An Adaptive and Robust Method for Multi-trait Analysis of Genome-wide Association Studies Using Summary Statistics

Qiaolan Deng¹,², Chi Song², and Shili Lin¹,³

¹Interdisciplinary Ph.D. Program in Biostatistics, ²Division of Biostatistics, College of Public Health, ³Department of Statistics, The Ohio State University, Columbus, Ohio

Abstract

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with human traits or diseases in the past decade. Nevertheless, much of the heritability of many traits is still unaccounted for. Commonly used single-trait analysis methods are conservative, while multi-trait methods improve statistical power by integrating association evidence across multiple traits. In contrast to individual-level data, GWAS summary statistics are usually publicly available, and thus methods using only summary statistics have greater usage. Although many methods have been developed for joint analysis of multiple traits using summary statistics, there are many issues, including inconsistent performance, computational inefficiency, and numerical problems when considering lots of traits. To address these challenges, we propose a multi-trait adaptive Fisher method for summary statistics (MTAFS), a computationally efficient method with robust power performance. We applied MTAFS to two sets of brain imaging derived phenotypes (IDPs) from the UK Biobank, including a set of 58 Volumetric IDPs and a set of 212 Area IDPs. Together with results from a simulation study, MTAFS shows its advantage over existing multi-trait methods, with robust performance across a range of underlying settings. It controls type 1 error well, and can efficiently handle a large number of traits.

Keywords: GWAS summary statistics, multiple traits, adaptive test, deep phenotyping data

1 Introduction

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with complex diseases (Visscher et al. 2017). However, for many complex traits, the heritability attributed to the genetic variants identified is still quite limited and a large proportion of the heritability remains unexplained (Manolio et al. 2009; Visscher et al. 2017). In GWAS, it is typical to test the association between a single trait and a single variant one at a time, the so-called single-trait analysis. In reality, a common phenomenon is pleiotropy, in which a genetic variant is associated with multiple traits (Solovieff et al. 2013). As such, conducting single-trait analyses may lose statistical power when genetic variants are truly associated with multiple traits. Therefore, there is an increasing need for methods that jointly analyze multiple traits together.

Although there are numerous existing multi-trait methods, many require individual-level genotype data (O’Reilly et al. 2012; Wu and Pankow 2016; Zhang et al. 2014). Due to privacy concern and data logistics, individual-level genotype data require permissions for access, limiting the applicability of methods relying on such data. In contrast, GWAS summary statistics such as effect sizes,
standard errors, z-scores, and p-values, are publicly available for most published studies. With increasing availability of GWAS summary statistics, methods that only require such information for multi-trait analysis undoubtedly will see greater usage.

Although relatively limited compared to other types of data and methods, a number of multi-trait methods using only summary statistics have been proposed. We categorize them into two groups. The first group consists of non-adaptive method. The method using the sum of squared z scores, denoted as SSU, is a special case of SPU proposed by Pan (Pan, 2009; Kim et al., 2015). He et al. (He et al., 2013) proposed a method using the sum of z scores, called SUM. Zhu et al. (Zhu et al., 2015) proposed a meta analysis method, called HOM, that is particularly suited for testing homogeneous effects across all traits. A chi-squared test, metaMANOVA, was proposed by Xu et al. (Xu et al., 2003). Cauchy’s method proposed by Liu and Xie (Liu and Xie, 2020) provides a general way to combine dependent p-values after appropriate transformation. The second group contains adaptive methods, which evaluate evidence adaptively and are particularly suited for heterogeneous situations where not all traits are associated, nor in the same directions or with the same effect size. An early example is TATES (Van der Sluis et al., 2013) which combines the p-value of each trait to obtain an overall p-value and it has similar power performance to the SSU (Liu and Lin, 2019). aSPU adaptively combines powered score test statistics and requires permutation to obtain p-values (Kim et al., 2015). Another method, metaUSAT, adaptively combines SSU and metaMANOVA (Ray and Boehnke, 2018). MixAda (Liu and Lin, 2018) adaptively combines SUM and a squared score test statistic, which is less powerful than metaUSAT under many scenarios (Ray and Boehnke, 2018). On the other hand, MTAR is an adaptive principal component (PC)-based association test (Guo and Wu, 2019). Wu recently proposed aMAT, claimed to be feasible for any number of traits (Wu, 2020).

Increased availability of GWAS summary statistics in recent years further points to the need for considering many traits simultaneously without accessing raw data. For example, in recent years, the UK Biobank has made thousands of functional and structural brain imaging phenotypes available, thus, a joint analysis of a large number of such traits may help better understand the biological mechanism of complex brain functions and diseases (Bycroft et al., 2018; Elliott et al., 2018). However, many previous methods using summary statistics as discussed above have only explored settings with a small number of traits (Zhu et al., 2015; Ray and Boehnke, 2018; Guo and Wu, 2019), rendering their performance of analyzing a large number of traits unknown. Our preliminary simulation study indicates that methods such as SSU and aMAT are sensitive to sparsity of signals and underlying correlation structures, whereas metaUSAT and SSU may not control type 1 error well at small significance levels. Computational issues also exist in some methods: aSPU is extremely time-consuming when the significance level is small due to its use of permutations; metaUSAT also becomes time-consuming when the p-values are extremely small, and it may return invalid values when the number of traits is large (e.g. over 200). Therefore, there is a need for robust and computationally efficient methods for settings where the number of traits is large, in the hundreds.

In this paper, we take up this challenge and propose a Multi-Trait Adaptive Fisher method for Summary statistics (MTAFS), a computationally efficient and statistically powerful method. In particular, MTAFS has three advantages over many existing methods. First, it controls type 1 error well compared to SSU and metaUSAT, regardless of the number of traits and significance levels. Second, it is robust, with good statistical power under different sparse and dense scenarios. Third, it is computationally efficient compared to metaUSAT and aSPU by avoiding permutations and is feasible for settings with a large number of traits.
2 Method

2.1 Setup

Let \( Z = (z_1, \ldots, z_T)'_{(1 \times T)} \) be the GWAS summary statistics, the z scores, across \( T \) traits for a given SNP. Our goal is to test whether the SNP is associated with at least one of the \( T \) traits. Under the null hypothesis of no association between the SNP and any of the traits, we assume \( Z \sim \mathcal{N}(0, R) \). Here, \( R \) is referred to as the trait correlation matrix, and can be estimated by the sample correlation of \( Z \) based on the independent and identically distributed assumption across SNPs (Zhu et al., 2015). Linkage disequilibrium score regression (LDSC) is another option (Turley et al., 2018; Guo and Wu, 2019). We denote the estimated correlation matrix by either method as \( \hat{R} \). For computational efficiency, we used sample correlation estimates in the simulation studies and LDSC in the real data applications.

First, we use eigen-decomposition to decorrelate the z scores. Let \( \hat{R} = Q \Lambda Q' \), where the columns of \( Q \) are eigenvectors in decreasing order of their corresponding eigenvalues given in the corresponding diagonal elements of \( \Lambda \). We denote the proportion of variance explained by the first two eigenvalues as \( v_0 \% \). Then let \( v_1 \%, v_2 \%, \text{ and } v_3 \% \) be the three percentages evenly distributed between \( v_0 \% \text{ and } 100\% \), with \( q_1, q_2, \text{ and } q_3 \) denoting the corresponding number of eigenvalues achieving the percent of variance explained for the first time.

