Multi-trait genome-wide analyses of the brain imaging phenotypes in UK Biobank

Chong Wu

Department of Statistics, Florida State University, Tallahassee, FL, USA

Abstract

Many genetic variants identified in genome-wide association studies (GWAS) are associated with multiple, sometimes seemingly unrelated traits. This motivates multi-trait association analyses, which have successfully identified novel associated loci for many complex diseases. While appealing, most existing methods focus on analyzing a relatively small number of traits and may yield inflated Type I error rates when a large number of traits need to be analyzed jointly. As deep phenotyping data are becoming rapidly available, we develop a novel method, referred to as aMAT (adaptive multi-trait association test), for multi-trait analysis of any number of traits. We applied aMAT to GWAS summary statistics for a set of 58 volumetric imaging derived phenotypes from the UK Biobank. aMAT had a genomic inflation factor of 1.04, indicating the Type I error rates were well controlled. More important, aMAT identified 24 distinct risk loci, 13 of which were ignored by standard GWAS. In comparison, the competing methods either had a suspicious genomic inflation factor or identified much fewer risk loci. Finally, four additional sets of traits have been analyzed and provided similar conclusions.

Keywords: Adaptive test; deep phenotyping data; GWAS summary data; multivariate trait; statistical power

1 Introduction

Genome-wide association studies (GWAS), which analyze a single trait each time, have identified thousands of genetic variants associated with an impressive number of complex traits and diseases [1]. While appealing, the identified genetic variants explain only a small proportion of the overall heritability, known as the 'missing heritability' problem [2]. In addition, many genetic variants are associated with multiple, sometimes seemingly unrelated traits [3]. This motivates multi-trait association tests [4–9] that detect genetic variant associations with multiple traits by jointly analyzing these traits. Multi-trait association tests have successfully improved the statistical power and enhanced biological interpretations.

There is an exciting opportunity to develop novel and powerful multi-trait association tests that can analyze any number of traits jointly, as deep phenotyping data from epidemiological studies and electronic health records are becoming rapidly available [10, 11]. For example, GWAS of 3,144 brain image-derived phenotypes have been carried out to provide insights into the genetic architecture of the brain structure and function [11]. However, most existing multi-trait methods focus on a relatively small number (i.e., less than ten) of traits and it is unclear whether existing multi-trait association testing methods can handle any number of traits.

In this work, we develop a novel computationally efficient multi-trait association testing method, termed adaptive multi-trait association test (aMAT). Compared with many existing methods, aMAT has two compelling features that make it potentially useful in many settings. First, aMAT yields well-controlled Type I error rates when analyzing any number (e.g., hundreds) of traits. This is achieved by taking the potential singularity of the trait correlation matrix into account. In contrast, many competing methods yield incorrect Type I error rates. Second, aMAT maintains high statistical power (usually more powerful than competing methods) over a wide range of scenarios. Because the association pattern varies between SNPs, the most optimal test will also vary and is an unknown prior. To maintain high power, aMAT first constructs a class of test such that hopefully one of them may have good power for a given scenario and then combines the testing results data-adaptively. Through simulations, we demonstrate these two compelling features. Additionally, by analyzing several brain image-derived phenotypes GWAS summary results jointly, we demonstrate that our approach can reproducibly identify additional associated genetic variants that have been ignored by several existing methods. These newly identified genetic variants provide additional biological insights into brain structure.

2 Methods

2.1 Overview of aMAT

We build upon previous works [12–15] to develop a novel multi-trait association test for jointly testing the association between a single genetic variant and any number of (potentially hundreds) traits. We use one single nucleotide polymorphism (SNP) for illustration, and the same method can be applied across the whole genome. Suppose we have its Z scores across p traits of interest, $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_p)'$. Under the null hypothesis H_0 that there is no association between the SNP and any traits of interest, $\boldsymbol{Z}$ follows a multivariate normal distribution [7–9] with mean zero and correlation matrix $\boldsymbol{R} = (r_{jk})$, i.e., $\boldsymbol{Z} \sim N(0, \boldsymbol{R})$.

aMAT involves three steps. First, we apply LD score regression (LDSC) [13, 14] to obtain an estimate of the correlation matrix, denoted by $\hat{\boldsymbol{R}}$ [9, 16]. Second, we construct a class of multi-trait association tests (MAT) such that hopefully each of them will be powerful under a certain scenario. To deal with potential (near) singularity problem of $\hat{\boldsymbol{R}}$, we calculate a modified pseudoinverse of $\hat{\boldsymbol{R}}$, denoted by $\hat{\boldsymbol{R}}_\gamma^+$. The test statistics of MAT is defined as follows:

$$T_{\mathrm{MAT}(\gamma)} = \boldsymbol{Z}'\hat{\boldsymbol{R}}_\gamma^+\boldsymbol{Z}.$$

Third, because the uniformly most powerful test does not exist and the optimal value of γ is data-dependent and unknown, we propose adaptive MAT (aMAT) to combine the results from a class of MAT tests:

$$T_{\mathrm{aMAT}} = \min_{\gamma\in\Gamma} p_{\mathrm{MAT}(\gamma)},$$

where $p_{\mathrm{MAT}(\gamma)}$ is the p-value of the MAT(γ) test and the default setting of Γ is $\Gamma = \{1, 10, 30, 50\}$, which performs well in both simulations and real applications. In the end, we apply a Gaussian copula approximation to calculate the p-value of aMAT efficiently.

2.2 Details of aMAT

Suppose we have Z score across p traits of interest for a SNP, $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_j)'$. Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ be the true marginal effect sizes for p traits. Note that Z score is either directly provided or can be calculated by $Z_j = \hat{\beta}_j/\mathrm{SE}_j$, where $\hat{\beta}_j$ is the estimated effect size for trait j and SE_j is its corresponding standard deviation. We are interested in testing whether the SNP is associated with any trait of interest, i.e., $H_0 : \boldsymbol{\beta} = 0$ vs $H_1 : \beta_j \neq 0$ for at least one $j \in \{1, 2, \ldots, p\}$. Of note, testing multi-trait effect ($H_0 : \boldsymbol{\beta} = 0$) is different from testing cross phenotype or pleiotropy effect [3], where the null hypothesis is at most one of the β_j, $j = 1, \ldots, p$, is nonzero.

Under the null hypothesis $H_0 : \boldsymbol{\beta} = 0$, $\boldsymbol{Z}$ asymptotically follows a multivariate normal distribution with mean zero and correlation matrix $\boldsymbol{R} = (r_{jk})$, i.e. $\boldsymbol{Z} \sim N(0, \boldsymbol{R})$ [7–9]. Then we construct aMAT by the following three steps: 1) estimating trait correlation matrix $\boldsymbol{R}$; 2) constructing a class of multi-trait association tests (MATs); and 3) constructing adaptive MAT (aMAT) to maintain high power across a wide range of scenarios.

Estimating trait correlation matrix $\boldsymbol{R}$ Following others [9, 16], aMAT applies LD Score Regression (LDSC) [13, 14] to obtain an estimate of $\boldsymbol{R}$, denoted by $\hat{\boldsymbol{R}}$. Specifically, we apply bivariate LDSC [14] to estimate the off-diagonal elements and univariate LDSC [13] to estimate the diagonal elements of $\boldsymbol{R}$. Because trait correlation matrix is the same for all the SNPs across the genome under the null, $\boldsymbol{R}$ only needs to be estimated once.

By applying LDSC, aMAT considers estimation error, including population stratification, cryptic relatedness, unknown sample overlap, and technical artifacts [16]. As demonstrated by others [9], LDSC provides a more accurate estimate of $\boldsymbol{R}$ than a commonly used approach that estimates $\boldsymbol{R}$ by sample correlation of genome-wide Z scores for each pair of traits [7, 8]. Thus, we choose LDSC to estimate the trait correlation matrix.

