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

Since many genetic variants identified in genome-wide association studies (GWAS) are associated with multiple, sometimes seemingly unrelated traits, one can improve the statistical power of GWAS by testing the association of multiple traits simultaneously. While appealing, most existing methods focus on analyzing a relatively small number of traits and may yield inflated Type I error rates when analyzing a large number of traits. We introduce a new method called aMAT for multi-trait analysis of GWAS summary statistics of an arbitrary number (e.g. hundreds) of traits. We first conduct extensive simulations and demonstrate that aMAT yields well-controlled Type I error rates and achieves robust statistical power when analyzing a large number of traits. Next, we apply aMAT to summary statistics for a group of 58 volume-related imaging phenotypes in UK Biobank. aMAT yields a genomic inflation factor of 1.04 and identifies 28 lead SNPs spanning in 24 distinct risk loci, 13 of which are missed by any individual univariate GWAS. In comparison, the competing methods either yield a suspicious genomic inflation factor or identify much fewer risk loci. Finally, four additional groups 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) have identified thousands of genetic variants associated with an impressive number of complex traits and diseases [1] by analyzing a single trait each time. However, the identified genetic variants explain only a small proportion of the overall heritability, known as "missing heritability" problem [2]. On the other hand, an interesting observation has been that many genetic variants are associated with multiple, sometimes seemingly unrelated traits [3]. To improve statistical power and offer new biological insights, several methods for multi-trait association test have been proposed [4–9].

While appealing, most existing multi-trait studies focus on analyzing a relatively small number of phenotypes (say, less than ten) jointly. However, 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 (IDPs) have been carried out to provide insights into the genetic architecture of the brain structure and function [11]. Thus, a method that can handle an arbitrary number of traits is urgently needed.

In this paper, we develop a method, adaptive multi-trait association test (aMAT), that enables a joint analysis of an arbitrary number (e.g. hundreds) of traits. Compared with many existing multi-trait association testing methods, aMAT has several compelling features that make it potentially useful in many settings. First, by taking the potential singularity of the trait correlation matrix into account, aMAT yields well-controlled Type I error rates when analyzing a large number (e.g. hundreds) of traits. In contrast, many competing methods may yield incorrect Type I error rates. Second, aMAT offers a general way that combines different methods to achieve robust statistical power over a wide range of scenarios. Because the uniformly most powerful test does not exist for multi-trait association testing problems in general, aMAT first constructs a class of test such that one of them may have good power for a given scenario. Then aMAT combines the testing results data-adaptively to achieve robust power performance. Third, like many competing methods [8, 9], aMAT can be applied to GWAS summary statistics (i.e. marginal association statistics), which are often publicly available. Fourth, aMAT applies the cross-trait LD Score regression (LDSC) [12] and thus captures all relevant sources of estimation error (including sampling variation, population stratification, unknown sample overlap, and cryptic relatedness) on estimating the correlation matrix. Finally, aMAT is computationally efficient even for hundreds of traits because the p-value of aMAT can be calculated analytically. Through extensive simulations and analysis of UK Biobank Image related phenotypes [11], we demonstrate that aMAT maintains well-controlled Type I error rates and achieves greater power than competing methods across a wide range of scenarios.

2 Methods

2.1 New method: aMAT

We propose a new test called adaptive multi-trait association test (aMAT) for testing the overall association between a single SNP with an arbitrary number of (potentially hundreds) traits. For illustration, we focus on a particular SNP j and repeat the same procedure for every SNP in the genome. Suppose we have its Z scores across p traits of interest, $\boldsymbol{Z_j} = (Z_{1j}, Z_{2j}, \ldots, Z_{jp})'$. Let $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ be the marginal effect sizes of the SNP j for p traits. Without confusion, we may skip j to simplify the notation. Note that Z score is either directly provided by a publicly available GWAS summary data or can be calculated by $Z_{sj} = \hat{\beta}_{sj}/\mathrm{SE}_{sj}$, where $\hat{\beta}_{sj}$ is the estimated β_{sj} and SE_{sj} is its standard deviation for trait s. We are interested in testing whether the SNP is associated with any traits, i.e., $H_0 : \boldsymbol{\beta} = 0$ vs $H_1 : \beta_j \neq 0$ for at least one $j \in \{1, 2, \ldots, p\}$. Note that 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 our aMAT by the following three steps: estimating trait correlation matrix $\boldsymbol{R}$; constructing a class of multi-trait association tests; constructing an adaptive test called aMAT to maintain robust power across different scenarios. In the following, we will describe these three steps in detail.

Estimating trait correlation matrix $\boldsymbol{R}$ Like MTAG [13], aMAT applies LD score regression (LDSC) [12, 14] to estimate $\boldsymbol{R}$ (estimated one is denoted by $\hat{\boldsymbol{R}}$). Specifically, bivariate LDSC [14] partitions the product of Z score from traits t and s as

$$E(Z_{tj}Z_{sj}) = C_g l_j + r_{ts},$$

where C_g is proportional to the genetic correlation between traits t and s; l_j is the LD score for SNP j, which can be estimated in a reference sample (e.g. 1000 Genomes Project data with European ancestry [15]). When $t = s$, it reduces to the univariate LDSC for a single trait [12]. Similar to MTAG [13], aMAT first runs a univariate LDSC [14] for each trait and uses the estimated intercepts to construct the diagonal elements of $\hat{\boldsymbol{R}}$. Then bivariate LDSCs [12] are applied for each pair of traits, and the estimated intercepts are used to construct the off-diagonal elements of $\hat{\boldsymbol{R}}$. Note that $\hat{\boldsymbol{R}}$ is the same for all the SNPs across the genome and thus only need to be estimated once.

