

Common genetic variation influencing human white matter microstructure

Running title: GWAS of brain white matter

Bingxin Zhao¹, Tengfei Li^{2,3}, Yue Yang¹, Xifeng Wang¹, Tianyou Luo¹, Yue Shan¹, Ziliang Zhu¹, Di Xiong¹, Mads E. Hauberg^{4,5,7,8}, Jaroslav Bendl⁴⁻⁶, John F. Fullard⁴⁻⁶, Panagiotis Roussos^{4-6,9}, Yun Li^{1,10,11}, Jason L. Stein^{10,12}, and Hongtu Zhu^{1,3*}

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

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

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

⁴Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁵Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁶Department of Genetics and Genomic Science and Institute for Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

⁷iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Denmark

⁸Centre for Integrative Sequencing (iSEQ), Aarhus University, Aarhus, Denmark

⁹Mental Illness Research, Education, and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, NY, USA

¹⁰Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

¹¹Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

¹²UNC 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

Brain regions communicate with each other via tracts of myelinated axons, commonly referred to as white matter. White matter microstructure can be measured in the living human brain using diffusion based magnetic resonance imaging (dMRI), and has been found to be altered in patients with neuropsychiatric disorders. Although under strong genetic control, few genetic variants influencing white matter microstructure have ever been identified. Here we identified common genetic variants influencing white matter microstructure using dMRI in 42,919 individuals (35,741 in the UK Biobank). The dMRIs were summarized into 215 white matter microstructure traits, including 105 measures from tract-specific functional principal component analysis. Genome-wide association analysis identified many novel white matter microstructure associated loci ($P < 2.3 \times 10^{-10}$). We identified shared genetic influences through genetic correlations between white matter tracts and 62 other complex traits, including stroke, neuropsychiatric disorders (e.g., ADHD, bipolar disorder, major depressive disorder, schizophrenia), cognition, neuroticism, chronotype, as well as non-brain traits. Common variants associated with white matter microstructure alter the function of regulatory elements in glial cells, particularly oligodendrocytes. White matter associated genes were enriched in pathways involved in brain disease pathogenesis, neurodevelopment process, 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 genetic architecture of white matter and its genetic links to a wide spectrum of clinical outcomes.

Keywords: White Matter Microstructure; dMRI; Diffusion Tensor Imaging; GWAS; Functional Principal Component Analysis; UK Biobank.

Brain functions depend on effective communication across brain regions¹. White matter comprises roughly half of the human brain and contains most of the brain's long-range communication pathways². White matter tracts build a complex network of structural connections, which keeps the brain globally connected and shapes communication and connectivity patterns³⁻⁵. Cellular microstructure in white matter tracts plays a pivotal role in maintaining the integrity of connectivity and mediating signal transitions among distributed brain regions⁶. Evidence from neuroscience has further suggested that white matter microstructure may underpin brain function and dysfunction^{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, schizophrenia¹¹, bipolar disorder¹², multiple sclerosis¹³, Alzheimer's disease¹⁴, corticobasal degeneration¹⁵, and Parkinson's disease¹⁶. 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) models¹⁷ 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 studies¹⁸, 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 changes¹⁹. In addition, MO can characterize the anisotropy type, describing whether the shape of the diffusion tensor is more linear or planar^{20,21}. 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 measurements of white matter microstructure have in general high heritability with estimates varying across different age groups²² and tracts²³. For example, heritability estimates of tract-averaged FA ranged from 53% to 90% in twin study of the Human Connectome Project (HCP)²⁴. Recent genome-wide association studies (GWAS) of UK Biobank reported an average SNP-based heritability of 48.7% across different tracts²⁵.

Several GWAS^{23,25-29} have been performed to identify loci associated with inter-individual variation in white matter microstructure but shared at least two major limitations: (i) sample size and (ii) spatial specificity. First, the current largest published GWAS of dMRI phenotypes has sample size 17,706 in Zhao, et al.²⁵. Similar to other brain-related traits³⁰, white matter has a complex and extremely polygenic genetic architecture^{25,31}. Large sample size is essential to boost GWAS power in order to identify many common risk variants with small effect sizes. Second, previous GWAS mainly focused on global dMRI measures of the whole brain^{26,27} or tract-averaged (mean) values^{23,25}. Global and tract-averaged 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 may lose lots of information, as microstructural differences and changes may not have a uniformly consistent pattern across the whole tract. Heterogeneous variation patterns typically exist within voxel-wise DTI maps of the 3D tract curve, which may be more relevant to specific underlying biological processes. For example, previous study found that the association between bipolar disorder and FA is specific to one given segment of the long anterior limb of internal capsule (ALIC) tract connecting prefrontal cortex with the thalamus and brain stem³². Due to these limitations, a large number of genetic factors influencing white matter may still be undiscovered. Consequently, with few exceptions (e.g., stroke²⁶ and cognitive traits²⁵), the shared genetic influences between white matter and other complex traits are unknown. Uncovering these potential genetic links may identify important brain regions that are involved in clinical outcomes, especially for brain-related disorders.

To overcome these limitations, here we collected individual-level dMRI from five data resources: the UK Biobank³³, Adolescent Brain Cognitive Development (ABCD³⁴), HCP³⁵, Pediatric Imaging, Neurocognition, and Genetics (PING³⁶), and Philadelphia Neurodevelopmental Cohort (PNC³⁷). We harmonized image processing by using the ENIGMA-DTI pipeline^{38,39} and obtained voxel-wise DTI maps for 42,919 subjects (after quality controls), including 35,741 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 parameters for FA, AD, MD, MO and RD in 21 tracts and across the

whole brain. Second, we applied functional principal component analysis (FPCA⁴⁰) to generate 105 tract-specific principal components (PCs) for FA by taking the top five PCs of the voxel-wise map within each tract. FPCA is a data-driven approach to characterize the strongest variation components of FA within each tract, which are expected to provide additional microstructural details about axonal organization and myelination omitted by tract-averaged values^{41,42}, while limiting multiple testing. More importantly, these PCs may represent FA changes that are more relevant to specific clinical outcomes. We then performed a genome-wide association analysis for these 215 phenotypes to discover the genetic architecture of white matter and explore the genetic links to a plethora of clinical endpoints 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 (BIG-KP) <https://bigkp.web.unc.edu/>.

RESULTS

GWAS Discovery and Validation for 215 DTI parameters.

Our discovery analysis utilized data from UKB subjects of British ancestry ($n = 33,292$). All of the 110 DTI mean parameters had significant SNP heritability⁴³ (h^2) after Bonferroni adjustment (215 tests, $P < 9.4 \times 10^{-31}$, **Fig. 1a** 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^{23,25}. 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 1b**. Different from tract-averaged value, these PCs captured more specific FA 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 FA

1 PCs are also generally under genetic control. As illustrated in later sections, those
2 heritable local FA variation patterns may also have higher power to identify the shared
3 genetic influences with other complex traits.

4

5 We performed GWAS for these 215 DTI parameters using 9,023,710 common genetic
6 variants after quality controls (Methods). All Manhattan and QQ plots can be browsed in
7 our BIG-KP server. At a stringent significance level 2.3×10^{-10} (i.e., $5 \times 10^{-8}/215$,
8 additionally adjusted for the 215 phenotypes studied), FUMA⁴⁴ clumped 595 partially
9 independent significant variants (Methods) involved in 1,101 significant associations
10 with 86 FA measures (21 mean and 65 PC parameters, **Supplementary Figs. 2-3** and
11 **Supplementary Table 2**). Genetic variants had broad effects across all white matter
12 tracts, and one variant often influenced multiple FA measures, such as rs12146713 in
13 region 12q23.3, rs309587 in 5q14.3, rs55705857 in 8q24.21, and rs1004763 in 22q13.1.
14 Of the 595 significant variants, 302 were only detected by PC parameters. On average,
15 the number of FA-associated significant variants was 37.0 in each tract (range = (4, 72),
16 **Fig. 2** and **Supplementary Table 3**), 50.3% of which were solely discovered by PC
17 parameters (range = (26.3%, 100%)). For example, all of the 22 significant variants
18 associated with CST were detected by PC parameters. Moreover, 66.7% (32/48) of the
19 variants in posterior corona radiata (PCR), 64.9% (37/57) in posterior thalamic radiation
20 (PTR), 59.7% (43/72) in SLF, and 56.3% (18/32) in cingulum cingulate gyrus (CGC) were
21 only associated with PC parameters. These results clearly illustrate the unique
22 contribution of tract-specific PC parameters in identifying genetic variants for FA
23 variations within white matter tract.

24

25 In addition, 770 significant variants were associated with 83 mean parameters of AD,
26 MD, MO and RD (2,069 significant associations), 565 of these 770 variants (with 967
27 associations) were not identified by FA measures (**Fig. 2**, **Supplementary Figs. 2-3**, and
28 **Supplementary Table 2**). The mean number of significant variants in each tract moved
29 up to 93.3 (range = (41, 160)), and rs13198474 in 6p22.2, rs2267161 in 22q12.2,
30 rs55705857 in 8q24.21, rs7935166 in 11p11.2, and rs7225002 in 7q21.31 were
31 associated with multiple non-FA measures. Of note, more than 70% of significant
32 variants in cingulum (CGH (90.7%) and CGC (73.3%)) were detected by non-FA measures

(**Supplementary Table 4**), which may suggest that FA is less useful in the thin line-like C-shaped cingulum region than in other tracts. Based on a second and more strict LD clumping ($LD\ r^2 < 0.1$), FUMA⁴⁴ defined independent lead variants from the above independent significant variants and then genetic loci were characterized (Methods). The 3,170 (1,101 + 2,609) significant variant-trait associations were summarized as 994 significant locus-trait associations (**Supplementary Tables 5-6**). We then performed functionally informed fine mapping for these locus-level signals using SuSiE⁴⁵ via PolyFun⁴⁶ 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 7**), suggesting 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^{30,47}.

15

We aimed to find independent replication of our discovery GWAS in five independent validation datasets, all consisting of individuals 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 LDSC⁴⁸ (Methods). The mean gc estimate 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 (**Supplementary Table 8**). Genetic correlation estimates near 1 indicates a consistent genetic basis for these phenotypes measured in different cohorts and MRI scanners. 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 and direction of the top variants in independent cohorts^{49,50}. Additionally, we tested for replication by using polygenic risk scores⁵¹ (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

1 **Supplementary Table 9**). Almost all (214/215) DTI parameters had significant PRS in at
2 least one dataset and 165 had significant PRS in all of them, showing the high
3 generalizability of our discovery GWAS results. Across the five validation datasets, the
4 mean additional variance that can be explained by PRS (i.e., incremental R-squared) was
5 1.7% (range = (0.4%, 4.2%)) for the 165 consistently significant DTI parameters. The
6 largest mean (incremental) R-squared was on the 2nd PC of EC (range = (2.2%, 6.5%), P
7 range = (7.2×10^{-24} , 1.5×10^{-9})).

8

9 Finally, we constructed PRS on four non-European validation datasets: the UKB Asian
10 (UKBA, $n = 419$), UKB Black (UKBBL, $n = 211$), ABCD Hispanic (ABCDH, $n = 768$), and ABCD
11 African American (ABCDAA, $n = 1,257$). The number of significant PRS was 158 and 40 in
12 UKBA and UKBBL, respectively (B-H adjustment at 0.05 level, **Supplementary Table 10**).
13 In addition, UKBW and UKBA had similar prediction performance (mean 2.38% vs.
14 2.33%, $P = 0.67$), but the accuracy became significantly smaller in UKBBL (mean 2.38%
15 vs. 1.67%, $P = 3.9 \times 10^{-9}$). For the two non-European non-UKB datasets, the number of
16 significant PRS was 121 and 114 in ABCDH and ABCDA, respectively (B-H adjustment at
17 0.05 level, **Supplementary Table 11**), which were much smaller than the ones observed
18 in ABCDE. The R-squared were similar between ABCDH and ABCDE (mean 0.74% vs.
19 0.69%, $P = 0.28$), but the accuracy significantly decreased in ABCDA (mean 0.48% vs.
20 0.69%, $P = 1.9 \times 10^{-7}$). These findings show that UKB British GWAS findings have high
21 generalizability in European cohorts, but the generalizability is reduced in
22 cross-population applications, especially in Black/African-American cohorts, highlighting
23 the importance of recruiting sufficient samples from global diverse populations in future
24 genetics discovery of white matter.