Constructing a class of MATs Because the underlying association pattern is unknown, we construct a class of multi-trait association tests (MATs) that hopefully each of them would be powerful for a given scenario.

Many existing methods involve the inverse of estimated trait correlation matrix $\hat{\boldsymbol{R}}$. We observe that $\hat{\boldsymbol{R}}$ is often near singular when analyzing a large number of traits, leading to incorrect Type I error rates. To address this challenge, we propose a modified pseudoinverse approach. Specifically, we first apply the singular value decomposition (SVD) to $\hat{\boldsymbol{R}}$ as $\hat{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}'$, where $\boldsymbol{U}$ is a $p \times p$ orthogonal matrix and $\boldsymbol{\Sigma}$ is a $p \times p$ diagonal matrix that contains the descending ordered singular values σ_i of $\hat{\boldsymbol{R}}$ on its diagonal. We next calculate a modified pseudoinverse of $\hat{\boldsymbol{R}}$ by

$$\hat{\boldsymbol{R}}_\gamma^+ = \boldsymbol{U}\boldsymbol{\Sigma}_\gamma^+\boldsymbol{U}',$$

where $\boldsymbol{\Sigma}_\gamma^+$ is formed from $\boldsymbol{\Sigma}$ by taking the reciprocal of the largest k singular values $\sigma_1, \ldots, \sigma_k$, and setting all other elements to zero, where k is the largest integer that satisfies $\sigma_1/\sigma_k < \gamma$. This is analogous to the principal components analysis, which restricts the analysis to top k axes of the largest variation. Next, we construct a class of multi-trait association tests (MATs) as:

$$T_{\mathrm{MAT}(\gamma)} = \boldsymbol{Z}'\hat{\boldsymbol{R}}_\gamma^+\boldsymbol{Z}.$$

Of note, $T_{\mathrm{MAT}(\gamma)}$ follows a chi-squared distribution χ_k^2 with k degrees of freedom under the null, and the p-value of MAT(γ) can be calculated analytically. MAT(γ) covers many existing tests

as special cases. For example, MAT(1) equals to a principal component (PC)-based association test called ET [9]. MAT(50) equals to chi-squared test approximately when the estimated trait correlation matrix $\hat{\boldsymbol{R}}$ is not near singular.

Constructing aMAT Because the association pattern varies between SNPs, the most optimal test will also vary and is an unknown prior. For example, MAT(1) achieves high power when the first principal component (PC) captures the majority association signals across the p traits. In contrast, when most PCs have weak signals, MAT(1) will lose power and MAT with larger γ will be more powerful. To maintain high power over a wide range of scenarios, we propose aMAT. Specifically, aMAT uses the smallest p-value from a class of MATs as test statistics:

$$T_{\text{aMAT}} = \min_{\gamma \in \Gamma} p_{\text{MAT}(\gamma)},$$

where $p_{\text{MAT}(\gamma)}$ is the p-value of the MAT(γ) test, and a sensible default choice of Γ is $\Gamma = \{1, 10, 30, 50\}$. Similar methods have been widely used in gene-level association tests [12, 17], pathway-based analysis [18], and microbiome association analysis [19].

Next, to calculate the p-value of aMAT analytically, we apply a Gaussian copula approximation method, which has been successfully used in gene-level association tests [15]. Specifically, the Gaussian copula approximation-based method involves the following two steps.

- First, we estimate the correlation matrix $\boldsymbol{\Omega}$ of the MAT(γ)s [$\gamma \in \Gamma$] test statistics through parametric bootstrap under the null. Specifically, we first simulate a new Z score vector $\boldsymbol{Z}^{\text{null}}$ under the null. In other words, we generate $\boldsymbol{Z}^{\text{null}}$ by a multivariate normal distribution with mean zero and covariance matrix $\hat{\boldsymbol{R}}$. Next, we apply a class of MATs to the simulated Z score $\boldsymbol{Z}^{\text{null}}$ and obtain their p-values. Under the null hypothesis, the p-values of MATs are uniformly distributed, and thus their inverse-normal transformed values ($q_\gamma = \Phi^{-1}(1 - p_{\text{MAT}(\gamma)})$) follow a multivariate normal distribution with mean zero and covariance $\boldsymbol{\Omega}$. We repeat this procedure B (e.g., 10,000) times and then estimate $\boldsymbol{\Omega}$ by the sample correlation matrix of q_γ, denoted by $\hat{\boldsymbol{\Omega}}$. Because the null distribution for each SNP remains the same, this step only needs to be implemented once.

- Second, we apply a Gaussian copula approximation for the joint distribution of q_γ for $\gamma \in \Gamma$. Specifically, the p-value of aMAT (p_{aMAT}) can be calculated as

$$p_{\text{aMAT}} = 1 - \Phi_M\Big[\big\{\Phi^{-1}(1 - T_{\text{aMAT}}), \ldots, \Phi^{-1}(1 - T_{\text{aMAT}})\big\}_{|\Gamma| \times 1}; \hat{\boldsymbol{\Omega}}\Big],$$

  where Φ_M denotes joint distribution function of a multivariate normal distribution and $|\Gamma|$ is the size of Γ.

2.3 Simulations

To speed up and simplify computations, we directly generated Z scores from its asymptotic normal distribution, $N(\Delta, \boldsymbol{R})$. This simulation procedure has been widely used by others [9, 16]. The estimated trait correlation matrix of brain image-derived phenotypes (IDPs) [11] was used as the true trait correlation matrix $\boldsymbol{R}$ such that the simulation settings are similar to real applications. We mainly considered two scenarios: 1) the trait matrix for the set of 58 volumetric IDPs, denoted by Volume; and 2) the trait matrix for the set of 472 structural MRI related IDPs, denoted by Freesurfer (because those IDPs were derived by Freesurfer). To evaluate the impact of estimation error on $\boldsymbol{R}$, two simulation settings were considered. First, we conducted tests with the true trait correlation matrix $\boldsymbol{R}$. Second, we assumed $\boldsymbol{R}$ is unknown and conducted tests with the estimated trait correlation matrix $\hat{\boldsymbol{R}}(s)$. Specifically, similar to [16], independent normal noise with mean zero and variance s was added to each element of the matrix $\boldsymbol{R}$ to simulate the estimated trait correlation matrix, denoted by $\hat{\boldsymbol{R}}(s)$. Based on our real data applications, the sampling variance was roughly 5×10^{-5}. Thus we varied s from 10^{-5} to 10^{-4} to evaluate the impact of estimation error on $\boldsymbol{R}$.

We simulated 500 million or one billion Z score vectors under the null ($\Delta = 0$) to evaluate Type I error rates with different significance level α and simulated $10,000$ Z score vectors under the alternative ($\Delta \neq 0$) to evaluate statistical power with the genome-wide significance level 5×10^{-8}. Under the alternative, a wide range of scenarios were considered. For example, similar to [9], we generated $\Delta = \sum_{j=1}^{k} c\sigma_j u_j$, where u_j is the singular vector of the trait correlation matrix $\boldsymbol{R}$, σ_j is the jth singular value of $\boldsymbol{R}$, c is the effect size, and k is the largest integer that satisfies $\sigma_1/\sigma_k < \gamma$. We varied γ and c to simulate different scenarios. We also considered other situations such as $\Delta = cu_j$ and $\Delta = c$. To save space, the majority simulation results were relegated to the supplementary.

