Subject Section

Gene- and pathway-based association tests for multiple traits with GWAS summary statistics

Il-Youp Kwak 1, Wei Pan 1,*

1Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455.
*To whom correspondence should be addressed.

Abstract

Summary: To identify novel genetic variants associated with complex traits and to shed new insights on underlying biology, in addition to the most popular single SNP-single trait association analysis, it would be useful to explore multiple correlated (intermediate) traits at the gene- or pathway-level by mining existing single GWAS or meta-analyzed GWAS data. For this purpose, we present an adaptive gene-based test and a pathway-based test for association analysis of multiple traits with GWAS summary statistics. The proposed tests are adaptive at both the SNP- and trait-levels; that is, they account for possibly varying association patterns (e.g. signal sparsity levels) across SNPs and traits, thus maintaining high power across a wide range of situations. Furthermore, the proposed methods are general: they can be applied to mixed types of traits, and to Z-statistics or p-values as summary statistics obtained from either a single GWAS or a meta-analysis of multiple GWAS. Our numerical studies with simulated and real data demonstrated the promising performance of the proposed methods.

Availability: The methods are implemented in R package aSPU, freely and publicly available at: https://cran.r-project.org/web/packages/aSPU/.
Contact: weip@biostat.umn.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 Introduction

In spite of the success of genome-wide association studies (GWAS) in identifying thousands of reproducible associations between single nucleotide polymorphism (SNPs) and complex diseases/traits, in general the identified genetic variants can explain only a small portion of heritability (Manolio et al. 2009). A main reason is due to small effect sizes of genetic variants, raising both challenges and opportunities in developing more powerful analysis strategies. Among others, endeavors in the following three directions have been undertaken. First, due to polygenic effects (with small effect sizes) on complex traits, instead of the popular single SNP-single trait analysis, it may be more powerful to conduct gene- and pathway-level association tests (Lin and Tang, 2011; Wu et al. 2010; Pan et al. 2014; Li, et al. 2011; Guo et al. 2011; Li et al. 2012; Pan et al. 2015). However, most of the existing association tests are based on the use of individual-level genotypic and phenotypic data, while quite often only summary statistics for single SNPs are available. Thus, some association tests for a single trait but applicable to GWAS summary statistics have appeared, including GATES (Li et al. 2011), GATES-Simes (Gao et al. 2013), HYST (Li et al. 2012), and aSPUs and aSPUsPath (Kwak and Pan, 2016). Second, while many GWAS have collected multiple (intermediate) traits, due to pleiotropic effects, multiple correlated (intermediate) traits, e.g. neuroimaging endophenotypes (Shen et al. 2010, Zhang et al. 2014), can be used to boost power and illuminate on underlying biological mechanisms as compared to popular disease-based single trait analyses; see a review by Yang and Wang (2013). Most of the existing association tests for multiple traits are based on individual-level data (Basu et al. 2013; Tang and Ferreira 2012; Yang et al 2010; Zhang et al. 2014; Wang et al. 2015; Fan et al. 2015, 2016; Wang et al. 2016) with only few exceptions such as MGAS (Sluis et al. 2015) and metaCCA (Cichonska et al. 2016). Third, to increase the sample size, large consortia are being formed, aiming for meta analysis of multiple GWAS, for which often only summary statistics for single SNP-single trait associations, rather than individual-level genotypic and phenotypic data, are available. Hence it is necessary to develop methods that are applicable...
to only summary statistics. Motivated by the above three considerations, here we present such tests.

To our knowledge, there are only two existing tests that are for gene- or pathway-based analysis of multiple traits and applicable to summary statistics. MGAS (Shin et al. 2015) uses an extended Simes procedure and behaves like a univariate minimum p-value approach, while metaCCA (Cichon et al. 2016) is based on canonical correlation analysis (CCA) of multiple traits and multiple SNPs, which is related to MANOVA and the GEE-score test (Zhang et al. 2014; Kim et al. 2016), the latter two may lose power in some situations with multiple but relatively sparse and weak association signals between the traits and SNPs (Pan et al. 2014; Zhang et al. 2014).

Accordingly, it would be useful to extend adaptive tests for multiple trait-single SNP (Kim et al. 2015) or for single trait-multiple SNP associations (Kuwak and Pan, 2016) with summary statistics, or for multiple trait-multiple SNP associations with individual-level data (Kim et al. 2016), to the current case of multiple trait-multiple SNP associations with only GWAS summary statistics, which is the aim here.