Of note, by using LDSC, aMAT considers estimation error, including population stratification, cryptic relatedness, unknown sample overlap, and technical artifacts [13]. As demonstrated previously [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]. Importantly, previous research [8, 9] has provided theoretical justification for using trait correlation matrix $\boldsymbol{R}$ to conduct multi-trait association tests.

Constructing a class of MATs Given the estimated correlation matrix $\hat{\boldsymbol{R}}$, we construct a class of multi-trait association tests (MATs) that hopefully each of them would be powerful for a given scenario.

One popular method for testing $\boldsymbol{\beta} = 0$ is the chi-squared test that uses the Z score vector $\boldsymbol{Z}$ to construct its statistic as

$$T_{\chi^2} = \boldsymbol{Z}'\hat{\boldsymbol{R}}^{-1}\boldsymbol{Z},$$

where $\hat{\boldsymbol{R}}^{-1}$ is the inverse of $\hat{\boldsymbol{R}}$. Note that T_{χ^2} follows a chi-squared distribution χ^2_p with p (equals to the number of traits) degrees of freedom under the null and thus the p-value of chi-squared test can be calculated analytically. Many GWAS summary statistics-based methods [8, 9] have been developed based on the above chi-squared test. Because most existing studies only analyze less than ten ($p < 10$) traits jointly, $\hat{\boldsymbol{R}}^{-1}$ is usually well-defined, leading to well-controlled Type I error rates for chi-squared based tests. However, when analyzing hundreds of traits or highly correlated traits jointly, $\hat{\boldsymbol{R}}$ is often near singular, leading to numerical problems for $\hat{\boldsymbol{R}}^{-1}$ and thus incorrect Type I error for chi-square based methods.

Here, we propose a modified pseudoinverse approach to address the near singular problem of $\hat{\boldsymbol{R}}$. 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 the singular vector and $\boldsymbol{\Sigma}$ is a diagonal matrix with ordered singular values σ_i of $\hat{\boldsymbol{R}}$ in 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 zeros. We use a condition number threshold of $\sigma_1/\sigma_i < \gamma$ to select the largest k singular values. This is analogous to the principal components analysis, restricting the analysis to top k axes of the largest variation. We next construct a class of multi-trait association tests (MATs) as:

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

Similar to chi-squared test, $T_{\mathrm{MAT}(\gamma)}$ follows a chi-squared distribution χ^2_k with k degrees of freedom under the null, where k is the number of singular values that meet $\sigma_1/\sigma_i < \gamma$. Then we can calculate the p-value for MAT(γ) analytically and denote by $p_{\mathrm{MAT}(\gamma)}$.

Note that MAT(1) equals to a principal component (PC)-based association test called ET [9]. MAT(50) corresponds to chi-squared test approximately when the trait correlation matrix $\boldsymbol{R}$ is not near singular and can be viewed as a minor adaption of chi-squared test.

An adaptive test (aMAT) to maintain robust power In general, there is no uniformly most powerful test; different tests will be powerful under different scenarios. Since the association pattern varies between SNPs, the most optimal test will also change from SNP to SNP, but it is unknown prior before conducting the test. For example, MAT(1) achieves high power when the first principal component (PC) captures the majority association signals across p traits we considered. In contrast, when most PCs have weak signals, MAT(1) will lose power and MAT with larger γ will be more powerful. To maintain robust power over a wide range of scenarios, we propose an adaptive test called aMAT. Specifically, aMAT uses the smallest p-value from a class of MAT tests as the test statistics:

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

where $p_{\mathrm{MAT}(\gamma)}$ is the p-value of the MAT(γ) test and the default setting of Γ is $\Gamma = \{1, 10, 30, 50\}$. It is worth noting that using the minimum p-value of multiple tests as test statistics to maintain robust performance has been widely used in gene-based tests [16, 17], pathway-based tests [18], and microbiome association analysis [19].

