# Atlas of Proteomic Signatures of Brain Structure and Its Links to Brain Disorders

## **3 Supplementary Material**

- 4 Supplementary Methods
- 5 Supplementary Figures
- 6 Figure S1 Flowchart of participant inclusion and exclusion.
- Figure S2 Gene Ontology (GO) enrichment of associated proteins from all the five structural
  measures.
- 9 Figure S3 Tissue enrichment of positively and negatively associated proteins for volume and mean
- 10 diffusivity (MD) measures.
- **Figure S4** Tissue enrichment of positively and negatively associated proteins for area and thickness measures
- 12 measures.
- Figure S5 The causal effect of protein on brain structure in the forward MR at strict clumping Pthreshold.
- 15 Figure S6 The causal effect of protein on disease in the forward MR at strict clumping *P* threshold.
- 16 Figure S7 The causal effect of brain structure on protein in the reverse MR.
- 17 Figure S8 The causal effect of disease on protein in the reverse MR.
- 18 Figure S9 The expression of coding genes across tissues for proteins exhibiting mediation effect.

### 20 Supplementary Methods

#### 21 Preprocessing of proteomic data

The information regarding how the quality control was undertaken and how the protein data was 22 23 normalized have been thoroughly documented in the Supplementary Information in previous 24 publication<sup>1</sup>, as well as on UK Biobank (UKB) official website (https://biobank.ndph.ox.ac.uk/ showcase/refer.cgi?id=4658). The counts of known sequences were converted into Normalized 25 Protein eXpression (NPX) values, which were derived through within-batch and across-batch 26 27 normalization, using Olink's MyData Cloud Software. Within batch normalization centers data at NPX=0 by subtracting the plate-specific median per assay from all samples and assays in the same 28 29 plate. Across-batch normalization calculates adjustment factors by determining the difference in 30 assay-specific median NPX values for each batch. This process involves two steps: the first addresses plate-to-plate variation within a batch, while the second accounts for batch-to-batch variation across 31 32 the study. Both steps involve shifting by an assay-specific fixed factor on the NPX scale: the plate median in the first step and the difference between assay-specific medians across batches in the 33 34 second step.

35 The Olink workflow includes a inbuilt quality control system consisting of three engineered 36 internal controls that are spiked into every sample and each abundance block. Olink's internal quality control (QC) assessment is performed at two levels; run QC and sample QC. For run QC, each 37 abundance block per panel and sample plate should fulfil the mean absolute deviation (MAD) in both 38 internal controls (Inc Ctrl and Amp Ctrl) which should not exceed 0.3 NPX, the deviation of sample 39 40 QC level is allowed for up to 1/6 samples and in each panel the median of 90% assays in plate and negative controls should be in the accepted range from predefined values set during validation. 41 The sample QC evaluates each sample individually using the internal controls (Inc Ctrl and Amp Ctrl), 42 43 which should fall within  $\pm 0.3$  NPX of the plate median across the abundance block. Additionally, the mean assay count for a sample must not be less than 500 counts. Samples that do not meet these 44 45 criteria will receive a warning for the corresponding abundance block in the dataset.

Outliers were identified using two approaches applied to each protein panel: (1) principal 46 47 component analysis (PCA), and (2) examining the median and interquartile range (IQR) of NPX 48 across proteins by sample. Data points were removed if (1) a standardized PC1 (the component that 49 captures the most variation) or PC2 (second largest component) value more than 5 standard deviations from the mean (which is zero in standardized PCA), or (2) a median NPX greater than 5 standard 50 deviations from the mean median, or an IQR of NPX greater than 5 standard deviations from the 51 mean IQR. We excluded outliers, data points with a QC or assay warning, and likely sample swaps, 52 removing the sample across all panels if half or more of the panels were affected; the remaining data 53 contained 56,695 samples and 52,790 individuals. Suspected sample swaps were identified by 54 examining discrepancies between the proteomic-predicted sex and outliers from cis protein 55 quantitative trait loci (pQTLs), where the standardized squared residuals for all proteins were summed 56 57 for each individual and divided by the sum of squared protein levels. Samples with incorrect genotypes were expected to show larger values than those with correct genotypes. 58

## 60 Supplementary Figures



#### 61

62 Figure S1 Flowchart of participant inclusion and exclusion. The details of inclusion and exclusion

63 criteria for association analysis between proteins and brain structures, with the number of volume

64 measure shown as the example.





