Genetics and population analysis

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

Il-Youp Kwak and Wei Pan*

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

*To whom correspondence should be addressed.

Associate Editor: Oliver Stegle

Received on April 29, 2016; revised on August 8, 2016; accepted on August 29, 2016

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 and Implementation: 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/trait, in general the identified genetic variants can explain only a small proportion 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 (Gui et al., 2011; Li et al., 2011, 2012; Lin and Tang, 2011; Pan et al., 2014, 2015; Wu et al., 2010). 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 (Gui et al., 2011), HYST (Li et al., 2012), MAGMA(de Leeuw et al., 2015), VEGAS (Liu et al., 2010) 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; Fan et al., 2015, 2016; Tang and Ferreira, 2012; Wang et al., 2015; Wang et al., 2016; Yang et al., 2010; Zhang et al., 2014) with only few exceptions such as MGAS (Van der 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 (Van der Sluis et al., 2015) uses an extended Simes procedure and behaves like a univariate minimum P value approach, while metaCCA (Cichonska 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 (Kim correlation analysis (CCA) of multiple traits and multiple SNPs, which with multiple but relatively sparse and weak association signals be- trait–single SNP (Kim 1

Methods

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

\[ SPUs(\gamma; Z; h) = \left| \left| Z_{(h)} \right| \right| = \left( \sum_{i=1}^{d} Z_{(h)}^{i} \right)^{1/2}. \]

\[ MTSPUsSet(\gamma_1; \gamma_2; Z) = \sum_{h=1}^{m} (SPUs(\gamma_1; Z_{(h)}))^{2/2}. \]

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 \( \gamma_2 \geq 1 \) control the ex- tents 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 \( \beta_{si} \neq 0 \). As discussed in more detail 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 regres- sion (Maitry et al., 2012) and non-parametric MANOVA or distance-based regression (McArdle and Anderson, 2001; Schaid, 2005; Wessel and Schork, 2006).

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

\[ SPUs(\gamma; Z; h) = \left| \left| Z_{(h)} \right| \right| = \left( \sum_{i=1}^{d} Z_{(h)}^{i} \right)^{1/2}. \]

\[ MTSPUsSet(\gamma_1; \gamma_2; Z) = \sum_{h=1}^{m} (SPUs(\gamma_1; Z_{(h)}))^{2/2}. \]

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 \( \gamma_2 \geq 1 \) control the ex- tents 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 \( \beta_{si} \neq 0 \). As discussed in more detail 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 regres- sion (Maitry et al., 2012) and non-parametric MANOVA or distance-based regression (McArdle and Anderson, 2001; Schaid, 2005; Wessel and Schork, 2006).
Since the optimal values of \((\gamma_1, \gamma_2)\) are unknown, we propose an adaptive test to data-adaptively choose \((\gamma_{1*, \gamma_2*})\):
\[
\text{MTaSPUsSet}(Z) = \min_{\gamma_{1*} \in \{1, 2, 4, 8\}} P_{(1*, \gamma_2*)}(Z),
\]
where \(P_{(1*, \gamma_2*)}\) 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 \(Z_{ij}\) as the \(i\)th row vector, and \(Z_i\) as the \(i\)th column vector (i.e. the \(Z\) scores for \(j\)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:
\[
Z \sim \text{MN}_{m \times d}(0_{m \times d}, P, R),
\]
where \(0_{m \times d}\) is the \(m \times d\) matrix with \(0\)s. It is equivalent to saying that
\[
\text{vec}(Z) \sim \text{MN}_{md}(0_{md}, R \otimes P), \quad (1)
\]
where \(\text{vec}(Z)\) is formed by stacking the columns of \(Z\), \(\otimes\) is the Kronecker product and \(0_{md}\) is a \(0\) vector of length \(md\).

