OXFORD

Genetics and population analysis

Integrate multiple traits to detect novel trait–gene association using GWAS summary data with an adaptive test approach

Bin Guo and Baolin Wu*

Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455, USA

*To whom correspondence should be addressed.

Abstract

Motivation: Genetics hold great promise to precision medicine by tailoring treatment to the individual patient based on their genetic profiles. Toward this goal, many large-scale genome-wide association studies (GWAS) have been performed in the last decade to identify genetic variants associated with various traits and diseases. They have successfully identified tens of thousands of disease-related variants. However they have explained only a small proportion of the overall trait heritability for most traits and are of very limited clinical use. This is partly owing to the small effect sizes of most genetic variants, and the common practice of testing association between one trait and one genetic variant at a time in most GWAS, even when multiple related traits are often measured for each individual. Increasing evidence suggests that many genetic variants can influence multiple traits simultaneously, and we can gain more power by testing association of multiple traits simultaneously. It is appealing to develop novel multi-trait association test methods that need only GWAS summary data, since it is generally very hard to access the individual-level GWAS phenotype and genotype data.

Results: Many existing GWAS summary data-based association test methods have relied on *ad hoc* approach or crude Monte Carlo approximation. In this article, we develop rigorous statistical methods for efficient and powerful multi-trait association test. We develop robust and efficient methods to accurately estimate the marginal trait correlation matrix using only GWAS summary data. We construct the principal component (PC)-based association test from the summary statistics. PC-based test has optimal power when the underlying multi-trait signal can be captured by the first PC, and otherwise it will have suboptimal performance. We develop an adaptive test by optimally weighting the PC-based test and the omnibus chi-square test to achieve robust performance under various scenarios. We develop efficient numerical algorithms to compute the analytical *P*-values for all the proposed tests without the need of Monte Carlo sampling. We illustrate the utility of proposed methods through application to the GWAS meta-analysis summary data for multiple lipids and glycemic traits. We identify multiple novel loci that were missed by individual trait-based association test.

Availability and implementation: All the proposed methods are implemented in an R package available at http://www.github.com/baolinwu/MTAR. The developed R programs are extremely efficient: it takes less than 2 min to compute the list of genome-wide significant single nucleotide polymorphisms (SNPs) for all proposed multi-trait tests for the lipids GWAS summary data with 2.5 million SNPs on a single Linux desktop.

Contact: baolin@umn.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Genetics hold great promise to precision medicine that tailors the treatment to the individual patient by considering their genetic profiles. Many large-scale genome-wide association studies (GWAS) performed in the past decade have successfully identified thousands of common genetic variants associated with various traits and diseases. But for many traits, in total they have explained a small proportion of the overall heritability (Manolio *et al.*, 2009; Visscher *et al.*, 2017). It is expected that many more genetic variants remain to be identified. These GWAS are primarily based on the paradigm of 'single trait single variant association test' even with multiple traits measured for each individual. There is increasing evidence showing that genetic variants can influence multiple traits simultaneously, and we can gain great power by studying multiple traits simultaneously (see, e.g . Broadaway *et al.*, 2016; Ferreira and Purcell, 2009; Galesloot *et al.*, 2014; Maity *et al.*, 2012; O'Reilly *et al.*, 2012; Seoane *et al.*, 2014; Stephens, 2013; Tang and Ferreira, 2012; Wu and Pankow, 2016; Zhu *et al.*, 2015). Ideally, we can re-analyze those existing GWAS data published in the last decade using the multi-trait association test approach to identify more novel genetic variants. However due to privacy concerns and various logistical considerations, it is generally very hard to access the individual-level GWAS phenotype and genotype data, which creates a barrier to further mine these existing data to extract more information.

Nevertheless most published GWAS have made the association test summary statistics publicly available. They include, e.g. the minor allele frequency (MAF), the estimated effect sizes with their SEs, and significance *P*-values for each single nucleotide polymorphism (SNP) analyzed in a GWAS. One viable solution is to develop new association test methods that depend on only these GWAS summary data (Pasaniuc and Price, 2017). For example, for the single variant-based association test, the GWAS meta-analysis (Evangelou and Ioannidis, 2013) is typically conducted based on the summary statistics, which can be as efficient as analyzing individual-level data across all studies (Lin and Zeng, 2010). Similar methods have been developed for meta-analysis of the rare variant set association across studies (Hu *et al.*, 2013; Lee *et al.*, 2014). For joint association test of a single variant with multiple traits, Stephens (2013) and Zhu *et al.* (2015) proposed methods using only individual GWAS summary statistics and GWAS meta-analysis summary results. The key insight of these approaches is that for a single variant, the summary Z-statistics across different traits share the same correlation as the traits (Stephens, 2013). This has motivated the widely used estimate based on the sample correlations of genome-wide summary Z-statistics, which implies that ideally only independent SNPs should be used in calculating the sample correlation matrix, and further nearly all of them are null SNPs. Therefore some forms of variant filtering based on the linkage disequilibrium (LD) and *P*-value are needed. We have found that using the empirical correlations can produce significantly biased estimates for polygenic traits. Recently there have been some research adopting the LD score regression to estimate the trait correlation using GWAS summary data (Baselmans *et al.*, 2017; Turley *et al.*, 2018). In this work, we conduct thorough investigation of the impact of trait correlation estimation, and show that the LD score regression can produce much more accurate estimates than the empirical correlation matrix. Using the empirical correlations can lead to increased false positives in the downstream association test. We further propose some efficient and adaptive multi-trait association test methods that have robust performance under a wide range of trait–gene association models.

