The comprehensive genetic architecture of brain white matter

Running title: GWAS of brain white matter

Bingxin Zhao[1], Tengfei Li[2,3], Yue Yang[1], Xifeng Wang[1], Tianyou Luo[1], Yue Shan[1], Ziliang Zhu[1], Di Xiong[1], Yun Li[1,4,5], Jason L. Stein[4,6], and Hongtu Zhu[1,3]*

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

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

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

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

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

[6]UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Corresponding author:

Hongtu Zhu

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

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

List of Pediatric Imaging, Neurocognition and Genetics (PING) authors provided in the supplemental materials.

Abstract

White matter keeps human brain globally connected and shapes communication and connectivity patterns among brain regions. White matter microstructure influences brain structural integrity and may underpin brain functions and disorders. Although under strong genetic control, a large number of genetic variants of white matter remain undiscovered. Here we analyzed the genetic architecture of white matter using diffusion magnetic resonance image (dMRI) of 42,279 individuals (35,101 in the UK Biobank). The dMRIs were consistently processed to generate 215 neuroimaging traits, including 105 measures from tract-specific functional principal component analysis. Genome-wide association analysis identified hundreds of novel independent risk variants ($P < 2.3 \times 10^{-10}$) for white matter microstructural differences. We uncovered 760 pairs of significant genetic correlation between white matter tracts and 60 other complex traits ($P < 2.3 \times 10^{-3}$), including stroke, brain-related disorders (e.g., ADHD, bipolar disorder, major depressive disorder), cognition, neuroticism, chronotype, as well as non-brain traits, such as hypertension, type 2 diabetes, lung function, coronary artery disease, and bone mineral density. Hi-C coupled gene-based analysis identified a large number of pleiotropic genes associated with both white matter and the above complex traits. Gene-set analysis indicated pathways involved in brain disease pathogenesis, developments and abnormalities of neural cells, and repair of white matter damage ($P < 1.5 \times 10^{-8}$). In summary, this large-scale tract-specific study provides a big step forward in understanding the genetics of white matter and its genetic links to other complex traits.

Keywords: Brain White Matter; dMRI; Diffusion Tensor Imaging; GWAS; FPCA; UK Biobank.

Brain functions depend on effective communication among brain regions[1]. White matter comprises roughly half of the human brain and contains most of the brain's long-range communication pathways[2]. White matter tracts build a complex network of structural connections, which keeps the brain globally connected and shapes communication and connectivity patterns[3-5]. Cellular microstructure in white matter tracts influences the integrity of connectivity and mediates signal transitions among distributed brain regions[6]. Evidence from neuroscience has suggested that white matter microstructure may underpin brain functions and dysfunctions[1,7,8], and connectivity differences or changes are relevant to a wide variety of neurological and psychiatric disorders, such as attention-deficit/hyperactivity disorder (ADHD)[9], major depressive disorder (MDD)[10], schizophrenia[11], bipolar disorder[12], multiple sclerosis[13], Alzheimer's disease[14], corticobasal degeneration[15], and Parkinson's disease[16]. White matter microstructural differences and abnormalities can be captured *in vivo* by diffusion magnetic resonance imaging (dMRI). Using dMRI data, microstructural connectivity can be quantified in diffusion tensor imaging (DTI) models[17] and measured by several DTI-derived parameters, including fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), radial diffusivity (RD), and mode of anisotropy (MO). Among them, FA serves as the primary metric of interest in many studies[18], which is a robust global measure of integrity/directionality and is highly sensitive to general connectivity changes. On the other hand, MD, AD, and RD directly quantify the abstract magnitude of directionalities, and thus are more sensitive to specific types of microstructural changes[19]. In addition, MO can describe whether the region of anisotropy is linear anisotropic, orthotropic, or planar anisotropic[20]. See **Supplementary Note** for a global overview of these commonly used DTI parameters.

White matter differences in general population cohorts are under strong genetic control. Both family and population-based studies have reported that DTI parameters can have high heritability with estimates varying across different age groups[21] and tracts[22]. For example, FA in adolescents twins can have heritability larger than 70%[23]. Recent genome-wide association studies (GWAS) of UK Biobank reported an average of 48.7% heritability across different tracts, with 68% in cingulum cingulate gyrus (CGC) and 27% in corticospinal tract (CST)[24]. Several GWAS[22,24-28] have been performed to

identify risk variants for white matter but shared at least two major limitations. The first one is small sample size. For most of dMRI measures, the largest GWAS sample size is 17,706 in Zhao, et al. [24]. Similar to other brain-related traits[29], white matter has a complex and extremely polygenic genetic architecture[24,30]. Large sample size is essential to boost the GWAS power to identify many common risk variants with small or medium effect size. The second one is that previous GWAS mainly focused on global dMRI measures of the whole brain[25,26] or tract-averaged mean values[22,24]. Global and tract-averaged mean measures can capture the largest variations in white matter, while reducing the burden to test multiple neuroimaging traits, particularly suitable for GWAS with limited sample size; however, these measures largely ignore the complicated geometric characteristics of voxel-wise maps in 3D tracts. These limitations result in that a large number of genetic factors of white matter may still be undiscovered. Consequently, with few exceptions (e.g., stroke[25] and cognitive traits[24]), the shared genetic influences between white matter and other complex traits remain unclear.

To overcome these limitations, here we collected individual-level dMRIs from five data resources: the UK Biobank[31], Adolescent Brain Cognitive Development (ABCD[32]), Human Connectome Project (HCP[33]), Pediatric Imaging, Neurocognition, and Genetics (PING[34]), and Philadelphia Neurodevelopmental Cohort (PNC[35]). We consistently processed these images by using the ENIGMA-DTI pipeline[36,37] and obtained voxel-wise DTI maps for 42,279 subjects (after quality controls), including 35,101 in UK Biobank. We mainly focused on 21 predefined white matter tracts and generated two groups of phenotypes. The first group contains 110 tract-averaged mean parameters for FA, AD, MD, MO and RD in 21 tracts and the whole brain. Second, we applied functional principal component analysis (FPCA) to generate 105 tract-specific principal components (PCs) by taking the top five PCs of the voxel-wise FA map within each tract. These 105 tract-specific FA PC parameters are expected to provide additional microstructural details omitted by tract-averaged mean values[38,39], while keeping the multiple testing burden not too heavy. We then performed a comprehensive genome-wide analysis for these 215 phenotypes to discover the genetic architecture of white matter and explore the genetic links to other complex traits in different trait domains. Our GWAS results have been

made publicly available at https://github.com/BIG-S2/GWAS and can be easily browsed through our Brain Imaging Genetics Knowledge Portal https://bigkp.web.unc.edu/.

RESULTS

GWAS Discovery and Validation for 215 DTI parameters.

We mainly used the UKB subjects of British ancestry for GWAS discovery (n = 33,292). All of the 110 DTI mean parameters had significant SNP heritability[40] (h^2) after Bonferroni adjustment (215 tests, $P < 9.4 \times 10^{-31}$, **Fig. 1** and **Supplementary Table 1**). The h^2 estimates varied from 24.8% to 65.4% (mean h^2 = 46.3%), which were comparable with previous results[22,24]. For the 105 tract-specific FA PC parameters, we found that 102 had significant h^2 (mean h^2 = 34.1%, h^2 range = (8.6%, 65.8%), $P < 1.1 \times 10^{-5}$). The 4th PC of corticospinal tract (CST, 6.2%), 5th PC of cingulum hippocampus (CGH, 4.4%), and 4th PC of superior fronto-occipital fasciculus (SFO, 3.7%) had nominally significant h^2 estimates ($P < 0.03$), which became insignificant after Bonferroni adjustment. The top five PCs in external capsule (EC) were highlighted in bottom panels of **Figure 1.** Different from tract-averaged mean, these PCs captured more specific variations in distinct subfields of EC, all of which had high h^2 (mean h^2 = 47.9%, h^2 range = (42.9%, 52.6%), $P < 1.8 \times 10^{-89}$). Another illustration was given in **Supplementary Figure 1** for the PCs of superior longitudinal fasciculus (SLF). These h^2 results show that the additional microstructural variations captured by unconventional tract-specific PC parameters are also generally under genetic control.

We performed linear mixed model-based association analysis for 9,023,710 common genetic variants via fastGWA[41] (Methods). At the stringent significance level 2.3×10^{-10} (i.e., $5 \times 10^{-8}/215$, additionally adjusted for the 215 phenotypes studied), we identified 595 independent (LD $r^2 < 0.6$, Methods) significant variants that had 1,101 significant associations with 86 FA measures (21 mean and 65 PC parameters, **Supplementary Figs. 2-3** and **Supplementary Table 2**). Of the 595 independent significant variants, 302 can only be detected by PC parameters. On average, the number of FA-associated independent significant variants was 37.0 in each tract (range = (4, 72), **Fig. 2** and **Supplementary Table 3**), 50.3% of which were solely discovered by PC parameters (range = (26.3%, 100%)). For example, all of the 22 independent significant variants

associated with CST were detected by PC parameters. Moreover, 66.7% (32/48) of the variants in posterior corona radiata (PCR), 64.9% (37/57) in posterior thalamic radiation (PTR), 59.7% (43/72) in SLF, and 56.3% (18/32) in CGC were only associated with PC parameters. These results clearly illustrate the unique contribution of tract-specific PC parameters in identifying genetic risk variants for white matter.

In addition, 770 independent significant variants were associated with 83 mean parameters of AD, MD, MO and RD (2,069 significant associations), 565 of these 770 variants (with 967 associations) were not identified by FA measures (**Fig. 2**, **Supplementary Figs. 2-3**, and **Supplementary Table 2**). The mean number of independent significant variants in each tract moved up to 93.3 (range = (41, 160)). Of note, more than 70% of independent significant variants in CGH (90.7%), SFO (78.0%), and CGC (73.3%) were detected by non-FA measures (**Supplementary Table 4**). On the other hand, all autosomal chromosomes had risk variants except for chromosome 21 (**Supplementary Table 5**). Interestingly, though only 9.1% of the genome was occupied by chromosomes 5 and 17 (6.5% and 2.6%, respectively), the two chromosomes contributed 47.8% associations of independent significant variants (36.2% and 11.6%, respectively). Previously identified genetic signals of DTI parameters were enriched on chromosomes 5. For example, Rutten-Jacobs, et al. [25] found strong associations for global FA and MD measures in chr5q14 locus, and 502 of the 696 (72.1%) associations identified in Zhao, et al. [24] located on chromosome 5. Our results confirm the enrichment of signals on chromosome 5 and also uncover widespread novel variants across the genome.

Based on LD structures in the 1000 Genomes reference panel[42], independent lead SNPs and genetic loci were defined by FUMA[43] (Methods), and the 3,170 (1,101 + 2,609) significant variant-trait associations can be characterized as 994 significant locus-trait associations (**Supplementary Tables 6-7**). We then performed functionally informed fine mapping for these locus-level signals using SuSiE[44] via PolyFun[45] framework (Methods). PolyFun + SuSiE identified 6,882 variant-trait pairs that had posterior causal probability (i.e., PIP) > 0.95 for 2,299 variants (**Supplementary Table 8**). Frequently fine-mapped variants for these DTI parameters included rs309587, rs2445874, rs3852188, rs3931702,

and rs7733216 in *VCAN*; rs10451283 and rs9897399 in *LINC02210-CRHR1*; rs2770573 and rs9393777 in *LINC00240*; rs242561 and rs62062794 in *MAPT*; and rs199510, rs70600, and rs70602 in *WNT3*, may suggest the existence of multiple causal effects in associated loci. In summary, our results illuminate the broad genetics control on white matter microstructural differences. The genetic effects are spread across a large number of variants, consistent with the observed extremely polygenic genetic architecture of many brain-related traits[29,46].

