

A cerebrospinal fluid synaptic protein biomarker for prediction of cognitive resilience versus decline in Alzheimer's disease

In the format provided by the
authors and unedited

SUPPLEMENTARY METHODS

A+T₁+ VERSUS A-T₁- CLASSIFICATION

Typically, “A” positivity is defined by levels of A β 42 and “T₁” positivity by PTau181, using a separate Gaussian Mixture model for each biomarker to derive value cutoffs¹ (**Supplementary Fig. 1a**). This leads to four possible groups A-T₁-, A+T₁-, A-T₁+, and A+T₁+. However, this classification system does not fit the “shape” of the data and artificially increases the number of A-T₁+ individuals² (**Supplementary Fig. 1a**), as the frequency of A-T+ individuals based on PET imaging biomarkers (the gold standards) are extremely rare². To overcome this limitation, we use the CSF PTau181:A β 42 ratio, which better fits the shape of the data (**Supplementary Fig. 1b**), to define A-T₁- versus A+T₁+ status ($\log_{10}$ PTau181:A β 42 cutoff = -1; **Supplementary Fig. 1c**). Previous studies have also shown that PTau181:A β 42 appropriately captures A-T₁- versus A+T₁+ status^{3,4}. A β positivity, regardless of T status, was determined using gold-standards CSF A β 42:A β 40 ratio or A β PET.

SOMASCAN VS IMMUNOASSAYS FOR “N” BIOMARKERS

We sought to compare CSF YWHAG:NPTX2 with CSF NfL, GAP43, and Neurogranin (Ng), which were also measured on the SomaScan assay. We confirmed that the levels of these biomarkers based on SomaScan and established immunoassays used for AD research were highly correlated⁵ (**Supplementary Fig. 2a-b**). For GAP43, SomaScan levels were compared with the immunoassay from the Blennow lab⁶.

Since CSF GAP43 and Ng were highly correlated ($r=0.98$, **Supplementary Fig. 2c**) we removed GAP43 from multivariate Cox models in **Figures 4 and 5** to avoid multicollinearity.

DERIVATION OF PLASMA SIGNATURE OF COGNITIVE IMPAIRMENT (CI).

We first sought to determine whether plasma YWHAG:NPTX2 could faithfully reflect CSF YWHAG:NPTX2. We used the set of 1,194 paired CSF and plasma samples ($n=597$ each) that were collected within 6 months of each other from the Knight-ADRC and Stanford cohorts. We found no significant correlation (**Supplementary Fig. 2a**). We then tested a series of frameworks to develop a plasma signature that could more accurately reflect CSF YWHAG:NPTX2 and CI.

First, we aimed to develop a plasma signature of CSF YWHAG:NPTX2 using the 240 plasma synapse proteins that changed with CI in the CSF, with the idea that perhaps many more synapse proteins beyond YWHAG and NPTX2 are needed to train a robust model from plasma. We split the 597 matched CSF and plasma samples into 478 for training (80%) and 119 for testing (20%). We used the LassoCV function from the scikit-learn⁷ Python package to train a penalized linear model to predict CI severity based on the levels of the 240 significant synapse proteins. Though the correlation between this plasma signature and CSF YWHAG:NPTX2 was decent, the correlations with CI were mild (**Supplementary Fig. 2b**).

To improve the robustness of the signature, we sought to use the same set of proteins to train a predictor of CI instead of CSF YWHAG:NPTX2 directly. This allowed us to dramatically increase training sample size as CI, but not matched CSF, was measured across many more samples and cohorts. We used Knight-ADRC and ROSMAP for training and Stanford for testing. This improved correlations with CI but weakened correlations with CSF YWHAG:NPTX2 (**Supplementary Fig. 2c**).

