

ARTICLE



An adaptive and robust method for multi-trait analysis of genome-wide association studies using summary statistics

Qiaolan Deng ^{1,2}, Chi Song¹ and Shili Lin ²✉

© The Author(s), under exclusive licence to European Society of Human Genetics 2023

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. Through annotation analysis, the underlying genes of the SNPs identified by MTAFS were found to exhibit higher expression and are significantly enriched in brain-related tissues. 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.

European Journal of Human Genetics; <https://doi.org/10.1038/s41431-023-01389-7>

INTRODUCTION

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with complex diseases [1]. In single-trait analysis, it is typical to test the association between a single trait and a single variant one at a time. However, a common phenomenon is pleiotropy, in which a genetic variant is associated with multiple traits [2]. As such, conducting single-trait analyses may lose statistical power when genetic variants are associated with multiple traits. Therefore, there is an increasing need for methods that analyze multiple traits jointly.

Although there are numerous existing multi-trait methods, many require individual-level genotype data such as MultiPhen [3]. 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 are publicly available for most published studies. There are a number of multi-trait methods applicable to summary statistics. We categorize them into two groups: non-adaptive [4–10] or adaptive methods [11–17]. The latter 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. Among the methods in these two categories, frequently compared or computationally efficient ones are: Cauchy [9], HOM [6], metaMANOVA [7], SSU [4], SUM [5], aMAT [17], metaUSAT [14], and MTAR [16].

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 [18, 19]. However, many previous methods using summary statistics have only explored settings with a small number of traits [6, 14, 16], rendering their performance for analyzing a large number of traits unknown. Our preliminary simulation study indicates that methods such as SSU and aMAT are sensitive to signal sparsity and trait correlation structures, whereas metaUSAT and SSU may not control type 1 error well. 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 is also time-consuming, and it may return invalid values when the number of traits is large (e.g., over 200).

In this paper, we propose a Multi-Trait Adaptive Fisher method for Summary statistics (MTAFS, workflow in Supplementary Fig. S1), a computationally efficient and statistically powerful method for joint analysis of hundreds of phenotypes. We evaluated the proposed method in simulation studies and compared it with existing methods. We also applied MTAFS to brain imaging data

¹Division of Biostatistics, College of Public Health, The Ohio State University, Columbus, OH, USA. ²Department of Statistics, College of Arts and Sciences, The Ohio State University, Columbus, OH, USA. ✉email: shili@stat.osu.edu

Received: 19 May 2022 Revised: 1 May 2023 Accepted: 10 May 2023

Published online: 26 May 2023

from the UK Biobank involving volume and area and identified genetic loci associated with brain functions.

MATERIALS AND METHODS

Setup

Let $\mathbf{Z} = (z_1, \dots, z_T)_{(1 \times T)}$ 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 $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{R})$. Here, \mathbf{R} is referred to as the trait covariance matrix, which can be estimated by the sample covariance of \mathbf{Z} based on the independent and identically distributed assumption across SNPs [6, 13]. Linkage disequilibrium score regression is another option, which may be potentially suited when there are sample overlaps across studies [8, 16]. The estimated covariance matrix is denoted as $\hat{\mathbf{R}}$. We used sample covariance matrix estimates in the simulation studies and the real data applications.

First, we use eigen-decomposition to decorrelate the z scores. Let $\hat{\mathbf{R}} = \mathbf{Q}\mathbf{A}\mathbf{Q}^T$, where the columns of \mathbf{Q} are eigenvectors in decreasing order of their corresponding eigenvalues given in the corresponding diagonal elements of \mathbf{A} . We denote the proportion of variance explained by the first two eigenvalues as $v_0\%$. Then let $v_1\%$, $v_2\%$, and $v_3\%$ be the three percentages evenly distributed between $v_0\%$ and 100% , with q_1 , q_2 , 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 \mathbf{U}_E : $\mathbf{U}_E^T = \mathbf{Z}\mathbf{Q}_E\mathbf{A}_E^{-\frac{1}{2}}$, where $\mathbf{Q}_{E(T \times E)}$ consists of the first E columns of \mathbf{Q} and $\mathbf{A}_{E(E \times E)}$ is a submatrix of \mathbf{A} containing only the first E eigenvalues. As a result, \mathbf{U}_E is a column vector of length E , and $\mathbf{U}_E \sim \mathcal{N}(0, \mathbf{I})$ under the null hypothesis.

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 [20]. 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 \mathbf{U}_E , we obtain a vector of independent (two-sided) p -values, denoted as $\mathbf{p}_E = (p_1, \dots, p_E)$, such that $\mathbf{p}_E = 2[1 - \Phi(|\mathbf{U}_E|)]$, 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 $k \in \{1, \dots, E\}$. We can rewrite s_k as a weighted sum of independent χ^2 variables [21], for which Davies method (R package CompQuadForm) or the saddlepoint approximation method (R package Survey) can efficiently approximate its p -value [22], denoted as p_{s_k} . We define the test statistic of our adaptive method for level E as follows:

$$AF(E) = \text{Cauchy}(p_{s_k}; k = 1, \dots, E) = \sum_{k=1}^E \omega_k \tan\{(0.5 - p_{s_k})\pi\},$$

where $\omega_k = \frac{1}{E}$ for all k 's. This way of combining the evidence from p -values follows the Cauchy's method in the literature [9], and the p -value of the test statistic can be calculated analytically:

$$p_{AF(E)} = 0.5 - \frac{\arctan(AF(E))}{\pi}. \quad (1)$$

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 [9]. Nevertheless, since the p -values are calculated analytically, Cauchy's method is much more computationally efficient than minP.