25

26 **Concordance with previous GWAS.**

27 Of the 33,292 subjects in our UKB British discovery GWAS, 17,706 had been used in the
28 largest previous GWAS²⁵ for 110 mean parameters. To examine the robustness of their
29 findings, we used the other 15,214 individuals (also removed the relatives⁵² of previous
30 GWAS subjects) to perform a new validation GWAS and then evaluated the strength of
31 replication (Methods). We calculated the replication slope, which was the correlation of
32 the standardized effect size of variants estimated from two independent GWAS⁵³. 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 FINDOR⁵³ 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 used LDSC⁴⁸ to calculate genetic correlation between measurements from the two GWAS. The mean gc estimate 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.²⁵, 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 catalog⁵⁴, our results tagged many variants that had been implicated with brain structures, including 7 in van der Meer, et al.⁵⁵ for hippocampal subfield volumes, 7 in Verhaaren, et al.⁵⁶ for cerebral white matter hyperintensity (WMH) burden, 5 in Vojinovic, et al.⁵⁷ for lateral ventricular volume, 5 in Rutten-Jacobs, et al.²⁶ for WMH and white matter integrity, 2 in Klein, et al.⁵⁸ for intracranial volume, 2 in Hibar, et al.⁵⁹ for subcortical brain region volumes, 2 in Fornage, et al.²⁸ for WMH burden, 1 in Elliott, et al.²³ for brain imaging measurements, 1 in Luo, et al.⁶⁰ for voxel-wise brain imaging measurement, 1 in Hashimoto, et al.⁶¹ for superior frontal gyrus grey matter volume, 1 in Ikram, et al.⁶² for intracranial volume, and 1 in Sprooten, et al.⁶³ 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.⁶⁴ for brain imaging measurements, 2 in Chung, et al.⁶⁵ for hippocampal volume in dementia, 1 in Chen, et al.⁶⁶ for putamen volume, and 1 in Christopher, et al.⁶⁷ for posterior cingulate cortex (**Supplementary Table 15**). For example, we observed colocalizations in region 5q14.3 with previously reported variants for WMH volume and white matter integrity²⁶, in 10q26.13 with hippocampal

1 volumes⁵⁵, in 17q21.31 with subcortical⁵⁹ and intracranial⁶² volumes, and in 17q25.1
2 with WMH volume²⁶/burden^{28,56} (**Supplementary Fig. 7**).

3

4 Moreover, we found lots of previous associations with other complex traits in different
5 domains (**Supplementary Table 16**). We highlighted 190 variants with psychological
6 traits (e.g., neuroticism⁶⁸, well-being spectrum⁶⁹, general risk tolerance⁷⁰), 179 with
7 cognitive/educational traits (e.g., cognitive ability⁷¹, educational attainment⁷²), 99 with
8 psychiatric disorders (e.g., schizophrenia⁷³, MDD⁷⁴, bipolar disorder⁷⁵, ADHD⁷⁶, autism
9 spectrum disorder⁷⁷), 95 with anthropometric traits (e.g., height⁷⁸, body mass index
10 (BMI)⁵³), 68 with bone mineral density^{79,80}, 54 with smoking/drinking (e.g., smoking⁸¹,
11 alcohol use disorder⁸²), 20 with neurological disorders (e.g., corticobasal degeneration⁸³,
12 Parkinson's disease⁸⁴, Alzheimer's disease⁸⁵, multiple sclerosis⁸⁶), 18 with sleep (e.g.,
13 sleep duration⁸⁷, chronotype⁸⁸), 11 with glioma (glioblastoma or non-glioblastoma)
14 tumors^{89,90}, and 6 with stroke⁹¹⁻⁹³. For example, white matter associated variants
15 colocalized with many risk variants of cognitive/educational traits as well as
16 brain-related disorders in regions 17q21.31, 6p22.1, and 6p22.2 (**Supplementary Fig. 8**).
17 Strong colocalizations were also found in 7p22.3 with anthropometric traits and bone
18 mineral density, in 10p12.31 with smoking/drinking and anthropometric traits, in 9p21.3
19 with glioma and stroke, and in 8q24.12 with bone mineral density (**Supplementary Fig.**
20 **9**).

21

22 To further explore these overlaps, we summarized the number of previously reported
23 variants of other traits that can be tagged by any DTI parameters in each white matter
24 tract (**Supplementary Table 17**). We found that variants associated with psychological,
25 cognitive/educational, smoking/drinking traits and neurological and psychiatric
26 disorders were globally linked to many white matter tracts (**Supplementary Fig. 10**). For
27 traits in other domains, the overlaps may have some tract-specific patterns. For
28 example, 3 of the 6 variants associated with stroke were linked to both SFO and ALIC,
29 and the other 3 were found in superior corona radiata (SCR), anterior corona radiata
30 (ACR), genu of corpus callosum (GCC), body of corpus callosum (BCC), EC, posterior limb
31 of internal capsule (PLIC), and posterior limb of internal capsule (RLIC). In addition, 7 of
32 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). **Figure 3** illustrated their genomic locations and voxel-wise effect size patterns in spatial brain 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 psychiatric⁹⁴ and neurological⁹⁵ disorders, as well as psychological⁹⁶ and cognitive/educational⁹⁷ traits. On the other hand, rs77126132 highlighted in SCC and BCC was particularly linked to glioma⁸⁹, and rs798510 in SCR, FX, and PLIC was associated with several anthropometric traits⁹⁸.

14 **An atlas of genetic correlations with other complex traits.**

15 Because of the shared loci associated with both white matter microstructure and other
16 complex traits, we systematically examined their pairwise genetic correlations by using
17 our discovery GWAS summary statistics ($n = 33,292$) and publicly available
18 summary-level data of other 76 complex traits via LDSC (Methods, **Supplementary Table**
19 **18**). There were 760 significant pairs between 60 complex traits and 175 DTI parameters
20 after B-H adjustment at 0.05 level (76×215 tests, P range = $(8.6 \times 10^{-12}, 2.3 \times 10^{-3})$,
21 **Supplementary Table 19**), 38.3% (291/760) of which were detected by PC parameters.
22 We found that DTI parameters were widely correlated with subcortical and WMH
23 volumes (**Supplementary Fig. 11**), brain-related traits (**Supplementary Fig. 12**), and
24 other non-brain traits (**Supplementary Fig. 13**). To validate these results, we performed
25 cross-trait PRS separately on our five European validation GWAS datasets and LDSC on
26 their meta-analyzed summary statistics ($n = 6,962$, Methods). We found that 681
27 (89.6%) of these 760 significant pairs can be validated in at least one of the six validation
28 analyses after B-H adjustment at 0.05 level (760 tests, P range = $(1.7 \times 10^{-10}, 2.9 \times 10^{-2})$,
29 **Supplementary Table 20**), indicating the robustness of our findings. We then reran LDSC
30 after meta-analyzed our UKB British discovery GWAS with these European validation
31 GWAS ($n = 40,254$). The number of significant pairs increased to 855 between 62

1 complex traits and 178 DTI parameters (**Fig. 4, Supplementary Figs. 14-16 and**
2 **Supplementary Table 21**).

3

4 We replicated previously reported genetic correlations with cognitive/educational
5 traits²⁵, drinking behavior²⁵, stroke^{23,26}, and MDD^{25,26}, and more tract-specific details
6 were revealed. For example, stroke (any subtypes) and ischemic stroke subtypes⁹² (large
7 artery stroke, cardioembolic stroke, and small vessel stroke) showed broad genetic
8 correlations with corpus callosum (GCC and BCC), corona radiata (ACR, SCR, and PCR),
9 limb of internal capsule (PLIC, ALIC), EC, SLF, SFO, and UNC ($|gc|$ range = (0.16, 0.42), $P <$
10 2.5×10^{-3}), matching findings in our association lookups. We further observed that small
11 vessel stroke subtype had specific but higher genetic correlations with ALIC and SFO
12 ($|gc|$ range = (0.52, 0.69), $P < 1.2 \times 10^{-3}$). In contrast, there were no significant genetic
13 correlations detected for large artery and cardioembolic stroke, demonstrating the
14 potentially much stronger genetic links between white matter tracts and small vessel
15 stroke subtype.

16

17 More importantly, many new genetic correlations were uncovered for brain-related
18 traits, such as Alzheimer's disease, ADHD, bipolar disorder, schizophrenia, chronotype,
19 insomnia, neuroticism, and risk tolerance. For example, significant genetic correlation
20 was found between PTR and Alzheimer's disease ($|gc| = 0.30$, $P = 1.7 \times 10^{-3}$), EC and
21 ADHD ($|gc| = 0.18$, $P = 4.5 \times 10^{-5}$), UNC and bipolar disorder ($|gc| > 0.15$, $P < 4.0 \times 10^{-4}$),
22 and SLF and schizophrenia ($|gc| = 0.11$, $P = 2.3 \times 10^{-3}$), matching previously reported
23 case-control differences^{12,99-101} on these tracts. We also found novel significant
24 correlations for non-brain traits, including high blood pressure, height, BMI, bone
25 mineral density, number of non-cancer illnesses and treatments, heavy manual or
26 physical work, smoking, coronary artery disease, lung function, and type 2 diabetes
27 (T2D). For example, high blood pressure was genetically correlated with 19 tracts
28 including SFO, SLF, UNC, EC, and ALIC ($|gc|$ range = (0.09, 0.25), $P < 2.4 \times 10^{-3}$). Previous
29 research found widespread associations between human brain and these traits, such as
30 bone mineral density¹⁰², hypertension¹⁰³, T2D¹⁰⁴, lung function¹⁰⁵, heart disease¹⁰⁶, and
31 anthropometric traits¹⁰⁷. Our findings further illuminate their underlying genetic links.
32 We summarized significant genetic correlations identified in each tract and found that

1 32.3% (120/372) of these tract-trait genetic correlations can only be detected by PC
2 parameters (**Supplementary Fig. 17** and **Supplementary Table 22**). For example, most of
3 the significant genetic correlations in EC were solely detected by its PC parameters, such
4 as ADHD, BMI, cognitive function, neuroticism, and insomnia.

5

6 We explored partial genetic causality among these traits using the latent causal
7 variable¹⁰⁸ (LCV) model (Methods). As suggested, we conservatively restricted the LCV
8 analysis to pairs with at least nominally significant genetic correlation ($P < 0.05$),
9 significant evidence of genetic causality (B-H adjustment at 0.01 level, 76×215 tests),
10 and large genetic causality proportion estimate ($|GCP| > 0.6$), which were extremely
11 unlikely to be false positives¹⁰⁸. The LCV model suggested that high blood pressure was
12 partially genetically causal for white matter ($|GCP| > 0.67$, $P < 2.2 \times 10^{-5}$,
13 **Supplementary Fig. 18** and **Supplementary Table 23**). On the other hand, white matter
14 may have partially genetically causal effects on insomnia, under sleep, and neuroticism
15 ($|GCP| > 0.64$, $P < 7.1 \times 10^{-8}$). These findings may lead to plausible biological hypotheses
16 in future research and suggest the existence of different biological mechanisms
17 underlying the atlas of genetic correlations. More efforts are required to explore causal
18 relationships and the shared biological processes¹⁰⁹ among these genetically correlated
19 traits.

20

21 **Gene-level analysis.**

22 We carried out MAGMA¹¹⁰ gene-based association analysis for the 215 DTI parameters
23 using our discovery GWAS summary statistics (Methods). There were 3,903 significant
24 gene-level associations ($P < 1.2 \times 10^{-8}$, adjusted for 215 phenotypes) between 620 genes
25 and 179 DTI parameters (**Supplementary Table 24**), 153 of the associated genes can
26 only be discovered by PC parameters. We replicated 99 of 112 MAGMA genes reported
27 in Zhao, et al.²⁵, 8 white matter-associated genes (*SH3PXD2A*, *NBEAL1*, *C1QL1*, *COL4A2*,
28 *TRIM47*, *TRIM65*, *UNC13D*, *FBF1*) in Verhaaren, et al.⁵⁶, 4 (*VCAN*, *TRIM47*, *XRCC4*,
29 *HAPLN1*) in Rutten-Jacobs, et al.²⁶, 3 (*ALDH2*, *PLEKHG1*, *TRIM65*) in T aylor, et al.²⁷, 3
30 (*ALDH2*, *PLEKHG1*, *TRIM65*) in Hofer, et al.¹¹¹, 2 (*TRIM47*, *TRIM65*) in Fornage, et al.²⁸,
31 and 2 (*GNA12*, *GNA13*) in Sprooten, et al.¹¹². Most of the other genes had not been
32 implicated with white matter. Many of our MAGMA genes had been linked to other

1 complex traits (**Supplementary Table 25**), such as 70 genes in Anney, et al.⁹⁴ for autism
 2 spectrum disorder or schizophrenia, 50 in Morris, et al.⁷⁹ for heel bone mineral density,
 3 38 in Hoffmann, et al.¹¹³ for blood pressure variation, 51 in Linnér, et al.⁷⁰ for risk
 4 tolerance, 36 in Rask-Andersen, et al.⁹⁸ for body fat distribution, and 26 in Hill, et al.¹¹⁴
 5 for neuroticism.