We compared aMAT with the following popular tests: 1) the sum of Z score vector ($\sum_{j=1}^{p} Z_j$; denoted as SUM) [6]; 2) the sum of squared Z score vector ($\sum_{j=1}^{p} Z_j^2$; denoted as SSU) [5, 20]; 3) a standard chi-squared test; 4) a modified chi-squared test called Hom [8] that is powerful for the homogeneous effect situation; and 5) a PC-based association test [9], which equals to MAT(1). For chi-squared tests, we applied generalized inverse. When the trait correlation matrix is near singular (our focus here), two popular tests Het [8] and AT [9] had some numerical issues, failing to provide p-values and thus having been ignored in our comparisons. The minP test [4] that takes the minimum marginal p-values across p traits as the test statistics is highly conservative and time-consuming when p is large and thus has only been briefly discussed.

2.4 Analysis of UK Biobank brain imaging GWAS summary data

We reanalyzed GWAS summary results of image-derived phenotypes (IDPs) with up to 8,428 individuals [11] in UK Biobank to gain additional insights into the genetic architecture of brain imaging. Specifically, we conducted multi-trait association tests to detect the genetic association between an SNP and five sets of highly related IDPs with up to 472 IDPs per set. These five sets were determined by [11], covering a wide range of the IDP classes with significant trait correlations in each grouping. To be concrete, we focused on the set of 58 brain volumetric measures [11] and briefly discussed the additional four IDP sets.

Data pre-process and multi-trait association tests. We removed variants that were either non-biallelic or strand ambiguous (SNPs with A/T or C/G alleles) and analyzed the remaining 9,971,805 SNPs. LDSC was applied to estimate the trait correlation matrix. For each set of IDPs, we applied aMAT with default setting ($\Gamma = \{1, 10, 30, 50\}$) and many competing methods, including SUM, SSU, MAT(1) (equals to ET), Chi-squared, and Hom tests to test the overall association between a SNP and a pre-defined IDP set.

Result analyses. We used the genomic inflation factor to evaluate the empirical Type I error rates. The genomic inflation factor is defined as the ratio of the median of the observed test statistic to the expected median, which quantifies the extent of inflation and false positive rate. An ideal genomic inflation factor is one. However, because of the polygenic nature of GWAS signals, a genomic inflation factor is usually slightly larger than one. However, a genomic inflation factor greater than 1.2 or even 1.5 shows an indication of inflated Type I error rates, and a genomic inflation factor less than 0.9 shows evidence for conservative Type I error rates.

In addition, we used Functional Mapping and Annotation (FUMA) [21] v1.3.4c to identify independent genomic risk loci and lead SNPs in these loci. Via FUMA, We further obtained the functional consequences for these SNPs by matching SNPs to many functional annotation databases, including CADD scores [22], REgulomeDB scores [23], chromatin states [24, 25], and ANNOVAR categories [26]. For lead SNPs identified by aMAT, replication was tested by an independent dataset obtained from ENIGMA consortium [27]. Specifically, we applied aMAT to the GWAS summary statistics of seven brain subcortical volumetric IDPs, including the volume of mean putamen, the volume of mean thalamus, the volume of mean pallidum, the volume of mean caudate, Intracranial Volume, the volume of mean amygdala, and the volume of mean accumbens. Then the replication rate was evaluated by the two-tailed binomial test. In the end, we filtered genome-wide significant SNPs identified by aMAT based on their functional annotations and then mapped them to genes by the following three strategies (via FUMA): positional mapping, eQTL mapping, and chromatin interaction mapping.

2.5 Availability of data and materials

The UK Biobank summary data is available in the `http://big.stats.ox.ac.uk`. The LDSC software and its required LD scores are available in the `https://github.com/bulik/ldsc`. The validation data from NIGMA Consortium can be obtained from `http://enigma.ini.usc.edu/research/download-enigma-gwas-results/`. The software for our proposed method aMAT is available at `https://github.com/ChongWuLab/aMAT`. Supplementary files available at `https://figshare.com/articles/aMAT_supplementary/9764519`.

3 Results

3.1 aMAT yields well-controlled Type I error rates

Table 1 shows the Type I error rates of different multi-trait methods with the Freesurfer trait correlation matrix. Our proposed methods MAT and aMAT yielded well controlled Type I error rates at different significance levels α. Of note, the Freesurfer set contained 472 IDPs, and 27 eigenvalue values of the trait correlation matrix were smaller than 0.01 (Supplementary Figure S1). As expected, chi-squared and Hom tests yielded inflated Type I error rates because corresponding test statistics involve the inverse of a near singular correlation matrix. Also, SSU yielded inflated Type I error rates. All methods except SSU yielded well-controlled Type I error rates under simulations with Volume trait correlation matrix and thus were relegated to Supplementary (Supplementary Table S1).

α	5×10^{-2}	1×10^{-2}	1×10^{-4}	1×10^{-6}	5×10^{-7}	5×10^{-8}
SUM	5.0×10^{-2}	1.0×10^{-2}	1.0×10^{-4}	9.7×10^{-7}	5.2×10^{-7}	4.5×10^{-8}
SSU	4.6×10^{-2}	1.0×10^{-2}	$1.8 \times 10^{-4}*$	$3.5 \times 10^{-6}*$	$2.0 \times 10^{-6}*$	$2.7 \times 10^{-7}*$
Chi-squared	4.7×10^{-2}	1.0×10^{-2}	$1.7 \times 10^{-4}*$	$4.3 \times 10^{-6}*$	2.6×10^{-6}	$4.9 \times 10^{-7}*$
Hom	$5.6 \times 10^{-1}*$	$4.5 \times 10^{-1}*$	$2.5 \times 10^{-1}*$	$1.5 \times 10^{-1}*$	$1.4 \times 10^{-1}*$	$1.1 \times 10^{-1}*$
MAT(1)	5.0×10^{-2}	1.0×10^{-2}	1.0×10^{-4}	9.8×10^{-7}	4.8×10^{-7}	4.8×10^{-8}
MAT(10)	5.0×10^{-2}	1.0×10^{-2}	1.0×10^{-4}	1.0×10^{-6}	4.9×10^{-7}	4.5×10^{-8}
MAT(30)	5.0×10^{-2}	1.0×10^{-2}	1.0×10^{-4}	1.0×10^{-6}	5.2×10^{-7}	5.4×10^{-8}
MAT(50)	5.0×10^{-2}	1.0×10^{-2}	9.9×10^{-5}	9.6×10^{-7}	4.4×10^{-7}	4.6×10^{-8}
aMAT	4.7×10^{-2}	9.4×10^{-3}	9.8×10^{-5}	1.1×10^{-6}	5.5×10^{-7}	4.7×10^{-8}

Table 1: **Type I error rates of different methods with the Freesurfer trait correlation matrix.** We simulated one billion (1×10^9) replications under the null and estimated Type I error rates as the proportions of p-values less than significance level α. * inflated Type I error rates.

