

1 **Unraveling the role of plasma proteins in dementia: insights from two cohort studies in**
2 **the UK, with causal evidence from Mendelian randomization**

3
4 Jessica Gong^{1,2}, Dylan M. Williams³, Shaun Scholes¹, Sarah Assaad¹, Feifei Bu⁴, Shabina
5 Hayat¹, Paola Zaninotto¹, Andrew Steptoe⁴

6
7 ¹ Department of Epidemiology and Public Health, University College London, London, UK

8 ² George Institute for Global Health, Imperial College London, London, UK

9 ³ MRC Unit for Lifelong Health & Ageing, University College London, London, UK

10 ⁴ Department of Behavioural Science and Health, University College London, London, UK

11

12 **Corresponding author:**

13 Dr Jessica Gong, jessica.gong@ucl.ac.uk

14 Department of Epidemiology and Public Health, University College London, London, UK

15

16 **Word count:** 7509 (Main text: 3955; Online methods: 3554)

17 **Number of references:** 92

18 **Number of figures:** 6

19 **Number of tables:** 0

1 **ABSTRACT**

2 Population-based proteomics offer a groundbreaking avenue to predict dementia onset. This
3 study employed a proteome-wide, data-driven approach to investigate protein-dementia
4 associations in 229 incident all-cause dementia (ACD) among 3,249 participants from the
5 English Longitudinal Study of Ageing (ELSA) over a median 9.8-year follow-up, then
6 validated in 1,506 incident ACD among 52,745 individuals from the UK Biobank (UKB) over
7 median 13.7 years. NEFL and RPS6KB1 were robustly associated with incident ACD; MMP12
8 was associated with vascular dementia in ELSA. Additional markers EDA2R and KIM1
9 (HAVCR1) were identified from sensitivity analyses. Combining NEFL and RPS6KB1 with
10 other factors yielded high predictive accuracy (area under the curve (AUC)=0.871) for incident
11 ACD. Replication in the UKB confirmed associations between identified proteins with various
12 dementia subtypes. Results from reverse Mendelian Randomization also supported the role of
13 several proteins as early dementia biomarkers. These findings underscore proteomics' potential
14 in identifying novel risk screening targets for dementia.

1 INTRODUCTION

2 The understanding of Alzheimer's disease and related dementia (ADRD) is increasingly
3 shifting towards a systemic and multifactorial perspective.[1, 2] Circulating proteins, as pivotal
4 agents in biological processes, offer direct insights into disease mechanisms and can serve as
5 early indicators, regulators, and effectors in disease pathways. This renders their studies
6 indispensable in both drug discovery and the development of diagnostics.[3, 4]
7 Mounting evidence supports the significance of proteomics profiling in exploring pathways
8 involved in ADRD.[3, 5-7] At the molecular level, deviations in protein function or expression
9 play a role in the pathogenesis of prodromal dementia,[3, 8, 9] while protein biomarkers can
10 forecast disease onset several years before symptoms manifest.[3, 8-10] Remarkably,
11 approximately 96% of currently approved drugs target proteins,[4, 11] underscoring the
12 substantial added value of proteomic analysis in ADRD drug discovery.
13 Integration of large-scale proteomics data into population studies represents a recent
14 development,[3, 12] enabling cost-effective simultaneous measurement of multiple proteins on
15 many samples.[3, 13, 14] This has led to the identification of distinct protein signatures relevant
16 to ADRD susceptibility.[4, 9, 10, 15, 16] For instance, longitudinal analyses in the UK
17 Whitehall II study spanning across two decades, demonstrated associations between 15 non-
18 amyloid/non-tau-related proteins and cognitive decline and dementia.[4] Similarly, the
19 Atherosclerosis Risk in Communities (ARIC) study in the US identified significant protein
20 signatures for dementia, including immune and proteostasis/autophagy pathways.[10]
21 Intriguingly, some of these associations were independent of known Alzheimer's disease (AD)
22 risk factors, suggesting novel potential targets for intervention.[4, 9, 10] Recent analyses based
23 on data from the UK Biobank, which identified GFAP, NEFL, GDF15 and LTBP2 were most
24 strongly associated with incident all-cause dementia (ACD), AD, and vascular dementia
25 (VAD).[16] However, previous studies utilized the aptamer-based SomaScan platform in ARIC

1 and Whitehall II,[4, 9, 10] and the SomaScan platform is deemed to have lower specificity
2 compared to the Olink platform, which employs multiplexed antibody-based immunoassays
3 proximity extension assay (PEA) technology.[17] Moreover, the UK Biobank study faced
4 limitations such as a lack of external validation cohorts and evidence triangulation with causal
5 inference.[16]

6 In this current study, we employed the large-scale Olink proteomics platform and a robust
7 dementia algorithm to assess the proteomic signature of dementia risk in over 3,000 older adults
8 using data from the English Longitudinal Study of Ageing (ELSA) as the discovery cohort. We
9 validated these findings using Olink proteomics data from over 50,000 participants from the
10 UK Biobank.[18] Two-sample bi-directional Mendelian randomization (MR) and drug target
11 MR (or cis-MR) were utilized to infer causality between protein concentration and dementia
12 outcomes, leveraging summary statistics from large genome-wide association study (GWAS)
13 consortia.

14

15 **RESULTS**

16 **Analysis 1: Protein-dementia associations in the ELSA discovery cohort**

17 The participant selection for the proteomics assay in ELSA is depicted in Supplementary Fig.
18 1. In 3,262 samples with proteomics assayed, based on the dementia algorithm, prevalent
19 dementia cases were excluded (N = 13), resulting in a final sample of 3,249 in the analysis.
20 The mean age was 63.4 years (SD = 9.2), 55% were women, and 97.2% were of white ethnicity
21 (Supplementary Table 1). A total of 229 incident dementia cases were documented over a
22 median follow-up of 9.8 years (min – max: 0.4 – 10.9 years). Specific details on the data
23 sources where these cases were extracted from in the first instance are included in
24 Supplementary Fig. 2. The normalized protein expression (NPX) levels for participants with

1 no dementia, incident dementia, and prevalent dementia for each protein are presented as box
2 plots in Supplementary Fig. 3.

3 We initially assessed the relationship between the NPX value of 276 plasma proteins and ACD
4 risk in the ELSA cohort, using Cox Proportional Hazard regression models. Unadjusted
5 analyses revealed that 95 measured proteins were significantly associated with ACD
6 (Supplementary Fig. 4). Among these, **NEFL** exhibited the strongest association with ACD, as
7 indicated by the false discovery rate (FDR)-corrected p-value (denoted as $P_{FDR} = 8.66 \times 10^{-37}$,
8 hazard ratio (HR) [95% confidence intervals (CI)]: **3.01 [2.63, 3.44]**), followed by **EDA2R**,
9 **SCARF2**, **LAYN**, **PGF**, **DCN**, **GFR-alpha-1**, **BNP**, **UNC5C**, **Dkk-4**, **KIM1** (also known as
10 **HAVCR1**), **TNFRSF12A**, **CADM3**, **TRAIL-R2**, **VWC2**, and **MMP12**.

11 In the minimally adjusted models (adjusted for age, sex, and ethnicity), **NEFL** ($P_{FDR} = 0.0002$;
12 HR [95% CI]: **1.55 [1.30, 1.85]**), **RPS6KB1** ($P_{FDR} = 0.003$; HR [95% CI]: **1.34 [1.18, 1.53]**),
13 **EDA2R** ($P_{FDR} = 0.046$; HR [95% CI]: **1.43 [1.19, 1.72]**) and **KIM1** ($P_{FDR} = 0.049$; HR [95%
14 CI]: **1.31 [1.14, 1.50]**) were significantly associated with ACD (Supplementary Fig. 5).

15 In the fully adjusted models (adjusted for age, sex, ethnicity, education, smoking status,
16 depression, presence of cardiovascular diseases, body mass index (BMI), systolic blood
17 pressure, low-density lipoprotein (LDL) cholesterol), **NEFL** ($P_{FDR} = 0.0008$; HR [95% CI]:
18 **1.54 [1.29, 1.84]**) and **RPS6KB1** ($P_{FDR} = 0.01$; HR [95% CI]: **1.33 [1.16, 1.52]**) remained
19 significantly associated with ACD (Fig. 1).

20 Sensitivity analyses demonstrated the robustness of the association between **NEFL** and **ACD**,
21 with significance persisting after excluding participants in other ethnic groups other than white
22 participants ($P_{FDR} = 0.001$; HR [95% CI]: **1.53 [1.28, 1.83]**) (Supplementary Fig. 6), APOE $\epsilon 4$
23 carriers ($P_{FDR} = 0.009$; HR [95% CI]: **1.61 [1.29, 2.01]**) (Supplementary Fig. 7), cases of
24 dementia occurring within the first year of follow-up ($P_{FDR} = 0.005$; HR [95% CI]: **1.50 [1.25,**
25 **1.80]**) (Supplementary Fig. 8), participants aged <60 years ($P_{FDR} = 0.011$; HR [95% CI]: **1.47**

1 [1.22, 1.78]) (Supplementary Fig. 9), and when death was considered as a competing risk in
2 Fine-Gray regression models ($P_{\text{FDR}} = 0.002$; HR [95% CI]: **1.50 [1.26, 1.80]**) (Supplementary
3 Fig. 10). Similarly, **RPS6KB1** exhibited a robust association with **ACD**, which remained
4 significant after excluding other ethnic groups ($P_{\text{FDR}} = 0.004$; HR [95% CI]: **1.34 [1.18, 1.53]**),
5 cases of dementia occurring within the first year of follow-up ($P_{\text{FDR}} = 0.025$; HR [95% CI]:
6 **1.31 [1.15, 1.50]**), participants aged <60 years ($P_{\text{FDR}} = 0.031$; HR [95% CI]: **1.31 [1.14, 1.50]**),
7 and it was significantly associated with ACD in the Fine-Gray competing risk model ($P_{\text{FDR}} =$
8 0.014 ; HR [95% CI]: **1.33 [1.16, 1.52]**). However, the significance in association between
9 RPS6KB1 and ACD based on P_{FDR} attenuated after excluding APOE $\epsilon 4$ carriers ($P_{\text{FDR}} = 0.961$;
10 HR [95% CI]: **1.28 [1.08, 1.50]**). After excluding other ethnic groups, **KIM1** was additionally
11 associated with **ACD** ($P_{\text{FDR}} = 0.027$; HR [95% CI]: **1.32 [1.15, 1.53]**). No evidence was found
12 for any sex differences in the protein-dementia associations when sex was fitted as an
13 interaction term in the Cox models (all P_{FDR} for interaction > 0.05).

14 When assessed by dementia subtypes, 89 incident AD and 41 cases of VAD were documented.
15 After full adjustment, no protein was found to be significantly associated with AD indicated by
16 $P_{\text{FDR}} < 0.05$ (Supplementary Fig. 11). **MMP12** was found to be associated with **VAD** ($P_{\text{FDR}} =$
17 0.046 ; HR [95% CI]: **2.06 [1.41, 2.99]**) (Supplementary Fig. 12). Albeit being non-significant
18 after FDR correction, of the proteins significantly associated with ACD, based on nominal
19 statistical significance (denoted as $P_{\text{uncorrected}} < 0.05$), **RPS6KB1** was associated with **AD**
20 ($P_{\text{uncorrected}} = 0.006$; HR [95% CI]: **1.29 [1.07, 1.55]**); and **NEFL** was associated with **VAD**
21 ($P_{\text{uncorrected}} = 0.001$; HR [95% CI]: **1.98 [1.31, 2.99]**).

22 For predicting incident ACD, plasma NEFL and RPS6KB1 parsimonious models yielded
23 modest Area Under the Receiver Operating Characteristic Curve (AUC) values [95% CI] of
24 $0.787 [0.757, 0.815]$, and $0.609 [0.571, 0.647]$, respectively (Fig. 2). We also evaluated the
25 performance of these two plasma proteins in combination with other measures, including

1 demographic predictors (age, sex, ethnicity, education), APOE ϵ 4 status, and memory score (a
2 combined test score of immediate- and delayed-recall). When NEFL was combined with these
3 predictors, the model achieved an accuracy of AUC [95% CI] = 0.866 [0.840, 0.888].
4 Comparatively, when RPS6KB1 was combined with other predictors, the model achieved a
5 comparable accuracy of AUC [95% CI] = 0.866 [0.842, 0.891]. NEFL and RPS6KB1 in
6 combination with all the other predictors yielded AUC [95% CI] = 0.871 [0.845, 0.894].
7 The XGBoost models revealed that age (mean |SHAP| = 0.047) and memory score (mean
8 |SHAP| = 0.017) were the most important features contributing to the prediction ACD onset.
9 Additionally, the protein markers RPS6KB1 (mean |SHAP| = 0.009) and NEFL (mean |SHAP|
10 = 0.008) emerged as the most prominent protein markers in predicting ACD (Fig. 3). The SHAP
11 plot also illustrated that individuals with elevated levels of RPS6KB1 were more predisposed
12 to developing ACD, while those with lower levels were more likely to remain ACD-free.

13

14 **Analysis 2: Protein-dementia associations in the UK Biobank validation cohort**

15 Based on the results from the main, sensitivity, and dementia subtype analyses conducted in
16 ELSA discovery cohort, all identified proteins were selected for further validation analyses
17 using proteomics data from UK Biobank validation cohort. However, RPS6KB1 was not
18 assayed in the UK Biobank.

19 In the UK Biobank, which included 52,745 participants with proteomics assayed and without
20 dementia at study baseline (53.9% women, 93.3% white ethnicity), the mean age was 56.8
21 years (SD = 8.2) (Supplementary Table 1). UK Biobank participants with proteomics assayed
22 were, on average, younger than participants in ELSA. Over a median of 13.7 years (min – max:
23 0.03 – 16.8 years) of follow-up, a total of 1,506 incident ACD, 732 incident AD, 281 incident
24 VAD, and 111 incident frontotemporal dementia (FTD) cases were recorded.