We aimed to find out-of-sample supports for our discovery GWAS in five independent validation GWAS of European ancestry: the UKB White but Non-British (UKBW, n = 1,809), ABCD European (ABCDE, n = 3,821), HCP (n = 334), PING (n = 461), and PNC (n = 537). First, for each DTI parameter, we checked the genetic correlation (gc) between discovery GWAS and the meta-analyzed European validation GWAS (total n = 6,962) by LDSC[47] (Methods). The mean gc estimates was 0.95 (standard error = 0.35) across the 215 DTI parameters, 121 of which were significant after adjusting for multiple testing by the Benjamini-Hochberg (B-H) procedure at 0.05 level (P < 0.03, **Supplementary Table 9**). Genetic correlation estimates ≈ 1 indicates consistent genetic basis of the same phenotype sampled in different GWAS. Next, we meta-analyzed our discovery GWAS with these European validation GWAS and found that 79.6% significant associations had smaller P-values after meta-analysis, suggesting similar effect size of the top variants in independent cohorts[48,49]. Additionally, we tested for replication by using polygenic risk scores[50] (PRS) derived from discovery GWAS (Methods). After B-H adjustment at 0.05 level (215 × 5 tests), the mean number of significant PRS in the five validation GWAS datasets was 195 (range = (193, 211), P range = (8.5×10^{-27}, 4.5×10^{-2}), **Supplementary Figs. 4-5** and **Supplementary Table 10**). Almost all (214/215) DTI parameters had significant PRS in at least one dataset and 165 had significant PRS in all of them, showing the high generalizability of our discovery GWAS results. Across the five validation datasets, the mean additional variance that can be explained by PRS (i.e., incremental R-squared) was 1.7% (range = (0.4%, 4.2%)) for the 165 consistently significant DTI parameters. The largest mean (incremental) R-squared was on the 2nd PC of EC (range = (2.2%, 6.5%), P range = (7.2×10^{-24}, 1.5×10^{-9})).

Finally, we constructed PRS on two non-European validation GWAS datasets: the ABCD Hispanic (ABCDH, n = 768) and ABCD African American (ABCDA, n = 1,257). The number of significant PRS became 121 and 114 in ABCDH and ABCDA, respectively (B-H adjustment at 0.05 level, **Supplementary Table 11**), which were much smaller than the ones observed in the above European validation GWAS. ABCDH and ABCDE had similar prediction accuracy (mean 0.74% vs. 0.69%, P = 0.28), but the prediction performance was significantly reduced in ABCDA (mean 0.48% vs. 0.69%, $P = 1.9 \times 10^{-7}$). These findings show that UKB British GWAS findings have high generalizability in European cohorts, but the generalizability can be reduced in cross-population applications, highlighting the importance of recruiting sufficient samples from global diverse populations in future genetics discovery of white matter.

Concordance with previous GWAS.

Of the 33,292 subjects in our UKB British discovery GWAS, 17,706 had been used in the largest previous GWAS[24] for 110 mean parameters. To examine the robustness of their findings, we used the other 15,214 individuals (also removed the relatives[51] of previous GWAS subjects) to perform a new validation GWAS and then evaluated the strength of replication (Methods). We calculated the replication slope, which was the correlation of the standardized effect size of variants estimated from two independent GWAS[52]. This analysis was restricted to top ($P < 1 \times 10^{-6}$ in previous GWAS) independent lead variants after LD-based clumping (window size 250, LD r^2 = 0.01). The replication slope was 0.84 (standard error = 0.02, $P < 2 \times 10^{-16}$), indicating strong similarity between these top variant effect size estimates. We also applied FINDOR[52] to reweight P-values by leveraging functional enrichments, after which the replication slope increased to 0.86 (standard error = 0.02, $P < 2 \times 10^{-16}$). In addition, for each of the 110 mean parameters, we checked the LDSC genetic correlation of the values sampled in the two GWAS. The mean gc estimates was 1.03 (standard error = 0.14, **Supplementary Fig. 6** and **Supplementary Table 12**) across these parameters, all of which were significant after B-H adjustment at 0.05 level ($P < 1.4 \times 10^{-5}$). In conclusion, these findings indicate that previous UKB GWAS results can be strongly validated in the new UKB British cohort.

Next, we carried out association lookups for 1,160 (595 + 565) independent significant variants (and variants within LD) detected in our UKB British discovery GWAS (Methods). Of the 213 variants (with 696 associations) identified in Zhao, et al. [24], 202 (with 671 associations) were in LD ($r^2 \geq 0.6$) with our independent significant variants (**Supplementary Table 13**). On the NHGRI-EBI GWAS catalog[53], our results tagged many variants that had been implicated with brain structures, including 7 in van der Meer, et al. [54] for hippocampal subfield volumes, 7 in Verhaaren, et al. [55] for cerebral white matter hyperintensity (WMH) burden, 5 in Vojinovic, et al. [56] for lateral ventricular volume, 5 in Rutten-Jacobs, et al. [25] for WMH and white matter integrity, 2 in Klein, et al. [57] for intracranial volume, 2 in Hibar, et al. [58] for subcortical brain region volumes, 2 in Fornage, et al. [27] for WMH burden, 1 in Elliott, et al. [22] for brain imaging measurements, 1 in Luo, et al. [59] for voxel-wise brain imaging measurement, 1 in Hashimoto, et al. [60] for superior frontal gyrus grey matter volume, 1 in Ikram, et al. [61] for intracranial volume, and 1 in Sprooten, et al. [62] for global FA (**Supplementary Table 14**). When the significance threshold was relaxed to 5×10^{-8}, we tagged variants reported in more previous studies, such as 2 in Shen, et al. [63] for brain imaging measurements, 2 in Chung, et al. [64] for hippocampal volume in dementia, 1 in Chen, et al. [65] for putamen volume, and 1 in Christopher, et al. [66] for posterior cingulate cortex (**Supplementary Table 15**).

Moreover, we found lots of previous associations with other complex traits in different domains (**Supplementary Table 16**). We highlighted 190 variants with psychological traits (e.g., neuroticism[67], well-being spectrum[68], general risk tolerance[69]), 179 with cognitive/educational traits (e.g., cognitive ability[70], educational attainment[71]), 99 with psychiatric disorders (e.g.,, schizophrenia[72], MDD[73], bipolar disorder[74], ADHD[75], autism spectrum disorder[76]), 95 with anthropometric traits (e.g., height[77], body mass index (BMI)[52]), 68 with bone mineral density[78,79], 54 with smoking/drinking (e.g., smoking[80], alcohol use disorder[81]), 20 with neurological disorders (e.g., corticobasal degeneration[82], Parkinson's disease[83], Alzheimer's disease[84], multiple sclerosis[85]), 18 with sleep (e.g., sleep duration[86], chronotype[87]), 11 with glioma (glioblastoma or non-glioblastoma) tumors[88,89], and 6 with stroke[90-92].

To further explore these overlaps, we summarized the number of previously reported variants of other traits that can be tagged by any DTI parameters in each white matter tract (**Supplementary Table 17**). We found that variants associated with psychological, cognitive/educational, smoking/drinking traits and neurological and psychiatric disorders were globally linked to many white matter tracts (**Fig. 3**). For traits in other domains, the overlaps may have some tract-specific patterns. For example, 3 of the 6 variants associated with stroke were linked to both SFO and anterior limb of internal capsule (ALIC), and another 3 were found in superior corona radiata (SCR), anterior corona radiata (ACR), genu of corpus callosum (GCC), body of corpus callosum (BCC), EC, posterior limb of internal capsule (PLIC), and posterior limb of internal capsule (RLIC). In addition, 7 of the 11 risk variants of glioma were associated with splenium of corpus callosum (SCC), 12 of the 18 variants reported for sleep were related to PLIC or inferior fronto-occipital fasciculus (IFO), and 26 of the 68 variants associated with bone mineral density were linked to CST. In addition, more than half of the variants tagged by uncinate fasciculus (UNC) and fornix (FX) had been implicated with anthropometric traits. We carried out voxel-wise association analysis for four representative pleiotropic variants (Methods). The right panel of **Figure 3** illustrated their voxel-wise effect size patterns in spatial maps. rs593720 and rs13198474 had strong effects in corpus callosum (GCC, BCC, and SCC), corona radiata (ACR and SCR), and EX, and the two variants widely tagged psychiatric[93] and neurological[94] disorders, as well as psychological[95] and cognitive/educational[96] traits. On the other hand, rs77126132 highlighted in SCC and BCC was particularly linked to glioma[88], and rs798510 in SCR, FX, and PLIC was associated with several anthropometric traits[97].

An atlas of genetic correlations with other complex traits.

Inspired by massive shared risk variants between DTI parameters and other complex traits, we systematically examined their pairwise genetic correlations by using our discovery GWAS summary statistics and publicly available summary-level data of other 76 complex traits via LDSC (Methods, **Supplementary Table 18**). There were 760 significant pairs between 60 complex traits and 175 DTI parameters after B-H adjustment at 0.05 level (76 × 215 tests, P range = (8.6×10^{-12}, 2.3×10^{-3}), **Supplementary Table 19**), 38.3% (291/760) of which were detected by PC parameters.

We found that DTI parameters were widely correlated with subcortical and WMH volumes (**Supplementary Fig. 7**), brain-related traits (**Supplementary Fig. 8**), and other non-brain traits (**Supplementary Fig. 9**). We replicated previously reported genetic correlations with cognitive/educational traits[24], drinking behavior[24], stroke[22,25], and MDD[24,25], and more tract-specific details were revealed. For example, stroke (any subtypes) and ischemic stroke subtypes[91] (large artery stroke, cardioembolic stroke, and small vessel stroke) showed broad genetic correlations with corpus callosum (GCC and BCC), corona radiata (ACR, SCR, and PCR), limb of internal capsule (PLIC, ALIC), EC, PTR, SLF, SFO, and UNC ($|gc|$ range = (0.17, 0.39), $P < 1.7 \times 10^{-3}$), matching findings in our association lookups. We further observed that small vessel stroke subtype had specific but higher genetic correlations with ALIC and SFO ($|gc|$ range = (0.56, 0.72), $P < 1.7 \times 10^{-3}$). In contrast, there were no significant genetic correlations detected for large artery and cardioembolic stroke, demonstrating the potentially much stronger genetic links between white matter tracts and small vessel stroke subtype.

More importantly, many new genetic correlations were uncovered for brain-related traits, such as Alzheimer's disease, ADHD, bipolar disorder, chronotype, insomnia, neuroticism, and risk tolerance. For example, significant genetic correlation was found between PTR and Alzheimer's disease ($|gc| = 0.32$, $P = 1.2 \times 10^{-3}$), EC and ADHD ($|gc| = 0.18$, $P = 7.1 \times 10^{-5}$), and UNC and bipolar disorder ($|gc| > 0.15$, $P < 7.9 \times 10^{-4}$). These results matched previously reported phenotypical relationships[12,98,99] in neuroscience. We also found novel significant correlations for non-brain traits, including high blood pressure, height, BMI, bone mineral density, number of non-cancer illnesses and treatments, heavy manual or physical work, smoking, coronary artery disease, lung function, and type 2 diabetes (T2D). For example, high blood pressure was genetically correlated with many tracts including SFO, SLF, UNC, EC, and ALIC ($|gc|$ range = (0.10, 0.26), $P < 2.2 \times 10^{-3}$). Previous research found widespread associations between human brain and these traits, such as bone mineral density[100], hypertension[101], T2D[102], lung function[103], heart disease[104], and anthropometric traits[105]. Our findings further illuminate their underlying genetic links. **Figure 4** summaries significant genetic correlations identified in each tract. We found that 33.3% (112/336) of these tract-trait genetic correlations can only be detected by PC parameters (**Supplementary Table 20**).

For example, most of the significant genetic correlations in EC were solely detected by its PC parameters, such as ADHD, BMI, cognitive function, neuroticism, and insomnia. In addition, 12 of the 20 significant genetic correlations with height can only be found by PC parameters.

To validate these results, we performed cross-trait PRS separately on our five European validation GWAS datasets and LDSC on their meta-analyzed summary statistics (n = 6,962, Methods). We found that 681 (89.6%) of these 760 significant pairs can be validated in at least one of the six validation analyses after B-H adjustment at 0.05 level (760 tests, P range = (1.7×10^{-10}, 2.9×10^{-2}), **Supplementary Table 21**), indicating the robustness of our findings. We then reran LDSC after meta-analyzed our UKB British discovery GWAS with these European validation GWAS (n = 40,254). The number of significant pairs increased to 855 (**Supplementary Table 22** and **Supplementary Figs. 10-12**). One example of new findings was the genetic correlation between SLF and schizophrenia[106] (|gc| = 0.11, $P = 2.3 \times 10^{-3}$).