For each of the 5 levels of percentage of variance explained, we use the corresponding \( E \) eigenvalues, \( E \in \{2, q_1, q_2, q_3, T\} \), along with their eigenvectors to construct the transformed z score vector \( U_E: U_E' = Z'Q\hat{E}_{EE}\Lambda_{EE}^{-\frac{1}{2}} \), where \( Q_{EE} \) consists of the first \( E \) columns of \( Q \) and \( \Lambda_{EE} \) is a submatrix of \( \Lambda \) containing only the first \( E \) eigenvalues. As a result, \( U_E \) is a column vector of length \( E \), and \( U_E \sim \mathcal{N}(0, I) \) under the null hypothesis. We then propose an adaptive method, as described in the following, in the spirit of the adaptive Fisher’s method (Song et al., 2016) for each of the five levels of variance explained. The resulting five p-values are then combined to construct an omnibus test statistic based on our proposed MTAFS method; the steps are depicted in a flow chart (supplementary Figure S1).

2.2 Adaptive Method

Unlike the traditional Fisher’s method which directly combines the \((-\log)\)-transformed p-values, the adaptive Fisher’s method considers ordered p-values and combines them adaptively (Song et al., 2016). The method we are proposing here also considered ordered p-values, but they are combined adaptively using a different strategy for computational efficiency. Specifically, based on an \( U_E \), we obtain a vector of independent (two-sided) p-values, denoted as \( \mathbf{p}_E = (p_1, \ldots, p_E) \), such that \( p_E = 2[1 - \Phi(|U_E|)] \), where \( \Phi(\cdot) \) is the cumulative distribution of standard normal distribution, and is a component-wise operation. We calculate the sum of the ordered negative log p-values and let \( s_k = \sum_{j=1}^{k} -(\log p(j)) \), where \( p(j) \) is the \( j \)th smallest p-value and \( k \in \{1, \ldots, E\} \). We can rewrite \( s_k \) as a weighted sum of independent \( \chi^2 \) variables (David and Nagaraja, 2004; Nagaraja, 2006), for which Davies method (R package CompQuadForm) or the saddlepoint approximation method (R package Survey) can efficiently approximate its p-value (Wu et al., 2016), denoted as \( p_k \). We define the test statistic of our adaptive method for level \( E \) as follows:

\[
AF(E) = Cauchy(p_{sk}; k = 1, \ldots, E) = \sum_{k=1}^{E} \omega_k \tan\{(0.5 - p_{sk})\pi\}, \tag{1}
\]

where \( \omega_k = \frac{1}{E} \) for all \( k \)'s. This way of combining the evidence from p-values follows what was referred to as the Cauchy’s method in the literature (Liu and Xie, 2020), and the p-value of the
test statistic can be calculated analytically:

\[ p_{AF(E)} = 0.5 - \frac{\arctan(AF(E))}{\pi}. \]  

(2)

We note that Cauchy’s method is similar to the minP method because only a few of the smallest p-values would typically dominate the overall significance. Nevertheless, since the p-values are calculated analytically, Cauchy’s method is much more computationally efficient than minP.

2.3 MTAFS

From the literature [Aschard et al. 2014] and our own preliminary study (Figure S4-S6), it is shown that using either the first few or all eigenvectors would lead to unstable power performance. Therefore, we propose MTAFS, which integrates evidence from five levels of variance explained, for robust consideration. Specifically, MTAFS constructs a test statistic that combines the \( \{p_{AF(E)}, E \in \{2, q_1, q_2, q_3, T\}\} \) obtained from Equation (2) for each of the 5 levels of variance explained. We define the test statistics of MTAFS as

\[ T_{MTAFS} = \text{Cauchy}(p_{AF(E)}; E \in \{2, q_1, q_2, q_3, T\}) = \sum_{E \in \{2, q_1, q_2, q_3, T\}} \omega_E \tan \{(0.5 - p_{AF(E)})\pi\}, \]  

(3)

where \( \omega_E = \frac{1}{5} \) for all \( E \)'s. As described above, the p-value of \( T_{MTAFS} \) is

\[ p_{MTAFS} = 0.5 - \frac{\arctan(T_{MTAFS})}{\pi}. \]

Since we have vectorized the R function of MTAFS, it can simultaneously analyze a large number of SNPs without using the “for” loop, which further increases its computational efficiency. MTAFS is implemented in an R package available at [http://www.github.com/Qiaolan/MTAFS](http://www.github.com/Qiaolan/MTAFS).

3 Simulations and results

3.1 Simulation Setup

We simulated z scores from \( \mathcal{N}(\mu, R) \) following previous studies [Guo and Wu 2019; Liu and Xie 2020; Wu 2020]. Various scenarios were constructed by setting different correlation matrices, association models and strengths, and levels of signal sparsity. For \( R \), we considered two realistic correlation matrices estimated from real data and two commonly used structures. Specifically, we used the UK Biobank brain image-derived phenotypes (IDPs): the set of 58 volumetric IDPs, with the resulting estimated correlation matrix referred to as UKCOR1 (Figure S2); and the T1 FAST region of interests containing 139 IDPs, denoted as UKCOR2 its correlation matrix (Figure S3). Moreover, we also examined two commonly-used correlation structures, compound symmetry (CS) and autocorrelation structure of order 1 (AR), each with two levels of correlation — weak (0.3) or strong (0.7). This leads to a total of 6 correlation matrices (Table 1). For analyzing data simulated, we re-estimated the correlation matrix instead of using the one for simulating the data.

We considered using two association models. In model 1, denoted as M1, we generated \( \mu = \sum_{j=1}^{J} c\lambda_j u_j \), where \( c \) is the parameter denoting the effect size, \( \lambda_j \) and \( u_j \) are the \( j^{th} \) eigenvalue and eigenvector of \( R \) respectively, and \( J \) represents the top \( J \) eigenvectors. We simulated different level of sparsity by varying \( J \) (Table 1). This association model was also simulated in other studies.
In the second association model (M2), we generated scenarios by directly setting some elements of \( \mu \) to be nonzero, with fewer non-zeros denoting greater sparsity. We note that when \( c = 0 \) in M1 or all elements of \( \mu \) are 0 in M2, we are in fact investigating the type 1 error.

Finally, we considered three levels of sparsity, high, intermediate, and low. For the highly sparse scenarios, in M1, only the top 2% or 5% of the eigenvectors had nonzero effect sizes, depending on the correlation structures; in M2, either 2%, 4%, or 5% of the traits had nonzero effect sizes, also depending on the correlation structures. The proportion of nonzero effect sizes was 20% in both models for the intermediate level of sparsity. The low sparsity scenarios had the proportion equal to 50% in both models. The specific eigenvectors (for M1) or the specific traits (for M2) that correspond to an nonzero effect size are given in Table 1.

We include eight competing methods in our simulation study for comparison with MTAFS: SUM, SSU, metaUSAT, metaMANOVA, Cauchy, HOM, MTAR, and aMAT. These are representatives of currently available multi-trait methods using summary statistics.

### 3.2 Type 1 errors

We first evaluated the type 1 error of MTAFS and the comparison methods at various significance levels, from \( 5 \times 10^{-2} \) to \( 1 \times 10^{-5} \), for multiple correlation matrices. Table 2 shows the results for UKCOR1. The estimated correlation matrix is obtained using the sample correlation over \( 10^5 \) replicates. One can see that all the methods except SSU and metaUSAT controlled type 1 error well. For SSU and metaUSAT, their type 1 errors are inflated when the significance levels are smaller (bolded values in Table 2). Because metaUSAT adaptively combines metaMANOVA and SSU, the inflation of metaUSAT could be caused by SSU. On the other hand, we see that Cauchy and MTAFS controlled type 1 error better with increasing significance level, consistent with a previous study (Liu and Xie 2020). Since MTAFS used Cauchy to combine p-values, MTAFS naturally shared the characteristics of Cauchy. We evaluated the type 1 error with UKCOR2 and observed similar findings (Table S1). We still observed type 1 error inflation for SUM and metaUSAT with the CS and AR correlation structure under either weak or strong correlation, although the magnitude were not as severe (Table S2-S9).