Figure S2 Gene Ontology (GO) enrichment of associated proteins from all the five structural 67 measures. Positively and negatively associated proteins were enriched separately. Positively 68 associated proteins were significantly enriched in biological processes such as immune system 69 process, inflammatory response, and cell adhesion (left panel). Negatively associated proteins were 70 significantly enriched in biological processes such as catabolic process, cellular catabolic process, and 71 response to organonitrogen (right panel). The 2,920 proteins were used as the background of 72 enrichment. An FDR-corrected P < 0.05 was considered significant. Source data are provided as a 73 74 Source Data file.



76

Figure S3 Tissue enrichment of positively and negatively associated proteins for volume and mean diffusivity (MD) measures. (A) Tissue enrichment results for proteins positively and negatively associated with brain volume. (B) Tissue enrichment results for proteins positively and negatively associated with MD measures. The 2,920 proteins were used as the background of enrichment. Significant enrichment at FDR-corrected P < 0.05 are colored in red. The brain tissue is also highlighted with rectangle and red color. Source data are provided as a Source Data file.



84

Figure S4 Tissue enrichment of positively and negatively associated proteins for area and thickness measures. (A) Tissue enrichment results for proteins positively and negatively associated with area. (B) Tissue enrichment results for proteins positively and negatively associated with thickness. The 2,920 proteins were used as the background of enrichment. Significant enrichment at FDR-corrected P < 0.05 are colored in red. The brain tissue is also highlighted with rectangle and red color. Source data are provided as a Source Data file