MTAFS

From the literature [23] and our own preliminary study (Figs. S2–S4), it is shown that using either the first few or all eigenvectors would lead to unstable power. Therefore, we propose MTAFS, which integrates evidence from five levels of variance explained for robustness. Specifically, MTAFS constructs a test statistic combining $\{p_{AF(E)}, E \in \{2, q_1, q_2, q_3, T\}\}$

obtained from Eq. (1). We define the test statistics of MTAFS as

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

where $\omega_E = \frac{1}{5}$ for all E 's, and the corresponding p -value is

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

We have implemented MTAFS in an R package available at (<http://www.github.com/Qiaolan/MTAFS>). By vectorizing the main R function, MTAFS can simultaneously analyze a large number of SNPs, thereby increasing its computational efficiency.

RESULTS

Simulation setup

We simulated z scores from $\mathcal{N}(\boldsymbol{\mu}, \mathbf{R})$ following previous studies [9, 16, 17]. Various scenarios were constructed by setting different correlation structures, association models and strengths, and levels of signal sparsity. For \mathbf{R} , we considered two realistic covariance 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 covariance matrix referred to as UKCOR1 (Fig. S5); and the T1 FAST region of interests containing 139 IDPs, with its estimated covariance matrix denoted as UKCOR2 (Fig. S6). We also examined two commonly-used correlation structures, compound symmetry (CS) and first-order autocorrelation structure (AR), each with two levels of correlation—weak (0.3) or strong (0.7). This leads to a total of 6 covariance matrices (Table S1). For the analysis, we re-estimated the covariance matrix instead of using the one for simulating the data.

We considered two association models. In model 1 [16, 17], denoted as M1, we generated $\boldsymbol{\mu} = \sum_{j=1}^J c\lambda_j \mathbf{u}_j$, with c denoting the effect size, λ_j and \mathbf{u}_j the j^{th} eigenvalue and eigenvector of \mathbf{R} , respectively, and J the top J eigenvectors. We simulated different level of sparsity by varying J (Table S1). In the second association model (M2), we generated scenarios by directly setting some elements of $\boldsymbol{\mu}$ to be nonzero, with fewer non-zeros denoting greater sparsity. We note that when $c = 0$ in M1 or all elements of $\boldsymbol{\mu}$ 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 corresponds to a nonzero effect size are given in Table S1. We included aMAT, Cauchy, HOM, metaMANOVA, metaUSAT, MTAR, SSU, and SUM for comparison following the literature [9, 14, 16, 17].

Type 1 errors

We first evaluated the type 1 error of MTAFS and the comparison methods at various significance levels, from 1×10^{-4} to 5×10^{-8} , for UKCOR1, UKCOR2, CS(0.7), and AR(0.3) with 50 traits. Type 1 errors were evaluated with 10^9 replications. Each covariance matrix was estimated using the sample covariance. Table 1 shows the results for UKCOR1. At a lower significance level, Cauchy and MTAFS—which used Cauchy's method to combine p -values—are slightly above the upper bound. However, as the significance level becomes more stringent, the type 1 error for both MTAFS and Cauchy are within the bounds, consistent with a previous study [9]. Similar behaviors are observed for metaMANOVA, aMAT, and

Table 1. Type 1 error with covariance matrix UKCOR1^a.

	Significance levels				
	1×10^{-4}	1×10^{-5}	1×10^{-6}	1×10^{-7}	5×10^{-8}
Lower bound	9.94×10^{-5}	9.80×10^{-6}	9.37×10^{-7}	8.00×10^{-8}	3.59×10^{-8}
Upper bound	1.01×10^{-4}	1.02×10^{-5}	1.06×10^{-6}	1.20×10^{-7}	6.41×10^{-8}
metaMANOVA	1.03×10^{-4}	1.04×10^{-5}	1.04×10^{-6}	1.06×10^{-7}	5.10×10^{-8}
metaUSAT	1.72×10^{-4}	1.97×10^{-5}	2.14×10^{-6}	2.28×10^{-7}	1.27×10^{-7}
SUM	1.00×10^{-4}	1.01×10^{-5}	1.00×10^{-6}	1.03×10^{-7}	5.50×10^{-8}
SSU	1.90×10^{-4}	2.82×10^{-5}	4.20×10^{-6}	6.19×10^{-7}	3.57×10^{-7}
HOM	1.01×10^{-4}	1.01×10^{-5}	1.01×10^{-6}	9.70×10^{-8}	4.80×10^{-8}
Cauchy	1.04×10^{-4}	1.03×10^{-5}	1.07×10^{-6}	1.15×10^{-7}	5.90×10^{-8}
aMAT	1.00×10^{-4}	1.10×10^{-5}	1.22×10^{-6}	1.30×10^{-7}	5.70×10^{-8}
MTAR	1.03×10^{-4}	1.04×10^{-5}	1.06×10^{-6}	1.05×10^{-7}	5.50×10^{-8}
MTAFS	1.06×10^{-4}	1.05×10^{-5}	1.07×10^{-6}	1.12×10^{-7}	5.40×10^{-8}

^aThe values in the table are empirical type 1 errors calculated by using 10^9 replications. Lower bound and upper bound correspond to the acceptance bounds of $\alpha \pm 2\sqrt{\alpha(1-\alpha)}/10^9$, where α is the significance level. The values outside the bounds are in bold. UKCOR1 is the estimated covariance matrix based on the 58 Volumetric image-derived phenotypes of the UK Biobank data.

MTAR. On the other hand, metaUSAT and SSU remain greatly outside the upper bound at each significance level considered, indicating that these two methods have consistently inflated type 1 error rates. We evaluated the type 1 error with UKCOR2 and observed similar trends (Table S2). For AR(0.3) and CS(0.7), type 1 error rates are better controlled by all methods except SSU for AR(0.3) (Tables S3, S4).

Power comparisons

For power comparisons, we simulated 1000 z scores and set the significance level at 5×10^{-8} . First, we evaluated the power with UKCOR1 model M1 (Fig. 1). When only the top two eigenvectors were informative (high sparsity), SUM and SSU were the most powerful methods, followed by metaUSAT and MTAFS. As more eigenvectors become informative (intermediate and low sparsity), 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. 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 for a fairer comparison of power. 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 the less sparse scenarios. We note that HOM had no power at all three sparsity levels, an observation consistent with previous studies [17].