### 3.3 Power comparisons

For power comparisons, we simulated 1000 z scores and the significance level was set to be \( 5 \times 10^{-5} \). First, we evaluated the power of the different methods with UKCOR1. For the association model M1 (Figure 1), when only the top two eigenvectors were informative, SUM and SSU were the most powerful methods, followed by metaUSAT and MTAFS (Figure 1a). As more eigenvectors become informative, the power of SUM decreased, while SSU, metaUSAT, and MTAFS continue to perform well, and aMAT also joined this group for the less sparse scenarios (Figure 1(b,c)). Considering the type 1 error inflation of SSU and metaUSAT, receiver operating characteristic (ROC) curves (with a particular effect size for each of the three sparsity settings) restricted to a small type 1 error range were used to measure the performance of the top 4 methods in each sparsity level, for a fairer comparison of power (Figure 1(d-f)). Due to the inflated type 1 error of SSU and metaUSAT, they in fact have smaller power compared to SUM and MTAFS when the empirical type 1 errors are the same at a very small level, especially with less sparse scenarios (Figure 1(e-f)). We note that HOM had no power at all three sparsity levels, an observation consistent with previous studies (Wu 2020).

For M2 with UKCOR1, MTAFS was seen to be the most powerful methods at all three sparsity
levels (Figure 2). It is interesting to see that, other than MTAFS, the other methods have unstable performance, depending on the sparsity levels. For example, Cauchy was competitive in the high sparsity setting, but its power dropped down to zero at intermediate and low sparsity levels. Comparing across models M1 and M2, we see that SSU was among the powerful for M1, but its power dropped down to zero for M2. Whereas MTAFS performs well consistently across the association models, effect sizes, and sparsity levels.

Next, we compare the results when using the correlation matrix UKCOR2 (Figure S7, S8). For both M1 and M2, the results were similar to those for UKCOR1. Considering all the results together, the main qualitative observation for UKCOR1 remains the same for the UKCOR2 correlation matrix: the performance of the other methods are unstable, and MTAFS is extremely consistent across all settings and was always among the top performers, whereas all the other methods are less stable.

For the CS covariance matrix with model M2 (Figure S9, S10), MTAFS remains among the group of most powerful methods. This is also true for M2 with the AR structure (Figure S11, S12, except that in the high sparsity setting, Cauchy outperformed all other methods by a large margin. Considering all results from the simulation study with two different association models, effect sizes, sparsity levels, and covariance structures. It is clear that MTAFS is the most robust method. Although metaUSAT is also among the leaders in all settings in terms of power, we would argue that MTAFS is preferred since its type 1 error is well controlled while metaUSAT has been seen to have severely inflated type 1 error in some settings. Further, MTAFS may outperform metaUSAT in some scenarios (Figure S8), whereas MTAFS was never greatly outperformed by metaUSAT.

4 Real data application

4.1 Data and pre-processing

Regional brain morphology such as surface area and thickness of the cerebral cortex, and volume of subcortical structures has a complex genetic architecture involving many common genetic variants with small effect sizes and the strongly overlapped genetic architectures of sets of regional brain features (van der Meer et al., 2020). van der Meer et al. (2020) applied an multi-trait method to 171 regional brain morphology measures and identified much more significant SNPs than the single-trait analysis, suggesting that multi-trait analysis of regional measures can be powerful to discover genetic variants. Also, brain imaging data (e.g., functional magnetic resonance imaging (fMRI) data) have been proved useful for investigating connections between brain function and genetics (Liu et al., 2009).

UK Biobank is a rich and long-term prospective epidemiological study of 500,000 volunteers (Sudlow et al., 2015). Participants were 40–69 years old at recruitment, with one aim being to acquire as rich data as possible before disease onset. Elliott et al. (2018) investigated the genetic architecture of brain structure and function by conducting GWAS of 3,144 functional and structural brain imaging phenotypes from the UK Biobank (http://big.stats.ox.ac.uk/), which cover the entire brain and including multimodal information on grey matter volume, area and thickness, white matter connections and functional connectivity. The single-trait analyses were mainly applied in the study, thus we would like to apply our multi-trait method to potentially discovery more genetic variants. We carried out two multi-trait analyses, one with a moderate number of traits: 58 Volumetric IDPs, and one with a large number of traits: 212 Area IDPs of grey matter (Figure S13).

The summary statistics included the z scores from measuring the associations between each of the 11,734,353 SNPs and each of the 58 or 212 IDPs. LDSC was applied to estimate the volume
and the area IDP correlation matrices. To obtain independent SNPs to satisfy the assumption of our method, we applied LD clumping to remove SNPs whose correlation with index SNPs were above 0.2 in each window of 250kb, leading to 593,416 SNPs remaining. MTAFS and a subset of the competing methods (those performed well in some settings in the simulation study) were applied to identify significant SNPs that are associated with at least one IDP in each of the two sets of traits. We used a genome-wide significance threshold of $5 \times 10^{-8}$ for each of the multi-trait analysis methods. The genes corresponding to the significant SNPs were identified using NCBI dbSNP (Sayers et al., 2021). To investigate gene annotations, we used Functional Mapping and Annotation (FUMA) (Watanabe et al., 2017) to show tissue specific expression patterns of genes identified by MTAFS and other methods.

4.2 Results of 58 Volumetric IDPs

MTAFS identified 264 SNPs with p-values less than $5 \times 10^{-8}$ (Figure 3a), followed by metaMANOVA with 90 SNPs (Table S10). The rest of the methods (metaUSAT, aMAT, MTAR) identified even fewer SNPs (Figure S14). We also carried out single-trait analysis as a comparison, which identified only 6 significant SNPs (at the significance level of $5 \times 10^{-8}/58$), all of which were also identified by each of the multi-trait methods.

Many of the unique genes found by MTAFS, including ATP8A2, DPP6, ERBB4, and GRID2, have been reported previously to be associated with brain structure and function. Several studies showed that ATP8A2 was closely related to cerebellum, and its mutation could cause cognitive impairment and intellectual disability (Martín-Hernández et al., 2016; McMillan et al., 2018; Onat et al., 2013). DPP6 has been reported to be associated with human neural diseases (Cacace et al., 2019; Clark et al., 2008) and thalamus volume (Alliey-Rodriguez et al., 2019). A knockout of it in mice led to impaired hippocampal-dependent learning and memory and smaller brain size (Lin et al., 2020). ERBB4 is a candidate risk gene for schizophrenia (Silberberg et al., 2006; Law et al., 2007) and an essential regulator of central neural system (Gassmann et al., 1995). It was also reported to be associated with total intracranial volume (Alliey-Rodriguez et al., 2019). GRID2 is known to be differentially expressed in Purkinje cells in the cerebellum and the deletion of it causes cerebellar ataxia (Hills et al., 2013; Van Schil et al., 2015). PAPPA was also found by both single-trait analysis and MTAFS. It was reported to be associated with brain region volumes in previous studies (Elliott et al., 2018; Zhao et al., 2019).

To further investigate the biological mechanism, we used FUMA to annotate the genes identified in terms of biological context. Figure 3c shows the gene expression heatmap of significant genes found by MTAFS. The expression value depends on the genotype-tissue expression (GTEx) project (Lonsdale et al., 2013) including 54 human tissues. There were 14 tissues specifically related to brain such as amygdala and caudate basal ganglia. There was a cluster of genes close to the top left with higher relative expression; this cluster includes 13 of the 14 brain-related tissues. In FUMA, we also tested if the gene set was significantly enriched in tissues. Especially, we were interested in whether the set of genes uniquely identified by MTAFS are biologically relevant. Figure 3c shows that those genes were enriched significantly in most brain-related tissues (red bars in the top plot showing up-regulation). In contrast, we found the gene set consisting of genes identified by the comparison methods was not significantly enriched in any of the brain-related tissues (Figure S16).

