

Proteomic Analysis Identifies Circulating Proteins Associated With Plasma Amyloid β and Incident Dementia

Supplement 1

Contents:

Supplemental Figures, Methods, and Results

List of supplemental figures

Supplemental Figure 1. Flowchart of participant inclusion at the midlife baseline (1993-95, visit 3)

Supplemental Figure 2. Flowchart of participant inclusion at late life (visit 5, 2011-13)

Supplemental Figure 3. Volcano plots of protein associated with plasma amyloid beta peptides at midlife: A β 42 (A), A β 40 (B), A β 42/A β 42 (C)

Supplemental Figure 4. Volcano plots of protein associated with plasma amyloid beta peptides at late life: A β 42 (A), A β 40 (B), A β 42/A β 42 (C)

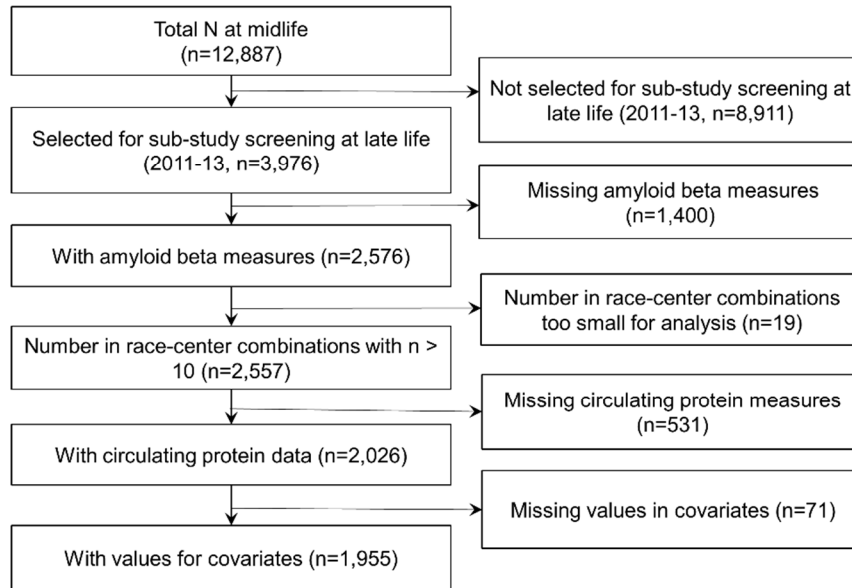
Supplemental Figure 5. Unadjusted Spearman correlations of the circulating proteins with significant associations with amyloid beta measures at midlife or late-life. The dash line indicates p-value=0.05

Supplemental Figure 6. Association of circulating proteins with Ab42, Ab40, and Ab42/Ab40 at midlife with and without controlling for cystatin C

Supplemental Figure 7. Association of circulating proteins with Ab42, Ab40, and Ab42/Ab40 at midlife with and without controlling for cystatin C and beta-2 microglobulin

Supplemental Figure 8. GO terms significantly enriched for A β 42-associated protein signals at late life based on gene-set enrichment analysis (GSEA, FDR < 0.05). Abbreviation. CC, cellular component; BP, biological process; MF, molecular function; FDR, false discovery rate

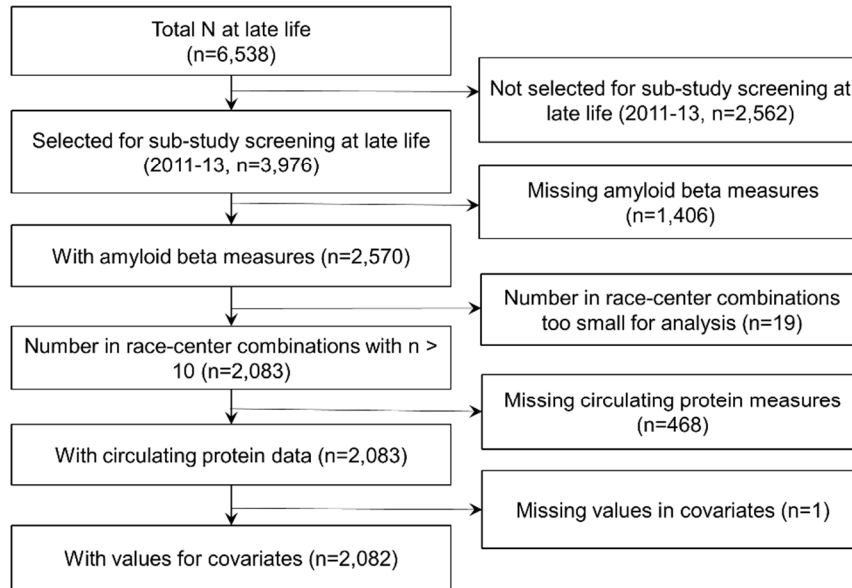
Supplemental Figure 9. Relevancy of the PEER factors. Stegle et al. Nat Protocol 2012 recommended using the number of PEER factors up to the “elbow” (n=4 in these plots) as covariates in the analysis to control for potential systemic confounding in high throughput expression data.



1

Supplemental Figure 1. Flowchart of participant inclusion at midlife (1993-95, visit 3)

