Serum proteomics reveals APOE dependent and independent protein signatures in Alzheimer’s disease

Elisabet A. Frick¹, Valur Emilsson¹,², Thorarinn Jonmundsson², Anna E. Steindorssdottr², Erik C. B. Johnson³,⁴, Raquel Puerta⁶, Eric B. Dammer³,⁵, Anantharaman Shantaraman³,⁵, Amanda Cano⁶,⁷, Mercè Boada⁶,⁷, Sergio Valero⁶,⁷, Pablo García-González⁶,⁷, Elias F. Gudmundsson¹, Alexander Gudjonsson¹, Joseph J. Loureiro⁶, Anthony P. Orth⁹, Nicholas T. Seyfried⁴, Allan I. Levey³,⁴, Agustín Ruiz⁶,⁷, Thor Aspelund¹,², Lori L. Jennings⁸, Lenore J. Launer¹⁰, Valborg Gudmundsdottir*¹,², Vilmundur Gudnason*¹,²

¹ Icelandic Heart Association, Kopavogur, 200, Iceland
² Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland
³ Goizueta Alzheimer’s Disease Research Center, Emory University School of Medicine, Atlanta, 30329, GA, USA
⁴ Department of Neurology, Emory University School of Medicine, Atlanta, 30329, GA, USA
⁵ Department of Biochemistry, Emory University School of Medicine, Atlanta, 30329, GA, USA
⁶ Research Center and Memory Clinic. Ace Alzheimer Center Barcelona – Universitat Internacional de Catalunya, Barcelona, 08028, Spain, Barcelona.
⁷ CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, 28029, Spain.
⁸ Novartis Biomedical Research, Cambridge, 02139, MA, USA
⁹ Novartis Biomedical Research, San Diego, 10675, CA, USA
¹⁰ Laboratory of Epidemiology and Population Sciences, Intramural Research Program, National Institute on Aging, Bethesda, 20892, MD, USA

Correspondence: valborgg@hi.is, v.gudnason@hjarta.is

Summary

The current demand for early intervention, prevention, and treatment of late onset Alzheimer’s disease (LOAD) warrants deeper understanding of the underlying molecular processes which could contribute to biomarker and drug target discovery. Utilizing high-throughput proteomic measurements in serum from a prospective population-based cohort of older adults (n=5,294), we identified 303 unique proteins associated with incident LOAD (median follow-up 12.8 years). Over 40% of these proteins were associated with LOAD independently of APOE-ε4 carrier status. These proteins were implicated in neuronal processes and overlapped with protein signatures of LOAD in brain and cerebrospinal fluid. We found 17 proteins which LOAD-association was strongly dependent on APOE-ε4 carrier status. Most of them showed consistent associations with LOAD in cerebrospinal fluid and a third had brain-specific gene expression. Remarkably, four proteins in this group (TBCA, ARL2, S100A13 and IRF6) were downregulated by APOE-ε4 yet upregulated as a consequence of LOAD as determined in a bi-directional Mendelian randomization analysis, reflecting a potential response to the disease onset. Accordingly, the direct association of these proteins to LOAD was reversed upon APOE-ε4 genotype adjustment, a finding which we replicate in an external cohort (n=719). Our findings provide an insight into the dysregulated...
pathways that may lead to the development and early detection of LOAD, including those both
independent and dependent on \( \text{APOE-}\varepsilon4 \). Importantly, many of the LOAD-associated proteins we find in
the circulation have been found to be expressed - and have a direct link with AD - in brain tissue. Thus,
the proteins identified here, and their upstream modulating pathways, provide a new source of
circulating biomarker and therapeutic target candidates for LOAD.

Keywords: Alzheimer’s disease, APOE, Proteomics, Mendelian randomization, Longitudinal study, Cross-
sectional study, CSF, Brain, Network.

Introduction

Alzheimer’s disease (AD) is the most common cause of dementia, accounting for up to 80% of all
dementia cases\(^1\), of which late onset Alzheimer’s disease (LOAD) is most common\(^2\). As of 2022,
approximately 55 million individuals worldwide had dementia, representing 1 out of 9 people aged 65
and over\(^3\). While promising advances have been made in amyloid-targeting therapeutic options for early-
stage LOAD\(^4,5\), they still have limited benefit and identification of additional risk pathways that can be
used for early detection and intervention is highly needed. To meet these demands, a variety of
biologically relevant circulating molecules have been broadly associated with LOAD risk. The proteome in
particular has the potential to reveal circulating markers of disease-related molecular pathways from
different tissues, and studies assessing the circulating proteomic signatures between non-demented
older adults and individuals suffering from LOAD have been described\(^6-17\). Modest sample sizes, low-
throughput proteomics and lack of longitudinal measurements have, however, been limiting factors in
these studies. However, a recent large-scale longitudinal study identified promising blood-based markers
for all-cause incident dementia although it is unknown how specific the results are to LOAD\(^18\).
Information on the global circulating proteomic profile preceding the onset of LOAD, and how well it
reflects AD-related processes in brain and CSF, is thus scarce.

Alzheimer’s disease has a considerable genetic component, and both common\(^19\) and rare risk
variants have been identified\(^20\), of which the strongest effects are conferred by variants in the well-
known apolipoprotein E (\( \text{APOE} \)) gene. Approximately 25% of the general population carries the \( \text{APOE-}\varepsilon4 \)
variant while it is present in over 50% of AD cases\(^21,22\). The \( \text{APOE-}\varepsilon4 \) allele increases the risk of LOAD by
threelfold in heterozygous carriers and up to twelvefold in homozygous carriers\(^23\). Although the link
between the \( \varepsilon4 \) allele and LOAD has been extensively researched, light has yet to be shed on the precise
mechanism by which the \( \text{APOE} \) gene affects LOAD onset and/or progression. Importantly, recent large-
scale proteogenomic studies have consistently established the \textit{APOE} locus as a protein-regulatory hotspot, regulating levels of hundreds of proteins in both circulation\textsuperscript{24–27} and cerebrospinal fluid (CSF)\textsuperscript{28,29}. Yet, it remains unknown to what extent these proteins relate to LOAD and if they can provide new information on the mechanisms through which \textit{APOE}-\textepsilon{4} mediates risk. Identifying LOAD-associated circulatory proteins and whether their association is \textit{APOE}-dependent or independent is crucial for the understanding of AD more generally as well as for gaining insight into potential pathways suitable for targeting in personalized treatment.

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

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

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

As hazard ratio variability can arise with lengthy follow-up time, secondary analyses were implemented with a 10-year follow-up cut-off, which revealed mostly overlapping results (Supplementary Note 1, Supplementary Tables 3 and 4, Supplementary Figure 1). We did however detect protein associations specific to the shorter follow-up time, which potentially reflect processes that take place closer to the LOAD diagnosis. As there may be further differences in proteomic profiles depending on whether protein sampling occurred before or after LOAD diagnosis, we additionally considered the protein profile of the 167 individuals with prevalent LOAD at baseline (Supplementary Note 2, Supplementary Figure 2A-C, Supplementary Tables 5-7). Interestingly, many of the proteins associated with increased risk of incident LOAD showed the opposite direction of effect for prevalent LOAD, although generally not statistically significant (Supplementary Figure 2D). These contrasting results
suggest an important temporal element in the LOAD-associated proteome. In total there were 346 aptamers (329 unique proteins) associated with LOAD when all outcomes (incident and prevalent LOAD), follow-up times and models were considered (Supplementary Tables 2, 3, 5, 6, and 7).

To evaluate which biological processes are reflected by the overall incident LOAD-associated protein signature in AGES, we performed a gene set enrichment analysis (GSEA). The strongest enrichment for protein associations in model 1 was observed for gene ontology (GO) terms related to neuron development and morphogenesis (Figure 2H-I, Supplementary Table 8). The proteins driving the enrichment included Neural Cell Adhesion Molecules 1 and 2 (NCAM1, and NCAM2), Netrin 1 (NTN1), Contactin 1 (CNTN1), Neurexins 1-3 (NRXN1, NRXN2, NRXN3), Neuronal Differentiation Gene 1 (NDGA), Phosphatidylinositol-4-phosphate 5-kinase 1alpha (PI4K1A1), MatriX MetalloProteinase 2 (MMP2), Frizzled Family Member 7 (FZD7), and Cell Adhesion Molecule L1 like (CHL1). GSEA of the protein profiles of model 2, where APOE-ε4 carrier status was adjusted for, showed similar enrichment results (Supplementary Table 8), demonstrating that these terms were mainly driven by the APOE-ε4 independent component of the LOAD-associated protein profile.