For M2 with UKCOR1, MTAFS was seen to be the most powerful methods at all three sparsity levels (Fig. 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 most 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 covariance matrix UKCOR2 (Figs. S7, S8). For both M1 and M2, the results were similar to those for UKCOR1. Considering all the results together, the main observation for UKCOR1 remains qualitatively the same for the UKCOR2 covariance matrix: the performance of the other methods is unstable, whereas MTAFS is extremely consistent across all settings and was always among the top performers.

For the CS covariance matrix with model M2 (Figs. S9, S10), MTAFS remains among the group of most powerful methods. This is also true for M2 with the AR structure (Figs. 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 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 (Table 2 contains power results with one effect size for each setting, with all results provided in Tables S5–S10, Figs. 2 and 3, S7–S12), 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 was the best in six of the 28 settings in Table 2 (tied for the most among all methods), and was within 5% of the best in 13 of the settings and within 10% of the best in all but 4 settings; even in the worst case, MTAFS was only 28% below the best performer. This robust performance characteristic of MTAFS was unmatched by any of the other methods: for metaMANOVA, metaUSAT, and MTAR, their power can be more than 88, 59, and 81% below the best performer; whereas the rest of the methods may have 0 power in some settings.

Real data application

Data and pre-processing. Brain functions and the underlying mechanisms are still largely unknown, despite considerable effort and investigation of connections between brain function and genetics using imaging data [24, 25]. It has been found that regional brain morphology, such as surface area and thickness of the cerebral cortex, and volume of subcortical structures, has a complex genetic architecture, where many SNPs, some having small effect sizes, may be associated with sets of regional brain features [10]. Thus, conventional single-trait analysis may have little power for detecting genetic variants due to its ignorance of underlying correlation structures. In contrast, some recent studies applied multi-trait methods to the UK Biobank [26] brain imaging data and achieved success [10]. For example, a study applied aMAT to 58 volumetric IDPs and identified SNPs failed to be detected by single-trait analyses [17].

UK Biobank study with 500,000 volunteers [26]. 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. [19] investigated the genetic architecture of brain structure and