The rest of the article is organized as following. We introduce the proposed multi-trait association test methods in Section 2.

Section 3 is devoted to simulation studies and Section 4 presents comprehensive analyses results of GWAS meta-analysis summary data for multiple lipids and glycemic traits. We end the article with a discussion in Section 5. All technical derivations are delegated to the Supplementary Material.

2 Materials and methods

Throughout the following discussion, we mainly focus on analyzing summary association statistics for multiple traits from a single cohort (either a single GWAS or GWAS meta-analysis). The proposed methods can be readily extended to partially overlapped GWAS (see Supplementary Section S1.3 for technical details). We take a two-step approach to conduct the summary statistics-based multi-trait association test: first, we estimate the trait correlation matrix using the summary statistics of all SNPs; second, with the estimated correlation matrix, we test the association of each SNP with multiple traits using developed methods detailed as follows.

2.1 Estimate trait correlation from summary statistics

When testing the marginal association of a SNP with multiple continuous traits, its summary Z-statistics across traits asymptotically follow the multivariate normal distribution with correlation, denoted as $\Sigma = (\rho_{ik})$, equal to the trait correlation matrix (see Stephens, 2013, Zhu *et al.*, 2015 and Supplementary Section S1.1). This has motivated the commonly used approach of empirically estimating Σ using the sample correlation of genome-wide summary Z-statistics, which implies that ideally only independent SNPs should be used in calculating the sample correlation matrix (therefore some LD pruning is required), and further nearly all of them are null SNPs (hence some *P*-value filtering is required). For traits of polygenic nature, as shown in Bulik-Sullivan *et al.* (2015a), in addition to the variant LD and trait dependence, both the trait heritability and genetic correlation contribute to the correlation of summary Z-statistics.

Consider a cohort of N samples with K quantitative outcomes, and assume we have the summary Z-statistics $(z_{1j}, \ldots, z_{Kj})$ for testing the marginal association of each trait with SNP $j = 1, \ldots, M$. We can check that

$$\mathbb{E}(z_{ij}z_{kj}) = \frac{(N+1)\rho_{g,ik}}{M}l_j + \rho_{ik}, \tag{1}$$

where ρ_{ik} is the correlation between the *i*th and *k*th outcomes, l_j denotes the LD score for SNP j (sum of its LD r^2 with all other SNPs), $\rho_{g,ik}$ measures the between trait genetic covariance. For traits of polygenic nature, there could be significant genetic heritability and covariance, leading to potentially non-ignorable $\rho_{g,ik}$. Naively using the sample correlation of summary Z-statistics will then lead to biased estimates of Σ. We can obtain more accurate estimates of Σ by regressing the pairwise product of summary statistics ($z_{ij}z_{kj}$) on l_j, as shown previously (Baselmans *et al.*, 2017; Turley *et al.*, 2018).

To reduce the impact of large summary statistics, it has been a common practice to filter out those large summary statistics (Bulik-Sullivan *et al.*, 2015a), which however is less efficient and can potentially lead to biased estimation. We propose to use a robust linear regression to incorporate all summary statistics: instead of removing those large summary statistics, we minimize their absolute differences in the regression.

2.2 Multi-trait association test

For simplicity of notation, denote the estimated correlation matrix as Σ, and consider a single SNP with the across-trait summary

statistics vector $Z = (z_1, \ldots, z_K)^T$. We can construct the following principal component (PC)-based association test (denoted as ET), $B = (Z^T u_1)^2 / d_1$, where d_1 is the largest eigenvalue and u_1 is the corresponding eigenvector computed from Σ. We can check that B asymptotically follows the χ_1^2 distribution. The ET performs well when the first PC captures majority of the association signals across multiple traits. Alternatively we can use the omnibus test (OT), $Q = Z^T \Sigma^{-1} Z$, which can detect any deviation from the null. We note that Turley *et al.* (2018) also adopted the OT Q to detect multi-trait SNP association. Multi-trait test can boost the association test power by leveraging the widespread pleiotropy (Solovieff *et al.*, 2013). We can show that the proposed ET can capture the mediated pleiotropy, while the OT performs well under general pleiotropy (Please see Supplementary Section S3 for more details). Q asymptotically follows the χ_K^2 distribution. Q generally has robust performance under different disease models. To obtain optimal test power, we can consider the adaptive test (denoted as AT) based on their weighted average, $A_\rho = (1 - \rho)Q + \rho B$, and use the minimum *P*-value, $T = \min_\rho P(A_\rho)$, as the test statistic. Here $P(A_\rho)$ denotes the *P*-value of A_ρ. We develop an efficient algorithm to quickly and *exactly* compute the analytical *P*-value of AT. The key idea is that we can decompose Q into two orthogonal components $Q = B + (Q - B)$, where $Q - B$ follows the χ_{K-1}^2 distribution. Hence $A_\rho = (1 - \rho)(Q - B) + B$, and we can then compute the AT *P*-value efficiently based on an 1D numerical integration (see Supplementary Section S2 for more details).