1 Using the same adjustment strategy for the Cox regression models in ELSA, **NEFL** was
2 replicated in the UK Biobank for **ACD** ($P_{\text{FDR}} = 1.02 \times 10^{-81}$; HR [95% CI]: **1.87 [1.75, 1.99]**).
3 **NEFL** was also associated with **AD** ($P_{\text{FDR}} = 1.89 \times 10^{-35}$; HR [95% CI]: **1.81 [1.65, 1.99]**), **VAD**
4 ($P_{\text{FDR}} = 1.59 \times 10^{-17}$; HR [95% CI]: **1.90 [1.64, 2.19]**), and **FTD** ($P_{\text{FDR}} = 1.10 \times 10^{-21}$; HR [95%
5 CI]: **2.97 [2.30, 3.70]**) (Fig. 4; Supplementary Table 2). **KIM1** was replicated for **ACD** (P_{FDR}
6 $= 3.15 \times 10^{-4}$; HR [95% CI]: **1.13 [1.06, 1.20]**); and was also associated with **AD** ($P_{\text{FDR}} = 0.077$;
7 HR [95% CI]: **1.11 [1.02, 1.21]**), and **VAD** ($P_{\text{FDR}} = 1.13 \times 10^{-6}$; HR [95% CI]: **1.44 [1.25,**
8 **1.66]**). **MMP12** was replicated for **VAD** ($P_{\text{FDR}} = 6.85 \times 10^{-5}$; HR [95% CI]: **1.36 [1.18, 1.56]**)
9 and it was also significantly associated with **ACD** ($P_{\text{FDR}} = 2.00 \times 10^{-6}$; HR [95% CI]: **1.17**
10 **[1.10, 1.24]**). **EDA2R** was replicated for **ACD** ($P_{\text{FDR}} = 3.18 \times 10^{-13}$; HR [95% CI]: **1.31 [1.22,**
11 **1.40]**); and was also associated with **AD** ($P_{\text{FDR}} = 6.06 \times 10^{-5}$; HR [95% CI]: **1.25 [1.13, 1.39]**),
12 and with **VAD** ($P_{\text{FDR}} = 0.001$; HR [95% CI]: **1.34 [1.15, 1.58]**).

13

14 **Analysis 3: Two-sample bi-directional MR**

15 We then assessed the potential causal relationships between the circulating protein
16 concentration, identified as significant based on the analyses described above, in relation to
17 dementia outcomes, using two-sample bi-directional MR. Summary statistics for genetic
18 variants associated with the circulating protein levels, protein quantitative trait locus (pQTL),
19 that is also associated with dementia in GWAS were used to infer causality. Since **RPS6KB1**
20 was not assayed in the UK Biobank, MR analyses were therefore conducted for **NEFL**, **KIM1**,
21 **EDA2R**, and **MMP12**.

22 For the GWAS used for dementia outcomes, three separate GWAS for AD (denoted as Kunkle
23 2019, Bellenguez 2022, FinnGen 2023), one GWAS for ACD (FinnGen 2023), and one GWAS
24 for VAD (FinnGen 2023) were used.

1 In the forward direction MR (circulating protein concentration \rightarrow dementia), there was
2 evidence of a potential causal link from circulating **EDA2R** to **AD** (FinnGen 2023) (coefficient
3 (β) [standard error (se)] = 0.259 [0.096], $P = 0.007$, based on the inverse-variance weighted
4 (IVW) method), and **ACD** (β [se] = 0.232 [0.110], $P = 0.035$, based on IVW) (Fig. 5,
5 Supplementary Table 3). However, these findings were less robust when methods such as MR-
6 Egger were applied, suggesting the presence of horizontal pleiotropy.

7 MR analyses in the backward direction (dementia \rightarrow circulating protein concentration)
8 supported **AD** (Bellenguez 2022) (β [se] = 0.056 [0.014]; $P = 1.081 \times 10^{-4}$, based on IVW),
9 **AD** (FinnGen 2023) (β [se] = 0.033 [0.015]; $P = 0.024$, based on IVW), and **VAD** (β [se] =
10 0.036 [0.014]; $P = 0.008$, based on maximum likelihood) as cause of altered **NEFL** abundance
11 (Fig. 5, Supplementary Table 3). Furthermore, there was evidence supporting a causal link
12 between **AD** (Kunkle 2019) and **MMP12**, demonstrated by MR-Egger, weighted median, and
13 weighted mode methods (β [se] = 0.027 [0.012], $P = 0.043$; β [se] = 0.025 [0.012], $P = 0.037$,
14 and β [se] = 0.026 [0.012], $P = 0.037$, respectively). There was evidence suggesting that **AD**
15 (FinnGen 2023) (β [se] = 0.013 [0.049], $P = 0.007$ (based on IVW)), **ACD** (β [se] = -0.032
16 [0.012], $P = 0.007$ (based on IVW)), and **VAD** (β [se] = -0.036 [0.012], $P = 0.003$ (based on
17 weighted median)) might have a causal link to altered **EDA2R** abundance.

18

19 **Analysis 4: Two sample drug target MR (cis-MR)**

20 The instrument selection for drug target MR (cis-MR) relies on single nucleotide
21 polymorphisms (SNPs) within or near the gene encoding region that regulates the protein of
22 interest. However, the encoding region of **EDA2R** is located within the *X* chromosome, which
23 precluded the analysis of drug target MR on **EDA2R**, as the sex chromosomes were excluded
24 from GWAS summary statistics. Cis-MR was conducted for **NEFL**, **KIM1**, and **MMP12**.

1 In drug target MR, there was no causal evidence for any of the protein-dementia relationships
2 (Fig. 5, Supplementary Table 4).

3 In sensitivity analysis using a less stringent instrument selection approach, results were largely
4 consistent with the main analysis. There was some evidence indicating a causal relationship
5 between **KIM1** and **AD** (FinnGen 2023) (β [se] = -0.102 [0.041], $P = 0.037$) and **ACD** (and β
6 [se] = -0.094 [0.037], $P = 0.036$) based on MR-Egger.

7

8 **Analysis 5: Enrichment analysis**

9 In Fig. 6 (also depicted in Supplementary Table 5), the enrichment analyses revealed several
10 biological pathways potentially implicated for the identified proteins (NEFL, RPS6KB1,
11 KIM1, EDA2R, and MMP12), including the immune system, cancers, and insulin signaling.
12 Tissue expression analysis showed expression in the brain for NEFL and in the kidney for
13 KIM1. Notably, one drug, LY2584702, which is a selective, adenosine triphosphate (ATP)-
14 competitive p70S6K inhibitor, has been investigated in clinical trials for the treatment of renal
15 cell carcinoma, metastases, neoplasm, and neuroendocrine tumors – where RPS6KB1 was
16 shown to be implicated in the mechanisms of action of the drug.

17 Furthermore, upon searching the Open Targets platform for the identified proteins, we
18 identified ten known small molecule drugs in clinical trials (including LY2584702) that are
19 linked to two proteins (RPS6KB1 and MMP12), targeting various cancers, chronic hepatitis C
20 infection, and chronic obstructive pulmonary disease (Supplementary Table 6).

21

22 **DISCUSSION**

23 Through a broad proteomics study within the ELSA cohort, encompassing 276 proteins across
24 3,249 participants, we identified key plasma proteins linked to an elevated risk of incident ACD
25 (NEFL and RPS6KB1) and vascular dementia (MMP12), based on fully adjusted models.

1 Secondary findings from minimally adjusted models and sensitivity analyses revealed EDA2R
2 and KIM1 as additional significant markers. NEFL and RPS6KB1 individually displayed
3 moderate predictive accuracy for ACD risk (AUC = 0.787 and 0.609, respectively), which
4 yielded an AUC of 0.871 when combined with demographic, genetic, and cognitive factors.
5 Notably, the XGboost machine learning algorithm further underscored RPS6KB1 and NEFL
6 as the most important protein features in predicting ACD onset. These discoveries from ELSA
7 were robustly replicated in the UK Biobank, where NEFL, MMP12, KIM1, and EDA2R were
8 significantly associated with various ACD and dementia subtypes, including AD, VAD, and
9 FTD. Furthermore, employing diverse MR approaches, several causal relationships were
10 observed between AD and VAD with NEFL, AD with MMP12, and between AD, ACD, VAD,
11 with EDA2R in the reverse direction. There was no evidence supporting causal relationships
12 between proteins and dementia from cis-MR analyses.

13 Elevated NEFL were found to be associated with an increased risk of ACD in ELSA, and with
14 ACD, AD, VAD, and FTD in the UK Biobank. Consistently, a previous study from the UK
15 Biobank ranked NEFL as the most important protein associated with future dementia events
16 out of 1463 protein markers.[16] NEFL is a marker of axonal injury,[19, 20] and is implicated
17 in several biological mechanisms related to dementia,[21, 22] including
18 neurodegeneration,[23, 24] inflammation,[25] central nervous system (CNS) injury,[26-28]
19 and atherosclerosis.[29] It is a well-established (non-specific) marker of neurodegenerative
20 diseases. While NEFL was found in our study to be causally linked to dementia based on the
21 MR findings, the strongest indication was in the backwards direction, which points toward its
22 role as a manifestation of prodromal dementia and anomalies in the brain, rather than a cause
23 of dementia. This underscores the value of NEFL as an important diagnostic and early
24 identification marker, as also demonstrated by the prediction models. It should be noted that
25 the inconsistencies in MR findings from the current study across various AD GWAS could stem

1 from the fact that the chosen genetic variant serving as the instrumental variable might exert a
2 varied impact on the outcome within the represented population. Such differences may arise
3 from variations in genetic backgrounds, environmental influences, or other population-specific
4 factors.

5 RPS6KB1, functions as a serine/threonine-protein kinase, operating downstream of
6 phosphoinositide 3-kinase (PI3K)/mammalian target of rapamycin (mTOR) signaling in
7 response to growth factors and nutrients, promoting cell proliferation, growth, and progression
8 through the cell cycle.[30] The mTOR complex 1 (mTORC1) signaling was found in a previous
9 study to be implicated in the biological aging process,[31] such that the inhibition of mTOR
10 may extend lifespan given that the mTOR activity becomes abnormally high with age.[31] In
11 the nervous system, the mTOR pathway is implicated in the regulation of synaptic remodeling
12 and long term potentiation.[32-35] Importantly, mTOR plays a crucial role in autophagy
13 regulation in neurons,[35, 36] and the mTOR/p70S6K axis is shown to be essential in the early
14 phases of plasticity for synaptic modifications and the formation of enduring memory.[37]
15 Previous analyses of the ARIC study similarly highlighted the importance of autophagy
16 signaling pathways in the two decades before dementia onset.[10] Interestingly, our sensitivity
17 analysis which excluded APOE ϵ 4 carriers observed an attenuation in association between
18 RPS6KB1 and dementia after considering multiple testing. Previous literature highlighted the
19 mechanisms affected by APOE ϵ 4,[38] such that the presence of APOE ϵ 4 may be necessary
20 for the overactivation in mTOR pathway which subsequently lead to tau hyperphosphorylation
21 and reduced A β clearance.[39] At a lower expression level, RPS6KB1 can facilitate the growth
22 of damaged axons resulting from CNS injury.[40] In AD patients, there is a down-regulation
23 of RPS6KB1,[40] and RPS6KB1 has also been implicated in the etiology of other complex
24 neurological diseases including depression and autism.[41] Transcriptomic exploration has
25 revealed the central role of RPS6KB1, and alterations in its co-expression occur during the

1 initial stages of AD, which highlights its potential as a biomarker for the early diagnosis of
2 AD.[42] Several small clinical trials of rapamycin are underway for investigating age-related
3 diseases including AD,[31] with primary outcomes assessing the effects on cognitive
4 performance, and biomarkers of aging.[31] From the Proteomics Drug Atlas, one drug
5 (LY2584702) was found to target RPS6KB1, which is a highly selective adenosine triphosphate
6 competitive inhibitor against p70S6 Kinase. However, the clinical trials associated with
7 LY2584702 mostly target cancers. Further explorations are therefore needed to decipher the
8 relationship between RPS6KB1 and the protein's pharmacological properties.

