# nature communications

Article

https://doi.org/10.1038/s41467-025-60185-7

# Atlas of proteomic signatures of brain structure and its links to brain disorders

Received: 16 August 2024

Accepted: 16 May 2025

Published online: 02 June 2025

Check for updates

Peng Ren<sup>1,2,3,12</sup>, Xiao-He Hou<sup>1,2,3,4,12</sup>, Zeyu Li<sup>1,2,3</sup>, Jia You  $\textcircled{D}^{1,2,3}$ , Yuzhu Li<sup>1,2,3</sup>, Wei Zhang<sup>1,2,3</sup>, Weikang Gong  $\textcircled{D}^5$ , Bei Zhang  $\textcircled{D}^{1,2,3}$ , Bangsheng Wu<sup>1,2</sup>, Linbo Wang  $\textcircled{D}^{1,2,3}$ , Chun Shen<sup>1,2,3</sup>, Yujie Zhao<sup>1,2,3</sup>, Qing Ma<sup>1,2,3</sup>, Jujiao Kang  $\textcircled{D}^{1,2,3}$ , Yuchao Jiang  $\textcircled{D}^{1,2,3}$ , Neil Roberts  $\textcircled{D}^6$ , Fan Xu  $\textcircled{D}^7$ , Yong He  $\textcircled{D}^{8,9}$ , Jin-Tai Yu  $\textcircled{D}^{1,2}$ , Meiyun Wang<sup>10</sup>  $\boxtimes$  & Wei Cheng  $\textcircled{D}^{1,2,3,11}$ 

Individual variation in brain structure influences deterioration due to disease and comprehensive profiling of the associated proteomic signature advances mechanistic understanding. Here, using data from 4997 UK Biobank participants, we analyzed the associations between 2920 plasma proteins and 272 neuroimaging-derived brain structure measures. We identified 5358 associations between 1143 proteins and 256 brain structure measures, with NCAN and LEP proteins showing the most associations. Functional enrichment implicated these proteins in neurogenesis, immune/apoptotic processes and neurons. Furthermore, bidirectional Mendelian randomization revealed 33 associations between 32 proteins and 23 brain structure measures, and 21 associations between nine brain structure associated proteins and ten brain disorders. Moreover, the significant associations between the identified proteins and mental health were mediated by brain volume and surface area. In summary, this study generates a comprehensive atlas mapping the patterns of association between proteome and brain structure, highlighting their potential value for studying brain disorders.

Magnetic Resonance Imaging (MRI) continues to play a pivotal role in the study of individual variation in the deterioration of brain structure in a wide range of neurodegenerative and neuropsychiatric disorders<sup>1,2</sup>. Alterations in brain structure are among the earliest pathophysiological changes detectable in neurodegenerative disorders and can manifest decades before overt clinical symptoms appear<sup>3,4</sup>. Besides, longitudinal MRI can potentially be used to monitor disease progression across multiple stages and to predict the risk of developing brain disorders in healthy individuals<sup>4,5</sup>. Furthermore, analysis of brain structure heterogeneity may allow the classification of patients into subtypes with different clinical manifestations and temporal trajectories<sup>6,7</sup>. Altogether, MRI-derived measures of brain

<sup>1</sup>Institute of Science and Technology for Brain-Inspired Intelligence, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Fudan University, Shanghai, China. <sup>2</sup>Department of Neurology, Huashan Hospital, Fudan University, Shanghai, China. <sup>3</sup>Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence (Fudan University), Ministry of Education, Shanghai, PR China. <sup>4</sup>Department of Neurology, Qingdao Hospital, University of Health and Rehabilitation Sciences (Qingdao Municipal Hospital), Qingdao, China. <sup>5</sup>School of Data Science, Fudan University, Shanghai, China. <sup>6</sup>Centre for Reproductive Health (CRH), Institute for Regeneration and Repair (IRR), University of Edinburgh, 4-5 Little France Drive, Edinburgh EH15 4UU, UK. <sup>7</sup>Institute of Mechanics and Computational Engineering, Department of Aeronautics and Astronautics, Fudan University, Shanghai, China. <sup>8</sup>State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China. <sup>9</sup>Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing, China. <sup>10</sup>Department of Medical Imaging, Henan Provincial People's Hospital & Zhengzhou University People's Hospital, Zhengzhou, China. <sup>11</sup>Fudan ISTBI–ZJNU Algorithm Centre for Brain-inspired Intelligence, Zhejiang Normal University, Zhejiang, China. <sup>12</sup>These authors contributed equally: Peng Ren, Xiao-He Hou. <sup>1</sup>e-mail: jintai\_yu@fudan.edu.cn; mywang@zzu.edu.cn; wcheng@fudan.edu.cn structure may play a crucial role in diagnosing, stratifying, predicting and monitoring brain disorders.

Considering that proteins are the final products of gene expression that contribute directly to variation in brain structure, profiling of the proteomic signature of brain structure has great potential for revealing the microscale mechanisms that underly brain disorders. It is well established that the presence of proteins in cerebrospinal fluid (CSF) can inform on changes in brain structure that occur in a wide range of brain disorders. In particular, total tau (or isoforms) and neurofilament light chain proteins are related to the degree of neurodegeneration in Alzheimer's disease. Measurement of the reduction of complement proteins was associated with the extent of brain atrophy in mild cognitive impairment patients<sup>8</sup>, and  $\alpha$ -synuclein is the critical protein that precedes overt neurodegeneration in Parkinson's disease<sup>9,10</sup>. Furthermore, an association between concentration in CSF and plasma has been reported for many proteins<sup>11</sup>. So that analysis of plasma can also be used to determine the level of proteins in the brain. Accordingly, measurement of the concentration of amyloid beta 42 and phosphorylated tau at amino acid 181 in plasma has been validated, by corresponding studies of CSF and brain tissue, to be a biomarker of Alzheimer's disease<sup>12,13</sup>. An association between the expression of plasma proteins and alteration in brain structure in several brain disorders, as well as in older aging, has been indicated by several studies<sup>14-17</sup>. Nevertheless, these initial studies are limited by considering only specific proteins and a small number of brain structures. They do not fully capture the comprehensive and intricate connections between plasma proteins and various aspects of brain structure, leaving significant gaps in the understanding of the underlying biological mechanisms. Whether the associations between proteins and measures of brain structure exhibit specific patterns? We hypothesise that by performing a more comprehensive analysis to uncover the details of the intricate relationship between proteomics and brain structure, it will be possible to unveil the underlying pathogenesis of a wide range of brain disorders and obtain important information regarding potential targets for therapeutic intervention. However, this scientific pursuit has been restricted by the lack of relevant data for a long time. The absence of such data, including genetic information, made it challenging to systematically explore the intricate relationships between plasma proteins and brain structure. The recent availability of large-scale datasets makes it feasible to establish robust links between the genetic regulation of proteins and that of brain structure, opening new avenues for research into the pathogenesis of neurodegenerative and neuropsychiatric disorders.

The database that has been analysed refers to 4997 participants in the UK Biobank (UKB) for whom plasma proteomic and neuroimaging data, and genetic data were available. From neuroimages, a total of 272 different MRI measures of brain structure across five categories were extracted, and the corresponding proteomic signatures were determined. Subsequently, enrichment analysis was performed to elucidate the functional significance of the proteomic signatures. Utilising genetic data, the bidirectional Mendelian randomisation (MR) analysis was performed to enhance the potential associations between proteins and brain structure, as well as between proteins and brain disorders. Finally, the potential mediating effect of brain structure on the association between proteins and brain health was investigated. Overall, we are dedicated to constructing a comprehensive atlas depicting the pattern of association between large-scale proteins and brain structure, and their implications in neurodegenerative and neuropsychiatric disorders.

#### Results

A flow diagram of the study design is shown in Fig. 1. The present study included participants from UKB for whom both plasma proteomic and neuroimaging data are available. A total of 4900 healthy participants with a mean age of 63.0 were selected after excluding major medical

conditions (Table 1 and Supplementary Data 1). The study used neuroimaging measures derived from the UKB, comprising 222 regional grev matter (GM) metrics assessing volume, thickness, and surface area across brain regions, alongside 54 white matter (WM) microstructure parameters that captured axonal integrity through fractional anisotropy (FA) and mean diffusivity (MD) measurements in major WM tracts (Supplementary Data 2). To enable Mendelian randomisation analysis between protein and brain structure, and between protein and disorders, the imputed genetic dataset released by UKB in July 2017 was used for a genome-wide association study (GWAS) of brain structure. The publicly available GWAS summary statistics of brain disorders were curated (Supplementary Data 3). Furthermore, a total of 3270 participants with a mean age of 54.2 were included in the mediation analysis, which demonstrated brain structure mediates the association between proteins and brain health (Supplementary Data 4).

#### Plasma proteins associated with brain structure

The analysis began with an investigation of the association between 2920 plasma proteins and 272 brain structure measures in five categories in 4900 participants (2642 females and 2355 males). Only a few of the included patients were diagnosed with the brain diseases (Supplementary Data 5). Details of the inclusion and exclusion criteria and participant demography can be found in Table 1 and Supplementary Fig 1. Overall, 5358 significant associations were identified between 1143 proteins and 256 brain structure measures across the five categories, with multiple comparisons corrected using the false discovery rate (FDR) method ( $P_{FDR} < 0.05$ ). Significant positive and negative associations were found in each category of brain structure measures. However, the proteins that exhibited the most significant associations differed between categories (Fig. 2A and Supplementary Data 6). In particular, for volume, NCAN showed the strongest association with the left rostral middle frontal cortex (standardised beta = 0.089,  $P_{FDR}$  = 2.40 × 10<sup>-11</sup>); for thickness, OXT showed the strongest association with the superior temporal cortex (standardised beta = -0.085,  $P_{FDR} = 3.58 \times 10^{-5}$ ), and for surface area, MOG showed the strongest association with the rostral middle frontal cortex (standardised beta = 0.097,  $P_{FDR}$  = 3.75 × 10<sup>-15</sup>). For measures of the WM tracts, LEP showed the strongest association with FA in the right corticospinal tract (standardised beta = 0.105,  $P_{FDR}$  = 7.61 × 10<sup>-6</sup>) and MD in the right medial lemniscus tract (standardised beta = -0.142,  $P_{FDR} = 9.38 \times 10^{-15}$ ).

From the view of brain structure measures, significant variations were observed in the number of associations among the different brain structure categories, specifically, the highest number of significant associations were observed for the WM MD measure (2501), followed by GM volume (1370), surface area (719) and thickness (682), and the least were found for WM FA (86) (Supplementary Data 6). In addition, substantial variations were observed for different brain regions and WM tracts within the same category (Fig. 2B). For the five categories, the highest number of associated proteins were with volume of left medial orbitofrontal, surface area of right pars orbitalis, thickness of left temporal pole, FA of right corticospinal tract and MD of right anterior thalamic radiation tract, respectively (see Supplementary Data 7). From the perspective of proteins, NCAN, SLITRK1, MOG, PTPRN2 and SEZ6L were the top five proteins that showed the highest total number of significant positive associations, and LEP, OXT, PAEP, CCN5 and XG were the top five proteins that showed the highest total number of significant negative associations (Fig. 2C). Furthermore, NCAN, LEP, PTPRN2, SLITRK1 and GFAP are the top five proteins with the highest number of significant associations for each structural category (Fig. 2D), and each was represented in at least two categories. The largest intersection between the proteins associated with the five structural categories was that between volume and MD, followed by volume and surface area (Fig. 2E). In sensitivity analyses, compared to



**Fig. 1** | **Schematic overview of study design. A** The association between 2920 plasma proteins and measures of brain structure from five categories. **B** The annotation of preferred chromosome positions and functional categories for proteins associated with brain structure. Functional categories were annotated based on Gene Ontology (https://geneontology.org). **C** The protein-brain structure associations satisfied FDR-corrected P < 0.05 were further used for MR analysis that enhanced the relationships between plasma proteins and brain structures. Since the MR requires the cohort of exposure and outcome not to overlap, the GWAS of brain structure measures were conducted based on UKB participants for whom neuroimaging data was available but proteomic data was not. The proteins exhibited significant MR associations with brain structure measures (FDR-corrected P < 0.05) and were eligible for the following three sub-analyses, i.e., SMR, protein-disease MR and mediation analyses. **D** To explain the possible mechanism of why

associations with only protein expression, age at imaging visit, sex, total intracranial volume and scanning site as the predictors, the inclusion of additional covariates didn't change the pattern of significant associations (Supplementary Data 8). In addition, the imputation procedure did not affect the pattern of significant associations between protein and brain structure (Supplementary Data 9). Furthermore, we found that the effects of protein expression were highly correlated between females and males, with a correlation coefficient of 0.88 for GM volume, 0.86 for cortical area, 0.89 for cortical thickness, 0.91 for WM FA and 0.85 for WM MD (Supplementary Data 10).