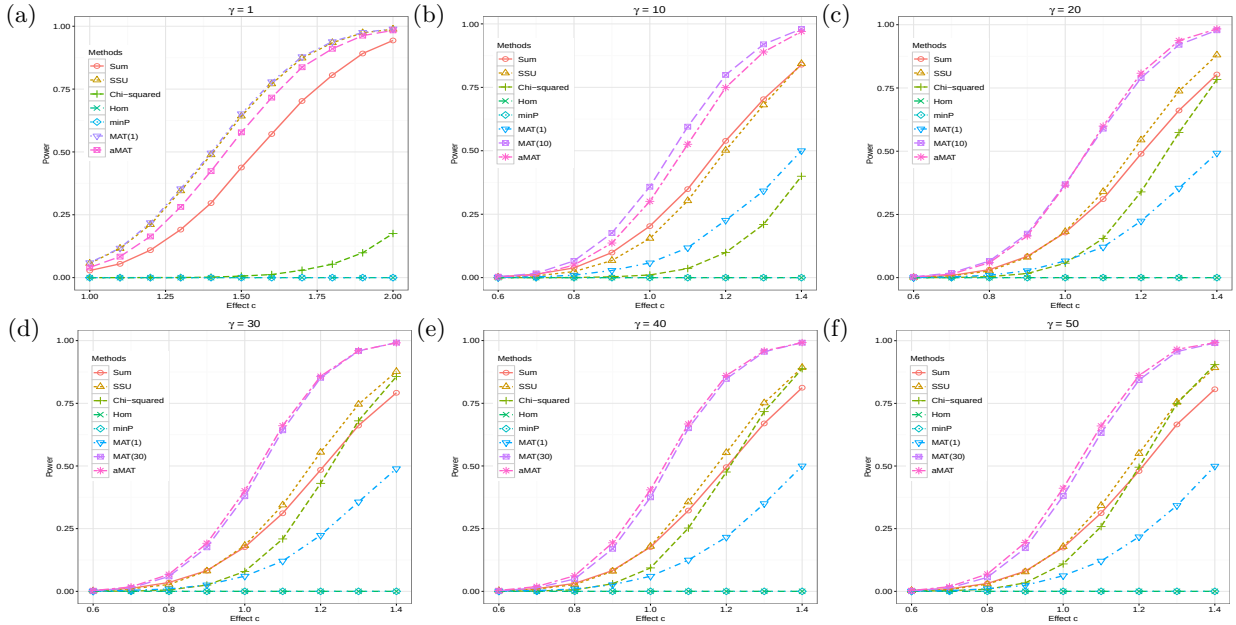

Figure 1: **Empirical power comparison with true Volume trait correlation matrix.** Under the alternative, we generated $\Delta = \sum_{j=1}^{k} c\sigma_j u_j$, where u_j is the jth singular vector of the Volume trait correlation matrix $\boldsymbol{R}$, σ_j is the jth singular value, c is the effect size, and k is the largest integer that satisfies $\sigma_1/\sigma_k < \gamma$. Empirical power was estimated as the proportions of p-values less than significance level 5×10^{-8}.

3.2 aMAT offers robust statistical power

Figure 1 shows the empirical power of different methods with true Volume trait correlation matrix $\boldsymbol{R}$. When the first principal component (PC) was informative, as expected, the MAT(1) performed best. However, MAT(1) was sensitive to the signal distribution and yielded much lower power when the top PC has weak or no signal (Supplementary Figure S2). The minP test was very conservative, and the power was almost zero for the situations considered here. Hom test [8] was known to be powerful for the homogeneous effect situation, but the power was close to zero when many principal components were informative (Figures 1d, e, and f). SUM and SSU tests ignored the relationships among traits and thus were less powerful than MAT tests. Because there is no uniformly optimal test for a composite alternative $H_1 : \boldsymbol{\beta} \neq 0$, different MAT tests achieved high power under different scenarios. However, aMAT maintained high power across a wide range of scenarios by combining the results from a class of MAT tests.

We considered several additional settings, including varying the way of selecting informative singular vectors and generating Δ, the trait correlation matrix (including five traits, 25 traits, Volume (58 traits), Area (206 traits), Thickness (208 traits), and Freesurf (472 traits)), and the generating distribution for Δ (Supplementary Figures S3–S13). We show that aMAT achieved a robust statistical power under all simulations considered. In contrast, competing methods that

achieved high power under one specific situation could lose power substantially under several other situations.

3.3 aMAT is robust to the estimation error of trait correlation matrix

In the previous two subsections, we assumed $\boldsymbol{R}$ was known. However, the trait correlation matrix $\boldsymbol{R}$ is unknown and has to be estimated. To evaluate the impact of estimation error on $\boldsymbol{R}$, we simulated Z score vectors from its asymptotic normal distribution $N(\Delta, \boldsymbol{R})$ and compared different testing methods with the estimated trait correlation matrix $\hat{\boldsymbol{R}}(s)$, which was constructed by adding independent normal noise with mean zero and variance s to each element of $\boldsymbol{R}$.

α	0.05	0.01	1×10^{-4}	1×10^{-6}	5×10^{-7}	5×10^{-8}
SUM	5.0×10^{-2}	1.0×10^{-2}	1.0×10^{-4}	1.0×10^{-6}	5.0×10^{-7}	4.7×10^{-8}
SSU	4.6×10^{-2}	1.0×10^{-2}	1.9×10^{-4}	4.1×10^{-6}	2.4×10^{-6}	3.9×10^{-7}
Chi-squared	6.4×10^{-2}	2.8×10^{-2}	1.1×10^{-2}	7.4×10^{-3}	7.1×10^{-3}	6.2×10^{-3}
Hom	1.7×10^{-1}	9.9×10^{-2}	5.0×10^{-2}	3.6×10^{-2}	3.4×10^{-2}	3.1×10^{-2}
MAT(1)	5.0×10^{-2}	1.0×10^{-2}	1.0×10^{-4}	1.0×10^{-6}	5.2×10^{-7}	4.9×10^{-8}
MAT(10)	5.0×10^{-2}	1.0×10^{-2}	9.9×10^{-5}	1.0×10^{-6}	5.0×10^{-7}	5.7×10^{-8}
MAT(30)	5.0×10^{-2}	1.0×10^{-2}	1.0×10^{-4}	1.0×10^{-6}	5.2×10^{-7}	4.5×10^{-8}
MAT(50)	5.0×10^{-2}	1.0×10^{-2}	9.9×10^{-5}	1.0×10^{-6}	5.0×10^{-7}	4.5×10^{-8}
aMAT	4.7×10^{-2}	9.4×10^{-3}	9.9×10^{-5}	1.2×10^{-6}	6.4×10^{-7}	4.7×10^{-8}

Table 2: **Type I error rates for different methods with the estimated Volume trait correlation matrix $\hat{\boldsymbol{R}}(10^{-4})$.** We simulated 500 million (5×10^8) replications with true Volume trait correlation matrix $\boldsymbol{R}$ under the null and constructed test statistics with $\hat{\boldsymbol{R}}(10^{-4})$. Type I error rates were estimated as the proportions of p-values less than significance level α.

Table 2 shows the Type I error rates of different methods with the estimated Volume trait correlation matrix $\hat{\boldsymbol{R}}(10^{-4})$. Both MAT and aMAT yielded well-controlled Type I error rates. In comparison, both chi-squared and Hom tests yielded inflated Type I error rates even though they had controlled (slightly conservative) Type I error rates with the true Volume trait correlation matrix $\boldsymbol{R}$ (Supplementary Table S1). This is because both chi-squared and Hom test statistics involve the inverse of the estimated trait correlation matrix $\hat{\boldsymbol{R}}(s)$ and small estimation error could lead to large deviation for small eigenvalues.

We next evaluated the impact of estimation errors on the statistical power. Figure 2 shows that the empirical power of MATs and aMAT with either true ($\boldsymbol{R}$) or estimated ($\boldsymbol{R}(\hat{10^{-4}})$) trait correlation matrix were almost the same, indicating estimation errors had little impact on the power of MATs and aMAT.