From Equation (1), we can see that \(Z_{ij}/\sqrt{R_{jj}}\) follows a normal distribution with mean 0 and covariance matrix \(R\), and that \(Z_{ij}/\sqrt{R_{jj}}\) follows a normal distribution with mean 0 and covariance matrix \(R\) (Zhou, 2014). Since \(P\) and \(R\) are correlation matrices with \(R_{jj} = P_{jj} = 1\), we obtain
\[
Z_{ij} \sim N_{md}(0_{md}, P), \quad \text{and} \quad Z_{ij} \sim N_{mj}(0_{mj}, R).
\]

Following Kim et al. (2015), we propose excluding the SNPs with small \(P\) values (e.g. \(< 0.05\)) and using a large subset of the remaining null SNPs to estimate \(P\) with the sample correlation matrix of the \(Z\) scores. For \(R\), as shown by Kwak and Pan (2016) and others, it can be approximated by the sample correlation matrix of the \(Z\) scores using a reference panel similar to the study population. For example, we used 1000G Phase I version 3 Shapeit2 Reference data downloaded from the KGG software website (Li et al., 2012); it contains about 81.2 million polymorphic markers on 2504 samples released in September 2014. By default, we used 379 CEU (Utah Residents with Northern and Western Ancestry) samples.

Finally we note that, based on the asymptotic null distribution of \(\text{vec}(Z)\) in (1), we can construct a score test (if \(Z\) is obtained by the univariate score test or its asymptotically equivalent Wald test):
\[
T_{\text{scg}} = \text{vec}(Z)'(R \otimes P)^{-1}\text{vec}(Z),
\]
in which if \(R \otimes P\) is not of full rank, a generalized inverse is used. Although \(T_{\text{scg}}\) has an asymptotic \(Z^2\) with degrees of freedom \(d = \text{rank}(R \otimes P)\), it may not work well for a high-dimensional \(R \otimes P\), thus, for MTaSPUsSet, we will use a single layer of Monte Carlo simulations to calculate its \(P\) value.

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

2.3 Pathway-based tests
We extend the pathway-based multitrait 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 \(Z = (Z_{11}, \ldots, Z_{m1})'\) with \(Z_{ij}\) being the \(ith\) partition the \(Z\) score matrix as \(Z = (Z_{11}, \ldots, Z_{ij})\) is further partitioned at the gene level to \(Z_{ij} = (Z_{ij1}, \ldots, Z_{ijd})\), and at the SNP level to \(Z_{ijg} = (Z_{ij1g}, \ldots, Z_{ijdg})\) for the \(d_g\) SNPs in gene \(g\).

We define the gene- and pathway-based tests for a single trait and then for multiple traits as
\[
\text{SPUs}((\gamma_1, \gamma_2); S) = \frac{(\sum_{g=1}^{d} P_{(\gamma_1, \gamma_2)}(Z_{ijg}))^{1/2}}{\sum_{g=1}^{d} P_{(\gamma_1, \gamma_2)}(Z_{ijg})},
\]
\[
\text{MTSPUsPath}((\gamma_1, \gamma_2); \gamma_3; S) = \sum_{g=1}^{d} \text{SPUs}((\gamma_1, \gamma_2); Z_{ijg}),
\]
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
\[
\text{MTaSPUsPath}(S, Z) = \min_{\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2, \gamma_3 \in \Gamma_3} P_{(\gamma_1, \gamma_2, \gamma_3)}(Z, S),
\]
where \(P_{(\gamma_1, \gamma_2, \gamma_3)}(Z, S)\) is the \(P\) value of MTSPUsPath(\(\gamma_1, \gamma_2, \gamma_3; Z, S\)), 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 MTaSPUsSet or MTaSPUsPath in a single layer of simulations. Briefly, after estimating \(P\) and \(R\), first we simulate null scores \(Z^{(b)} \sim \text{MN}_{md}(0_{md}, 0_{md}, P, R)\) 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 variate \(Z^{(b)}\) in the following way (Zhou, 2014). We first generate an \(n \times d\) matrix \(L\) with each element independently from a standard univariate normal distribution with mean 0 and variance 1; that is \(L \sim \text{MN}_{md}(0_{md}, I_{md}, I_{md})\). Then we obtain \(Z^{(b)} = \text{DLE}'\), where \(D\) and \(E\) are Cholesky decompositions of \(P\) and \(R\) with \(P = D D'\) and \(E = E E'\).

