

1 Serum proteomics reveals *APOE* dependent and independent protein signatures in Alzheimer's disease

2 Elisabet A. Frick¹, Valur Emilsson^{1,2}, Thorarinn Jonmundsson², Anna E. Steindorsdottir², Erik C. B.
3 Johnson^{3,4}, Raquel Puerta⁶, Eric B. Dammer^{3,5}, Anantharaman Shantaraman^{3,5}, Amanda Cano^{6,7}, Mercè
4 Boada^{6,7}, Sergi Valero^{6,7}, Pablo García-González^{6,7}, Elias F. Gudmundsson¹, Alexander Gudjonsson¹,
5 Joseph J. Loureiro⁸, Anthony P. Orth⁹, Nicholas T. Seyfried^{3,4}, Allan I. Levey^{3,4}, Agustin Ruiz^{6,7}, Thor
6 Aspelund^{1,2}, Lori L. Jennings⁸, Lenore J. Launer¹⁰, Valborg Gudmundsdottir^{*1,2}, Vilmundur Gudnason^{*1,2}

7 1 Icelandic Heart Association, Kopavogur, 200, Iceland

8 2 Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland

9 3 Goizueta Alzheimer's Disease Research Center, Emory University School of Medicine, Atlanta, 30329, GA, USA

10 4 Department of Neurology, Emory University School of Medicine, Atlanta, 30329, GA, USA

11 5 Department of Biochemistry, Emory University School of Medicine, Atlanta, 30329, GA, USA

12 6 Research Center and Memory Clinic. Ace Alzheimer Center Barcelona – Universitat Internacional de Catalunya, Barcelona,
13 08028, Spain, Barcelona.

14 7 CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National Institute of Health Carlos III,
15 Madrid, 28029, Spain.

16 8 Novartis Biomedical Research, Cambridge, 02139, MA, USA

17 9 Novartis Biomedical Research, San Diego, 10675, CA, USA

18 10 Laboratory of Epidemiology and Population Sciences, Intramural Research Program, National Institute
19 on Aging, Bethesda, 20892, MD, USA

20

21 Correspondence: valborgg@hi.is, v.gudnason@hiarta.is

22 Summary

23 The current demand for early intervention, prevention, and treatment of late onset Alzheimer's disease
24 (LOAD) warrants deeper understanding of the underlying molecular processes which could contribute to
25 biomarker and drug target discovery. Utilizing high-throughput proteomic measurements in serum from
26 a prospective population-based cohort of older adults (n=5,294), we identified 303 unique proteins
27 associated with incident LOAD (median follow-up 12.8 years). Over 40% of these proteins were
28 associated with LOAD *independently* of *APOE-ε4* carrier status. These proteins were implicated in
29 neuronal processes and overlapped with protein signatures of LOAD in brain and cerebrospinal fluid. We
30 found 17 proteins which LOAD-association was strongly *dependent* on *APOE-ε4* carrier status. Most of
31 them showed consistent associations with LOAD in cerebrospinal fluid and a third had brain-specific gene
32 expression. Remarkably, four proteins in this group (TBCA, ARL2, S100A13 and IRF6) were
33 downregulated by *APOE-ε4* yet upregulated as a consequence of LOAD as determined in a bi-directional
34 Mendelian randomization analysis, reflecting a potential response to the disease onset. Accordingly, the
35 direct association of these proteins to LOAD was reversed upon *APOE-ε4* genotype adjustment, a finding
36 which we replicate in an external cohort (n=719). Our findings provide an insight into the dysregulated

37 pathways that may lead to the development and early detection of LOAD, including those both
38 independent and dependent on *APOE-ε4*. Importantly, many of the LOAD-associated proteins we find in
39 the circulation have been found to be expressed - and have a direct link with AD - in brain tissue. Thus,
40 the proteins identified here, and their upstream modulating pathways, provide a new source of
41 circulating biomarker and therapeutic target candidates for LOAD.

42 **Keywords:** Alzheimer's disease, APOE, Proteomics, Mendelian randomization, Longitudinal study, Cross-
43 sectional study, CSF, Brain, Network.

44 **Introduction**

45 Alzheimer's disease (AD) is the most common cause of dementia, accounting for up to 80% of all
46 dementia cases¹, of which *late onset* Alzheimer's disease (LOAD) is most common². As of 2022,
47 approximately 55 million individuals worldwide had dementia, representing 1 out of 9 people aged 65
48 and over³. While promising advances have been made in amyloid-targeting therapeutic options for early-
49 stage LOAD^{4,5}, they still have limited benefit and identification of additional risk pathways that can be
50 used for early detection and intervention is highly needed. To meet these demands, a variety of
51 biologically relevant circulating molecules have been broadly associated with LOAD risk. The proteome in
52 particular has the potential to reveal circulating markers of disease-related molecular pathways from
53 different tissues, and studies assessing the circulating proteomic signatures between non-demented
54 older adults and individuals suffering from LOAD have been described⁶⁻¹⁷. Modest sample sizes, low-
55 throughput proteomics and lack of longitudinal measurements have, however, been limiting factors in
56 these studies. However, a recent large-scale longitudinal study identified promising blood-based markers
57 for all-cause incident dementia although it is unknown how specific the results are to LOAD¹⁸.
58 Information on the global circulating proteomic profile preceding the onset of LOAD, and how well it
59 reflects AD-related processes in brain and CSF, is thus scarce.

60 Alzheimer's disease has a considerable genetic component, and both common¹⁹ and rare risk
61 variants have been identified²⁰, of which the strongest effects are conferred by variants in the well-
62 known apolipoprotein E (*APOE*) gene. Approximately 25% of the general population carries the *APOE-ε4*
63 variant while it is present in over 50% of AD cases^{21,22}. The *APOE-ε4* allele increases the risk of LOAD by
64 threefold in heterozygous carriers and up to twelvefold in homozygous carriers²³. Although the link
65 between the $\epsilon4$ allele and LOAD has been extensively researched, light has yet to be shed on the precise
66 mechanism by which the *APOE* gene affects LOAD onset and/or progression. Importantly, recent large-

67 scale proteogenomic studies have consistently established the *APOE* locus as a protein-regulatory
68 hotspot, regulating levels of hundreds of proteins in both circulation^{24–27} and cerebrospinal fluid
69 (CSF)^{28,29}. Yet, it remains unknown to what extent these proteins relate to LOAD and if they can provide
70 new information on the mechanisms through which *APOE*- ϵ 4 mediates its risk. Identifying LOAD-
71 associated circulatory proteins and whether their association is *APOE*-dependent or independent is
72 crucial for the understanding of AD more generally as well as for gaining insight into potential pathways
73 suitable for targeting in personalized treatment.

74 The current study tests the hypotheses that specific proteomic signatures in the circulation
75 precede LOAD diagnosis and can reflect dysregulated biological pathways in the brain and CSF.
76 Furthermore, we expect that some of these protein signatures may be affected by the *APOE*- ϵ 4 genotype
77 and can thus provide molecular read-out of pathways directly affected by *APOE*- ϵ 4. To address these
78 hypotheses, we used a high throughput aptamer-based platform to characterize 4,137 serum proteins in
79 5,294 participants of the population-based Age, Gene/Environment Susceptibility-Reykjavik Study (AGES-
80 RS)³⁰ to identify protein signatures of incident LOAD (events occurring during follow-up) and prevalent
81 LOAD, taking an unbiased, longitudinal, and cross-sectional approach to the discovery of potential
82 biomarkers for LOAD (Figure 1). Considering *APOE*'s protein-regulatory influence and how it may impact
83 the way that serum-proteins are associated with LOAD, we disentangled the LOAD protein signature into
84 *APOE*- ϵ 4 dependent and independent components, by identifying proteins whose LOAD-association is
85 largely attenuated upon conditioning on *APOE*- ϵ 4 carrier status. We compared the serum protein
86 signature of LOAD to those observed in cerebrospinal fluid (CSF) and brain and finally, used genetic
87 variation as anchors to determine the potential causal direction between serum proteins and disease
88 state.

89

90 Results

91 **The AGES study cohort.** This prospective population-based study was based on 5,127 participants free of
92 dementia at baseline, after the exclusion of 163 individuals with prevalent non-AD dementia, and 167
93 individuals with prevalent LOAD. During a potential follow-up of 12.8 years (median), 655 individuals
94 were diagnosed with incident LOAD, with the last individual being diagnosed 16 years from baseline.
95 Participants with incident LOAD were older at entry, were more likely to carry an *APOE-ε4* allele, had
96 lower BMI, and had lower education levels compared to healthy individuals (Supplementary Table 1). See
97 Figure 1 for study overview.

98 **Serum protein profile of incident LOAD in AGES.** To investigate the LOAD-associated circulatory
99 proteomic patterns which occur prior to disease onset, we used Cox proportional hazards models and
100 found 320 aptamers (303 proteins) to be significantly (FDR < 0.05) associated with incident LOAD
101 diagnosis after adjusting for age and sex (model 1), with hazard ratios (HRs) ranging from 0.78 for TBCA
102 to 1.47 for NTN1 (Figure 2D, Supplementary Table 2). To account for variability related to *APOE-ε4*
103 carrier status, we adjusted for the genotype in an additional model (model 2, Supplementary Table 2),
104 which resulted in 140 significant aptamers (130 proteins, HR 0.79 (CD4) – 1.25 (CGA/FSHB), FDR < 0.05)
105 (Figure 2E), all of which overlapped with model 1 (Figure 2F). When comparing the two models, 43% of
106 the serum proteins remained significant after *APOE-ε4* adjustment, indicating that their LOAD
107 association is independent of the *APOE-ε4* genotype (Table 1). Adjusting for additional AD risk factors
108 and eGFR (see Methods) retained 38 significant LOAD-associated aptamers (35 proteins, HR 0.80 (CD4) –
109 1.26 (SMOC1), FDR < 0.05) (model 3, Supplementary Table 2), which may reflect specific processes
110 affecting risk of LOAD but not driven by currently established risk factors.

111 As hazard ratio variability can arise with lengthy follow-up time, secondary analyses were
112 implemented with a 10-year follow-up cut-off, which revealed mostly overlapping results
113 (Supplementary Note 1, Supplementary Tables 3 and 4, Supplementary Figure 1). We did however detect
114 protein associations specific to the shorter follow-up time, which potentially reflect processes that take
115 place closer to the LOAD diagnosis. As there may be further differences in proteomic profiles depending
116 on whether protein sampling occurred before or after LOAD diagnosis, we additionally considered the
117 protein profile of the 167 individuals with prevalent LOAD at baseline (Supplementary Note 2,
118 Supplementary Figure 2A-C, Supplementary Tables 5-7). Interestingly, many of the proteins associated
119 with increased risk of incident LOAD showed the opposite direction of effect for prevalent LOAD,
120 although generally not statistically significant (Supplementary Figure 2D). These contrasting results

121 suggest an important temporal element in the LOAD-associated proteome. In total there were 346
122 aptamers (329 unique proteins) associated with LOAD when all outcomes (incident and prevalent LOAD),
123 follow-up times and models were considered (Supplementary Tables 2, 3, 5, 6, and 7).

124 To evaluate which biological processes are reflected by the overall incident LOAD-associated
125 protein signature in AGES, we performed a gene set enrichment analysis (GSEA). The strongest
126 enrichment for protein associations in model 1 was observed for gene ontology (GO) terms related to
127 neuron development and morphogenesis (Figure 2H-I, Supplementary Table 8). The proteins driving the
128 enrichment included Neural Cell Adhesion Molecules 1 and 2 (NCAM1, and NCAM2), Netrin 1 (NTN1),
129 Contactin 1 (CNTN1), Neuropilin 1 (NRP1), Fibronectin Leucine Rich Transmembrane Protein 2 (FLRT2),
130 Matrix Metalloproteinase 2 (MMP2) and Cell Adhesion Molecule L1 like (CHL1). GSEA of the protein
131 profiles of model 2, where *APOE-ε4* carrier status was adjusted for, showed similar enrichment results
132 (Supplementary Table 8), demonstrating that these terms were mainly driven by the *APOE-ε4*
133 independent component of the LOAD-associated protein profile.

134 ***Serum proteins with APOE-dependent association to incident LOAD.*** As previously mentioned, 43% of
135 the protein associations with incident LOAD were independent of *APOE-ε4*. Of the remaining 57% that
136 were affected by *APOE-ε4* adjustment, we identified 17 proteins whose associations with incident LOAD
137 were particularly strongly affected by *APOE-ε4* carrier status (Table 2, Figure 3A, Supplementary Figure 3,
138 Supplementary Table 2). These proteins, hereafter referred to as *APOE*-dependent proteins, were
139 defined as proteins significantly ($FDR < 0.05$) associated with incident LOAD in *model 1* but whose
140 nominal significance was attenuated ($P > 0.05$) or direction of effect changed upon *APOE-ε4* adjustment
141 in *model 2*. These *APOE*-dependent proteins included those with the strongest associations to LOAD
142 prior to adjusting for the *APOE-ε4* allele (Figure 2A, D). The levels of the *APOE* protein itself were not
143 associated with either incident or prevalent LOAD. Figure 3B shows the intra-correlations among the 17
144 *APOE*-dependent proteins. All the 17 *APOE*-dependent proteins were strongly regulated by the *APOE-ε4*
145 allele (Figure 3C, Table 2, Supplementary Figure 4, Supplementary Table 9), with the $\epsilon 4$ allele increasing
146 the levels of five of the proteins and decreasing the levels of the other 12. Accordingly, we observed that
147 *increased* levels of the five *APOE-ε4* upregulated proteins and *decreased* levels of the 12 *APOE-ε4*
148 downregulated proteins were also associated with higher risk of LOAD, yielding a hazard ratio above and
149 below one, respectively (Figure 3D). As per definition, most of the *APOE*-dependent proteins lost
150 significance upon *APOE-ε4* adjustment yet interestingly, the direction of effect inverted for five proteins
151 after *APOE-ε4* adjustment (ARL2, IRF6, NEFL, S100A13, TBCA) (Figure 3E).

152 The HR conferred by *APOE-ε4* for incident LOAD in AGES was 2.1 ($P = 1.23e-27$) per each $\epsilon 4$
153 allele. To evaluate if any of the 17 *APOE*-dependent proteins might mediate the effect of *APOE-ε4* on
154 incident LOAD, we considered the change in HR for *APOE-ε4* on risk of incident LOAD when adjusting for
155 individual proteins. We found that adjustment for most proteins resulted in a minor effect decrease,
156 suggesting they do not mediate the *APOE-ε4* effect on LOAD (Supplementary Figure 5A). Intriguingly,
157 however, the adjustment for four proteins (NEFL, ARL2, TBCA and S100A13) caused an increase of ~10%
158 in *APOE-ε4* effect size (Supplementary Figure 5A-B). Thus, the effect of *APOE-ε4* on LOAD is partly
159 masked by secondary opposing associations between these proteins and LOAD, which are further
160 explored below. The effect of *APOE-ε4* carrier status on LOAD risk was largely unchanged in a
161 multivariable model containing all 17 *APOE*-dependent proteins, thus not supporting a mediating effect
162 of these proteins for the LOAD risk conferred by *APOE-ε4* (base model: HR = 2.1, $P = 1.23e-27$,
163 multivariable model: HR = 2.2, $P = 3.2e-10$). However, although not direct mediators, the 17 proteins
164 could be blood-based readouts of a true mediator within tissue-specific pathological processes occurring
165 prior to LOAD diagnosis.

166 To map out potential tissues of origin for the circulating levels of the 17 *APOE*-dependent
167 proteins, we considered gene expression data from the Human Protein Atlas³¹. We observed that five
168 (LRRN1, TMCC3, FAM159B, NEFL, GSTM1) of the *APOE*-dependent proteins had elevated gene
169 expression in brain compared to other tissues and additional two (IFIT2, NDE1) clustered with brain-
170 specific genes (Table 2). Of the remaining *APOE*-dependent proteins, six were universally expressed,
171 including in brain tissue, and four were enriched in other tissues. We did not detect any significantly
172 enriched molecular signatures nor GO terms for the 17 *APOE*-dependent proteins (Supplementary Table
173 7). However, a network analysis of measured and inferred physical protein-protein interactions³²
174 revealed that the *APOE*-dependent proteins interact directly with proteins involved in neuronal
175 response- and development, neuroinflammation and AD (Figure 4, Supplementary Table 10-12,
176 Supplementary Note 3).

177 Given the well-established relationship between *APOE* and cholesterol³³ we explored the
178 potential effect of serum lipid levels on the association between LOAD and the 17 *APOE*-dependent
179 proteins (Supplementary Table 13, Supplementary Figures 6-7, Supplementary Note 4). Our findings
180 suggest that, while many of the *APOE*-dependent proteins are associated with cholesterol levels, it is not
181 the driver of their link to LOAD.

182 **External validation of protein associations with incident LOAD in the ACE cohort.** We set to externally
183 evaluate our observations in an independent cohort, the ACE - Alzheimer Center Barcelona (n=1,341),
184 with SOMAscan platform (7k) measurements from plasma of individuals who were referred to the
185 center. The longitudinal component of ACE consists of individuals who had been diagnosed with mild
186 cognitive impairment (MCI) at the center and had been followed up. A total of 719 participants had
187 follow-up information and 266 converted to LOAD over a median follow-up of 3.14 years (Supplementary
188 Table 14). Despite the fundamentally different cohorts, with AGES being population-based and using the
189 5K SOMAscan platform and ACE being based on individuals with established symptoms and the 7K
190 SOMAscan platform, we replicated 36 protein associations with LOAD at nominal significance ($P < 0.05$)
191 in the smaller ACE cohort (Table 3, Figure 2F-G). Of those, 30 proteins were nominally significant in
192 model 1 with 97% being directionally consistent with the observations in AGES (Figure 2F). In model 2, 21
193 proteins were nominally significant, 86% of which were directionally consistent (Figure 2G). After
194 multiple testing correction, seven proteins remained statistically significant ($FDR < 0.05$), all of which
195 were directionally consistent (Table 3, Figure 2F-G). Six were statistically significant ($FDR < 0.05$) in model
196 1 (NEFL, LRRN1, TBCA, CTF1, C1orf56 and TIMP4) and one in model 2 (S100A13) (Supplementary Table
197 9). Of all 332 tested aptamers, 213 (64%) were directionally consistent regardless of significance in
198 model 1 (Exact binomial test $P = 2.0e-05$) and 202 (61%) were directionally consistent in model 2 (Exact
199 binomial test $P = 0.002$), demonstrating an enrichment for consistency in direction of effect. The protein
200 associations replicated in the ACE cohort are of particular interest as they represent potentially clinically
201 relevant candidates for LOAD that are consistent in two different contexts, in both a general population
202 and a clinically derived symptomatic sample set. However, our results suggest that many of the proteins
203 that associate with long-term LOAD risk are not strongly associated with the conversion from MCI to AD,
204 which is further into the AD trajectory and may also explain the limited overlap between the proteins
205 associated with prevalent and incident LOAD in AGES.