# Physical and functional annotation of brain structure associated proteins

To investigate whether brain structure-associated proteins favour specific chromosomes and functional categories, the density of plasma proteins and brain structure were connected, SMR analysis was performed to examine whether the pQTL of brain structure-related plasma proteins were associated with eQTL of brain cortex because of pleiotropy. **E** The MR analysis investigating whether proteins exhibiting significant association with brain structure also affect the risk of brain disorders. **F** The mediation analysis investigates whether the brain structure could mediate the association between plasma proteins associated with brain structure and mental health. The mediator and outcome were constructed as latent variables. FA fractional anisotropy, MD mean diffusivity, SNP single- nucleotide polymorphism, pQTL protein quantitative trait loci, GWAS genome-wide association study, eQTL expression quantitative trait loci, MR mendelian randomisation. Created in BioRender. Li, Z. (2025) https://BioRender.com/ ghla3ez.

discovered proteins in each chromosome and the functional enrichment of the proteins in each tissue type, brain cellular types and Gene Ontology (GO) biological processes (Fig. 3A) were examined. The brain structure associated proteins tend to be disproportionately located in chromosomes 19, 17 and 22 compared to the short length of these three chromosomes (Fig. 3B). In terms of function, the brain structure associated proteins did not exhibit any preference for specific UKB panels (Fig. 3C). With regard to GO enrichment, the proteins that exhibited significant positive and negative associations with brain structure in the five categories exhibited separate enrichment in immune and catabolic functions (Supplementary Fig. 2). Interestingly, when focused on a single category, obvious differences were observed in GO biological processes. In particular, for volume, the proteins that exhibited positive associations were enriched in GO processes relating to nervous system development and neurogenesis, while the proteins

#### Table 1 | Demography of included UKB participants for association analysis

	Volume (N = 4908)	Area ( <i>N</i> = 4997)	Thickness (N = 4997)	FA (N = 4632)	MD (N = 4632)	Р
Age at imaging visit (year)	63.0 (7.96)	63.0 (7.96)	63.0 (7.96)	63.1 (7.92)	63.1 (7.92)	0.990
Age at baseline (year)	54.0 (7.79)	54.1 (7.80)	54.1 (7.80)	54.0 (7.75)	54.0 (7.75)	0.990
Interval (year)	8.94 (1.76)	8.93 (1.75)	8.93 (1.75)	9.01 (1.76)	9.01 (1.76)	0.047
Gender	-	-	-	-	-	0.982
Female	2603 (53.0%)	2642 (52.9%)	2642 (52.9%)	2470 (53.3%)	2470 (53.3%)	-
Male	2305 (47.0%)	2355 (47.1%)	2355 (47.1%)	2162 (46.7%)	2162 (46.7%)	-
eTIV (cm <sup>3</sup> )	1551 (153)	1550 (152)	1550 (152)	1550 (153)	1550 (153)	0.997
SiteID	-	-	-	-	-	0.421
Site1	1238 (25.2%)	1264 (25.3%)	1264 (25.3%)	1202 (25.9%)	1202 (25.9%)	-
Site2	785 (16.0%)	785 (15.7%)	785 (15.7%)	781 (16.9%)	781 (16.9%)	-
Site3	2885 (58.8%)	2948 (59.0%)	2948 (59.0%)	2649 (57.2%)	2649 (57.2%)	-
Education (year)	17.4 (3.97)	17.4 (3.98)	17.4 (3.98)	17.4 (3.97)	17.4 (3.97)	0.995
Townsed index	-1.74 (2.82)	-1.74 (2.82)	-1.74 (2.82)	-1.74 (2.82)	-1.74 (2.82)	1.000
Smoking	-	-	-	-	-	1.000
Never	3038 (61.9%)	3093 (61.9%)	3093 (61.9%)	2880 (62.2%)	2880 (62.2%)	-
Previous	1679 (34.2%)	1709 (34.2%)	1709 (34.2%)	1573 (34.0%)	1573 (34.0%)	-
Current	191 (3.89%)	195 (3.90%)	195 (3.90%)	179 (3.86%)	179 (3.86%)	-
Drinking	-	-	-	-	-	1.000
Never	155 (3.16%)	158 (3.16%)	158 (3.16%)	146 (3.15%)	146 (3.15%)	-
Previous	135 (2.75%)	137 (2.74%)	137 (2.74%)	124 (2.68%)	124 (2.68%)	-
Current	4618 (94.1%)	4702 (94.1%)	4702 (94.1%)	4362 (94.2%)	4362 (94.2%)	-
Ethnicity	-	-	-	-	-	1.000
White	4755 (96.9%)	4842 (96.9%)	4842 (96.9%)	4489 (96.9%)	4489 (96.9%)	-
Mixed	34 (0.69%)	34 (0.68%)	34 (0.68%)	33 (0.71%)	33 (0.71%)	-
Asian or Asian British	42 (0.86%)	43 (0.86%)	43 (0.86%)	37 (0.80%)	37 (0.80%)	-
Black or Black British	36 (0.73%)	37 (0.74%)	37 (0.74%)	34 (0.73%)	34 (0.73%)	-
Chinese	10 (0.20%)	10 (0.20%)	10 (0.20%)	10 (0.22%)	10 (0.22%)	-
Others	31 (0.63%)	31 (0.62%)	31 (0.62%)	29 (0.63%)	29 (0.63%)	-

The unadjusted P-values represent the statistical significance of overall differences in demographic characteristics across different brain structural metrics. Categorical variables were compared between groups using Fisher's exact test, and One-way ANOVA was used for continuous variables. All statistical tests were two-sided. FA fractional anisotropy, MD mean diffusivity.

that exhibited negative associations were enriched in GO processes related to immune function, and for MD, the proteins that exhibited positive associations were enriched in GO processes related to apoptosis, while the proteins that exhibited negative associations were enriched in GO processes related to morphogenesis. No significant enrichments in GO processes were observed for the other three brain structure categories (i.e., thickness, surface area and FA) (Fig. 3D).

In terms of tissue enrichment, the proteins that exhibited association with volume and surface area were enriched in up-regulated differential expression gene (DEG) of brain tissue and the proteins associated with MD were enriched in down-regulated DEG of brain tissues (Fig. 3E). In addition, separate analysis revealed that the tissue enrichment observed for volume and area was attributable to positively associated proteins, while the observed enrichment for MD was attributable to negatively associated proteins (Supplementary Fig 3 and 4). Given the significant enrichment of brain structureassociated proteins in brain tissue, consideration was given to whether the coding gene of these proteins prefers to be expressed in specific brain cell types. The volume-associated proteins were enriched in excitatory and inhibitory neurons, while the thickness-associated proteins were enriched in astrocytes and endothelial cell types (Fig. 3F).

# Associations between proteins and brain structure identified through MR analysis

To further enhance the observed extensive associations between proteins and brain structure measures, bidirectional MR analyses were

performed for 5358 protein-brain structure pairs that exhibited significant associations. Utilising inverse-variance weighted (IVW) methods, the forward MR results revealed 33 significant MR relationships between 32 proteins and brain structure across the five categories (15 GM regions and 8 WM tracts) at FDR-corrected p < 0.05 $(p < 3.08 \times 10^{-4})$ . Notably, the cingulate gyrus part of the cingulum, parahippocampal part of the cingulum and thalamic radiation were the tracts that exhibited the most MR associations with proteins. Of the 33 significant forward MR relationships, MD showed not only the highest number of significant MR relationships, but also the highest effect size of associations. The increased protein expression was found, through MR analysis, to be significantly associated with lower MD of WM tracts, with the most notable finding being the association between increased BTN2A1 protein expression and lower MD of the right parahippocampal part of the cingulum (beta = -0.05, 95% CI -0.08 to -0.03,  $P_{FDR} < 0.001$ ). On the contrary, increased protein expression was associated with higher FA of WM tracts, among which the most significant one is that increased expression of REN protein was associated with higher FA of left anterior thalamic radiation (beta = 0.12, 95% CI 0.06 to 0.19,  $P_{FDR} < 0.05$ ). (Fig. 4). When a less strict significance threshold of nominal p < 0.001 was applied, a total of 59 proteins, 25 GM regions and 16 WM tracts were involved in 71 potential MR relationships. Interestingly, more brain-enriched proteins, including NCAM1, OXT, and NPTXR, emerged as having a significant association with the measures of brain structure. Specifically, higher expression of NCAM1 and OXT proteins were associated with lower volume of the right precentral and lateral orbitofrontal cortex, and a



Fig. 2 | Association between plasma proteins and brain structures. A For each modality, each dot represents the most significant association between the specific protein and all corresponding metrics (i.e., regional or tract-wise) in Fig. 2B. Positive associations are shown in red and negative associations are shown in blue. Standardised coefficients are shown. **B** The distribution of the count of significantly associated proteins across different brain regions and tracts. Positive and negative associations are shown together. **C** The top 30 proteins with the highest total

number of significantly positive (top panel) and negative (bottom panel) associations across modalities. **D** The top five proteins with the highest number of significant associations for each modality. **E** The UpSet plot showing the relationship of significantly associated proteins between different modalities. The counts of associations are shown in rows, and the categories of proteins are shown in columns. Source data are provided as a Source Data file. FA fractional anisotropy, MD mean diffusivity.

higher expression of NPTXR protein was associated with higher FA of left posterior thalamic radiation (Fig. 4). We found that most of the significant protein-brain structure remained significant when analysed using at least one of the other MR methods (Supplementary Data 11). In sensitivity analysis, most of the MR analyses showed no evidence of directional pleiotropy (MR-Egger intercept) or horizontal pleiotropy (MR-PRESSO). For three out of five MR relationships exhibiting directional pleiotropy, in which the assumption of the instrumental variable was violated, the results from the other three MR



methods remained highly significant (Supplementary Data 11 and 12). To confirm the reliability of the MR relationships, the forward MR analyses were repeated with a stricter clumping at a *p* threshold of  $5 \times 10^{-8}$ , the results of which remained significant and showed a consistent direction of effect with those using clumping at the *p* threshold of  $5 \times 10^{-6}$  (Supplementary Fig. 5 and Supplementary Data 13 and 14).

In reverse MR analyses, we identified 42 significant associations between brain structural measures (15 GM regions and 8 WM tracts) and 34 proteins that survived FDR correction p < 0.05 ( $p < 3.75 \times 10^{-4}$ ). Of the significant reverse MR relationships, we observed 11 MR associations for GM volume (most significant, Left precentral-CA14, beta = 0.15,  $p = 1.48 \times 10^{-5}$ ), 9 for GM area (most significant, Left

Fig. 3 | Physical and functional annotation of brain structure associated proteins. A The overview of the annotation categories, including protein-coding gene position, tissue enrichment, cellular enrichment and GO biological process enrichment. B The distribution of coding genes for brain structure association proteins across different chromosomes. The density of the associated proteins is shown, which accounts for the length of the chromosomes. C The distribution of brain structure association proteins across different UKB-defined protein categories. D The enrichment of coding genes for brain structure-associated proteins across GO biological processes. The positively and negatively associated proteins are shown separately, only volume and MD measures showed significant enrichment. E The enrichment of coding genes for brain structure-associated proteins across tissues. The GTEx v8 gene expression data and FUMA GENE2FUN were used.

fusiform-MOG, beta = 0.21,  $p = 3.77 \times 10^{-7}$ ), 7 for GM thickness (most significant, Right superiortemporal-TGFA, beta = -0.16,  $p = 7.92 \times 10^{-5}$ ) and 15 for WM MD (most significant, forceps minor-KLK6, beta = 0.11,  $p = 2.30 \times 10^{-7}$ ). No bidirectional MR relationships were found between proteins and brain structure measures (Supplementary Fig. 7 and Supplementary Data 15). We found that most of the significant brain structure-protein relationships remained significant when analysed using at least one of the other MR methods (Supplementary Data 15). In sensitivity analysis, only one MR relationship showed evidence of directional pleiotropy, for which the results from the other three MR methods remained highly significant (Supplementary Data 16). The reverse MR analyses repeated with a stricter clumping at the *p* threshold of  $5 \times 10^{-8}$  showed a significant and generally consistent direction of effect with those using clumping at the *p* threshold of  $5 \times 10^{-6}$  (Supplementary Data 17 and 18).