We explored partial genetic causality among these traits using the latent causal variable[107] (LCV) model (Methods). As suggested, we conservatively restricted the LCV analysis to pairs with at least nominally significant genetic correlation ($P < 0.05$), significant evidence of genetic causality (B-H adjustment at 0.01 level, 76 × 215 tests), and large genetic causality proportion estimate (|GCP| > 0.6), which were extremely unlikely to be false positives[107]. We observed that high blood pressure was partially genetically causal for white matter (|GCP| > 0.67, $P < 2.2 \times 10^{-5}$, **Supplementary Fig. 13** and **Supplementary Table 23**). On the other hand, white matter had partially genetically causal effects on insomnia, under sleep, and neuroticism (|GCP| > 0.64, $P < 7.1 \times 10^{-8}$). These findings may lead to plausible biological hypotheses in future research and suggest the existence of different biological mechanisms underlying the atlas of genetic correlations. More efforts are required to explore causal relationships and the shared biological processes[108] among these genetically correlated traits.

Gene-level analysis.

We carried out MAGMA[109] gene-based association analysis for the 215 DTI parameters using our discovery GWAS summary statistics (Methods). There were 3,903 significant gene-level associations ($P < 1.2 \times 10^{-8}$, adjusted for 215 phenotypes) between 620 genes and 179 DTI parameters (**Supplementary Table 24**), 153 of the associated genes can only be discovered by PC parameters. We replicated 99 of 112 MAGMA genes reported in Zhao, et al. [24], 8 white matter-associated genes (*SH3PXD2A, NBEAL1, C1QL1, COL4A2, TRIM47, TRIM65, UNC13D, FBF1*) in Verhaaren, et al. [55], 4 (*VCAN, TRIM47, XRCC4, HAPLN1*) in Rutten-Jacobs, et al. [25], 3 (*ALDH2, PLEKHG1, TRIM65*) in Traylor, et al. [26], 3 (*ALDH2, PLEKHG1, TRIM65*) in Hofer, et al. [110], 2 (*TRIM47, TRIM65*) in Fornage, et al. [27], and 2 (*GNA12, GNA13*) in Sprooten, et al. [111]. Most of the other genes had not been implicated with white matter. Many of our MAGMA genes had been linked to other complex traits (**Supplementary Table 25**), such as 70 genes in Anney, et al. [93] for autism spectrum disorder or schizophrenia, 50 in Morris, et al. [78] for heel bone mineral density, 38 in Hoffmann, et al. [112] for blood pressure variation, 51 in Linnér, et al. [69] for risk tolerance, 36 in Rask-Andersen, et al. [97] for body fat distribution, and 26 in Hill, et al. [113] for neuroticism.

Next, we mapped significant variants ($P < 2.3 \times 10^{-10}$) to genes according to physical position, expression quantitative trait loci (eQTL) association, and 3D chromatin (Hi-C) interaction via FUMA[43] (Methods). FUMA yielded 1,189 new associated genes (1,630 in total) that were not discovered in MAGMA analysis (**Supplementary Table 26**), replicating 286 of the 292 FUMA genes identified in Zhao, et al. [24] and more other genes in previous studies of white matter, such as *PDCD11*[55], *ACOX1*[55], *CLDN23*[110], *EFEMP1*[25,26,55], and *IRS2*[110]. More overlapped genes were also observed between white matter and other traits (**Supplementary Table 27).** Particularly, 876 FUMA genes were solely mapped by significant Hi-C interactions in brain tissues (**Supplementary Table 28**), demonstrating the power of integrating chromatin interaction profiles in GWAS of white matter.

We then explored the gene-level pleiotropy between white matter and 79 complex traits, including nine neurological and psychiatric disorders[114] studied in Sey, et al. [114] and (other) traits studied in our genetic correlation analysis. For brain-related traits, the

associated genes were predicted by the recently developed Hi-C-coupled MAGMA[114] (H-MAGMA) tool (Methods). Traditional MAGMA[109] was used for non-brain GWAS. H-MAGMA prioritized 737 significant genes for white matter ($P < 6.3 \times 10^{-9}$, adjusted for 215 phenotypes and two brain tissue types, **Supplementary Table 29**), and we focused on 329 genes that can be replicated in our meta-analyzed European validation GWAS (n = 6,962) at nominal significance level ($P < 0.05$, **Supplementary Table 30**). We found that 298 of these 329 genes were associated with at least one of 57 complex traits (**Supplementary Table 31**). **Figure 5** and **Supplementary Table 32** display the number of overlapped genes between 57 complex traits and 21 white matter tracts. Most white matter tracts have many pleiotropic genes with other complex traits, aligning with patterns in association lookups and genetic correlation analysis. For example, schizophrenia had 80 overlapped genes with SLF, 71 with CGC, 68 with EC, and 65 with SCR. Global white matter changes in schizophrenia patients had been observed[106,115,116]. Particularly, 230 white matter H-MAGMA genes had been identified in Sey, et al. [114] for nine neurological and psychiatric disorders (**Supplementary Table 33**). *NSF*[117], *GFAP*[118], *TRIM27*[72], *HLA-DRA*[117,119], and *KANSL1*[76,95] were associated with five of these disorders, and another 69 genes were linked to at least three different disorders (**Supplementary Fig. 14**). In summary, our analysis largely expands the overview of gene-level pleiotropy, informing the shared genetic influences between white matter and other complex traits.

Biological annotations.

To gain more insights into biological mechanisms, we performed several analyses to explore biological interpretations of our discovery GWAS results. First, MAGMA gene-property[109] analysis was performed for 13 GTEx[120] (v8) brain tissues to examine whether the tissue-specific gene expression levels were related to significance between genes and DTI parameters (Methods). After Bonferroni adjustment (13 × 215 tests), we detected 57 significant associations for gene expression in brain cerebellar hemisphere and cerebellum tissues ($P < 1.8 \times 10^{-5}$, **Supplementary Fig. 15** and **Supplementary Table 34**), suggesting that genes with higher transcription levels on white matter-presented regions also had stronger genetic associations with DTI parameters. In contrast, no signals were observed on regions primarily dominated by grey matter, such as basal ganglia and cortex. We also performed chromatin-based annotation analysis by

stratified LDSC[121] for 490 cell-type/tissue-specific DNase I hypersensitivity and activating histone marker annotations (Methods). After Bonferroni adjustment (490 × 215 tests), nine brain tissue histone annotations had significantly enriched contributions to per-SNP heritability for seven tracts ($P < 4.5 \times 10^{-7}$, **Supplementary Table 35**).

MAGMA[109] competitive gene-set analysis was performed for 15,496 gene sets (5,500 curated gene sets and 9,996 GO terms, Methods). We found 180 significant gene sets after Bonferroni adjustment (15,496 × 215 tests, $P < 1.5 \times 10^{-8}$, **Supplementary Table 36**). The top five frequently prioritized gene sets were “dacosta uv response via ercc3 dn” (M4500), “dacosta uv response via ercc3 common dn” (M13522), “graessmann apoptosis by doxorubicin dn” (M1105), “gobert oligodendrocyte differentiation dn” (M2369), and “blalock alzheimers disease up” (M12921). M4500 and M13522 are *ERCC3*-associated gene sets related to xeroderma pigmentosum (XP) and trichothiodystrophy (TTD) syndromes, which are genetic disorders caused by a defective nucleotide excision repair system[122,123]. In addition to skin symptoms, patients of XP and TTD often reported various neurological deteriorations and white matter abnormalities, such as intellectual impairment[124], myelin structures degradation[125], and diffuse dysmyelination[126]. M1105 regulates the apoptosis of breast cancer cells in response to doxorubicin treatment. Clinical research found that breast cancer chemotherapy like doxorubicin was neurotoxic[127] and can cause therapy-induced brain structural changes and decline in white matter integrity[128]. M2369 plays a critical role in oligodendrocyte differentiation, which mediates the repair of white matter after damaging events[129], and M12921 is related to the pathogenesis of Alzheimer's disease[130].

Several gene sets of rat sarcoma (Ras) proteins, small GTPases, and rho family GTPases were also prioritized by MAGMA, such as “go regulation of small gtpase mediated signal transduction” (GO: 0051056), “go small gtpase mediated signal transduction” (GO: 0007264), “go re gelation of ras protein signal transduction” (GO: 0046578), “go ras protein signal transduction” (GO: 0007265), and “reactome signaling by rho gtpases” (M501). Ras proteins activity is involved in developmental processes and abnormalities of neural cells in central nervous system[131,132]; small and rho family GTPases play crucial roles in basic cellular processes during the entire neurodevelopment process and are

closely connected to several neurological disorders[133-135]. We also observed significant enrichment in neuron-related pathways, including "go neurogenesis" (GO: 0022008), "go neuron differentiation" (GO: 0030182), "go neuron development" (GO: 0048666), "go regulation of neuron differentiation" (GO: 0045664), and "go regulation of nervous system development" (GO: 0051960). Finally, we applied DEPICT[136] gene-set enrichment testing for 10,968 pre-constituted gene sets (Methods), 7 of which survived Bonferroni adjustment (10,968 × 215 tests, $P < 2.1 \times 10^{-8}$), such as two gene sets involved in Ras proteins and small GTPases (GO: 0046578 and GO: 0005083) and another two for vasculature and blood vessel developments (GO: 0001944 and GO: 0001568, **Supplementary Table 37**). More MAGMA enriched gene sets can also be detected by DEPICT when the significance threshold was relaxed to 6.5×10^{-6} (i.e., not adjusted for testing 215 phenotypes), such as GO: 0051960, GO: 0045664, GO: 0007264, and GO: 0051056. In summary, our results provide many insights into the underlying biological processes of white matter, suggesting that DTI measures could be useful in understanding the shared pathophysiological pathways between white matter and multiple diseases and disorders.

DISCUSSION

In this study, we analyzed the genetic architecture of brain white matter using dMRI scans of 42,279 subjects collected from five publicly accessible data resources. Through a genome-wide analysis, we identified hundreds of previously unknown variants and genes for white matter microstructural differences. Many previously reported genetic hits were confirmed in our discovery GWAS, and we further validated our discovery GWAS in a few replication cohorts. We evaluated the genetic relationships between white matter and a wide variety of complex traits in association lookups, genetic correlation estimation, and gene-level analysis. A large proportion of our findings were revealed by unconventional tract-specific PC parameters. Bioinformatics analyses found tissue-specific functional enrichments and lots of enriched biological pathways. Together, these results suggest the value of large-scale neuroimaging data integration and the application of tract-specific FPCA in studying the genetics of human brain.

One limitation of the present study is that the majority of publicly available dMRI data are from subjects of European ancestry and our discovery GWAS focused on UKB British individuals. Such GWAS strategy can efficiently avoid false discoveries due to population stratifications and heterogeneities across studies[22,137], but may raise the question that to what degree the research findings can be generalized and applied on global populations[138,139]. In our analysis, we found that the UKB British-derived PRS were still widely significant in Hispanic and African American testing cohorts but had reduced performances, especially in African American cohort. This may indicate that the genetic architecture of white matter is similar but not the same across different populations. Identifying the cross-population and population-specific components of genetic factors for human brain could be an interesting future topic. As more non-European neuroimaging data become available (e.g., the ongoing CHIMGEN project[140] in Chinese population), global integration efforts are needed to study the comparative genetic architectures and to explore the multi-ethnic genetics relationships among brain and other human complex traits.

URLs.

Brain Imaging GWAS Summary Statistics, https://github.com/BIG-S2/GWAS;

Brain Imaging Genetics Knowledge Portal, https://bigkp.web.unc.edu/;

UKB Imaging Pipeline, https://git.fmrib.ox.ac.uk/falmagro/UK_biobank_pipeline_v_1;

ENIGMA-DTI Pipeline, http://enigma.ini.usc.edu/protocols/dti-protocols/;

PLINK, https://www.cog-genomics.org/plink2/;

GCTA & fastGWA, http://cnsgenomics.com/software/gcta/;

METAL, https://genome.sph.umich.edu/wiki/METAL;

Michigan Imputation Server, https://imputationserver.sph.umich.edu/;

FUMA, http://fuma.ctglab.nl/;

MGAMA, https://ctg.cncr.nl/software/magma;

H-MAGMA, https://github.com/thewonlab/H-MAGMA;

LDSC, https://github.com/bulik/ldsc/;

LCV, https://github.com/lukejoconnor/LCV/;

DEPICT, https://github.com/perslab/depict;

FINDOR, https://github.com/gkichaev/FINDOR;

SuSiE, https://github.com/stephenslab/susieR;

PolyFun, https://github.com/omerwe/polyfun;

NHGRI-EBI GWAS Catalog, https://www.ebi.ac.uk/gwas/home;

The atlas of GWAS Summary Statistics, http://atlas.ctglab.nl/;

METHODS

Methods are available in the ***Methods*** section.

