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, M.S.[1], Huiling Liao. B.S.[8], Thomas E. Nichols, Ph.D.[9,10], and Hongtu Zhu, Ph.D.*[1]

[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

[10] Department of Statistics, University of Warwick, Coventry, UK, CV4 7AL

**Corresponding author*:
Hongtu Zhu
Department of Biostatistics, University of North Carolina at Chapel Hill,
3105C McGavran-Greenberg Hall, Chapel Hill, NC 27599.
Email address: htzhu@email.unc.edu Phone: (919) 966-7250

Key Words: Genetic architecture; Genetic overlap; Pleiotropy; UK Biobank; DTI, White matter microstructure.

Number of words in the abstract: 227, Number of words in the main text: 3386,
Number of figures: 6, Number of tables: 0, Number of supplemental files: 2.

Abstract

Background

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 opportunity to identify novel genetic variants of WM tracts and explore the genetic overlap with other brain-related complex traits.

Method

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.

Results

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.

Conclusions

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 (1-5) as well as neurodegenerative/neuropsychiatric traits, such as Alzheimer's disease (6), Parkinson's disease (7), schizophrenia (SCZ) (8), and attention-deficit/hyperactivity disorder (ADHD) (9). A better understanding of genetic factors influencing integrity of WM tracts could have important implication for understanding the etiology of these diseases as well as individual variation in intelligence. To reveal the underlying genetic contributions to brain structural development and disease/disorder processes, imaging genetics studies of WM microstructure has been an active research area over the past fifteen years. The structural changes of WM tracts are typically measured and quantified in diffusion tensor imaging (DTI) (10). Brain diffusivity can be influenced by many aspects of its micro- or macro-structures (11). To reconstruct the WM pathways and tissue microstructure, DTI models the diffusion properties of WM using random movement of water. Specifically, DTI quantifies diffusion magnetic resonance imaging (dMRI) in a tensor model and analyzes diffusions in all directions. A typical DTI diagonalizes the tensor and calculates three pairs of eigenvalues/eigenvectors that respectively represent one primary and two secondary diffusion directions. Within each voxel, several DTI parameters can be derived: fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), and mode of anisotropy (MO). As a summary measure of WM integrity (12, 13), higher FA indicates stronger directionality in this voxel. MD quantifies the magnitude 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 (14). 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) (15), as well as tractography and graph theory analysis; see Tamnes, Roalf (16) 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 (17, 18), young children (19), older children (20, 21), adolescents (22), and young adults (23) to middle aged (24) and older adults (25). Participants in these studies are typically monozygotic and dizygotic twins or family members. Table 1 of Vuoksimaa, Panizzon (24) 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 the limitation of family-based studies, for which the heritability estimation has relied on contrasting the phenotypic similarity between monozygotic and dizygotic twins. Genetic architecture denotes the characteristics of genetic variations that contribute to the broad-sense heritability of a phenotype (26). Based on the number of genetic variants contributing to phenotypic variance, genetic architecture can be described as monogenic (one variant), oligogenic (few variants), polygenic (many variants), or omnigenic, which hypothesizes that almost all genetic variants have small but non-zero genetic contributions (27, 28). Uncovering the genetic architecture and discovering the associated genetic variants are essential steps to delineate the functional mechanisms and understand the genetic overlap between white matter structures and neuropsychiatric traits.

Recent developments have enabled heritability estimation and genetic variants discovery with using the common single-nucleotide polymorphisms (SNPs) data collected in general populations. Instead of using the expected genetic correlation based 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) (29, 30). The architecture of genetic influences can be assessed by SNP annotation and partition (31, 32). Genome-wide association studies (GWAS) and post-GWAS analysis can further identify causal genetic variants at SNP, locus and gene levels (33, 34), and assess the genetic overlap of complex traits in different domains (35, 36). With these methods, the availability of genomic and imaging data from recent large population-based United Kingdom (UK) Biobank resource (37) offers the opportunity to

uncover the genetic basis of brain WM tracts in one large-scale, relatively homogeneous population. The UK Biobank 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 (38).

Here we generated 110 tract-based DTI parameters using the British ancestry UKB sample including 17,706 participants. For each of the 110 phenotypes, we estimated the SNP-heritability, assessed the distribution of genetic effects by SNP annotation and partition, and carried out GWAS to identify the associated genetic variants at SNP and locus levels. We further discovered gene-level associations via MAGMA (39), and explored the functional consequences of the significant SNPs by functional mapping and annotation analysis (FUMA, (34)). To detect genetic overlap and pleiotropy in WM tracts and other complex traits, we performed association lookups at SNP and gene levels on the NHGRI-EBI GWAS catalog (40) and estimated genetic correlations via LD score regression (LDSC, (36)). 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 DTI data (38) 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 ID and full names of the 21 WM tracts are listed in **Supplementary Table 1.** A full description of the DTI registration and quality controls is documented in supplementary information and an overview is given in **Supplementary Figure 1**. We removed values greater than five times the median absolute deviation from the median in each continuous variable. All individuals were aged between 40 and 80 years and the proportion of females was 52.9%.

Genotyping and quality control

We downloaded the imputed SNP data from UKB data resource (41). We further performed the following SNP data quality controls using PLINK (42): excluding subjects with more than 10% missing genotypes, only including SNPs with MAF > 0.01, genotyping rate > 90%, and passing Hardy-Weinberg test (p-value>$1*10^{-7}$). We also removed SNPs with imputation INFO score less than 0.8.

