Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n=17,706)

Running title: GWAS of brain white matter

Bingxin Zhao, M.S.[1], Jingwen Zhang, Ph.D.[2], Joseph G. Ibrahim, Ph.D.[1], Tianyou Luo, B.S.[1], Rebecca C. Santelli, Ph.D.[3], Yun Li, Ph.D.[1,4,5], Tengfei Li, Ph.D.[6,7], Yue Shan, M.S.[1], Ziliang Zhu, B.S.[1], Fan Zhou, Ph.D.[1], Huiling Liao. B.S.[8], Thomas E. Nichols, Ph.D.[9], and Hongtu Zhu, Ph.D.*[1,7]

[1] Departments of Biostatistics, [3] Psychiatry, [4] Genetics, [5] Computer Science and [6] Radiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 27599

[2] Department of Biostatistics, T.H. Chan School of Public Health, Harvard University, Boston, MA, USA, 02115

[7] Biomedical Research Imaging Center, School of Medicine, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 27599

[8] Department of Statistics, Texas A&M University, College Station, TX, USA, 77843

[9] Wellcome Trust Centre for Integrative Neuroimaging, Big Data Institute, University of Oxford, Oxford, UK, OX1 2JD

Corresponding author:

Hongtu Zhu, Ph.D.

Full postal address: Department of Biostatistics, University of North Carolina at Chapel Hill, 135 Dauer Drive, 3101 McGavran-Greenberg Hall, Chapel Hill, NC 27599.

Email address: htzhu@email.unc.edu, Telephone number: (919) 966-7250, Fax number: (919) 966-7256.

Abstract

Individual variations of white matter (WM) tracts are known to be associated with various cognitive and neuropsychiatric traits. Diffusion tensor imaging (DTI) and genome-wide single-nucleotide polymorphism (SNP) data from 17,706 UK Biobank participants offer the opportunity to identify novel genetic variants of WM tracts and explore the genetic overlap with other brain-related complex traits. We analyzed the genetic architecture of 110 tract-based DTI parameters, carried out genome-wide association studies (GWAS), and performed post-GWAS analyses, including association lookups, gene-based association analysis, functional gene mapping, and genetic correlation estimation. We found that DTI parameters are substantially heritable for all WM tracts (mean heritability 48.7%). We observed a highly polygenic architecture of genetic influence across the genome (p-value=$1.67*10^{-05}$) as well as the enrichment of genetic effects for active SNPs annotated by central nervous system cells (p-value=$8.95*10^{-12}$). GWAS identified 213 independent significant SNPs associated with 90 DTI parameters (696 SNP-level and 205 locus-level associations; p-value<$4.5*10^{-10}$, adjusted for testing multiple phenotypes). Gene-based association study prioritized 112

significant genes, most of which are novel. More importantly, association lookups found that many of the novel SNPs and genes of DTI parameters have previously been implicated with cognitive and mental health traits. In conclusion, the present study identifies many new genetic variants at SNP, locus and gene levels for integrity of brain WM tracts and provides the overview of pleiotropy with cognitive and mental health traits.

Introduction

Complex brain functions rely on dynamic interactions between distributed brain areas operating in large-scale networks. Consequently, the integrity of white matter connections between brain areas is critical to proper function. Microstructural differences in white matter (WM) tracts are phenotypically associated with information processing speed and intelligence [1-4] as well as neurodegenerative/neuropsychiatric traits, such as Alzheimer's disease [5], Parkinson's disease [6], schizophrenia (SCZ) [7], and attention-deficit/hyperactivity disorder (ADHD) [8]. A better understanding of genetic factors influencing integrity of WM tracts could have important implication for

64 understanding the etiology of these diseases as well as individual variation in
65 intelligence. To reveal the underlying genetic contributions to brain structural
66 development and disease/disorder processes, imaging genetics studies of WM
67 microstructure has been an active research area over the past fifteen years. The
68 structural changes of WM tracts are typically measured and quantified in diffusion
69 tensor imaging (DTI) [9]. Brain diffusivity can be influenced by many aspects of its micro-
70 or macro-structures [10]. To reconstruct the WM pathways and tissue microstructure, DTI
71 models the diffusion properties of WM using random movement of water. Specifically,
72 DTI quantifies diffusion magnetic resonance imaging (dMRI) in a tensor model and
73 analyzes diffusions in all directions. A typical DTI diagonalizes the tensor and calculates
74 three pairs of eigenvalues/eigenvectors that respectively represent one primary and
75 two secondary diffusion directions. Within each voxel, several DTI parameters can be
76 derived: fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), radial
77 diffusivity (RD), and mode of anisotropy (MO). As a summary measure of WM integrity
78 [11], higher FA indicates stronger directionality in this voxel. MD quantifies the magnitude
79 of absolute directionality, AD is the eigenvalue of the principal direction, RD is the

average of the eigenvalues of the two secondary directions, and MO is the third moment of a tensor. Positive MO reflects narrow tubular water diffusion, whereas a negative value denotes planar water diffusion [12]. There are several approaches to analyze DTI data across the whole brain, including manual region-of-interest (ROI) analysis, automated ROI analysis, voxel-based analysis, such as tract-based spatial statistics (TBSS) [13], as well as tractography and graph theory analysis; see Tamnes, Roalf [14] for a survey.

In family-based studies, the magnitude of genetic influences (i.e., heritability) in various DTI parameters of WM tracts, including FA, MD, AD, and RD, has been examined across a wide age range, from neonates [15], young children [16], older children [17], adolescents [18], and young adults [19] to middle aged [20] and older adults [21]. Participants in these studies are typically monozygotic and dizygotic twins or family members. Table 1 of Vuoksimaa, Panizzon [20] lists 14 studies that illustrated that a substantial proportion of variance in DTI parameters (FA, MD, AD, and RD) was explained by additive genetic effects. However, the genetic architecture of DTI parameters remains largely unknown due to

96 the limitation of family-based studies, for which the heritability estimation has relied on
97 contrasting the phenotypic similarity between monozygotic and dizygotic twins. Genetic
98 architecture denotes the characteristics of genetic variations that contribute to the
99 broad-sense heritability of a phenotype [22]. Based on the number of genetic variants
100 contributing to phenotypic variance, genetic architecture can be described as
101 monogenic (one variant), oligogenic (few variants), polygenic (many variants), or
102 omnigenic, which hypothesizes that almost all genetic variants have small but non-zero
103 genetic contributions [23, 24]. Uncovering the genetic architecture and discovering the
104 associated genetic variants are essential steps to delineate the functional mechanisms
105 and understand the genetic overlap between white matter structures and
106 neuropsychiatric traits.
107
108 Recent developments have enabled heritability estimation and genetic variants
109 discovery with using the common single-nucleotide polymorphisms (SNPs) data
110 collected in general populations. Instead of using the expected genetic correlation based
111 on pedigree information, SNP heritability is estimated by adding up the genetic effects

across a large number of common SNPs (minor allele frequency [MAF]>0.05 or 0.01) [25, 26]. The architecture of genetic influences can be assessed by SNP annotation and partition [27, 28]. Genome-wide association studies (GWAS) and post-GWAS analysis can further identify causal genetic variants at SNP, locus and gene levels [29, 30], and assess the genetic overlap of complex traits in different domains [31, 32]. With these methods, the availability of genomic and imaging data from recent large population-based United Kingdom (UK) Biobank resource [33] offers the opportunity to uncover the genetic basis of brain WM tracts in one large-scale, relatively homogeneous population. The UK Biobank (UKB) has captured data from over 500,000 original participants of middle or elderly ages (age range 40-69), and is currently in the process of following up with 100,000 of these participants to perform brain MRI screening [34].

Rutten-Jacobs, Tozer [35] and Elliott, Sharp [36] performed GWAS for brain MRI phenotypes using the UKB brain imaging data released in 2017 (n~8500). Elliott, Sharp [36] showed the ubiquitous impact of genetics in various brain imaging measures, and Rutten-Jacobs, Tozer [35] focused on the DTI parameters and examined their genetic overlaps with

128 stroke, depression, and dementia. However, the simple size in these GWAS was far from
129 being sufficient, for which only a few novel loci were detected. Here we generated 110
130 tract-based DTI parameters using the British ancestry UKB sample including 17,706
131 participants. For each of the 110 phenotypes, we estimated the SNP-heritability,
132 assessed the distribution of genetic effects by SNP annotation and partition, and carried
133 out GWAS to identify the associated genetic variants at SNP and locus levels. In addition,
134 we discovered gene-level associations via MAGMA [37], and explored the functional
135 consequences of the significant SNPs by functional mapping and annotation analysis
136 (FUMA [30]). To detect genetic overlap and pleiotropy in WM tracts and other complex
137 traits, we performed association lookups at SNP and gene levels on the NHGRI-EBI
138 GWAS catalog [38] and estimated genetic correlations via linkage disequilibrium (LD) score
139 regression (LDSC [32]). As demonstrated later, hundreds of novel genetic associations
140 were detected in the present GWAS and a much clearer picture of widespread
141 pleiotropy with cognitive and mental health traits was found in our post-GWAS analysis.
142 The UKB GWAS results were further validated in an independent imaging genetics

dataset. The GWAS summary statistics have been made publicly available at https://med.sites.unc.edu/bigs2/data/gwas-summary-statistics/.