Note: One supplementary information pdf file and one supplementary zip file are available.

ACKNOWLEDGEMENTS

This research was partially supported by U.S. NIH grants MH086633 (H.Z.) and MH116527 (TF.L.). We thank Sophia Cui, Xiaopeng Zong, and Peter Vandehaar for helpful conversations. We thank the individuals represented in the UK Biobank, ABCD, HCP, PING, and PNC studies for their participation and the research teams for their work in collecting, processing and disseminating these datasets for analysis. We gratefully acknowledge all the studies and databases that made GWAS summary data available. This research has been conducted using the UK Biobank resource (application number 22783), subject to a data transfer agreement. Part of the data collection and sharing for this project was funded by the Pediatric Imaging, Neurocognition and Genetics Study (PING) (U.S. National Institutes of Health Grant RC2DA029475). PING is funded by the National Institute on Drug Abuse and the Eunice Kennedy Shriver National Institute of Child Health & Human Development. PING data are disseminated by the PING Coordinating Center at the Center for Human Development, University of California, San Diego. Support for the collection of the PNC datasets was provided by grant RC2MH089983 awarded to Raquel Gur and RC2MH089924 awarded to Hakon Hakonarson. All PNC subjects were recruited through the Center for Applied Genomics at The Children's Hospital in Philadelphia. Part of the data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National

Institutes of Health and additional federal partners under award numbers U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/scientists/workgroups/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. HCP data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

AUTHOR CONTRIBUTIONS

B.Z., H.Z., Y.L., and J.S. designed the study. B.Z., TF. L, Y.Y., X.W., and TY. L analyzed the data. TF. L, Y.S., Z.Z., Y.Y., X.W., TY. L, and D.X., downloaded the datasets, preprocessed MRI data, and undertook the quantity controls. B.Z. and H.Z. wrote the manuscript with feedback from all authors.

COMPETETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

REFERENCES

1. van den Heuvel, M.P. & Sporns, O. A cross-disorder connectome landscape of brain dysconnectivity. *Nature reviews neuroscience* **20**, 435-446 (2019).
2. Hagmann, P. *et al.* Mapping the structural core of human cerebral cortex. *PLoS biology* **6**(2008).

3. Zielinski, B.A., Gennatas, E.D., Zhou, J. & Seeley, W.W. Network-level structural covariance in the developing brain. *Proceedings of the National Academy of Sciences* **107**, 18191-18196 (2010).
4. Fields, R.D. White matter in learning, cognition and psychiatric disorders. *Trends in neurosciences* **31**, 361-370 (2008).
5. Filley, C.M. & Fields, R.D. White matter and cognition: making the connection. *Journal of neurophysiology* **116**, 2093-2104 (2016).
6. Kuceyeski, A., Maruta, J., Relkin, N. & Raj, A. The Network Modification (NeMo) Tool: elucidating the effect of white matter integrity changes on cortical and subcortical structural connectivity. *Brain connectivity* **3**, 451-463 (2013).
7. Bathelt, J., Scerif, G., Nobre, A. & Astle, D. Whole-brain white matter organization, intelligence, and educational attainment. *Trends in neuroscience and education* **15**, 38-47 (2019).
8. Vaquero, L., Rodríguez-Fornells, A. & Reiterer, S.M. The left, the better: white-matter brain integrity predicts foreign language imitation ability. *Cerebral Cortex* **27**, 3906-3917 (2017).
9. Wu, Z.-M. *et al.* White matter microstructural alterations in children with ADHD: categorical and dimensional perspectives. *Neuropsychopharmacology* **42**, 572-580 (2017).
10. Zou, K. *et al.* Alterations of white matter integrity in adults with major depressive disorder: a magnetic resonance imaging study. *Journal of psychiatry & neuroscience: JPN* **33**, 525 (2008).
11. Cetin-Karayumak, S. *et al.* White matter abnormalities across the lifespan of schizophrenia: a harmonized multi-site diffusion MRI study. *Molecular psychiatry*, 1-12 (2019).
12. Versace, A. *et al.* Elevated left and reduced right orbitomedial prefrontal fractional anisotropy in adults with bipolar disorder revealed by tract-based spatial statistics. *Archives of general psychiatry* **65**, 1041-1052 (2008).
13. De Santis, S. *et al.* Evidence of early microstructural white matter abnormalities in multiple sclerosis from multi-shell diffusion MRI. *NeuroImage: Clinical* **22**, 101699 (2019).

14. Lee, S. *et al.* White matter hyperintensities are a core feature of Alzheimer's disease: evidence from the dominantly inherited Alzheimer network. *Annals of neurology* **79**, 929-939 (2016).
15. Hess, C., Christine, C., Apple, A., Dillon, W. & Aminoff, M. Changes in the thalamus in atypical parkinsonism detected using shape analysis and diffusion tensor imaging. *American Journal of Neuroradiology* **35**, 897-903 (2014).
16. Veselý, B. & Rektor, I. The contribution of white matter lesions (WML) to Parkinson's disease cognitive impairment symptoms: a critical review of the literature. *Parkinsonism & related disorders* **22**, S166-S170 (2016).
17. Basser, P.J., Mattiello, J. & LeBihan, D. Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B* **103**, 247-254 (1994).
18. Grieve, S., Williams, L., Paul, R., Clark, C. & Gordon, E. Cognitive aging, executive function, and fractional anisotropy: a diffusion tensor MR imaging study. *American Journal of Neuroradiology* **28**, 226-235 (2007).
19. Basser, P. & Pierpaoli, C. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of magnetic resonance. Series B* **111**, 209 (1996).
20. Ennis, D.B. & Kindlmann, G. Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* **55**, 136-146 (2006).
21. Vuoksimaa, E. *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* **38**, 2026-2036 (2017).
22. Elliott, L.T. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216 (2018).
23. Chiang, M.-C. *et al.* Genetics of white matter development: a DTI study of 705 twins and their siblings aged 12 to 29. *Neuroimage* **54**, 2308-2317 (2011).
24. Zhao, B. *et al.* Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n = 17,706). *Molecular Psychiatry* (2019).

25. Rutten-Jacobs, L.C. *et al.* Genetic study of white matter integrity in UK Biobank (N= 8448) and the overlap with stroke, depression, and dementia. *Stroke* **49**, 1340-1347 (2018).
26. Traylor, M. *et al.* Genetic variation in PLEKHG1 is associated with white matter hyperintensities (n= 11,226). *Neurology* **92**, e749-e757 (2019).
27. Fornage, M. *et al.* Genome-wide association studies of cerebral white matter lesion burden: the CHARGE consortium. *Annals of neurology* **69**, 928-939 (2011).
28. Adib-Samii, P. *et al.* 17q25 Locus is associated with white matter hyperintensity volume in ischemic stroke, but not with lacunar stroke status. *Stroke* **44**, 1609-1615 (2013).
29. O'Connor, L.J. *et al.* Extreme polygenicity of complex traits is explained by negative selection. *The American Journal of Human Genetics* **105**, 456-476 (2019).
30. Adib-Samii, P. *et al.* Genetic architecture of white matter hyperintensities differs in hypertensive and nonhypertensive ischemic stroke. *Stroke* **46**, 348-353 (2015).
31. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine* **12**, e1001779 (2015).
32. Casey, B. *et al.* The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. *Developmental cognitive neuroscience* **32**, 43-54 (2018).
33. Somerville, L.H. *et al.* The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5–21 year olds. *NeuroImage* **183**, 456-468 (2018).
34. Jernigan, T.L. *et al.* The pediatric imaging, neurocognition, and genetics (PING) data repository. *Neuroimage* **124**, 1149-1154 (2016).
35. Satterthwaite, T.D. *et al.* Neuroimaging of the Philadelphia neurodevelopmental cohort. *Neuroimage* **86**, 544-553 (2014).
36. Jahanshad, N. *et al.* Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: A pilot project of the ENIGMA–DTI working group. *Neuroimage* **81**, 455-469 (2013).

37. Kochunov, P. *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* **95**, 136-150 (2014).
38. Mattingsdal, M. *et al.* Pathway analysis of genetic markers associated with a functional MRI faces paradigm implicates polymorphisms in calcium responsive pathways. *Neuroimage* **70**, 143-149 (2013).
39. Azadeh, S. *et al.* Integrative Bayesian analysis of neuroimaging-genetic data through hierarchical dimension reduction. in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* 824-828 (IEEE, 2016).
40. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**, 76-82 (2011).
41. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of large-scale data. *Nature genetics* **51**, 1749 (2019).
42. Consortium, G.P. A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
43. Watanabe, K., Taskesen, E., Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826 (2017).
44. Wang, G., Sarkar, A.K., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine-mapping. *bioRxiv*, 501114 (2019).
45. Weissbrod, O. *et al.* Functionally-informed fine-mapping and polygenic localization of complex trait heritability. *BioRxiv*, 807792 (2019).
46. Martin, A.R., Daly, M.J., Robinson, E.B., Hyman, S.E. & Neale, B.M. Predicting polygenic risk of psychiatric disorders. *Biological psychiatry* **86**, 97-109 (2019).
47. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**, 1236-1241 (2015).
48. Jansen, P.R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nature Genetics* **51**, 394-403 (2019).

49. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genetics* **38**, 209-213 (2006).
50. Jansen, I.E. *et al.* Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature genetics* **51**, 404-413 (2019).
51. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
52. Kichaev, G. *et al.* Leveraging polygenic functional enrichment to improve GWAS power. *The American Journal of Human Genetics* **104**, 65-75 (2019).
53. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research* **47**, D1005-D1012 (2018).
54. van der Meer, D. *et al.* Brain scans from 21,297 individuals reveal the genetic architecture of hippocampal subfield volumes. *Molecular Psychiatry*, in press. (2018).
55. Verhaaren, B.F. *et al.* Multiethnic genome-wide association study of cerebral white matter hyperintensities on MRI. *Circulation: Cardiovascular Genetics* **8**, 398-409 (2015).
56. Vojinovic, D. *et al.* Genome-wide association study of 23,500 individuals identifies 7 loci associated with brain ventricular volume. *Nature communications* **9**, 1-11 (2018).
57. Klein, M. *et al.* Genetic markers of ADHD-related variations in intracranial volume. *American Journal of Psychiatry* **176**, 228-238 (2019).
58. Hibar, D.P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224-229 (2015).
59. Luo, Q. *et al.* Association of a Schizophrenia-Risk Nonsynonymous Variant With Putamen Volume in Adolescents: A Voxelwise and Genome-Wide Association Study. *JAMA psychiatry* **76**, 435-445 (2019).
60. Hashimoto, R. *et al.* Common variants at 1p36 are associated with superior frontal gyrus volume. *Translational psychiatry* **4**, e472-e472 (2014).

61. Ikram, M.A. *et al.* Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nature Genetics* **44**, 539-544 (2012).
62. Sprooten, E. *et al.* Common genetic variants and gene expression associated with white matter microstructure in the human brain. *Neuroimage* **97**, 252-261 (2014).
63. Shen, L. *et al.* Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort. *Neuroimage* **53**, 1051-1063 (2010).
64. Chung, J. *et al.* Genome-wide association study of Alzheimer's disease endophenotypes at prediagnosis stages. *Alzheimer's & Dementia* **14**, 623-633 (2018).
65. Chen, C.-H. *et al.* Leveraging genome characteristics to improve gene discovery for putamen subcortical brain structure. *Scientific Reports* **7**, 15736 (2017).
66. Christopher, L. *et al.* A variant in PPP4R3A protects against alzheimer-related metabolic decline. *Annals of neurology* **82**, 900-911 (2017).
67. Luciano, M. *et al.* Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nature Genetics* **50**, 6-11 (2018).
68. Baselmans, B.M. *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nature genetics* **51**, 445-451 (2019).
69. Linnér, R.K. *et al.* Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics* **51**, 245-257 (2019).
70. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications* **9**, 2098 (2018).
71. Lee, J.J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* **50**, 1112–1121 (2018).
72. Ikeda, M. *et al.* Genome-wide association study detected novel susceptibility genes for schizophrenia and shared trans-populations/diseases genetic effect. *Schizophrenia bulletin* **45**, 824-834 (2019).