6

7 Next, we mapped significant variants ($P < 2.3 \times 10^{-10}$) to genes according to physical
 8 position, expression quantitative trait loci (eQTL) association, and 3D chromatin (Hi-C)
 9 interaction via FUMA⁴⁴ (Methods). FUMA yielded 1,189 new associated genes (1,630 in
 10 total) that were not discovered in MAGMA analysis (**Supplementary Table 26**),
 11 replicating 286 of the 292 FUMA genes identified in Zhao, et al.²⁵ and more other genes
 12 in previous studies of white matter, such as *PDCD11*⁵⁶, *ACOX1*⁵⁶, *CLDN23*¹¹¹,
 13 *EFEMP1*^{26,27,56}, and *IRS2*¹¹¹. More overlapped genes were also observed between white
 14 matter and other traits (**Supplementary Table 27**). Particularly, 876 FUMA genes were
 15 solely mapped by significant Hi-C interactions in brain tissues (**Supplementary Table 28**),
 16 demonstrating the power of integrating chromatin interaction profiles in GWAS of white
 17 matter.

18

19 We then explored the gene-level pleiotropy between white matter and 79 complex
 20 traits, including nine neurological and psychiatric disorders¹¹⁵ studied in Sey, et al.¹¹⁵
 21 and (other) traits studied in our genetic correlation analysis. For brain-related traits, the
 22 associated genes were predicted by the recently developed Hi-C-coupled MAGMA¹¹⁵
 23 (H-MAGMA) tool (Methods). Traditional MAGMA¹¹⁰ was used for non-brain GWAS.
 24 H-MAGMA prioritized 737 significant genes for white matter ($P < 6.3 \times 10^{-9}$, adjusted for
 25 215 phenotypes and two brain tissue types, **Supplementary Table 29**), and we focused
 26 on 329 genes that can be replicated in our meta-analyzed European validation GWAS (n
 27 = 6,962) at nominal significance level ($P < 0.05$, **Supplementary Table 30**). We found
 28 that 298 of these 329 genes were associated with at least one of 57 complex traits
 29 (**Supplementary Table 31**). **Supplementary Figure 19** and **Supplementary Table 32**
 30 display the number of overlapped genes between 57 complex traits and 21 white matter
 31 tracts. Most white matter tracts have many pleiotropic genes with other complex traits,
 32 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^{101,116,117}. Particularly, 230 white matter H-MAGMA genes had been identified in Sey, et al.¹¹⁵ for nine neurological and psychiatric disorders (**Supplementary Table 33**). *NSF*¹¹⁸, *GFAP*¹¹⁹, *TRIM27*⁷³, *HLA-DRA*^{118,120}, and *KANSL1*^{77,96} were associated with five of these disorders, and another 69 genes were linked to at least three different disorders (**Supplementary Fig. 20**). 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.

In order to identify tissues and cell types where genetic variation leads to changes in white matter microstructure, we performed partitioned heritability analyses¹²¹ from the GWAS of global FA and MD within tissue type and cell type specific regulatory elements. First, we utilized regulatory elements across multiple adult and fetal tissues¹²². As expected, both FA and MD had the most significant enrichment of heritability in active gene regulation regions of brain tissues (**Fig. 5a**, **Supplementary Fig. 21**, and **Supplementary Table 34**). To identify gross cell types, we again performed partitioned heritability using chromatin accessibility data of two brain cell types, neurons (NeuN+) and glia (NeuN-) sampled from 14 brain regions, including both cortical and subcortical¹²³. For all regions, we found that significant enrichment of FA and MD heritability existed in glial but not neuronal regulatory elements after B-H adjustment at 0.05 level (**Fig. 5b**). These results are expected as white matter is largely composed of glial cell types. For further resolution on cell types, we tested partitioned heritability enrichment within differentially accessible chromatin of glial cell subtypes, oligodendrocyte (NeuN-/Sox10+), microglia and astrocyte (NeuN-/Sox10-) and two neuronal cell subtypes GABAergic (NeuN+/Sox6+) and glutamatergic neurons (NeuN+/Sox6-) (Methods). Heritability of FA and MD was significantly enriched in oligodendrocyte, microglia, and astrocyte annotations ($P < 4.8 \times 10^{-3}$). The oligodendrocyte annotation accounted for 10.4% (standard error = 2.6%, $P = 9.5 \times 10^{-5}$) of the FA heritability while only composed 0.3% of the variants. In contrast, no significant enrichment was observed in neurons (**Fig. 5c**). These analyses imply that

1 common variants associated with white matter microstructure alter the function of
2 regulatory elements in glial cells, particularly oligodendrocytes, the cell type expected to
3 influence white matter microstructure, providing strong support of the biological
4 validity of the genetic associations.

5

6 To gain more insights into biological mechanisms, we performed several analyses to
7 explore biological interpretations of white matter associated genes. First, MAGMA gene
8 property¹¹⁰ analysis was carried out for 13 GTEx¹²⁴ (v8) brain tissues to examine whether
9 the tissue-specific gene expression levels were related to significance between genes
10 and DTI parameters (Methods). After Bonferroni adjustment (13×215 tests), we
11 detected 57 significant associations for gene expression in brain cerebellar hemisphere
12 and cerebellum tissues ($P < 1.8 \times 10^{-5}$, **Supplementary Fig. 22** and **Supplementary Table**
13 **35**), suggesting that genes with higher transcription levels on white matter-presented
14 regions also had stronger genetic associations with DTI parameters. In contrast, no
15 signals were observed on regions primarily dominated by grey matter, such as basal
16 ganglia and cortex. Next, we performed drug target lookups in a recently established
17 drug target network¹²⁵, which included 273 nervous system drugs (ATC code starts with
18 “N”) and 241 targeted genes. We found that 19 white matter associated genes were
19 targets for 104 drugs, 43 of which were anti-psychotics (ATC: N05A, target such as
20 *DRD4*) to manage psychosis like schizophrenia and bipolar, 40 were anti-depressants
21 (ATC: N06A, target such as *SLC6A4*) to treat MDD and other conditions, 14 were
22 anti-Parkinson drugs (ATC: N04B, target such as *HTR2B*), and 14 were anti-convulsants
23 (ATC: N03A, target such as *SCN5A*) used in the treatment of epileptic seizures
24 (**Supplementary Table 36**). In addition, we treated white matter associated genes as an
25 annotation and performed partitioned heritability enrichment analysis¹²¹ for the other
26 76 complex traits (Methods). After B-H adjustment at 0.05 level, heritability of 54
27 complex traits was significantly enriched in regions influencing DTI parameters
28 (**Supplementary Fig. 23** and **Supplementary Table 37**). These results suggest the
29 potential clinical values of the genes identified for white matter microstructure.

30

31 MAGMA¹¹⁰ competitive gene-set analysis was performed for 15,496 gene sets (5,500
32 curated gene sets and 9,996 GO terms, Methods). We found 180 significant gene sets

1 after Bonferroni adjustment ($15,496 \times 215$ tests, $P < 1.5 \times 10^{-8}$, **Supplementary Table 38**).
2 The top five frequently prioritized gene sets were “dacosta uv response via ercc3 dn”
3 (M4500), “dacosta uv response via ercc3 common dn” (M13522), “graessmann
4 apoptosis by doxorubicin dn” (M1105), “gobert oligodendrocyte differentiation dn”
5 (M2369), and “blalock alzheimers disease up” (M12921). M4500 and M13522 are
6 *ERCC3*-associated gene sets related to xeroderma pigmentosum (XP) and
7 trichothiodystrophy (TTD) syndromes, which are genetic disorders caused by a defective
8 nucleotide excision repair system^{126,127}. In addition to skin symptoms, patients of XP and
9 TTD often reported various neurological deteriorations and white matter abnormalities,
10 such as intellectual impairment¹²⁸, myelin structures degradation¹²⁹, and diffuse
11 dysmyelination¹³⁰. M1105 regulates the apoptosis of breast cancer cells in response to
12 doxorubicin treatment. Clinical research found that breast cancer chemotherapy like
13 doxorubicin was neurotoxic¹³¹ and can cause therapy-induced brain structural changes
14 and decline in white matter integrity¹³². M2369 plays a critical role in oligodendrocyte
15 differentiation, which mediates the repair of white matter after damaging events¹³³, and
16 M12921 is related to the pathogenesis of Alzheimer's disease¹³⁴.

17
18 Several gene sets of rat sarcoma (Ras) proteins, small GTPases, and rho family GTPases
19 were also prioritized by MAGMA, such as “go regulation of small gtpase mediated signal
20 transduction” (GO: 0051056), “go small gtpase mediated signal transduction” (GO:
21 0007264), “go regulation of ras protein signal transduction” (GO: 0046578), “go ras
22 protein signal transduction” (GO: 0007265), and “reactome signaling by rho gtpases”
23 (M501). Ras proteins activity is involved in developmental processes and abnormalities
24 of neural cells in central nervous system^{135,136}; small and rho family GTPases play crucial
25 roles in basic cellular processes during the entire neurodevelopment process and are
26 closely connected to several neurological disorders¹³⁷⁻¹³⁹. We also observed significant
27 enrichment in pathways related to nervous system, including “go neurogenesis” (GO:
28 0022008), “go neuron differentiation” (GO: 0030182), “go neuron development” (GO:
29 0048666), “go regulation of neuron differentiation” (GO: 0045664), and “go regulation
30 of nervous system development” (GO: 0051960). Finally, we applied DEPICT¹⁴⁰ gene-set
31 enrichment testing for 10,968 pre-constituted gene sets (Methods), 7 of which survived
32 Bonferroni adjustment ($10,968 \times 215$ 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 39**). 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). 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 microstructure and multiple diseases and disorders.

DISCUSSION

In this study, we analyzed the genetic architecture of brain white matter using dMRI scans of 42,919 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 and cell-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^{23,141}, but may raise the question that to what degree the research findings can be generalized and applied to global populations^{142,143}. In our analysis, we found that the UKB British-derived PRS were still widely significant in Hispanic, Asian, and Black/African American testing cohorts but had reduced performances, especially in Black/African American cohorts. 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 project¹⁴⁴ 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, one supplementary figure pdf file, and one supplementary table zip file are available.

1

2 **ACKNOWLEDGEMENTS**

3 This research was partially supported by U.S. NIH grants MH086633 (H.Z.), MH116527
 4 (TF.L.), and HD079124 (Y.L.). We thank Sophia Cui, Xiaopeng Zong, and Peter Vandehaar
 5 for helpful conversations. We thank the individuals represented in the UK Biobank,
 6 ABCD, HCP, PING, and PNC studies for their participation and the research teams for
 7 their work in collecting, processing and disseminating these datasets for analysis. We
 8 gratefully acknowledge all the studies and databases that made GWAS summary data
 9 available. This research has been conducted using the UK Biobank resource (application
 10 number 22783), subject to a data transfer agreement. Part of the data collection and
 11 sharing for this project was funded by the Pediatric Imaging, Neurocognition and
 12 Genetics Study (PING) (U.S. National Institutes of Health Grant RC2DA029475). PING is
 13 funded by the National Institute on Drug Abuse and the Eunice Kennedy Shriver
 14 National Institute of Child Health & Human Development. PING data are disseminated
 15 by the PING Coordinating Center at the Center for Human Development, University of
 16 California, San Diego. Support for the collection of the PNC datasets was provided by
 17 grant RC2MH089983 awarded to Raquel Gur and RC2MH089924 awarded to Hakon
 18 Hakonarson. All PNC subjects were recruited through the Center for Applied Genomics
 19 at The Children's Hospital in Philadelphia. Part of the data used in the preparation of this
 20 article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study
 21 (<https://abcdstudy.org>), held in the NIMH Data Archive (NDA). This is a multisite,
 22 longitudinal study designed to recruit more than 10,000 children age 9-10 and follow
 23 them over 10 years into early adulthood. The ABCD Study is supported by the National
 24 Institutes of Health and additional federal partners under award numbers
 25 U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106,
 26 U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156,
 27 U01DA041174, U24DA041123, U24DA041147, U01DA041093, and U01DA041025. A full
 28 list of supporters is available at <https://abcdstudy.org/federal-partners.html>. A listing of
 29 participating sites and a complete listing of the study investigators can be found at
 30 <https://abcdstudy.org/scientists/workgroups/>. ABCD consortium investigators designed
 31 and implemented the study and/or provided data but did not necessarily participate in
 32 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.L.S. designed the study. B.Z., T.F. L, Y.Y., X.W., and T.Y. L analyzed the data. T.F. L, Y.S., Z.Z., Y.Y., X.W., T.Y. L, and D.X., downloaded the datasets, preprocessed dMRI data, and undertook the quantity controls. P.R., M.E.H., J.B., and J.F.F. analyzed brain cell chromatin accessibility data. B.Z. and H.Z. wrote the manuscript with feedback from all authors.