Materials and Methods

Participants and image preprocessing

We used data from 17,706 UKB individuals of British ancestry (self-reported ethnic background, Data-Field 21000). The ancestry information was checked and confirmed by the top genetic principal components provided by UKB [39] (GPCs, Data-Field 22009) (**Supplementary Figure 1**). The dMRI data [34] and covariates were downloaded from the UKB data resource. We generated 110 DTI parameters: FA, AD, MD, MO and RD of 21 WM tracts, and their average values across these tracts. The WM tracts were labelled by the ENIGMA-DTI pipeline [40, 41], which was widely applied to measure the variation of microstructural integrity [42-44]. The ID and full names of these 21 WM tracts are listed in **Supplementary Table 1.** A full description of the DTI preprocessing and analysis, imaging quality controls, white matter tracts, and formulas to calculate the DTI parameters are documented in supplementary information. An overview of the ENIGMA-DTI pipeline

159 applied in this study is given in **Supplementary Figure 2** and a few examples are shown
160 in **Supplementary Figures 3-6**. We removed values greater than five times the median
161 absolute deviation from the median in each continuous variable. All individuals were
162 aged between 40 and 80 years and the proportion of females was 52.9%.
163
164 **Genotyping and quality control**
165 We downloaded the imputed SNP data from UKB data resource [39]. We further
166 performed the following SNP data quality controls using PLINK [45]: excluding subjects
167 with more than 10% missing genotypes, only including SNPs with MAF > 0.01,
168 genotyping rate > 90%, and passing Hardy-Weinberg test (p-value>$1*10^{-7}$). We also
169 removed SNPs with imputation INFO score less than 0.8.
170
171 **SNP-heritability analysis and genome-wide association analysis**
172 For each of the 110 DTI parameters, we estimated the proportion of variation explained
173 by all autosomal SNPs with using univariate GCTA-GREML analysis [25]. We considered the
174 fixed effects of age (at imaging), age-squared, gender, age-gender interaction,

age-squared-gender interaction, as well as the top 40 genetic principal components. We also estimated the proportion of variation explained by SNPs in each chromosome. In addition, we performed cell-type-specific SNP heritability analysis. SNPs were grouped according to their functional activeness in various cell groups [28] and specifically in the central nervous system (CNS) cell group: CNS_active, CNS_inactive, and Always_inactive, see supplementary information for detailed definitions. We performed GWAS for each DTI parameter separately with PLINK [45]. The same set of covariates as in GCTA-GREML analysis were adjusted in GWAS and all other analyses unless stated otherwise.

Genomic risk loci characterization and comparison with previous findings

We characterized genomic risk loci by using FUMA [30] online platform (v1.3.4). FUMA first identified independent significant SNPs, which were defined as significant SNPs that were independent of each other (R^2<0.6). FUMA then constructed LD block for independent significant SNPs by tagging all SNPs that had a MAF $\geq$ 0.0005 and were in LD ($R^2 \geq 0.6$) with at least one of the independent significant SNPs. If LD blocks of independent significant SNPs were closed (<250 kb based on the closest boundary SNPs

191 of LD blocks), they were merged to a single genomic locus. More details of FUMA
192 analysis can be found in Watanabe, Taskesen [30]. Independent significant SNPs and all
193 SNPs in LD with them were subsequently searched on NHGRI-EBI GWAS catalog [38]
194 (v2019-01-31) to look for reported associations (p-value<$9*10^{-6}$) with any traits.
195
196 **Gene-based association analysis and functional annotation**
197 We carried out gene-based association analysis for 18,796 protein-coding candidate
198 genes via MAGMA [37] (v1.07). Gene-based p-values were calculated by summarizing the
199 GWAS results of corresponding SNPs, which were mapped to genes according to their
200 physical positions. Significant genes were searched on NHGRI-EBI GWAS catalog [38]
201 (v2019-01-31) to look for their previously reported associations with any traits. We
202 focused on brain-related complex traits and characterized them into six groups:
203 cognitive (e.g., general cognitive ability, cognitive performance, math ability, and
204 intelligence), education (e.g., years of education and college completion), reaction time,
205 neuroticism, neurodegenerative diseases (e.g., Alzheimer's disease, Parkinson's disease
206 and corticobasal degeneration), and neuropsychiatric disorders (e.g., major depressive

disorder [MDD], SCZ, bipolar disorder [BD], ADHD, alcohol use disorder, and autism spectrum disorder).

We also performed functional gene annotation and mapping via FUMA. SNPs were annotated with their biological functionality and then were linked to genes by a combination of positional, expression quantitative trait loci (eQTL) association, and 3D chromatin interaction mappings. Specifically, independent significant SNPs and all SNPs in LD with them were annotated for gene functional consequences by ANNOVAR [46]. The annotated SNPs were mapped to 35,808 candidate genes based on physical position on the genome (tissue/cell types for 15-core chromatin state: brain), eQTL associations (tissue types: GTEx v7 brain [47], BRAINEAC [48], and CommonMind Consortium [49]) and chromatin interaction mapping (built-in chromatin interaction data: dorsolateral prefrontal cortex, hippocampus [50]; annotate enhancer/promoter regions: E053-E082 brain [51]). We used default values for all other parameters in FUMA.

Genetic correlation estimation with LDSC

We used LDSC (v1.0.0, https://github.com/bulik/ldsc) to estimate the pairwise genetic correlation between DTI parameters and other traits by their GWAS summary statistics. In LDSC, we used the pre-calculated LD scores provided by LDSC (https://data.broadinstitute.org/alkesgroup/LDSCORE/), which were computed using 1000 Genomes European data. We used HapMap3 SNPs and removed all SNPs in chromosome 6 in the MHC region.

Genome-wide polygenic risk scores

Genome-wide polygenic risk scores [52] were created to examine the out-of-sample prediction ability of the UKB GWAS results. Two procedures were used to adjust for the LD structure: 1) LD-based pruning (window size 50, step 5, R-squared = 0.2); and 2) posterior effect size estimation under continuous shrinkage prior with an external LD reference panel [53]. We tried five p-value thresholds for predictor selection in each of the two procedures: 1, 0.5, 0.05, $5*10^{-4}$ and $5*10^{-8}$. Thus, ten polygenic scores were generated via PLINK and we reported the best prediction accuracy that can be achieved by a single score of these ten. Besides same-trait prediction, we also used cross-trait

polygenic risk scores[54] to validate the observed significant genetic correlations between DTI parameters and other brain-related traits. The association between polygenic score and phenotype 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 score (i.e., the incremental R-squared) was used to measure the prediction accuracy.

Results

SNP heritability estimation

Figures 1-2, **Supplementary Figures 7-12, and Supplementary Video 1** display the SNP heritability of DTI parameters estimated by all common autosomal SNPs. The associated standard errors, raw and Bonferroni-adjusted p-values from the one-sided likelihood ratio tests are given in **Supplementary Table 2**. All SNP heritability estimates were significantly larger than zero (Bonferroni-adjusted p-value<0.004). Genetic factors accounted for a moderate or large portion of the variance of DTI parameters in all WM tracts (mean heritability 0.487, standard errors are around 0.041). For example, genetic effects explained more than 60% of the total variance of FA in the posterior limb of the

internal capsule (PLIC), anterior corona radiata (ACR), superior longitudinal fasciculus (SLF), and cingulum cingulate gyrus (CGC). The lowest SNP heritability of FA across all WM tracts were found in fornix (FX, 37%) and corticospinal tract (CST, 27%). According to the functions of WM tracts (Connectopedia Knowledge Database, http://www.fmritools.com/kdb/white-matter/), we clustered them into four communities including complex fibers (C1: ACR, ALIC, PCR, PLIC, PTR, RLIC, SCR, EC, SS), associative fibers (C2: CGC, CGH, FX, FXST, IFO, SFO, SLF, UNC), commissural fibers (C3: BCC, GCC, SCC) and projection fibers (C4: CST) (**Figure 1,** see **Supplementary Table 1** for IDs). We found that the set of WM tracts in C1 and C3 (mean=0.512) tended to have higher SNP heritability than those in C2 and C4 (mean=0.440, p-value=$2.16*10^{-04}$).