9 Matrix metalloproteinases (MMPs) belong to a multigenic family of membrane-bound or
10 secreted zinc-containing endopeptidases, which indirectly modulate the cellular processes
11 through activation and inactivation of signaling molecules such as trophic factors cytokines,
12 and receptor.[43, 44] MMPs play important roles in cell proliferation and death,
13 neuroinflammation, neurodegeneration, and glial reactivity,[45] and are linked to their
14 proteolytic disruption action on the blood-brain barrier.[46] Based on experimental models,
15 $A\beta_{40}$ contribute to the changes in blood-brain barrier (BBB) permeability, and increased
16 expression of MMPs in transgenic human amyloid precursor protein (hAPP)-overexpressing
17 mice, in turn compromises BBB integrity.[47] There is also an increased expression of MMPs
18 in cerebrospinal fluid of AD patients.[48] Selective inhibitors for MMP12[49] was shown to
19 reduce inflammation and delay of atherosclerosis progression.[50, 51] While observational
20 studies demonstrated an association between high levels of plasma MMP12 and recurrent
21 cardiovascular disease, MR yielded the inverse finding, such that genetic variants associated
22 with higher MMP12 levels are associated with decreased risks of coronary artery disease and
23 large artery atherosclerotic stroke, discouraging potential clinical trials of MMP12
24 inhibitors.[3, 52] For dementia, the associations between MMP12 and VAD risk in the UK
25 Biobank, and MMP12 and AD risk in the ARIC cohort was similarly highlighted.[10, 16] At

1 an elevated dosage, this medication can penetrate the BBB and manifest an inhibitory effect on
2 metalloproteinase activity within the brain,[53] and was shown to decrease some seizure-
3 related parameters.[53]

4 In our study, although the significant findings from minimally adjusted models for KIM1 and
5 EDA2R, attenuated after full adjustments, results from the MR analyses showed some possible
6 causal links between these proteins and dementia. We acknowledge however that some of these
7 MR results may be biased by horizontal pleiotropy with the proteins affecting multiple
8 diseases,[54] possibly via immune, renal, and metabolic disease pathways,[54-56] which
9 subsequently contribute to the risk of dementia.[54] There was also evidence from previous
10 studies indicating higher levels of EDA2R were associated a smaller total brain volume,
11 smaller grey matter volume, and less normal-appearing white matter volume.[55]

12 The current study exhibits robustness through several key strengths. Firstly, it draws upon two
13 extensive population-based cohorts with prolonged follow-up, employing data-driven
14 proteome-wide methods that yield high-throughput and reliable proteomics data. The selection
15 of the protein panel in ELSA is also noteworthy for its focused curation on dementia-related
16 markers, enhancing the study's precision in investigating associations with dementia risk in a
17 nationally representative sample of older adults. The findings through the inclusion of the UK
18 Biobank cohort encompassing a broader selection of proteins enhance the validity of our
19 results. Both cohorts are well-characterized longitudinal cohorts, which enabled adjustment for
20 wide range of factors. Furthermore, our study benefits from applying a robust and
21 comprehensive dementia algorithm in ELSA, which integrated information from various
22 sources, bolstered by details on medication use and informant-solicited information, which has
23 been reported to correlates better with objective cognitive performance than self-report alone
24 as well as medication which particularly captured those with younger onset dementia,[57]
25 allowing for accurate identification of incident dementia cases and the exclusion of prevalent

1 cases. Another significant strength lies in our approach to assessing protein-dementia
2 associations through a range of established methods for evidence triangulation. Additionally,
3 the utilization of Olink antibody-based PEA is recognized for its superior specificity in
4 proteomics assays.[17] The integration of proteogenomic in MR analyses was an additional
5 strength.[12]

6 Some limitations should be acknowledged. First, several circulating protein markers potentially
7 relevant to dementia, for example, proteins such as GFAP and GDF-15, were not assayed in
8 ELSA. There is also limited specification, or a lack of protein measurements from the A/T/N
9 classification framework,[58] such as beta-amyloid, p-tau217 and p-tau181. Third, another
10 source of limitation is that the algorithm used for dementia ascertainment lacked information
11 extracted from primary care data, and uncertainties exist ascertaining dementia subtypes.
12 Nevertheless, it is important to note that AD and VAD pathology often co-exist on a population
13 level, and many dementia patients exhibit mixed neuropathology.[59] Fourth, it is important to
14 note that we lacked external validation cohorts for RPS6KB1. Lastly, there are inherent
15 assumptions in MR analyses, and for drug target MR specifically, genetics might not directly
16 inform on specific pharmacological aspects of drug exposure.

17 In conclusion, in our proteome-wide study from two UK-based large-scale population-based
18 cohorts, we demonstrated the utility of data-driven proteomics analyses in identifying novel
19 targets for dementia. MR analyses leveraging extensive GWAS data substantiated some of
20 these protein-dementia relationships with causal evidence. Looking forward, given the cost-
21 effectiveness and minimal invasiveness of peripheral blood-based biomarkers, developing a
22 panel of blood-based protein markers could greatly enhance dementia diagnostics in routine
23 clinical practice. Integrating large-scale population-based proteomics with other omics, such
24 as genomics,[12] offers immense potential for deeper biological insights into diseases.
25 Additionally, several proteins identified in our study emerge as promising candidates for future

1 clinical trials, addressing the urgent need for more treatment options amid the growing global
2 burden of dementia.

3

4 **ONLINE METHODS**

5 **ELSA discovery cohort study population**

6 ELSA is a nationally representative sample of men and women aged 50 years and over living
7 in the community in England. Data collection started in 2002-2003, with participants re-
8 interviewed every two years. Details of survey design are available elsewhere.[60] Blood
9 sample collection in ELSA took place for the first time in wave 2 nurse visit in 2004-2005 and
10 subsequently in every four-year interval. The blood collected from wave 4 nurse visits in 2008-
11 2009 were used for the proteomics profiling (thus forming the baseline sample of this study),
12 thereby affording a unique temporal perspective that allows for the exploration of the
13 relationship between protein concentrations and ADRD over a span exceeding 10 years. The
14 following exclusion criteria were applied: 1) participants who died within 2 years of the wave
15 4 nurse visit; or 2) participants lost to follow-up (missing at ≥ 2 waves). A total of 3,305
16 available plasma samples from wave 4 were retrieved for the proteomics profiling.

17

18 **Measurement of plasma proteins in ELSA**

19 The proteomics dataset in ELSA was curated with a specific focus on investigating the
20 underlying biological processes associated with ADRD and cognitive decline. The
21 comprehensive assays encompassed an extensive array of cardiovascular and inflammatory
22 markers, in addition to markers integral to neurological processes such as axon guidance,
23 neurogenesis, and synapse assembly. These analyses were conducted utilizing the Olink
24 proteomics platform, the antibody based Olink PEA technology.[61]

1 We used Olink Target 96 Neurology, Cardiovascular II and Neurology Exploratory arrays in
2 these analyses. The cardiovascular array includes major inflammatory and cardiometabolic
3 pathways relevant to dementia risk, while the neurology array has been designed to include
4 proteins implicated in brain aging. Frozen samples were shipped to Olink for aliquoting,
5 plating, and assays. These assays include a built-in quality control based on four internal
6 controls that are spiked into all samples, and external controls.

7 Following stringent data quality control (Supplementary Methods), proteins were measured
8 across three panels containing 276 proteins. Proteins were presented as NPX values, the
9 arbitrary unit on \log^2 scale from Olink proteomics.

10

11 **Covariates assessment in ELSA**

12 Baseline (at wave 4, 2008-2009) sociodemographic and socioeconomic covariates included age
13 (in years), biological sex (male vs female) and ethnicity (white vs other ethnic groups) were
14 self-reported. The age that participants left formal education was coded as follows: none, age
15 14 or under, 15, 16, 17, 18, 19 or over. Smoking status was self-reported and was categorized
16 as never, former, and current smoker. Physician-diagnosed cardiovascular disease (heart attack,
17 angina, or heart failure) was self-reported. Depression was also self-reported. BMI was
18 calculated using participant's height and weight measured during the wave 4 nurse visit. Three
19 measurements were taken of systolic and diastolic pressure on the respondent's right arm while
20 they were seated, and the average of the three measurements was used. LDL cholesterol was
21 assayed using the blood sample collected by the nurse. The APOE genotype in ELSA was
22 derived from the analysis of two specific SNPs, namely rs7412 and rs429358. To determine
23 these genotypes, two TaqMan assays from Assay-On-Demand, a product of Applied
24 Biosystems and Gene service Ltd in Cambridge, UK, were employed. These assays were
25 conducted on a 7900HT analyzer, manufactured by Applied Biosystems, and the genotypes

1 were determined using the Sequence Detection Software (version 2.0), also from Applied
2 Biosystems. The quality control of genome-wide genotyping has been described
3 elsewhere.[62] Episodic memory was assessed at wave 4, evaluated through the immediate and
4 delayed recall tasks of the Consortium to Establish a Registry for AD (CERAD). Participants
5 were presented with a ten-word list and tasked with recalling it both immediately and after a
6 delay. The scores from these tests were then aggregated to compute a memory score.

7

8 **Dementia algorithm in ELSA**

9 The standardized algorithm for identifying dementia cases relied on five primary data sources:
10 1) coded information extracted from interviews across all waves using participant self-reported
11 physician diagnosis of AD and dementia; 2) caregivers who completed a modified short-form
12 Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE); 3) medication data
13 collected during nurse visits (wave 6, 8, and 9); 4) linked data from hospital admissions (NHS
14 Hospital Episode Statistics) and 5) mortality records (Office for National Statistics Mortality
15 Statistics). All data sources were integrated into the algorithm development process.

16 Regarding self-reported diagnosis of dementia, during each interview, participants were asked
17 whether a medical professional had informed them of a diagnosis of AD, dementia, organic
18 brain senility, or any other serious memory condition. A positive response indicates the
19 presence of dementia.

20 Caregivers completed a modified short-form IQCODE on behalf of individuals unable to
21 respond independently. Caregivers (acting as proxies) were instructed to assess the current
22 functional performance of the participant, comparing it to that of two years prior, instead of the
23 standard 10-year interval. Consistent with prior research, individuals with an IQCODE score
24 exceeding 3.38 were classified as having dementia.[63]

1 Medication data, encoded for analysis, was collected during nurse visits in waves 6 (2012-
2 2013) and wave 8 (2016-2017), and wave 9 (2018-2019). In waves 8 and 9, two mutually
3 exclusive subsets of the sample underwent nurse visits. As specified in the British National
4 Formulary (BNF), we specifically considered four common drugs for dementia: Donepezil
5 hydrochloride (0411000D0), Galantamine (0411000F0), Memantine hydrochloride
6 (0411000G0), and Rivastigmine (0411000E0).

7 The 10th revision of the International Statistical Classification of Diseases and Related Health
8 Problems (ICD-10) was employed to extract dementia-related outcomes from the National
9 Health Service (NHS) Hospital Episode Statistics (HES) data (see Supplementary Table 7A for
10 ICD-10 codes used for ascertaining all-cause dementia); HES data covered the period between
11 March 1997 and January 2018. Mortality statistics provided by the Office for National Statistics
12 (ONS) were accessible for the period between April 2002 and April 2018. The methodology
13 used to determine ACD using the mortality-linked data mirrored the approach taken with HES
14 data based on ICD codes.

15 Dementia subtypes, specifically AD and VAD, were also derived (Supplementary Table 7B).
16 In the case of AD, all four data sources were utilized: self-reported physician-diagnosis of AD
17 during each study interview, medication usage based on BNF from nurse visits, and the
18 presence of ICD-10 codes in HES and mortality data. For ascertaining VAD, ICD-10 codes
19 from HES and mortality data were used.

20 The computation of time-to-event for dementia cases depended on the data source where the
21 event of dementia was first recorded. The date of censoring for non-dementia cases varied
22 across participant groups. Details can be found in the Supplementary Methods.

23

24 **UK Biobank for protein validation**

1 UK Biobank is a large population-based cohort from the UK with over half a million
2 participants aged 40 – 69 years, recruited between 2006 and 2010.[18] The UK Biobank
3 Pharma Proteomics Project (UKB-PPP) is a consortium of 13 biopharmaceutical companies
4 which funded the proteomic profiling on blood plasma samples. Proteomics profiling was
5 conducted in 54,219 participants at study baseline, with 2,923 unique proteins assayed using
6 the antibody based Olink Explore 3072 PEA, across eight Olink panels (Cardiometabolic I,
7 Cardiometabolic II, Inflammation I, Inflammation II, Neurology I, Neurology II, and Oncology
8 I, Oncology II). Consortium members opted for samples enriched in specific diseases of
9 interest, while the remaining population was randomly sampled using a stratified approach
10 based on age, sex, and recruitment center.[18] The current analysis excluded those with
11 dementia at study baseline, which yielded a total sample of 52,745 individuals.

12 We have attempted, where possible, to derive similar variables for both the ELSA and biobank
13 cohorts with consideration of the level of missingness, to maximize comparability. Participant's
14 age was derived based on date of birth and date of attending an initial assessment center.
15 Participant's biological sex was acquired from central registry at recruitment and contains a
16 mixture of the sex recorded by the NHS and self-reported sex. Ethnicity was self-reported and
17 categorized into White, Mixed, Asian, or Asian British, and Black or Black British. Highest
18 qualification was determined by the answers provided to the question: "Which of the following
19 qualifications do you have?", with options included: College or University degree; NVQ
20 (National Vocational Qualification) or HND (Higher National Diploma) or HNC (Higher
21 National Certificate) or equivalent; other professional qualifications e.g.: nursing; A levels/AS
22 levels; O levels/GCSEs (General Certificate of Secondary Education) or equivalent; CSEs
23 (General Certificate of Secondary Education) or equivalent; or none of the above. Smoking
24 status was self-reported and constructed using information about the current/past smoking
25 status of the participant and categorized as never, former, and current smoker. Self-reported

1 medical conditions were solicited through the touchscreen questionnaire as well as during
2 verbal interview conducted by a trained nurse, and the presence of cardiovascular disease (heart
3 attack, angina, or heart failure) were defined if the participant reported any of these conditions.
4 Depression was affirmatory if the participant confirmed to have any of probable recurrent
5 major depression (severe), probable recurrent major depression (moderate), or single probable
6 major depression episode, if reported on the questionnaire or nurse-administered verbal
7 interview. BMI was constructed from height and weight measured during the initial assessment
8 center visit using an Omron device. Two automatic readings of blood pressure were taken a
9 few moments apart, using an Omron HEM-7015IT digital blood pressure monitor, and the
10 average of the readings were used. LDL cholesterol was measured by enzymatic protective
11 selection analysis on a Beckman Coulter AU5800 from the blood sample collected at
12 recruitment.

13 ACD, AD, VAD and FTD were defined by the UK Biobank dementia algorithm.[64] The last
14 date of censoring was 31st December 2022 (last date of linkage to death and inpatient records).
15 Baseline dementia was removed.

