Identification of Disease-Sensitive Brain Imaging Phenotypes and Genetic Factors Using GWAS Summary Statistics

Duo Xi, Dingnan Cui, Jin Zhang, Muheng Shang, Minjianan Zhang, Lei Guo, Junwei Han(✉), and Lei Du(✉)

School of Automation, Northwestern Polytechnical University, Xi'an 710072, China
{jhan,dulei}@nwpu.edu.cn

Abstract. Brain imaging genetics is a rapidly growing neuroscience area that integrates genetic variations and brain imaging phenotypes to investigate the genetic underpinnings of brain disorders. In this field, using multi-modal imaging data can leverage complementary information and thus stands a chance of identifying comprehensive genetic risk factors. Due to privacy and copyright issues, many imaging and genetic data are unavailable, and thus existing imaging genetic methods cannot work. In this paper, we proposed a novel multi-modal brain imaging genetic learning method that can study the associations between imaging phenotypes and genetic variations using genome-wide association study (GWAS) summary statistics. Our method leverages the powerful multi-modal of brain imaging phenotypes and GWAS. More importantly, it does not need to access the imaging and genetic data of each individual. Experimental results on both Alzheimer's Disease Neuroimaging Initiative (ADNI) database and GWAS summary statistics suggested that our method has the same learning ability, including identifying associations between genetic biomarkers and imaging phenotypes and selecting relevant biomarkers, as those counterparts depending on the individual data. Therefore, our learning method provides a novel methodology for brain imaging genetics without individual data.

Keywords: Brain imaging genetics · GWAS summary statistics · Multi-modal brain image analysis

L. Du—This work was supported in part by National Natural Science Foundation of China [61973255, 62136004, 61936007] at Northwestern Polytechnical University.

Alzheimer's Disease Neuroimaging Initiative: Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 622–631, 2023.
https://doi.org/10.1007/978-3-031-43904-9_60

1 Introduction

Nowadays, brain imaging genetics has gained increasing attention in the neuroscience area. This interdisciplinary field refers to integrates genetic variations (single nucleotide polymorphisms, SNPs) and structural or functional neuroimaging quantitative traits (QTs). Different imaging technologies can capture different knowledge of the brain and thus are a better choice in brain imaging genetics [12,17]. Over the past decade, genome-wide association studies (GWAS) have proven to be a powerful tool in finding the genetic effects on imaging phenotypes in single SNP level [7,10,14]. However, GWAS can only investigate the single-SNP-single-QT relationship, and thus may lose the information among multiple SNPs and/or multiple QTs due to the polygenic inheritance of brain disorders [9].

To leverage the multi-modal brain imaging QTs and identify the joint effect of multiple SNPs, many learning methods were proposed for multi-modal brain imaging genetics [5,6,15]. The dirty multi-task sparse canonical correlation analysis (DMTSCCA) is a bi-multivariate learning method for multi-modal brain imaging genetics [4]. DMTSCCA can disentangle the specific patterns of multi-modal imaging QTs from shared ones and thus is state-of-the-art. However, DMTSCCA depends on individual-level imaging and genetic data, and cannot work when the original imaging and genetic data are unavailable.

Since GWAS studies usually release their summary statistics results for academic use, we here developed a novel DMTSCCA method using GWAS summary statistics rather than individual data. The method, named S-DMTSCCA, has the same ability as DMTSCCA in modeling the association between multi-modal imaging QTs and SNPs and does not require raw imaging and genetic data. We investigated the performance of S-DMTSCCA based on two kinds of experiments. Firstly, we applied S-DMTSCCA to GWAS summary statistics from Alzheimer's Disease Neuroimaging Initiative (ADNI) and compared it to DMTSCCA which directly ran on the original imaging genetic data of ADNI. Results suggested that S-DMTSCCA and DMTSCCA obtained equivalent results. Secondly, we applied S-DMTSCCA to a GWAS summary statistics from the UK Biobank. The experiment results showed that S-DMTSCCA can identify meaningful genetic markers for brain imaging QTs. More importantly, the structure information of SNPs was also captured which was usually missed by GWAS. It is worth noting that all these results were obtained without assessing the original neuroimaging genetic data. This demonstrates that our method is a powerful tool and provides a novel method for brain imaging genetics.

