Transcriptome-wide association analysis of 211 neuroimaging traits identifies new genes for brain structures and yields insights into the gene-level pleiotropy with other complex traits

Running title: TWAS of brain structures

Bingxin Zhao[1,6], Yue Shan[1,6], Yue Yang[1], Tengfei Li[2,3], Tianyou Luo[1], Ziliang Zhu[1], Yun Li*[1,4,5,7], and Hongtu Zhu*[1,3,7]

[1]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[2]Department of Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[3]Biomedical Research Imaging Center, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[4]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[5]Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[6] These authors contributed equally: Bingxin Zhao, Yue Shan.

[7] These authors jointly supervised this work: Yun Li, Hongtu Zhu.

**Corresponding authors*:

Yun Li

5090 Genetic Medicine Building, 120 Mason Farm Road, Chapel Hill, NC 27599.

E-mail address: yun_li@med.unc.edu Phone: (919) 843-2832

Hongtu Zhu

3105C McGavran-Greenberg Hall, 135 Dauer Drive, Chapel Hill, NC 27599.

E-mail address: htzhu@email.unc.edu Phone: (919) 966-7250

List of Alzheimer's Disease Neuroimaging Initiative (ADNI) and Pediatric Imaging, Neurocognition and Genetics (PING) authors provided in the supplemental materials.

Abstract

Structural and microstructural variations of human brain are heritable and highly polygenic traits, with hundreds of associated genes founded in recent genome-wide association studies (GWAS). Using gene expression data, transcriptome-wide association studies (TWAS) can prioritize these GWAS findings and also identify novel gene-trait associations. Here we performed TWAS analysis of 211 structural neuroimaging phenotypes in a discovery-validation analysis of six datasets. Using a cross-tissue approach, TWAS discovered 204 associated genes (86 new) exceeding Bonferroni significance threshold of $1.37*10^{-8}$ (adjusted for testing multiple phenotypes) in the UK Biobank (UKB) cohort, and validated 18 TWAS or previous GWAS-detected genes. The TWAS-significant genes of brain structures had been linked to a wide range of complex traits in different domains. Additional TWAS analysis of 11 cognitive and mental health traits detected 69 overlapping significant genes with brain structures, further characterizing the genetic overlaps among these brain-related traits. Through TWAS gene-based polygenic risk scores (PRS) prediction, we found that TWAS PRS gained substantial power in association analysis compared to conventional variant-based PRS, and up to 6.97% of phenotypic variance (p-value=$7.56*10^{-31}$) in testing datasets can be explained by UKB TWAS-derived PRS. In conclusion, our study illustrates that TWAS can be a powerful supplement to traditional GWAS in imaging genetics studies for gene discovery-validation, genetic co-architecture analysis, and polygenic risk prediction.

Keywords: Gene expression; Cross-tissue TWAS; Regional brain volumes; Diffusion tensor imaging; UK Biobank.

Brain structural and microstructural differences are phenotypically associated with many other complex traits across different categories, such as cognitive measures[1-5], neurodegenerative/neuropsychiatric traits[6-9], alcohol and tobacco consumption[10], and physical bone density[11]. Structural variations of human brain can be quantified by multimodal magnetic resonance imaging (MRI). Specifically, the T1-weighted MRI (T1-MRI) can provide basic morphometric information of brain tissues, such as volume, surface area, sulcal depth, and cortical thickness. In region of interest (ROI)-based T1-MRI analysis, images are annotated onto ROIs of pre-defined brain atlas, and then both global (e.g., whole brain, gray matter, white matter) and local (e.g., basal ganglia structures, limbic and diencephalic regions) markers can be generated to measure the brain anatomy. On the other hand, diffusion MRI (dMRI) can capture local tissue microstructure through the random movement of water. Using diffusion tensor imaging (DTI) models, brain structural connectivity can be quantified by using white matter tracts extracted from dMRI, which build psychical connections among brain ROIs and are involved in connected networks for various brain functions[12,13]. See Miller, et al. [11] and Elliott, et al. [14] for a global overview and more information about neuroimaging modalities used in the present study.

Structural neuroimaging traits have shown moderate to high degree of heritability in both twin and population-based studies[14-24]. In the past ten years, genome-wide association studies (GWAS)[3,14,24-33] have been conducted to identify the associated genetic variants (typically single-nucleotide polymorphisms [SNPs]) for brain structures. A highly polygenic[34,35] genetic architecture has been observed, indicating that a large number of genetic variants contribute to the brain structure variations measured by neuroimaging biomarkers[21,36]. Particularly, using data from the UK Biobank (UKB[39]) cohort, two recent large-scale GWAS have identified 578 associated genes for 101 regional brain volumes derived from T1-MRI[37] (referred as ROI volumes, n=19,629) and 110 DTI parameters of dMRI[38] (referred as DTI parameters, n=17,706). Some of these discovered genes had been implicated with the same or other traits such as cognition and mental health diseases/disorders in previous GWAS. However, most of them have not been verified and need further investigations. As a supplement to traditional GWAS, recent advances of gene expression imputation methods[40-46] and developments of

reference databases (e.g., the Genotype-Tissue Expression (GTEx) project[47]) have put the transcriptome-wide association studies (TWAS) forward for gene-trait association analysis. Despite some challenges[48] such as interpreting causality, TWAS has successfully discovered novel gene-trait associations and provided new insights into biological mechanisms for many complex traits[49]. Through imputed transcriptomes, TWAS can reduce the multiple testing burden and leverage gene expression data to increase testing power for gene-trait association detection. This is a particularly desirable feature for imaging genetics studies, for which most of neuroimaging GWAS datasets continue to have small sample sizes and heavy multiple testing burden[50].

Here we applied TWAS methods to 211 structural neuroimaging traits including 101 ROI volumes and 110 DTI parameters. As these brain-related traits tend to be highly polygenic[21,36] and are related with many traits across different categories[11], we used a cross-tissue (panel) TWAS approach (UTMOST[42]) in our main analysis. UTMOST first performs single-tissue gene-trait association analysis in each reference panel with both within-tissue and cross-tissue statistical penalties, and then combines these single-tissue results using the Generalized Berk-Jones (GBJ) test[51], which is aware of tissue-dependence and can account for the potential sharing of local expression regulation across tissues. The UKB dataset was used in the discovery phase (n=19,629 for ROI volumes and 17,706 for DTI parameters, respectively). For the same UKB cohort, we compared TWAS-significant genes to previous GWAS findings in gene-based association analysis via MAGMA[52] and gene-level functional mapping and annotation results by FUMA[53]. The UKB TWAS results were validated in five independent data sources, including Philadelphia Neurodevelopmental Cohort (PNC[54],n=537), Alzheimer's Disease Neuroimaging Initiative (ADNI[55], n=860), Pediatric Imaging, Neurocognition, and Genetics (PING[56], n=461), the Human Connectome Project (HCP[57], n=334), and the ENIGMA2[24] and ENIGMA-CHARGE collaboration[33] (n=13,193, for 8 ROI volume traits, referred as ENIGMA in this paper). Additional TWAS analysis was performed on 11 cognitive and mental traits to explore their genetics overlaps with brain structures. Chromatin interaction enrichment analysis and drug-target lookups were conducted for TWAS-significant genes. Finally, we developed TWAS gene-based polygenic risk scores[58]

(PRS) using FUSION[40] to fully assess polygenic architecture and examine the predictive ability of the UKB TWAS results.

RESULTS

Overview of TWAS discovery-validation in the six datasets

We conducted a two-phase discovery-validation TWAS analysis for 211 neuroimaging traits by using the UKB cohort for discovery and the other datasets (ADNI, HCP, PING, PNC, and ENIGMA) for validation. We applied the UTMOST gene expression imputation models trained on 44 GETx (v6) reference panels, and used GWAS summary statistics generated from previous GWAS as inputs. In the rest of this paper, we refer $1.37*10^{-8}$ (that is, $5*10^{-2}$/17,290/211, adjusted for all candidate genes and traits performed) as the significance threshold for gene-trait associations unless otherwise stated.

The UKB discovery phase identified 614 significant gene-trait associations (**Supplementary Table 1**) between 204 genes and 135 neuroimaging traits (53 ROI volumes, 82 DTI parameters). Of the 204 TWAS-significant genes, 61 (29.9%) had significant associations with more than two neuroimaging traits, 25 (12.3%) had more than five significant associations, and 12 (5.9%) had at least ten, including *OSER1, XRCC4, PLEKHM1, ZKSCAN4, EIF4EBP3, MAPT, LRRC37A, CRHR1, FOXF1, TREH, ARHGAP27,* and *C6orf100*. These 12 genes together contributed 195 (31.8%) of the 614 gene-trait associations, indicating their widespread influences on brain structures. Specifically, we identified 123 genes whose imputed gene expression levels were significantly associated with one of more of the 53 ROI volumes (215 associations in total, 115 new, **Supplementary Fig. 1**), and 103 significantly associated genes (22 overlapping) for one or more of the 82 DTI parameters (399 associations in total, 219 new, **Supplementary Fig. 2**). **Figure 1** illustrates that TWAS prioritized previous GWAS findings of MAGMA and FUMA and also discovered many new associations and genes. Moreover, some genes were associated with both ROI volumes and DTI parameters, while others were more specifically related to certain structures (**Supplementary Fig. 3**). For example, *XRCC4, ZKSCAN4, EIF4EBP3,* and *CD14* were associated with DTI parameters but not ROI volumes, *DEFB124, COX4I2, HCK, HM13*, and *REM1* showed associations with putamen and pallidum volumes, and the associations of *PLEKHM1, LRRC37A, MAPT, CNNM2,*

NT5C2, ARHGAP27, and *CRHR1* were spread widely across DTI parameters and total brain volume.