Intuitively, the p-value of aMAT can be calculated by a one layer Monte Carlo simulation [16], which is generally time-consuming. Here, we apply a Gaussian copula approximation-based method to calculate its p-value analytically as follows:

- First, we estimate the correlation matrix $\boldsymbol{\Omega}$ of several component MAT(γ) test statistics through a parametric bootstrap method under the null. Specifically, we first simulate a new Z score vector $\boldsymbol{Z}^{\mathrm{null}}$ under the null. To consider potential singularity problems, we apply singular value decomposition to the trait correlation matrix $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}'$ and generate $\boldsymbol{Z}^{\mathrm{null}}$ by $\boldsymbol{N}\boldsymbol{U}\boldsymbol{D}^{1/2}\boldsymbol{U}'$, where $\boldsymbol{N}$ is $p \times 1$ vector and each element is generated by a standard normal distribution and $\boldsymbol{D}^{1/2}$ is formed by setting each negative element of $\boldsymbol{D}$ to zero and taking the square root of each element. Then we apply each MAT(γ) to the simulated Z score $\boldsymbol{Z}^{\mathrm{null}}$ and obtain its p-values. Under the null hypothesis, the p-values of those MATs are uniformly distributed, and thus its inverse-normal transformed values ($q_\gamma = qnorm(1 - p_{\mathrm{MAT}(\gamma)})$) follow normal distributions with mean

zero marginally. We repeat this procedure for B times and then estimate the correlation matrix of q_γ by its sample correlations and denote it by $\mathbf{\Omega}$.

- Once learned $\mathbf{\Omega}$, we apply a Gaussian copula approximation for the joint distribution of q_γ for $\gamma \in \Gamma$. In other words, we assume q_γ for $\gamma \in \Gamma$ follows a joint multivariate normal distribution. Then the p-value of aMAT (p_{aMAT}) can be calculated as

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

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

In practice, the above procedure with $B = 10,000$ provides a reliable estimate of $\mathbf{\Omega}$ and only needs to be run once for computing the p-values for all the SNPs across the whole genome. Furthermore, Gaussian copula approximation-based method has been applied to combine different gene-based tests for GWAS summary data and showed a good performance [17].

2.2 Simulations

To speed up and simplify computations, we directly generated Z scores from its asymptotic normal distribution, $N(\Delta, \boldsymbol{R})$, which has been widely used by others [9, 13]. To evaluate the performance with different type of trait correlation, we used two estimated trait correlation matrices from the GWAS of brain image-derived phenotypes (IDPs) [11]: the trait matrix for the group of Freesurfer volume IDPs, which included 58 volume related IDPs and was denoted by Volume; and the trait matrix for the group of all the available Freesurfer IDPs, which included 472 structural MRI related IDPs and was denoted by Freesurfer.

We set $\Delta = 0$ under the null and set $\Delta \neq 0$ under the alternative. We simulated 500 million (5×10^8) Z score vectors under the null to evaluate Type I error rates with different significance level α and simulated $10,000$ Z score vectors under the alternative to evaluate statistical power under the genome-wide significance level 5×10^{-8}. Under the alternative, we considered a wide range of scenarios. 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 $\boldsymbol{R}$, c is the effect size and k is the largest integer that stratifies $\sigma_1/\sigma_k < \gamma$. 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.

First, we compared different testing methods in terms of Type I error rates and power with the true trait correlation matrix $\boldsymbol{R}$. Second, we evaluated the impact of estimation

error on $\boldsymbol{R}$. Specifically, we simulated Z-score vectors with the true trait correlation matrix $\boldsymbol{R}$ and then conducted tests with the estimated trait correlation matrix $\hat{\boldsymbol{R}}(s)$. Specifically, similar to [13], we added independent normal noise with mean zero and variance s to each element of the matrix $\boldsymbol{R}$ to simulate the estimated trait correlation matrix $\hat{\boldsymbol{R}}(s)$. Based on our real data analysis, the sampling variance is roughly 5×10^{-5}, and thus we varied s from 10^{-5} to 10^{-4} to evaluate the impact of estimation error on $\hat{\boldsymbol{R}}$.

We compared the performance of aMAT with the following popular tests: (i) the sum of Z score vector ($\sum_{j=1}^{p} Z_j$; denoted as SUM) [6]; (ii) the sum of squared Z score vector ($\sum_{j=1}^{p} Z_j^2$; denoted as SSU) [5, 20]; (iii) a standard chi-squared test; (iv) a modified chi-squared test called Hom [8] that is powerful for the homogeneous effect situation; (v) 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 (focused here), two popular tests Het [8] and AT [9] had some numerical problems and failed to provide p-values and thus has been ignored in our simulations. The minP test [4] that takes the minimum marginal p-values across p traits as test statistics is highly conservative and time-consuming when p is high and thus has been only briefly discussed.

2.3 Analysis of UK Biobank brain imaging GWAS summary data

We reanalyzed GWAS summary data for thousands of image-derived phenotypes (IDPs) with up to 8,428 UK Biobank individuals [11]. To uncover the genetic components of brain structure and function, we conducted multi-trait association tests to detect the genetic association between an individual SNP and a group of highly related IDPs. Following [11], we analyzed five groups of IDPs with up to 472 IDPs per group.

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. For each SNP, the summary Z score for a single IDP was calculated by dividing the regression coefficient estimate by its standard deviation. Based on their Z score, we applied LDSC to estimate the trait correlation matrix. For each group of IDPs, we applied aMAT 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 group of IDPs.

Result analyses. We used FUMA [21] v1.3.4c to identify independent genomic risk loci and relevant SNPs in these loci. Via FUMA, We further obtained the functional consequences for these SNPs by matching SNPs to some 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 in the independent ENIGMA [27] sample. Specifically, we applied aMAT to the GWAS summary statistics of seven brain subcortical volume related 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 we evaluated the replication rate by the two-tailed binomial test.