We believed there was additional potential for improvement. We sought to expand upon our modeling framework to include non-synapse plasma proteins, while also filtering out proteins that were unstable across cohorts and were likely affected by ApoE-binding (**Supplementary Fig. 3a**). To elaborate, we first applied the Weighted Gene Co-expression Network Analysis⁸ (WGCNA) package in R to derive 14 correlation modules from the plasma proteome. Knight-ADRC and ROSMAP cohorts were used for module training and Stanford for testing. We used the “blockwiseModules” function with the following parameters: power = 5, networkType = “unsigned”, corType=“bicor”, maxBlockSize=10000, deepSplit = 2, pamRespectsDendro = F, minModuleSize = 20, reassignThreshold=1e-6, mergeCutHeight = 0.25. Proteins not assigned to any module were assigned to the “grey” module (M0). We confirmed that modules were robust within and across cohorts, as proteins within a module tended to correlate with their respective module eigenproteins more than other module eigenproteins, and this was very consistent across train and test data (**Supplementary Fig. 3b**).

We next regressed each module eigenprotein with age, sex, cohort, *APOE4* dose, and CI (**Supplementary Fig. 4a**). We identified several modules with strong cohort-effects (**Supplementary Fig. 4b**), suggesting proteins in these modules are likely sensitive to differences in blood processing between cohorts/centers. We filtered out proteins in these modules. We also identified a single module with a strong *APOE4* effect (**Supplementary Fig. 4c**). The strong “block-like” association between *APOE* genotype and proteins in this module suggest that *APOE* genotype may alter the binding affinity between these proteins and their respective Somalogic aptamer probes, potentially due to apoE-protein interactions. Notably, NfL is a member of this module and has a strong negative *APOE4* effect, decreasing in signal with increasing *APOE4* dose. This aligns with previous reports of discordance between SomaScan and antibody-based assays for NfL⁹. We filtered out proteins in this *APOE* module. Lastly, for each of the remaining seven modules and the “grey” module (M0), we performed a Fisher’s exact test to identify modules that were enriched for synapse proteins that changed with CI independent of A β and tau in the CSF (**Fig. 1b**). Module 3 was the only module that was strongly enriched for these synapse proteins (**Supplementary Fig. 4d**), so we trained our plasma signature of CI using the 745 proteins in module 3.

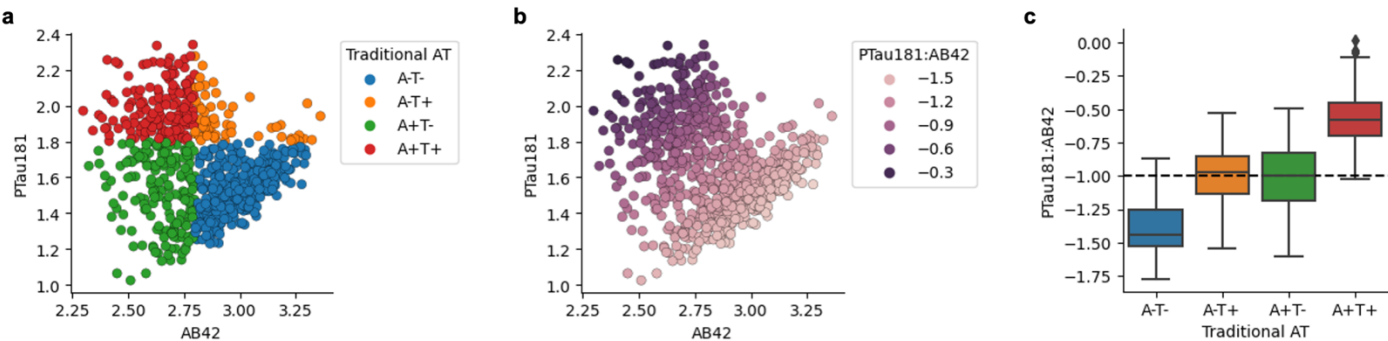
The LassoCV function from the scikit-learn⁷ Python package was used to train a penalized linear model to predict cognitive impairment severity based on the levels of the 745 plasma proteins in module 3. The same train/test split as for WGCNA was used, with Knight-ADRC and ROSMAP cohorts for training and Stanford for testing. 5-fold cross validation was implemented to identify the optimal lambda parameter. We call this model the “plasma signature” throughout the manuscript (**Fig. 4**).