3 Results

3.1 Simulation study

3.1.1 Marginal trait correlation estimation
We first evaluate the estimation of between trait correlations using the GWAS summary statistics. We consider three continuous polygenic outcomes and set the genetic and environment correlation matrices based on the lipids data as $\Sigma_g = (0.08, -0.59, 0.39)$, $\Sigma_e = (-0.19, 0.77, -0.50)$ (here we only list the three pairwise correlations; see Supplementary Section S3.1 for details). We assume the same heritability h^2 for three traits and the marginal trait correlation matrix Σ is then $h^2 \Sigma_g + (1 - h^2)\Sigma_e$.

To mimic a true genome-wide LD structure in the simulation, we use the 9713 European GWAS samples from the Atherosclerosis Risk in Communities Study (ARIC; dbGaP: phs000280.v3.p1) and consider approximately 1.2 million HapMap3 common SNPs. We select $M = 6100$ independent causal SNPs that are ~400 KB apart. We further divide these causal SNPs into two sets: M_b of them have genetic covariance $h^2 \Sigma_g / (2M_b)$ with the rest having $h^2 \Sigma_g / (2M_s)$. Here $M_s = M - M_b$ and we consider $M_b = 200$ in our numerical studies. Hence, a small subset of SNPs will have relatively large effect sizes, while the majority of them have modest effect sizes.

We consider $h^2 = 0.1, 0.3, 0.5$ in the simulations, and compare two approaches for estimating Σ: (i) the sample correlation matrix of summary Z-statistics (excluding genome-wide significant SNPs), denoted as $\hat{\Sigma}^s$; (ii) the proposed robust LD score regression, denoted as $\hat{\Sigma}^r$; The LD scores are pre-computed based on the 1000 Genomes Project European samples (Abecasis *et al.*, 2012). The LD score regression of Bulik-Sullivan *et al.* (2015a) performs slightly worse than the robust LD score regression for estimating both trait and genetic correlations. We provided complete results at the Supplementary Section S3.1. Table 1 summarizes the bias and root-mean-square error (RMSE) computed over 100 simulations. The robust LD score regression-based approach performs much better than the naive sample correlation-

Table 1. Bias and RMSE of estimating marginal correlations Σ among three traits

		Bias			RMSE		
		Σ_{12}	Σ_{13}	Σ_{23}	Σ_{12}	Σ_{13}	Σ_{23}
$h^2 = 0.1$	$\hat{\Sigma}^s$	0.002	−0.027	0.018	0.012	0.030	0.022
	$\hat{\Sigma}^r$	−0.002	0.000001	−0.0003	0.011	0.016	0.014
$h^2 = 0.3$	$\hat{\Sigma}^s$	0.009	−0.058	0.039	0.017	0.062	0.044
	$\hat{\Sigma}^r$	−0.001	−0.004	0.003	0.014	0.020	0.018
$h^2 = 0.5$	$\hat{\Sigma}^s$	0.010	−0.061	0.043	0.022	0.068	0.049
	$\hat{\Sigma}^r$	−0.0005	−0.008	0.006	0.019	0.028	0.024

Note: h^2 is the overall heritability; $\Sigma = h^2 \Sigma_g + (1 - h^2)\Sigma_e$. $\hat{\Sigma}^s$ is the sample correlation-based estimate, and $\hat{\Sigma}^r$ is the LD score regression-based estimate.

Table 2. Ratio of empirical Type I errors divided by the significance level α estimated over 10^{10} simulations for the three proposed tests (OT, ET and AT)

	α	10^{-6}	10^{-7}	5×10^{-8}	10^{-8}
OT	$r = 0$	1.01 (0.01)	1.02 (0.03)	1.03 (0.05)	0.94 (0.10)
	$r = 0.2$	1.01 (0.01)	1.02 (0.03)	1.03 (0.05)	1.16 (0.11)
	$r = 0.5$	1.02 (0.01)	0.99 (0.03)	1.05 (0.05)	1.13 (0.11)
	$r = 0.8$	1.01 (0.01)	1.00 (0.03)	1.07 (0.05)	1.11 (0.11)
ET	$r = 0$	1.01 (0.01)	0.99 (0.03)	0.96 (0.04)	0.96 (0.10)
	$r = 0.2$	1.00 (0.01)	0.97 (0.03)	0.97 (0.04)	0.90 (0.09)
	$r = 0.5$	1.00 (0.01)	0.97 (0.03)	0.95 (0.04)	0.89 (0.09)
	$r = 0.8$	1.00 (0.01)	0.97 (0.03)	0.99 (0.04)	1.04 (0.10)
AT	$r = 0$	0.96 (0.01)	0.99 (0.03)	1.00 (0.04)	0.98 (0.10)
	$r = 0.2$	0.96 (0.01)	0.99 (0.03)	1.01 (0.04)	0.89 (0.09)
	$r = 0.5$	0.96 (0.01)	1.01 (0.03)	1.08 (0.05)	0.88 (0.09)
	$r = 0.8$	0.97 (0.01)	1.05 (0.03)	1.11 (0.05)	0.89 (0.09)

