

- 1 **The circulating proteome and brain health: Mendelian randomisation and cross-sectional analyses**
- 2 Rosie M. Walker PhD^{a,b,c,*}
- 3 Michael Chong PhD^{a,d*}
- 4 Nicolas Perrot PhD^a
- 5 Marie Pigeyre MD^{a,e}
- 6 Danni A. Gadd PhD^f
- 7 Aleks Stolicyn PhD^c
- 8 Liu Shi PhD^{g,h}
- 9 Archie Campbell MAⁱ
- 10 Xueyi Shen PhD^c
- 11 Heather C. Whalley PhD^{c,i}
- 12 Alejo Nevado-Holgado PhD^g
- 13 Andrew M. McIntosh MD^c
- 14 Stefan Heitmeier^j
- 15 Sumathy Rangarajan MSc^a
- 16 Martin O'Donnell^{a,k}
- 17 Eric E. Smith MD^{a,l,m,n}
- 18 Salim Yusuf DPhil^{a,e,†}
- 19 William N. Whiteley FRCP^{a,c,o,†}
- 20 Guillaume Paré MD^{a,d,p,†}

21 ^aPopulation Health Research Institute, Hamilton Health Sciences and McMaster University, Hamilton,
22 ON, Canada

23 ^bSchool of Psychology, University of Exeter, Perry Road, Exeter, UK

24 ^cCentre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK

25 ^dDepartment of Pathology and Molecular Medicine, Faculty of Health Sciences, McMaster University,
26 Hamilton, ON, Canada

27 ^eDepartment of Medicine, Michael G DeGroote School of Medicine, Faculty of Health Sciences,
28 McMaster University, Hamilton, ON, Canada

29 ^fCentre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of
30 Edinburgh, Edinburgh, UK

31 ^gDepartment of Psychiatry, University of Oxford, Oxford, UK

32 ^hNxera Pharma UK Limited, Cambridge, UK

33 ⁱGeneration Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and
34 Cancer, University of Edinburgh, Edinburgh, EH4 2XU, UK

35 ^jBayer AG, Pharmaceuticals, R&D, 42113, Wuppertal, Germany

36 ^kHealth Research Board Clinical Research Facility, University of Galway, Galway, Ireland

37 ^lDepartment of Clinical Neurosciences and Hotchkiss Brain Institute, Cumming School of Medicine,
38 Calgary, Alberta, Canada

39 ^mUniversity of Calgary, Calgary, Alberta, Canada

40 ⁿDepartment of Clinical Neurosciences, University of Calgary, Calgary, Alberta, Canada

41 ^oMRC Centre for Population Health, University of Oxford, Oxford, UK

42 ^PThrombosis and Atherosclerosis Research Institute, Hamilton Health Sciences and McMaster

43 University, Hamilton, ON, Canada

44 *Joint first authors

45 [†]Joint final authors

46 Corresponding authors: pareg@mcmaster.ca and r.walker4@exeter.ac.uk

47

48 **Abstract**

49 Decline in cognitive function is the most feared aspect of ageing. Poorer midlife cognitive function is
50 associated with increased dementia and stroke risk. The mechanisms underlying variation in
51 cognitive function are uncertain. Here, we assessed associations between 1160 proteins' plasma
52 levels and two measures of cognitive function, the digit symbol substitution test (DSST) and the
53 Montreal Cognitive Assessment in 1198 PURE-MIND participants. We identified five DSST
54 performance-associated proteins (NCAN, BCAN, CA14, MOG, CDCP1), with NCAN and CDCP1
55 showing replicated association in an independent cohort, GS (N=1053). MRI-assessed structural
56 brain phenotypes partially mediated (8-19%) associations between NCAN, BCAN, and MOG, and
57 DSST performance. Mendelian randomisation analyses suggested higher CA14 levels might cause
58 larger hippocampal volume and increased stroke risk, whilst higher CDCP1 levels might increase
59 intracranial aneurysm risk. Our findings highlight candidates for further study and the potential for
60 drug repurposing to reduce risk of stroke and cognitive decline.

61 **Acronyms:** Average causal mediation effect (ACME); brevicin (BCAN); carbonic anhydrase 14 (CA14);
62 cluster of differentiation 6 (CD6); CUB-domain containing protein 1 (CDCP1); confidence interval (CI);
63 cerebral microbleed (CMB); cerebrospinal fluid (CSF); digit symbol substitution test (DSST);
64 extracellular matrix (ECM); false discovery rate (FDR); Generation Scotland imaging subsample (GS);
65 Generation Scotland: Scottish Family Health Study (GS:SFHS); instrumental variable (IV); inverse
66 variance weighted (IVW); myelin oligodendrocyte glycoprotein (MOG); Montreal Cognitive
67 Assessment (MoCA); Mendelian randomisation (MR); magnetic resonance imaging (MRI); neurocan
68 (NCAN); perineuronal net (PNN); odds ratio (OR); posterior probability (PP); protein quantitative trait
69 loci (pQTL); pairwise conditional analysis and co-localisation analyses (PWCoCo); Prospective Urban
70 and Rural Epidemiology (PURE); robust adjusted profile score (RAPS); silent brain infarct (SBI);
71 standard deviation (SD); small vessel disease (SVD); white matter hyperintensity (WMH)

72 **Key words:** cognition, dementia, Alzheimer's disease, stroke, proteins, brain structure, MRI,

73 Mendelian randomisation

74 **1. Introduction**

75 Decline in cognitive ability and dementia are the most feared aspects of ageing ^[1], providing a strong
76 rationale for investigating the mechanisms underlying cognitive function. Poorer cognitive function
77 is associated with a greater risk of Alzheimer’s dementia and stroke ^[2, 3]. This may be due to reduced
78 “cognitive reserve”, which postulates that lower premonitory cognitive function leads to worse
79 cognitive impairment for a given degree of neuropathology ^[4]. Better understanding of these
80 mechanisms could inform strategies for the prevention and treatment of dementia and stroke.

81 Recent studies have highlighted the potential for investigating cognition and structural brain
82 phenotypes through the study of plasma proteins ^[5-9]. These studies identified associations between
83 cognitive function and proteins involved in biological functions previously implicated in dementia,
84 including synaptic function, inflammation, immune function, and blood-brain barrier integrity ^[5-7, 9].
85 At the time of carrying out this study, previous studies were limited by a focus on a restricted
86 number of proteins and/or purely observational analyses.

87 Here, we investigated associations between 1160 plasma proteins and cognitive function in the
88 Prospective Urban and Rural Epidemiology (PURE)-MIND cohort ^[10], and we sought replication in the
89 independent imaging subsample of the Generation Scotland cohort (henceforth, referred to as
90 “GS”). The proteins assessed represent a wide range of biological processes, permitting a
91 hypothesis-free approach to investigating cognitive function. Subsets of these proteins have been
92 assessed in previous studies, allowing assessment of cross-study replication.

93 Using a simple measure of processing speed, the digit symbol substitution task (DSST), and a
94 cognitive screening tool, the Montreal Cognitive Assessment (MoCA), we carried out a screen for
95 cognition-associated proteins, and then employed mediation analyses to assess the proportion of
96 the protein expression-cognition relationship that could be explained by structural brain
97 phenotypes, including measures of brain volume and white matter hyperintensity (WMH) volume.
98 WMH is an MRI marker of white matter damage, and is one of the manifestations of age-related

99 cerebral small vessel disease. Two sample Mendelian randomisation (MR) analyses were performed
100 to assess potentially causal effects of genetically predicted protein levels on genetically predicted
101 cognitive function, brain structure, stroke subtypes, and Alzheimer’s disease (see **Figure 1** for an
102 overview of the study design).

103

104 2. Results

105 Cohort characteristics

106 Key demographic, cognitive, brain MRI, and health variables for the participants in PURE-MIND
107 (N=1198) and GS (N=1053) are summarised in **Table 1**. Participants in PURE-MIND were significantly
108 younger than participants in GS (PURE-MIND: mean = 54.5 years (SD = 8.05 years); GS: mean = 59.9
109 years (SD = 9.59 years); $p < 2.2 \times 10^{-16}$), and a small, but significant, between-cohort difference in
110 DSST score was observed (PURE-MIND: mean = 69.5 (SD = 15.3); GS: mean = 68.1 (SD = 15.2); $p =$
111 0.0298). Differences in the levels/types of education received by the two cohorts were observed ($p =$
112 5×10^{-4}). The two cohorts also differed significantly on several brain volume measurements: PURE-
113 MIND participants have a smaller total brain volume (PURE-MIND: mean = 1058 cm³ (SD = 110 cm³);
114 GS: mean = 1069 cm³ (SD = 109 cm³); $p = 0.0212$), cerebral white matter volume (PURE-MIND: mean
115 = 447 cm³ (SD = 60.5 cm³); GS: mean = 455 cm³ (SD = 56.9 cm³); $p = 1.87 \times 10^{-3}$), and hippocampal
116 volume (PURE-MIND: mean = 3.92 cm³ (SD = 0.433 cm³); GS: mean = 4.18 cm³ (SD = 0.437 cm³); $p <$
117 2.2×10^{-16}) than GS participants, whilst GS participants have a smaller intracranial volume (ICV) than
118 PURE-MIND participants (PURE-MIND: mean = 1491 cm³ (SD = 159 cm³); GS: mean = 1400 cm³ (SD =
119 225 cm³); $p < 2.2 \times 10^{-16}$). The two cohorts also differed significantly on other health-related
120 variables that were not directly assessed in this study (**Table 1**).