206 **External validation of reversed LOAD association conditional on APOE-ε4 for a subset of proteins.**

207 Specifically considering the APOE-dependent proteins, the association between the APOE-ε4 allele and
208 the proteins was replicated for 13 of 17 proteins in the ACE cohort (Figure 3F). Furthermore, the change
209 in direction of effect for incident LOAD upon APOE-ε4 adjustment was replicated in the ACE cohort for 4
210 of 5 proteins (ARL2, NEFL, S100A13 and TBCA) (Figure 3G-H) (Supplementary Table 9), with even larger
211 effects observed in the ACE cohort compared to AGES in the APOE-ε4 adjusted model and three proteins
212 (ARL2, S100A13 and TBCA) becoming statistically significant ($P < 0.05$). Thus, the attenuation of the
213 primary LOAD associations for these proteins upon APOE-ε4 adjustment meet the criteria of APOE-ε4

214 dependence (see Methods). No significant interaction between protein and *APOE-ε4* carrier status on AD
215 risk was observed in either the AGES or ACE cohorts. Taken together, our results show that these
216 proteins are strongly downregulated by *APOE-ε4*, and consequently show an inverse relationship with
217 incident LOAD, but when adjusting for the *APOE-ε4* allele, their association to LOAD is still significant but
218 reversed – suggesting a secondary non-*APOE-ε4*-mediated process affecting these same proteins in
219 relation to LOAD in the opposite direction that is more strongly observed a cohort of individuals with
220 MCI than in the population-based AGES cohort.

221 ***Mendelian randomization to identify potential causal associations between proteins and LOAD.*** The
222 proteins associated with LOAD could include proteins causally related to the disease, or proteins whose
223 serum level changes reflect a response to prodromal or genetic liability to LOAD. To test this hypothesis,
224 we performed a bi-directional two-sample MR analysis, including the targets of all 346 aptamers
225 associated with LOAD in our study. Genetic variant associations for serum protein levels were obtained
226 from a catalog of cis-protein quantitative trait loci (pQTLs) from AGES²⁴ while variant associations with
227 LOAD were extracted from a recent GWAS on 39,106 clinically diagnosed LOAD cases, 46,828 proxy-
228 LOAD and dementia cases and 401,577 controls of European ancestry¹⁹. In total, 117 (34%) of the LOAD-
229 associated serum aptamers had cis-pQTLs that were suitable as genetic instruments and were included
230 in the protein-LOAD MR analysis (Supplementary Table 15).

231 In the forward MR analysis, two proteins, integrin binding sialoprotein (IBSP) and amyloid
232 precursor protein (APP), had support for causality (Supplementary Table 16). IBSP had a risk-increasing
233 effect for LOAD in both the causal (OR = 1.26, FDR = 0.03) and observational analysis (incident LOAD full
234 follow up, HR = 1.13, FDR = 0.04). APP had a protective effect for LOAD in both the causal (OR = 0.76, FDR
235 = 0.03) and observational analysis (incident LOAD full follow up, HR = 0.87, FDR = 0.047). Notably, while
236 not statistically significant, we observed suggestive support for a protective effect of genetically
237 determined serum levels of acetylcholinesterase (ACHE, OR = 0.92, P = 0.061), a target of clinically used
238 therapeutic agent for dementia³⁴ (Supplementary Table 16, Supplementary Figure 8). In a forward MR
239 analysis of the *APOE*-dependent protein interaction partners, two proteins, APP and MAPK3, had support
240 for causality (Supplementary Tables 10-12, Supplementary Note 3).

241 As most of the observational protein associations in the current study were detected for incident
242 LOAD, and thus reflect changes that take place before the onset of clinically diagnosed disease, it is
243 unlikely that their levels and effects are direct downstream consequences of the disease after it reaches
244 a clinical stage. However, they may reflect a response to a prodromal stage of the disease. We therefore

245 performed a reverse MR to test if the observed changes in serum protein levels are likely to occur
246 downstream of the genetic liability to LOAD, which may capture processes both at the prodromal and
247 clinical stage. The *APOE* locus is likely to have a dominant pleiotropic effect in the reverse MR analysis
248 (Supplementary Table 17, Supplementary Figure 9, Supplementary Note 5), as it has a disproportionately
249 strong effect on LOAD risk compared to all other common genetic variants, while also being a well-
250 established pQTL trans-hotspot, affecting circulating levels of up to hundreds of proteins^{24,25,27}. We
251 therefore performed the primary reverse MR analysis using only LOAD-associated genetic variants
252 outside of the *APOE* locus as instruments. We found two proteins (S100A13 and ARL2) that were
253 significantly (FDR < 0.05) affected by LOAD or its genetic liability (Supplementary Table 17,
254 Supplementary Figures 9-10). Interestingly, both were among the 17 previously identified *APOE*-
255 dependent LOAD proteins, together with two additional proteins that were nominally significant in the
256 reverse MR (TBCA, $P = 4.4e-4$, FDR = 0.051 and IRF6, $P = 7.9e-4$, FDR = 0.055). Thus intriguingly, these
257 findings suggest that these four proteins are upregulated by LOAD, in contrast to the observed *APOE-ε4*
258 downregulation of the same proteins (Figure 5). This supports our findings of competing biological
259 effects described above (Figure 3E, Supplementary Figure 5) and collectively our results indicate that
260 simultaneous opposing effects of *APOE-ε4* on one hand and LOAD on the other result in differential
261 regulation of these proteins in serum (Figure 5B).

262 We performed a replication analysis of the effect of *APOE-ε4* on protein levels and the reverse
263 MR results for these four proteins using published protein GWAS summary statistics from two recent
264 studies^{25,35}. In the external datasets, the downregulation of all four proteins by *APOE-ε4* (as determined
265 by the rs429358 C allele) was replicated. In the reverse MR analysis (excluding the *APOE* locus), the
266 upregulation of protein levels by LOAD liability observed in AGES was also detected for two proteins
267 (S100A13 and TBCA) in both validation cohorts, reaching significance ($P < 0.05$) in the study by
268 Ferkingstad et al. (Supplementary Figure 11, Supplementary Table 18). While the two proteins changed
269 direction in a similar manner as in AGES, the effect size was considerably smaller in the validation
270 cohorts. Importantly, however, individuals in these two cohorts are much younger than those in AGES,
271 with mean ages of 55 and 48 years for the Ferkingstad et al. and Sun et al. studies, respectively,
272 compared to 76 years in AGES. Therefore, we conducted an age-stratified reverse MR analysis in AGES,
273 which showed a strong age-dependent effect, with a much larger effect of LOAD genetic liability on
274 protein levels in individuals over 80 years old compared to those younger than 80 years (Supplementary
275 Figure 11). The effect size in AGES individuals under 80 years old was in line with the effect observed in
276 the validation cohorts. Thus, if the upregulation of these proteins reflects a response to prodromal or

277 preclinical LOAD, an older cohort may be needed to detect an association of the same degree as we
278 found in AGES. However, the observed support in the validation cohorts for the discordant effects of
279 *APOE* vs non-*APOE* LOAD-associated genetic variants on the same serum proteins strongly implicates
280 these proteins as directly relevant to LOAD, potentially as readouts of biological processes that are both
281 disrupted by *APOE-ε4* and modulated in the opposite manner as a response to genetic predisposition to
282 LOAD or the disease onset in general.

283 Together, these results indicate that LOAD or its general genetic liability causally affects the
284 levels of some *APOE*-dependent proteins, but this effect is simultaneously masked by the strong effects
285 of the *APOE* locus in the other direction (Figure 5A). These outcomes strengthen results described above,
286 showing that the levels of these four proteins are strongly downregulated in *APOE-ε4* carriers and lower
287 levels of these proteins are therefore associated with increased risk of LOAD in an *APOE*-dependent
288 manner (Figure 5B). Simultaneously, the reverse MR analysis shows that the collective effect of the other
289 non-*APOE* LOAD risk variants is to upregulate the serum levels of these same proteins, possibly reflecting
290 a response mechanism to LOAD pathogenesis (Figure 5C). Again, this is in line with the observational
291 analysis, where all four proteins changed direction of effect when adjusting for *APOE-ε4* (Figure 5A,
292 Figure 2D-E).

293 **Overlap with the AD brain and CSF proteome.** To evaluate to what extent our LOAD-associated serum
294 proteins reflect the proteomic profile of AD in relevant tissues, we queried data from recent proteomic
295 studies of AD in cerebrospinal fluid (CSF)³⁶ and brain³⁷ which also describe tissue specific co-regulatory
296 modules. We observed that of our LOAD-associated serum proteins, 51 proteins were also associated
297 with AD in brain as measured by mass-spectrometry, with 32 (63%) being directionally consistent (Figure
298 6A-B) (Supplementary Tables 19-20). Higher directional consistency was observed within the *APOE*-
299 independent protein group, or 15 (71%) of 21 proteins associated with AD in brain tissue. Additionally,
300 60 proteins were directly associated with AD in CSF as measured with SOMAscan (7k) (Figure 6A) with 46
301 (77%) being directionally consistent (Figure 6B)²¹. The proportion of directionally consistent associations
302 between serum and CSF was higher in both the *APOE*-independent and dependent protein groups, or
303 88% (22 of 25 and 7 of 8 for *APOE*-independent and dependent proteins, respectively) (Figure 6B,
304 Supplementary Table 19). However, directional inconsistency between plasma and CSF AD proteomic
305 profiles has been reported before in a similar comparison³⁸. Fourteen proteins overlapped between all
306 three tissues in the context of AD (Figure 6A) (Supplementary Table 19). Many of these proteins have
307 established links or are highly relevant to LOAD, such as Spondin 1 (SPON1), involved in the processing of

308 amyloid precursor protein (APP)³⁹; Secreted Modular Calcium-Binding Protein 1 (SMOC1) previously
309 proposed as a biomarker of LOAD in postmortem brains and CSF⁴⁰; Netrin-1 (NTN1), an interactor of APP
310 and regulator of amyloid-beta production⁴¹; Neurofilament light (NEFL), previously proposed as a
311 plasma biomarker for LOAD and axon injury^{42,43} and Von Willebrand factor (VWF), known for its role in
312 blood clotting and associations with LOAD⁴⁴ (Supplementary Table 19). Notably, some of the *APOE*-
313 dependent proteins were associated with AD across all three tissues such as TBCA and TP53I11.

314 We have previously described the co-regulatory structure of the serum proteome, which can
315 broadly be defined as 27 modules of correlated proteins²⁶ (Supplementary Table 21). In the current study
316 we found that among the 346 aptamers (329 proteins) associated with LOAD (prevalent or incident, any
317 model), five serum protein modules (M27, M3, M11, M2 and M24) were overrepresented (Figure 6C,
318 Supplementary Table 22). In particular, the 140 *APOE*-independent aptamers were specifically
319 overrepresented in module M27, enriched for proteins involved in neuron development and the
320 extracellular matrix, and module M3 that is associated with growth factor signaling pathways
321 (Supplementary Table 22). By contrast, the 17 *APOE*-dependent proteins were specifically enriched in
322 protein module M11 (Supplementary Table 22), which is strongly enriched for lipid pathways and is
323 under strong genetic control of the *APOE* locus²⁶. Serum modules M27, M24 and M11 were all enriched
324 for AD-associations in CSF (Figure 6C). We next sought to understand to what extent our LOAD-
325 associated proteins identified in serum might reflect AD protein signatures in CSF and brain tissue.
326 Among the LOAD-associated proteins measured in serum, we found the *APOE*-dependent and *APOE*-
327 independent proteins to be enriched in different CSF modules, most of which were also linked to AD
328 (Figure 6D, Supplementary Table 22). In brain tissue, the serum *APOE*-independent LOAD proteins were
329 particularly enriched in brain module M42 (Matrisome), which is enriched for extracellular matrix (ECM)
330 proteins³⁷. Strikingly, M42 was strongly enriched for the AD-proteomic profiles of all three tissues (Figure
331 6E, Supplementary Table 22). Interestingly, members of this module (SMOC1, APP, SPON1, NTN1,
332 GPNMB) showed some of the strongest associations in serum to incident LOAD in our study (Figure 2D,
333 Supplementary Table 2) as well as in brain (Figure 6B, Supplementary Table 22). This module has
334 furthermore been demonstrated to be correlated with amyloid beta ($A\beta$) deposition in the brain and
335 some of its protein constituents (e.g MDK, NTN1 and SMOC1) have been shown to colocalize with and
336 bind to $A\beta$ ³⁷. Additionally, the *APOE* locus regulates M42 levels in the brain (mod-qTL), and while the
337 *APOE* protein is a member of module M42, this regulation was found to not be solely driven through the
338 levels of the *APOE* protein itself³⁷. Our results simultaneously show that other members of the module,
339 such as SPON1 and SMOC1, exhibit an *APOE*-independent association to incident LOAD in serum.

340 Interestingly, these same two proteins are increased in CSF thirty years prior to symptom onset in
341 autosomal dominant early onset AD⁴⁵. In summary, we demonstrate significant overlaps in LOAD-
342 associated protein expression across blood, CSF and brain on both an individual protein level and on
343 protein module level.

344 **Discussion**

345 We describe a comprehensive mapping of the serum protein profile of LOAD that provides insight into
346 processes that are independent or dependent on the genetic control of *APOE-ε4* (Figure 7). We
347 identified 329 proteins in total that differed in the incident or prevalent LOAD cases compared to non-
348 LOAD participants in a population-based cohort with long-term follow-up. Among these, we identified a
349 novel grouping of proteins based on their primary LOAD-association being statistically independent of
350 (130 proteins), or dependent on (17 proteins) *APOE-ε4* carrier status. Many of the *APOE*-independent
351 proteins are implicated in neuronal pathways and are shared with the LOAD-associated CSF and brain
352 proteome. The 17 *APOE*-dependent proteins overlap with AD-associated protein modules in CSF and
353 interact directly with protein partners involved in LOAD, including APP. Another key finding is, amongst
354 these 17 proteins, four proteins change LOAD-associated direction of effect both observationally and
355 genetically when taking *APOE-ε4* carrier status into account. Collectively, our results suggest that while
356 their primary association with LOAD reflects the risk conferred by *APOE-ε4*, there exists a secondary
357 causal effect of LOAD itself on the protein levels in the reverse direction as supported by the MR
358 analysis, possibly reflecting a response to the disease onset.

359 Previous studies identifying proteins associated with LOAD have been limited to a cross-sectional
360 cohort or are based on all-cause dementia^{18,46-48}. Here we extend those findings by distinguishing LOAD
361 cases from other types of dementia in a prospective cohort study to identify LOAD-specific serum protein
362 signatures preceding clinical onset. Furthermore, our comparative approach of statistical models with
363 and without *APOE-ε4* adjustment provides a novel compartmentalized view of the LOAD serum protein
364 profile and demonstrates how protein effects can differ depending on genetic confounders which are
365 imperative to take into consideration. We found that the proteins associated with incident LOAD in our
366 study, in particular those independently of *APOE-ε4* such as GPNMB, NTN1, SMOC1 and SPON1, overlap
367 with the proteomic profile of LOAD in CSF³⁸ and brain³⁷, and are enriched for neuronal pathways, which
368 may reflect an altered abundance of neuronal proteins in the circulation during the prodromal stage of
369 LOAD. These overlaps that we find across independent cohorts and different proteomics technologies
370 suggest that the serum levels of some proteins have a direct link to the biological systems involved in

371 LOAD pathogenesis and may even provide a peripheral readout of neurodegenerative processes prior to
372 clinical diagnosis of LOAD. In particular, the proteins that show directionally consistent effect sizes
373 suggest exceptional AD-specific robustness as the measurements vary by tissue, methodology and
374 populations.

375 We identified 17 proteins with a particularly strong *APOE*-dependent association to incident LOAD,
376 of which eight were also associated with prevalent AD in CSF. The association between *APOE*- $\epsilon 4$ and
377 circulating levels of these proteins has been reported by our group^{24,26,27} and others⁴⁹, but their direct
378 association with incident LOAD has to our knowledge not been previously described. These *APOE*-
379 dependent proteins may point directly to the processes through which *APOE*- $\epsilon 4$ mediates its risk on
380 LOAD and provide a readout of the pathogenic process in the circulation of the approximately 50% of
381 LOAD patients worldwide carrying the variant^{21,22}. While our data does not provide information on the
382 tissue-origin of the *APOE*-dependent proteins, nine either exhibit brain-specific gene expression, cluster
383 with brain-specific genes⁵⁰ or have been associated with LOAD at the transcriptomic or protein level in
384 brain tissue or CSF (Table 2). At the genetic level, a lookup in the GWAS catalog⁵¹ shows that an intron
385 variant in the *IRF6* gene has a suggestive GWAS association with LOAD via *APOE*- $\epsilon 4$ carrier status
386 interaction⁵². In addition, variants in the *TMCC3* gene have been linked to educational attainment⁵³ and
387 caudate volume change rate⁵⁴ and variants in the *TBCA* gene have been suggestively associated with
388 reaction time⁵⁵ and PHF-tau levels⁵⁶. Collectively, the gene expression patterns for these proteins in the
389 brain, interactions with proteins involved in neuronal processes and suggestive associations between
390 genetic markers in or near these genes and brain-related outcomes suggest that these *APOE*-dependent
391 proteins may reflect brain-specific processes affected by *APOE*- $\epsilon 4$ carrier status that affect the risk of
392 developing LOAD. Importantly, the association patterns for *ARL2*, *S100A13* and *TBCA* suggest the
393 presence of a pathway that is downregulated by *APOE*- $\epsilon 4$ early in life, given the consistent effect of
394 *APOE*- $\epsilon 4$ on the same proteins in younger cohorts, but upregulated at the onset of LOAD, as supported
395 by the larger observed effects in the *APOE*- $\epsilon 4$ adjusted analysis in the ACE cohort of individuals who are
396 closer to diagnosis on the AD trajectory than those in AGES. Additional studies are required to expand on
397 these interpretations and dissect the complex mechanisms at play and to determine if the modulation of
398 the process represented by these proteins has therapeutic potential.

399 Two proteins, *IBSP* and *APP*, were identified to potentially have a causal role in LOAD. *IBSP* was
400 previously associated with plasma amyloid- β and incident dementia⁵⁷, while *APP* is the precursor protein
401 for amyloid- β ⁵⁸. Based on the MR analysis for the LOAD-associated proteins that could be tested, the

402 majority do not appear to be causal in and of themselves but their association with incident LOAD may
403 still reflect changes that occur years before the onset of LOAD that could be of interest to target before
404 irreversible damage accumulates.

