)\Sigma + \theta J_K] \Sigma^{-1} Z = (1 - \theta) Z^T \Sigma^{-1} Z + \theta (1_K^T \Sigma^{-1} Z)^2.$$

From the perspective of maximizing power, we should try to use the true correlation structure of the effect sizes for each individual variant, which however are unknown in practice, and our assumed structure may not hold in practice leading to suboptimal test power. We want to emphasize that the proposed tests will still preserve the type I errors, even though the modeling assumptions for the effect sizes are wrong.

References

- Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. (2012) An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, 56–65. [PubMed: 23128226]
- Andreassen O, Djurovic S, Thompson W, Schork A, Kendler K, ODonovan M, et al. (2013a) Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *The American Journal of Human Genetics*, 92, 197–209. [PubMed: 23375658]
- Andreassen OA, Thompson WK, Schork AJ, Ripke S, Mattingsdal M, Kelsoe JR, et al. (2013b) Improved detection of common variants associated with schizophrenia and bipolar disorder using pleiotropy-informed conditional false discovery rate. *PLoS Genetics*, 9, e1003455. [PubMed: 23637625]
- Botev ZI (2017) The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 1–24.
- Boyle EA, Li YI, Pritchard JK (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell*, 169, 1177–1186. [PubMed: 28622505]
- Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. (2015) An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, 47, 1236–1241. [PubMed: 26414676]
- Chung D, Yang C, Li C, Gelernter J, Zhao H (2014) GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. *PLoS Genetics*, 10, e1004787. [PubMed: 25393678]

- Davies RB (1980) Algorithm AS 155: the distribution of a linear combination of χ^2 random variables. *Applied Statistics*, 29, 323–333.
- Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, Jackson AU, et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genetics*, 42, 105–116. [PubMed: 20081858]
- Ferreira MAR and Purcell SM (2009) A multivariate test of association. *Bioinformatics*, 25, 132–133. [PubMed: 19019849]
- Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, et al. (2018). mvtnorm: multivariate normal and t distributions. R package version 1.0-8. <https://cran.r-project.org/web/packages/mvtnorm> (accessed August 1, 2018).
- He Q, Avery CL, Lin DY (2013) A general framework for association tests with multivariate traits in large-scale genomics studies. *Genetic Epidemiology*, 37, 759–767. [PubMed: 24227293]
- Kim J, Bai Y, Pan W (2015) An adaptive association test for multiple phenotypes with GWAS summary statistics. *Genetic Epidemiology*, 39, 651–663. [PubMed: 26493956]
- Kodama K, Horikoshi M, Toda K, Yamada S, Hara K, Irie J, et al. (2012) Expression-based genome-wide association study links the receptor CD44 in adipose tissue with type 2 diabetes. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 7049–7054. [PubMed: 22499789]
- Kodama K, Toda K, Morinaga S, Yamada S, Butte AJ (2015) Anti-CD44 antibody treatment lowers hyperglycemia and improves insulin resistance, adipose inflammation, and hepatic steatosis in diet-induced obese mice. *Diabetes*, 64, 867–875. [PubMed: 25294945]
- Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR (2012) Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, 28, 2540–2542. [PubMed: 22843982]
- Liu J, Pei Y, Papasian CJ, Deng HW (2009) Bivariate association analyses for the mixture of continuous and binary traits with the use of extended generalized estimating equations. *Genetic Epidemiology*, 33, 217–227. [PubMed: 18924135]
- Liu LF, Kodama K, Wei K, Tolentino LL, Choi O, Engleman EG, et al. (2015) The receptor CD44 is associated with systemic insulin resistance and proinflammatory macrophages in human adipose tissue. *Diabetologia*, 58, 1579–1586. [PubMed: 25952479]
- Maity A, Sullivan PF, Tzeng JY (2012) Multivariate phenotype association analysis by marker-set kernel machine regression. *Genetic Epidemiology*, 36, 686–695. [PubMed: 22899176]
- Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nature Genetics*, 44, 659–669. [PubMed: 22581228]
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. (2009) Finding the missing heritability of complex diseases. *Nature*, 461, 747–753. [PubMed: 19812666]
- O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, Jarvelin MR, Coin LJM (2012) MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS One*, 7, e34861. [PubMed: 22567092]
- Pasaniuc B and Price AL (2017) Dissecting the genetics of complex traits using summary association statistics. *Nature Reviews Genetics*, 18, 117–127.
- Pickrell JK, Berisa T, Liu JZ, Sagurel L, Tung JY, Hinds DA (2016) Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics*, 48, 709–717. [PubMed: 27182965]
- Rasmussen-Torvik LJ, Alonso A, Li M, Kao W, Kattgen A, Yan Y, et al. (2010) Impact of repeated measures and sample selection on genome-wide association studies of fasting glucose. *Genetic Epidemiology*, 34, 665–673. [PubMed: 20839289]
- Ray D and Boehnke M (2018). Methods for meta-analysis of multiple traits using gwas summary statistics. *Genetic Epidemiology*, 42, 134–145. [PubMed: 29226385]
- Schifano E, Li L, Christiani D, Lin X (2013) Genome-wide association analysis for multiple continuous secondary phenotypes. *The American Journal of Human Genetics*, 92, 744–759. [PubMed: 23643383]