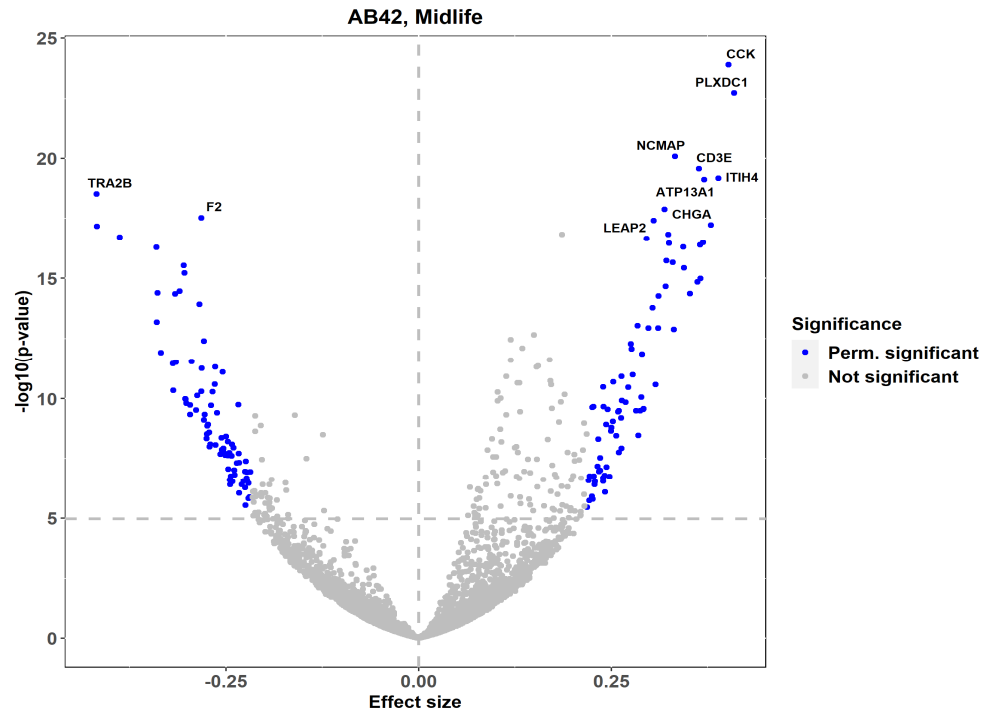
The 19 participants excluded due to race-center combination too small were: 4 Asian ancestry and 12 Black participants from Minnesota and Washington County, and 3 American Indian or Alaska Native from Forsyth and Washington County. Between midlife and late-life, 1,834 participants were in both analyzed samples.



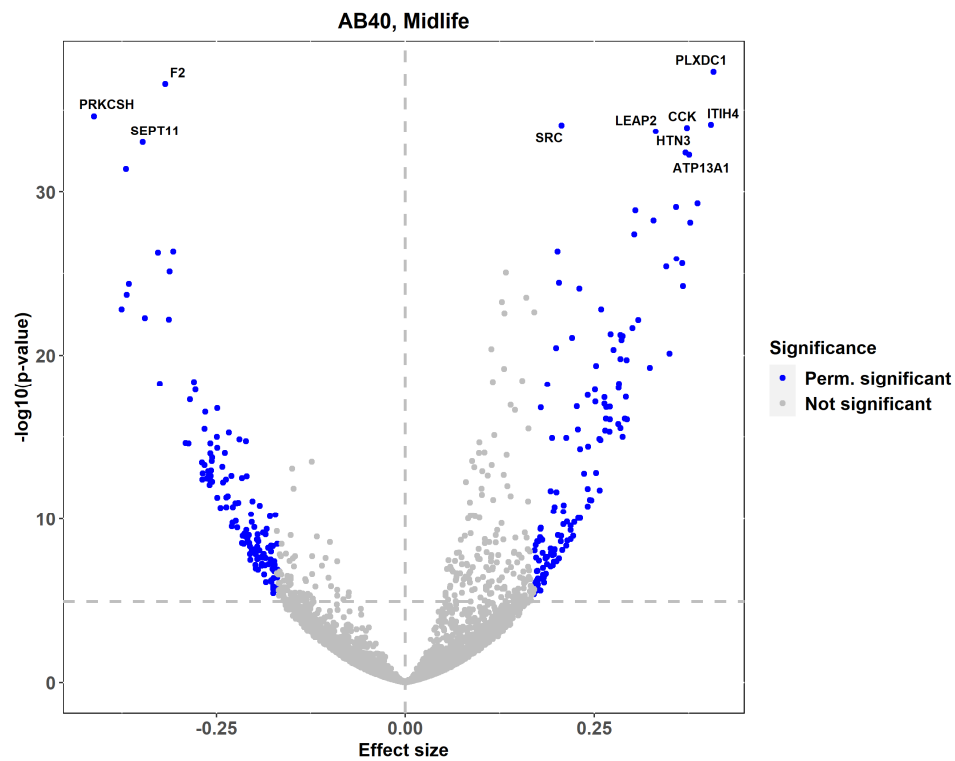
2

Supplemental Figure 2. Flowchart of participant inclusion at late life (visit 5, 2011-13).

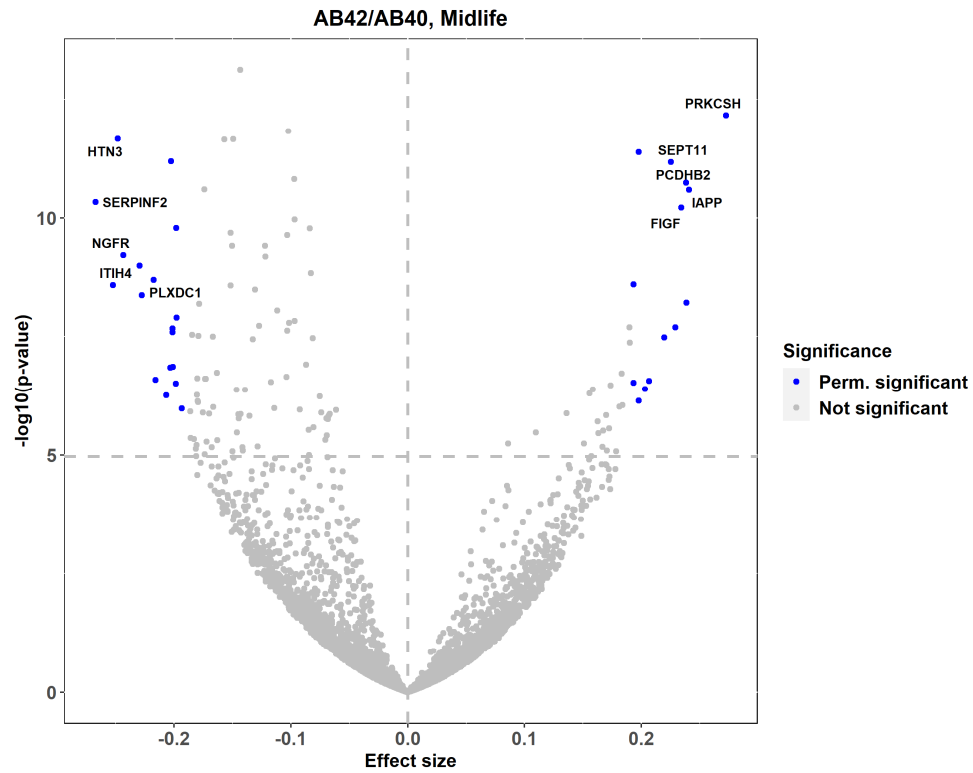
The 19 participants excluded due to race-center combination too small were: 4 Asian ancestry and 12 Black participants from Minnesota and Washington County, and 3 American Indian or Alaska Native from Forsyth and Washington County. Between midlife and late-life, 1,834 participants were in both analyzed samples.