To further control the false positive rate in the presence of pleiotropic effects, we performed Causal Analysis Using Summary Effect Estimates (CAUSE) analysis to distinguish causality and genetic correlation. Almost no significant difference was found between the genetic correlation model and causal model at FDR-corrected P < 0.05. However, at a clumping *P* threshold of  $5 \times 10^{-6}$ , 57 out of 71 proteinbrain structure associations tended to be better characterised by causal model than genetic correlation models (i.e., delta\_elpd < 0). Similarly, at a clumping *P* threshold of  $5 \times 10^{-8}$ , 39 out of 59 proteinbrain structure associations tended to be better characterised by a causal model than genetic correlation models (Supplementary Data 11 and 13).

# Pleiotropic association between plasma proteins and brain gene expression

Given the discovered MR relationship between 32 plasma proteins and brain structures, an intuitive question would be how these peripheral plasma proteins connect to the brain structural measures. One possible mechanism is that the peripheral plasma and brain share the same regulatory relationships. This possibility was investigated by first examining the gene expression of 32 proteins that were associated with brain structure in different tissues. While ENPP6, POXDL2, and HPCAL1 were highly enriched in brain tissues, some proteins were expressed in brain tissues without showing enrichment, and others were present only at very low expression (Fig. 5). Next, to test for pleiotropic association between the protein expression and gene expression. The Summary-based Mendelian Randomisation (SMR) analysis found that SULT1A1, REN, AGER, F11R, LEPR, and PAMR1 in the plasma exhibited pleiotropic association with the expression of their protein-coding gene in the brain, with  $P_{\text{SMR}} < 0.05/32$  and *P*-value of heterogeneity in dependent instruments (HEIDI) > 0.05. In other words, these plasma proteins shared regulation mechanism with the brain.

# Associations between proteins and disorders identified through MR analysis

To investigate whether the proteins associated with brain structure also exhibited associations with the risk of disorders, the bidirectional The enrichment in brain tissue is highlighted in the grey bar. The significant enrichments after the FDR correction are shown with a black border. **F** The enrichment of brain structure-associated proteins across brain cell types. The cell type gene sets are obtained elsewhere, the significant enrichments are shown with coloured circles. For all the enrichment analyses, the statistical significance was determined with hypergeometric tests and the significance threshold was set to FDR-corrected P < 0.05. The UKB 2,920 proteins are used as a background. All statistical tests were two-sided. Source data are provided as a Source Data file. GO go ontology, OPC oligodendrocyte precursor cell, FA fractional anisotropy, MD mean diffusivity, DEG differential expression gene. Created in BioRender. Li, Z. (2025) https://BioRender.com/ghla3ez.

MR analyses were performed between the 32 proteins that exhibited significant MR associations with brain structure and ten central nervous system disorders, with a significance threshold at FDR correction p < 0.05 ( $p < 3.40 \times 10^{-3}$ ). For the associations between protein and disease, a total of 21 significant protein-disease relationships were observed, corresponding to nine proteins and ten disorders, with the effect size expressed as odds ratio (OR). The strongest MR relationships were discovered between FKBP5 and amyotrophic lateral sclerosis (ALS; OR = 0.48, 95% CI = 0.30 to 0.78,  $p = 2.81 \times 10^{-3}$ ) and between INHBC and ALS (OR = 1.32, 95% CI = 1.10 to 1.59,  $p = 2.85 \times$  $10^{-3}$ ). Significant associations were found for all ten central nervous system disorders. Except in the case of attention deficit hyperactivity disorder, increased expression of BTN3A2 was associated with higher risk of Alzheimer's disease (AD; OR = 1.09, 95% CI = 1.06 to 1.12, p = $2.04 \times 10^{-9}$ ), anxiety (OR = 1.03, 95% CI = 1.01 to 1.04,  $p = 1.51 \times 10^{-5}$ ), bipolar disorder (BP; OR = 1.08, 95% CI = 1.05 to 1.12,  $p = 3.27 \times 10^{-6}$ ), major depressive disorder (MDD; OR = 1.07, 95% CI = 1.05 to 1.09, p = $2.66 \times 10^{-9}$ ) and schizophrenia (SCZ; OR = 1.23, 95% CI = 1.21 to 1.25,  $p = 1.65 \times 10^{-93}$ ). On the contrary, increased levels of BTN2A1 proteins exhibited a protective effect on the risk of these disorders. BTN2A1 and BTN3A2 proteins showed the strongest associations with BP and SCZ. Moreover, the increased expression of ENPP6 was associated with a lower risk of Parkinson's disease (PD; OR = 0.86, 95% CI = 0.79 to 0.93,  $p = 1.72 \times 10^{-4}$ ) and multiple sclerosis (MS; OR = 0.68, 95% CI = 0.56 to 0.82,  $p = 4.82 \times 10^{-5}$ ). (Fig. 6 and Supplementary Data 19). We found that most of the significant protein-disease relationships remained significant when analysed using at least one of the other MR methods (Supplementary Data 19).

Sensitivity analysis revealed that most of the MR analyses showed no evidence of directional pleiotropy (MR-Egger intercept) or horizontal pleiotropy (MR-PRESSO). Furthermore, for four MR relationships exhibiting directional pleiotropy, which violated the assumptions of instrument variables for MR, the results from the other three MR methods (weighted mean, weighted and MR-Egger) remained highly significant (Supplementary Data 19 and 20). In order to confirm the reliability of the MR relationships that have been discovered, the MR analysis was repeated using a stricter clumping *p* threshold of  $5 \times 10^{-8}$ , in which case 17 out of 21 MR relationships remained significant and showed consistent direction of effect (see Supplementary Fig. 6, Supplementary Data 21 and 22).

In reverse MR analyses, we identified 11 significant associations between seven brain disorders and nine proteins that survived FDR correction p < 0.05 ( $p < 1.96 \times 10^{-3}$ ). Of the significant reverse MR relationships (Supplementary Fig. 8), the brain disorders demonstrated associations with the increased expression of six proteins (PAMR1, MAVS, F11R, REN, GER and LEPR) and decreased expression of three proteins (TNFRSF4, BTN2A1, ENPP6). The bidirectional relationships were only found between ENPP6 and MS, and between BTN2A1 and BP. We found that most of the significant disease-protein relationships remained significant when analysed using at least one of the other MR methods (Supplementary Data 23). In sensitivity analysis, No MR relationship showed evidence of directional pleiotropy

	Exposure (Protein)	Outcome (Imaging metric)	/Vs N		Beta (95%CI)	<i>P</i> value
Volume	ITGB6 ENPP6 AGER BLMH NCAM1 OXT NELL2 PTPRZ1 ADM ZNRD2 NCAM1 CHAD OMD	Left middletemporal Left putamen Left bankssts Right postcentral Right precentral Right lateralorbitofrontal Left supramarginal Right pallidum Right putamen Right thalamus Right superiortemporal Left putamen Right thalamus	56 69 101 76 64 55 85 76 29 44 64 43 54		$\begin{array}{c} 0.07 \ (0.04, \ 0.10) \\ 0.06 \ (0.03, \ 0.08) \\ 0.05 \ (0.02, \ 0.07) \\ -0.05 \ (-0.08, \ -0.02) \\ -0.05 \ (-0.07, \ -0.02) \\ 0.05 \ (0.02, \ 0.09) \\ 0.06 \ (0.03, \ 0.10) \\ -0.11 \ (-0.17, \ -0.05) \\ 0.07 \ (0.03, \ 0.11) \\ -0.04 \ (-0.06, \ -0.02) \\ 0.07 \ (0.03, \ 0.12) \\ 0.06 \ (0.03, \ 0.10) \end{array}$	3.36e-05 * 3.88e-05 * 1.08e-04 * 3.08e-04 * 5.23e-04 6.22e-04 6.64e-04 7.41e-04 7.81e-04 8.15e-04 9.08e-04
Area	LRRC37A2 LEPR PODXL2 ITGA11 VWC2L CRLF1	Right lingual Right supramarginal Left rostralmiddlefrontal Left inferiorparietal Left lateralorbitofrontal Right insula	82 38 69 39 13 49	₩  +=   +1  +=   -=-	-0.04 (-0.05, -0.03) 0.09 (0.05, 0.13) 0.05 (0.03, 0.08) -0.07 (-0.10, -0.03) 0.17 (0.08, 0.27) -0.08 (-0.13, -0.03)	4.15e-09 <b>*</b> 3.17e-05 <b>*</b> 5.38e-05 <b>*</b> 2.74e-04 <b>*</b> 2.95e-04 <b>*</b> 7.82e-04
Thick	INHBC FAS CA3 PAMR1 PLAT CA14 IL31RA PRSS8 TGFBR2 NHLRC3	Left superiortemporal Right lateraloccipital Left superiorfrontal Left temporalpole Left lateraloccipital Right parahippocampal Right precuneus Left temporalpole Right lingual Left medialorbitofrontal	56 39 39 83 26 50 54 46 63 46	H=1   ↓=1   ↓=1   ↓=1   ↓=1   ↓=1   ↓=1   ↓=1	$\begin{array}{c} -0.07 \ (-0.10, \ -0.04) \\ -0.14 \ (-0.20, \ -0.08) \\ 0.10 \ (0.05, \ 0.15) \\ -0.06 \ (-0.09, \ -0.03) \\ 0.16 \ (0.07, \ 0.24) \\ -0.10 \ (-0.15, \ -0.05) \\ -0.04 \ (-0.06, \ -0.02) \\ -0.10 \ (-0.16, \ -0.05) \\ 0.08 \ (0.03, \ 0.13) \\ -0.08 \ (-0.13, \ -0.03) \end{array}$	4.41e-06 <b>*</b> 8.69e-06 <b>*</b> 1.10e-04 <b>*</b> 2.07e-04 <b>*</b> 2.54e-04 <b>*</b> 3.73e-04 4.23e-04 6.51e-04 8.59e-04
FA	REN NPTXR REN	Left anterior thalamic radiation Left posterior thalamic radiation Left superior thalamic radiation	43 40 43	·	0.13 (0.06, 0.19) 0.09 (0.04, 0.15) 0.12 (0.05, 0.18)	2.79e-04 <b>*</b> 3.66e-04 5.38e-04
GW	BTN3A2 BTN3A2 BTN3A2 BTN3A2 BTN3A2 BTN3A2 BTN3A2 BTN3A2 STN3A2 STN3A2 BTN3A2 FAM20A PARP1 F11R COMT FKBP5 MAVS MSRA CCL28 IKBKG HPCAL1 COMT SETMAR IKBKG SETMAR IKBKG SETMAR IKBKG SETMAR IKBKG SETMAR IKBKG IRAK1 CD300LG HEXIM1 PTPRF PARP1 RHOC HEXIM1 IKBKG IRAK1 PTPRF PARP1 RHOC HEXIM1 FYB1 SCAMP3 ATOX1 PAK4	Left cingulate gyrus part of cingulum Right parahippocampal part of cingulum Right parahippocampal part of cingulum Left cingulate gyrus part of cingulum Right anterior thalamic radiation Left cingulate gyrus part of cingulum Right corticospinal tract Left anterior thalamic radiation Left cingulate gyrus part of cingulum Right corticospinal tract Left anterior thalamic radiation Left cingulate gyrus part of cingulum Right concent factor Left cingulate gyrus part of cingulum Right acoustic radiation Left cingulate gyrus part of cingulum Right accustic radiation Left cingulate gyrus part of cingulum Right accustic radiation Left cingulate gyrus part of cingulum Right accustic radiation Left cingulate gyrus part of cingulum Right actor thalamic radiation Left inferior fronto-occipital fasciculus Right cingulate gyrus part of cingulum Left superior thalamic radiation Left inferior longitudinal fasciculus Right cingulate gyrus part of cingulum Right superior thalamic radiation Left cingulate gyrus part of cingulum Right inferior longitudinal fasciculus Right superior thalamic radiation Left cingulate gyrus part of cingulum Right ringulate gyrus	<ul> <li>43</li> <li>73</li> <li>97</li> <li>60</li> <li>29</li> <li>32</li> <li>41</li> <li>21</li> <li>53</li> <li>9</li> <li>33</li> <li>31</li> <li>28</li> <li>21</li> <li>28</li> <li>78</li> <li>19</li> <li>27</li> <li>29</li> <li>28</li> <li>18</li> <li>30</li> <li>61</li> <li>11</li> <li>25</li> <li>29</li> <li>18</li> <li>19</li> <li>65</li> <li>66</li> <li>29</li> <li>26</li> <li>18</li> <li>20</li> </ul>		$\begin{array}{c} 0.12 (0.00, 0.16) \\ 0.08 (0.06, 0.09) \\ -0.06 (-0.08, -0.03) \\ -0.11 (-0.16, -0.06) \\ -0.14 (-0.21, -0.08) \\ -0.15 (-0.22, -0.08) \\ 0.08 (0.04, 0.13) \\ -0.18 (-0.27, -0.09) \\ 0.06 (0.03, 0.09) \\ 0.17 (0.08, 0.26) \\ -0.13 (-0.19, -0.06) \\ -0.12 (-0.18, -0.06) \\ -0.12 (-0.18, -0.05) \\ -0.15 (-0.23, -0.07) \\ 0.07 (0.03, 0.11) \\ -0.20 (-0.31, -0.09) \\ -0.12 (-0.19, -0.06) \\ -0.12 (-0.19, -0.06) \\ -0.12 (-0.19, -0.06) \\ -0.12 (-0.19, -0.06) \\ -0.12 (-0.19, -0.06) \\ -0.12 (-0.19, -0.06) \\ -0.12 (-0.19, -0.06) \\ -0.12 (-0.19, -0.06) \\ -0.12 (-0.19, -0.06) \\ -0.12 (-0.19, -0.06) \\ -0.12 (-0.19, -0.06) \\ -0.12 (-0.19, -0.06) \\ -0.09 (-0.14, -0.04) \\ -0.21 (-0.32, -0.09) \\ -0.10 (-0.15, -0.04) \\ -0.06 (-0.09, -0.02) \\ -0.13 (-0.20, -0.06) \\ 0.06 (0.03, 0.10) \\ -0.07 (-0.11, -0.03) \\ 0.15 (0.06, 0.24) \\ 0.14 (0.06, 0.23) \\ -0.20 (-0.31, -0.08) \\ -0.21 (-0.32, -0.05) \\ -0.20 (-0.31, -0.08) \\ -0.13 (-0.20, -0.05) \\ -0.13 (-0.21, -0.05) \\ -0.13 (-0.21, -0.05) \\ -0.15 (-0.24, -0.06) \\ -0.16 (-0.25, -0.06) \\ \end{array}$	1.24e-04 1.24e-26 * 5.92e-07 * 9.29e-06 * 2.16e-07 * 5.58e-05 * 5.58e-05 * 6.69e-05 * 1.12e-04 * 1.32e-04 * 1.32e-04 * 2.36e-04 * 2.28e-04 * 2.43e-04 * 2.43e-04 * 2.51e-04 * 2.78e-04 * 2.78e-04 * 3.37e-04 * 3.52e-04 * 5.52e-04
			-0.4	4 -0.2 0 0.2	0.4	

