# nature medicine

Article

https://doi.org/10.1038/s41591-025-03565-2

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

#### Received: 5 July 2024

Accepted: 4 February 2025

Published online: 31 March 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

Rates of cognitive decline in Alzheimer's disease (AD) are extremely heterogeneous. Although biomarkers for amyloid-beta (AB) and tau proteins, the hallmark AD pathologies, have improved pathology-based diagnosis, they explain only 20-40% of the variance in AD-related cognitive impairment (CI). To discover novel biomarkers of CI in AD, we performed cerebrospinal fluid (CSF) proteomics on 3,397 individuals from six major prospective AD case-control cohorts. Synapse proteins emerged as the strongest correlates of CI, independent of AB and tau. Using machine learning, we derived the CSF YWHAG:NPTX2 synapse protein ratio, which explained 27% of the variance in CI beyond CSF pTau<sub>181</sub>:A $\beta_{42}$ , 11% beyond tau positron emission tomography, and 28% beyond CSF neurofilament, growth-associated protein 43 and neurogranin in A $\beta^+$  and phosphorylated tau<sup>+</sup> (A+T<sub>1</sub>+) individuals. CSF YWHAG:NPTX2 also increased with normal aging and 20 years before estimated symptom onset in carriers of autosomal dominant AD mutations. Regarding cognitive prognosis, CSF YWHAG:NPTX2 predicted conversion from  $A+T_1+$  cognitively normal to mild cognitive impairment (standard deviation increase hazard ratio = 3.0,  $P = 7.0 \times 10^{-4}$ ) and A+T<sub>1</sub>+ mild cognitive impairment to dementia (standard deviation increase hazard ratio =  $2.2, P = 8.2 \times 10^{-16}$ ) over a 15-year follow-up, adjusting for CSF pTau<sub>181</sub>:A $\beta_{42}$ , CSF neurofilament, CSF neurogranin, CSF growth-associated protein 43, age, APOE4 and sex. We also developed a plasma proteomic signature of CI, which we evaluated in 13,401 samples, which partly recapitulated CSF YWHAG:NPTX2. Overall, our findings underscore CSF YWHAG:NPTX2 as a robust prognostic biomarker for cognitive resilience versus AD onset and progression, highlight the potential of plasma proteomics in replacing CSF measurement and further implicate synapse dysfunction as a core driver of AD dementia.

Alzheimer's disease (AD) is the most common age-related neurodegenerative disease characterized by decades-long buildup of amyloid-beta (A $\beta$ ) plaques and neurofibrillary tau tangles followed by dementia<sup>1</sup>. Rates of cognitive decline in AD are extremely heterogeneous, with symptom onset occurring between the ages of 40 and 100 years and conversion from mild cognitive impairment (MCI) to AD dementia occurring in 2-20 years<sup>2</sup>. While the development of cerebrospinal fluid (CSF) and positron emission tomography (PET) biomarkers of Aß and tau have begun to untangle this heterogeneity and have thereby improved AD diagnosis, patient stratification and drug development<sup>3-7</sup>, Aß and tau still only explain 20-40% of the variance in cognitive impairment (CI) in AD<sup>8-11</sup> (Extended Data Fig. 1a), suggesting the existence of additional drivers of AD dementia. The prevalence of AB<sup>+</sup> cognitively normal aged individuals further underscores the need for an increased understanding of what drives AD dementia versus cognitive resilience12,13.

The 'A/T/N' (A $\beta$ /tau/neurodegeneration) AD biomarker framework<sup>14</sup>, developed by the National Institute on Aging (NIA) and the Alzheimer's Association, structures the integration of biomarkers. Among CSF biomarkers, A $\beta_{42}$  is typically used to define 'A' positivity and pTau<sub>181</sub> to define 'T<sub>1</sub>' (phosphorylated secreted tau) positivity<sup>14</sup>. The CSF pTau<sub>181</sub>:A $\beta_{42}$  ratio captures both aspects simultaneously<sup>15,16</sup>. 'T<sub>2</sub>' includes emerging biomarkers of fibrillary tau proteinopathy, such as CSF pT205, CSF MTBR-243 (ref. 17) and tau PET<sup>18</sup>. The 'N' category includes A $\beta$ -independent and tau-independent biomarkers of AD, such as neurofilament light (NfL) for axon degeneration and neurogranin (Ng) for synapse dysfunction<sup>5</sup>. However, these 'N' biomarkers explain only a small additional proportion of variance in CI beyond A $\beta$  and tau<sup>5</sup>.

To discover new robust Aβ-independent and tau-independent correlates of CI in AD, we performed large-scale proteomics (SomaScan, mass spectrometry (MS)) on the CSF of 3,397 individuals across six deeply phenotyped case-control cohorts with AD spanning both sporadic and autosomal dominant AD (ADAD): Stanford (includes the Stanford Alzheimer's Disease Research Center (ADRC), Stanford Aging and Memory Study (SAMS) and Poston cohort); Knight-ADRC; Alzheimer's Disease Neuroimaging Initiative (ADNI); Dominantly Inherited Alzheimer's Network (DIAN); BioFINDER2; and Kuopio University Hospital (Fig. 1a and Supplementary Table 1). We integrated these CSF proteomics data with CSF and PET biomarkers of AB and tau, cognitive function, age, sex, APOE4 genotype and ADAD mutation status to derive a robust CSF biomarker of CI that explains CI beyond existing A/T/N biomarkers. Lastly, we derived a plasma surrogate of the CSF biomarker based on the SomaScan plasma proteomics data from 2,829 individuals from the Knight-ADRC and Religious Order Study/Memory Aging Project (ROSMAP) cohorts and evaluate the signature in an additional 9,502 individuals from Stanford, the Global Neurodegeneration Proteomics Consortium (GNPC) (https://www.neuroproteome.org/) and the Atherosclerosis Risk in Communities (ARIC) study.

# Multicohort CSF proteomics for AD biomarker discovery

We performed proteomics on 3,397 CSF samples (3,187 with a complete CI diagnosis) from six independent cohorts. To identify CSF proteins that explained additional variance of CI beyond AD pathology, we regressed the global clinical dementia rating (CDR) (a clinical CI score) against CSF protein levels, while adjusting for CSF pTau<sub>181</sub>:A $\beta_{42}$ , age, sex, *APOE4*, cohort and principal component 1 (PC1) of the proteome (Methods). We analyzed the SomaScan proteomics data (7,289 protein measurements per sample) from the Knight-ADRC (*n* = 756) and ADNI (*n* = 716) cohorts for discovery.

We identified 675 significantly (Benjamini–Hochberg q < 0.05) upregulated and 721 significantly downregulated proteins with CI (Fig. 1b and Supplementary Tables 2 and 3). The most significant proteins were enriched at the synapse (based on the SynGO database<sup>19</sup>; Fig. 1c). The most upregulated synapse proteins included YWHAG, YWHAZ, YWHAH, NEFL, NEFH, DLG2, HOMER1, MAP1LC3A, PPP3CA and PPP3R1. The YWHA family (which encode 14-3-3) proteins, DLG2 and calcineurin subunits (PPP3CA and PPP3R1) were strongly associated with pTau<sub>181</sub>:A $\beta_{42}$  (ref. 20) (Extended Data Fig. 1b,c). In line with these observations, A $\beta_{42}$  signaling promoted calcineurin activity<sup>21</sup>, and inhibition of calcineurin activity protected from A $\beta$ -induced and tau-induced synapse loss and CI in mice<sup>22,23</sup>. Notably, SMOC1, an extracellular matrix protein previously linked to AD and A $\beta$  plaques<sup>24,25</sup>, was not associated with CI after pTau<sub>181</sub>:A $\beta_{42}$  adjustment (Extended Data Fig. 1b).

The most downregulated proteins included NPTX2, NPTXR, SLI-TRK1, CBLN4, LRFN2 and EPHA4. These proteins were weakly negatively associated with pTau<sub>181</sub>:A $\beta_{42}$  (ref. 20) (Extended Data Fig. 1b,c). The most downregulated protein was NPTX2, encoded by an immediate early gene that regulates homeostatic scaling of excitatory synapses on parvalbumin interneurons<sup>26</sup> to prevent neuronal network hyperactivity<sup>27</sup>. In line with its reduction in AD CSF, NPTX2 mRNA and protein are downregulated in AD neurons based on brain single-cell RNA sequencing, immunohistochemistry and bulk proteomic studies<sup>28,29</sup>. Interestingly, overexpression of NPTX2 tau P301S in the mouse hippocampus protects synapses from complement-mediated glial engulfment<sup>30</sup>, suggesting that it may be a synaptic resilience factor.