SNP-heritability analysis and genome-wide association analysis

For each of the 110 DTI parameters, we estimated the proportion of variation explained by all autosomal SNPs with using linear mixed-effects model (LMM) via GCTA (29). We considered the 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 (GPCs, Data-Field 22009). We also estimated the proportion of variation explained by SNPs in each chromosome. In addition, we partitioned the SNPs according to cell-type-specific annotations (32) to perform enrichment analysis. SNPs were grouped according to their functional activeness in various cell groups 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 (42). The same set of covariates were adjusted in all heritability and GWAS analyses.

Genomic risk loci characterization and comparison with previous findings

We characterized genomic risk loci by using FUMA (34) 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 of LD blocks), they were merged to a single genomic locus. More details of FUMA analysis can be found in (34). Independent significant SNPs and all SNPs in LD with them were subsequently searched on NHGRI-EBI GWAS catalog (40) (v2019-01-31) to look for reported associations (p-value<$9*10^{-6}$) with any traits.

Gene-based association analysis and functional annotation

We carried out gene-based association analysis for 18,796 protein-coding candidate genes via MAGMA (39) (v1.07). Gene-based p-values were calculated by summarizing the GWAS results of corresponding SNPs, which were mapped to genes according to their psychical positions. Significant genes were searched on NHGRI-EBI GWAS catalog (40) (v2019-01-31) to look for their previously reported associations with any traits. We focused on brain-related complex traits and characterized them into six groups: cognitive (e.g., general cognitive ability, cognitive performance, math ability, and intelligence), education (e.g., years of education and college completion), reaction time, neuroticism, neurodegenerative diseases (e.g., Alzheimer's disease, Parkinson's disease 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 (43). 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 (44), BRAINEAC (45), and CommonMind Consortium (46)) and chromatin interaction mapping (built-in chromatin interaction data: dorsolateral prefrontal cortex, hippocampus (47); annotate enhancer/promoter regions: E053-E082 brain (48)). 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.

Results

SNP heritability estimation

Figures 1-2 and **Supplementary Figures 2-7** 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). SNP heritability of FA decreased to 37% and 27%, respectively, in the fornix (FX) and corticospinal tract (CST). According to the functions of the WM tracts (Connectopedia Knowledge Database, http://www.fmritools.com/kdb/white-matter/), we clustered them into four communities: complex fibers (C1), associative fibers (C2), commissural fibers (C3) and projection fibers (C4). 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 the chromosome **(Figure 3(a)**, R^2=61.2%, p-value=$1.67*10^{-05}$). This finding reveals a highly polygenic or omnigenic genetic architecture (28) of WM tracts. The large number of SNPs that contribute to the variation in DTI parameters are widely spread across the whole genome. To further illustrate this architecture, we ordered and clustered the 22 chromosomes into three groups by their lengths: long, medium, and short. The long

group had 4 chromosomes (CHRs 2, 1, 6, 3), which together accounted for 33% of the length of the whole genome; the medium group had 6 chromosomes (CHRs 4, 5, 7, 8, 10, 11), which accounted for another 33% of the length of the whole genome; and the short group consisted of the remaining 12 chromosomes. **Figure 3(b)** shows the SNP heritability estimates grouped by chromosomal length. It is clear that longer chromosomes tended to have higher SNP heritability estimates than medium (p-value=$3.82*10^{-13}$) or shorter (p-value<$2.20*10^{-16}$) ones for DTI parameters.

To compare the contribution of SNPs with different activity level, we partitioned the genetic variation according to CNS-cell-specific annotations: CNS_active, CNS_inactive, 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 8**. 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 9, 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 associations with 90 DTI parameters **(Figure 4, Supplementary Tables 6-7)**. RD and FA of splenium of corpus callosum (SCC) had the largest number of independent significant SNPs. 502 of the 696 independent significant associations located in chromosome 5 **(Supplementary Table 8, Supplementary Figure 10)**. The 696 independent significant SNP-level associations can be further characterized as 205 locus-level associations

(Supplementary Table 9). FA and RD of SCC, FA and AD of FX, and RD of ALIC had at least five genetic risk loci **(Supplementary Table 10)**. Each chromosome had at least one genetic risk locus except for chromosomes 13 ,20 and 21, and chromosome 5 had the largest number of risk loci **(Supplementary Tables 11)**. GWAS results at $5*10^{-9}$ and $5*10^{-8}$ significance levels are also reported in above tables and figures.

Concordance with previous GWAS results

Association lookups on the NHGRI-EBI GWAS catalog (40) 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 (49) (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 (50) (n=8,428) and found that 212 of the 368 significant associations (Supplementary Table 6 of (50)) were replicated in the present study. We note that the both (49) and (50) analyzed a subset of the sample presented in this 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 Rutten-Jacobs, Tozer (49) and Elliott, Sharp (50), including *VCAN, C16orf95, NBEAL1, SH3PXD2A, CACNB2, SRA1, GNA12, CPED1,* and *EPHA3,* but most of the identified genes were not previously linked to DTI parameters. Association lookups found that 51 of the 112 significant genes were implicated with cognitive, education, reaction, neuroticism, neuropsychiatric and neurodegenerative traits in previous studies, such as *CRHR1 (51-54)*, *MAPT (55-58)*, *KANSL1 (59-61)*, and *MSRA* (62-64) **(Supplementary Table 15, Figure 5)**. We also annotated the SNPs by functional consequences on gene functions **(Supplementary Figure 11)** and performed functional gene mapping. Gene mapping

discovered 292 genes **(Supplementary Table 16)**, 218 of which were not detected in the gene-based association analysis.