73. Hyde, C.L. *et al.* Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature genetics* **48**, 1031 (2016).
74. Ruderfer, D.M. *et al.* Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Molecular psychiatry* **19**, 1017 (2014).
75. Lesch, K.-P. *et al.* Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *Journal of neural transmission* **115**, 1573-1585 (2008).
76. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics* **51**, 431 (2019).
77. Allen, H.L. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832-838 (2010).
78. Morris, J.A. *et al.* An atlas of genetic influences on osteoporosis in humans and mice. *Nature genetics* **51**, 258-266 (2019).
79. Medina-Gomez, C. *et al.* Life-course genome-wide association study meta-analysis of total body BMD and assessment of age-specific effects. *The American Journal of Human Genetics* **102**, 88-102 (2018).
80. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature genetics* **51**, 237-244 (2019).
81. Sanchez-Roige, S. *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*, appi. ajp. 2018.18040369 (2018).
82. Kouri, N. *et al.* Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. *Nature communications* **6**, 7247 (2015).
83. Simon-Sanchez, J. *et al.* Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature genetics* **41**, 1308 (2009).
84. Kunkle, B.W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature genetics* **51**, 414 (2019).
85. Beecham, A.H. *et al.* Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature genetics* **45**, 1353 (2013).

86. Dashti, H. *et al.* GWAS in 446,118 European adults identifies 78 genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *bioRxiv*, 274977 (2018).
87. Jones, S.E. *et al.* Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nature communications* **10**, 343 (2019).
88. Melin, B.S. *et al.* Genome-wide association study of glioma subtypes identifies specific differences in genetic susceptibility to glioblastoma and non-glioblastoma tumors. *Nature genetics* **49**, 789 (2017).
89. Kinnersley, B. *et al.* Genome-wide association study identifies multiple susceptibility loci for glioma. *Nature communications* **6**, 1-9 (2015).
90. Traylor, M. *et al.* Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. *The Lancet Neurology* **11**, 951-962 (2012).
91. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature genetics* **50**, 524-537 (2018).
92. Dichgans, M. *et al.* Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke* **45**, 24-36 (2014).
93. Anney, R.J.L. *et al.* Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism* **8**, 21 (2017).
94. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nature genetics* **49**, 1511 (2017).
95. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nature Genetics* **50**, 920 (2018).
96. Lam, M. *et al.* Large-Scale Cognitive GWAS Meta-Analysis Reveals Tissue-Specific Neural Expression and Potential Nootropic Drug Targets. *Cell reports* **21**, 2597-2613 (2017).

97. Rask-Andersen, M., Karlsson, T., Ek, W.E. & Johansson, Å. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nature communications* **10**, 339 (2019).

98. Onnink, A.M.H. *et al.* Deviant white matter structure in adults with attention-deficit/hyperactivity disorder points to aberrant myelination and affects neuropsychological performance. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **63**, 14-22 (2015).

99. Wen, Q. *et al.* White matter alterations in early-stage Alzheimer's disease: A tract-specific study. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **11**, 576-587 (2019).

100. Minn, Y., Suk, S. & Do, S. Osteoporosis as an independent risk factor for silent brain infarction and white matter changes in men and women: the PRESENT project. *Osteoporosis International* **25**, 2465-2469 (2014).

101. Hannawi, Y. *et al.* Hypertension is associated with white matter disruption in apparently healthy middle-aged individuals. *American Journal of Neuroradiology* **39**, 2243-2248 (2018).

102. Tan, X. *et al.* Micro-structural white matter abnormalities in type 2 diabetic patients: a DTI study using TBSS analysis. *Neuroradiology* **58**, 1209-1216 (2016).

103. Dodd, J.W. *et al.* Brain structure and function in chronic obstructive pulmonary disease: a multimodal cranial magnetic resonance imaging study. *American journal of respiratory and critical care medicine* **186**, 240-245 (2012).

104. Berry, C. *et al.* Small-vessel disease in the heart and brain: current knowledge, unmet therapeutic need, and future directions. *Journal of the American Heart Association* **8**, e011104 (2019).

105. Xu, J., Li, Y., Lin, H., Sinha, R. & Potenza, M.N. Body mass index correlates negatively with white matter integrity in the fornix and corpus callosum: a diffusion tensor imaging study. *Human brain mapping* **34**, 1044-1052 (2013).

106. Seok, J.-H. *et al.* White matter abnormalities associated with auditory hallucinations in schizophrenia: a combined study of voxel-based analyses of diffusion tensor imaging and structural magnetic resonance imaging. *Psychiatry Research: Neuroimaging* **156**, 93-104 (2007).

107. O'Connor, L.J. & Price, A.L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nature genetics* **50**, 1728-1734 (2018).
108. Cortes, A., Albers, P.K., Dendrou, C.A., Fugger, L. & McVean, G. Identifying cross-disease components of genetic risk across hospital data in the UK Biobank. *Nature Genetics* **52**, 126-134 (2020).
109. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Computational Biology* **11**, e1004219 (2015).
110. Hofer, E. *et al.* White matter lesion progression: genome-wide search for genetic influences. *Stroke* **46**, 3048-3057 (2015).
111. Sprooten, E. *et al.* White matter integrity as an intermediate phenotype: exploratory genome-wide association analysis in individuals at high risk of bipolar disorder. *Psychiatry Research* **206**, 223-231 (2013).
112. Hoffmann, T.J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nature genetics* **49**, 54 (2017).
113. Hill, W.D. *et al.* Genetic contributions to two special factors of neuroticism are associated with affluence, higher intelligence, better health, and longer life. *Molecular psychiatry*, 1-19 (2019).
114. Sey, N.Y. *et al.* A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. (Nature Publishing Group, 2020).
115. Kelly, S. *et al.* Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group. *Molecular psychiatry* **23**, 1261 (2018).
116. Stämpfli, P. *et al.* Subtle white matter alterations in schizophrenia identified with a new measure of fiber density. *Scientific reports* **9**, 1-11 (2019).
117. Hamza, T.H. *et al.* Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. *Nature genetics* **42**, 781 (2010).
118. Hol, E. *et al.* Neuronal expression of GFAP in patients with Alzheimer pathology and identification of novel GFAP splice forms. *Molecular psychiatry* **8**, 786-796 (2003).

119. Jakkula, E. *et al.* Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. *The American Journal of Human Genetics* **86**, 285-291 (2010).
120. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *BioRxiv*, 787903 (2019).
121. Finucane, H.K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature Genetics* **50**, 621-629 (2018).
122. Uribe-Bojanini, E., Hernandez-Quiceno, S. & Cock-Rada, A.M. Xeroderma Pigmentosum with Severe Neurological Manifestations/De Sanctis–Cacchione Syndrome and a Novel XPC Mutation. *Case reports in medicine* **2017**(2017).
123. Kraemer, K.H. *et al.* Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype–phenotype relationship. *Neuroscience* **145**, 1388-1396 (2007).
124. Anttinen, A. *et al.* Neurological symptoms and natural course of xeroderma pigmentosum. *Brain* **131**, 1979-1989 (2008).
125. Kassubek, J. *et al.* The cerebro-morphological fingerprint of a progeroid syndrome: white matter changes correlate with neurological symptoms in xeroderma pigmentosum. *PloS one* **7**(2012).
126. Harreld, J., Smith, E., Prose, N., Puri, P. & Barboriak, D. Trichothiodystrophy with dysmyelination and central osteosclerosis. *American journal of neuroradiology* **31**, 129-130 (2010).
127. Menning, S. *et al.* Changes in brain white matter integrity after systemic treatment for breast cancer: a prospective longitudinal study. *Brain imaging and behavior* **12**, 324-334 (2018).
128. Deprez, S., Billiet, T., Sunaert, S. & Leemans, A. Diffusion tensor MRI of chemotherapy-induced cognitive impairment in non-CNS cancer patients: a review. *Brain imaging and behavior* **7**, 409-435 (2013).
129. Hamanaka, G., Ohtomo, R., Takase, H., Lok, J. & Arai, K. White-matter repair: Interaction between oligodendrocytes and the neurovascular unit. *Brain circulation* **4**, 118 (2018).

130. Blalock, E.M. *et al.* Incipient Alzheimer's disease: microarray correlation analyses reveal major transcriptional and tumor suppressor responses. *Proceedings of the National Academy of Sciences* **101**, 2173-2178 (2004).
131. Koini, M., Rombouts, S., Veer, I., Van Buchem, M. & Huijbregts, S. White matter microstructure of patients with neurofibromatosis type 1 and its relation to inhibitory control. *Brain imaging and behavior* **11**, 1731-1740 (2017).
132. Kang, M. & Lee, Y.-S. The impact of RASopathy-associated mutations on CNS development in mice and humans. *Molecular brain* **12**, 1-17 (2019).
133. Shikanai, M., Yuzaki, M. & Kawauchi, T. Rab family small GTPases-mediated regulation of intracellular logistics in neural development. *Histology and Histopathology* **33**, 765-771 (2017).
134. Qu, L. *et al.* The Ras Superfamily of Small GTPases in Non-neoplastic Cerebral Diseases. *Frontiers in Molecular Neuroscience* **12**, 121 (2019).
135. Govek, E.-E., Newey, S.E. & Van Aelst, L. The role of the Rho GTPases in neuronal development. *Genes & development* **19**, 1-49 (2005).
136. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nature Communications* **6**, 5890 (2015).
137. Smith, S.M. & Nichols, T.E. Statistical challenges in “big data” human neuroimaging. *Neuron* **97**, 263-268 (2018).
138. Martin, A.R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics* **51**, 584 (2019).
139. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in diverse human populations. *Nature communications* **10**, 1-9 (2019).
140. Xu, Q. *et al.* CHIMGEN: a Chinese imaging genetics cohort to enhance cross-ethnic and cross-geographic brain research. *Molecular Psychiatry*, 1-13 (2019).
141. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575 (2007).
142. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).

143. Consortium, I.H. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-58 (2010).
144. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
145. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527 (2016).
146. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).
147. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
148. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011).

METHODS

GWAS design and Imaging phenotypes. We performed the following GWAS separately: 1) the UKB British discovery GWAS, which used data of individuals of British ancestry[51] from the UKB study (n = 33,292); 2) five validation GWAS performed on individuals of European ancestry, including UKB White but Non-British (UKBW, n = 1,809), ABCD European (ABCDE, n = 3,821), HCP (n = 334), PING (n = 461), and PNC (n = 537); 3) two non-European validation GWAS, including ABCD Hispanic (ABCDH, n = 768) and ABCD African American (ABCDA, n = 1,257); and 4) a UKB British GWAS with subjects not present in previous GWAS[24] (also removed the relatives of previous GWAS subjects, n = 15,214). See **Supplementary Table 38** for a summary of these GWAS and demographic information of study cohorts. The raw dMRI, covariates and genetic data were downloaded from each data resource. We processed the dMRI data locally using consistent procedures via ENIGMA-DTI pipeline[36,37] to generate 215 mean and PC DTI phenotypes for 21 predefined white matter tracts (**Supplementary Table 39**). A full description of image acquisition and preprocessing, quality controls, ENIGMA-DTI pipeline, white matter tracts, principle component extraction, and formulas of DTI parameters are detailed in **Supplementary Note**. An overview of imaging procedures is shown in **Supplementary Figs. 16-17** and a few image examples are given in

Supplementary Figs. 18-21. For each continuous phenotype or covariate variable, we removed values greater than five times the median absolute deviation from the median value.