We validated the UKB results in the other five independent cohorts. For each dataset, we applied the Bonferroni-corrected significance threshold accounting for all candidate genes and traits analyzed (that is, $5*10^{-2}$/17,290/number of traits, **Supplementary Tables 2-6**). We found that 13 UKB TWAS-significant genes and 5 more previous GWAS-significant genes can be validated in one or more of the five validation datasets (**Supplementary Fig. 4**) including *ANKRD42, DCC, DCTPP1, DLGAP5, HCK, LGALS3, UBE2C, KLRD1, LRRC37A, OSER1, PRPF3, TREH, TGM7, NUP210L, DOK5, KRTAP5-1, C20orf166,* and *DPP4*. The TWAS novel findings and validated genes were discussed further in details below.

Novel TWAS discoveries and validated genes

Of the 204 UKB TWAS-significant genes, 90 were not discovered in previous GWAS of the same UKB dataset (**Supplementary Table 7**). TWAS resulted in 60 new associated genes for 53 ROI volumes (106 associations, **Supplementary Fig. 5**), and 52 new genes for 82 DTI parameters (139 associations, **Supplementary Fig. 6**). According to NHGRI-EBI GWAS catalog[59], the 90 TWAS-significant genes replicated four previous findings on brain structures, including *JPH3*[60] for hippocampal volume in mild cognitive impairment, *CNNM2*[61] for white matter lesion progression, *FOXF1*[62] for hippocampal volume in Alzheimer's disease progression, and *C1QL1*[63] for white matter hyperintensity burden. The other 86 genes had not been linked to brain structure previously and thus can be regarded as novel genes for these 211 neuroimaging traits. To explore the genetic overlaps with other traits in different domains, we performed association lookups for the 90 TWAS-significant genes on the NHGRI-EBI GWAS catalog (**Supplementary Table 8**)**. Figure 2** shows that these genes were widely associated with physical measures (e.g., height, waist-to-hip ratio, heel bone mineral density, body mass index), cognitive traits (e.g., cognitive function, intelligence, math ability), neuropsychiatric and neurodegenerative diseases/disorders (e.g., schizophrenia, bipolar disorder, Alzheimer's disease), coronary artery disease, mean corpuscular hemoglobin, neuroticism,

education, reaction time, chronotype, smoking behavior and alcohol use, such as *CDK2AP1*[64-67], *ELL*[68-70], *CTTNBP2*[71-73], and *SH2B1*[72,74-76].

For the 18 TWAS-validated genes shown in **Supplementary Fig. 4**, 8 (*ANKRD42, DCC, LRRC37A, NUP210L, DOK5, KRTAP5-1, C20orf166,* and *DPP4*) of them had been discovered in the previous UKB GWAS and were implicated in brain-related complex traits, such as neuroticism[64], major depression[77], schizophrenia[78-80], Intelligence[81], math ability[73], reaction time[75], and insomnia[82]. The left ten genes, which were novel findings of TWAS, also had known associations with many cognitive and mental health traits. For example, previous GWAS reported that *HCK* was associated with chronotype[82], *LGALS3* with schizophrenia[83], *UBE2C* with reaction time[75], *KLRD1* with adolescent idiopathic scoliosis[84], *OSER1* with cognitive performance[77] and Alzheimer's disease[76], and *PRPF3* with chronotype[76,85] and neuropsychiatric disorders[86]. In summary, TWAS novel and validated genes expand the overview of gene-level pleiotropy across these traits, suggesting that neuroimaging-derived biomarkers could be useful in studying a wide range of complex traits.

Compared to brain tissue-specific TWAS analysis

As a comparison, we performed a brain tissue-specific version of TWAS that only combines brain tissues in UTMOST (Method). This brain tissue-specific TWAS detected 308 significant gene-trait associations (**Supplementary Table 9**) between 107 unique genes and 96 neuroimaging traits, including 64 associated genes for one or more of 37 ROI volumes (104 associations, **Supplementary Fig. 7**), and 53 genes (10 overlapping) for one or more of 59 DTI parameters (204 associations, **Supplementary Fig. 8**).

Most (101/107) of the tissue-specific genes have been identified by either the cross-tissue TWAS (95/107) or previous GWAS (70/107). The 6 genes that were uniquely identified by tissue-specific analysis included *KNCN, LHFPL3, MBD2, TBK1, C3orf62,* and *TMEM173. LHFPL3* showed associations with education[87], social behavior[88,89], cognitive ability[75], schizophrenia[90], and bipolar disorder[91]. *MBD2* was associated with reaction time[75], *TBK1* with amyotrophic lateral sclerosis[92,93], and *C3orf62* with intelligence[82]. Compared to tissue-specific TWAS, cross-tissue analysis clearly identified more signals.

For example, of the 215 gene-trait associations identified by cross-tissue analysis of ROI volumes, 100 had been identified in GWAS, 28 can be additionally identified by tissue-specific TWAS, and 87 can only be detected by cross-tissue analysis (**Supplementary Fig. 9**). Similarly, 180 of the 399 cross-tissue TWAS associations for DTI can be identified in GWAS, 69 can be additionally identified by tissue-specific TWAS, and 150 were cross-tissue TWAS only (**Supplementary Fig. 10**). These results illustrate the advantage of cross-tissue analysis over brain tissue-specific TWAS for discovering association signals that are difficult to be identified in traditional GWAS. We further compared their results in a few follow-up analyses below.

Comparison with GWAS variant-level signals and conditional analysis

For each of the 614 gene-trait associations detected in cross-trait TWAS, we used previous GWAS summary statistics to check the most significant variant within the gene region (with a 1MB window on each side) that was pinpointed in the same UKB dataset (Method). The GWAS p-value of the most significant variant was greater than $1*10^{-6}$ for any associations of 13 genes (**Supplementary Table 10**). None of them had been identified by MAGMA or FUMA, indicating that it can be difficult to detect these genes by GWAS or post-GWAS screening for any of these neuroimaging traits. Of the 13 genes, 7 (*OSER1, TREH, PRPF3, KLRD1, TGM7, DCTPP1, UBE2C*) were validated in one or more of the five validation datasets and were discussed in previous section. For the other 6 genes (*CELSR3, MYO9A, DNAJC24, GYPE, TMEM136, MOB4*) genes, *MOB4* was reported for major depression[94] and autism spectrum disorder/schizophrenia[95], *DNAJC24* was linked to adolescent idiopathic scoliosis[84], and *CELSR3* was associated with education[65] and cognitive ability[64,81]. The same checking was then performed for the 308 significant gene-trait associations of brain tissue-specific TWAS. We found that only one gene *DCTPP1* had minimum GWAS p-value greater than $1*10^{-6}$ (**Supplementary Table 11**).

We next performed a conditional analysis to see whether the TWAS signals remained significant after adjustment for the most significant genetic variant used in UTMOST gene expression imputation models (Method). Although our cross-tissue analysis combined information from many genetic variants across various human tissues, we found that 418 of the 614 associations may indeed be dominated by the strongest

GWAS signal of the imputation model, as their conditional p-values were larger than 0.05 (**Supplementary Table 12**). However, the conditional p-values of four genes (*XRCC4, OBFC1, C15orf56, NMT1*) were smaller than $1*10^{-6}$ for 18 gene-trait associations, suggesting that these associations were unlikely to be driven by a signal genetic variant. When the p-value threshold was relaxed to $1*10^{-3}$, 66 associations of 20 genes persisted after conditional analysis. The conditional analysis was also performed on significant associations of brain tissue-specific TWAS. Their conditional p-values were smaller than $1*10^{-6}$ for three genes (*XRCC4, C15orf56, NMT1*) with 15 associations, and were smaller than $1*10^{-3}$ for 10 genes with 42 associations (**Supplementary Table 13**).

Additional TWAS analysis for cognitive and mental health traits

To further explore the gene-level genetic overlaps among brain structure and other brain-related traits, we performed cross-tissue TWAS analysis for 11 cognitive and mental health traits (**Supplementary Table 14**). We found that 69 of the 204 TWAS-significant genes of neuroimaging traits were also significantly associated with one or more of the 11 cognitive and mental health traits (**Figure 3**). These results suggest the genes involved in brain structure changes are often also active in brain functions and mental disorder/diseases. For example, we found 33 overlapping genes with cognitive function, 32 with education, 26 with numerical reasoning, 25 with intelligence, 23 with neuroticism, 19 with drinking behavior, and 13 with schizophrenia. A large proportion (48/69) of these genes were associated with more than one cognitive or mental health traits, and 11 genes were linked to at least five traits, including *SCML4, C16orf54, DCC, NFATC2IP, NPIPB7, NPIPB9, SH2B1, CRHR1, LRRC37A, HIST1H2BO,* and *NKAPL,* indicating the high degree of statistical pleiotropy[96] of these genes.