Serum proteins with APOE-dependent association to incident LOAD. As previously mentioned, 43% of the protein associations with incident LOAD were independent of APOE-ε4. Of the remaining 57% that were affected by APOE-ε4 adjustment, we identified 17 proteins whose associations with incident LOAD were particularly strongly affected by APOE-ε4 carrier status (Table 2, Figure 3A, Supplementary Figure 3, Supplementary Table 2). These proteins, hereafter referred to as APOE-dependent proteins, were defined as proteins significantly (FDR < 0.05) associated with incident LOAD in model 1 but whose nominal significance was attenuated (P > 0.05) or direction of effect changed upon APOE-ε4 adjustment in model 2. These APOE-dependent proteins included those with the strongest associations to LOAD prior to adjusting for the APOE-ε4 allele (Figure 2A, D). The levels of the APOE protein itself were not associated with either incident or prevalent LOAD. Figure 3B shows the intra-correlations among the 17 APOE-dependent proteins. All the 17 APOE-dependent proteins were strongly regulated by the APOE-ε4 allele (Figure 3C, Table 2, Supplementary Figure 4, Supplementary Table 9), with the ε4 allele increasing the levels of five of the proteins and decreasing the levels of the other 12. Accordingly, we observed that increased levels of the five APOE-ε4 upregulated proteins and decreased levels of the 12 APOE-ε4 downregulated proteins were also associated with higher risk of LOAD, yielding a hazard ratio above and below one, respectively (Figure 3D). As per definition, most of the APOE-dependent proteins lost significance upon APOE-ε4 adjustment yet interestingly, the direction of effect inverted for five proteins after APOE-ε4 adjustment (ARL2, IRF6, NEFL, S100A13, TBCA) (Figure 3E).
The HR conferred by APOE-ε4 for incident LOAD in AGES was 2.1 (P = 1.23e-27) per each ε4 allele. To evaluate if any of the 17 APOE-dependent proteins might mediate the effect of APOE-ε4 on incident LOAD, we considered the change in HR for APOE-ε4 on risk of incident LOAD when adjusting for individual proteins. We found that adjustment for most proteins resulted in a minor effect decrease, suggesting they do not mediate the APOE-ε4 effect on LOAD (Supplementary Figure 5A). Intriguingly, however, the adjustment for four proteins (NEFL, ARL2, TBCA and S100A13) caused an increase of ~10% in APOE-ε4 effect size (Supplementary Figure 5A-B). Thus, the effect of APOE-ε4 on LOAD is partly masked by secondary opposing associations between these proteins and LOAD, which are further explored below. The effect of APOE-ε4 carrier status on LOAD risk was largely unchanged in a multivariable model containing all 17 APOE-dependent proteins, thus not supporting a mediating effect of these proteins for the LOAD risk conferred by APOE-ε4 (base model: HR = 2.1, P = 1.23e-27, multivariable model: HR = 2.2, P = 3.2e-10). However, although not direct mediators, the 17 proteins could be blood-based readouts of a true mediator within tissue-specific pathological processes occurring prior to LOAD diagnosis.

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

Given the well-established relationship between APOE and cholesterol we explored the potential effect of serum lipid levels on the association between LOAD and the 17 APOE-dependent proteins (Supplementary Table 13, Supplementary Figures 6-7, Supplementary Note 4). Our findings suggest that, while many of the APOE-dependent proteins are associated with cholesterol levels, it is not the driver of their link to LOAD.
External validation of protein associations with incident LOAD in the ACE cohort. We set to externally evaluate our observations in an independent cohort, the ACE - Alzheimer Center Barcelona (n=1,341), with SOMAscan platform (7k) measurements from plasma of individuals who were referred to the center. The longitudinal component of ACE consists of individuals who had been diagnosed with mild cognitive impairment (MCI) at the center and had been followed up. A total of 719 participants had follow-up information and 266 converted to LOAD over a median follow-up of 3.14 years (Supplementary Table 14). Despite the fundamentally different cohorts, with AGES being population-based and using the 5K SOMAscan platform and ACE being based on individuals with established symptoms and the 7K SOMAscan platform, we replicated 36 protein associations with LOAD at nominal significance (P < 0.05) in the smaller ACE cohort (Table 3, Figure 2F-G). Of those, 30 proteins were nominally significant in model 1 with 97% being directionally consistent with the observations in AGES (Figure 2F). In model 2, 21 proteins were nominally significant, 86% of which were directionally consistent (Figure 2G). After multiple testing correction, seven proteins remained statistically significant (FDR < 0.05), all of which were directionally consistent (Table 3, Figure 2F-G). Six were statistically significant (FDR < 0.05) in model 1 (NEFL, LRRN1, TBCA, CTF1, C1orf56 and TIMP4) and one in model 2 (S100A13) (Supplementary Table 9). Of all 332 tested aptamers, 213 (64%) were directionally consistent regardless of significance in model 1 (Exact binomial test P = 2.0e-05) and 202 (61%) were directionally consistent in model 2 (Exact binomial test P = 0.002), demonstrating an enrichment for consistency in direction of effect. The protein associations replicated in the ACE cohort are of particular interest as they represent potentially clinically relevant candidates for LOAD that are consistent in two different contexts, in both a general population and a clinically derived symptomatic sample set. However, our results suggest that many of the proteins that associate with long-term LOAD risk are not strongly associated with the conversion from MCI to AD, which is further into the AD trajectory and may also explain the limited overlap between the proteins associated with prevalent and incident LOAD in AGES.

External validation of reversed LOAD association conditional on APOE-ε4 for a subset of proteins. Specifically considering the APOE-dependent proteins, the association between the APOE-ε4 allele and the proteins was replicated for 13 of 17 proteins in the ACE cohort (Figure 3F). Furthermore, the change in direction of effect for incident LOAD upon APOE-ε4 adjustment was replicated in the ACE cohort for 4 of 5 proteins (ARL2, NEFL, S100A13 and TBCA) (Figure 3G-H) (Supplementary Table 9), with even larger effects observed in the ACE cohort compared to AGES in the APOE-ε4 adjusted model and three proteins (ARL2, S100A13 and TBCA) becoming statistically significant (P < 0.05). Thus, the attenuation of the primary LOAD associations for these proteins upon APOE-ε4 adjustment meet the criteria of APOE-ε4
dependence (see Methods). No significant interaction between protein and APOE-ε4 carrier status on AD risk was observed in either the AGES or ACE cohorts. Taken together, our results show that these proteins are strongly downregulated by APOE-ε4, and consequently show an inverse relationship with incident LOAD, but when adjusting for the APOE-ε4 allele, their association to LOAD is still significant but reversed – suggesting a secondary non-APOE-ε4-mediated process affecting these same proteins in relation to LOAD in the opposite direction that is more strongly observed a cohort of individuals with MCI than in the population-based AGES cohort.

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

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

As most of the observational protein associations in the current study were detected for incident LOAD, and thus reflect changes that take place before the onset of clinically diagnosed disease, it is unlikely that their levels and effects are direct downstream consequences of the disease after it reaches a clinical stage. However, they may reflect a response to a prodromal stage of the disease. We therefore
performed a reverse MR to test if the observed changes in serum protein levels are likely to occur downstream of the genetic liability to LOAD, which may capture processes both at the prodromal and clinical stage. The APOE locus is likely to have a dominant pleiotropic effect in the reverse MR analysis (Supplementary Table 17, Supplementary Figure 9, Supplementary Note 5), as it has a disproportionately strong effect on LOAD risk compared to all other common genetic variants, while also being a well-established pQTL trans-hotspot, affecting circulating levels of up to hundreds of proteins\textsuperscript{24,25,27}. We therefore performed the primary reverse MR analysis using only LOAD-associated genetic variants outside of the APOE locus as instruments. We found two proteins (S100A13 and ARL2) that were significantly (FDR < 0.05) affected by LOAD or its genetic liability (Supplementary Table 17, Supplementary Figures 9-10). Interestingly, both were among the 17 previously identified APOE-dependent LOAD proteins, together with two additional proteins that were nominally significant in the reverse MR (TBCA, P = 4.4e-4, FDR = 0.051 and IRF6, P = 7.9e-4, FDR = 0.055). Thus intriguingly, these findings suggest that these four proteins are upregulated by LOAD, in contrast to the observed APOE-\varepsilon4 downregulation of the same proteins (Figure 5). This supports our findings of competing biological effects described above (Figure 3E, Supplementary Figure 5) and collectively our results indicate that simultaneous opposing effects of APOE-\varepsilon4 on one hand and LOAD on the other result in differential regulation of these proteins in serum (Figure 5B).