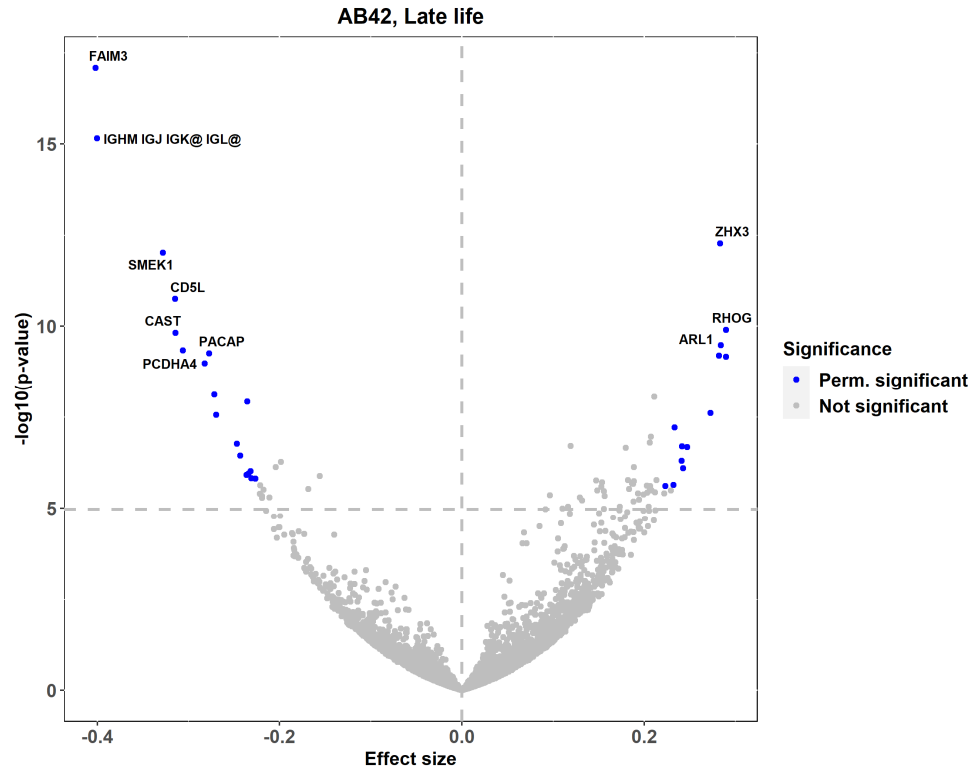
Supplemental Figure 3A. Effect size and regression p-value of the association of A β 42 and circulating proteins at midlife. Blue dots represent proteins that were significant based on permutation test (p -value $< 1.078E-5$). The 10 proteins that were permutation significant and had the lowest regression p-value were labeled. The dotted line represents regression p-value $< 1.078E-5$. Abbreviation. Perm, permutation.



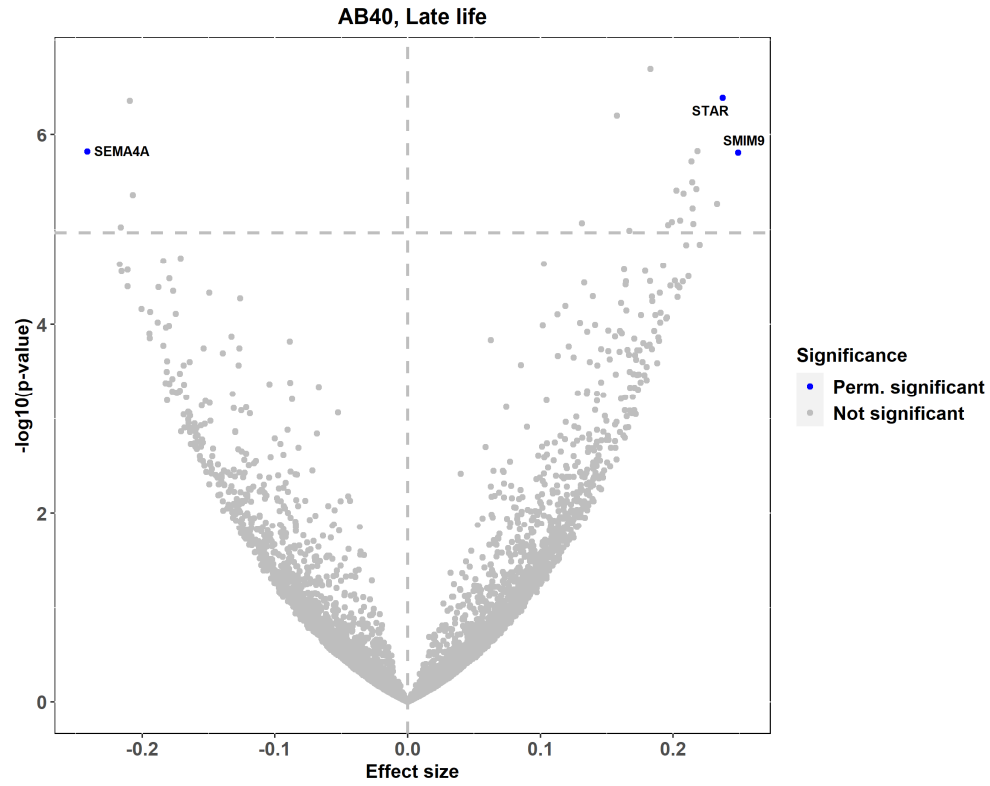
Supplemental Figure 3B. Effect size and regression p-value of the association of A β 40 and circulating proteins at midlife. Blue dots represent proteins that were significant based on permutation test (p-value < 1.078E-5). The 10 proteins that were permutation significant and had the lowest regression p-value were labeled. The dotted line represents regression p-value < 1.078E-5. Abbreviation. Perm, permutation.