Genetic correction with other traits

We estimated the pairwise genetic correlation between 110 DTI parameters and 14 other complex traits **(Supplementary Table 17)**. We focused on traits showing evidence of pleiotropy in association lookups. 43 pairs of phenotypes had significant genetic correlation after adjusting for multiple testing (1,540 tests) by the Benjamini-Hochberg (B-H) procedure (65) at 0.05 level **(Supplementary Table 18, Supplementary Figure 12).** 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) **(Figure 6)**. 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.

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 (49, 50) 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). With the DTI and common autosomal SNP data from 17,706 UKB participants, the present study made novel contributions to 1) understand the genetic landscape of WM tract; 2) identify novel genetic risk variants at SNP, locus, and gene levels; and 3) provide the overview of statistical pleiotropy (35, 66) with other brain-related complex traits.

Our SNP heritability estimates are close to the ones reported in previous family-based studies (e.g., Table 1 of (24)), and are also within a similar range as those reported in (50), where the mean heritability is around 0.450. These results suggest that studies of DTI phenotypes using common SNPs may be more informative than studies focused on

rare variants. Our results partitioning the genetic variation in chromosomes or SNP functional sets shed light on the distribution of genetic signals across the genome and different functional consequences. These findings suggest a highly polygenic genetic architecture of DTI parameters and also provide evidence for stronger genetic signals from SNPs in active chromatin regions, especially for those active in the CNS cell type. For such highly polygenic traits, large sample size is essential for GWAS to discover the widespread genetic signals (67). Compared to previous GWAS (49, 50), our study with larger sample size identifies many newly associated genetic variants for DTI parameters. More importantly, these novel findings uncover the widespread pleiotropy between DTI parameters and cognitive and metal health 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 (23), 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 (68). 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.

Funding

This research was partially supported by U.S. NIH grants MH086633 and MH116527, and a grant from the Cancer Prevention Research Institute of Texas.

Acknowledgements

We thank the individuals represented in the UK Biobank dataset 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 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.

Conflict of interest

The authors declare no competing financial interests.

REFERENCES

1. Penke L, Maniega SM, Murray C, Gow AJ, Hernández MCV, Clayden JD, et al. (2010): A general factor of brain white matter integrity predicts information processing speed in healthy older people. *Journal of Neuroscience*. 30:7569-7574.
2. Penke L, Maniega SM, Bastin M, Hernández MV, Murray C, Royle N, et al. (2012): Brain-wide white matter tract integrity is associated with information processing speed and general intelligence. *Molecular psychiatry*. 17:955.
3. Tamnes CK, Østby Y, Walhovd KB, Westlye LT, Due-Tønnessen P, Fjell AM (2010): Intellectual abilities and white matter microstructure in development: a diffusion tensor imaging study. *Human brain mapping*. 31:1609-1625.
4. Ritchie SJ, Bastin ME, Tucker-Drob EM, Maniega SM, Engelhardt LE, Cox SR, et al. (2015): Coupled changes in brain white matter microstructure and fluid intelligence in later life. *Journal of Neuroscience*. 35:8672-8682.

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

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

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

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

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

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

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

12. Thomason ME, Thompson PM (2011): Diffusion imaging, white matter, and psychopathology. *Annual review of clinical psychology*. 7.

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

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

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

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

17. Geng X, Gouttard S, Sharma A, Gu H, Styner M, Lin W, et al. (2012): Quantitative tract-based white matter development from birth to age 2years. *Neuroimage*. 61:542-557.

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

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

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

21. Brouwer RM, Mandl RC, Schnack HG, van Soelen IL, van Baal GC, Peper JS, et al. (2012): White matter development in early puberty: a longitudinal volumetric and diffusion tensor imaging twin study. *PloS one*. 7:e32316.

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

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

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

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

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

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

28. Boyle EA, Li YI, Pritchard JK (2017): An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. 169:1177-1186.
29. Yang J, Lee SH, Goddard ME, Visscher PM (2011): GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*. 88:76-82.
30. Loh P-R, Bhatia G, Gusev A, Finucane HK, Bulik-Sullivan BK, Pollack SJ, et al. (2015): Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature genetics*. 47:1385-1392.
31. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. (2011): Genome partitioning of genetic variation for complex traits using common SNPs. *Nature genetics*. 43:519.
32. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. (2015): Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics*. 47:1228-1235.
33. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. (2017): 10 years of GWAS discovery: biology, function, and translation. *The American Journal of Human Genetics*. 101:5-22.
34. Watanabe K, Taskesen E, Bochoven A, Posthuma D (2017): Functional mapping and annotation of genetic associations with FUMA. *Nature communications*. 8:1826.
35. Watanabe K, Stringer S, Frei O, Mirkov MU, Polderman TJ, van der Sluis S, et al. (2018): A global view of pleiotropy and genetic architecture in complex traits. *bioRxiv*.500090.
36. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, et al. (2015): An atlas of genetic correlations across human diseases and traits. *Nature genetics*. 47:1236.
37. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. (2015): UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*. 12:e1001779.
38. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JL, Griffanti L, Douaud G, et al. (2018): Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *NeuroImage*. 166:400-424.
39. de Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015): MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology*. 11:e1004219.

40. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. (2018): The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research*. 47:D1005-D1012.
41. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. (2017): Genome-wide genetic data on~ 500,000 UK Biobank participants. *BioRxiv*.166298.
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. (2007): PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*. 81:559-575.
43. Wang K, Li M, Hakonarson H (2010): ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*. 38:e164-e164.
44. Consortium G (2015): The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 348:648-660.
45. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. (2014): Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature neuroscience*. 17:1418.
46. Fromer M, Roussos P, Sieberts SK, Johnson JS, Kavanagh DH, Perumal TM, et al. (2016): Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature neuroscience*. 19:1442.
47. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL, et al. (2016): A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell reports*. 17:2042-2059.
48. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, et al. (2015): Integrative analysis of 111 reference human epigenomes. *Nature*. 518:317.
49. Rutten-Jacobs LC, Tozer DJ, Duering M, Malik R, Dichgans M, Markus HS, et al. (2018): Genetic study of white matter integrity in UK Biobank (N= 8448) and the overlap with stroke, depression, and dementia. *Stroke*. 49:1340-1347.
50. Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, et al. (2018): Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature*. 562:210.
51. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. (2018): Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics*. 50:1112.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

67. Zhao B, Zou F (2018): Is PRS Good for Predicting Complex Polygenic Traits? *bioRxiv*.447797.

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

Captions to figures

Figure 1. SNP heritability estimates grouped by white matter tract functions.

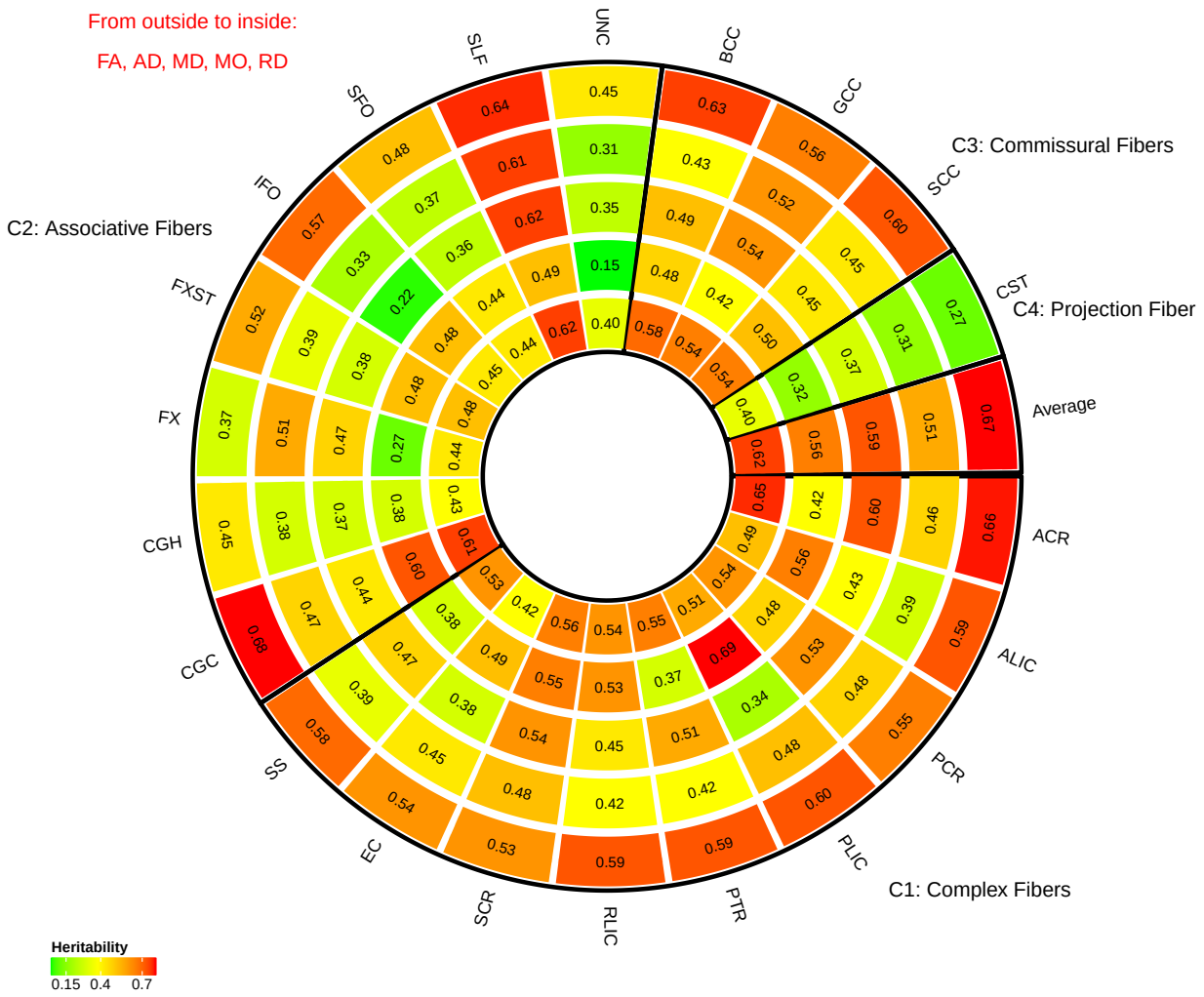
Figure 2. Distribution of SNP heritability estimates of the 21 white matter tracts in brain.