Effect estimate (Beta) and 95% CIs

# **Fig. 4** | **The association between protein and brain structure in the forward MR.** The forest plot shows the significant MR relationships of the IVW method at a clumping *P* threshold of $5 \times 10^{-6}$ . The significance was determined with a Wald test,

and all MR results at a nominal P < 0.001 are shown. Raw P-values are shown in the

rightmost column. The MR relationships that meet the significance threshold of

FDR-corrected P < 0.05 are marked with an asterisk. The centre of the error bar means the estimated effect of the MR relationship using the IVW method. All statistical tests were two-sided. Source data are provided as a Source Data file. FA fractional anisotropy, MD mean diffusivity, IV independent variant.



Fig. 5 | Association between brain structure-associated proteins and gene expression of brain. The expression of coding genes across tissues for proteins that exhibited significant MR associations with brain structure. Source data are provided as a Source Data file. TPM transcripts per million.

(Supplementary Data 24). The reverse MR analyses repeated with a stricter clumping at the p threshold of  $5 \times 10^{-8}$  showed generally consistent results with those using clumping at the p threshold of  $5 \times 10^{-6}$  (Supplementary Data 25 and 26).

To further control the false positive rate in the presence of pleiotropic effects, the CAUSE analysis found almost no significant difference between the genetic correlation model and causal model at FDR-corrected P < 0.05. However, at a clumping P threshold of  $5 \times$  $10^{-6}$ . 16 out of 21 protein-disease associations tended to be better characterised by the causal model than the genetic correlation models (i.e., delta\_elpd < 0). At a clumping P threshold of  $5 \times 10^{-8}$ , 27 out of 34 protein-disease associations tended to be better characterised by causal model than the genetic correlation models (Supplementary Data 19 and 21)

#### Brain structure mediated association between protein and mental health

Since the plasma proteins were associated with the brain structures and the brain structure-mental health connections were already welldocumented, we further asked whether the brain structure could mediate the association between protein and mental health. A total of 3270 participants with a mean age of 54.2 were eligible for inclusion in mediation analysis (Supplementary Data 4). The proteins that exhibited significant MR associations with brain structure at FDR p < 0.05were considered, and by using a structural equation model (SEM) model, we constructed latent mediators with protein-associated brain structures and latent outcomes with mental health phenotypes<sup>18</sup>. A total of four significant mediation relationships were identified, including one full mediation and three partial mediations (FDR-corrected p < 0.05). In particular, the regional volume of right precentral and left putamen could fully mediate 65% of the negative association between ENPP6 protein expression and latent mental health phenotype. Furthermore, the surface area of the right supramarginal mediated 16% of the negative association between ITGA11 protein expression and latent mental health phenotype. For all four mediation relationships, depression and anxiety were the observable variables contributing most to the latent mental health phenotype (Fig. 7 and Supplementary Data 27). For exploratory purposes, a mediation analysis was conducted for those proteins that exhibited significant protein-neuroimaging MR relationships at a less strict significance threshold of nominal P < 0.001. An additional six mediation relationships were revealed, among which half (NCAM1, OMD, and PTPRZ1) were full mediations. In particular, the regional volume of the left supramarginal gyrus and right thalamus could fully mediate 63% of the negative relationship between NCAM1 protein expression and mental health, as well as 53% of the negative relationship between OMD protein expression and mental health. The regional volume of the right precentral gyrus and basal ganglia could fully mediate 26% of the negative relationship between PTPRZ1 protein and mental health. Furthermore, the regional volume of right medial orbitofrontal and left middle temporal mediates the positive association between OXT protein expression and mental health phenotype (Fig. 7 and Supplementary Data 27). Interestingly, ENPP6, NCAM1, NELL2, PTPRZ1, and OXT proteins were found to be highly enriched in brain tissues (Supplementary Fig. 9).

## Discussion

This study of the UK Biobank database has provided comprehensive insights regarding the plasma proteomic signature of brain structure. Altogether, 256 out of 272 brain structure measures from all five categories demonstrated significant associations with plasma proteins. The MD of WM tracts not only exhibited the highest number of associated proteins but also the highest strength of association. The highest number of common proteins was shared by the MD of WM tracts and volume measures, NCAN, SLITRK1, and MOG had the highest number of positive associations, while LEP, OXT, and PAEP had the highest number of negative associations. Functional enrichment analysis revealed that the proteins associated with brain structure were enriched in neurogenesis, brain tissues, and neuronal cell types, as well as immune and apoptotic-related processes. The proteins exhibited significant associations with brain structure measures also demonstrated associations with the risk of neurodegenerative and psychiatric disorders. Moreover, significant associations were discovered between identified proteins and mental health phenotypes, which could be mediated by proteinrelated brain structural measures. Collectively, the findings form a comprehensive atlas delineating the association pattern between largescale proteins and brain structure, as well as their implications in neurodegenerative and neuropsychiatric disorders.

Exposure (Protein)	Outcome (Disease)	/Vs N		OR (95%CI)	P value
BTN3A2		97	н	1.09 (1.06, 1.12)	2.04e-09
BTN2A1	AD	102	H=1	0.91 (0.87, 0.95)	8.43e-05
FAM20A		50	++	0.88 (0.82, 0.95)	9.03e-04
BTN3A2		70	н	0.95 (0.93, 0.98)	1.05e-04
BTN2A1	ADHD	89	H	1.08 (1.04, 1.12)	1.80e-04
FKBP5	ALS	29	⊢1	0.48 (0.30, 0.78)	2.81e-03
INHBC	, LO	64	<b>⊢</b> -	1.32 (1.10, 1.59)	2.85e-03
BTN2A1	Anxietv	82	н	0.96 (0.95, 0.98)	1.01e-05
BTN3A2	,	55	ж	1.03 (1.01, 1.04)	1.51e-05
LRRC37A2	ASD	72	H	1.06 (1.03, 1.10)	6.18e-05
BTN2A1		71	H	0.86 (0.81, 0.90)	1.11e-09
BTN3A2	BP	73	H	1.08 (1.05, 1.12)	3.27e-06
FAM20A		44	┝╼┥	1.13 (1.06, 1.21)	3.58e-04
BTN3A2	MDD	69	н	1.07 (1.05, 1.09)	2.66e-09
ENPP6	MS	54	┝╼╌┥	0.68 (0.56, 0.82)	4.82e-05
ENPP6	PD	74	++	0.86 (0.79, 0.93)	1.72e-04
BTN3A2		63	н	1.23 (1.21, 1.25)	1.65e-93
LRRC37A2		72	н	0.96 (0.94, 0.97)	4.43e-10
BTN2A1	SCZ	67	H	0.86 (0.82, 0.90)	7.12e-10
TNFRSF4		42	⊦ <del>∎</del> -	0.87 (0.82, 0.94)	1.08e-04
HEXIM1		23	<b>⊦</b> ∎-	0.90 (0.83, 0.96)	3.40e-03
		0	0.5 1 Effect estimate (OR) and 95%	1.5 2 Cls	

**Fig. 6** | **The association between protein and disease in the forward MR.** The forest plot shows the significant MR relationships of the IVW method at a clumping *P* threshold of  $5 \times 10^{-6}$ . Raw *P*-values are shown in the rightmost column. The significance was determined with a Wald test, and the MR relationships meet the significance threshold of FDR-corrected *P* < 0.05 are shown. The centre of the error bar means the estimated effect of the MR relationship using the IVW method. All

statistical tests were two-sided. Source data are provided as a Source Data file. ADHD attention deficit hyperactivity disorder, ASD autism spectrum disorder, BP bipolar disorder, MDD major depressive disorder, SCZ schizophrenia, AD Alzheimer's disease, PD Parkinson's disease, MS multiple sclerosis, ALS amyotrophic lateral sclerosis, IV independent variant, OR odd ratio, CI confidence interval.

Brain structure and its individual variance manifest a pivotal role in brain disorders. Nevertheless, only a few studies have attempted to find a potential link between plasma protein expression and brain structure, among which are reports that neurology-related proteins are associated with brain and total GM volume in normal aging<sup>17</sup> and the MR associations have been reported between 11 blood proteins and volumes of subcortical brain structures<sup>16</sup>. Instead of limiting the association and MR analysis to a narrow scope of proteins or brain structure measures, a more comprehensive atlas of association patterns between proteins and brain structure has been performed in the present study, which has the potential to reveal insights that are both multi-scale and multi-faceted. Accordingly, much more extensive associations and MR relationships between plasma proteins and brain structure were observed in our study. Altogether, 5358 associations were observed between 1143 proteins and 79 GM volume, 65 GM surface area, 65 GM thickness, 20 FA of WM tracts, and 27 MD of WM tracts. MD of WM tracts exhibited the largest number of associations and MR relationships, followed by regional GM volume and surface area. Though MD and FA are both metrics that reflect the microstructural integrity of white matter, MD demonstrated the most significant associations with proteins, whereas FA showed the least significant associations. The higher number of associations between MD and CSF proteins might stem from the fact that CSF proteins reflect early pathophysiological changes in the brain, to which MD is sensitive, especially in the early stages of neurological diseases such as AD<sup>19</sup>. In patients with MS, MD was more strongly associated with widespread brain degeneration than FA as well<sup>20</sup>. FA is primarily influenced by directional water diffusion and is more specific to structural coherence and fibre integrity<sup>21</sup>. This specificity may make FA less sensitive to diffuse or generalised changes and more reliant on pronounced structural disruptions.