Chromatin interaction enrichment analysis and drug-target lookups

To explore the biological interpretations of TWAS and GWAS-significant genes, we performed enrichment analysis in promoter-related chromatin interactions of four types of brain neurons[97] (iPSC-induced excitatory neurons, iPSC-derived hippocampal DG-like neurons, iPSC-induced lower motor neurons, and primary astrocytes), and also in high confident interactions of adult and fetal cortex[98] (Method). The raw p-values of Wilcoxon rank test for enrichment were summarized in **Supplementary Table 15.** We

found that cross-tissue TWAS-significant genes of the 11 cognitive and mental health were significantly enriched in chromatin interactions from all of the five validation datasets (p-value range=[$4.91*10^{-11,}$ $3.03*10^{-5}$]), suggesting that TWAS-significant genes actively interacted with other chromatin regions and played a more important role in regulating gene expressions as compared with other genes. The cross-tissue TWAS-significant genes of neuroimaging traits also showed significant enrichments (p-value range= [$1.38*10^{-3}$,$2.44*10^{-2}$]). Merging the two sets of genes resulted in smaller p-value in each dataset (p-value range=[$2.93*10^{-11,}$ $2.77*10^{-5}$]). The most significant enrichment was observed in iPSC-induced lower motor neurons. These results remained significant after adjusting for multiple testing by using Benjamini-Hochberg (B-H) procedure at 0.05 level (**Supplementary Table 16**). In contrast, GWAS-significant genes were only significantly enriched in primary astrocytes and high confident interactions (p-value range=[$5.11*10^{-3,}$ $1.48*10^{-2}$]), and brain tissue-specific TWAS-significant genes did not show any significant enrichments after B-H adjustment.

We carried out drug-target lookups using a recently published drug-target database[99] to see whether any of the TWAS and GWAS-significant genes were known targets of existing drugs. We focused on nervous system drugs with Anatomical Therapeutic Chemical (ATC) code started with "N", yielding 2,285 drug-gene pairs between 273 drugs and 241 targeted genes. We found that 12 TWAS-significant genes of the 11 cognitive and mental health traits were known targets for 64 drugs, including *CACNA1I, ESR1, ALDH2, CACNA1C, GRM2, KCNJ3, SCN3A, CACNA1D, KCNK3, CHRNA3, CHRNA6,* and *SLC6A4*. Of the 64 drugs, 27 were anti-depressants (ATC: N06A) to treat major depressive disorder and other conditions, and 10 were anti-psychotics (ATC: N05A) to manage psychosis such as schizophrenia and bipolar disorder (**Supplementary Table 17**). In addition, 3 more drug-target genes (*GABBR1, HTR2B, CREB1*) were detected by GWAS or TWAS of neuroimaging traits (**Supplementary Table 18**). These 3 genes were targets for 19 more drugs, 6 of which were anti-Parkinson drugs (ATC: N04) for treatment of Parkinson's disease and related conditions, and 5 were anti-migraine preparations (ATC: N02C) used in prophylaxis and treatment of migraine. These results may suggest that TWAS-significant genes could be considered as new targets in future drug development.

TWAS gene-based polygenic risk scores analysis

To fully assess the polygenic genetic architecture of neuroimaging traits and examine the predictive ability of UKB TWAS results, we constructed TWAS gene-based PRS on subjects in PNC, HCP, PING, and ADNI cohorts for all of the 211 neuroimaging traits (Method). The prediction analysis was conducted separately on 52 reference panels (13 GETx v7 brain tissues, 35 GTEx v7 other tissues, 1 non-GETx brain tissue, and 3 non-GETx other tissues) using the FUSION[40] software and database. We found that genetically predicted profiles for 28 ROI volumes (**Figure 5**) and 23 DTI parameters (**Supplementary Fig. 11**) were significantly associated with the corresponding observed traits in all testing datasets after Bonferroni correction (that is, 101*4+3*110=734 tests). Compared to previous SNP-based PRS analysis that yielded significant PRS profiles for 11 ROI volumes[37], gene-based PRS profiles were significant for more ROI volumes, such as left/right insula, left/right pallidum, left/right ventral DC, left/right fusiform, and left/right transverse temporal, suggesting the substantial power gain in association analysis of PRS. The significant TWAS PRS can account for 0.97%-6.97% phenotypic variance (p-value range=[$8.0*10^{-29}$, $6.81*10^{-5}$]) (**Supplementary Tables 19-20**), which was within the similar range to SNP-based PRS analysis. For example, the (incremental) R-squared of TWAS PRS of Cerebellar vermal lobules VIII–X was 6.97% in PNC and 6.48% in HCP, and the R-squared of SFO MD-derived TWAS PRS was 3.8% in PING and 2.41% in PNC. We also examined the performance of each reference panel on these significant traits. There was a significant linear relationship between the panel sample size and average prediction R-squared (48 GTEx reference panels, simple correlation=0.53, p-value=$1.21*10^{-4}$, **Supplementary Fig. 12**), which means that currently panel sample size may dominate the performance of TWAS PRS analysis regardless of the tissue specificity[58]. Among the brain tissue panels, we found that cerebellum tissue had the largest sample size and also showed the highest average R-squared (**Supplementary Table 21**), further supporting the importance of reference panel sample size.

DISCUSSION

In this study, we applied TWAS methods on 211 neuroimaging traits to identify genes, whose imputed expression levels were associated with brain structure variations. Using

a cross-tissue approach, our main discovery analysis identified 86 novel genes and validated 18 significant genes at stringent Bonferroni-correction p-value thresholds. Conditional analysis and comparison with GWAS variant-level results suggested that the identification and validation of new genes reflect the ability of TWAS to reduce the testing burden and to combine the small genetic variant effects. We also performed brain tissue-specific TWAS and illustrated the unique strengths of cross-tissue TWAS in conditional and enrichment analyses. Lots of brain structure-related genes were known genetic factors for a wide range of complex traits, ranging from physical traits, cognition, mental disease/disorders, blood assays, to lifestyle, which extend the potential applications of neuroimaging traits. Some of these genetic overlaps were additionally highlighted by a TWAS analysis of 11 cognitive and mental health traits.

The present study faces some limitations. First, since these results are purely based on statistical associations, it is hard to draw conclusions about the underlying causality and prioritize causal genes[42,100]. This is also one of the main challenges for most of the current TWAS approaches[48]. Follow-up experimental validation is a clear need to confirm TWAS results and pinpoint the causal genes of brain structure changes. Second, the brain tissue-specific TWAS did not yield much new results compared to the previous GWAS and brain tissue panels did not show better prediction accuracy than non-brain tissues in gene-based PRS analysis. Both of the two observations support the use of multiple tissues in our analysis to increase testing power for association analysis, but making the causality interpretation of TWAS results even more complicated. In addition, though gene-based PRS had much better power in association tests than SNP-based polygenic scores, their prediction accuracies were similar. These limitations may be due to the fact that currently brain tissue reference panels do not have large sample size and/or the associated gene expression imputations may have low quality. Despite these limitations, it is clear that TWAS have the potential to become a powerful supplement to traditional GWAS in imaging genetics studies. In our study, many new gene-trait associations were discovered and the underlying genetic overlaps among complex traits were largely expanded. With better brain tissue gene expression reference panels and more neuroimaging GWAS datasets available, future TWAS analyses of neuroimaging

traits are expected to show the value of tissue specificity and improve our understanding for the genetic basis of human brain.

ACKNOWLEDGEMENTS

This research was partially supported by U.S. NIH grants MH086633 (H.Z.) and MH116527 (TF.L.). We thank Quan Wang, Bingshan Li, and Jia Wen for helpful conversations. We thank the individuals represented in the UK Biobank, ADNI, HCP, PING, PNC, ENIGMA2, and ENIGMA-CHARGE datasets for their participation and the research teams for their work in collecting, processing and disseminating these datasets for analysis. This research has been conducted using the UK Biobank resource (application number 22783), subject to a data transfer agreement. We gratefully acknowledge all the studies and databases that made GWAS summary data available. Part of data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd; Janssen Alzheimer Immunotherapy Research & Development, LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Part of the data collection and

sharing for this project was funded by the Pediatric Imaging, Neurocognition and Genetics Study (PING) (U.S. National Institutes of Health Grant RC2DA029475). PING is funded by the National Institute on Drug Abuse and the Eunice Kennedy Shriver National Institute of Child Health & Human Development. PING data are disseminated by the PING Coordinating Center at the Center for Human Development, University of California, San Diego. Support for the collection of the PNC datasets was provided by grant RC2MH089983 awarded to Raquel Gur and RC2MH089924 awarded to Hakon Hakonarson. All PNC subjects were recruited through the Center for Applied Genomics at The Children's Hospital in Philadelphia. HCP data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

AUTHOR CONTRIBUTIONS

B.Z., Y.S., Y.L., and H.Z. designed the study. B.Z., Y.S., Y.Y., TF.L., TY.L., and Z.Z performed the experiments and analyzed the data. B.Z., Y.S., Y.L., and H.Z. wrote the manuscript with feedback from all authors.

COMPETETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. Ritchie, S.J. *et al.* Beyond a bigger brain: Multivariable structural brain imaging and intelligence. *Intelligence* **51**, 47-56 (2015).
2. Davies, G. *et al.* Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N= 112 151). *Molecular Psychiatry* **21**, 758–767 (2016).
3. van der Meer, D. *et al.* Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes. *Molecular Psychiatry*, in press. (2018).

4. Caldiroli, A. *et al.* The relationship of IQ and emotional processing with insula volume in schizophrenia. *Schizophrenia Research* **202**, 141-148 (2018).
5. Vreeker, A. *et al.* The relationship between brain volumes and intelligence in bipolar disorder. *Journal of Affective Disorders* **223**, 59-64 (2017).
6. Nir, T.M. *et al.* Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging. *NeuroImage: clinical* **3**, 180-195 (2013).
7. Bohnen, N.I. & Albin, R.L. White matter lesions in Parkinson disease. *Nature Reviews Neurology* **7**, 229 (2011).
8. Voineskos, A.N. Genetic underpinnings of white matter 'connectivity': heritability, risk, and heterogeneity in schizophrenia. *Schizophrenia research* **161**, 50-60 (2015).
9. Sudre, G. *et al.* Estimating the Heritability of Structural and Functional Brain Connectivity in Families Affected by Attention-Deficit/Hyperactivity Disorder. *JAMA psychiatry* **74**, 76-84 (2017).
10. Peng, P. *et al.* Brain structure alterations in respect to tobacco consumption and nicotine dependence: a comparative voxel-based morphometry study. *Frontiers in neuroanatomy* **12**, 43 (2018).
11. Miller, K.L. *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nature Neuroscience* **19**, 1523-1536 (2016).
12. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059-1069 (2010).
13. Hu, W., Zhang, A., Cai, B., Calhoun, V. & Wang, Y.-P. Distance canonical correlation analysis with application to an imaging-genetic study. *Journal of Medical Imaging* **6**, 026501 (2019).
14. Elliott, L.T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216 (2018).
15. Wen, W. *et al.* Distinct genetic influences on cortical and subcortical brain structures. *Scientific Reports* **6**, 32760 (2016).
16. den Braber, A. *et al.* Heritability of subcortical brain measures: a perspective for future genome-wide association studies. *NeuroImage* **83**, 98-102 (2013).

17. Eyler, L.T. *et al.* Conceptual and data-based investigation of genetic influences and brain asymmetry: a twin study of multiple structural phenotypes. *Journal of Cognitive Neuroscience* **26**, 1100-1117 (2014).
18. Blokland, G.A., de Zubicaray, G.I., McMahon, K.L. & Wright, M.J. Genetic and environmental influences on neuroimaging phenotypes: a meta-analytical perspective on twin imaging studies. *Twin Research and Human Genetics* **15**, 351-371 (2012).
19. Kremen, W.S. *et al.* Genetic and environmental influences on the size of specific brain regions in midlife: the VETSA MRI study. *Neuroimage* **49**, 1213-1223 (2010).
20. Jansen, A.G., Mous, S.E., White, T., Posthuma, D. & Polderman, T.J. What twin studies tell us about the heritability of brain development, morphology, and function: a review. *Neuropsychology Review* **25**, 27-46 (2015).
21. Zhao, B. *et al.* Heritability of regional brain volumes in large-scale neuroimaging and genetic studies. *Cerebral Cortex* **29**, 2904-2914 (2018).
22. Biton, A. *et al.* Polygenic architecture of human neuroanatomical diversity. *bioRxiv*, 592337 (2019).
23. Toro, R. *et al.* Genomic architecture of human neuroanatomical diversity. *Molecular Psychiatry* **20**, 1011-1016 (2015).
24. Hibar, D.P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224-229 (2015).
25. Hibar, D.P. *et al.* Novel genetic loci associated with hippocampal volume. *Nature Communications* **8**, 13624 (2017).
26. Franke, B. *et al.* Genetic influences on schizophrenia and subcortical brain volumes: large-scale proof of concept. *Nature Neuroscience* **19**, 420-431 (2016).
27. Guadalupe, T. *et al.* Human subcortical brain asymmetries in 15,847 people worldwide reveal effects of age and sex. *Brain imaging and behavior* **11**, 1497-1514 (2017).
28. Ikram, M.A. *et al.* Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nature Genetics* **44**, 539-544 (2012).
29. Bis, J.C. *et al.* Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nature Genetics* **44**, 545-551 (2012).

30. Grasby, K.L. *et al.* The genetic architecture of the human cerebral cortex. *bioRxiv*, 399402 (2018).
31. Hofer, E. *et al.* Genetic Determinants of Cortical Structure (Thickness, Surface Area and Volumes) among Disease Free Adults in the CHARGE Consortium. *bioRxiv*, 409649 (2018).
32. Satizabal, C.L. *et al.* Genetic Architecture of Subcortical Brain Structures in Over 40,000 Individuals Worldwide. *bioRxiv*, 173831 (2017).
33. Adams, H.H. *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nature neuroscience* **19**, 1569 (2016).
34. Boyle, E.A., Li, Y.I. & Pritchard, J.K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186 (2017).
35. Timpson, N.J., Greenwood, C.M.T., Soranzo, N., Lawson, D.J. & Richards, J.B. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nature Reviews Genetics* **19**, 110-124 (2017).
36. O'Connor, L.J. *et al.* Extreme polygenicity of complex traits is explained by negative selection. *The American Journal of Human Genetics* **105**, 456-476 (2019).
37. Zhao, B. *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. *Nature Genetics* **51**, 1637-1644 (2019).
38. Zhao, B. *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). *Molecular Psychiatry* (2019).
39. Sudlow, C. *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 Medicine* **12**, e1001779 (2015).
40. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nature genetics* **48**, 245 (2016).

41. Barbeira, A.N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nature communications* **9**, 1825 (2018).
42. Hu, Y. *et al.* A statistical framework for cross-tissue transcriptome-wide association analysis. (Nature Publishing Group, 2019).
43. Gamazon, E.R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics* **47**, 1091 (2015).
44. Zeng, P. & Zhou, X. Non-parametric genetic prediction of complex traits with latent Dirichlet process regression models. *Nature communications* **8**, 456 (2017).
45. Zhou, X. & Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nature methods* **11**, 407 (2014).
46. Nagpal, S. *et al.* TIGAR: An Improved Bayesian Tool for Transcriptomic Data Imputation Enhances Gene Mapping of Complex Traits. *The American Journal of Human Genetics* (2019).
47. Consortium, G. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660 (2015).
48. Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nature genetics* **51**, 592 (2019).
49. Zhang, W. Advancements of transcriptome imputation and related transcriptome-wide association studies. *Current Research in Biochemistry and Molecular Biology* **1**, 14-16 (2019).
50. Smith, S.M. & Nichols, T.E. Statistical challenges in “big data” human neuroimaging. *Neuron* **97**, 263-268 (2018).
51. Sun, R. & Lin, X. Set-based tests for genetic association using the generalized Berk-Jones statistic. *arXiv preprint arXiv:1710.02469* (2017).
52. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Computational Biology* **11**, e1004219 (2015).
53. Watanabe, K., Taskesen, E., Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).

54. Satterthwaite, T.D. *et al.* Neuroimaging of the Philadelphia neurodevelopmental cohort. *Neuroimage* **86**, 544-553 (2014).
55. Weiner, M.W. *et al.* The Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's & Dementia* **9**, e111-e194 (2013).
56. Jernigan, T.L. *et al.* The pediatric imaging, neurocognition, and genetics (PING) data repository. *Neuroimage* **124**, 1149-1154 (2016).
57. Somerville, L.H. *et al.* The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5–21 year olds. *NeuroImage* **183**, 456-468 (2018).
58. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nature Genetics* **50**, 538–548 (2018).
59. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research* **47**, D1005-D1012 (2018).
60. Chung, J. *et al.* Genome-wide association study of Alzheimer's disease endophenotypes at prediagnosis stages. *Alzheimer's & Dementia* **14**, 623-633 (2018).
61. Hofer, E. *et al.* White matter lesion progression: genome-wide search for genetic influences. *Stroke* **46**, 3048-3057 (2015).
62. Scelsi, M.A. *et al.* Genetic study of multimodal imaging Alzheimer's disease progression score implicates novel loci. *Brain* **141**, 2167-2180 (2018).
63. Verhaaren, B.F. *et al.* Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circulation: Cardiovascular Genetics* **8**, 398-409 (2015).
64. Lam, M. *et al.* Large-Scale Cognitive GWAS Meta-Analysis Reveals Tissue-Specific Neural Expression and Potential Nootropic Drug Targets. *Cell reports* **21**, 2597-2613 (2017).
65. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics* **48**, 624–633 (2016).