Given the enrichment of synapse proteins, we sought to derive a multiprotein synaptic signature of CI. Using the ADNI data, we trained a penalized linear model to predict CI severity based on the levels of 214 synapse proteins that significantly changed with CI. We used recursive feature elimination (RFE) to further simplify the model to facilitate clinical applications (Fig. 1d). The model identified a near 1:1 difference between YWHAG and NPTX2 to be a suitable signature of CI (Fig. 1d). As we log-normalized the *z*-scored protein levels before the analyses, the difference between normalized protein levels represents a normalized ratio. The figures refer to YWHAG.1, a specific YWHAG proteoform detected by the SomaLogic aptamer with Seqld: 4179-57. Notably, ratios between CSF YWHA family proteins and NPTX2 based on MS have been previously associated with general AD phenotypes<sup>31,32</sup>, suggesting reproducibility across cohorts and proteomic platforms.

We validated the association of YWHAG:NPTX2 with CI across all cohorts with SomaScan data (ADNI r = 0.54; Knight-ADRC r = 0.55; Stanford r = 0.62; DIAN r = 0.66; total n = 2,067) including both sporadic AD and ADAD (Fig. 1e and Supplementary Table 4). Correlations were consistent across sexes (Extended Data Fig. 1d) and slightly exceeded the correlations of pTau<sub>181</sub>:A $\beta_{42}$  with CI (Extended Data Fig. 1e). The YWHAG:NPTX2 ratio was not significantly affected by the cohort (Fig. 1f), making it advantageous for clinical applications.

To assess the potential of using YWHAG:NPTX2 for cognitive diagnosis, we tested logistic regression models that aimed to distinguish individuals across different CI stages. We found that while pTau<sub>181</sub>:A $\beta_{42}$  had better performance in distinguishing individuals with MCI versus cognitively normal individuals, YWHAG:NPTX2 had better performance in the later stages (Fig. 1g, Extended Data Fig. 1f and Supplementary Table 5). When tasked to distinguish individuals with dementia versus cognitively normal individuals, YWHAG:NPTX2 with an area under the curve (AUC) of 0.97 was predictive (Fig. 1g and Extended Data Fig. 1g).

We next sought to determine the robustness of YWHAG:NPTX2 in explaining CI beyond AD pathology. We observed that while YWHAG:NPTX2 and pTau<sub>181</sub>:A $\beta_{42}$  were correlated (r = 0.61), low and high levels of YWHAG:NPTX2 further separated A+T<sub>1</sub>-positive (A+T<sub>1</sub>+) (log<sub>10</sub> pTau<sub>181</sub>:A $\beta_{42}$  > -1; Methods) individuals into no impairment versus dementia, respectively (Fig. 1h). Among A+T<sub>1</sub>+ individuals (n = 898), 62% of individuals with low levels of YWHAG:NPTX2 (bottom 25th percentile) were cognitively normal and 37% had only MCI, whereas only 4% of individuals with high YWHAG:NPTX2 (top 25th percentile) were cognitively normal and 46% had dementia (Extended Data Fig. 2a). This pattern was consistent across cohorts and both sporadic AD and ADAD. Using linear regression, we found that pTau<sub>181</sub>:A $\beta_{42}$ 



Fig. 1| The CSF YWHAG:NPTX2 ratio explains a substantial proportion of variance in CI beyond amyloid and tau in AD. a, Study design. Integration of CSF proteomics, AD pathology biomarkers and clinical cognitive scoring from six independent cohorts to identify the molecular correlates of CI, independently of AD pathology. b, Volcano plot showing the change with CI independent of age, sex, cohort, APOE4 dose and pTau<sub>181</sub>:A $\beta_{42}$ , and PC1 of the CSF proteome, in the Knight-ADRC and ADNI cohorts (n = 1,472). Bold indicates synapse proteins based on the SynGO database; q values are Benjamini-Hochberg-corrected P values. c, Rank-based pathway enrichment heatmap of differentially abundant proteins. Cells are color-coded according to  $-\log_{10}(q)$ . **d**, A penalized linear model was trained to predict CI severity using synaptic proteins that significantly changed with CI. RFE showed that two proteins sufficiently captured 83% of the full model performance. Model coefficients show the normalized ratio between YWHAG.1 and NPTX2. e, Box plot showing YWHAG.1:NPTX2 versus CI severity across cohorts with the SomaScan data (n = 2,067). The box bounds are the Q1, median and Q3; the whiskers show  $Q1 - 1.5 \times$  the interquartile range (IQR) and  $Q3 + 1.5 \times$  the IQR. f, CSF YWHAG.1:NPTX2 regressed against age, CI, sex and cohort in a linear model (n = 2,067). The points and error bars represent the standardized effect sizes and 95% confidence intervals. g, AUC results from the logistic regression to classify CI stage based on YWHAG.1:NPTX2 or pTau<sub>181</sub>:A $\beta_{42}$ . All SomaScan cohorts are included (n = 2,067). **h**, YWHAG.1:NPTX2 versus pTau<sub>181</sub>:A $\beta_{42}$ , color-coded according to CI. Linear correlation and P value are shown.  $i_r$  results from a linear model regressing CI against covariates displayed on the x axis in  $A+T_1+$ individuals (n = 898). The difference in  $r^2$  values between the full model and the model with only pTau<sub>181</sub>:A $\beta_{42}$  is shown. **j**, As in **h** but for YWHAZ:NPTX2 versus tau PET in  $A\beta^+$  individuals in BioFINDER2. **k**, As in **i** but with different covariates in BioFINDER2 (n = 512). The difference of  $r^2$  values between the full model and the model with only A  $\beta_{42}$  : A  $\beta_{40}$  + tau PET is shown. The bars and error bars in  $\boldsymbol{g},\boldsymbol{i}$ and k represent bootstrapped (n = 1,000) means and 95% confidence intervals. Two-sided P values were calculated using the empirical distribution of the bootstrapped test statistic. \*\*P < 0.01, \*\*\*P < 0.001. Graphics in a created with BioRender.com.

explained 10% of the variance in CI in A+T<sub>1</sub>+ individuals, YWHAG:NPTX2 explained 36% and YWHAG:NPTX2 explained an additional 27% beyond pTau<sub>181</sub>:A $\beta_{42}$  (Fig. 1i). YWHAG:NPTX2 was significantly associated with CI independently of pTau<sub>181</sub>:A $\beta_{42}$  and age in A+T<sub>1</sub>+ individuals across all cohorts and proteomic platforms (Extended Data Fig. 2b). Notably, in the DIAN ADAD cohort, YWHAG:NPTX2 was associated with CI even

after accounting for the estimated age at symptom onset (Extended Data Fig. 2b).

While  $pTau_{181}$ :  $A\beta_{42}$  is a robust biomarker of  $A\beta$  plaques and phosphorylated secreted tau, it is not well representative of tau tangle load (T<sub>2</sub>), which is known to correlate with CI more strongly<sup>18</sup>. Therefore, we analyzed the BioFINDER2 cohort; we performed time-point-matched

targeted CSF synapse protein MS proteomics, tau PET imaging and CSF  $A\beta_{42}$ :  $A\beta_{40}$  measurement<sup>32</sup>. As YWHAG was not measured, we analyzed YWHAZ, a related protein (r = 0.92; Extended Data Fig. 2c), which was also associated with CI albeit not as strongly (Fig. 1b and Extended Data Fig. 2d). We confirmed the association of YWHAZ:NPTX2 with CI based on MS in BioFINDER2 (r = 0.63; Extended Data Fig. 2e). We plotted YWHAZ:NPTX2 versus tau PET in  $A\beta^+$  individuals (*n* = 512), color-coded according to CI severity (Fig. 1j). We observed several interesting patterns. First, we found a moderate correlation between YWHAZ:NPTX2 and tau PET (r = 0.45). Second, we observed that all individuals with above moderate levels of tau had above moderate levels of YWHAZ:NPTX2, but not vice versa, suggesting that YWHAZ:NPTX2 may change before tau during AD progression. Third, we observed that YWHAZ:NPTX2 and tau PET independently explained CI severity. Using linear regression, we found that  $A\beta_{42}$ :  $A\beta_{40}$  and tau PET together explained 35% of the variance in CI in A $\beta^+$  individuals; YWHAZ:NPTX2 explained an additional 11% beyond  $A\beta_{42}$ :  $A\beta_{40}$  and tau PET (Fig. 1k). The association of YWHAZ:NPTX2 with CI was robust to additional adjustment with age, APOE4 dose, sex and CSF NfL (Extended Data Fig. 2f,g). Although YWHAG:NPTX2 and YWHAZ:NPTX2 were highly correlated (r = 0.94) in SomaScan (Extended Data Fig. 2c), additional studies are needed to confirm the association of YWHAG:NPTX2 with Cl independently of tau PET.