In addition, we propose a novel Monte Carlo simulation method based on a matrix normal distribution to estimate the p-values for our proposed tests, which is well justified by known asymptotic theory that is suitable for large GWAS. In our proposed approach, we use a reference panel to estimate linkage disequilibrium (LD) among physically nearby SNPs; in contrast, metaCCA uses a similar method to estimate a joint covariance matrix for both the multiple traits and multiple SNPs, possibly explaining why it requires a large sample size of the reference panel to perform well, as to be confirmed in our later simulations. We also note that in MGAS, instead of individual-level genotypic data in a reference panel, p-values as summary statistics are used to empirically estimate LD among SNPs, which may lead to non-positive definite correlation matrices as numerically shown in Kuwak and Pan (2016).

Finally we note that our proposed methods are general with a wide range of applications. For example, the multiple traits can be mixed types: some may be quantitative while others binary; the summary statistics for single SNP-single trait associations, as either Z-statistics or p-values, can be obtained from either a single GWAS or a meta-analysis of multiple GWAS (with any valid test being applied). It is noteworthy to point out that the current version of metaCCA requires an equal sample size for all SNP-trait summary statistics in a real dataset to be confirmed in our later simulations. We also note that in MGAS, instead of individual-level genotypic data in a reference panel, p-values as summary statistics are used to empirically estimate LD among SNPs, which may lead to non-positive definite correlation matrices as numerically shown in Kuwak and Pan (2016).

2 Methods

2.1 Notation

Suppose there are $d$ SNPs (e.g., in a gene for gene-based testing) with additive genotype scores $g = (g_1, \ldots, g_d)'$, where $g_j$ is the number of minor alleles of the $j$th SNP; there are $n > 1$ quantitative or binary phenotypes $Y = (Y_1, \ldots, Y_m)'$; let $e = (e_1, \ldots, e_d)'$ denotes a set of covariates. We first consider one phenotype $Y_h$ by applying a generalized linear model:

$$g[Y_h] = \beta_0g + \sum_{j=1}^d g_j\beta_{hj} + \alpha_0e_h,$$

where $g[\cdot]$ is a canonical link function (i.e., the identity function for a quantitative trait, or a logit function for a binary trait). We are interested in testing $H_0^h: \beta_{hj} = 0$ for all $h = 1, \ldots, m$ and $j = 1, \ldots, d$.

For a given dataset $\{(Y_{ih}, E_{ih}, c_i): i = 1, \ldots, n\}$ with $n$ subjects, the score vector $U_h = (U_{h1}, \ldots, U_{hn})'$ for $\beta_h$ is

$$U_h = \sum_{i=1}^n (Y_{ih} - \hat{\mu}_{ih})E_{ih}:$$

where $\mu_{ih} = E(Y_{ih}|H_0) = g^{-1}(\hat{\beta}_{h0} + \alpha_0e_h)$ is the estimated mean of $Y_{ih}$ in the null model (under $H_0$).

Kim et al. (2016) constructed an adaptive test for multi-trait and multi-SNP association using the score vector. However, in the current context without individual-level data, we cannot directly calculate $U_{hj}$'s as given in the formula.

Here we assume that we only have summary statistics, say an $m \times d$ matrix of Z-scores, $Z$. Each element $Z_{ij}$ from the $i$th row and $j$th column of $Z$, represents a $Z$ score for testing association between the $i$th phenotype and the $j$th SNP. A $Z$ score is (asymptotically) a weighted version of an element in the score vector: $Z_{ij} = \hat{\beta}_{ij}/se(\hat{\beta}_{ij}) \approx U_{ij}/se(U_{ij})$; the approximation is based on the asymptotic equivalence between the Wald test and the Score test. Taking the $Z$ scores in place of the score vector has been proposed to test for multivariate single SNP associations (Kim et al. 2015) and single trait–multiple SNP associations (Kuwak and Pan, 2016).

2.2 Gene-based tests

We extend the gene-based tests based on individual-level data (Kim et al. 2016) to those based on summary statistics. Specifically, we define a test statistic for single trait-multiple SNP association and that for multiple trait–multiple SNP association as

$$\text{MTSPUsSet}(\gamma_1, Z_{h1}) = ||Z_{(h)}|| \gamma_1 = \left( \sum_{j=1}^d Z_{hj}^2 \right)^{1/2},$$

$$\text{MTSPUsSet}(\gamma_2, Z) = \sum_{h=1}^m (\text{MTSPUsSet}(\gamma_1, Z_{h1}))^{\gamma_2},$$