We performed a replication analysis of the effect of APOE-\varepsilon4 on protein levels and the reverse MR results for these four proteins using published protein GWAS summary statistics from two recent studies\textsuperscript{25,35}. In the external datasets, the downregulation of all four proteins by APOE-\varepsilon4 (as determined by the rs429358 C allele) was replicated. In the reverse MR analysis (excluding the APOE locus), the upregulation of protein levels by LOAD liability observed in AGES was also detected for two proteins (S100A13 and TBCA) in both validation cohorts, reaching significance (P < 0.05) in the study by Ferkingstad et al. (Supplementary Figure 11, Supplementary Table 18). While the two proteins changed direction in a similar manner as in AGES, the effect size was considerably smaller in the validation cohorts. Importantly, however, individuals in these two cohorts are much younger than those in AGES, with mean ages of 55 and 48 years for the Ferkingstad et al. and Sun et al. studies, respectively, compared to 76 years in AGES. Therefore, we conducted an age-stratified reverse MR analysis in AGES, which showed a strong age-dependent effect, with a much larger effect of LOAD genetic liability on protein levels in individuals over 80 years old compared to those younger than 80 years (Supplementary Figure 11). The effect size in AGES individuals under 80 years old was in line with the effect observed in the validation cohorts. Thus, if the upregulation of these proteins reflects a response to prodromal or
preclinical LOAD, an older cohort may be needed to detect an association of the same degree as we found in AGES. However, the observed support in the validation cohorts for the discordant effects of APOE vs non-APOE LOAD-associated genetic variants on the same serum proteins strongly implicates these proteins as directly relevant to LOAD, potentially as readouts of biological processes that are both disrupted by APOE-ε4 and modulated in the opposite manner as a response to genetic predisposition to LOAD or the disease onset in general.

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

Overlap with the AD brain and CSF proteome. To evaluate to what extent our LOAD-associated serum proteins reflect the proteomic profile of AD in relevant tissues, we queried data from recent proteomic studies of AD in cerebrospinal fluid (CSF) and brain which also describe tissue specific co-regulatory modules. We observed that of our LOAD-associated serum proteins, 51 proteins were also associated with AD in brain as measured by mass-spectrometry, with 32 (63%) being directionally consistent (Figure 6A-B) (Supplementary Tables 19-20). Higher directional consistency was observed within the APOE-independent protein group, or 15 (71%) of 21 proteins associated with AD in brain tissue. Additionally, 60 proteins were directly associated with AD in CSF as measured with SOMAscan (7k) (Figure 6A) with 46 (77%) being directionally consistent (Figure 6B). The proportion of directionally consistent associations between serum and CSF was higher in both the APOE-independent and dependent protein groups, or 88% (22 of 25 and 7 of 8 for APOE-independent and dependent proteins, respectively) (Figure 6B, Supplementary Table 19). However, directional inconsistency between plasma and CSF AD proteomic profiles has been reported before in a similar comparison. Fourteen proteins overlapped between all three tissues in the context of AD (Figure 6A) (Supplementary Table 19). Many of these proteins have established links or are highly relevant to LOAD, such as Spondin 1 (SPON1), involved in the processing of
amyloid precursor protein (APP)\textsuperscript{39}; Secreted Modular Calcium-Binding Protein 1 (SMOC1) previously proposed as a biomarker of LOAD in postmortem brains and CSF\textsuperscript{40}; Netrin-1 (NTN1), an interactor of APP and regulator of amyloid-beta production\textsuperscript{41}; Neurofilament light (NEFL), previously proposed as a plasma biomarker for LOAD and axon injury\textsuperscript{42,43} and Von Willebrand factor (VWF), known for its role in blood clotting and associations with LOAD\textsuperscript{44} (Supplementary Table 19). Notably, some of the APOE-dependent proteins were associated with AD across all three tissues such as TBCA and TP53I11.

We have previously described the co-regulatory structure of the serum proteome, which can broadly be defined as 27 modules of correlated proteins\textsuperscript{26} (Supplementary Table 21). In the current study we found that among the 346 aptamers (329 proteins) associated with LOAD (prevalent or incident, any model), five serum protein modules (M27, M3, M11, M2 and M24) were overrepresented (Figure 6C, Supplementary Table 22). In particular, the 140 APOE-independent aptamers were specifically overrepresented in module M27, enriched for proteins involved in neuron development and the extracellular matrix, and module M3 that is associated with growth factor signaling pathways (Supplementary Table 22). By contrast, the 17 APOE-dependent proteins were specifically enriched in protein module M11 (Supplementary Table 22), which is strongly enriched for lipid pathways and is under strong genetic control of the APOE locus\textsuperscript{26}. Serum modules M27, M24 and M11 were all enriched for AD-associations in CSF (Figure 6C). We next sought to understand to what extent our LOAD-associated proteins identified in serum might reflect AD protein signatures in CSF and brain tissue. Among the LOAD-associated proteins measured in serum, we found the APOE-dependent and APOE-independent proteins to be enriched in different CSF modules, most of which were also linked to AD (Figure 6D, Supplementary Table 22). In brain tissue, the serum APOE-independent LOAD proteins were particularly enriched in brain module M42 (Matrisome), which is enriched for extracellular matrix (ECM) proteins\textsuperscript{37}. Strikingly, M42 was strongly enriched for the AD-proteomic profiles of all three tissues (Figure 6E, Supplementary Table 22). Interestingly, members of this module (SMOC1, APP, SPON1, NTN1, GPNMB) showed some of the strongest associations in serum to incident LOAD in our study (Figure 2D, Supplementary Table 2) as well as in brain (Figure 6B, Supplementary Table 22). This module has furthermore been demonstrated to be correlated with amyloid beta (Aβ) deposition in the brain and some of its protein constituents (e.g MDK, NTN1 and SMOC1) have been shown to colocalize with and bind to Aβ\textsuperscript{37}. Additionally, the APOE locus regulates M42 levels in the brain (mod-qTL), and while the APOE protein is a member of module M42, this regulation was found to not be solely driven through the levels of the APOE protein itself\textsuperscript{37}. Our results simultaneously show that other members of the module, such as SPON1 and SMOC1, exhibit an APOE-independent association to incident LOAD in serum.
Interestingly, these same two proteins are increased in CSF thirty years prior to symptom onset in autosomal dominant early onset AD\(^45\). In summary, we demonstrate significant overlaps in LOAD-associated protein expression across blood, CSF and brain on both an individual protein level and on protein module level.

**Discussion**

We describe a comprehensive mapping of the serum protein profile of LOAD that provides insight into processes that are independent or dependent on the genetic control of \(APOE-\varepsilon4\) (Figure 7). We identified 329 proteins in total that differed in the incident or prevalent LOAD cases compared to non-LOAD participants in a population-based cohort with long-term follow-up. Among these, we identified a novel grouping of proteins based on their primary LOAD-association being statistically independent of (130 proteins), or dependent on (17 proteins) \(APOE-\varepsilon4\) carrier status. Many of the \(APOE\)-independent proteins are implicated in neuronal pathways and are shared with the LOAD-associated CSF and brain proteome. The 17 \(APOE\)-dependent proteins overlap with \(AD\)-associated protein modules in CSF and interact directly with protein partners involved in LOAD, including APP. Another key finding is, amongst these 17 proteins, four proteins change LOAD-associated direction of effect both observationally and genetically when taking \(APOE-\varepsilon4\) carrier status into account. Collectively, our results suggest that while their primary association with LOAD reflects the risk conferred by \(APOE-\varepsilon4\), there exists a secondary causal effect of LOAD itself on the protein levels in the reverse direction as supported by the MR analysis, possibly reflecting a response to the disease onset.

Previous studies identifying proteins associated with LOAD have been limited to a cross-sectional cohort or are based on all-cause dementia\(^{18,46-48}\). Here we extend those findings by distinguishing LOAD cases from other types of dementia in a prospective cohort study to identify LOAD-specific serum protein signatures preceding clinical onset. Furthermore, our comparative approach of statistical models with and without \(APOE-\varepsilon4\) adjustment provides a novel compartmentalized view of the LOAD serum protein profile and demonstrates how protein effects can differ depending on genetic confounders which are imperative to take into consideration. We found that the proteins associated with incident LOAD in our study, in particular those independently of \(APOE-\varepsilon4\) such as GPNMB, NTN1, SMOC1 and SPON1, overlap with the proteomic profile of LOAD in CSF\(^38\) and brain\(^37\), and are enriched for neuronal pathways, which may reflect an altered abundance of neuronal proteins in the circulation during the prodromal stage of LOAD. These overlaps that we find across independent cohorts and different proteomics technologies suggest that the serum levels of some proteins have a direct link to the biological systems involved in
LOAD pathogenesis and may even provide a peripheral readout of neurodegenerative processes prior to clinical diagnosis of LOAD. In particular, the proteins that show directionally consistent effect sizes suggest exceptional AD-specific robustness as the measurements vary by tissue, methodology and populations.