NCAN, SLITRK1, MOG, PTPRN2, and SEZ6L were the top proteins that showed the highest total number of significant positive associations with brain structures. The links between these proteins and brain disorders have been reported in previous studies. NCAN, an extracellular matrix protein, is closely related to the development, neuronal



**Fig. 7** | **The mediation effect of brain structure on association between protein and mental health.** The SEM model was established for assessing the mediation effect, and the significance of the paths were examined with bootstrapping and *t* test. The observable variables are shown with rectangles, and the latent variables are shown with ellipses in the top row. Standardised coefficients of path a and path b, as well as the corresponding significance values, are shown. The proportion mediated by brain structure is shown on the rightmost column. The potential high

collinearity of observable variables in constructing the latent variables is addressed with an iterative selection strategy. As an exploratory analysis, all proteins that exhibited MR associations with brain structure at nominal P < 0.001 were included. The proteins exhibiting significant MR association with brain structure at FDR-corrected P < 0.05 were highlighted in bold. All statistical tests were two-sided. Source data are provided as a Source Data file. Prop proportion.

senescence, and apoptosis of neurons. NCAN has been found to be associated with neuropsychiatric and neurodegenerative diseases<sup>22</sup>. In addition, NCAN demonstrated a predominant correlation with AD pathology<sup>23</sup>. Existing evidence has linked the plasma expression of NCAN protein to total brain volume<sup>17</sup>. The findings of the present study have expanded the significant associations of NCAN protein to the volume of 50 regions, surface area of 43 regions, FA of three tracts, and MD of five tracts. Apart from NCAN, the other proteins are also relevant to the central neural system and brain diseases. SLITRK1 function is associated with the control of neurite development and synaptogenesis<sup>24</sup>. SLITRK has been linked to Tourette's syndrome, trichotillomania, and obsessive-compulsive disorder<sup>25</sup>. MOG, expressed in oligodendrocytes in the central nervous system, is an encephalitogenic protein that can trigger a demyelinating immune response<sup>26</sup>. Moreover, PTPRN2 has been implicated in AD and FTD<sup>23</sup>, while SEZ6L has been associated with BP<sup>27</sup>. The proteins with significant negative associations with brain structure include LEP, OXT, PAEP, CCN5, and XG. LEP is well-known for its role in metabolic regulation and has been implicated in neurodegenerative diseases, potentially linking metabolic dysfunction to brain pathology<sup>28</sup>. OXT is associated with social behaviours and stress responses<sup>29</sup>. It has also been implicated in anxiety, depression, and AD<sup>30</sup>, although its relationship with brain structure remains unclear. Previous studies have highlighted associations between these proteins and CNS or brain disorders, and our findings further reveal their associations with brain structure. While the roles of PAEP, CCN5, and XG in brain diseases are not yet fully understood, our study identifies significant associations between these proteins and brain structure, offering a valuable basis for future studies.

Thirty-three significant MR relationships between plasma proteins and brain structures were revealed through MR analysis. The strongest associations with proteins were observed in the cingulate gyrus part of the cingulum, parahippocampal part of the cingulum, and thalamic radiation. The cingulate gyrus, part of the cingulum, is involved in emotional regulation, attention, and memory, and disruptions in this tract have been implicated in conditions such as AD and major depressive disorder<sup>31</sup>. The parahippocampal cingulum is closely associated with episodic memory and spatial navigation, and alterations in this tract are also observed in the early stages of AD<sup>31</sup>. Our results highlight the potential involvement of the cingulum in the pathophysiology of neurodegenerative and psychiatric disorders through associations with specific proteins. The thalamic radiation serves as a critical pathway for relaying sensory and motor signals between the thalamus and cortical regions. Studies have demonstrated that alteration in the anterior thalamic radiation integrity is linked to SCZ and depression<sup>32,33</sup>. Taken together, these findings suggest that the associated proteins may play key roles in maintaining the structural integrity and functional regulation of these tracts, thereby contributing to the pathophysiology of neurodegenerative and psychiatric disorders.

Our analysis revealed that brain structure-associated proteins are disproportionately localised to chromosomes 19, 17, and 22, relative to the relatively short length of these chromosomes. The expression of brain structure-associated proteins on chromosomes 19, 17, and 22 is particularly noteworthy given the established relevance of these chromosomes to brain health and disease. Chromosome 19 has a high gene density and contains key genetic factors related to brain aging and neurodegenerative diseases, including APOE, which is strongly linked to AD<sup>34</sup>. Chromosome 17 includes *MAPT*, a gene associated with tau-related pathologies, such as FTD<sup>35</sup>. Chromosome 22 has been implicated in neurodevelopmental and psychiatric disorders, with alterations in this region contributing to schizophrenia and 22q11.2 deletion syndrome<sup>36,37</sup>. Our findings show that many of the proteins identified on these chromosomes, including those encoded on chromosome 17 (OMG<sup>38</sup>, BAIAP2<sup>39</sup>, GFAP<sup>40</sup>), chromosome 19 (NCAN<sup>23</sup>, APLP1<sup>41</sup>, GDF15<sup>42</sup>), and chromosome 22 (NPTXR<sup>43</sup>, OSM<sup>44</sup>, RTN4R<sup>45</sup>) are strongly associated with brain structure and have been previously implicated in various neurological and psychiatric disorders. These results suggest that these chromosomes may serve as hotspots for genes influencing brain morphology and function, reflecting their broader role in neurodevelopmental and disease pathways.

From the point of view of function, the proteins which showed an association with the five measures of brain structure were enriched in relatively broad functions, including immune and catabolic processes. On the other hand, the proteins associated with volume and MD showed a more specific functional relationship. As volume and MD are negatively correlated<sup>46,47</sup>, the positively associated proteins should be compared to the negatively associated proteins, and the same is true for the negatively associated proteins. Specifically, the proteins that were positively associated with volume and negatively associated with MD were both enriched in synthesis functions, such as nervous system development, neurogenesis and cell projection organisation, whereas the proteins that were negatively associated with volume and positively associated with MD were enriched in clearance related immune. and apoptotic function, respectively. The immune and apoptotic functions are different forms that are relevant to cell death, among which apoptotic cell death could eliminate cells no longer needed without initiating an immune response48. These findings greatly extend knowledge regarding protein-brain structure relationships and highlight the potential functional preference of associated proteins for different brain structures.

With regard to the MR association between brain structureassociated plasma proteins and brain disorders, significant relationships were observed for all ten brain disorders. Among the nine proteins with significant MR associations, BTN3A2, a member of the immunoglobulin superfamily playing a role in T-cell responses within anxiety, attention deficit hyperactivity disorder (ADHD), BP, MDD, and SCZ. This finding is supported by prior evidence that the genetic variants of BTN3A2 showed pleiotropic effects on SCZ, BP, and MDD<sup>49</sup>, as well as that increased expression of BTN3A2 might confer risk for SCZ by altering excitatory synaptic function<sup>50</sup>. The findings of the present study extend knowledge of the role of BTN3A2 to a broader range of disorders. On the contrary, we found that BTN2A1 was associated with decreased risk of AD, anxiety, and SCZ, which is consistent with a report that BTN2A1 is involved in the immunomodulation of the activity of vo T-cells and has significant anti-neuroinflammatory and neuroprotective effects<sup>51</sup>. Although BTN3A2 and BTN2A1 are known to interact<sup>52</sup>, given that BTN3A2 and BTN2A1 were separately identified as risk and protective factors for brain disorders in the present study, further studies are needed to elucidate the underlying mechanisms of how BTN3A2 and BTN2A1 proteins are oppositely involved in brain disorders. Furthermore, the ENPP6 protein exhibited a significant association with the risk of PD. In PD, the putamen, which is part of the basal ganglia network that is crucial for motor control, is heavily impacted<sup>53</sup>, and this is potentially related to the observed association between ENPP6 and the volume of left putamen. Altogether, although the exact mechanism underlying the protein-disease relationships remains to be elucidated, the findings of the present study do emphasise the potential of brain-associated proteins as novel risk factors of brain disorders. Given the smaller sample sizes for anxiety and ALS compared to other disorders, the non-significant MR-relationships may be potentially due to the impact of sample size rather than indicating a lack of significance. Therefore, further studies based on a larger case count are warranted to validate our findings.

the adaptive immune response, exhibited a broad association with AD,

Now that proteins manifest an effect on both brain structure and brain disorders, elucidating the intricate relationship of proteins, brain structure, and brain disorders is of major clinical importance. However, due to the limited number of patients with both proteomic and neuroimaging data in UKB, direct evaluation of the mediation relationships corresponding to particular brain disorders was not possible. Instead, the information on mental health phenotypes available in UKB was used to investigate whether the protein-associated brain structure measures could serve as potential mediators of the association between proteins and mental health. A total of ten proteins were found to be associated with mental health, that was mediated by brain structure. Three of the ten proteins were highly enriched in brain tissues, namely ENPP6, NCAM1, and OXT, and the mediators with high loadings, including supramarginal cortex, thalamus, putamen, precentral cortex, OFC, were brain regions frequently involved in depression and anxiety<sup>54,55</sup>. In particular, a strong mediating effect of the precentral cortex and putamen was observed on the relationship between ENPP6 and mental health, with a mediation proportion of 65%. The precentral cortex, primarily involved in motor control and coordination, has been increasingly recognised for its role in working memory and emotional regulation<sup>56</sup>. Cortical thinning in the precentral gyrus has been associated with an increased risk of depressive symptomatology<sup>57</sup>. In addition, advanced control relevant activation in the precentral gyrus has been linked with suicide risk in mood disorders<sup>58</sup>. The high mediation proportion suggests that ENPP6 may influence mental health indirectly by modulating the structure or function of these regions. NCAM1 plays a significant role in the development of the nervous system by regulating neurogenesis and neurite outgrowth, while OXT is a precursor protein that produces oxytocin and neurophysin I, which are transported axonally to the nerve endings and secreted into the bloodstream. The association between NCAM1 and mental health was mediated by the supramarginal gyrus and thalamus, whereas the association between OXT and mental health was mediated by the medial orbitofrontal, middle temporal cortex, accumbens, and pallidum. These findings are consistent

prefrontal cortex relating to reward in depression<sup>55,59,60</sup>, as well as the disturbance of the large-scale cortico-striato-thalamo-cortical circuitry involved in anxiety<sup>54</sup>. Although several proteins, including BTN3A2, BTN2A1 and INHBC, did not exhibit significant mediation relationships for mental health phenotypes in UKB, given that BTN3A2 and BTN2A1 were associated with MD of different parts of the cingulum and INHBC exhibited an association with ALS-related superior temporal cortex, the possibility remains that certain brain structure is potential mediator for associations between BTN3A2, BTN2A1, INHBC and brain disorders in patients with confirmed diagnosis. In summary, our findings do pave the way for understanding the intricate interplay between microscale proteins, brain structure endophenotypes, and behaviour phenotypes.

Several potential limitations should be considered when interpreting the results of the present study. Firstly, although the study is of the largest cohort to have so far been investigated to characterise the association pattern between proteins and brain structures, further replication in independent cohorts and cohorts with different ethnicities is vital. Secondly, due to the acquisition schedule of the UK Biobank, there is a time gap between plasma collection and neuroimaging scanning. Given that the plasma proteome could change with age, this may induce potential bias<sup>61,62</sup>. The interval has been accounted for as a covariate. However, future study designs with more synchronised data collection could address this concern. Thirdly, although this research has utilised comprehensive proteomic databases, it is still far from measuring all circulating proteins. Future advances in proteomic measurement technologies will increase the possibility of discovering novel linkages. Fourthly, the included participants who underwent both the imaging visit and proteomic analysis exhibited slight differences in age and sex compared to the other UKB participants. Although we have partially controlled the effects of age and sex through regression, validation with future releases of proteome-neuroimaging data covering a broader range could strengthen our findings. Fifthly, as the proteins are highly correlated with one another, and the brain measures also exhibit correlations, the significance of protein-MR associations presented little inflation. Although we have performed multiple comparison corrections and validated the results with MR and mediation analyses, future independent validation could enhance the discovered associations between proteins and brain structures. Sixthly, the plasma proteomic profile in younger individuals may differ from that observed in our cohort. The disease-associated proteins identified in this study may reflect later-stage plasma protein markers, rather than early biomarkers, especially considering that disorders such as autism spectrum disorder (ASD), ADHD, schizophrenia, BPD, and anxiety typically have onset in childhood or early adulthood. Consequently, future validation is needed to identify if the disease-associated protein changes are longitudinally predictive of incident diseases.