Note: Listed within parentheses are the SEs.

based estimates, which had much larger biases. The biased trait correlation estimates can lead to inflated or conservative Type I errors for the down-stream multi-trait association tests (see Supplementary Section S3.1)

3.1.2 Multi-trait association test
We evaluate the performance of proposed tests compared with the following tests: (i) the minimum marginal *P*-value across traits (denoted as minP); (ii) the sum of Z-statistics (denoted as SZ) along the line of fixed effects meta-analysis; (iii) the sum of squared Z-statistics (denoted as SZ2) along the line of heterogeneity effects meta-analysis; (iv) an adaptive test based on weighting SZ and SZ2 (denoted as AZ) in the same vein as the proposed AT. Specifically AZ is defined based on the minimum *P*-values of $(1 - \rho)SZ2 + \rho SZ^2$ over $\rho \in [0, 1]$. We develop an efficient algorithm to quickly compute its analytical *P*-value without the need of resampling (see Supplementary Material for details); (v) the metaUSAT method of Ray and Boehnke (2018), which is based on adaptively weighting the OT and SZ2 tests. We found inflated Type I errors of metaUSAT in our numerical studies (see Supplementary Material for details) and used a revised implementation with proper Type I error control in our comparison (denoted as MUSAT in the following discussion).

We first evaluate the Type I errors of proposed tests by simulating 10^{10} random vectors from $\mathcal{N}(0, \Sigma)$ for $K = 5$ traits. Here Σ is a compound symmetry correlation matrix with correlation $r \in (0, 0.2, 0.5, 0.8)$. Table 2 summarizes the empirical Type I errors at the significance

levels $\alpha = 10^{-6}, 10^{-7}, 5 \times 10^{-8}, 10^{-8}$. Overall all proposed tests controlled the Type I errors well. We have done more studies under various simulation settings and checked that the proposed tests properly controlled the Type I errors (please see Supplementary Section S5).

We evaluate the power of different methods under 5×10^{-8} significance level based on simulating 10^7 random vectors from $\mathcal{N}(\Delta, \Sigma)$, where Σ is a correlation matrix with $\Sigma_{12} = -0.11, \Sigma_{13} = -0.36$, $\Sigma_{23} = 0.23$. We consider setting Δ as fixed values, and randomly simulating Δ from $\mathcal{N}(0, 3)$ and uniform distributions, $U(-5, 5), U(0, 5)$. For fixed Δ, we decompose $\Delta = \sum_{j=1}^{3} \delta_j u_j$ with (u_1, u_2, u_3) being the three eigenvectors of Σ. When the signal vector Δ is completely captured by the first PC (i.e. $\delta_2 = \delta_3 = 0$), the PC-based test ET has the optimal power. Table 3 shows the estimated power under various settings. The first setting favors ET, which performs much better than OT. However ET is sensitive to the signal distribution, and performs much worse when the top PC has weak association signal. The minimum *P*-value-based test (minP) performs well when one of the traits dominates the association signal, but otherwise it generally has suboptimal performance. The SZ test performs well when all marginal effects follow the same directions. The SZ2 test is relatively more robust and performs well with multiple large marginal effects. In contrast, both OT and AT have very robust and consistent performance over all settings. The

adaptive test AT can truly combine the strength of both OT and ET and outperform both tests when they have comparable powers.

Overall, the first four tests (minP, SZ, SZ2 and AZ) do not explicitly account for the trait correlations, and generally have less favorable performance. We can compute their analytical *P*-values very efficiently. The PC-based test ET performs well when the top PC captures majority of the association signals. The SZ test has good performance when all marginal trait effects follow the same direction. In contrast, both OT and SZ2 are quadratic tests and have more robust performance. The adaptive test AT has very robust and consistent performance. We have conducted more simulation studies investigating the performance of different methods under various settings of different number of traits and trait dependence. The overall conclusions remain the same. We provide the complete results at the Supplementary Section S5

3.2 Application

We conduct comprehensive analysis of GWAS meta-analysis results for multiple lipids and glycemic traits. The proposed tests (OT, ET and AT) have performed better than the other three competing methods (minP, SZ, SZ2 and AZ). In the following, we mainly focus on results for the proposed methods, and leave the complete results