Plasma signature protein weights and intercepts for both versions of the plasma signature are provided in **Supplementary Table 7**. To apply these weights to new independent plasma SomaScan datasets, follow the GitHub tutorial <https://github.com/hamiltonoh/synapsesignature>. Briefly, the steps are as follows. Plasma protein levels should be log₁₀ normalized, then z-score normalized using means and standard deviations from our training data (**Supplementary Table 7**). Additionally, if using the v4 assay and signature, simple multiplicative scale factors should be applied before log₁₀ and z-score normalization (**Supplementary Table 7**). If using the v5 assay, the v4.1 signature should be used. v5 data should be normalized to v4.1 assay space using v4 to v4.1 multiplicative scale factors (request from Somalogic) before log₁₀ and z-score normalization. After deriving the signature (linear combination of protein values and weights with final addition of an intercept value), it should be z-score normalized based on the mean and standard deviation in our training data (plasma signature mean=0.7445, std=0.5015).

REFERENCES

1. Timsina, J. *et al.* Harmonization of CSF and imaging biomarkers in Alzheimer's disease: Need and practical applications for genetics studies and preclinical classification. *Neurobiol. Dis.* **190**, 106373 (2024).
2. Karlsson, L. *et al.* Cerebrospinal fluid reference proteins increase accuracy and interpretability of biomarkers for brain diseases. *Nat. Commun.* **15**, 3676 (2024).
3. Martínez-Dubarbíe, F. *et al.* Accuracy of plasma A β 40, A β 42, and p-tau181 to detect CSF Alzheimer's pathological changes in cognitively unimpaired subjects using the Lumipulse automated platform. *Alzheimers Res. Ther.* **15**, 163 (2023).
4. Motta, C. *et al.* Different associations between amyloid- β 242, amyloid- β 40, and amyloid- β 42/40 with soluble phosphorylated-tau and disease burden in Alzheimer's disease: a cerebrospinal fluid and fluorodeoxyglucose-positron emission tomography study. *Alzheimers Res. Ther.* **15**, 144 (2023).
5. Timsina, J. *et al.* Comparative Analysis of Alzheimer's Disease Cerebrospinal Fluid Biomarkers Measurement by Multiplex SOMAscan Platform and Immunoassay-Based Approach 1. *J. Alzheimers Dis.* **89**, 193–207 (2022).
6. Öhrfelt, A. *et al.* Association of CSF GAP-43 With the Rate of Cognitive Decline and Progression to Dementia in Amyloid-Positive Individuals. *Neurology* **100**, e275–e285 (2023).
7. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
8. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
9. Eldjarn, G. H. *et al.* Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348–358 (2023).