2 Method

In this article, we represent scalars with italicized letters, column vectors with boldface lowercase letters, and matrices with boldface capitals. For $\mathbf{X} = (x_{ij})$, the ith row is denoted as $\mathbf{x}^i$, jth column as $\mathbf{x}_j$, and the ith matrix as $\mathbf{X}_i$. $\|\mathbf{X}\|_2$ denotes the Euclidean norm, $\|\mathbf{X}\|_{2,1}$ denotes the sum of the Euclidean norms of the rows of $\mathbf{X}$. Suppose $\mathbf{X} \in \mathbb{R}^{n \times p}$ load the genetic data with n subjects and p biomarkers, and $\mathbf{Y}_c \in \mathbb{R}^{n \times q} (c = 1, ..., C)$ load the cth modality of phenotype data, where q and C is the number of imaging QTs and imaging modalities (tasks) respectively.

2.1 DMTSCCA

DMTSCCA identifies the genotype-phenotype associations between SNPs and multi-modal imaging QTs using the following model [4],

$$\min_{\mathbf{S}, \mathbf{W}, \mathbf{v}_c} \sum_{c=1}^{C} \left[\kappa_c \|\mathbf{X}\left(\mathbf{s}_c + \mathbf{w}_c\right) - \mathbf{Y}_c \mathbf{v}_c\|_2^2 + \lambda_v \|\mathbf{v}_c\|_1 \right] + \lambda_s \|\mathbf{S}\|_{2,1} + \lambda_w \|\mathbf{W}\|_{1,1} \\ s.t. \ \|\mathbf{X}\left(\mathbf{s}_c + \mathbf{w}_c\right)\|_2^2 = 1, \|\mathbf{Y}_c \mathbf{v}_c\|_2^2 = 1, \forall c. \tag{1}$$

In this model, $\kappa \in \mathbb{R}^{1 \times C} \, (0 \leq \kappa_c \leq 1, \sum_c \kappa_c = 1)$ is a weight vector to balance among multiple sub-tasks. In this paper, κ ensures an equal optimization for each imaging modality. $\mathbf{v}_c$ is the cth vector in $\mathbf{V}$, where $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_C] \in \mathbb{R}^{q \times C}$ denotes the canonical weight for phenotypic data. $\mathbf{S} \in \mathbb{R}^{p \times C}$ and $\mathbf{W} \in \mathbb{R}^{p \times C}$ are the canonical weights for genotypes, where $\mathbf{S}$ is the task-consistent component being shared by all tasks and $\mathbf{W}$ is the task-dependent component being associated with a single task. λ_v, λ_s and λ_w are nonnegative tuning parameters.

2.2 Summary-DMTSCCA (S-DMTSCCA)

Now we propose S-DMTSCCA only using summary statistics from GWAS. It does not need individual-level imaging and genetic data.

For ease of presentation, we derive our method by first introducing GWAS. GWAS uses linear regression to study the effect of a single SNP on a single imaging QT. Let $\mathbf{x}_d$ denotes the genotype data with p SNPs and $\mathbf{y}_l$ denotes the phenotype data with q imaging QTs, a typical GWAS model can be defined as

$$\mathbf{y}_l = \alpha + \mathbf{x}_d b_{dl} + \epsilon, \tag{2}$$

where b_{dl} is the effect size of the d-th SNP on the l-th imaging QT. α is the y-intercept, and ϵ is the error term which is independent of $\mathbf{x}_d$. When the SNPs and imaging QTs were normalized to have zero mean and unit variance, b_{dl} will equal to the covariance between $\mathbf{x}_d$ and $\mathbf{y}_l$, i.e., $b_{dl} = \frac{\mathbf{x}_d^T \mathbf{y}_l}{n-1}$. On this account, we can construct $\mathbf{B} \in \mathbb{R}^{p \times q}$ by loading $p \times q$ summary statistics of GWAS. Obviously, $\mathbf{B}$ will be the covariance between multiple SNPs and multiple imaging QTs since its element is the covariance of a single SNP and a single imaging QT. Let $\mathbf{B}_c$ denotes covariance of the c-th modality from GWAS, we have

$$\mathbf{X}^T \mathbf{Y}_c = (n-1) \, \mathbf{B}_c. \tag{3}$$

Further, we use $\hat{\Sigma}_{XX}$ denote an estimated covariance of genetic data, i.e., $\hat{\Sigma}_{XX} = \frac{\overline{\mathbf{X}}^T \overline{\mathbf{X}}}{n-1}$. $\overline{\mathbf{X}}$ can be obtained from n subjects of the same or similar population since we do not have the original data.