where $Z_{(h)}$ represents the $h$th row vector of matrix $Z$, i.e. the $Z$ scores for the $h$th trait. Two scalars $\gamma_1 \geq 1$ and $\gamma_2 \geq 1$ controls the extents of weighting on the SNPs and traits respectively. For example, a larger $\gamma_1$ (or $\gamma_2$) is expected to yield higher power if there are a smaller number of the SNPs (or traits) with truly non-zero associations (i.e. with the corresponding $\beta_{hj} \neq 0$). As discussed in more details in Kim et al. (2016), MTSPUsSet($1, 1$) is like a burden test (Shen et al. 2010), while MTSPUsSet($\gamma_1, \gamma_2$) for large values of $\gamma_1$ and $\gamma_2$ is effectively equivalent to a univariate minimum p-value test on all single SNP-single trait pairs; MTSPUsSet($2, 2$) is closely related to a variance-component score test in kernel machine regression (Maty et al. 2012) and nonparametric MANOVA or distance-based regression (McArule and Anderson 2001; Wessel and Schork 2006; Schaid 2005).

Since the optimal values of ($\gamma_1, \gamma_2$) are unknown, we propose an adaptive test to data-adaptively choose ($\gamma_1, \gamma_2$):

$$\text{MTAdSPUsSet}(Z) = \min_{\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2} p(\gamma_1, \gamma_2; Z),$$

where $p(\gamma_1, \gamma_2; Z)$ is the $p$-value for MTSPUsSet($\gamma_1, \gamma_2$; $Z$), and by default we use $\Gamma_1 = \{1, 2, 4, 8\}$ and $\Gamma_2 = \{1, 2, 4, 8\}$.

A main innovation here is to use a matrix normal distribution (Gupta and Nagar 1999; Zhou 2014) to obtain $p$-values based on the known asymptotic normal distribution of the $Z$ scores under $H_0$. Specifically, denote $Z_{(h)}$ as the $h$th row vector, and $Z_j$ as the $j$th column vector.
the Z scores for $i$th SNP of $Z$. If the sample size is large (with relatively small numbers of traits and SNPs), by the standard asymptotics for the Z scores, it is reasonable to assume that the null distribution of $Z$ is a matrix normal distribution:

$$Z \sim MN_m[d, \mathbf{0}_{m \times d}, \mathbf{P}, \mathbf{R}],$$

where $\mathbf{0}_{m \times d}$ is the $m \times d$ matrix with 0's. It is equivalent to saying that

$$\text{vec}(Z) \sim MN_m[d, \mathbf{0}_{m \times d}, \mathbf{R} \otimes \mathbf{P}],$$

where vec$(Z)$ is formed by stacking the columns of $Z \otimes$ is the Kronecker product, and $\mathbf{0}_{m \times d}$ is a 0 vector of length $m \times d$.

From equation (1), we see that $Z_1/\sqrt{m}$ follows a normal distribution with mean 0 and covariance matrix $\mathbf{P}$, and that $Z_1/\sqrt{m}$ follows a normal distribution with mean 0 and covariance matrix $\mathbf{P}$. Hence, we expect that

$$|Z_1/\sqrt{m}| \sim |T| \sim \chi_d^2$$

with degrees of freedom $d = \text{rank}(\mathbf{R} \otimes \mathbf{P})$. We can construct a score test (if $Z$ is obtained by the univariate score test or its asymptotically equivalent Wald test):

$$T_{\text{Wald}} = \text{vec}(Z)'(\mathbf{R} \otimes \mathbf{P})^{-1}\text{vec}(Z),$$

in which if $\mathbf{R} \otimes \mathbf{P}$ is not of full rank, a generalized inverse is used. Although $T_{\text{Wald}}$ has an asymptotic $\chi_d^2$ with degrees of freedom $d = \text{rank}(\mathbf{R} \otimes \mathbf{P})$, it may not work well for a high-dimensional $\mathbf{R} \otimes \mathbf{P}$, thus, as for MTSPUSet, we will use a single layer of Monte Carlo simulations to calculate its $p$-value.

As discussed in Zhou et al. (2014) and Kim et al. (2016), the score test is similar to CCA and MANOVA, hence we expect that $T_{\text{Wald}}$ will perform similarly to metaCCA, as to be confirmed.

2.3 Pathway-based tests

We extend the pathway-based multi-trait association tests of Kim et al. (2016) to the case with only GWAS summary statistics. Given a pathway $S$ with $|S|$ genes, we partition the Z score matrix as $Z = [Z_{ij(1)}, \ldots, Z_{ij(n)}]'$ with $Z_{ij(r)}$ as the rth row vector (i.e., Z scores for the rth trait), $Z_{ij(r)}$ is further partitioned at the gene level to $Z_{ij(r)} = [Z_{(i)(1)}, Z_{(i)(2)}, \ldots, Z_{(i)(|G|)}]'$, and at the SNP level to $Z_{ij(r)} = [Z_{(i)(1)}Z_{(i)(2)} \cdots Z_{(i)(d)}]'$ (for the $d_i$ SNPs in gene $i$).