405 A major strength of this study is the high-quality data from a prospective longitudinal population-
406 based cohort study with detailed follow-up, broad coverage of circulating proteins and a comprehensive
407 comparison to the AD-proteome in CSF and brain. The limitations of our study include that our results
408 are based on a Northern European cohort and cannot necessarily be transferred directly to other
409 populations or ethnicities. Additionally, while we partly replicate our overall findings in an external
410 cohort, a greater replication proportion could be anticipated in a more comparable cohort. The ACE
411 cohort consists of clinically referred individuals with MCI and proteomic measurements performed on a
412 different version of the SOMAscan platform. Additionally, different normalization procedures were
413 applied by SomaLogic for the two SOMAscan versions, which may have an effect on the LOAD
414 associations⁴⁸. Further studies are required to determine the impact of time to event, platform and
415 normalization approaches on the associations between circulating proteins and LOAD. Regardless of
416 these differences, we did replicate the majority of the *APOE*-dependent LOAD associations, including the
417 *APOE*-dependent change in effect for *ARL2*, *S100A13* and *TBCA*. We could not test all LOAD-associated
418 proteins for causality, including most of the *APOE*-dependent proteins, due to lack of significant cis-
419 pQTLs for two thirds of the proteins, thus we cannot exclude the possibility that some could be causal
420 but missed by our analysis. Finally, despite our LOAD diagnosis criteria it is possible that some of our
421 findings reflect processes related to dementia in general. As a result, it is critical that these findings be
422 validated in individuals with established amyloid-beta and tau deposits, as well as in experimental
423 settings.

424 The proteins highlighted in this study and the mechanisms they point to may be used as a source of
425 biomarkers or therapeutic targets that can be modulated for the prevention or treatment of LOAD. This
426 large prospective cohort study, using both a longitudinal and cross-sectional design, represents a unified
427 and comprehensive reference analysis with which past and future serum protein biomarkers and drug
428 targets can be considered, compared, and evaluated.

429

430 **Acknowledgements**

431 The authors acknowledge the contribution of the Icelandic Heart Association (IHA) staff to AGES-
432 Reykjavik, as well as the involvement of all study participants. The protein measurements in AGES were
433 supported by Novartis Biomedical Research. The National Institute on Aging (NIA) contracts N01-AG-
434 12100 and HHSN271201200022C for V.G. financed the study. IHA received a grant from Althingi (the
435 Icelandic Parliament), V.G. received funding from the NIA (1R01AG065596-01A1) and E.J. from the NIA
436 (1K08AG068604). A.I.L. was also funded from the NIA (P30AG066511 and U01AG061357). The content is
437 solely the responsibility of the authors and does not necessarily represent the official views of the
438 National Institutes of Health. The Genome Research @ Ace Alzheimer Center Barcelona project
439 (GR@ACE) is supported by Grifols SA, Fundación bancaria 'La Caixa', Ace Alzheimer Center Barcelona and
440 CIBERNED. Ace Alzheimer Center Barcelona is one of the participating centers of the Dementia Genetics
441 Spanish Consortium (DEGESCO). The FACEHBI study is supported by funds from Ace Alzheimer Center
442 Barcelona, Grifols, Life Molecular Imaging, Araclon Biotech, Alkahest, Laboratorio de análisis Echevarne
443 and IrsiCaixa. Authors acknowledge the support of the Spanish Ministry of Science and Innovation,
444 Proyectos de Generación de Conocimiento grants PID2021-122473OA-I00, PID2021-123462OB-I00 and
445 PID2019-106625RB-I00. ISCIII, Acción Estratégica en Salud, integrated in the Spanish National R+D+I Plan
446 and financed by ISCIII Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional
447 (FEDER "Una manera de hacer Europa") grants PI13/02434, PI16/01861, PI17/01474, PI19/00335,
448 PI19/01240, PI19/01301, PI22/01403, PI22/00258 and the ISCIII national grant PMP22/00022, funded by
449 the European Union (NextGenerationEU). The support of CIBERNED (ISCIII) under the grants
450 CB06/05/2004 and CB18/05/00010. The support from the ADAPTED and MOPEAD projects, European
451 Union/EFPIA Innovative Medicines Initiative Joint (grant numbers 115975 and 115985, respectively);
452 from PREADAPT project, Joint Program for Neurodegenerative Diseases (JPND) grant N° AC19/00097;
453 from HARPONE project, Agency for Innovation and Entrepreneurship (VLAIO) grant N° PR067/21 and
454 Janssen. DESCARTES project is funded by German Research Foundation (DFG). Amanda Cano received
455 support from the ISCIII under the grant Sara Borrell (CD22/00125).

456 **Author contributions**

457 Conceptualization – E.A.F., Va.G., V.E, Vi.G.
458 Formal Analysis – E.A.F., Va.G., T.J., A.E.S., E.F.G., A.G., T.A., E.B.D., A.S.
459 Resources – Vi.G., V.E., A.I.L., A.R., A.P.O., J.J.L., L.L.J., N.T.S., E.C.B.J., L.J.L.
460 Data Curation – R.P., A.C., M.B., P.G.G., S.V., Va.G., T.A., E.F.G.
461 Writing original draft – E.A.F., Va.G.

462 Writing review & editing – All authors

463 Visualization – E.A.F., Va.G.

464 Supervision – Va.G., Vi.G.

465 Funding Acquisition – Vi.G., L.J.L.

466 **Declaration of interests**

467 L.L.J., A.P.O and J.J.L are employees and stockholders of Novartis. N.T.S and A.I.L are co-founders of
468 Emtherapro. No other potential conflicts of interest relevant to this article were reported.

469 **Figure titles and legends**

470 **Fig. 1 – Study overview** Flowchart of the current study. **A)** Overview of the AGES cohort and study
471 participants. Prevalent non-AD dementia cases were excluded from the analysis. **B)** Overview of the
472 aptamers tested and their associations with LOAD. Serum measurements of 4782 aptamers were
473 associated to prevalent and incident LOAD status, using logistic and Cox proportional hazards regression
474 models, respectively. From the proteins associated with incident LOAD, sets of 140 proteins with an
475 *APOE*-independent associations and 17 proteins with an *APOE*-dependent association were defined. The
476 *APOE*-dependent proteins were further expanded to first degree protein-protein interaction (PPI)
477 partners. All sets of proteins were subjected to functional enrichment analysis and bi-directional
478 Mendelian Randomization (MR) analysis. **C)** Overview of the replication cohorts used in the study which
479 include proteins measured in the circulation (ACE) as well as in brain and CSF (Emory).

480 **Fig. 2 – Proteins associated with LOAD in AGES. A-B)** Volcano plots showing the protein association
481 profile for incident LOAD **A)** without *APOE-ε4* adjustment (model 1) and **B)** with *APOE-ε4* adjustment
482 (model 2). **C)** Venn diagram for the overlap between models 1 and 2 for incident LOAD. **D-E)** Enrichment
483 of top Gene Ontology terms from GSEA analysis for incident LOAD (model 1) shown as **D)** dotplot,
484 stratified by ontology and **E)** gene-concept network. **F-G)** Comparison of effect sizes (HR) for incident
485 LOAD between the AGES and the ACE cohorts for all proteins reaching nominal significance ($P < 0.05$) in
486 ACE for **F)** model 1 and **G)** model 2.

487 **Fig. 3 – Proteins with *APOE-ε4* dependent association to incident LOAD. A)** Spaghetti plot showing the
488 statistical significance of protein associations with incident LOAD across the three models, highlighting a
489 set of 17 unique proteins (green) whose association with incident LOAD is attenuated upon *APOE-ε4*
490 adjustment. The horizontal lines indicate $FDR < 0.05$ (dashed) and $P < 0.05$ (dot-dashed). **B)** Pairwise

491 Pearson's correlation between the 17 *APOE*-dependent proteins. **C)** Forest plot showing the linear
492 associations between *APOE* genotype and the 17 *APOE*-dependent proteins. The beta coefficient shows
493 the change in protein levels per $\epsilon 4$ allele count. **D-E)** Forest plots showing the associations between the
494 17 *APOE*-dependent proteins and incident LOAD **D)** without *APOE*- $\epsilon 4$ adjustment (model 1) and **E)** with
495 *APOE*- $\epsilon 4$ adjustment (model 2). LOAD-HR indicates risk per SD increase of protein levels. Proteins that
496 change direction of effect between the two models are highlighted in red.

497 **Fig. 4 – Functional enrichment analysis of *APOE*-dependent protein-protein interaction partners.** **A)** A
498 scheme of the PPI partners selection, where first degree partners of the *APOE*-dependent proteins were
499 extracted from the InWeb database. **B-C)** Enrichment of selected Gene Ontology terms for the PPI
500 partner proteins shown as **B)** dotplot and **C)** gene-concept network. **D-E)** Enrichment of top seven unique
501 Wikipathways shown as **D)** dotplot and **E)** gene-concept network.

502 **Fig. 5 – Reverse Mendelian randomization analysis.** **A)** Comparison of hazard ratios for incident LOAD
503 with and without *APOE*- $\epsilon 4$ adjustment in the observational analysis, the effects of *APOE*- $\epsilon 4$ on protein
504 levels and reverse MR odds ratios (excluding the *APOE* locus) for the four *APOE*-dependent proteins that
505 change direction of effect in both observational and causal analyses when *APOE* is accounted for. **B-C)**
506 Visual summaries of the observed data. **B)** Mediation diagrams showing 3 possible hypotheses that could
507 explain the relationship between *APOE*- $\epsilon 4$, LOAD and the four proteins shown in **A)**. Our analyses do not
508 support the hypothesis that LOAD mediates the effect of *APOE*- $\epsilon 4$ on proteins (Hypothesis 1) nor the
509 other way around (Hypothesis 2). However, our results from both the observational and causal analyses
510 support the hypothesis that two mechanisms are at play that affect the same proteins in the opposite
511 direction (Hypothesis 3). **C)** The *APOE*- $\epsilon 4$ mutation leads to increased risk of LOAD via its effects in brain
512 tissue. The same mutation results in a downregulation of serum levels of four proteins that are
513 themselves negatively associated with incident LOAD. Additionally, other non-*APOE* LOAD risk variants
514 lead to upregulation of the same proteins in the reverse MR analysis, possibly reflecting a response to
515 LOAD or its genetic liability.

516 **Fig. 6 – Overlap between AD protein signatures in serum, brain and CSF.** **A)** A Venn diagram showing
517 the overlap of AD-associated proteins in serum, brain and CSF. **B)** A comparison of the effect sizes for AD
518 associated proteins that overlap between serum and brain (top) and serum and CSF (bottom). The
519 proteins are stratified based on the *APOE*-dependence in AGES for incident LOAD. The effect size in AGES
520 is shown for incident LOAD model 1, except for proteins that were uniquely identified using the shorter
521 10-year follow-up or prevalent LOAD, in which case the respective effect size from the significant

522 association is shown. **C-E)** Heatmap showing the enrichment of AD-associated proteins by tissue type (x-
 523 axis) in **C)** the AGES serum protein modules, **D)** Emory CSF protein modules and **E)** Emory brain protein
 524 modules (y-axis).

525 **Fig. 7 – Graphical summary of the results.**

526 **Tables with titles and legends**

527 **Table 1 –** A summary table of the top 20 significant *APOE-independent* proteins associated with incident
 528 LOAD in AGES. The effect size (Hazard Ratio (Confidence Interval)) and level of significance (P, FDR) is
 529 shown for model 2, adjusting for age, sex and *APOE-ε4*. The final column indicates if the aptamers have
 530 been orthogonally validated by mass-spectrometry²⁶. The *APOE-independency* is defined as proteins
 531 remaining significantly (FDR < 0.05) associated with incident LOAD after *APOE-ε4* adjustment.

Entrez symbol	HR	P	FDR	MS validated
CD4	0.79 (0.73-0.87)	1.9E-07	9.3E-04	
BRD4	0.82 (0.75-0.89)	1.4E-06	3.4E-03	
SPON1	1.24 (1.13-1.36)	3.6E-06	3.9E-03	
CLEC2L	0.83 (0.77-0.9)	5.7E-06	3.9E-03	
SOD3	1.21 (1.11-1.31)	5.2E-06	3.9E-03	Yes
ACHE	1.22 (1.12-1.33)	3.9E-06	3.9E-03	
IL1R1	1.23 (1.12-1.34)	4.7E-06	3.9E-03	
ADAM30	0.83 (0.77-0.9)	1.1E-05	6.0E-03	
PLA2G2E	0.82 (0.75-0.89)	1.2E-05	6.0E-03	
TFPI	1.2 (1.11-1.31)	1.3E-05	6.0E-03	Yes
SVEP1	1.21 (1.11-1.33)	2.5E-05	0.011	Yes
SVEP1	1.21 (1.11-1.32)	3.3E-05	0.011	Yes
LCORL	0.82 (0.75-0.9)	3.2E-05	0.011	
PROK1	0.84 (0.77-0.91)	3.1E-05	0.011	
NCAM1	1.2 (1.1-1.3)	4.0E-05	0.011	Yes
SMOC1	1.21 (1.11-1.33)	3.9E-05	0.011	
SOD3	1.19 (1.09-1.29)	4.0E-05	0.011	
NLGN1	1.18 (1.09-1.28)	4.5E-05	0.011	
SURF1	1.19 (1.1-1.3)	4.3E-05	0.011	
IGF2	0.84 (0.77-0.92)	5.4E-05	0.012	

532

533 **Table 2 –** A summary table of the 17 *APOE-dependent* LOAD associated proteins, describing their tissue
 534 specificity in the Human Protein Atlas v22, results from the association analyses in AGES, references for
 535 previous associations with *APOE* or LOAD and whether the aptamers have been orthogonally validated

536 via MS²⁶. The APOE-dependency is defined as being significant (FDR<0.05) in model 1 and fully non-
 537 significant in model 2 (P>0.05) for incident LOAD.

					Incident LOAD (Cox)				APOE-e4 (linear regression)				MS validation
					Model 1		Model 2						
Entrez Symbol	Tissue specificity	Tissue cluster	LOAD ref	APOE ref	HR	FDR	HR	P	Beta	SE	P	FDR	
LRRN1	Brain	Brain	59	49	1.258	1.E-05	1.018	0.711	1.061	0.023	<1E-300	<1E-300	
S100A13	Low	Mitochondria		49	0.799	7.E-05	1.099	0.092	-1.090	0.021	<1E-300	<1E-300	
TBCA	Low	Vesicular transport		49	0.782	7.E-06	1.086	0.145	-1.183	0.020	<1E-300	<1E-300	Yes
CTF1	Low	Fibroblasts, ECM	60	49	1.262	2.E-04	1.077	0.147	0.611	0.021	1.E-171	5.E-171	
ARL2	Low	Striated muscle			0.835	0.002	1.093	0.077	-0.991	0.022	<1E-300	<1E-300	
C1orf56	Testis	Spermatogenesis			1.200	0.002	1.070	0.113	0.513	0.026	6.E-85	9.E-85	
MSN	Low	Neutrophils, Inflammatory response	61		1.186	0.003	1.039	0.388	0.608	0.025	9.E-122	2.E-121	
TMCC3	Brain	Brain, Nervous system development	62		0.849	0.005	0.962	0.364	-0.517	0.026	4.E-86	7.E-86	
HBQ1	Bone marrow	Neutrophils, Humoral immune response			0.860	0.013	0.966	0.423	-0.455	0.025	1.E-69	1.E-69	
IRF6	Esophagus, Skin	Squamous epithelial cells			0.876	0.032	1.033	0.458	-0.675	0.025	4.E-153	1.E-152	
IFIT2	Bone marrow	Brain & bone marrow, Chromatin organization			0.875	0.035	0.949	0.214	-0.275	0.026	2.E-26	2.E-26	
TP53I11	Low	Adipose tissue, ECM organization			0.879	0.036	0.981	0.653	-0.436	0.026	1.E-60	1.E-60	
FAM159B	Brain and other	Unknown function			0.882	0.039	0.977	0.591	-0.419	0.026	8.E-56	9.E-56	
NDE1	Low	Brain, Nervous system development	63		0.876	0.039	1.001	0.985	-0.489	0.025	3.E-80	5.E-80	Yes
NEFL	Brain	Brain, Neuronal signaling	64		0.878	0.041	1.031	0.516	-0.578	0.024	2.E-121	4.E-121	
GSTM1	Brain	Liver, Metabolism	65	66	1.136	0.042	1.077	0.087	0.276	0.026	1.E-26	1.E-26	
GGT2	Kidney, Thyroid	Kidney, Transmembrane transport			0.887	0.048	0.981	0.646	-0.384	0.026	2.E-47	3.E-47	

538

539 **Table 3** – Replication of the LOAD associated proteins from AGES in the ACE cohort. All proteins with
 540 nominal P<0.05 in either model are shown. P and FDR values <0.05 are highlighted in bold.