In the end, we considered several additional settings with different s ($s = 10^{-5}$ or $s = 5 \times 10^{-5}$) and/or with the estimated Freesurf trait correlation matrix. The results were similar to those in Table 2 and Figure 2, thus were relegated to the Supplementary (Supplementary Tables S2–S6 and

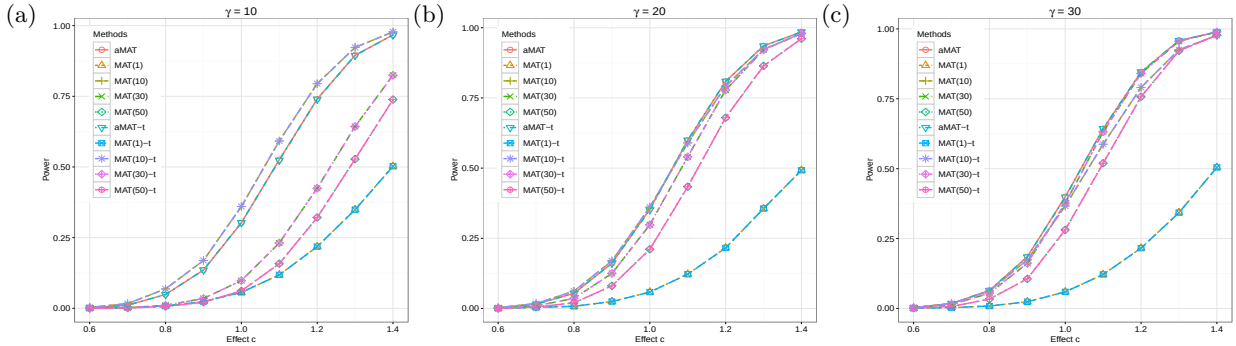

Figure 2: **Empirical power comparison between true and estimated Volume trait correlation matrix.** Under the alternative, we generated $\Delta = \sum_{j=1}^{k} c\sigma_j u_j$, where u_j is the singular vector of the true Volume correlation matrix $\boldsymbol{R}$, c is the effect size and k is the largest integer that satisfies $\sigma_1/\sigma_k < \gamma$. We simulated 10,000 replications with $\boldsymbol{R}$ and constructed test statistics with $\hat{\boldsymbol{R}}(10^{-4})$. We further estimated empirical power as the proportions of p-values less than significance level 5×10^{-8}. MAT(1), MAT(10), MAT(30), MAT(50), aMAT represent the results with $\hat{\boldsymbol{R}}(10^{-4})$, while MAT(1)-t, MAT(10)-t, MAT(30)-t, MAT(50)-t, aMAT-t represent the results with $\boldsymbol{R}$.

Supplementary Figures S14–S18).

3.4 aMAT identifies novel risk loci for brain volumetric measures

To demonstrate aMAT's effectiveness in real multi-trait association studies, we performed a multi-stage study for a set of 58 volumetric image-derived phenotypes (IDPs), which was defined by [11]. In the discovery stage, we applied aMAT to the GWAS summary statistics from UK Biobank [11]. aMAT had a genomic inflation factor of 1.04 (Supplementary Figure S19), indicating the type I error rates of aMAT were well controlled. aMAT identified 801 significant SNPs (with $P < 5 \times 10^{-8}$) across the whole genome, 453 of which were ignored by any individual IDP tests at the 5×10^{-8} genome-wide significance level. These 801 GWAS variants were represented by 28 lead SNPs, located in 24 distinct risk loci (Table 3 and Figure 3). Among these 28 lead SNPs, 13 SNPs (46.4%) were missed by any individual IDP tests at the 5×10^{-8} genome-wide significance level.

Next, we replicated our findings in an independent dataset obtained from the ENIGMA consortium [27], which contains GWAS summary statistics of seven subcortical volumes in up to 13,171 subjects. Despite seven brain subcortical volumetric IDPs reported in ENIGMA consortium only represent a subset of the volumetric IDPs in UK Biobank, the replication rate was high. Out of the 24 distinct risk loci, four loci (rs73123652, rs77956314, rs10129414, and rs1187159) were successfully replicated (two-tailed binomial test $P = 6.25 \times 10^{-30}$) under the genome-wide significance threshold ($P < 5 \times 10^{-8}$)). The numbers of replicated SNPs rose to 13 under a relaxed cutoff of 0.05 (two-tailed binomial test $P = 2.24 \times 10^{-10}$).

To link the identified variants to functional annotation, we applied Functional Mapping and An-

Locus	SNP	CHR	BP	A1	A2	MAF	aMAT P	Nearest Gene
1	**rs72688340**	1	47983839	G	A	0.171	2.5×10^{-9}	*AL356458.1*
2	**rs6665134**	1	180949628	C	A	0.401	1.7×10^{-10}	*STX6*
3	rs476557	1	215140251	A	C	0.475	4.3×10^{-8}	*RP11-323K10.1*
4	**rs1504**	2	37066018	T	G	0.447	1.4×10^{-8}	*AC007382.1*
5	rs13070564	3	190629975	G	T	0.383	8.8×10^{-13}	*GMNC*
6	**rs10031823**	4	103125031	T	C	0.395	1.3×10^{-8}	*SLC39A8*
6	rs34333163	4	103283117	A	G	0.075	4.4×10^{-16}	*SLC39A8*
7	**rs1922930**	6	1364691	A	C	0.118	3.8×10^{-8}	*RP4-668J24.2*
8	**rs77126132**	7	54966738	G	A	0.089	1.8×10^{-8}	*SNORA73*
9	rs2707521	7	120940436	C	T	0.377	2.2×10^{-16}	*CPED1*
10	**rs2974298**	8	42376477	T	C	0.493	2.5×10^{-9}	*SLC20A2*
11	rs10217651	9	118923652	A	G	0.386	1.7×10^{-14}	*PAPPA*
11	rs35565319	9	119065043	C	T	0.072	6.7×10^{-16}	*PAPPA*
11	**rs7030607**	9	119245183	G	A	0.36	6.7×10^{-9}	*ASTN2*
12	**rs12783517**	10	21878407	C	T	0.299	1.3×10^{-8}	*MLLT10*
13	rs4962692	10	126424823	G	A	0.414	2.7×10^{-10}	*FAM53B*
14	**rs1187159**	11	92009792	T	C	0.413	5.1×10^{-11}	*NDUFB11P1*
15	rs73123652	12	65874956	T	C	0.106	1.1×10^{-12}	*MSRB3*
16	rs4301837	12	102336310	T	C	0.496	6.9×10^{-9}	*DRAM1*
16	rs17797222	12	102913946	A	G	0.224	3.6×10^{-13}	*RP11-210L7.1*
17	rs12146713	12	106476805	T	C	0.093	3.2×10^{-9}	*NUAK1*
18	rs77956314	12	117323367	T	C	0.083	1.2×10^{-9}	*HRK*
19	rs10129414	14	56193272	G	A	0.438	2.0×10^{-9}	*RP11-813I20.2*
20	rs2464469	15	58362025	G	A	0.412	2.1×10^{-13}	*ALDH1A2*
21	**rs13330163**	16	70660243	A	G	0.453	3.0×10^{-12}	*IL34*
22	rs12920553	16	87227046	G	T	0.418	3.5×10^{-10}	*C16orf95*
23	**rs6121038**	20	30254773	T	G	0.291	5.2×10^{-10}	*BCL2L1*
24	**rs1004764**	22	38474852	G	A	0.377	4.2×10^{-9}	*SLC16A8*

Table 3: **Summary statistics of significantly associated regions identified by aMAT in the multi-trait analysis of a set of volume related IDPs.** Independent lead SNPs are defined by $r^2 < 0.1$ and distinct loci are > 250kb apart. SNP, CHR, BP, A1, A2 are the lead SNP, chromosome, position, effect allele, and non-effect allele, respectively. The MAF is the minor allele frequency and obtained from the 1000 genomes reference panel (Phase 3). aMAT P is the p-value for the aMAT test. The bolded SNPs correspond to the novel SNPs that were ignored by any individual IDP tests at the genome-wide significance level 5×10^{-8}.