121 Scores on the DSST were normally distributed in both PURE-MIND and GS (**Supplementary Figure 1**),
122 but scores on the MoCA showed a leftward skew in PURE-MIND (**Supplementary Figure 1**).

123 Identification of protein biomarkers of cognitive function and enrichment analyses

124 Five proteins were associated with DSST performance in PURE-MIND (**Figure 2**; **Figure 3**;
125 **Supplementary Table 1**; **Supplementary Figure 2**). Higher plasma levels of neurocan (NCAN; $\beta = 2.03$
126 (indicating a 2.03 higher DSST score per a standard deviation higher NCAN level), $p = 9.11 \times 10^{-8}$),
127 brevican (BCAN; $\beta = 1.91$, $p = 5.56 \times 10^{-7}$), carbonic anhydrase 14 (CA14; $\beta = 1.90$, $p = 5.90 \times 10^{-7}$),

128 and myelin-oligodendrocyte glycoprotein (MOG; $\beta = 1.82$, $p = 2.29 \times 10^{-6}$), and lower levels of CUB
129 domain-containing protein 1 (CDCP1; $\beta = -1.57$, $p = 3.97 \times 10^{-5}$) were associated with significantly
130 better DSST performance. Adjustment for educational attainment modestly attenuated the effect
131 estimate for all five proteins (**Supplementary Table 1**). Levels of NCAN, BCAN and MOG were
132 positively correlated ($0.251 \leq r \leq 0.615$; all $p < 2.20 \times 10^{-16}$), whilst CDCP1 and CA14 expression levels
133 were negatively correlated ($r = -1.01$, $p = 4.82 \times 10^{-4}$; **Supplementary Table 2**). Three proteins (NCAN,
134 BCAN and CDCP1) proteins were also measured in GS, of which two replicated their association with
135 DSST performance: NCAN ($\beta = 1.40$, $p = 1.07 \times 10^{-3}$) and CDCP1 ($\beta = -1.99$, $p = 9.21 \times 10^{-6}$; **Figure 3**).
136 MoCA performance was not associated with the level of any protein (all $p \geq 2.34 \times 10^{-4}$;
137 **Supplementary Table 3; Supplementary Figure 3**). When considering the 129 proteins that were
138 nominally significantly associated ($p < 0.05$) with DSST score in PURE-MIND and measured in GS
139 (**Supplementary Table 4**), their effect estimates showed a strong, statistically significant between-
140 cohort correlation ($r = 0.625$, $p = 2.34 \times 10^{-15}$).

141 Proteins nominally associated ($p < 0.05$) with DSST performance (N = 184) were enriched for brain-
142 expressed proteins, most significantly for proteins with hippocampal expression (FDR-corrected $p =$
143 0.0154 ; **Supplementary Table 5**). Better DSST performance was nominally associated with lower
144 levels of 90 proteins. These proteins mapped to the following immune pathways “interleukin-10
145 signalling”, “glomerulonephritis”, “regulation of granulocyte chemotaxis”, “positive regulation of
146 leukocyte chemotaxis”, “positive regulation of leukocyte migration”, and “inflammation” (FDR-
147 corrected $p \leq 0.0337$; **Supplementary Table 6**).

148 **Structural brain phenotypes as mediators of protein biomarker-DSST performance associations**

149 In PURE-MIND, better DSST performance was associated with greater cerebral white matter volume
150 ($\beta = 0.0615$, $p = 4.34 \times 10^{-7}$), greater total brain volume ($\beta = 0.0349$, $p = 9.64 \times 10^{-6}$), greater
151 hippocampal volume ($\beta = 2.97$, $p = 4.79 \times 10^{-3}$), and lower log-transformed WMH volume ($\beta = -3.20$,
152 $p = 1.18 \times 10^{-6}$). These associations replicated in GS (**Supplementary Table 7**).

153 Assessment of the relationships between protein levels and DSST-associated structural brain
154 phenotypes in PURE-MIND revealed systematic differences between those proteins for which higher
155 levels were associated with better DSST performance (NCAN, BCAN, CA14, and MOG), and CDCP1,
156 which was negatively associated with DSST performance (**Figure 4**). Whilst NCAN, BCAN, CA14, and
157 MOG showed a positive direction of association with total brain, cerebral white matter, and
158 hippocampal volume measurements and a negative association with WMH volume, the converse
159 was true for CDCP1. The associations between NCAN levels and total brain, cerebral white matter,
160 and hippocampal volumes reached statistical significance ($p \leq 2.56 \times 10^{-5}$) and were replicated in GS
161 ($p \leq 6.70 \times 10^{-3}$). BCAN levels were significantly associated with all four brain volumes ($p \leq 4.36 \times 10^{-4}$),
162 with the associations with total brain and cerebral white matter volumes replicating in GS ($p \leq 3.44 \times 10^{-3}$).
163 The associations between MOG levels and total brain and cerebral white matter
164 volumes attained statistical significance ($p \leq 2.63 \times 10^{-9}$), but could not be assessed in GS. We did not
165 identify any significant associations with CA14 or CDCP1 levels after multiple testing correction.

166 In PURE-MIND, cerebral white matter volume explained a significant proportion of variance in the
167 relationship between MOG (19.2%), BCAN (14.9%), and NCAN (12.7%) levels and DSST performance
168 (all $p < 2 \times 10^{-16}$) (**Supplementary Table 8**). After controlling for cerebral white matter volume, the
169 average effect estimates (based on 1000 bootstrap resamples) for these proteins were reduced from
170 1.81 to 1.47 (MOG), 1.91 to 1.62 (BCAN), and 2.03 to 1.77 (NCAN). Log-transformed WMH volume
171 was a significant partial mediator of the association between BCAN levels and DSST performance ($p = 0.002$).
172 Controlling for log-transformed WMH resulted in a reduction in the effect estimate from
173 1.91 to 1.75 (8% mediation).

174 **Identification of potentially causal relationships between protein levels and cognitive function,** 175 **structural brain phenotypes, and disease outcomes**

176 Inverse variance weighted (IVW) MR analyses were performed to assess the effects of genetically
177 predicted CA14, CDCP1, and MOG levels on cognitive function, structural brain phenotypes, and

178 Alzheimer's disease, and stroke. Protein quantitative trait loci (pQTLs) located in *cis* to the genes
179 encoding the proteins-of-interest acted as instrumental variables (IVs) for plasma protein levels
180 (**Supplementary Table 9**). An insufficient number of pQTLs precluded the assessment of BCAN
181 and NCAN with any of the outcomes-of-interest. For MOG, limited overlap between the pQTLs and
182 SNPs included in the outcome GWASs meant only a subset of the outcomes-of-interest could be
183 assessed.

184 A one standard deviation higher level of genetically predicted plasma CA14 was associated with a
185 larger hippocampal volume ($\beta = 0.0971$ [95% CI: 0.0300 to 0.164], $p = 4.58 \times 10^{-3}$), and a greater risk
186 of all stroke (odds ratio (OR) = 1.08 [95% CI: 1.02 to 1.14], $p = 6.97 \times 10^{-3}$; **Supplementary Table 10**).
187 A one standard deviation higher level of genetically predicted plasma CDCP1 was associated with an
188 increased risk of intracranial aneurysm (OR = 1.22 [95% CI: 1.02 to 1.47], $p = 0.0280$). These
189 associations were corroborated by similar effect estimates from weighted median and MR-RAPS
190 analyses. No evidence of directional or horizontal pleiotropy were observed, and the correct causal
191 direction was assessed. No significant associations were observed between the genetically predicted
192 levels of CA14 or CDCP1 and risk of Alzheimer's disease ($p \geq 0.125$) (NB. A lack of significant pQTLs
193 precluded the assessment of BCAN, MOG, or NCAN).

194 Sensitivity analyses were performed in which instrumental variables (IVs) were selected using a
195 stricter threshold for independence. For genetically predicted CA14, these analyses supported the
196 association with hippocampal volume ($\beta = 0.144$ [95% CI: 0.0435 to 0.244], $p = 4.97 \times 10^{-3}$), and
197 produced a consistent, although non-significant, effect estimate for the association with risk for all
198 stroke (**Supplementary Table 10**). For CDCP1, the sensitivity analyses identified a consistent,
199 although non-significant, effect estimate for the association with risk for intracranial aneurysm.

200 To assess whether between-population heterogeneity in pQTL effects could have affected our
201 findings, we sought to replication using random effects meta-analysis (**Supplementary Table 10**). For
202 genetically predicted CA14, this approach supported the significant associations with hippocampal

203 volume ($\beta = 0.0953$ [95% CI: 0.0586 to 0.132], $p = 3.63 \times 10^{-7}$), and risk for all stroke (OR = 1.06 [95%
204 CI: 1.03 to 1.09], $p = 2.56 \times 10^{-4}$). For genetically predicted CDCP1, it was only possible to assess the
205 association with risk for intracranial aneurysm using a single IV; this identified a significant
206 association (OR = 1.32 [95% CI: 1.08 to 1.61], $p = 6.93 \times 10^{-3}$).

207 Pairwise Conditional Analysis and Co-localisation Analyses (PWCoCo) were performed to assess the
208 presence of a shared variant for each of the five proteins-of-interest and the same outcomes as
209 assessed by two-sample MR analyses. We were only adequately powered to assess co-localisation
210 between SNPs associated with one pair of traits: MOG plasma level and cognitive function. We did
211 not observe any evidence in support of co-localisation or conditional co-localisation (posterior
212 probability (PP)₄/PP₃ $\leq 4.81 \times 10^{-4}$).