Partitioning and annotating genetic variation

To examine the distribution of SNP heritability across the genome, we partitioned SNP data into 22 chromosomes and estimated the SNP heritability by each chromosome **(Supplementary Table 3)**. We found that the mean heritability across all 110 DTI parameters explained by each chromosome was linearly associated with the length of

271 the chromosome **(Figure 3(a)**, R^2=61.2%, p-value=1.67*10^{-05}). This finding reveals a
272 highly polygenic or omnigenic genetic architecture [24] of WM tracts. The large number of
273 SNPs that contribute to the variation in DTI parameters are widely spread across the
274 whole genome. To further illustrate this architecture, we ordered and clustered the 22
275 chromosomes into three groups by their lengths: long, medium, and short. The long
276 group had 4 chromosomes (CHRs 2, 1, 6, 3), which together accounted for 33% of the
277 length of the whole genome; the medium group had 6 chromosomes (CHRs 4, 5, 7, 8,
278 10, 11), which accounted for another 33% of the length of the whole genome; and the
279 short group consisted of the remaining 12 chromosomes. **Figure 3(b)** shows the SNP
280 heritability estimates grouped by chromosomal length. It is clear that longer
281 chromosomes tended to have higher SNP heritability estimates than medium
282 (p-value=3.82*10^{-13}) or shorter (p-value<2.20*10^{-16}) ones for DTI parameters.
283
284 To compare the contribution of SNPs with different activity level, we partitioned the
285 genetic variation according to CNS-cell-specific annotations: CNS_active, CNS_inactive,
286 and Always_inactive **(Supplementary Table 4)**. Heritability estimated by SNPs residing in

chromatin regions inactive across all cell groups (Always_inactive) was clearly much smaller than the heritability estimated by SNPs residing in chromatin regions active in CNS cell (CNS_active, p-value<$2.20*10^{-16}$). The heritability estimated by CNS_inactive SNPs (inactive in CNS cell but active in other cells) was also significantly smaller than that of CNS_active SNPs (p-value=$8.95*10^{-12}$) **(Figure 3(c))**. This pattern remained consistent across all the five types of DTI parameters, though larger variance was observed for the MO parameters.

GWAS results of 110 DTI parameters

We carried out GWAS for the 110 DTI parameters with using 8,955,960 SNPs after genotyping quality controls. All Manhattan and QQ plots are shown in **Supplementary Figure 13**. 19,530 significant associations were detected at the $4.5*10^{-10}$ significance level (that is, $5*10^{-8}/110$, adjusted for testing multiple phenotypes) **(Supplementary Figure 14, Supplementary Table 5)**. RD and MD of anterior limb of internal capsule (ALIC) had more than 3,000 significant associations. Significant SNPs were summarized into 213 independent significant SNPs by FUMA, which had 696 independent significant

303 associations with 90 DTI parameters **(Figure 4, Supplementary Tables 6-7)**. RD and FA of
304 splenium of corpus callosum (SCC) had the largest number of independent significant
305 SNPs. Of the 696 independent significant associations, 502 located in chromosome 5
306 **(Supplementary Table 8, Supplementary Figure 15)**. The 696 independent significant
307 SNP-level associations can be further characterized as 205 locus-level associations
308 **(Supplementary Table 9)**. FA and RD of SCC, FA and AD of FX, and RD of ALIC had at
309 least five genetic risk loci **(Supplementary Table 10)**. Each chromosome had at least one
310 genetic risk locus except for chromosomes 13 ,20 and 21, and chromosome 5 had the
311 largest number of risk loci **(Supplementary Tables 11)**. Enrichment of GWAS signals in
312 chromosome 5 for DTI parameters has been found in Rutten-Jacobs, Tozer [35],
313 particularly in the chr5q14 locus. Further research is needed to explore the biological
314 role of chromosome 5 for microstructural integrity changes that can be measured by
315 dMRI. GWAS results at $5*10^{-9}$ and $5*10^{-8}$ significance levels are also reported in above
316 tables and figures.
317
318 **Concordance with previous GWAS results**

Association lookups on the NHGRI-EBI GWAS catalog [38] found that 122 of the 213 independent significant SNPs (associated with 83 DTI parameters) were reported to be associated with any traits **(Supplementary Table 12)**. Our study replicated many SNPs reported in previous GWAS of WM hyperintensity measures and other brain structural measures **(Supplementary Table 13)**, most of which were recently detected in Rutten-Jacobs, Tozer [35] (n=8,448). In addition, we tagged 15 different SNPs associated with neuropsychiatric disorders, 40 with cognitive traits, 12 with education, 47 with neuroticism, 17 with neurodegenerative diseases, and 2 with reaction time**.** We also compared our results with the those reported in Elliott, Sharp [36] (n=8,428) and found that 212 of the 368 significant associations (Supplementary Table 6 of Elliott, Sharp [36]) were replicated in the present study.

Gene-based association analysis and functional mapping

Gene-based association analysis identified 508 significant gene-level associations (p-value<$2*10^{-8}$, adjusted for testing multiple phenotypes) between 112 genes and 96 DTI parameters **(Supplementary Table 14).** Our results replicated genes discovered in

335 Rutten-Jacobs, Tozer [35] and Elliott, Sharp [36], including *VCAN, C16orf95, NBEAL1,*

336 *SH3PXD2A, CACNB2, SRA1, GNA12, CPED1,* and *EPHA3,* but most of the identified genes

337 were not previously linked to DTI parameters. Association lookups found that 51 of the

338 112 significant genes were implicated with cognitive, education, reaction, neuroticism,

339 neuropsychiatric and neurodegenerative traits in previous studies, such as *CRHR1* [55-58],

340 *MAPT* [59-62], *KANSL1* [63-65], and *MSRA* [66-68] **(Supplementary Table 15, Figure 5)**. We also

341 annotated the SNPs by functional consequences on gene functions **(Supplementary**

342 **Figure 16)** and performed functional gene mapping. Gene mapping discovered 292

343 genes **(Supplementary Table 16)**, 218 of which were not detected in the gene-based

344 association analysis.

345

346 **Genetic correction with other traits**

347 We estimated the pairwise genetic correlation between 110 DTI parameters and 14

348 other complex traits **(Supplementary Table 17)**. We focused on traits showing evidence

349 of pleiotropy in association lookups. 43 pairs of phenotypes had significant genetic

350 correlation after adjusting for multiple comparisons (1,540 tests) by using the

Benjamini-Hochberg (B-H) procedure [69] at 0.05 level **(Supplementary Table 18, Supplementary Figure 17).** Reaction time had significant negative correlations with FA parameters (mean=-0.181), and had widespread positive correlations with AD, MO, MD and RD (mean=0.165) **(Supplementary Figure 18)**. Education, cognitive, intelligence, and numerical reasoning also had positive genetic correlations with AD, FA, and MO. On the other hand, depression, MDD and drink frequency showed negative genetic correlations with FA. Other pairs were insignificant after multiple testing adjustment.

We also estimated the pairwise genetic correlation between 110 DTI parameters and 100 regional brain volume measures (ROIs, **Supplementary Table 19**). We found widespread genetic overlaps between DTI parameters and brain volumes (**Supplementary Figures 19-23**), and 490 pairs were significant after adjusting for multiple comparisons by using the B-H procedure at 0.05 level (11,000 tests). For example, white matter volume had significantly positive genetic correlations with FA of BCC, CGC, FX, FXST, and GCC WM tracts. All genetic correlation estimates and the associated p-values can be found in **Supplementary Table 20**.