We identified 17 proteins with a particularly strong APOE-dependent association to incident LOAD, of which eight were also associated with prevalent AD in CSF. The association between APOE-ε4 and circulating levels of these proteins has been reported by our group and others, but their direct association with incident LOAD has to our knowledge not been previously described. These APOE-dependent proteins may point directly to the processes through which APOE-ε4 mediates its risk on LOAD and provide a readout of the pathogenic process in the circulation of the approximately 50% of LOAD patients worldwide carrying the variant. While our data does not provide information on the tissue-origin of the APOE-dependent proteins, nine either exhibit brain-specific gene expression, cluster with brain-specific genes or have been associated with LOAD at the transcriptomic or protein level in brain tissue or CSF (Table 2). At the genetic level, a lookup in the GWAS catalog shows that an intron variant in the IRF6 gene has a suggestive GWAS association with LOAD via APOE-ε4 carrier status interaction. In addition, variants in the TMCC3 gene have been linked to educational attainment and caudate volume change rate and variants in the TBCA gene have been suggestively associated with reaction time and PHF-tau levels. Collectively, the gene expression patterns for these proteins in the brain, interactions with proteins involved in neuronal processes and suggestive associations between genetic markers in or near these genes and brain-related outcomes suggest that these APOE-dependent proteins may reflect brain-specific processes affected by APOE-ε4 carrier status that affect the risk of developing LOAD. Importantly, the association patterns for ARL2, S100A13 and TBCA suggest the presence of a pathway that is downregulated by APOE-ε4 early in life, given the consistent effect of APOE-ε4 on the same proteins in younger cohorts, but upregulated at the onset of LOAD, as supported by the larger observed effects in the APOE-ε4 adjusted analysis in the ACE cohort of individuals who are closer to diagnosis on the AD trajectory than those in AGES. Additional studies are required to expand on these interpretations and dissect the complex mechanisms at play and to determine if the modulation of the process represented by these proteins has therapeutic potential.

Two proteins, IBSP and APP, were identified to potentially have a causal role in LOAD. IBSP was previously associated with plasma amyloid-β and incident dementia, while APP is the precursor protein for amyloid-β. Based on the MR analysis for the LOAD-associated proteins that could be tested, the
majority do not appear to be causal in and of themselves but their association with incident LOAD may still reflect changes that occur years before the onset of LOAD that could be of interest to target before irreversible damage accumulates.

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

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

Acknowledgements
The authors acknowledge the contribution of the Icelandic Heart Association (IHA) staff to AGES-Reykjavik, as well as the involvement of all study participants. The protein measurements in AGES were supported by Novartis Biomedical Research. The National Institute on Aging (NIA) contracts N01-AG-12100 and HHSN271201200022C for V.G. financed the study. IHA received a grant from Althingi (the Icelandic Parliament), V.G. received funding from the NIA (1R01AG065596-01A1) and E.J. from the NIA (1K08AG068604). A.I.L was also funded from the NIA (P30AG066511 and U01AG061357). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The Genome Research @ Ace Alzheimer Center Barcelona project (GR@ACE) is supported by Grifols SA, Fundación bancaria ‘La Caixa’, Ace Alzheimer Center Barcelona and CIBERNED. Ace Alzheimer Center Barcelona is one of the participating centers of the Dementia Genetics Spanish Consortium (DEGESCO). The FACEHBI study is supported by funds from Ace Alzheimer Center Barcelona, Grifols, Life Molecular Imaging, Araclon Biotech, Alkahest, Laboratorio de análisis Echevarne and InsCaixa. Authors acknowledge the support of the Spanish Ministry of Science and Innovation, Proyectos de Generación de Conocimiento grants PID2021-122473OA-I00, PID2021-123462OB-I00 and PID2019-106625RB-I00. ISCIII, Acción Estratégica en Salud, integrated in the Spanish National R+D+i Plan and financed by ISCIII Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER “Una manera de hacer Europa”) grants PI13/02434, PI16/01861, PI17/01474, PI19/00335, PI19/01240, PI19/01301, PI22/01403, PI22/00258 and the ISCIII national grant PMP22/00022, funded by the European Union (NextGenerationEU). The support of CIBERNED (ISCIII) under the grants CB06/05/2004 and CB18/05/00010. The support from the ADAPTED and MOPEAD projects, European Union/EFPIA Innovative Medicines Initiative Joint (grant numbers 115975 and 115985, respectively); from PREADAPT project, Joint Program for Neurodegenerative Diseases (JPND) grant Nº AC19/00097; from HARPONE project, Agency for Innovation and Entrepreneurship (VLAIO) grant Nº PR067/21 and Janssen. DESCARTES project is funded by German Research Foundation (DFG). Amanda Cano received support from the ISCIII under the grant Sara Borrell (CD22/00125).

Author contributions

Conceptualization – E.A.F., Va.G., V.E, Vi.G.
Writing original draft – E.A.F., Va.G.
Writing review & editing – All authors
Visualization – E.A.F., Va.G.
Supervision – Va.G., Vi.G.
Funding Acquisition – Vi.G., LJ.L.

Declaration of interests

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 Emtherapro. No other potential conflicts of interest relevant to this article were reported.

Figure titles and legends

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

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

**Fig. 3 – Proteins with APOE-ε4 dependent association to incident LOAD. A)** Spaghetti plot showing the statistical significance of protein associations with incident LOAD across the three models, highlighting a set of 17 unique proteins (green) whose association with incident LOAD is attenuated upon APOE-ε4 adjustment. The horizontal lines indicate FDR < 0.05 (dashed) and P < 0.05 (dot-dashed). **B)** Pairwise
Pearson’s correlation between the 17 APOE-dependent proteins. C) Forest plot showing the linear associations between APOE genotype and the 17 APOE-dependent proteins. The beta coefficient shows the change in protein levels per ε4 allele count. D-E) Forest plots showing the associations between the 17 APOE-dependent proteins and incident LOAD D) without APOE-ε4 adjustment (model 1) and E) with APOE-ε4 adjustment (model 2). LOAD-HR indicates risk per SD increase of protein levels. Proteins that change direction of effect between the two models are highlighted in red.

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

**Fig. 5 – Reverse Mendelian randomization analysis.** A) Comparison of hazard ratios for incident LOAD with and without APOE-ε4 adjustment in the observational analysis, the effects of APOE-ε4 on protein levels and reverse MR odds ratios (excluding the APOE locus) for the four APOE-dependent proteins that change direction of effect in both observational and causal analyses when APOE is accounted for. B-C) Visual summaries of the observed data. B) Mediation diagrams showing 3 possible hypotheses that could explain the relationship between APOE-ε4, LOAD and the four proteins shown in A). Our analyses do not support the hypothesis that LOAD mediates the effect of APOE-ε4 on proteins (Hypothesis 1) nor the other way around (Hypothesis 2). However, our results from both the observational and causal analyses support the hypothesis that two mechanisms are at play that affect the same proteins in the opposite direction (Hypothesis 3). C) The APOE-ε4 mutation leads to increased risk of LOAD via its effects in brain tissue. The same mutation results in a downregulation of serum levels of four proteins that are themselves negatively associated with incident LOAD. Additionally, other non-APOE LOAD risk variants lead to upregulation of the same proteins in the reverse MR analysis, possibly reflecting a response to LOAD or its genetic liability.

**Fig. 6 – Overlap between AD protein signatures in serum, brain and CSF.** A) A Venn diagram showing the overlap of AD-associated proteins in serum, brain and CSF. B) A comparison of the effect sizes for AD associated proteins that overlap between serum and brain (top) and serum and CSF (bottom). The proteins are stratified based on the APOE-dependence in AGES for incident LOAD. The effect size in AGES is shown for incident LOAD model 1, except for proteins that were uniquely identified using the shorter 10-year follow-up or prevalent LOAD, in which case the respective effect size from the significant
association is shown. C-E) Heatmap showing the enrichment of AD-associated proteins by tissue type (x-axis) in C) the AGES serum protein modules, D) Emory CSF protein modules and E) Emory brain protein modules (y-axis).

Fig. 7 – Graphical summary of the results.

