

## Transcriptomic and macroscopic architectures of intersubject functional variability in human brain white-matter

Jiao Li<sup>1,2,3</sup>, Guo-Rong Wu<sup>4</sup>, Bing Li<sup>1,3</sup>, Feiyang Fan<sup>1,3</sup>, Xiaopeng Zhao<sup>1,3</sup>, Yao Meng <sup>1,3</sup>, Peng Zhong<sup>1,3</sup>, Siqi Yang<sup>1,3</sup>, Bharat B. Biswal<sup>1,3,5</sup>, Huafu Chen <sup>1,2,3</sup>✉ & Wei Liao <sup>1,3</sup>✉

Intersubject variability is a fundamental characteristic of brain organizations, and not just “noise”. Although intrinsic functional connectivity (FC) is unique to each individual and varies across brain gray-matter, the underlying mechanisms of intersubject functional variability in white-matter (WM) remain unknown. This study identified WMFC variabilities and determined the genetic basis and macroscale imaging in 45 healthy subjects. The functional localization pattern of intersubject variability across WM is heterogeneous, with most variability observed in the heteromodal cortex. The variabilities of heteromodal regions in expression profiles of genes are related to neuronal cells, involved in synapse-related and glutamic pathways, and associated with psychiatric disorders. In contrast, genes over-expressed in unimodal regions are mostly expressed in glial cells and were related to neurological diseases. Macroscopic variability recapitulates the functional and structural specializations and behavioral phenotypes. Together, our results provide clues to intersubject variabilities of the WMFC with convergent transcriptomic and cellular signatures, which relate to macroscale brain specialization.

<sup>1</sup>The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Laboratory for Neuroinformation, University of Electronic Science and Technology of China, Chengdu 611731, P.R. China. <sup>2</sup>The Center of Psychosomatic Medicine, Sichuan Provincial Center for Mental Health, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu 611731, P.R. China. <sup>3</sup>School of Life Science and Technology, Center for Information in BioMedicine, University of Electronic Science and Technology of China, Chengdu 611731, P.R. China. <sup>4</sup>Key Laboratory of Cognition and Personality, Faculty of Psychology, Southwest University, Chongqing 400715, P.R. China. <sup>5</sup>Department of Biomedical Engineering, New Jersey Institute of Technology, Newark, NJ 07103, USA. ✉email: [chenhf@uestc.edu.cn](mailto:chenhf@uestc.edu.cn); [weiliao.wl@gmail.com](mailto:weiliao.wl@gmail.com)

**B**lood-oxygenation-level-dependent-functional magnetic resonance imaging (BOLD-fMRI) has become the method of choice for evaluating the coherent signal fluctuations across the brain, because it has high spatial and temporal resolutions and is noninvasive<sup>1–3</sup>. This intrinsically inter-regional functional connectivity (FC) architecture is unique among individuals<sup>4–7</sup> and reflects cognitive or population variabilities<sup>8,9</sup>. A higher intersubject variability of gray-matter (GM) FC reliably exists in higher-order cognitive networks, and lower intersubject variability exists in lower-order perceptual networks, suggesting intersubject variability gradients<sup>10–12</sup>. There is intersubject variability in cortical organization, even after accurate brain structure alignment of brain functions to structures<sup>8</sup>. Although white-matter (WM) comprises half of the human brain, its variabilities in connectome organization across individuals have been relatively unexplored<sup>13,14</sup>.

Brain WM signals measured using BOLD-fMRI provide functional information about intrinsic activity and can be used to characterize its connectivity<sup>15,16</sup>. WM functional networks that are aligned and interact across long-distance WM tracts<sup>17</sup> and functional organization of human corpus callosum that is closely adjacent to resting-state GM networks<sup>18</sup>, have highlighted the functional roles of WM. In a recent study, WM functional topography (e.g., small-worldness and nonrandom modularity) was computed and its reliability and reproducibility demonstrated, suggesting that WM functional connectivity (WMFC) cannot be simply attributed to noise<sup>19</sup>. In addition, it has been reported that intersubject variation in WMFC can be used to estimate individual general fluid intelligence<sup>20</sup>. Therefore, WMFC is unlikely to be generated by noise<sup>21</sup>, and may provide important information that contributes to the understanding of the underlying mechanisms of individual variabilities in cognition and behavior.

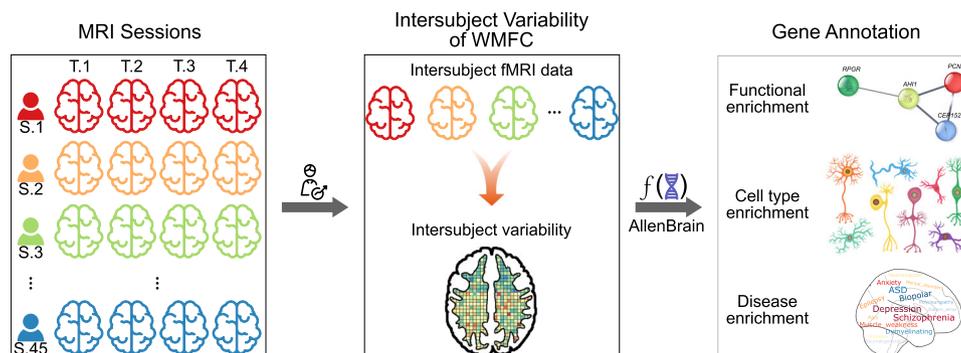
The function of WM often overlaps with and is constrained by the physical substrate's known anatomical pathways/microstructures<sup>22</sup>. WM microstructure is a potential predictor of intersubject variations, because it is under strong genetic control. The WM associated genes regulate pathways involved in brain disease pathogenesis and neurodevelopmental processes<sup>23</sup>. Several genome-wide association studies<sup>24–27</sup> have been conducted to identify loci associated with interindividual variations in WM microstructure, but these studies have been impeded by spatial specificities, which have been the major limitation<sup>28</sup>. In this study, we used an approach for combining neuroimaging and spatial patterned gene expression to investigate the molecular mechanisms of individual variabilities in FC architecture<sup>29,30</sup>. The Allen Human Brain Atlas (AHBA) microarray

dataset was used to identify transcriptomes associated with human neuroimaging with multi-modal evidence, suggesting a link between conserved gene expression and relevant functional circuitries<sup>31–33</sup>. To the best of our knowledge, no study has previously characterized intersubject variabilities in WMFC-related genes using transcriptome-neuroimaging association analyses, therefore, the genetic influences on intersubject variabilities in WMFC are, thus far, unknown.

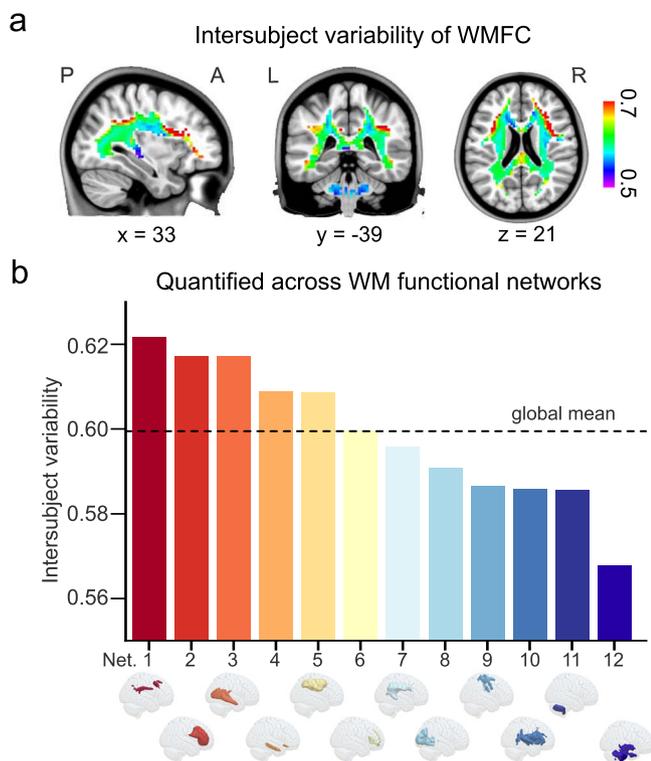
Our objective was to identify the intersubject variability of WMFC-related transcriptional architecture among individuals. Considering intersubject variabilities in functional organizations, researchers have further characterized brain networks responsible for the etiologies of disorders<sup>34–36</sup>. More recently, Sun et al.<sup>37</sup> have reported alterations of intersubject variabilities of the FC in schizophrenia, suggesting potential implications for understanding the high clinical heterogeneity of this disorder. We therefore hypothesized that this information would increase our understanding of the genetic causes of intersubject variabilities of WMFC in healthy subjects, and may be associated with psychiatric and neurological brain disorders. We collected resting-state BOLD-fMRI data on 45 healthy subjects, and each scanned four times over ~6 months. This unique data set allowed us to assess the interindividual variabilities in WMFC, which controlled intrasubject variabilities<sup>10,38</sup>. We determined the following: (i) whether intersubject variabilities in WMFC were hierarchically organized, and whether the degree of variability varied from unimodal to heteromodal association networks; (ii) how individual variabilities in WMFC were affected by the transcriptome and specific cell types; (iii) whether clinical significance could be calculated by the overlap between intersubject variability-related genes and those involved in diseases; and (iv) whether the intersubject variability was related with known macroscale indices, including structural and functional organizations. Our findings will facilitate the study of both the spatial distribution of intersubject variabilities in WMFC and the multiscale associations in functional variabilities in brain WM across subjects. The data generation and analysis schematic are shown in Fig. 1.

## Results

**Study design and image quality assessment.** The current work used a hybrid design (i.e., within + between design) in which scans were repeated one or more times on the same day and across one or more sessions<sup>39</sup>. All subjects participated in four sessions (each scanning included two runs) over ~6 months (Supplementary Result 1 and Supplementary Fig. 1). To measure



**Fig. 1 Data generation and analysis pipeline.** This study was performed using neuroimaging and six AHBA transcriptomic data. Gene expression maps were extracted from NeuroSynth (<https://www.neurosynth.org>). Neuroimaging data were collected from 45 subjects who underwent four sessions within approximately 6 months. After establishing the relationship between intersubject variability in white-matter functional connectivity and gene expression, the variability-related highly-ranked gene set was identified. Enrichment analysis of the gene set was used to identify the functions of these genes. The identified intersubject variability-related genes were annotated by functional, cell types, and disease enrichments.



**Fig. 2** Nonuniform distribution of intersubject variability of white-matter functional connectivity (WMFC). **a** Variability distribution across voxels.