213 3. Discussion

214 In this large-scale analysis of the associations between the plasma levels of 1160 proteins and
215 cognitive function, we identify CA14 and CDCP1 as being associated with processing speed, as
216 measured by the DSST, and having potentially causal effects on hippocampal volume and stroke
217 (CA14)and intracranial aneurysm (CDCP1).

218 Other proteins (BCAN, NCAN, and MOG) were associated with DSST performance and important
219 structural brain phenotypes, with cerebral white matter volume mediating a significant proportion
220 (13-19%) of the relationship between the levels of all three proteins and DSST performance, and
221 WMH volume mediating 8% of the relationship between BCAN levels and DSST performance. A lack
222 of genetic instruments precluded the assessment of potentially causal effects of BCAN and NCAN
223 with any outcome-of-interest, and MOG with several outcomes-of-interest.. Enrichment analyses of
224 proteins that were nominally significantly associated with DSST performance revealed a significant
225 enrichment for brain-expressed proteins.

226 There were no significant associations between plasma protein levels and performance on the
227 MoCA. This might reflect the fact that the MoCA is a screening tool for mild cognitive impairment^[11],
228 meaning its sensitivity to detect variation in cognitive function in non-clinical groups is likely to be
229 limited. The maximum MoCA score is 30 and scores higher than 26 indicate normal function. A high
230 mean score (26.5) and a left-skewed distribution indicate a ceiling effect, which likely limited power
231 to detect associations between protein levels and MoCA score in PURE-MIND.

232 CA14 is one of fifteen isoforms of the carbonic anhydrase family of zinc metalloprotease enzymes,
233 which catalyse the reversible hydration of carbon dioxide^[12]. CA14 is expressed by neurons^[13] and
234 involved in regulating extracellular pH following synaptic transmission^[14, 15]. Consistent with our
235 findings, acute inhibition of CA14 leads to impaired performance on cognitive tasks in mice^[16].
236 Carbonic anhydrase activation may lead to beneficial cognitive effects in rodents^[17]. In keeping with
237 our MR results, there are neuroprotective effects of carbonic anhydrase inhibition in models of

238 amyloidosis, Huntington’s disease, and ischaemic and haemorrhagic stroke ^[17]. The mechanisms by
239 which carbonic anhydrase inhibition and activation exert their effects are uncertain ^[16, 17]. FDA-
240 approved carbonic anhydrase inhibitors, and thus the majority of carbonic anhydrase inhibitors
241 investigated to date, are pan-carbonic anhydrase inhibitors. Of the carbonic anhydrase family
242 members measured in our study (CA1, 2, 3, 4, 5A, 6, 9, 12, 13, and 14), only CA14 levels were
243 significantly associated with DSST performance. Further studies are required to determine the
244 therapeutic potential for carbonic anhydrase modulation in the context of cognitive impairment,
245 Alzheimer’s disease, and stroke.

246 The extracellular matrix (ECM) proteins NCAN and BCAN are brain-specific chondroitin sulfate
247 proteoglycans, which are expressed by neurons and astrocytes (NCAN and BCAN), and
248 oligodendrocytes (BCAN). They contribute to the formation of a specialised structure, the
249 perineuronal net (PNN), which plays a key role in memory and neuronal plasticity, and which is
250 disrupted in Alzheimer’s disease ^[18]. Our findings are consistent with those of Harris et al. (2020) ^[5],
251 who found plasma levels of NCAN and BCAN were positively associated with brain volume. Plasma
252 levels of both NCAN and BCAN have previously been shown to be positively associated with general
253 cognitive function and DSST performance ^[9], whilst BCAN levels have been found to be positively
254 associated with Mini Mental State Examination performance and reduced in patients with
255 Alzheimer’s disease or mild cognitive impairment ^[7]. Mice that are lacking either NCAN or BCAN
256 expression show normal development and memory function but reduced hippocampal long term
257 potentiation ^[19, 20], whilst quadruple knock-outs, which lack NCAN, BCAN, and two additional ECM
258 proteins (tenascin-C and tenascin-R) show an altered ratio of excitatory to inhibitory synapses and a
259 reduction in the number and complexity of hippocampal PNNs ^[21]. Genetic variation in the gene
260 encoding A Disintegrin and Metalloproteinase with Thrombospondin Motifs 4 (ADAMTS4), which
261 degrades the four members of the lectican family (including NCAN and BCAN), has been implicated
262 in Alzheimer’s disease ^[22]. Taken together, the evidence suggests NCAN, BCAN and their regulators
263 as molecules-of-interest in Alzheimer’s disease.

264 MOG is an oligodendrocyte-expressed membrane glycoprotein, the exact function of which is
265 unknown^[23].

266 CDCP1 is a widely expressed transmembrane glycoprotein that acts as a ligand for T cell-expressed
267 Cluster of Differentiation 6 (CD6), and is implicated in autoimmune conditions^[24]. CDCP1 is
268 amenable to modulation by approved drug treatments: Itolizumab, which is used to treat psoriasis,
269 disrupts CDCP1-CD6 binding and downregulates T-cell-mediated inflammation^[25], whilst
270 atomoxetine, a treatment for attention deficit hyperactivity disorder, which is being considered for
271 the treatment of mild cognitive impairment, reduced cerebrospinal fluid (CSF) CDCP1 levels^[26].
272 Intriguingly, findings in mice suggest a functional link between CDCP1 and MOG^[6].

273 Our study has several strengths. We measured 1160 proteins, associated with a wide range of
274 physiological processes, in a large, well-characterised cohort. Replication analyses, where possible,
275 were performed in an independent cohort in which proteins were measured using an independent
276 methodology. The availability of genetic and brain MRI data permitted an exploration of causality
277 and putative causal pathways. The use of MR to identify potentially causal associations will have
278 offered protection against some of the common confounders of observational analyses^[27], with the
279 use of multiple MR methods, which generally gave concordant estimates of effect, mitigating against
280 the individual biases of different MR methodologies^[28]. Moreover, by requiring instrumental
281 variables to be located in *cis* to their target protein, we limited the chance of pleiotropic effects^[29].

282 There are also several limitations to consider.

283 First, the 1160 proteins measured represent a small subset of the circulating proteome^[30]. Although
284 these proteins are involved in a wide range of biological functions represented by all 13 Olink Target
285 96 panels, limitations to our understanding of the proteome mean that it is not possible to assess the
286 extent to which these proteins are representative of the rest of the proteome. Replication analyses
287 were only performed for those proteins for which data were available in the GS cohort, meaning
288 that we did not assess replication of CA14 or MOG.

289 Second, the availability of suitable IVs mean that our primary MR analyses were only performed for
290 CA14, CDCP1, and MOG. Whilst we required a minimum of three IVs for the primary MR analyses,
291 our sensitivity analyses, in which a stricter threshold for independence was applied to the IVs
292 necessitated the use of fewer than three IVs in each analysis. As such, the results of the sensitivity
293 analyses should be interpreted with this caveat in mind.

294 Third, for all but one pair of traits, we were insufficiently powered to assess co-localisation between
295 genetic variants associated with protein level and cognition, structural brain phenotypes, and
296 disease outcomes. This means that it is possible that significant MR findings might reflect the
297 presence of separate causal variants in linkage disequilibrium (LD) with one another^[31]

298 Fourth, we measured protein levels in the plasma, rather than in the brain or CSF. It is, however,
299 important to note the striking enrichment for brain-expressed proteins amongst the DSST-associated
300 proteins. Previous analyses of the GS cohort, in which replication was sought in the present study,
301 have identified the levels of several plasma proteins as being associated with multiple markers of
302 brain health^[8]. These findings support the use of the plasma to assess brain-related phenotypes and
303 emphasise the need for additional research to explain the mechanisms controlling the efflux of
304 brain-expressed proteins into the bloodstream in non-clinical populations. Moreover, the use of *cis*
305 pQTLs, which are likely to be shared across tissues^[32], as IVs in our MR analyses, supports the
306 possibility that the MR-identified associations reflect the actions of the proteins-of-interest in the
307 brain.

308 In summary, we identified protein biomarkers of cognitive function that may causally affect brain
309 structure and risk for stroke and intracranial aneurysm. Notwithstanding the need for replication,
310 our findings prompt several hypotheses that should be assessed by future studies. Our apparently
311 paradoxical findings of higher CA14 levels being associated with both better cognitive function and
312 increased stroke risk suggest that molecular findings can inform a more nuanced understanding of
313 the relationship between premorbid cognitive function and neurological disease risk. It is possible

314 that improved risk stratification may be achieved through the combination of cognitive assessment
315 and biomarker measurement. The availability of approved drugs targeting our identified proteins
316 raises the possibility of drug repurposing for novel therapeutic interventions to prevent cognitive
317 decline, stroke, and intracranial aneurysm.

318

319 4. Methods

320 Sample information

321 This study used data from participants of self-reported European (N = 3514), Latin (N = 4309), or
322 Persian (N = 1332) ancestry from the Population Urban Rural Epidemiology (PURE) biomarker sub-
323 study ^[33] (**Supplementary Information**). African (n=659), South Asian (n=604), East Asian (n=314),
324 and Arab (n=204) participants were excluded to align PURE genetic data with external genetic
325 datasets, which are predominantly European. Participants of Latin and Persian ancestry were
326 included due to their genetic overlap with European participants ^[33].