66. Hill, W. *et al.* A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry* **24**, 169–181 (2019).
67. Morris, J.A. *et al.* An atlas of genetic influences on osteoporosis in humans and mice. *Nature genetics* **51**, 258 (2019).
68. Kunkle, B.W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature genetics* **51**, 414 (2019).
69. Astle, W.J. *et al.* The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415-1429. e19 (2016).
70. Kim, S.K. Identification of 613 new loci associated with heel bone mineral density and a polygenic risk score for bone mineral density, osteoporosis and fracture. *PloS one* **13**, e0200785 (2018).
71. Howard, D.M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature neuroscience* **22**, 343 (2019).
72. Linnér, R.K. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics* **51**, 245-257 (2019).
73. Lee, J.J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* **50**, 1112–1121 (2018).
74. Shungin, D. *et al.* New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187 (2015).
75. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications* **9**, 2098 (2018).
76. Kichaev, G. *et al.* Leveraging polygenic functional enrichment to improve GWAS power. *The American Journal of Human Genetics* **104**, 65-75 (2019).
77. Wray, N.R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics* **50**, 668 (2018).

78. Li, Z. *et al.* Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nature Genetics* **49**, 1576-1583 (2017).
79. Periyasamy, S. *et al.* Association of schizophrenia risk with disordered niacin metabolism in an Indian genome-wide association study. *JAMA psychiatry* **76**, 1026-1034 (2019).
80. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421 (2014).
81. Savage, J.E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics* **50**, 912-919 (2018).
82. Jansen, P.R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nature Genetics* **51**, 394-403 (2019).
83. Lam, M. *et al.* Pleiotropic Meta-Analysis of Cognition, Education, and Schizophrenia Differentiates Roles of Early Neurodevelopmental and Adult Synaptic Pathways. *bioRxiv*, 519967 (2019).
84. Liu, J. *et al.* The coexistence of copy number variations (CNVs) and single nucleotide polymorphisms (SNPs) at a locus can result in distorted calculations of the significance in associating SNPs to disease. *Human genetics* **137**, 553-567 (2018).
85. Jones, S.E. *et al.* Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nature communications* **10**, 343 (2019).
86. Consortium, C.-D.G.o.t.P.G. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet* **381**, 1371-1379 (2013).
87. Rietveld, C.A. *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences* **111**, 13790-13794 (2014).
88. St Pourcain, B. *et al.* Variability in the common genetic architecture of social-communication spectrum phenotypes during childhood and adolescence. *Molecular autism* **5**, 18 (2014).

89. Day, F.R., Ong, K.K. & Perry, J.R. Elucidating the genetic basis of social interaction and isolation. *Nature communications* **9**, 2457 (2018).
90. Goes, F.S. *et al.* Genome-wide association study of schizophrenia in Ashkenazi Jews. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **168**, 649-659 (2015).
91. Hou, L. *et al.* Genetic variants associated with response to lithium treatment in bipolar disorder: a genome-wide association study. *The Lancet* **387**, 1085-1093 (2016).
92. Van Rheenen, W. *et al.* Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nature genetics* **48**, 1043 (2016).
93. Nicolas, A. *et al.* Genome-wide analyses identify KIF5A as a novel ALS gene. *Neuron* **97**, 1268-1283. e6 (2018).
94. Zhou, H. *et al.* Genetic risk variants associated with comorbid alcohol dependence and major depression. *JAMA psychiatry* **74**, 1234-1241 (2017).
95. Anney, R.J.L. *et al.* Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism* **8**, 21 (2017).
96. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics*, in press. (2019).
97. Song, M. *et al.* Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nature genetics* **51**, 1252 (2019).
98. Giusti-Rodriguez, P.M. & Sullivan, P.F. Using three-dimensional regulatory chromatin interactions from adult and fetal cortex to interpret genetic results for psychiatric disorders and cognitive traits. *BioRxiv*, 406330 (2019).
99. Wang, Q. *et al.* A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data. *Nature neuroscience* **22**, 691 (2019).
100. Ioannidis, N.M. *et al.* Gene expression imputation identifies candidate genes and susceptibility loci associated with cutaneous squamous cell carcinoma. *Nature communications* **9**, 4264 (2018).

101. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).

METHODS

GWAS summary statistics datasets

We made use of GWAS summary statistics to test for gene-trait associations in our TWAS study. The GWAS summary-level were from six studies, including the UK Biobank (UKB, http://www.ukbiobank.ac.uk/resources/) study, the Human Connectome Project (HCP, https://www.humanconnectome.org/) study, the Pediatric Imaging, Neurocognition, and Genetics (PING, http://www.chd.ucsd.edu/research/ping-study.html) study, the Philadelphia Neurodevelopmental Cohort (PNC, https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000607.v1.p1) study, the Alzheimer's Disease Neuroimaging Initiative (ADNI, http://adni.loni.usc.edu/data-samples/) study, and ENIGMA2 (GWAS of subcortical volumes) and the ENIGMA-CHARGE collaboration (http://enigma.ini.usc.edu/research/). For discovery, we used the GWAS summary statistics of the UKB study. Then the GWAS results of the other studies were used for validation, see **Supplementary Table 22** for a summary of sample size and the analyzed neuroimaging traits of each GWAS. More information about study cohorts and neuroimaging traits can be found in the original GWAS[24,33,37,38]. We also performed TWAS analysis for 11 cognitive and mental health traits, see **Supplementary Table 23** for these data resources.

Cross-tissue TWAS analysis by UTMOST

Cross-tissue TWAS analysis was performed for each trait using the UTMOST software (https://github.com/Joker-Jerome/UTMOST). We first run single-tissue association test for each of the 44 GTEx (v6) reference panels using the above GWAS summary statistics as input. There were 17,290 candidate genes considered in UTMOST. Second, the gene-trait associations in 44 panels (tissues) were combined by the GBJ test (https://cran.r-project.org/web/packages/GBJ/). We used the pre-trained cross-tissue imputation models and pre-calculated covariance matrices provided by UTMOST. For

the 211 neuroimaging traits in the UKB cohort, we also performed a brain-tissue specific version of UTMOST analysis that only combined brain tissues.

Comparison with previous GWAS findings

We compared TWAS-significant genes with those identified in the same UKB cohort by MAGMA gene-based association analysis and FUMA functional gene mapping analysis, which can be found in previous GWAS (Supplementary Tables 12 and 15 of Zhao, et al. [37] for ROI volumes and Supplementary Tables 14 and 16 of Zhao, et al. [38] for DTI parameters, respectively). For each significant gene-trait association, we also explored whether any genetic variant of this gene region (with 1MB window on both sides) had been linked to this neuroimaging trait by checking the smallest p-value in corresponding GWAS. For TWAS-significant genes that were not identified in GWAS, we used NHGRI-EBI GWAS catalog (version 2019-10-14, https://www.ebi.ac.uk/gwas/) to look for their reported associations with brain structure traits and any other traits. We summarized the traits that frequently reported for these genes, such as physical measures (e.g., height, waist-to-hip ratio, heel bone mineral density, body mass index), cognitive functions (such as general cognitive ability, cognitive performance), intelligence, educational attainment, math ability (such as highest math class taken and self-reported math ability), reaction time, neuroticism, neurodegenerative diseases (such as Alzheimer's disease and Parkinson's disease), neuropsychiatric disorders (such as major depressive disorder, schizophrenia, and bipolar disorder), coronary artery disease, and mean corpuscular hemoglobin.

Cross-tissue analysis conditional on the most significant GWAS signal

The TWAS gene expression imputation model can be viewed as a weighted sum of multiple genetic variants. If certain variant has a relatively large weight, the imputed gene expression could be driven by a single GWAS signal. In order to look at how many significant TWAS signals could be dominated by a single genetic variant, we rerun TWAS analysis in UKB cohort conditional on the most significant variant used in the UTMOST imputation model. First, for each reference panel, we considered a simple linear model

Phenotype ~ imputed gene expression + variant,

where the variant conditioned on was the most significant variant in previous GWAS of this phenotype in the same UKB cohort. Then, single-tissue conditional p-values of the imputed gene expression were combined by the GBJ test across the 44 GTEx reference panels.

Enrichment analyses and drug-target lookups

The chromatin interaction enrichments between significant and non-significant genes were tested using the Wilcoxon rank sum test. For the adult neural Promoter Capture Hi-C (PCHi-C), the enrichment of each gene was measured as the number of interactions overlapping gene with CHiCAGO Enrichment Score greater than 5[97]. The enrichment was tested separately in four cell types, including induced pluripotent stem cells (iPSC)-induced excitatory neurons, iPSC-derived hippocampal DG-like neurons, iPSC-induced lower motor neurons, and primary astrocytes. For the high confident interactions of adult and fetal cortex, the enrichment of each gene was measured as the sum of $-\log_{10}$(P-value) of all significant interactions overlapping the gene[98]. The drug-target lookups were conducted using the drug-gene associations reported in Wang, et al. [99]. We focused on nervous system drugs whose Anatomical Therapeutic Chemical code starts with "N" according to the DrugBank database (version 2019-07-02, https://www.drugbank.ca/atc).