Association discovery and validation. Genotyping and quality controls are documented in **Supplementary Note**. We estimated the SNP heritability by all autosomal SNPs in UKB British discovery GWAS data using GCTA-GREML analysis[40]. The adjusted covariates included age (at imaging), age-squared, sex, age-sex interaction, age-squared-sex interaction, imaging site, as well as the top 40 genetic principle components (PCs) provided by UKB[51] (Data-Field 22009). The heritability estimates were tested in one-sided likelihood ratio tests. We performed linear mixed model-based association analysis using fastGWA[41]. The same set of covariates as in GCTA-GREML analysis were adjusted. To replicate previous findings, we also performed another UKB British GWAS with subjects not present in previous GWAS[24]. In addition, GWAS were separately performed on UKBW, ABCDE, HCP, PING, PNC, ABCDA, and ABCDH datasets using Plink[141]. In the seven validation GWAS, we adjusted for age, age-squared, sex, age-sex interaction, age-squared-sex interaction, and top ten genetic PCs estimated from genetic variants. We also adjusted for imaging sites in ABCD GWAS (ABCDA, ABCDH, and ABCDE).

We applied a few analyses to support the findings in UKB British discovery GWAS. First, the LDSC[47] software (version 1.0.0) was then used to estimate the pairwise genetic correlation between DTI parameter values in discovery GWAS and the meta-analyzed five European validation GWAS (n = 6,962). The meta-analysis was performed using METAL[142] with the sample-size weighted approach. We used the pre-calculated LD scores provided by LDSC, which were computed using 1000 Genomes European data. We used HapMap3[143] variants and removed all variants in the major histocompatibility complex (MHC) region. In addition, we also performed another meta-analysis for the UKB British discovery GWAS and the five European validation GWAS to check whether the *P*-values became smaller after combining these results. Next, polygenic risk scores (PRS) were created on the seven validation datasets using the BLUP effect sizes estimated from GCTA-GREML analysis of UKB British discovery GWAS. We used PLINK to

generate risk scores in each testing data by summarizing across genome-wide variants, weighed by their BLUP effect sizes. We tried 17 *P*-value thresholds for variant selection using their marginal *P*-values from fastGWAS: 1, 0.8, 0.5, 0.4, 0.3, 0.2, 0.1, 0.08, 0.05, 0.02, 0.01, 1×10^{-3}, 1×10^{-4}, 1×10^{-5}, 1×10^{-6}, 1×10^{-7}, and 1×10^{-8}. Then, we generated 17 polygenic profiles for each phenotype and reported the best prediction power that can be achieved by a single profile. The association between polygenic profile and phenotype was estimated and tested in linear models, adjusting for the effects of covariates used in the corresponding validation GWAS. The additional phenotypic variation that can be explained by polygenic profile (i.e., the incremental R-squared) was used to measure the prediction accuracy.

Genomic risk loci characterization and comparison with previous findings. We defined genomic risk loci by using FUMA (version 1.3.5e). We input the UKB British discovery GWAS summary statistics after reweighting the *P*-values using functional information via FINDOR[52]. Specifically, FUMA first identified independent significant variants, which were variants with a *P*-value smaller than the predefined threshold and independent of other significant variants (LD $r^2 < 0.6$). FUMA constructed LD blocks for these independent significant variants by tagging all variants in LD ($r^2 \geq 0.6$) with at least one independent significant variant and had a MAF ≥ 0.0005. These variants included those from the 1000 Genomes reference panel that may not have been included in the GWAS. Based on these independent significant variants, lead variants were identified as those that were independent from each other (LD $r^2 < 0.1$). If LD blocks of independent significant variants were closed (<250 kb based on the closest boundary variants of LD blocks), they were merged to a single genomic locus. Thus, each genomic risk locus could contain more than one independent significant variants and lead variants. We performed functionally-informed fine-mapping by using SuSiE[44] method via PolyFun[45] framework for risk loci. The summary statistics from UKB British discovery GWAS were used as input. As suggested, we estimated the LD matrix using our training GWAS individuals. To validate previous findings reported in Zhao, et al. [24], we estimated the pairwise genetic correlation between DTI parameter values in previous GWAS and the UKB British GWAS with subjects not included in previous GWAS. We also estimated the replication slope[52] between two groups of standardized effect sizes. We focused on

previously reported top ($P < 1 \times 10^{-6}$) independent SNPs after LD-based clumping (window size 250, LD $r^2 = 0.01$). Independent significant variants and all their tagged variants were searched by FUMA in the NHGRI-EBI GWAS catalog (version 2019-09-24) to look for previously reported associations ($P < 9 \times 10^{-6}$) with any traits. In our UKB British discovery GWAS data, we performed voxel-wise association analysis to illustrate spatial maps for several selected pleiotropic variants. The same set of covariates used in the above tract-based GWAS analysis were adjusted in this voxel-wise analysis.

Genetic correlation estimation and validation. We used LDSC to estimate the pairwise genetic correlation between DTI parameters and other complex traits. The summary statistics of DTI parameters were from the UKB British discovery GWAS and the summary statistics of other traits were collected from publicly accessible data resources listed in **Supplementary Table 18**. To replicate the significant associations, we reran LDSC using the meta-analyzed summary statistics from the five European validation GWAS. In addition, we also constructed PRS for other complex traits on each of the five validation datasets and tested whether the PRS had significant association with DTI parameters. We used the LD-based pruning (window size 50, step 5, LD $r^2 = 0.2$) procedure to account for the LD structure in this cross-trait PRS analysis. We also applied the 17 GWAS *P*-value thresholds for variants selection and reported the smallest *P*-value observed in validation data. We applied the LCV[107] (version 2019-03-14) to explore the genetical causal relationships between DTI parameters and other complex traits. We used the summary statistics from the UKB British discovery GWAS and the pre-calculated LD scores provided by LDSC.

Gene-level analysis. We first performed gene-based association analysis in UKB British discovery GWAS for 18,796 protein-coding genes using MAGMA[109] (version 1.07). Default MAGMA settings were used with zero window size around each gene. We then carried out FUMA functional annotation and mapping analysis, in which variants were annotated with their biological functionality and then were linked to 35,808 candidate genes by a combination of positional, eQTL, and 3D chromatin interaction mappings. We chose brain-related tissues/cells in all options and used default values for all other parameters. For the detected genes in MAGMA and FUMA, we performed lookups in

the NHGRI-EBI GWAS catalog (version 2020-02-08) again to explore their previously reported associations. We also applied H-MAGMA[114] (version 2019-11-29) to perform Hi-C coupled gene-based association analysis by integrating Hi-C profiles from fetal and adult brain tissues[144,145].

Biological annotations. We performed gene property analysis for the 13 GTEx[120] v8 brain tissues via MAGMA. Specifically, we tested whether the tissue-specific gene expression levels can be linked to the strength of the gene-trait association. We also performed heritability enrichment analysis via stratified LDSC[121] for cell-type/tissue-specific annotations of DNase I hypersensitivity and activating histone[146,147]. MAGMA and DEPICT (version 1 rel194) were separately used to explore the implicated biological pathways. MAGMA gene-set analysis examined 5,500 curated gene sets and 9,996 Gene Ontology (GO) terms from the Molecular Signatures Database[148] (MSigDB, version 7.0) and DEPICT tested 10,968 pre-constructd gene sets using GWAS summary statistics with *P*-value < 10^{-5} as input. All other parameters were set as default.

Data availability

Our GWAS summary statistics have been shared at https://github.com/BIG-S2/GWAS. The individual-level raw data used in this study can be obtained from five publicly accessible data resources: UK Biobank (http://www.ukbiobank.ac.uk/resources/), ABCD (https://abcdstudy.org/), PING (https://www.chd.ucsd.edu/research/ping-study.html), PNC (https://www.med.upenn.edu/bbl/philadelphianeurodevelopmentalcohort.html), and HCP (https://www.humanconnectome.org/).

Code availability

We made use of publicly available software and tools listed in URLs. All codes used to generate our results are available upon reasonable request.

Figure legends

Figure 1. SNP heritability estimates of 215 DTI parameters (n=33,292 subjects) and illustration of the top five FA principal components (PCs) of external capsule (EC).

a) The SNP heritability was estimated as the proportion of variation explained by all autosomal genetic variants using GCTA-GREML analysis[40] after adjusting the effects of age (at imaging), age-squared, sex, age-sex interaction, age-squared-sex interaction, as well as the top 40 genetic principle components (PCs). The 110 mean DTI parameters and 105 FA PC DTI parameters (n=33,292 subjects) are displayed on the left and right panels, respectively. The *x*-axis lists the names of white matter tracts. b) The functional principal component (PC) coefficients for the top five FA PCs of EC.

Figure 2. Annotation of white matter tracts and the number of independent significant variants identified in UKB British discovery GWAS at 2.3×10^{-10} significance level (n=33,292 subjects).

a) Annotation of the 21 white matter tracts in human brain; b) The first three columns are the number of independent significant variants identified in each white matter tract by 1) any DTI parameters; 2) any FA parameters; 3) FA PC parameters, respectively. The fourth column in b) displays the proportion of FA-associated variants that can only be identified by PC parameters. We performed linear mixed model-based association analysis using fastGWA[41].

Figure 3. Number of previously reported GWAS variants for other complex traits that are associated with white matter tracts and the spatial map of voxel-wise effect size patterns for four selected pleiotropic variants (n=33,292 subjects).

The *y*-axis lists the 21 white matter tracts. The *x*-axis provides the name of different trait domains. The displayed numbers are the number of pleiotropic variants that have been linked to other traits. Details of these overlaps can be found in **Supplementary Table 16.**

Figure 4. Significant pairwise genetic correlations between white matter tracts and other complex traits (n=33,292 subjects).

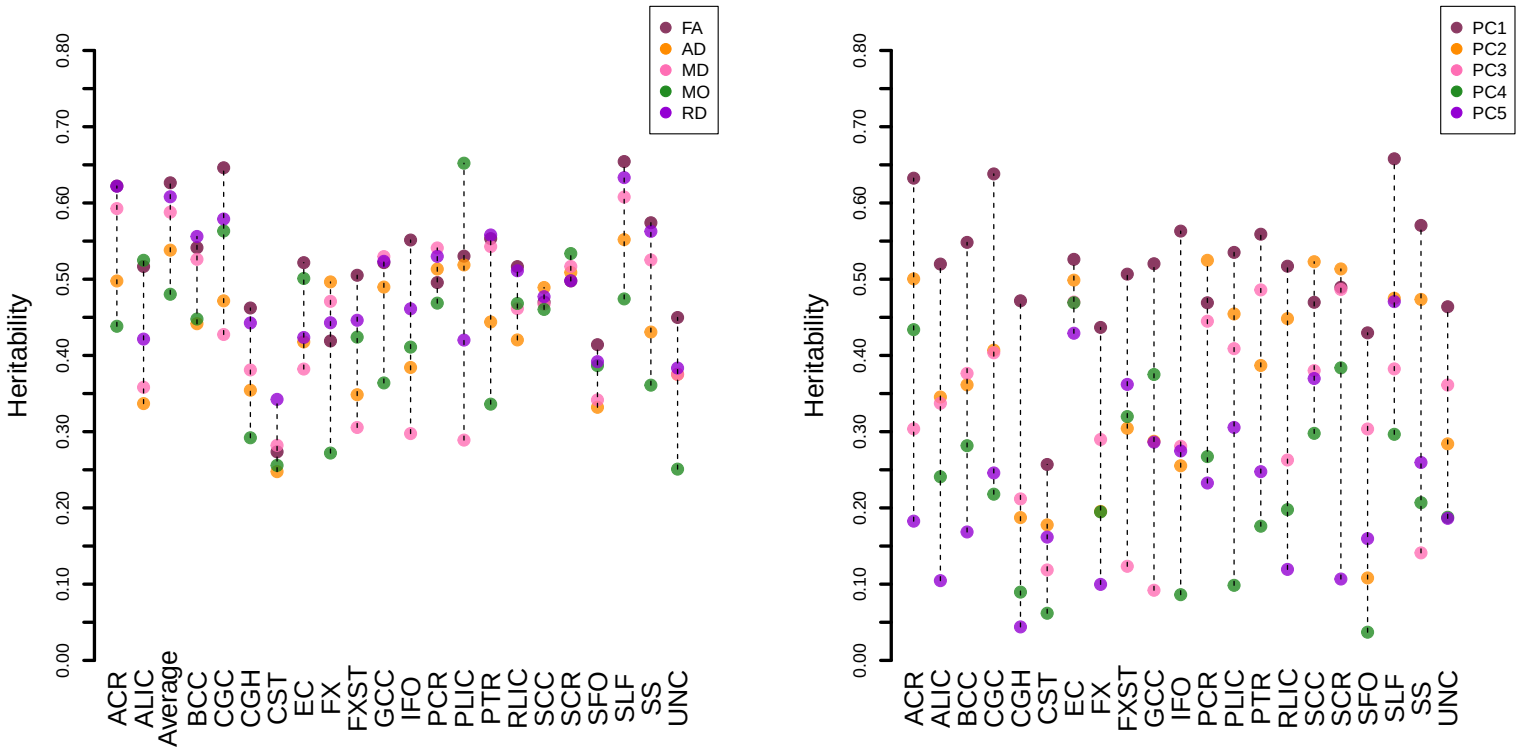
The pairwise genetic correlations were estimated and tested by LDSC[47]. We adjusted for multiple testing by the Benjamini-Hochberg procedure at 0.05 significance level (215 × 76 tests). The *y*-axis lists the 21 white matter tracts. The *x*-axis provides the name of

other traits, whose sample size and detailed information can be found in **Supplementary Table 18.** The pairs were divided into three groups and labeled with three different colors: “Mean only” (orange) represent pairs solely identified by mean DTI parameters; “PC only” (red) are pairs uniquely detected by FA PC DTI parameters; and “Both” (blue) correspond to pairs that can be found by both mean and PC parameters.