4.3 Results of 212 Area IDPs

This analysis considered a much larger set of traits. In this case, our preliminary analysis found that metaUSAT has numerical issues; thus, it was excluded from consideration. MTAFS identified 55
SNPs with p-values less than $5 \times 10^{-8}$ (Figure 4a). On the other hand, metaMANOVA, aMAT, and MTAR had identified smaller sets of similar number of SNPs (Figure S17). Single-trait analysis only identified 1 SNP (Table S10), and that SNP was identified by all multi-trait methods, indicating that multi-trait methods were more powerful than single-trait analysis in real data applications, mostly because they are less conservative compared to Bonferroni correction for the number of SNPs × traits combinations.

Among the genes corresponding to significant SNPs identified by MTAFS only, we found that two genes, DPP6 and LINC02210-CRHR1, were previously identified in our first analysis with 58 Volumetric IDPs. We further investigated LINC02210-CRHR1 and found that it was reported to be associated with several brain structures and functions. (Zhao et al., 2019) reported that LINC02210-CRHR1 was significantly associated with brain volume, and (Hibar et al., 2015) found that it was associated specifically with subcortical brain region volumes. Two recent studies showed its relevance in the cortical surface area (Shin et al., 2020; Grasby et al., 2020). Several genes identified by MTAFS, such as C16orf95 and DGKI, also appeared in other studies using the UK Biobank data: van der Meer et al. (2020) analyzed the structural brain imaging data and identified genes C16orf95, DGKI, SYT1, and VCAN; Hofer et al. (2020) conducted association studies of brain cortical thickness, surface area, and volume, and they also identified gene C16orf95, DAAM1, NR2F1, NSF, and VCAN.

In the expression heatmap (Figure 4c), it shows a cluster of genes that had higher relative expression in the brain-related tissues than other tissues (the cluster locating at the same position as in the Volume analysis). In particular, we saw that PHACTR3 was highly expressed in all brain-related tissues. Many studies showed that this gene is important in intelligence, cognitive function, and schizophrenia (Goes et al., 2015; Turley et al., ???; Davies et al., 2018; Hill et al., 2019). Figure 4b shows that the gene set consisting of genes identified by MTAFS were significantly enriched in brain three tissues substantia nigra, cortex, and anterior cingulate cortex. In contrast, the gene set consisting of the genes identified by only the other methods was not significantly enriched in any brain tissues (Figure S19), although the expression level in a small cluster had relatively higher expression in brain-related tissues (Figure S18).

5 Discussion

GWAS have successfully identified a large number of genetic variants associated with traits or diseases. However, for many traits, a large portion of the heritability is still unaccounted for. In contrast to individual-level data, GWAS summary statistics are usually publicly available and have more potentials for achieving greater statistical power through combining a large amount of information. Our method utilizes z scores which are usually available in GWAS summary statistics along with their p-values. In rare cases where only p-values are available, we can transform the p-values to z scores by using the normality assumption. Although methods are available for joint analyses of a large number of traits from deep phenotyping data, inconsistent performance, computational inefficiency, and numerical issues when a large number of traits is considered are issues that are yet to be resolved. Our proposed MTAFS is an attempt in this direction. Our simulation study shows that MTAFS can control type 1 error well and has consistent performance under a variety of settings, underscoring its robustness. In real data applications, we see that, in contrast to single-trait analysis, MTAFS identified many more significant SNPs without omitting any detected by the former. Further, MTAFS identified more significant SNPs than the existing multi-trait analysis methods, and the genes identified by MTAFS are supported by evidence in the literature. Moreover, the expression of the gene set identified by MTAFS are more highly
expressed in a biologically relevant manner. In contrast, the expression of gene sets identified by
the existing methods do not lead to significant enrichment in brain tissues. Taken together, the
two analyses show the power of MTAFS and provide some insights on genes that may be related
to brain Volumetric and Area IDPs.

In general, MTAFS exhibits desirable properties, and have several advantages over existing
methods as a whole. First, MTAFS controls type 1 error well, even with small significance levels.
Second, MTAFS has robust performance given various correlation matrices, underlying association
models, and different levels of signal sparsity. Third, MTAFS is an efficient method in practice,
though it is not as computationally efficient as some existing methods (Table S12). It is much faster
than methods using permutation tests like minP and aSPU. For example, it took about 2 hours
to analyze 593,416 SNPs for 58 Volumetric IDPs with a single core of 4GB memory. Moreover,
parallel computing can greatly reduce its computational time making it acceptable in practice. As
a demonstration, we analyzed the Area IDP data for 593,416 SNPs and 212 traits, and MTAFS
finished the analysis in only 10 minutes by using 60 cores of 4GB memory.

The advantages notwithstanding, there are limitations of the proposed method. First, because
we transform raw z score vectors by eigendecomposition, it is difficult to interpret the association
between one SNP and one single trait. Second, our choice of the levels of variance explained and
the number of levels are both ad hoc. Third, MTAFS currently only considers common variants;
thus further development is warranted for including rare ones.
| Correlation | # Traits | Association Models | Effect Sizes<sup>a</sup> | Settings<sup>b</sup> |
|-------------|----------|--------------------|--------------------------|---------------------|
| UKCOR1      | 58       | M1                 | [0.6, 1.6]               | J=2                 |
|             |          |                    | [0.6, 1.6]               | J=11                |
|             |          |                    | [0.6, 1.6]               | J=25                |
|             |          | M2                 | [3, 6]                   | \(\mu_{56} - \mu_{58}\) |
|             |          |                    | [0.07, 1.3]              | \(\mu_{46} - \mu_{58}\) |
|             |          |                    | [0.3, 0.7]               | \(\mu_{29} - \mu_{58}\) |
| UKCOR2      | 139      | M1                 | [0.5, 1.4]               | J=2                 |
|             |          |                    | [0.5, 1.1]               | J=27                |
|             |          |                    | [0.5, 1.1]               | J=69                |
|             |          | M2                 | [1.3, 2.5]               | \(\mu_{133}, \mu_{134}\) |
|             |          |                    | [1.5, 3]                 | \(\mu_{107} - \mu_{134}\) |
|             |          |                    | [1, 2]                   | \(\mu_{70} - \mu_{139}\) |
| CS(0.3)     | 50       | M2                 | [2, 6]                   | \(\mu_1, \mu_2\) |
|             |          |                    | [1, 3]                   | \(\mu_1 - \mu_{10}\) |
|             |          |                    | [4, 7]                   | \(\mu_1, \mu_2\) |
|             | 100      |                    | [1.5, 2.5]               | \(\mu_1 - \mu_{20}\) |
| CS(0.7)     | 50       | M2                 | [2, 4]                   | \(\mu_1, \mu_2\) |
|             |          |                    | [1, 2]                   | \(\mu_1 - \mu_{10}\) |
|             | 100      |                    | [2.5, 4.5]               | \(\mu_1, \mu_2\) |
|             |          |                    | [1, 1.5]                 | \(\mu_1 - \mu_{20}\) |
| AR(0.3)     | 50       | M2                 | [4, 7]                   | \(\mu_{25}, \mu_{26}\) |
|             |          |                    | [1, 4]                   | \(\mu_{20} - \mu_{30}\) |
|             | 100      |                    | [5, 8]                   | \(\mu_{50}, \mu_{51}\) |
|             |          |                    | [1, 2.5]                 | \(\mu_{40} - \mu_{60}\) |
| AR(0.7)     | 50       | M2                 | [2, 6]                   | \(\mu_{25}, \mu_{26}\) |
|             |          |                    | [2, 5]                   | \(\mu_{20} - \mu_{30}\) |
|             | 100      |                    | [4, 6.7]                 | \(\mu_{50}, \mu_{51}\) |
|             |          |                    | [2, 3.5]                 | \(\mu_{40} - \mu_{60}\) |

<sup>a</sup> For M1, the effect size refers to \(c\); for M2, it is the nonzero value of the \(\mu\) components. For all, 10 different effect sizes are considered, which are evenly distributed in the range specified, inclusive.