	Exposure (Protein)	Outcome (Imaging metric)	/Vs N			Beta (95%CI)	P value
	ENPP6 ITGB6 RGMB	Left putamen Left middletemporal Right middletemporal	29 29 11		┝╾┤ ┝╺╾┤	0.07 (0.04, 0.10) 0.08 (0.04, 0.12) 0.10 (0.05, 0.15)	3.10e-06 <b>*</b> 4.59e-05 <b>*</b> 5.23e-05 <b>*</b>
	MDK	Left thalamus	27	┝╾┤		-0.08 (-0.12, -0.04)	5.79e-05 *
me		Left insula Right supramarginal	38		ilei Laul	$0.04 \ (0.02, 0.06)$	1.60e-04 *
olu	ITGA11	Left inferiorparietal	20	لحا		-0.07 (-0.11 -0.03)	2.44e-04 ×
Š	TFPI	Left precuneus	25	1-1	H=I	0.05 (0.02, 0.08)	5.83e-04
	KIAA0319	Left medialorbitofrontal	18	H=1		-0.05 (-0.09, -0.02)	6.04e-04
	NCAM1	Right precentral	26	H=1		-0.05 (-0.08, -0.02)	6.07e-04
	OMG	Right superiortemporal	17		┝╼┥	0.09 (0.04, 0.15)	8.40e-04
	IL6	Right insula	6	⊢⊷⊣		-0.09 (-0.15, -0.04)	8.61e-04
	SEZ6L	Left inferiortemporal	23		H=-1	0.08 (0.03, 0.13)	9.38e-04
	LRRC37A2	Right lingual	56	н		-0.05 (-0.06, -0.04)	2.94e-21 *
	LEPR	Right supramarginal	13		i ⊨=⊣	0.11 (0.07, 0.16)	3.48e-06 *
	MOG	Left medialorbitofrontal	17		╞╼┥	0.09 (0.05, 0.14)	7.19e-05 <b>*</b>
ŋ	MOG		21	H=H		-0.07 (-0.11, -0.04)	1.54e-04 <b>*</b>
Are	MOG	Left lingual	15	F=		-0.09 (-0.14, -0.04)	2.56e-04 *
	NCAN	Left narsonercularis	24				4 67e=04 +
	NCAN	Right parsorbitalis	24			0.08 (0.03, 0.13)	6.68e-04
	LEPR	Right superiorparietal	14			0.08 (0.03, 0.13)	7.50e-04
	INHBB	Left rostralanteriorcingulate	18		H=-1	0.06 (0.02, 0.09)	9.54e-04
	INHBC	Left superiortemporal	31	H		-0.07 (-0.10, -0.04)	2.69e-06 *
	FAS	Right lateraloccipital	11	⊢		-0.18 (-0.26, -0.10)	1.16e-05 <b>*</b>
	NOMO1	Right lingual	14	⊢⊷⊣		-0.17 (-0.25, -0.09)	1.67e-05 <b>*</b>
	CA14	Right parahippocampal	18	┝╼┥	1	-0.13 (-0.20, -0.06)	1.27e−04 <b>★</b>
ic,	IL31RA	Right precuneus	35	H		-0.04 (-0.06, -0.02)	2.15e-04 <b>*</b>
É	PAMR1	Left temporalpole	45	H=+		-0.06 (-0.09, -0.03)	2.72e-04 <b>*</b>
	IL31RA	Right superiorparietal	35	H=1		-0.04 (-0.06, -0.02)	3.08e-04 *
	IL31RA	Left medialerhitefrantel	34	H=1		-0.04 (-0.06, -0.02)	3.79e-04 *
	IL 31RA	Left medialorbitorrontal	25			-0.09(-0.15, -0.04)	4.000-04 *
	REN	Left anterior thalamic radiation	10	r=1		0.25 (0.14, 0.36)	5.81e-06 <b>*</b>
A	REN	Left superior thalamic radiation	13			0.21 (0.09, 0.32)	3.29e-04 *
	REN	Right superior thalamic radiation	13		i i e e e i i i	0.17 (0.08, 0.27)	5.36e-04
	BTN3A2	Left cingulate gyrus part of cingulum	42		H .	0.08 (0.06, 0.09)	2.35e-22 *
	BTN2A1	Right parahippocampal part of cingulum	40	H		-0.07 (-0.09, -0.04)	4.85e-08 <b>*</b>
	TNFRSF4	Right acoustic radiation	22	⊢⊷⊣		-0.17 (-0.23, -0.11)	5.59e-08 <b>*</b>
	EDA2R	Right inferior fronto-occipital fasciculus	11	⊢-•		-0.30 (-0.43, -0.18)	3.19e-06 <b>*</b>
	FAP	Left inferior fronto-occipital fasciculus	13		⊢1	0.13 (0.07, 0.20)	4.78e-05 <b>*</b>
	LACTB2	Left cingulate gyrus part of cingulum	9	⊢		-0.17 (-0.25, -0.09)	7.26e-05 <b>*</b>
	GUSB	Right acoustic radiation	42		i H=-1	0.08 (0.04, 0.11)	9.23e-05 *
	CCL28	Right tract corticospinal tract	33		=	0.09 (0.04, 0.13)	9.53e-05 *
	ADSP	Right medial lemniscus	15			0.09 (0.04, 0.13)	1.39e-04 *
₽	COMT	Right anterior thalamic radiation	20		1-1	-0.13 (-0.20, -0.06)	1.000-04 *
Z	CTSD	Right superior thalamic radiation	27			0.09 (0.04 0.13)	1 99e-04 +
	CERT	Left anterior thalamic radiation	6	<b>⊢</b> ∎–-1		-0.20 (-0.31, -0.09)	2.53e-04 <b>*</b>
	CRELD2	Right cingulate gyrus part of cingulum	29	 H=H		-0.07 (-0.11, -0.03)	2.63e-04 *
	DSG3	Right anterior thalamic radiation	36			-0.07 (-0.10, -0.03)	2.89e-04 <b>*</b>
	COMT	Left cingulate gyrus part of cingulum	14	⊢1		-0.13 (-0.20, -0.06)	3.27e-04 <b>*</b>
	SETMAR	Right uncinate fasciculus	7	⊢		-0.11 (-0.17, -0.05)	4.37e-04 <b>*</b>
	SMOC1	Left corticospinal tract	32	⊢∎⊣		-0.07 (-0.11, -0.03)	5.16e-04
	ATOX1	Left anterior thalamic radiation	11	⊢		-0.17 (-0.27, -0.07)	7.93e-04
	CTF1	Left anterior thalamic radiation	6	<b>H--1</b>		-0.21 (-0.34, -0.09)	8.18e-04
	ATOX1	Left cingulate gyrus part of cingulum	11	⊢		-0.18 (-0.29, -0.08)	8.52e-04
	FAP	Right inferior fronto-occipital fasciculus	15		<b>   </b>	0.12 (0.05, 0.18)	8.70e-04
	WUK	Right superior thatamic radiation	22			0.10 (0.04, 0.16)	9.08e-04
			-0.6	-0.4 -0.2	0 0.2 0.4		
			Eff	ect estimate (Beta)	and 95% CIs		