We define the gene- and pathway-based tests for a single trait and then for multiple traits as

$$\text{SPU}(1, \gamma_1; Z_{(i)}) = ||Z_{(i)}||_1 = \left(\sum_{i=1}^{d_i} Z_{(i)(i)}^2/d_i\right)^{1/2},$$

$$\text{SPUPath}(1, \gamma_1; Z_{(i)}, S) = \sum_{i \in S} \text{SPU}(1, \gamma_1; Z_{(i)})^{1/2}/|S|^{1/2},$$

$$\text{MTSPUPath}(1, \gamma_1; Z, S) = \sum_{i \in S} \text{MTSPU}(1, \gamma_1; Z_{(i)}, S)^{1/2},$$

where the three integers $\gamma_1 \geq 1, \gamma_2 \geq 1$ and $\gamma_3 \geq 1$ are used to adaptively weight the SNPs, genes and traits respectively. For example, a larger $\gamma_1$ (or $\gamma_2$, or $\gamma_3$) is more effective when there are a smaller number of truly associated SNPs (or genes, or traits).

To adaptively choose $\gamma_1, \gamma_2, \gamma_3$, we propose a pathway-based adaptive test as

$$\text{MTSPUPath}(\gamma_1, \gamma_2, \gamma_3; Z, S) = \min_{\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_3 \in \Gamma_3} p_{\text{MTSPU}(1, \gamma_1, \gamma_2, \gamma_3; Z, S)},$$

where $p_{\text{MTSPU}(1, \gamma_1, \gamma_2; Z, S)}$ is the p-value of MTSPUPath($\gamma_1, \gamma_2; Z, S$), and by default we use $\Gamma_1 = \{1, 2, 4, 8\}, \Gamma_2 = \{1, 2, 4, 8\}$ and $\Gamma_3 = \{1, 2, 4, 8\}.

2.4 P-value calculations

Monte Carlo simulations are used to obtain the p-values for MTSPUSet or MTSPUPath in a single layer of simulations. Briefly, after estimating $\mathbf{P}$ and $\mathbf{R}$, first we simulate null scores $Z^{(b)} \sim MN_m[d, \mathbf{0}_{m \times d}, \mathbf{R}, \mathbf{P}]$ for $b = 1, \ldots , B$. Then we use the null scores to calculate the null test statistics, from which the p-values can be calculated (Kwak and Pan 2016). A larger $B$ is needed to estimate a smaller p-value.

We generate a matrix normal variable $Z^{(b)}$ in the following way (Zhou 2014). We first generate an $n \times d$ matrix $\mathbf{L}$ with each element independently from a standard univariate normal distribution with mean 0 and variance 1; that is, $\mathbf{L} \sim MN_m(d, \mathbf{0}_{d \times d}, \mathbf{I}_d, I_d)$. Then we obtain $Z^{(b)} = \mathbf{D}\mathbf{L}\mathbf{E}'$, where $\mathbf{D}$ and $\mathbf{E}$ are Cholesky decompositions of $\mathbf{P}$ and $\mathbf{R}$ with $\mathbf{P} = \mathbf{DD}'$ and $\mathbf{R} = \mathbf{EE}'$.

Specifically, for MTSPUSet, we consider the $b$th pathway $S$ with $|S|$ genes, we partition the Z score matrix as $Z = \{Z_{ij}^{(1)}, \ldots , Z_{ij}^{(S)}\}'$ with $Z_{ij}^{(s)}$ as the rth row vector (i.e., Z scores for the rth trait), $Z_{ij}^{(s)}$ is further partitioned at the gene level to $Z_{ij}^{(s)} = \{Z_{(i)(1)}, Z_{(i)(2)}, \ldots , Z_{(i)(|G|)}\}'$, and at the SNP level to $Z_{ij}^{(s)} = \{Z_{(i)(1)}Z_{(i)(2)} \cdots Z_{(i)(d)}\}'$ (for the $d_i$ SNPs in gene $i$).
• Step 5: Finally the $p$-value for the MTaSPUsSet test, $\text{MTaSPUsSet}_W$, is given by:

$$\text{MTaSPUsSet}_W = \sum_{k=1}^{B} \mathbf{1}(\frac{R}{\text{MTaSPUsSet}(Z)^{W}}) \leq \text{MTaSPUsSet}(Z) + \mathbf{1}(B + 1).$$

A similar procedure is used to obtain the $p$-values for MTaSPUsPath and MTaSPUsPath. When only $p$-values for single SNP-single trait associations, instead of $Z$ statistics, are available as summary statistics, we use $|Z| = \Phi^{-1}(1 - F/2)$, where $\Phi$ is the cumulative distribution function of the standard univariate normal distribution; we replace all $Z$'s with $|Z|$'s to calculate the test statistics.