16

17 **Statistical analysis**

18 **Analysis 1: Protein-dementia associations in the ELSA discovery cohort**

19 We used Cox proportional hazards regression models to evaluate the associations between each
20 plasma protein NPX value (scaled) and incident dementia using the ‘survival’ R package. Rank-
21 based inverse normal transformation was first applied to the protein levels and scaled to have
22 a mean of 0 and standard deviation of 1 prior to all analyses.[65]

23 All proteins had $\leq 6\%$ missing. Missing protein measurements were imputed using the K-
24 nearest neighbor ($k = 57$) imputation using the ‘impute’ R package,[66] which works by
25 identifying the nearest 57 individuals defined using Euclidean distances and imputing with

1 their medians, with k calculated from the square root of the total sample size ($N = 3,249$).

2 Missingness was imputed for proteomics and clinical data independently, where clinical data

3 was imputed using the Multiple Imputation by Chained Equations procedure ('mice' R

4 package),[67] with 30 imputed datasets and 10 iterations.

5 Based on the normalized and imputed protein data, we first assessed the protein-dementia

6 associations without any model adjustments by pooling the estimates from all 30 imputed

7 datasets. The models were further adjusted for age, sex, and ethnicity for the minimally-

8 adjusted model; and adjusted for age, sex, ethnicity, education, smoking status, depression,

9 cardiovascular disease, BMI, systolic blood pressure, and LDL cholesterol for the fully-

10 adjusted models for each protein NPX value associated with the risk of incident dementia, with

11 P_{FDR} set at a cut-off of 0.05, this translates to an uncorrected p-value of 0.00018 (with 276 tests

12 for significance for 276 proteins). All p-values were two-sided. FDR-corrected p-values

13 (denoted as P_{FDR}) were reported and displayed using volcano plot, accompanying the HR.

14 The Cox proportional hazard regression analyses were repeated for AD and VAD as the

15 outcome of interest separately.

16 A series of sensitivity analyses were conducted using the same adjustment methods as the main

17 Cox regression for incident ACD, by excluding the following: 1) dementia cases that occurred

18 during the first year of follow-up to reduce the possibility of reverse causation bias; 2) other

19 ethnic groups other than white ethnicity; 3) APOE $\epsilon 4$ carriers; and 4) participants <60 years of

20 age. We also conducted competing risk of death using Fine-Gray regression models[68] with

21 the same set of covariate adjustments, given that death may preclude dementia from occurring.

22 The competing risk models, which estimated sub-distribution HRs, were conducted for incident

23 ACD accounting for all-cause mortality as a competing risk, incorporating time to event

24 data.[68, 69] Lastly, sex differences in the protein-dementia associations were assessed by

25 fitting sex as an interaction term in the Cox regression models.[70]

1 Next, receiver operating characteristic (ROC) analyses were conducted to assess the precision
2 of the identified proteins from the fully adjusted Cox regression models in predicting incident
3 ACD. These analyses were performed independently and in combination with additional
4 factors including age, sex, ethnicity, education, APOE ϵ 4 status, and memory score. To evaluate
5 the performance of the Cox models, bootstrapping was performed to assess the stability of the
6 AUC estimates. A total of 2000 bootstrap resamples were generated. Bootstrapping involves
7 resampling with replacement from the original dataset to estimate the sampling distribution of
8 a statistic. We used the bootstrapped samples to compute 95% CI for the AUC of each model,
9 utilizing the R packages ‘caret’[71], ‘boot’[72] and ‘pROC’.[73] The mean AUC value across
10 all bootstrap resamples was calculated to estimate the overall predictive performance of the
11 survival model.

12 We utilized eXtreme Gradient Boosting (XGBoost), an advanced decision-tree ensemble
13 machine learning algorithm within a gradient boosting framework, to identify the most
14 important features predictive of future ACD onset.[74] To ensure comprehensive analysis and
15 validation, we incorporated all available protein, demographic, and clinical predictors into our
16 models, using imputed data to handle any missing values. The XGBoost model feature
17 importance scores were used to select key predictors. To explain the outputs of our XGBoost
18 models and to provide interpretability to the feature importance rankings, we employed
19 SHapley Additive exPlanations (SHAP) values. SHAP values offer a consistent and
20 theoretically sound method for interpreting the contributions of individual features to the
21 model’s predictions.[75] Furthermore, this process of evidence triangulation with XGBoost
22 and SHAP values was conducted to validate and complement the findings from our Cox
23 regression models. By comparing and cross validating, the results from these two different
24 methodological approaches, we aimed to enhance the robustness and reliability of our
25 predictive model for future ACD onset.

1

2 **Analysis 2: Protein-dementia associations in the UK Biobank validation cohort**

3 The significant proteins from the main and sensitivity analyses based on ELSA were then
4 validated using the UK Biobank proteomics data, if assayed. The reason for choosing ELSA as
5 the discovery cohort and UK Biobank as the validation cohort is that ELSA's target panels were
6 specifically designed to investigate proteomic signatures related to cognitive decline and
7 dementia, resulting in a more focused selection of proteins. In contrast, UK Biobank offers a
8 broader protein selection, allowing for effective validation.

9 The UKB-PPP proteomics samples underwent sample selection, processing, and quality
10 control procedures.[18] Missing protein measurements for the remaining individuals were
11 imputed using K-nearest neighbor imputation ($k = 230$), the protein levels were normalized
12 and scaled akin to the analyses in ELSA proteomics data described above.[65] Clinical data
13 was imputed with 30 imputed datasets and 10 iterations, and was imputed separately to the
14 proteomics data.

15 Data was processed and analyzed using the R Studio Workbench on the UK Biobank Research
16 Analysis Platform, under application No. 71702.

17

18 **Analysis 3: Two-sample bi-directional MR**

19 MR pertains to employing genetic variants to explore causal connections regarding the
20 influence of modifiable exposures on a range of outcomes. Anchored in Mendel's laws of
21 inheritance and instrumental variable estimation techniques, MR principles facilitate the
22 inference of causal effects, even in the presence of unobserved confounding factors. MR studies
23 are often compared to naturally occurring randomized trials, resembling a randomized control
24 trial where genetic factors are inherently assigned at random by nature.[76] The MR approach
25 rests on three key assumptions: 1) the instruments are correlated with the exposure; 2) the

1 instruments are linked to the outcome solely through the studied exposure (exclusion restriction
2 assumption); and 3) the instruments are independent of other factors influencing the outcome
3 (independence assumption).

4 To infer causality between dementia-associated proteins and AD, ACD, and VAD, we
5 performed bidirectional two-sample MR, leveraging summary statistics from different GWAS
6 consortia. For AD, we used GWAS summary statistics derived from three consortia:
7 (International Genomics of Alzheimer's Project (IGAP) (Kunkle et al 2019),^[77] European
8 Alzheimer & Dementia Biobank (EADB) consortium (Bellenguez et al 2022),^[78] and the
9 FinnGen 2023 study);^[79] For ACD and vascular dementia GWAS were both derived from the
10 FinnGen 2023 study.^[79] Summary details of the dementia GWAS used were included in
11 Supplementary Table 8.

12 Selection of instruments to proxy for altered protein abundance were derived using pQTL
13 mapping proteins that identifies genetic associations in participants of European ancestry from
14 the UK Biobank based on Olink data (<https://doi.org/10.7303/syn51364943>).^[18] The effects
15 of protein pQTL were standardized to align with the same effect allele.

16 To ensure the three MR assumptions were not violated, instruments for the exposure were
17 selected by using association at genome-wide significance ($P < 5 \times 10^{-8}$). SNP with a minor
18 allele frequency (MAF) $< 5\%$ were excluded. linkage disequilibrium (LD) clumping was done
19 with a window size of 10,000 kilobases [kb] at an R^2 threshold of 0.001. In cases where a
20 requested SNP from the exposure GWAS is not found in the outcome GWAS, a proxy SNP
21 with a LD coefficient of $R^2 > 0.8$ to the requested missing target SNP were sought as a
22 substitute, if available, by querying the LDLink web server. LD proxies are determined using
23 data from the 1000 Genomes European sample. The returned information includes the effect
24 of the proxy SNP on the outcome, along with details such as the proxy SNP itself, the effect
25 allele of the proxy SNP, and the corresponding allele for the target SNP. The effect of a SNP

1 on an outcome and exposure were then harmonized to be relative to the same allele. We specify
2 the inference of positive strand alleles by utilizing allele frequencies for palindromes instead
3 of eliminating palindromic variants. F-statistics were used to assess the SNP-exposure strength
4 ($F > 10$).

5 We first performed the MR in the forward direction (circulating protein concentration \rightarrow
6 dementia) on the harmonized effects to estimate the effect of genetically proxied protein
7 abundance on genetic liability to the dementia outcome of interest to be relative to the same
8 allele. Given the study outcomes were AD and related dementia, variants in the *APOE* region
9 (human genome reference builds GRCh38 - chromosome: 19, base pair position: 44,407,913 –
10 45,408,821) were removed due to its pleiotropic nature and large effect size. We then calculated
11 the effects for each individual variant by employing a two-term Taylor series expansion of the
12 Wald ratio. Subsequently, we utilized the weighted delta inverse-variance weighted (IVW)
13 method to conduct a meta-analysis of individual SNP effects, to estimate the combined effect
14 of the Wald ratios. Sensitivity analyses included using various MR methods, including MR-
15 Egger, Weighted Median, Maximum Likelihood, Weighted Mode, and leave-one-out
16 methods.[76]

17 The same method was then applied to the backward MR (dementia \rightarrow circulating protein
18 concentration), with SNPs extracted from the same GWAS as described above, except with
19 dementia being the exposure data and protein being the outcome.

20 All analyses were conducted using human genome reference build GRCh38. In instances where
21 the genome build was based on GRCh37 assembly Hg19, a lift-over process was executed to
22 convert genome coordinates and annotations to the GRCh38 using a designated alignment. All
23 MR analyses were conducted using ‘TwoSampleMR’, ‘MendelianRandomization’ and
24 ‘LDlinkR’ R packages.[80]

25

1 **Analysis 4: Two sample drug target MR (cis-MR)**

2 Next, we used a two-sample MR study design, based predominantly on genetic variants located
3 in or near genes that encode the relevant drug targets, to infer causality from protein
4 concentration \rightarrow dementia (cis-MR).[81-84] Cis-MR is considered to be less susceptible to
5 pleiotropy, and the potential effect of a drug by analyzing the genomic locus encoding protein
6 targets, which may be informative for drug trial design.[82, 84]

7 The selection of the primary cis-MR models included SNPs successfully harmonized with the
8 gene encoding regions, with the flanking region being within 10-kilobase (kB) in either
9 direction of the start and stop coordinates for the genes, according to Human Genome reference
10 release GRCh38. LD clumping was then done to remove the excess of most highly correlated
11 variants at each locus, with R^2 threshold set at < 0.001 , retaining the SNPs with the strongest
12 associations with the protein of interest filtered by $P < 5 \times 10^{-5}$. The sensitivity analyses which
13 took a more liberal approach, selecting valid instruments at the threshold of $P < 5 \times 10^{-5}$ and
14 clumped at $R^2 < 0.01$.

15 All analyses were done on human genome reference builds - GRCh38, lift over was performed
16 from GRCh 37 assembly Hg19 to GRCh 38, using CrossMap 0.7.0 in Python (version 3.12).
17 Analysis 1 to 4 were done using R Studio (version 4.4.0).

18

19 **Analysis 5: Enrichment analysis**

20 Enrichment analysis was conducted by searching open-source databases to further characterize
21 the identified proteins from the Cox proportional hazard regressions. We employed the
22 Enrichr,[85] which is a computational method infers knowledge about an input gene set by
23 comparing it to annotated gene sets that represent existing biological knowledge. It determines
24 if the input set of genes significantly overlaps with these annotated gene sets. We used the full
25 set of ELSA proteins as the background gene set, to glean a deeper biological understanding.