			Model 1				Model 2			
Aptamer	Gene symbol	APOE proteins AGES	HR	95% CI	P	FDR	HR	95% CI	P	FDR
11293_14_3	LRRN1	Dependent	1.410	1.256-1.582	5.4E-09	1.8E-06	1.142	0.969-1.346	0.112	0.839
5744_12_3	C1orf56	Dependent	1.354	1.190-1.541	4.3E-06	5.2E-04	1.175	1.022-1.352	0.024	0.565
13732_79_3	CTF1	Dependent	1.324	1.174-1.494	4.7E-06	5.2E-04	1.034	0.888-1.205	0.666	0.958
10082_251_3	NEFL	Dependent	0.771	0.683-0.870	2.5E-05	0.002	1.084	0.917-1.282	0.346	0.958
12501_10_3	TBCA	Dependent	0.789	0.695-0.895	2.3E-04	0.015	1.279	1.063-1.538	0.009	0.336
6462_12_3	TIMP4	Independent	1.261	1.109-1.434	4.0E-04	0.022	1.236	1.087-1.405	1.2E-03	0.198
8819_3_3	IGFBP2	Independent	1.253	1.087-1.445	0.002	0.085	1.250	1.082-1.443	0.002	0.204

2570_72_5	IGFBP2	Independent	1.237	1.080-1.415	0.002	0.085	1.234	1.077-1.414	0.002	0.204
8469_41_3	IGFBP2	Independent	1.230	1.066-1.418	0.004	0.164	1.222	1.059-1.410	0.006	0.316
10978_39_3	GH2	Independent	1.182	1.043-1.340	0.009	0.298	1.131	0.998-1.281	0.053	0.751
2813_11_2	AGRP	Other	1.187	1.039-1.355	0.011	0.346	1.196	1.048-1.364	0.008	0.323
7100_31_3	CD2	Independent	1.166	1.031-1.317	0.014	0.353	1.102	0.975-1.246	0.121	0.839
6168_11_3	TP53	Other	1.190	1.034-1.369	0.015	0.353	1.127	0.976-1.302	0.104	0.839
4964_67_1	ERAP1	Independent	0.861	0.763-0.972	0.015	0.353	0.841	0.742-0.953	0.007	0.316
7223_60_3	S100A13	Dependent	0.853	0.749-0.971	0.016	0.353	1.500	1.240-1.816	3.1E-05	0.010
4930_21_1	STC1	Independent	1.173	1.024-1.343	0.021	0.391	1.174	1.024-1.346	0.021	0.565
9539_25_3	COL26A1	Other	0.865	0.764-0.979	0.022	0.391	0.890	0.788-1.005	0.059	0.756
7239_9_3	GSTM1	Dependent	1.159	1.021-1.316	0.023	0.391	1.070	0.941-1.217	0.304	0.951
4929_55_1	SHBG	Other	1.168	1.022-1.335	0.023	0.391	1.149	1.005-1.313	0.042	0.751
13118_5_3	SMOC1	Independent	1.172	1.022-1.345	0.024	0.391	1.157	1.006-1.331	0.041	0.751
5660_51_3	SOD3	Independent	1.162	1.018-1.326	0.026	0.411	1.162	1.018-1.325	0.026	0.565
5581_28_3	FGL1	Other	1.149	1.015-1.301	0.028	0.430	1.118	0.988-1.266	0.077	0.757
13022_20_3	TP53I11	Dependent	0.877	0.777-0.991	0.035	0.458	1.059	0.930-1.206	0.388	0.958
8235_48_3	CHGB	Other	1.151	1.010-1.311	0.035	0.458	1.123	0.983-1.282	0.088	0.797
7016_12_3	GCNT1	Independent	0.864	0.754-0.990	0.036	0.458	0.874	0.763-1.001	0.052	0.751
13731_14_3	C7	Independent	1.159	1.010-1.331	0.036	0.458	1.147	1.001-1.316	0.049	0.751
2925_9_1	SERPINE1	Other	0.875	0.770-0.995	0.042	0.511	0.870	0.762-0.992	0.038	0.751
3290_50_2	CD109	Independent	0.884	0.784-0.996	0.043	0.511	0.896	0.794-1.012	0.076	0.757
6626_81_3	CHST12	Independent	0.871	0.761-0.997	0.045	0.518	0.904	0.790-1.036	0.146	0.848
6604_59_3	NDNF	Other	0.879	0.774-0.999	0.048	0.528	0.897	0.790-1.019	0.096	0.817
14129_1_3	IFNA7	Other	1.131	1.000-1.279	0.051	0.543	1.134	1.003-1.282	0.045	0.751
8072_19_3	MZT1	Independent	1.118	0.996-1.256	0.058	0.604	1.146	1.020-1.287	0.022	0.565
4982_54_1	PI3	Other	1.139	0.989-1.311	0.070	0.608	1.185	1.027-1.366	0.020	0.565
4297_62_3	SPON1	Independent	1.127	0.989-1.285	0.073	0.608	1.167	1.019-1.336	0.025	0.565
8242_9_3	CLEC2L	Independent	1.102	0.974-1.248	0.124	0.686	1.136	1.003-1.287	0.045	0.751
12587_65_3	ARL2	Dependent	0.909	0.804-1.028	0.130	0.686	1.252	1.075-1.457	0.004	0.255

541

542

543 **Methods**

544 *AGES study population*

545 Participants aged 66 through 96 were from the Age, Gene/Environment Susceptibility (AGES)-Reykjavik
546 Study cohort. AGES is a single-center prospective population-based study of deeply phenotyped subjects
547 (n = 5,764, mean age 76.6 \pm 5.6 years) and survivors of the 40-year-long prospective Reykjavik study, an
548 epidemiologic study aimed to understand aging in the context of gene/environment interaction by
549 focusing on four biologic systems: vascular, neurocognitive (including sensory), musculoskeletal, and
550 body composition/metabolism³⁰. Of the AGES participants, 3,411 attended a 5-year follow-up visit. LOAD
551 diagnosis at AGES baseline and follow-up visits was carried out using a three-step procedure described in
552 Sigurdsson et al.⁶⁷. Cognitive assessment was carried out on all participants. Neuropsychological testing
553 was performed on individuals with suspected dementia. Individuals remaining suspect for dementia
554 underwent further neurologic and proxy examinations in the second diagnosis step. Thirdly, a panel
555 comprising of a neurologist, geriatrician, neuroradiologist, and neuropsychologist assessed the positive
556 scoring participants according to international guidelines³⁰ and gave a dementia diagnosis. The
557 participants were followed up for incident dementia through medical and nursing home reports and
558 death certificates. The follow-up time was up to 16.9 years, with the last individual being diagnosed 16
559 years from baseline. Nursing home reports were based on intake exams upon entry or standardized
560 procedures carried out in all Icelandic nursing homes⁶⁸. The participants diagnosed at baseline were
561 defined as prevalent LOAD cases while individuals diagnosed with LOAD during the follow-up period
562 were defined as incident LOAD cases. All prevalent non-AD dementia cases (n=163) were excluded from
563 analyses.

564 Age, sex, education, and lifestyle variables were assessed via questionnaires at baseline.
565 Education was categorized as primary, secondary, college, or university degree. Smoking was
566 characterized as current, former, or never smoker. *APOE* genotyping was assessed via microplate array
567 diagonal gel electrophoresis (MADGE)⁶⁹. BMI and hypertension were assessed at baseline. BMI was
568 calculated as weight (kg) divided by height (m) squared, and hypertension was defined as
569 antihypertensive treatment or BP >140/90 mm Hg. Type 2 diabetes was defined from self-reported
570 diabetes, diabetes medication use, or fasting plasma glucose \geq 7 mmol/L. Serum creatinine was measured
571 via the Roche Hitachi 912 instrument and estimated glomerular filtration rate (eGFR) derived with the
572 four-variable MDRD study equation⁷⁰. The AGES study was approved by the NBC in Iceland (approval

573 number VSN-00-063), and by the National Institute on Aging Intramural Institutional Review Board, and
574 the Data Protection Authority in Iceland.

575 *Proteomic measurements*

576 The proteomic measurements in AGES have been described in detail elsewhere^{27,71} and was available for
577 5,457 participants. Briefly, a custom version of the SOMAscan platform (Novartis V3-5K) was applied
578 based on the slow-off rate modified aptamer (SOMAmer) protein profiling technology^{72,73} including 4,782
579 aptamers that bind to 4,137 human proteins. Serum was prepared using a standardized protocol⁷⁴ from
580 blood samples were collected after an overnight fast, stored in 0.5 ml aliquots at -80°C and serum
581 samples that had not been previously thawed were used for the protein measurements. All samples
582 were run as a single set at SomaLogic Inc. (Boulder, CO, US). Hybridization controls were used to adjust
583 for systematic variability in detection and calibrator samples of three dilution sets (40%, 1%, and 0.005%)
584 were included so that the degree of fluorescence was a quantitative reflection of protein concentration.
585 All aptamers that passed quality control had median intra-assay and inter-assay coefficient of variation
586 (CV) < 5%. Finally, intraplate median signal normalization was applied to individual samples by
587 SomaLogic instead of normalization to an external reference of healthy individuals, as is done for later
588 versions of the SOMAscan platform ([https://somalogic.com/wp-](https://somalogic.com/wp-content/uploads/2022/07/SL00000048_Rev-3_2022-01_-Data-Standardization-and-File-Specification-Technical-Note-v2.pdf)
589 [content/uploads/2022/07/SL00000048_Rev-3_2022-01_-Data-Standardization-and-File-Specification-](https://somalogic.com/wp-content/uploads/2022/07/SL00000048_Rev-3_2022-01_-Data-Standardization-and-File-Specification-Technical-Note-v2.pdf)
590 [Technical-Note-v2.pdf](https://somalogic.com/wp-content/uploads/2022/07/SL00000048_Rev-3_2022-01_-Data-Standardization-and-File-Specification-Technical-Note-v2.pdf)).

591 *ACE cohort*

592 ACE Alzheimer Center Barcelona was founded in 1995 and has collected and analyzed roughly 18,000
593 genetic samples, diagnosed over 8,000 patients, and participated in nearly 150 clinical trials to date. For
594 more details, visit www.fundacioace.com/en. The syndromic diagnosis of all subjects of the ACE cohort
595 was established by a multidisciplinary group of neurologists, neuropsychologists, and social workers.
596 Healthy controls (HCs), including individuals with subjective cognitive decline (SCD) diagnosis, were
597 assigned a Clinical Dementia Rating (CDR) of 0, and mild cognitive impairment (MCI) individuals a CDR of
598 0.5. For MCI diagnoses, the classification of López *et al.*, 2003, and Petersen's criteria were used⁷⁵⁻⁷⁸. The
599 2011 National Institute on Aging and Alzheimer's Association (NIA-AA) guidelines were used for AD
600 diagnosis⁷⁹. All ACE clinical protocols have been previously published⁸⁰⁻⁸². Paired plasma and CSF
601 samples⁸³, following consensus recommendations, were stored at -80°C. A subset of ACE cohort was
602 analyzed with the SOMAscan 7k proteomic platform⁸⁴ (n = 1,370), (SomaLogic Inc., Boulder, CO, US). The

603 proteomic data underwent standard quality control procedures at SomaLogic and was median
604 normalized to reference using the Normalization by Maximum Likelihood (ANML) method
605 ([https://somalogic.com/wp-content/uploads/2022/07/SL00000048_Rev-3_2022-01_-Data-](https://somalogic.com/wp-content/uploads/2022/07/SL00000048_Rev-3_2022-01_-Data-Standardization-and-File-Specification-Technical-Note-v2.pdf)
606 [Standardization-and-File-Specification-Technical-Note-v2.pdf](https://somalogic.com/wp-content/uploads/2022/07/SL00000048_Rev-3_2022-01_-Data-Standardization-and-File-Specification-Technical-Note-v2.pdf)). Additionally, *APOE* genotyping was
607 assessed using TaqMan genotyping assays for rs429358 and rs7412 SNPs (Thermo Fisher). Genotypes
608 were furthermore extracted from the Axiom 815K Spanish Biobank Array (Thermo Fisher) performed by
609 the Spanish National Center for Genotyping (CeGen, Santiago de Compostela, Spain).

610 *Statistical analysis*

611 Protein measurement data was centered, scaled and Box-Cox transformed, and extreme outliers
612 excluded as previously described⁷¹. Sample size was not predetermined by any statistical method but
613 rather by available data. The associations of serum protein profiles with prevalent AD (n = 167) were
614 examined cross-sectionally via logistic regression at baseline. The associations of serum protein profiles
615 with incident LOAD (n = 655) were examined longitudinally via Cox proportional-hazards models.
616 Participants who died or were diagnosed with incident non-AD dementia were censored at date of death
617 or diagnosis. To account for hazard ratio variability which may arise with lengthy follow-ups, a secondary
618 analysis using 10-year follow-up cut-off of incident LOAD was performed (n_{LOAD} = 432). To compare the
619 fits of the two follow-up times and test for time-dependence of the coefficients we used anova and the
620 `survsplit` function from the Survival R package⁸⁵. For both prevalent and incident LOAD, we examined
621 three covariate-adjusted models. The primary model (model 1) included the covariates sex and age.
622 Model 2 included as an additional covariate the *APOE-ε4* allele count. The third model (model 3)
623 included additional adjustment for cardiovascular risk factors, lifestyle, and kidney function (BMI, type 2
624 diabetes, education, hypertension, smoking history, eGFR) as they have been associated with risk of
625 LOAD⁸⁶. Benjamini-Hochberg false discovery rate (FDR) was used to account for multiple hypothesis
626 testing. Analyses were conducted using R version 4.2.1. ACE SomaLogic proteomics data was similarly
627 Box-Cox transformed and association analysis performed in the same manner as in AGES.

628 *APOE-dependence criteria* of the proteins were defined as serum proteins that met FDR significance of <
629 0.05 in association with incident LOAD in model 1, thus unadjusted for the *APOE-ε4* allele, but whose
630 nominal significance was abolished upon *APOE-ε4* correction in model 2 (P > 0.05). Serum proteins that
631 remained nominally significantly associated with incident LOAD (P < 0.05) upon *APOE-ε4* correction but
632 changed direction of effect were also considered to meet the *APOE* dependence criteria, as a reversal of
633 the effect indicates that the primary association is driven by *APOE-ε4*.

634 Functional enrichment analyses were performed using Over-Representation Analysis (ORA) and Gene Set
635 Enrichment Analysis (GSEA) using the R packages ClusterProfiler and fgsea^{87,88}. The association
636 significance cut-off for inclusion in ORA was FDR < 0.05. Background for both methods was specified as
637 all proteins tested from the analysis leading up to enrichment testing. The investigated gene sets were
638 the following: Gene Ontology, Human Phenotype Ontology, KEGG, Wikipathways, Reactome, Pathway
639 Interaction Database (PID), MicroRNA targets (MIRDB and Legacy), Transcription factor targets (GTRD
640 and Legacy), ImmuneSigDB and the Vaccine response gene set⁸⁹. Finally, we looked into tissue gene
641 expression signatures via the same methods (ORA and GSEA) using data from GTEX⁹⁰ and The Human
642 Protein Atlas, where gene expression patterns across tissues were categorized in a similar manner as
643 described by Uhlen et al⁵⁰ and tissue-elevated expression considered as gene expression in any of the
644 categories 'tissue-specific', 'tissue-enriched' or 'group-enriched'. MinGSSize was set at 2 when
645 investigating the LOAD-associated serum proteins directly. When investigating the *APOE*-dependent
646 protein interaction partners, minGSSize was set to 15 and maxGSSize was set to 500. Overrepresentation
647 of brain cell type markers among LOAD-associated proteins was tested using a Fisher's exact test, with
648 the SOMApanel protein set as background. Tissue specificity lookup for the top LOAD associated proteins
649 was based on the Human Protein Atlas version 22 (<https://v22.proteinatlas.org/>). For the protein-protein
650 interaction (PPI) network analysis, PPIs from InWeb³² (n = 14,448, after Entrez ID filtering) were used to
651 obtain the first-degree interaction partners of the *APOE*-dependent proteins.

652 *Protein comparisons across serum, CSF and brain*

653 To compare protein modules and AD associations across tissues, protein modules and protein
654 associations to AD were obtained from brain³⁷ and CSF³⁶. The brain data, from the Banner Sun Health
655 Research Institute⁹¹ and ROSMAP⁹², included TMT-MS-based quantitative proteomics for 106 controls,
656 200 asymptomatic AD cases and 182 AD cases. The CSF samples were collected under the auspices of the
657 Emory Goizueta Alzheimer's Disease Research Center (ADRC) and Emory Healthy Brain Study (EHBS)³⁶.
658 The cohort consisted of 140 healthy controls and 160 patients with AD as defined by the NIA research
659 framework⁹³. Protein measurements were performed using TMT-MS and SomaScan (7k). Only
660 SomaLogic protein measurements were included in the comparison between CSF and serum, which were
661 median normalized. Proteins were matched on SomaLogic aptamer ID when possible but otherwise by
662 Entrez gene symbol. Overlaps between modules and AD-associated (FDR<0.05) proteins across tissues
663 were evaluated with Fisher's exact test.

664 *Mendelian randomization*

665 A two-sample bi-directional Mendelian randomization (MR) analysis was performed to first evaluate the
666 potential causal effects of serum protein levels on AD (forward MR), and second to evaluate the
667 potential causal effects of AD or its genetic liability on serum protein levels (reverse MR). All aptamers
668 significantly ($FDR < 0.05$) associated with LOAD (incident or prevalent) were included in the MR analyses,
669 or a total of 346 unique aptamers (Supplementary Tables 2, 3 and 5), of which 320 aptamers were
670 significant in the full follow-up incident LOAD analysis (models 1-3), 106 aptamers were significant in the
671 10-year follow-up incident LOAD analysis (model 1-3) and ten aptamers were significant in the prevalent
672 LOAD analysis (models 1-3). Genetic instruments for serum protein levels were obtained from a GWAS of
673 serum protein levels in AGES²⁴ and defined as follows. All variants within a 1 Mb (± 500 kb) cis-window
674 for the protein-encoding gene were obtained for a given aptamer. A cis-window-wide significance level
675 $P_{\text{cis}} = 0.05/N$, where N equals the number of SNPs within a given cis-window, was computed and
676 variants within the cis window for each aptamer were clumped ($r^2 \geq 0.2$, $P \geq P_{\text{cis}}$). The effect of the
677 genetic instruments for serum protein levels on LOAD risk was obtained from a GWAS on GWAS on
678 39,106 clinically diagnosed LOAD cases, 46,828 proxy-LOAD and dementia cases and 401,577 controls of
679 European ancestry¹⁹. Genetic instruments for the serum protein levels not found in the LOAD GWAS data
680 set were replaced by proxy SNPs ($r^2 \geq 0.8$) when possible, to maximize SNP coverage. Genetic
681 instruments for LOAD in the reverse causation MR analysis were obtained from the same LOAD GWAS²⁴,
682 where genome-wide significant variants were extracted ($P < 5 \times 10^{-8}$) and clumped at a more stringent LD
683 threshold ($r^2 \geq 0.01$) than for the protein instruments to limit overrepresentation of SNPs from any
684 given locus across the genome. In the reverse causation MR analysis, cis variants (± 500 kb) for the given
685 protein were excluded from the analysis to avoid including pleiotropic instruments affecting the
686 outcome (protein levels) through other mechanisms than the exposure (LOAD). A secondary reverse
687 causation MR analysis was performed excluding any variants in the *APOE* locus (chr19:45,048,858-
688 45,733,201). Causal estimate for each protein was obtained by the generalized weighted least squares
689 (GWLS) method⁹⁴, which accounts for correlation between instruments. Causality for proteins with single
690 cis-acting variants was assessed with the Wald ratio estimator. For the reverse causation MR analysis,
691 the inverse variance weighted method was applied due to a more stringent LD filtering of the
692 instruments. Instrument heterogeneity was evaluated with Cochran's Q test and horizontal pleiotropy
693 with the MR Egger test.

694 **Supplemental information titles and legends**

695 **Supplementary Figure 1 – A-B)** Volcano plots showing the protein association profile for incident LOAD,
696 restricted to 10-year follow-up **A)** without *APOE e4* adjustment (model 1) and **B)** with *APOE e4*
697 adjustment (model 2). **C)** Venn diagram for the aptamer overlap between models 1 and 2 for incident
698 AD, restricted to 10-year follow-up.

699 **Supplementary Figure 2 – A-B)** Volcano plots showing the protein association profile for prevalent LOAD,
700 **A)** without *APOE e4* adjustment (model 1) and **B)** with *APOE e4* adjustment (model 2). **C)** Venn diagram
701 for the aptamer overlap between models 1 and 2 for prevalent LOAD. **D)** Heatmap comparing the effect
702 sizes for models 1 and 2 and both prevalent (odds ratio) and incident LOAD (hazard ratio). Serum protein
703 module membership and *APOE*-dependence (see Methods for definition) are annotated at the top.