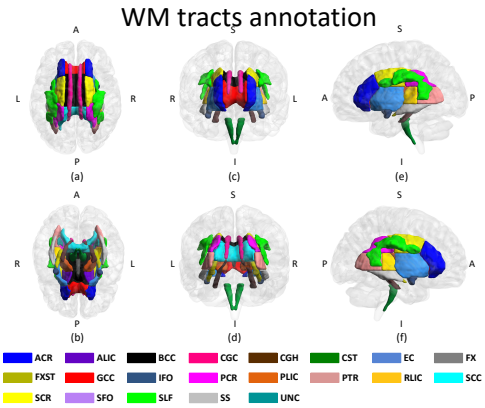
Figure 3. Heritability estimated by SNPs in each chromosome or in functionally annotated SNP categories.

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 $<4.5*10^{-10}$.

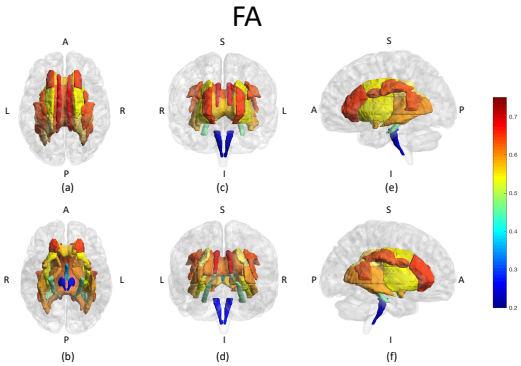
Figure 5. Genes identified in gene-based association analysis of DTI parameters that have been implicated with traits of neuroticism, neurodegenerative diseases, neuropsychiatric disorders, education, cognitive, and reaction time in previous GWAS.

Figure 6. Selected pairwise genetic correlations between DTI parameters and other traits.

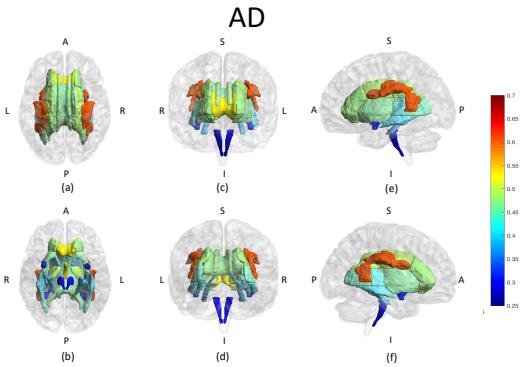

From outside to inside:
FA, AD, MD, MO, RD
C1: Complex Fibers
C2: Associative Fibers
C3: Commissural Fibers
C4: Projection Fiber
Average
UNC
BCC
GCC
SCC
CST
ACR
ALIC
PCR
PLIC
PTR
RLIC
SCR
EC
SS
CGC
CGH
FX
FXST
IFO
SFO
SLF
Heritability
0.15 0.4 0.7



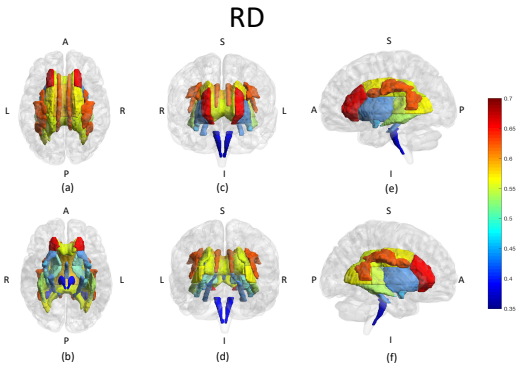
Anatomical ROIs of TBSS: (a) Superior view (b) Inferior view (c) Anterior view (d) Posterior view (e) Left view (f) Right view



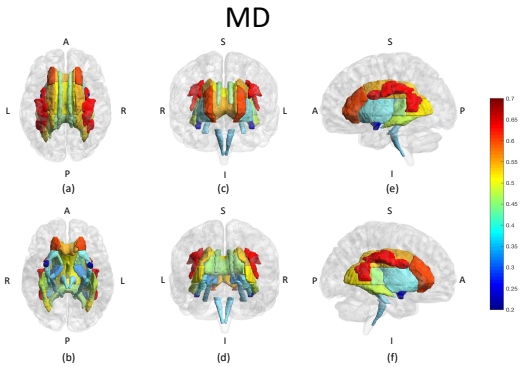
Estimated Heritability of FA on TBSS ROIs: (a) Superior view (b) Inferior view (c) Anterior view (d) Posterior view (e) Left view (f) Right view