According the phenotype correlation [8], the covariance of phenotype data of the c-th modality can be calculated by

$$\hat{\Sigma}_{YYc} = \frac{\mathbf{Y}_c^T \mathbf{Y}_c}{n-1} = corr\left(\mathbf{B}_c\right). \tag{4}$$

Let $\mathbf{s}_c^*$, $\mathbf{w}_c^*$ and $\mathbf{v}_c^*$ denote the final results, we will present how to solve them only using GWAS results ($\mathbf{B}$) and several subjects of a public reference database.

Solving S and W: Since our method is bi-convex, we can solve one variable by fixing the remaining variables as constants. The model of S-DMTSCCA and DMTSCCA are the same [4], and thus we can solve each $\hat{\mathbf{s}}_c$ and $\hat{\mathbf{w}}_c$ by substituting Σ_{XX}, Σ_{YY}, Σ_{XY} and Σ_{YX}. Specifically, we have the following closed-form solution,

$$\hat{\mathbf{s}}_c = \frac{(n-1)\mathbf{B}_c\mathbf{v}_c}{(n-1)\hat{\Sigma}_{XX}+\frac{\lambda_s}{\kappa_c}\mathbf{D}} = \frac{\mathbf{B}_c\mathbf{v}_c}{\hat{\Sigma}_{XX}+\frac{\lambda_s}{\kappa_c}\tilde{\mathbf{D}}}, \tag{5}$$

$$\hat{\mathbf{w}}_c = \frac{(n-1)\mathbf{B}_c\mathbf{v}_c}{(n-1)\hat{\Sigma}_{XX}+\frac{\lambda_w}{\kappa_c}\breve{\mathbf{D}}_c} = \frac{\mathbf{B}_c\mathbf{v}_c}{\hat{\Sigma}_{XX}+\frac{\lambda_w}{\kappa_c}\tilde{\mathbf{D}}_c}. \tag{6}$$

In both equations, $\mathbf{D}$ and $\tilde{\mathbf{D}}$ are diagonal matrices, and their i-th diagonal element are $\frac{1}{2\|\mathbf{s}^i\|_2}$ and $\frac{1}{2(n-1)\|\mathbf{s}^i\|_2}$ for $i = 1, ..., p$ respectively. $\breve{\mathbf{D}}_c$ and $\tilde{\mathbf{D}}_c$ are also diagonal matrices, and their i-th diagonal element are $\frac{1}{2|w_{ic}|}$ and $\frac{1}{2(n-1)|w_{jc}|}$ for $i = 1, ..., p$.

Finally, to satisfy the equality constraints, $\mathbf{S}$ and $\mathbf{W}$ are respectively scaled by

$$\mathbf{s}_c^* = \frac{\hat{\mathbf{s}}_c}{\sqrt{(n-1)(\hat{\mathbf{s}}_c+\hat{\mathbf{w}}_c)^T\hat{\Sigma}_{XX}(\hat{\mathbf{s}}_c+\hat{\mathbf{w}}_c)}},\ \mathbf{w}_c^* = \frac{\hat{\mathbf{w}}_c}{\sqrt{(n-1)(\hat{\mathbf{s}}_c+\hat{\mathbf{w}}_c)^T\hat{\Sigma}_{XX}(\hat{\mathbf{s}}_c+\hat{\mathbf{w}}_c)}}. \tag{7}$$