327 The PURE biomarker study also included European participants enrolled in PURE-MIND (N = 1198) ^[10]
328 (**Supplementary Information**).

329 The European, Latin, and Persian PURE biomarker cohort participants were used to identify protein
330 biomarker pQTLs ^[33], for use in MR analyses, whilst data from the European PURE-MIND biomarker
331 participants were used for observational association analyses.

332 We sought replication of our observational findings in GS ^[34, 35], which was recruited through re-
333 contact of the Generation Scotland: Scottish Family Health Study (GS:SFHS) ^[36, 37]. GS:SFHS is a
334 population- and family-based cohort of >24,000 individuals from Scotland. GS:SFHS participants
335 were recruited between 2006 and 2011. Upon recruitment, participants attended a clinic where
336 detailed health, cognitive, and lifestyle information, and biological samples were collected. Between
337 2015 and 2018, a subset of the GS:SFHS participants completed additional health and cognitive
338 assessments, brain MRI, and provided blood samples for proteomic analysis. Up to 1053 GS
339 participants were available for replication analyses.

340 Ethical approval

341 All centres contributing to PURE were required to obtain approval from their respective ethics
342 committees (Institutional Review Boards). Participant data is confidential and only authorized

343 individuals can access study-related documents. The participants' identities are protected in
344 documents transmitted to the Coordinating Office, as well as biomarker and genetic data.
345 Participants provided informed consent to obtain baseline information, and to collect and store
346 genetic and other biological specimens.

347 The GS:SFHS obtained ethical approval from the NHS Tayside Committee on Medical Research
348 Ethics, on behalf of the National Health Service (reference: 05/S1401/89). All participants provided
349 broad and enduring written informed consent for biomedical research. GS:SFHS has Research Tissue
350 Bank Status (reference: 15/ES/0040), providing generic ethical approval for a wide range of uses
351 within medical research. The imaging subsample of GS:SFHS (referred to as "GS" herein) received
352 ethical approval from the NHS Tayside committee on research ethics (reference 14/SS/0039). All
353 experimental methods were in accordance with the Helsinki declaration.

354 **Assessment of between-cohort differences**

355 Between cohort differences in quantitative variables (age, DSST score, BMI, blood pressure, and
356 brain MRI volumes) were assessed using two-sample t-tests. Categorical variables (sex, education
357 type, and disease status) where the smallest count in any cell of the contingency table was greater
358 than five were assessed using a chi-squared test; otherwise, a Fisher's Exact Test was employed.
359 Statistical significance was defined as $p < 0.05$.

360 **Assessment of cognitive function**

361 General cognitive ability was measured in PURE-MIND and GS by trained assessors using the DSST
362 (Wechsler Adult Intelligence Scale, 3rd Edition) ^[38]. The DSST is a pencil and paper test in which
363 participants must match symbols to numbers according to a key. Participants were scored according
364 to the number of correct matches made within two minutes (maximum score: 133). The DSST
365 measures several cognitive functions, including associative learning and executive function ^[39], and
366 DSST performance is highly correlated with the general intelligence factor, g . PURE-MIND

367 participants completed the Montreal Cognitive Assessment (MoCA) ^[11], a questionnaire-based test
368 with scores 0 to 30. A score of 26 or higher is considered normal ^[11].

369 **Measurement of plasma protein expression**

370 In the PURE biomarker cohort, 1196 plasma protein levels were measured by proximity extension
371 assay using the Olink Proseek Target 96 reagent kit (Olink, Uppsala, Sweden) in 12066 participants
372 (including 3735 European, 4695 Latin, and 1436 Persian). Following pre-processing and quality
373 control steps (**Supplementary Information**), measurements were available for 1160 biomarkers in
374 8369-9154 European, Latin, or Persian participants (depending on biomarker-specific missingness).

375 In GS, plasma protein levels were measured with the SOMAscan assay platform (SomaLogic Inc.), as
376 described previously ^[40]. Following initial data processing and quality control steps, measures of
377 4058 proteins were available in 1095 participants. Prior to analysis, protein abundance
378 measurements were log-transformed and rank-based inverse normalised.

379 **Brain imaging**

380 PURE-MIND participants enrolled in the PURE biomarker cohort were scanned at four sites in Canada
381 (three at 1.5T (two on General Electric (GE) scanners, one on a Phillips scanner), one at 3T (GE)). The
382 brain imaging phenotypes assessed in this study were total brain volume (excluding ventricles), total
383 white matter volume, hippocampal volume, average cortical thickness, a multi-region composite
384 thickness measure designed to differentiate Alzheimer's disease patients from clinically normal
385 participants ^[41], silent brain infarcts (SBI), cerebral microbleeds (CMB), and WMH volumes. These will
386 henceforth be referred to as the "structural brain phenotypes". Further information about the
387 derivation of the structural brain phenotypes is available in the **Supplementary Information**.

388 **Genotyping and imputation of PURE-MIND**

389 PURE participant genotypes (ThermoFisher Axiom Precision Medicine Research Array r.3) were called
390 using Axiom Power Tools and in-house scripts. Quality control steps are described in the
391 **Supplementary Information**.

392 Imputation was performed on the 749,783 genotyped variants following the TOPMed Imputation
393 server pipeline (<https://imputation.biodatacatalyst.nih.gov/>). Further details are in the
394 **Supplementary Information**.

395 **Assessment of the association between protein biomarkers and cognitive function and structural** 396 **imaging phenotypes**

397 We assessed the association between standardised protein levels and cognitive and structural brain
398 phenotypes using two-tailed linear (DSST, MoCA, total brain volume, white matter volume,
399 hippocampal volume, WMH volume, cortical thickness) or logistic (CMB, SBI) regression. The
400 cognitive or structural brain phenotype-of-interest was the dependent variable with the
401 standardised protein expression level, age, age², sex, and the first ten genetic principal components
402 as independent variables. A sensitivity analysis was performed for DSST-associated proteins in which
403 we further adjusted for education (a categorical variable with levels: (i) no education; (ii) high school
404 or less; (iii) trade school; and (iv) college or university). We calculated Pearson's correlation
405 coefficient to assess the pairwise correlations between DSST-associated proteins. Within each
406 analysis, we applied a Bonferroni correction to determine statistical significance, yielding the
407 following significance thresholds: $p < 4.31 \times 10^{-5}$ when assessing associations with 1160 proteins; $p <$
408 2.5×10^{-3} when assessing associations with the five DSST-associated proteins across four DSST-
409 associated structural brain phenotypes; and $p < 5 \times 10^{-3}$ when assessing 10 pairwise correlations
410 between proteins.

411 We performed replication analyses in GS for the significant proteins identified in PURE-MIND. Two-
412 tailed mixed effects models were fitted using the `lmekin` function from the R package `coxme` v.2.2.17
413 ^[42] to assess the association of the outcome variable (DSST performance, total brain volume

414 (excluding ventricles), cerebral white matter volume, hippocampal volume, and WMH volume) with
415 standardised protein expression, covarying for age, age², sex, study site (Dundee or Aberdeen), the
416 delay between blood sampling and protein extraction, depression (a binary variable representing
417 lifetime depression status), and a kinship matrix. When a brain volume phenotype was the outcome
418 variable, additional covariates were included to account for ICV, the interaction between ICV and
419 study site (to account for a site-associated batch effect on ICV measurement), and whether there
420 was manual intervention using tools within Freesurfer during the quality control process. Replication
421 was defined as a concordant direction of effect, meeting a Bonferroni-corrected threshold of $p <$
422 1.67×10^{-2} (accounting for the assessment of three DSST-associated proteins) or $p < 7.14 \times 10^{-3}$
423 (accounting for the assessment of seven structural brain phenotype-protein combinations).

424 **Assessment of the association of DSST performance with MRI-derived structural brain phenotypes**

425 To identify mediators of the association between protein expression and DSST performance, we first
426 established the structural brain phenotypes that satisfied the requirements of potential mediators
427 (i.e. associated with both DSST performance and at least one DSST-associated protein), and then
428 formally tested the mediation relationship by bootstrap mediation analyses.

429 We estimated the association between DSST performance and structural brain phenotypes in PURE-
430 MIND using linear models. All brain volume measurements were normalised to ICV and the models
431 included covariates for age, age², sex, and the first ten genetic principal components. We defined
432 statistical significance as $p < 0.00625$ (Bonferroni correction for eight phenotypes; two-tailed) and
433 sought replication of significant associations ($N = 5$) in GS. In GS, brain volumes were residualised for
434 ICV, scanner location, the interaction between ICV and scanner location, and whether there was
435 manual intervention during the quality control process. The resultant residuals were included as the
436 dependent variable in a mixed effects model with DSST score, age, age², sex, depression, and a
437 kinship matrix as independent variables. Statistical significance was defined as $p < 0.0125$ (two-
438 tailed).

439 The DSST performance-associated brain MRI phenotypes (N = 4) were assessed as potential
440 mediators of the protein level-DSST associations (N = 3, yielding a total of N = 9 mediations to assess)
441 using bootstrap mediation analysis in PURE-MIND. Analyses were performed using the R package
442 “mediation”^[43] with 1000 bootstraps. We corrected for the nine potential mediation relationships
443 assessed using a Bonferroni-corrected threshold of $p < 5.56 \times 10^{-3}$.