Figure 5. Number of overlapped genes between 57 complex traits and 21 white matter tracts in gene-based association analysis (n=33,292 subjects).

The *y*-axis lists the 21 white matter tracts. The *x*-axis provides the name of other traits, whose sample size and detailed information can be found in **Supplementary Table 18**. The displayed numbers are the number of overlapped genes between each tract-trait pair in gene-based association analysis. For DTI parameters and brain-related complex traits, the association analysis was performed using Hi-C coupled MAGMA analysis; and we performed traditional MAGMA analysis for non-brain traits.

a)



b)

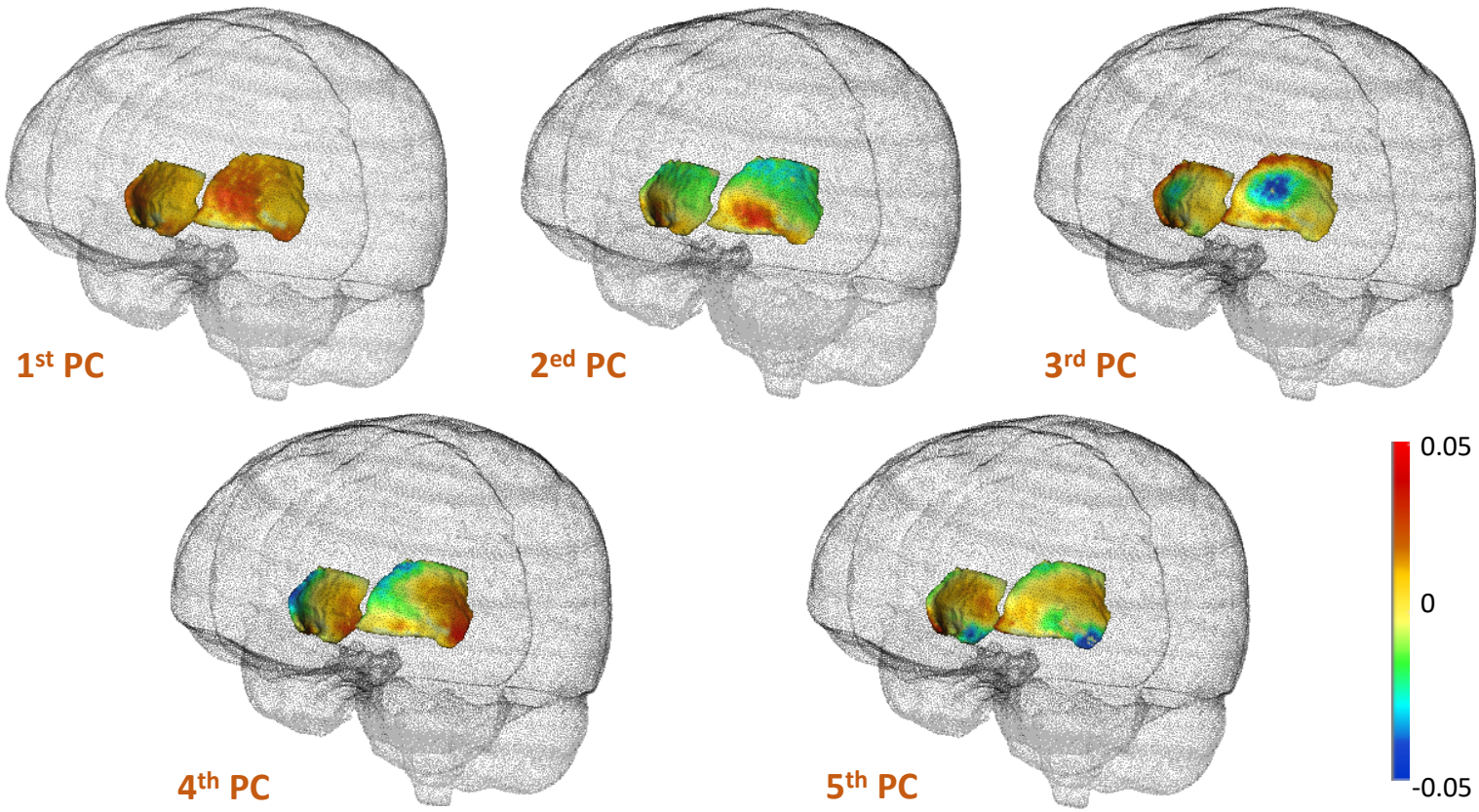
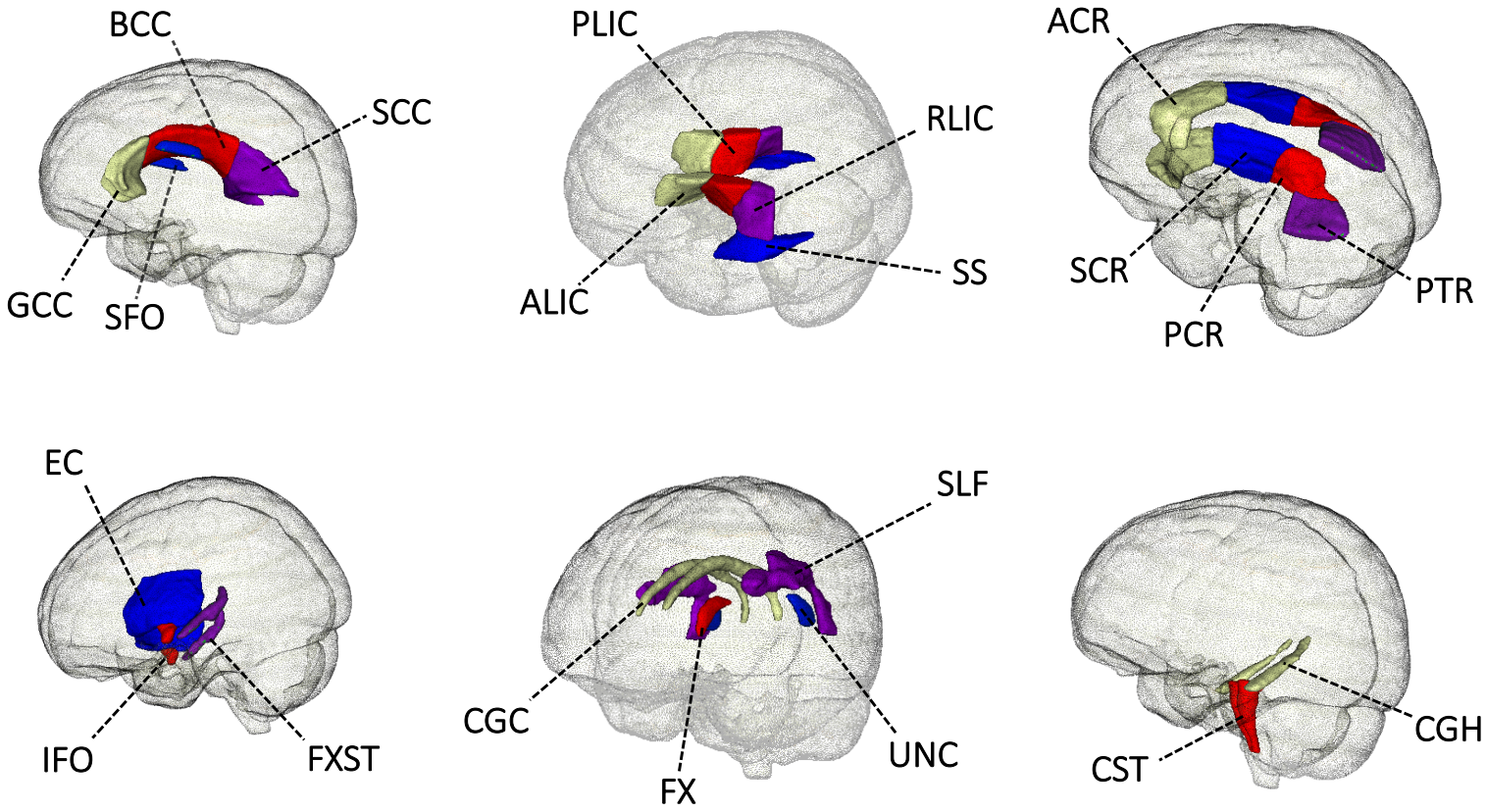


Figure 1

a)



b)