Solving V: If $\mathbf{S}$ and $\mathbf{W}$ are solved, we can fix them to solve for $\mathbf{V}$. In line with Eqs. (5–6), substituting Eq. (3) and Eq. (4) into the equation of $\hat{\mathbf{v}}_c$ in [4], we can get the solution formulas for $\mathbf{V}$, i.e.,

$$\hat{\mathbf{v}}_c = \frac{\mathbf{B}_c^T(\mathbf{s}_c+\mathbf{w}_c)}{\hat{\Sigma}_{YYc}+\frac{\lambda_v}{\kappa_c}\mathbf{Q}_c},\ \mathbf{v}_c^* = \frac{\hat{\mathbf{v}}_c}{\sqrt{(n-1)\hat{\mathbf{v}}_c^T\hat{\Sigma}_{YYc}\mathbf{v}_c}}. \tag{8}$$

$\mathbf{Q}$ is a diagonal matrix with the j-th element being $\frac{1}{2(n-1)|v_{jc}|}$ $(j = 1, ..., q)$.

Now we have obtained all the solutions to DMTSCCA without using original imaging and genetic data. In contrast, we use the GWAS summary statistics to obtain the covariance between imaging QTs and SNPs. The in-set covariance Σ_{YY} can also be calculated based on the results of GWAS. The in-set covariance Σ_{XX} can be approximated using subjects of the same population. In this paper, we used the public 1000 genome project (1kGP) database to generate $\hat{\Sigma}_{XX}$. In practice, using $\hat{\Sigma}_{XX}$ of the same population could yield acceptable results [1,13]. Therefore, our S-DMTSCCA is quite meaningful since it does not depend on raw neuroimaging genetic data.

3 Experiment and Results

We conducted two kinds of experiments to evaluate S-DMTSCCA. First, we used the ADNI data set where the original brain imaging phenotypes and genotypes are available. Specifically, our method cannot access individual-level imaging and genetic data. Instead, S-DMTSCCA can only work on the GWAS summary statistics obtained from this ADNI data set. At the same time, DMTSCCA directly ran on the original imaging and genetic data. By comparison, we can observe the performance difference between S-DMTSCCA and DMTSCCA. This can help evaluate the usefulness of our method. Second, we ran our method on a public GWAS result which studied the associations of imaging phenotypes and SNPs from the UK Biobank.

To find the best parameters for λ_s, λ_w and λ_v in DMTSCCA, we employed the grid search strategy with a moderate candidate parameter range $10^i(i = -5, -4, ..., 0, ..., 4, 5)$. Since S-DMTSCCA takes summary statistics as the input data without using the individual data, the conventional regularization parameters procedure is impracticable. Therefore, we used the grid search method with the same range based on the data set whose individual-level data was accessible to find the optimal parameters for S-DMTSCCA. Besides, to ensure equal optimization for each imaging modality, we used the same constants for the task weight parameters κ_c in different sub-tasks.

We used the 1kGP data sets as the reference samples. By conducting whole-genome sequencing on individuals from a range of ethnicity, the 1kGP institute obtains an extensive collection of human prevalent genetic variants [3]. We used individuals of British in England and Scotland (GBR) from 1kGP (release 20130502) to compute $\hat{\Sigma}_{XX}$ in S-DMTSCCA since UK Biobank GWAS results from the European ancestors. All experiments ran on the same platform, and all methods employed the same stopping condition, i.e. both $\max_c |(\mathbf{s}_c + \mathbf{w}_c)^{t+1} - (\mathbf{s}_c + \mathbf{w}_c)^t| \leq 10^{-5}$ and $\max_c |\mathbf{v}_c^{t+1} - \mathbf{v}_c^t| \leq 10^{-5}$.

3.1 Study on the ADNI Dataset

Data Source. The individual-level brain genotype and imaging data we used were downloaded from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). One goal of ADNI is to investigate the feasibility of utilizing a multi-modal approach that combines serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment to measure the progression of Alzheimer's disease (AD). For the latest information, see https://www.adni-info.org.