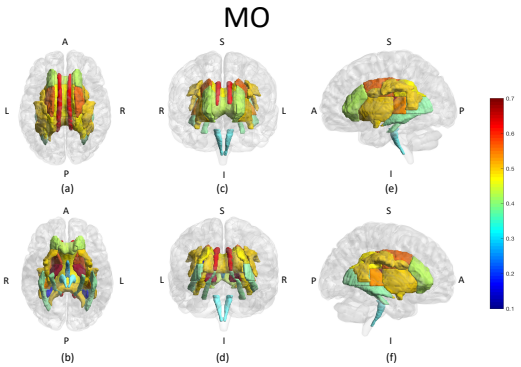
Estimated Heritability of AD on TBSS ROIs: (a) Superior view (b) Inferior view (c) Anterior view (d) Posterior view (e) Left view (f) Right view



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

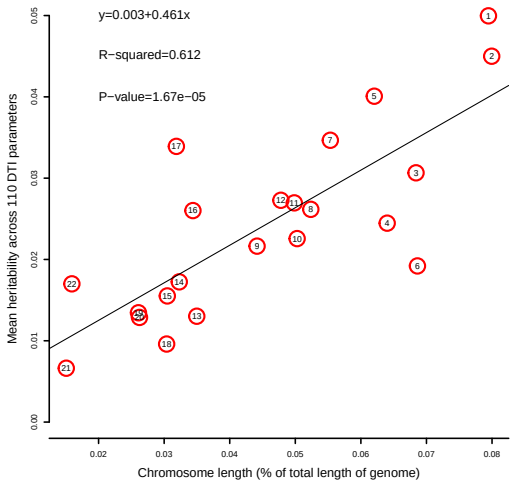


Estimated Heritability of MD on TBSS ROIs: (a) Superior view (b) Inferior view (c) Anterior view (d) Posterior view (e) Left view (f) Right view



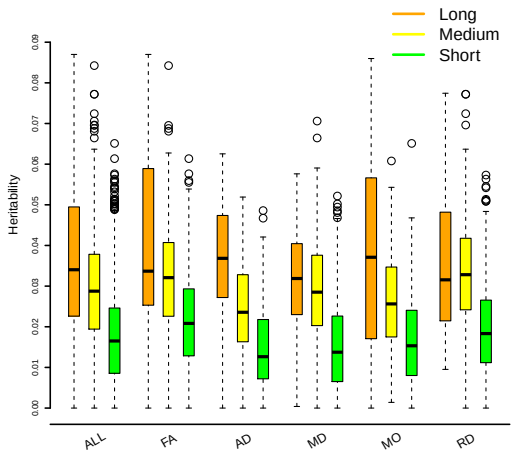
Estimated Heritability of MO on TBSS ROIs: (a) Superior view (b) Inferior view (c) Anterior view (d) Posterior view (e) Left view (f) Right view

A



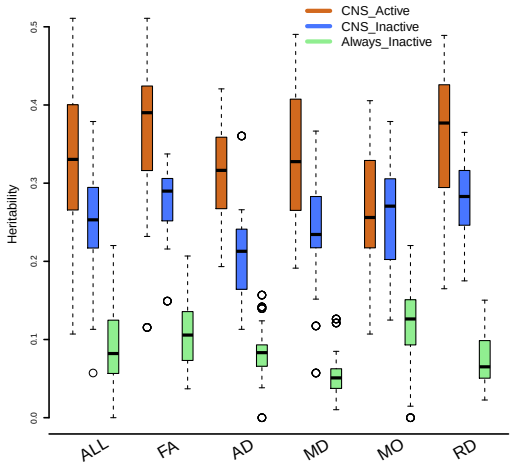
(a) Mean SNP heritability of each chromosome.