In the end, we filtered genome-wide significant SNPs obtained by aMAT based on its functional annotations and then mapped them to genes by the following three strategies (via FUMA): positional mapping, eQTL mapping, and chromatin interaction mapping.

2.4 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 software for our proposed method aMAT is available at `https://github.com/ChongWu-Biostat/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. Note that the Freesurfer group 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 due to near singular correlation matrix that need to be inverted for their test statistics. Also, SSU yielded inflated Type I error rates. However, aMAT(50), which can be viewed as a minor adaption of Hom test, yielded well-controlled Type I error rates. Importantly, our proposed methods MAT and aMAT yielded well controlled Type I error rates at different significance levels α. All methods 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}	9.8×10^{-7}	5.0×10^{-7}	4.0×10^{-8}
SSU	4.6×10^{-2}	1.0×10^{-2}	1.7×10^{-4}	3.7×10^{-6}	2.0×10^{-6}	3.4×10^{-7}
Chi-squared	4.7×10^{-2}	1.0×10^{-2}	1.7×10^{-4}	4.5×10^{-6}	2.7×10^{-6}	5.4×10^{-7}
Hom	5.6×10^{-1}	4.5×10^{-1}	2.5×10^{-1}	1.5×10^{-1}	1.4×10^{-1}	1.1×10^{-1}
MAT(1)	5.0×10^{-2}	1.0×10^{-2}	1.0×10^{-4}	1.0×10^{-6}	5.7×10^{-7}	5.0×10^{-8}
MAT(10)	5.0×10^{-2}	1.0×10^{-2}	1.0×10^{-4}	9.8×10^{-7}	5.1×10^{-7}	4.8×10^{-8}
MAT(30)	5.0×10^{-2}	1.0×10^{-2}	9.9×10^{-5}	1.0×10^{-6}	5.4×10^{-7}	6.0×10^{-8}
MAT(50)	5.0×10^{-2}	1.0×10^{-2}	1.0×10^{-4}	1.0×10^{-6}	5.3×10^{-7}	6.4×10^{-8}
aMAT	4.7×10^{-2}	9.4×10^{-3}	9.7×10^{-5}	1.2×10^{-6}	6.3×10^{-7}	6.4×10^{-8}

Table 1: **Type I error rates of different methods with the Freesurfer trait correlation matrix.** We simulated 50 million (5×10^8) replications under the null and estimated Type I error rates as the proportions of p-values less than significance level α.