3 Results
3.1 Simulations 1: choice of the reference panel
To demonstrate the validity and performance of our proposed methods, we designed a "control-control" experiment using the Welcome Trust Case Control Consortium (WTCCC) GWAS genotypic data (Consortium 2007; Kwak and Pan 2016). The WTCCC GWAS dataset contains about 3,000 controls with a total of 500,568 SNPs. Following the WTCCC's quality control (QC) recommendations, we removed subjects and SNPs that did not pass the QC criteria, resulting in 469,612 SNPs in 2,938 control subjects. We further pruned SNPs with MAF < 5% since we had only 379 samples in our reference panel to infer the LD structure for a set of SNPs. We considered 4,572 unique genes in 186 KEGG pathways to check type I error rates of all the tests. A total of 64,557 SNPs were mapped to these genes.

We simulated multiple traits using a multivariate normal distribution with mean 0 and correlation matrix in Equation (3) of Figure S1, which was estimated based on the GIANT data for women. We generated a set of six traits for each of the 2938 control subjects. Then we calculated the univariate $Z$ scores for all 64,557-6 SNP-trait pairs. A Monte Carlo simulation size of $B = 10^5$ was used to calculate the $p$-values. The results were based on 1000 simulation replicates.

For each gene (or pathway), $R$ was estimated from the 1000 Genome Project CEU samples. To estimate $P$, we excluded the SNPs with p-values < 0.05 and used the remaining 48,669 SNPs. Equation (1) of Figure S1 is the estimate for $P$. This estimate is close to the true value shown in Equation (3) of Figure S1, $P_w$. We pruned SNPs in high LD by removing any SNP if it was correlated with another SNP with an absolute value of Pearson’s correlation coefficient larger than 0.95 to avoid a nearly singular matrix $R$, causing numerical problems.

3.1.1 Gene-based tests
We first investigated the effects of the choice of the reference panel on estimating LD among SNPs, i.e., $R$ for each gene. We considered three scenarios: 1) using the whole 2938 WTCCC controls as the reference panel as an ideal case; 2) using only a random set of 100 WTCCC control samples as the reference panel to see whether a sample size as low as 100, close to that of many published reference panels, was sufficient to obtain accurate estimates; 3) using the 1000 Genomes Project CEU samples with 379 individuals as the reference panel, a more realistic scenario without individual level data.

Figure S2 shows the QQ plots of the $p$-values of the MTaSPUsSet test based on each of the three ways to estimate the SNP correlation matrix. We can see that all three plots looked reasonable with the estimated inflation factor $\lambda$ as 1.01, 0.99 and 0.99 respectively, all close to 1. It was confirmed that the type I error rates seemed to be well controlled in all cases.

Next we further compared the results as shown in Figure 1. By comparing the results between using the WTCCC whole control samples and using only 100 samples as reference panel, we conclude that taking only 100 samples from the original whole dataset seemed to perform well; the Pearson correlation ($r$) between the two was 0.99. The top right and bottom left panels compare the results between using the WTCCC whole data, WTCCC 100 samples and 1000 Genome Project CEU samples as the reference panel; again they showed high degrees of mutual agreement with a Pearson correlation coefficient as high as 0.97 and 0.98 respectively. In the bottom right panel, we further compared the results of MTaSPUsPath with only summary statistics (using the 1000 Genome Project CEU samples as the reference panel) to a similar GEE-based adaptive test with individual-level data (Kim et al. 2016). Although the agreement was reasonably high with a Pearson correlation coefficient of 0.9, there were some differences, indicating that caution is needed when using summary statistics.

We also tried metaCCA (Cichonska et al. 2016) and $T_{Xo}$ on the simulated data, and found that both might not work well when the sample size of the reference panel was small. We used 1) the whole 2938 WTCCC controls as an ideal case; 2) 100-2000 samples from the WTCCC control data; 3) using the 1000 Genome Project CEU samples, respectively, as the reference panel. We used “metaCcaGp” function in the R version of metaCCA at: https://bioconductor.org/packages/devel/bioc/html/metaCCA.html. Figures S3 and S4 show the QQ plots for each scenario. In particular, it showed that even a sample size of 500 drawn from the WTCCC control data or of 379 for the 1000 Genome Project CEU samples might not be large enough, because of this reason, we would not apply the tests to the real data.

Importantly, it was confirmed that metaCCA and $T_{Xo}$ gave almost the same $p$-values, as shown in Figure S5.