<sup>b</sup> For M1, the number specified is for \(J\), the number of eigenvectors having a nonzero effect; for M2, we list the range of the \(\mu\) components that have nonzero values, inclusive.
| Methods       | $5 \times 10^{-2}$ | $1 \times 10^{-2}$ | $1 \times 10^{-3}$ | $1 \times 10^{-4}$ | $1 \times 10^{-5}$ |
|--------------|------------------|------------------|-----------------|-----------------|-----------------|
| metaMANOVA   | 1.01             | 1.01             | 1.02            | 0.963           | 0.95            |
| metaUSAT     | 1                | 1.08             | 1.17            | **1.72**        | **1.87**        |
| SUM          | 0.92             | 1.01             | **1.34**        | 2               | **3.16**        |
| SSU          | 1.01             | 1                | 1.01            | 1.08            | 1.05            |
| HOM          | 1.14             | 1.13             | 1.07            | 1.03            | 1.03            |
| Cauchy       | 1.16             | 1.14             | 1.1             | 1.13            | 1.03            |
| aMAT         | 1.01             | 0.99             | 0.981           | 0.95            | 1.04            |
| MTAR         | 1.01             | 1.08             | 1.17            | 2               | **3.16**        |
| MTAFS        | 1.01             | 1.08             | 1.17            | **2**           | **3.16**        |

A The values in the table are ratios of empirical Type I errors divided by the corresponding significance levels. Values larger than 1.3 are bold (let $\alpha = 1 \times 10^{-5}$, $\alpha + 3\sqrt{\frac{\alpha(1-\alpha)}{10^{-\alpha}}} \approx 1.3 \times 10^{-5}$).
Figure 1: Comparison of methods for model M1 using the UKCOR1 correlation matrix. (a) high sparsity, with only top 2 eigenvectors informative; (b) intermediate sparsity, with top 11 eigenvectors informative; (c) low sparsity, with top 25 eigenvectors informative; (d) partial ROC curves for the four best methods with comparable power in (a); (e) partial ROC curves for the four best methods with comparable power in (b); (f) partial ROC curves for the four best methods with comparable power in (c).
Figure 2: Comparison of methods for model M2 using the UKCOR1 correlation matrix. (a) high sparsity, with only 3 nonzero components of $\mu$; (b) intermediate sparsity, with 13 nonzero components of $\mu$; (c) low sparsity, with 30 nonzero components of $\mu$ out of a total of 58; (d) partial ROC curves for the three best methods with comparable power in (a); (e) partial ROC curves for the three best methods with comparable power in (b); (f) partial ROC curves for the three best methods with comparable power in (c).
Figure 3: Analysis results of the 58 Volumetric IDPs. (a) Manhattan plot of the SNPs identified by MTAFS. For (b) and (c), we use the GTEx data over 54 tissue types. (b) Tissue expression analysis for genes uniquely identified by MTAFS for volume. Significant enrichment are in red with p-values less than 0.05 after Bonferroni correction; (c) The expression heatmap of all genes identified by MTAFS for volume. The red clusters at the top of the figure close to the left have higher relative expression.
Figure 4: Analysis results of the 212 Area IDPs. (a) Manhattan plot of the SNPs identified by MTAFS. For (b) and (c), we use the GTEx data over 54 tissue types. (b) Tissue expression analysis for genes uniquely identified by MTAFS for volume. Significant enrichment are in red with p-values less than 0.05 after Bonferroni correction; (c) The expression heatmap of all genes identified by MTAFS for volume. The red clusters have higher relative expression.
References

Alliey-Rodriguez, N. et al. (2019). Nrnx1 is associated with enlargement of the temporal horns of the lateral ventricles in psychosis. Translational psychiatry, 9(1), 1–7.

Aschard, H. et al. (2014). Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. The American Journal of Human Genetics, 94(5), 662–676.

Bycroft, C. et al. (2018). The uk biobank resource with deep phenotyping and genomic data. Nature, 562(7726), 203–209.

Cacace, R. et al. (2019). Loss of dpp6 in neurodegenerative dementia: a genetic player in the dysfunction of neuronal excitability. Acta neuropathologica, 137(6), 901–918.

Clark, B. D. et al. (2008). Dpp6 localization in brain supports function as a kv4 channel associated protein. Frontiers in molecular neuroscience, 1, 8.

David, H. A. and Nagaraja, H. N. (2004). Order statistics. John Wiley & Sons.

Davies, G. et al. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. Nature communications, 9(1), 1–16.

Elliott, L. T. et al. (2018). Genome-wide association studies of brain imaging phenotypes in uk biobank. Nature, 562(7726), 210–216.

Gassmann, M. et al. (1995). Aberrant neural and cardiac development in mice lacking the erbb4 neuregulin receptor. Nature, 378(6555), 390–394.

Goes, F. S. et al. (2015). Genome-wide association study of schizophrenia in ashkenazi jews. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 168(8), 649–659.

Grasby, K. L. et al. (2020). The genetic architecture of the human cerebral cortex. Science, 367(6484).

Guo, B. and Wu, B. (2019). Integrate multiple traits to detect novel trait–gene association using gwas summary data with an adaptive test approach. Bioinformatics, 35(13), 2251–2257.

He, Q. et al. (2013). A general framework for association tests with multivariate traits in large-scale genomics studies. Genetic epidemiology, 37(8), 759–767.

Hibar, D. P. et al. (2015). Common genetic variants influence human subcortical brain structures. Nature, 520(7546), 224–229.

Hill, W. D. et al. (2019). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. Molecular psychiatry, 24(2), 169–181.

Hills, L. B. et al. (2013). Deletions in grid2 lead to a recessive syndrome of cerebellar ataxia and tonic upgaze in humans. Neurology, 81(16), 1378–1386.

Hofer, E. et al. (2020). Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. Nature communications, 11(1), 1–16.
Kim, J. et al. (2015). An adaptive association test for multiple phenotypes with gwas summary statistics. Genetic epidemiology, 39(8), 651–663.

Law, A. J. et al. (2007). Disease-associated intronic variants in the erbb4 gene are related to altered erbb4 splice-variant expression in the brain in schizophrenia. Human molecular genetics, 16(2), 129–141.

Lin, L. et al. (2020). A novel structure associated with aging is augmented in the dpp6-ko mouse brain. Acta Neuropathologica Communications, 8(1), 1–18.

Liu, J. et al. (2009). Combining fmri and snp data to investigate connections between brain function and genetics using parallel ica. Human brain mapping, 30(1), 241–255.

Liu, Y. and Xie, J. (2020). Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. Journal of the American Statistical Association, 115(529), 393–402.

Liu, Z. and Lin, X. (2018). Multiple phenotype association tests using summary statistics in genome-wide association studies. Biometrics, 74(1), 165–175.

Liu, Z. and Lin, X. (2019). A geometric perspective on the power of principal component association tests in multiple phenotype studies. Journal of the American Statistical Association.

Lonsdale, J. et al. (2013). The genotype-tissue expression (gtex) project. Nature genetics, 45(6), 580–585.

Manolio, T. A. et al. (2009). Finding the missing heritability of complex diseases. Nature, 461(7265), 747–753.