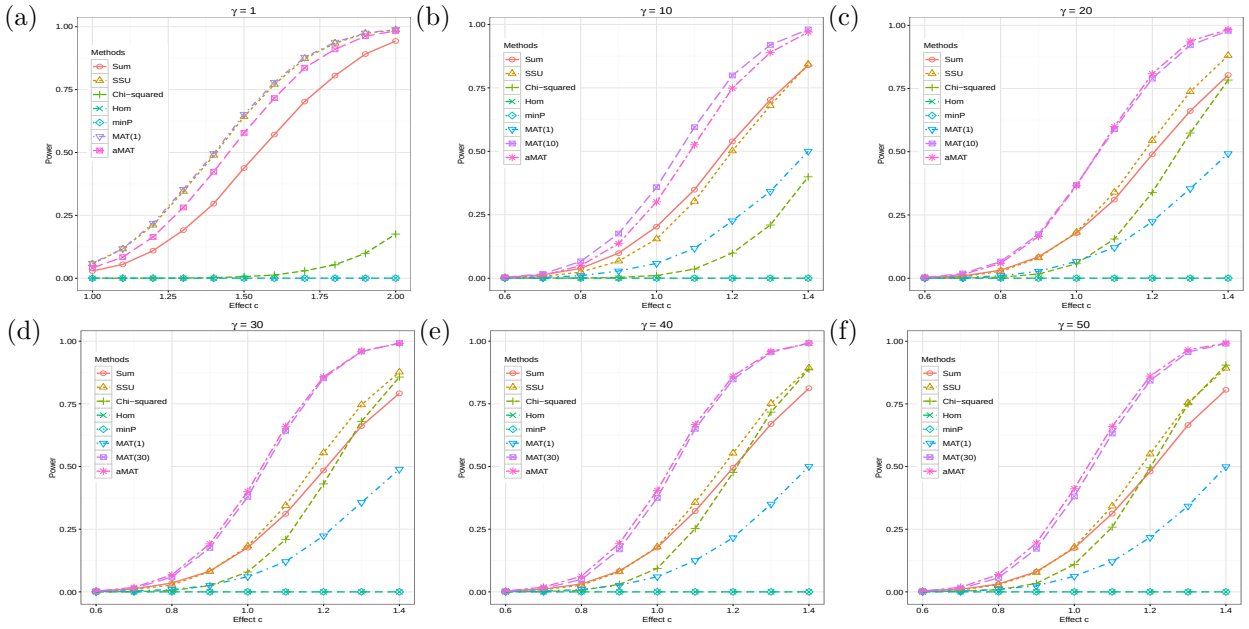

Figure 1: **Empirical power comparison with true Volume trait correlation matrix.** Under the alternative, we generated $\Delta = \sum_{j=1}^{k} c\sigma_j u_j$, where u_j is the singular vector of the Volume correlation matrix $\boldsymbol{R}$, c is the effect size and k is the largest integer that stratifies $\sigma_1/\sigma_k < \gamma$. We further estimated empirical power 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). Perhaps because the Volume trait correlation matrix was near singular, the minP test was very conservative and the power was almost zero for the situations considered here. Hom test [8] was powerful for the homogeneous effect situation and the power was close to zero as the effect was heterogeneous. SUM and SSU tests ignored the relationships among traits and were less powerful than MAT tests. Because there is no uniformly optimal test for a composite alternative $H_1 : \boldsymbol{\beta} \neq 0$, different NAT tests achieved high power under different scenarios. However, aMAT maintained a robust and 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, and the generating distribution for Δ (Supplementary Figures S3–S9). Across all simulation settings, a test achieved high power under one situation but might lose power substantially under another situation. In contrast, aMAT achieved a robust statistical power under all simulations considered.

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 we have to estimate the trait correlation matrix to construct test statistics. 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 $\boldsymbol{R}$.

Table 2 shows the Type I error rates of different methods with the estimated Volume trait correlation matrix $\hat{\boldsymbol{R}}(10^{-4})$. Note that standard Chi-squared test and Hom test yielded controlled (slightly conservative) Type I error rates with the true Volume trait correlation matrix $\boldsymbol{R}$ (Supplementary Table S1). However, when small estimation errors were considered, both the Chi-squared test and Hom test yielded inflated Type I error rates. Because both the Chi-squared test and Hom test construct test statistics with the inverse of the estimated trait correlation matrix $\hat{\boldsymbol{R}}(s)$, small estimation error on $\hat{\boldsymbol{R}}(s)$ may lead to a relatively large

deviation for small eigenvalues, leading to inflated Type I error rates. In contrast, both MAT and aMAT yielded well-controlled Type I error rates since only several largest eigenvalues of the estimated correlation matrix have been used.

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 50 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.

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

In the end, we considered several additional settings with different estimation error strength ($s = 10^{-5}$ or $s = 5 \times 10^{-5}$) and 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 Figures S10–S14).

3.4 Application to a UK Biobank brain imaging GWAS summary data

For our empirical application of aMAT, we built on a recent study by UK Biobank of thousands of image-derived phenotypes (IDPs) [11]. Similar to [11], we analyzed five groups of IDPs that are strongly genetically related (Methods). We mainly focused on the results of our proposed method aMAT and compared some competing methods at the end. Figure 3 illustrates the estimated trait correlation matrix for each group of IDPs, showing that many IDPs within each group are highly correlated.

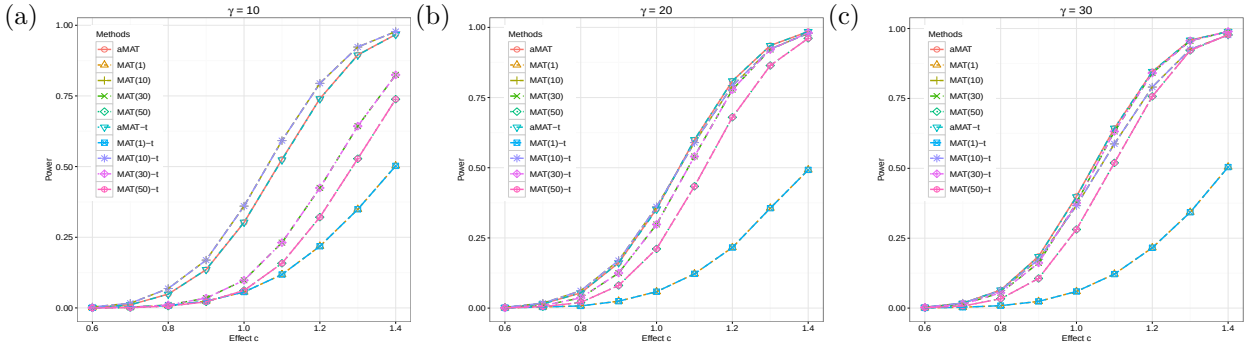

Figure 2: **Empirical power comparison between true and estimated Volume trait correlation matrix.** Under the alternative, we generated $\Delta = \sum_{j=1}^{k} c\sigma_j u_j$, where u_j is the singular vector of the true Volume correlation matrix $\boldsymbol{R}$, c is the effect size and k is the largest integer that stratifies $\sigma_1/\sigma_k < \gamma$. 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 $\boldsymbol{R}$.