Tables with titles and legends

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

<table>
<thead>
<tr>
<th>Entrez symbol</th>
<th>HR</th>
<th>P</th>
<th>FDR</th>
<th>MS validated</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD4</td>
<td>0.79 (0.73-0.87)</td>
<td>1.9E-07</td>
<td>9.3E-04</td>
<td></td>
</tr>
<tr>
<td>BRD4</td>
<td>0.82 (0.75-0.89)</td>
<td>1.4E-06</td>
<td>3.4E-03</td>
<td></td>
</tr>
<tr>
<td>SPON1</td>
<td>1.24 (1.13-1.36)</td>
<td>3.6E-06</td>
<td>3.9E-03</td>
<td></td>
</tr>
<tr>
<td>CLEC2L</td>
<td>0.83 (0.77-0.9)</td>
<td>5.7E-06</td>
<td>3.9E-03</td>
<td></td>
</tr>
<tr>
<td>SOD3</td>
<td>1.21 (1.11-1.3)</td>
<td>5.2E-06</td>
<td>3.9E-03</td>
<td></td>
</tr>
<tr>
<td>ACHB</td>
<td>1.22 (1.12-1.33)</td>
<td>3.9E-06</td>
<td>3.9E-03</td>
<td></td>
</tr>
<tr>
<td>IL1R1</td>
<td>1.23 (1.12-1.34)</td>
<td>4.7E-06</td>
<td>3.9E-03</td>
<td></td>
</tr>
<tr>
<td>ADAM80</td>
<td>0.83 (0.77-0.9)</td>
<td>1.1E-05</td>
<td>6.0E-03</td>
<td></td>
</tr>
<tr>
<td>PLA2G2E</td>
<td>0.82 (0.75-0.89)</td>
<td>1.2E-05</td>
<td>6.0E-03</td>
<td></td>
</tr>
<tr>
<td>TFF1</td>
<td>1.2 (1.11-1.31)</td>
<td>1.3E-05</td>
<td>6.0E-03</td>
<td></td>
</tr>
<tr>
<td>SVEP1</td>
<td>1.21 (1.11-1.33)</td>
<td>2.5E-05</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>SVEP1</td>
<td>1.21 (1.11-1.32)</td>
<td>3.3E-05</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>LCORL</td>
<td>0.82 (0.75-0.9)</td>
<td>3.2E-05</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>PROK1</td>
<td>0.84 (0.77-0.91)</td>
<td>3.1E-05</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>NCAM1</td>
<td>1.2 (1.1-1.3)</td>
<td>4.0E-05</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>SMOC1</td>
<td>1.21 (1.11-1.13)</td>
<td>3.9E-05</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>SOD3</td>
<td>1.39 (1.09-1.29)</td>
<td>4.0E-05</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>NLGN1</td>
<td>1.38 (1.09-1.28)</td>
<td>4.5E-05</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>SURF1</td>
<td>1.19 (1.1-1.3)</td>
<td>4.3E-05</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>KGF2</td>
<td>0.84 (0.77-0.92)</td>
<td>5.4E-05</td>
<td>0.012</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 – A summary table of the 17 APOE-dependent LOAD associated proteins, describing their tissue specificity in the Human Protein Atlas v22, results from the association analyses in AGES, references for previous associations with APOE or LOAD and whether the aptamers have been orthogonally validated.
The APOE-dependency is defined as being significant (FDR<0.05) in model 1 and fully non-significant in model 2 (P>0.05) for incident LOAD.