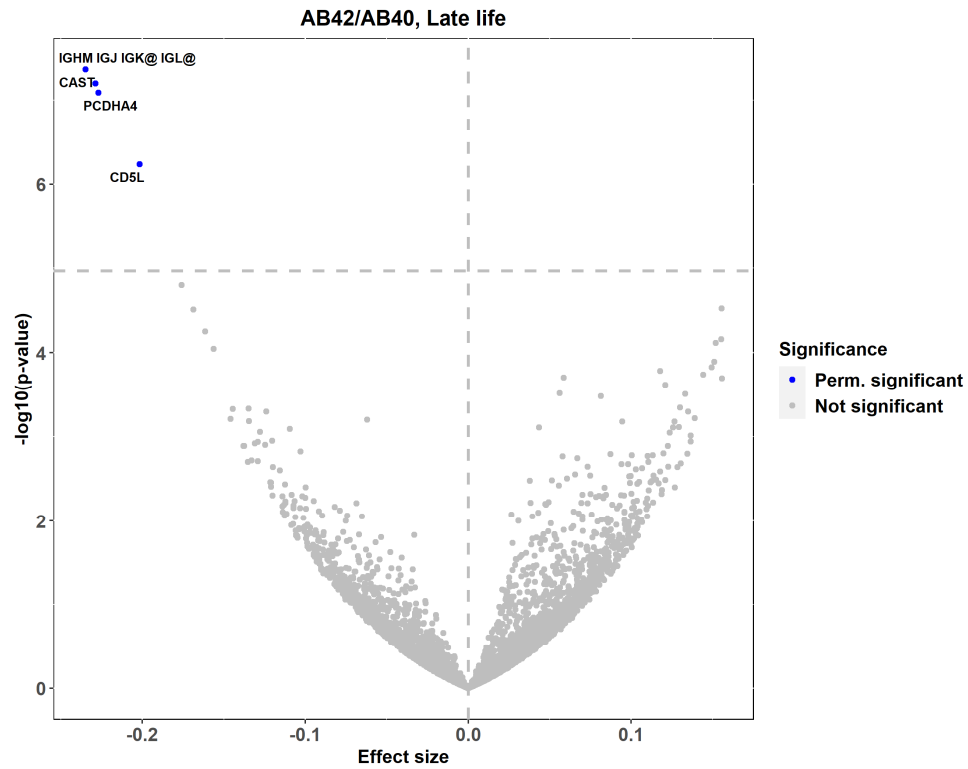
Supplemental Figure 3C. Effect size and regression p-value of the association of A β 42/A β 40 ratio and circulating proteins at midlife. Blue dots represent proteins that were significant based on permutation test (p-value < 1.078E-5). The 10 proteins that were permutation significant and had the lowest regression p-value were labeled. The dotted line represents regression p-value < 1.078E-5. Abbreviation. Perm, permutation.



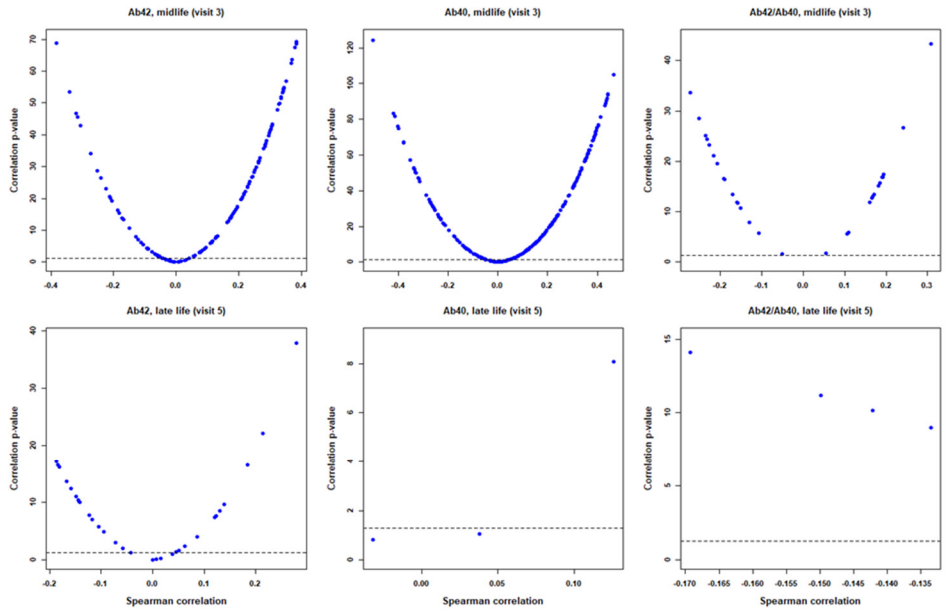
Supplemental Figure 4A. Effect size and regression p-value of the association of A β 42 and circulating proteins at late life. Blue dots represent proteins that were significant based on permutation test (p -value < 1.078E-5). The 10 proteins that were permutation significant and had the lowest regression p-value were labeled. The dotted line represents regression p-value < 1.078E-5. Abbreviation. Perm, permutation.



Supplemental Figure 4B. Effect size and regression p-value of the association of A β 40 and circulating proteins at late life. Blue dots represent proteins that were significant based on permutation test (p-value < 1.078E-5). The dotted line represents regression p-value < 1.078E-5. Abbreviation. Perm, permutation.