Together, these results show that CSF synapse proteins, some with established causal roles in synaptic and cognitive resilience to AD pathology in mouse models (that is, calcineurin, NPTX2), are among the strongest correlates of CI severity independent of A $\beta$  and tau in humans, and that the CSF YWHAG:NPTX2 ratio is a synapse protein signature that explains a major proportion of variance in CI in AD beyond gold standard biomarkers of A $\beta$  and tau.

# CSF YWHAG:NPTX2 versus established neurodegeneration AD biomarkers

While we observed that CSF YWHAG:NPTX2 explained CI beyond A $\beta$  and tau, whether it also explained CI beyond established biomarkers of neurodegeneration and synapse dysfunction was unknown. Therefore, we compared CSF YWHAG:NPTX2 with CSF NfL, growth-associated protein 43 (GAP-43) and Ng<sup>5</sup>, which were also measured on the SomaScan assay. We confirmed that the levels of these biomarkers based on SomaScan were highly correlated with levels based on established immunoassays used in AD research<sup>33,34</sup> (Supplementary Fig. 2). We further confirmed that the levels of these biomarkers were higher in AD and A+T<sub>1</sub>+ individuals compared to controls in the Knight-ADRC and ADNI cohorts (Fig. 2a).

We examined pairwise correlations between YWHAG:NPTX2 and these 'N' biomarkers in the Knight-ADRC and ADNI cohorts (total n = 1,472) and found that YWHAG:NPTX2 was notably distinct from the rest (Fig. 2b-d). To our surprise, CSF YWHAG:NPTX2 was slightly negatively correlated with GAP-43 (r = -0.07) and Ng (r = -0.09), even though all biomarkers were positively correlated with AD.

Regarding associations with CI, we observed that NfL, GAP-43 and Ng each and together explained only a small proportion of the variance in CI (1–5%) in A+T<sub>1</sub>+ individuals (Fig. 2e). Conversely, YWHAG:NPTX2 explained 31% of the variance, 28% beyond NfL, GAP-43 and Ng.

These results suggest that the levels of YWHAG:NPTX2 represent a 'pathology' that is distinct from previously reported AD neurodegeneration and synapse dysfunction biomarkers and one that is much more closely related to CI.

## CSF YWHAG:NPTX2 in normal aging and ADAD

As age is the strongest risk factor for AD onset, we next asked whether YWHAG:NPTX2 increases during normal aging before CI. Surprisingly, we found that YWHAG:NPTX2 increased with age not only in later decades, but also in the earliest decades of adulthood, -30 years before changes in pTau<sub>181</sub>:A $\beta_{42}$  (Fig. 3a). This pattern was replicated in the BioFINDER2 cohort (Extended Data Fig. 3a).



**Fig. 2**|**CSF YWHAG:NPTX2 versus established neurodegeneration AD biomarkers. a**, Levels of CSF NfL, GAP-43, Ng and YWHAG.1:NPTX2 between AD and healthy and A+T<sub>1</sub>+ and A-T<sub>1</sub>- in the Knight-ADRC and ADNI cohorts are shown. \*\*\**P* < 0.001 based on a standard two-sided *t*-test. The box bounds show the Q1, median and Q3; the whiskers show the Q1 – 1.5× the IQR and Q3 + 1.5× the IQR. **b**, Linear correlation and *P* value between CSF YWHAG.1:NPTX2 and CSF NfL in the Knight-ADRC and ADNI cohorts. The colors indicate the CI stage as shown in Fig. 1e. **c**, As in **b** but for CSF GAP-43. **d**, As in **b** but for CSF Ng. **e**, *r*<sup>2</sup> results from linear models regressing CI against covariates displayed on the *x* axis in the Knight-ADRC and ADNI cohorts (*n* = 1,472). The bars and error bars represent bootstrapped (*n* = 1,000) means and 95% confidence intervals. Two-sided *P* values were calculated using the empirical distribution of the bootstrapped test statistic. The difference in *r*<sup>2</sup> values between the full model and the model with NfL, GAP-43 and Ng is shown. \*\*\**P* < 0.001.

To determine whether changes with age in YWHAG:NPTX2 precede AD symptom onset, we leveraged data from ADAD mutation carriers in the DIAN cohort who have genetically determined early-onset AD. Specifically, we tested whether YWHAG:NPTX2 had a steeper age-related increase in presymptomatic carriers versus noncarriers. We tested a linear model regressing YWHAG:NPTX2 against carrier status, age and their interaction, among cognitively normal individuals aged under 55 years, the age range where noncarriers are A-T<sub>1</sub>-(Fig. 3a). We found that carriers had significantly higher YWHAG:NPTX2 (ref. 35)



Age group to CI stage

Fig. 3 | CSF YWHAG:NPTX2 ratio increases with normal aging and presymptomatic ADAD. a, Changes with age of YWHAG.1:NPTX2 and pTau<sub>181</sub>:A $\beta_{42}$  in cognitively normal non-ADAD mutation carriers. Locally weighted scatterplot smoothing regression lines with 95% confidence intervals are shown. b, Changes with age of YWHAG.1:NPTX2 in cognitively normal individuals aged under 55 years stratified according to ADAD mutation carrier status. *P* values from a linear model regressing YWHAG.1:NPTX2 against ADAD carrier status, age and their interaction are shown. Linear regression lines with 95% confidence intervals for carriers versus noncarriers are shown. c, Association between mean EAO and slope of YWHAG.1:NPTX2 change with age. Spearman correlation and *P* value are shown. Data from noncarriers are shown for comparison. The linear regression line with the 95% confidence intervals is shown. d, Changes with age of YWHAG.1:NPTX2 in cognitively normal non-ADAD mutation carriers stratified according to *APOE* genotype. *P* values from a linear model regressing YWHAG.1:NPTX2 against *APOE4* dose, age and their interaction are shown.

 $(P = 7.21 \times 10^{-13})$  and a steeper age-related increase (two times increase

in slope, interaction P = 0.032) (Fig. 3b). ADAD mutations have varying degrees of severity, with estimated ages at symptom onset ranging from 25 to 65 years depending on the mutation<sup>25,36</sup>. To determine whether age-related slopes of YWHAG:NPTX2 correlate with ages at symptom onset, we grouped presymptomatic carriers into bins based on estimated age at onset (EAO) (<35, 35–45, 45–55, 55–65) and calculated age-related YWHAG:NPTX2 slopes per bin. We observed a strong negative correlation between mean EAO per bin and age-related YWHAG:NPTX2 slopes (Spearman r = -0.9, P = 0.037), whereby those with earlier ages at symptom onset had steeper age-related increases in YWHAG:NPTX2 (Fig. 3c and Extended Data Fig. 3b).

We then examined the effects of the *APOE* genotype, the leading genetic risk factor for sporadic AD, on age-related YWHAG:NPTX2 slopes. We regressed YWHAG:NPTX2 against *APOE4* (high-risk allele) dose, age and their interaction in cognitively normal individuals across the lifespan from the Knight-ADRC, ADNI and Stanford SomaScan cohorts. Like the ADAD mutation carrier status, *APOE4* was significantly associated with higher YWHAG:NPTX2 ( $P = 7.50 \times 10^{-6}$ ) and a steeper increase in YWHAG:NPTX2 with age (33% increase in slope compared to *APOE3*/3 homozygotes,  $P = 3.40 \times 10^{-3}$ ; Fig. 3d). *APOE2* (protective allele) carriers showed no significant differences, although we suspect this may be because of the limited sample size.