Rationale for using PTau181:AB42 to define AT status

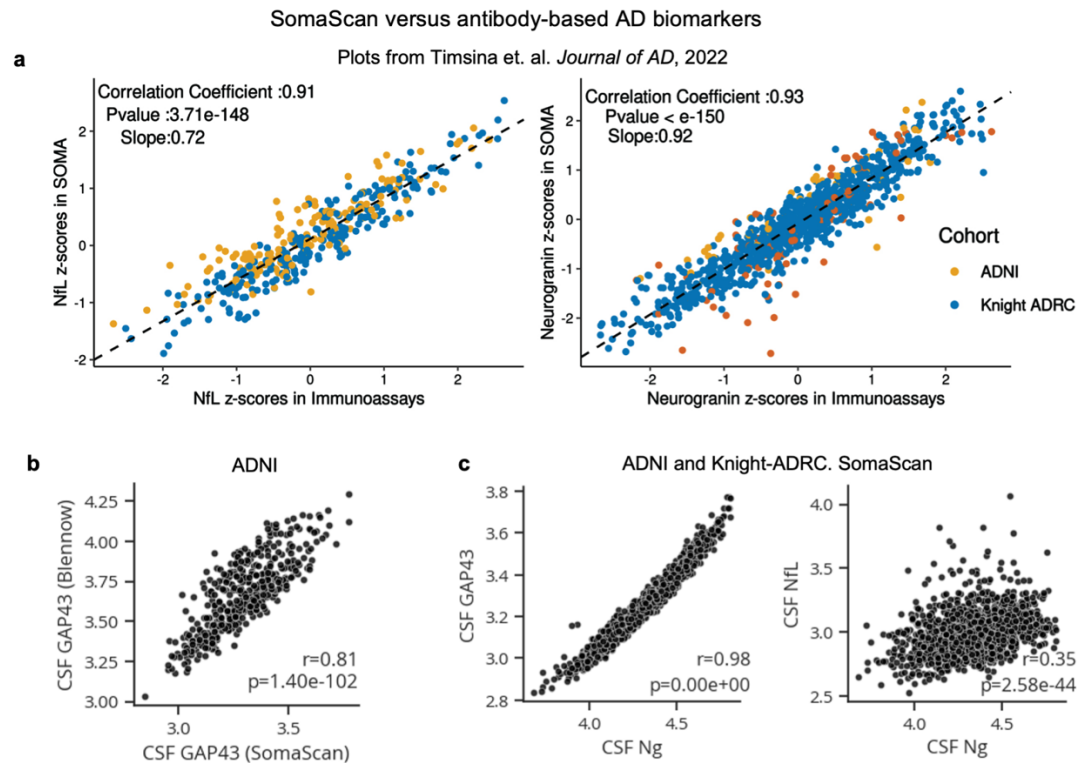


Supplementary Figure 1. AT₁ status classification using the CSF PTau181:Aβ42 ratio.

a, CSF PTau181 versus Aβ42, colored by AT₁ status based on conventional Gaussian Mixture model cutoffs per biomarker.

b, CSF PTau181 versus Aβ42, colored by CSF PTau181:Aβ42 ratio.

c, Boxplot showing PTau181:Aβ42 versus traditional AT₁ status. For this study, CSF PTau181:Aβ42 > -1 was used to define A+T₁+ individuals, and CSF PTau181:Aβ42 < -1 for A-T₁- individuals.



Supplementary Figure 2. SomaScan versus established immunoassays for “N” biomarkers.

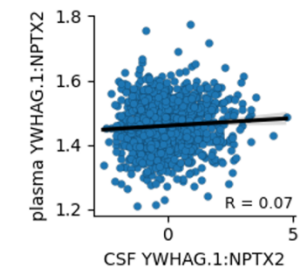
a, Directly adapted from Timsina et. al. *Journal of AD*, 2022. Plots show correlations between SomaScan and established AD immunoassays for NfL and Neurogranin (Ng) in the Knight ADRC and ADNI cohorts. Pearson correlations and p-values are shown.

b, Correlation and p-value between SomaScan GAP43 versus immunoassay GAP43 in the ADNI cohort.

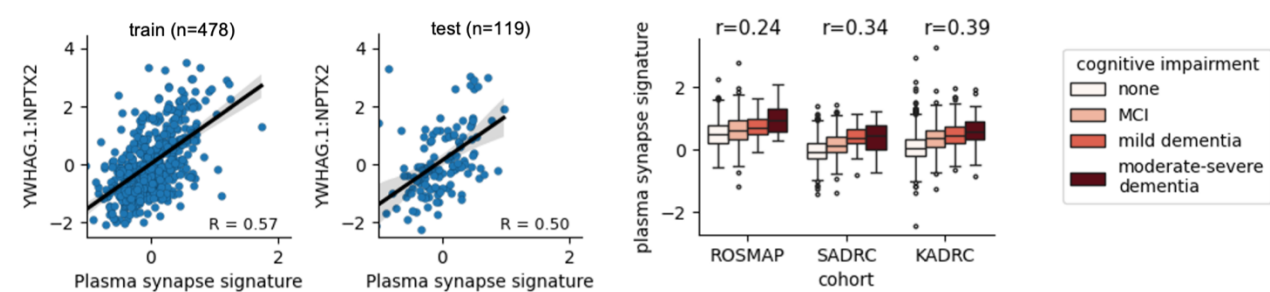
c, Correlations and p-values between CSF GAP43 versus CSF Ng (left) and CSF NfL versus CSF Ng (right).