Specifically, for MTaSPUsSet,

- \(\text{Step 1.}\) Generate independent \(Z^{(b)} \sim \text{MN}_{md}(0_{md}, P, R)\) for \(b = 1, \ldots, B\).
- \(\text{Step 2.}\) Calculate the null test statistics MTSPUsSet(\(\gamma_1, \gamma_2, Z^{(b)}\)).
- \(\text{Step 3.}\) The \(P\) value for MTSPUsSet(\(\gamma_1, \gamma_2; Z\)) is
\[
P_{(\gamma_1, \gamma_2)} = \frac{\sum_{b=1}^{B} I(\text{MTSPUsSet}(\gamma_1, \gamma_2; Z^{(b)}))}{B},
\]
and that for MTSPUsSet(\(\gamma_1, \gamma_2; Z^{(b)}\)) is
\[
P_{(\gamma_1, \gamma_2)} = \frac{\sum_{b=1}^{B} I(\text{MTSPUsSet}(\gamma_1, \gamma_2; Z^{(b)}))}{B},
\]
- \(\text{Step 4.}\) Calculate the null and observed test statistics MTaSPUsSet(\(Z^{(b)}\)) = \min_{\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2} P_{(\gamma_1, \gamma_2)}(Z^{(b)}),
MTaSPUsSet(\(Z\)) = \min_{\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2} P_{(\gamma_1, \gamma_2)}(Z);
- \(\text{Step 5.}\) Finally the \(P\) value for the MTaSPUsSet test, \(P_{\text{MTaSPUsSet}}\), is
\[
\sum_{b=1}^{B} I(\text{MTaSPUsSet}(Z^{(b)})) \leq \text{MTaSPUsSet}(Z) + 1/(B + 1).
\]
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 $|Z| = \Phi^{-1}(1 - P/2)$, 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 3000 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 2938 control subjects. We further removed 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 4572 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 our reference panel to infer the LD structure for a set of SNPs. We designed a “control–control” experiment using the WTCCC whole control data; (2) using only a random set of 100 WTCCC controls as an ideal case; (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. Supplementary 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 $T_{sca}$ gave almost the same $P$ values, as shown in Supplementary Figure S3.

3.1.2 Pathway-based tests
For evaluations, we designed a control–control experiment using the WTCCC CD data. We randomly chose 3–15 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.

Supplementary Figure S6 compares the results of MTaSPUsPath with various reference panels, and of a similar pathway-based
adaptive test called GEE-aSPUsParh 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 realistic scenarios. To mimic real data, the genotype data, $G_{i,e,d}$, 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 Z score matrices, $Z_i$; let $Z_1$, $Z_2$ and $Z_3$, for the three genes in the GIANT data, representing various association patterns (Fig. 2 and Supplementary Fig. S15). The covariance structure among the multiple traits, $V$, was the correlation matrix in Equation (3) of Supplementary Figure S1 for women in the GIANT data. We simulated the traits for each gene as

$$y_i = MN(aZ_{ig}, V), i = 1, 2, 3,$$

where $a$ is a constant to be varied to control the effect size, and $g_i$ is the genotype data for subject $i$ (i.e. $i$th row vector of $G_{i,e,d}$) for $i = 1, 2, \ldots, 379$. The numbers of the SNPs were 8, 6 and 19 for TMEM110, ATP2A1 and RPS10-NUDT3, respectively.

To compare 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) ($S_{na}$) 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 SNPs in a gene.

Supplementary 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 anthropometrics 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 0.05/17 562 = 2.85 × 10^{-6} 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, $R$, 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 Supplementary Figure S1.