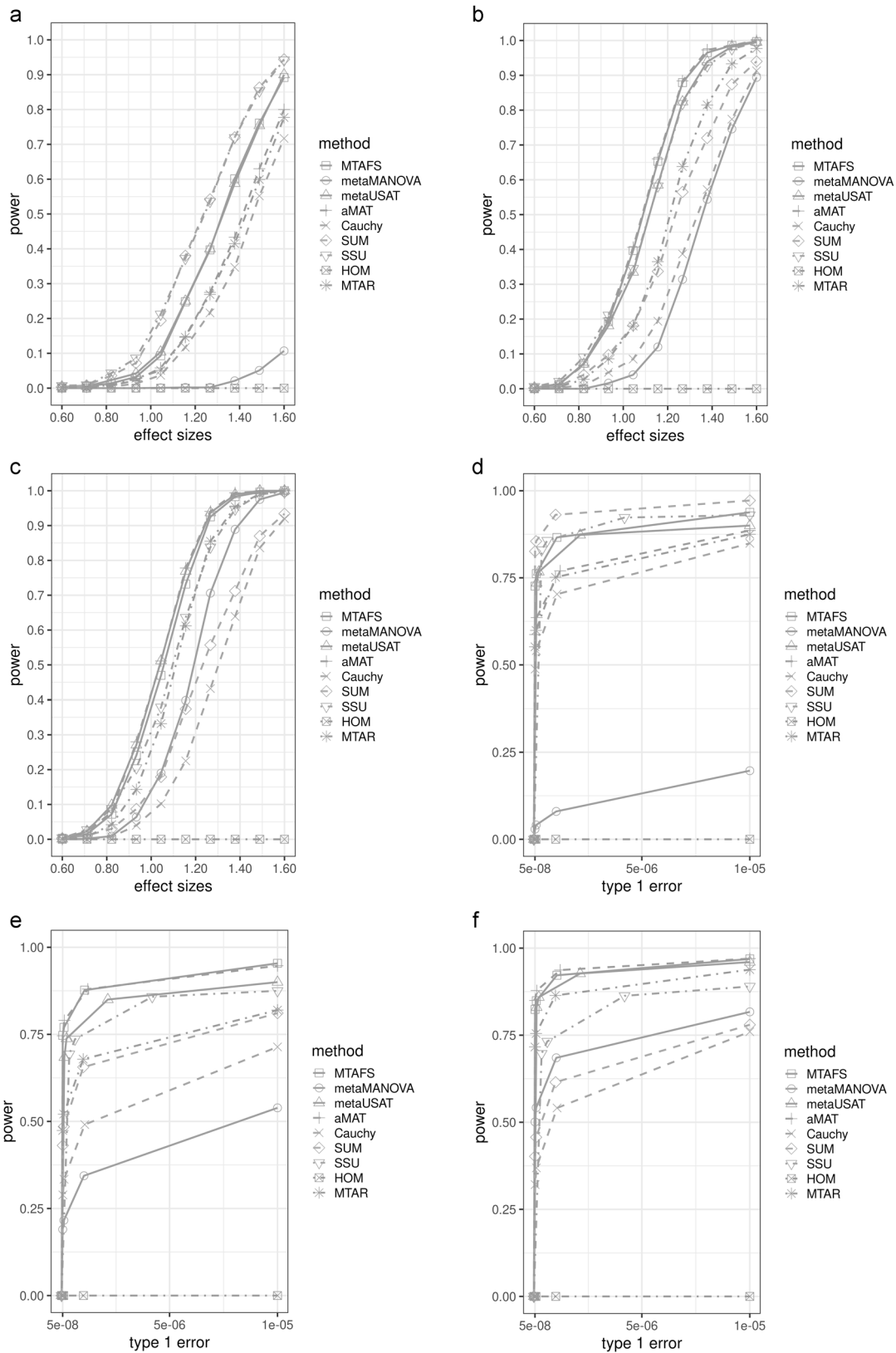


Fig. 1 Power comparison and ROC curves of UKCOR1 and M1. Comparison of methods for model M1 using the UKCOR1 covariance matrix in terms of power over a range of effect sizes (a–c) and partial ROC curves comparing power versus type 1 error (d–f) for a particular effect size. The power results (a–c) are plotted at significance level of 5×10^{-8} , while the ROC curves (d–f) are shown for the range of type 1 error rate from 5×10^{-8} to 1×10^{-5} . **a, d** High sparsity, with only top 2 eigenvectors informative; **b, e** Intermediate sparsity, with top 11 eigenvectors informative; **c, f** Low sparsity, with top 25 eigenvectors informative. MTAR only returns the number of significant SNPs given a significance level, thus excluded in the ROC plots. The effect size was set to be 1.2 for the high, and 1.0 for the intermediate and low sparsity settings.

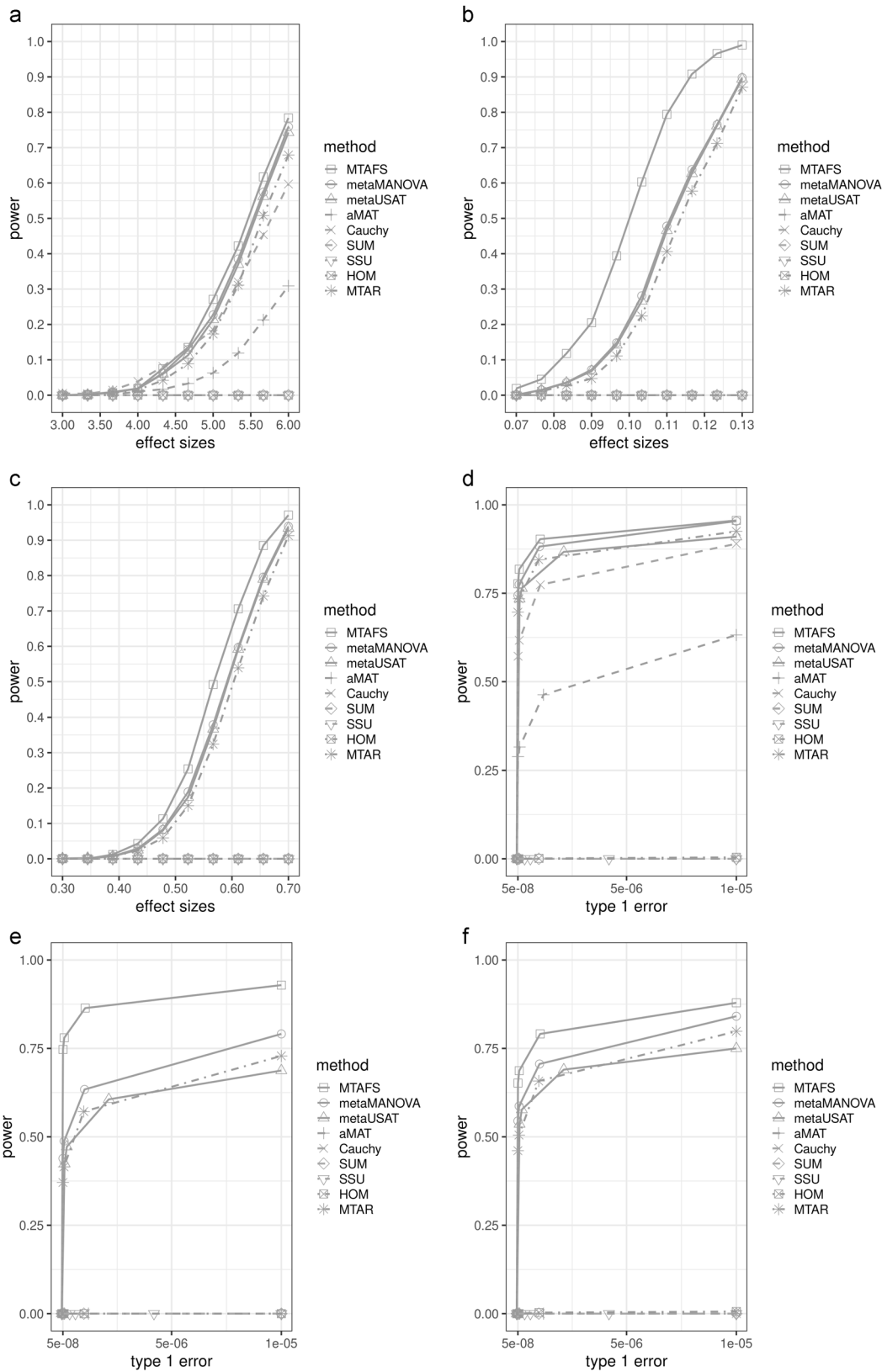


Fig. 2 Power comparison and ROC curves of UKCOR1 and M2. Comparison of methods for model M2 using the UKCOR1 covariance matrix in terms of power over a range of effect sizes (a–c) and partial ROC curves comparing power versus type 1 error (d–f) for a particular effect size. The power results (a–c) are plotted at significance level of 5×10^{-8} , while the ROC curves (d–f) are shown for the range of type 1 error rate from 5×10^{-8} to 1×10^{-5} . **a, d** High sparsity, with 3 nonzero components of μ ; **b, e** Intermediate sparsity, 13 nonzero components of μ ; **c, f** Low sparsity, with 30 nonzero components of μ . MTAR only returns the number of significant SNPs given a significance level, thus excluded in the ROC plots. The effect size was set to be 5.0 for the high, 0.1 for the intermediate, and 0.55 for the low sparsity settings.