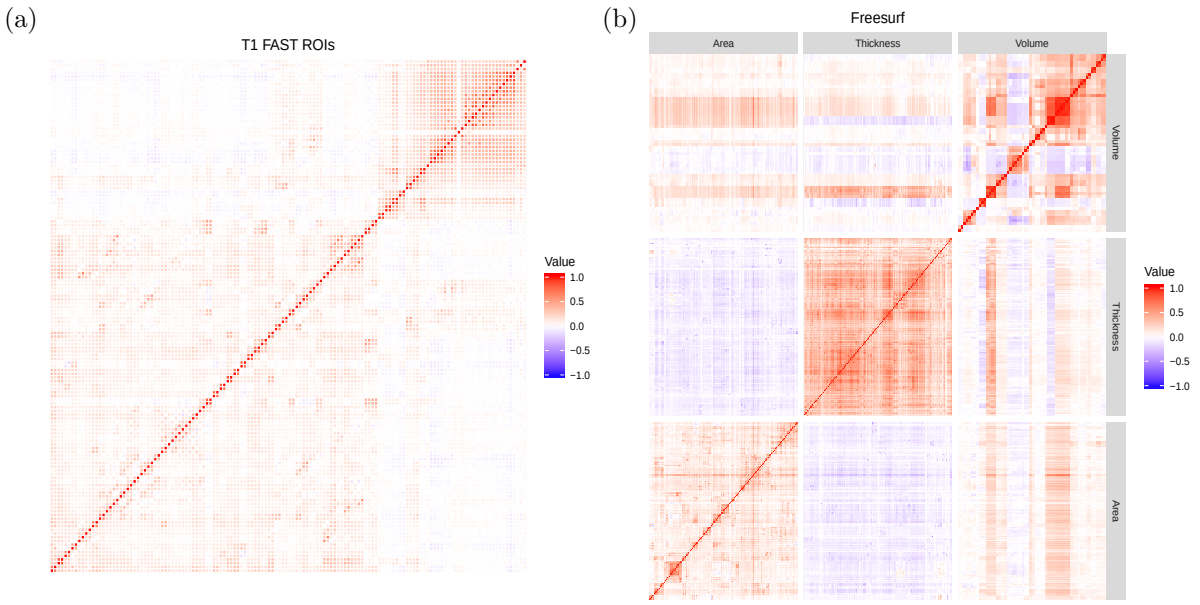

Figure 3: **The estimated trait correlation matrix for different groups of IDPs.** The subfigure (a) was for T1 FAST region of interests (ROIs) group, which contained 138 IDPs, and subfigure (b) was for Freesurf group, which contained 472 IDPs. Note that the trait correlation estimated here was the correlation 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].

aMAT results. We applied aMAT to the GWAS summary statistics from the UK Biobank [11]. To be concrete, we mainly focus on the results of analyzing the group of 58 Freesurfer volume IDPs. The aMAT results had a genomic inflation factor of 1.04 (Supplementary Figure S15). Among 9,971,805 SNPs to be tested, aMAT identified 801 significant SNPs, 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 4). 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. These results indicate that conducting aMAT test made it possible to uncover significant SNPs and loci that were ignored by an individual IDP test.

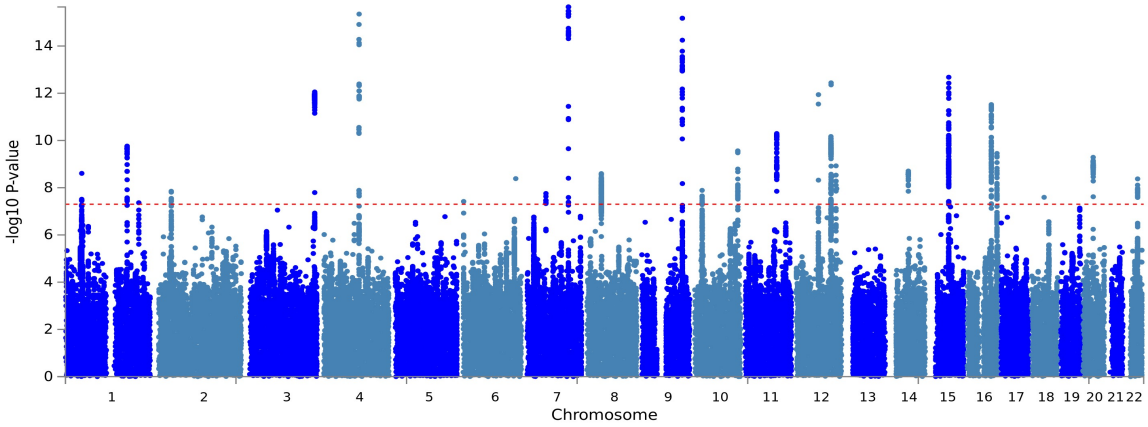

Figure 4: **Multi-trait analysis for the Volume group.** Manhattan plot displays the association results of aMAT per variant ordered by their genomic position on the x axis and showing the strength with the $-\log_{10}(p)$ on the y axis.

Replication of aMAT-identified loci. To test the lead SNPs for replication, we compared aMAT result for the volume group with those obtained by the ENIGMA consortium [27]. Specifically, we obtained GWAS summary statistics of seven subcortical volumes in up to 13,171 subjects. We then applied aMAT to the group of seven subcortical volumes from the ENIGMA, denoted by ENIGMA group. The aMAT had a genomic inflation factor of 1.06 and identified 927 genome-wide significant SNPs.