704 **Supplementary Figure 3 –** Spaghetti plots showing the statistical significance of protein associations with
705 **A)** prevalent LOAD and **B)** incident LOAD restricted to 10 year follow-up across the three models,
706 highlighting the 17 *APOE*-dependent proteins (green) defined as those whose association with incident
707 LOAD is attenuated with *APOE e4* adjustment for incident AD using full follow-up (Fig 2a).

708 **Supplementary Figure 4 –** Boxplots visualizing protein levels in **A)** AGES serum samples, see
709 Supplementary Table 9 for statistical evaluation and **B)** ACE plasma samples, stratified by *APOE*-*e4* allele
710 carrier and incident LOAD status, for the 17 proteins with *APOE*-dependent association to incident LOAD
711 in AGES.

712 **Supplementary Figure 5 – A)** Histogram showing the incident LOAD HR for *APOE-e4* allele count in AGES
713 from Cox proportional hazard regression models adjusted for a single LOAD-associated protein at a time,
714 in addition to age and sex. The dashed line indicates the HR for *APOE-e4* when only adjusting for age and
715 sex. The proteins with the largest effect on the *APOE-e4* HR when included in the model are denoted. **B)**
716 Forest plot showing the incident LOAD HR with 95% confidence intervals for *APOE-e4* when adjusting for
717 each of the 17 *APOE*-dependent model in the Cox proportional hazard regression.

718 **Supplementary Figure 6 – A-B)** Spaghetti plots showing the statistical significance of protein associations
719 with incident LOAD across three statistical models, adjusting for age and sex, and then additionally either
720 serum **A)** total cholesterol or **B)** LDL cholesterol, and finally *APOE e4* carrier status. The LOAD association
721 of the 17 *APOE*-dependent proteins (green) is not attenuated when adjusting for total cholesterol or LDL
722 but only when adjusting for *APOE e4* carrier status. **C-H)** Protein associations with serum total
723 cholesterol (**C-E)** and LDL cholesterol (**F-H)**. The density plots show the effect size for the associations
724 between cholesterol levels and all measured proteins stratified by their AD association, adjusting for age

725 and sex (C and F), and additionally *APOE-e4* carrier status (D and G). E) and H) Comparisons of the effect
726 sizes for association with cholesterol levels in the models with (y-axis) and without (x-axis) *APOE-e4*
727 adjustment shows that the effect size is often increased for the *APOE*-dependent proteins (red) after
728 *APOE-e4* adjustment.

729 **Supplementary Figure 7** – Reverse MR analysis (excluding the *APOE* locus) for the causal effect of A)
730 total cholesterol and B) LDL cholesterol on serum levels of the 17 *APOE*-dependent proteins. C-E) For
731 each protein, a comparison of the effect sizes for the association with the *APOE-ε4* genotype (x-axis) and
732 C) incident LOAD, D) total cholesterol and E) LDL cholesterol (y-axis). TC, total cholesterol; LDL, low
733 density lipoprotein cholesterol.

734 **Supplementary Figure 8** – MR scatterplots for all proteins with more than one genetic instrument and
735 with $P < 0.05$ in the MR analysis for a causal effect of protein levels on AD. Each point represents a genetic
736 instrument (SNP) and shows its effect on serum protein levels (x-axis) and AD (y-axis). The slope of the
737 blue line indicates the inverse variance weighted MR effect.

738 **Supplementary Figure 9** – Example MR scatterplots from the reverse MR analysis for a causal effect of
739 AD on serum protein levels, showing discordant direction of effect (inverse variance weighted method
740 indicated by slope of dashed line) when including (left) or excluding (right) *APOE* genetic variants. Each
741 point represents a genetic instrument (SNP) and shows its effect on AD (x-axis) and serum protein levels
742 (y-axis).

743 **Supplementary Figure 10** – A forest plot showing the results for individual SNPs in the reverse MR
744 analysis for a causal effect of AD on serum protein levels, together with the full multi-SNP IVW estimate
745 (red). Plots are shown for the five proteins with $FDR < 0.1$ in the primary reverse MR analysis, excluding
746 variants from the *APOE* locus.

747 **Supplementary Figure 11** – Comparison of the effects of *APOE-ε4* (left) versus AD, as evaluated through
748 reverse MR analysis excluding *APOE* variants, (right) on serum protein levels in AGES and two additional
749 cohorts. Results are shown for the four *APOE*-dependent proteins with opposite direction of effect in
750 the two analyses. The AGES results are shown for the full cohort and two age strata.

751 **Supplementary Table 1** - Baseline characteristics of the AGES cohort stratified by LOAD status. P values
752 are obtained from t-test for continuous variables, and chi-square test for categorical variables. P and FDR
753 values < 0.05 are highlighted in bold.

754 **Supplementary Table 2** – Proteins significantly associated with incident LOAD after FDR (Benjamini-
755 Hochberg) correction are displayed. The full Cox proportional hazards regression model (Model 3) was
756 adjusted for baseline age, sex, *APOE-e4* genotype, body mass index, diabetes, education, hypertension,
757 smoking history, and eGFR as denoted below. P and FDR values <0.05 are highlighted in bold.
758 Abbreviations: HR, Hazard ratio; CI, Confidence Interval, Zph: proportional hazard test statistic. Estimates
759 represent difference in LOAD risk per standard deviation increase of aptamer level.

760 **Supplementary Table 3** – Proteins significantly associated with incident LOAD with 10-year follow-up
761 after FDR (Benjamini-Hochberg) correction are displayed. The full Cox proportional hazards regression
762 model (Model 3) was adjusted for baseline age, sex, *APOE-e4* genotype, body mass index, diabetes,
763 education, hypertension, smoking history, and eGFR. P and FDR values <0.05 are highlighted in bold.
764 Abbreviations: HR, Hazard ratio; CI, Confidence Interval, Zph: proportional hazard test statistic. Estimates
765 represent difference in LOAD risk per standard deviation increase of aptamer level.

766 **Supplementary Table 4** – Comparisons of the Cox models with full follow-up vs 10 year follow up using
767 ANOVA. P and FDR values <0.05 are highlighted in bold. Abbreviations: df, degrees of freedom

768 **Supplementary Table 5** – Proteins significantly associated with prevalent LOAD after FDR (Benjamini-
769 Hochberg) correction. The full logistic regression model (Model 3) was adjusted for baseline age, sex,
770 *APOE-e4*, body mass index, diabetes, education, hypertension, smoking history, and eGFR. P and FDR
771 values <0.05 are highlighted in bold. Abbreviations: OR, odds ratio; SE, standard error. Estimates
772 represent difference in LOAD risk per standard deviation increase of aptamer level.

773 **Supplementary Table 6** – Replication of the LOAD associated proteins from Supplementary Table 5 in the
774 ACE cohort. Proteins significantly associated with prevalent LOAD after FDR (Benjamini-Hochberg)
775 correction are displayed. Logistic regression models were adjusted for baseline age + sex (Model 1) and
776 additionally for *APOE-e4* in Model 2. P and FDR values <0.05 are highlighted in bold. Abbreviations: OR,
777 odds ratio; SE, standard error. Estimates represent difference in LOAD risk per standard deviation
778 increase of aptamer level.

779 **Supplementary Table 7** – Results from Gene Set Enrichment Analysis (GSEA) and Overrepresentation
780 Analysis (ORA) of the prevalent LOAD associated proteins against protein-coding gene sets labeled below
781 using the R packages listed below. P values were adjusted for multiple comparisons using FDR (Benjamini
782 Hochberg). Abbreviations: ES, Enrichment score; NES, Enrichment score normalized

783 **Supplementary Table 8** – Gene Set Enrichment Analysis and Overrepresentation Analysis of incident
784 LOAD associated proteins in AGES against protein-coding gene sets labeled below using the R packages
785 listed below. P values were adjusted for multiple comparisons using FDR (Benjamini Hochberg). P and
786 FDR values <0.05 are highlighted in bold. Abbreviations: ES, Enrichment score; NES, Enrichment score
787 normalized; size of the pathway after removing missing genes; leadingEdge, vector of genes that drive
788 the enrichment.

789 **Supplementary Table 9** – Replication of the incident LOAD associated proteins from Supplementary
790 Table 2 in the ACE cohort. Proteins significantly associated with incident LOAD after FDR (Benjamini-
791 Hochberg) correction are displayed. Linear and Cox proportional regression models were adjusted for
792 baseline age, sex (Model 1) and additionally for APOE-e4 genotype in Model 2. P and FDR values <0.05
793 are highlighted in bold. Abbreviations: HR, Hazard ratio; CI, Confidence Interval, Zph: proportional hazard
794 test statistic. Estimates represent difference in LOAD risk per standard deviation increase of aptamer
795 level.

796 **Supplementary Table 10** – First-degree protein partners of the APOE-dependent proteins from the
797 InWeb database (Li et al. (2016), Nature Methods 2016 14:1 14, 61–64. 10.1038/nmeth.4083).
798 Abbreviations: CS, confidence score.

799 **Supplementary Table 11** – Gene Set Enrichment Analysis and Overrepresentation Analysis of the first-
800 degree protein partners of the APOE-dependent LOAD associated proteins against protein-coding gene
801 sets labeled below using the R packages listed below. P values were adjusted for multiple comparisons
802 using FDR (Benjamini Hochberg).

803 **Supplementary Table 12** – Forward MR analysis examining the causal effects of the protein partners of
804 the AGES-defined APOE-dependent LOAD-associated proteins on AD. P and FDR values <0.05 are
805 highlighted in bold. Abbreviations: OR, odds ratio; SE, standard error; eggerPass, passes Egger pleiotropy
806 sensitivity analysis; wmPass, passes weighted median sensitivity analysis

807 **Supplementary Table 13** – Reverse MR analysis examining the causal effects of blood cholesterol (total
808 and LDL) on the AGES-defined APOE-dependent LOAD-associated proteins. P and FDR values <0.05 are
809 highlighted in bold. Abbreviations: OR, odds ratio; SE, standard error.

810 **Supplementary Table 14** – Overview of the genetic instruments and their associations with serum levels
811 in AGES

812 **Supplementary Table 15** - Baseline characteristics of the ACE cohort stratified by LOAD status. P values
813 are obtained from t-test for continuous variables, and chi-square test for categorical variables. P and FDR
814 values <0.05 are highlighted in bold.

815 **Supplementary Table 16** – Forward MR analysis examining the causal effects of all LOAD-associated
816 proteins in AGES on AD. P and FDR values <0.05 are highlighted in bold. Abbreviations: OR, odds ratio;
817 SE, standard error; eggerPass, passes Egger pleiotropy sensitivity analysis; wmPass, passes weighted
818 median sensitivity analysis

819 **Supplementary Table 17** – Reverse MR analysis examining the causal effects of AD or its genetic liability
820 on all LOAD-associated proteins in AGES. The primary analysis was performed without the APOE locus
821 included, but results including the APOE locus are also shown. P and FDR values <0.05 are highlighted in
822 bold. Abbreviations: OR, odds ratio; SE, standard error; Q_pval, Cochran's Q p-value for instrument
823 heterogeneity

824 **Supplementary Table 18** – Replication of reverse MR analysis results in data from Ferkingstad et al. and
825 Sun et al, examining the causal effects of AD on all AGES-defined LOAD-associated proteins. The analysis
826 was performed with and without the APOE locus included for instrument selection. One protein, ARL2,
827 was not available in the Sun et al. data. P and FDR values <0.05 are highlighted in bold. Abbreviations: SE,
828 standard error.

829 **Supplementary Table 19** – Proteins associated with prevalent or incident LOAD in AGES that are also
830 associated with AD in either CSF or brain. AGES results are shown for all outcomes and models from
831 which significant proteins were obtained from. P and FDR values <0.05 are highlighted in bold.
832 Abbreviations: HR, Hazard ratio; CI, Confidence Interval; OR, Odds ratio; SE, Standard error.

833 **Supplementary Table 20** – Protein association with AD in brain when adjusting for APOE-e4 count (0,1,2)
834 in multinomial logistic regression. Data was obtained from Johnson et al, Nature Neuroscience 2022.
835 Proteins with any association to LOAD in AGES and available data are included in the table, and the HR
836 for incident LOAD in Model 1 and Model 2 are shown for comparison and used to evaluate directional
837 consistency. P and FDR values <0.05 are highlighted in bold. Abbreviations: OR, Odds ratio; SE, Standard
838 error; APOE, APOE-dependent proteins in AGES.

839 **Supplementary Table 21** – Serum, CSF and brain protein module membership for all proteins measured
840 in serum in the AGES study.

841 CSF modules were obtained from an extended analysis compared to Dammer et al, Alzheimer's Research
842 & Therapy 2022 (manuscript in preparation), and brain modules from Johnson et al., Nature
843 Neuroscience 2022. The same protein can occur in more than one module when measured by more than
844 one aptamer (serum and CSF), by different platforms (CSF) or when different peptides are quantified by
845 MS (CSF, brain)

846 **Supplementary Table 22** – Overlaps between protein modules (defined from serum, CSF or brain) and
847 proteins associated with AD (in serum, protein or brain) tested by Fisher's exact test. Serum modules
848 were additionally tested for overlap with proteins detected/measured in CSF and brain, irrespective of
849 their AD association. FDR was calculated using Benjamini-Hochberg. P and FDR values <0.05 are
850 highlighted in bold. Abbreviations: OR, odds ratio; CI, Confidence Interval.

851 **References**

- 852 1. Gatz, M. *et al.* Role of genes and environments for explaining Alzheimer disease. *Arch Gen*
853 *Psychiatry* **63**, 168–174 (2006).
- 854 2. Reitz, C., Rogaeva, E. & Beecham, G. W. Late-onset vs nonmendelian early-onset Alzheimer
855 disease: A distinction without a difference? *Neurol Genet* **6**, (2020).
- 856 3. Rajan, K. B. *et al.* Population Estimate of People with Clinical AD and Mild Cognitive Impairment in
857 the United States (2020–2060). *Alzheimers Dement* **17**, 1966 (2021).
- 858 4. van Dyck, C. H. *et al.* Lecanemab in Early Alzheimer's Disease. *New England Journal of Medicine*
859 **388**, 9–21 (2023).
- 860 5. Mintun, M. A. *et al.* Donanemab in Early Alzheimer's Disease. *New England Journal of Medicine*
861 **384**, 1691–1704 (2021).
- 862 6. Sattlecker, M. *et al.* Longitudinal Protein Changes in Blood Plasma Associated with the Rate of
863 Cognitive Decline in Alzheimer's Disease. *Journal of Alzheimer's Disease* **49**, 1105–1114 (2016).
- 864 7. Kiddle, S. J. *et al.* Candidate Blood Proteome Markers of Alzheimer's Disease Onset and
865 Progression: A Systematic Review and Replication Study. *Journal of Alzheimer's Disease* **38**, 515–
866 531 (2014).
- 867 8. Sattlecker, M. *et al.* Alzheimer's disease biomarker discovery using SOMAscan multiplexed
868 protein technology. *Alzheimer's & Dementia* **10**, 724–734 (2014).
- 869 9. O'Bryant, S. E. *et al.* A Serum Protein-Based Algorithm for the Detection of Alzheimer's Disease.
870 *Arch Neurol* **67**, 1077 (2010).

- 871 10. Ijsselstijn, L. *et al.* Serum levels of pregnancy zone protein are elevated in presymptomatic
872 Alzheimer's disease. *J Proteome Res* **10**, 4902–4910 (2011).
- 873 11. Ray, S. *et al.* Classification and prediction of clinical Alzheimer's diagnosis based on plasma
874 signaling proteins. *Nature Medicine* 2007 **13**:11 **13**, 1359–1362 (2007).
- 875 12. Henkel, A. W. *et al.* Multidimensional plasma protein separation technique for identification of
876 potential Alzheimer's disease plasma biomarkers: A pilot study. *J Neural Transm* **119**, 779–788
877 (2012).
- 878 13. Choi, J., Malakowsky, C. A., Talent, J. M., Conrad, C. C. & Gracy, R. W. Identification of oxidized
879 plasma proteins in Alzheimer's disease. *Biochem Biophys Res Commun* **293**, 1566–1570 (2002).
- 880 14. Cutler, P. *et al.* Proteomic identification and early validation of complement 1 inhibitor and
881 pigment epithelium-derived factor: Two novel biomarkers of Alzheimer's disease in human
882 plasma. *Proteomics Clin Appl* **2**, 467–477 (2008).
- 883 15. Hye, A. *et al.* Proteome-based plasma biomarkers for Alzheimer's disease. *Brain* **129**, 3042–3050
884 (2006).
- 885 16. Doecke, J. D. *et al.* Blood-Based Protein Biomarkers for Diagnosis of Alzheimer Disease. *Arch*
886 *Neurol* **69**, 1318 (2012).
- 887 17. Kiddle, S. J. *et al.* Candidate Blood Proteome Markers of Alzheimer's Disease Onset and
888 Progression: A Systematic Review and Replication Study. *Journal of Alzheimer's Disease* **38**, 515–
889 531 (2014).
- 890 18. Walker, K. A. *et al.* Large-scale plasma proteomic analysis identifies proteins and pathways
891 associated with dementia risk. *Nature Aging* 2021 **1**:5 **1**, 473–489 (2021).
- 892 19. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related
893 dementias. *Nature Genetics* 2022 **1–25** (2022) doi:10.1038/s41588-022-01024-z.
- 894 20. Khani, M., Gibbons, E., Bras, J. & Guerreiro, R. Challenge accepted: uncovering the role of rare
895 genetic variants in Alzheimer's disease. *Mol Neurodegener* **17**, 3 (2022).
- 896 21. Frisoni, G. B. *et al.* The prevalence of apoE-ε4 in Alzheimer's disease is age dependent. *J Neurol*
897 *Neurosurg Psychiatry* **65**, 103 (1998).
- 898 22. Gharbi-Meliani, A. *et al.* The association of APOE ε4 with cognitive function over the adult life
899 course and incidence of dementia: 20-year follow-up of the Whitehall II study. *Alzheimers Res*
900 *Ther* **13**, (2021).
- 901 23. Corder, E. H. *et al.* Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer's
902 Disease in Late Onset Families. *Science* (1979) **261**, 921–923 (1993).