In conclusion, utilising the unique opportunity provided by large-scale genetic, proteomic, and neuroimaging data, this study represents a comprehensive atlas of the proteomic signature of human brain structure. Our findings underscore the widespread associations between plasma proteins and structural changes in the brain, revealing how plasma proteins may reflect alterations in the central nervous system. The identified proteins exhibited associations with a wide range of neurodegenerative and neuropsychological disorders, highlighting the potential of plasma proteins as critical biomarkers for assessing brain health. Furthermore, brain structure was found to play a mediating role in the relationship between plasma proteins and brain health. By uncovering these associations, this research bridges the gap between plasma protein markers, brain structure, and brain disorders, offering an opportunity to unravel the mechanisms underlying brain disorders and to identify potential therapeutic targets.

The UK Biobank is a large-scale population-based cohort of approximately 500,000 participants aged 40 to 69 years (https://biobank. ndph.ox.ac.uk/showcase/), who provided fully informed written consent for the collection of genomic, whole-body imaging, electronic health record linkage, body fluid biomarker, and physical and anthropometric measurements<sup>63</sup>. UKB received approval from the National Information Governance Board for Health and Social Care and the National Health Service North West Centre for Research Ethics Committee (Ref: 11/NW/0382). In particular, approximately 40,000 participants received neuroimaging scanning, and a cohort of 54,265 participants for whom plasma samples were collected at the commencement of the study. The present study refers to the participants in UKB for whom both plasma proteomic and neuroimaging data are available, and excludes participants that present obvious brain structural abnormalities, including all cause dementia, stroke, Parkinson's disease, parkinsonism, Huntington's disease, multiple sclerosis, epilepsy, intracranial injury, cerebrovascular diseases, encephalitis, myelitis, encephalomyelitis, bacterial meningitis, intracranial and intraspinal abscess and granuloma, intracranial and intraspinal phlebitis and thrombophlebitis, cerebral infarction, vascular syndromes of brain in cerebrovascular diseases, other demyelinating diseases of central nervous system, congenital malformations of the nervous system, malignant neoplasm of brain, intracranial laceration and haemorrhage due to birth injury or other degenerative disorders of nervous system (Supplementary Data 1). Finally, approximately 4997 participants were used in our primary association analysis (see Table 1 and Supplementary Fig. 1).

#### Protein measurements in plasma

The plasma samples were processed using a NovaSeq 6000 Sequencing System (Illumina Inc., San Diego, USA), and the concentrations of 2923 unique proteins were measured using Olink Explore 3072 with eight panels (Olink Proteomics AB, Uppsala, Sweden). Subsequently, guality control, outlier detection, and normalisation are applied to produce Normalised Protein eXpression (NPX) values in relative units on a log2 scale for each protein and each participant<sup>64</sup>. Proteins with a missing rate of 30% or higher were excluded (GLIPR1, NPM1, and PCOLCE), and the remaining missing values were imputed based on median values. All remaining 2920 proteins showed a missing rate of less than 25%, of which 2911 proteins actually showed a missing rate of less than 20%, and 1215 proteins had a missing rate of less than 5%. The value of the expression of each protein was inverse-rank normalised. The effects of sex and age were regressed out from protein expression before the protein-brain structure association analysis because they exhibited high collinearity with the protein. In this way, we ruled out the possibility that the observed protein-brain structure associations were simply the effect of age and sex. Finally, the NPX values of 2920 proteins were available for further analysis.

#### Imaging derived phenotypes

Given that grey and white matter are crucial components that serve distinct functions for overall brain health, here we included brain structural imaging derived phenotypes (IDPs) that were previously made available by the UKB neuroimaging group<sup>63</sup>. Similar to previous researches, regional grey matter measures (volume, thickness, and area) derived from T1-weighted images and microstructural measures derived from diffusion weighted imaging, including FA and MD, were chosen to be used in the present study<sup>65</sup>. Specifically, 86 GM volume, 68 GM surface area, and 68 cortical GM thickness measurements were computed for brain regions corresponding to the Desikan-Killiany atlas by using FreeSurfer v6.0.0 (https://surfer.nmr.mgh.harvard.edu), and 27 FA and 27 MD values were available for 27 major WM tracts in

UKB (Supplementary Data 2). Totally, 218 GM measurements of regions and 54 WM measurements of tracts were used. Measurements of the whole brain were only included as covariates. All the IDP data were inverse-rank normalised before statistical analyses.

#### Genetic data processing and analysis

With regard to genetics, the imputed genetic dataset released by UKB in July 2017 was used for MR analysis in the present study. The samples and information extracted regarding genotype were subject to strict quality control<sup>66</sup>. Specifically, single-nucleotide polymorphisms (SNPs) with a missing rate > 5%, minimum minor allele frequency < 0.1%, or Hardy-Weinberg equilibrium test  $P < 1 \times 10^{-10}$  had been excluded<sup>18,67</sup>. To reduce the difference in population structure, all analyses were based on unrelated individuals. The unrelated individuals were selected with the following criteria: (1) used to compute principal components, (2) not identified as outliers for heterozygosity and missing rates, (3) without putative sex chromosome aneuploidy, (4) with no more than ten putative third-degree relatives. (5) have a genetic background of 'White British'. Since the cohort for the publicly available GWAS summary of IDPs overlaps with the cohort used for proteomic analysis, to satisfy the no-overlapping assumption of the cohort for two-sample MR analysis of the relationship between plasma proteins and IDPs, we performed GWAS analysis based on 25,576 UKB participants for whom neuroimaging data was available but proteomic data was not. A linear association test was then performed for each IDP with PLINK 268, with age at imaging visit, sex, education years, scanning site, total intracranial volume, and 40 genetic principal components provided by UKB as covariates. All IDPs were quantile normalised before the GWAS analysis was performed.

#### GWAS summary of brain disorders

To conduct Mendelian randomisation analysis between protein and disorders, we curated GWAS summary statistics of brain disorders. Given their significance in public health and the relatively large number of cases. GWAS summary statistics were collected from publicly available databases for six psychiatric disorders (ADHD (cases/total = 19,099/53,293)<sup>69</sup>, anxiety (cases/total = 7016/18,186)<sup>70</sup>, ASD (cases/ total=18,381/46,350)<sup>71</sup>, BP (cases/total = 20,352/51,710)<sup>72</sup>, MDD (cases/ total = 45,396/142,646)<sup>73</sup>, SCZ (cases/total=67,390/161,405)<sup>74</sup>) and four neurodegenerative disorders75 (AD (cases/total=10,520/401,661), PD (cases/total=4,681/407,500), MS (cases/total = 2409/408,561) and ALS (cases/total = 531/184,000)). In particular, the psychiatric disorders GWAS summary was downloaded from the Psychiatric Genomics Consortium (https://www.med.unc.edu/pgc/download-results), and the degenerative disease GWAS summary was downloaded from FinnGen (https://www.finngen.fi/fi). All studies were of European ancestry and had no overlapping individuals with the UKB. A detailed summary can be found in Supplementary Data 3.

#### Gene expression and quantitative trait loci

Here, we used the protein quantitative trait loci (pQTL) and gene expression quantitative trait loci (eQTL) summary data to characterise genetic variants of protein and gene expression separately. A "pQTL" stands for "protein quantitative trait locus," which refers to a genetic variant associated with the abundance of a specific protein, while an "eQTL" stands for "expression quantitative trait locus," indicating a genetic variant linked to the expression level of a gene (mRNA transcript) at a specific locus on the chromosome. The protein quantitative trait loci relationships were obtained from the previous UK Biobank Pharma Proteomics Project (http://ukbppp.gwas.eu), which provides comprehensive pQTL mapping of 2923 proteins and identifies 14,287 primary genetic associations, in addition to ancestry-specific pQTL mapping in non-Europeans. Consistent with our primary analysis, we restricted the pQTL mapping relationships to participants of European ancestry. Details of data processing and genetic association analysis have been described elsewhere in previous work  $^{\rm 64}.$ 

The gene expression data used to characterise the distribution of protein-coding genes were obtained for 34 different tissues from the database of the GTEx project v8 (https://www.gtexportal.org/home/)<sup>76</sup>. All variant-eQTL associations that were tested in brain cortex from GTEx v8, and in particular cis-eQTL, were used for pleiotropy association analysis between pQTL and eQTL.

#### Statistical analyses

The present study contains six different research contents, three main analyses, and three sub-analyses. A flow diagram of the analyses that were performed is shown in Fig. 1. In the main analysis, the association analysis between proteins and brain structure measures was first performed, after which, the physical position and functions of the associated proteins were annotated. Furthermore, the association relationships were further enhanced with MR analysis. The proteins exhibited significant MR relationships at FDR-corrected P < 0.05 were eligible for the following three sub-analyses, i.e., SMR, protein-disease MR, and mediation analyses. The goal of these analyses was to investigate the potential mechanism by which the plasma proteins were associated with brain structure, whether brain structure-associated proteins would contribute to the risk of brain disorders, and whether brain structure could mediate the association between proteins that are associated with brain structures and mental health.

Association analysis. A generalised linear model was used to investigate the association between plasma protein expression and imaging derived measures of brain structure, including regional GM volume, area, and thickness, as well as FA and MD of WM tracts. Protein expression, age at imaging visit, sex, years of education, Townsend index, ethnicity, smoking and drinking status at imaging visit, the interval between baseline and imaging visit, scanning site, and total intracranial volume were used as predictors of the structural measures. The significance was determined by FDR correction across all proteinbrain structure pairs. We conducted three sensitivity analyses. First, an additional association analysis with only protein expression, age at imaging visit, sex, total intracranial volume, and scanning site as the predictors was performed to explore the influence of the covariates. Second, we conducted an additional association analysis without performing an imputation of protein expression. Third, to examine potential sex differences, the association analysis was further conducted in female and male separately.

Physical and functional annotation. To demonstrate potential differences in associated proteins related to different structural measures, physical and multiple functional annotations were performed for brain structure-associated proteins, including chromosome distribution annotation, UKB panel enrichment, tissue enrichment, cellular enrichment, and biological process enrichment. The physical annotation was performed by first mapping the protein-coding genes to chromosome positions with a previously established topr R software<sup>77</sup>. Subsequently, the number of protein-coding genes on each chromosome was divided by the length of that chromosome, which generated the density of associated proteins in each chromosome that accounted for the length of the chromosome. UKB panel and cellular enrichment annotations were performed using over-representation analysis from Python scipy. The four UKB panels, namely, cardiometabolic, inflammation, neurology, and oncology, used to classify proteins were downloaded from the UKB website (https://biobank.ndph. ox.ac.uk/showcase/ukb/auxdata/olink\_assay.dat). The marker genes of seven central nervous system cell types, including astrocytes, endothelial cells, microglia, excitatory neurons, inhibitory neurons, oligodendrocytes, and oligodendrocyte progenitor cells, were obtained from a prior study<sup>78</sup>. Tissue enrichment was performed with a

hypergeometric test from the FUMA GENE2FUNC module<sup>79</sup>, using GTEx v8 as the source of gene expression. Specifically, DEG sets were pre-calculated by performing a two-sided *t* test for each of the tissue labels in comparison to all the others in turn. Genes with Bonferroni-corrected p < 0.05 and absolute log fold change  $\ge 0.58$  were defined as DEG in a given tissue compared to others. For each structural measure, tissue enrichment was first conducted for all associated proteins, followed by separate tissue enrichment for positively and negatively associated proteins. Similarly, for each structural measure, GO biological process enrichment was performed separately for positively and negatively associated proteins by using the over-representation analysis function of the WebGestalt toolkit<sup>80</sup>. For all enrichment analyses, the encoding genes of 2920 proteins of UKB were used as a customised background gene set, and the significance threshold was set to FDR corrected p < 0.05, unless otherwise specified.