We next tested replication of the 28 lead SNPs in 24 distinct risk loci, 13 of which showed nominally significant association results ($P < 0.05$) for the same SNPs within the same locus (two-tailed binomial test $P = 2.24 \times 10^{-10}$). Importantly, four loci (rs73123652, rs77956314, rs10129414, and rs1187159) showed genome-wide significant association results ($P < 5 \times 10^{-8}$) for the same SNPs in the ENIGMA data (two-tailed binomial test $P = 6.25 \times 10^{-30}$).

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 group 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}.

Perhaps because seven brain subcortical volume IDPs reported in ENIGMA consortium only represent a subset of the Volume group (which contains 58 IDPs), 15 out of 28 lead SNPs are not replicated. To validation this argument, we reran multi-trait analyses for the group of 14 (including both left and right) subcortical volume as determined by Freesurfer. We identified 11 independent loci after LD clumping. Among these 11 loci, nine showed nominally significant association results ($P < 0.05$) for the same SNPs (two-tailed binomial test $P = 9.80 \times 10^{-11}$). Again, the same four loci (rs10502971, rs1187160, rs12302173, and rs8017172) showed genome-wide significant association results ($P < 5 \times 10^{-8}$) for the same SNPs within the same locus (two-tailed binomial test $P = 6.25 \times 10^{-30}$).

Functional annotation of genetic variants. We applied Functional Mapping and Annotation (FUMA) [21] (Methods) to link the associated variants to functional annotation. Functional annotation of all genome=wide significant SNPs (excluding the one not available in FUMA database, $n = 772$) in the associated loci showed that SNPs were mostly (90.8%) located in intronic/intergenic areas (Supplementary Table S7 and Figure 5b). Those relevant SNPs were also enriched for chromatin states 4 (33.2%) and 5 (40.0%), indicating effects on active transcription (Figure 5a). Importantly, six SNPs were exonic non-synonymous that 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] $\geq$ 20). Overall, these results suggest that most GWAS relevant SNPs are located in non-coding regions despite some non-synonymous variants have been found.

Next, we linked the significantly associated variants to genes via three gene-mapping strategies used in FUMA (Methods). Figure 5d 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 on chromosome 20 (rs6121038) is particularly notable. Many genes in this locus such as *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 5c). Therefore, those genes might affect volume related 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 in the brain related tissues. By searching GWAS catalog [1], we found that three out of six (*METTL10*, *LEMD3*, and *MYLK2*) genes