A stage-wise simulation strategy was used to calculate the $P$ values for each gene. We started with the simulation number $B = 10^3$; we sequentially increased $B$ to $10^5$, then $10^6$ and finally $10^7$ 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 or 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 (Supplementary Table S1). In particular, to demonstrate the sex differences of genetic effects, the new test

![Fig. 2. Log-transformed P values of univariate SNP–trait associations for some genes. TMEM110 and VTA1 (left two) uniquely detected using MTaSPUsSet compared with other single SNP–multiple traits approaches (MTaSPUs and $S_{na}$). ATP2A1 and RFWD2 (right two) uniquely detected using MTaSPUsSet compared with other multiple SNPs–single trait approach (aSPUs). P value less than genome-wide significance threshold of $5 \times 10^{-8}$ colored in dark color (blue). (Color version of this figure is available at Bioinformatics online.)](https://academic.oup.com/bioinformatics/article-abstract/33/1/64/2525686)
pinpointed 12 and 56 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 sex-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, although 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 cool) 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 cool 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 18 genes for women. Supplementary 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–single trait test was unable to do so; see genes TMEM110 and VTA1 in Figure 2 as examples (also shown in Supplementary Figures S10 and S11, respectively).

3.3.3 Comparison with gene-based multiple SNP–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 $0.05/(17562 \times 6) = 4.75 \times 10^{-7}$ 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. Supplementary Figures S12 and S13 show 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, MTaSPUsSet 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 Supplementary Figs. S12 and S13).

3.3.4 Comparison with another gene-based multiple trait test

We applied MGAS of Van der Sluis 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 $0.05/6424 = 7.78 \times 10^{-6}$ 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 MTaSPUsSet 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 6424 “kgg” genes starting with “LOC” and “LINC” were not in the MTaSPUsSet 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 MTaSPUsSet, respectively. Although there were some shared and general patterns between the results of the two methods, MTaSPUsSet 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 MTaSPUsSet was more powerful, although further validations are needed.

To further contrast the differences between the two tests, Supplementary Table S2 lists the 17 significant genes identified by MGAS with the corresponding $P$ values from the two tests. Genes LCORL, VTA1, BICD2, RASA2, GNA12, NCOA1, TNS1, CEP112, DNM3 and RFWD2 were significant for women by both MGAS and MTaSPUsSet, and LCORL, RASA2 and NDUFS3 were significant for men by both tests, while LCORL and RASA2 were significant for both men and women by both tests. Gene LCORL was known to be associated with anthropometric traits, including body height in African Americans (Carty et al., 2012), birth weight...
and adult height (Horikoshi et al., 2013); it is also a candidate gene for body weight in sheep (Al-Mamun 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 MTaSPUsSet, along with the \((\gamma_1, \gamma_2)\) values for the most significant MTSPUsSet \((\gamma_1, \gamma_2)\) 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 \((\gamma_1, \gamma_2)\) 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 \((\gamma_1, \gamma_2) = (1, 2)\) or \((2, 1)\) gave the most significant results. On one hand, 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 \((\gamma_1, \gamma_2) = (4, 8)\) 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 \((\gamma_1, \gamma_2)\) 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-control 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. In simulation study 2, we compared power of MTaSPUsSet with that based on single SNP-multiple trait and multiple SNP-single trait methods. When there were only weak and multiple association across multiple SNPs and multiple traits such as gene RPS10-NUDT3, MTaSPUsSet worked best.

We applied our gene-based MTaSPUsSet 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 MTaSPUsSet 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. MTaSPUsSet identified more genes. For more comparison, we also applied MGAS (Van der Sluis 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 MTaSPUsSet and MGAS to a common set of 5197 genes. MTaSPUsSet identified 27 and 39 significant genes for men and women, respectively, compared with only 7 and 14 genes by MGAS, suggesting possible power gains by MTaSPUsSet. 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.

Acknowledgement

The authors are grateful to the reviewers for constructive comments.

Funding

This work was supported by National Institutes of Health [R01GM113250, R01HL105397 and R01HL116720], and the Minnesota Supercomputing Institute.

References