MR associations between protein expression and brain structure.

Traditional observational epidemiological studies have long been hindered by challenges such as confounding. Instead, the MR method utilises Mendel's laws of segregation and independent assortment, which ensure that genetic variants are distributed independently of environmental influences and other genetic factors. Hence, the genetic associations should therefore be largely free from confounding<sup>81</sup>. This could enhance the reliability of the results by leveraging genetic variants as instrumental variables to assess the MR relationship between protein expression and brain structure. To further enhance the significant protein-brain structure associations in the first analysis, bidirectional MR analyses were performed by using the TwoSampleMR R package (version: 0.5.8)<sup>82</sup>. For forward MR analysis with protein expression as exposure and structural measure as outcome, the IVW method was used as the primary inference. The independent SNPs from the GWAS summary of proteins were selected by using the clumping technique programmed in PLINK 1.9 software<sup>68</sup>, with an r2 threshold of 0.01, a window size of 1 Mb, and a p threshold of  $5 \times 10^{-6}$ . Putative outliers were detected with Cochran's O test for the IVW method and Rucker's O' test for the MR-Egger model. The outliers with a nominal p-value < 0.05 were excluded. If the final number of suitable genetic instruments was less than three, we excluded the protein-brain structure pairs from MR analyses. The significance of the relationship was determined using two criteria: FDR-corrected p < 0.05 and a less strict significance threshold of nominal p < 0.001. For reverse MR analyses with structural measures as exposure and protein expression as outcome, genetic instruments were selected from the GWAS summary statistics of brain structure measures at a P threshold of  $5 \times 10^{-6}$ . The clumping and outlier detection was performed using the same procedure as the forward MR. Moreover, to enhance the reliability of the MR results, the bidirectional MR analyses were repeated with a stricter clumping at a p threshold of  $5 \times 10^{-8}$ . In addition, we conducted another four MR methods: MR-Egger, weighted median, weighted mode, and Wald ratio, with the former three to complement the potential bias in the IVW results, and the last one when only one genetic instrument was valid.

To consider a potential violation of MR assumptions, three sensitivity analyses were further conducted to verify the significant associations between proteins and brain structures. Firstly, a leave-one-out analysis was utilised to check whether the observed MR relationship was driven by a single SNP<sup>83</sup>. Secondly, the MR-PRESSO technique was used to detect possible horizontal pleiotropy<sup>84</sup>, when a genetic variant influences the brain structure (directly or indirectly, through other traits) independently of the hypothesised protein. Thirdly, MR-Egger regression analysis was used to detect possible directional pleiotropy<sup>85</sup>, when a genetic variant associated with protein levels influences brain structure through a shared pathway rather than a direct effect of the protein on the brain structure. To further control the false positive rate in the presence of pleiotropic effects, we performed CAUSE analysis to distinguish causality and genetic correlation using the cause R package. Similar to that of MR analysis, two clumping thresholds (i.e.,  $5 \times 10^{-6}$  and  $5 \times 10^{-8}$ ), were used, and the significance threshold was set to FDR-corrected P < 0.05.

Pleiotropic association between plasma pQTL and brain eQTL. If an association is found between the expression of proteins in plasma and brain structure, it is important to try to determine the mechanisms by which the relevant peripheral plasma proteins influence brain structure. In the first sub-analysis, the investigation may be begun by characterising the gene expression of the brain structure-associated proteins across the 34 main tissues represented in the GTEx v8 database. Next, the possibility that the plasma protein-brain structure relationship is attributable to shared genetic regulation between plasma protein and brain gene expression can be examined. Specifically, for each brain structure-associated protein surviving MR analysis, we used the SMR method (version: 1.3.1)<sup>86</sup> to integrate summarylevel pQTL data with eQTL data to identify genes whose protein expression in plasma are associated with their expression levels in the brain because of pleiotropy. The HEIDI test was conducted, and the significance threshold was set to  $P_{\text{SMR}} < 0.05/32$  and  $P_{\text{HEIDI}} > 0.05$ .

MR analysis between proteins and disorders. In the second subanalysis, to investigate whether proteins exhibited potential effects on brain structure that could contribute to the risk of central nervous system disorders, the bidirectional MR analyses were performed between proteins that exhibited significant MR associations with brain structure at FDR-corrected p < 0.05 and ten disorders. In other words, we leveraged genetic variants as instrumental variables to assess the MR relationship between protein expression and common diseases. A similar MR procedure to that of protein-brain structure measures was utilised, among which the clumping at a p threshold of  $5 \times 10^{-6}$  was performed in the main analysis, and clumping at a *p* threshold of  $5 \times 10^{-8}$  was conducted for validation. The significance threshold for the bidirectional MR analysis was set to FDR-corrected p < 0.05, and details of the GWAS summary statistics for six psychiatric disorders and four neurodegenerative disorders are included in the relevant section. For case-control GWAS summary statistics, the OR statistics were converted to log odds before MR analysis. CAUSE analysis was also conducted to distinguish causality and genetic correlation.

Mediation analysis. Finally, In the third sub-analysis, it is crucial to examine whether the plasma proteins could contribute to phenotypes relevant to mental health through the brain structure. A total of 3270 participants with a mean age of 54.2 were eligible for inclusion in mediation analysis (Supplementary Data 4). Accordingly, by using the SEM in the lavaan R package (version: 0.6.17)<sup>87</sup>, a mediation analysis study was performed for a subset of UKB participants for whom proteomic, neuroimaging, and also mental health data were available. In particular, the protein-associated brain structural measures were used to construct the latent structural measure, the mental health phenotypes derived from UKB, including depression, anxiety, mania, wellbeing, psychotic experience, self-harm, mental distress, and trauma, were used to construct the latent mental health phenotype<sup>18</sup>. An iterative strategy was used to remove observable variables of high collinearity when constructing latent variables. Specifically, given a specific protein, for a pair of protein-associated structure measures with a correlation greater than the threshold, the structure measures with the less significant protein-structure association were removed. By requiring the weights of all surviving observational variables contributing to latent structural measures to fall within the range [-1,1], while minimising the removal of observational brain structure

measurements, a correlation threshold of 0.55 was chosen. All variables were inverse-rank normalised before SEM analysis, and the SEM models for which there was a significant association between protein and brain structure (i.e., model a), brain structure and mental health (model b), protein and mental health (model c), as well as significant indirect effect (model a x b), were considered as valid mediations. The exploratory mediation analysis was performed for proteins that survived protein-brain structure MR analysis, and investigations were performed for both a strict FDR-corrected threshold of p < 0.05 and a less strict threshold of nominal p < 0.001.

#### **Reporting summary**

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

### Data availability

All plasma proteomic, neuroimaging, genomic and mental health phenotype data are publicly available at the UK Biobank (http://www. ukbiobank.ac.uk/) and could be accessed with a reasonable request. The data in the present study were used according to the application no. 19542 and 202239. The GWAS summary data for the disease can be found at the Psychiatric Genomics Consortium (https://www.med.unc. edu/pgc/download-results) and the FinnGen website (https://www. finngen.fl/fi). The in vitro tissue gene expression data are available at GTEx (https://www.gtexportal.org/home/). The GTEx v8 cis-summary data of the brain cortex tissue can be downloaded from https:// yanglab.westlake.edu.cn/software/smr/#eQTLsummarydata. Source data are provided in this paper.

## **Code availability**

All software and analytical methods used in this study are publicly available. The code used to run the analyses in this study is publicly available at https://github.com/hitrp/proteomicSignatureOfBrain Structure/. Brain visualisations were generated by the authors using the ENIGMA Toolbox (https://enigma-toolbox.readthedocs.io/), which is distributed under the open-source license permitting academic use. Additional figures were created using the ggplot2 package in R and the BioRender platform. Figures created with BioRender are in compliance with BioRender's Academic License Terms and are intended for publication purposes.

## References

- Winter, N. R. et al. Quantifying deviations of brain structure and function in major depressive disorder across neuroimaging modalities. JAMA Psychiatry 79, 879–888 (2022).
- Han, L. K. M. et al. Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group. *Mol. Psychiatry* 26, 5124–5139 (2021).
- 3. Bateman, R. J. et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N. Engl. J. Med.* **367**, 795–804 (2012).
- Mitchell, T. et al. Emerging neuroimaging biomarkers across disease stage in Parkinson disease: A review. JAMA Neurol. 78, 1262–1272 (2021).
- Wang, L. et al. Association of specific biotypes in patients with Parkinson disease and disease progression. *Neurology* 95, e1445–e1460 (2020).
- Jiang, Y. et al. Neuroimaging biomarkers define neurophysiological subtypes with distinct trajectories in schizophrenia. *Nat. Ment. Health* 1, 186–199 (2023).
- Chen, D. et al. Neurophysiological stratification of major depressive disorder by distinct trajectories. *Nat. Ment. Health* 1, 863–875 (2023).
- 8. Li, M. et al. Associations of cerebrospinal fluid complement proteins with Alzheimer's pathology, cognition, and brain structure in non-dementia elderly. *Alzheimers Res. Ther.* **16**, 12 (2024).

- 9. Abbasi, N. et al. Relationship between cerebrospinal fluid biomarkers and structural brain network properties in Parkinson's disease. *Mov. Disord.* **33**, 431–439 (2018).
- Fujishima, M., Kawasaki, Y., Mitsuhashi, T. & Matsuda, H. Impact of amyloid and tau positivity on longitudinal brain atrophy in cognitively normal individuals. *Alzheimers Res. Ther.* 16, 77 (2024).
- Dayon, L. et al. Proteomes of paired human cerebrospinal fluid and plasma: Relation to blood-brain barrier permeability in older adults. *J. Proteome Res.* 18, 1162–1174 (2019).
- 12. Sung, Y. J. et al. Proteomics of brain, CSF, and plasma identifies molecular signatures for distinguishing sporadic and genetic Alzheimer's disease. *Sci. Transl. Med.* **15**, eabq5923 (2023).
- Hampel, H. et al. Developing the ATX(N) classification for use across the Alzheimer disease continuum. *Nat. Rev. Neurol.* 17, 580–589 (2021).
- Benedet, A. L. et al. Stage-specific links between plasma neurofilament light and imaging biomarkers of Alzheimer's disease. *Brain* 143, 3793–3804 (2020).
- Tissot, C. et al. Plasma pTau181 predicts cortical brain atrophy in aging and Alzheimer's disease. *Alzheimers Res. Ther.* 13, 69 (2021).
- Jain, P. R. et al. Multiomic approach and Mendelian randomization analysis identify causal associations between blood biomarkers and subcortical brain structure volumes. *Neuroimage* 284, 120466 (2023).
- 17. Harris, S. E. et al. Neurology-related protein biomarkers are associated with cognitive ability and brain volume in older age. *Nat. Commun.* **11**, 800 (2020).
- Li, Y. et al. The brain structure and genetic mechanisms underlying the nonlinear association between sleep duration, cognition and mental health. *Nat. Aging* 2, 425–437 (2022).
- Nir, T. M. et al. Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging. *NeuroImage*. *Clin.* 3, 180–195 (2013).
- Senda, J. et al. MRI mean diffusivity detects widespread brain degeneration in multiple sclerosis. J. Neurol. Sci. 319, 105–110 (2012).
- Kochunov, P. et al. Fractional anisotropy of water diffusion in cerebral white matter across the lifespan. *Neurobiol. Aging* 33, 9–20 (2012).
- Lin, J. Z., Duan, M. R., Lin, N. & Zhao, W. J. The emerging role of the chondroitin sulfate proteoglycan family in neurodegenerative diseases. *Rev. Neurosci.* **32**, 737–750 (2021).
- 23. Mravinacová, S. et al. CSF protein ratios with enhanced potential to reflect Alzheimer's disease pathology and neurodegeneration. *Mol. Neurodegener.* **19**, 15 (2024).
- Hatayama, M. & Aruga, J. Developmental control of noradrenergic system by SLITRK1 and its implications in the pathophysiology of neuropsychiatric disorders. *Front. Mol. Neurosci.* 15, 1080739 (2022).
- Puranik, N. & Song, M. Insight into the association between slitrk protein and neurodevelopmental and neuropsychiatric conditions. *Biomolecules* 14, https://doi.org/10.3390/biom14091060 (2024).
- Cobo-Calvo, A., Vukusic, S. & Marignier, R. Clinical spectrum of central nervous system myelin oligodendrocyte glycoprotein autoimmunity in adults. *Curr. Opin. Neurol.* **32**, 459–466 (2019).
- Xu, C. et al. Polymorphisms in seizure 6-like gene are associated with bipolar disorder I: evidence of gene × gender interaction. J. Affect. Disord. 145, 95–99 (2013).
- 28. Zou, X. et al. Role of leptin in mood disorder and neurodegenerative disease. *Front. Neurosci.* **13**, 378 (2019).
- Froemke, R. C. & Young, L. J. Oxytocin, neural plasticity, and social behavior. Annu. Rev. Neurosci. 44, 359–381 (2021).
- Rana, T. et al. Exploring the role of neuropeptides in depression and anxiety. Prog. Neuro Psychopharmacol. Biol. Psychiatry 114, 110478 (2022).