Martín-Hernández, E. et al. (2016). New atp8a2 gene mutations associated with a novel syndrome: encephalopathy, intellectual disability, severe hypotonia, chorea and optic atrophy. Neurogenetics, 17(4), 259–263.

McMillan, H. J. et al. (2018). Recessive mutations in atp8a2 cause severe hypotonia, cognitive impairment, hyperkinetic movement disorders and progressive optic atrophy. Orphanet journal of rare diseases, 13(1), 1–10.

Nagaraja, H. N. (2006). Order statistics from independent exponential random variables and the sum of the top order statistics. In Advances in Distribution Theory, Order Statistics, and Inference, pages 173–185. Springer.

Onat, O. E. et al. (2013). Missense mutation in the atpase, aminophospholipid transporter protein atp8a2 is associated with cerebellar atrophy and quadrupedal locomotion. European Journal of Human Genetics, 21(3), 281–285.

O’Reilly, P. F. et al. (2012). Multiphen: joint model of multiple phenotypes can increase discovery in gwas. PloS one, 7(5), e34861.

Pan, W. (2009). Asymptotic tests of association with multiple snps in linkage disequilibrium. Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society, 33(6), 497–507.

Ray, D. and Boehnke, M. (2018). Methods for meta-analysis of multiple traits using gwas summary statistics. Genetic epidemiology, 42(2), 134–145.
Sayers, E. W. et al. (2021). Database resources of the national center for biotechnology information. *Nucleic acids research*, 49(D1), D10.

Shin, J. et al. (2020). Global and regional development of the human cerebral cortex: Molecular architecture and occupational aptitudes. *Cerebral Cortex*, 30(7), 4121–4139.

Silberberg, G. et al. (2006). The involvement of erbb4 with schizophrenia: association and expression studies. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 141(2), 142–148.

Solovieff, N. et al. (2013). Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics*, 14(7), 483–495.

Song, C. et al. (2016). The screening and ranking algorithm for change-points detection in multiple samples. *The annals of applied statistics*, 10(4), 2102.

Sudlow, C. et al. (2015). Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLOS medicine*, 12(3), e1001779.

Turley, P. et al. (2018). Multi-trait analysis of genome-wide association summary statistics using mtag. *Nature genetics*, 50(2), 229–237.

Turley16, P. et al. (????). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals.

van der Meer, D. et al. (2020). Understanding the genetic determinants of the brain with mostest. *Nature communications*, 11(1), 1–9.

Van der Sluis, S. et al. (2013). Tates: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS genetics*, 9(1), e1003235.

Van Schil, K. et al. (2015). Early-onset autosomal recessive cerebellar ataxia associated with retinal dystrophy: new human hotfoot phenotype caused by homozygous grid2 deletion. *Genetics in Medicine*, 17(4), 291–299.

Visscher, P. M. et al. (2017). 10 years of gwas discovery: biology, function, and translation. *The American Journal of Human Genetics*, 101(1), 5–22.

Watanabe, K. et al. (2017). Functional mapping and annotation of genetic associations with fuma. *Nature communications*, 8(1), 1–11.

Wu, B. and Pankow, J. S. (2016). Sequence kernel association test of multiple continuous phenotypes. *Genetic epidemiology*, 40(2), 91–100.

Wu, B. et al. (2016). On efficient and accurate calculation of significance p-values for sequence kernel association testing of variant set. *Annals of human genetics*, 80(2), 123–135.

Wu, C. (2020). Multi-trait genome-wide analyses of the brain imaging phenotypes in uk biobank. *Genetics*, 215(4), 947–958.

Xu, X. et al. (2003). Combining dependent tests for linkage or association across multiple phenotypic traits. *Biostatistics*, 4(2), 223–229.
Zhang, Y. et al. (2014). Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *NeuroImage*, 96, 309–325.

Zhao, B. et al. (2019). Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nature genetics*, 51(11), 1637–1644.

Zhu, X. et al. (2015). Meta-analysis of correlated traits via summary statistics from gwass with an application in hypertension. *The American Journal of Human Genetics*, 96(1), 21–36.
An Adaptive and Robust Method for Multi-trait Analysis of Genome-wide Association Studies Using Summary Statistics
Supplementary Material

Qiaolan Deng,\textsuperscript{1,3} Chi Song,\textsuperscript{2} and Shili Lin\textsuperscript{1,3}

\textsuperscript{1}Interdisciplinary Ph.D. Program in Biostatistics
\textsuperscript{2}Division of Biostatistics, College of Public Health
\textsuperscript{3}Department of Statistics, The Ohio State University, Columbus, Ohio
Figure S1: workflow of MTAFS

Table S1: Type 1 error\textsuperscript{a} with UKCOR2

| Methods   | $5 \times 10^{-2}$ | $1 \times 10^{-2}$ | $1 \times 10^{-3}$ | $1 \times 10^{-4}$ | $1 \times 10^{-5}$ |
|-----------|-------------------|-------------------|-------------------|-------------------|-------------------|
| metaMANOVA| 1.03              | 1.03              | 1.05              | 1.014             | 0.93              |
| metaUSAT  | 1                 | 1.1               | \textbf{1.22}     | \textbf{1.79}     | \textbf{1.95}     |
| SUM       | 1                 | 1                | 1.01              | 1.06              | 1.01              |
| SSU       | 0.94              | 1.04             | \textbf{1.41}     | \textbf{1.98}     | \textbf{3.25}     |
| HOM       | 1                 | 1.01             | 1.02              | 1.04              | 1.06              |
| Cauchy    | 1.14              | 1.1               | 1.05              | 1.06              | 1.15              |
| aMAT      | 0.97              | 0.97             | 0.99              | 1                 | 1.17              |
| MTAR      | 1.02              | 1.03             | 0.99              | 1.062             | 0.99              |
| MTAFS     | 1.19              | 1.16             | 1.1               | 1.04              | 0.95              |

\textsuperscript{a} Values are ratios of empirical Type I errors divided by the corresponding significance levels. Inflated values are bold.

Figure S17: Venn diagram of number of significantly associated SNPs for Area identified by different methods at $5 \times 10^{-8}$. 
Figure S2: The LDSC estimated trait correlation matrix of Volume. Volume consists of 58 IDPs.
Figure S3: The LDSC estimated trait correlation matrix of T1FAST. T1FAST consists of 138 IDPs
Figure S4: Power comparisons with UKCOR1 and M1. The first eigenvector explained at least 30%, thus the lines for 10% and 20% were excluded in the plots. (a) In the sparse scenario, the model using eigenvectors which explained 40% of variance gave the maximum power. (b) In the intermediate scenario, including eigenvectors which explained 70% of variance gave the maximum power. (c) In the dense scenario, including eigenvectors corresponding to 80% or 90% of variance gave comparable power.
Figure S5: Power comparisons with UKCOR2 and M1. (a) In the sparse scenario, the model using only the first eigenvector (10% line) was not the most powerful. (b) In the intermediate scenario, using eigenvectors which explained 50% or 60% of variance gave the maximum power. (c) In the dense scenario, including eigenvectors corresponding to 80% of variance gave the maximal power.
Figure S6: Power comparisons with UKCOR2 and M1. (a) and (b) has CS(0.3), and (c) and (d) has CS(0.7). Regardless of sparse ((a) and (c)) or dense scenarios ((b) and (d)), including all eigenvectors had the maximal power.
Table S2: Type 1 error$^a$ with correlation matrix CS(0.3) and 50 traits

| Methods     | $5 \times 10^{-2}$ | $1 \times 10^{-2}$ | $1 \times 10^{-3}$ | $1 \times 10^{-4}$ | $1 \times 10^{-5}$ |
|-------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| metaMANOVA  | 1.01              | 1.01              | 1.01              | 1.02              | 1.17              |
| metaUSAT    | 1                 | 1.06              | 1.09              | **1.48**          | **1.66**          |
| SUM         | 1                 | 1                 | 1                 | 1.02              | 1.16              |
| SSU         | 0.96              | 1.01              | 1.13              | **1.3**           | **1.76**          |
| HOM         | 1                 | 1                 | 1                 | 1                 | 1.16              |
| Cauchy      | 1.27              | 1.23              | 1.1               | 1.02              | 1.02              |
| aMAT        | 0.99              | 1                 | 0.99              | 1.05              | 1.2               |
| MTAR        | 1                 | 0.99              | 0.97              | 1.01              | 0.91              |
| MTAFS       | 1.05              | 0.99              | 0.91              | 0.87              | 1.03              |

$^a$ The values in the table are ratios of empirical Type I errors divided by the corresponding significance levels. Values larger than 1.3 are bold.