444 **Functional and tissue-specific expression enrichment analyses**

445 Proteins associated with DSST performance at $p < 0.05$ in PURE-MIND were included in functional
446 and tissue-specific expression analyses in three groups: (i) all proteins; (ii) positively associated
447 proteins; and (iii) negatively associated proteins. Enrichment was assessed relative to all proteins in
448 our dataset that passed quality control (N =1160). Functional enrichment analyses were performed
449 using WebGestalt (<http://www.webgestalt.org/>)^[44] using default parameter settings for the over-
450 representation analysis method to assess enrichment for: (i) gene ontology categories (biological
451 processes, molecular functions, and cellular compartments); (ii) Reactome pathways; and (iii)
452 disease-associated genes (Disgenet). Tissue-specific enrichment analyses were performed using the
453 “GTEx v8: 54 tissue types” and “GTEx v8: 30 general tissue types” gene expression datasets in
454 FUnctional Mapping and Annotation (FUMA)^[45]. For both the functional enrichment and tissue
455 expression analyses, enrichment was assessed using a hypergeometric test and significant
456 enrichment was defined as a Benjamini-Hochberg-adjusted $p < 0.05$, correcting for the number of
457 tests performed within each analysis platform. Analyses were performed using web interfaces
458 accessed on 18/04/2022 (WebGestalt and FUMA) and 14/01/2023 (FUMA).

459 **Two-sample forward MR analyses**

460 We performed two-sample forward MR analyses to identify potentially causal associations between
461 genetically predicted plasma protein levels and: (i) cognitive function; (ii) structural brain
462 phenotypes (total brain volume, cerebral white matter volume, hippocampal volume, WMH volume,

463 and CMB); and (iii) disease outcomes (Alzheimer's disease, all stroke, stroke subtypes (ischaemic,
464 cardioembolic, large artery, and small vessel), and intracranial aneurysm).

465 Associations between single nucleotide polymorphisms (SNPs) and plasma protein expression levels
466 were calculated in PURE (**Supplemental Information**). Following quality control, a set of pQTLs that
467 was independent ($r^2 < 0.1$) in all three populations was retained. Sensitivity analyses were performed
468 in which the pruning threshold was adjusted to $r^2 < 0.01$.

469 The independent set of pQTLs were assessed for their associations with cognitive function, structural
470 brain phenotypes, and disease outcomes using summary statistics from published studies ^[46-53].

471 MR analyses were performed using the R packages MRBase for TwoSample MR v.0.5.6 ^[54], mr.raps
472 v.0.4.1 ^[55], and MRPRESSO v.1.0 ^[56]. We employed several complementary MR approaches: IVW ^[57],
473 weighted median ^[58], robust adjusted profile scores (RAPS) ^[55], MR-Egger ^[59], and MR-PRESSO ^[56]. We
474 adopted the IVW approach as our primary methodology and defined statistical significance using a
475 liberal within-outcome variable Bonferroni correction for the proteins (CA14, CDCP1, and MOG) that
476 could be assessed, yielding a significance threshold of $p < 0.0167$ (or $p < 0.025$ or 0.05 when an
477 outcome could only be assessed for two or one protein(s)). Further details of the MR analyses are
478 included in the **Supplementary Information**.

479 **Pairwise Conditional Analysis and Co-localisation Analysis (PWCoCo)**

480 PWCoCo ^[31, 60] was performed to assess the existence of a shared causal variant between (i) pQTLs
481 for each of the five proteins-of-interest and (ii) variants associated with the outcomes assessed in
482 the two-sample MR analyses (**Supplementary Information**).

483 **Software**

484 Statistical analyses and plot generation were performed in R (versions 3.6.0, 4.1.1, 4.1.2, 4.2.0,
485 4.2.1).

486 **Data availability**

487 The terms of consent for PURE participants preclude the sharing of individual-level data. Individual
488 level data is available through collaboration with PURE researchers
489 (<https://www.phri.ca/research/pure/>). Summary-statistics for the analyses presented here are
490 available in the supplementary materials.

491 According to the terms of consent for GS participants, applications for individual-level data must be
492 reviewed by the GS Access Committee (access@generationscotland.org). Complete summary
493 statistics are available in the supplementary materials for the protein-DSST score associations
494 assessed in this study.

495 **Code availability**

496 The code used to generate the results in this study is available on reasonable request from the
497 corresponding author.

498 **Funding and acknowledgements**

499 *PURE*

500 The PURE study is an investigator-initiated study that is funded by the Population Health Research
501 Institute, the Canadian Institutes of Health Research (CIHR), Heart and Stroke Foundation of Ontario,
502 support from CIHR's Strategy for Patient Oriented Research, through the Ontario SPOR Support Unit,
503 as well as the Ontario Ministry of Health and Long-Term Care and through unrestricted grants from
504 several pharmaceutical companies (with major contributions from AstraZeneca [Canada], Sanofi-
505 Aventis [France and Canada], Boehringer Ingelheim [Germany and Canada], Servier, and
506 GlaxoSmithKline), and additional contributions from Novartis and King Pharma and from various
507 national or local organisations in participating countries as follows: Argentina—Fundacion ECLA;
508 Bangladesh—Independent University, Bangladesh and Mitra and Associates; Brazil—Unilever Health
509 Institute, Brazil; Canada—Public Health Agency of Canada and Champlain Cardiovascular Disease
510 Prevention Network; Chile—Universidad de la Frontera; Colombia—Colciencias (grant number 6566–

511 04–18062); South Africa—The North-West University, SANPAD (SA and Netherlands Programme for
512 Alternative Development), National Research Foundation, Medical Research Council of South Africa,
513 The South Africa Sugar Association, Faculty of Community and Health Sciences; Sweden—grants
514 from the Swedish State under the Agreement concerning research and education of doctors, the
515 Swedish Heart and Lung Foundation, the Swedish Research Council, the Swedish Council for Health,
516 Working Life and Welfare, King Gustaf V's and Queen Victoria Freemasons Foundation, AFA
517 Insurance, Swedish Council for Working Life and Social Research, Swedish Research Council for
518 Environment, Agricultural Sciences and Spatial Planning, grant from the Swedish State under
519 (LäkarUtbildningsAvtalet) Agreement, and grant from the Västra Götaland Region; and United Arab
520 Emirates—Sheikh Hamdan Bin Rashid Al Maktoum Award for Medical Sciences, Dubai Health
521 Authority, Dubai. The PURE biomarker project was supported by Bayer and the CIHR. The biomarker
522 project was led by PURE investigators at the Population Health Research Institute (Hamilton,
523 Canada) in collaboration with Bayer scientists. Bayer directly compensated the Population Health
524 Research Institute for measurement of the biomarker panels, scientific, methodological, and
525 statistical work. Genetic analyses were supported by CIHR (G-18–0022359) and Heart and Stroke
526 Foundation of Canada (application number 399497) in the form of funding to GP.

527 GS

528 This work was supported by the Wellcome Trust [104036/Z/14/Z, 220857/Z/20/Z, and
529 216767/Z/19/Z] and an MRC Mental Health Data Pathfinder Grant [MC_PC_17209] to AMM. DAG is
530 funded by the Wellcome Trust Translational Neuroscience PhD Programme at the University of
531 Edinburgh [108890/Z/15/Z]. LS and ANH are supported by Medical Research Council
532 [MR/L023784/2]: Dementias Platform UK. LS is also supported by a Medical Research Council Award
533 to the University of Oxford [MC_PC_17215]. AS is supported through the Wellcome-University of
534 Edinburgh Institutional Strategic Support Fund (Reference 204804/Z/16/Z), and indirectly through
535 the Lister Institute of Preventive Medicine award with reference 173096. Generation Scotland

536 received core support from the Chief Scientist Office of the Scottish Government Health Directorates
537 [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was
538 carried out by the Genetics Core Laboratory at the Clinical Research Facility, Edinburgh, Scotland,
539 and was funded by the UK's Medical Research Council and the Wellcome Trust [104036/Z/14/Z].

540 We are grateful to all the families who took part in Generation Scotland, the general practitioners
541 and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation
542 Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers,
543 research scientists, volunteers, managers, receptionists, healthcare assistants and nurses.

544 We thank Dr Alison Offer for assistance in producing the forest plots.

545 **Conflicts of interest**

546 MC is supported by a Canadian Institute of Health Research doctoral award and has received
547 consulting fees from Bayer AG. MP is supported by the EJ Moran Campbell Internal Career Research
548 Award from McMaster University. DAG is a part-time employee of Optima partners, a health data
549 consultancy based at the Bayes centre, The University of Edinburgh. SH is an employee of Bayer AG.
550 AMM has previously received speaker's fees from Illumina and Janssen and research grant funding
551 from The Sackler Trust. SY is supported by the Heart and Stroke Foundation/Marion W Burke Chair in
552 Cardiovascular Disease. GP is supported by the CISCO Professorship in Integrated Health Systems.
553 The other authors declare no competing interests.