Validation in an independent dataset

To validate the UKB GWAS results, we repeated GWAS of 110 DTI parameters on data obtained from the Philadelphia Neurodevelopmental Cohort [70] (PNC) study (n=520). More details about PNC dataset and GWAS can be found in the supplementary information. Due to the small sample size, the probability of reaching GWAS significance level was low in the PNC data. Therefore, we focused on the 3,954,646 overlapped SNPs and checked whether the effect signs of top UKB GWAS SNPs were concordant in the two studies [71]. For the 5,625 significant UKB associations (88 DTI parameters, $4.5*10^{-10}$ significance level), 85.4% (4,803) associations had the same effect signs in the two studies. In addition, 64 of the 88 DTI parameters have larger than 95% effect sign matching rate (**Supplementary Table 21**). We also assessed the prediction accuracy of UKB GWAS results on the PNC data with the genome-wide polygenic risk scores prediction [52]. After adjusting for multiple comparisons by using the B-H procedure at 0.05 level, 104 of the 110 UKB-derived polygenic scores were significantly associated with the corresponding DTI parameter of the PNC dataset **(Supplementary Table 22).**

The significant polygenic scores can account for up to 2.95% phenotypic variation, and the largest R-squared 2.95% was found in AD of ALIC (p-value=$5.15*10^{-10}$). Other DTI parameters with R-squared larger than 2% included AD of FX (R-squared=2.36%, p-value=$3.06*10^{-8}$), AD of SLF (R-squared=2.36%, p-value=$7.48*10^{-9}$), average AD (R-squared=2.21%, p-value=$5.89*10^{-10}$), MO of SLF (R-squared=2.12%, p-value=$5.92*10^{-8}$), MO of ACR (R-squared=2.06%, p-value=$2.51*10^{-7}$), and FA of CGH (R-squared=2.00%, p-value=$1.22*10^{-7}$). In summary, the joint analysis with PNC datasets shows moderate to high level of agreement in term of GWAS effect signs, and indicates that the UKB GWAS summary statistics have widespread out-of-sample prediction power across WM tracts. We also constructed cross-trait polygenic risk scores [54, 72, 73] for PNC subjects to validate the genetic overlap between DTI parameters and brain-related behavioral traits. Of the 43 significant genetic correlation pairs observed in UKB LDSC analysis, 26 pairs were significant (p-value range=[$4.16*10^{-11}$,$2.67*10^{-2}$]) after adjusting for multiple comparisons by using the B-H procedure at 0.05 level **(Supplementary Table 23)**. Particularly, reaction time-derived polygenic scores replicated significant

genetic correlations with 18 DTI parameters, and depression and educational attainment-derived polygenic scores each validated 2 DTI parameters.

Discussion

Heritability and GWAS analyses can provide guidance for downstream analyses to model the functional mechanisms and pathways involved in the phenotype of interest or its pleiotropy traits. A large number of family-based neuroimaging studies have documented that WM tracts are essentially heritable across the lifespan. Two recent GWAS [35, 36] have made attempts to explore the genetic risk variants of DTI parameters, however, they were less powered due to the limited sample size (n<9,000). Compared to the previous GWAS, the present study made novel contributions to 1) understand the genetic landscape of WM tract via chromosome-specific SNP heritability analysis; 2) identify novel genetic risk variants for many DTI parameters; 3) perform gene-based association analysis and conduct functional gene mapping with eQTL and chromatin interaction data; 4) uncover the statistical pleiotropy [31, 74] with other brain-related

413 complex traits; and 5) examine the out-of-sample prediction ability of UKB GWAS
414 results.
415
416 Our SNP heritability estimates are close to the ones reported in previous family-based
417 studies (e.g., Table 1 of Vuoksimaa, Panizzon [20]), and are also within a similar range as
418 those reported in Elliott, Sharp [36], where the mean heritability is around 0.450. These
419 results suggest that studies of DTI phenotypes using common SNPs may be more
420 informative than studies focused on rare variants. Our results partitioning the genetic
421 variation in chromosomes or SNP functional sets shed light on the distribution of genetic
422 signals across the genome and different functional consequences. These findings
423 suggest a highly polygenic genetic architecture of DTI parameters and also provide
424 evidence for stronger genetic signals from SNPs in active chromatin regions, especially
425 for those active in the CNS cell type. For such highly polygenic traits, large sample size is
426 essential for GWAS to discover the widespread genetic signals. Our study with larger
427 sample size identifies hundreds of new genetic associations at variant, locus, and gene
428 levels. More importantly, these novel findings lead to uncover the widespread

pleiotropy between DTI parameters and cognitive and mental health traits. Small but significant genetic correlations were quantified between DTI parameters and other brain-related complex traits. As the UKB releases more imaging data, it can be expected that better powered genetic studies on heritable WM tracts will continue facilitating gene exploration and helping understand the causal relationships of brain-related complex traits.

Our analyses reflect several methodological limitations of the current approaches on population-based imaging genetic studies. First, similar to previous studies [19], CST and FX were reported to have low SNP heritability, which may be due to the fact that such small, tubular tracts cannot be well registered and reliably resolved with current techniques [75]. Second, heritability estimated by SNP data reflects narrow-sense heritability, which only considers the additive genetic effects of common variants. The genetic architecture may change as we broadly consider all genetic contributions (such as rare variants, non-additive effects and gene-gene interactions) in future studies. However, it is notable that with common SNPs in the UK Biobank, we have gained

heritability estimates comparable to those reported in family-based studies. Finally, it is worth mentioning that the UKB data used in this study were sampled from a specific cohort (British ancestry) with a specific age-range. Since genetic ancestries are common confounding effects and aging can play an important role in brain WM structure changes, one should be careful to generalize these findings to general populations or to specific clinical cohorts. With more data from diverse imaging genetics studies, future research will be required to overcome these limitations and advance our biological understanding of the human brain.

Acknowledgements

This research was partially supported by U.S. NIH grants MH086633 and MH116527, and a grant from the Cancer Prevention Research Institute of Texas. We thank the individuals represented in the UK Biobank and the Philadelphia Neurodevelopmental Cohort (PNC) 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. Ethics approval for the UK Biobank study was obtained from the North West Centre for Research Ethics Committee (11/NW/0382). For the PNC study, the institutional review boards of both the University of Pennsylvania and the Children's Hospital of Philadelphia approved all study procedures. Informed consent was obtained from all subjects. We gratefully acknowledge all the studies and databases that made their GWAS summary data available. The authors acknowledge the Texas Advanced Computing Center (TACC, http: //www.tacc.utexas.edu) at The University of Texas at Austin for providing HPC and storage resources that have contributed to the research results reported within this paper.

Code availability

We made use of publicly available software and tools. All codes used to generate results that are reported in this paper are available upon request.

Competing of interest

The authors declare no competing financial interests.

Supplementary Information accompanies this paper is available on the Molecular Psychiatry website.

REFERENCES

1. Penke L, Maniega SM, Bastin M, Hernández MV, Murray C, Royle N *et al.* Brain-wide white matter tract integrity is associated with information processing speed and general intelligence. *Molecular psychiatry* 2012; **17**(10)**:** 955.

2. Tamnes CK, Østby Y, Walhovd KB, Westlye LT, Due-Tønnessen P, Fjell AM. Intellectual abilities and white matter microstructure in development: a diffusion tensor imaging study. *Human brain mapping* 2010; **31**(10)**:** 1609-1625.

3. Ritchie SJ, Bastin ME, Tucker-Drob EM, Maniega SM, Engelhardt LE, Cox SR *et al.* Coupled changes in brain white matter microstructure and fluid intelligence in later life. *Journal of Neuroscience* 2015; **35**(22)**:** 8672-8682.

4. Ritchie SJ, Booth T, Hernández MdCV, Corley J, Maniega SM, Gow AJ *et al.* Beyond a bigger brain: Multivariable structural brain imaging and intelligence. *Intelligence* 2015; **51:** 47-56.

5. Nir TM, Jahanshad N, Villalon-Reina JE, Toga AW, Jack CR, Weiner MW *et al.* Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging. *NeuroImage: clinical* 2013; **3:** 180-195.

6. Bohnen NI, Albin RL. White matter lesions in Parkinson disease. *Nature Reviews Neurology* 2011; **7**(4)**:** 229.

7. Voineskos AN. Genetic underpinnings of white matter 'connectivity': heritability, risk, and heterogeneity in schizophrenia. *Schizophrenia research* 2015; **161**(1)**:** 50-60.

8. Sudre G, Choudhuri S, Szekely E, Bonner T, Goduni E, Sharp W *et al.* Estimating the Heritability of Structural and Functional Brain Connectivity in Families Affected by Attention-Deficit/Hyperactivity Disorder. *JAMA psychiatry* 2017; **74**(1)**:** 76-84.

9. Basser PJ, Mattiello J, LeBihan D. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B* 1994; **103**(3)**:** 247-254.

10. Beaulieu C. The basis of anisotropic water diffusion in the nervous system–a technical review. *NMR in Biomedicine* 2002; **15**(7-8)**:** 435-455.

11. Jones DK, Knösche TR, Turner R. White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. *Neuroimage* 2013; **73:** 239-254.

12. Cox SR, Ritchie SJ, Tucker-Drob EM, Liewald DC, Hagenaars SP, Davies G *et al.* Ageing and brain white matter structure in 3,513 UK Biobank participants. *Nature communications* 2016; **7:** 13629.

13. Smith SM, Jenkinson M, Johansen-Berg H, Rueckert D, Nichols TE, Mackay CE *et al.* Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* 2006; **31**(4)**:** 1487-1505.

14. Tamnes CK, Roalf DR, Goddings A-L, Lebel C. Diffusion MRI of white matter microstructure development in childhood and adolescence: methods, challenges and progress. *Developmental cognitive neuroscience* 2017.

15. Lee SJ, Steiner RJ, Luo S, Neale MC, Styner M, Zhu H *et al.* Quantitative tract-based white matter heritability in twin neonates. *NeuroImage* 2015; **111:** 123-135.