Table 2. Summary of power for all simulation settings based on one effect size per setting^a.

Correlation	Association Model	Sparsity	MTAFS	metaMANOVA	metaUSAT	aMAT	Cauchy	SUM	SSU	HOM	MTAR
UKCOR1	M1	H	0.892	0.107	0.900	0.800	0.717	0.944	0.942	0	0.777
		M	0.879	0.314	0.824	0.884	0.387	0.565	0.819	0	0.638
		L	0.919	0.700	0.942	0.947	0.483	0.537	0.836	0	0.855
UKCOR1	M2	H	0.784	0.759	0.743	0.309	0.597	0	0	0	0.679
		M	0.966	0.765	0.762	0	0	0	0	0	0.712
		L	0.885	0.796	0.79	0	0	0	0	0	0.742
UKCOR2	M1	H	0.89	0.014	0.922	0.445	0.183	0.914	0.974	0.134	0.455
		M	0.972	0.387	0.935	0.977	0.379	0.443	0.949	0.005	0.6
		L	0.956	0.765	0.958	0.959	0.279	0.266	0.909	0.001	0.937
UKCOR2	M2	H	0.917	0.391	0.379	0	0	0	0	0	0.323
		M	0.926	0.954	0.96	0.239	0.011	0.001	0.127	0	0.956
		L	0.762	0.844	0.891	0.795	0.002	0.029	0.184	0.033	0.864
CS(0.3)	M2 ^b	H	0.818	0.885	0.868	0.631	0	0	0	0	0.848
		L	0.881	0.914	0.9	0.896	0.014	0	0	0	0.88
		H	0.922	0.959	0.94	0.943	0.93	0	0	0	0.93
CS(0.7)	M2 ^b	L	0.945	0.963	0.957	0.957	0.002	0	0	0	0.956
		H	0.857	0.92	0.888	0	0.034	0	0	0	0.88
		L	0.897	0.941	0.918	0	0	0	0	0	0.91
AR(0.3)	M2 ^b	H	0.874	0.931	0.894	0	0.067	0	0	0	0.891
		L	0.766	0.831	0.796	0	0	0	0	0	0.787
		H	0.683	0.735	0.74	0.693	0.95	0	0.692	0	0.689
AR(0.7)	M2 ^b	L	0.999	0.948	0.997	0.946	0.176	0.198	0.998	0.184	0.987
		H	0.712	0.736	0.749	0.697	0.995	0	0.718	0	0.679
		L	0.794	0.18	0.66	0.016	0.003	0.06	0.706	0.066	0.15
AR(0.7)	M2 ^c	H	0.942	0.965	0.959	0.961	0.592	0	0.002	0	0.95
		L	0.967	0.97	0.986	0.97	0.606	0.019	0.941	0.016	0.977
		H	0.905	0.95	0.939	0.953	0.795	0	0.002	0	0.918
AR(0.7)	M2 ^c	L	0.516	0.228	0.712	0.321	0.060	0.01	0.707	0.006	0.416

^aEffect sizes are: UKCOR1 M1 (H 1.6; M 1.27; L 1.27), UKCOR1 M2 (H 6; M 0.12; L 0.66), UKCOR2 M1 (H 1.4; M 1.1; L 1.03), UKCOR2 M2 (H 2.5; M 3; L 2), CS (0.3) (50 traits: H 6, L 3; 100 traits: H 7, L 2.5), CS (0.7) (50 traits: H 4, L 2; 100 traits: H 4.5, L 1.5), AR (0.3) (50 traits: H 7, L 4; 100 traits: H 8, L 2.5), AR (0.7) (50 traits: H 6, L 5; 100 traits: H 6.7, L 3.5). Results for all settings are in Tables S4–S9 and Figs. 1 and 2, S7–S12.

^b50 traits.

^c100 traits.