Linear regression lines with 95% confidence intervals for specified *APOE* genotypes are shown. **e**, Box plot showing changes in YWHAG.1:NPTX2 across different age groups and CI stages (*n* = 1,846). The box bounds show the Q1, median and Q3; the whiskers show Q1 – 1.5× the IQR and Q3 + 1.5× the IQR; s.d. changes in YWHAG.1:NPTX2 with cognitively normal aging and cognitive decline are shown. **f**, Changes with estimated years to symptom onset (EYO) of YWHAG.1:NPTX2 stratified according to ADAD carrier status. ADAD carrier points are color-coded according to CI stage as in Fig. 1e. Linear regression lines with 95% confidence intervals for carriers versus noncarriers before and after estimated symptom onset are shown. Slopes for carriers are shown. **g**, Changes with age and CI stage of YWHAG.1:NPTX2 for all individuals. Points are colorcoded according to CI stage as in Fig. 1e and sized according to Aβ positivity. **h**, Schematic of proposed model showing that changes in YWHAG.1:NPTX2 with cognitively normal aging underlie age at AD onset.

Our analyses thus far revealed YWHAG:NPTX2 increases with normal aging and presymptomatic AD as well as CI severity during AD progression. We next compared the degrees to which YWHAG:NPTX2 increases during these two phases. We found that YWHAG:NPTX2 increased by 1.95 s.d. over 60 years of normal aging and then 2.66 s.d. from cognitively normal aged to late-stage dementia (Fig. 3e). Although cognitively normal versus dementia groups were age-matched given the nature of case-control studies, assuming -20 years for progression from A $\beta^+$  cognitively normal to AD dementia based on population-based studies<sup>37,38</sup>, our data suggest that AD progression mimics 82 years of 'normal' age-related increases in YWHAG:NPTX2, representing a stark 4.1-time increase in slope during AD progression compared to normal aging.

We examined this phenomenon in ADAD by plotting YWHAG:NPTX2 versus EYO. We compared YWHAG:NPTX2 slopes before and after estimated symptom onset and, similar to our estimates in sporadic AD, we observed a 3.4-time increase in the YWHAG:NPTX2 slope during ADAD symptom progression compared to the presymptomatic phase (Fig. 3f). Notably, YWHAG:NPTX2 increased in ADAD ~22 years before estimated symptom onset (Fig. 3f).

To obtain a bird's-eye view of all these data, we plotted YWHAG:NPTX2 versus age color-coded according to CI stage and sized according to  $AT_1$  status for ADAD carriers versus noncarriers (Fig. 3g). We confirmed the extremely accelerated increase in YWHAG:NPTX2 among ADAD mutation carriers leading up to early-onset AD, as well as





in A+T<sub>1</sub>+ individuals with MCI to mild dementia, while adjusting for pTau<sub>181</sub>:A $\beta_{42}$ , CSF NfL, CSF Ng, APOE4, age, sex and CI stage. Results from a cross-cohort, fixed effects meta-analysis is shown (total n = 520). The points and error bars represent HRs and 95% confidence intervals. g, As in f but for predicting dementia onset in A+T<sub>1</sub>+ cognitively normal individuals (total n = 171). CI stage was not included as a covariate because all individuals were cognitively normal. h. Cox proportional hazards regression was used to associate YWHAG.1:NPTX2 with future cognitive decline in A+T<sub>1</sub>+ individuals, while adjusting for  $pTau_{181}$ :AB<sub>42</sub>, CSF NfL, CSF Ng. APOE4, age, sex and CI stage in the combined sample (n = 697). The points and error bars represent the HRs and 95% confidence intervals for each covariate. i, Kaplan–Meier curves with 95% confidence intervals showing the rates of future cognitive decline in A+T<sub>1</sub>+ versus A-T<sub>1</sub>- individuals. The HR and 95% confidence intervals is shown. **j**, As in **i** but for YWHAG.1:NPTX2<sub>high</sub> (top 25th percentile) versus YWHAG.1:NPTX2<sub>low</sub> (bottom 25th percentile) individuals. k, As in i but for A+T<sub>1</sub>+YWHAG.1:NPTX2<sub>high</sub> versus A-T<sub>1</sub>-YWHAG.1:NPTX2<sub>low</sub> individuals. Graphics in a and e created with BioRender.com.

the widespread heterogeneity in noncarriers leading to sporadic AD in some and cognitive maintenance in others, despite amyloid positivity and old age (Fig. 3g,h).

Collectively, these results demonstrate that YWHAG:NPTX2, a robust correlate of CI severity in AD, substantially increases with cognitively normal aging and presymptomatic AD.

# CSF YWHAG:NPTX2 associations with future AD progression

We next sought to determine the potential clinical utility of YWHAG:NPTX2 in predicting future AD onset and progression. First,

we leveraged A $\beta$  and tau PET imaging data collected 4–15 years after CSF draw in the ADNI cohort to assess whether YWHAG:NPTX2 predicts A $\beta$ -driven tau accumulation (Fig. 4a). Using linear regression, we found that YWHAG:NPTX2 modified the future association between A $\beta$  and tau PET (YWHAG:NPTX2 × A $\beta$  PET interaction,  $P = 6.84 \times 10^{-4}$ ), adjusting for baseline CI, pTau<sub>181</sub>:A $\beta_{42}$ , age, sex and *APOE4* (n = 120). Among individuals with high future A $\beta$  load, high baseline YWHAG:NPTX2 was associated with higher future tau PET, while low YWHAG:NPTX2 was associated with limited A $\beta$ -related tau PET increase (Fig. 4b). These results align with previous studies showing that A $\beta$  combined with synapse dysfunction and neuronal hyperactivity drives tau accumulation



**Fig. 5** | **Defined YWHAG:NPTX2 groups predict future cognitive resilience versus decline. a**, YWHAG.1:NPTX2 subgroups were defined using logistic regression to predict CI stage in individuals in adjacent CI stages. The Youden index was used to define cutoffs. Moderate and severe dementia were combined because of the limited sample size. An additional –1 group was defined based on 95% sensitivity in the model classifying MCI versus cognitively normal. The colors of the distributions indicate the CI stage as shown in Fig. 1e. b, Kaplan–Meier curves with 95% confidence intervals showing the probabilities of cognitive

maintenance (no change in CI stage) among MCI A+T<sub>1</sub>+ individuals (Knight-ADRC and ADNI cohorts) stratified according to the YWHAG.1:NPTX2 groups. HRs and 95% confidence intervals from a cross-cohort, fixed effects meta-analysis comparing groups to the 'correctly' classified group, adjusted for pTau<sub>181</sub>:A $\beta_{42}$ , CSF NfL, CSF Ng, age, *APOE4* and sex are shown (total *n* = 397). **c**, As in **b** but in cognitively normal A+T<sub>1</sub>+ individuals (total *n* = 168). CI stage was not included as a covariate because all individuals were cognitively normal.

and propagation  $^{39,40}$ , and a study showing that levels of CSF GAP-43 modify the rate of A\beta-driven tau accumulation  $^{41}$ .

More important than predicting future tau tangle buildup is predicting future cognitive decline. We plotted future Alzheimer's Disease Assessment Scale-13 (ADAS13) cognitive score versus future tau load, color-coded according to baseline YWHAG:NPTX2 in future A $\beta$  PET<sup>+</sup> individuals (n = 70) (Fig. 4c). ADAS13 was chosen for its sensitivity and dynamic range over global CDR. Among individuals with low-to-mild tau buildup, we observed that YWHAG:NPTX2 distinguished cognitively normal versus impaired individuals (Fig. 4c). All individuals with high tau PET had high YWHAG:NPTX2. Using linear regression, we found that  $pTau_{181}$ :A $\beta_{42}$ , A $\beta$  PET and tau PET together explained 41% of the variance in ADAS13 in A+T<sub>1</sub>+ individuals, and YWHAG:NPTX2 explained an additional 13% (Fig. 4d). We confirmed that YWHAG:NPTX2 was significantly associated with future ADAS13, while adjusting for tau tangle load and several additional covariates (Extended Data Fig. 4a).

To more broadly assess whether YWHAG:NPTX2 could predict future cognitive decline independent of A $\beta$  and tau, we used data from all cohorts with longitudinal cognitive follow-up (ADNI, Knight-ADRC, Stanford; Fig. 4e). Global CDR CI staging was used because it was measured across all cohorts. We analyzed both dementia progression from an MCI to mild dementia baseline, as well as dementia onset from a cognitively normal baseline.