Intersubject variability was quantified at each white-matter (WM) voxel across subjects after correcting for intrasubject variability and technical noise. **b** Intersubject variation in WMFC at the network level based on previous parcellation of the WM into 12 functional networks<sup>17</sup>. Network (Net) 1, frontoparietal control network; Net 2, deep frontal network; Net 3, inferior longitudinal fasciculus system; Net 4, temporal-orbitofrontal network; Net 5, dorsal frontoparietal network; Net 6, forceps minor system; Net 7, superior longitudinal fasciculus system; Net 8, visual network; Net 9 & 10, sensorimotor networks; and Net 11 & 12, cerebellum networks. Source data provided as Source Data file.

the quality of the resting-state BOLD-fMRI data, we calculated a series of common image quality metrics using MRIQC<sup>40</sup>. Data were first visually inspected for the whole brain field of view coverage, signal artifacts, and proper alignment to the corresponding anatomical images. The image quality metrics for the spatial information included Entropy Focus Criterion, Foreground to Background Energy Ratio, Signal-to-noise ratio (SNR). The image quality metrics for the temporal information included Temporal SNR (tSNR) and Mean Fractional Displacement (mean FD; Supplementary Result 2 and Supplementary Fig. 2).

**Nonuniform distribution of intersubject variability.** To study the organizational patterns of WMFC, we first estimated voxel-wise interindividual variabilities in the brain WM after controlling for intrasubject variability and technical noise (Supplementary Result 3 and Supplementary Fig. 3). We found that FC interindividual variability exhibited a nonuniform spatial distribution across brain WM voxels (Fig. 2a).

We further assessed the variabilities of the WMFC organization within each of the specific brain WM functional networks<sup>17</sup>. The hierarchy of the functional variations in WM across 12 specific networks was similar to that obtained from the prior gray-matter FC results<sup>10–12,41</sup>. Specifically, the frontoparietal control and default mode networks showed high levels of functional variability, while the sensorimotor, and visual networks exhibited low interindividual

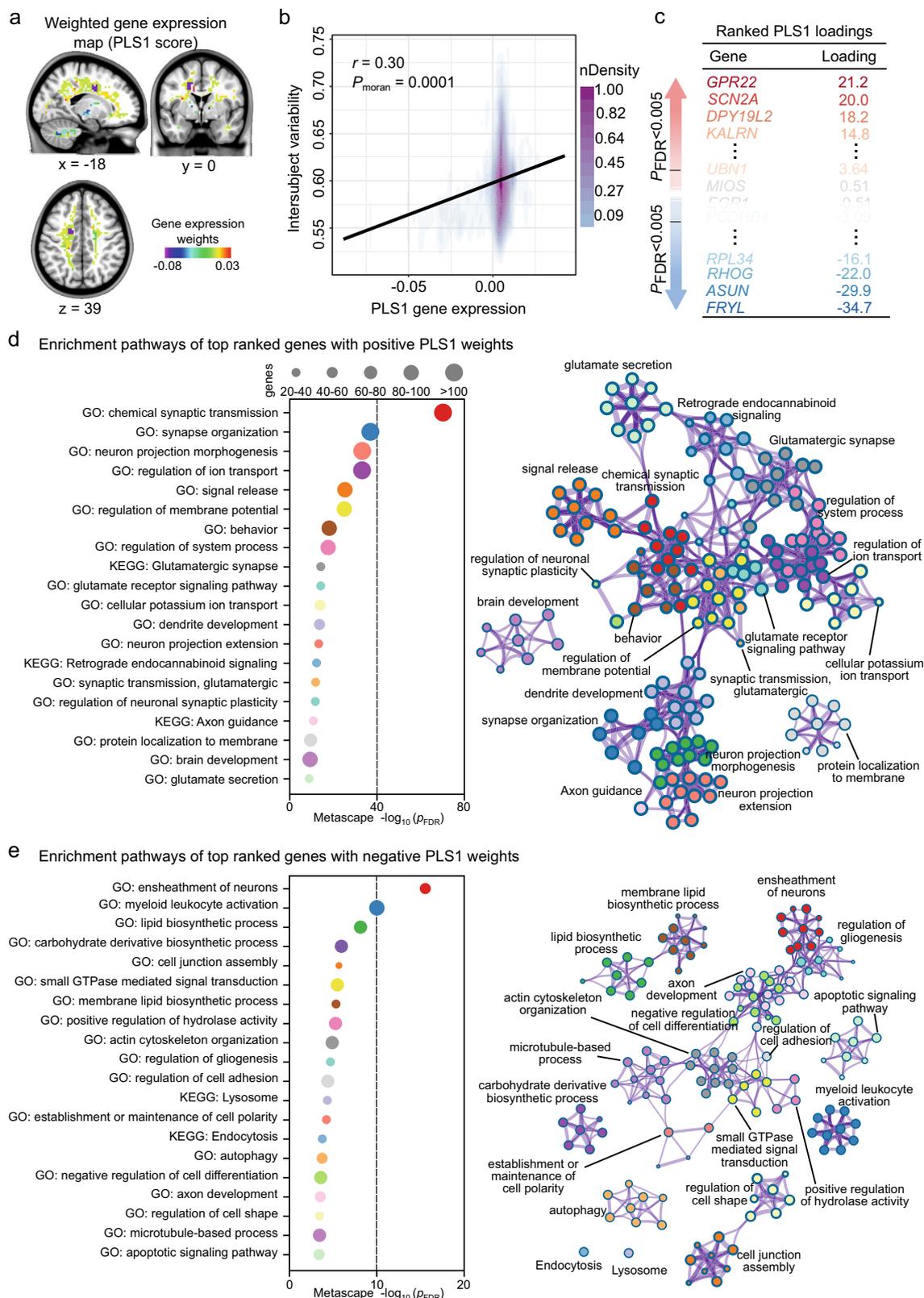
variations (Fig. 2b). In contrast to the gray-matter FC variations, the lowest variation was found in the visual and sensorimotor networks, where the cerebellum network showed the lowest level of functional variability in WM.

**The gene expression patterns associated with intersubject variability of WMFC.** We used the AHBA, a whole-brain transcriptomic dataset, to obtain patterns of gene expression in the brain for examining the transcriptomes associated with WMFC intersubject variability (Supplementary Result 4 and Supplementary Table 1). We aligned the nonuniform distribution map to the atlas of gene expression for ~9000 genes in the adult human cortex available in the AHBA dataset. Gene expression maps were obtained from the Neurosynth-Gene database (Supplementary Result 5). We performed partial least square (PLS) correlation analyses to identify the dominant gene expression patterns of functional variability (Fig. 3a). PLS correlation analyses ranked all ~9000 genes by their multivariate correlation with the variability distributions, resulting in one ranked gene list. These findings were corrected for spatial autocorrelation using the Moran Spectral Randomization (MSR)<sup>42</sup> implemented in BrainSpace<sup>43</sup>, which generated spatially constrained null models for irregularly spaced data. Clear spatial correlation was evident between the first PLS (PLS1) component and the intersubject variability of WMFC ( $r = 0.30$ ,  $P_{\text{moran}} = 0.0001$ ; Fig. 3b), with high expressions (red areas) in the frontal networks and lower expressions (blue) in the cerebellum areas. In the gene list, expression patterns that were more strongly correlated with the corresponding intersubject variability map had large positive and negative PLS weights, and therefore, occupied more extreme ranks (Fig. 3c). Genes with strongly positive PLS weights showed positive spatial correlations between their expressions and variable WMFC organizations, and vice versa.

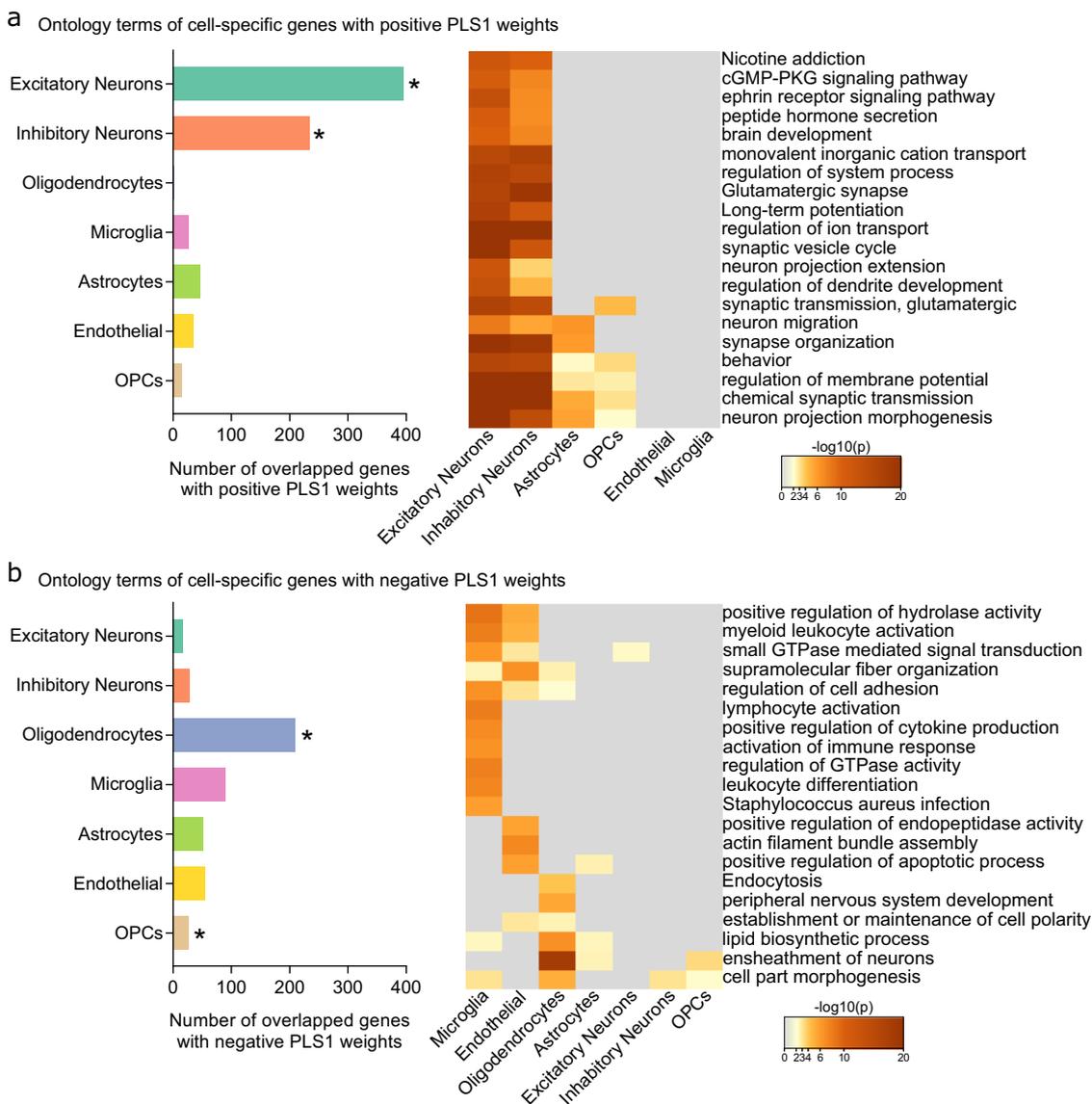
To investigate whether molecular biological signatures influenced the variable WMFC organization, we aligned the gene ontology (GO) biological processes and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways with the positive and negative PLS1 gene lists, respectively, using Metascape<sup>44</sup>. The top-ranked genes with positive PLS1 (PLS1+) weights were significantly enriched for several synapse-related terms and glutamatergic pathways (Fig. 3d), such as “chemical synaptic transmission”, “synapse organization”, “Glutamatergic synapse”, and “glutamate secretion”. However, the top ranked genes with negative PLS1 (PLS1-) weights were significantly enriched for glial cells-related pathways (Fig. 3e), including “ensheathment of neurons”, “regulation of gliogenesis”, and “axon development”. All these enrichment pathways were corrected with the false discovery ratio (FDR) with  $P < 0.05$ .