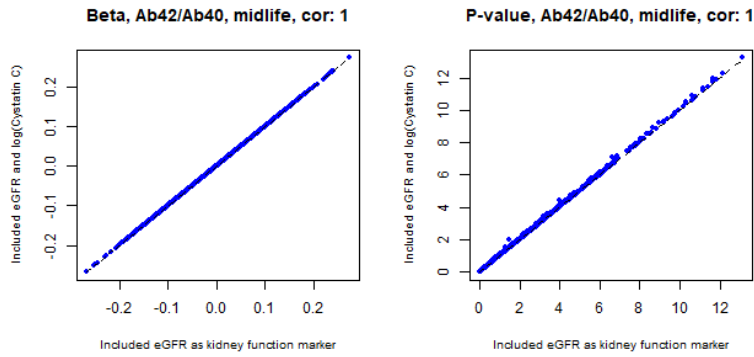
Supplemental Figure 4C. Effect size and regression p-value of the association of A β 42/A β 40 ratio and circulating proteins at late life. Blue dots represent proteins that were significant based on permutation test (p-value < 1.078E-5). The dotted line represents regression p-value < 1.078E-5. Abbreviation. Perm, permutation.



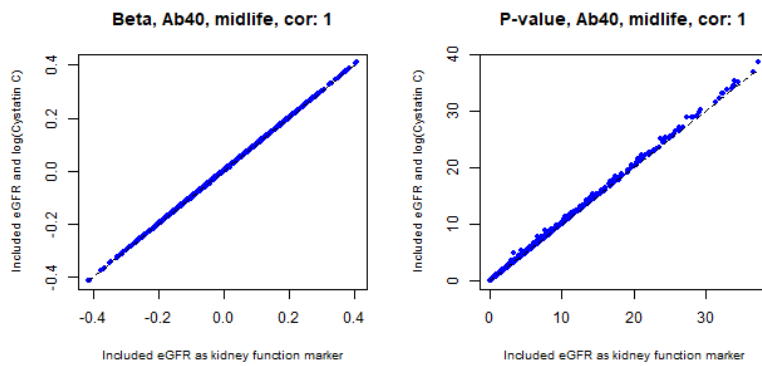
3

Supplemental Figure 5. Unadjusted Spearman correlations of the circulating proteins with significant associations with amyloid beta measures at midlife or late-life. The dash line indicates p-value=0.05

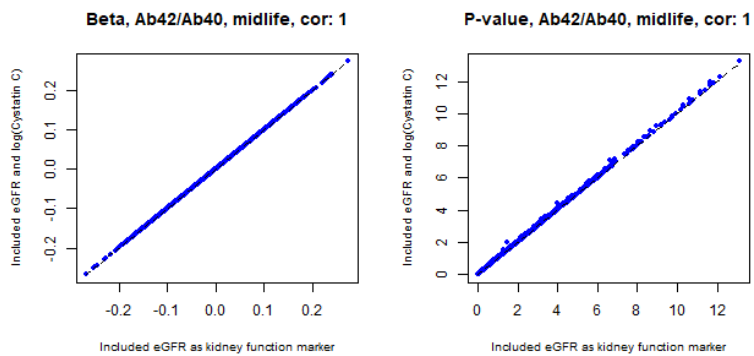
A



B

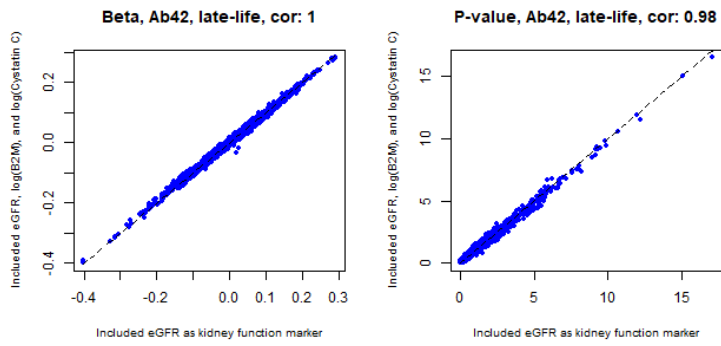


C

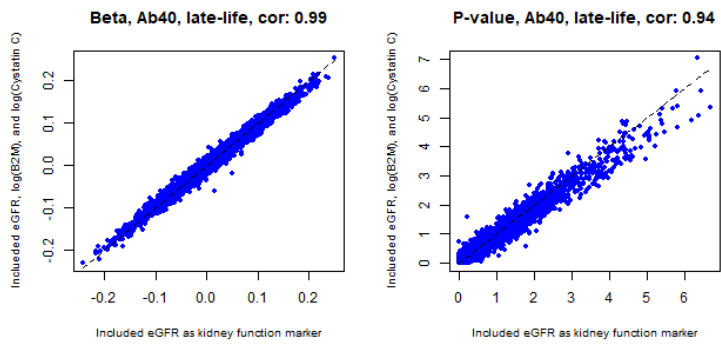


Supplemental Figure 6. Association of circulating proteins with Ab42, Ab40, and Ab42/Ab40 at midlife with and without controlling for cystatin C. Pearson correlations of the betas and p-values with and without controlling for cystatin C were 1.0 for all.