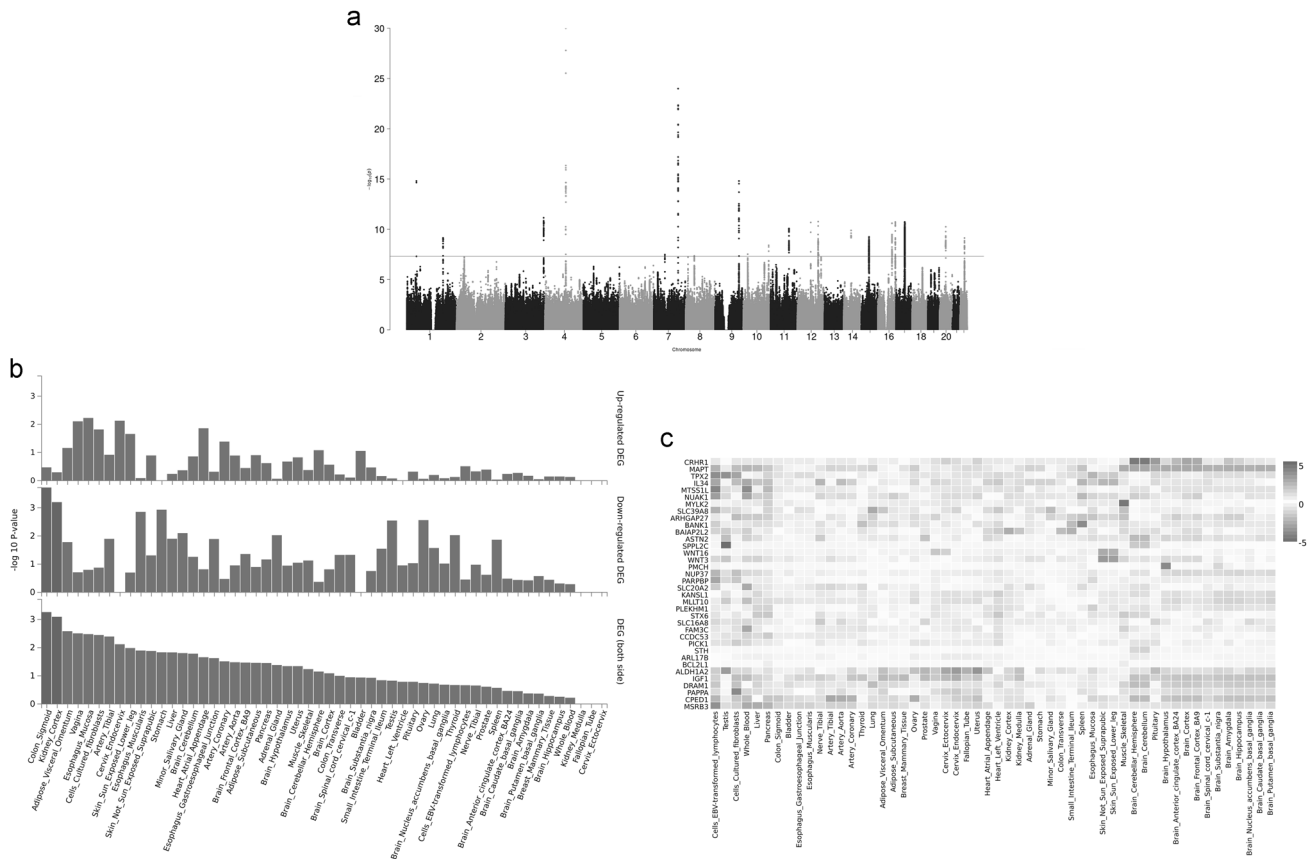


Fig. 3 Analysis results of the 58 Volumetric IDPs. **a** Manhattan plot of the SNPs identified by MTAFS. For $-\log_{10} p$ -values > 30 , they are censored at 10^{-30} for a better visualization. For **b**, **c**, we use the GTEx data over 54 tissue types. **b** Tissue expression analysis for genes identified by MTAFS. Red bars denote differential expression with the Bonferroni corrected p -values less than 0.05. Blue bars denote no differential expression. Top panel: up-regulation; middle panel: down-regulation; bottom panel: two-sided results. **c** The expression heatmap of all genes identified by MTAFS. Each row denotes a gene and each column denotes a tissue.

Table 3. The number of significant SNPs identified at 5×10^{-8} significance level^a.

	MTAFS	metaMANOVA	aMAT	MTAR	Single-trait
58 Volumetric IDPs	36	36	44	36	15
212 Area IDPs	35	17	35	26	7

^aThe resulting SNPs were LD (linkage disequilibrium) pruned using PLINK following Schulz et al. (2017).

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 include multimodal information on grey matter volume, area, and thickness. In particular, the 58 Volumetric IDPs and the 212 Area IDPs are both related to grey matter, with a sample size of 8428 individuals. Imputation of genotypes and quality controls for genotyping data are described in the literature [18, 19]. We carried out two multi-trait analyses, one with a moderate number of the 58 Volumetric IDPs, and the other with a large number of the 212 Area IDPs of grey matter (Fig. S13). Note that we only used the summary statistics not the raw genotypes nor the IDPs in our analysis.

The summary statistics included the z scores calculated by Wald test from measuring the associations between each of the 11,734,353 SNPs and each of the 58 or 212 IDPs [19]. We focused on common variants and filtered out SNPs with $MAF < 0.05$, leading to 4,590,290 SNPs remaining. The covariance matrix for each of the two sets of IDPs was estimated based on the z scores for SNPs with $MAF > 0.05$, following the literature [6, 13]. MTAFS

and a subset of the competing methods that performed reasonably well 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. Note that the methods excluded had almost no power in some settings or have inflated type I errors in most settings. For convenience of reference, hereafter, competing methods refer to metaMANOVA, aMAT, and MTAR, unless specified otherwise. We used a genome-wide significance threshold of 5×10^{-8} for each of the multi-trait analysis methods. The SNPs identified were summarized into significant loci using LD-pruning [27]. The genes corresponding to the significant SNPs were identified using NCBI dbSNP [28]. To investigate gene annotations, we used Functional Mapping and Annotation (FUMA) [29] to show tissue specific expression patterns of genes identified by MTAFS and other methods.

Results of 58 Volumetric IDPs. MTAFS identified 2,157 SNPs with p -values less than 5×10^{-8} (Figs. 3a, S13a), leading to 36 LD-pruned significant loci (Table 3). aMAT identified 44 significant loci, while the rest of the competing methods identified the same

number of loci as MTAFS (Table 3, Fig. S14). We also carried out single-trait analysis as a comparison and corrected for multiple comparison based on the effective number of phenotypes [30], a method less conservative than Bonferroni. All LD-pruned loci identified by the single-trait analysis were also identified by MTAFS and the comparison methods.

We first investigated several genes containing significant SNPs stacked as “towers” in the Manhattan plot (Fig. 3a). Gene *SLC39A8*, living in the chromosome 4 tower, has been found in the literature to possess important functions in the brain. In particular, an investigation studying the association between common variants and cerebellar volume states that *SLC39A8* is associated with a wide-range of traits including inferior posterior and flocculonodular lobule, striatum and putamen volumes, schizophrenia, neurodevelopmental outcomes and intelligence test performance, and numerous other factors [31]. Genes *FAM3C* and *WNT16*, located in the right (much taller) tower in chromosome 7, are both reported to be associated with brain volume by a study using the UK Biobank data [32]. *PAPPA*, located in chromosome 9, is associated with the volume of brainstem and brain region according to multiple studies [32, 33].

MTAFS uniquely identified two genes, *BAIAP2L2* and *TPX2*, located in the chromosome 22 tower and the chromosome 20 tower, respectively. There are also a tower in chromosome 22 and another in chromosome 20 in the aMAT Manhattan plot—no towers in these two chromosomes from the other methods (Figs. S15–S17)—but they contain neither of these two genes. A recent study investigating brain white matter reported an association between *BAIAP2L* and white matter microstructure [34]. Gene *TPX2* has also been reported to be associated with brain volume and sulcal depth [34, 35].