16. Lee SJ, Steiner RJ, Yu Y, Short SJ, Neale MC, Styner MA *et al.* Common and heritable components of white matter microstructure predict cognitive function at 1 and 2 y. *Proceedings of the National Academy of Sciences* 2017; **114**(1)**:** 148-153.

17. Brouwer RM, Mandl RC, Peper JS, van Baal GCM, Kahn RS, Boomsma DI *et al.* Heritability of DTI and MTR in nine-year-old children. *Neuroimage* 2010; **53**(3)**:** 1085-1092.

18. Chiang M-C, Barysheva M, Toga AW, Medland SE, Hansell NK, James MR *et al.* BDNF gene effects on brain circuitry replicated in 455 twins. *Neuroimage* 2011; **55**(2)**:** 448-454.

19. Kochunov P, Jahanshad N, Marcus D, Winkler A, Sprooten E, Nichols TE *et al.* Heritability of fractional anisotropy in human white matter: a comparison of Human Connectome Project and ENIGMA-DTI data. *Neuroimage* 2015; **111:** 300-311.

20. Vuoksimaa E, Panizzon MS, Hagler Jr DJ, Hatton SN, Fennema-Notestine C, Rinker D *et al.* Heritability of white matter microstructure in late middle age: A twin study of tract-based fractional anisotropy and absolute diffusivity indices. *Human brain mapping* 2017; **38**(4)**:** 2026-2036.

21. Kanchibhotla SC, Mather KA, Wen W, Schofield PR, Kwok JB, Sachdev PS. Genetics of ageing-related changes in brain white matter integrity–A review. *Ageing research reviews* 2013; **12**(1)**:** 391-401.

22. Timpson NJ, Greenwood CM, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nature Reviews Genetics* 2017.

23. Badano JL, Katsanis N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nature reviews Genetics* 2002; **3**(10)**:** 779.

24. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 2017; **169**(7)**:** 1177-1186.

25. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* 2011; **88**(1)**:** 76-82.

26. Loh P-R, Bhatia G, Gusev A, Finucane HK, Bulik-Sullivan BK, Pollack SJ *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature genetics* 2015; **47**(12)**:** 1385-1392.

27. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nature genetics* 2011; **43**(6)**:** 519.

28. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* 2015; **47**(11)**:** 1228-1235.

29. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA *et al.* 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics* 2017; **101**(1)**:** 5-22.

30. Watanabe K, Taskesen E, Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* 2017; **8**(1)**:** 1826.

31. Watanabe K, Stringer S, Frei O, Mirkov MU, Polderman TJ, van der Sluis S *et al.* A global view of pleiotropy and genetic architecture in complex traits. *bioRxiv* 2018**:** 500090.

32. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* 2015; **47**(11)**:** 1236-1241.

33. 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 medicine* 2015; **12**(3)**:** e1001779.

34. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JL, Griffanti L, Douaud G *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage* 2018; **166:** 400-424.

35. Rutten-Jacobs LC, Tozer DJ, Duering M, Malik R, Dichgans M, Markus HS *et al.* Genetic study of white matter integrity in UK Biobank (N= 8448) and the overlap with stroke, depression, and dementia. *Stroke* 2018; **49**(6)**:** 1340-1347.

36. Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* 2018; **562**(7726)**:** 210-216.

37. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Computational Biology* 2015; **11**(4)**:** e1004219.

38. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research* 2018; **47**(D1)**:** D1005-D1012.

39. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K *et al.* Genome-wide genetic data on~ 500,000 UK Biobank participants. *BioRxiv* 2017**:** 166298.

40. Jahanshad N, Kochunov PV, Sprooten E, Mandl RC, Nichols TE, Almasy L *et al.* Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the ENIGMA–DTI working group. *Neuroimage* 2013; **81:** 455-469.

41. Kochunov P, Jahanshad N, Sprooten E, Nichols TE, Mandl RC, Almasy L *et al.* Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: comparing meta and megaanalytical approaches for data pooling. *Neuroimage* 2014; **95:** 136-150.

42. Kim MJ, Elliott ML, d'Arbeloff TC, Knodt AR, Radtke SR, Brigidi BD *et al.* Microstructural integrity of white matter moderates an association between childhood adversity and adult trait anger. *Aggressive behavior* 2019; **45**(3)**:** 310-318.

43. Kelly S, Jahanshad N, Zalesky A, Kochunov P, Agartz I, Alloza C *et al.* Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group. *Molecular psychiatry* 2018; **23**(5)**:** 1261.

44. Dennison MJ, Rosen ML, Sambrook KA, Jenness JL, Sheridan MA, McLaughlin KA. Differential associations of distinct forms of childhood adversity with neurobehavioral measures of reward processing: a developmental pathway to depression. *Child development* 2019; **90**(1)**:** e96-e113.

45. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 2007; **81**(3)**:** 559-575.

46. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* 2010; **38**(16)**:** e164-e164.

47. Consortium G. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 2015; **348**(6235)**:** 648-660.

48. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature Neuroscience* 2014; **17**(10)**:** 1418-1428.

49. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature Neuroscience* 2016; **19**(11)**:** 1442-1453.

50. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL *et al.* A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Reports* 2016; **17**(8)**:** 2042-2059.

51. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 2015; **518**(7539)**:** 317.

52. Consortium IS. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**(7256)**:** 748-752.

53. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic Prediction via Bayesian Regression and Continuous Shrinkage Priors. *bioRxiv* 2019**:** 416859.

54. Pasaniuc B, Price AL. Dissecting the genetics of complex traits using summary association statistics. *Nature Reviews Genetics* 2017; **18**(2)**:** 117.

55. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* 2018; **50**(8)**:** 1112–1121.

56. Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics* 2018; **50**(7)**:** 920.

57. Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications* 2018; **9**(1)**:** 2098.

58. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke T-K, Shirali M *et al.* Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics* 2018; **50**(1)**:** 6-11.

59. Kouri N, Ross OA, Dombroski B, Younkin CS, Serie DJ, Soto-Ortolaza A *et al.* Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. *Nature communications* 2015; **6:** 7247.

60. Lam M, Trampush JW, Yu J, Knowles E, Davies G, Liewald DC *et al.* Large-Scale Cognitive GWAS Meta-Analysis Reveals Tissue-Specific Neural Expression and Potential Nootropic Drug Targets. *Cell reports* 2017; **21**(9)**:** 2597-2613.

61. Chang D, Nalls MA, Hallgrímsdóttir IB, Hunkapiller J, van der Brug M, Cai F *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nature genetics* 2017; **49**(10)**:** 1511.

62. Okbay A, Baselmans BM, De Neve J-E, Turley P, Nivard MG, Fontana MA *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics* 2016; **48**(6)**:** 624–633.

63. Sanchez-Roige S, Palmer AA, Fontanillas P, Elson SL, Team aR, Consortium SUDWGotPG *et al.* Genome-wide association study meta-analysis of the Alcohol Use Disorders Identification Test (AUDIT) in two population-based cohorts. *American Journal of Psychiatry* 2018**:** appi. ajp. 2018.18040369.

64. Jun G, Ibrahim-Verbaas CA, Vronskaya M, Lambert J-C, Chung J, Naj AC *et al.* A novel Alzheimer disease locus located near the gene encoding tau protein. *Molecular psychiatry* 2016; **21**(1)**:** 108.

65. Trampush JW, Yang M, Yu J, Knowles E, Davies G, Liewald D *et al.* GWAS meta-analysis reveals novel loci and genetic correlates for general cognitive function: a report from the COGENT consortium. *Molecular psychiatry* 2017; **22**(3)**:** 336.

66. Li Z, Chen J, Yu H, He L, Xu Y, Zhang D *et al.* Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nature genetics* 2017; **49**(11)**:** 1576.

67. Bergen S, O'dushlaine C, Ripke S, Lee P, Ruderfer D, Akterin S *et al.* Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Molecular psychiatry* 2012; **17**(9)**:** 880.

68. Kramer PL, Xu H, Woltjer RL, Westaway SK, Clark D, Erten-Lyons D *et al.* Alzheimer disease pathology in cognitively healthy elderly: a genome-wide study. *Neurobiology of aging* 2011; **32**(12)**:** 2113-2122.

69. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society Series B (Methodological)* 1995**:** 289-300.

70. Satterthwaite TD, Elliott MA, Ruparel K, Loughead J, Prabhakaran K, Calkins ME *et al.* Neuroimaging of the Philadelphia neurodevelopmental cohort. *Neuroimage* 2014; **86:** 544-553.

71. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genetics* 2006; **38**(2)**:** 209-213.