CORRESPONDENCE AND REQUESTS FOR MATERIALS should be addressed to H.Z.

COMPETING 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).

- 1 6. Kuceyeski, A., Maruta, J., Relkin, N. & Raj, A. The Network Modification (NeMo)
2 Tool: elucidating the effect of white matter integrity changes on cortical and
3 subcortical structural connectivity. *Brain connectivity* **3**, 451-463 (2013).
- 4 7. Bathelt, J., Scerif, G., Nobre, A. & Astle, D. Whole-brain white matter
5 organization, intelligence, and educational attainment. *Trends in neuroscience*
6 *and education* **15**, 38-47 (2019).
- 7 8. Vaquero, L., Rodríguez-Fornells, A. & Reiterer, S.M. The left, the better:
8 white-matter brain integrity predicts foreign language imitation ability. *Cerebral*
9 *Cortex* **27**, 3906-3917 (2017).
- 10 9. Wu, Z.-M. *et al.* White matter microstructural alterations in children with ADHD:
11 categorical and dimensional perspectives. *Neuropsychopharmacology* **42**,
12 572-580 (2017).
- 13 10. Zou, K. *et al.* Alterations of white matter integrity in adults with major depressive
14 disorder: a magnetic resonance imaging study. *Journal of psychiatry &*
15 *neuroscience: JPN* **33**, 525 (2008).
- 16 11. Cetin-Karayumak, S. *et al.* White matter abnormalities across the lifespan of
17 schizophrenia: a harmonized multi-site diffusion MRI study. *Molecular*
18 *psychiatry*, 1-12 (2019).
- 19 12. Versace, A. *et al.* Elevated left and reduced right orbitomedial prefrontal
20 fractional anisotropy in adults with bipolar disorder revealed by tract-based
21 spatial statistics. *Archives of general psychiatry* **65**, 1041-1052 (2008).
- 22 13. De Santis, S. *et al.* Evidence of early microstructural white matter abnormalities
23 in multiple sclerosis from multi-shell diffusion MRI. *NeuroImage: Clinical* **22**,
24 101699 (2019).
- 25 14. Lee, S. *et al.* White matter hyperintensities are a core feature of Alzheimer's
26 disease: evidence from the dominantly inherited Alzheimer network. *Annals of*
27 *neurology* **79**, 929-939 (2016).
- 28 15. Hess, C., Christine, C., Apple, A., Dillon, W. & Aminoff, M. Changes in the
29 thalamus in atypical parkinsonism detected using shape analysis and diffusion
30 tensor imaging. *American Journal of Neuroradiology* **35**, 897-903 (2014).

- 1 16. Veselý, B. & Rektor, I. The contribution of white matter lesions (WML) to
2 Parkinson's disease cognitive impairment symptoms: a critical review of the
3 literature. *Parkinsonism & related disorders* **22**, S166-S170 (2016).
- 4 17. Bassar, P.J., Mattiello, J. & LeBihan, D. Estimation of the effective self-diffusion
5 tensor from the NMR spin echo. *Journal of Magnetic Resonance, Series B* **103**,
6 247-254 (1994).
- 7 18. Grieve, S., Williams, L., Paul, R., Clark, C. & Gordon, E. Cognitive aging, executive
8 function, and fractional anisotropy: a diffusion tensor MR imaging study.
9 *American Journal of Neuroradiology* **28**, 226-235 (2007).
- 10 19. Bassar, P. & Pierpaoli, C. Microstructural and physiological features of tissues
11 elucidated by quantitative-diffusion-tensor MRI. *Journal of magnetic resonance.*
12 *Series B* **111**, 209 (1996).
- 13 20. Ennis, D.B. & Kindlmann, G. Orthogonal tensor invariants and the analysis of
14 diffusion tensor magnetic resonance images. *Magnetic Resonance in Medicine:*
15 *An Official Journal of the International Society for Magnetic Resonance in*
16 *Medicine* **55**, 136-146 (2006).
- 17 21. Mascalchi, M. *et al.* Progression of microstructural damage in spinocerebellar
18 ataxia type 2: a longitudinal DTI study. *American Journal of Neuroradiology* **36**,
19 1096-1101 (2015).
- 20 22. Vuoksima, E. *et al.* Heritability of white matter microstructure in late middle
21 age: A twin study of tract-based fractional anisotropy and absolute diffusivity
22 indices. *Human brain mapping* **38**, 2026-2036 (2017).
- 23 23. Elliott, L.T. *et al.* Genome-wide association studies of brain imaging phenotypes
24 in UK Biobank. *Nature* **562**, 210-216 (2018).
- 25 24. Kochunov, P. *et al.* Heritability of fractional anisotropy in human white matter: a
26 comparison of Human Connectome Project and ENIGMA-DTI data. *Neuroimage*
27 **111**, 300-311 (2015).
- 28 25. Zhao, B. *et al.* Large-scale GWAS reveals genetic architecture of brain white
29 matter microstructure and genetic overlap with cognitive and mental health
30 traits (n = 17,706). *Molecular Psychiatry* (2019).

- 1 26. Rutten-Jacobs, L.C. *et al.* Genetic study of white matter integrity in UK Biobank
2 (N= 8448) and the overlap with stroke, depression, and dementia. *Stroke* **49**,
3 1340-1347 (2018).
- 4 27. Traylor, M. *et al.* Genetic variation in PLEKHG1 is associated with white matter
5 hyperintensities (n= 11,226). *Neurology* **92**, e749-e757 (2019).
- 6 28. Fornage, M. *et al.* Genome-wide association studies of cerebral white matter
7 lesion burden: the CHARGE consortium. *Annals of neurology* **69**, 928-939 (2011).
- 8 29. Adib-Samii, P. *et al.* 17q25 Locus is associated with white matter hyperintensity
9 volume in ischemic stroke, but not with lacunar stroke status. *Stroke* **44**,
10 1609-1615 (2013).
- 11 30. O'Connor, L.J. *et al.* Extreme polygenicity of complex traits is explained by
12 negative selection. *The American Journal of Human Genetics* **105**, 456-476
13 (2019).
- 14 31. Adib-Samii, P. *et al.* Genetic architecture of white matter hyperintensities differs
15 in hypertensive and nonhypertensive ischemic stroke. *Stroke* **46**, 348-353 (2015).
- 16 32. Safadi, Z. *et al.* Functional segmentation of the anterior limb of the internal
17 capsule: linking white matter abnormalities to specific connections. *Journal of*
18 *Neuroscience* **38**, 2106-2117 (2018).
- 19 33. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes
20 of a wide range of complex diseases of middle and old age. *PLoS Medicine* **12**,
21 e1001779 (2015).
- 22 34. Casey, B. *et al.* The adolescent brain cognitive development (ABCD) study:
23 imaging acquisition across 21 sites. *Developmental cognitive neuroscience* **32**,
24 43-54 (2018).
- 25 35. Somerville, L.H. *et al.* The Lifespan Human Connectome Project in Development:
26 A large-scale study of brain connectivity development in 5–21 year olds.
27 *NeuroImage* **183**, 456-468 (2018).
- 28 36. Jernigan, T.L. *et al.* The pediatric imaging, neurocognition, and genetics (PING)
29 data repository. *Neuroimage* **124**, 1149-1154 (2016).
- 30 37. Satterthwaite, T.D. *et al.* Neuroimaging of the Philadelphia neurodevelopmental
31 cohort. *Neuroimage* **86**, 544-553 (2014).

- 1 38. Jahanshad, N. *et al.* Multi-site genetic analysis of diffusion images and voxelwise
2 heritability analysis: A pilot project of the ENIGMA–DTI working group.
3 *Neuroimage* **81**, 455-469 (2013).
- 4 39. Kochunov, P. *et al.* Multi-site study of additive genetic effects on fractional
5 anisotropy of cerebral white matter: comparing meta and megaanalytical
6 approaches for data pooling. *Neuroimage* **95**, 136-150 (2014).
- 7 40. Hall, P., Müller, H.-G. & Wang, J.-L. Properties of principal component methods
8 for functional and longitudinal data analysis. *The annals of statistics*, 1493-1517
9 (2006).
- 10 41. Mattingsdal, M. *et al.* Pathway analysis of genetic markers associated with a
11 functional MRI faces paradigm implicates polymorphisms in calcium responsive
12 pathways. *Neuroimage* **70**, 143-149 (2013).
- 13 42. Azadeh, S. *et al.* Integrative Bayesian analysis of neuroimaging-genetic data
14 through hierarchical dimension reduction. in *2016 IEEE 13th International*
15 *Symposium on Biomedical Imaging (ISBI)* 824-828 (IEEE, 2016).
- 16 43. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide
17 complex trait analysis. *The American Journal of Human Genetics* **88**, 76-82
18 (2011).
- 19 44. Watanabe, K., Taskesen, E., Bochoven, A. & Posthuma, D. Functional mapping
20 and annotation of genetic associations with FUMA. *Nature Communications* **8**,
21 1826 (2017).
- 22 45. Wang, G., Sarkar, A.K., Carbonetto, P. & Stephens, M. A simple new approach to
23 variable selection in regression, with application to genetic fine-mapping.
24 *bioRxiv*, 501114 (2019).
- 25 46. Weissbrod, O. *et al.* Functionally-informed fine-mapping and polygenic
26 localization of complex trait heritability. *BioRxiv*, 807792 (2019).
- 27 47. Martin, A.R., Daly, M.J., Robinson, E.B., Hyman, S.E. & Neale, B.M. Predicting
28 polygenic risk of psychiatric disorders. *Biological psychiatry* **86**, 97-109 (2019).
- 29 48. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases
30 and traits. *Nature Genetics* **47**, 1236-1241 (2015).

- 1 49. Jansen, P.R. *et al.* Genome-wide analysis of insomnia in 1,331,010 individuals
2 identifies new risk loci and functional pathways. *Nature Genetics* **51**, 394-403
3 (2019).
- 4 50. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient
5 than replication-based analysis for two-stage genome-wide association studies.
6 *Nature Genetics* **38**, 209-213 (2006).
- 7 51. Jansen, I.E. *et al.* Genome-wide meta-analysis identifies new loci and functional
8 pathways influencing Alzheimer's disease risk. *Nature genetics* **51**, 404-413
9 (2019).
- 10 52. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic
11 data. *Nature* **562**, 203-209 (2018).
- 12 53. Kichaev, G. *et al.* Leveraging polygenic functional enrichment to improve GWAS
13 power. *The American Journal of Human Genetics* **104**, 65-75 (2019).
- 14 54. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide
15 association studies, targeted arrays and summary statistics 2019. *Nucleic Acids*
16 *Research* **47**, D1005-D1012 (2018).
- 17 55. van der Meer, D. *et al.* Brain scans from 21,297 individuals reveal the genetic
18 architecture of hippocampal subfield volumes. *Molecular Psychiatry*, in press.
19 (2018).
- 20 56. Verhaaren, B.F. *et al.* Multiethnic genome-wide association study of cerebral
21 white matter hyperintensities on MRI. *Circulation: Cardiovascular Genetics* **8**,
22 398-409 (2015).
- 23 57. Vojinovic, D. *et al.* Genome-wide association study of 23,500 individuals
24 identifies 7 loci associated with brain ventricular volume. *Nature*
25 *communications* **9**, 1-11 (2018).
- 26 58. Klein, M. *et al.* Genetic markers of ADHD-related variations in intracranial
27 volume. *American Journal of Psychiatry* **176**, 228-238 (2019).
- 28 59. Hibar, D.P. *et al.* Common genetic variants influence human subcortical brain
29 structures. *Nature* **520**, 224-229 (2015).
- 30 60. Luo, Q. *et al.* Association of a Schizophrenia-Risk Nonsynonymous Variant With
31 Putamen Volume in Adolescents: A Voxelwise and Genome-Wide Association
32 Study. *JAMA psychiatry* **76**, 435-445 (2019).