**Highly ranked genes expressed in specific cell types.** The above findings indicated that the intrinsic transcriptomic organizations across the WM in healthy subjects were correlated with intersubject variabilities of WMFC. However, brain-wide transcriptional variations may reflect bulk samples of cells such as neurons, oligodendrocytes, astrocytes, microglia, endothelial cells, and oligodendrocyte precursors (OPCs)<sup>45–47</sup>. Therefore, we predicted that variable WMFC organization may be related to the expression patterns of variability-related genes across different cell types with varying spatial distributions. Because spatially comprehensive maps of cell type densities across the human brain were not available, we used previously defined cell-class gene sets to test whether the observed intersubject variabilities of WMFC were organized to broad cell-classes in the human WM<sup>45</sup>.

We showed that the top ranked genes in the PLS1 + gene list were expressed primarily in neurons, both inhibitory (234 genes,



**Fig. 3 Transcriptomic decoding of the intersubject variability of white-matter functional connectivity (WMFC).** **a** Weighted gene expression map of the first component of partial least squares (PLS1). The combination of genes, the most explained variance of intersubject variability in WMFC, was obtained by the PLS correlation method. **b** Scatterplot of the relationship between PLS1 scores and intersubject variability in WMFC. **c** Ranked PLS1 loading. Genes that were strongly positively weighted on PLS1 (e.g., *GPR22*) correlated positively with intersubject variability, whereas genes that were strongly negatively weighted on PLS1 (e.g., *FRYL*) correlated negatively with intersubject variability. Functional enrichment analysis of top ranked genes (from **c**) with PLS1+ (**d**) and PLS1- (**e**) weights. All terms were retained after  $P < 0.05$ , false discovery ratio-corrected. nDensity, density estimate, scaled to a maximum of 1. Source data provided as Source Data file.

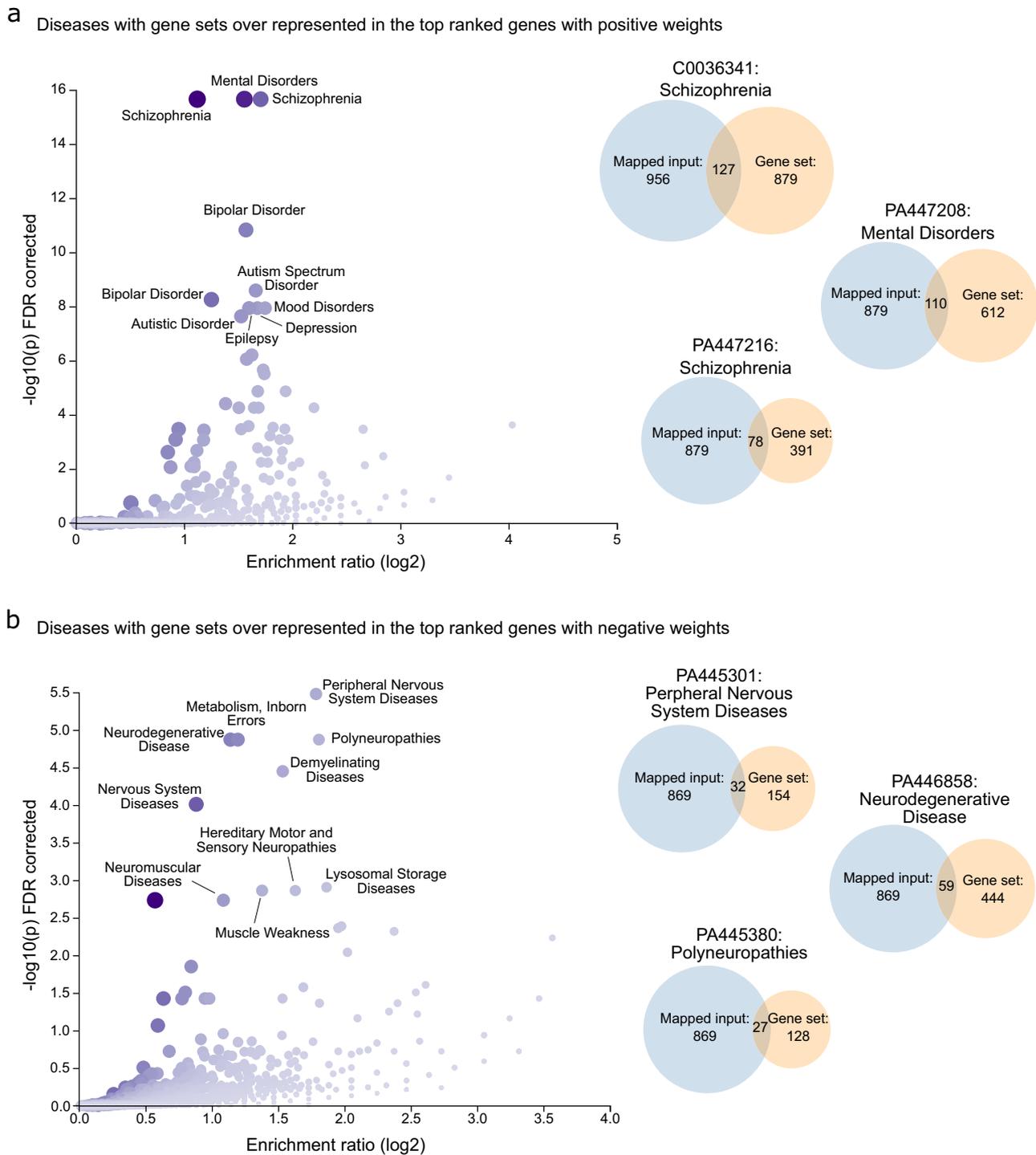


**Fig. 4 Cell type specificity of intersubject variability of white-matter functional connectivity (WMFC).** Enrichment pathways of cell-specific highly-ranked genes with PLS1+ (a) and PLS1- (b) weights. Left panel is the number of highly-ranked genes in each cell type. Right panel is the enrichment analysis in each cell type. A key of heatmap indicates the statistically enriched terms in each cell type. \*,  $P_{perm} < 0.05$ , false discovery rate-corrected. Source data provided as Source Data file.

$P_{perm} = 0.0003$ , FDR-corrected) and excitatory (395 genes,  $P_{perm} = 0.0003$ , FDR-corrected) (Fig. 4a). Genes specific to oligodendrocytes (209 genes,  $P_{perm} = 0.0007$ , FDR-corrected) and OPCs (27 genes,  $P_{perm} = 0.0007$ , FDR-corrected) were also overrepresented in the top ranked PLS1- gene list (Fig. 4b). Confirming our approach, enrichment analyses using cell type specific genes in the PLS1+ gene list showed that intersubject variability of WMFC was significantly enriched for biological processes associated with neuronal cells. These were enriched for GO terms such as “Glutamatergic synapse”, “chemical synaptic transmission”, “synaptic vesicle cycle”, “synaptic transmission, glutamatergic”, and “synapse organization” (Fig. 4a). In a similar manner, enrichment pathways identified in cell type genes in the PLS1- gene list were usually concentrated in glial cells (Fig. 4b). Notably, these enrichments were more abundant in microglia, endothelial cells, and oligodendrocytes, suggesting a more pronounced physiological mechanism in these cells than that of the astrocytes and OPCs. Together, these results supported the validity of our indirect approach in assigning gene expressions to unique cell types, and

allowed us to detect specific cell types encoded in variable WMFC organizations.

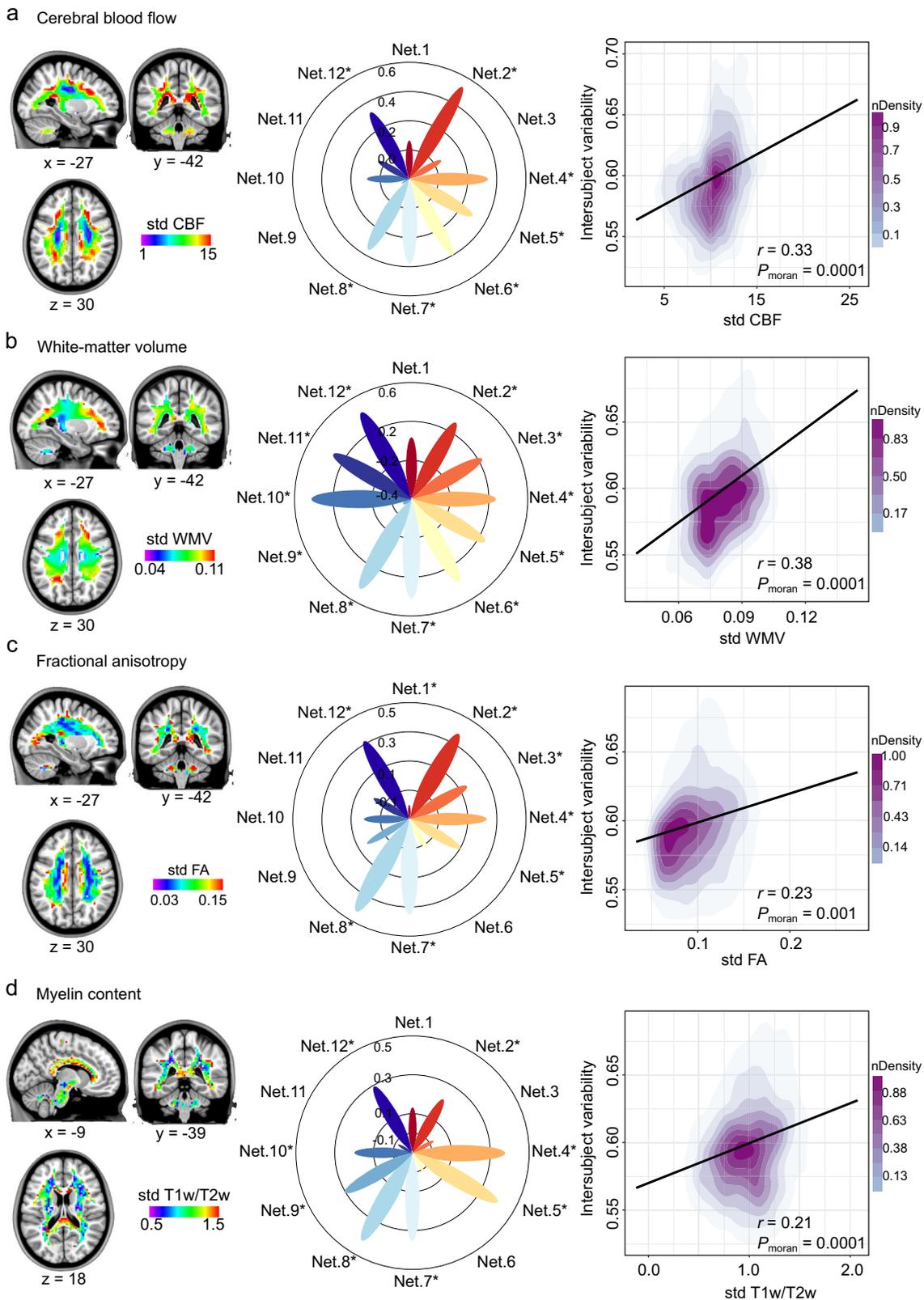
**Genes enriched for brain diseases.** To further characterize the importance of the putative variability of genes within the top ranked PLS1+ and PLS1- gene sets, we tested the hypothesis that genes related to brain WM disorders were enriched among genes associated with variable WMFC organizations. Using WebGestalt<sup>48</sup>, we combined Over-Representation Analysis (ORA) and Gene Set Enrichment Analysis (GSEA) of Disease Ontology (DO) terms based on DisGeNET, OMIM, and GLAD4U databases in the top-ranked PLS1+ and PLS1- gene sets. We found a significant overlap between the top-ranked PLS1+ gene list and genes associated with psychiatric disorders, such as schizophrenia, bipolar disorder, autism spectrum disorder, and depression, all of which were FDR-corrected with  $P < 0.05$  (Fig. 5a; Supplementary Result 6 and Supplementary Table 2). The results were consistent with previous studies that reported functional abnormalities of WM in major depressive