Gene-based TWAS polygenic risk prediction

Gene-based polygenic profiles were created to assess the out-of-sample prediction power of the UKB TWAS results. In this analysis, we used the individual-level phenotype and genetic data, whose processing steps were detailed in previous GWAS[37,38]. The FUSION software and database (http://gusevlab.org/projects/fusion/) were used to impute gene expression levels in UKB, ADNI, HCP, PNC, and PING datasets using individual-level genetic data. We performed imputation for 52 different reference panels (**Supplementary Table 21**). In training data (UKB), we estimated the effect size of each imputed gene expression in a linear regression model, while adjusting for the age (at imaging), age-squared, sex, age-sex interaction, age-squared-sex interaction, as well as the top 40 genetic principle components (PCs) provided by UKB[101] (Data-Field 22009). For ROI volumes, we also included total brain volume (for ROIs other than total brain

volume itself) as a covariate. The gene-based PRS were generated in testing data by summarizing across imputed gene expressions, weighed by their effect sizes estimated from the training data. We tried a series of p-value thresholds for predictor selection: 1, 0.8, 0.5, 0.4, 0.3, 0.2, 0.1, 0.08, 0.05, 0.02, 0.01, 0.001, $1*10^{-4}$, $1*10^{-5}$, $1*10^{-6}$, $1*10^{-7}$, and $5*10^{-8}$. Thus, seventeen polygenic profiles were generated for each neuroimaging traits and we reported the best prediction power that can be achieved by a single profile of them in the single reference panel. The association between polygenic profile and trait was estimated and tested in linear regression model, adjusting for the effects of age and sex. The additional phenotypic variation that can be explained by polygenic profile (i.e., the incremental R-squared) was used to measure the prediction power.

Data availability

The individual-level data used in this work was obtained from five publicly available datasets: the UK Biobank (UKB) study, the Human Connectome Project (HCP) study, the Pediatric Imaging, Neurocognition, and Genetics (PING) study, the Philadelphia Neurodevelopmental Cohort (PNC) study, and the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. The GWAS summary statistics of UKB study have been shared at https://github.com/BIG-S2/GWAS, and the summary statistics of other validation datasets will also be shared at https://github.com/BIG-S2/GWAS upon acceptance of this paper. We also used the summary-level data of ENIGMA2 and ENIGMA-CHARGE collaboration, which can be obtained at http://enigma.ini.usc.edu/research/. In addition, we used other 11 sets of publicly available GWAS summary statistics shared by several GWAS databases. These data resources were summarized in **Supplementary Table 23**.

Code availability

We made use of publicly available software and tools, especially the UTMOST (https://github.com/Joker-Jerome/UTMOST) and the FUSION (http://gusevlab.org/projects/fusion/). All codes used to generate results that are reported in this paper are available upon request.

Figure legends

Figure 1. Selected significant gene-trait associations discovered in UKB (UK Biobank) cross-tissue TWAS analysis of 211 neuroimaging traits (n=19,629 subjects for ROI volumes and 17,706 for DTI parameters).

The gene-level associations were estimated and tested by the cross-tissue UTMOST approach (https://github.com/Joker-Jerome/UTMOST). We used the p-value threshold of $1.37*10^{-8}$, corresponding to adjusting for testing 211 imaging phenotypes with the Bonferroni correction. The *x* axis provides the IDs of the neuroimaging traits, and the *y* axis lists the detected genes in TWAS. The new (UTMOST new) and previously reported GWAS-significant associations (MAGMA, FUMA, and FUMA&MAGMA) were labeled with different colors (orange, purple, green, and red, respectively).

Figure 2. TWAS-significant genes of neuroimaging traits (n=19,629 subjects for ROI volumes and 17,706 for DTI parameters) that have been linked to other complex traits in previous GWAS.

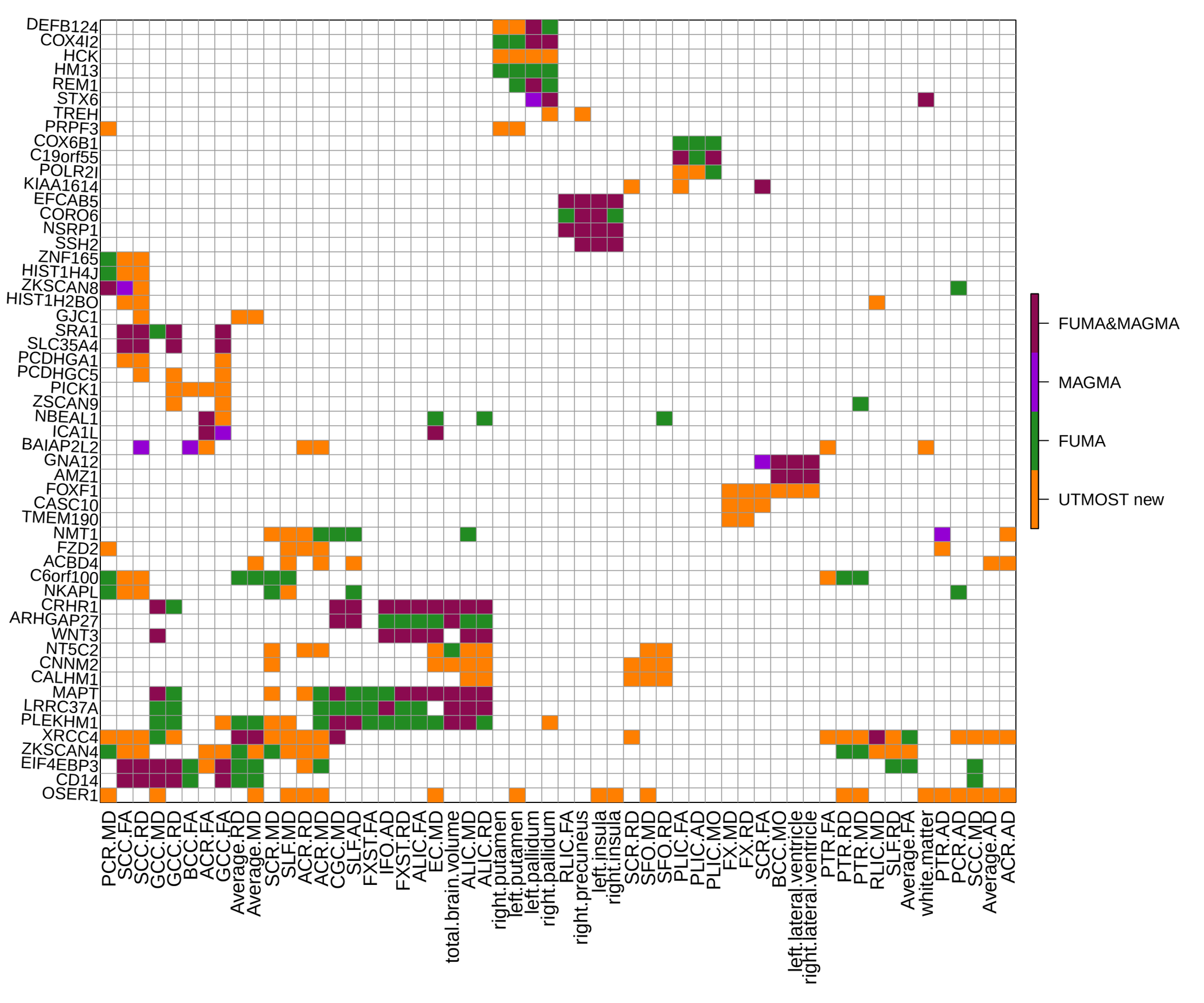
For each of the TWAS-significant genes listed in the *x* axis, we manually checked the previously reported associations on the NHGRI-EBI GWAS catalog (https://www.ebi.ac.uk/gwas/). The genes associated with DTI parameters (DTI), ROI volumes (Volume), and both of them (Both) were labeled with three different colors (blue, orange, and green, respectively).