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

<table>
<thead>
<tr>
<th>Entrez Symbol</th>
<th>Tissue specificity</th>
<th>Tissue cluster</th>
<th>LOAD ref</th>
<th>APOE ref</th>
<th>HR</th>
<th>FDR</th>
<th>95 % CI</th>
<th>P</th>
<th>Beta</th>
<th>SE</th>
<th>P</th>
<th>FDR</th>
<th>MS validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRRN1</td>
<td>Brain</td>
<td>Brain</td>
<td>59</td>
<td>49</td>
<td>1.258</td>
<td>1.0-05</td>
<td>1.0-018</td>
<td>0.711</td>
<td>1.061</td>
<td>0.023</td>
<td>&lt;1E-300</td>
<td>&lt;1E-300</td>
<td>Yes</td>
</tr>
<tr>
<td>S100A13</td>
<td>Low</td>
<td>Mitochondria</td>
<td>49</td>
<td>49</td>
<td>0.799</td>
<td>7.0-05</td>
<td>1.0-089</td>
<td>0.092</td>
<td>-1.090</td>
<td>0.021</td>
<td>&lt;1E-300</td>
<td>&lt;1E-300</td>
<td>Yes</td>
</tr>
<tr>
<td>TBCA</td>
<td>Low</td>
<td>Vesicular transport</td>
<td>49</td>
<td>49</td>
<td>0.782</td>
<td>7.0-06</td>
<td>1.0-086</td>
<td>0.145</td>
<td>-1.183</td>
<td>0.020</td>
<td>&lt;1E-300</td>
<td>&lt;1E-300</td>
<td>Yes</td>
</tr>
<tr>
<td>CT1F</td>
<td>Low</td>
<td>Fibroblasts, ECM</td>
<td>60</td>
<td>49</td>
<td>1.62</td>
<td>2.0-04</td>
<td>1.0-077</td>
<td>0.147</td>
<td>0.611</td>
<td>0.021</td>
<td>1.0-171</td>
<td>5.0-171</td>
<td></td>
</tr>
<tr>
<td>ARL1</td>
<td>Low</td>
<td>Striated muscle</td>
<td>59</td>
<td>49</td>
<td>0.835</td>
<td>0.002</td>
<td>1.0-093</td>
<td>0.077</td>
<td>-0.991</td>
<td>0.022</td>
<td>&lt;1E-300</td>
<td>&lt;1E-300</td>
<td>Yes</td>
</tr>
<tr>
<td>Clorf56</td>
<td>Testis</td>
<td>Spermatogenesis</td>
<td>60</td>
<td>60</td>
<td>1.200</td>
<td>0.002</td>
<td>1.0-070</td>
<td>0.113</td>
<td>0.513</td>
<td>0.026</td>
<td>6.0-85</td>
<td>9.0-85</td>
<td>Yes</td>
</tr>
<tr>
<td>MSN</td>
<td>Low</td>
<td>Neutrophils, Inflammatory response</td>
<td>61</td>
<td></td>
<td>1.186</td>
<td>0.003</td>
<td>1.0-039</td>
<td>0.388</td>
<td>0.608</td>
<td>0.025</td>
<td>9.0-122</td>
<td>2.0-121</td>
<td></td>
</tr>
<tr>
<td>TMCC3</td>
<td>Brain</td>
<td>Brain, Nervous system development</td>
<td>62</td>
<td></td>
<td>0.849</td>
<td>0.005</td>
<td>0.9-062</td>
<td>0.364</td>
<td>-0.517</td>
<td>0.026</td>
<td>4.0-66</td>
<td>7.0-66</td>
<td>Yes</td>
</tr>
<tr>
<td>HBQ1</td>
<td>Bone marrow</td>
<td>Neutrophils, Humoral immune response</td>
<td>60</td>
<td></td>
<td>0.860</td>
<td>0.013</td>
<td>0.9-066</td>
<td>0.423</td>
<td>-0.495</td>
<td>0.025</td>
<td>1.0-69</td>
<td>1.0-69</td>
<td>Yes</td>
</tr>
<tr>
<td>HIF6</td>
<td>Esophagus, Skin</td>
<td>Squamous epithelial cells</td>
<td>63</td>
<td></td>
<td>0.876</td>
<td>0.032</td>
<td>1.0-083</td>
<td>0.458</td>
<td>-0.675</td>
<td>0.025</td>
<td>4.0-153</td>
<td>1.0-152</td>
<td></td>
</tr>
<tr>
<td>IFT1</td>
<td>Bone marrow</td>
<td>Brain &amp; bone marrow, Chromatin organization</td>
<td>64</td>
<td></td>
<td>0.875</td>
<td>0.035</td>
<td>0.9-049</td>
<td>0.214</td>
<td>-0.275</td>
<td>0.026</td>
<td>2.0-26</td>
<td>2.0-26</td>
<td>Yes</td>
</tr>
<tr>
<td>TIMP3H1</td>
<td>Low</td>
<td>Adipose tissue, ECM organization</td>
<td>65</td>
<td></td>
<td>0.879</td>
<td>0.036</td>
<td>0.9-081</td>
<td>0.853</td>
<td>-0.436</td>
<td>0.026</td>
<td>1.0-60</td>
<td>1.0-60</td>
<td>Yes</td>
</tr>
<tr>
<td>FAM159B</td>
<td>Brain and other</td>
<td>Unknown function</td>
<td>66</td>
<td></td>
<td>0.882</td>
<td>0.039</td>
<td>0.9-077</td>
<td>0.591</td>
<td>-0.439</td>
<td>0.026</td>
<td>8.0-66</td>
<td>9.0-66</td>
<td>Yes</td>
</tr>
<tr>
<td>NDE1</td>
<td>Low</td>
<td>Brain, Nervous system development</td>
<td>63</td>
<td></td>
<td>0.876</td>
<td>0.039</td>
<td>1.0-095</td>
<td>0.985</td>
<td>-0.489</td>
<td>0.025</td>
<td>3.0-80</td>
<td>5.0-80</td>
<td>Yes</td>
</tr>
<tr>
<td>NEFL</td>
<td>Brain</td>
<td>Brain, Neuronal signaling</td>
<td>64</td>
<td></td>
<td>0.878</td>
<td>0.041</td>
<td>1.0-031</td>
<td>0.516</td>
<td>-0.578</td>
<td>0.024</td>
<td>2.0-121</td>
<td>4.0-121</td>
<td></td>
</tr>
<tr>
<td>GSTM1</td>
<td>Brain</td>
<td>Liver, Metabolism</td>
<td>65</td>
<td></td>
<td>1.136</td>
<td>0.042</td>
<td>1.0-077</td>
<td>0.087</td>
<td>0.276</td>
<td>0.026</td>
<td>1.0-26</td>
<td>1.0-26</td>
<td></td>
</tr>
<tr>
<td>GG72</td>
<td>Kidney, Thyroid</td>
<td>Kidney, Transmembrane transport</td>
<td>66</td>
<td></td>
<td>0.887</td>
<td>0.048</td>
<td>0.9-081</td>
<td>0.646</td>
<td>-0.384</td>
<td>0.026</td>
<td>2.0-47</td>
<td>3.0-47</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 3 – Replication of the LOAD associated proteins from AGES in the ACE cohort. All proteins with nominal P<0.05 in either model are shown. P and FDR values <0.05 are highlighted in bold.
<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Gene Name</th>
<th>Category</th>
<th>p-val</th>
<th>Effect Size</th>
<th>95% CI</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>other p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>2570_72_5</td>
<td>IGFBP2</td>
<td>Independent</td>
<td>0.002</td>
<td>1.020-1.415</td>
<td>1.080-1.414</td>
<td>1.237</td>
<td>1.085-1.414</td>
<td>0.002</td>
</tr>
<tr>
<td>8469_41_3</td>
<td>IGFBP2</td>
<td>Independent</td>
<td>0.004</td>
<td>1.066-1.418</td>
<td>1.099-1.410</td>
<td>1.230</td>
<td>1.066-1.418</td>
<td>0.006</td>
</tr>
<tr>
<td>10976_39_3</td>
<td>GH2</td>
<td>Independent</td>
<td>0.009</td>
<td>1.043-1.340</td>
<td>0.988-1.281</td>
<td>1.182</td>
<td>1.043-1.340</td>
<td>0.053</td>
</tr>
<tr>
<td>2813_11_2</td>
<td>AGRF</td>
<td>Other</td>
<td>0.011</td>
<td>1.039-1.355</td>
<td>1.048-1.364</td>
<td>1.187</td>
<td>1.039-1.355</td>
<td>0.008</td>
</tr>
<tr>
<td>7106_31_3</td>
<td>CD2</td>
<td>Independent</td>
<td>0.014</td>
<td>1.031-1.317</td>
<td>0.975-1.246</td>
<td>1.166</td>
<td>1.031-1.317</td>
<td>0.121</td>
</tr>
<tr>
<td>6168_11_2</td>
<td>TP53</td>
<td>Other</td>
<td>0.015</td>
<td>1.034-1.369</td>
<td>0.976-1.302</td>
<td>1.190</td>
<td>1.034-1.369</td>
<td>0.104</td>
</tr>
<tr>
<td>4964_67_1</td>
<td>ERAP1</td>
<td>Independent</td>
<td>0.016</td>
<td>0.763-0.972</td>
<td>0.742-0.953</td>
<td>0.861</td>
<td>0.763-0.972</td>
<td>0.007</td>
</tr>
<tr>
<td>7289_9_3</td>
<td>S100A13</td>
<td>Dependent</td>
<td>0.016</td>
<td>0.749-0.971</td>
<td>1.240-1.816</td>
<td>0.853</td>
<td>0.749-0.971</td>
<td>3.1E-05</td>
</tr>
<tr>
<td>4930_21_1</td>
<td>STC1</td>
<td>Independent</td>
<td>0.021</td>
<td>1.024-1.343</td>
<td>1.024-1.346</td>
<td>1.173</td>
<td>1.024-1.343</td>
<td>0.021</td>
</tr>
<tr>
<td>5939_26_3</td>
<td>CD26G1</td>
<td>Other</td>
<td>0.022</td>
<td>0.764-0.979</td>
<td>0.788-1.005</td>
<td>0.865</td>
<td>0.764-0.979</td>
<td>0.059</td>
</tr>
<tr>
<td>7239_9_3</td>
<td>GSTM1</td>
<td>Dependent</td>
<td>0.023</td>
<td>1.021-1.316</td>
<td>0.941-1.217</td>
<td>1.179</td>
<td>1.021-1.316</td>
<td>0.304</td>
</tr>
<tr>
<td>4929_90_1</td>
<td>SHBG</td>
<td>Other</td>
<td>0.023</td>
<td>1.022-1.335</td>
<td>1.149-1.500</td>
<td>1.168</td>
<td>1.022-1.335</td>
<td>0.042</td>
</tr>
<tr>
<td>13118_5_3</td>
<td>SMOC1</td>
<td>Independent</td>
<td>0.024</td>
<td>1.023-1.345</td>
<td>1.006-1.331</td>
<td>1.172</td>
<td>1.023-1.345</td>
<td>0.041</td>
</tr>
<tr>
<td>5660_51_3</td>
<td>SOD3</td>
<td>Independent</td>
<td>0.026</td>
<td>1.018-1.326</td>
<td>1.018-1.325</td>
<td>1.162</td>
<td>1.018-1.326</td>
<td>0.026</td>
</tr>
<tr>
<td>5581_28_3</td>
<td>PG1</td>
<td>Other</td>
<td>0.028</td>
<td>1.015-1.301</td>
<td>0.988-1.266</td>
<td>1.149</td>
<td>1.015-1.301</td>
<td>0.077</td>
</tr>
<tr>
<td>13022_20_3</td>
<td>TP53H1</td>
<td>Dependent</td>
<td>0.035</td>
<td>0.777-0.991</td>
<td>0.930-1.206</td>
<td>0.877</td>
<td>0.777-0.991</td>
<td>0.388</td>
</tr>
<tr>
<td>8235_48_3</td>
<td>CHS8</td>
<td>Other</td>
<td>0.035</td>
<td>1.010-1.311</td>
<td>0.983-1.282</td>
<td>1.151</td>
<td>1.010-1.311</td>
<td>0.088</td>
</tr>
<tr>
<td>7016_12_3</td>
<td>GCNT1</td>
<td>Independent</td>
<td>0.036</td>
<td>0.754-0.990</td>
<td>0.768-1.001</td>
<td>0.864</td>
<td>0.754-0.990</td>
<td>0.052</td>
</tr>
<tr>
<td>13731_14_3</td>
<td>C7</td>
<td>Independent</td>
<td>0.036</td>
<td>1.010-1.331</td>
<td>1.001-1.316</td>
<td>1.159</td>
<td>1.010-1.331</td>
<td>0.049</td>
</tr>
<tr>
<td>2925_9_1</td>
<td>STFP1</td>
<td>Other</td>
<td>0.042</td>
<td>0.770-0.995</td>
<td>0.762-0.992</td>
<td>0.875</td>
<td>0.770-0.995</td>
<td>0.038</td>
</tr>
<tr>
<td>3290_50_2</td>
<td>CD109</td>
<td>Independent</td>
<td>0.043</td>
<td>0.784-0.966</td>
<td>0.794-1.012</td>
<td>0.884</td>
<td>0.784-0.966</td>
<td>0.076</td>
</tr>
<tr>
<td>6626_81_3</td>
<td>CHST12</td>
<td>Independent</td>
<td>0.045</td>
<td>0.761-0.997</td>
<td>0.790-1.036</td>
<td>0.871</td>
<td>0.761-0.997</td>
<td>0.146</td>
</tr>
<tr>
<td>6604_50_3</td>
<td>NDNF</td>
<td>Other</td>
<td>0.048</td>
<td>0.774-0.999</td>
<td>0.790-1.019</td>
<td>0.879</td>
<td>0.774-0.999</td>
<td>0.096</td>
</tr>
<tr>
<td>14159_1_3</td>
<td>IFNA7</td>
<td>Other</td>
<td>0.050</td>
<td>1.000-1.279</td>
<td>1.003-1.282</td>
<td>1.131</td>
<td>1.000-1.279</td>
<td>0.045</td>
</tr>
<tr>
<td>8072_19_3</td>
<td>MZT1</td>
<td>Independent</td>
<td>0.051</td>
<td>0.996-1.256</td>
<td>1.020-1.267</td>
<td>1.118</td>
<td>0.996-1.256</td>
<td>0.022</td>
</tr>
<tr>
<td>4982_54_1</td>
<td>FB</td>
<td>Other</td>
<td>0.051</td>
<td>0.989-1.311</td>
<td>1.027-1.366</td>
<td>1.139</td>
<td>0.989-1.311</td>
<td>0.020</td>
</tr>
<tr>
<td>4297_62_3</td>
<td>SPON1</td>
<td>Independent</td>
<td>0.073</td>
<td>0.989-1.265</td>
<td>1.019-1.336</td>
<td>1.127</td>
<td>0.989-1.265</td>
<td>0.025</td>
</tr>
<tr>
<td>8242_9_3</td>
<td>CLEC2L</td>
<td>Independent</td>
<td>0.124</td>
<td>0.974-1.248</td>
<td>1.136</td>
<td>1.103-1.287</td>
<td>0.102</td>
<td>0.974-1.248</td>
</tr>
<tr>
<td>12587_65_3</td>
<td>ARL2</td>
<td>Dependent</td>
<td>0.130</td>
<td>0.804-1.028</td>
<td>1.252</td>
<td>1.076-1.457</td>
<td>0.899</td>
<td>0.804-1.028</td>
</tr>
</tbody>
</table>