- 903 24. Gudjonsson, A. *et al.* A genome-wide association study of serum proteins reveals shared loci with
904 common diseases. *Nature Communications* 2022 13:1 **13**, 1–13 (2022).
- 905 25. Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* 2018 558:7708 **558**, 73–79
906 (2018).
- 907 26. Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease.
908 *Science* **361**, (2018).
- 909 27. Emilsson, V. *et al.* Coding and regulatory variants are associated with serum protein levels and
910 disease. *Nature Communications* 2022 13:1 **13**, 1–11 (2022).
- 911 28. Cruchaga, C. *et al.* Proteogenomic analysis of human cerebrospinal fluid identifies neurologically
912 relevant regulation and informs causal proteins for Alzheimer’s disease. *Res Sq* (2023)
913 doi:10.21203/rs.3.rs-2814616/v1.
- 914 29. Yang, C. *et al.* Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins
915 implicated in neurological disorders. *Nat Neurosci* **24**, 1302–1312 (2021).
- 916 30. Harris, T. B. *et al.* Age, Gene/Environment Susceptibility – Reykjavik Study: Multidisciplinary
917 Applied Phenomics. *Am J Epidemiol* **165**, 1076 (2007).
- 918 31. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, (2015).
- 919 32. Li, T. *et al.* A scored human protein–protein interaction network to catalyze genomic
920 interpretation. *Nature Methods* 2016 14:1 **14**, 61–64 (2016).
- 921 33. Postmus, I. *et al.* Pharmacogenetic meta-analysis of genome-wide association studies of LDL
922 cholesterol response to statins. *Nat Commun* **5**, (2014).
- 923 34. Marucci, G. *et al.* Efficacy of acetylcholinesterase inhibitors in Alzheimer’s disease.
924 *Neuropharmacology* **190**, (2021).
- 925 35. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and disease.
926 *Nat Genet* **53**, 1712–1721 (2021).
- 927 36. Dammer, E. B. *et al.* Proteomic Network Analysis of Alzheimer’s Disease Cerebrospinal Fluid
928 Reveals Alterations Associated with APOE ϵ 4 Genotype and Atomoxetine Treatment. *medRxiv*
929 2023.10.29.23297651 (2023) doi:10.1101/2023.10.29.23297651.
- 930 37. Johnson, E. C. B. *et al.* Large-scale deep multi-layer analysis of Alzheimer’s disease brain reveals
931 strong proteomic disease-related changes not observed at the RNA level. *Nature Neuroscience*
932 2022 25:2 **25**, 213–225 (2022).
- 933 38. Dammer, E. B. *et al.* Multi-platform proteomic analysis of Alzheimer’s disease cerebrospinal fluid
934 and plasma reveals network biomarkers associated with proteostasis and the matrisome.
935 *Alzheimers Res Ther* **14**, (2022).

- 936 39. Porter, T. *et al.* Cognitive gene risk profile for the prediction of cognitive decline in
937 presymptomatic Alzheimer's disease. *Pers Med Psychiatry* **7–8**, 14–20 (2018).
- 938 40. Zhou, M. *et al.* Targeted mass spectrometry to quantify brain-derived cerebrospinal fluid
939 biomarkers in Alzheimer's disease. *Clin Proteomics* **17**, 1–14 (2020).
- 940 41. Lourenço, F. C. *et al.* Netrin-1 interacts with amyloid precursor protein and regulates amyloid-
941 beta production. *Cell Death Differ* **16**, 655–663 (2009).
- 942 42. Zetterberg, H. Neurofilament Light: A Dynamic Cross-Disease Fluid Biomarker for
943 Neurodegeneration. *Neuron* **91**, 1–3 (2016).
- 944 43. Graham, N. S. N. *et al.* Axonal marker neurofilament light predicts long-term outcomes and
945 progressive neurodegeneration after traumatic brain injury. *Sci Transl Med* **13**, (2021).
- 946 44. Wolters, F. *et al.* Von Willebrand Factor and the Risk of Dementia: A Population-Based Study
947 (P1.092). *Neurology* **86**, (2016).
- 948 45. Johnson, E. C. B. *et al.* Cerebrospinal fluid proteomics define the natural history of autosomal
949 dominant Alzheimer's disease. *Nat Med* (2023) doi:10.1038/s41591-023-02476-4.
- 950 46. Sattlecker, M. *et al.* Alzheimer's disease biomarker discovery using SOMAScan multiplexed
951 protein technology. *Alzheimer's & Dementia* **10**, 724–734 (2014).
- 952 47. Lindbohm, J. V. *et al.* Plasma proteins, cognitive decline, and 20-year risk of dementia in the
953 Whitehall II and Atherosclerosis Risk in Communities studies. *Alzheimer's & Dementia* **18**, 612
954 (2022).
- 955 48. Eldjarn, G. H. *et al.* Large-scale plasma proteomics comparisons through genetics and disease
956 associations. *Nature* **622**, 348–358 (2023).
- 957 49. Sebastiani, P. *et al.* A serum protein signature of APOE genotypes in centenarians. *Aging Cell* **18**,
958 e13023 (2019).
- 959 50. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, (2015).
- 960 51. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies,
961 targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005–D1012 (2019).
- 962 52. Jun, G. *et al.* A NOVEL ALZHEIMER DISEASE LOCUS LOCATED NEAR THE GENE ENCODING TAU
963 PROTEIN. *Mol Psychiatry* **21**, 108 (2016).
- 964 53. Okbay, A. *et al.* Polygenic prediction of educational attainment within and between families from
965 genome-wide association analyses in 3 million individuals. *Nat Genet* **54**, 437–449 (2022).
- 966 54. Brouwer, R. M. *et al.* Genetic variants associated with longitudinal changes in brain structure
967 across the lifespan. *Nature Neuroscience* 2022 25:4 **25**, 421–432 (2022).

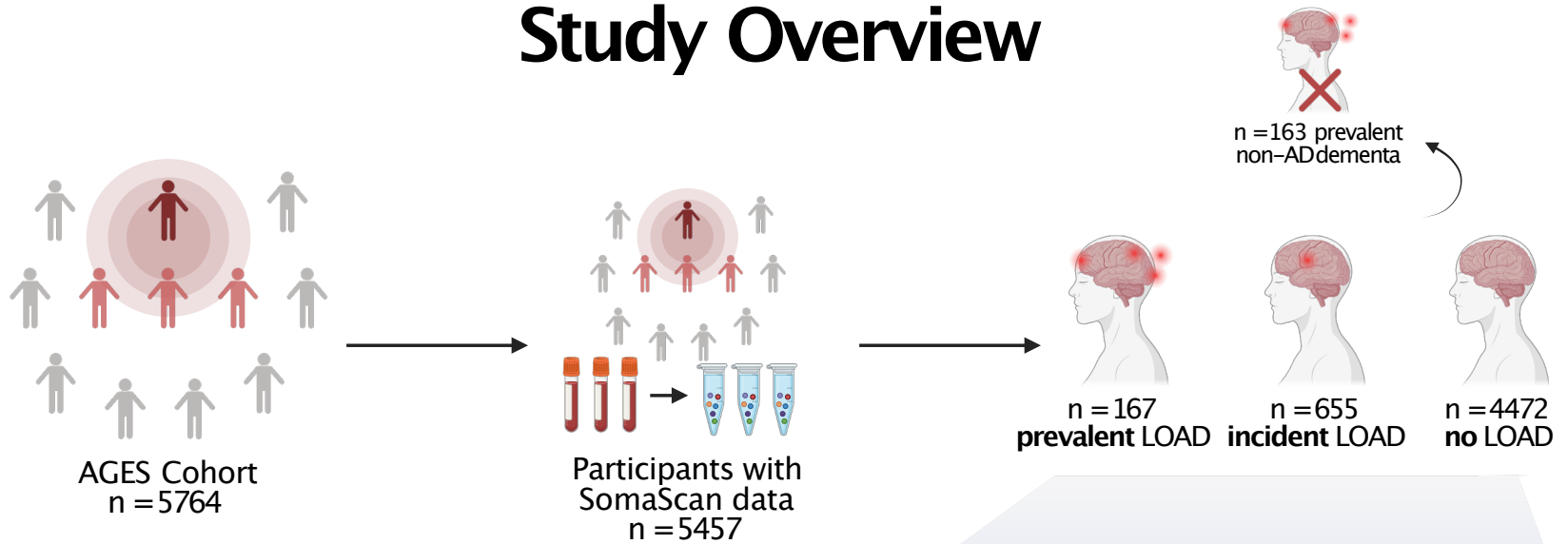
- 968 55. Davies, G. *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing
969 general cognitive function. *Nat Commun* **9**, (2018).
- 970 56. Wang, H. *et al.* Genome-wide interaction analysis of pathological hallmarks in Alzheimer's
971 disease. *Neurobiol Aging* **93**, 61–68 (2020).
- 972 57. Tin, A. *et al.* Proteomic Analysis Identifies Circulating Proteins Associated With Plasma Amyloid β
973 and Incident Dementia. *Biological Psychiatry Global Open Science* (2022)
974 doi:10.1016/J.BPSGOS.2022.04.005.
- 975 58. Tanzi, R. E. *et al.* Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near
976 the Alzheimer locus. *Science* **235**, 880–884 (1987).
- 977 59. Bai, Z. *et al.* Distinctive RNA Expression Profiles in Blood Associated with Alzheimer's Disease after
978 Accounting for White Matter Hyperintensities. *Alzheimer Dis Assoc Disord* **28**, 226–233 (2014).
- 979 60. Wang, D. *et al.* Cardiotrophin-1 (CTF1) ameliorates glucose-uptake defects and improves memory
980 and learning deficits in a transgenic mouse model of Alzheimer's disease. *Pharmacol Biochem*
981 *Behav* **107**, 48–57 (2013).
- 982 61. Rayaprolu, S. *et al.* Flow-cytometric microglial sorting coupled with quantitative proteomics
983 identifies moesin as a highly-abundant microglial protein with relevance to Alzheimer's disease.
984 *Mol Neurodegener* **15**, 28 (2020).
- 985 62. Chen, H.-H. *et al.* Genetically regulated expression in late-onset Alzheimer's disease implicates
986 risk genes within known and novel loci. *Transl Psychiatry* **11**, 618 (2021).
- 987 63. Sügis, E. *et al.* HENA, heterogeneous network-based data set for Alzheimer's disease. *Sci Data* **6**,
988 151 (2019).
- 989 64. Zetterberg, H. *et al.* Association of Cerebrospinal Fluid Neurofilament Light Concentration With
990 Alzheimer Disease Progression. *JAMA Neurol* **73**, 60 (2016).
- 991 65. Bernardini, S. *et al.* Glutathione S-Transferase P1 *C Allelic Variant Increases Susceptibility for
992 Late-Onset Alzheimer Disease: Association Study and Relationship with Apolipoprotein E ϵ 4 Allele.
993 *Clin Chem* **51**, 944–951 (2005).
- 994 66. Pinhel, M. A. S. *et al.* Glutathione S-transferase variants increase susceptibility for late-onset
995 Alzheimer's disease: association study and relationship with apolipoprotein E ϵ 4 allele. *Clin Chem*
996 *Lab Med* **46**, (2008).
- 997 67. Sigurdsson, S. *et al.* Incidence of Brain Infarcts, Cognitive Change, and Risk of Dementia in the
998 General Population: The AGES-Reykjavik Study (Age Gene/Environment Susceptibility-Reykjavik
999 Study). *Stroke* **48**, 2353–2360 (2017).

- 1000 68. Jørgensen, L. M., El Kholy, K., Damkjær, K., Deis, A. & Schroll, M. »RAI«- Et internationalt system
1001 til vurdering af beboere på plejehjem. *Ugeskr Laeger* **159**, 6371–6376 (1997).
- 1002 69. Gudnason V, S. J. S. L. H. S. S. G. Association of apolipoprotein E polymorphism with plasma levels
1003 of high density lipoprotein and lipoprotein(a), and effect of diet in healthy men and women.
1004 *NUTRITION METABOLISM AND CARDIOVASCULAR DISEASES* **3**, 136–141 (1993).
- 1005 70. Levey, A., Greene, T., Kusek, J. & Beck, G. A simplified equation to predict glomerular filtration
1006 rate from serum creatinine. *Journal of the American Society of Nephrology* **11**, 155A (2000).
- 1007 71. Gudmundsdottir, V. *et al.* Circulating Protein Signatures and Causal Candidates for Type 2
1008 Diabetes. *Diabetes* **69**, 1843 (2020).
- 1009 72. Lamb, J. R., Jennings, L. L., Gudmundsdottir, V., Gudnason, V. & Emilsson, V. It's in Our Blood: A
1010 Glimpse of Personalized Medicine. *Trends Mol Med* **27**, 20–30 (2021).
- 1011 73. Gold, L. *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery.
1012 doi:10.1371/journal.pone.0015004.
- 1013 74. Tuck, M. K. *et al.* Standard operating procedures for serum and plasma collection: Early detection
1014 research network consensus statement standard operating procedure integration working group.
1015 *J Proteome Res* **8**, 113–117 (2009).
- 1016 75. Jessen, F. *et al.* A conceptual framework for research on subjective cognitive decline in preclinical
1017 Alzheimer's disease. *Alzheimer's & Dementia* **10**, 844–852 (2014).
- 1018 76. Lopez, O. L. *et al.* Risk Factors for Mild Cognitive Impairment in the Cardiovascular Health Study
1019 Cognition Study. *Arch Neurol* **60**, 1394 (2003).
- 1020 77. Petersen, R. C. *et al.* Mild cognitive impairment: a concept in evolution. *J Intern Med* **275**, 214–
1021 228 (2014).
- 1022 78. Petersen, R. C. *et al.* Mild Cognitive Impairment. *Arch Neurol* **56**, 303 (1999).
- 1023 79. Jack, C. R. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer's
1024 disease. *Alzheimer's & Dementia* **14**, 535–562 (2018).
- 1025 80. Orellana, A. *et al.* Establishing In-House Cutoffs of CSF Alzheimer's Disease Biomarkers for the
1026 AT(N) Stratification of the Alzheimer Center Barcelona Cohort. *Int J Mol Sci* **23**, 6891 (2022).
- 1027 81. Rodriguez-Gomez, O. *et al.* FACEHBI: A PROSPECTIVE STUDY OF RISK FACTORS, BIOMARKERS AND
1028 COGNITION IN A COHORT OF INDIVIDUALS WITH SUBJECTIVE COGNITIVE DECLINE. STUDY
1029 RATIONALE AND RESEARCH PROTOCOLS. *J Prev Alzheimers Dis* 1–9 (2016)
1030 doi:10.14283/jpad.2016.122.

- 1031 82. Moreno-Grau, S. *et al.* Genome-wide association analysis of dementia and its clinical
1032 endophenotypes reveal novel loci associated with Alzheimer’s disease and three causality
1033 networks: The GR@ACE project. *Alzheimer’s & Dementia* **15**, 1333–1347 (2019).
- 1034 83. Vanderstichele, H. *et al.* Standardization of preanalytical aspects of cerebrospinal fluid biomarker
1035 testing for Alzheimer’s disease diagnosis: A consensus paper from the Alzheimer’s Biomarkers
1036 Standardization Initiative. *Alzheimer’s & Dementia* **8**, 65–73 (2012).
- 1037 84. Candia, J., Daya, G. N., Tanaka, T., Ferrucci, L. & Walker, K. A. Assessment of variability in the
1038 plasma 7k SomaScan proteomics assay. *Sci Rep* **12**, 17147 (2022).
- 1039 85. Therneau, T., Crowson, C. & Clinic, M. Using Time Dependent Covariates and Time Dependent
1040 Coefficients in the Cox Model. (2014).
- 1041 86. Gottesman, R. F. *et al.* Associations Between Midlife Vascular Risk Factors and 25-Year Incident
1042 Dementia in the Atherosclerosis Risk in Communities (ARIC) Cohort. *JAMA Neurol* **74**, 1246 (2017).
- 1043 87. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R Package for Comparing Biological
1044 Themes Among Gene Clusters. *OMICS* **16**, 284 (2012).
- 1045 88. Korotkevich, G. *et al.* Fast gene set enrichment analysis. *bioRxiv* 060012 (2021)
1046 doi:10.1101/060012.
- 1047 89. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for
1048 interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550 (2005).
- 1049 90. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues.
1050 *Science* **369**, 1318 (2020).
- 1051 91. Beach, T. G. *et al.* Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body
1052 Donation Program. *Neuropathology* **35**, 354–389 (2015).
- 1053 92. Bennett, D. A. *et al.* Religious Orders Study and Rush Memory and Aging Project. *Journal of*
1054 *Alzheimer’s Disease* **64**, S161–S189 (2018).
- 1055 93. Jack, C. R. *et al.* NIA-AA Research Framework: Toward a biological definition of Alzheimer’s
1056 disease. *Alzheimer’s & Dementia* **14**, 535–562 (2018).
- 1057 94. Burgess, S., Dudbridge, F. & Thompson, S. G. Combining information on multiple instrumental
1058 variables in Mendelian randomization: comparison of allele score and summarized data methods.
1059 *Stat Med* **35**, 1880 (2016).
- 1060

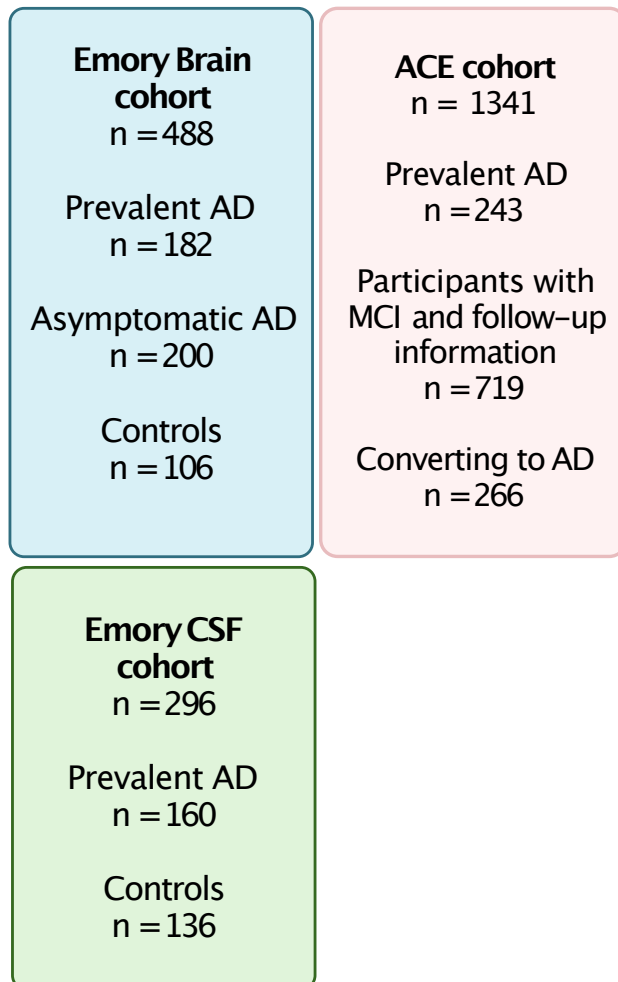
Study Overview

A



Replication cohorts

C



B

Somalogic serum proteins (n = 4782 aptamers)