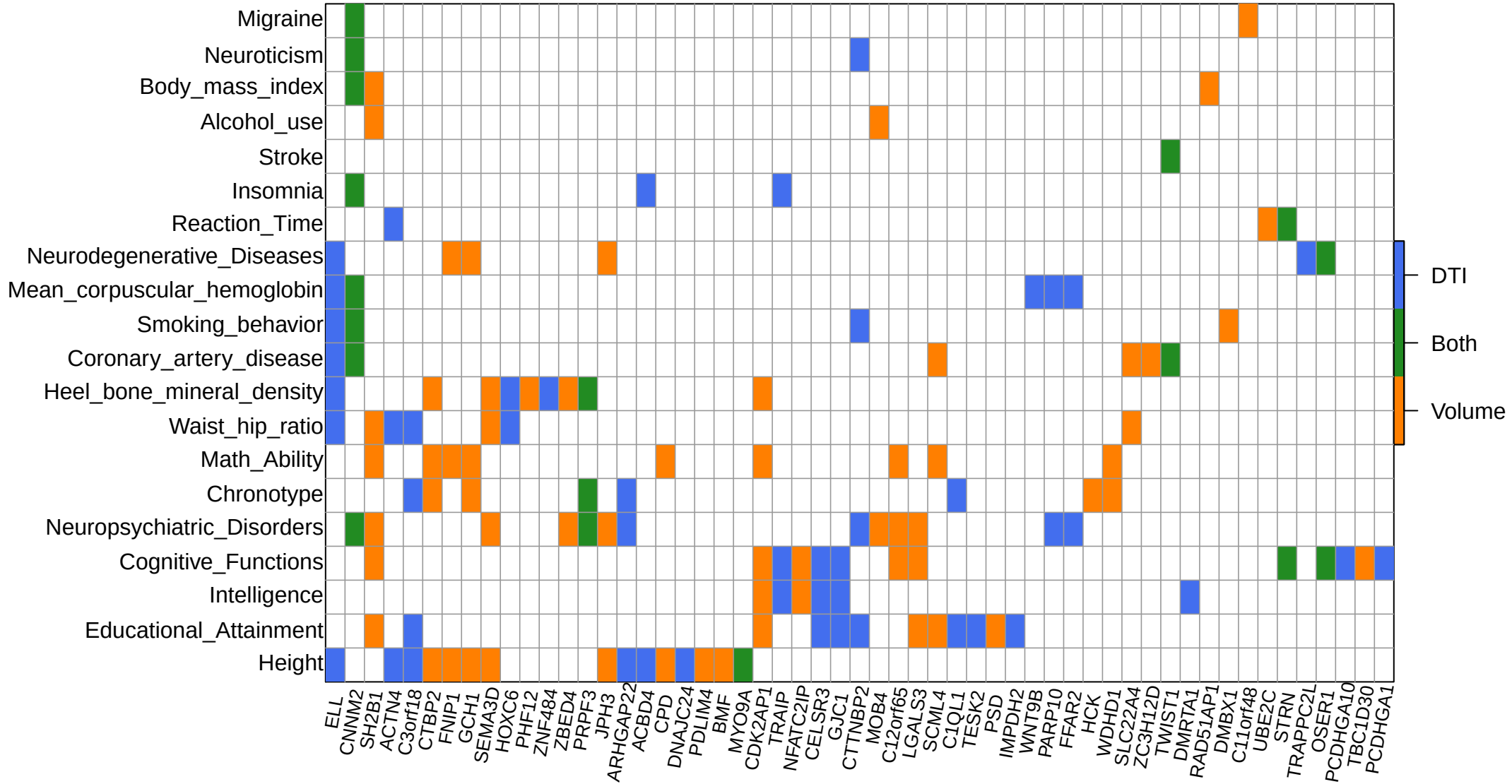
Figure 3. Overlapping TWAS-significant genes between neuroimaging traits (n=19,629 subjects for ROI volumes and 17,706 for DTI parameters) and 11 cognitive and mental health traits.

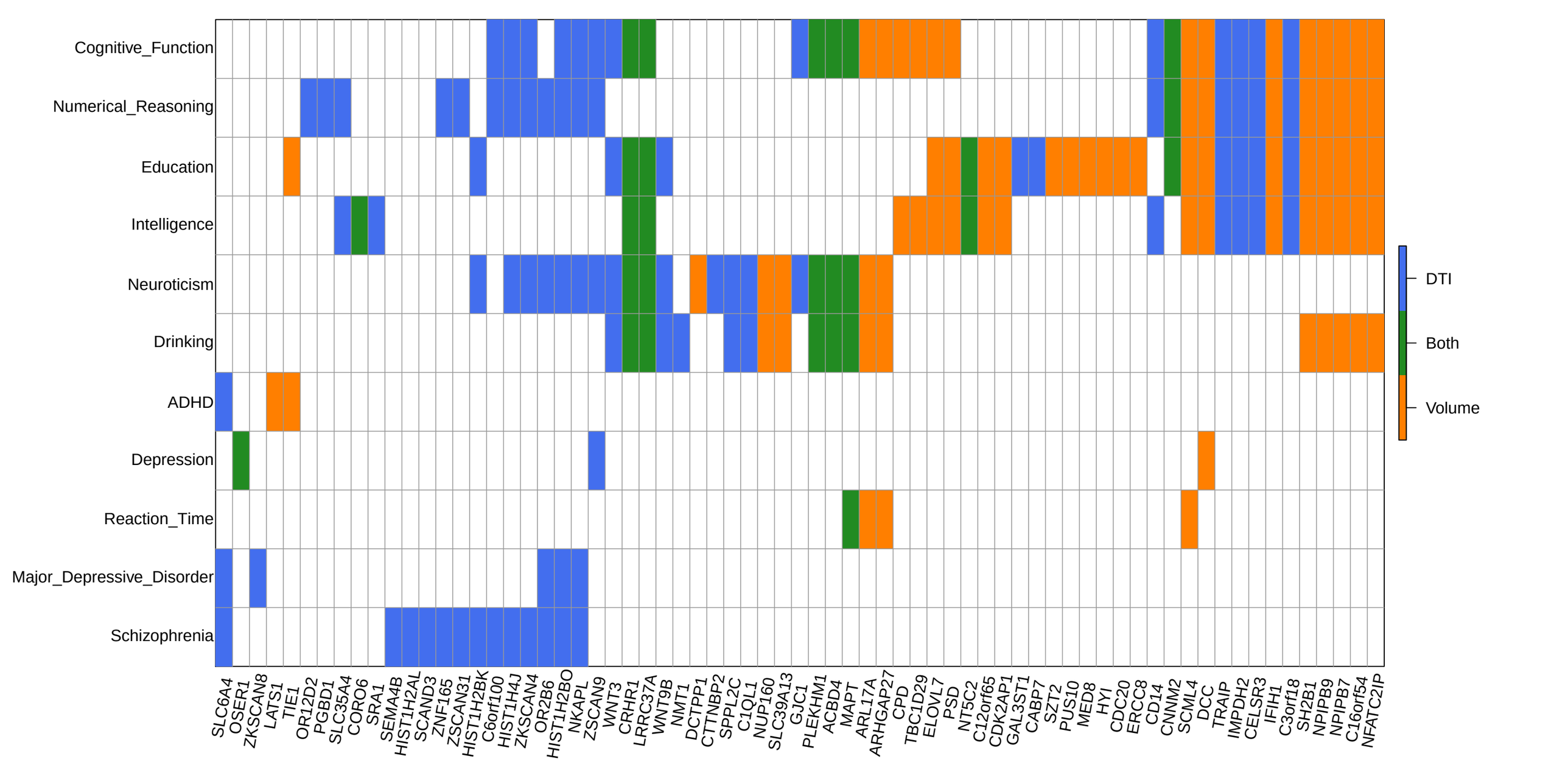
The gene-level associations were estimated and tested by the cross-tissue UTMOST approach (https://github.com/Joker-Jerome/UTMOST). We adjusted for testing 211 neuroimaging traits (p-value threshold $1.37*10^{-8}$) and 11 cognitive traits (p-value threshold $2.63*10^{-7}$) with the Bonferroni correction, respectively. The *x* axis provides the IDs of the neuroimaging traits. The *y* axis lists the 11 cognitive and mental health traits, and **Supplementary Table 23** details the resources of their GWAS summary statistics and the sample sizes of corresponding studies.