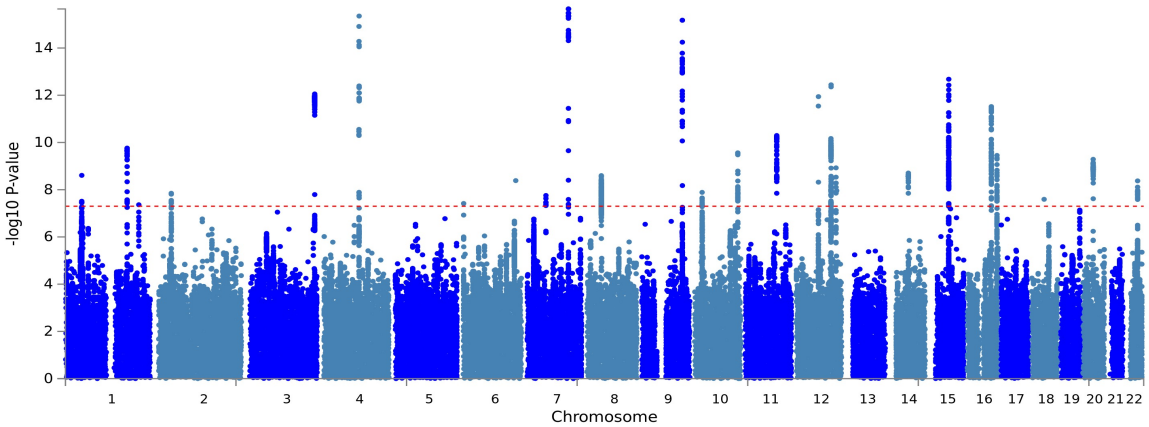

Figure 3: **Multi-trait analysis for the Volume set.** Manhattan plot displays the association results of aMAT per variant ordered by their genomic position on the x axis and showing the strength with the $-\log_{10}(p)$ on the y axis.

notation (FUMA) [21]. First, functional annotation of all genome-wide significant SNPs ($n = 772$, excluding those not available in FUMA database) showed that significant SNPs were mostly (90.8%) located in intronic/intergenic areas (Supplementary Table S7 and Figure 4b). Those relevant SNPs were also enriched for chromatin states 4 (33.2%) and 5 (40.0%), indicating effects on active transcription (Figure 4a). More important, six SNPs were exonic non-synonymous, which lead to a probably deleterious change in the sequence of the encoded protein (Supplementary Table S8). Five genome-wide significant SNPs (rs10507144, rs3789362, rs4646626, rs6680541, and rs2845871) had a high observed probability of a deleterious variant effect (CADD score [22] ≥ 20). Overall, these results suggest that most GWAS relevant SNPs are located in non-coding regions despite some non-synonymous variants have been found.

Second, we linked the associated variants to genes via three gene-mapping strategies used in FUMA. Figure 4d shows the Venn diagram of the number of mapped genes by three mapping strategies. Positional, eQTL, and chromatin interaction gene mapping strategies linked SNPs to 54, 50, and 76 genes, respectively. This resulted in 133 unique mapped genes (Supplementary Table S9), 13 of which were identified by all three mapping strategies (Supplementary Table S10). The locus (rs6121038) on chromosome 20 is particularly notable. Many genes in this locus (including *FOXS1*, *MYLK2*, *TPX2*, *BCL2L1*, and *COX412*) might interact physically or through eQTL, as indicated by chromatin interaction data in mesenchymal stem cell line and eQTL data (Figure 4c). Therefore, those genes might affect volumetric IDPs via a similar biological mechanism. Six genes (*STX6*, *EGFR*, *SMIM19*, *METTL10*, *LEMD3*, and *MYLK2*) are of particular interest as they are implicated via eQTL mapping with brain-related tissues. By searching GWAS Catalog [1], we found that three out of six (*METTL10*, *LEMD3*, and *MYLK2*) genes were reported to be associated with

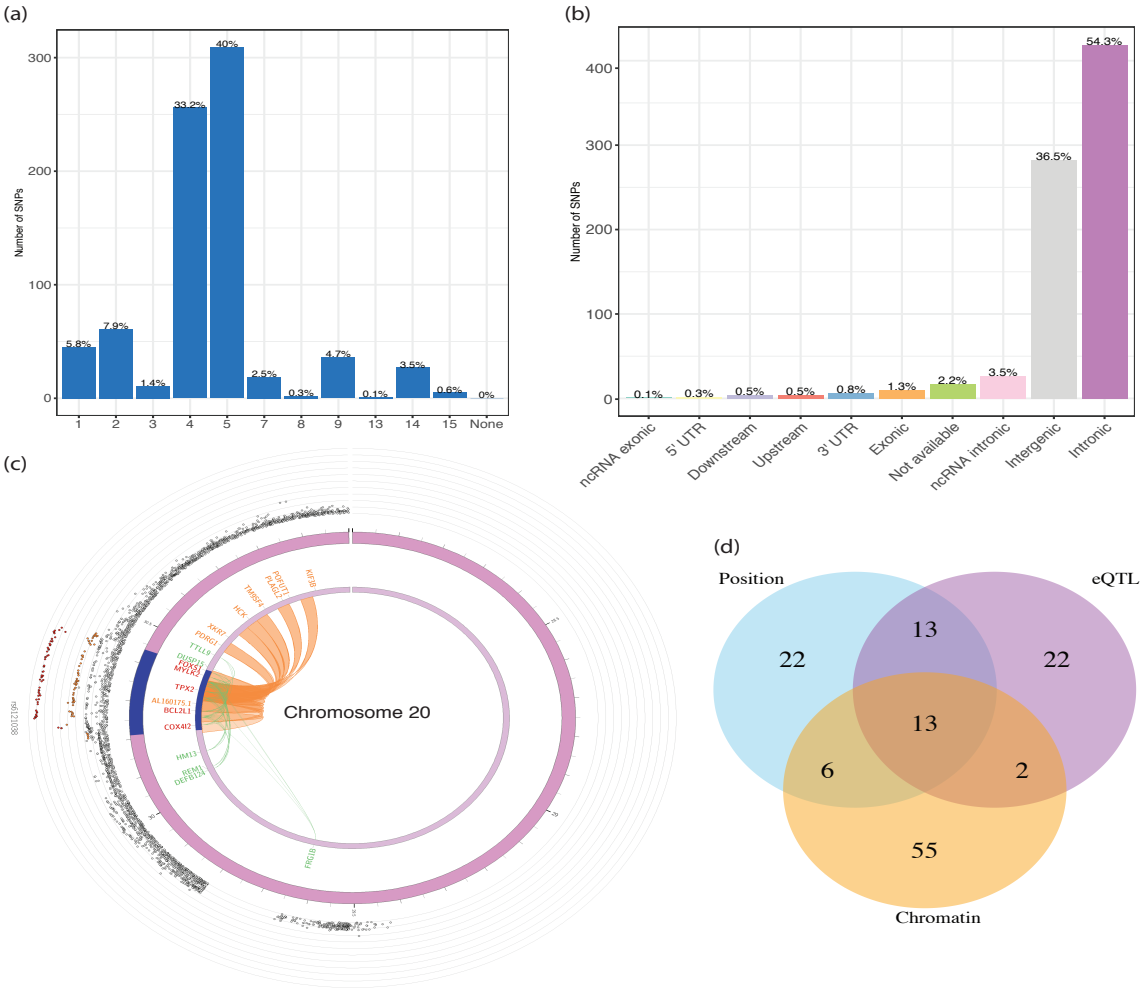

Figure 4: **Functional annotation and implications of aMAT results. a** and **b**, Distribution of (**a**) functional effects and (**b**) minimum chromatin state across 127 tissue/cell types for variants in aMAT identified genomic risk loci. **c**, Zoomed-in Circos plot of chromosome 20. Circos plots show implicated genes by either chromatin interaction (colored orange) or eQTLs (colored green) or both (colored red). The dark blue areas are genomic risk loci identified by aMAT. Chromatin interactions and eQTL associations are colored orange and green, respectively. The most outer layer shows a Manhattan plot, displaying $-\log_{10}(p)$ for SNPs with $p < 0.05$. **d**, Venn diagram of number of linked genes by three different strategies.