Derivation of plasma signature of cognitive impairment

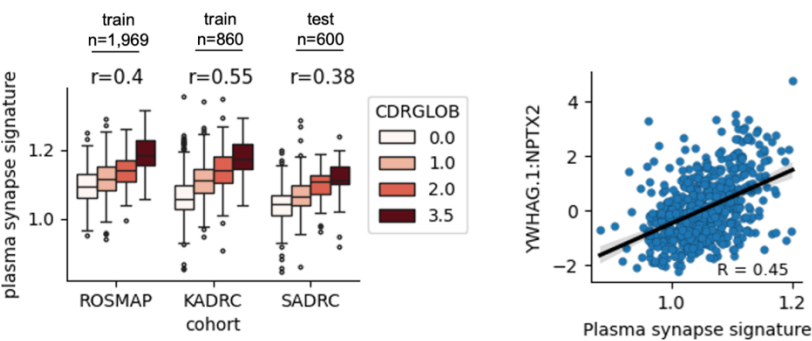
a Plasma YWHAG.1:NPTX2 versus CSF YWHAG.1:NPTX2



b Predict CSF YWHAG.1:NPTX2 using plasma synapse proteins that change with CI in the CSF (Fig. 1b)



c Predict CI stage using plasma synapse proteins that change with CI in the CSF (Fig. 1b)



Supplementary Figure 3. WGCNA of the plasma proteome across cohorts.

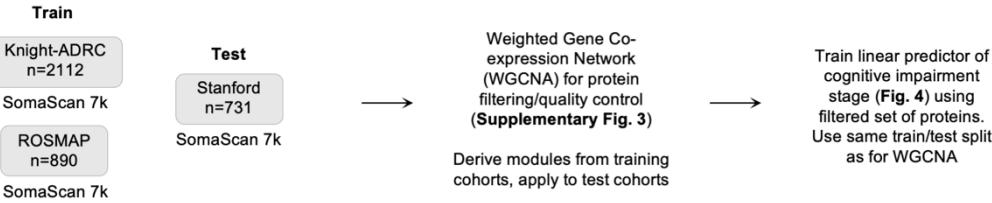
a, Plasma versus CSF YWHAG:NPTX2 in samples collected within 6 months in the Knight-ADRC and Stanford cohorts.

b, A linear model was trained to predict CSF YWHAG:NPTX2 using plasma synapse proteins that changed with CI in the CSF (Fig. 1b). Correlations between actual (y-axis) and predicted (x-axis) CSF YWHAG:NPTX2 values are shown (left). Correlations between predicted CSF YWHAG:NPTX2 values and CI stage across cohorts are shown (right).