**Fig. 5 Highly-ranked genes enriched for brain diseases.** Disease-association analysis in top ranked genes with PLS1+ (a) and PLS1− (b) weights. Volcano plots show disorders with gene sets that were overexpressed in top ranked genes of PLS1. Only the top 10 terms were annotated ( $P_{FDR} < 0.05$ ). Venn diagrams show the top three overlapped diseases between the top gene set and genes sets for diseases. Source data provided as Source Data file.

disorders<sup>49</sup>, schizophrenia<sup>50,51</sup>, and autism spectrum disorder<sup>52</sup>. The PLS1− ranking genes were also enriched for neurological diseases, including peripheral nervous system diseases, neurodegenerative disease, polyneuropathies, demyelinating diseases, hereditary motor and sensory neuropathies, muscle weakness, and neuromuscular diseases (all FDR-corrected  $P < 0.05$ ; Fig. 5b; Supplementary Table 3). These results were accompanied by enrichment pathways in cell type specific genes, indicating the different etiological mechanisms in PLS1+ and PLS1− gene sets

and the potential clinical applications of intersubject variability in WMFC. Considering multiple sclerosis as a typical disorder with abnormalities in WM, we also determined the relationships between WMFC intersubject variabilities and differentially expressed genes in multiple sclerosis<sup>53</sup>. We found that WMFC intersubject variability was correlated with dysregulated gene expression ( $r = -0.29$ ,  $P_{\text{moran}} = 0.0002$ ; Supplementary Result 7 and Supplementary Fig. 4), suggesting that multiple sclerosis involved differential impacts to areas along the WM distribution.



**Relationship between macroscale architecture and intersubject variability of WMFC.** At the macroscale level, we sought to understand how variations of WMFC related to the hierarchical neurobiological organization of brain anatomical topography and the corticocortical connectivity pattern. Cerebral blood flow (CBF) regulation, with individual variability, is essential for brain function<sup>8,54</sup>. The frontoparietal control regions exhibit higher

variability in CBF (Fig. 6a). Higher WMFC variability has been spatially associated with the higher variable of CBF ( $r = 0.33$ ,  $P_{\text{moran}} = 0.0001$ , FDR-corrected). Specifically, the deep frontal WM network, temporal-orbitofrontal network, default mode network, dorsal attention network, superior longitudinal fasciculus system, visual superficial WM system and cerebellum system exhibited spatial correlations. These results indicated that the

**Fig. 6 Intersubject variability of white-matter functional connectivity (WMFC) aligning with macroscopic imaging.** **a** The distribution of intersubject variability of cerebral blood flow (CBF), and the relationship of intersubject variability between WMFC and CBF. **b** The distribution of intersubject variability of white-matter volume (WMV), and the relationship of intersubject variability between WMFC and WMV. **c** The distribution of intersubject variability of fractional anisotropy (FA), and the relationship of intersubject variability between WMFC and FA. **d** The distribution of intersubject variability of myelin content, and the relationship of intersubject variability between WMFC and myelin content. The left panel is the distribution of intersubject variability in each property. The middle panel is the association between functional variability in WM and variability of each neuroimaging index at the network level. Right panel shows the association between functional variability in WM and variability of each neuroimaging index at the whole brain WM level. The asterisk represents that WMFC intersubject variability correlated with the neuroimaging index variability within the WM functional network. Source data provided as Source Data file.

association regions with the greatest variabilities of WMFC involved a disproportionate degree of CBF across subjects.

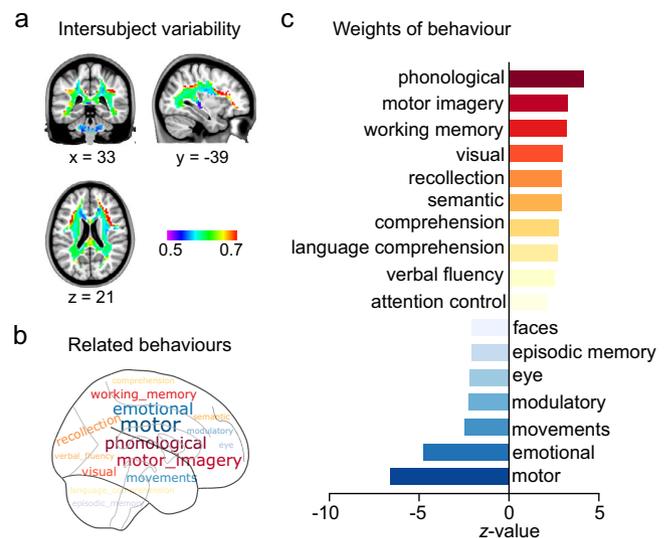
Because almost half of the brain volume is WM, we computed the variability of WM volumes<sup>55</sup>, which was spatially correlated with the variability of WMFC in whole ( $r = 0.38$ ,  $P_{\text{moran}} = 0.0001$ , FDR-corrected) and 11 of 12 networks, except for the frontoparietal control network (Fig. 6b).

Fractional anisotropy (FA) reflects fiber organization and axonal diameter in WM<sup>56,57</sup>. The structural variability of WM was consistent with the previous study reporting that superficial WM exhibited higher variability, whereas deep WM showed lower variability (Fig. 6c)<sup>58</sup>. Analysis of the association between structural variability of WM findings revealed a positive correlation in the default mode network, deep frontal WM network, temporal-orbitofrontal network, dorsal attention network, superior longitudinal fasciculus system, visual superficial WM system, and cerebellum system. When quantified on the whole brain surface, structural variability in WM showed a moderate but significant spatial correlation with the variability of WMFC ( $r = 0.23$ ,  $P_{\text{moran}} = 0.001$ , FDR-corrected; Fig. 6c).

The T1w/T2w mapping has been proposed as a reliable method to measure the myelin content<sup>59</sup> and appears related to the anatomical hierarchy in brain GM architecture<sup>60</sup>. In this study, visual and sensorimotor networks in the WM exhibited higher variable T1w/T2w ratios (Fig. 6d). We also assessed the relationships between variable WMFC organizations and variable T1w/T2w maps. We estimated the spatial correlations between variable T1w/T2w and WMFC and found that the deep frontal WM network, temporal-orbitofrontal network, default mode network, dorsal attention network, superior longitudinal fasciculus system, visual network, sensorimotor network, and cerebellum system exhibited correlations. Notably, a moderate positive correlation was also observed on a whole voxel level ( $r = 0.21$ ,  $P_{\text{moran}} = 0.0001$ , FDR-corrected).

The consistent correlations of brain organizations between variabilities of the four neuroimaging indices and WMFC intersubject variability involved the deep frontal WM network, the temporal-orbitofrontal network/default mode network, the dorsal attention network, the superior longitudinal fasciculus system, and the visual network. The specificity of correlations in some brain WM functional organizations among the neuroimaging indices might indicate the different roles of brain properties in WMFC intersubject variability<sup>8</sup>.

We also assessed the short-range (i.e., local) and long-range (i.e., distant) corticocortical connectivities at each voxel tested for associations with intersubject variabilities of WMFC, as previously reported<sup>10</sup>. The percentage of local connectivity showed a moderate but significant correlation with the intersubject variability of WMFC ( $r = -0.24$ ,  $P_{\text{moran}} = 0.02$ , FDR-corrected) across the entire voxels in WM (Supplementary Result 8 and Supplementary Fig. 5a), while the percentage of distant connectivity was not correlated with the intersubject variability of WMFC ( $r = -0.03$ ,  $P_{\text{moran}} = 0.43$ ; Supplementary Result 8 and Supplementary Fig. 5b).



**Fig. 7 Functional variability in white-matter functional connectivity (WMFC) highly predictive of individual differences in cognitive traits.**

**a** Intersubject variability of WMFC. **b** Distribution of intersubject variability-related behaviors. **c** Weights of behavior. Source data provided as Source Data file.

**The role of intrinsic functional variations in behaviors.** To determine whether the variable WMFC organizations (Fig. 7a) were associated with behavioral performances, we performed a NeuroSynth meta-analyses. The relationships between cognition and interindividual variabilities is shown in Fig. 7b. Notably, we found a positive correlation between the functional variation and higher-order cognitive qualities (Fig. 7c), such as “phonological,” “motor imagery,” and “working memory”. Functional variability was negatively associated with lower-perceptual behavioral measures. Overall, these results supported the inference that inter-individual WMFC variabilities had important implications for cognitive domains.

**Reproducibility of intersubject variabilities in WMFC.** Because PLS is a model usually used for a predictive approach<sup>61</sup>, we validated the intersubject variability of WMFC based on the trained PLS predictive model. We used the variability data from four sessions, and estimated the intersubject variability in each session based on two runs. We then trained the PLS model on the first three sessions and tested it on the fourth session in a leave-one-out cross-validation scheme. We found that the intersubject variability based on the fourth session correlated with the predictive intersubject variability based on the first three sessions ( $r = 0.88$ ,  $P_{\text{moran}} = 0.0001$ ; Supplementary Result 9 and Supplementary Fig. 6).

In addition, we also verified the intersubject variability of WMFC using the four resting-state fMRI scans during 2 days, of

the Human Connectome Project (HCP) 7 T dataset. We found that the intersubject variability of WMFC based on the HCP 7 T dataset showed a similar pattern with the intersubject variability across 12 WM functional networks in this work ( $r=0.60$ ,  $P=0.04$ ; Supplementary Fig. 7).

## Discussion

The current study characterized the endophenotypic intersubject variability of WMFC, revealing how brain function in WM was associated with phenotypic gene expression. These two biological parameters are complementary and mutually reinforcing. First, the spatial distribution was organized along a broadly hierarchical axis, anchored in unimodal regions and extending into heteromodal association regions. Second, the transcriptomic and cellular organizations of high and low variabilities of WMFC were distinct and specialized. Genes overexpressed in heteromodal association regions were expressed predominantly in neuronal cells, involved in synapse-related terms and glutamic pathways, and associated with common psychiatric disorders. However, genes overexpressed in unimodal regions were mostly expressed in glial cells and were associated with neurological diseases. Furthermore, the spatial distributions of functional variabilities were recapitulated in multi-modularity brain structural, functional, and behavioral organizations. Our findings provided a perspective on intersubject variability in brain function, which has potential implications for understanding brain evolution and development.