3.1.2 Pathway-based tests
For evaluations, we designed a control-control experiment using the WTCCC CD data. We randomly chose 3 to 15 genes from the WTCCC data to form a pathway. We applied the MTaSPUsPath test to each of 319 pathways. Simulations were conducted with different reference panels used to estimate $R$, similar to what was done for gene-based testing.

Figure S6 compares the results of MTaSPUsPath with various reference panels, and of a similar pathway-based adaptive test called GEE-aSPUs.
based on individual-level data (Kim et al. 2016). Similar conclusions to those for the gene-based MTaSPUsSet test can be drawn.

3.2 Simulations 2: power

The second simulation study was designed to compare the power of MTaSPUsSet with that based on single SNP-multiple trait and multiple SNP-single trait methods under realistic scenarios. To mimic real data, the genotype data, $\mathbf{G}_{i,n,d}$, were taken from the 1000 Genome Project CEU samples for three representative genes: TMEM110, ATP2A1 and RPS10-NUDT3 with their true effect sizes as the three $\mathbf{Z}$ score matrices, $\mathbf{Z}$. Let $\mathbf{Z}_i$, $\mathbf{Z}_g$ and $\mathbf{Z}_a$, for the three genes in the GIANT data, representing various association patterns (Figures 2 and S15). The covariance structure among the multiple traits, $\mathbf{V}$, was the correlation matrix in Equation (3) of Figure S1 for women in the GIANT data. We simulated the traits for each gene as

$$y_i \sim MN(\mathbf{aZ}_i, \mathbf{V}), \quad l = 1, 2, 3, $$

where $\mathbf{a}$ was a constant to be varied to control the effect size, and $\mathbf{g}_i$ was the genotype data for subject $i$ (i.e. ith row vector of $\mathbf{G}_{i,n,d}$ for $i = 1, 2, \ldots, 379$). The numbers of the SNPs were 8, 6 and 19 for TMEM110, ATP2A1 and RPS10-NUDT3 respectively.

To compared our multiple SNP-multiple trait method with single SNP-multiple trait and multiple SNP-single trait methods, we applied a gene-based aSPUs test (Kwak and Pan, 2016, multiple SNP-single trait method) for each trait, then combined the results across the multiple traits with a significance threshold of $0.05/6 = 0.0083$ based on the Bonferroni correction; we also considered two single SNP-multiple trait association tests proposed by Zhu et al. (2015) ($S_{het}$) and Kim et al. (2015) (MTaSPUs). The two tests were applied to each SNP and then we used a significance threshold of 0.05/$d$ based on the Bonferroni correction, where $d$ is the number of the SNPs in a gene.

Figure S16 shows empirical type I error rates and power curves based on 1000 replicates in each case. For genes TMEM110 and ATP2A1, there were large association signals for single trait height and one SNP (SNP 6) respectively, several other methods also performed as well as MTaSPUsSet; however, for gene RPS10-NUDT3, there were only weak and multiple association across multiple SNPs and multiple traits, MTaSPUsSet was the clear winner with the highest power. In summary, due to its adaptive combination of information across SNPs and traits, our proposed MTaSPUsSet test could maintain high power across all the scenarios.

3.3 Analysis of the GIANT data

We applied the MTaSPUsSet test to the summary statistics for sex stratified anthropometric data from The Genetic Investigation of Anthropometric Traits (GIANT) consortium (Randall et al. 2013). The data contain the $p$-values of univariate testing on single SNP-single trait associations on 2.7 million SNPs with each of six anthropometric traits that are well established to represent the body size and shape: height, weight, BMI, waist circumference (WC), hip circumference (HIP), and waist-hip circumference ratio (WHR).

The original study was based on a single SNP-single trait association analysis (Randall et al. 2013). Instead, we applied gene-based and other association tests on the six traits (height, weight, BMI, WC, HIP and WHR) for men and for women separately. Since all study participants were of European ancestry, we used the 1000 Genome Project CEU samples as the reference panel for both methods.

First, for MTaSPUsSet, in total 2,722,976 SNPs were mapped to 17,562 genes (plus 2-kb upstream and 2-kb downstream regions for each gene). We set the genome-wide significance threshold at $0.05/(17562 \times 6) = 2.85 \times 10^{-5}$ based on the Bonferroni correction. We pruned SNPs in high LD by removing any SNP if it was correlated with another SNP with an absolute value of Pearson’s correlation coefficient larger than 0.95. For each gene, the correlations among the SNPs, $\mathbf{R}$, were estimated from the 1000 Genome Project CEU samples. The correlations among the six traits were estimated based on 1,454,615 null SNPs with non-significant $Z$ scores for men and women respectively as shown in Figure S1.