Using Cox proportional hazards regression, we tested the association of YWHAG:NPTX2 with a future CI stage increase among A+T<sub>1</sub>+ individuals with MCI to mild dementia over 1–15 years, adjusting for baseline CI, pTau<sub>181</sub>:A $\beta_{42}$ , CSF NfL, CSF Ng (CSF GAP-43 was redundant with Ng; Supplementary Fig. 2c), age, sex and *APOE4* dose in each cohort (total *n* = 520). YWHAG:NPTX2 significantly predicted future cognitive decline across all cohorts; in a meta-analysis, an s.d. deviation increase in YWHAG:NPTX2 conferred a 124% increased risk of cognitive decline (meta hazard ratio (HR) = 2.24, meta *P* = 8.16 × 10<sup>-16</sup>; Fig. 4f and Supplementary Table 6).

We then tested whether YWHAG:NPTX2 could predict dementia onset in A+T<sub>1</sub>+ cognitively normal individuals, adjusting for pTau<sub>181</sub>:A $\beta_{42}$ , CSF NfL, CSF Ng, age, sex and *APOE4* dose (the Stanford cohort was not included because of a low event sample size, total *n* = 171). YWHAG:NPTX2 significantly predicted dementia onset across all cohorts; in a meta-analysis, an s.d. increase in YWHAG:NPTX2 conferred a 197% increased risk of conversion from cognitively normal to dementia (meta HR = 2.97, meta P = 7.03 × 10<sup>-4</sup>; Fig. 4g and Supplementary Table 7).

Aggregating data from all A+T<sub>1</sub>+ cognitively normal, MCI and mild dementia individuals across cohorts (total *n* = 697), YWHAG:NPTX2 was the strongest predictor of future cognitive decline among covariates (HR = 2.26, P = 2.24 × 10<sup>-22</sup>; Fig. 4h). Notably, YWHAG:NPTX2 was more strongly associated with future cognitive decline than the established biomarkers pTau<sub>181</sub>:A $\beta_{42}$ , CSF NfL and CSF Ng in the multivariate model.

To aid patient stratification, we binned individuals into binary high and low groups (upper and lower 25th percentiles) and tested AT<sub>1</sub> status and YWHAG:NPTX2 status in predicting future cognitive decline, individually and together. As done previously, we aggregated data from all A+T<sub>1</sub>+ cognitively normal, MCI and mild dementia across cohorts. Based on AT<sub>1</sub> status alone, we found that A+T<sub>1</sub>+ individuals had a four-time increased risk of future cognitive decline compared to A-T<sub>1</sub>-negative  $(A-T_1-)$  individuals (HR = 3.96,  $P = 5.94 \times 10^{-16}$ ; Fig. 4i). Surprisingly, based on YWHAG:NPTX2 status alone, YWHAG:NPTX2<sub>high</sub> individuals had a striking 15-time increased risk of future cognitive decline compared to YWHAG:NPTX2<sub>low</sub> individuals (HR = 15.36,  $P = 8.04 \times 10^{-48}$ ; Fig. 4j). Combining both biomarkers, A+T<sub>1</sub>+ and YWHAG:NPTX2<sub>high</sub> individuals had a 19-time increased risk of future cognitive decline compared to A-T<sub>1</sub>-YWHAG:NPTX2<sub>low</sub> individuals (HR = 18.87,  $P = 3.74 \times 10^{-25}$ ; Fig. 4k). No additional covariates were included in these Cox models, demonstrating the power of these biomarkers alone in predicting future cognitive decline versus maintenance.

Together, these results demonstrate that YWHAG:NPTX2 provides additional prognostic clinical utility beyond gold standard AD biomarkers.

# Five defined CSF YWHAG:NPTX2 groups for cognitive prognosis

Given that percentiles may vary depending on the cohorts analyzed, we aimed to establish YWHAG:NPTX2 thresholds for stratifying individuals into defined low and high groups, which can be validated in future studies and potentially applied in clinical trials. While A/T/N status is typically binary (positive versus negative), we hypothesized that YWHAG:NPTX2, which correlates with increasing Cl stages (five groups), could define multiple groups with varying risk of cognitive decline.

We used the Youden index from binary logistic regression models (MCI versus normal, mild dementia versus MCI, moderate versus mild



Fig. 6 | The plasma proteomic signature of CI partly recapitulates the CSF YWHAG:NPTX2 ratio, predicting AD onset and progression. a, Box plot showing plasma signature versus CI severity across cohorts (total n = 6,301). The box represent the Q1, median and Q3; the whiskers show Q1-1.5× the IQR and Q3+1.5× the IQR. Pearson correlations are shown. b, Protein coefficients for the plasma signature. c, Correlations between the plasma signature and CSF YWHAG.1:NPTX2 in the Knight-ADRC and Stanford cohorts (n = 518). The colors indicate the CI stage as shown in a. Linear regression lines with 95% confidence intervals for cognitively normal versus cognitively impaired individuals are shown. d, Plasma signature versus neurofibrillary tau tangle load in the ROSMAP cohort, color-coded according to CI as in **a**. **e**, r<sup>2</sup> values from linear models regressing CI against covariates displayed on the x axis in the ROSMAP cohort (n = 110). The bars and error bars represent bootstrapped (n = 1,000) means and 95% confidence intervals. Two-sided P values were calculated via the empirical distribution of the bootstrapped test statistic. The difference in r values between the two models is shown. \*\*\*P < 0.001. f, Results from a multivariate linear model regressing CI against the displayed covariates in the ROSMAP cohort

(n = 110). The points and error bars represent the standardized effect size and 95% confidence interval. g, Cox proportional hazards regression was used to associate the plasma signature with future cognitive decline in individuals with MCI to mild dementia, while adjusting for APOE4, age, sex and CI stage. The results from a cross-cohort, fixed effects meta-analysis are shown (total n = 1,877). The points and error bars represent the HRs and 95% confidence intervals. h, As in g but for predicting dementia onset in cognitively normal individuals (total n = 4,753). CI stage was not included as a covariate because all individuals were cognitively normal. i, Cox proportional hazards regression was used to associate the plasma signature with future cognitive decline in all individuals across the Knight-ADRC, ROSMAP and Stanford cohorts, while adjusting for APOE4, age, sex and CI stage (n = 2,292). The points and error bars represent the HRs and 95% confidence intervals for each covariate. j, Kaplan-Meier curves with 95% confidence intervals showing the rates of future cognitive decline in plasma signature<sub>high</sub> (top 25th percentile) versus plasma signature<sub>low</sub> (bottom 25th percentile) individuals. HR and 95% confidence interval are shown.

dementia; Fig. 1g) to define four YWHAG:NPTX2 groups tracking CI stages (0, no CI; 1, MCI; 2, mild dementia; 3, moderate-to-severe dementia; Fig. 5a). Moderate and severe dementia were combined because of the limited sample size. Additionally, we defined a '-1' group representing predominantly cognitively normal individuals (95% sensitivity for MCI versus normal).

Interestingly, we observed that many individuals were classified into YWHAG:NPTX2 groups that did not match their diagnosis (Fig. 5a). We hypothesized these 'misclassifications' may provide prognostic insights into future cognitive maintenance or decline. We focused on MCI and cognitively normal  $A+T_1+$  individuals, which are key populations for AD dementia prevention clinical trials. First, we stratified MCI A+T<sub>1</sub>+ individuals in the Knight-ADRC and ADNI cohorts (n = 397) into these YWHAG:NPTX2 groups and determined their rates of future cognitive decline (defined as an increase in CI stage) over 15 years using a Cox proportional regression meta-analysis. Interestingly, we found that relative to 'correctly' classified individuals with MCI (group 1), those classified into the 'cognitively normal' group (group 0) had a 73% reduced risk of cognitive decline (meta HR = 0.27, meta  $P = 7.35 \times 10^{-6}$ ), while those classified into the 'mild dementia' group (group 2) had a 2.3-time increased risk of cognitive decline (meta HR = 2.30, meta  $P = 2.74 \times 10^{-6}$ ), adjusted for pTau<sub>181</sub>:A $\beta_{42}$ , CSF NfL, CSF Ng, age, sex and *APOE4* (Fig. 5b, Extended Data Fig. 5a and Supplementary Table 8). Specifically, only 17 of 77 (22%) individuals in group 0 experienced cognitive decline, while 93 of 133 (70%) in group 2 declined over 15 years (Fig. 5b and Supplementary Table 8).