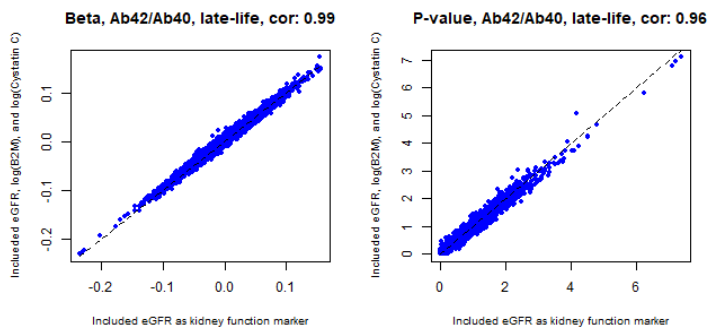
A



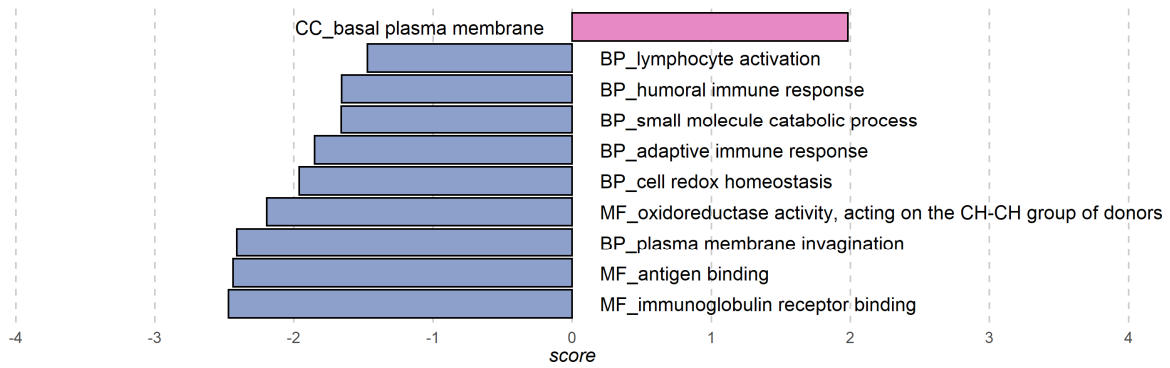
B



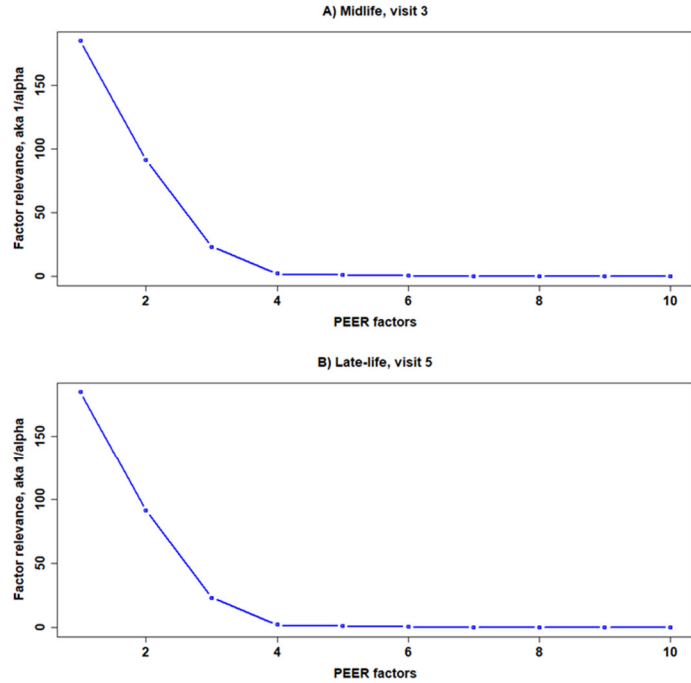
C



Supplemental Figure 7. Association of circulating proteins with Ab42, Ab40, and Ab42/Ab40 at midlife with and without controlling for cystatin C and beta-2 microglobulin. Pearson correlations of the betas and p-values with and without controlling for cystatin C ranged from 0.94 to 0.99.



Supplemental Figure 8. GO terms significantly enriched for A β 42-associated protein signals at late life based on gene-set enrichment analysis (GSEA, FDR < 0.05). Abbreviation. CC, cellular component; BP, biological process; MF, molecular function; FDR, false discovery rate.



5

Supplemental Figure 9. Relevancy of the PEER factors. Stegle et al. Nat Protocol 2012 recommended using the number of PEER factors up to the “elbow” (n=4 in these plots) as covariates in the analysis to control for potential systemic confounding in high throughput expression data.

Supplemental Methods

Measurement of variables used in the reporting of population characteristics, sensitivity analysis, and not in primary association analysis

The following variables were used in the reporting of population characteristics and sensitivity analysis. Education levels (< high school, high school graduate or vocational school, and at least some college, graduate or professional school), and current smoking status were self-reported. Body mass index (BMI) was calculated using height and weight measured at study visit. Systolic blood pressure was the average of the second and the third measures of three measures. Total plasma cholesterol, triglycerides, and glucose were measured on the Cobas autoanalyzer at midlife and the OLYMPUS analyzer at late life. C-reactive protein (CRP) levels were quantified using an immunoturbidimetric assay on a Beckman Coulter instrument. Prevalent diabetes mellitus was defined as having a fasting glucose level ≥ 126 mg/dl, non-fasting glucose level ≥ 200 mg/dl, self-reported diabetes medication use, or self-reported physician diagnosis of diabetes. Hypertension was defined as systolic blood pressure ≥ 140 mm Hg, diastolic blood pressure ≥ 90 mm Hg, or self-reported use of anti-hypertension medications. Prevalent cardiovascular disease was defined as a composite of prevalent coronary heart disease (CHD), stroke, or heart failure. Prevalent CHD at enrollment (visit 1) was determined based on echocardiogram at study visit or self-reported history of myocardial infarction or heart or arterial procedures. Prevalent stroke at enrollment was based on self-report. Prevalent heart failure at enrollment was determined based on the Gothenburg score and self-reported use of heart failure medications (1). Subsequent CHD, stroke, or heart failure events were ascertained by cohort surveillance adjudicated by expert committees (since visit 1 for CHD and stroke and since 2005 for heart failure) based on information from hospital records and annual telephone interview with ARIC participants (2-4). Genotyping of the two single nucleotide polymorphisms (rs429358, rs7412) that define the *APOE* $\epsilon 4$ genotypes was performed using the TaqMan assay (Thermo Fisher Scientific, Waltham MA).