554

555 **References**

- 556 1. Martin, G.M., *Defeating Dementia*. Nature, 2004. **431**: p. 247-248.
- 557 2. Rostamian, S., S. Mahinrad, T. Stijnen, B. Sabayan, and A.J. de Craen, *Cognitive impairment*
558 *and risk of stroke: a systematic review and meta-analysis of prospective cohort studies*.
559 Stroke, 2014. **45**(5): p. 1342-8.
- 560 3. Valenzuela, M.J. and P. Sachdev, *Brain reserve and dementia: a systematic review*. Psychol
561 Med, 2006. **36**(4): p. 441-54.
- 562 4. Pettigrew, C. and A. Soldan, *Defining Cognitive Reserve and Implications for Cognitive Aging*.
563 Curr Neurol Neurosci Rep, 2019. **19**(1): p. 1.
- 564 5. Harris, S.E., S.R. Cox, S. Bell, R.E. Marioni, B.P. Prins, A. Pattie, et al., *Neurology-related*
565 *protein biomarkers are associated with cognitive ability and brain volume in older age*. Nat
566 Commun, 2020. **11**(1): p. 800.
- 567 6. Lindbohm, J.V., N. Mars, K.A. Walker, A. Singh-Manoux, G. Livingston, E.J. Brunner, et al.,
568 *Plasma proteins, cognitive decline, and 20-year risk of dementia in the Whitehall II and*
569 *Atherosclerosis Risk in Communities studies*. Alzheimers Dement, 2022. **18**(4): p. 612-624.
- 570 7. Whelan, C.D., N. Mattsson, M.W. Nagle, S. Vijayaraghavan, C. Hyde, S. Janelidze, et al.,
571 *Multiplex proteomics identifies novel CSF and plasma biomarkers of early Alzheimer's*
572 *disease*. Acta Neuropathol Commun, 2019. **7**(1): p. 169.
- 573 8. Gadd, D.A., R.F. Hillary, D.L. McCartney, L. Shi, A. Stolicyn, N.A. Robertson, et al., *Integrated*
574 *methylome and phenome study of the circulating proteome reveals markers pertinent to*
575 *brain health*. Nat Commun, 2022. **13**(1): p. 4670.
- 576 9. Tin, A., A.E. Fohner, Q. Yang, J.A. Brody, G. Davies, J. Yao, et al., *Identification of circulating*
577 *proteins associated with general cognitive function among middle-aged and older adults*.
578 Communications Biology, 2023. **6**(1): p. 1117.

- 579 10. Smith, E.E., M. O'Donnell, G. Dagenais, S.A. Lear, A. Wielgosz, M. Sharma, et al., *Early*
580 *cerebral small vessel disease and brain volume, cognition, and gait*. *Ann Neurol*, 2015. **77**(2):
581 p. 251-61.
- 582 11. Nasreddine, Z.S., N.A. Phillips, V. Bedirian, S. Charbonneau, V. Whitehead, I. Collin, et al., *The*
583 *Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment*. *J*
584 *Am Geriatr Soc*, 2005. **53**(4): p. 695-9.
- 585 12. Lindskog, S., *Structure and mechanism of carbonic anhydrase*. *Pharmacol Ther*, 1997. **74**(1):
586 p. 1-20.
- 587 13. Parkkila, S., A.K. Parkkila, H. Rajaniemi, G.N. Shah, J.H. Grubb, A. Waheed, et al., *Expression*
588 *of membrane-associated carbonic anhydrase XIV on neurons and axons in mouse and human*
589 *brain*. *Proc Natl Acad Sci U S A*, 2001. **98**(4): p. 1918-23.
- 590 14. Chen, J.C. and M. Chesler, *pH transients evoked by excitatory synaptic transmission are*
591 *increased by inhibition of extracellular carbonic anhydrase*. *Proc Natl Acad Sci U S A*, 1992.
592 **89**(16): p. 7786-90.
- 593 15. Shah, G.N., B. Ulmasov, A. Waheed, T. Becker, S. Makani, N. Svichar, et al., *Carbonic*
594 *anhydrase IV and XIV knockout mice: roles of the respective carbonic anhydrases in buffering*
595 *the extracellular space in brain*. *Proc Natl Acad Sci U S A*, 2005. **102**(46): p. 16771-6.
- 596 16. Provensi, G., F. Carta, A. Nocentini, C.T. Supuran, F. Casamenti, M.B. Passani, et al., *A New*
597 *Kid on the Block? Carbonic Anhydrases as Possible New Targets in Alzheimer's Disease*. *Int J*
598 *Mol Sci*, 2019. **20**(19).
- 599 17. Lemon, N., E. Canepa, M.A. Ilies, and S. Fossati, *Carbonic Anhydrases as Potential Targets*
600 *Against Neurovascular Unit Dysfunction in Alzheimer's Disease and Stroke*. *Front Aging*
601 *Neurosci*, 2021. **13**: p. 772278.
- 602 18. Sorg, B.A., S. Berretta, J.M. Blacktop, J.W. Fawcett, H. Kitagawa, J.C. Kwok, et al., *Casting a*
603 *Wide Net: Role of Perineuronal Nets in Neural Plasticity*. *J Neurosci*, 2016. **36**(45): p. 11459-
604 11468.

- 605 19. Brakebusch, C., C.I. Seidenbecher, F. Asztely, U. Rauch, H. Matthies, H. Meyer, et al.,
606 *Brevican-deficient mice display impaired hippocampal CA1 long-term potentiation but show*
607 *no obvious deficits in learning and memory.* Mol Cell Biol, 2002. **22**(21): p. 7417-27.
- 608 20. Zhou, X.H., C. Brakebusch, H. Matthies, T. Oohashi, E. Hirsch, M. Moser, et al., *Neurocan is*
609 *dispensable for brain development.* Mol Cell Biol, 2001. **21**(17): p. 5970-8.
- 610 21. Gottschling, C., D. Wegrzyn, B. Denecke, and A. Faissner, *Elimination of the four extracellular*
611 *matrix molecules tenascin-C, tenascin-R, brevican and neurocan alters the ratio of excitatory*
612 *and inhibitory synapses.* Sci Rep, 2019. **9**(1): p. 13939.
- 613 22. Marioni, R.E., S.E. Harris, Q. Zhang, A.F. McRae, S.P. Hagenaars, W.D. Hill, et al., *GWAS on*
614 *family history of Alzheimer's disease.* Transl Psychiatry, 2018. **8**(1): p. 99.
- 615 23. Peschl, P., M. Bradl, R. Hoftberger, T. Berger, and M. Reindl, *Myelin Oligodendrocyte*
616 *Glycoprotein: Deciphering a Target in Inflammatory Demyelinating Diseases.* Front Immunol,
617 2017. **8**: p. 529.
- 618 24. Enyindah-Asonye, G., Y. Li, J.H. Ruth, D.S. Spassov, K.E. Hebron, A. Zijlstra, et al., *CD318 is a*
619 *ligand for CD6.* Proc Natl Acad Sci U S A, 2017. **114**(33): p. E6912-E6921.
- 620 25. Dogra, S., S. Uprety, and S.H. Suresh, *Itolizumab, a novel anti-CD6 monoclonal antibody: a*
621 *safe and efficacious biologic agent for management of psoriasis.* Expert Opin Biol Ther, 2017.
622 **17**(3): p. 395-402.
- 623 26. Levey, A.I., D. Qiu, L. Zhao, W.T. Hu, D.M. Duong, L. Higginbotham, et al., *A phase II study*
624 *repurposing atomoxetine for neuroprotection in mild cognitive impairment.* Brain, 2022.
625 **145**(6): p. 1924-1938.
- 626 27. Lawlor, D.A., R.M. Harbord, J.A. Sterne, N. Timpson, and G. Davey Smith, *Mendelian*
627 *randomization: using genes as instruments for making causal inferences in epidemiology.*
628 *Stat Med*, 2008. **27**(8): p. 1133-63.
- 629 28. Slob, E.A.W. and S. Burgess, *A comparison of robust Mendelian randomization methods*
630 *using summary data.* Genet Epidemiol, 2020. **44**(4): p. 313-329.

- 631 29. Swerdlow, D.I., K.B. Kuchenbaecker, S. Shah, R. Sofat, M.V. Holmes, J. White, et al., *Selecting*
632 *instruments for Mendelian randomization in the wake of genome-wide association studies.*
633 *Int J Epidemiol*, 2016. **45**(5): p. 1600-1616.
- 634 30. Omenn, G.S., L. Lane, C.M. Overall, F.J. Corrales, J.M. Schwenk, Y.K. Paik, et al., *Progress on*
635 *Identifying and Characterizing the Human Proteome: 2018 Metrics from the HUPO Human*
636 *Proteome Project.* *J Proteome Res*, 2018. **17**(12): p. 4031-4041.
- 637 31. Zheng, J., V. Haberland, D. Baird, V. Walker, P.C. Haycock, M.R. Hurle, et al., *Phenome-wide*
638 *Mendelian randomization mapping the influence of the plasma proteome on complex*
639 *diseases.* *Nat Genet*, 2020. **52**(10): p. 1122-1131.
- 640 32. Yang, C., F.H.G. Farias, L. Ibanez, A. Suhy, B. Sadler, M.V. Fernandez, et al., *Genomic atlas of*
641 *the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological*
642 *disorders.* *Nat Neurosci*, 2021. **24**(9): p. 1302-1312.
- 643 33. Narula, S., S. Yusuf, M. Chong, C. Ramasundarahettige, S. Rangarajan, S.I. Bangdiwala, et al.,
644 *Plasma ACE2 and risk of death or cardiometabolic diseases: a case-cohort analysis.* *Lancet*,
645 2020. **396**(10256): p. 968-976.
- 646 34. Habota, T., A.L. Sandu, G.D. Waiter, C.J. McNeil, J.D. Steele, J.A. Macfarlane, et al., *Cohort*
647 *profile for the STRatifying Resilience and Depression Longitudinally (STRADL) study: A*
648 *depression-focused investigation of Generation Scotland, using detailed clinical, cognitive,*
649 *and neuroimaging assessments.* *Wellcome Open Res*, 2021. **4**: p. 185.
- 650 35. Navrady, L.B., M.K. Wolters, D.J. MacIntyre, T.K. Clarke, A.I. Campbell, A.D. Murray, et al.,
651 *Cohort Profile: Stratifying Resilience and Depression Longitudinally (STRADL): a questionnaire*
652 *follow-up of Generation Scotland: Scottish Family Health Study (GS:SFHS).* *Int J Epidemiol*,
653 2018. **47**(1): p. 13-14g.
- 654 36. Smith, B.H., A. Campbell, P. Linksted, B. Fitzpatrick, C. Jackson, S.M. Kerr, et al., *Cohort*
655 *Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its*