We used three modalities, i.e. the 18-F florbetapir PET (AV45) scans, fluorodeoxyglucose positron emission tomography (FDG) scans, and structural MRI (sMRI) scans. These data had been aligned to the same visit of each subject. The sMRI data were analyzed with voxel-based morphometry (VBM) by SPM. All scans were aligned to a T1-weighted template image, segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) maps, normalized to the standard MNI space, and smoothed with an 8 mm FWHM kernel. Additionally, the FDG and AV45 scans were registered into the same MNI space. Then 116 regions of interest (ROIs) level measurements were extracted based on the MarsBaR automated anatomical labeling (AAL) atlas. These 116 imaging QTs were pre-adjusted to remove the effects of the baseline age, gender, education, and handedness by the regression weights generated from healthy controls. The genotype data were also from the ADNI database. Specifically, we studied 5000 SNPs of chromosome 19: 46670909 - 46167652 including the well-known AD risk genes such as *APOE* [11]. As a methodology paper, we aimed to develop a GWAS summary statistics-based imaging genetics method that could handle GWAS summary statistics rather than individual-level imaging and genetic data. Thus, although using a large number of SNPs could be more interesting, it might go beyond the key research topic of this paper. The goal of the method was to explore the relationships between the multiple modalities of QTs (GM densities for VBM scans, amyloid values for AV45 scans and glucose utilization for FDG scans) and SNPs. For clarity, we respectively denoted the canonical weights of imaging QTs for AV45-PET, FDG-PET, and VBM-MRI as $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$.

Biomarkers Identification. We presented the identified SNPs and imaging QTs for each imaging modality based on the estimated canonical weights in Fig. 1 (a, b, and c). DMTSCCA and S-DMTSCCA decomposed the canonical weight of SNPs into two components, i.e., the multi-modal shared component $\mathbf{S}$ and modality-specific component $\mathbf{W}$. Thus, we presented both of them here. Both $\mathbf{S}$ and $\mathbf{W}$ of S-DMTSCCA identified the famous AD-related SNP rs429358, and the top SNPs marked in this figure were all related to AD, demonstrating the effectiveness of our method. In addition, S-DMTSCCA presented a task-consistent pattern, indicating that SNPs such as rs12721051 (*APOC1*), rs56131196 (*APOC1*) and rs44203638 (*APOC1*) could contribute to all three imaging modalities. S-DMTSCCA also presented a task-specific pattern, including rs7256200 (*APOC1*), rs73052335 (*APOC1*) and rs10414043 (*APOC1*) which were associated with Av45 and FDG scans, rs483082 (*APOC1*) which was associated with Av45 and VBM-MRI scans, rs12721046 (*APOC1*) and rs5117 (*APOC1*) which were only associated with VBM imaging scans. Importantly, our method identified the same top SNPs as the conventional DMTSCCA, suggesting that S-DMTSCCA possesses an equivalent feature selection capacity to DMTSCCA. In the heat maps of imaging QTs, S-DMTSCCA was able to identify different biomarkers for different scans as the conventional one. For example, the *Frontal-Med-Orb-Right* and *Frontal-Med-Orb-Left* of AV45-PET scans, *Cingulum-Post-Right* of FDG-PET scans and *Hippocampus-Left* of VBM-MRI scans. All in all, S-DMTSCCA presented a good agreement with the conventional method in feature selection. These results demonstrated that our method could be a very promising and meaningful method in multi-modal brain imaging genetics since it did not use individual-level brain imaging genetic data.