Use of probabilistic estimation of expression residuals (PEER) factors as covariates for the association analysis between the A β peptides and circulating proteins

The PEER method uses a 3-component model to capture the systematic variability of a large number of continuous variables and has been applied in proteomic studies (5, 6). The three components correspond to variability specific to each protein, variability due to known correlates of the proteins, and hidden systematic factors, which can be used as covariates to control for confounding. The selection of four hidden factors was based on the relevancy criteria where the relevancy is the posterior variance of the hidden factor. Hidden factors with higher variance capture broader systematic variation. A recommended method for selecting the number of hidden factors as covariates is to include those up to the 'elbow' of the relevancy plot (**Figures S9A and B** for midlife and late life, respectively) (7).

Sensitivity analysis of the association between the A β peptides and circulating proteins

We conducted two sensitivity analyses. First, given that circulating proteins might be effectors, mediators, or biomarkers of clinical risk factors of dementia (5), we evaluated the significant A β associations by including additional covariates: *APOE* ϵ 4 carrier status, body mass index (BMI), current smoking status, prevalent diabetes, hypertension, and cardiovascular disease, blood glucose, systolic blood pressure, total cholesterol, and log₂ transformed triglycerides and C-reactive protein levels. Second, additional available biomarkers of kidney function were added as covariates (cystatin C at midlife and late-life and beta-2 microglobulin at late-life) after log transformation.

Association analysis of A β 42-associated proteins with scores of the Mini Mental State Examination (MMSE) at late-life

Of the 344 participants with prevalent dementia at late-life (visit 5), we excluded one participant without MMSE score, 106 without protein measures, and one without eGFR. The association analysis included 236 participants and was conducted using linear regression with MMSE as the outcome and inverse normal transformed protein levels as the independent variable controlling for age, sex, race-center, eGFR, and four PEER factors.

Incident dementia ascertainment in the ARIC study

Dementia status was determined by an expert committee, which included physicians and neuropsychologists, based on the criteria from the National Institute on Aging–Alzheimer’s Association (NIA-AA) workgroups and Diagnostic and Statistical Manual of Mental Disorders (DSM), 5th Edition, as described previously (8, 9). The data included detailed cognitive and functional assessments (**Table S27**) collected at ARIC visits (visit 5, 2011–2013; visit 6, 2016–18, and visit 7, 2018–19) and cognitive tests at visits 2 (1990–92) and 4 (1996–98). The Clinical Dementia Rating (CDR) interview and the Functional Assessment Questionnaire (FAQ) were used in in-person and telephone interviews of participants and informants who could not attend the clinical visit. Participants who did not attend visit 5 were administered the modified Telephone Interview for Cognitive Status (TICS)(10, 11). TICS scores were education-adjusted (8). After visit 5, the Six Item Screener (SIS) was offered annually to all participants, and the Alzheimer’s Dementia 8-Item (AD8) Informant Questionnaire was administered to informants by phone (12). Dementia status was also ascertained using ICD-9 dementia codes at hospitalization discharge and on death certificates obtained by cohort surveillance.

Dementia date was first set as the earliest of either the hospitalization date with an ICD-9 code for dementia, death date if a dementia code was listed on the death certificate, date of telephone communication with the participant or proxy with indication of dementia, or date of the first visit when dementia was indicated. Dementia onset date ascertained from informant interviews, hospitalization and death certificate was subtracted by six months to account for the expected lag in the reporting of the event. Participants who were classified as not having dementia were censored at the last study contact date when there was no indication of dementia or the date of death obtained by cohort surveillance.

Study population for the association analysis of plasma proteins and incident dementia in ARIC

The participant inclusion criteria for these two analyses have been reported previously (5). For the analysis of incident dementia with baseline at midlife, the baseline visit (visit 3, 1993-95) had 12,887 participants. After excluding participants who were not self-reported Black or White (n=38) or self-reported Black in Minnesota or Washington County (n=42), had prevalent dementia or were censored before baseline (n=8), missing covariates (n=422), or missing measures of circulating proteins (n=1,308), 11,069 participants were included. Participants were followed up to 2011 (visit 5, 2011-13). The median follow-up time was 17.2 years. We observed 1,131 incident dementia events.

For the analysis of incident dementia with baseline at late-life, the baseline visit (visit 5, 2011-13) had 6,538 participants. After excluding participants who were not self-reported Black or White (n=18) or self-reported Black in Minnesota or Washington County (n=24), had prevalent dementia (n=341), missing cognitive status (n=64), missing covariates (n=636), measures of circulating proteins (n=974), or incident dementia status not ascertained (n=371), 4,110 participants were included. Participants were followed up to 2017 (visit 6, 2016-17). The median follow-up time was 4.9 years. We observed 428 incident dementia events.

Incident dementia ascertainment in the Age, Gene/Environment Susceptibility (AGES) - Reykjavik Study

Dementia classification was in accordance with the Diagnostic and Statistical Manual of Mental Disorders-IV criteria (13) and conducted using a three-step procedure (14). All participants were administered the MMSE and the Digit Symbol Substitution Test. Participants who received a low score on either measure were administered a more comprehensive battery of cognitive measures. Participants who received a low score on the Trails A and B measures or the Rey Auditory Verbal Learning Test received an additional assessment, which included a neurologic examination and a proxy interview. Dementia diagnoses were adjudicated based on consensus during conferences that included a neurologist, geriatrician, neuropsychologist and a neuroradiologist who provided a clinical reading of available MRIs.