Table S3: Type 1 error$^a$ with correlation matrix CS(0.3) and 100 traits

| Methods     | $5 \times 10^{-2}$ | $1 \times 10^{-2}$ | $1 \times 10^{-3}$ | $1 \times 10^{-4}$ | $1 \times 10^{-5}$ |
|-------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| metaMANOVA  | 1.01              | 1.02              | 1.04              | 1.09              | 1.23              |
| metaUSAT    | 1.02              | 1.06              | 1.02              | **1.38**          | **1.52**          |
| SUM         | 1                 | 1                 | 1.01              | 1.02              | 0.97              |
| SSU         | 0.97              | 1                 | 1.08              | 1.2               | **1.33**          |
| HOM         | 1                 | 1                 | 1.02              | 1.03              | 0.96              |
| Cauchy      | **1.33**          | 1.3               | 1.11              | 1.04              | 1.12              |
| aMAT        | 1                 | 1.01              | 1.03              | 1.06              | 1.1               |
| MTAR        | 1.01              | 0.99              | 1.01              | 1.07              | 0.88              |
| MTAFS       | 0.97              | 0.89              | 0.81              | 0.75              | 0.73              |

$^a$ The values in the table are ratios of empirical Type I errors divided by the corresponding significance levels. Values larger than 1.3 are bold.

Table S4: Type 1 error$^a$ with correlation matrix CS(0.7) and 50 traits

| Methods     | $5 \times 10^{-2}$ | $1 \times 10^{-2}$ | $1 \times 10^{-3}$ | $1 \times 10^{-4}$ | $1 \times 10^{-5}$ |
|-------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| metaMANOVA  | 1.01              | 1.01              | 1.03              | 1.08              | 1.14              |
| metaUSAT    | 1.02              | 1.05              | 1.03              | 1.19              | **1.41**          |
| SUM         | 1                 | 1                 | 1                 | 1.01              | 1.06              |
| SSU         | 1                 | 1                 | 1                 | 1.03              | 1.06              |
| HOM         | 1                 | 1                 | 1                 | 1.01              | 1.05              |
| Cauchy      | 1.26              | 1.28              | 1.23              | 1.18              | 1.15              |
| aMAT        | 1                 | 1                 | 1                 | 1.01              | 1.06              |
| MTAR        | 1                 | 0.98              | 1.03              | 1                 | 0.86              |
| MTAFS       | 1.05              | 0.99              | 0.93              | 0.86              | 0.9               |

$^a$ The values in the table are ratios of empirical Type I errors divided by the corresponding significance levels. Values larger than 1.3 are bold.
### Table S5: Type 1 error\(^a\) with correlation matrix CS(0.7) and 100 traits

| Methods     | \(5 \times 10^{-2}\) | \(1 \times 10^{-2}\) | \(1 \times 10^{-3}\) | \(1 \times 10^{-4}\) | \(1 \times 10^{-5}\) |
|-------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| metaMANOVA  | 1.03                   | 1.04                   | 1.06                   | 1.1                    | 1.09                   |
| metaUSAT    | 1.04                   | 1.04                   | 0.93                   | 1.16                   | \textbf{1.35}          |
| SUM         | 1                      | 1                      | 1.01                   | 0.97                   | 1.05                   |
| SSU         | 1                      | 1                      | 1.02                   | 0.97                   | 1.06                   |
| HOM         | 1                      | 1                      | 1.02                   | 0.97                   | 1.02                   |
| Cauchy      | 1.28                   | \textbf{1.31}          | 1.28                   | 1.22                   | \textbf{1.31}          |
| aMAT        | 1                      | 1                      | 0.99                   | 1.03                   | 1.02                   |
| MTAR        | 1                      | 1                      | 0.97                   | 1.03                   | 1.02                   |
| MTAFS       | 0.97                   | 0.89                   | 0.81                   | 0.77                   | 0.81                   |

\(^a\) The values in the table are ratios of empirical Type I errors divided by the corresponding significance levels. Values larger than 1.3 are bold.

### Table S6: Type 1 error\(^a\) with correlation matrix AR(0.3) and 50 traits

| Methods     | \(5 \times 10^{-2}\) | \(1 \times 10^{-2}\) | \(1 \times 10^{-3}\) | \(1 \times 10^{-4}\) | \(1 \times 10^{-5}\) |
|-------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| metaMANOVA  | 1.01                   | 1.01                   | 1.01                   | 1.02                   | 1.04                   |
| metaUSAT    | 0.72                   | 0.77                   | 0.83                   | 0.96                   | 1.06                   |
| SUM         | 1                      | 1                      | 0.99                   | 0.96                   | 1.18                   |
| SSU         | 0.99                   | 1.01                   | 1.06                   | 1.15                   | \textbf{1.36}          |
| HOM         | 1                      | 1                      | 0.99                   | 0.99                   | 1.22                   |
| Cauchy      | 1.02                   | 1.01                   | 1.01                   | 1.06                   | 0.93                   |
| aMAT        | 0.98                   | 0.98                   | 0.97                   | 0.96                   | 1.01                   |
| MTAR        | 1                      | 1                      | 1.04                   | 0.99                   | 1                      |
| MTAFS       | 1.16                   | 1.14                   | 1.09                   | 1.06                   | 1.02                   |

\(^a\) The values in the table are ratios of empirical Type I errors divided by the corresponding significance levels. Values larger than 1.3 are bold.

### Table S7: Type 1 error\(^a\) with correlation matrix AR(0.3) and 100 traits

| Methods     | \(5 \times 10^{-2}\) | \(1 \times 10^{-2}\) | \(1 \times 10^{-3}\) | \(1 \times 10^{-4}\) | \(1 \times 10^{-5}\) |
|-------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| metaMANOVA  | 1.02                   | 1.02                   | 1.04                   | 1                      | 0.94                   |
| metaUSAT    | 0.72                   | 0.77                   | 0.83                   | 0.88                   | 0.94                   |
| SUM         | 1                      | 1                      | 1                      | 0.99                   | 1.06                   |
| SSU         | 1                      | 1                      | 1.02                   | 1.02                   | 1.06                   |
| HOM         | 1.01                   | 1                      | 1.02                   | 0.99                   | 1.08                   |
| Cauchy      | 1.02                   | 1.01                   | 1                      | 0.96                   | 1.1                    |
| aMAT        | 0.98                   | 0.97                   | 0.97                   | 0.94                   | 0.84                   |
| MTAR        | 1                      | 0.95                   | 1.15                   | 1.07                   | 1.1                    |
| MTAFS       | 1.17                   | 1.13                   | 1.07                   | 1.03                   | 1.02                   |