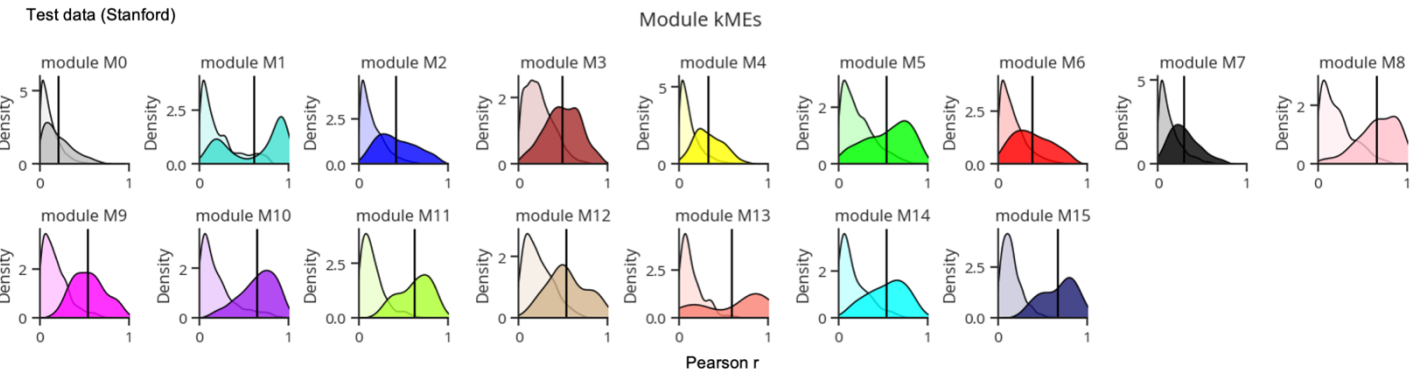
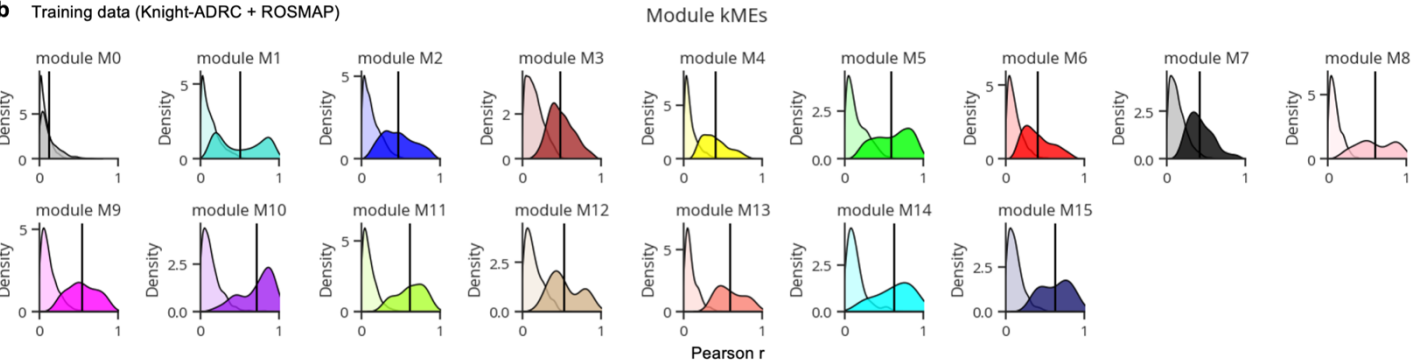
c, A linear model was trained to predict CI stage using plasma synapse proteins that changed with CI in the CSF (Fig. 1b). Correlations between actual (x-axis) and predicted (y-axis) CI stages are shown (left). Correlation between CSF YWHAG:NPTX2 and predicted CI values is shown (right).

a Derivation of plasma signature of cognitive impairment

Split into train and test cohorts.
Remove repeat visit samples.



b Training data (Knight-ADRC + ROSMAP)

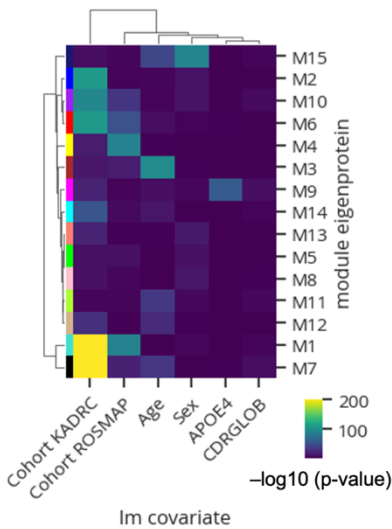
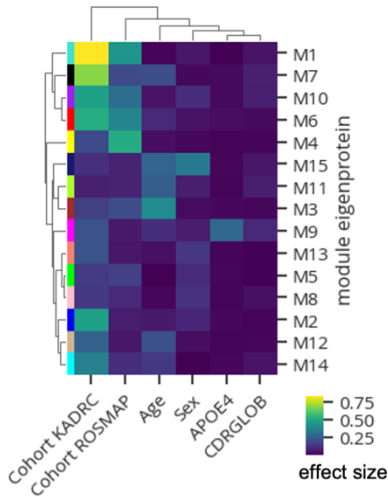


Supplementary Figure 4. WGCNA of the plasma proteome across cohorts.

a, Workflow to derive robust plasma signature of cognitive impairment.