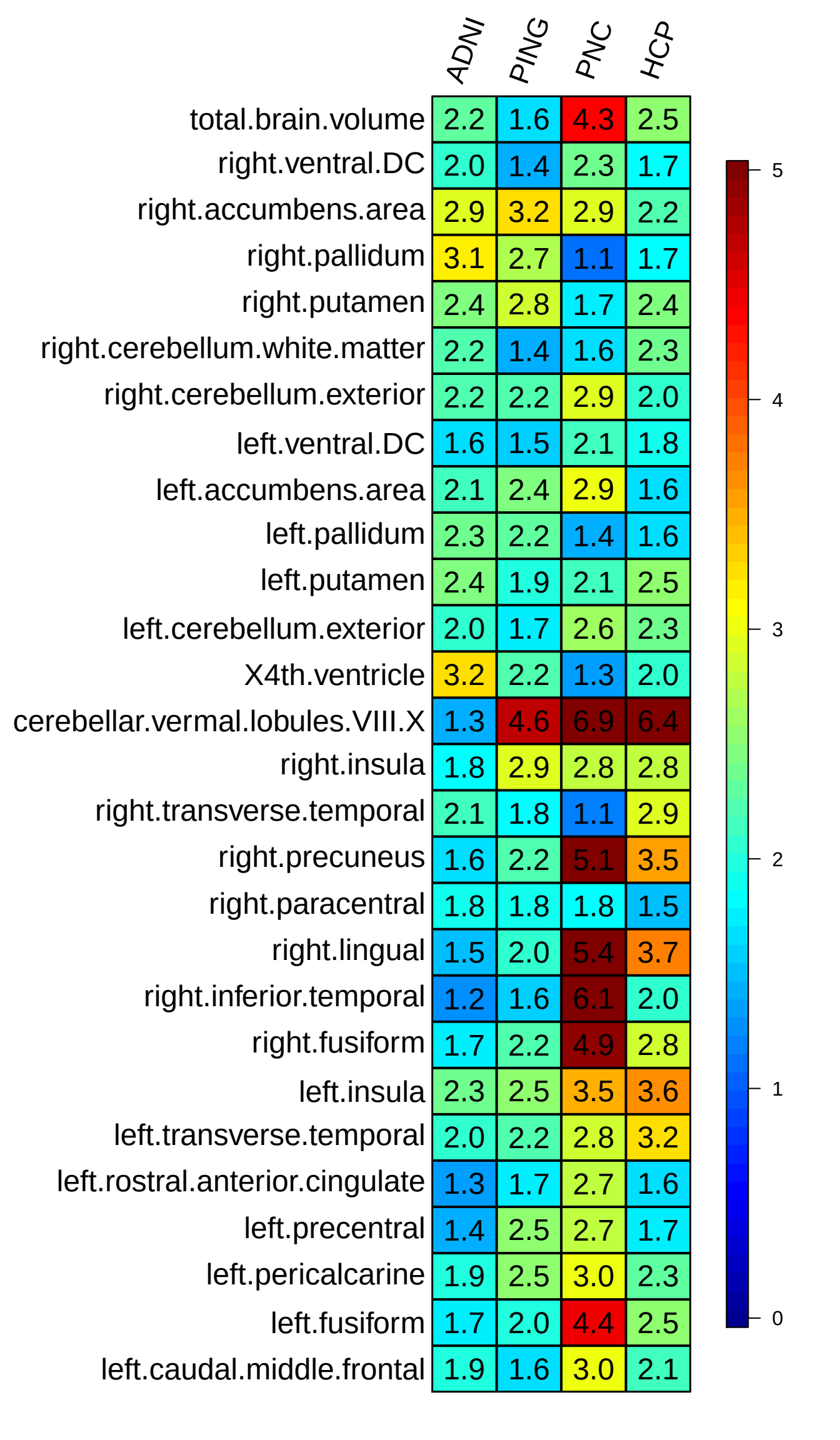
Figure 4. Prediction accuracy (incremental R-squared) of gene-based polygenic risk scores constructed by UKB TWAS results (n=19,629 subjects) on the four independent datasets.

The *x* axis lists the four independent cohorts (ADNI, HCP, PING and PNC) and the *y* axis lists the ROI volumes. The displayed numbers are the proportions of phenotypic variation that can be additionally explained by UKB TWAS-derived gene-based PRS.


DEFB124
COX4I2
HCK
HM13
REM1
STX6
TREH
PRPF3
COX6B1
C19orf55
POLR2I
KIAA1614
EFCAB5
CORO6
NSRP1
SSH2
ZNF165
HIST1H4J
ZKSCAN8
HIST1H2BO
GJC1
SRA1
SLC35A4
PCDHGA1
PCDHGC5
PICK1
ZSCAN9
NBEAL1
ICA1L
BAIAP2L2
GNA12
AMZ1
FOXF1
CASC10
TMEM190
NMT1
FZD2
ACBD4
C6orf100
NKAPL
CRHR1
ARHGAP27
WNT3
NT5C2
CNNM2
CALHM1
MAPT
LRRC37A
PLEKHM1
XRCC4
ZKSCAN4
EIF4EBP3
CD14
OSER1
PCR.MD
SCC.FA
SCC.RD
GCC.MD
GCC.RD
BCC.FA
ACR.FA
GCC.FA
Average.RD
Average.MD
SCR.MD
SLF.MD
ACR.RD
ACR.MD
CGC.MD
SLF.AD
FXST.FA
IFO.AD
FXST.RD
ALIC.FA
EC.MD
total.brain.volume
ALIC.MD
ALIC.RD
right.putamen
left.putamen
left.pallidum
right.pallidum
RLIC.FA
right.precuneus
left.insula
right.insula
SCR.RD
SFO.MD
SFO.RD
PLIC.FA
PLIC.AD
PLIC.MO
FX.MD
FX.RD
SCR.FA
BCC.MO
left.lateral.ventricle
right.lateral.ventricle
PTR.FA
PTR.RD
PTR.MD
RLIC.MD
SLF.RD
Average.FA
white.matter
PTR.AD
PCR.AD
SCC.MD
Average.AD
ACR.AD
FUMA&MAGMA
MAGMA
FUMA
UTMOST new


Migraine
Neuroticism
Body_mass_index
Alcohol_use
Stroke
Insomnia
Reaction_Time
Neurodegenerative_Diseases
Mean_corpuscular_hemoglobin
Smoking_behavior
Coronary_artery_disease
Heel_bone_mineral_density
Waist_hip_ratio
Math_Ability
Chronotype
Neuropsychiatric_Disorders
Cognitive_Functions
Intelligence
Educational_Attainment
Height
ELL
CNNM2
SH2B1
ACTN4
C3orf18
CTBP2
FNIP1
GCH1
SEMA3D
HOXC6
PHF12
ZNF484
ZBED4
PRPF3
JPH3
ARHGAP22
ACBD4
CPD
DNAJC24
PDLIM4
BMF
MYO9A
CDK2AP1
TRAIP
NFATC2IP
CELSR3
GJC1
CTTNBP2
MOB4
C12orf65
LGALS3
SCML4
C1QL1
TESK2
PSD
IMPDH2
WNT9B
PARP10
FFAR2
HCK
WDHD1
SLC22A4
ZC3H12D
TWIST1
DMRTA1
RAD51AP1
DMBX1
C11orf48
UBE2C
STRN
TRAPPC2L
OSER1
PCDHGA10
TBC1D30
PCDHGA1
DTI
Both
Volume


Cognitive_Function
Numerical_Reasoning
Education
Intelligence
Neuroticism
Drinking
ADHD
Depression
Reaction_Time
Major_Depressive_Disorder
Schizophrenia
SLC6A4
OSER1
ZKSCAN8
LATS1
TIE1
OR12D2
PGBD1
SLC35A4
CORO6
SRA1
SEMA4B
HIST1H2AL
SCAND3
ZNF165
ZSCAN31
HIST1H2BK
C6orf100
HIST1H4J
ZKSCAN4
OR2B6
HIST1H2BO
NKAPL
ZSCAN9
WNT3
CRHR1
LRRC37A
WNT9B
NMT1
DCTPP1
CTTNBP2
SPPL2C
C1QL1
NUP160
SLC39A13
GJC1
PLEKHM1
ACBD4
MAPT
ARL17A
ARHGAP27
CPD
TBC1D29
ELOVL7
PSD
NT5C2
C12orf65
CDK2AP1
GAL3ST1
CABP7
SZT2
PUS10
MED8
HYI
CDC20
ERCC8
CD14
CNNM2
SCML4
DCC
TRAIP
IMPDH2
CELSR3
IFIH1
C3orf18
SH2B1
NPIPB9
NPIPB7
C16orf54
NFATC2IP
DTI
Both
Volume


ADNI
PING
PNC
HCP
total.brain.volume 2.2 1.6 4.3 2.5
right.ventral.DC 2.0 1.4 2.3 1.7
right.accumbens.area 2.9 3.2 2.9 2.2
right.pallidum 3.1 2.7 1.1 1.7
right.putamen 2.4 2.8 1.7 2.4
right.cerebellum.white.matter 2.2 1.4 1.6 2.3
right.cerebellum.exterior 2.2 2.2 2.9 2.0
left.ventral.DC 1.6 1.5 2.1 1.8
left.accumbens.area 2.1 2.4 2.9 1.6
left.pallidum 2.3 2.2 1.4 1.6
left.putamen 2.4 1.9 2.1 2.5
left.cerebellum.exterior 2.0 1.7 2.6 2.3
X4th.ventricle 3.2 2.2 1.3 2.0
cerebellar.vermal.lobules.VIII.X 1.3 4.6 6.9 6.4
right.insula 1.8 2.9 2.8 2.8
right.transverse.temporal 2.1 1.8 1.1 2.9
right.precuneus 1.6 2.2 5.1 3.5
right.paracentral 1.8 1.8 1.8 1.5
right.lingual 1.5 2.0 5.4 3.7
right.inferior.temporal 1.2 1.6 6.1 2.0
right.fusiform 1.7 2.2 4.9 2.8
left.insula 2.3 2.5 3.5 3.6
left.transverse.temporal 2.0 2.2 2.8 3.2
left.rostral.anterior.cingulate 1.3 1.7 2.7 1.6
left.precentral 1.4 2.5 2.7 1.7
left.pericalcarine 1.9 2.5 3.0 2.3
left.fusiform 1.7 2.0 4.4 2.5
left.caudal.middle.frontal 1.9 1.6 3.0 2.1
5
4
3
2
1
0