B

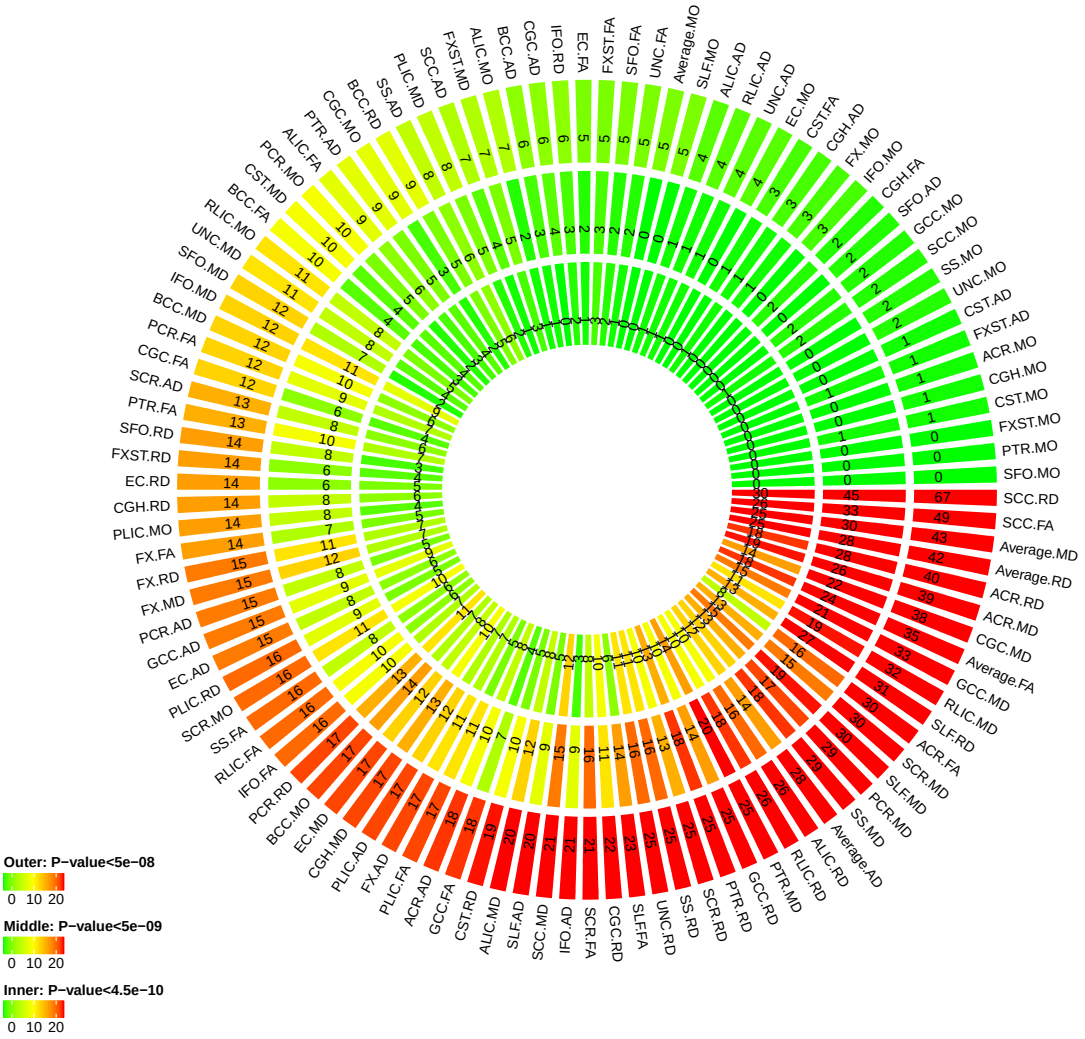


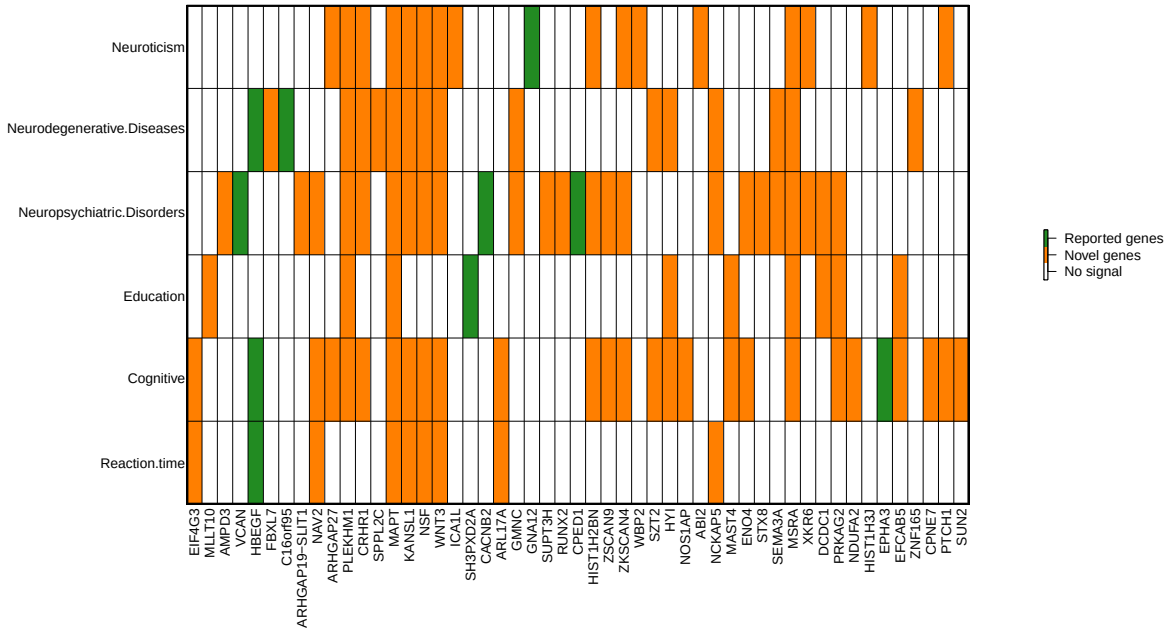
(b) SNP heritability grouped by chromosome length.