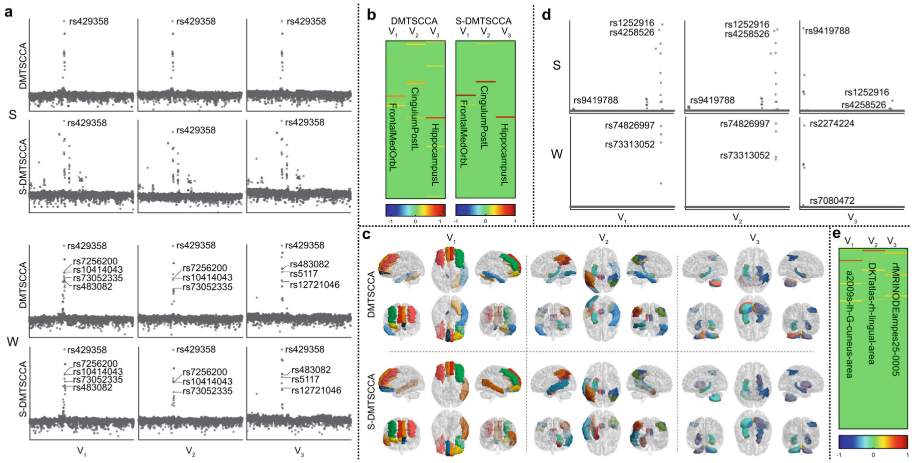

Fig. 1. Comparison of canonical weights from DMTSCCA and S-DMTSCCA when applied to ADNI data set (**a**, **b** and **c**), and Comparison of canonical weights when applied to IDPs GWAS (**d** and **e**). **a** The canonical weights ($70 + \log_2 |\mathbf{u}|$) of SNPs. **b** The canonical weights of QTs. **c** Visualization of the top 10 identified ROIs mapped on the brain, and the different ROIs marked with different colors. This sub-figure was drawn by BrainNet Viewer toolbox [16]. **d** The canonical weights of SNPs. **e** The canonical weights of QTs. Within each sub-figure, there are three columns for three modalities, i.e., $\mathbf{v}_1$ for AV45, $\mathbf{v}_2$ for FDG, and $\mathbf{v}_3$ for VBM.

Bi-multivariate Associations. We summarized the CCCs of conventional DMTSCCA and our method in Table 1. The values here represented the strength of the identified correlations. Since we have 3 imaging modalities in this section, 3 groups of CCCs were shown. Firstly, we can see that all CCCs of S-DMTSCCA were comparable to those of DMTSCCA, and all CCCs were relatively high (>0.2). These results indicated that our method can identify equivalent bi-multivariate associations between genetic variants and multi-modal phenotypes without using individual-level data.

Table 1. CCCs values between SNPs and three modalities imaging QTs.

Model	SNP-$\mathbf{v}_1$	SNP-$\mathbf{v}_2$	SNP-$\mathbf{v}_3$
DMTSCCA	0.4722	0.3306	0.2450
S-DMTSCCA	0.4559	0.3122	0.2018

3.2 Application to Summary Statistics from Brain Imaging GWAS

Summary Statistics from GWAS. The GWAS summary data were associations between brain imaging-derived phenotypes (IDPs) and SNPs. This GWAS studied a comprehensive set of imaging QTs derived from different types of brain imaging data of 8428 individuals from the UK Biobank database [7]. We used three modalities of imaging QTs, including two structural volumetric measurements obtained by FreeSurfer, i.e., the Desikan-Killiany-Tourville atlas and the Destrieux atlas, and one from resting-state functional MRI (rfMRI). We used summary statistics of these three modalities of the whole brain and denoted their canonical weights as $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ respectively. In particular, there were 148,64 and 76 imaging QTs for three imaging modalities. Genotype data for these imaging QTs were obtained from the UKB database [2]. We used 5000 SNPs in chromosome 10: 95952196 - 96758136 and chromosome 14: 59072418 - 59830880. We aimed to evaluate our method with the expectation to gain a comprehensive understanding of the genetic basis of multi-modal brain imaging phenotypes.

Biomarkers Identification. Figure 1 (d and e) presented the heat maps of the identified SNPs and imaging QTs for all three modalities, with modality-consistent and modality-specific SNPs separately shown. In this figure, **S** identified rs9419788 (*PLCE1*) in chromosome 10, and rs1252916 (*DAAM1*) and rs4258526 (*LINC01500*) in chromosome 14. This implied that these loci are shared by all three modalities. In addition, in heat maps of **W**, rs7080472 (*PLCE1*) only can be identified by rfMRI scans, and rs74826997 (*LINC01500*) and rs73313052 only can be identified by the other two scans. All these identified biomarkers shared by three modalities or related to a specific modality were consistent with the results of GWAS. This suggested that S-DMTSCCA can select leading genetic variations contributing to related imaging QTs. From the heat maps of imaging QTs, our method simultaneously identified the specific related imaging QTs for a specific task while GWAS cannot. These specific patterns included the *a2009s-lh-G-cuneus-area* for a2009s imaging QTs, the *DKTatlas-rh-lingual-area* for DKT IDPs, and *rfMRI-NODEampes25-0005* for rfMRI scan. In summary, these results demonstrated that S-DMTSCCA can simultaneously identify the important genetic variants and imaging phenotypes of multiple modalities only depending on summary statistics data.