Table 3. Power (%) under 5×10^{-8} significance level

Δ	$(\delta_1, \delta_2, \delta_3)$	minP	SZ	SZ2	AZ	MUSAT	ET	OT	AT
$(3.59, -2.71, -3.97)$	$(6, 0, 0)$	6.45	0.02	28.76	4.32	22.70	30.15	17.00	26.42
$(2.56, -4.42, -3.73)$	$(6, 2, 0)$	13.21	2.69	38.25	15.15	32.31	30.12	29.86	34.25
$(4.82, -3.25, -2.49)$	$(6, 0, 2)$	21.14	0.00	38.10	12.70	35.27	30.13	36.46	39.11
$(5.33, -2.40, -2.61)$	$(6, -1, 2)$	37.72	0.00	40.59	14.83	38.27	30.13	40.14	41.95
$(4.63, -0.83, -2.81)$	$(5, -2, 1)$	15.60	0.00	15.47	2.96	13.21	9.01	13.77	14.57
$(3.22, 2.31, -3.21)$	$(3, -4, -1)$	1.49	0.00	7.09	5.20	15.24	0.15	19.97	16.59
$(1.52, -4.26, 1.26)$	$(2, 3, 3)$	8.32	0.00	2.30	7.29	19.86	0.01	25.41	21.40
$(5, 0, 0)$	$(2.99, 2.55, -3.09)$	25.94	1.07	5.36	12.69	23.63	0.14	29.76	25.42
$(2.71, -5.16, -0.07)$	$(4, 3, 3)$	31.43	0.01	21.50	22.78	44.40	1.51	51.68	46.74
$(-0.49, -2.71, -4.63)$	$(4, 2, -3)$	15.64	30.21	11.83	36.49	27.21	1.51	33.51	29.40
$\Delta \sim U(0, 5)$		9.09	35.11	13.55	40.12	35.41	0.17	39.19	36.53
$\Delta \sim U(-5, 5)$		9.08	10.11	13.94	24.75	30.43	2.93	33.30	31.60
$\Delta \sim N(0, 3)$		20.71	11.14	20.20	27.64	32.08	4.06	34.33	32.91

Note: ET is the PC-based test; OT is the omnibus chi-square test; AT is the adaptive test; minP is the minimum *P*-value-based test, SZ is the sum of Z-statistics, SZ2 is the sum of squared Z-statistics, and AZ is the test based on adaptively weighting SZ and SZ2; MUSAT is the revised metaUSAT test. Data are simulated from $\mathcal{N}(\Delta, \Sigma)$, where $\Delta = \sum_{j=1}^{3} \delta_j u_j$ with (u_1, u_2, u_3) being the three eigenvectors of Σ.
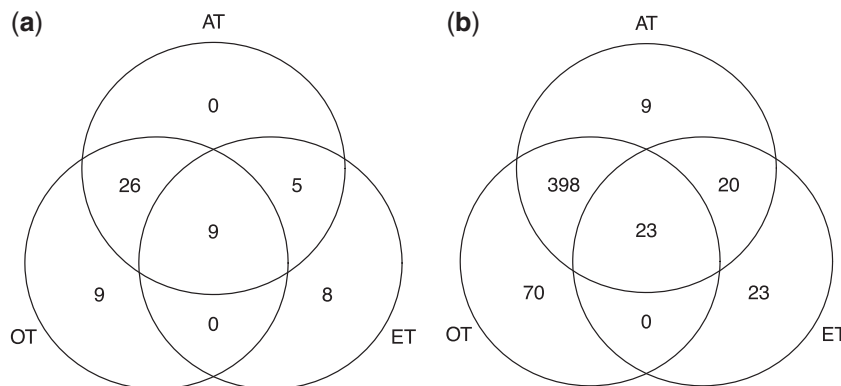

Fig. 1. Venn diagram of number of significantly associated loci and SNPs identified by three joint association test methods based on the lipids GWAS summary data. **(a)** Number of detected significant loci. **(b)** Number of detected significant SNPs

to the Supplementary Material. We note that it is more productive to treat the proposed tests as a complementary approach to the existing single trait-based test. We thus present joint association test results excluding genome-wide significant SNPs for any trait. The analysis results including all SNPs are provided at the Supplementary Section S5

3.2.1 Analysis of lipids GWAS results

We analyze the GWAS meta-analysis results for three plasma lipids (low-density lipoproteins (LDL) cholesterol, triglyceride and total cholesterol) based on around 100 000 European individuals from the Global Lipids Consortium (Teslovich *et al.*, 2010). We note that the Global Lipids Consortium conducted a followup study using a

Metabochip with a small panel of pre-selected SNPs based on 190 000 European samples at Willer *et al.* (2013), which will be used for partial validation in our analysis

At the 5×10^{-8} genome-wide significance level, the omnibus chi-square test OT identified 44 significant loci, and the PC-based test ET identified 22 significant loci. The adaptive test AT identified 40 significant loci, including the majority of significant loci identified by OT and ET. Figure 1a and b compare the number of significant loci and SNPs identified by the proposed joint tests.