1 We searched the following bioinformatics databases: Gene Ontology (GO): GO Molecular
2 Function, GO Biological Process, and GO Cellular Component [86], Kyoto Encyclopedia of
3 Genes and Genomes (KEGG) [87], Reactome Pathway Database (REACTOME) [88],
4 Illuminating the Druggable Genome (IDG) [89], Proteomics Drug Atlas (PDA) [90], and
5 Genotype-Tissue Expression (GTEx) [91]. Statistical significance was indicated if $P_{\text{FDR}} < 0.05$,
6 considering multiple testing.
7 Furthermore, we utilized the Open Targets platform (<https://www.opentargets.org/>) for the
8 systematic identification of potential therapeutic drug targets among the identified proteins.
9 [92]

1 **References**

- 2 1. Shi Y, Holtzman DM. Interplay between innate immunity and Alzheimer disease:
3 APOE and TREM2 in the spotlight. *Nature Reviews Immunology*. 2018;18(12):759-72.
- 4 2. Yousef H, Czupalla CJ, Lee D, Chen MB, Burke AN, Zera KA, et al. Aged blood
5 impairs hippocampal neural precursor activity and activates microglia via brain endothelial cell
6 VCAM1. *Nature medicine*. 2019;25(6):988-1000.
- 7 3. Suhre K, McCarthy MI, Schwenk JM. Genetics meets proteomics: perspectives for
8 large population-based studies. *Nature Reviews Genetics*. 2021;22(1):19-37.
- 9 4. Lindbohm JV, Mars N, Walker KA, Singh-Manoux A, Livingston G, Brunner EJ, et al.
10 Plasma proteins, cognitive decline, and 20-year risk of dementia in the Whitehall II and
11 Atherosclerosis Risk in Communities studies. *Alzheimer's & dementia*. 2022;18(4):612-24.
- 12 5. Gomes B, Ashley EA. Artificial intelligence in molecular medicine. *New England*
13 *Journal of Medicine*. 2023;388(26):2456-65.
- 14 6. Teunissen CE, Verberk IM, Thijssen EH, Vermunt L, Hansson O, Zetterberg H, et al.
15 Blood-based biomarkers for Alzheimer's disease: towards clinical implementation. *The Lancet*
16 *Neurology*. 2022;21(1):66-77.
- 17 7. Zetterberg H. Biofluid-based biomarkers for Alzheimer's disease-related pathologies:
18 An update and synthesis of the literature. *Alzheimer's & Dementia*. 2022;18(9):1687-93.
- 19 8. Pietzner M, Wheeler E, Carrasco-Zanini J, Cortes A, Koprulu M, Wörheide MA, et al.
20 Mapping the proteo-genomic convergence of human diseases. *Science*.
21 2021;374(6569):eabj1541.
- 22 9. Walker KA, Chen J, Zhang J, Fornage M, Yang Y, Zhou L, et al. Large-scale plasma
23 proteomic analysis identifies proteins and pathways associated with dementia risk. *Nature*
24 *Aging*. 2021;1(5):473-89.
- 25 10. Walker KA, Chen J, Shi L, Yang Y, Fornage M, Zhou L, et al. Proteomics analysis of
26 plasma from middle-aged adults identifies protein markers of dementia risk in later life.
27 *Science translational medicine*. 2023;15(705):eadf5681.
- 28 11. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive
29 map of molecular drug targets. *Nature reviews Drug discovery*. 2017;16(1):19-34.
- 30 12. Sun BB, Suhre K, Gibson BW. Promises and challenges of populational proteomics in
31 health and disease. *Molecular & Cellular Proteomics*. 2024:100786.
- 32 13. Assarsson E, Lundberg M, Holmquist G, Björkstén J, Bucht Thorsen S, Ekman D, et
33 al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and
34 excellent scalability. *PloS one*. 2014;9(4):e95192.
- 35 14. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody E, et al. Aptamer-based multiplexed
36 proteomic technology for biomarker discovery. *Nature Precedings*. 2010:1-.
- 37 15. Kivimäki M, Walker KA, Pentti J, Nyberg ST, Mars N, Vahtera J, et al. Cognitive
38 stimulation in the workplace, plasma proteins, and risk of dementia: three analyses of
39 population cohort studies. *bmj*. 2021;374.
- 40 16. Guo Y, You J, Zhang Y, Liu W-S, Huang Y-Y, Zhang Y-R, et al. Plasma proteomic
41 profiles predict future dementia in healthy adults. *Nature Aging*. 2024.
- 42 17. Eldjarn GH, Ferkingstad E, Lund SH, Helgason H, Magnusson OT, Gunnarsdottir K,
43 et al. Large-scale plasma proteomics comparisons through genetics and disease associations.
44 *Nature*. 2023;622(7982):348-58.
- 45 18. Sun BB, Chiou J, Traylor M, Benner C, Hsu Y-H, Richardson TG, et al. Plasma
46 proteomic associations with genetics and health in the UK Biobank. *Nature*.
47 2023;622(7982):329-38.
- 48 19. Ljungqvist J, Zetterberg H, Mitsis M, Blennow K, Skoglund T. Serum neurofilament
49 light protein as a marker for diffuse axonal injury: results from a case series study. *Journal of*
50 *neurotrauma*. 2017;34(5):1124-7.

- 1 20. Graham NS, Zimmerman KA, Moro F, Heslegrave A, Maillard SA, Bernini A, et al.
2 Axonal marker neurofilament light predicts long-term outcomes and progressive
3 neurodegeneration after traumatic brain injury. *Science translational medicine*.
4 2021;13(613):eabg9922.
- 5 21. Olsson B, Portelius E, Cullen NC, Sandelius Å, Zetterberg H, Andreasson U, et al.
6 Association of cerebrospinal fluid neurofilament light protein levels with cognition in patients
7 with dementia, motor neuron disease, and movement disorders. *JAMA neurology*.
8 2019;76(3):318-25.
- 9 22. Ashton NJ, Janelidze S, Al Khleifat A, Leuzy A, van der Ende EL, Karikari TK, et al.
10 A multicentre validation study of the diagnostic value of plasma neurofilament light. *Nature*
11 *communications*. 2021;12(1):3400.
- 12 23. Marks JD, Syrjanen JA, Graff-Radford J, Petersen RC, Machulda MM, Campbell MR,
13 et al. Comparison of plasma neurofilament light and total tau as neurodegeneration markers:
14 associations with cognitive and neuroimaging outcomes. *Alzheimer's Research & Therapy*.
15 2021;13:1-14.
- 16 24. Illán-Gala I, Lleo A, Karydas A, Staffaroni AM, Zetterberg H, Sivasankaran R, et al.
17 Plasma tau and neurofilament light in frontotemporal lobar degeneration and Alzheimer
18 disease. *Neurology*. 2021;96(5):e671-e83.
- 19 25. Novakova L, Zetterberg H, Sundström P, Axelsson M, Khademi M, Gunnarsson M, et
20 al. Monitoring disease activity in multiple sclerosis using serum neurofilament light protein.
21 *Neurology*. 2017;89(22):2230-7.
- 22 26. Gisslén M, Price RW, Andreasson U, Norgren N, Nilsson S, Hagberg L, et al. Plasma
23 concentration of the neurofilament light protein (NFL) is a biomarker of CNS injury in HIV
24 infection: a cross-sectional study. *EBioMedicine*. 2016;3:135-40.
- 25 27. Zetterberg H, Hietala MA, Jonsson M, Andreasen N, Styurd E, Karlsson I, et al.
26 Neurochemical aftermath of amateur boxing. *Archives of neurology*. 2006;63(9):1277-80.
- 27 28. Kuhle J, Gaiottino J, Leppert D, Petzold A, Bestwick JP, Malaspina A, et al. Serum
28 neurofilament light chain is a biomarker of human spinal cord injury severity and outcome.
29 *Journal of Neurology, Neurosurgery & Psychiatry*. 2015;86(3):273-9.
- 30 29. Wingo AP, Fan W, Duong DM, Gerasimov ES, Dammer EB, Liu Y, et al. Shared
31 proteomic effects of cerebral atherosclerosis and Alzheimer's disease on the human brain.
32 *Nature neuroscience*. 2020;23(6):696-700.
- 33 30. Bahrami-B F, Ataie-Kachoei P, Pourgholami MH, Morris DL. p70 Ribosomal protein
34 S6 kinase (Rps6kb1): an update. *Journal of clinical pathology*. 2014;67(12):1019-25.
- 35 31. Mannick JB, Lamming DW. Targeting the biology of aging with mTOR inhibitors.
36 *Nature aging*. 2023;3(6):642-60.
- 37 32. Ma T, Hoeffler CA, Capetillo-Zarate E, Yu F, Wong H, Lin MT, et al. Dysregulation of
38 the mTOR pathway mediates impairment of synaptic plasticity in a mouse model of
39 Alzheimer's disease. *PloS one*. 2010;5(9):e12845.
- 40 33. Jaworski J, Spangler S, Seeburg DP, Hoogenraad CC, Sheng M. Control of dendritic
41 arborization by the phosphoinositide-3'-kinase–Akt–mammalian target of rapamycin pathway.
42 *Journal of Neuroscience*. 2005;25(49):11300-12.
- 43 34. Hoeffler CA, Klann E. mTOR signaling: at the crossroads of plasticity, memory and
44 disease. *Trends in neurosciences*. 2010;33(2):67-75.
- 45 35. Yates SC, Zafar A, Hubbard P, Nagy S, Durant S, Bicknell R, et al. Dysfunction of the
46 mTOR pathway is a risk factor for Alzheimer's disease. *Acta neuropathologica*
47 *communications*. 2013;1:1-15.
- 48 36. Wang Y, Zhang H. Regulation of autophagy by mTOR signaling pathway. *Autophagy:*
49 *Biology and diseases: Basic science*. 2019:67-83.

- 1 37. Magnuson B, Ekim B, Fingar DC. Regulation and function of ribosomal protein S6
2 kinase (S6K) within mTOR signalling networks. *Biochemical Journal*. 2012;441(1):1-21.
- 3 38. Hunsberger HC, Pinky PD, Smith W, Suppiramaniam V, Reed MN. The role of APOE4
4 in Alzheimer's disease: strategies for future therapeutic interventions. *Neuronal signaling*.
5 2019;3(2):NS20180203.
- 6 39. Jung CH, Ro S-H, Cao J, Otto NM, Kim D-H. mTOR regulation of autophagy. *FEBS*
7 *letters*. 2010;584(7):1287-95.
- 8 40. Al-Ali H, Ding Y, Slepak T, Wu W, Sun Y, Martinez Y, et al. The mTOR substrate S6
9 kinase 1 (S6K1) is a negative regulator of axon regeneration and a potential drug target for
10 central nervous system injury. *Journal of neuroscience*. 2017;37(30):7079-95.
- 11 41. Venkatasubramani JP, Subramanyam P, Pal R, Reddy BK, Srinivasan DJ, Chattarji S,
12 et al. N-terminal variant Asp14Asn of the human p70 S6 Kinase 1 enhances translational
13 signaling causing different effects in developing and mature neuronal cells. *Neurobiology of*
14 *Learning and Memory*. 2020;171:107203.
- 15 42. Upadhyaya Y, Xie L, Salama P, Cao S, Nho K, Saykin AJ, et al. Differential co-
16 expression analysis reveals early stage transcriptomic decoupling in alzheimer's disease. *BMC*
17 *Medical Genomics*. 2020;13:1-10.
- 18 43. Sternlicht MD, Werb Z. How matrix metalloproteinases regulate cell behavior. *Annual*
19 *review of cell and developmental biology*. 2001;17(1):463-516.
- 20 44. Chen Q, Jin M, Yang F, Zhu J, Xiao Q, Zhang L. Matrix metalloproteinases:
21 inflammatory regulators of cell behaviors in vascular formation and remodeling. *Mediators of*
22 *inflammation*. 2013;2013.
- 23 45. Brkic M, Balusu S, Libert C, Vandenbroucke RE. Friends or foes: matrix
24 metalloproteinases and their multifaceted roles in neurodegenerative diseases. *Mediators of*
25 *inflammation*. 2015;2015.
- 26 46. Lakhan SE, Kirchgessner A, Tepper D. Matrix metalloproteinases and blood-brain
27 barrier disruption in acute ischemic stroke. *Frontiers in neurology*. 2013;4:42109.
- 28 47. Hartz AM, Bauer B, Soldner EL, Wolf A, Boy S, Backhaus R, et al. Amyloid- β
29 contributes to blood-brain barrier leakage in transgenic human amyloid precursor protein mice
30 and in humans with cerebral amyloid angiopathy. *Stroke*. 2012;43(2):514-23.
- 31 48. Horstmann S, Budig L, Gardner H, Koziol J, Deuschle M, Schilling C, Wagner S.
32 Matrix metalloproteinases in peripheral blood and cerebrospinal fluid in patients with
33 Alzheimer's disease. *International psychogeriatrics*. 2010;22(6):966-72.
- 34 49. Hunninghake GM, Cho MH, Tesfaigzi Y, Soto-Quiros ME, Avila L, Lasky-Su J, et al.
35 MMP12, lung function, and COPD in high-risk populations. *New England Journal of*
36 *Medicine*. 2009;361(27):2599-608.
- 37 50. Olejarz W, Łacheta D, Kubiak-Tomaszewska G. Matrix metalloproteinases as
38 biomarkers of atherosclerotic plaque instability. *International journal of molecular sciences*.
39 2020;21(11):3946.
- 40 51. Iyer RP, Patterson NL, Zouein FA, Ma Y, Dive V, de Castro Brás LE, Lindsey ML.
41 Early matrix metalloproteinase-12 inhibition worsens post-myocardial infarction cardiac
42 dysfunction by delaying inflammation resolution. *International journal of cardiology*.
43 2015;185:198-208.
- 44 52. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic
45 atlas of the human plasma proteome. *Nature*. 2018;558(7708):73-9.
- 46 53. Pijet B, Konopka A, Rejmak E, Stefaniuk M, Khomiak D, Bulska E, et al. The matrix
47 metalloproteinase inhibitor marimastat inhibits seizures in a model of kainic acid-induced
48 status epilepticus. *Scientific Reports*. 2020;10(1):21314.
- 49 54. You J, Guo Y, Zhang Y, Kang J-J, Wang L-B, Feng J-F, et al. Plasma proteomic profiles
50 predict individual future health risk. *Nature Communications*. 2023;14(1):7817.