72. Clarke T, Lupton M, Fernandez-Pujals A, Starr J, Davies G, Cox S *et al.* Common polygenic risk for autism spectrum disorder (ASD) is associated with cognitive ability in the general population. *Molecular psychiatry* 2016; **21**(3)**:** 419.

73. Mistry S, Harrison JR, Smith DJ, Escott-Price V, Zammit S. The use of polygenic risk scores to identify phenotypes associated with genetic risk of bipolar disorder and depression: A systematic review. *Journal of affective disorders* 2018.

74. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. *Nature Reviews Genetics* 2013; **14**(7)**:** 483.

75. Bach M, Laun FB, Leemans A, Tax CM, Biessels GJ, Stieltjes B *et al.* Methodological considerations on tract-based spatial statistics (TBSS). *Neuroimage* 2014; **100:** 358-369.

Figure legends

Figure 1. SNP heritability estimates grouped by white matter tract functions. The white matter tracts are clustered into four communities including complex fibers (C1), associative fibers (C2), commissural fibers (C3), and projection fibers (C4) according to the Connectopedia Knowledge Database, http://www.fmritools.com/kdb/white-matter/

Figure 2. Distribution of SNP heritability estimates of the 21 white matter tracts in brain.

Figure 3. Heritability estimated by SNPs in each chromosome or in functionally annotated SNP categories.

Figure 4. Number of independent significant SNPs discovered for each DTI parameter at different GWAS significance levels. Outer layer: p-value $<5*10^{-8}$; middle layer: p-value $<5*10^{-9}$; and inner layer: p-value $<4.5*10^{-10}$.

Figure 5. Genes identified in gene-based association analysis of DTI parameters that have been implicated with traits of neuroticism, neurodegenerative diseases, neuropsychiatric disorders, education, cognitive, and reaction time in previous GWAS.

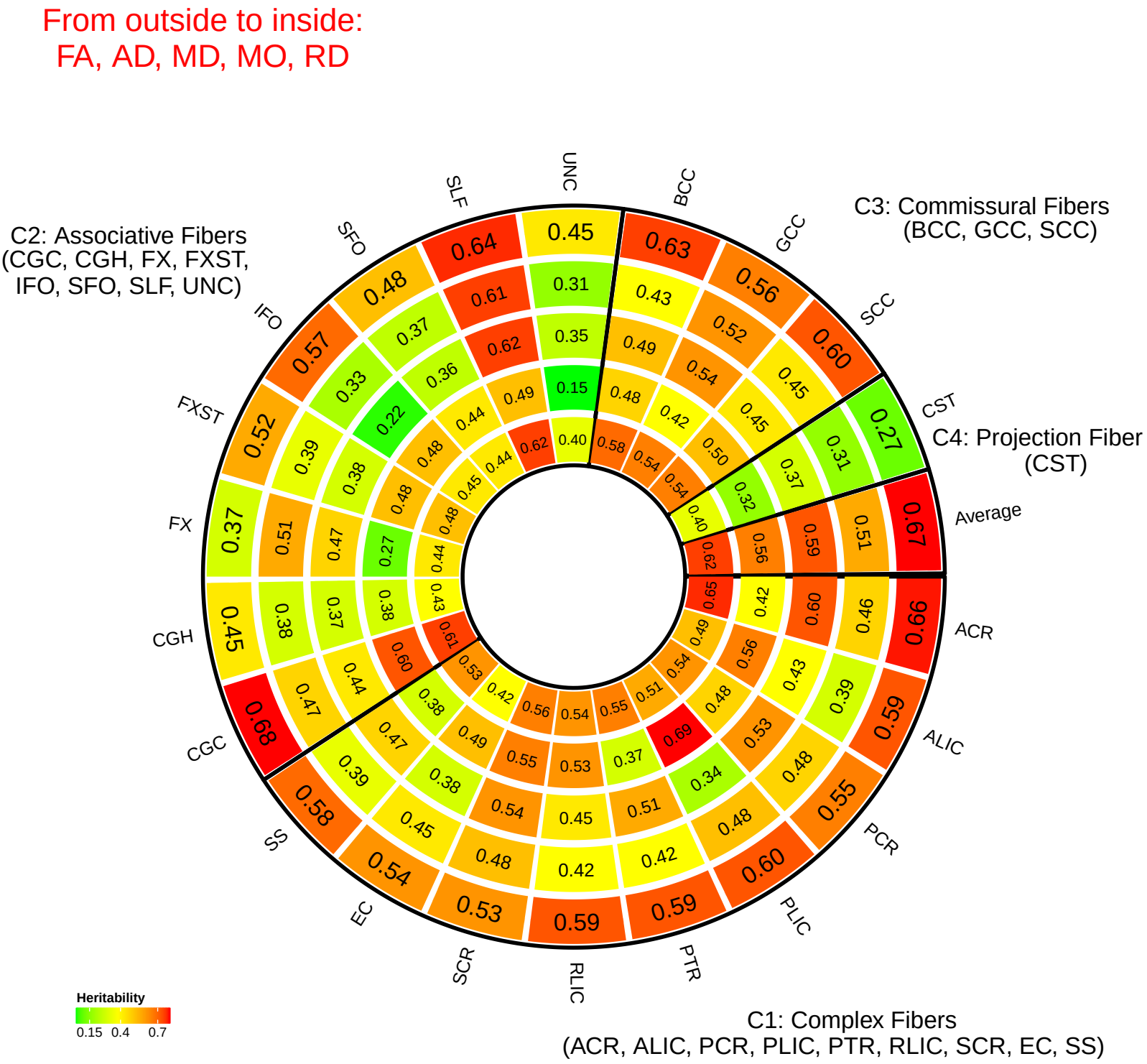
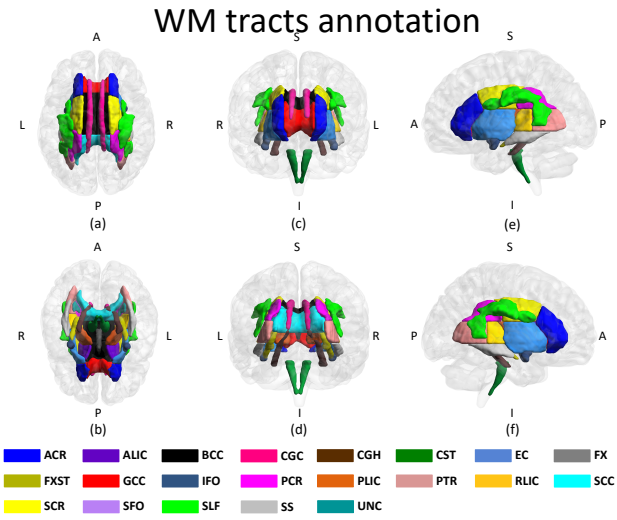
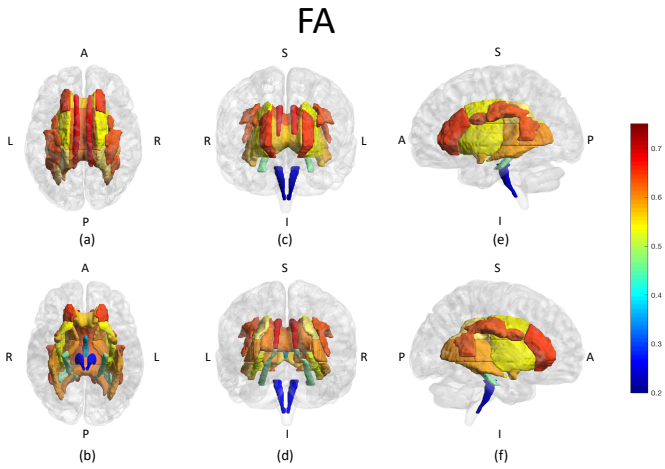


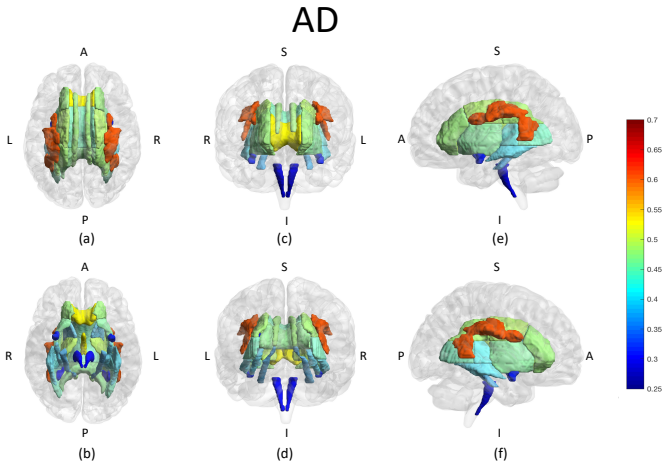
Figure 1: SNP heritability estimates grouped by white matter tract functions. The white matter tracts are clustered into four communities: complex fibers (C1), associative fibers (C2), commissural fibers (C3) and projection fibers (C4) according to the Connectopedia Knowledge Database, `http://www.fmritools.com/kdb/white-matter/`.