- Article
- Bubb, E. J., Metzler-Baddeley, C. & Aggleton, J. P. The cingulum bundle: Anatomy, function, and dysfunction. *Neurosci. Biobehav. Rev.* 92, 104–127 (2018).
- Cai, J. et al. White matter changes and its relationship with clinical symptom in medication-naive first-episode early onset schizophrenia. Asian J. psychiatry 82, 103482 (2023).
- Deng, F. et al. Abnormal segments of right uncinate fasciculus and left anterior thalamic radiation in major and bipolar depression. *Prog. Neuro Psychopharmacol. Biol. Psychiatry* 81, 340–349 (2018).
- 34. Serrano-Pozo, A., Das, S. & Hyman, B. T. APOE and Alzheimer's disease: advances in genetics, pathophysiology, and therapeutic approaches. *Lancet Neurol.* **20**, 68–80 (2021).
- Manzoni, C. et al. Genome-wide analyses reveal a potential role for the MAPT, MOBP, and APOE loci in sporadic frontotemporal dementia. Am. J. Hum. Genet. 111, 1316–1329 (2024).
- Nehme, R. et al. The 22q11.2 region regulates presynaptic geneproducts linked to schizophrenia. *Nat. Commun.* 13, 3690 (2022).
- Gur, R. C. et al. Neurocognitive profiles of 22q11.2 and 16p11.2 deletions and duplications. *Mol. Psychiatry* **30**, 379–387 (2025).
- Gerhards, R. et al. Oligodendrocyte myelin glycoprotein as a novel target for pathogenic autoimmunity in the CNS. Acta Neuropathol. Commun. 8, 207 (2020).
- Kang, J., Park, H. & Kim, E. IRSp53/BAIAP2 in dendritic spine development, NMDA receptor regulation, and psychiatric disorders. *Neuropharmacology* **100**, 27–39 (2016).
- 40. Pereira, J. B. et al. Plasma GFAP is an early marker of amyloid-β but not tau pathology in Alzheimer's disease. *Brain* **144**, 3505–3516 (2021).
- Begcevic, I. et al. Brain-related proteins as potential CSF biomarkers of Alzheimer's disease: A targeted mass spectrometry approach. J. Proteom. 182, 12–20 (2018).
- 42. Guo, Y. et al. Plasma proteomic profiles predict future dementia in healthy adults. *Nat. Aging* **4**, 247–260 (2024).
- 43. Kvartsberg, H. et al. Characterization of the postsynaptic protein neurogranin in paired cerebrospinal fluid and plasma samples from Alzheimer's disease patients and healthy controls. *Alzheimers Res. Ther.* **7**, 40 (2015).
- 44. Huang, J. et al. Inflammation-related plasma and CSF biomarkers for multiple sclerosis. *Proc. Natl. Acad. Sci. USA* **117**, 12952–12960 (2020).
- Lo Bianco, L. et al. Genetic variation is associated with RTN4R expression and working memory processing in healthy humans. *Brain Res. Bull.* 134, 162–167 (2017).
- Ding, W. et al. Association of cortical and subcortical microstructure with disease severity: impact on cognitive decline and language impairments in frontotemporal lobar degeneration. *Alzheimer's Res. Ther.* **15**, 58 (2023).
- Poos, J. M. et al. Longitudinal brain atrophy rates in presymptomatic carriers of genetic frontotemporal. *Dementia* **99**, e2661–e2671 (2022).
- Arandjelovic, S. & Ravichandran, K. S. Phagocytosis of apoptotic cells in homeostasis. *Nat. Immunol.* 16, 907–917 (2015).
- Pineda-Cirera, L., Cabana-Domínguez, J., Lee, P. H., Fernàndez-Castillo, N. & Cormand, B. Identification of genetic variants influencing methylation in brain with pleiotropic effects on psychiatric disorders. *Prog. Neuro-Psychopharmacol. Biol. Psychiatry* **113**, 110454 (2022).
- Wu, Y. et al. Identification of the primate-specific gene BTN3A2 as an additional schizophrenia risk gene in the MHC loci. *EBioMedicine* 44, 530–541 (2019).
- Herrmann, T. & Karunakaran, M. M. Butyrophilins: γδ T Cell Receptor Ligands, Immunomodulators and More. *Front. Immunol.* 13, 876493 (2022).
- 52. Cano, C. E. et al. BTN2A1, an immune checkpoint targeting  $V\gamma 9V\delta 2$ T cell cytotoxicity against malignant cells. *Cell Rep.* **36**, 109359 (2021).

- 53. Rong, D. et al. Consistent abnormal activity in the putamen by dopamine modulation in Parkinson's disease: A resting-state neuroimaging meta-analysis. *Brain Res. Bull.* **210**, 110933 (2024).
- 54. Zhang, X. et al. Structural and functional deficits and couplings in the cortico-striato-thalamo-cerebellar circuitry in social anxiety disorder. *Transl. Psychiatry* **12**, 26 (2022).
- 55. Zhang, B. et al. Roles of the medial and lateral orbitofrontal cortex in major depression and its treatment. *Mol. Psychiatry* (2024).
- 56. Fang, P. et al. Effects of computerized working memory training on neuroplasticity in healthy individuals: A combined neuroimaging and neurotransmitter study. *Neuroimage* **298**, 120785 (2024).
- Bos, M. G. N., Peters, S., van de Kamp, F. C., Crone, E. A. & Tamnes, C. K. Emerging depression in adolescence coincides with accelerated frontal cortical thinning. *J. Child Psychol. Psychiatry Allied Discip.* 59, 994–1002 (2018).
- Minzenberg, M. J. et al. Control-related frontal-striatal function is associated with past suicidal ideation and behavior in patients with recent-onset psychotic major mood disorders. J. Affect. Disord. 188, 202–209 (2015).
- 59. Rupprechter, S. et al. Blunted medial prefrontal cortico-limbic reward-related effective connectivity and depression. *Brain* **143**, 1946–1956 (2020).
- Cheng, W. et al. Sleep duration, brain structure, and psychiatric and cognitive problems in children. *Mol. Psychiatry* 26, 3992–4003 (2021).
- 61. Niu, L. et al. Plasma proteome variation and its genetic determinants in children and adolescents. *Nat. Genet.* **57**, 635–646 (2025).
- 62. Lehallier, B. et al. Undulating changes in human plasma proteome profiles across the lifespan. *Nat. Med.* **25**, 1843–1850 (2019).
- 63. Littlejohns, T. J. et al. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. *Nat. Commun.* **11**, 2624 (2020).
- 64. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
- 65. Tian, Y. E. et al. Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. *Nat. Med.* **29**, 1221–1231 (2023).
- Ge, Y.-J. et al. Genetic architectures of cerebral ventricles and their overlap with neuropsychiatric traits. *Nat. Hum. Behav.* 8, 164–180 (2024).
- 67. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
- 68. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
- Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* 51, 63–75 (2019).
- Otowa, T. et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry* 21, 1391–1399 (2016).
- 71. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
- 72. Stahl, E. A. et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
- Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* 50, 668–681 (2018).
- 74. Trubetskoy, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
- 75. Kurki, M. I. et al. FinnGen provides genetic insights from a wellphenotyped isolated population. *Nature* **613**, 508–518 (2023).
- Consortium, T. G. et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
- 77. Juliusdottir, T. topr: an R package for viewing and annotating genetic association results. *BMC Bioinform*. **24**, 268 (2023).

- Ren, P. et al. Regional transcriptional vulnerability to basal forebrain functional dysconnectivity in mild cognitive impairment patients. *Neurobiol. Dis.* 177, 105983 (2023).
- Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 8, 1826 (2017).
- Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 47, W199–W205 (2019).
- Richmond, R. C. & Davey Smith, G. Mendelian randomization: Concepts and scope. Cold Spring Harb. Perspect. Med. 12, https:// doi.org/10.1101/cshperspect.a040501 (2022).
- Guo, J. et al. Mendelian randomization analyses support causal relationships between brain imaging-derived phenotypes and risk of psychiatric disorders. *Nat. Neurosci.* 25, 1519–1527 (2022).
- Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife* 7, https://doi.org/10. 7554/elife.34408 (2018).
- Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* 50, 693–698 (2018).
- Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* 44, 512–525 (2015).
- Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* 48, 481–487 (2016).
- Rosseel, Y. lavaan: An R package for structural equation modeling. J. Stat. Softw 48, 1–36 (2012).

# Acknowledgements

This study used the UK Biobank Resource under application numbers 19542 and 202239. We thank all participants and team members of the UK Biobank. We thank the participants and investigators of the FinnGen study. W.C. was supported by grants from the National Key R&D Programme of China (No. 2023YFC3605400) and the National Natural Science Foundation of China (No. 82472055). J.-T.Y. was supported by grants from the Science and Technology Innovation 2030 Major Projects (no. 2022ZD0211600), the National Natural Science Foundation of China (nos 82071201, 81971032 and 92249305), the Shanghai Municipal Science and Technology Major Project (nos 2018SHZDZX01 and 2023SHZDZX02), the Research Start-Up Fund of Huashan Hospital (no. 2022QD002), the Excellence 2025 Talent Cultivation Programme at Fudan University (no. 3030277001), Shanghai Talent Development Funding for the Project (no. 2019074), and the Zhangjiang Lab, Tianqiao and Chrissy Chen Institute, and the State Key Laboratory of Neurobiology and Frontiers Centre for Brain Science of Ministry of Education, Fudan University. J.-F.F. was supported by the National Key R&D Programme of China (nos 2018YFC1312904 and 2019YFA0709502), the Shanghai Municipal Science and Technology Major Project (no. 2018SHZDZX01), the 111 Project (no. B18015), the Shanghai Centre for Brain Science and Brain-Inspired Technology and the Zhangjiang Lab. The funders had no role in study design, data collection and analysis,

decision to publish or preparation of the manuscript. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

## **Author contributions**

All authors had full access to the data in the study and accepted responsibility to submit it for publication. Conceptualisation, W.C., J.-T.Y. and M.-Y.W.; Methodology, P.R., X.-H.H., B.-S.W., W.-K.G., C.S. Y.-J.Z., Q.M. and W.Z., Formal Analysis, P.R., X.-H.H., Z.-Y.L., J.Y. and Y.-Z.L.; Data Curation, W.Z., Y.-Z.L., J.Y. and Z.-Y.L.; Writing–Original Draft, P.R. and X.-H.H.; Writing–Review & Editing, P.R., X.-H.H., B.Z., L.-B.W., J.-J.K., Y.-C.J., N.R., Y.H., F.X. and Y.M.; Visualisation, P.R. and X.-H.H.; Supervision, W.C., J.-T.Y. and M.-Y.W.; Project Administration, W.C., J.-T.Y. and M.-Y.W.; Funding Acquisition, W.C., J.-T.Y. and M.-Y.W.

## **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/s41467-025-60185-7.

**Correspondence** and requests for materials should be addressed to Jin-Tai Yu, Meiyun Wang or Wei Cheng.

**Peer review information** *Nature Communications* thanks Weichen Song, AmanPreet Badhwar, who co-reviewed with Manpreet Singh, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit http://creativecommons.org/ licenses/by/4.0/.

© The Author(s) 2025