Figure S5 The causal effect of protein on brain structure in the forward mendelian randomization (MR) at a strict clumping P threshold. The forest plot shows the significant MR relationships of IVW method with a strict clumping P threshold of  $5 \times 10^{-8}$ . All MR results with a nominal P < 0.001 are shown. Raw P values are shown in the right most column. The MR relationships meet the significance threshold of FDR-corrected P < 0.05 are marked with arterisk. All statistical tests were two-sided. Source data are provided as a Source Data file

	Exposure (Protein)	Outcome (Disease)	/Vs N		OR (95%CI)	P value
	BTN3A2		69	H	1.09 (1.06, 1.12)	1.31e-09
I.	BTN2A1	AD	48	н	0.90 (0.85, 0.95)	4.18e-05
I.	EDA2R		10	· · · · · · · · · · · · · · · · · · ·	1.40 (1.11, 1.77)	5.16e-03
i	BTN2A1		47	H	1.09 (1.04, 1.13)	6.25e-05
	BTN3A2	ADHD	53	H	0.95 (0.93, 0.98)	1.51e-04
	LACTB2		8	<b>⊢</b>	0.79 (0.68, 0.93)	3.77e-03
	INHBC	ALS	35	F	1.34 (1.11, 1.62)	2.61e-03
ī	BTN2A1		36	H	0.96 (0.94, 0.98)	2.77e-06
	BTN3A2	Anxiety	41	ji i	1.02 (1.01, 1.04)	2.88e-05
	FAP		20	H+1	1.09 (1.04, 1.13)	9.13e-05
	LRRC37A2		45	H	1.06 (1.03, 1.10)	6.17e-05
	FAP	ASD	18	<u>⊢</u>	1.19 (1.08, 1.32)	4.44e-04
	RGMB		13	<b>⊢</b> 1	1.16 (1.05, 1.29)	4.40e-03
	TNFRSF4		20	⊢I	0.87 (0.79, 0.96)	4.84e-03
Ī	BTN2A1		32	н	0.85 (0.80, 0.89)	1.65e-09
	BTN3A2		49	H	1.08 (1.05, 1.12)	4.60e-06
	CRELD2		24	H-4	1.13 (1.06, 1.20)	1.70e-04
	CA14	PD	13	F+-4	0.79 (0.70, 0.89)	1.80e-04
	NCAN	DP	19	<u>⊢</u>	1.19 (1.08, 1.32)	6.93e-04
	CERT		6		0.75 (0.61, 0.91)	3.41e-03
	SETMAR		5		1.16 (1.05, 1.29)	4.90e-03
	ITGB6		24	F+4	0.88 (0.80, 0.96)	4.95e-03
	BTN3A2		46	H	1.07 (1.05, 1.09)	1.43e-09
	BTN2A1	MDD	35	H	0.91 (0.88, 0.95)	1.72e-06
	ITGB6		25	}+-	1.07 (1.03, 1.13)	2.53e-03
	ENPP6	MS	13	<b>⊢−−−1</b>	0.62 (0.47, 0.82)	8.41e-04
	ENPP6	PD	33	H-1	0.78 (0.71, 0.85)	3.16e-08
	BTN3A2		36	н	1.23 (1.21, 1.26)	7.12e-90
	LRRC37A2		45	H	0.96 (0.94, 0.97)	3.01e-10
	BTN2A1		21	н	0.82 (0.76, 0.88)	2.20e-08
	TNFRSF4	SCZ	9	H-4	0.81 (0.72, 0.90)	2.28e-04
	FAP		4	<b>⊢</b> ₊–	1.22 (1.10, 1.36)	2.69e-04
	NHLRC3		4		1.52 (1.21, 1.91)	3.13e-04
	PAMR1		25	H-H	1.10 (1.04, 1.17)	2.04e-03
			Г 0	0.5 1 1.5 2 Effect estimate (OR) and 95% CIs		