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 [36] including 54 human tissues. Although there are no major clusters, there is a small cluster in the upper right consisting of two genes that are more highly expressed in brain related tissues. In FUMA, we also tested if the gene set was significantly enriched in tissues. Figure 3b shows that the genes were not significantly enriched in brain-related tissues. The competing methods have similar findings (Figs. S15–S17).

Results of 212 Area IDPs. This analysis considered a much larger set of traits. MTAFS identified 1,170 SNPs with p -values less than 5×10^{-8} (Fig. S19a, b). Both MTAFS and aMAT identified 35 significant loci after LD pruning, while the rest of the competing methods identified fewer (Table 3, Fig. S18). Single-trait analysis only identified 7 loci and all were identified by the multi-trait methods.

Similarly, we started with genes located in the “towers” of the Manhattan plot (Fig. S19a). The strongest signals from the tall tower in chromosome 15 correspond to gene *THBS1*, which was reported to be associated with cortical surface area [37]. *DAAM1*, located in chromosome 14, has been shown to affect brain volumes [38]. Several studies have found that *ZIC4* of chromosome 3 is associated with brain-related traits such as parietal lobe volume and total cortical area [32, 34, 38]. The rest of the towers contain genes *RPL21P24*, *STRN*, *C16orf95*, *NSF*, and *NFIX*, which have been reported to be associated with brain volume and cortical area by many studies [10, 32].

The expression heatmap (Fig. S19c) shows a cluster of three genes (top-left corner) that has much higher expression levels in two brain-related tissues: cerebellar hemisphere and cerebellum. There is also a broader cluster (also in the top left) containing genes that have relatively higher expression in the remaining brain tissues compared to the other tissues. Figure S19b shows that the gene set consisting of genes identified by MTAFS are

significantly enriched in brain hypothalamus with up-regulation. The results of the competing methods are in the supplementary materials (Figs. S20–S22).

DISCUSSION

GWAS have successfully identified a large number of genetic variants associated with traits or diseases. 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 with stringent significance levels, including the genome-wide significance level at 5×10^{-8} , and has consistent performance under a variety of settings, underscoring its robustness. Although it is a common practice to directly simulate summary statistics for evaluating multi-trait methods [16, 17], we nevertheless also carry out a brief simulation study to further gauge the type I error rate when genotypes rather than summary statistics are generated. Since our real data analysis considers only variants with $MAF > 0.05$, here we considered genotypes whose MAFs are randomly distributed between 0.05 and 0.5. Our results indicate that the type I error is also well controlled for, albeit a bit more conservative when the significance level is set to be stringent (Table S11, next to last column in the segment on Type I Error). As mentioned above, for the two real data examples, we restricted our analyses to variants with MAF greater than 0.05 and we calculated the genomic inflation factor (GIF) to further validate our findings. For the *Volumetric* (first analysis) and the *Area* (second analysis) data, the GIF lambda value was 1.05 for both analyses, indicating little evidence of any potential inflation of type I error.

Like all the other methods, the performance of MTAFS is dependent on a number of underlying factors that are explored in the simulation studies. Nevertheless, although MTAFS is not always the best, it keeps up with the top performers in all combinations of four underlying correlation structures, two association models, and three sparsity levels. Further, it is the only method, among the nine studied, that has this property. For example, SSU and SUM performed reasonably well under UKCOR1 model M1 but had absolutely no power under M2. On the other hand, metaMANOVA is among the top performers for the high sparsity setting under UKCOR1 model M2, but much less power under UKCOR2 model M2, and there is a big discrepancy in its relative power for the high sparsity and low sparsity settings in UKCOR1 M1 but not so much in M2. In contrast, although MTAFS was seen to be a clear winner in only a few scenarios, its power is always up there with the top group regardless in all 28 combinations considered. For a real data analysis, since the underlying correlation structure, the model, and the sparsity level are all unknown, yet existing methods are seen to be rather sensitive to these primary factors, we believe that a method like MTAFS, which is robust to these unknown underlying features, would be desirable.

In general, MTAFS exhibits desirable properties, and have several advantages over existing methods. First, MTAFS controls type 1 error well. Second, MTAFS has robust performance given various covariance matrices, underlying association models, and different levels of signal sparsity. Third, MTAFS is an efficient method in practice. Although 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. Parallel computing can greatly reduce its computational time, making it acceptable in practice. As a demonstration, we analyzed the *Area* IDP data for 4,590,290 SNPs and 212 traits, and MTAFS finished the analysis in less than one hour by using 100 cores of 4GB memory.

The advantages notwithstanding, there are limitations in the proposed method. First, because we transform raw z score vectors by eigen decomposition, 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, although MTAFS can analyze rare variants as long as their GWAS summary statistics are available, we are concerned about the statistical power of single rare variants analysis methods as well as the normality assumption. Indeed, both rare variants and non-normality of the trait distribution can affect the power and the type I error of MTAFS, as we demonstrated type I error inflation in two bins with $MAF < 0.05$ or in the use of a t-distribution for generating the trait values (Table S11). Therefore, we suggest that MTAFS be only used for studying common variants – multiple traits associations. Further, the normality of the trait distribution and the summary statistics should be carefully checked before applying the method. If normality is questionable, then an appropriate transformation, such as an inverse-normal quantile transformation, may be first applied before running MTAFS.

DATA AVAILABILITY

The author states that all data necessary for confirming the conclusions presented in the article are represented fully within the article. The UK Biobank summary statistics are available in the <https://open.win.ox.ac.uk/ukbiobank/big40/BIGv2/>.

CODE AVAILABILITY

The software for our proposed method MTAFS is available at <https://github.com/Qiaolan/MTAFS>. All results for the real data analyses are at https://drive.google.com/drive/folders/1VpKHkT4mHnNXVesygkbt8K5X_pio0En9zusp=share_link.