This study built on an earlier hypothesis that functional brain connectivity varied along with GM cortical gradients<sup>8,10,11,38,62</sup>. This variability has potential evolutionary significance, shaped by genetic and environmental factors<sup>10</sup>. The evolutionary trajectories are similar between brain GM and WM, which both originate from limbic specific and subcortical areas. They later include more associated networks for higher cognitive functions in WM<sup>58,63</sup>. Thus, it is reasonable to conclude that there are the similar spatially varying gradients of functional variabilities between GM and WM. However, the absence of an evolutionary WM expansion map has prevented us from directly verifying similar principles of functional variability in WM.

The transcriptome revealed variability-related biological processes, based on the complex pattern of functional topographies. Genes overexpressed in high and low variability areas were distinct and specialized, while genes overexpressed in high variability areas were enriched for synapse-related terms and glutamatergic pathways. Synapses in the central nervous system represent the classic mechanism through which neural cells communicate. Synaptic-style release of glutamate, the brain's major excitatory neurotransmitter, occurs deep in the WM<sup>64</sup>. Here, glutamatergic pathways permit communication between axons and glial cells, enabling axon activity to couple with high fidelity to glial physiology<sup>65–67</sup>. The synaptic density changes with age<sup>68</sup>. Its overproduction is highest in prefrontal areas and lowest in primary sensory areas during development<sup>69</sup>. Thus, a gene overexpressed in synapse-related pathways may relate to highly functional variability in heteromodal association areas.

Genes with negatively PLS1 weights were mostly enriched for glia-related pathways. WM contains numerous glial cells. The biological process of regulation of gliogenesis results in the generation of glial cells<sup>70</sup>. In addition, ensheathment occurs in neurons in which glial cells envelop neuronal cell bodies and/or axons to form an insulating layer. This can take the form of myelinating or non-myelinating ensheathment. Myelin may serve as an inhibitor of brain plasticity<sup>71</sup>. This suggests that early sensory areas may require less plasticity, and therefore more myelin, whereas higher-order areas have less myelination,

which might enable greater plasticity<sup>72</sup>. In the present study, we found that genes, overexpressed and underexpressed in areas with low and high variability in WMFC, respectively, were mostly enriched for glia-related biological processes. Together, this phenomenon may be caused by more myelination of neurons in low-perceptual areas, and more non-myelination of neurons in higher-cognitive areas.

The gene expression profiles characteristic of different cell-types arise because these cells have distinct sets of transcription regulators. We thus exploited how the cell type-specific manner affected gene regulation leading to intersubject variabilities in WMFC. When integrated across levels of cell types, these variability-related genes from bulk tissues were assigned to seven canonical cell classes<sup>45</sup>. Functional variability in WM was associated with genes expressed primarily in neurons (both inhibitory and excitatory) and oligodendrocytes, suggesting that intersubject variability may partially depend on synaptic transmission and myelinated processes<sup>10,72</sup>. The cellular organizations refined our transcriptomic analyses, suggesting the distinct contributions of neurons and glial cells, along with an inverse functional variability gradient. Critically, our analytic workflow identified cell classes and genes without relying on postmortem tissue from participants; thus, enabling investigators to make predictions regarding the human neuroscience or the biology of distinct human disorders, by using data from native human tissue<sup>45</sup>.

The clinical relevance of intersubject variability-related genes was further confirmed by over-representation analyses<sup>73</sup>. Genes enriched for synapse-related terms and glutamatergic pathways were mostly associated with various psychiatric disorders, such as depression, schizophrenia, and autism spectrum disorder. This was supported by a previous study linking gene transcription to synaptic activity processes to major psychiatric disorders<sup>74</sup>. Dysregulated synaptic development, properties, and plasticity have been hypothesized to result in altered neuronal function in various neuropsychiatric disorder<sup>75,76</sup>, suggesting that the disrupted neurobiological mechanisms are shared across common disorders<sup>77</sup>, and intersubject variability in WMFC may be a candidate intermediate phenotype for the association between these genes and some mental disorders.

Genes associated with microglia were significantly overlapped with neurological diseases. Microglia, the resident immune cells of the central nervous system, play an important role in maintaining tissue homeostasis and contribute to normal brain development<sup>78</sup>. Therefore, studying microglia provides unprecedented insight into mechanisms involved in neurological diseases<sup>79,80</sup>. Thus, microglial cells have the potential to act as diagnostic markers of disease onset or progression, and could contribute to the outcome of neurodegenerative diseases<sup>81</sup>. In this study, genes associated with microglia were overexpressed in low variability areas, indicating that these brain areas may contribute to neurological diseases. Our results showed that variability-related gene data can relate to knowledge about transcriptome-neuroimaging, and may be clinically translatable.

Leveraging macroscale neuroimaging, we found that intersubject variability in WMFC recapitulated the variability of structural (i.e., WMV, FA, and T1w/T2w) and functional (i.e., CBF) organizations. From an evolutionary point of view, the heteromodal association areas that have undergone dramatic cortical expansion showed higher metabolic demands, more structural variability, and greatly reduced myelination, when compared with unimodal areas<sup>8,58</sup>. It is striking how well the CBF and structural variability maps matched the distribution of intersubject variability in WMFC, as revealed in this study. These findings suggested that the different scales of human brain organization are likely not independent from each other during brain development<sup>29</sup>.

The reasons for differences in the molecular hierarchies observed in the variable functional organizations between WM and GM are not completely understood. Using transcriptomic analysis and cellular specificity of the intersubject variability in WMFC, we found specific and common enrichment pathways of intersubject variability-related genes between WMFC and gray-matter FC (Supplementary Result 10 and Supplementary Fig. 8). Although WM contains a significantly lower neuron-to-glia ratio than GM<sup>82</sup>, synapse-related terms are commonly enriched for functional variabilities both in WM and GM, indicating the important roles of whole brain functional variability for psychiatric disorders. In addition, intersubject variabilities of WMFC organization exhibited specific enrichment pathways in glial cell-related terms, such as ensheathment of neurons and regulation of gliogenesis, highlighting the importance of functional variabilities in WM, as compared with GM, for neurological diseases. Understanding the nuances of shared and specific pathways of variabilities between WM and GM will increase our understanding of human neuroscience and may provide avenues of treatment for several diseases.

Several methodological considerations are noteworthy. First, we collected longitudinal, multi-session resting-state BOLD-fMRI data to ensure stable and reliable intersubject variability in WMFC. In addition, considering the spatial and temporal limitations of BOLD-fMRI data based on a 3 T machine, we further used four sessions of the HCP 7 T dataset to verify the intersubject variability of WMFC. Second, we used several complementary analysis strategies to obtain clean BOLD-fMRI signals from the WM, by strictly controlling the boundary between WM and GM, by separating WM and GM functional signal in pre-processing, and by identifying participants' voxels only in WM to create a WM mask<sup>17,19–21,49</sup>. Furthermore, from the architecture of brain venous systems, the possibility that deoxygenated blood was drained from cortical GM to deep WM was significantly small<sup>83</sup>. Moreover, there are two venous systems in normal neuroanatomy; one is the superficial venous system, which drains deoxygenated blood in the GM cortex and superficial WM into pial veins. The other is a deep system draining deoxygenated blood in deep WM into subependymal veins<sup>83,84</sup>. Furthermore, the brain venous architectures are spatially non-overlapping. Deoxygenated blood drainage from GM cortex to deep venous system through WM does exist, but the probability of draining is less than 3%<sup>83,84</sup>. Collectively, based on brain venous system architecture, these approaches ensured that BOLD-fMRI signals analyzed in the present study were from WM.

Several limitations in this study need to be considered. First, all subjects were college students, so a larger sample from a community-based population of varying ages is needed to generalize the current findings. Second, the AHBA gene data were measured postmortem in six subjects, which limited examinations of transcriptome-neuroimaging associations across groups and possibly placed individual effects out of scope. In addition, the AHBA gene data only included the right hemisphere for two subjects, and contained relatively small probes in WM, limiting our investigation between genes and functional variability in whole brain WM. Finally, the PLS analysis may result in inaccurate associations when the number of samples per feature is relatively small. In future studies, we will first perform data reduction to solve this problem<sup>85</sup> by selecting genes based on prior hypotheses.

In this study, we have proposed a landscape of the transcriptomic decoding and cellular specificity of intersubject variability in WMFC—high-to-low functional-varying gradients that were associated with gene expressions, associated with neurons to glial cells. Notably, synapse-related and glia-related genes were overexpressed in association and low-order perceptual

networks, respectively, which were specifically related to several neuropsychiatric illnesses and neurological diseases. In addition, the shared and specific biological processes between WM and GM highlighted the complementary and reverberating functional roles. These findings emphasized the clinically translational relevance for incorporating the specific intersubject variabilities in WM and GM to determine effective therapeutic targets.

## Methods

**Study design and participants.** This longitudinal study was approved by the Local Medical Ethics Committee of the University of Electronic Science and Technology of China (UESTC), China. Written informed consent was obtained from all subjects prior to scanning. Forty-five healthy subjects [age (mean  $\pm$  SD): 23.67  $\pm$  1.65 years, 20 females, all right-handed] were recruited from and the data were collected at the UESTC. All subjects had no history of neurological or psychiatric conditions, and no gross abnormalities on brain MRI.

Each subject underwent four MRI scanning sessions within 6 months (~14, 30, and 180 days from enrollment). All subjects performed two resting-state fMRI runs per session to estimate the WMFC. After quality control, 43 subjects who had two good runs (mean FD < 0.2 mm) in each session and two subjects who had one good run in one of four sessions were included for subsequent analyses.

**Data acquisition.** Multimodal neuroimaging battery (resting-state functional, structural, and diffusion data) were acquired on a 3.0 Tesla MRI scanner (GE Discovery 750 MRI) at the UESTC. During the resting-state functional scanning, subjects were instructed not to think of anything particular, to keep their eyes open, and to stay in a fixed position.

**Resting-state BOLD-fMRI scan:** To characterize intersubject variability of WMFC, each subject first underwent two resting-state BOLD-fMRI runs (30 min 10 s with 905 volumes per run) using an echo-planar imaging sequence: repetition time (TR)/echo time (TE) = 2000 ms/30 ms, flip angle = 90°, field of view (FOV) = 240  $\times$  240 mm<sup>2</sup>, in-plane matrix = 64  $\times$  64, slice thickness = 3.2 mm, no interslice gap, and 43 transverse slices. The identical protocol was used for each session.

**Resting-state arterial spin labeling (ASL) scan:** To associate the spatial pattern of intersubject variability of WMFC, 34 out of 45 subjects underwent one resting-state ASL run during the last session using a 3D pulsed continuous ASL technique. Interleaved 30 pairs control and label images were acquired (7.5 min with 50 of control and label images). We used the following settings: TR/TE = 4632 ms/10.536 ms, flip angle = 111°, label duration = 1450 ms, post-labeling delay = 1525 ms, FOV = 240  $\times$  240 mm<sup>2</sup>, in-plane matrix = 128  $\times$  128, slice thickness = 4 mm, no interslice gap, slices = 36, acquisition time = 269 s.