Associations with LOAD status

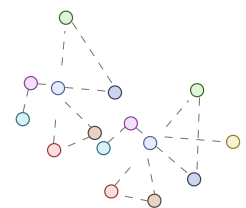
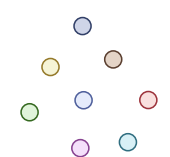
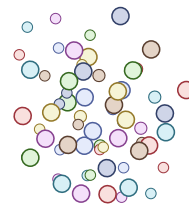
341 aptamers associated with incident LOAD

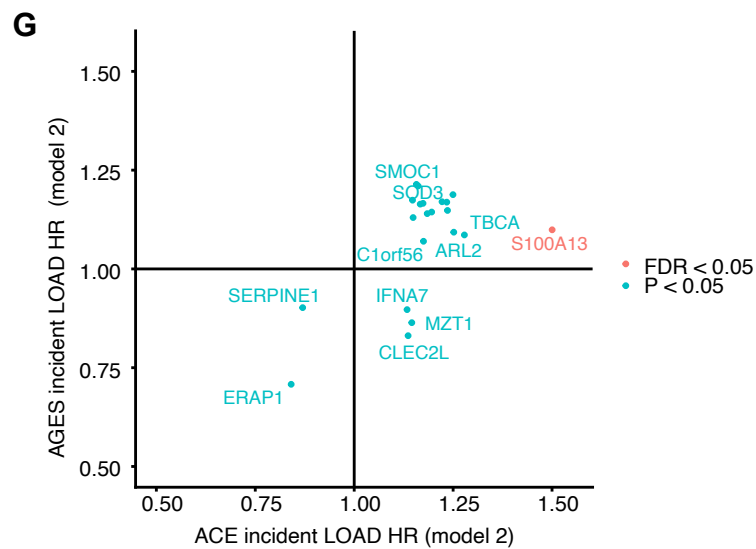
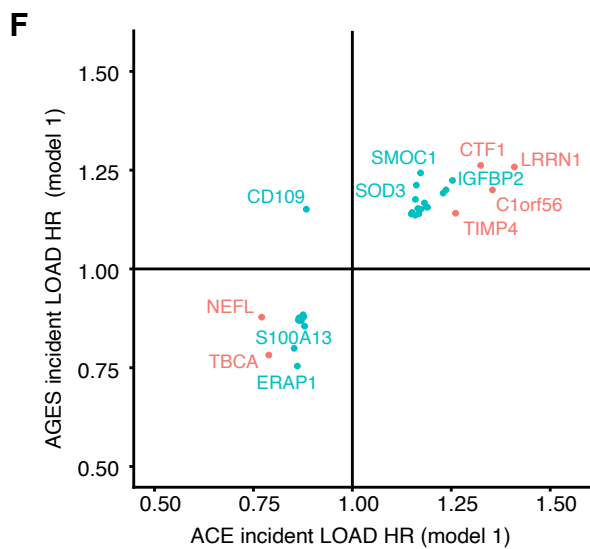
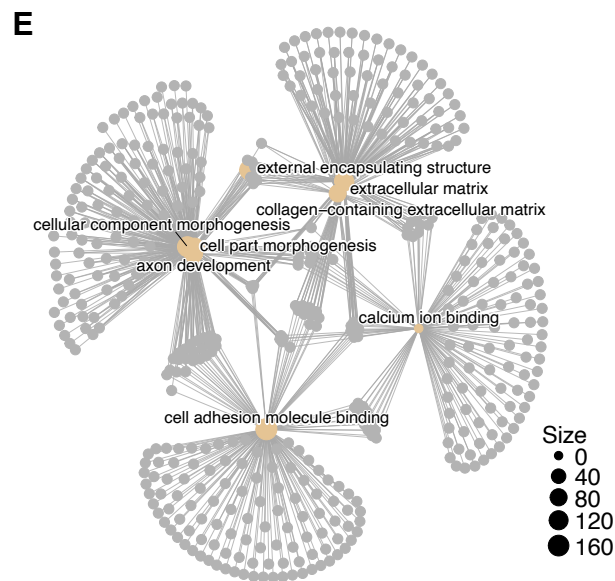
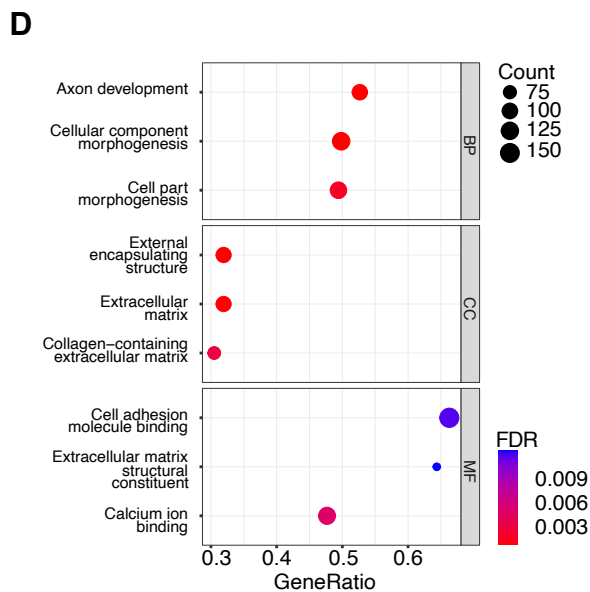
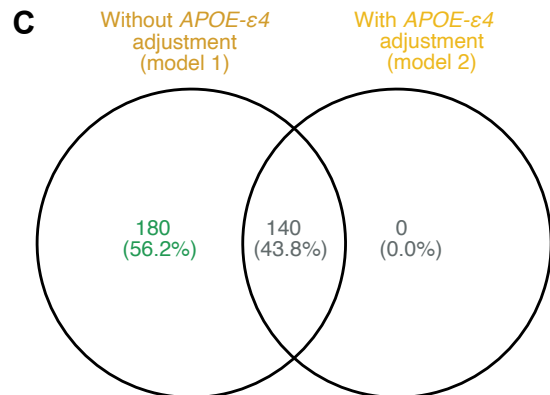
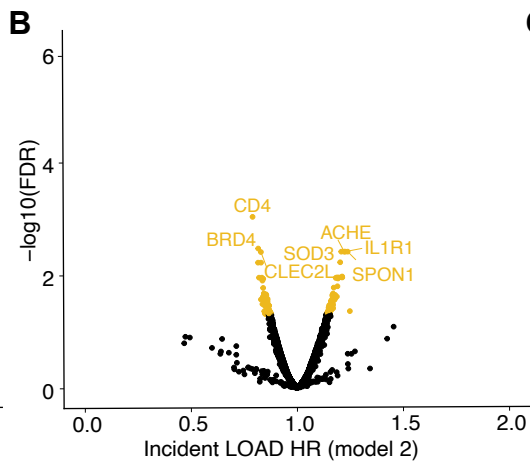
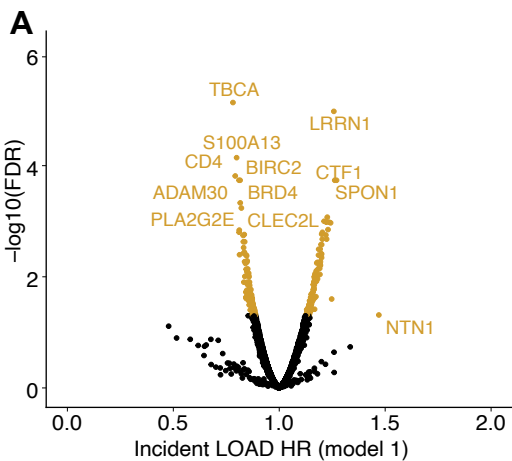
10 aptamers associated with prevalent LOAD

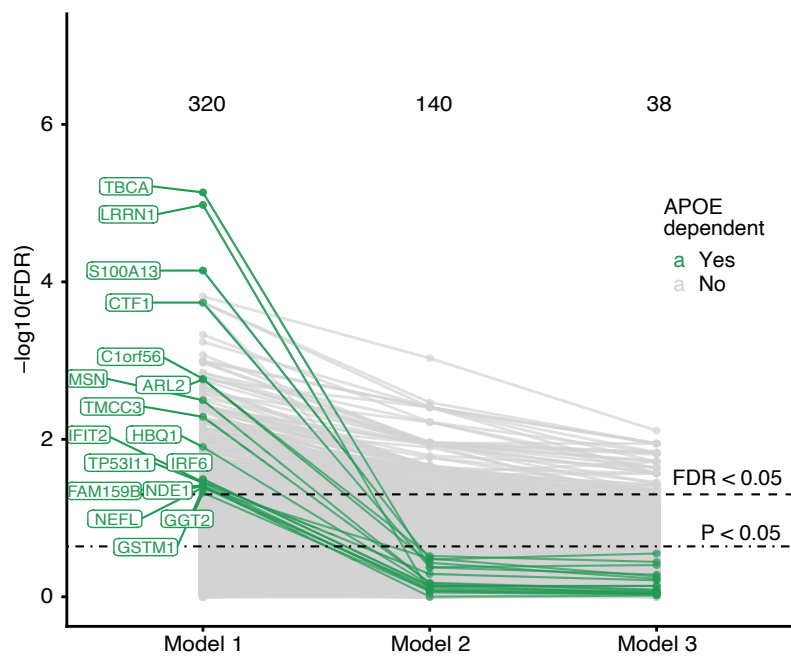
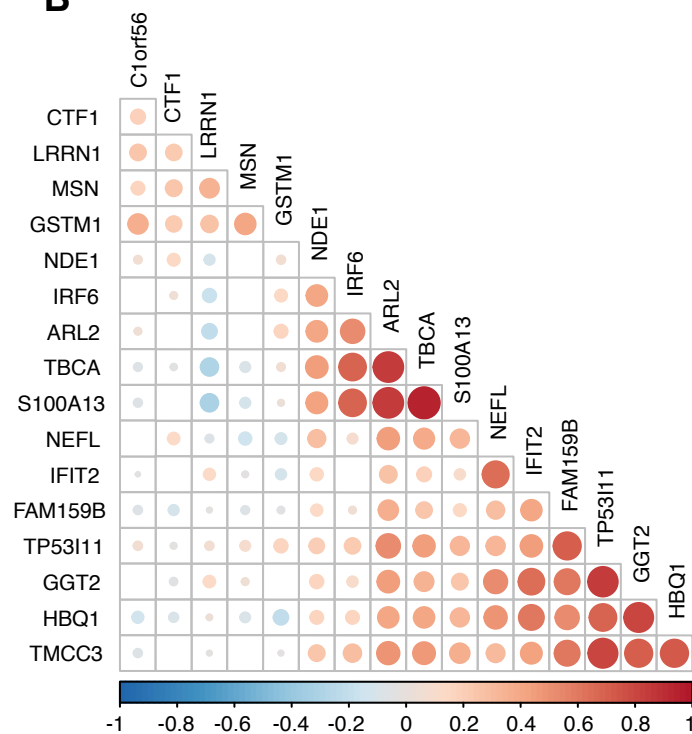
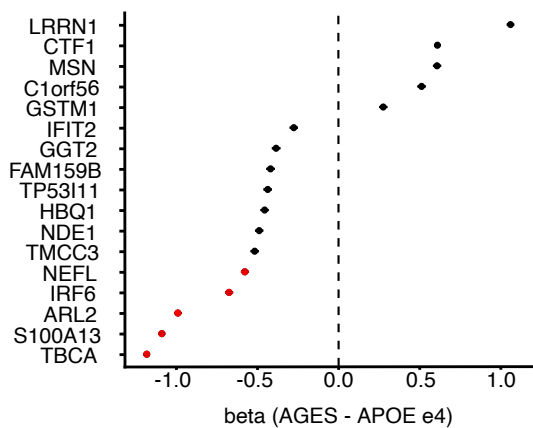
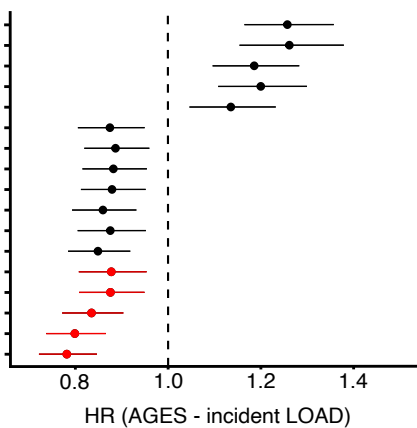
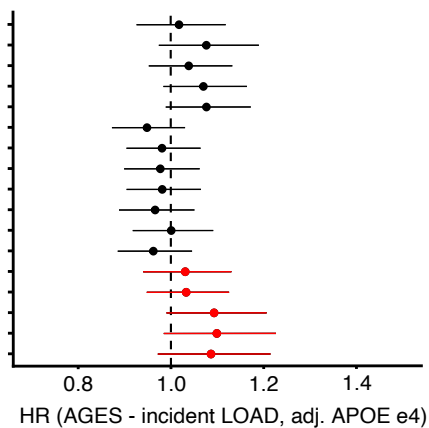
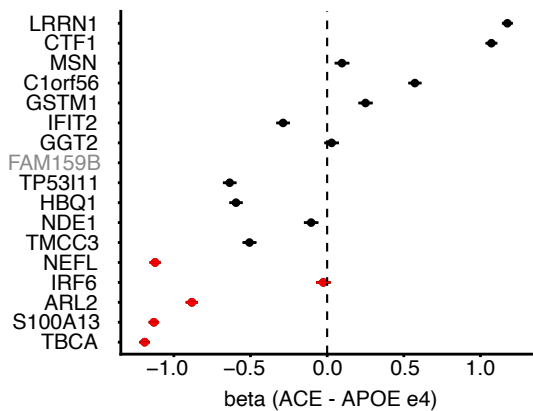
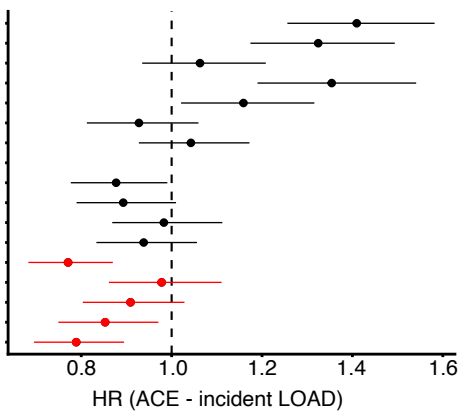
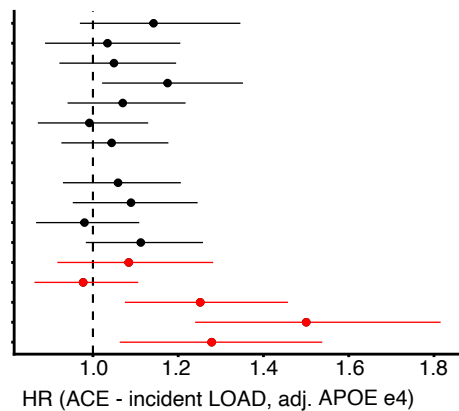
140 APOE independent proteins

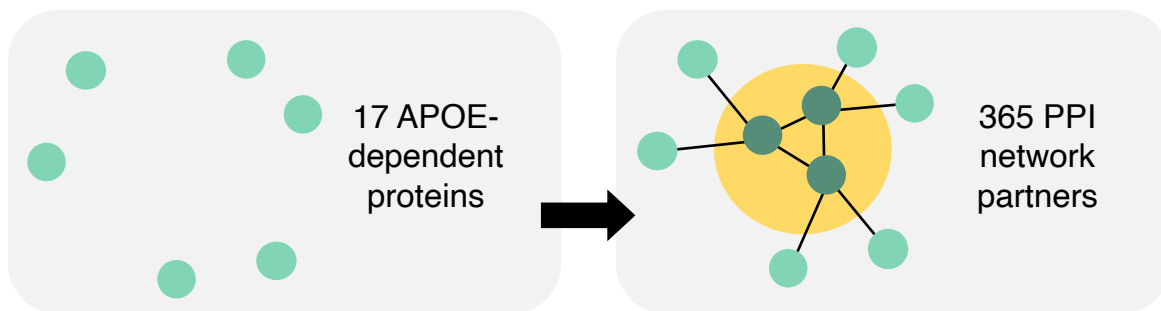
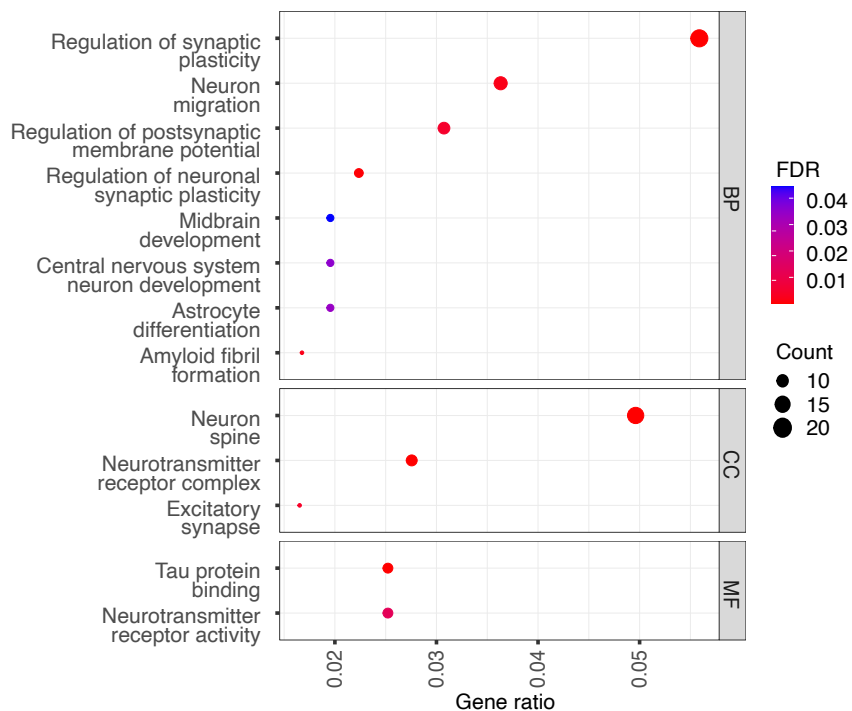
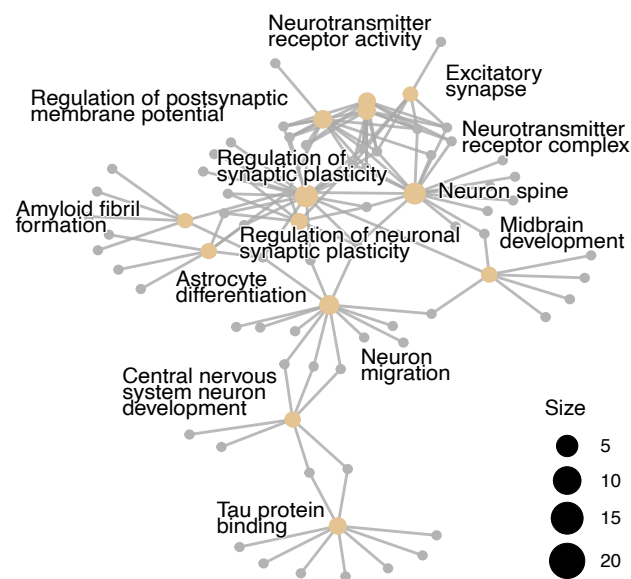
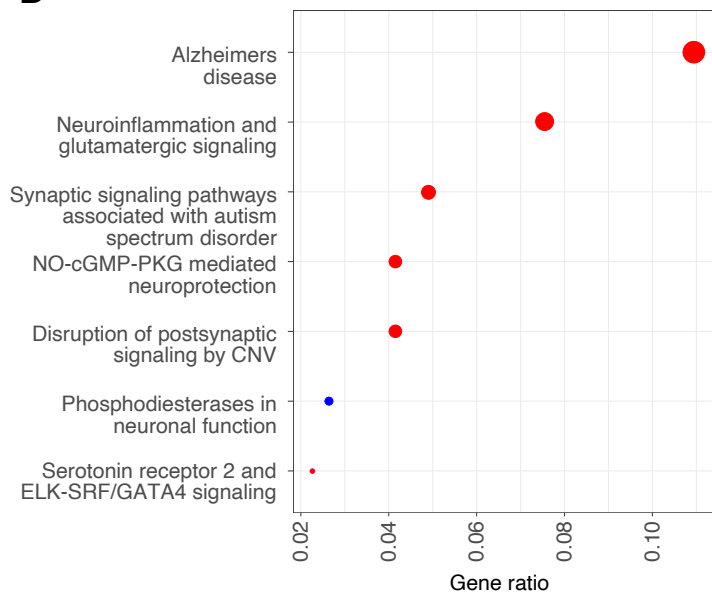
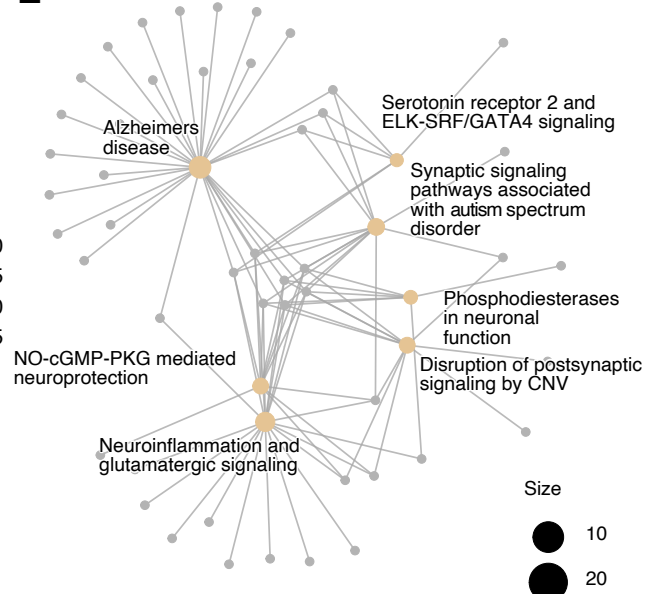
17 APOE dependent proteins