- Scott RA, Lagou V, Welch RP, Wheeler E, Montasser ME, Luan J, et al. (2012) Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nature Genetics*, 44, 991–1005. [PubMed: 22885924]
- Seoane JA, Campbell C, Day INM, Casas JP, Gaunt TR (2014) Canonical correlation analysis for gene-based pleiotropy discovery. *PLoS Computational Biology*, 10, e1003876. [PubMed: 25329069]
- Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, Krauss RM, Stephens M (2015). A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS One*, 10, e0120758. [PubMed: 25898129]
- Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW (2013) Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics*, 14, 483–495.
- Stephens M (2013) A unified framework for association analysis with multiple related phenotypes. *PLoS One*, 8, e65245. [PubMed: 23861737]
- Tang CS and Ferreira MAR (2012) A gene-based test of association using canonical correlation analysis. *Bioinformatics*, 28, 845–850. [PubMed: 22296789]
- van der Sluis S, Posthuma D, Dolan CV (2013) TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genetics*, 9, e1003235. [PubMed: 23359524]
- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, and Yang J (2017) 10 years of GWAS discovery: biology, function, and translation. *American Journal of Human Genetics* 101, 5–22. [PubMed: 28686856]
- Visscher PM and Yang J (2016) A plethora of pleiotropy across complex traits. *Nature Genetics*, 48, 707–708. [PubMed: 27350602]
- Wu B, Guan W, Pankow JS (2016) On efficient and accurate calculation of significance p-values for sequence kernel association test of variant set. *Annals of Human Genetics*, 80, 123–135. [PubMed: 26757198]
- Wu B and Pankow JS (2015) Statistical methods for association tests of multiple continuous traits in genome-wide association studies. *Annals of Human Genetics*, 79, 282–293. [PubMed: 25857693]
- Yang Q, Wu H, Guo CY, Fox CS (2010) Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. *Genetic Epidemiology*, 34, 444–454. [PubMed: 20583287]
- Zhu X, Feng T, Tayo B, Liang J, Young JH, Franceschini N, et al. (2015) Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *The American Journal of Human Genetics*, 96, 21–36. [PubMed: 25500260]
- Zhu X and Stephens M (2017). Bayesian large-scale multiple regression with summary statistics from genome-wide association studies. *The Annals of Applied Statistics*, 11, 1561–1592. [PubMed: 29399241]

Figure 1. Venn diagram for the total number of identified significant loci by the proposed joint association test methods.

Figure 2.
Distribution of estimated degree of pleiotropy for the glycemic traits data.

Table 1

Ratio of empirical type I errors divided by the significance level α estimated over 10^{10} null simulations: listed within parenthesis are the standard errors.

	10^{-6}	$K = 5$ 10^{-7}	5×10^{-8}	10^{-6}	$K = 10$ 10^{-7}	5×10^{-8}
OT	1.01 (0.01)	0.99 (0.03)	1.07 (0.05)	1.01 (0.01)	1.02 (0.03)	0.98 (0.04)
PT	0.99 (0.01)	0.97 (0.03)	0.98 (0.05)	1.01 (0.01)	1.00 (0.03)	0.96 (0.04)
AT	1.00 (0.01)	1.03 (0.03)	1.06 (0.05)	1.02 (0.01)	0.98 (0.03)	0.99 (0.04)

Table 2

Power (%) under 5×10^{-8} significance level for testing $K = 5$ traits: all traits have 0.5 pairwise correlation except the first trait, which has negative correlation -0.5 with all the other traits. PT is the pleiotropy based test; OT is the omnibus chi-square test; AT is the adaptive test; ST is the sum test, SQT is the quadratic sum test, and minP is based on the minimum marginal test p-value. Data are simulated from $\mathcal{N}(\Delta, \Sigma)$.

	minP	ST	SQT	PT	OT	AT
U(1,4)	3.02	32.9	2.69	68.4	62.4	68.8
U(0,4)	2.32	17.1	1.13	47.6	47.4	51.5
$\mathcal{N}(2,1)$	2.30	16.1	0.88	47.5	43.9	49.6
U(-4,4)	2.15	2.29	0.24	15.3	58.0	56.3
$\mathcal{N}(0, 3)$	31.8	6.19	10.2	23.5	74.0	72.9

Table 3

Power (%) under 5×10^{-8} significance level for testing $K = 10$ traits: all traits have 0.5 pairwise correlation except the first trait, which has negative correlation -0.5 with all the other traits. PT is the pleiotropy based test; OT is the omnibus chi-square test; AT is the adaptive test; ST is the sum test, SQT is the quadratic sum test, and minP is based on the minimum marginal test p-value. Data are simulated from $\mathcal{N}(\Delta, \Sigma)$.

	minP	ST	SQT	PT	OT	AT
U(1,4)	4.16	11.1	1.89	87.4	81.3	90.3
U(0,4)	2.36	3.17	0.60	56.7	67.9	73.0
$\mathcal{N}(2,1)$	2.89	2.77	0.44	58.1	59.2	69.1
U(-4,4)	2.67	0.04	0.08	7.87	94.1	93.2
$\mathcal{N}(0, 3)$	49.3	0.33	7.78	18.2	97.5	97.1

Table 4

Association analysis of MAGIC GWAS summary data: significant loci identified by AT, but missed by individual trait based tests.

SNP	Chr	Gene	Joint test p-values				Dupuis et al. (2010) Z-scores			Scott et al. (2012) FG test p-value
			minP	PT	OT	AT	FG	FI	HOMA-B	
rs340874	1	PROX1	2.1×10^{-7}	8.3×10^{-2}	3.5×10^{-9}	6.6×10^{-9}	-5.38	2.07	4.12	4.1×10^{-10}
rs958444	5	SGCD	1.9×10^{-1}	3.5×10^{-5}	2.6×10^{-8}	4.9×10^{-8}	1.84	-1.25	0.86	-
rs4836906	9	OR1L1	3.6×10^{-2}	4.7×10^{-8}	2.1×10^{-9}	3.9×10^{-9}	-2.08	-0.47	-2.55	-
rs7116432	11	CD44	6.3×10^{-5}	2.7×10^{-8}	4.5×10^{-7}	4.99×10^{-8}	-3.33	-4.15	-3.61	-