Subsequently, structural, myelin, and diffusion data were obtained. All subjects, during each session, underwent high-resolution T1-weighted (T1w) structural scans using a 3D Fast spoiled gradient-recalled echo sequence in the sagittal orientation (TR/TE = 5.952 ms/1.96 ms, flip angle = 12°, FOV = 256  $\times$  256 mm<sup>2</sup>, in-plane matrix = 256  $\times$  256, slice thickness = 1 mm, and 154 slices) and diffusion-weighted images (DWI) using a spin echo-based echo-planar imaging sequence (TR/TE = 8500/64.7 ms, flip angle = 90°, FOV = 256  $\times$  256 mm<sup>2</sup>, in-plane matrix = 128  $\times$  128, slice thickness = 2 mm, slices = 78, four volumes without diffusion weighting  $b = 0$  s/mm<sup>2</sup>, and 60 non-collinear directions  $b = 1000$  s/mm<sup>2</sup>). In addition, 44 out of 45 subjects also underwent T2-weighted structural scans during the last session using a 3D Cube T2 sequence in the sagittal orientation with TR/TE = 2500 ms/60.97 ms, flip angle = 90°, FOV = 256  $\times$  256 mm<sup>2</sup>, in-plane matrix = 256  $\times$  256, slice thickness = 1 mm, and 256 slices.

**Data preprocessing.** Both BOLD-fMRI and ASL images were preprocessed using the statistical parametric mapping toolbox (SPM12; <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>), as outlined below. The resting-state BOLD-fMRI data were preprocessed according to our previously described WM functional procedures<sup>19,20,49</sup> based on DPARSF (v4.3, <http://rfmri.org/DPARSF>). For each run, slice-timing correction and realignment were applied to the 900 functional volumes after excluding the first five volumes (10 s). For each subject, the mean FD was measured to control data quality. Each subject's structural image was co-registered to the preprocessed functional images, and was then segmented into GM, WM, and cerebrospinal fluid (CSF) using DARTEL<sup>86</sup>. The mean signals from CSF, 24 head motion parameters (Friston 24-parameter model) were regressed out by multiple linear regression analysis. To avoid elimination of important neural signals, we did not regress out WM and brain global signals, as previous studies have suggested<sup>17,19–21,49</sup>.

To minimize mixing signal (and noise) components from the WM and GM due to partial volume effect, subsequent processing of the functional images was performed separately for WM and GM, in accordance with our previous studies<sup>19,20,49</sup>. First, each individual mask was generated using a 60% threshold on the probability map of WM (i.e., produced by structural segmentation). Second, functional images were spatially divided into WM functional images using the dot product between functional images and the individual WM mask. Third, the WM functional images were spatially normalized to a standard space (Montreal

Neurologic Institute, MNI) by structural segmentation and were resampled to  $3 \times 3 \times 3 \text{ mm}^3$ . Then, only voxels identified as WM across all of the participants were used to create the group-level WM mask. To minimize the impact of deep brain structures, the probability (25% threshold) Harvard-Oxford Atlas was used to remove subcortical structures (i.e., the bilateral thalamus, putamen, caudate, pallidum, and accumbens) from the group-level WM mask as previous studies suggested<sup>17,19,20,49</sup>. Subsequently, functional data preprocessing included smoothing with 6 mm full-width half-maximum (FWHM) isotropic Gaussian kernel, and a band-pass filtering (0.01–0.10 Hz).

For ASL images, the absolute quantified rCBF maps were first obtained by subtracting the label from control images using Functool (v 12.2.01) embedded in the GE MR-750 scanner system. Then, the z-scored rCBF maps were performed by SPM12 according to the following steps: firstly, T1w images were coregistered to their own rCBF maps, and then segmented into GM, WM, and CSF. Second, the WM maps were nonlinear co-registered to the T1 template in MNI space. Finally, each rCBF map was normalized into the standard MNI space with 3 mm isotropic voxels based on the transform parameters that were estimated during nonlinearly co-registration. The normalized rCBF maps were spatially smoothed by a 6 mm FWHM Gaussian kernel.

The DWI data were preprocessed on volumetric space using FSL (v6.0, <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki>). For each subject's native space, the DWI images were corrected for the eddy-current-induced distortions and head movements, and the diffusion tensor model was estimated using the linear least-squares fitting method, resulting in the fractional anisotropy (FA) map. Individual T1-weighted images were co-registered to the images in the DTI native space. Then, the co-registered T1-weighted image was mapped to the T1 image in MNI space by applying an affine transformation to obtain the transformation matrix<sup>87</sup>. Finally, FA maps were normalized into the standard MNI space and resampled with 3-mm isotropic voxels based on the inverse transformation matrix.

To obtain myelin enhanced contrast images, the T1w/T2w ratio was calculated for the same subject using MRTOOL (<https://www.nitrc.org/projects/mrtool/>)<sup>56</sup>. To standardize the T1/T2w image, we created two subject-specific masks by warping predefined masks in the MNI standard space to individual space. One mask contained relatively low values on the T1w image and high values on the T2w image, and the other mask had reversed characteristics. We implemented this specification by selecting two masks covering the eyeballs and temporal muscles, respectively<sup>56</sup>. These masks were defined directly in the MNI space by segmenting and thresholding the ICBM152 template images<sup>56</sup>. In parallel, the T2w image was co-registered to the T1w image. Then, the T1w and T2w images were jointly subjected to bias correction to ensure that the sensitivity profile was spatially equalized. After calibrating T1w and T2w images, their ratio was calculated to produce the calibrated T1w/T2w image as a myelin enhanced contrast image.

**Intersubject variability of WMFC.** After preprocessing, BOLD-fMRI data were concatenated into a single run (900 × 2 volumes) for each session<sup>10</sup> if the subject had two good runs. After obtaining an individual WM mask for each subject, these individual masks were averaged across participants to obtain the percentage of participants in which it was classified as WM for each voxel. Voxels identified in all sessions and all participants as WM were considered as a group-level WM mask, resulting in 7151 voxels<sup>19,20,49</sup>.

Intersubject variability of WMFC was estimated in line with previous studies on functional variability in GM<sup>10,11,38</sup>. To obtain individual WMFC maps, we used each of the 7151 voxels as the seed, and computed the FC between the seed and the remaining voxels, resulting in 7,151 maps for each subject and each session. This individual correlation map based on each seed voxel was vectored as  $WMFC_v(s, t)$ , where  $v = 1, 2, \dots, 7,151$ , the values represented the FC between the seed  $v$  and the remaining voxels;  $s (s \in 1, 2, \dots, 45)$  denoted the number of subjects; and  $t (t \in 1, 2, \dots, 4)$  denoted the number of sessions.

For each session, 45 maps derived from 45 subjects were obtained. For a given seed voxel,  $v$ , the spatial similarity between the 45 maps derived from 45 subjects was quantified by averaging Pearson's correlation coefficients between any two WMFC vectors (total,  $C_{45}^2 = 990$  combinations):

$$R_v(t) = E[\text{corr}(WMFC_v(s_p, t), WMFC_v(s_q, t))],$$

where,  $p$  and  $q$  indicate two different subjects,  $p, q = 1, 2, \dots, 45; p \neq q$ . This resulted in an intersubject similarity map.

To obtain the intersubject variability of WMFC, we first adjusted spatial similarities between the 45 maps for each session:

$$R'_v(t) = 1 - R_v(t)$$

Finally, four original (unadjusted) intersubject variability maps were obtained.

The intra-subject variability was measured based on repeated multi-sessions ( $t = 4$ ) of a subject. For each subject, four WMFC vectors were obtained from four sessions. Accordingly, the spatial similarities of intra-subjects were estimated using any two WMFC vectors (total,  $C_4^2 = 6$  combinations). Considering the values of WMFC maps ranging from -1 to 1, we inverted the similarity map by subtracting from one (i.e.,  $1 - R_{\text{spatial similarity}}$ ) as the intra-subject variability of each subject:

$$N_v(s) = 1 - E[\text{corr}(WMFC_v(s, t_m), WMFC_v(s, t_n))],$$

where  $m, n = 1, 2, \dots, 4; m \neq n$ .

The intra-subject variability map was then averaged across 45 subjects and assigned to the seed voxel,  $v$ :

$$N_v = E[N_v(s)]$$

"Technical noise" of the voxel  $v$  was reflected by the tSNR of the BOLD-fMRI signal:

$$tSNR_v = \frac{\langle S \rangle_t}{\sigma_t}$$

Where  $\langle S \rangle_t$  is the average BOLD signal across time, and  $\sigma_t$  is the corresponding temporal standard-deviation map. The technical noise was the inverse tSNR (i.e.,  $1/tSNR$ ).

To minimize the effects of confounding factors on intersubject variability, the intra-subject variability and "technical noise" were regressed out from original (unadjusted)  $R'_v(t)$  using a general linear model (GLM) for adjusting intersubject variability:

$$V_v(t) = R'_v(t) - \beta_1 \times N_v - \beta_2 \times \text{Noise} - c,$$

where  $\beta_1, \beta_2$ , and  $c$  are parameters determined via a GLM. Finally, resulting adjusted maps derived from each session  $t$  were averaged for intersubject variabilities of the WMFC map (Fig. 2).

**Functional organization of intersubject variability.** To quantify the distribution of intersubject variability within each of functional network in WM, we used the previous parcellations based on healthy controls to identify large-scale networks as reported by Peer et al.<sup>17</sup>. Considering non-zero voxels of the WM mask in this study, we first overlapped the WM mask with the previous functional parcellations in WM. We then assigned each voxel to one of the functional networks based on the MNI coordinate, and rearranged the network label. According to the previous study<sup>17</sup>, the function of each WM network was described as follows: Network 1 was the frontoparietal control network and default mode network ( $n = 160$  voxels); Network 2 was the deep frontal WM network ( $n = 1293$  voxels); Network 3 was the inferior longitudinal fasciculus system ( $n = 777$  voxels); Network 4 was the temporal-orbitofrontal network and default mode network ( $n = 112$  voxels); Network 5 was the dorsal attention network ( $n = 717$  voxels); Network 6 was the forceps minor system ( $n = 233$  voxels); Network 7 was the superior longitudinal fasciculus system ( $n = 2135$  voxels); Network 8 was the visual superficial WM system ( $n = 40$  voxels); Network 9 was the sensorimotor superficial WM system ( $n = 90$  voxels); Network 10 was the sensorimotor and ventral attention system ( $n = 475$  voxels); and Network 11 ( $n = 224$  voxels) and Network 12 ( $n = 873$  voxels) were cerebellum systems. Finally, the intersubject variability of voxels within each WM functional network was averaged as the intersubject variability of the WM functional networks.