CC-BY-NC-ND 4.0 International license
It is made available under a CC-BY-NC-ND 4.0 International license.
Methods

AGES study population

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

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

The proteomic measurements in AGES have been described in detail elsewhere and was available for 5,457 participants. Briefly, a custom version of the SOMAscan platform (Novartis V3-5K) was applied based on the slow-off rate modified aptamer (SOMAmer) protein profiling technology including 4,782 aptamers that bind to 4,137 human proteins. Serum was prepared using a standardized protocol from blood samples were collected after an overnight fast, stored in 0.5 ml aliquots at −80°C and serum samples that had not been previously thawed were used for the protein measurements. All samples were run as a single set at SomaLogic Inc. (Boulder, CO, US). Hybridization controls were used to adjust for systematic variability in detection and calibrator samples of three dilution sets (40%, 1%, and 0.005%) were included so that the degree of fluorescence was a quantitative reflection of protein concentration. All aptamers that passed quality control had median intra-assay and inter-assay coefficient of variation (CV) ≤ 15%. Finally, intraplate median signal normalization was applied to individual samples by SomaLogic instead of normalization to an external reference of healthy individuals, as is done for later versions of the SOMAscan platform.

ACE cohort

ACE Alzheimer Center Barcelona was founded in 1995 and has collected and analyzed roughly 18,000 genetic samples, diagnosed over 8,000 patients, and participated in nearly 150 clinical trials to date. For more details, visit www.fundacioace.com/en. The syndromic diagnosis of all subjects of the ACE cohort was established by a multidisciplinary group of neurologists, neuropsychologists, and social workers. Healthy controls (HCs), including individuals with subjective cognitive decline (SCD) diagnosis, were assigned a Clinical Dementia Rating (CDR) of 0, and mild cognitive impairment (MCI) individuals a CDR of 0.5. For MCI diagnoses, the classification of López et al., 2003, and Petersen’s criteria were used. The 2011 National Institute on Aging and Alzheimer’s Association (NIA-AA) guidelines were used for AD diagnosis. All ACE clinical protocols have been previously published. Paired plasma and CSF samples, following consensus recommendations, were stored at −80°C. A subset of ACE cohort was analyzed with the SOMAscan 7k proteomic platform (n = 1,370), (SomaLogic Inc., Boulder, CO, US).
proteomic data underwent standard quality control procedures at SomaLogic and was median normalized to reference using the Normalization by Maximum Likelihood (ANML) method (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 assessed using TaqMan genotyping assays for rs429358 and rs7412 SNPs (Thermo Fisher). Genotypes were furthermore extracted from the Axiom 815K Spanish Biobank Array (Thermo Fisher) performed by the Spanish National Center for Genotyping (CeGen, Santiago de Compostela, Spain).

Statistical analysis

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

APOE-dependence criteria of the proteins were defined as serum proteins that met FDR significance of < 0.05 in association with incident LOAD in model 1, thus unadjusted for the APOE-ε4 allele, but whose nominal significance was abolished upon APOE-ε4 correction in model 2 (P > 0.05). Serum proteins that remained nominally significantly associated with incident LOAD (P < 0.05) upon APOE-ε4 correction but changed direction of effect were also considered to meet the APOE dependence criteria, as a reversal of the effect indicates that the primary association is driven by APOE-ε4.
Functional enrichment analyses were performed using Over-Representation Analysis (ORA) and Gene Set Enrichment Analysis (GSEA) using the R packages ClusterProfiler and fgsea\textsuperscript{87,88}. The association significance cut-off for inclusion in ORA was FDR < 0.05. Background for both methods was specified as all proteins tested from the analysis leading up to enrichment testing. The investigated gene sets were the following: Gene Ontology, Human Phenotype Ontology, KEGG, Wikipathways, Reactome, Pathway Interaction Database (PID), MicroRNA targets (MIRDB and Legacy), Transcription factor targets (GTRD and Legacy), ImmuneSigDB and the Vaccine response gene set\textsuperscript{89}. Finally, we looked into tissue gene expression signatures via the same methods (ORA and GSEA) using data from GTEx\textsuperscript{90} and The Human Protein Atlas, where gene expression patterns across tissues were categorized in a similar manner as described by Uhlen et al\textsuperscript{50} and tissue-elevated expression considered as gene expression in any of the categories ‘tissue-specific’, ‘tissue-enriched’ or ‘group-enriched’. MinGSize was set at 2 when investigating the LOAD-associated serum proteins directly. When investigating the APOE-dependent protein interaction partners, minGSize was set to 15 and maxGSize was set to 500. Overrepresentation of brain cell type markers among LOAD-associated proteins was tested using a Fisher’s exact test, with the SOMApanel protein set as background. Tissue specificity lookup for the top LOAD associated proteins was based on the Human Protein Atlas version 22 (\url{https://v22.proteinatlas.org/}). For the protein-protein interaction (PPI) network analysis, PPIs from InWeb\textsuperscript{32} (n = 14,448, after Entrez ID filtering) were used to obtain the first-degree interaction partners of the APOE-dependent proteins.

Protein comparisons across serum, CSF and brain

To compare protein modules and AD associations across tissues, protein modules and protein associations to AD were obtained from brain\textsuperscript{37} and CSF\textsuperscript{36}. The brain data, from the Banner Sun Health Research Institute\textsuperscript{91} and ROSMAP\textsuperscript{92}, included TMT-MS-based quantitative proteomics for 106 controls, 200 asymptomatic AD cases and 182 AD cases. The CSF samples were collected under the auspices of the Emory Goizueta Alzheimer’s Disease Research Center (ADRC) and Emory Healthy Brain Study (EHBS)\textsuperscript{36}. The cohort consisted of 140 healthy controls and 160 patients with AD as defined by the NIA research framework\textsuperscript{93}. Protein measurements were performed using TMT-MS and SomaScan (7k). Only SomaLogic protein measurements were included in the comparison between CSF and serum, which were median normalized. Proteins were matched on SomaLogic aptamer ID when possible but otherwise by Entrez gene symbol. Overlaps between modules and AD-associated (FDR<0.05) proteins across tissues were evaluated with Fisher’s exact test.