volumetric IDPs. For example, *METTL10* was associated with mean platelet volume [28], total hippocampal volume [29], and dentate gyrus granule cell layer volume [29]. *STX6* was identified by all three mapping strategies and contained a significant SNP (rs3789362) with CADD score higher than 20. *STX6* was reported to be associated with progressive supranuclear palsy (PSP) [30, 31] and the volume of many brain regions (including cerebellum, thalamus, putamen, pallidum, hippocampus, and brainstem) were significantly reduced in PSP compared to control subjects [32]. Future studies might thus consider the potential role of *STX6* in PSP.

Third, the identified genes were enriched in many GWAS Catalog reported volumetric gene sets, including dentate gyrus granule cell layer volume ($P = 1.46 \times 10^{-13}$), hippocampal subfield CA4 volume ($P = 1.46 \times 10^{-13}$), and hippocampal subfield CA3 volume ($P = 5.96 \times 10^{-12}$) (Supplementary Figure S20). We further performed gene-set analysis for tissue expression and biological pathways via FUMA. Two related tissues (Brain Cerebellar Hemisphere and Brain Cerebellum) were nominally associated with the volumetric IDPs set when only correcting the number of tissues being tested (Figure S21).

In summary, aMAT identified several novel and replicable risk loci that have been ignored by standard GWAS analysis. These newly identified loci not only showcase the power of our proposed approach but also deepens our understanding of the genetic basis of brain volumetric measures.

3.5 aMAT identifies novel risk loci for four additional IDP sets.

We further applied aMAT to four additional IDP sets, which were defined by [11] and represented many IDP classes with significant trait correlations in each grouping (Figure 5). These sets include 206 region area related IDPs (denoted by Area), 208 region thickness related IDPs (denoted by Thickness), 472 structural MRI related IDPs constructed by Freesurfer (denoted by Freesurfer), and 138 T1 ROIs related IDPs (denoted by ROIs). In total, we identified 84 independent risk loci and 97 lead SNPs, 59 of which were ignored by any individual IDP tests at the genome-wide significance level (Table 4, Supplementary Tables S11–S14, and Supplementary Figures S22–S25).

	IDPs	Sig SNPs	Loci	Lead SNPs	Novel SNPs
Area	206	2264	25	28	19
Thickness	208	39	2	2	0
Freesurf	472	2820	23	29	15
ROIs	138	2721	34	38	25

Table 4: **Summary statistics of aMAT for analyzing four additional IDP sets.** IDPs, Sig SNPs, Loci, Lead SNPs are the number of IDPs in the set, number of genome-wide significant SNPs identified by aMAT, number of independent risk loci, number of lead SNPs in the identified risk loci, respectively. Novel SNPs is the number of lead SNPs that were ignored by any individual GWAS at genome-wide significance level (5×10^{-8}).

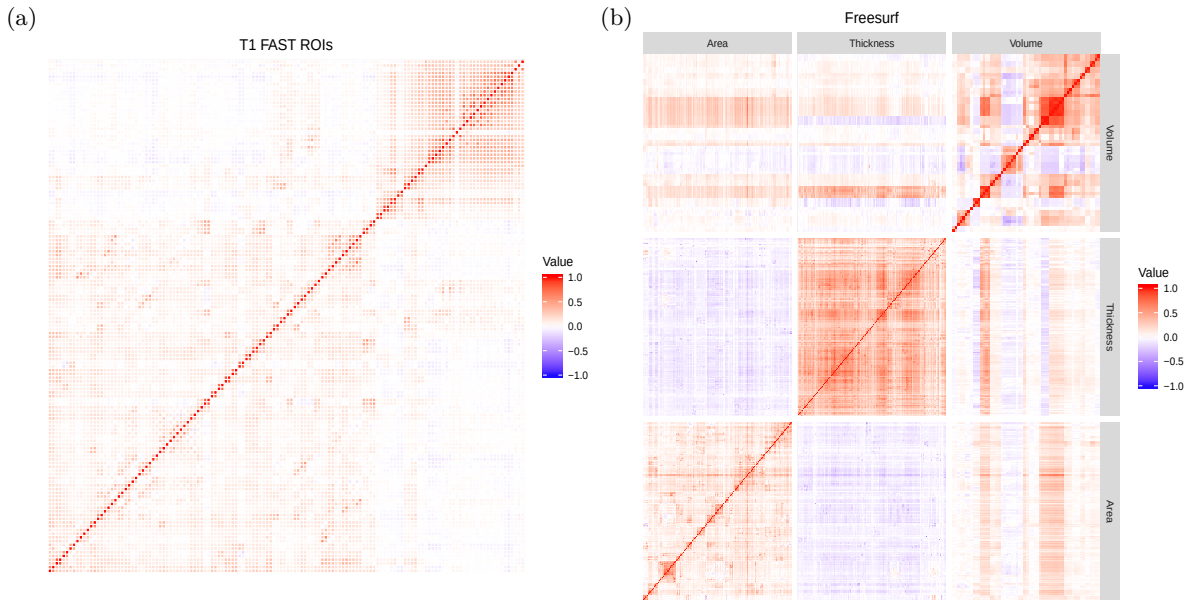

Figure 5: **The estimated trait correlation matrix for different sets of IDPs.** The subfigure (a) was for T1 FAST region of interests (ROIs) set, which contained 138 IDPs, and subfigure (b) was for Freesurf set, which contained 472 IDPs. Note that the trait correlation estimated here was the correlation of the raw phenotypes conditional on the covariates (e.g., confoundings), which was different from the raw trait correlation provided by Elliott et al. (2018) [11].

3.6 aMAT performs better than competing methods in real data applications

We compared aMAT to many competing methods. First, aMAT had a good genomic inflation factor in general (from 0.967 to 1.097) (Supplementary Table S15). In comparison, chi-squared test and Hom test had unstable genomic inflation factors. For example, chi-squared test had a genomic inflation factor of 0.338 for analyzing the Volume IDP set, while Hom test had a genomic inflation factor of 3.102 for analyzing the Freesurf IDP set and 0.755 for analyzing the Area IDP set. SUM had a very good genomic inflation factor, and SSU had a slightly inflated genomic inflation factor (between 1.186 and 1.250). These results are in line with our simulation results.

Next, we compared the statistical power of different tests that have a good genomic inflation factor. Supplementary Figure S26 shows the Venn diagram for the numbers of the significant SNPs and associated loci identified by the tests at the 5×10^{-8} genome-wide significance level for analyzing the Volume IDP set. Across the whole genome, aMAT, MAT(50), MAT(1), and SUM identified 801, 671, 3, and 25 significant SNPs, respectively. More important, aMAT, MAT(50), MAT(1), and SUM detected 453, 347, 0, and 18 significant and novel SNPs, respectively; a significant and novel SNP is defined as the genome-wide significant SNP but not detected by any individual GWAS analysis at the 5×10^{-8} genome-wide significance level. We further compared the results at the risk region level; a risk region is defined by LDetect [33]. aMAT detected 23 risk regions, while MAT(1) and SUM identified 1 and 3 risk regions, respectively. Each test detected some risk regions that

have been ignored by other tests. For example, aMAT and SUM tests identified five and three risk regions that were ignored by other methods, respectively. This illustrates that each test can be more powerful than the others under certain genetic architectures. For analyzing other IDP sets, we obtained similar results (Supplementary Figure S27–S30).