Next, among cognitively normal A+T<sub>1</sub>+ individuals (n = 168), those in the 'MCI' group (group 1) had a 2.7-time increased risk of cognitive decline (meta HR = 2.72, meta  $P = 7.16 \times 10^{-4}$ ), and those in the 'mild dementia' group (group 2) had a 4.9-time increased risk of cognitive decline (meta HR = 4.92, meta  $P = 2.39 \times 10^{-3}$ ) compared to 'correctly' classified cognitively normal individuals (group 0), adjusted for pTau<sub>181</sub>:A $\beta_{42}$ , CSF NfL, CSF Ng, age, sex and *APOE4* (Fig. 5c, Extended Data Fig. 5b and Supplementary Table 9). Notably, all cognitively normal individuals or individuals with MCI in the high confidence cognitively normal group (group -1) maintained cognition over 10 years, while almost all in the 'moderate-to-severe dementia' group (group 3) declined within 1 year, although sample sizes were too small for rigorous statistics (Fig. 5b,c).

Overall, we demonstrate that individuals with the same clinical diagnosis and  $AT_1$  positivity can be further stratified into defined YWHAG:NPTX2 groups, which are strongly associated with future cognitive outcomes. Thresholds defining these groups are provided in Methods for future validation and exploration in clinical trials.

## Partial plasma proteomic surrogate of CSF YWHAG:NPTX2

While CSF biomarkers provide important insights for AD research and in the clinic, the invasiveness of CSF extraction limits widespread clinical use. Thus, we sought to derive a plasma proteomics-based biomarker of CI that could recapitulate CSF YWHAG:NPTX2. We performed SomaScan plasma proteomics on 4,245 samples from the Knight-ADRC, Stanford and ROSMAP cohorts (Supplementary Table 10); 3,899 samples had complete CI diagnosis and 519 samples from the Knight-ADRC and Stanford cohorts were collected within 6 months of CSF samples from the same individuals, enabling direct plasma–CSF comparisons.

We first tested the correlations between plasma YWHAG:NPTX2 with CI and CSF YWHAG:NPTX2 and found no significant correlations (Supplementary Fig. 3a). We then systematically tested several frameworks to optimize correlations between the plasma signature with CI and CSF YWHAG:NPTX2 (Supplementary Methods and Supplementary Figs. 3-5). Briefly, the optimal framework used a penalized linear model trained on plasma protein levels to predict CI based on a subset of plasma proteins that were (1) enriched for synapse proteins associated with CI in CSF, (2) unaffected by cohort and (3) not subject to putative APOE genotype-based proteoform-aptamer binding alterations. We trained the plasma signature on unique patient samples from the Knight-ADRC (n = 1,969) and ROSMAP (n = 860) cohorts and tested on unique patient samples from the Stanford cohort (n = 600). We also tested the signature in the GNPC (F. B. Imam et al., manuscript in preparation), which encompasses over 40,000 patient samples from over 20 international research groups. Specifically, we tested the signature on a subset of 2,872 unique patient samples with complete Clinformation (global CDR) and without a diagnosis of non-AD neurodegeneration (five independent cohorts; Knight-ADRC, Stanford and ROSMAP not included).

The plasma signature (Supplementary Table 11) correlated with CI across all training and test cohorts (Knight-ADRC r = 0.66; ROSMAP r = 0.62; Stanford r = 0.51; GNPC-L r = 0.54; GNPC-E r = 0.47; GNPC-P r = 0.23; GNPC-I r = 0.68; GNPC-N r = 0.54; total n = 6,301; Fig. 6a and Supplementary Table 12). Proteins with strong weights included CPLX2, PTPRD, PI3, MAG and PTGDS (increased with CI), and NPTXR, SEZ6L, CD93, TPPP3 and PIANP (decreased with CI) (Supplementary Table 11). Notably, CPLX2, PTPRD, NPTXR (the receptor for NPTX2) and SEZ6L are synaptic proteins, confirming synapse protein associations with CI across both CSF and plasma (Fig. 6b). The plasma signature correlated with CSF YWHAG:NPTX2 (Knight-ADRC r = 0.57; Stanford r = 0.53; Fig. 6c), with stronger correlations observed in individuals with some degree of CI (CI ≥ MCI r = 0.64; CI = none r = 0.28; Fig. 6c).

To assess whether the plasma signature, like CSF YWHAG:NPTX2, explained CI beyond A $\beta$  and tau in AD, we used the ROSMAP cohort, which has collected comprehensive neuropathological and cognitive data from participants. We analyzed 110 individuals whose blood draws were within 2 years of death; autopsies confirmed a neuropathological diagnosis of AD (neuritic plaques Consortium to Establish a Registry for Alzheimer's Disease score = probable or definite; Braak stage  $\geq$  III). Plotting the plasma signature versus neurofibrillary tau tangle load, color-coded according to CI severity, we observed that high plasma signature levels were correlated with CI beyond tau levels (Fig. 6d). Linear regression showed that the plasma signature explained 36% of the variance in CI beyond neuritic A $\beta$  plaque and tau tangle load (Fig. 6e). The association was robust to additional adjustment with age, sex, *APOE4* dose and postmortem interval (Fig. 6f).

Next, we examined whether the plasma signature could be used to predict future cognitive decline, as with CSF YWHAG:NPTX2. In addition to the ROSMAP, Knight-ADRC and Stanford cohorts, we analyzed the ARIC study, a large independent cohort with SomaScan plasma proteomics and MCI and dementia diagnosis follow-up<sup>42</sup>. For each cohort (total n = 1,877), we used a Cox proportional hazards regression model to test the association between the plasma signature and a future increase in CI stage over 1–15 years among individuals with MCI to mild dementia, adjusting for baseline CI, age, sex and *APOE4* dose. We did not have sufficient data to include plasma or CSF biomarkers of AD pathology. The plasma signature significantly predicted future cognitive decline across all cohorts; in a meta-analysis, an s.d. increase in the plasma signature conferred a 56% increased risk of cognitive decline (meta HR = 1.56, meta  $P = 3.16 \times 10^{-19}$ ; Fig. 6g and Supplementary Table 13).

We then examined conversion from cognitively normal to dementia (total n = 4,753) and found that an s.d. increase in the plasma signature conferred an 86% increased risk, adjusting for age, sex and *APOE4* dose (meta HR = 1.86, meta  $P = 9.97 \times 10^{-17}$ ; Fig. 6h and Supplementary Table 14). Aggregating data across the ROSMAP, Knight-ADRC and Stanford cohorts (total n = 2,292), the plasma signature was among the strongest predictors among covariates (HR = 1.59,  $P = 1.00 \times 10^{-24}$ ; Fig. 6i), with age, baseline CI and *APOE4* dose also having significant effects. Binary high and low groups based on the upper and lower 25th percentiles of the aggregated sample, as done with YWHAG:NPTX2, revealed that plasma signature<sub>high</sub> individuals had a seven-time increased risk of future cognitive decline compared to plasma signature<sub>low</sub> individuals (HR = 7.17,  $P = 2.12 \times 10^{-64}$ ; Fig. 6j), with no additional covariate adjustment.

Together, these data show that plasma proteomics combined with machine learning can be used to derive a plasma-based protein signature that correlates with AD dementia independently of A $\beta$  and tau and partly recapitulates CSF YWHAG:NPTX2.

## Discussion

Overall, our findings reveal that synapse proteins in the CSF and plasma are among the strongest A $\beta$ -independent and tau-independent correlates of CI in AD, and that from these synapse proteins emerges the

CSF YWHAG:NPTX2 ratio, a sparse and robust correlate of CI. We found that YWHAG:NPTX2 increases with cognitively normal aging starting early in life and predicts AD onset and progression in both sporadic AD and ADAD across six independent deeply phenotyped AD cohorts. Most notably, we established YWHAG:NPTX2 thresholds that define five groups which predict future cognitive resilience versus decline among early AD A+ $T_1$ + individuals of the same cognitive diagnosis.

Although these findings indicate that CSF YWHAG:NPTX2 represents a biological process in the brain that is central to cognitive function in AD, what that process is remains unclear. Our data suggest it reflects a 'pathology' distinct from existing neurodegeneration and synaptic AD biomarkers NfL, GAP-43 and Ng. Based on the literature, we speculate that it relates to synapse dysfunction and neuronal hyperactivity-induced synapse loss caused by reduced expression of NPTX2. In human brains, the NPTX2 mRNA and protein are downregulated in AD neurons based on single-cell RNA sequencing, immunohistochemistry and bulk proteomics<sup>28,29</sup>, suggesting that its decrease with CI in AD CSF may reflect decreased expression. Furthermore, genetic loss of NPTX2 and NPTXR causes major GluA4 loss<sup>27</sup>, increased network hyperactivity<sup>27</sup> and increased complement-mediated microglial engulfment of synapses<sup>30</sup>. While NPTX2 has been studied exclusively in neurons, it is worth noting that the gene is also highly expressed in the oligodendrocyte lineage in humans<sup>43</sup>.