365 PPI network partners

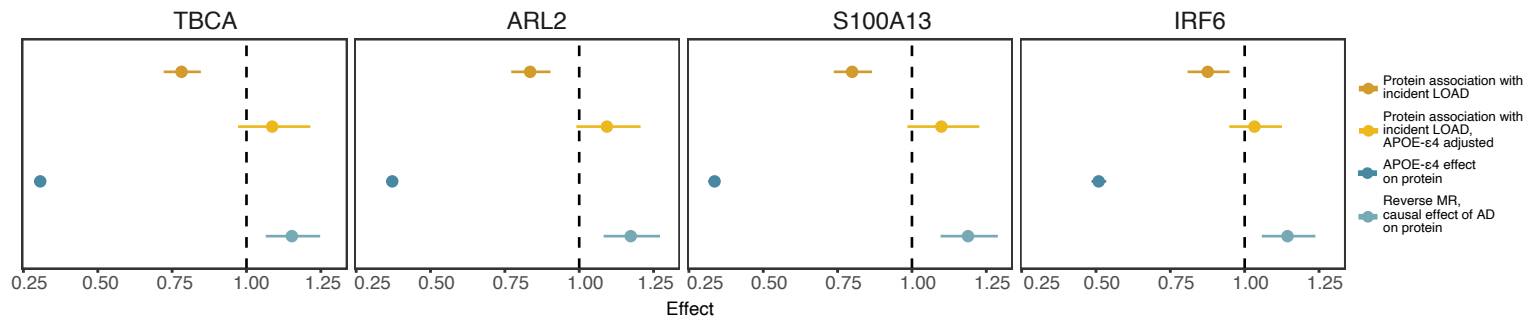




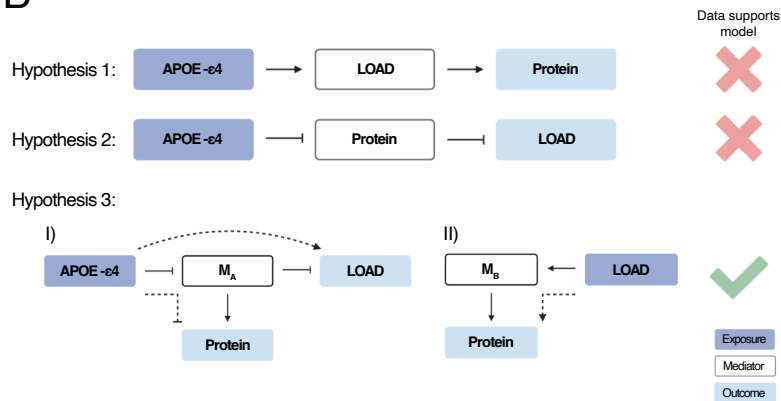
A**B****C****D****E****F****G****H**

A**B****C****D****E**

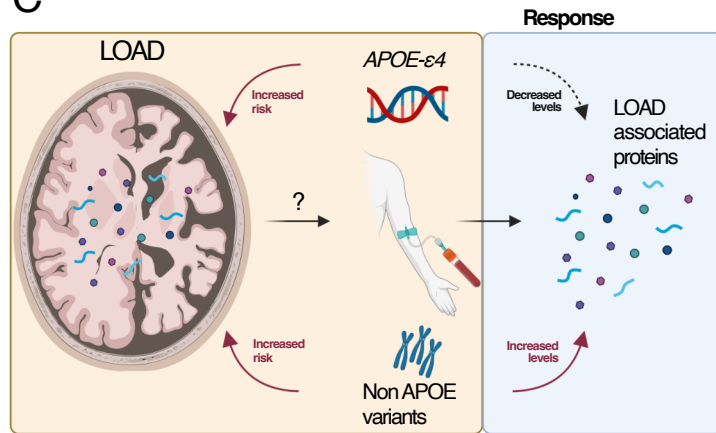
A



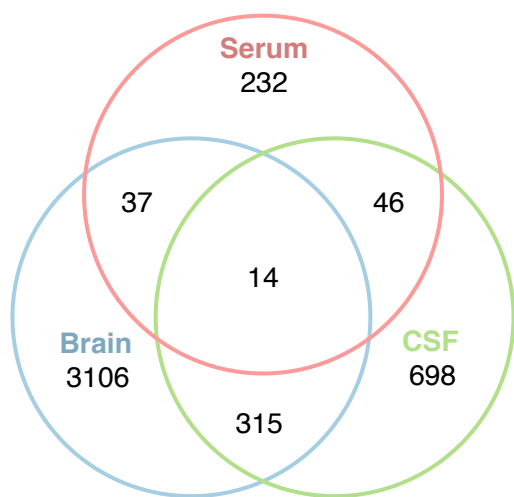
B



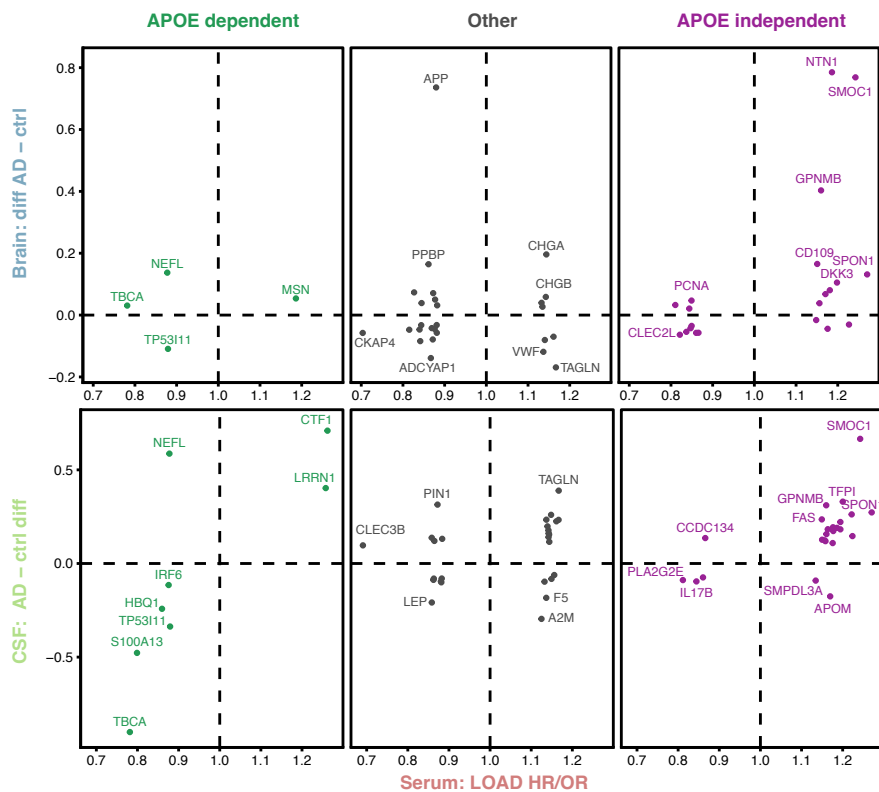
C



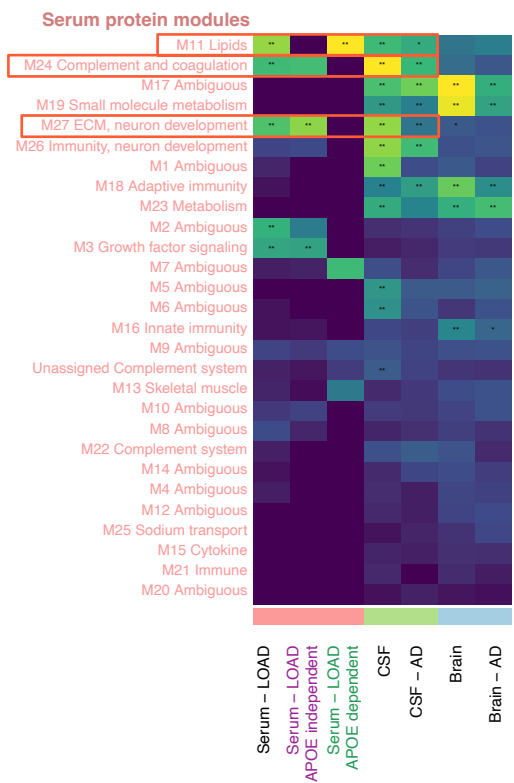
A



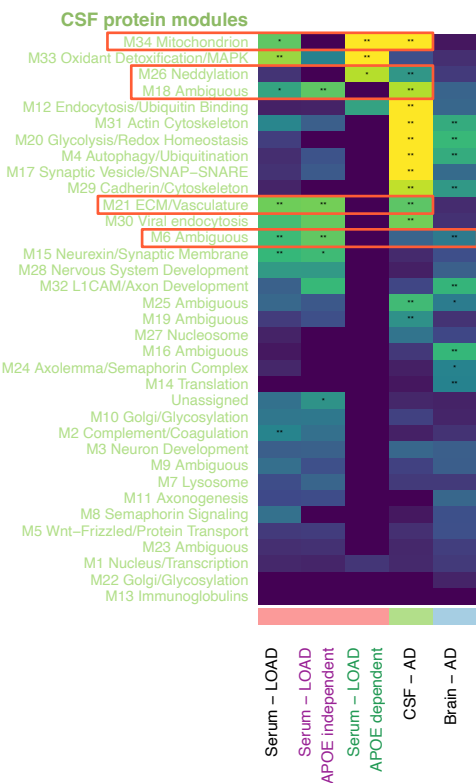
B



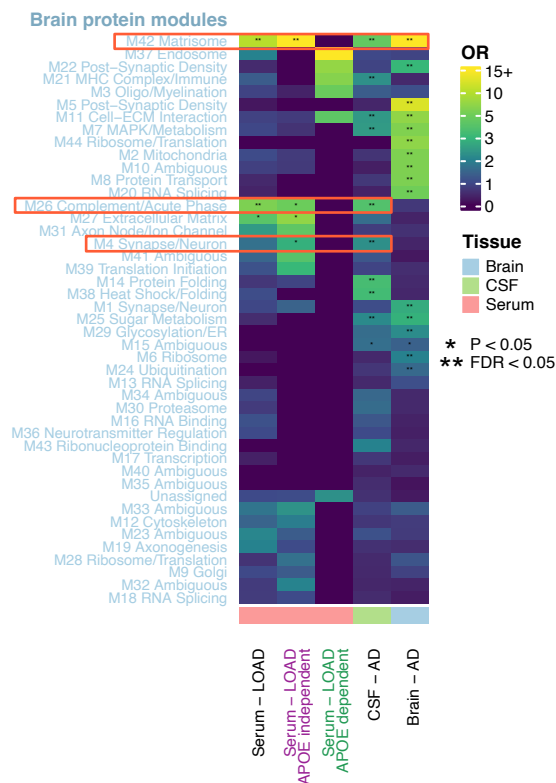
C

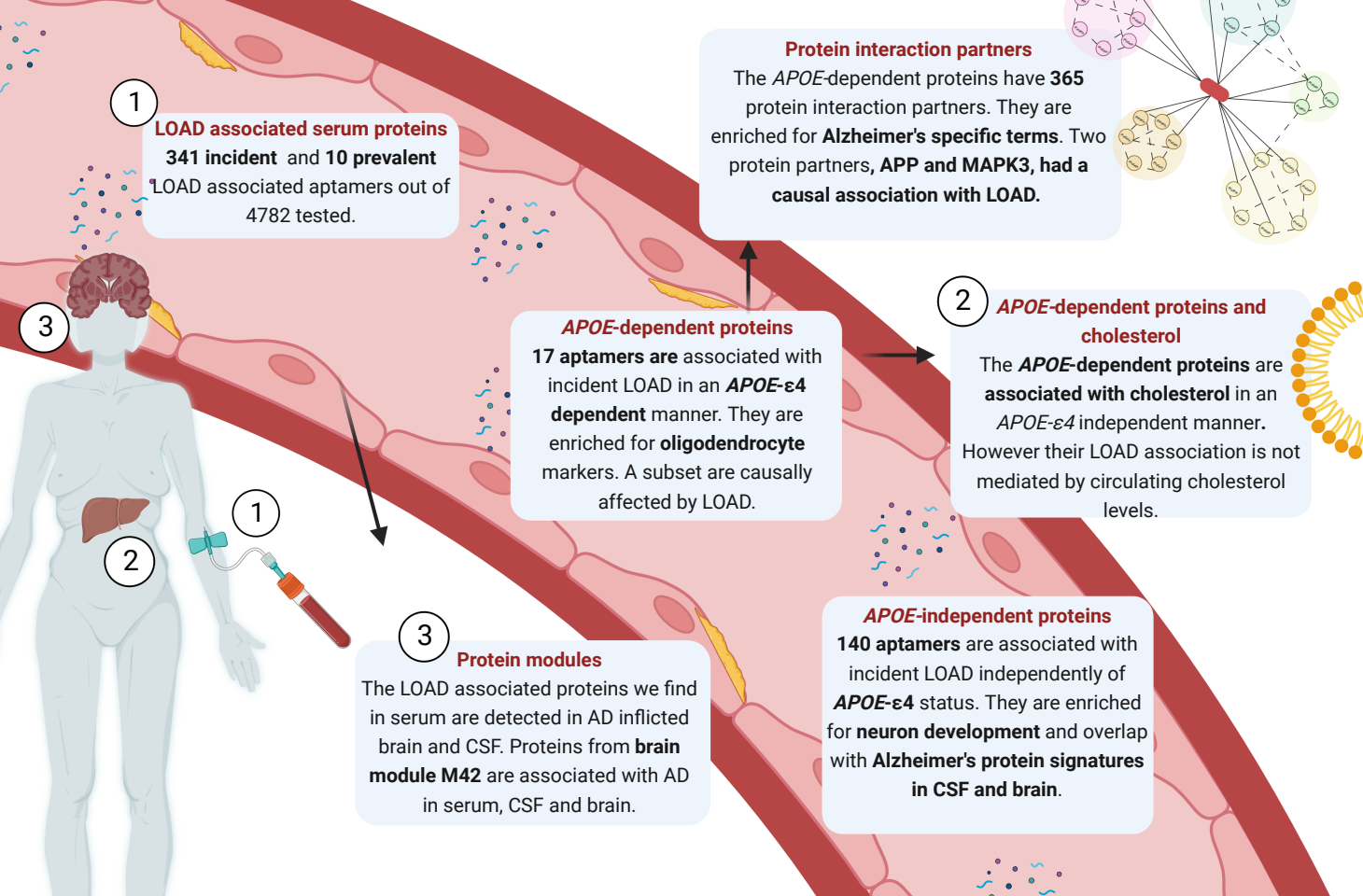


D



E





1 **LOAD associated serum proteins**
341 incident and 10 prevalent
LOAD associated aptamers out of
4782 tested.

Protein interaction partners
The *APOE*-dependent proteins have 365
protein interaction partners. They are
enriched for **Alzheimer's specific terms**. Two
protein partners, **APP and MAPK3**, had a
causal association with LOAD.

3

1

2

3 **Protein modules**
The LOAD associated proteins we find
in serum are detected in AD inflicted
brain and CSF. Proteins from **brain
module M42** are associated with AD
in serum, CSF and brain.

***APOE*-dependent proteins**
17 aptamers are associated with
incident LOAD in an ***APOE*- ϵ 4
dependent** manner. They are
enriched for **oligodendrocyte**
markers. A subset are causally
affected by LOAD.

2 ***APOE*-dependent proteins and
cholesterol**
The ***APOE*-dependent proteins** are
associated with cholesterol in an
APOE- ϵ 4 independent manner.
However their LOAD association is not
mediated by circulating cholesterol
levels.

***APOE*-independent proteins**
140 aptamers are associated with
incident LOAD independently of
***APOE*- ϵ 4** status. They are enriched
for **neuron development** and overlap
with **Alzheimer's protein signatures**
in CSF and brain.