- 1 55. Harris SE, Cox SR, Bell S, Marioni RE, Prins BP, Pattie A, et al. Neurology-related
2 protein biomarkers are associated with cognitive ability and brain volume in older age. *Nature*
3 *Communications*. 2020;11(1):800.
- 4 56. Llaurador-Coll M, Rios S, García-Gavilán JF, Babio N, Vilella E, Salas-Salvadó J.
5 Plasma levels of neurology-related proteins are associated with cognitive performance in an
6 older population with overweight/obesity and metabolic syndrome. *GeroScience*.
7 2023;45(4):2457-70.
- 8 57. Nichols E, Gross AL, Zhang YS, Meijer E, Hayat S, Steptoe A, et al. Considerations
9 for the use of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) in
10 cross-country comparisons of cognitive aging and dementia. *Alzheimer's & Dementia*. 2024.
- 11 58. van der Flier WM, Scheltens P. The ATN framework—moving preclinical Alzheimer
12 disease to clinical relevance. *JAMA neurology*. 2022;79(10):968-70.
- 13 59. Langa KM, Foster NL, Larson EB. Mixed dementia: emerging concepts and therapeutic
14 implications. *Jama*. 2004;292(23):2901-8.
- 15 60. Steptoe A, Breeze E, Banks J, Nazroo J. Cohort profile: the English longitudinal study
16 of ageing. *International journal of epidemiology*. 2013;42(6):1640-8.
- 17 61. Wik L, Nordberg N, Broberg J, Björkstén J, Assarsson E, Henriksson S, et al.
18 Proximity extension assay in combination with next-generation sequencing for high-
19 throughput proteome-wide analysis. *Molecular & Cellular Proteomics*. 2021;20.
- 20 62. Ajnakina O, Cadar D, Steptoe A. Interplay between socioeconomic markers and
21 polygenic predisposition on timing of dementia diagnosis. *Journal of the American Geriatrics*
22 *Society*. 2020;68(7):1529-36.
- 23 63. Almeida-Meza P, Steptoe A, Cadar D. Markers of cognitive reserve and dementia
24 incidence in the English Longitudinal Study of Ageing. *The British Journal of Psychiatry*.
25 2021;218(5):243-51.
- 26 64. Wilkinson T, Schnier C, Bush K, Rannikmäe K, Henshall DE, Lerpiniere C, et al.
27 Identifying dementia outcomes in UK Biobank: a validation study of primary care, hospital
28 admissions and mortality data. *European journal of epidemiology*. 2019;34:557-65.
- 29 65. Gadd DA, Hillary RF, Kuncheva Z, Mangelis T, Cheng Y, Dissanayake M, et al. Blood
30 protein levels predict leading incident diseases and mortality in UK Biobank. *medRxiv*.
31 2023:2023.05.01.23288879. Preprint at <https://doi.org/10.1101/2023.05.01.23288879>.
- 32 66. Hastie T, Tibshirani R, Narasimhan B, Chu G. impute: impute: Imputation for
33 microarray data. R package version; 2020.
- 34 67. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained
35 equations in R. *Journal of statistical software*. 2011;45:1-67.
- 36 68. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing
37 risk. *Journal of the American statistical association*. 1999;94(446):496-509.
- 38 69. Austin PC, Ibrahim M, Putter H. Accounting for Competing Risks in Clinical Research.
39 *JAMA*. 2024.
- 40 70. Gong J, Harris K, Peters SA, Woodward M. Sex differences in the association between
41 major cardiovascular risk factors in midlife and dementia: a cohort study using data from the
42 UK Biobank. *BMC medicine*. 2021;19:1-11.
- 43 71. Kuhn M. Building predictive models in R using the caret package. *Journal of statistical*
44 *software*. 2008;28:1-26.
- 45 72. Canty A, Ripley B. Package ‘boot’. *Bootstrap Functions CRAN R Proj*. 2017.
- 46 73. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an
47 open-source package for R and S+ to analyze and compare ROC curves. *BMC bioinformatics*.
48 2011;12:1-8.
- 49 74. Chen T, Guestrin C, editors. Xgboost: A scalable tree boosting system. *Proceedings of*
50 *the 22nd acm sigkdd international conference on knowledge discovery and data mining*; 2016.

- 1 75. Sundararajan M, Najmi A, editors. The many Shapley values for model explanation.
2 International conference on machine learning; 2020: PMLR.
- 3 76. Burgess S, Smith GD, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines
4 for performing Mendelian randomization investigations: update for summer 2023. Wellcome
5 open research. 2019;4.
- 6 77. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-
7 analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A β , tau,
8 immunity and lipid processing. *Nature genetics*. 2019;51(3):414-30.
- 9 78. Bellenguez C, Küçükali F, Jansen IE, Kleindam L, Moreno-Grau S, Amin N, et al.
10 New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature*
11 *Genetics*. 2022;54(4):412-36.
- 12 79. Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, et al. FinnGen
13 provides genetic insights from a well-phenotyped isolated population. *Nature*.
14 2023;613(7944):508-18.
- 15 80. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing
16 Mendelian randomization analyses using summarized data. *International journal of*
17 *epidemiology*. 2017;46(6):1734-9.
- 18 81. Williams DM, Finan C, Schmidt AF, Burgess S, Hingorani AD. Lipid lowering and
19 Alzheimer disease risk: a Mendelian randomization study. *Annals of neurology*. 2020;87(1):30-
20 9.
- 21 82. Swerdlow DI, Kuchenbaecker KB, Shah S, Sofat R, Holmes MV, White J, et al.
22 Selecting instruments for Mendelian randomization in the wake of genome-wide association
23 studies. *International journal of epidemiology*. 2016;45(5):1600-16.
- 24 83. Burgess S, Mason AM, Grant AJ, Slob EA, Gkatzionis A, Zuber V, et al. Using genetic
25 association data to guide drug discovery and development: Review of methods and
26 applications. *The American Journal of Human Genetics*. 2023;110(2):195-214.
- 27 84. Patel A, Gill D, Newcombe P, Burgess S. Conditional inference in cis-Mendelian
28 randomization using weak genetic factors. *Biometrics*. 2023.
- 29 85. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive
30 and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*.
31 2013;14(1):128.
- 32 86. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene
33 ontology: tool for the unification of biology. *Nature genetics*. 2000;25(1):25-9.
- 34 87. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking
35 genomes to life and the environment. *Nucleic acids research*. 2007;36(suppl_1):D480-D4.
- 36 88. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome
37 pathway knowledgebase. *Nucleic acids research*. 2020;48(D1):D498-D503.
- 38 89. Rodgers G, Austin C, Anderson J, Pawlyk A, Colvis C, Margolis R, Baker J. Glimmers
39 in illuminating the druggable genome. *Nature reviews Drug discovery*. 2018;17(5):301-2.
- 40 90. Mitchell DC, Kuljanin M, Li J, Van Vranken JG, Bulloch N, Schweppe DK, et al. A
41 proteome-wide atlas of drug mechanism of action. *Nature Biotechnology*. 2023;41(6):845-57.
- 42 91. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across
43 human tissues. *Science*. 2020;369(6509):1318-30.
- 44 92. Ochoa D, Hercules A, Carmona M, Suveges D, Gonzalez-Uriarte A, Malangone C, et
45 al. Open Targets Platform: supporting systematic drug–target identification and prioritisation.
46 *Nucleic acids research*. 2021;49(D1):D1302-D10.

47

1 **Acknowledgement**

2 The National Institute of Aging (NIA) (grant No. [R01AG17644]) funded the proteomics data
3 curation in ELSA. J.G. is supported by the NIA (grant No. [R01AG17644]). D.M.W. is
4 supported by the British Medical Research Council (MRC) (grant No. [MC_UU_00019/3]).

5 This research has been conducted using the UK Biobank Resource under application number
6 71702. This work uses data provided by patients and collected by the NHS as part of their care
7 and support. Copyright © 2023, NHS England. Re-used with the permission of the NHS
8 England and UK Biobank. All rights reserved.

9 We would like to thank the participants in ELSA for their contribution to the research. We
10 additionally want to acknowledge the participants and investigators of the FinnGen study. We
11 would like to thank Olink representatives and the Newcastle laboratory for their support for
12 data procurement and data pre-processing.

13

14 **Conflict of interests**

15 Olink had no part in designing the study or analyzing the data. No conflicts of interest to be
16 declared from any of the authors.

17

18 **Author contributions**

19 The study's conceptualization and design were undertaken by A.S., P.Z., and J.G. Proteomics
20 data curation in ELSA was a joint effort between J.G. and A.S. J.G. handled all data processing
21 and analysis and wrote the initial manuscript draft. A.S. and D.M.W. provided oversight and
22 guidance for statistical analyses and theoretical frameworks. D.M.W. provided access to the
23 UK Biobank data, under project application No. 71702. J.G. led the analysis and interpretation
24 of data, as well as manuscript revisions. J.G. led the development of the dementia algorithm in
25 ELSA, with contributions from A.S., S.S., S.A., F.B., and S.H. All authors participated in the

1 manuscript's revision process and collectively accepted responsibility for its submission for
2 publication.

3

4 **Data availability**

5 The ELSA data is available on UK Data Service. The proteomics data in ELSA will be
6 deposited on UK Data Service upon publication. All data from the UK Biobank, including the
7 proteomics data is available by directly submitting a project application to the UK Biobank.

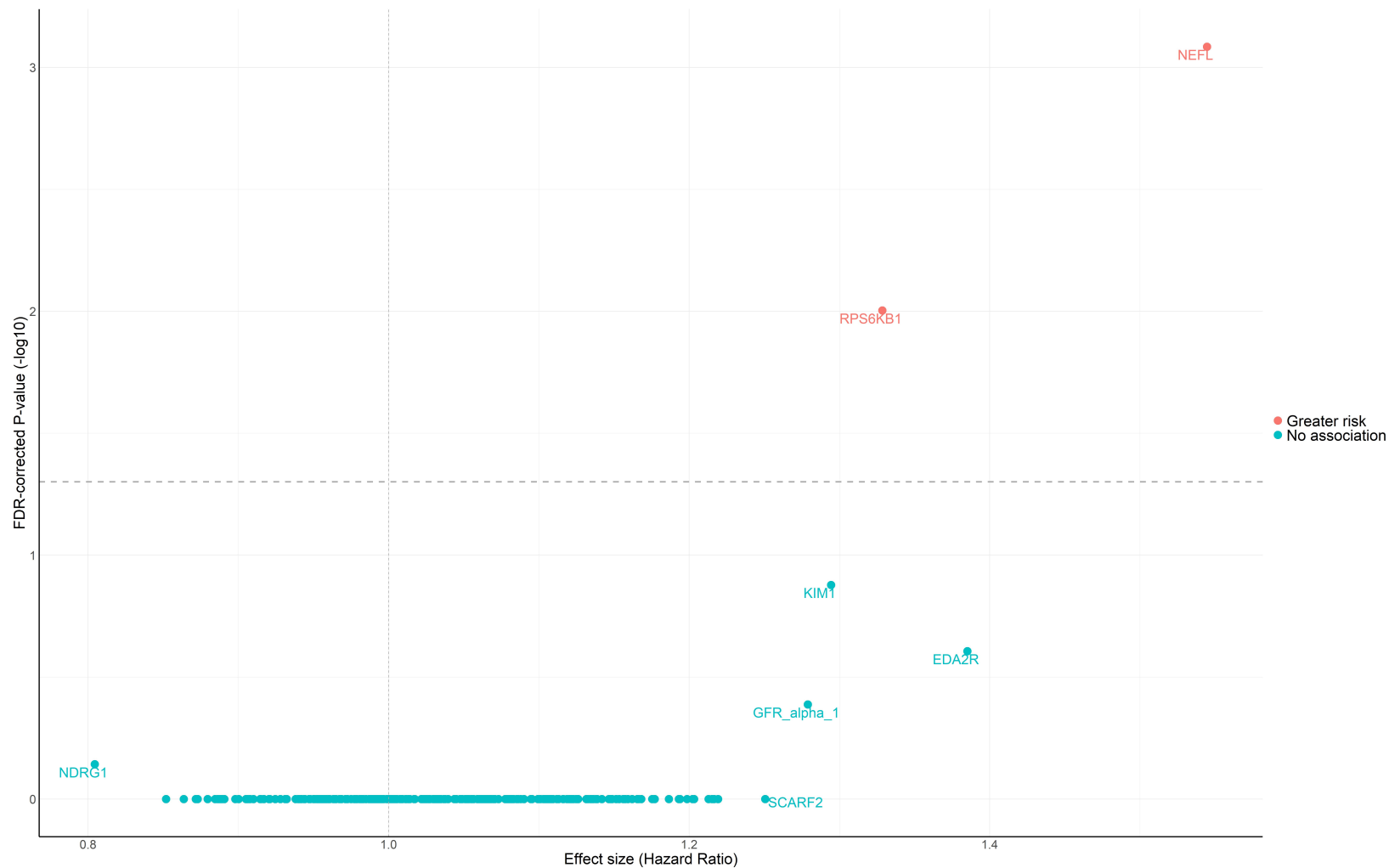
8 All GWAS summary statistics are available online: <https://doi.org/10.7303/syn51364943>;
9 https://www.finngen.fi/en/access_results; <https://gwas.mrcieu.ac.uk/>.

10

11 **Code availability**

12 The codes used for all analyses are available on GitHub repository:
13 https://github.com/jgong94/ELSA_proteomics_dementia.