Many of those identified significant SNPs have been found genome-wide significant in the followup study of Willer *et al.* (2013). Table 4 listed the total number of SNPs and loci identified by three joint association test methods.

Table 5 listed the test results for those identified novel loci that did not harbor any significant SNPs in both Teslovich *et al.* (2010) and Willer *et al.* (2013). Table 6 listed the significant SNPs identified only by one of the proposed tests (AT, OT and ET) and their minimum *P*-values across three traits in the Teslovich *et al.* (2010) and Willer *et al.* (2013) study (denoted as minP-2010 and minP-2013, respectively).

3.2.2 Analysis of GWAS results for glycemic traits

We also analyze the GWAS meta-analysis results for two glycemic traits: fasting glucose and indices of *β*-cell function (HOMA-B) based on 46 186 non-diabetic European samples conducted by the international

Table 4. Number of significant SNPs and loci identified for the lipids GWAS summary data

	OT	ET	AT	Total
SNPs	491 (74%)	66 (58%)	450 (75%)	543 (72%)
Loci	44 (84%)	22 (64%)	40 (90%)	57 (75%)

Note: listed within parentheses are the percentage of SNPs and loci that have been found genome-wide significant in the followup study of Willer *et al.* (2013).

Table 5. Novel significant loci identified by the proposed joint test methods

SNP	Chr	Gene	OT	ET	AT	minP-2010	minP-2013
rs6730449	2	DYNC2LI1	5.75 e-07	1.72 e-08	3.72 e-08	1.83 e-07	2.43 e-06
rs762861	4	HGFAC	1.52 e-07	2.74 e-08	4.72 e-08	9.15 e-07	5.04 e-07
rs7705104	5	PELO	2.38 e-08	9.07 e-01	5.12 e-08	1.72 e-02	—
rs10462958	5	TIMD4	7.02 e-07	2.99 e-08	6.42 e-08	2.03 e-07	—
rs499921	6	FRK	4.99 e-07	2.70 e-08	5.80 e-08	1.81 e-07	5.07 e-08
rs12667186	7	SP4	9.69 e-07	3.14 e-08	6.75 e-08	1.36 e-07	1.08 e-07
rs17148062	7	POT1	4.66 e-08	7.65 e-01	9.98 e-08	6.14 e-02	—
rs4455806	8	SOX17	1.29 e-06	4.60 e-08	9.85 e-08	1.84 e-07	5.09 e-08
rs10090964	8	SDCBP	5.78 e-07	2.98 e-08	6.40 e-08	5.09 e-08	7.37 e-08
rs4573621	10	GPAM	4.54 e-08	4.14 e-02	9.72 e-08	2.70 e-04	2.09 e-04
rs4238103	12	LARP4	4.08 e-09	9.05 e-02	4.08 e-09	3.83 e-04	—
rs9600211	13	KLF12	1.57 e-09	1.88 e-01	1.57 e-09	1.52 e-05	—
rs4419034	15	FRMD5	4.87 e-08	4.30 e-01	1.04 e-07	5.05 e-08	1.51 e-06
rs16959082	17	RCVRN	3.67 e-08	5.32 e-02	7.87 e-08	7.44 e-05	4.32 e-01

Note: Listed are the joint test *P*-values and the minimum marginal test *P*-values across all traits in Teslovich *et al.* (2010) and Willer *et al.* (2013) (denoted as minP-2010 and minP-2013 respectively).

Table 6. Some significant SNPs identified only by one of the proposed tests (AT, OT or ET)

	SNP	Chr	Gene	OT	ET	AT	minP-2010	minP-2013
AT	rs3095340	6	KIAA1949	5.37 e-08	1.75 e-07	4.61 e-08	1.27 e-07	4.23 e-07
	rs505870	6	SLC22A3	7.41 e-08	5.27 e-08	4.12 e-08	3.34 e-07	1.05 e-09
	rs10806731	6	LPA	6.16 e-08	1.13 e-07	4.75 e-08	6.40 e-07	—
	rs4808957	19	GATAD2A	5.05 e-08	2.05 e-07	4.52 e-08	9.58 e-07	1.83 e-14
OT	rs9938020	16	NFATC3	3.58 e-08	1.74 e-01	7.68 e-08	7.61 e-04	5.32 e-33
	rs1529929	16	PPP4R1L	3.27 e-08	3.33 e-01	7.02 e-08	2.09 e-03	7.20 e-20
	rs16962767	16	PPP4R1L	4.78 e-08	7.86 e-01	1.02 e-07	1.74 e-02	3.33 e-29
	rs1016563	18	PCBP3	3.99 e-08	4.11 e-05	8.55 e-08	7.09 e-08	2.68 e-10
ET	rs499921	6	ABLIM1	4.99 e-07	2.70 e-08	5.80 e-08	1.81 e-07	5.07 e-08
	rs12667186	7	TOP3B	9.69 e-07	3.14 e-08	6.75 e-08	1.36 e-07	1.08 e-07
	rs12676593	8	MYO1E	9.40 e-07	4.73 e-08	1.01 e-07	6.16 e-08	9.13 e-08
	rs11075910	16	THSD4	5.66 e-07	3.63 e-08	7.80 e-08	5.93 e-08	2.16 e-08