Bi-multivariate Associations. In addition to feature selection, we calculated CCCs of S-DMTSCCA. The values for the three modalities were 0.1455, 0.1354, and 0.1474 respectively. This indicated that S-DMTSCCA could identify substantial bi-multivariate associations for each modality which might be attributed to its good modeling capability. These results again suggested that S-DMTSCCA can work well with summary statistics.

4 Conclusion

In brain imaging genetics, DMTSCCA can identify the genetic basis of multimodal phenotypes. However, DMTSCCA depends on individual-level genetic and imaging data and thus was infeasible when the raw data cannot be obtained. In this paper, we developed a source-free S-DMTSCCA method using GWAS summary statistics rather than the original imaging and genetic data. Our method had the same modeling ability as the conventional methods. When applied to multi-modal phenotypes from ADNI, S-DMTSCCA showed an agreement in feature selection and canonical correlation coefficient results with DMTSCCA. When applied to multiple modalities of GWAS summary statistics, our method can identify important SNPs and their related imaging QTs simultaneously. In the future, it is essential to consider the pathway and brain network information in our method to identify higher-level biomarkers of biological significance.

References

1. Barbeira, A.N., et al.: Exploring the phenotypic consequences of tissue specific gene expression variation inferred from gwas summary statistics. Nat. Commun. **9**(1), 1825 (2018)
2. Bycroft, C., et al.: The UK Biobank resource with deep phenotyping and genomic data. Nature **562**(7726), 203–209 (2018)
3. Consortium,G.P., et al.: A global reference for human genetic variation. Nature **526**(7571), 68 (2015)
4. Du, L., et al.: Associating multi-modal brain imaging phenotypes and genetic risk factors via a dirty multi-task learning method. IEEE Trans. Med. Imaging **39**(11), 3416–3428 (2020)
5. Du, L., et al.: Fast multi-task scca learning with feature selection for multi-modal brain imaging genetics. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 356–361. IEEE (2018)
6. Du, L., et al.: Multi-task sparse canonical correlation analysis with application to multi-modal brain imaging genetics. IEEE/ACM Trans. Comput. Biol. Bioinf. **18**(1), 227–239 (2019)
7. Elliott, L.T., et al.: Genome-wide association studies of brain imaging phenotypes in UK Biobank. Nature **562**(7726), 210–216 (2018)
8. Li, T., Ning, Z., Shen, X.: Improved estimation of phenotypic correlations using summary association statistics. Front. Genet. **12**, 665252 (2021)
9. Manolio, T.A., et al.: Finding the missing heritability of complex diseases. Nature **461**(7265), 747–753 (2009)
10. Marouli, E., et al.: Rare and low-frequency coding variants alter human adult height. Nature **542**(7640), 186–190 (2017)
11. Ramanan, V.K., et al.: APOE and BCHE as modulators of cerebral amyloid deposition: a florbetapir pet genome-wide association study. Mol. Psychiatry **19**(3), 351–357 (2014)
12. Shen, L., Thompson, P.M.: Brain imaging genomics: integrated analysis and machine learning. Proc. IEEE **108**(1), 125–162 (2019)
13. Turley, P., et al.: Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat. Genet. **50**(2), 229–237 (2018)

14. Uffelmann, E., et al.: Genome-wide association studies. Nat. Rev. Methods Primers **1**(1), 59 (2021)
15. Wei, K., Kong, W., Wang, S.: An improved multi-task sparse canonical correlation analysis of imaging genetics for detecting biomarkers of Alzheimer's disease. IEEE Access **9**, 30528–30538 (2021)
16. Xia, M., Wang, J., He, Y.: BrainNet Viewer: a network visualization tool for human brain connectomics. PLoS ONE **8**(7), e68910 (2013)
17. Zhuang, X., Yang, Z., Cordes, D.: A technical review of canonical correlation analysis for neuroscience applications. Hum. Brain Mapp. **41**(13), 3807–3833 (2020)