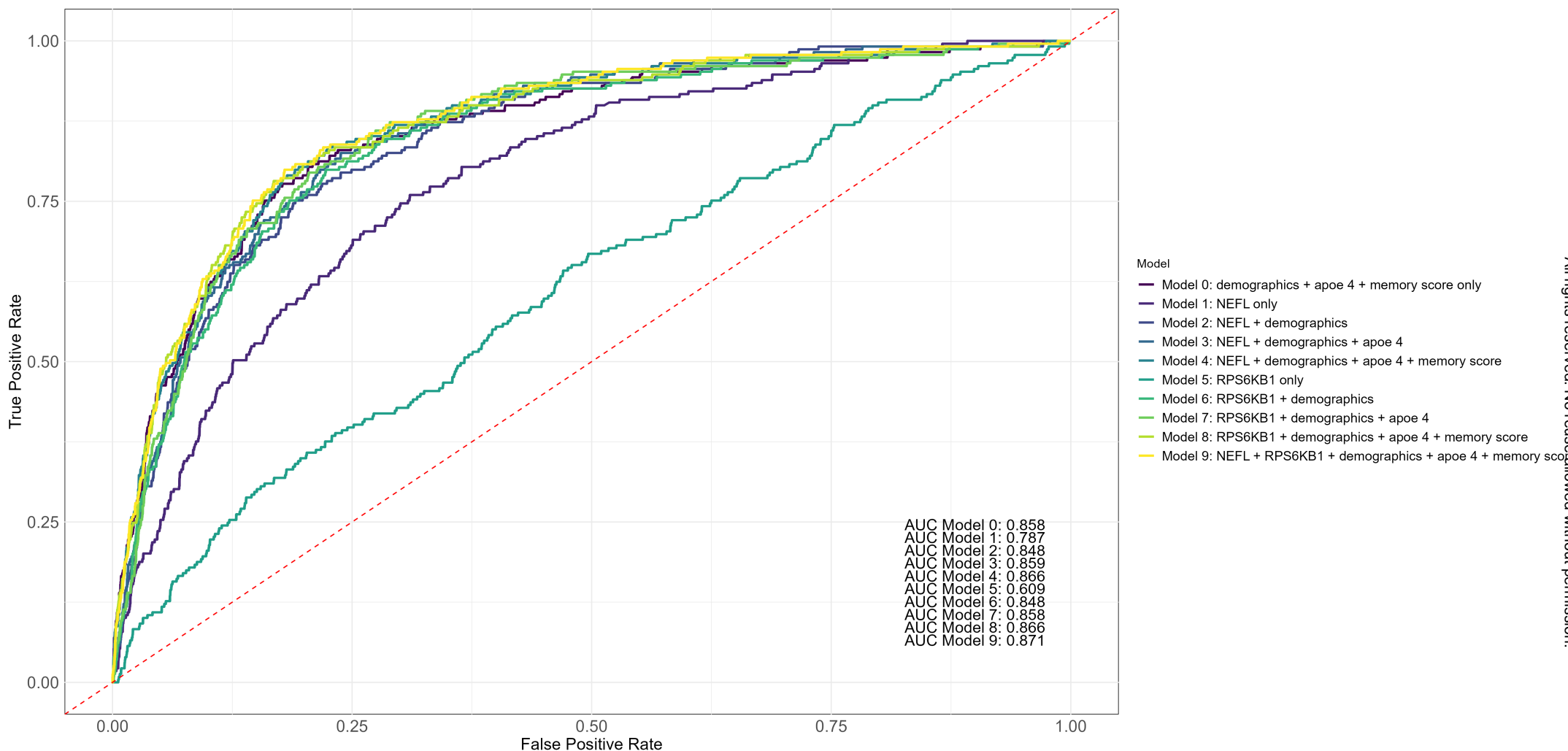
Figure 1. Volcano plot showing the Hazard ratio (x-axis) and two-sided false discovery rate (FDR)-corrected P values (y-axis) for the association of protein concentration with incident all-cause dementia using imputed data.



Hazard ratios from Cox Proportional Hazard Regression models adjusted for age, sex, education, race, smoking status, depression, cardiovascular disease, body mass index, systolic blood pressure, low-density lipoprotein (LDL) cholesterol.

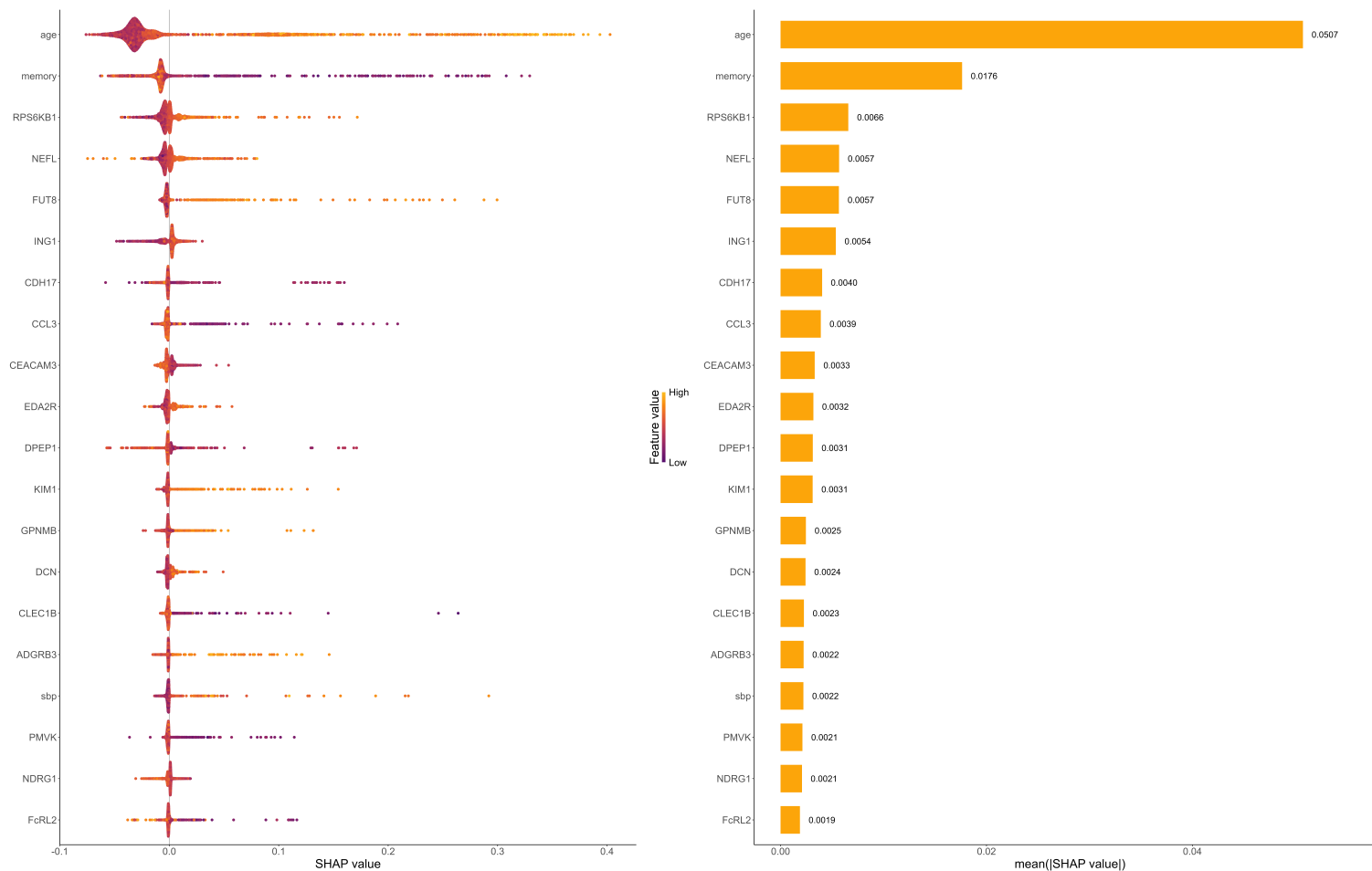
Proteins above the horizontal dotted grey line were significantly associated with incident dementia false discovery rate (FDR)-corrected p-value < 0.05.

Figure 2. Predictive accuracy of NEFL and RPS6KB1, alone or in combination with demographic variables, APOE 4 status, and memory score for all-cause dementia.



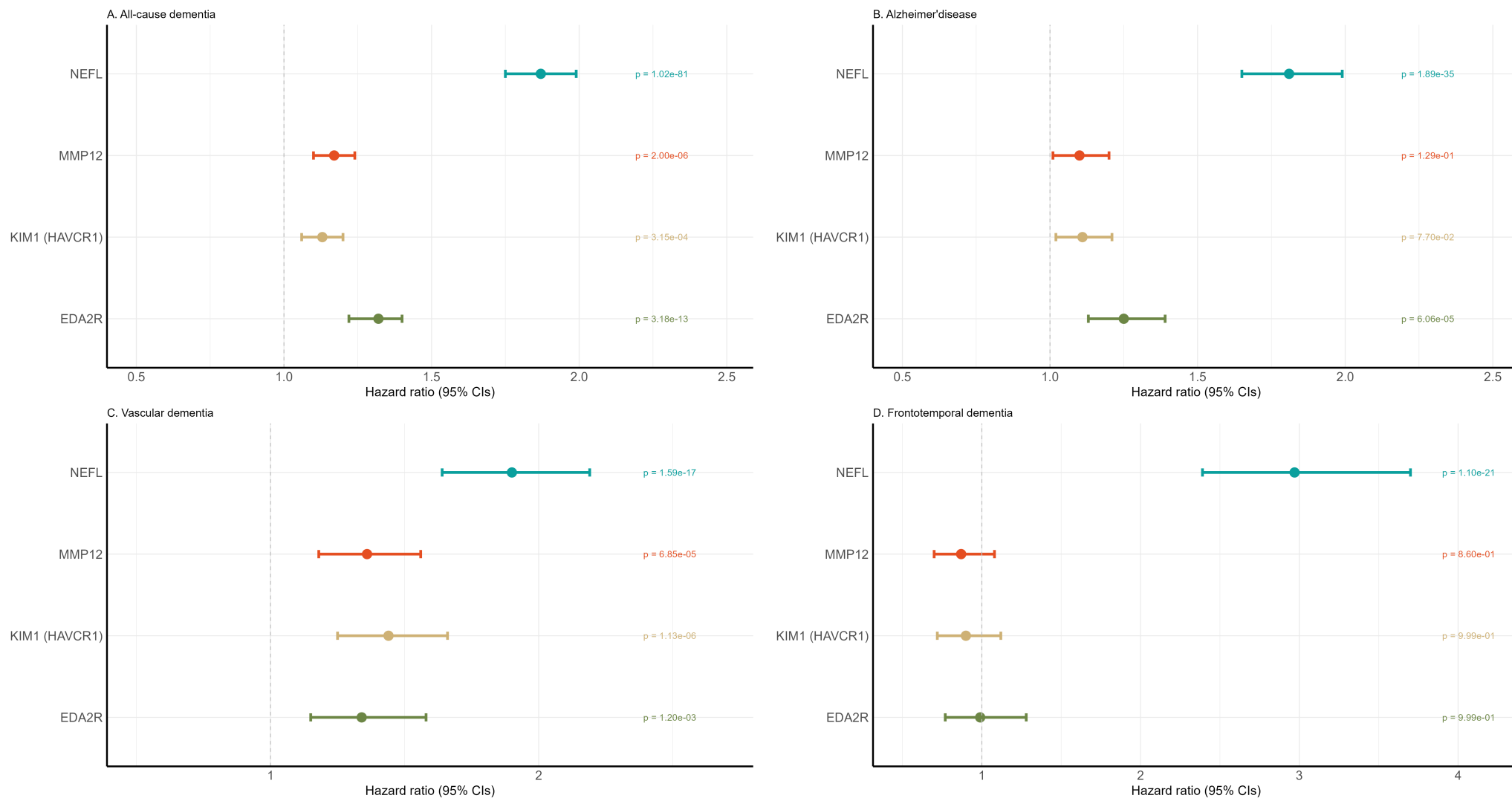
Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curves illustrate the performance of various variable models in predicting the incidence of all-cause dementia. Demographics variables included sex, age, education, and race. Memory score included a combined test score of immediate recall and delayed recall.

Figure 3. Protein importance ranking using XGBoost decision tree-based machine learning algorithm and SHAP visualization for selected features on all-cause dementia.



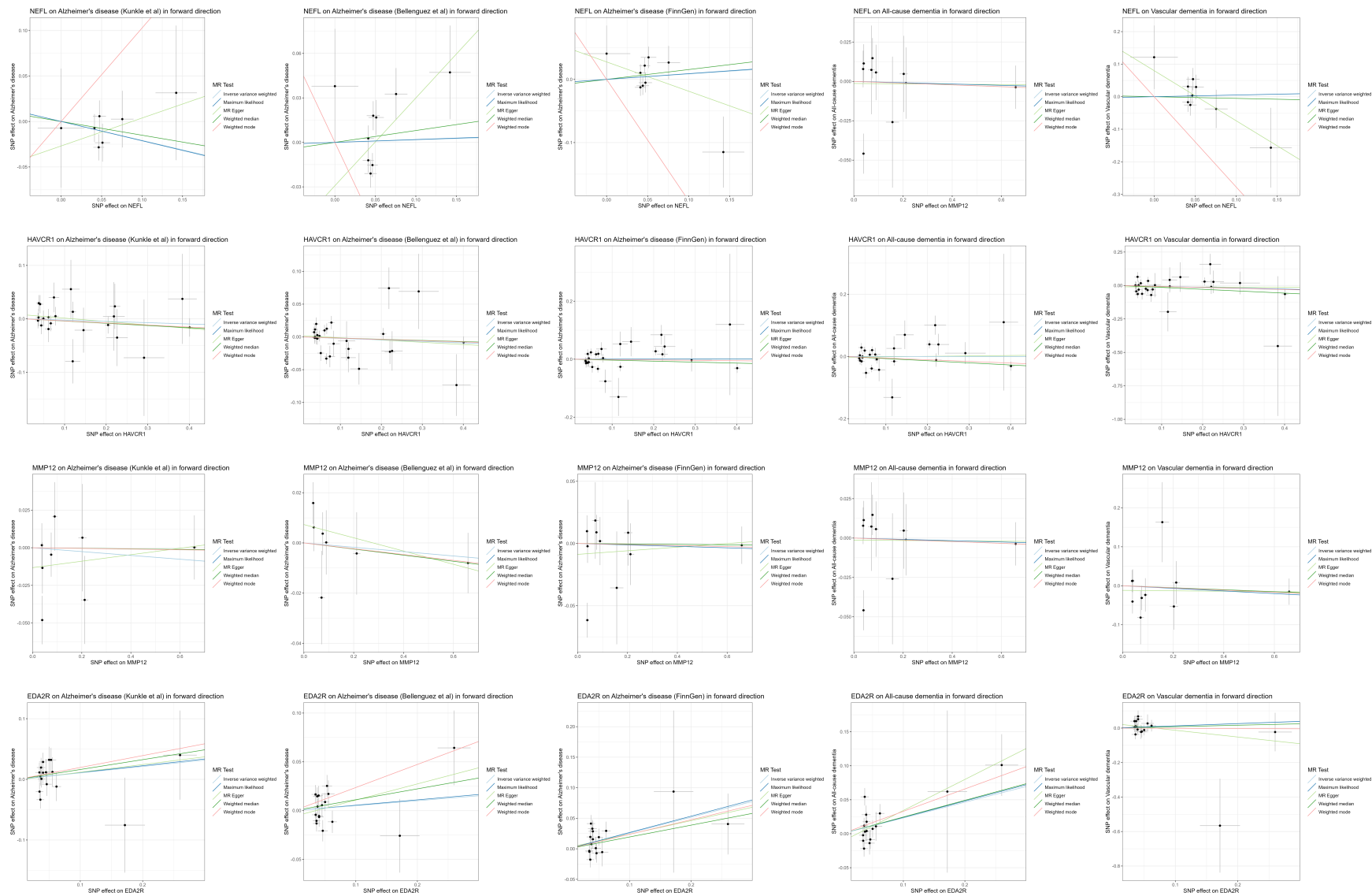
A) SHapley Additive exPlanations (SHAP) values from eXtreme Gradient Boosting (XGBoost) model displaying the selected features. The y-axis indicates the feature names in order of importance ranked from top to bottom. The x-axis represents the SHAP value, which indicates the degree of change in log odds. The width of the range of the horizontal bars showed the extent of the contribution to the prediction of all-cause dementia. The color of each point on the graph represents the value of the corresponding feature, with orange indicating high values and purple indicating low values. The direction on the x-axis indicates the likelihood of developing all-cause dementia towards the right, and likelihood of free from dementia towards the left. B) Mean absolute SHAP values for each feature derived from XGBoost model.