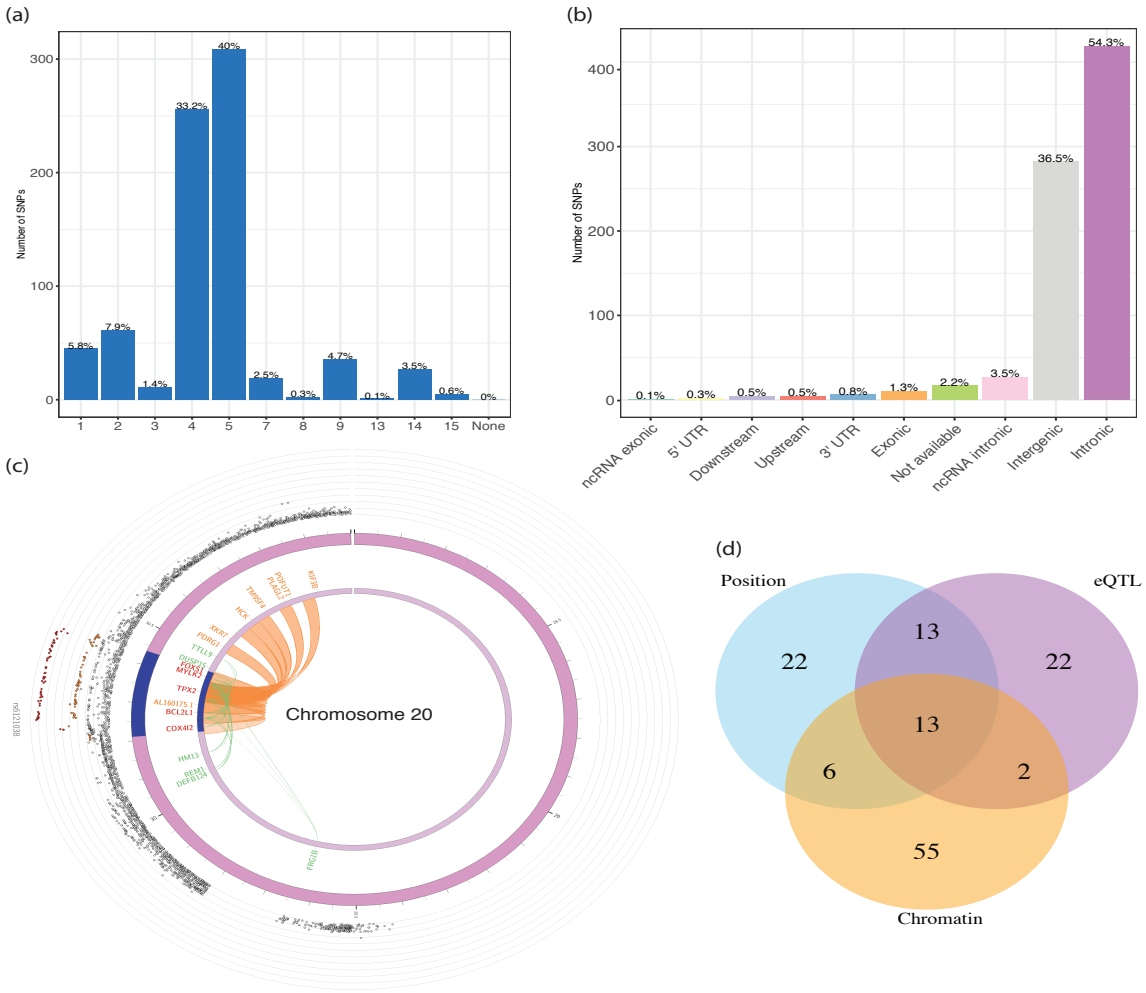

Figure 5: **Functional annotation and implications of aMAT results.** **a** and **b**, Distribution of (**a**) functional effects and (**b**) minimum chromatin state across 127 tissue/cell types for variants in aMAT identified genomic risk loci. **c**, Zoomed-in Circos plot of chromosome 20. Circos plots 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.

were reported to be associated with volume related 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.

Overall, the identified genes were enriched in many GWAS catalog reported volume-related 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 S16). 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 Volume IDPs group when only correcting the number of tissues being tested (Figure S17).

Results for analyzing other IDP groups. We further applied aMAT to four additional IDP groups, including 206 region area related IDPs (denoted by Area), 208 region thickness related IDPs (denoted by Thickness), 472 Freesurfer structural MRI related IDPs (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 ignored by any individual IDP tests at the genome-wide significance level (Table 4, Supplementary Tables S11–S14, and Supplementary Figures S18–S21).

	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 groups.** IDPs, Sig SNPs, Loci, Lead SNPs are the number of IDPs in the group, 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}).

Comparison to other Multi-trait association tests 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 a unstable genomic inflation factor. For example, Chi-squared test had a genomic inflation factor of 0.338 for analyzing the Volume IDP group, while Hom test had a genomic inflation factor of 3.102 for analyzing the Freesurf IDP group and 0.755 for analyzing the Area IDP group. SUM had a very good genomic inflation factor and SSU had a slightly inflated genomic inflation factor (between 1.186 and 1.250).

Next, we compared the statistical power of different tests that have a good genomic inflation factor. Supplementary Figure S22 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 group. Among 9,971,805 SNPs to be tests, aMAT, MAT(50), MAT(1), and SUM identified 801, 671, 3, and 25 significant SNPs, respectively. Importantly, 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 one that is not detected by any individual IDP tests 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. Note that each test detected some risk regions that other tests failed to detect. For example, aMAT identified 5 risk regions that were ignored by competing methods. This illustrates that each test can be more powerful than the others under certain situations because there is no most uniformly most powerful test in general. Importantly, our proposed adaptive test aMAT detect more SNPs and loci than competing methods, demonstrating its robust performance. For analyzing other IDP groups, we obtained similar results (Supplementary Figure S23–S26).

4 Discussion

We have introduced aMAT, a multi-trait method for conducting a joint analysis of GWAS summary statistics for an arbitrary number of traits. Both our simulation and real data results confirmed that aMAT could yield well-controlled Type I error rates and achieve robust statistical power across a wide range of scenarios. In our empirical application to the Volume group (of 58 volume related 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 can be well replicated by an independent study [27]. Importantly, by integrating functional annotation via FUMA, we found most GWAS relevant SNPs are located in non-coding regions, and the linked genes are over-enriched in volume related gene sets. The proposed method aMAT are implemented in R, which is

publicly available at https://github.com/ChongWu-Biostat/aMAT.

Our proposed method aMAT is a multi-trait method that aims to test the overall association between a SNP and an arbitrary number of traits. In other words, we are interested in testing whether the SNP is associated with none of the traits or not. Multi-trait analysis is different from cross phenotype or pleiotropy effect analysis [3], where the null hypothesis is at most one trait is associated with the SNP. Of note, the proposed adaptive test aMAT is a general framework and can be easily extended to incorporate other multi-trait methods such as MTAG [13], N-GWAMA [34], and HIPOn [35]. To be specific, 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 adaption and strict evaluation might be needed. We leave it to our future work.

We conclude with several limitations of our proposed method aMAT. First, because there is no uniformly most powerful test, aMAT first constructs a class of multi-trait association tests (MAT) with different γs such that hopefully at least one of them would be powerful for a given scenario. Then aMAT combines the results data-adaptively. Even though the default setting $\Gamma = \{1, 10, 30, 50\}$ performs good in both simulations and real data analyses and the testing result seems robust to the choice of Γ set (Supplementary Figure S27), we do not know the optimal choice of Γ and leave it to future research. Second, aMAT applies LDSC to estimate trait correlation matrix and therefore inherit both benefits and limitations from LDSC. For example, the univariate LDSC intercept, which was used for estimating the diagonal elements of the trait correlation matrix, can be biased for a trait with a large sample size and a high (SNP) heritability [36].

Funding

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

Acknowledgements

We thank 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] 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.

[13] 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.

[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] Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;526(7571):75–81.

[16] 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.

[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. Annotation 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.