The role of YWHAG in the brain is less understood, but the YWHA family of proteins localize in neuron bodies and synapses<sup>44</sup>, and YWHAG mutations cause childhood epilepsy<sup>45</sup>, implying regulation of neuronal activity. Furthermore, YWHA proteins physically regulate tau aggregation and phosphorylation<sup>46,47</sup>, and YWHA protein concentrations are dramatically increased in Creutzfeldt–Jakob disease CSF<sup>48</sup>, suggesting potential involvement in activity-dependent prion and tau propagation. YWHAG also binds to phosphatidylserine<sup>44</sup>, which is involved in synaptic pruning<sup>49</sup>. Future studies that collect CSF near death and perform molecular measurements from matched postmortem brains combined with mechanistic studies may illuminate what 'pathology' CSF YWHAG:NPTX2 and other CSF biomarkers represent.

In addition to reported roles of YWHAG and NPTX2 in neuronal hyperactivity and synapse dysfunction, our study shows that CSF YWHAG:NPTX2 is associated with several aspects of AD including CI, normal aging, genetically driven AB overproduction (ADAD) and tau accumulation, which together strongly implicate its relevance to synapse dysfunction. To elaborate, as with YWHAG:NPTX2, synapse loss is the most robust histological correlate of CI, beyond AB and tau<sup>50</sup>. Second, synapse dysfunction, morphological alterations and loss, rather than overt neuron loss, are major hallmarks of mammalian brain aging that are closely linked with age-related cognitive decline in nonhuman primates<sup>51</sup>. Third, Aβ oligomers cause synapse loss and neuronal hyperactivity<sup>50</sup>, akin to how ADAD mutations-which presumably lead to A $\beta$  overproduction-are associated with a faster increase in YWHAG:NPTX2 with age. Lastly, neuronal hyperactivity enhances tau propagation<sup>40</sup>, which aligns with the positive association between YWHAG:NPTX2 and future A $\beta$ -driven tau PET.

Together, these data suggest that CSF YWHAG:NPTX2 is probably a measure of synapse dysfunction perhaps related to loss of NPTX2-driven hyperactivity and YWHAG-driven tau toxicity, and point to synapse dysfunction as a promising therapeutic target to promote cognitive resilience in the presence of A $\beta$  and tau. Strategies to restore NPTX2 expression to youthful levels may be especially promising, as overexpression of NPTX2 in tau P301S mice protects synapses from complement-mediated microglial engulfment<sup>30</sup>. NPTX2 is also reduced in frontotemporal dementia (FTD) and dementia with Lewy bodies CSF<sup>52</sup>, suggesting that its restoration may have benefits across multiple neurodegenerative diseases. Future studies are needed to determine whether CSF YWHAG:NPTX2 is correlated with CI independent of age and disease-specific pathologies across non-AD dementias, including FTD, dementia with Lewy bodies and amyotrophic lateral sclerosis.

Beyond biological and therapeutic implications, we show comprehensive evidence that CSF YWHAG:NPTX2 provides major prognostic utility in early AD well beyond established A/T/N biomarkers. We demonstrate it can be used to stratify both  $A+T_1+$  cognitively normal patients and A+T<sub>1</sub>+ patients with MCI into five low-risk and high-risk groups based on the thresholds provided in Methods. This stratification may aid in clinical trial target patient selection (that is, by selecting high-risk patients) and drug efficacy evaluation (that is, treatment response in different groups). The lack of cohort effects is especially advantageous for biomarker applications because no batch correction is needed, at least with the SomaScan assay. It is important to further validate our SomaScan YWHAG:NPTX2 group thresholds in independent cohorts, evaluate YWHAG:NPTX2 in comparison to tau PET in predicting future cognitive outcomes and examine changes in YWHAG:NPTX2 longitudinally in both the normal aging population and the population with early AD. Furthermore, developing affordable YWHAG:NPTX2 assays will be essential to overcome the high cost of the SomaScan assay and enable widespread clinical use.

Lastly, we show the development of a plasma proteomic signature of CI that partly recapitulates the characteristics of CSF YWHAG:NPTX2. Notably, the highest weighted proteins in the plasma signature are synapse proteins previously implicated as brain-specific proteins linked to brain aging<sup>53</sup>. Although more efforts are needed to improve the plasma signature, we expect that future advances in proteomics and machine learning will lead to sparse, scalable plasma surrogates of CSF YWHAG:NPTX2 to be used broadly for patient monitoring, clinical trials and research.

## **Online content**

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-025-03565-2.

## References

- 1. Knopman, D. S. et al. Alzheimer disease. *Nat. Rev. Dis. Primers* 7, 33 (2021).
- 2. Dujardin, S. et al. Tau molecular diversity contributes to clinical heterogeneity in Alzheimer's disease. *Nat. Med.* **26**, 1256–1263 (2020).
- 3. Ossenkoppele, R. et al. Amyloid and tau PET-positive cognitively unimpaired individuals are at high risk for future cognitive decline. *Nat. Med.* **28**, 2381–2387 (2022).
- 4. Strikwerda-Brown, C. et al. Association of elevated amyloid and tau positron emission tomography signal with near-term development of Alzheimer disease symptoms in older adults without cognitive impairment. *JAMA Neurol.* **79**, 975–985 (2022).
- Zetterberg, H. & Bendlin, B. B. Biomarkers for Alzheimer's disease—preparing for a new era of disease-modifying therapies. *Mol. Psychiatry.* 26, 296–308 (2021).
- 6. van Dyck Christopher, H. et al. Lecanemab in early Alzheimer's disease. *N. Engl. J. Med.* **388**, 9–21 (2023).
- 7. Hansson, O. Biomarkers for neurodegenerative diseases. *Nat. Med.* **27**, 954–963 (2021).
- 8. Mostafavi, S. et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci.* **21**, 811–819 (2018).
- 9. Hanseeuw, B. J. et al. Association of amyloid and tau with cognition in preclinical Alzheimer disease: a longitudinal study. *JAMA Neurol.* **76**, 915–924 (2019).
- Boyle, P. A. et al. To what degree is late life cognitive decline driven by age-related neuropathologies? *Brain* 144, 2166–2175 (2021).

- 11. Tosun, D. et al. Contribution of Alzheimer's biomarkers and risk factors to cognitive impairment and decline across the Alzheimer's disease continuum. *Alzheimers Dement.* **18**, 1370–1382 (2022).
- 12. Andersen, S. L. Centenarians as models of resistance and resilience to Alzheimer's disease and related dementias. *Adv. Geriatr. Med. Res.* **2**, e200018 (2020).
- Zhang, M. et al. The correlation between neuropathology levels and cognitive performance in centenarians. *Alzheimers Dement*. 19, 5036–5047 (2023).
- Jack, C. R. Jr et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimers Dement.* 20, 5143–5169 (2024).
- 15. Martínez-Dubarbie, F. et al. Accuracy of plasma  $A\beta_{40}$ ,  $A\beta_{42}$ , and p-tau<sub>181</sub> to detect CSF Alzheimer's pathological changes in cognitively unimpaired subjects using the Lumipulse automated platform. *Alzheimers Res. Ther.* **15**, 163 (2023).
- Motta, C. et al. Different associations between amyloid-βeta 42, amyloid-βeta 40, and amyloid-βeta 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).
- Horie, K. et al. CSF MTBR-tau243 is a specific biomarker of tau tangle pathology in Alzheimer's disease. *Nat. Med.* 29, 1954–1963 (2023).
- van der Flier, W. M. & Scheltens, P. The ATN Framework—moving preclinical Alzheimer disease to clinical relevance. *JAMA Neurol.* 79, 968–970 (2022).
- 19. Koopmans, F. et al. SynGO: an evidence-based, expert-curated knowledge base for the synapse. *Neuron* **103**, 217–234 (2019).
- Cruchaga, C. et al. Multi-cohort cerebrospinal fluid proteomics identifies robust molecular signatures across the Alzheimer disease continuum. *Neuron* (in the press).
- 21. Stallings, N. R. et al. Pin1 mediates  $A\beta_{42}$ -induced dendritic spine loss. Sci. Signal. **11**, eaap8734 (2018).
- 22. Yin, Y. et al. Tau accumulation induces synaptic impairment and memory deficit by calcineurin-mediated inactivation of nuclear CaMKIV/CREB signaling. *Proc. Natl Acad. Sci. USA* **113**, E3773–E3781 (2016).
- Stallings, N. R., O'Neal, M. A., Hu, J., Shen, Z.-J. & Malter, J. S. Long-term normalization of calcineurin activity in model mice rescues Pin1 and attenuates Alzheimer's phenotypes without blocking peripheral T cell IL-2 response. *Alzheimers Res. Ther.* 15, 179 (2023).
- 24. Drummond, E. et al. The amyloid plaque proteome in early onset Alzheimer's disease and Down syndrome. *Acta Neuropathol. Commun.* **10**, 53 (2022).
- Johnson, E. C. B. et al. Cerebrospinal fluid proteomics define the natural history of autosomal dominant Alzheimer's disease. *Nat. Med.* 29, 1979–1988 (2023).
- Chang, M. C. et al. Narp regulates homeostatic scaling of excitatory synapses on parvalbumin-expressing interneurons. *Nat. Neurosci.* 13, 1090–1097 (2010).
- 27. Pelkey, K. A. et al. Pentraxins coordinate excitatory synapse maturation and circuit integration of parvalbumin interneurons. *Neuron* **85**, 1257–1272 (2015).
- 28. Mathys, H. et al. Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365–4385 (2023).
- 29. Hurst, C. et al. Integrated proteomics to understand the role of neuritin (NRN1) as a mediator of cognitive resilience to Alzheimer's disease. *Mol. Cell. Proteomics* **22**, 100542 (2023).
- Zhou, J. et al. The neuronal pentraxin Nptx2 regulates complement activity and restrains microglia-mediated synapse loss in neurodegeneration. Sci. Transl. Med. 15, eadf0141 (2023).