100 Figure S6 The causal effect of protein on disease in the forward MR at a strict clumping P

101 threshold. The forest plot shows the significant MR relationships of IVW method with a strict clumping P102 threshold of  $5 \times 10^{-8}$ . Raw P values are shown in the right most column. The MR relationships meet the 103 significance threshold of FDR-corrected P < 0.05 are shown. All statistical tests were two-sided. Source 104 data are provided as a Source Data file.

	Exposure (Imaging metric)	Outcome (Protein)	IVs N		Beta (95%CI)	P value
		0414	24		0.45 (0.08, 0.02)	4.40- 05
	Picht superiortemporal	DCAN	21		0.15 (0.08, 0.22)	1.48e-05
	Right superiortemporal	MOC	26		0.19 (0.10, 0.28)	2.066-05
	Left thalamus	MOG	46		0.14 (0.07, 0.20)	2.59e-05
	Left lateralorbitofrontal	LRIMZ	31		0.16 (0.08, 0.24)	7.54e-05
me	Left insula	CNTN2	31		0.12 (0.06, 0.18)	1.19e-04
olu	Right putamen	CSF1	55		-0.11(-0.16, -0.05)	1.40e-04
>	Right thalamus	KLK6	37		0.16 (0.08, 0.24)	1.52e-04
	Left insula	LRTM2	42		0.12 (0.06, 0.19)	1.67e-04
	Left middletemporal	RTN4R	36		-0.13(-0.20, -0.06)	1.81e-04
	Right pericalcarine	NCAN	69	<del>  •</del> ·	0.07 (0.03, 0.10)	1.88e-04
	Right superiortemporal	GDF2	24		0.15 (0.07, 0.24)	2.75e-04
	Left fusiform	MOG	28		0.21 (0.13, 0.29)	3.77e-07
	Right cuneus	MOG	60	++-	0.11 (0.06, 0.15)	8.08e-07
	Left fusiform	NXPH1	27		0.18 (0.10, 0.26)	4.24e-06
	Right rostralmiddlefrontal	PTPRR	44		0.16 (0.09, 0.23)	8.94e-06
Irea	Left superiorfrontal	NCAN	45	<b> </b>	0.15 (0.08, 0.22)	1.60e-05
4	Right superiortemporal	B4GAT1	40	<b> </b> -+-	0.14 (0.07, 0.21)	3.54e-05
	Right cuneus	NCAN	57	<b>⊦</b> ∎-	0.09 (0.05, 0.13)	3.89e-05
	Right pericalcarine	NCAN	71	┟╾┤	0.07 (0.03, 0.10)	1.41e-04
	Right rostralmiddlefrontal	KIAA0319	44	<b>⊢</b> ⊷⊣	0.13 (0.06, 0.19)	2.72e-04
	Right superiortemporal	TGFA	21	<b>⊢</b> ⊷⊣ !	-0.16(-0.23, -0.08)	7.92e-05
	Right lingual	MSR1	28	<b>  - -</b>	0.12 (0.06, 0.17)	9.42e-05
	Left insula	FSTL3	20		0.14 (0.07, 0.22)	1.10e-04
hic	Left insula	ADM	17	<u> ⊢</u>	0.15 (0.07, 0.23)	1.76e-04
	Left insula	WFIKKN2	19	<b>⊢</b> ⊷-	-0.13 (-0.20, -0.06)	1.93e-04
	Left lateraloccipital	MIA	30	++-	0.09 (0.04, 0.15)	3.64e-04
	Right cuneus	NOMO1	29	<b> </b> -	0.10 (0.04, 0.15)	3.75e-04
	forceps minor	KLK6	45	<b> </b> +∎-	0.11 (0.07, 0.16)	2.30e-07
	Left superior thalamic radiation	SH3BP1	36	<b>  + -</b>	0.14 (0.08, 0.20)	1.35e-06
	Right superior longitudinal fasciculus	CD93	56	<b> </b> ■	-0.10(-0.15, -0.06)	3.95e-06
	Right superior thalamic radiation	LY96	31		0.14 (0.08, 0.20)	4.39e-06
	Right superior longitudinal fasciculus	PRND	50	H=H	-0.10(-0.15, -0.06)	5.46e-06
	Right anterior thalamic radiation	DSG4	20		0.18 (0.10, 0.27)	9.48e-06
	Right superior thalamic radiation	ELOA	30	i Line i li	0.13 (0.07, 0.20)	5.03e-05
₽	Right superior thalamic radiation	CDSN	9		-0.22(-0.32, -0.11)	7.61e-05
	forceps minor	OMG	46		0.08 (0.04, 0.13)	8.48e-05
	Right inferior longitudinal fasciculus	SETMAR	46		0.10 (0.05, 0.14)	8.49e-05
	Right anterior thalamic radiation	ACAA1	31	⊢++I <sup>1</sup>	-0.12(-0.19, -0.06)	1.44e-04
	Right anterior thalamic radiation	CD300LG	30	<b>⊢</b> ⊷-	-0.12(-0.18, -0.06)	1.60e-04
	Left medial lemniscus	SOST	17		-0.15(-0.22, -0.07)	2.43e-04
	Right inferior fronto-occipital fasciculus	CD300LG	50	F=-1	-0.08(-0.13, -0.04)	3.61e-04
	Right superior thalamic radiation	BRME1	29		-0.12(-0.18, -0.05)	3.66e-04
					1	
			÷	0.4 -0.2 0 0.2 0	).4	