- 656 *participants and their potential for genetic research on health and illness.* Int J Epidemiol,
657 2013. **42**(3): p. 689-700.
- 658 37. Smith, B.H., H. Campbell, D. Blackwood, J. Connell, M. Connor, I.J. Deary, et al., *Generation*
659 *Scotland: the Scottish Family Health Study; a new resource for researching genes and*
660 *heritability.* BMC Med Genet, 2006. **7**: p. 74.
- 661 38. Wechsler, D., *Wechsler Adult Intelligence Scale-Third Edition (WAIS-III).* 1997, San Antonio:
662 Harcourt Assessment Inc.
- 663 39. Jaeger, J., *Digit Symbol Substitution Test: The Case for Sensitivity Over Specificity in*
664 *Neuropsychological Testing.* J Clin Psychopharmacol, 2018. **38**(5): p. 513-519.
- 665 40. Shi, L., N.J. Buckley, I. Bos, S. Engelborghs, K. Sleegers, G.B. Frisoni, et al., *Plasma Proteomic*
666 *Biomarkers Relating to Alzheimer's Disease: A Meta-Analysis Based on Our Own Studies.*
667 Front Aging Neurosci, 2021. **13**: p. 712545.
- 668 41. Schwarz, C.G., J.L. Gunter, H.J. Wiste, S.A. Przybelski, S.D. Weigand, C.P. Ward, et al., *A large-*
669 *scale comparison of cortical thickness and volume methods for measuring Alzheimer's*
670 *disease severity.* Neuroimage Clin, 2016. **11**: p. 802-812.
- 671 42. Therneau, T.M., *coxme: mixed effects Cox models.* 2012. p. R Package.
- 672 43. Tingley, D., T. Yamamoto, K. Hirose, L. Keele, and K. Imai, *mediation: R Package for Causal*
673 *Mediation Analysis.* Journal of Statistical Software, 2014. **59**(5).
- 674 44. Liao, Y., J. Wang, E.J. Jaehnig, Z. Shi, and B. Zhang, *WebGestalt 2019: gene set analysis toolkit*
675 *with revamped UIs and APIs.* Nucleic Acids Res, 2019. **47**(W1): p. W199-W205.
- 676 45. Watanabe, K., E. Taskesen, A. van Bochoven, and D. Posthuma, *Functional mapping and*
677 *annotation of genetic associations with FUMA.* Nat Commun, 2017. **8**(1): p. 1826.
- 678 46. Savage, J.E., P.R. Jansen, S. Stringer, K. Watanabe, J. Bryois, C.A. de Leeuw, et al., *Genome-*
679 *wide association meta-analysis in 269,867 individuals identifies new genetic and functional*
680 *links to intelligence.* Nat Genet, 2018. **50**(7): p. 912-919.

- 681 47. Smith, S.M., G. Douaud, W. Chen, T. Hanayik, F. Alfaro-Almagro, K. Sharp, et al., *An expanded*
682 *set of genome-wide association studies of brain imaging phenotypes in UK Biobank*. *Nat*
683 *Neurosci*, 2021. **24**(5): p. 737-745.
- 684 48. Hibar, D.P., H.H.H. Adams, N. Jahanshad, G. Chauhan, J.L. Stein, E. Hofer, et al., *Novel genetic*
685 *loci associated with hippocampal volume*. *Nat Commun*, 2017. **8**: p. 13624.
- 686 49. Persyn, E., K.B. Hanscombe, J.M.M. Howson, C.M. Lewis, M. Traylor, and H.S. Markus,
687 *Genome-wide association study of MRI markers of cerebral small vessel disease in 42,310*
688 *participants*. *Nat Commun*, 2020. **11**(1): p. 2175.
- 689 50. Knol, M.J., D. Lu, M. Traylor, H.H.H. Adams, J.R.J. Romero, A.V. Smith, et al., *Association of*
690 *common genetic variants with brain microbleeds: A genome-wide association study*.
691 *Neurology*, 2020. **95**(24): p. e3331-e3343.
- 692 51. Bellenguez, C., F. Kucukali, I.E. Jansen, L. Kleindam, S. Moreno-Grau, N. Amin, et al., *New*
693 *insights into the genetic etiology of Alzheimer's disease and related dementias*. *Nat Genet*,
694 2022. **54**(4): p. 412-436.
- 695 52. Mishra, A., R. Malik, T. Hachiya, T. Jurgenson, S. Namba, D.C. Posner, et al., *Stroke genetics*
696 *informs drug discovery and risk prediction across ancestries*. *Nature*, 2022. **611**(7934): p.
697 115-123.
- 698 53. Bakker, M.K., R.A.A. van der Spek, W. van Rheenen, S. Morel, R. Bourcier, I.C. Hostettler, et
699 al., *Genome-wide association study of intracranial aneurysms identifies 17 risk loci and*
700 *genetic overlap with clinical risk factors*. *Nat Genet*, 2020. **52**(12): p. 1303-1313.
- 701 54. Hemani, G., J. Zheng, B. Elsworth, K.H. Wade, V. Haberland, D. Baird, et al., *The MR-Base*
702 *platform supports systematic causal inference across the human phenome*. *Elife*, 2018. **7**.
- 703 55. Zhao, Q., J. Wang, G. Hemani, J. Bowden, and D.S. Small, *Statistical inference in two-sample*
704 *summary-data Mendelian randomization using robust adjusted profile score*. *The Annals of*
705 *Statistics*, 2020. **48**(3): p. 1742-1769, 28.

- 706 56. Verbanck, M., C.Y. Chen, B. Neale, and R. Do, *Detection of widespread horizontal pleiotropy*
707 *in causal relationships inferred from Mendelian randomization between complex traits and*
708 *diseases*. Nat Genet, 2018. **50**(5): p. 693-698.
- 709 57. Burgess, S., A. Butterworth, and S.G. Thompson, *Mendelian randomization analysis with*
710 *multiple genetic variants using summarized data*. Genet Epidemiol, 2013. **37**(7): p. 658-65.
- 711 58. Bowden, J., G. Davey Smith, P.C. Haycock, and S. Burgess, *Consistent Estimation in*
712 *Mendelian Randomization with Some Invalid Instruments Using a Weighted Median*
713 *Estimator*. Genet Epidemiol, 2016. **40**(4): p. 304-14.
- 714 59. Burgess, S. and S.G. Thompson, *Interpreting findings from Mendelian randomization using*
715 *the MR-Egger method*. Eur J Epidemiol, 2017. **32**(5): p. 377-389.
- 716 60. Robinson, J.W., G. Hemani, M.S. Babaei, Y. Huang, D.A. Baird, E.A. Tsai, et al., *An efficient*
717 *and robust tool for colocalisation: Pair-wise Conditional and Colocalisation (PWCoCo)*.
718 bioRxiv, 2022: p. 2022.08.08.503158.
- 719

720 **Figure legends**

721 **Figure 1.** Overview of the study design. This study involved European (N = 3514), Latin (N = 4309),
722 and Persian (N = 1332) PURE participants for whom genetic and plasma proteomic data were
723 available. Observational analyses to detect plasma biomarkers of cognitive function were performed
724 in the subset of these participants who were enrolled in the PURE-MIND sub-study (N = 1198), for
725 whom plasma protein (N = 1060 proteins) and MRI measurements were available. Mediation
726 analyses were performed to assess whether any observed associations between protein levels and
727 cognitive function were mediated by structural brain phenotypes ascertained by MRI. Finally, two-
728 sample Mendelian randomisation analyses were performed to assess potentially causal effects of
729 genetically-predicted cognition-associated protein levels on genetically-predicted neurological
730 outcomes. For these analyses, genetic instrumental variables for protein levels were identified in the
731 European, Latin, and Persian PURE participants, and associations with neurological outcomes were
732 assessed using external (non-PURE) datasets.

733 **Figure 2.** Manhattan plot indicating associations between the levels of plasma proteins and
734 performance on the DSST in participants from the PURE-MIND cohort (N = 1198). Each protein is
735 represented by a triangle with upwards-facing triangles indicating a positive association with DSST
736 performance and downwards-facing triangles indicating a negative association with DSST
737 performance. The position of each protein on the x-axis is determined by the genomic location of its
738 corresponding gene and the position on the y-axis is determined by the $-\log_{10} p$ -value. The dashed
739 horizontal line indicates the Bonferroni-corrected significance threshold ($p = 4.31 \times 10^{-5}$) required to
740 maintain a 5% type I error rate.

741 **Figure 3.** Forest plot indicating the association between protein levels and DSST performance for
742 significantly associated proteins. For each protein, the difference in DSST score associated with a
743 standard deviation higher level of protein is shown, together with the 95% confidence interval.
744 Abbreviations: BCAN: brevican; CA14: carbonic anhydrase 14; CDCP1: CUB-domain containing

745 protein 1; CI: confidence interval; GS: Generation Scotland imaging subsample; MOG: myelin
746 oligodendrocyte glycoprotein; NCAN: neurocan; PURE: Prospective Urban and Rural Epidemiology
747 study.