REFERENCES

- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature* 2009;461:747–53.
- Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet*. 2013;14:483–95.
- O'Reilly PF, Hoggart CJ, Pomyen Y, Calboli FCF, Elliott P, Jarvelin MR, et al. MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. *PLoS One*. 2012;7:e34861.
- Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society* 2009;33:497–507.
- He Q, Avery CL, Lin DY. A general framework for association tests with multivariate traits in large-scale genomics studies. *Genet Epidemiol*. 2013;37:759–67.
- 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. *Am J Hum Genet*. 2015;96:21–36.
- Xu X, Tian L, Wei LJ. Combining dependent tests for linkage or association across multiple phenotypic traits. *Biostatistics* 2003;4:223–9.
- Turley P, Walters RK, Maghzi A, Okbay A, Lee JJ, Fontana MA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet*. 2018;50:229–37.
- Liu Y, Xie J. Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *J Am Stat Assoc*. 2020;115:393–402.
- van der Meer D, Frei O, Kaufmann T, Shadrin AA, Devor A, Smeland OB, et al. Understanding the genetic determinants of the brain with MOSTest. *Nat Commun*. 2020;11:1–9.
- Van der Sluis S, Posthuma D, Dolan CV. TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genet*. 2013;9:e1003235.
- Liu Z, Lin X. A geometric perspective on the power of principal component association tests in multiple phenotype studies. *J Am Stat Assoc*. 2019;114:975–90.
- Kim J, Bai Y, Pan W. An adaptive association test for multiple phenotypes with GWAS summary statistics. *Genet Epidemiol*. 2015;39:651–63.
- Ray D, Boehnke M. Methods for meta-analysis of multiple traits using GWAS summary statistics. *Genet Epidemiol*. 2018;42:134–45.
- Liu Z, Lin X. Multiple phenotype association tests using summary statistics in genome-wide association studies. *Biometrics* 2018;74:165–75.
- Guo B, Wu B. Integrate multiple traits to detect novel trait–gene association using GWAS summary data with an adaptive test approach. *Bioinformatics* 2019;35:2251–7.
- Wu C. Multi-trait genome-wide analyses of the brain imaging phenotypes in UK Biobank. *Genetics* 2020;215:947–58.
- 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;562:203–9.
- 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;562:210–6.
- Song C, Min X, Zhang H. The screening and ranking algorithm for change-points detection in multiple samples. *Ann Appl Stat*. 2016;10:2102.
- David HA, Nagaraja HN. *Order statistics*: John Wiley & Sons; 2004.
- Wu B, Guan W, Pankow JS. On efficient and accurate calculation of significance p-values for sequence kernel association testing of variant set. *Ann Hum Genet*. 2016;80:123–35.
- Aschard H, Vilhjálmsdóttir BJ, Greliche N, Morange PE, Trégué DA, Kraft P. Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *Am J Hum Genet*. 2014;94:662–76.
- Sporns O, Tononi G, Kötter R. The human connectome: a structural description of the human brain. *PLoS Comput Biol*. 2005;1:e42.
- Bi Xa HuX, Xie Y, Wu H. A novel CERNNE approach for predicting Parkinson's Disease-associated genes and brain regions based on multimodal imaging genetics data. *Med Image Anal*. 2021;67:101830.
- Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.
- Schulz H, Ruppert AK, Herms S, Wolf C, Mirza-Schreiber N, Stegle O, et al. Genome-wide mapping of genetic determinants influencing DNA methylation and gene expression in human hippocampus. *Nat Commun*. 2017;8:1511.
- Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, et al. Database resources of the national center for biotechnology information. *Nucleic acids Res*. 2021;49:D10.
- Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8:1–11.
- Li MX, Yeung JMY, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet*. 2012;131:747–56.
- Chambers T, Escott-Price V, Legge S, Baker E, Singh KD, Walters JTR, et al. Genetic common variants associated with cerebellar volume and their overlap with mental disorders: a study on 33,265 individuals from the UK-Biobank. *Mol Psychiatry*. 2022;27:2282–90.
- Smith SM, Douaud G, Chen W, Hanayik T, Alfaro-Almagro F, Sharp K, et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat Neurosci*. 2021;24:737–45.
- Zhao B, Luo T, Li T, Li Y, Zhang J, Shan Y, et al. 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. *Nat Genet*. 2019;51:1637–44.
- Zhao B, Zhang J, Ibrahim JG, Luo T, Santelli RC, Li Y, et al. Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). *Mol Psychiatry*. 2021;26:3943–55.
- Van Der Meer D, Kaufmann T, Shadrin AA, Makowski C, Frei O, Roelfs D, et al. The genetic architecture of human cortical folding. *Sci Adv* 2021;7:eabj9446.
- Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45:580–5.
- Hofer E, Roshchupkin GV, Adams HHH, Knol MJ, Lin H, Li S, et al. Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. *Nat Commun*. 2020;11:1–16.
- van Der Lee SJ, Knol MJ, Chauhan G, Satizabal CL, Smith AV, Hofer E, et al. A genome-wide association study identifies genetic loci associated with specific lobar brain volumes. *Commun Biol*. 2019;2:285.

ACKNOWLEDGEMENTS

We would like to thank the Section Editor and two anonymous reviewers for their valuable comments and suggestions. We would also like to acknowledge the Ohio Supercomputer Center for providing resources that have contributed to the research results reported within this paper.

AUTHOR CONTRIBUTIONS

QD conceived the study design, performed analyses, developed methodology and drafted the manuscript. CS and SL provided extensive feedback regarding the study design, methodology and analyses, and revised the manuscript.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICAL APPROVAL

The UK Biobank brain imaging data (<https://open.win.ox.ac.uk/ukbiobank/big40/BIGv2/>) were publicly available without identifying information.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41431-023-01389-7>.

Correspondence and requests for materials should be addressed to Shili Lin.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.