106Figure S7 The causal effect of brain structure on protein. The forest plot shows the significant MR107relationships of IVW method with a clumping P threshold of  $5 \times 10^{-6}$ . Raw P values are shown in the right108most column. The MR relationships meet the significance threshold of FDR-corrected P < 0.05 are shown.109All statistical tests were two-sided. Source data are provided as a Source Data file.

Exposure (Disease)	Outcome (Protein)	/Vs N		OR (95%CI)	P value	
AD	TNFRSF4	83	н	0.97 (0.96, 0.99)	3.76e-05	
	PAMR1	52	┝╍┥	1.05 (1.02, 1.08)	3.18e-04	
ADHD	MAVS	54	⊢	1.05 (1.02, 1.08)	7.51e-04	
	F11R	52	┝╼╼┥	1.05 (1.02, 1.08)	8.52e-04	
ASD	REN	35	<b>⊢</b> 1	1.06 (1.02, 1.10)	1.17e-03	
PD	TNFRSF4	93	<b>⊢</b> ⊷4	0.95 (0.93, 0.97)	2.29e-06	
DF	BTN2A1	96	H=4	0.96 (0.94, 0.98)	5.07e-05	
MDD	BTN2A1	21	<b>⊢</b> -1	0.88 (0.83, 0.95)	5.00e-04	
MS	ENPP6	18	H=-1	0.92 (0.90, 0.94)	9.39e-14	
	AGER	35	H	1.04 (1.02, 1.05)	3.04e-06	
PD	LEPR	22	F+-1	1.04 (1.01, 1.06)	1.96e-03	
0.8 0.95 1.1 Effect estimate (OR) and 95% CIs						

Figure S8 The causal effect of disease on protein in the reverse MR. The forest plot shows the significant MR relationships of IVW method with a clumping P threshold of  $5 \times 10^{-6}$ . Raw P values are shown in the right most column. The MR relationships meet the significance threshold of FDR-corrected P < 0.05 are shown. All statistical tests were two-sided. Source data are provided as a Source Data file.



- 118 Figure S9 The expression of coding genes across tissues for proteins exhibiting mediation effect.
- 119 Source data are provided as a Source Data file.
- 120

## 121 **References**

Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* 622,
 329-338 (2023).