Supplemental References

1. Eriksson H, Caidahl K, Larsson B, Ohlson LO, Welin L, Wilhelmsen L, et al. (1987): Cardiac and pulmonary causes of dyspnoea--validation of a scoring test for clinical-epidemiological use: the Study of Men Born in 1913. *Eur Heart J.* 8:1007-1014.
2. Rosamond WD, Folsom AR, Chambless LE, Wang CH, McGovern PG, Howard G, et al. (1999): Stroke incidence and survival among middle-aged adults: 9-year follow-up of the Atherosclerosis Risk in Communities (ARIC) cohort. *Stroke.* 30:736-743.
3. Rosamond WD, Folsom AR, Chambless LE, Wang CH (2001): Coronary heart disease trends in four United States communities. The Atherosclerosis Risk in Communities (ARIC) study 1987-1996. *Int J Epidemiol.* 30 Suppl 1:S17-22.
4. Rosamond Wayne D, Chang Patricia P, Baggett C, Johnson A, Bertoni Alain G, Shahar E, et al. (2012): Classification of Heart Failure in the Atherosclerosis Risk in Communities (ARIC) Study. *Circulation: Heart Failure.* 5:152-159.
5. Walker KA, Chen J, Zhang J, Fornage M, Yang Y, Zhou L, et al. (2021): Large-scale plasma proteomic analysis identifies proteins and pathways associated with dementia risk. *Nature Aging.* 1:473-489.
6. Zhang J, Dutta D, Kottgen A, Tin A, Schlosser P, Grams M, et al. (2022): Large Bi-Ethnic Study of Plasma Proteome Leads to Comprehensive Mapping of cis-pQTL and Models for Proteome-wide Association Studies *Nat genet (in press)*.
7. Stegle O, Parts L, Piipari M, Winn J, Durbin R (2012): Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc.* 7:500-507.
8. Knopman DS, Gottesman RF, Sharrett AR, Wruck LM, Windham BG, Coker L, et al. (2016): Mild Cognitive Impairment and Dementia Prevalence: The Atherosclerosis Risk in Communities Neurocognitive Study (ARIC-NCS). *Alzheimers Dement (Amst).* 2:1-11.

9. Gottesman RF, Albert MS, Alonso A, Coker LH, Coresh J, Davis SM, et al. (2017): Associations Between Midlife Vascular Risk Factors and 25-Year Incident Dementia in the Atherosclerosis Risk in Communities (ARIC) Cohort. *JAMA Neurol.* 74:1246-1254.
10. Welsh KA, Breitner JCS, Magruder-Habib KM (1993): Detection of Dementia in the Elderly Using Telephone Screening of Cognitive Status. *Cognitive and Behavioral Neurology.* 6.
11. Knopman DS, Roberts RO, Geda YE, Pankratz VS, Christianson TJ, Petersen RC, et al. (2010): Validation of the telephone interview for cognitive status-modified in subjects with normal cognition, mild cognitive impairment, or dementia. *Neuroepidemiology.* 34:34-42.
12. Carpenter CR, DesPain B, Keeling TN, Shah M, Rothenberger M (2011): The Six-Item Screener and AD8 for the detection of cognitive impairment in geriatric emergency department patients. *Annals of emergency medicine.* 57:653-661.
13. American Psychiatric Association (2000): *Diagnostic and statistical manual of mental disorders (4th ed., Text Revision)*. Washington, DC.
14. Harris TB, Launer LJ, Eiriksdottir G, Kjartansson O, Jonsson PV, Sigurdsson G, et al. (2007): Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol.* 165:1076-1087.

The supplemental tables are in a separate Excel file.

List of Supplemental Tables

Supplemental Table 1. Characteristics of participants included and excluded from analysis at midlife and late life

Supplemental Table 2. The number in each category of the proteins excluded from analysis due to unspecific binding

Supplemental Table 3. Significance threshold for the interrogation of incident dementia association for the A β -associated proteins

Supplemental Table 4. Encoding genes of the A β -associated proteins that were included in the generation of PC to test for association with incident dementia

Supplemental Table 5. Characteristics of participants by levels of the A β measures at midlife

Supplemental Table 6. Characteristics of participants by levels of the A β measures at late life

Supplemental Table 7. Spearman correlations between the plasma amyloid beta measures (A β 42, A β 40, A β 42/A β 40) at the same time point and between time points

Supplemental Table 8. Circulating proteins having significant association with one or more plasma amyloid beta measures (A β 42, A β 40, A β 42/A β 40) at midlife (visit 3, n=1,955)

Supplemental Table 9. Circulating proteins having significant association with one or more plasma amyloid beta measures (A β 42, A β 40, A β 42/A β 40) at late life (visit 5, n=2,082)

Supplemental Table 10. A β -associated proteins that overlap between midlife and late-life

Supplemental Table 11. Quality control information of proteins that were associated with one or more A β measures at midlife or late life based on permutation test

Supplemental Table 12. Results of sensitivity analysis of the A β -associated proteins at midlife (visit 3, n=1,834)

Supplemental Table 13. Results of sensitivity analysis of the A β -associated proteins at late life (visit 5, n=1,802)

Supplemental Table 14. Association between A β -associated proteins with scores of Mini Mental State Examination (MMSE) among participants with dementia at late-life (n=236)

Supplemental Table 15. Association between A β -associated proteins with incident dementia in ARIC

Supplemental Table 16. Population characteristics of the AGES study

Supplemental Table 17. Results of the replication of the A β -associated proteins for dementia risk in AGES

Supplemental Table 18. Variance of A β measures explained by age, sex, race, eGFR and associated proteins

Supplemental Table 19. GO terms enriched for A β 42-associated proteins at midlife (visit 3) based on gene set enrichment analysis (GSEA)

Supplemental Table 20. GO terms enriched for A β 40-associated proteins at midlife (visit 3) based on gene set enrichment analysis (GSEA)

Supplemental Table 21. GO terms enriched for A β 42/A β 40-associated proteins at midlife (visit 3) based on gene set enrichment analysis (GSEA)

Supplemental Table 22. GO terms that were overlapped among the significant GO terms from the GSEA analysis of protein association signals of Ab42, Ab40, and Ab42/Ab40 at midlife

Supplemental Table 23. GO terms significantly enriched for A β 42-associated proteins at late life (visit 5) based on gene-set enrichment analysis (GSEA)

Supplemental Table 24. Association between PC1 of the A β -associated proteins in enriched GO terms with incident dementia.

Supplemental Table 25. Genes that are in significant GO terms related to the nervous system and have been linked to amyloid pathology based on literature search.

Supplemental Table 26. Encoding genes of A β 42-associated proteins with significant tissue enrichment in the liver at midlife and in the spleen at late-life

Supplemental Table 27. Ten cognitive and functional assessment instruments administered at ARIC visits 5, 6, and 7