- 1 61. Hashimoto, R. *et al.* Common variants at 1p36 are associated with superior
2 frontal gyrus volume. *Translational psychiatry* **4**, e472-e472 (2014).
- 3 62. Ikram, M.A. *et al.* Common variants at 6q22 and 17q21 are associated with
4 intracranial volume. *Nature Genetics* **44**, 539-544 (2012).
- 5 63. Sprooten, E. *et al.* Common genetic variants and gene expression associated with
6 white matter microstructure in the human brain. *Neuroimage* **97**, 252-261
7 (2014).
- 8 64. Shen, L. *et al.* Whole genome association study of brain-wide imaging
9 phenotypes for identifying quantitative trait loci in MCI and AD: A study of the
10 ADNI cohort. *Neuroimage* **53**, 1051-1063 (2010).
- 11 65. Chung, J. *et al.* Genome-wide association study of Alzheimer's disease
12 endophenotypes at prediagnosis stages. *Alzheimer's & Dementia* **14**, 623-633
13 (2018).
- 14 66. Chen, C.-H. *et al.* Leveraging genome characteristics to improve gene discovery
15 for putamen subcortical brain structure. *Scientific Reports* **7**, 15736 (2017).
- 16 67. Christopher, L. *et al.* A variant in PPP4R3A protects against alzheimer-related
17 metabolic decline. *Annals of neurology* **82**, 900-911 (2017).
- 18 68. Luciano, M. *et al.* Association analysis in over 329,000 individuals identifies 116
19 independent variants influencing neuroticism. *Nature Genetics* **50**, 6-11 (2018).
- 20 69. Baselmans, B.M. *et al.* Multivariate genome-wide analyses of the well-being
21 spectrum. *Nature genetics* **51**, 445-451 (2019).
- 22 70. Linnér, R.K. *et al.* Genome-wide association analyses of risk tolerance and risky
23 behaviors in over 1 million individuals identify hundreds of loci and shared
24 genetic influences. *Nature Genetics* **51**, 245-257 (2019).
- 25 71. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic
26 loci influencing general cognitive function. *Nature Communications* **9**, 2098
27 (2018).
- 28 72. Lee, J.J. *et al.* Gene discovery and polygenic prediction from a genome-wide
29 association study of educational attainment in 1.1 million individuals. *Nature*
30 *Genetics* **50**, 1112-1121 (2018).

- 1 73. Ikeda, M. *et al.* Genome-wide association study detected novel susceptibility
2 genes for schizophrenia and shared trans-populations/diseases genetic effect.
3 *Schizophrenia bulletin* **45**, 824-834 (2019).
- 4 74. Hyde, C.L. *et al.* Identification of 15 genetic loci associated with risk of major
5 depression in individuals of European descent. *Nature genetics* **48**, 1031 (2016).
- 6 75. Ruderfer, D.M. *et al.* Polygenic dissection of diagnosis and clinical dimensions of
7 bipolar disorder and schizophrenia. *Molecular psychiatry* **19**, 1017 (2014).
- 8 76. Lesch, K.-P. *et al.* Molecular genetics of adult ADHD: converging evidence from
9 genome-wide association and extended pedigree linkage studies. *Journal of*
10 *neural transmission* **115**, 1573-1585 (2008).
- 11 77. Grove, J. *et al.* Identification of common genetic risk variants for autism
12 spectrum disorder. *Nature genetics* **51**, 431 (2019).
- 13 78. Allen, H.L. *et al.* Hundreds of variants clustered in genomic loci and biological
14 pathways affect human height. *Nature* **467**, 832-838 (2010).
- 15 79. Morris, J.A. *et al.* An atlas of genetic influences on osteoporosis in humans and
16 mice. *Nature genetics* **51**, 258-266 (2019).
- 17 80. Medina-Gomez, C. *et al.* Life-course genome-wide association study
18 meta-analysis of total body BMD and assessment of age-specific effects. *The*
19 *American Journal of Human Genetics* **102**, 88-102 (2018).
- 20 81. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights
21 into the genetic etiology of tobacco and alcohol use. *Nature genetics* **51**, 237-244
22 (2019).
- 23 82. Sanchez-Roige, S. *et al.* Genome-wide association study meta-analysis of the
24 Alcohol Use Disorders Identification Test (AUDIT) in two population-based
25 cohorts. *American Journal of Psychiatry*, appi. ajp. 2018.18040369 (2018).
- 26 83. Kouri, N. *et al.* Genome-wide association study of corticobasal degeneration
27 identifies risk variants shared with progressive supranuclear palsy. *Nature*
28 *communications* **6**, 7247 (2015).
- 29 84. Simon-Sanchez, J. *et al.* Genome-wide association study reveals genetic risk
30 underlying Parkinson's disease. *Nature genetics* **41**, 1308 (2009).

- 1 85. Kunkle, B.W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease
2 identifies new risk loci and implicates A β , tau, immunity and lipid processing.
3 *Nature genetics* **51**, 414 (2019).
- 4 86. Beecham, A.H. *et al.* Analysis of immune-related loci identifies 48 new
5 susceptibility variants for multiple sclerosis. *Nature genetics* **45**, 1353 (2013).
- 6 87. Dashti, H. *et al.* GWAS in 446,118 European adults identifies 78 genetic loci for
7 self-reported habitual sleep duration supported by accelerometer-derived
8 estimates. *bioRxiv*, 274977 (2018).
- 9 88. Jones, S.E. *et al.* Genome-wide association analyses of chronotype in 697,828
10 individuals provides insights into circadian rhythms. *Nature communications* **10**,
11 343 (2019).
- 12 89. Melin, B.S. *et al.* Genome-wide association study of glioma subtypes identifies
13 specific differences in genetic susceptibility to glioblastoma and
14 non-glioblastoma tumors. *Nature genetics* **49**, 789 (2017).
- 15 90. Kinnersley, B. *et al.* Genome-wide association study identifies multiple
16 susceptibility loci for glioma. *Nature communications* **6**, 1-9 (2015).
- 17 91. Traylor, M. *et al.* Genetic risk factors for ischaemic stroke and its subtypes (the
18 METASTROKE collaboration): a meta-analysis of genome-wide association
19 studies. *The Lancet Neurology* **11**, 951-962 (2012).
- 20 92. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects
21 identifies 32 loci associated with stroke and stroke subtypes. *Nature genetics* **50**,
22 524-537 (2018).
- 23 93. Dichgans, M. *et al.* Shared genetic susceptibility to ischemic stroke and coronary
24 artery disease: a genome-wide analysis of common variants. *Stroke* **45**, 24-36
25 (2014).
- 26 94. Anney, R.J.L. *et al.* Meta-analysis of GWAS of over 16,000 individuals with autism
27 spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap
28 with schizophrenia. *Molecular Autism* **8**, 21 (2017).
- 29 95. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17
30 new Parkinson's disease risk loci. *Nature genetics* **49**, 1511 (2017).

- 1 96. Nagel, M. *et al.* Meta-analysis of genome-wide association studies for
2 neuroticism in 449,484 individuals identifies novel genetic loci and pathways.
3 *Nature Genetics* **50**, 920 (2018).
- 4 97. Lam, M. *et al.* Large-Scale Cognitive GWAS Meta-Analysis Reveals Tissue-Specific
5 Neural Expression and Potential Nootropic Drug Targets. *Cell reports* **21**,
6 2597-2613 (2017).
- 7 98. Rask-Andersen, M., Karlsson, T., Ek, W.E. & Johansson, Å. Genome-wide
8 association study of body fat distribution identifies adiposity loci and sex-specific
9 genetic effects. *Nature communications* **10**, 339 (2019).
- 10 99. Onnink, A.M.H. *et al.* Deviant white matter structure in adults with
11 attention-deficit/hyperactivity disorder points to aberrant myelination and
12 affects neuropsychological performance. *Progress in Neuro-Psychopharmacology*
13 *and Biological Psychiatry* **63**, 14-22 (2015).
- 14 100. Wen, Q. *et al.* White matter alterations in early-stage Alzheimer's disease: A
15 tract-specific study. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease*
16 *Monitoring* **11**, 576-587 (2019).
- 17 101. Seok, J.-H. *et al.* White matter abnormalities associated with auditory
18 hallucinations in schizophrenia: a combined study of voxel-based analyses of
19 diffusion tensor imaging and structural magnetic resonance imaging. *Psychiatry*
20 *Research: Neuroimaging* **156**, 93-104 (2007).
- 21 102. Minn, Y., Suk, S. & Do, S. Osteoporosis as an independent risk factor for silent
22 brain infarction and white matter changes in men and women: the PRESENT
23 project. *Osteoporosis International* **25**, 2465-2469 (2014).
- 24 103. Hannawi, Y. *et al.* Hypertension is associated with white matter disruption in
25 apparently healthy middle-aged individuals. *American Journal of Neuroradiology*
26 **39**, 2243-2248 (2018).
- 27 104. Tan, X. *et al.* Micro-structural white matter abnormalities in type 2 diabetic
28 patients: a DTI study using TBSS analysis. *Neuroradiology* **58**, 1209-1216 (2016).
- 29 105. Dodd, J.W. *et al.* Brain structure and function in chronic obstructive pulmonary
30 disease: a multimodal cranial magnetic resonance imaging study. *American*
31 *journal of respiratory and critical care medicine* **186**, 240-245 (2012).

- 1 106. Berry, C. *et al.* Small-vessel disease in the heart and brain: current knowledge,
2 unmet therapeutic need, and future directions. *Journal of the American Heart*
3 *Association* **8**, e011104 (2019).
- 4 107. Xu, J., Li, Y., Lin, H., Sinha, R. & Potenza, M.N. Body mass index correlates
5 negatively with white matter integrity in the fornix and corpus callosum: a
6 diffusion tensor imaging study. *Human brain mapping* **34**, 1044-1052 (2013).
- 7 108. O'Connor, L.J. & Price, A.L. Distinguishing genetic correlation from causation
8 across 52 diseases and complex traits. *Nature genetics* **50**, 1728-1734 (2018).
- 9 109. Cortes, A., Albers, P.K., Dendrou, C.A., Fugger, L. & McVean, G. Identifying
10 cross-disease components of genetic risk across hospital data in the UK Biobank.
11 *Nature Genetics* **52**, 126-134 (2020).
- 12 110. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized
13 gene-set analysis of GWAS data. *PLoS Computational Biology* **11**, e1004219
14 (2015).
- 15 111. Hofer, E. *et al.* White matter lesion progression: genome-wide search for genetic
16 influences. *Stroke* **46**, 3048-3057 (2015).
- 17 112. Sprooten, E. *et al.* White matter integrity as an intermediate phenotype:
18 exploratory genome-wide association analysis in individuals at high risk of
19 bipolar disorder. *Psychiatry Research* **206**, 223-231 (2013).
- 20 113. Hoffmann, T.J. *et al.* Genome-wide association analyses using electronic health
21 records identify new loci influencing blood pressure variation. *Nature genetics*
22 **49**, 54 (2017).
- 23 114. Hill, W.D. *et al.* Genetic contributions to two special factors of neuroticism are
24 associated with affluence, higher intelligence, better health, and longer life.
25 *Molecular psychiatry*, 1-19 (2019).
- 26 115. Sey, N.Y. *et al.* A computational tool (H-MAGMA) for improved prediction of
27 brain-disorder risk genes by incorporating brain chromatin interaction profiles.
28 (Nature Publishing Group, 2020).
- 29 116. Kelly, S. *et al.* Widespread white matter microstructural differences in
30 schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia
31 DTI Working Group. *Molecular psychiatry* **23**, 1261 (2018).