A stage-wise simulation strategy was used to calculate the $p$-values for each gene. We started with the simulation number $B = 10^6$; we sequentially increased $B$ to $10^7$, then $10^8$ and finally $10^9$ if a gene’s $p$-value was less than 0.003, 0.0003 and 0.00003 respectively.

The MTaSPUsSet test identified a total of 137 genes to be genome-wide significant for men and women: 81 for men, 125 for women and 69 for both.

3.3.1 Comparison with single SNP-single trait analysis

As a comparison, for single SNP-single trait analysis, we used a genome-wide significance threshold of $5 \times 10^{-8}$ based on a Bonferroni adjustment for six traits, yielding in total 1298 significant SNPs (with 623 SNPs mapped to 62 genes) for men, and 2072 significant SNPs (with 990 SNPs mapped to 97 genes) for women. Although there were many common genes (i.e. 53 and 85 for men and women) identified by both methods, the proposed MTaSPUsSet test identified more genes (Table S1). In particular, to demonstrate the sex differences of genetic effects, the new test pinpointed 12 and 36 significant genes uniquely and specifically for men and women respectively; in contrast, the popular and standard single SNP-single trait analysis identified 20 and 55 genes uniquely for men and women respectively. The smaller number of men-specific genes identified by the new test could be due to its higher power: it is reasonable to assume that some of the identified sex-specific genes are false positives due to inadequate power for either sex, though further validations are needed.

3.3.2 Comparison with single SNP-multiple trait analysis

We applied two other single SNP-multiple trait association tests proposed by Zhu et al. (2015) ($S_{het}$) and Kim et al. (2015) (MTaSPUs) to compare with MTaSPUsSet, a multiple SNP-multiple trait testing method.

The usual genome-wide significance threshold of $5 \times 10^{-8}$ was used in analysis. $S_{het}$ identified 2038 significant SNPs (with 1073 SNPs mapped to 137 genes) for men, and 2700 SNPs (with 1265 SNPs mapped to 133 genes) for women. MTaSPUs detected 2205 SNPs (with 1198 SNPs mapped to 153 genes) for men, and 2708 SNPs (with 1342 SNPs mapped to 143 genes) for women.

For a comparison, we focused on the genes that MTaSPUsSet uniquely and specifically detected for men and women: 4 genes for men and 8 genes for women. Figures S10 and S11 show the log-transformed $p$-values of univariate testing on single SNP-single trait associations for these genes. We can see that multiple SNPs were associated with one or more traits, for which MTaSPUsSet gained power by aggregating association signals across multiple SNPs while a single SNP-multiple trait test was unable to do so; see genes TMEM110 and YTA1 in Figure 2 as examples (also shown in Figures S10 and S11 respectively).

3.3.3 Comparison with gene-based multiple SNPs-single trait analysis

Next, we conducted a multiple SNP-single trait analysis. We applied a gene-based aSPUs test (Kwak and Pan, 2016) for each trait, then combining the results across the multiple traits with a genome-wide significance threshold of $0.05/(17562 \times 6) = 4.75 \times 10^{-5}$ based on the Bonferroni correction. The aSPUs test identified 81 significant genes for men and 111 genes for women (with 54 genes common for both sexes).

Again we focused on the genes uniquely and specifically identified by MTaSPUsSet for men and women respectively: 9 and 17 genes respectively. Figure S12 and S13 shows the log-transformed $p$-values of...
univariate testing on single SNP–single trait associations for these genes. This time we can see some strong association signals across multiple traits for some single SNPs, in which case MTaSPUSet could detect an overall association by aggregating information across the multiple traits, while multiple SNP–single trait tests might fail. As examples see genes ATP2A1 and RFWD2 in Figure 2 (as also shown in Figures S12 and Figure S13).

### 3.3.4 Comparison with another gene-based multiple trait test

We applied MGAS of Shiu et al. (2015) using "kgg" software. The same 2-kb upstream and 2-kb downstream regions were used in mapping the SNPs to each gene, and the same estimated trait correlation matrices were used. However, for unknown reasons, only in total 969,832 SNPs were mapped to 6,424 genes, compared to ours of mapping 2,722,976 SNPs to 17,562 genes. Accordingly, the genome-wide significance threshold was set at 0.05/6424 = 7.78 × 10^{-6} based on the Bonferroni correction. In total only 19 genes were identified by MGAS to be significant: 16 genes for women and 8 for men.

For a fair comparison between MTaSPUSet and MGAS, we examined more closely the 17,562 and 6,424 mapped genes for each method. There were 5197 shared genes commonly mapped by both methods; many of the 6,424 “kgg” genes starting with “LOC” and “LINC” were not in the MTaSPUSet set of the 17,562 genes. We decided to apply both methods to the common set of the 5197 genes. The genome-wide significance level was set at 0.05/5197 by the Bonferroni adjustment.