4 Discussion

We have introduced aMAT, a multi-trait association test, to conduct jointly analysis of any number of traits. Through simulations and real data analyses, we demonstrated that aMAT could yield well-controlled Type I error rates and achieve high statistical power across a wide range of scenarios. In our empirical application to the Volume set (of 58 volumetric IDPs), we identified 28 lead SNPs located in 24 distinct risk loci, 13 of which were missed by any individual IDP tests at the genome-wide significance level 5×10^{-8}. These identified lead SNPs were well replicated by an independent dataset [27].

aMAT is a computationally efficient method since p-values can be calculated analytically. We further improved the computational efficiency by pre-calculating the distribution under the null, which remains the same for all SNPs. For example, by a parallel computing strategy with 100 cores in a standard server (each core has 4 GB memory), aMAT completed calculating p-values for multi-trait analysis of volumetric set in 2.2 minutes (about 9.97 million SNPs in total, Supplementary Table S16). Of note, because aMAT involves combining results from different tests, it is excepted that aMAT is slightly slower than competing methods. To facilitate data analysis for both statistical and clinical investigators, we have implemented our proposed method aMAT into open-source software.

aMAT is a general framework and can be easily extended to incorporate other multi-trait methods such as MTAG [16], N-GWAMA [34], and HIPO [35]. Specifically, we can treat any other powerful tests as an individual MTA test and then apply aMAT to combine its results to further improve the power. However, since most methods are originally designed for analyzing a few phenotypes jointly and ignore the singularity problems of the trait correlation matrix, some minor adaptions and strict evaluations are needed. Thus, we leave it to our future work. More importantly, we view aMAT to be complementary (rather than superior) to standard GWAS. This is because standard GWAS focus on a single trait and may identify several risk loci that have been ignored by aMAT.

We conclude with two limitations of our proposed method aMAT. First, the optimal choice of Γ is unknown. Of note, the default setting $\Gamma = \{1, 10, 30, 50\}$ performs good in both simulations and real data analyses. We further evaluated several different choices of Γ and found that aMAT results were robust to the choice of Γ in many simulation settings (Supplementary Figure S31).

However, there is no theoretical guarantee for our default choice of Γ. Second, aMAT applies LDSC to estimate the trait correlation matrix and therefore inherits both benefits and limitations from LDSC. For example, aMAT applies the univariate LDSC to estimate the diagonal elements of the trait correlation matrix, and univariate LDSC is known to be biased for a trait with high (SNP) heritability [36]. We leave these interesting topics to future work.

Funding

This work was supported by the First Year Assistant Professor grant at Florida State University.

Acknowledgements

We thank the associate editor and two reviewers for helpful and insightful comments and the Research Computing Center at Florida State University for providing computing resources.

References

[1] Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Research. 2018;47(D1):D1005–D1012.

[2] Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747–753.

[3] Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nature Reviews Genetics. 2013;14(7):483–495.

[4] Conneely KN, Boehnke M. So many correlated tests, so little time! Rapid adjustment of P values for multiple correlated tests. The American Journal of Human Genetics. 2007;81(6):1158–1168.

[5] Yang Q, Wang Y. Methods for analyzing multivariate phenotypes in genetic association studies. Journal of Probability and Statistics. 2012;2012:652569.

[6] He Q, Avery CL, Lin DY. A general framework for association tests with multivariate traits in large-scale genomics studies. Genetic Epidemiology. 2013;37(8):759–767.

[7] Kim J, Bai Y, Pan W. An adaptive association test for multiple phenotypes with GWAS summary statistics. Genetic Epidemiology. 2015;39(8):651–663.

[8] Zhu X, Feng T, Tayo BO, Liang J, Young JH, Franceschini N, et al. Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. The American Journal of Human Genetics. 2015;96(1):21–36.

[9] Guo B, Wu B. Integrate multiple traits to detect novel trait–gene association using GWAS summary data with an adaptive test approach. Bioinformatics. 2018;.

[10] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–209.

[11] Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. Nature. 2018;562(7726):210–216.

[12] Pan W, Kim J, Zhang Y, Shen X, Wei P. A powerful and adaptive association test for rare variants. Genetics. 2014;197(4):1081–1095.

[13] Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nature Genetics. 2015;47(11):1236–1241.

[14] Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics. 2015;47(3):291–295.

[15] Sun R, Lin X. Genetic Variant Set-Based Tests Using the Generalized Berk?Jones Statistic With Application to a Genome-Wide Association Study of Breast Cancer. Journal of the American Statistical Association. 2019;0(0):1–13.

[16] Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nature Genetics. 2018;50(2):229–237.

[17] Sun R, Lin X. Set-Based Tests for Genetic Association Using the Generalized Berk-Jones Statistic. arXiv preprint arXiv:171002469. 2017;.

[18] Kwak IY, Pan W. Adaptive gene-and pathway-trait association testing with GWAS summary statistics. Bioinformatics. 2015;32(8):1178–1184.

[19] Wu C, Chen J, Kim J, Pan W. An adaptive association test for microbiome data. Genome Medicine. 2016;8(1):56.

[20] Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. Genetic Epidemiology. 2009;33(6):497–507.

[21] Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nature Communications. 2017;8(1):1826.

[22] Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics. 2014;46(3):310–315.

[23] Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annota-

tion of functional variation in personal genomes using RegulomeDB. Genome Research. 2012;22(9):1790–1797.

[24] Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nature Methods. 2012;9(3):215–216.

[25] Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317–330.

[26] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research. 2010;38(16):e164–e164.

[27] Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivières S, Jahanshad N, et al. Common genetic variants influence human subcortical brain structures. Nature. 2015;520(7546):224–229.

[28] Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. Cell. 2016;167(5):1415–1429.

[29] Van der Meer D, Rokicki J, Kaufmann T, Córdova-Palomera A, Moberget T, Alnæs D, et al. Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes. Molecular Psychiatry. 2018;p. 1.

[30] Höglinger GU, Melhem NM, Dickson DW, Sleiman PM, Wang LS, Klei L, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. Nature Genetics. 2011;43(7):699–705.

[31] Chen JA, Chen Z, Won H, Huang AY, Lowe JK, Wojta K, et al. Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. Molecular Neurodegeneration. 2018;13(1):41.

[32] Messina D, Cerasa A, Condino F, Arabia G, Novellino F, Nicoletti G, et al. Patterns of brain atrophy in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. Parkinsonism & Related Disorders. 2011;17(3):172–176.

[33] Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics. 2016;32(2):283–285.

[34] Baselmans BM, Jansen R, Ip HF, van Dongen J, Abdellaoui A, van de Weijer MP, et al. Multivariate genome-wide analyses of the well-being spectrum. Nature Genetics. 2019;51(3):445–451.

[35] Qi G, Chatterjee N. Heritability informed power optimization (HIPO) leads to enhanced detection of genetic associations across multiple traits. PLoS genetics. 2018;14(10):e1007549.

[36] de Vlaming R, Johannesson M, Magnusson PK, Ikram MA, Visscher PM. Equivalence of LD-score regression and individual-level-data methods. bioRxiv. 2017;p. 211821.