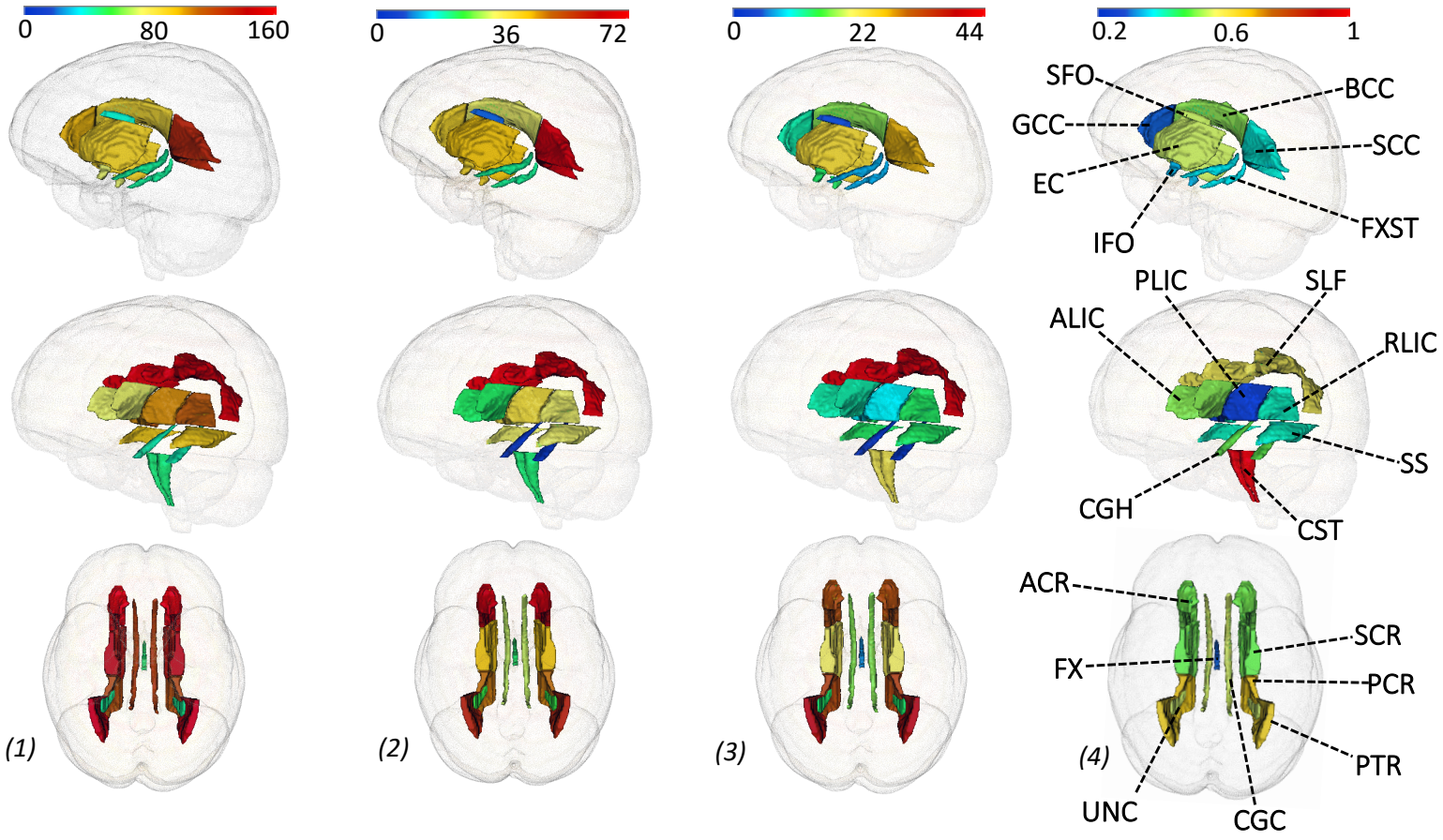


Figure 2

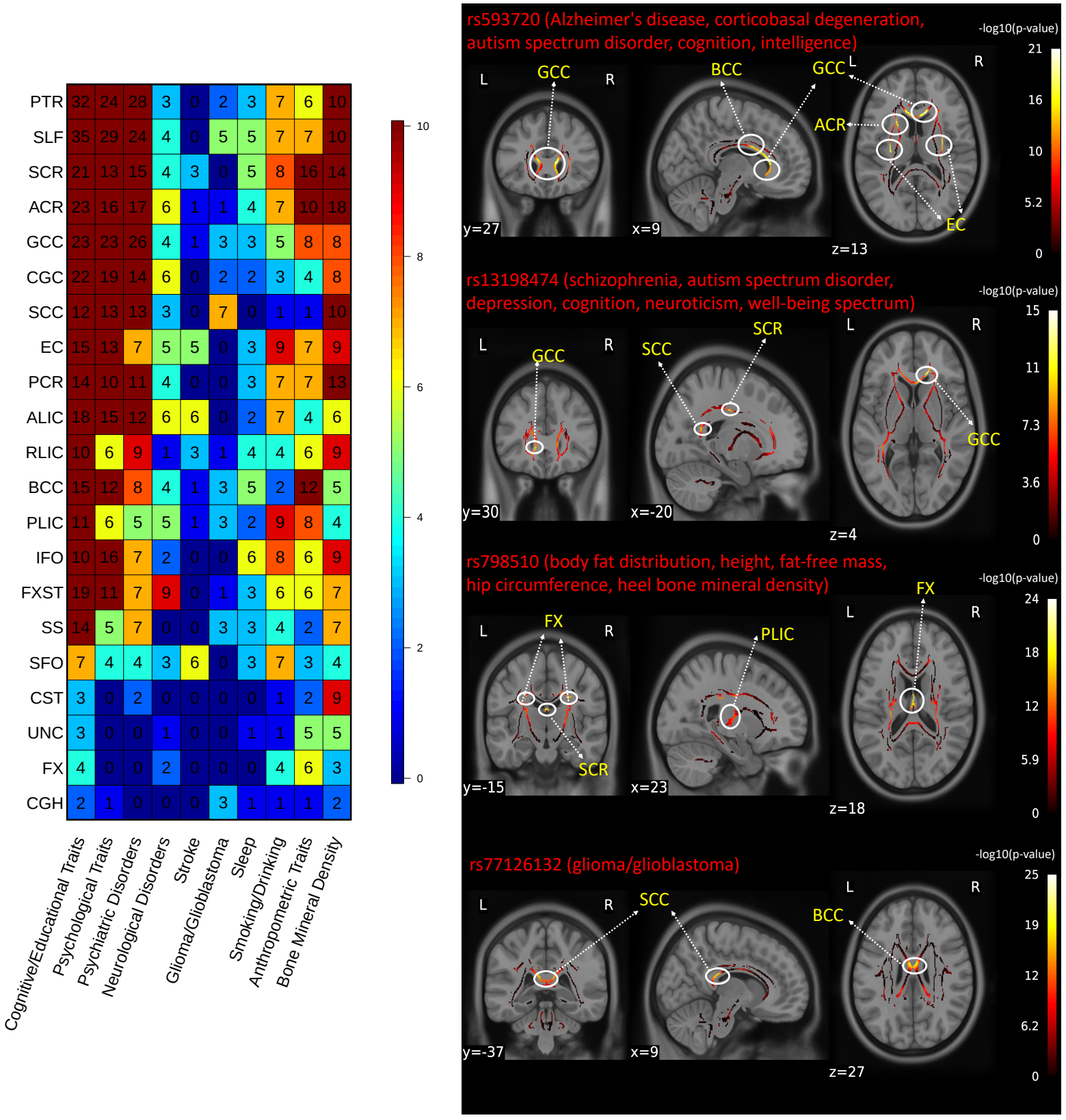


Figure 3

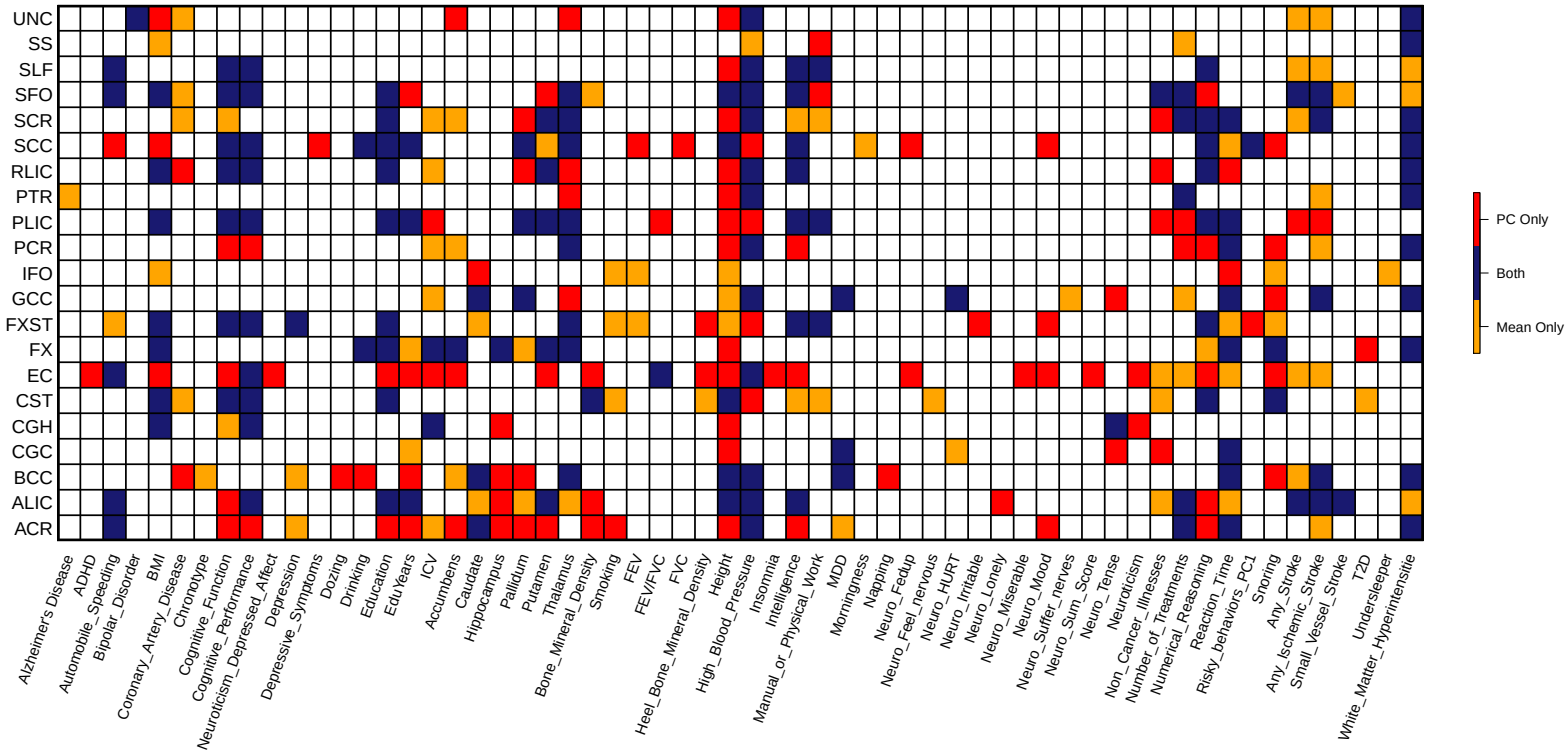


Figure 4

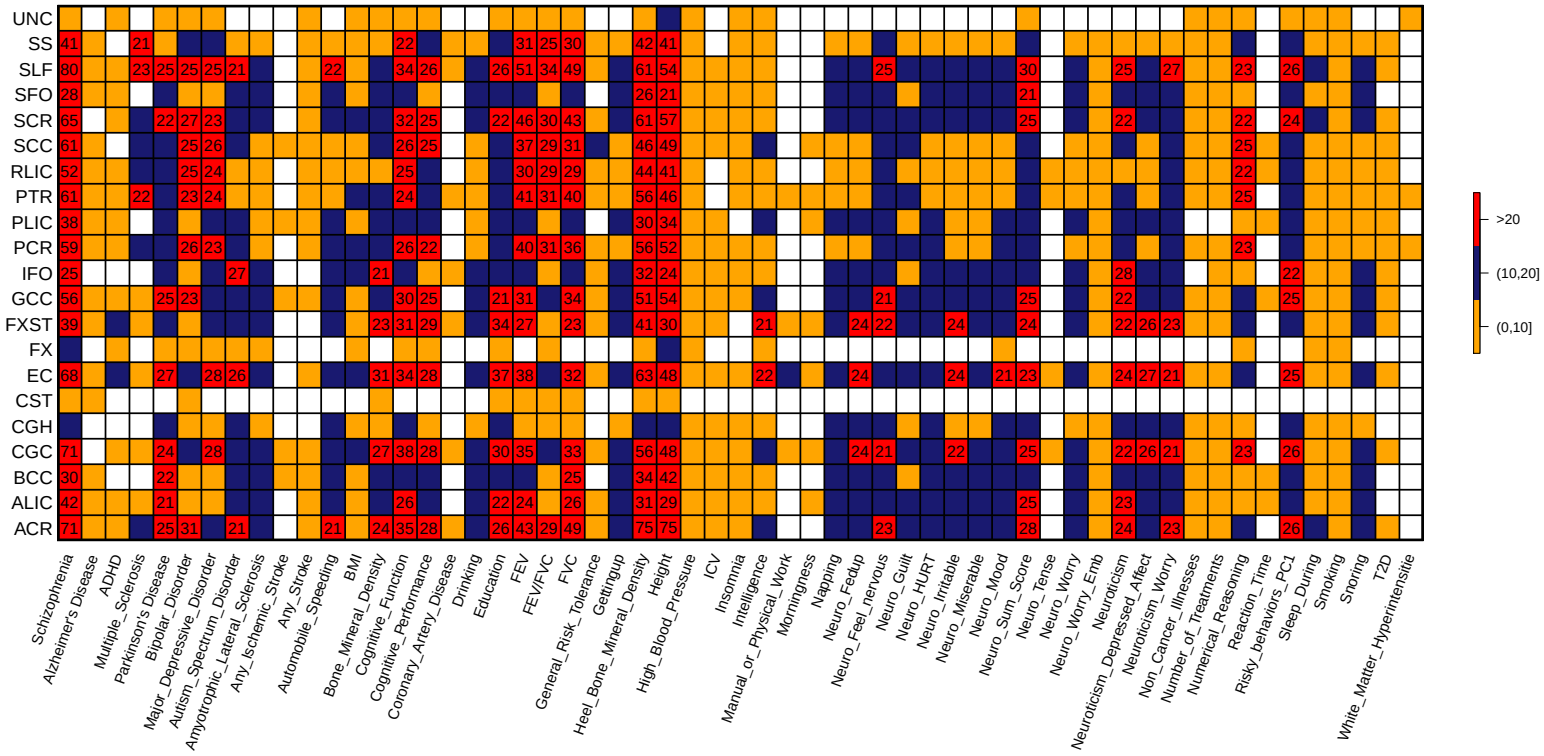


Figure 5