\(^a\) The values in the table are ratios of empirical Type I errors divided by the corresponding significance levels.
**Table S8:** Type 1 error<sup>a</sup> with correlation matrix AR(0.7) and 50 traits

| Methods    | $5 \times 10^{-2}$ | $1 \times 10^{-2}$ | $1 \times 10^{-3}$ | $1 \times 10^{-4}$ | $1 \times 10^{-5}$ |
|------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| metaMANOVA | 1.01                | 1.01                | 1.01                | 1                   | 1.08                |
| metaUSAT   | 0.93                | 1.03                | 1.14                | **1.43**            | **1.76**            |
| SUM        | 1                   | 1                   | 1                   | 1                   | 0.92                |
| SSU        | 0.98                | 1.02                | 1.15                | **1.37**            | **1.52**            |
| HOM        | 1                   | 1                   | 1                   | 0.98                | 1.04                |
| Cauchy     | 1.1                 | 1.09                | 1.05                | 1.04                | 0.98                |
| aMAT       | 0.96                | 0.96                | 0.98                | 1.03                | **1.29**            |
| MTAR       | 1                   | 1.01                | 1.02                | 1.11                | 0.95                |
| MTAFS      | 1.16                | 1.14                | 1.1                 | 1.06                | 1.13                |

<sup>a</sup> The values in the table are ratios of empirical Type I errors divided by the corresponding significance levels. Values larger than 1.3 are bold.

**Table S9:** Type 1 error<sup>a</sup> with correlation matrix AR(0.7) and 100 traits

| Methods    | $5 \times 10^{-2}$ | $1 \times 10^{-2}$ | $1 \times 10^{-3}$ | $1 \times 10^{-4}$ | $1 \times 10^{-5}$ |
|------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| metaMANOVA | 1.02                | 1.03                | 1.05                | 1.08                | 1.17                |
| metaUSAT   | 0.95                | 1.05                | 1.17                | **1.46**            | **1.73**            |
| SUM        | 1                   | 1                   | 0.98                | 0.95                | 1.02                |
| SSU        | 0.99                | 1.02                | 1.08                | 1.29                | **1.6**             |
| HOM        | 1                   | 1                   | 0.99                | 0.99                | 1.03                |
| Cauchy     | 1.09                | 1.07                | 1.04                | 1.03                | 1.18                |
| aMAT       | 0.98                | 0.98                | 0.99                | 1.05                | 1.19                |
| MTAR       | 1                   | 0.98                | 0.97                | 1.02                | 0.94                |
| MTAFS      | 1.18                | 1.15                | 1.1                 | 1.1                 | 0.94                |

<sup>a</sup> The values in the table are ratios of empirical Type I errors divided by the corresponding significance levels. Values larger than 1.3 are bold.
Figure S7: Comparison of methods for model M1 using the UKCOR2 correlation matrix. (a) high sparsity, with only top 2 eigenvectors informative; (b) intermediate sparsity, with top 27 eigenvectors informative; (c) low sparsity, with top 69 eigenvectors informative; (d) partial ROC curves for the four best methods with comparable power in (a); (e) partial ROC curves for the four best methods with comparable power in (b); (f) partial ROC curves for the four best methods with comparable power in (c).
Figure S8: Comparison of methods for model M2 using the UKCOR2 correlation matrix. (a) high sparsity, with only 2 nonzero components of $\mu$; (b) intermediate sparsity, with 28 nonzero components of $\mu$; (c) low sparsity, with 70 nonzero components of $\mu$ out of a total of 139; (d) partial ROC curves for the three best methods with comparable power in (a); (e) partial ROC curves for the four best methods with comparable power in (b); (f) partial ROC curves for the four best methods with comparable power in (c).
Figure S9: Comparison of methods for model M2 using the CS correlation matrix and 50 traits. (a) and (b) have CS(0.3), and (c) and (d) have CS(0.7). (a) and (c) have high sparsity, with only 2 nonzero components of $\mu$; (b) and (d) have intermediate sparsity, with 10 nonzero components of $\mu$ out of a total of 50 traits.
Figure S10: Comparison of methods for model M2 using the CS correlation matrix and 100 traits. (a) and (b) have CS(0.3), and (c) and (d) have CS(0.7). (a) and (c) have high sparsity, with only 2 nonzero components of $\mu$; (b) and (d) have intermediate sparsity, with 20 nonzero components of $\mu$ out of a total of 100 traits.
Figure S11: Comparison of methods for model M2 using the AR correlation matrix and 50 traits. (a) and (b) have AR(0.3), and (c) and (d) have AR(0.7). (a) and (c) have high sparsity, with only 2 nonzero components of $\mu$; (b) and (d) have intermediate sparsity, with 10 nonzero components of $\mu$ out of a total of 50 traits.
Figure S12: Comparison of methods for model M2 using the AR correlation matrix and 100 traits. (a) and (b) have AR(0.3), and (c) and (d) have AR(0.7). (a) and (c) have high sparsity, with only 2 nonzero components of $\mu$; (b) and (d) have intermediate sparsity, with 20 nonzero components of $\mu$ out of a total of 100 traits.
Figure S13: The LDSC estimated trait correlation matrix of Area. Area consists of 212 IDPs.
Figure S14: Venn diagram of number of significantly associated SNPs for Volume identified by different methods at $5 \times 10^{-8}$.

Figure S15: The expression heatmap of all genes identified by competing methods for volume. The red clusters have higher relative expression.
**Figure S16:** Tissue expression analysis for genes identified by competing methods for volume. Significant enrichment are in red with p-values less than 0.05 after Bonferroni correction.
Table S10: The number of significant SNPs identified by methods

| method     | MTAFS | metaMANOVA | metaUSAT | aMAT | MTAR | SUM | SSU | Single Trait |
|------------|-------|------------|----------|------|------|-----|-----|--------------|
| Volume     | 264   | 90         | 89       | 15   | 78   | 0   | 1   | 6            |
| Area       | 55    | 16         | -        | 11   | 14   | 0   | 3   | 1            |

Table S11: Type 1 error\(^a\) with Area trait correlation matrix

| Methods    | 5 × 10\(^{-2}\) | 1 × 10\(^{-2}\) | 1 × 10\(^{-3}\) | 1 × 10\(^{-4}\) | 1 × 10\(^{-5}\) |
|------------|----------------|----------------|----------------|----------------|----------------|
| metaMANOVA | 1.05           | 1.07           | 1.09           | 1.08           | 1.21           |
| aMAT       | 1.06           | 1.1            | 1.15           | 1.28           | 1.2            |
| MTAR       | 0.99           | 1              | 1.04           | 1.08           | 0.97           |
| MTAFS      | 1.2            | 1.16           | 1.11           | 1              | 1.07           |

\(^a\) The values in the table are ratios of empirical Type I errors divided by the corresponding significance levels.

Figure S18: The expression heatmap of all genes identified by competing methods for area. The red clusters have higher relative expression.
**Figure S19:** Tissue expression analysis for genes identified by competing methods for area. Significant enrichment are in red with p-values less than 0.05 after Bonferroni correction.

**Table S12:** Computing time of different methods (in seconds)

|                | MTAFS | metaMANOVA | metaUSAT | aMAT | MTAR |
|----------------|-------|------------|----------|------|------|
| 58 Volumetric IDPs | 7200  | 8.4        | 9000     | 540  | 0.48 |
| 212 Area IDPs    | 600   | 60         | -        | 960  | 2    |
| UKCOR1, M1, Power | 8.39  | 0.01       | 269      | 1.55 | 0.054|
| UKCOR1, M1, Type 1 | 13.93 | 0.01       | 12.7     | 1.51 | 0.08 |

Notes: (1) For 212 Area IDPs, MTAFS used 60 cores. (2) Power represents the power analysis which had 1000 SNPs and the effect size was 1. (3) Type 1 represents type 1 error analysis which had 1000 SNPs.