Note: Listed are the joint test *P*-values and the minimum marginal test *P*-values across all traits in Teslovich et al. (2010) and Willer et al. (2013) (denoted as minP-2010 and minP-2013, respectively).
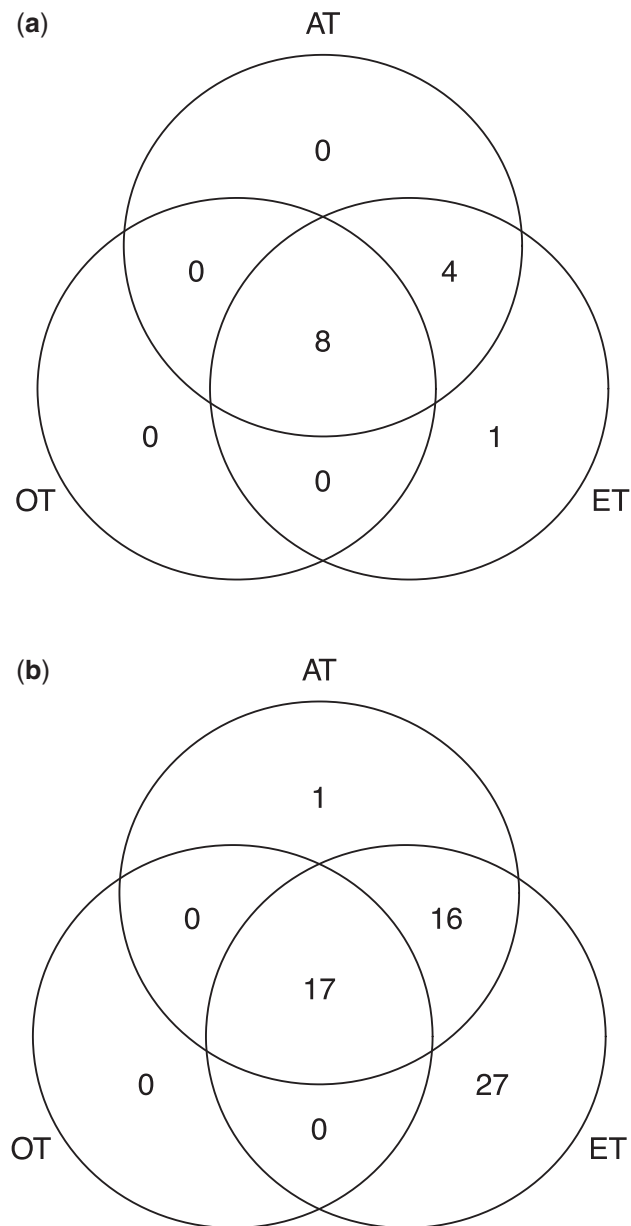
(a)

(b)

Fig. 2. Venn diagram of the total number of significant loci and SNPs identified by three multi-trait test methods based on glycemic GWAS summary data. **(a)** Number of detected significant loci. **(b)** Number of detected significant SNPs

MAGIC consortium (Dupuis *et al.*, 2010). Figure 2a and b shows the venn diagram for the total number of identified significant loci and SNPs by the proposed joint association test methods at the 5×10^{-8} genome-wide significance level. The chi-square test OT identified 8 significantly associated loci, and the PC test ET identified 13 significant loci. The adaptive association test AT identified all significant loci identified by OT and 4 additional significant locus identified by ET.

The comprehensive analysis results of the GWAS summary data for the lipids and glycemic traits are provided at the Supplementary Section S5.

4 Discussion

Many GWAS have been conducted in the past decade and successfully identified tens of thousands of variants related to various traits.

However there are still significant missing heritability for most traits and there remain many more genetic variants with small effects to be discovered. In the post-GWAS era, owing to many difficulties of sharing raw genotype and phenotype data, it is useful to develop statistical methods that can leverage the publicly available GWAS summary data to identify more novel genetic variants. In this article, we have focused on testing SNP association across multiple traits, which has been shown to have improved test power than individual trait-based association test. To properly control the false positives for multi-trait association test, we need to accurately estimate the across trait correlations. Our results show that the commonly used approach of using empirical correlation matrix of summary Z-scores should be avoided if possible, since it can produce highly biased estimates and lead to significantly inflated Type I errors for polygenic traits. The LD score regression-based approach could produce more accurate estimates, and has performed well in our numerical studies. In GWAS we typically need to test tens of millions of SNPs, and it is desirable to develop efficient statistical methods. All our proposed methods are scalable to genome-wide association test: we can quickly compute their analytical *P*-values without the need of resampling or permutation.