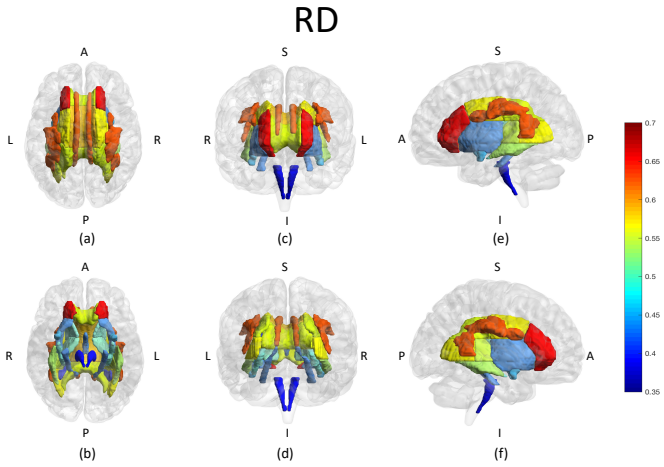
Anatomical ROIs of TBSS: (a) Superior view (b) Inferior view (c) Anterior view (d) Posterior view (e) Left view (f) Right view



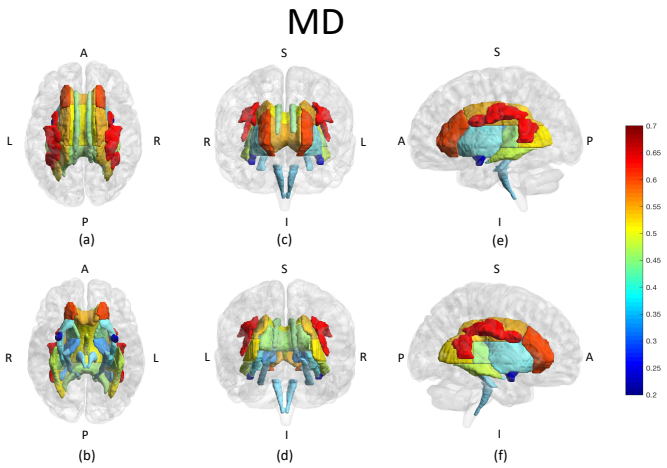
Estimated Heritability of FA on TBSS ROIs: (a) Superior view (b) Inferior view (c) Anterior view (d) Posterior view (e) Left view (f) Right view



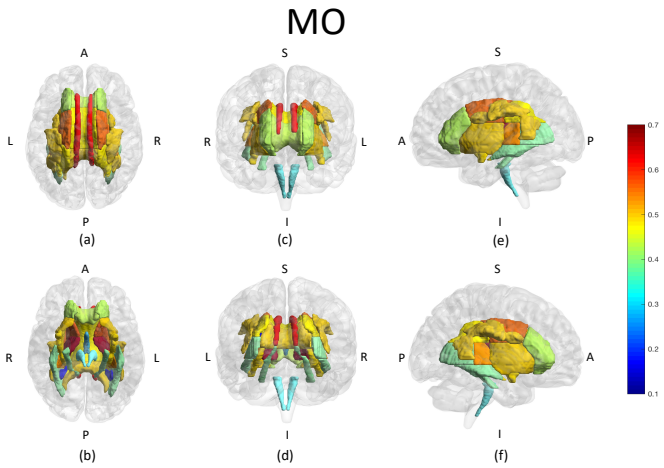
Estimated Heritability of AD on TBSS ROIs: (a) Superior view (b) Inferior view (c) Anterior view (d) Posterior view (e) Left view (f) Right view



Estimated Heritability of RD on TBSS ROIs: (a) Superior view (b) Inferior view (c) Anterior view (d) Posterior view (e) Left view (f) Right view

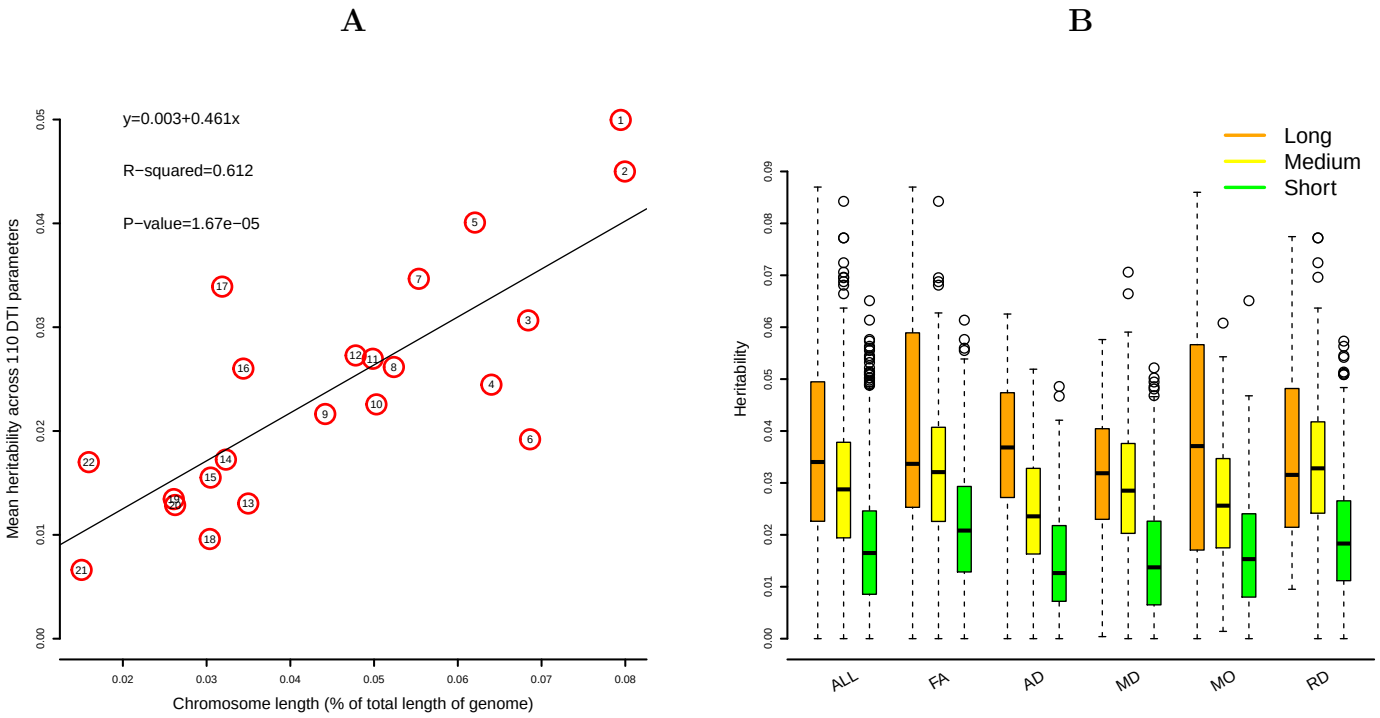


Estimated Heritability of MD on TBSS ROIs: (a) Superior view (b) Inferior view (c) Anterior view (d) Posterior view (e) Left view (f) Right view



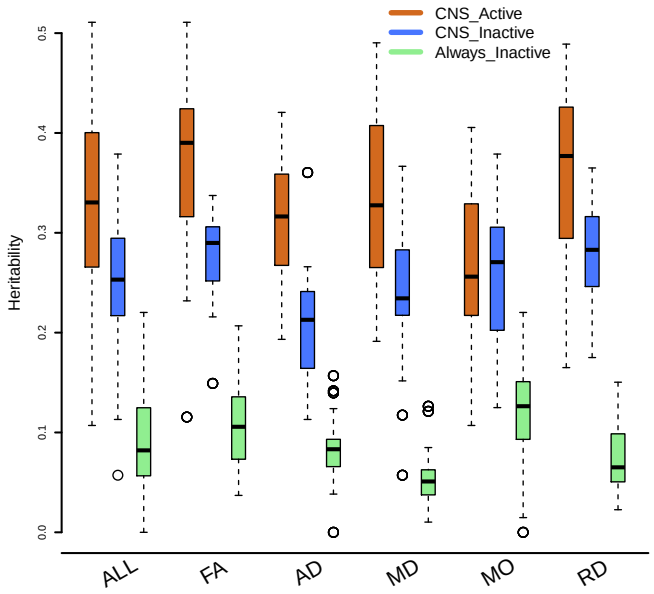
Estimated Heritability of MO on TBSS ROIs: (a) Superior view (b) Inferior view (c) Anterior view (d) Posterior view (e) Left view (f) Right view

Figure 2: Distribution of SNP heritability estimates of the 21 white matter tracts in brain.