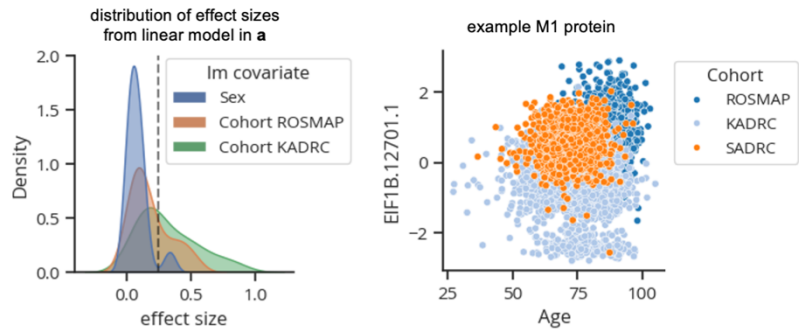
b, WGCNA resulted in 14 modules. Proteins not assigned to any modules were assigned to module zero (M0). Density plots showing the distribution of correlations between proteins and module eigenproteins are shown. Correlations between proteins and their respective module eigenproteins are included in the opaque distributions, while correlations between proteins and non-matching module eigenproteins are included in the transparent distributions. Correlations from training data are shown above and testing data below.

a Module eigenprotein ~ age + sex + cohort + *APOE4* dose + CDR-Global

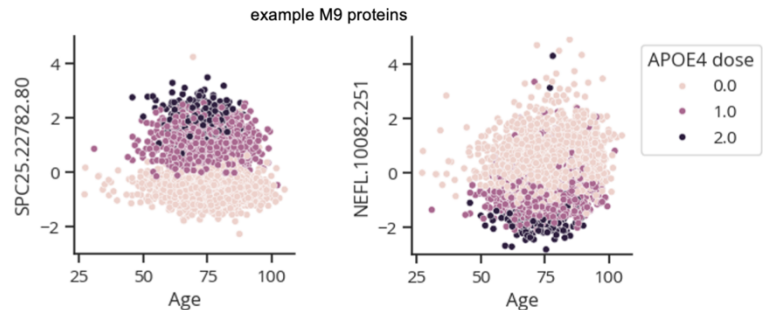


Plasma protein QC/filtering to derive signature of cognitive impairment

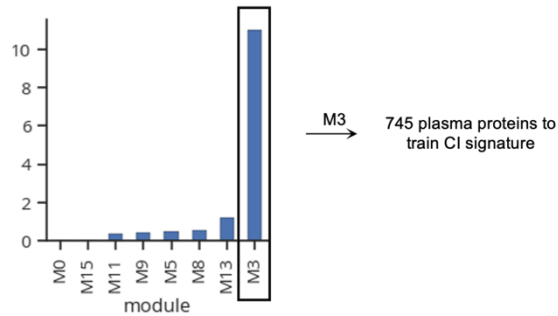
b 1. Remove proteins from modules with strong cohort effects (M1, M2, M4, M6, M7, M10, M12, M14)



c 2. Remove proteins from module with strong *APOE* genotype effect (M9)



d 3. Subset to proteins from module with strong enrichment for CI differentially expressed CSF synapse proteins from Fig. 1b



Supplementary Figure 5. Plasma protein filtering to derive plasma signature of cognitive impairment.

a, Heatmap showing results from linear models regressing each module eigenprotein with age, sex, cohort, *APOE4* dose, and cognitive impairment stage (CDR-Global). Upper heatmap shows beta effect sizes, and lower heatmap shows $-\log_{10} \text{p-values}$.

b, Distribution of effect sizes from sex and cohort covariates from the linear models in **a**. Proteins from modules with large cohort effects (>0.25) from at least one of the cohort covariates were filtered out (left). An example protein with a large cohort effect in module 1 is shown (right).

c, Proteins from the module (M9) with a large *APOE* effect (as shown in **a**) were filtered out. Two example proteins from M9 with large *APOE* effects are shown.

187 **d**, Remaining modules were assessed for their enrichment of synapse proteins that change with cognitive
188 impairment independent of A β and tau in the CSF.
189