Figure 3 shows the Manhattan plots for men and women based on MGAS and MTaSPUSet respectively. Although there were some shared and general patterns between the results of the two methods, MTaSPUSet identified a larger number of significant genes: a total of 49 genes with 27 and 39 for men and women respectively. In contrast, MGAS identified only a total of 17 genes with 7 and 14 for men and women respectively. It might suggest that MTaSPUSet was more powerful, though further validations are needed.

To further contrast the differences between the two tests, Table S2 lists the 17 significant genes identified by MGAS with the corresponding p-values from the two tests. Genes LCO1L, VTA1, BICD2, RASS1, GNA12, NCOA1, TNS1, EPN12, DNM3 and RFWD2 were significant for women by both MGAS and MTaSPUSet, and LCO1L, RASS1 and NCOA1 were significant for men by both tests, while LCO1L and RASS1 were significant for both men and women by both tests. Gene LCO1L was known to be associated with anthropometric traits, including body height in African Americans (Carty et al. 2012), birth weight and adult height (Horiuchi et al. 2013); it is also a candidate gene for body weight in sheep (Al-Maman et al. 2015) and body size in horse (Metzger et al. 2013).

Figure 4 shows the p-values of the univariate test on single trait–single SNP associations for some genes identified by MTaSPUSet, along with the \( \gamma_{1, 2} \) values for the most significant MTSPUSet(\( \gamma_{1, 2} \)) test (i.e. with the smallest p-value) for each gene. Note that, due to the use of a finite number of Monte Carlo simulations, multiple sets of \( \gamma_{1, 2} \) values might give equal (and smallest) p-values.

It can be seen that for genes RPGRIP1L and RPS10-NUDT3, since there were many moderately significant univariate p-values (for univariate trait-SNP associations) with a dense association pattern, small values \( \gamma_{1, 2} = (1, 2) \) or \( (2, 1) \) gave the most significant results. In contrast, for gene DNM3 with a larger number of SNPs, the association pattern was more sparse with main associations between some SNPs and trait height, larger values of \( \gamma_{1, 2} = (4, 8) \) or \( (8, 8) \) gave the most significant result. On the other hand, for gene ZCCHC2, due to the two or three highly significant univariate p-values between one or two SNPs and two traits, weight and BMI, any value of \( \gamma_{1, 2} \) would detect the overall association.

### 4 Discussion

We have presented new gene- and pathway-based adaptive association tests for multiple traits using only GWAS summary statistics. Our control- simulation experiments using the WTCCC genotype data with simulated multiple traits demonstrated that the type I error rates were well controlled. For the estimation of LD among SNPs (i.e. correlation matrix \( R \)), the choice of a reference panel (with individual-level genotypic data) would be a key for the performance. In the WTCCC control-control experiments, we compared three reference panels based on either the whole or a small subset of the original WTCCC control data, and the 1000 Genome Project CEU samples (with 379 subjects). The p-values calculated from the three reference panels were in general similar, but not exactly the same; the Pearson correlation coefficient of the log(p-values) between any two reference panels was at least 0.97, confirming that either the 1000 Genome Project CEU samples or a small subset of the control samples from the original population were sufficient for the WTCCC subject population.

We applied our gene-based MTaSPUSet test to the meta-analyzed GIANT data. Since the participants in the GIANT data were of European and European American descent, the use of the 1000 Genome Project CEU panel was expected to be reasonable. The MTaSPUSet test identified a total of 137 significant genes: 81 for men, 125 for women and 69 for both. As a comparison, for single SNP–single trait analysis identified 117 genes: 62 for men, 97 for women and 42 for both. MTaSPUSet identified more genes. For more comparison, we also applied MGAS (Shiu et al. 2015) using the same reference panel, identifying only 19 significant genes using
“kgg” software with a smaller set of the genes being mapped. For a fair comparison, we applied both MTaSPuSsSet and MGAS to a common set of 5197 genes. MTaSPuSsSet identified 27 and 39 significant genes for men and women respectively, compared to only 7 and 14 genes by MGAS, suggesting possible power gains by MTaSPuSsSet. We also note that the other method metaCCA could not be applied to the GIANT data because it required a common sample size for all SNP-trait pairs, while the sample size for some SNPs ranged from around 200 to about 70,000 across the traits.

**Acknowledgment**

The authors are grateful to the reviewers for constructive comments.

**Funding**

This work was supported by National Institutes of Health [R01/GM113329, R01/HL105397, R01/HL116720], and the Minnesota Supercomputing Institute.

**References**


Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678.