Mendelian randomization
A two-sample bi-directional Mendelian randomization (MR) analysis was performed to first evaluate the potential causal effects of serum protein levels on AD (forward MR), and second to evaluate the potential causal effects of AD or its genetic liability on serum protein levels (reverse MR). All aptamers significantly (FDR<0.05) associated with LOAD (incident or prevalent) were included in the MR analyses, or a total of 346 unique aptamers (Supplementary Tables 2, 3 and 5), of which 320 aptamers were significant in the full follow-up incident LOAD analysis (models 1-3), 106 aptamers were significant in the 10-year follow-up incident LOAD analysis (model 1-3) and ten aptamers were significant in the prevalent LOAD analysis (models 1-3). Genetic instruments for serum protein levels were obtained from a GWAS of serum protein levels in AGES\textsuperscript{24} and defined as follows. All variants within a 1\textsuperscript{Mb} (±500\textsuperscript{kb}) cis-window for the protein-encoding gene were obtained for a given aptamer. A cis-window-wide significance level $P_{cis}=0.05/N$, where $N$ equals the number of SNPs within a given cis-window, was computed and variants within the cis window for each aptamer were clumped ($r^2_{cis}=0.2$, $P_{cis}<P_b$). The effect of the genetic instruments for serum protein levels on LOAD risk was obtained from a GWAS on GWAS on 39,106 clinically diagnosed LOAD cases, 46,828 proxy-LOAD and dementia cases and 401,577 controls of European ancestry\textsuperscript{19}. Genetic instruments for the serum protein levels not found in the LOAD GWAS data set were replaced by proxy SNPs ($r^2_{cis}>0.8$) when possible, to maximize SNP coverage. Genetic instruments for LOAD in the reverse causation MR analysis were obtained from the same LOAD GWAS\textsuperscript{24}, where genome-wide significant variants were extracted ($P_{cis}<5\times10^{-8}$) and clumped at a more stringent LD threshold ($r^2_{cis}>0.01$) than for the protein instruments to limit overrepresentation of SNPs from any given locus across the genome. In the reverse causation MR analysis, cis variants (±500\textsuperscript{kb}) for the given protein were excluded from the analysis to avoid including pleiotropic instruments affecting the outcome (protein levels) through other mechanisms than the exposure (LOAD). A secondary reverse causation MR analysis was performed excluding any variants in the \textit{APOE} locus (chr19:45,048,858-45,733,201). Causal estimate for each protein was obtained by the generalized weighted least squares (GWLS) method\textsuperscript{94}, which accounts for correlation between instruments. Causality for proteins with single cis-acting variants was assessed with the Wald ratio estimator. For the reverse causation MR analysis, the inverse variance weighted method was applied due to a more stringent LD filtering of the instruments. Instrument heterogeneity was evaluated with Cochran’s Q test and horizontal pleiotropy with the MR Egger test.

\textbf{Supplemental information titles and legends}
**Supplementary Figure 1 – A-B** Volcano plots showing the protein association profile for incident LOAD, restricted to 10-year follow-up **A** without APOE e4 adjustment (model 1) and **B** with APOE e4 adjustment (model 2). **C** Venn diagram for the aptamer overlap between models 1 and 2 for incident AD, restricted to 10-year follow-up.

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

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

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

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

**Supplementary Figure 6 – A-B** Spaghetti plots showing the statistical significance of protein associations with incident LOAD across three statistical models, adjusting for age and sex, and then additionally either serum **A** total cholesterol or **B** LDL cholesterol, and finally APOE e4 carrier status. The LOAD association of the 17 APOE-dependent proteins (green) is not attenuated when adjusting for total cholesterol or LDL but only when adjusting for APOE e4 carrier status. **C-H** Protein associations with serum total cholesterol (C-E) and LDL cholesterol (F-H). The density plots show the effect size for the associations between cholesterol levels and all measured proteins stratified by their AD association, adjusting for age.
and sex (C and F), and additionally APOE-e4 carrier status (D and G). (E) and (H) Comparisons of the effect sizes for association with cholesterol levels in the models with (y-axis) and without (x-axis) APOE-e4 adjustment shows that the effect size is often increased for the APOE-dependent proteins (red) after APOE-e4 adjustment.

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

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

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

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

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

Supplementary Table 1 - Baseline characteristics of the AGES cohort stratified by LOAD status. P values are obtained from t-test for continuous variables, and chi-square test for categorical variables. P and FDR values <0.05 are highlighted in bold.
Supplementary Table 2 – Proteins significantly associated with incident LOAD after FDR (Benjamini-Hochberg) correction are displayed. The full Cox proportional hazards regression model (Model 3) was adjusted for baseline age, sex, APOE-e4 genotype, body mass index, diabetes, education, hypertension, smoking history, and eGFR as denoted below. P and FDR values <0.05 are highlighted in bold. Abbreviations: HR, Hazard ratio; CI, Confidence Interval; Zph: proportional hazard test statistic. Estimates represent difference in LOAD risk per standard deviation increase of aptamer level.

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

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

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

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

Supplementary Table 7 – Results from Gene Set Enrichment Analysis (GSEA) and Overrepresentation Analysis (ORA) of the prevalent LOAD associated proteins against protein-coding gene sets labeled below using the R packages listed below. P values were adjusted for multiple comparisons using FDR (Benjamini Hochberg). Abbreviations: ES, Enrichment score; NES, Enrichment score normalized.
Supplementary Table 8 – Gene Set Enrichment Analysis and Overrepresentation Analysis of incident LOAD associated proteins in AGES against protein-coding gene sets labeled below using the R packages listed below. P values were adjusted for multiple comparisons using FDR (Benjamini Hochberg). P and FDR values <0.05 are highlighted in bold. Abbreviations: ES, Enrichment score; NES, Enrichment score normalized; size of the pathway after removing missing genes; leadingEdge, vector of genes that drive the enrichment.

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

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

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

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

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

Supplementary Table 14 – Overview of the genetic instruments and their associations with serum levels in AGES
Supplementary Table 15 - Baseline characteristics of the ACE cohort stratified by LOAD status. P values are obtained from t-test for continuous variables, and chi-square test for categorical variables. P and FDR values <0.05 are highlighted in bold.

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

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

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

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

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

Supplementary Table 21 – Serum, CSF and brain protein module membership for all proteins measured in serum in the AGES study.
CSF modules were obtained from an extended analysis compared to Dammer et al., Alzheimer’s Research & Therapy 2022 (manuscript in preparation), and brain modules from Johnson et al., Nature Neuroscience 2022. The same protein can occur in more than one module when measured by more than one aptamer (serum and CSF), by different platforms (CSF) or when different peptides are quantified by MS (CSF, brain).

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

**References**


AGES Cohort
n = 5764

Participants with SomaScan data
n = 5457

Somalogic serum proteins (n = 4782 aptamers)

Study Overview

Replication cohorts

Emory Brain cohort
n = 488
Prevalent AD
n = 182
Asymptomatic AD
n = 200
Controls
n = 106

ACE cohort
n = 1341
Prevalent AD
n = 243
Participants with MCI and follow-up information
n = 719
Converting to AD
n = 266

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
Serotonin receptor 2 and ELK-SRF/GATA4 signaling

Phosphodiesterases in neuronal function

Disruption of postsynaptic signaling by CNV

NO-cGMP-PKG mediated neuroprotection

Neuroinflammation and glutamatergic signaling

Alzheimer's disease

Synaptic signaling pathways associated with autism spectrum disorder

NO-cGMP-PKG mediated neuroprotection

Disruption of postsynaptic signaling by CNV

Phosphodiesterases in neuronal function

Serotonin receptor 2 and ELK-SRF/GATA4 signaling

A

B

C

D

E
Increased risk of LOAD associated with protein levels.

**Hypothesis 1:**
- **APOE-ε4** → LOAD → Protein

**Hypothesis 2:**
- **APOE-ε4** → Protein → LOAD

**Hypothesis 3:**
- I) **APOE-ε4** → \( M_A \) → LOAD → Protein
- II) **APOE-ε4** → \( M_A \) → LOAD

**Data supports model:**
- Hypothesis 1: No
- Hypothesis 2: No
- Hypothesis 3: Yes

**LOAD**
- Increased risk
- Increased levels

**APOE-ε4**
- Decreased levels

**Non APOE variants**
- Increased risk
- Increased levels

**Response**
- Exposure
- Mediator
- Outcome

**Protein association with incident LOAD:**
- Protein association with incident LOAD, APOE-ε4 adjusted APOE-ε4 effect on protein
- Reverse MR, causal effect of AD on protein
LOAD associated serum proteins
341 incident and 10 prevalent
LOAD associated aptamers out of 4782 tested.

APOE-dependent proteins
17 aptamers are associated with incident LOAD in an APOE-ε4 dependent manner. They are enriched for oligodendrocyte markers. A subset are causally affected by LOAD.

APOE-independent proteins
140 aptamers are associated with incident LOAD independently of APOE-ε4 status. They are enriched for neuron development and overlap with Alzheimer's protein signatures in CSF and brain.

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.

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.

APOE-dependent proteins and cholesterol
The APOE-dependent proteins are associated with cholesterol in an APOE-ε4 independent manner. However their LOAD association is not mediated by circulating cholesterol levels.