748 **Figure 4.** Forest plots indicating the association between the levels of DSST-associated proteins and
749 DSST-associated structural brain phenotype. For each protein, the effect estimate (change in brain
750 volume (cm³ or %) per standard deviation increase in protein expression) is shown, together with the
751 95% confidence interval. Abbreviations: BCAN: brevican; CA14: carbonic anhydrase 14; CDCP1: CUB-
752 domain containing protein 1; CI: confidence interval; GS: Generation Scotland imaging subsample;
753 MOG: myelin oligodendrocyte glycoprotein; NCAN: neurocan; PURE: Prospective Urban and Rural
754 Epidemiology study; WMH: white matter hyperintensity.

755 **Tables**

756 **Table 1.** Demographic information for the discovery sample (PURE-MIND) and the replication sample (GS). Basic sample demographic information is
 757 presented together with information on relevant clinical characteristics, cognitive performance, and structural brain imaging measures. The number of
 758 participants for whom information was available for each variable are indicated in the “N” columns. In GS, there were participants with missing information
 759 for smoking (N = 31), hypertension (N = 2), diabetes (N = 1), stroke (N = 1), and depression (N = 2). *PURE-MIND: college or university; GS: university. For
 760 variables that were measured in both cohorts, a statistical comparison was performed and significant *p*-values (*p* < 0.05) are indicated in bold.

	PURE-MIND		GS		Test statistic (degrees of freedom)	<i>p</i> -value
	N (%)	Mean (SD)	N (%)	Mean (SD)		
Age (years)	1198	54.5 (8.05)	1053	59.9 (9.59)	<i>t</i> = 14.5 (df = 2249)	< 2.2 x 10⁻¹⁶
Sex						
Female	709 (59.2%)		627 (59.5%)		$\chi^2 = 0.0173$	0.895
Male	489 (40.8%)		426 (40.5%)			
Education (highest level achieved)						
High school or less	346 (28.9%)		316 (29.9%)		Fisher's exact test	5.0 x 10⁻⁴
Trade school	83 (6.93%)		328 (31.1%)			
College/university*	768 (64.1%)		295 (28.0%)			

Other	NA		21 (1.99%)			
Unknown	1 (0.0835%)		94 (8.93%)			
Clinical characteristics						
Current/former smoker	583 (48.7%)		466 (45.6%)		$\chi^2 = 1.96$	0.161
Non-smoker	615 (51.3%)		556 (54.4%)			
Body mass index (kg/m ²)	1193	27.1 (5.15)	1053	28.2 (5.72)	$t = 5.06$ (df = 2244)	4.92×10^{-7}
Waist hip ratio	1192	0.870 (0.0965)			N/A	
Systolic blood pressure	1197	130 (18.8)	1051	141 (18.9)	$t = -13.8$ (df = 2246)	$< 2.2 \times 10^{-16}$
Diastolic blood pressure	1197	80.9 (11.3)	1051	82.4 (10.8)	$t = -3.21$ (df = 2246)	1.37×10^{-3}
Hypertension	270 (22.5%)				N/A	
Total cholesterol (mmol/L)	1197	5.52 (1.11)			N/A	
High density lipoprotein cholesterol (mmol/L)	1197	1.47 (0.399)			N/A	
Low density lipoprotein cholesterol (mmol/L)	1187	3.35 (0.882)			N/A	
Diabetes	61 (5.09%)		73 (6.93%)		$\chi^2 = 3.09$ (df = 1)	0.0787
Stroke	3 (0.250%)		27 (2.56%)		Fisher's exact test	8.53×10^{-7}
Cardiovascular disease	36 (3.00%)		103 (9.78%)		$\chi^2 = 43.3$ (df = 1)	4.80×10^{-11}

Depression	NA		336 (31.9%)		N/A	
Cognitive tests						
Digit Symbol Substitution Test (no. correct in two minutes)	1198	69.5 (15.3)	1053	68.1 (15.2)	$t = -2.18$ (df = 2249)	0.0298
Montreal Cognitive Assessment (no. items correct)	1198	26.5 (2.39)			N/A	
Brain Imaging Measures						
Cortical thickness (mm)	1198	2.35 (0.0891)			N/A	
Total brain volume without ventricles (cm ³)	1198	1058 (110)	943	1069 (109)	$t = -2.31$ (df = 2139)	0.0212
Total cerebral white matter volume (cm ³)	1198	447 (60.5)	939	455 (56.9)	$t = -3.11$ (df = 2135)	1.87×10^{-3}
Total hippocampal volume (cm ³)	1198	3.92 (0.433)	941	4.18 (0.437)	$t = -13.7$ (df = 2137)	$< 2.2 \times 10^{-16}$
WMH volume (log-transformed cm ³)	1198	0.751 (0.693)	934	0.777 (0.789)	$t = -0.807$ (df = 2130)	0.419
Estimated intracranial volume (cm ³)	1198	1491 (159)	944	1400 (225)	$t = 11.0$ (df = 2140)	$< 2.2 \times 10^{-16}$
Silent brain infarct	95 (7.93%)				N/A	
Cerebral microbleeds	86 (7.18%)				N/A	

762 Abbreviations: GS: Generation Scotland; PURE: Prospective Urban and Rural Epidemiology study; SD: standard deviation; WMH: white matter
763 hyperintensities

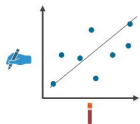
1 Study cohort



PURE participants with:

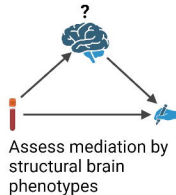


2 Observational association analyses

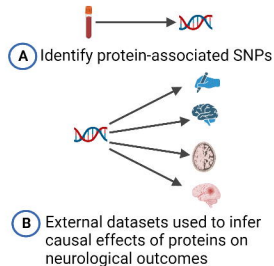


- A Identify cognition-associated proteins
- B Replication analyses in independent STRADL cohort

3 Mediation analyses

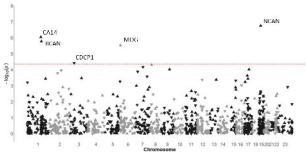


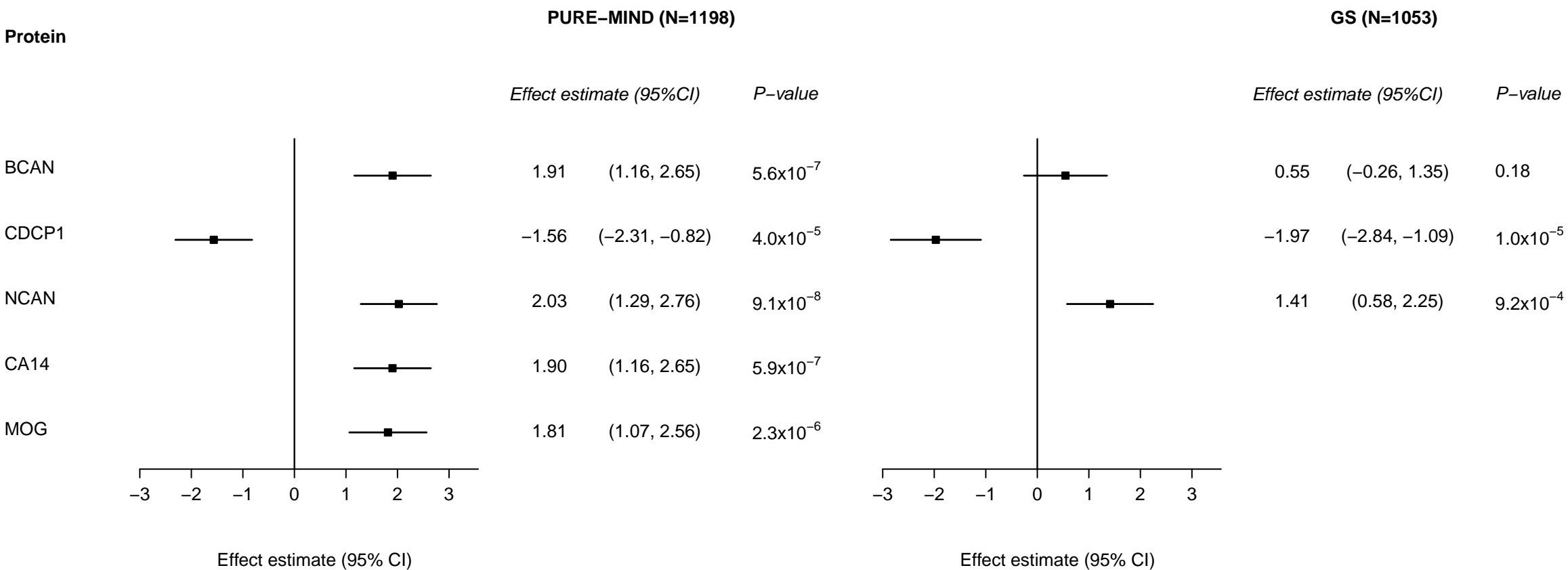
4 Mendelian randomisation



KEY

Alzheimer's disease		MRI brain imaging	
Cognitive tests		Plasma proteomics (x 1060)	
Genotyping		Stroke	





Phenotype

PURE-MIND (N = 1168–1198)

GS (N = 932–939)