(a) Mean SNP heritability of each chromosome.

(b) SNP heritability grouped by chromosome length.



(c) SNP heritability grouped by CNS functional annotations.

Figure 3: Heritability estimated by SNPs in each chromosome or in functionally annotated SNP categories.

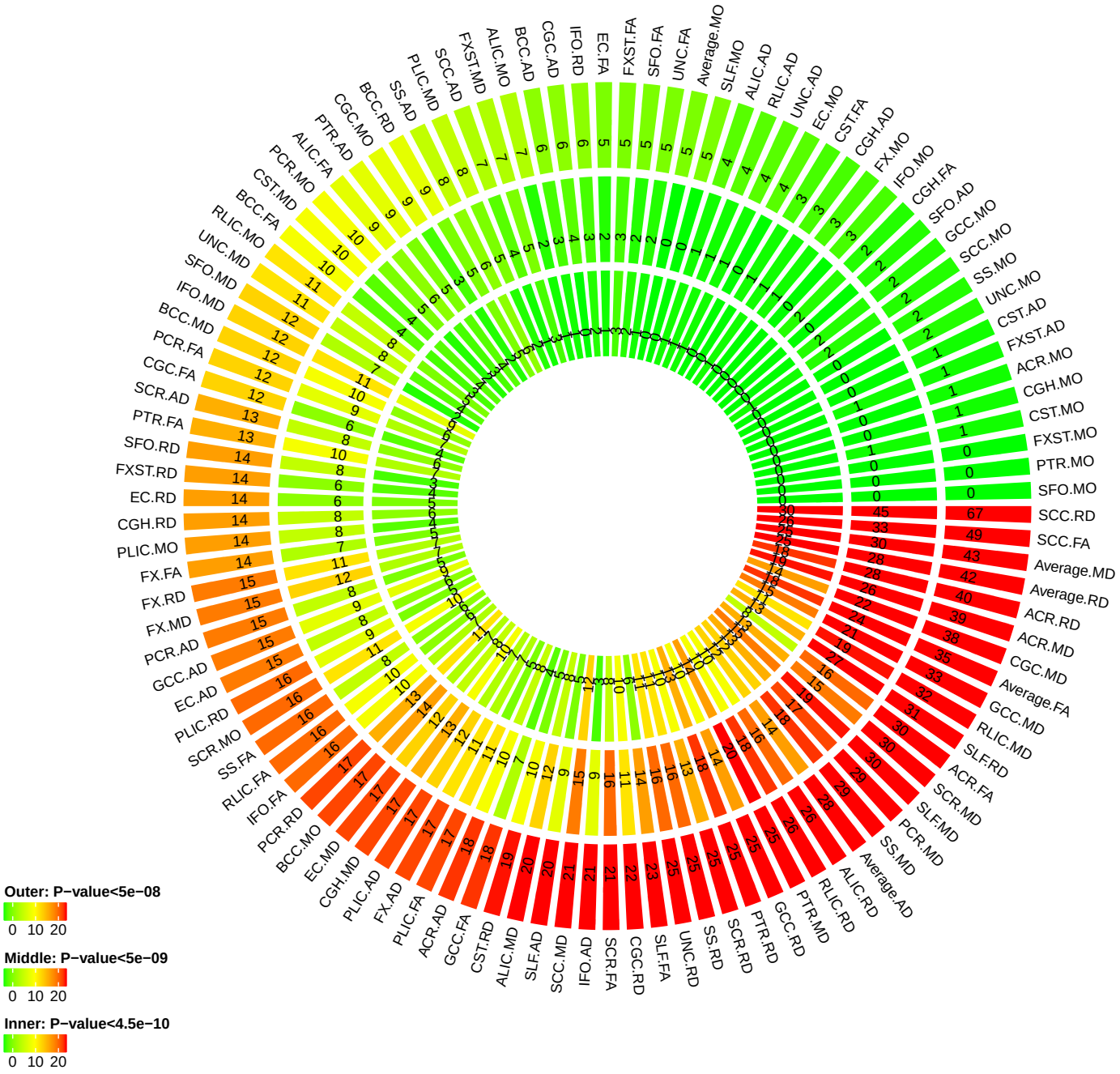


Figure 4: Number of independent significant SNPs discovered for each DTI parameter at different GWAS significance levels. Outer layer: p-value $< 5 \times 10^{-8}$; middle layer: p-value $< 5 \times 10^{-9}$; and inner layer: p-value $< 4.5 \times 10^{-10}$.

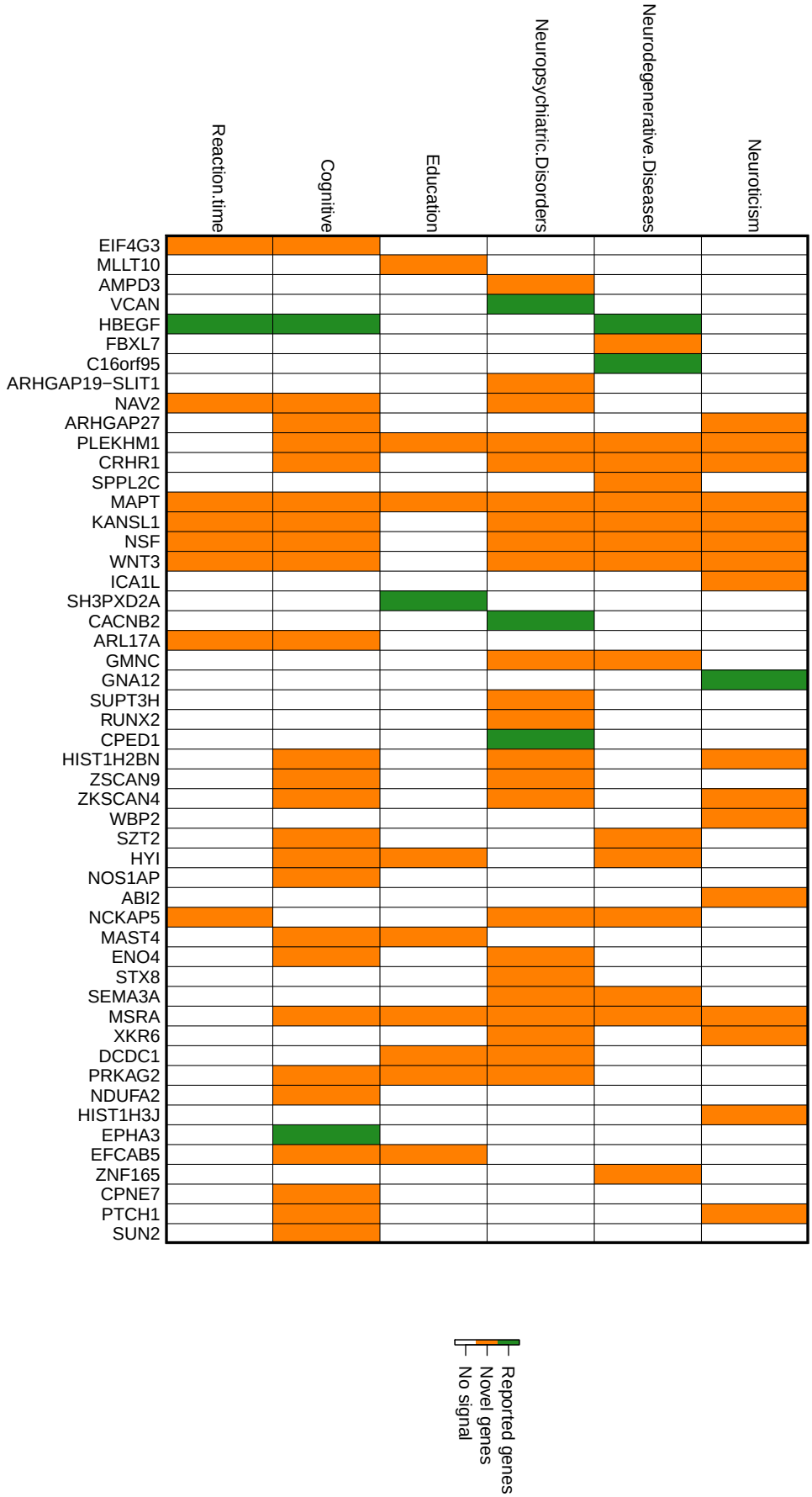


Figure 5: Genes identified in gene-based association analysis of DTI parameters that have been implicated with traits of neuroticism, neurodegenerative diseases, neuropsychiatric disorders, education, cognitive, and reaction time in previous GWAS.