- 1 117. Stämpfli, P. *et al.* Subtle white matter alterations in schizophrenia identified with
2 a new measure of fiber density. *Scientific reports* **9**, 1-11 (2019).
- 3 118. Hamza, T.H. *et al.* Common genetic variation in the HLA region is associated with
4 late-onset sporadic Parkinson's disease. *Nature genetics* **42**, 781 (2010).
- 5 119. Hol, E. *et al.* Neuronal expression of GFAP in patients with Alzheimer pathology
6 and identification of novel GFAP splice forms. *Molecular psychiatry* **8**, 786-796
7 (2003).
- 8 120. Jakkula, E. *et al.* Genome-wide association study in a high-risk isolate for multiple
9 sclerosis reveals associated variants in STAT3 gene. *The American Journal of*
10 *Human Genetics* **86**, 285-291 (2010).
- 11 121. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using
12 genome-wide association summary statistics. *Nature genetics* **47**, 1228-1235
13 (2015).
- 14 122. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes.
15 *Nature* **518**, 317 (2015).
- 16 123. Fullard, J.F. *et al.* An atlas of chromatin accessibility in the adult human brain.
17 *Genome research* **28**, 1243-1252 (2018).
- 18 124. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across
19 human tissues. *BioRxiv*, 787903 (2019).
- 20 125. Wang, Q. *et al.* A Bayesian framework that integrates multi-omics data and gene
21 networks predicts risk genes from schizophrenia GWAS data. *Nature*
22 *neuroscience* **22**, 691 (2019).
- 23 126. Uribe-Bojanini, E., Hernandez-Quiceno, S. & Cock-Rada, A.M. Xeroderma
24 Pigmentosum with Severe Neurological Manifestations/De Sanctis–Cacchione
25 Syndrome and a Novel XPC Mutation. *Case reports in medicine* **2017**(2017).
- 26 127. Kraemer, K.H. *et al.* Xeroderma pigmentosum, trichothiodystrophy and Cockayne
27 syndrome: a complex genotype–phenotype relationship. *Neuroscience* **145**,
28 1388-1396 (2007).
- 29 128. Anttinen, A. *et al.* Neurological symptoms and natural course of xeroderma
30 pigmentosum. *Brain* **131**, 1979-1989 (2008).

- 1 129. Kassubek, J. *et al.* The cerebro-morphological fingerprint of a progeroid
2 syndrome: white matter changes correlate with neurological symptoms in
3 xeroderma pigmentosum. *PLoS one* **7**(2012).
- 4 130. Harreld, J., Smith, E., Prose, N., Puri, P. & Barboriak, D. Trichothiodystrophy with
5 dysmyelination and central osteosclerosis. *American journal of neuroradiology*
6 **31**, 129-130 (2010).
- 7 131. Menning, S. *et al.* Changes in brain white matter integrity after systemic
8 treatment for breast cancer: a prospective longitudinal study. *Brain imaging and*
9 *behavior* **12**, 324-334 (2018).
- 10 132. Deprez, S., Billiet, T., Sunaert, S. & Leemans, A. Diffusion tensor MRI of
11 chemotherapy-induced cognitive impairment in non-CNS cancer patients: a
12 review. *Brain imaging and behavior* **7**, 409-435 (2013).
- 13 133. Hamanaka, G., Ohtomo, R., Takase, H., Lok, J. & Arai, K. White-matter repair:
14 Interaction between oligodendrocytes and the neurovascular unit. *Brain*
15 *circulation* **4**, 118 (2018).
- 16 134. Blalock, E.M. *et al.* Incipient Alzheimer's disease: microarray correlation analyses
17 reveal major transcriptional and tumor suppressor responses. *Proceedings of the*
18 *National Academy of Sciences* **101**, 2173-2178 (2004).
- 19 135. Koini, M., Rombouts, S., Veer, I., Van Buchem, M. & Huijbregts, S. White matter
20 microstructure of patients with neurofibromatosis type 1 and its relation to
21 inhibitory control. *Brain imaging and behavior* **11**, 1731-1740 (2017).
- 22 136. Kang, M. & Lee, Y.-S. The impact of RASopathy-associated mutations on CNS
23 development in mice and humans. *Molecular brain* **12**, 1-17 (2019).
- 24 137. Shikanai, M., Yuzaki, M. & Kawauchi, T. Rab family small GTPases-mediated
25 regulation of intracellular logistics in neural development. *Histology and*
26 *Histopathology* **33**, 765-771 (2017).
- 27 138. Qu, L. *et al.* The Ras Superfamily of Small GTPases in Non-neoplastic Cerebral
28 Diseases. *Frontiers in Molecular Neuroscience* **12**, 121 (2019).
- 29 139. Govek, E.-E., Newey, S.E. & Van Aelst, L. The role of the Rho GTPases in neuronal
30 development. *Genes & development* **19**, 1-49 (2005).
- 31 140. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies
32 using predicted gene functions. *Nature Communications* **6**, 5890 (2015).

- 1 141. Smith, S.M. & Nichols, T.E. Statistical challenges in “big data” human
2 neuroimaging. *Neuron* **97**, 263-268 (2018).
- 3 142. Martin, A.R. *et al.* Clinical use of current polygenic risk scores may exacerbate
4 health disparities. *Nature genetics* **51**, 584 (2019).
- 5 143. Duncan, L. *et al.* Analysis of polygenic risk score usage and performance in
6 diverse human populations. *Nature communications* **10**, 1-9 (2019).
- 7 144. Xu, Q. *et al.* CHIMGEN: a Chinese imaging genetics cohort to enhance
8 cross-ethnic and cross-geographic brain research. *Molecular Psychiatry*, 1-13
9 (2019).
- 10 145. Chen, C.-Y. *et al.* Improved ancestry inference using weights from external
11 reference panels. *Bioinformatics* **29**, 1399-1406 (2013).
- 12 146. Jiang, L. *et al.* A resource-efficient tool for mixed model association analysis of
13 large-scale data. *Nature genetics* **51**, 1749 (2019).
- 14 147. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and
15 population-based linkage analyses. *The American Journal of Human Genetics* **81**,
16 559-575 (2007).
- 17 148. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of
18 genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).
- 19 149. Consortium, I.H. Integrating common and rare genetic variation in diverse
20 human populations. *Nature* **467**, 52-58 (2010).
- 21 150. Wang, D. *et al.* Comprehensive functional genomic resource and integrative
22 model for the human brain. *Science* **362**, eaat8464 (2018).
- 23 151. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in
24 developing human brain. *Nature* **538**, 523-527 (2016).
- 25 152. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics*
26 **27**, 1739-1740 (2011).

27 28 **METHODS**

29
30 **GWAS design and Imaging phenotypes.** We analyzed the following GWAS datasets
31 separately: 1) the UKB British discovery GWAS, which used data of individuals of British
32 ancestry⁵² from the UKB study ($n = 33,292$); 2) five validation GWAS performed on

individuals of European ancestry: 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 UKB validation GWAS: UKB Asian (UKBA, $n = 419$) and UKB Black (UKBBL, $n = 211$); 4) two non-European non-UKB validation GWAS, including ABCD Hispanic (ABCDH, $n = 768$) and ABCD African American (ABCD A, $n = 1,257$); and 5) a UKB British GWAS with subjects not present in previous GWAS²⁵ (also removed the relatives of previous GWAS subjects, $n = 15,214$). See **Supplementary Table 40** 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^{38,39} to generate 215 mean and PC DTI phenotypes for 21 predefined white matter tracts (**Supplementary Table 41**). 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 tract annotation and imaging procedures is shown in **Supplementary Figures 24-26** and a few image examples are given in **Supplementary Figures 27-30**. For each continuous phenotype or covariate variable, we removed values greater than five times the median absolute deviation from the median value. The ancestry assignment in UKB was based on self-reported ethnic background (Data-Field 21000), whose accuracy was verified in Bycroft, et al.⁵² For ABCD, we assigned ancestry by a combination analysis using self-reported ethnicity and ancestry inference results from SNPweights¹⁴⁵, see **Supplementary Note** for details.

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 analysis⁴³. 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 UKB⁵² (Data-Field 22009). The heritability estimates were tested in one-sided likelihood ratio tests. We performed linear mixed model-based association analysis using fastGWA¹⁴⁶. 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 GWAS²⁵. In addition, GWAS were separately performed on European validation datasets UKBW, ABCDE, HCP, PING, and PNC using Plink¹⁴⁷. In the five 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 analysis. The meta-analysis was then performed on these validation datasets using METAL¹⁴⁸ with the sample-size weighted approach.

We applied a few analyses to support the findings in UKB British discovery GWAS. First, the LDSC⁴⁸ software (version 1.0.0) was used to estimate the pairwise genetic correlation between DTI parameter values in discovery GWAS and the meta-analyzed five European validation GWAS ($n = 6,962$). We used the pre-calculated LD scores provided by LDSC, which were computed using 1000 Genomes European data. We used HapMap3¹⁴⁹ variants and removed all variants in the major histocompatibility complex (MHC) region. In addition, we 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 nine 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 fastGWA: 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 age, gender, and top ten genetic PCs. 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 FINDOR⁵³. Specifically, FUMA first clumped partially independent significant variants, which were variants with a P -value smaller than the predefined threshold and independent of other significant variants ($LD\ r^2 < 0.6$, default value). 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 $MAF \geq 0.0005$. These variants included those from the 1000 Genomes reference panel that may not have been included in the GWAS. Based on these significant variants, independent 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 SuSiE⁴⁵ method via PolyFun⁴⁶ 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.²⁵, 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 slope⁵³ 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 LCV¹⁰⁸ (version 2019-03-14) to explore the genetical causal relationships between DTI parameters and other complex traits. We used meta-analyzed GWAS summary statistics 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 MAGMA¹¹⁰ (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-MAGMA¹¹⁵ (version 2019-11-29) to perform Hi-C coupled gene-based association analysis by integrating Hi-C profiles from fetal and adult brain tissues^{150,151}.

Biological annotations. We performed heritability enrichment analysis via partitioned LDSC¹²¹. Baseline models were included when estimating the enrichment scores for our tissue type and cell type specific annotations. Methods to prepare in-house chromatin data of three glial cell subtypes and two neuronal cell subtypes can be found in the **Supplementary Note**. We performed gene property analysis for the 13 GTEx¹²⁴ 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. In addition, we treated

1 DTI associated genes in MAGMA, H-MAGMA or FUMA analysis as an annotation and
2 tested whether the heritability of other complex traits was enriched in this DTI
3 annotation. MAGMA and DEPICT (version 1 rel194) were separately used to explore the
4 implicated biological pathways. MAGMA gene-set analysis examined 5,500 curated gene
5 sets and 9,996 Gene Ontology (GO) terms from the Molecular Signatures Database¹⁵²
6 (MSigDB, version 7.0) and DEPICT tested 10,968 pre-constructed gene sets using GWAS
7 summary statistics with P -value $< 10^{-5}$ as input. All other parameters were set as default.

9 **Code availability**

10 We made use of publicly available software and tools listed in URLs. Other codes used in
11 our analyses are available upon reasonable request.

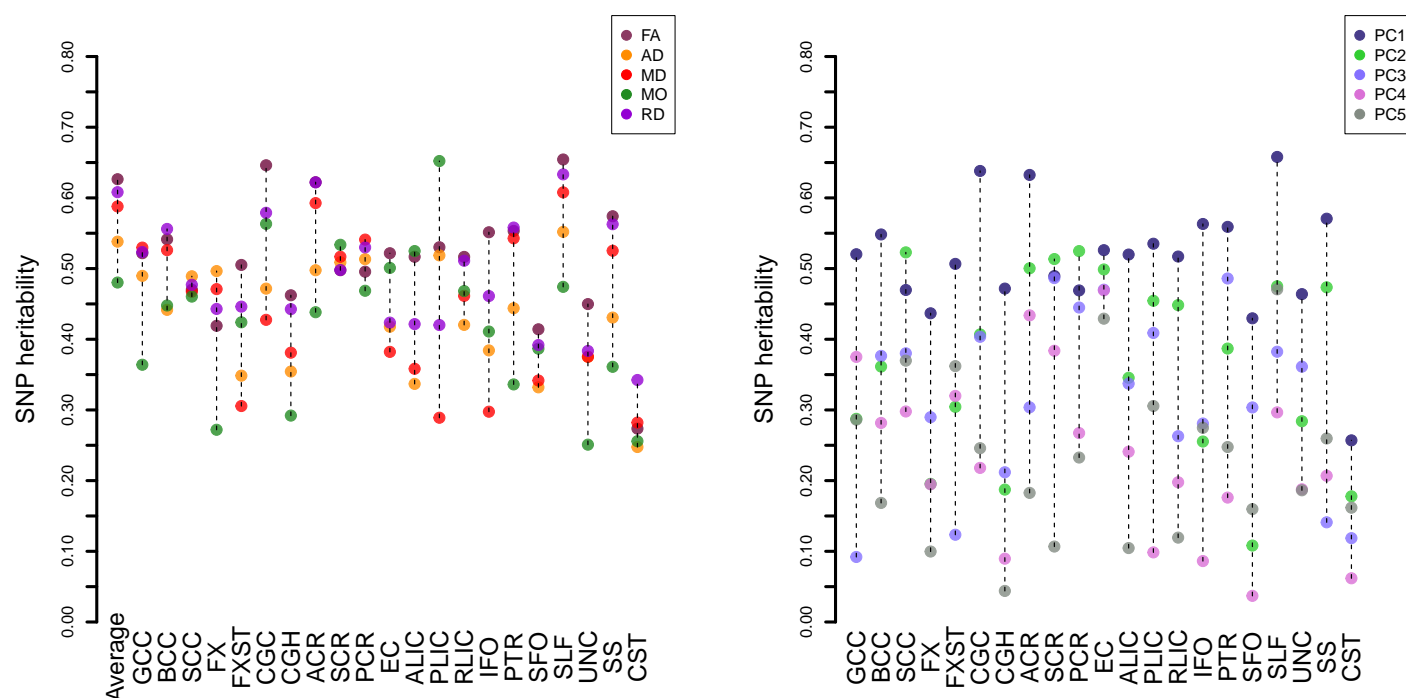
13 **Reporting summary**

14 Further information on research design is available in the Nature Research Reporting
15 Summary.

17 **Data availability**

18 Our GWAS summary statistics have been shared at <https://github.com/BIG-S2/GWAS>.
19 The individual-level raw data used in this study can be obtained from five publicly
20 accessible data resources: UK Biobank (<http://www.ukbiobank.ac.uk/resources/>), ABCD
21 (<https://abcdstudy.org/>), PING (<https://www.chd.ucsd.edu/research/ping-study.html>),
22 PNC (<https://www.med.upenn.edu/bbl/philadelphianeurodevelopmentalcohort.html>),
23 and HCP (<https://www.humanconnectome.org/>). Our results can also be easily browsed
24 through our knowledge portal <https://bigkp.web.unc.edu/>.