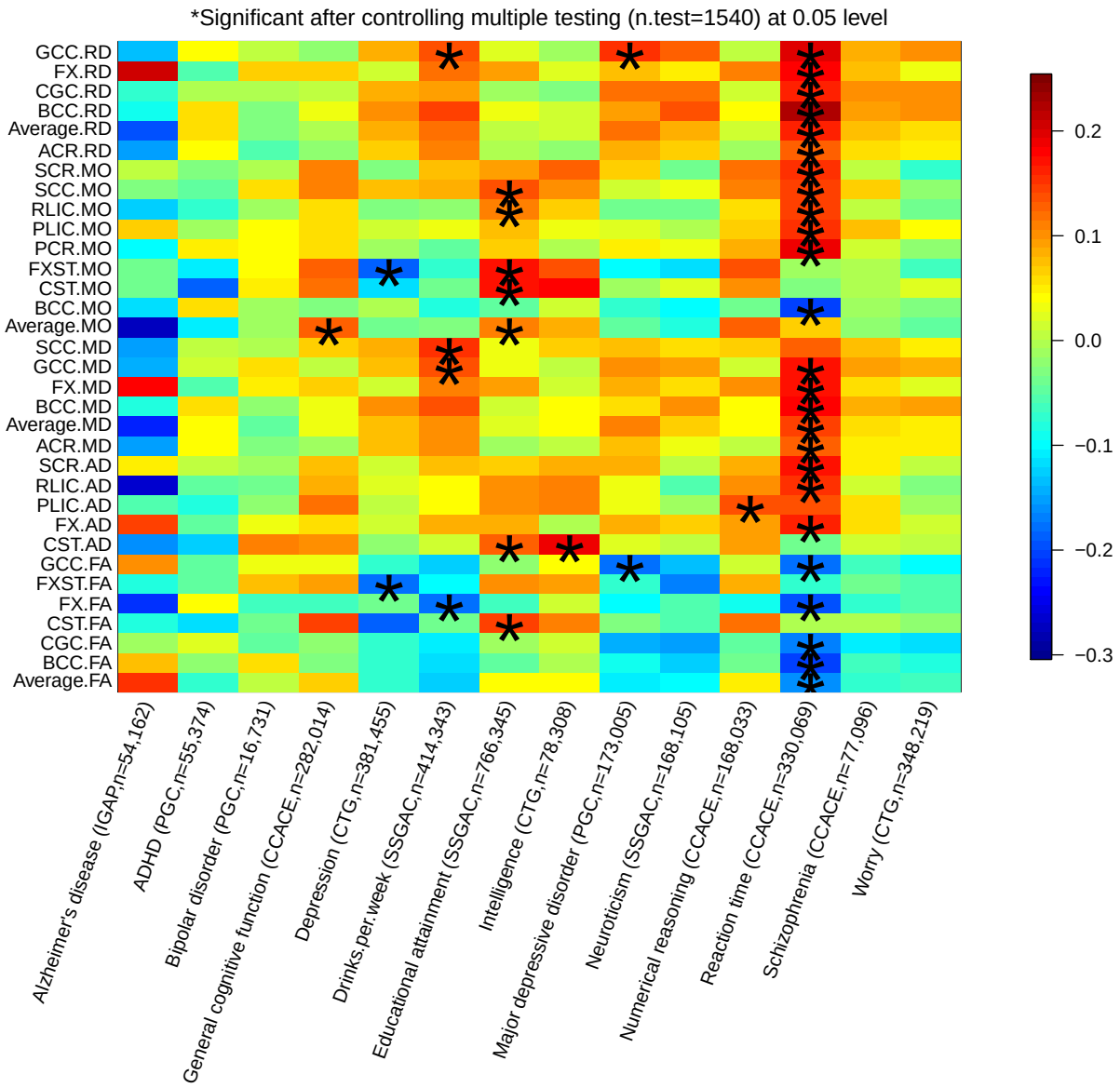
C



(c) SNP heritability grouped by CNS functional annotations.


Outer: P−value<5e−08
0 10 20
Middle: P−value<5e−09
0 10 20
Inner: P−value<4.5e−10
0 10 20


Neuroticism
Neurodegenerative.Diseases
Neuropsychiatric.Disorders
Education
Cognitive
Reaction.time
EIF4G3
MLLT10
AMPD3
VCAN
HBEGF
FBXL7
C16orf95
ARHGAP19−SLIT1
NAV2
ARHGAP27
PLEKHM1
CRHR1
SPPL2C
MAPT
KANSL1
NSF
WNT3
ICA1L
SH3PXD2A
CACNB2
ARL17A
GMNC
GNA12
SUPT3H
RUNX2
CPED1
HIST1H2BN
ZSCAN9
ZKSCAN4
WBP2
SZT2
HYI
NOS1AP
ABI2
NCKAP5
MAST4
ENO4
STX8
SEMA3A
MSRA
XKR6
DCDC1
PRKAG2
NDUFA2
HIST1H3J
EPHA3
EFCAB5
ZNF165
CPNE7
PTCH1
SUN2
Reported genes
Novel genes
No signal


*Significant after controlling multiple testing (n.test=1540) at 0.05 level
GCC.RD
FX.RD
CGC.RD
BCC.RD
Average.RD
ACR.RD
SCR.MO
SCC.MO
RLIC.MO
PLIC.MO
PCR.MO
FXST.MO
CST.MO
BCC.MO
Average.MO
SCC.MD
GCC.MD
FX.MD
BCC.MD
Average.MD
ACR.MD
SCR.AD
RLIC.AD
PLIC.AD
FX.AD
CST.AD
GCC.FA
FXST.FA
FX.FA
CST.FA
CGC.FA
BCC.FA
Average.FA
Alzheimer's disease (IGAP,n=54,162)
ADHD (PGC,n=55,374)
Bipolar disorder (PGC,n=16,731)
General cognitive function (CCACE,n=282,014)
Depression (CTG,n=381,455)
Drinks.per.week (SSGAC,n=414,343)
Educational attainment (SSGAC,n=766,345)
Intelligence (CTG,n=78,308)
Major depressive disorder (PGC,n=173,005)
Neuroticism (SSGAC,n=168,105)
Numerical reasoning (CCACE,n=168,033)
Reaction time (CCACE,n=330,069)
Schizophrenia (CCACE,n=77,096)
Worry (CTG,n=348,219)
0.2
0.1
0.0
−0.1
−0.2
−0.3