Figure 4. Forest Plots for the associations between identified proteins from ELSA and dementia and dementia subtypes validated in the UK Biobank.

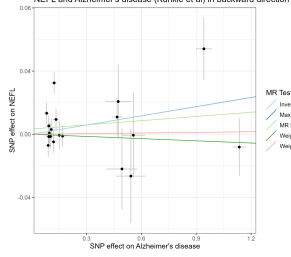


Multiple adjusted hazard ratios and 95% confidence intervals from Cox Proportional Hazard Regression models for NEFL, KIM1 (HAVCR1), MMP12, EDA2R and the associations with A) all-cause dementia; B) Alzheimer's disease; C) vascular dementia; D) frontotemporal dementia. All models adjusted for age, sex, education, ethnicity, smoking status, depression, cardiovascular disease, body mass index, systolic blood pressure, low-density lipoprotein (LDL) cholesterol. P values were false discovery rate (FDR) corrected.

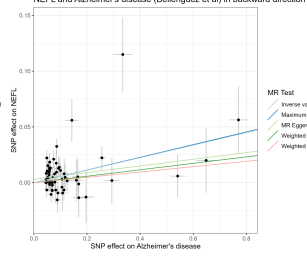
Figure 5. Two-sample bidirectional Mendelian randomisation and drug target Mendelian randomisation scatter plots for four proteins (NEFL, KIM1 (HAVCR1), MMP12, EDA2R) in five GWAS for Alzheimer's disease, all-cause dementia, and vascular dementia.



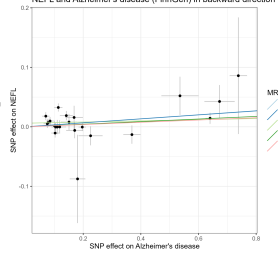
NEFL and Alzheimer's disease (Kunkle et al) in backward direction



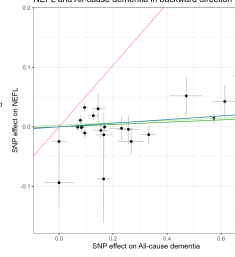
NEFL and Alzheimer's disease (Bellenguez et al) in backward direction



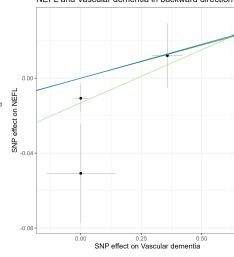
NEFL and Alzheimer's disease (FinnGen) in backward direction



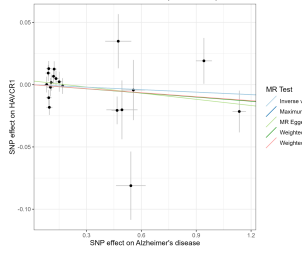
NEFL and All-cause dementia in backward direction



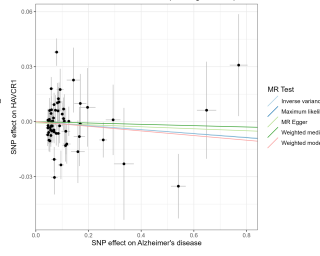
NEFL and Vascular dementia in backward direction



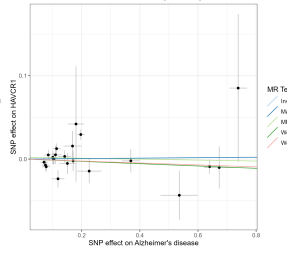
HAVCR1 and Alzheimer's disease (Kunkle et al) in backward direction



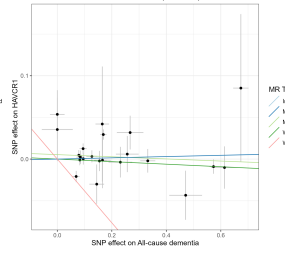
HAVCR1 and Alzheimer's disease (Bellenguez et al) in backward direction



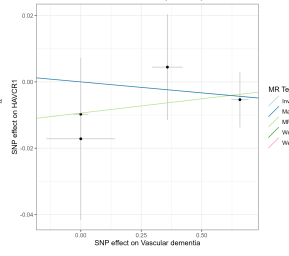
HAVCR1 and Alzheimer's disease (FinnGen) in backward direction



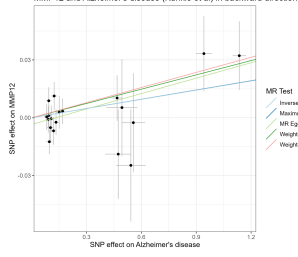
HAVCR1 and All-cause dementia (FinnGen) in backward direction



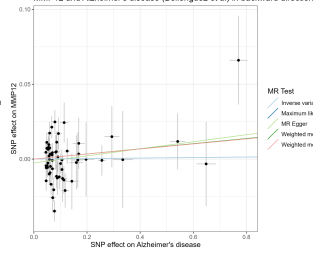
HAVCR1 and Vascular dementia (FinnGen) in backward direction



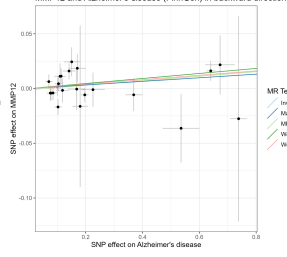
MMP12 and Alzheimer's disease (Kunkle et al) in backward direction



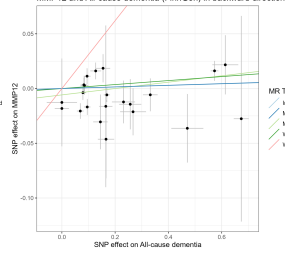
MMP12 and Alzheimer's disease (Bellenguez et al) in backward direction



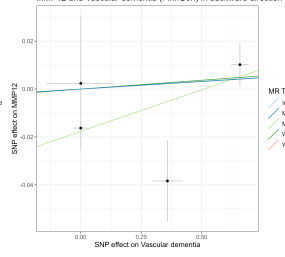
MMP12 and Alzheimer's disease (FinnGen) in backward direction



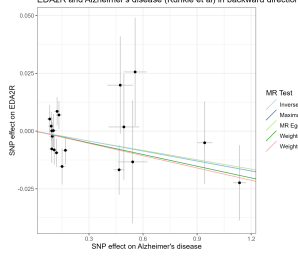
MMP12 and All-cause dementia (FinnGen) in backward direction



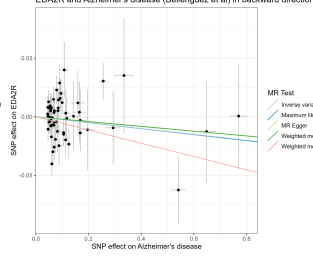
MMP12 and Vascular dementia (FinnGen) in backward direction



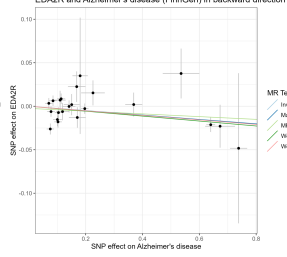
EDA2R and Alzheimer's disease (Kunkle et al) in backward direction



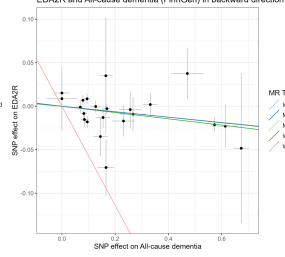
EDA2R and Alzheimer's disease (Bellenguez et al) in backward direction



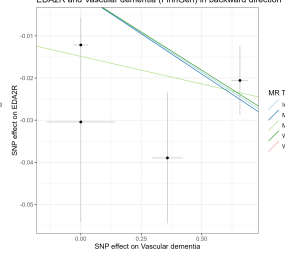
EDA2R and Alzheimer's disease (FinnGen) in backward direction

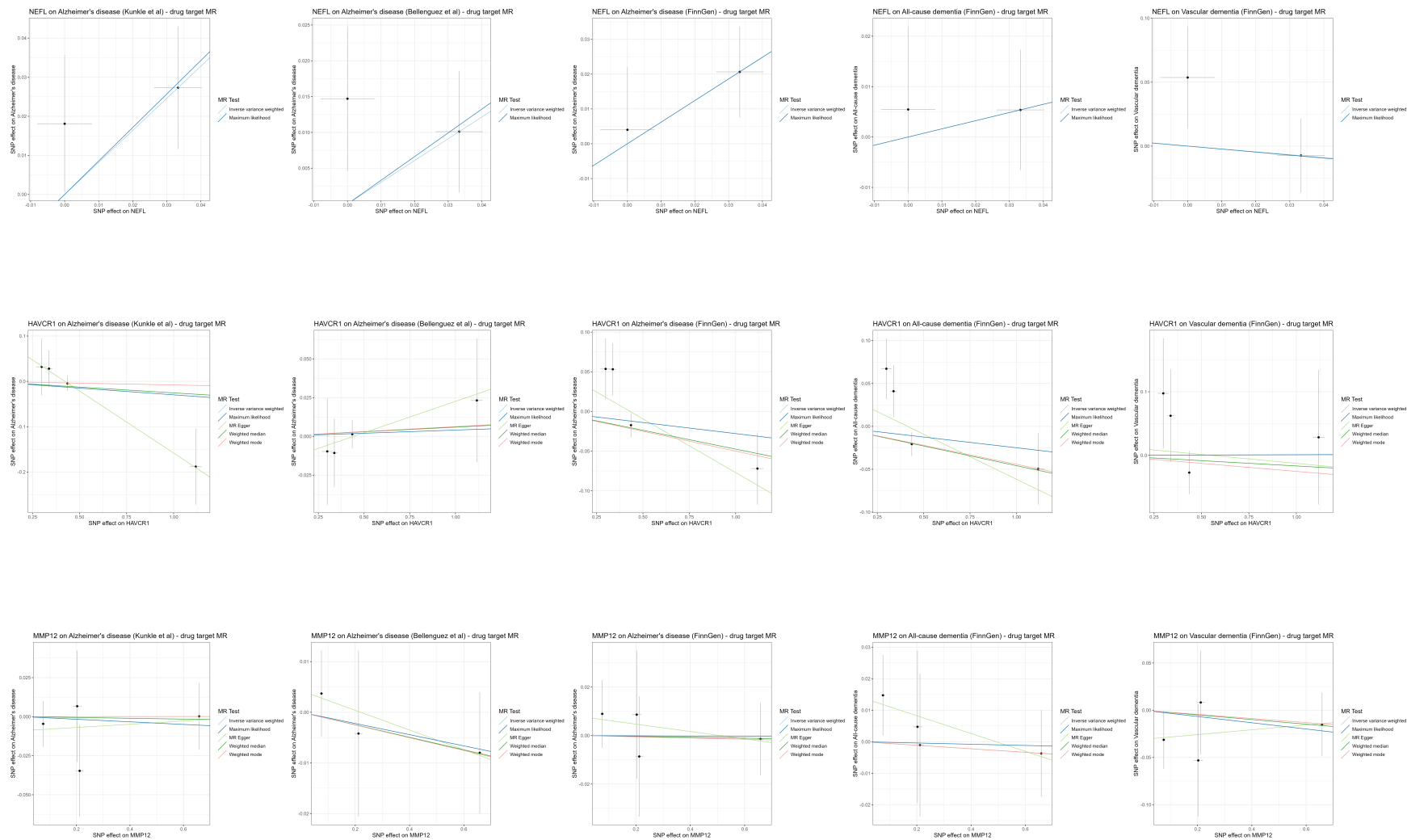


EDA2R and All-cause dementia (FinnGen) in backward direction



EDA2R and Vascular dementia (FinnGen) in backward direction





Scatter plots for Mendelian randomization (MR) analyses of the causal effect of proteins and dementia in forward and backward directions and drug target cis-MR. Analyses were conducted using the inverse variant weighted, maximum likelihood, MR-Egger, Weighted median, Weighted mode methods. The slope of each line corresponding to the estimated MR effect per method.