**Investigating the gene expression pattern associated with the intersubject variabilities of WMFC.** The AHBA dataset bridged the gap between macroscale intersubject variability of WMFC and microscale gene expression<sup>46</sup>. The AHBA microarray-based gene expression provided high resolution coverage of nearly the entire brain, with 3702 spatially distinct tissue samples taken from six neurotypical postmortem brains. The AHBA microarray data were processed including verifying probe-to-gene annotations, filtering of probes, and probe selection following the protocol of Arnatkevicius et al.<sup>88</sup>. After preprocessing, 10,027 genes lists were used to select gene expression maps from Neurosynth-Genes (<https://www.neurosynth.org/genet/>). These gene expression maps covered the whole brain and co-registered the locations of all microarray samples with the MNI stereotactic space. Consequently, the overlapped 9922 gene list were used for further analyses. In addition, the number of overlapped voxels between gene expression and intersubject variability maps was 3415, resulting in a matrix (3415 voxels × 9922 genes) of the brain-wide gene expression for the WM.

PLS correlation<sup>89</sup> was used to determine the relationship between the intersubject variability of WMFC and transcriptional activity for all 9,922 genes. Gene expression data were used as predictor variables of intersubject variability in the PLS regression. PLS1 was the linear combination of gene expression values, which was most strongly correlated with functional variability in WM. Findings were corrected for spatial autocorrelation with Moran Spectral Randomization (MSR)<sup>42</sup> implemented in BrainSpace<sup>43</sup>, which generated spatially constrained null models for irregularly spaced data. Bootstrapping was used to estimate the error of each gene PLS1 weight, and the ratio of the weight of each gene to its bootstrap standard error was used to calculate the Z scores and to rank the genes according to their contributions to PLS1<sup>90</sup>. The set of genes with a FDR of 5%, either positive (PLS1 +), or negative (PLS1 -), was identified for the intersubject variability gene list.

**Gene annotation analyses: functional enrichment.** Functional enrichment for GO gene sets and the KEGG pathway search were used to include over 40 independent knowledgebases within Metascape<sup>44</sup>. The lists of PLS1 + and PLS1 - genes were separately analyzed using Metascape (<https://metascape.org/gp/index.html#/main/step1>). Based on the well-adopted hypergeometric test, Metascape identified all ontology terms that contained a statistically greater number of genes in common with an input list than expected by chance. Briefly, pairwise similarities between

any two enriched terms were computed based on a  $Kappa$ -test score<sup>91</sup>. The similarity matrix was then hierarchically clustered to trim the constructed tree into separate clusters. The resulting enrichment pathways were thresholded for significance at 5%, corrected by the FDR.

**Genes expressed in specific cell types.** To further refine our analysis and considering cellular diversity in the brain, we took an indirect approach to assign PLS1+ and PLS1− genes to seven canonical cell classes. To obtain gene sets from each cell type, we compiled data from five different single-cell studies using postmortem cortical samples in human postnatal subjects. This approach avoided any bias based on acquisition methodology, analysis or thresholding, leading to the initial inclusion of 58 cell classes, which were provided by Seidlitz et al.<sup>45</sup>. Many of those cell classes were overlapping based on nomenclature and/or constituent genes. To avoid any bias based on acquisition, analysis, or thresholding, five different single-cell studies were included in this study<sup>92–96</sup>. We further organized cell types into seven canonical classes following Seidlitz et al.<sup>45</sup>. These cell types included microglia, endothelial cells, oligodendrocyte precursors, oligodendrocytes, astrocytes, and excitatory and inhibitory neurons. Only one study included the annotation of the pericyte type, thus this gene set was excluded. To assign inter-subject variability-related genes obtained by PLS analysis to cell types, we overlapped the gene set of each cell type with the PLS1+ or PLS1− rank gene list. We resampled the genes involved in cell types 5000 times to test the null hypothesis that the PLS1+ or PLS1− gene list was randomly assigned to different cell types. The  $P_{perm}$  scores were obtained by the occupied null models (<5th, or >95th centile) and corrected by FDR.

**Genes enriched for brain diseases.** Brain disease-related enrichment for inter-subject variability-related genes was determined using WebGestalt<sup>48</sup> (<http://www.webgestalt.org/>) to identify disease-association terms based on DisGeNET, OMIM, and GLAD4U databases. The gene lists of PLS1+ and PLS1− were curated for brain disease gene associations. The top 10 terms were listed according to the ranked  $P < 0.05$ , FDR-corrected.

**Relationship between macroscale functional and anatomical variability and WMFC variability.** The rCBF has received much attention in brain function-related studies due to its important role in maintaining normal brain function and its close relationships with brain metabolism and connectivity<sup>8</sup>. Canonical structural variability indices in WM were white matter volume (WMV) and FA variabilities<sup>58</sup>. The myelin content representing T1w/T2w was an important component in WM. Thus, intersubject variability in these functional and anatomical indices was estimated as voxel-wise using SD across subjects<sup>5,55,97</sup>. The Spearman's correlation coefficient was calculated between the intersubject variability map in WMFC and these macroscale indices for across-voxel in the WM.

**Behavior decoding of the intersubject variabilities of WMFC.** Functional activation probability maps were obtained for multiple cognitive terms using the Neurosynth meta-analytic database using volumetric “association term” maps<sup>98</sup> (<https://www.neurosynth.org>). Then, each cognitive map was overlapped with the intersubject variability of the WMFC map to obtain the non-zero voxels between the two maps. Pearson's correlation analysis was used to measure the similarity between the cognitive map and intersubject variability of the WMFC map. Terms were selected in this study if their thresholds of  $z$ -statistic values were above 2.1 (corresponding Pearson's correlation coefficient:  $P < 0.05$ ), and if they were not “noise” terms, which did not capture any coherent cognitive function<sup>99</sup>.

**Statistics and reproducibility.** Testing for linear associations between the intersubject variability of WMFC and other brain phenotypes may lead to biased testing statistics due to spatial autocorrelation of MRI data. Thus, we generated random datasets, with equivalent spatial autocorrelations as the response variables using MSR and the singleton procedure<sup>42</sup> implemented in BrainSpace<sup>43</sup>. The MSR methods generated spatially constrained null models for irregularly spaced data. All linear models were fitted for the original data as well as 10,000 corresponding simulated datasets.  $P_{moran}$ -values were obtained by the occupied null models (<5th, or >95th centile) and corrected by FDR.

Reproducible analyses were used to validate the intersubject variability in WMFC. We first built trained PLS predictive model based on the variability data from four sessions, where each session included two runs. The first three sessions were trained on the PLS model and the fourth session were used as test data in a leave-one-out cross-validation scheme. To test the predictive power, correlation analysis was used between the predictive and observed intersubject variability. In addition, HCP dataset was used to test the robustness of the intersubject variability of WMFC.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

Human gene expression maps that support the findings of this study are available Neurosynth-Gene database (<https://www.neurosynth.org/genes/>) based on the Allen Brain Atlas (<https://human.brain-map.org/static/download>). Compiled cell-specific gene set list from all available large-scale single-cell studies of the adult human cortex can be obtained from the raw Seidlitz et al.<sup>45</sup> dataset ([https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-020-17051-5/MediaObjects/41467\\_2020\\_17051\\_MOESM8\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-020-17051-5/MediaObjects/41467_2020_17051_MOESM8_ESM.xlsx)). Disease-association terms were obtained from WebGestalt website (<http://www.webgestalt.org/>). The 7 T Human connectome dataset is available at <https://db.humanconnectome.org/>. The neuroimaging data that support the findings of this study are available from the corresponding author (W.L.) upon reasonable request. The source data underlying Figs. 2, 3, 4, 5, 6, and 7 are provided as Supplementary Data 1.

## Code availability

The code for gene expression analysis can be found at <https://github.com/BMHLab/AHBAProcessing>. Gene enrichments were analyzed at <https://metascape.org/gp/index.html#/main/step1>. The code for spatial permutation testing was implemented in BrainSpace (<https://www.brainspace.com/>). The code for preprocessing of WM functional data can be found at <https://github.com/weilia081/WMPFPrep>.

Received: 1 July 2021; Accepted: 30 November 2021;  
Published online: 20 December 2021

## References

1. Biswal, B. et al. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34**, 537–541 (1995).
2. Biswal, B. B. et al. Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. USA* **107**, 4734–4739 (2010).
3. Ogawa, S. et al. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* **89**, 5951–5955 (1992).
4. Braga, R. M. & Buckner, R. L. Parallel interdigitated distributed networks within the individual estimated by intrinsic functional connectivity. *Neuron* **95**, 457–471 e455 (2017).
5. Gordon, E. M. et al. Individual variability of the system-level organization of the human brain. *Cereb. Cortex* **27**, 386–399 (2017).
6. Gordon, E. M. et al. Precision functional mapping of individual human brains. *Neuron* **95**, 791–807 e797 (2017).
7. Finn, E. S. et al. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671 (2015).
8. Cui, Z. X. et al. Individual variation in functional topography of association networks in youth. *Neuron* **106**, 340–353 e348 (2020).
9. Kong, R. et al. Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. *Cereb. Cortex* **29**, 2533–2551 (2019).
10. Mueller, S. et al. Individual variability in functional connectivity architecture of the human brain. *Neuron* **77**, 586–595 (2013).
11. Xu, T. et al. Interindividual variability of functional connectivity in awake and anesthetized rhesus macaque monkeys. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **4**, 543–553 (2019).
12. Seitzman, B. A. et al. Trait-like variants in human functional brain networks. *Proc. Natl. Acad. Sci. USA* **116**, 22851–22861 (2019).
13. Fields, R. D. White matter in learning, cognition and psychiatric disorders. *Trends Neurosci.* **31**, 361–370 (2008).
14. Fields, R. D. Map the other brain. *Nature* **501**, 25–27 (2013).
15. Marussich, L. et al. Mapping white-matter functional organization at rest and during naturalistic visual perception. *Neuroimage* **146**, 1128–1141 (2017).
16. Ding, Z. et al. Detection of synchronous brain activity in white matter tracts at rest and under functional loading. *Proc. Natl. Acad. Sci. USA* **115**, 595–600 (2018).
17. Peer, M. et al. Evidence for functional networks within the human brain's white matter. *J. Neurosci.* **37**, 6394–6407 (2017).
18. Wang, P. et al. The organization of the human corpus callosum estimated by intrinsic functional connectivity with white-matter functional networks. *Cereb. Cortex* **30**, 3313–3324 (2020).
19. Li, J. et al. Exploring the functional connectome in white matter. *Hum. Brain Mapp.* **40**, 4331–4344 (2019).
20. Li, J. et al. A neuromarker of individual general fluid intelligence from the white-matter functional connectome. *Transl. Psychiatry* **10**, 147 (2020).