In this work, we make several contributions to studies of multi-trait association test using GWAS summary data. We propose a robust LD score regression to accurately estimate the trait correlation using only GWAS summary data. Although not a major focus of this article, the proposed robust LD score regression also provides a good approach to estimating genetic correlation using GWAS summary data. We develop powerful, efficient and robust adaptive multi-trait association test methods based on the summary statistics. The proposed methods are extremely scalable to genome-wide association test: we can quickly and accurately compute analytical *P*-values without the need of Monte Carlo approximation. All our proposed methods have been implemented in an R package publicly available online. The Supplementary Material contain the detailed analysis results for the lipids and glycemic GWAS summary data, and sample codes to install and use the developed R package.

In this article, we have mainly focused on those efficient and genome-wide scalable joint association test methods with analytically computed *P*-values and proper control of Type I errors, and have not studied those methods that often require computationally intensive Monte Carlo simulations (see, e.g. Kim *et al.*, 2015; Shim *et al.*, 2015; Stephens, 2013). As a partial remedy, we have included in the comparison the efficient tests, SZ, SZ2 and AZ, along the line of Kim *et al.* (2015).

In summary, our developed multi-trait association test methods can be used to further identify novel variants and the robust LD score regression approach can be used to study the genetic correlation of human traits. Both approaches need only the GWAS summary data and can help to dissect the genetic architecture of complex human traits.

References

Abecasis,G.R. *et al.* (2012) An integrated map of genetic variation from 1, 092 human genomes. *Nature*, **491**, 56–65.

Baselmans,B.M. *et al.* (2017) Multivariate Genome-wide and integrated transcriptome and epigenome-wide analyses of the Well-being spectrum. https://doi.org/10.1101/115915. 115915.

Broadaway,K.A. *et al.* (2016) A statistical approach for testing cross-phenotype effects of rare variants. *Am. J. Hum. Genet.*, **98**, 525–540.

Bulik-Sullivan,B. *et al.* (2015a) An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, **47**, 1236–1241.

Bulik-Sullivan,B.K. *et al.* (2015b) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.*, **47**, 291–295.

Dupuis,J. *et al.* (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.*, **42**, 105–116.

Evangelou,E. and Ioannidis,J.P.A. (2013) Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.*, **14**, 379–389.

Ferreira,M.A.R. and Purcell,S.M. (2009) A multivariate test of association. *Bioinformatics*, **25**, 132–133.

Galesloot,T.E. *et al.* (2014) A comparison of multivariate genome-wide association methods. *PLoS One*, **9**, e95923.

Hu,Y.J. *et al.* (2013) Meta-analysis of gene-level associations for rare variants based on single-variant statistics. *Am. J. Hum. Genet.*, **93**, 236–248.

Kim,J. *et al.* (2015) An adaptive association test for multiple phenotypes with GWAS summary statistics. *Genet. Epidemiol.*, **39**, 651–663

Lee,S. *et al.* (2014) Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.*, **95**, 5–23.

Lin,D.Y. and Zeng,D. (2010) Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet. Epidemiol.*, **34**, 60–66.

Maity,A. *et al.* (2012) Multivariate phenotype association analysis by marker-set Kernel machine regression. *Genet. Epidemiol.*, **36**, 686–695.

Manolio,T.A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.

O'Reilly,P.F. *et al.* (2012) MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS One*, **7**, e34861.

Pasaniuc,B. and Price,A.L. (2017) Dissecting the genetics of complex traits using summary association statistics. *Nat. Rev. Genet.*, **18**, 117–127.

Ray,D. and Boehnke,M. (2018) Methods for meta-analysis of multiple traits using gwas summary statistics. *Genet. Epidemiol.*, **42**, 134–145.

Seoane,J.A. *et al.* (2014) Canonical correlation analysis for gene-based pleiotropy discovery. *PLoS Comput. Biol.*, **10**, e1003876.

Shim,H. *et al.* (2015) A Multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. *PLoS One,* **10**, e0120758.

Solovieff,N. *et al.* (2013) Pleiotropy in complex traits: challenges and strategies. *Nat. Rev. Genet.*, **14**, 483–495.

Stephens,M. (2013) A unified framework for association analysis with multiple related phenotypes. *PLoS One*, **8**, e65245

Tang,C.S. and Ferreira,M.A.R. (2012) A gene-based test of association using canonical correlation analysis. *Bioinformatics*, **28**, 845–850.

Teslovich,T.M. *et al.* (2010) Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*, **466**, 707–713.

Turley,P. *et al.* (2018) Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.*, **50**, 229–237.

Visscher,P.M. *et al.* (2017) 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.*, **101**, 5–22.

Willer,C.J. *et al.* (2013) Discovery and refinement of loci associated with lipid levels. *Nat. Genet.*, **45**, 1274–1283.

Wu,B. and Pankow,J.S. (2016) Sequence kernel association test of multiple continuous phenotypes. *Genet. Epidemiol.*, **40**, 91–100.

Zhu,X. *et al.* (2015) Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Hum. Genet.*, **96**, 21–36.