a



b

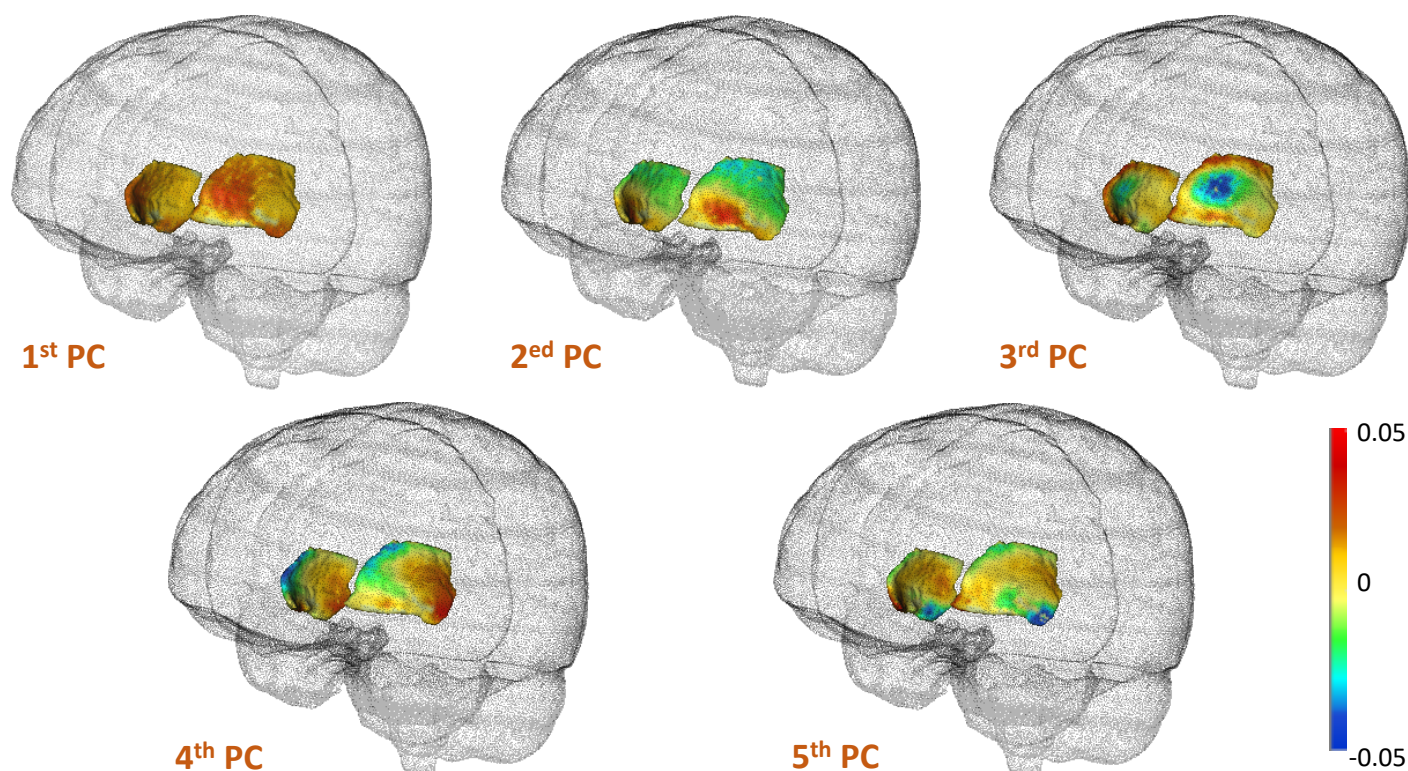


Figure 1: SNP heritability estimates of 215 DTI parameters ($n = 33,292$ subjects) and illustration of the top five FA principal components (PCs) of external capsule (EC). **a) The 110 mean DTI parameters and 105 FA PC DTI parameters are displayed on the left and right panels, respectively. The x-axis lists the names of white matter tracts. **b**) The functional principal component (PC) loading coefficients for the top five FA PCs of EC.**

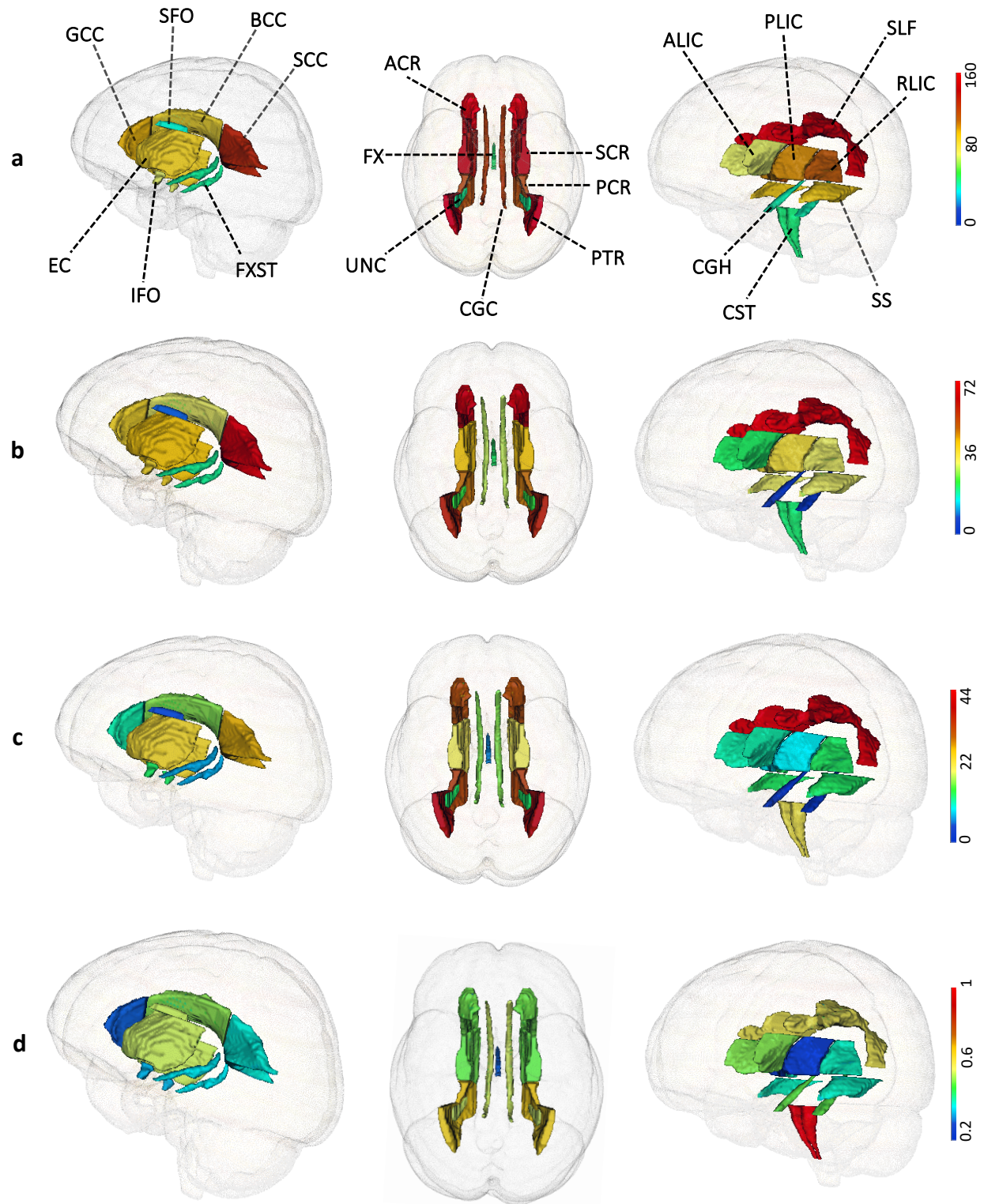


Figure 2: Number of independent significant variants identified in UKB British discovery GWAS at 2.3×10^{-10} significance level ($n = 33,292$ subjects). The first three rows are the number of independent significant variants identified in each white matter tract by **a)** any DTI parameters; **b)** any FA parameters; **c)** FA PC parameters, respectively. The last row **d)** displays the proportion of FA-associated variants that can only be identified by PC parameters.