21. Ji, G. J. et al. Low-frequency blood oxygen level-dependent fluctuations in the brain white matter: more than just noise. *Sci. Bull.* **62**, 656–657 (2017).
22. Seung, S. *Connectome: How the Brain's Wiring Makes Us Who We Are* (Houghton Mifflin Harcourt Publishing Company, New York, 2012).
23. Zhao, B. et al. Common genetic variation influencing human white matter microstructure. *Science* **372**, eabf3736 (2021).
24. Elliott, L. T. et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210–216 (2018).
25. Zhao, B. et al. Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits ( $n = 17,706$ ). *Mol. Psychiatry* **26**, 3943–3955 (2021).
26. Traylor, M. et al. Genetic variation in PLEKHG1 is associated with white matter hyperintensities ( $n=11,226$ ). *Neurology* **92**, E749–E757 (2019).
27. Fornage, M. et al. Genome-wide association Studies of cerebral white matter lesion burden: the charge consortium. *Ann. Neurol.* **69**, 928–939 (2011).
28. Fornito, A., Arnatkeviciute, A. & Fulcher, B. D. Bridging the gap between connectome and transcriptome. *Trends Cogn. Sci.* **23**, 34–50 (2019).
29. van den Heuvel, M. P., Scholtens, L. H. & Kahn, R. S. Multiscale neuroscience of psychiatric disorders. *Biol. Psychiatry* **86**, 512–522 (2019).
30. Li, J. et al. Cortical structural differences in major depressive disorder correlate with cell type-specific transcriptional signatures. *Nat. Commun.* **12**, 1647 (2021).
31. Hawrylycz, M. et al. Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015).
32. Richiardi, J. et al. BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241–1244 (2015).
33. Wang, G. Z. et al. Correspondence between resting-state activity and brain gene expression. *Neuron* **88**, 659–666 (2015).
34. Brennan, B. P. et al. Use of an individual-level approach to identify cortical connectivity biomarkers in obsessive-compulsive disorder. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **4**, 27–38 (2019).
35. Wang, D. et al. Individual-specific functional connectivity markers track dimensional and categorical features of psychotic illness. *Mol. Psychiatry* **25**, 2119–2129 (2020).
36. Lebois, L. A. M. et al. Large-scale functional brain network architecture changes associated with trauma-related dissociation. *Am. J. Psychiatry* **178**, 165–173 (2021).
37. Sun, X. et al. Disrupted Intersubject Variability Architecture in Functional Connectomes in Schizophrenia. *Schizophr. Bull.* **47**, 837–848 (2021).
38. Stoecklein, S. et al. Variable functional connectivity architecture of the preterm human brain: Impact of developmental cortical expansion and maturation. *Proc. Natl. Acad. Sci. USA* **117**, 1201–1206 (2020).
39. Zuo, X. N. et al. An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci. Data* **1**, 140049 (2014).
40. Esteban, O. et al. MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One* **12**, e0184661 (2017).
41. Li, M. et al. Performing group-level functional image analyses based on homologous functional regions mapped in individuals. *PLoS Biol.* **17**, e2007032 (2019).
42. Wagner, H. H. & Dray, S. Generating spatially constrained null models for irregularly spaced data using Moran spectral randomization methods. *Methods Ecol. Evol.* **6**, 1169–1178 (2015).
43. Vos de Wael, R. et al. BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Commun. Biol.* **3**, 103 (2020).
44. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).
45. Seidlitz, J. et al. Transcriptomic and cellular decoding of regional brain vulnerability to neurogenetic disorders. *Nat. Commun.* **11**, 3358 (2020).
46. Hawrylycz, M. J. et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).
47. Oldham, M. C. et al. Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11**, 1271–1282 (2008).
48. Wang, J. et al. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130–W137 (2017).
49. Li, J. et al. White-matter functional topology: a neuromarker for classification and prediction in unmedicated depression. *Transl. Psychiatry* **10**, 365 (2020).
50. Jiang, Y. et al. White-matter functional networks changes in patients with schizophrenia. *Neuroimage* **190**, 172–181 (2019).
51. Fan, Y. S. et al. Impaired interactions among white-matter functional networks in antipsychotic-naïve first-episode schizophrenia. *Hum. Brain Mapp.* **41**, 230–240 (2020).
52. Chen, H. et al. Atypical functional covariance connectivity between gray and white matter in children with autism spectrum disorder. *Autism Res.* **14**, 464–472 (2021).
53. Shang, Z. et al. Identification of key genes associated with multiple sclerosis based on gene expression data from peripheral blood mononuclear cells. *PeerJ* **8**, e8357 (2020).
54. Kisler, K. et al. Cerebral blood flow regulation and neurovascular dysfunction in Alzheimer disease. *Nat. Rev. Neurosci.* **18**, 419–434 (2017).
55. Laumann, T. O. et al. Functional system and areal organization of a highly sampled individual human brain. *Neuron* **87**, 657–670 (2015).
56. Ganzetti, M., Wenderoth, N. & Mantini, D. Whole brain myelin mapping using T1- and T2-weighted MR imaging data. *Front Hum. Neurosci.* **8**, 671 (2014).
57. Sampaio-Baptista, C. et al. White matter structure and myelin-related gene expression alterations with experience in adult rats. *Prog. Neurobiol.* **187**, 101770 (2020).
58. Crosson, P. L. et al. Structural variability across the primate brain: a cross-species comparison. *Cereb. Cortex* **28**, 3829–3841 (2018).
59. Glasser, M. F. & Essen, Van D. C. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* **31**, 11597–11616 (2011).
60. Burt, J. B. et al. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **21**, 1251–1259 (2018).
61. Abdi, H. Partial least squares regression and projection on latent structure regression (PLS Regression). *WIREs Comp. Stat.* **2**, 97–106 (2010).
62. Wang, D. et al. Parcellating cortical functional networks in individuals. *Nat. Neurosci.* **18**, 1853–1860 (2015).
63. Paus, T. et al. Maturation of white matter in the human brain: a review of magnetic resonance studies. *Brain Res Bull.* **54**, 255–266 (2001).
64. Alix, J. J. P. & Domingues, A. M. D. White matter synapses form, function, and dysfunction. *Neurology* **76**, 397–404 (2011).
65. Micu, I. et al. NMDA receptors mediate calcium accumulation in myelin during chemical ischaemia. *Nature* **439**, 988–992 (2006).
66. Salter, M. G. & Fern, R. NMDA receptors are expressed in developing oligodendrocyte processes and mediate injury. *Nature* **438**, 1167–1171 (2005).
67. Karadottir, R. et al. NMDA receptors are expressed in oligodendrocytes and activated in ischaemia. *Nature* **438**, 1162–1166 (2005).
68. Sakai, J. How synaptic pruning shapes neural wiring during development and, possibly, in disease. *Proc. Natl. Acad. Sci. USA* **117**, 16096–16099 (2020).
69. Tierney, A. L. & Nelson, C. A. Brain development and the role of experience in the early. *Zero Three* **30**, 9–13 (2009).
70. Bonni, A. et al. Regulation of gliogenesis in the central nervous system by the JAK-STAT signaling pathway. *Science* **278**, 477–483 (1997).
71. Glasser, M. F. et al. Trends and properties of human cerebral cortex: correlations with cortical myelin content. *Neuroimage* **93**, 165–175 (2014).
72. Hunt, B. A. E. et al. Relationships between cortical myeloarchitecture and electrophysiological networks. *Proc. Natl. Acad. Sci. USA* **113**, 13510–13515 (2016).
73. Pinerio, J. et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **45**, D833–D839 (2017).
74. Patel, Y. et al. Virtual histology of cortical thickness and shared neurobiology in 6 psychiatric disorders. *JAMA Psychiatry* **78**, 47–63 (2021).
75. Wang, X. Y. et al. Synaptic dysfunction in complex psychiatric disorders: from genetics to mechanisms. *Genome Med.* **10**, 9 (2018).
76. Tomoda, T., Yang, K. & Sawa, A. Neuronal autophagy in synaptic functions and psychiatric disorders. *Biol. Psychiatry* **87**, 787–796 (2020).
77. Anttila, V. et al. Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).
78. Subhramanyam, C. S. et al. Microglia-mediated neuroinflammation in neurodegenerative diseases. *Semin Cell Dev. Biol.* **94**, 112–120 (2019).
79. Hickman, S. et al. Microglia in neurodegeneration. *Nat. Neurosci.* **21**, 1359–1369 (2018).
80. Bachiller, S. et al. Microglia in neurological diseases: a road map to brain-disease dependent-inflammatory response. *Front Cell Neurosci.* **12**, 488 (2018).
81. Perry, V. H., Nicoll, J. A. R. & Holmes, C. Microglia in neurodegenerative disease. *Nat. Rev. Neurol.* **6**, 193–201 (2010).
82. Gore, J. C. et al. Functional MRI and resting state connectivity in white matter—a mini-review. *Magn. Reson Imaging* **63**, 1–11 (2019).
83. Huang, Y. et al. Voxel-wise detection of functional networks in white matter. *Neuroimage* **183**, 544–552 (2018).
84. Ruiz, D. S., Yilmaz, H. & Gailloud, P. Cerebral developmental venous anomalies: current concepts. *Ann. Neurol.* **66**, 271–283 (2009).
85. Wang, H. T. et al. Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists. *Neuroimage* **216**, 116745 (2020).
86. Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95–113 (2007).
87. Zhang, Z. et al. Altered functional-structural coupling of large-scale brain networks in idiopathic generalized epilepsy. *Brain* **134**, 2912–2928 (2011).
88. Arnatkeviciute, A., Fulcher, B. D. & Fornito, A. A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* **189**, 353–367 (2019).

89. Krishnan, A. et al. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* **56**, 455–475 (2011).
90. Morgan, S. E. et al. Cortical patterning of abnormal morphometric similarity in psychosis is associated with brain expression of schizophrenia-related genes. *Proc. Natl. Acad. Sci. USA* **116**, 9604–9609 (2019).
91. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**, 37–46 (1960).
92. Zhang, Y. et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* **89**, 37–53 (2016).
93. Lake, B. B. et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
94. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
95. Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. USA* **112**, 7285–7290 (2015).
96. Li, M. F. et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).
97. Mohr, P. N. C. & Nagel, I. E. Variability in brain activity as an individual difference measure in neuroscience? *J. Neurosci.* **30**, 7755–7757 (2010).
98. Yarkoni, T. et al. Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* **8**, 665–670 (2011).
99. Margulies, D. S. et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. USA* **113**, 12574–12579 (2016).

### Acknowledgements

We are grateful to all the participants in this study. This work was supported by the National Key Project of Research and Development of Ministry of Science and Technology (2018AAA0100705), the National Natural Science Foundation of China (82121003, 61871077, 62036003, U1808204, and 61876156), the Excellent Youth Foundation of Sichuan Scientific Committee (2020JDJQ0016), and the China Postdoctoral Science Foundation (BX2021057). The reproducible neuroimaging data were provided by the “WU-Minn-Ox” HCP 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

W.L. and H.C. led the project. J.L., W.L., and H.C. were responsible for the study concept and the design of the study. J.L. and W.L. created the figures and wrote the manuscript.

G-R.W. performed data analysis in the HCP dataset. B.L. analysed the neuroimaging data in the UESTC dataset. F.F., X.Z., Y.M., and P.Z. contributed to data acquisition. Y.M. and S.Y. analysed the transcriptional data. B.B.B. made substantial contributions to the manuscript and provided critical comments. All authors reviewed and commented on the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02952-y>.

**Correspondence** and requests for materials should be addressed to Huaifu Chen or Wei Liao.

**Peer review information** *Communications Biology* thanks Lorenzo Pini and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: Michel Thiebaut de Schotten and Karli Montague-Cardoso.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021