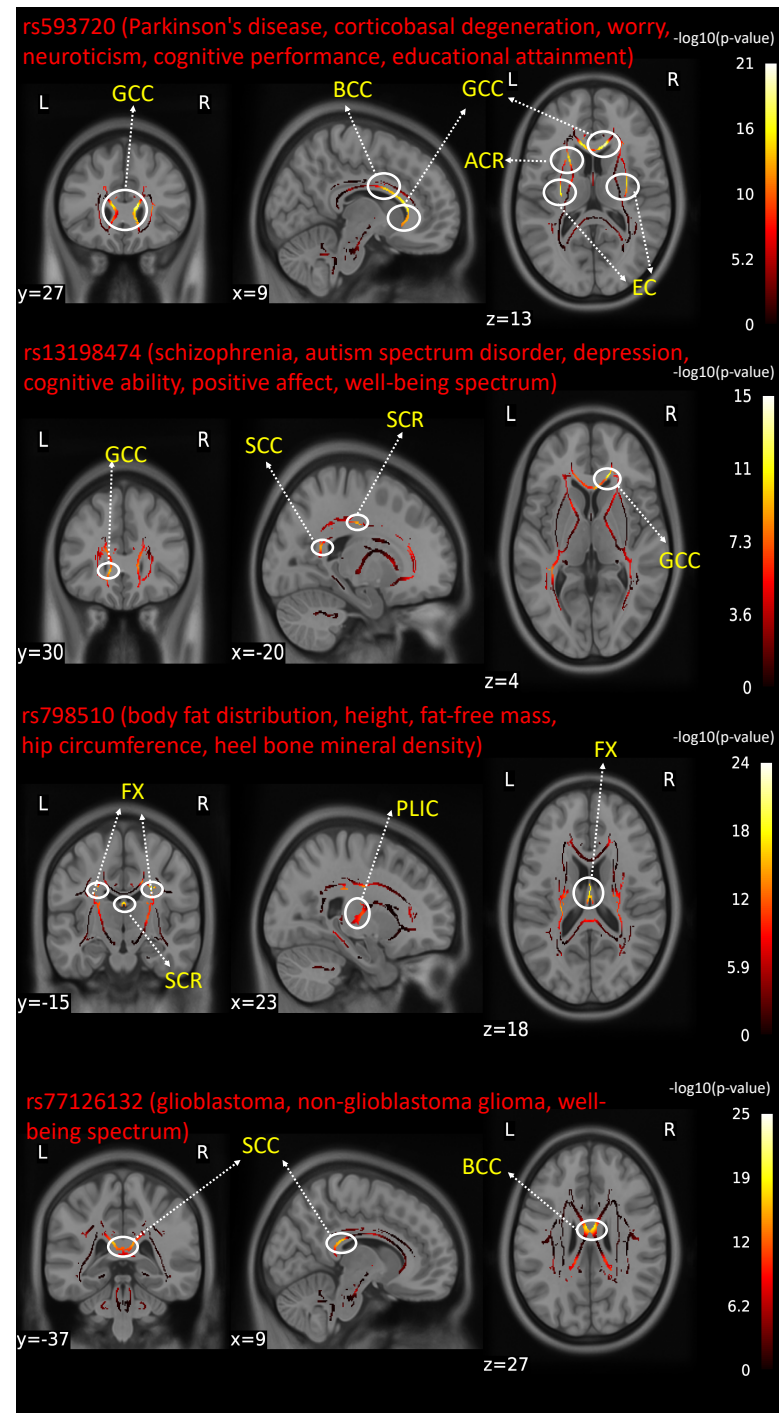
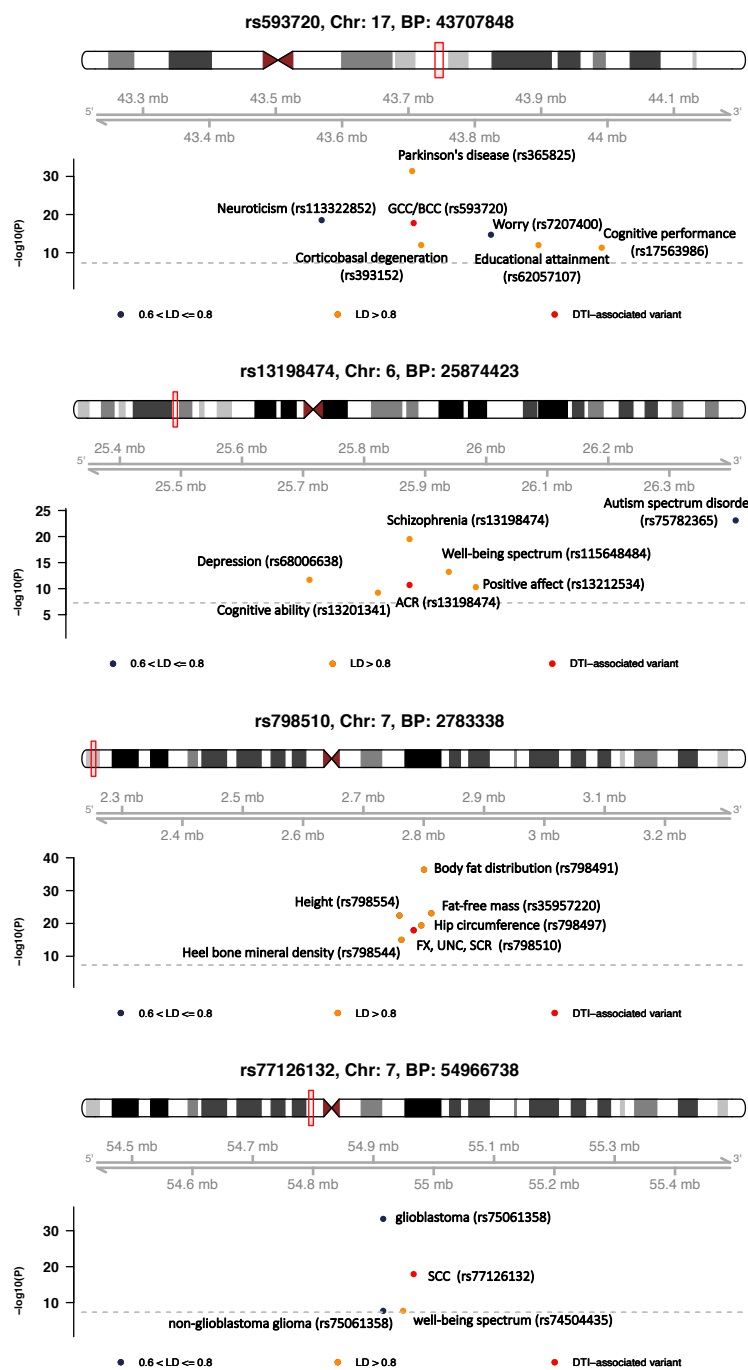


Figure 3: The genomic region and brain spatial map of voxel-wise effect size patterns for four selected pleiotropic variants (n = 33,292 subjects). We labeled previously reported GWAS variants for other complex traits in genomic regions influencing white matter microstructure (left). In spatial maps (right), we illustrate voxel-wise effect sizes of pleiotropic variants in white matter tracts.

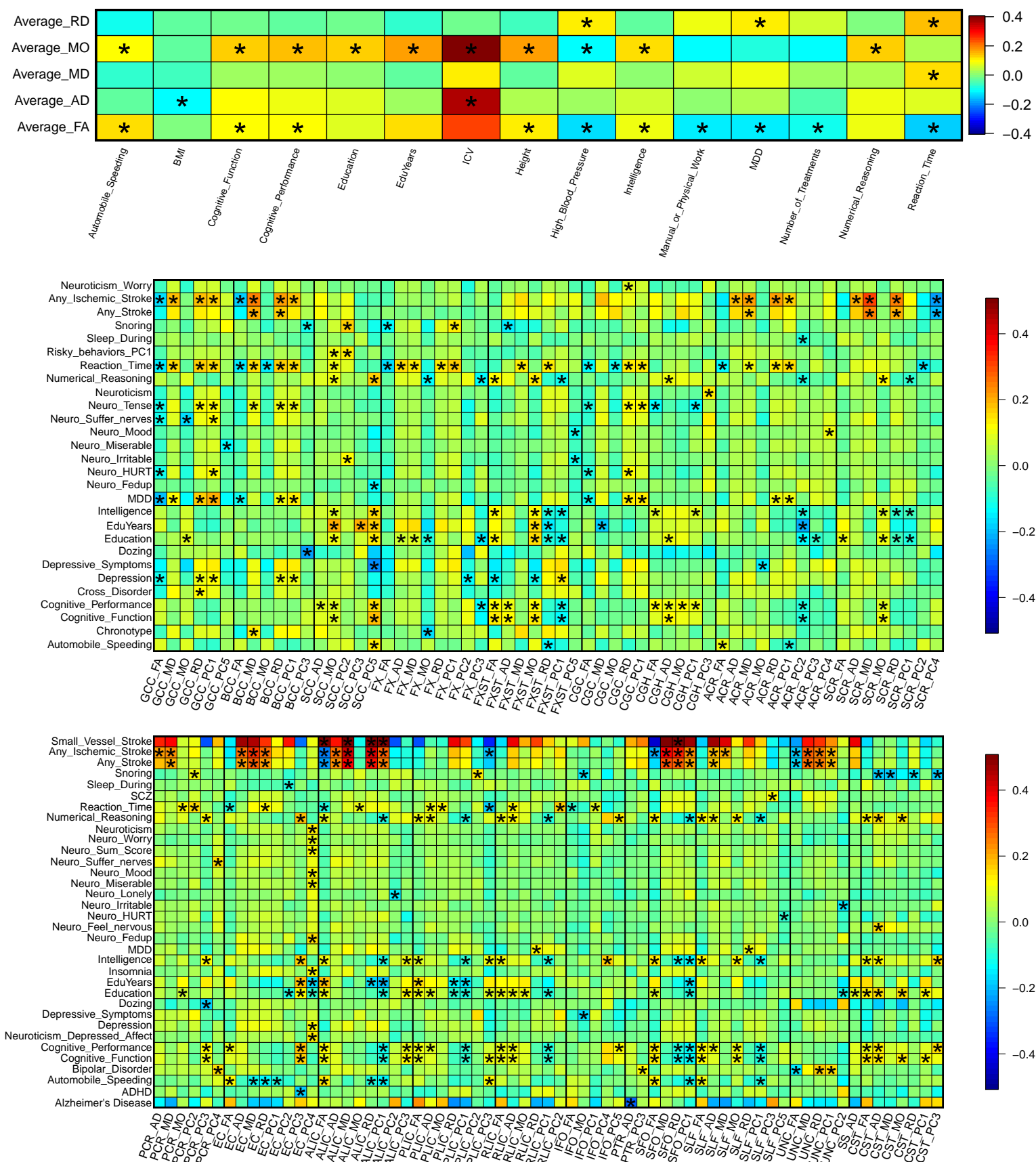


Figure 4: Selected pairwise genetic correlations between white matter microstructure and other complex traits (n = 40,254 subjects). We adjusted for multiple testing by the Benjamini-Hochberg procedure at 0.05 significance level (215 × 76 tests), while significant pairs are labeled with stars. Sample size and detailed information of complex traits can be found in Supplementary Table 18.

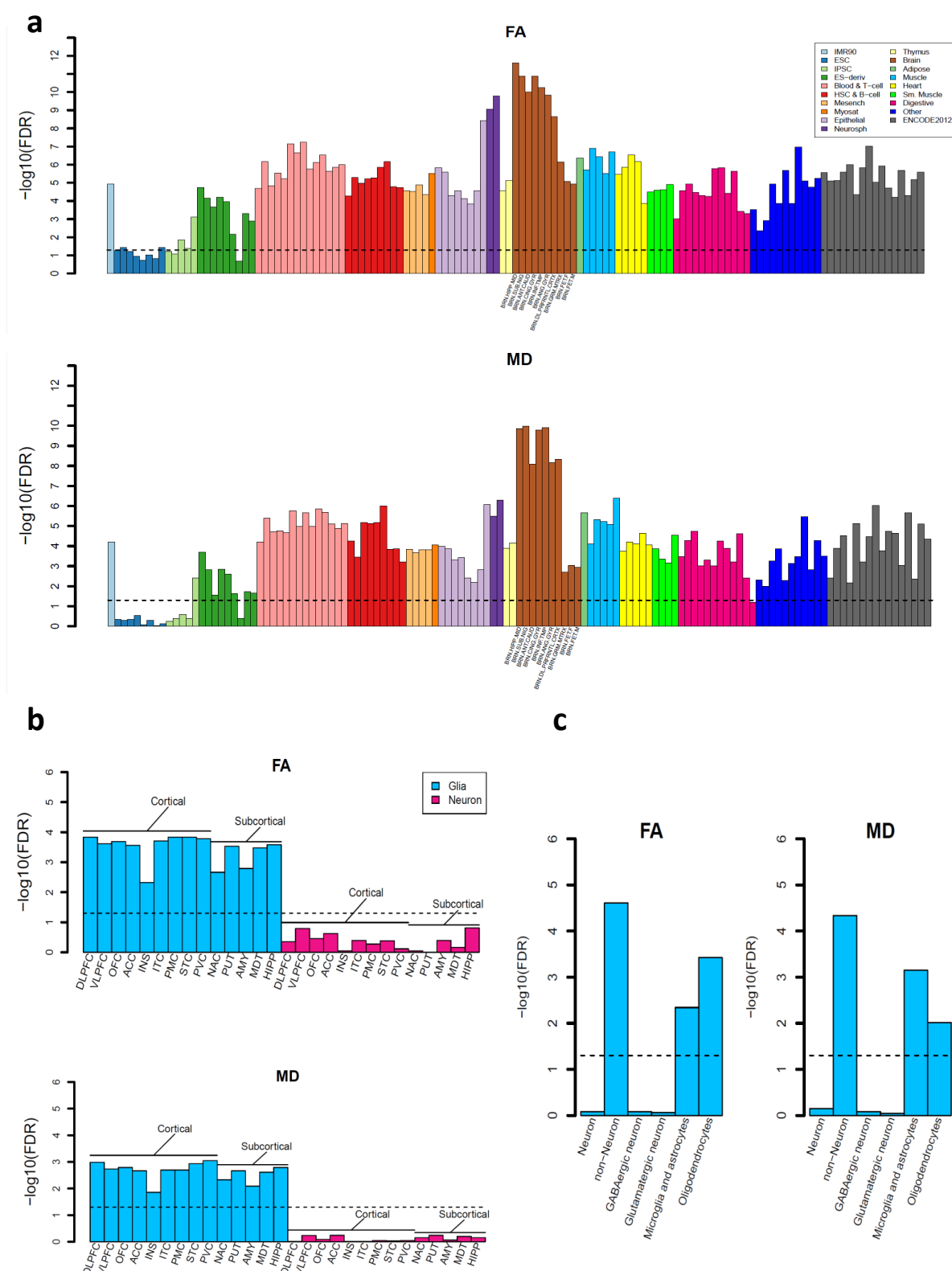


Figure 5: Partitioned heritability enrichment analysis (n = 33,292 subjects). **a)** Heritability enrichment in regulatory elements across tissues and cell types. Brain tissues are labelled in x-axis. **b)** Heritability enrichment in regulatory elements of two brain cell types (neuron and glia) sampled from 14 brain regions. **c)** Heritability enrichment in regulatory elements of glial cell subtypes (non-neuron, including oligodendrocyte and microglia & astrocyte) and neuronal cell subtypes (neuron, including GABAergic and glutamategic neurons).