- 31. Sathe, G. et al. Quantitative proteomic profiling of cerebrospinal fluid to identify candidate biomarkers for Alzheimer's disease. *Proteomics Clin. Appl.* **13**, 1800105 (2019).
- 32. Nilsson, J. et al. Cerebrospinal fluid biomarker panel for synaptic dysfunction in a broad spectrum of neurodegenerative diseases. *Brain* **147**, 2414–2427 (2024).
- Timsina, J. et al. Comparative analysis of Alzheimer's disease cerebrospinal fluid biomarkers measurement by multiplex SOMAscan platform and immunoassay-based approach. J. Alzheimers Dis. 89, 193–207 (2022).
- 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).
- 35. Shen, Y. et al. CSF proteomics identifies early changes in autosomal dominant Alzheimer's disease. *Cell* **187**, 6309–6326 (2024).
- 36. Ryman, D. C. et al. Symptom onset in autosomal dominant Alzheimer disease. *Neurology* **83**, 253–260 (2014).
- Mitchell, A. J. & Shiri-Feshki, M. Rate of progression of mild cognitive impairment to dementia—meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr. Scand.* **119**, 252–265 (2009).
- 38. Jia, J. et al. Biomarker changes during 20 years preceding Alzheimer's disease. *N. Engl. J. Med.* **390**, 712–722 (2024).
- Pooler, A. M. et al. Amyloid accelerates tau propagation and toxicity in a model of early Alzheimer's disease. Acta Neuropathol. Commun. 3, 14 (2015).
- 40. Wu, J. W. et al. Neuronal activity enhances tau propagation and tau pathology in vivo. *Nat. Neurosci.* **19**, 1085–1092 (2016).
- 41. Franzmeier, N. et al. Elevated CSF GAP-43 is associated with accelerated tau accumulation and spread in Alzheimer's disease. *Nat. Commun.* **15**, 202 (2024).
- 42. Walker, K. A. et al. Large-scale plasma proteomic analysis identifies proteins and pathways associated with dementia risk. *Nat. Aging* **1**, 473–489 (2021).
- 43. Gabitto, M. I. et al. Integrated multimodal cell atlas of Alzheimer's disease. *Nat. Neurosci.* **27**, 2366–2383 (2024).
- Zhang, J. & Zhou, Y. 14-3-3 proteins in glutamatergic synapses. Neural Plast. 2018, 8407609 (2018).
- 45. Ye, X.-G. et al. YWHAG mutations cause childhood myoclonic epilepsy and febrile seizures: molecular sub-regional effect and mechanism. *Front. Genet.* **12**, 632466 (2021).
- 46. Neves, J. F. et al. Phosphorylated full-length Tau interacts with 14-3-3 proteins via two short phosphorylated sequences, each occupying a binding groove of 14-3-3 dimer. *FEBS J.* **288**, 1918–1934 (2021).
- 47. Hochmair, J. et al. 14-3-3 binding regulates Tau assembly and microtubule association. Preprint at *bioRxiv* https://doi.org/10.1101/2024.03.15.585148 (2024).
- Hsich, G., Kenney, K., Gibbs, C. J., Lee, K. H. & Harrington, M. G. The 14-3-3 brain protein in cerebrospinal fluid as a marker for transmissible spongiform encephalopathies. *N. Engl. J. Med.* 335, 924–930 (1996).
- Scott-Hewitt, N. et al. Local externalization of phosphatidylserine mediates developmental synaptic pruning by microglia. *EMBO J.* **39**, e105380 (2020).
- 50. Colom-Cadena, M. et al. The clinical promise of biomarkers of synapse damage or loss in Alzheimer's disease. *Alzheimers Res. Ther.* **12**, 21 (2020).
- 51. Morrison, J. H. & Baxter, M. G. The ageing cortical synapse: hallmarks and implications for cognitive decline. *Nat. Rev. Neurosci.* **13**, 240–250 (2012).
- 52. Nilsson, J. et al. Cerebrospinal fluid biomarker panel of synaptic dysfunction in Alzheimer's disease and other neurodegenerative disorders. *Alzheimers Dement*. **19**, 1775–1784 (2023).

53. Oh, H. S. et al. Organ aging signatures in the plasma proteome track health and disease. *Nature* **624**, 164–172 (2023).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/ by-nc-nd/4.0/.

© The Author(s) 2025

Hamilton Se-Hwee Oh  $0^{1,2,3}$ , Deniz Yagmur Urey<sup>2,3</sup>, Linda Karlsson  $0^4$ , Zeyu Zhu<sup>5</sup>, Yuanyuan Shen<sup>6,7</sup>, Amelia Farinas<sup>2,3,8</sup>, Jigyasha Timsina<sup>6,7</sup>, Michael R. Duggan  $0^9$ , Jingsha Chen<sup>10</sup>, Ian H. Guldner<sup>11</sup>, Nader Morshed  $0^{12,13}$ , Chengran Yang  $0^{6,7}$ , Daniel Western  $0^{6,7}$ , Muhammad Ali  $0^{6,7}$ , Yann Le Guen  $0^{11,14}$ , Alexandra Trelle<sup>9</sup>, Sanna-Kaisa Herukka<sup>15</sup>, Tuomas Rauramaa<sup>16</sup>, Mikko Hiltunen<sup>17</sup>, Anssi Lipponen  $0^{17}$ , Antti J. Luikku<sup>18</sup>, Kathleen L. Poston  $0^{2,3,11}$ , Elizabeth Mormino<sup>11</sup>, Anthony D. Wagner  $0^{3,19}$ , Edward N. Wilson  $0^{2,3,11}$ , Divya Channappa<sup>2,3,11</sup>, Ville Leinonen<sup>18</sup>, Beth Stevens<sup>12,13,20</sup>, Alexander J. Ehrenberg<sup>21,22,23</sup>, Rebecca F. Gottesman  $0^{24}$ , Josef Coresh  $0^{25}$ , Keenan A. Walker  $0^9$ , Henrik Zetterberg  $0^{26,27,28,29,30,31}$ , David A. Bennett<sup>32</sup>, Nicolai Franzmeier  $0^{5,33,34}$ , Oskar Hansson  $0^{4,35}$ , Carlos Cruchaga  $0^{6,7}$  & Tony Wyss-Coray  $0^{2,3,11}$ 

<sup>1</sup>Graduate Program in Stem Cell and Regenerative Medicine, Stanford University, Stanford, CA, USA. <sup>2</sup>The Phil and Penny Knight Initiative for Brain Resilience, Stanford University, Stanford, CA, USA. <sup>3</sup>Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA. <sup>4</sup>Clinical Memory Research Unit, Department of Clinical Sciences in Malmö, Lund University, Lund, Sweden. 5Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, Munich, Germany. 6Department of Psychiatry, Washington University, St. Louis, MO, USA. 7NeuroGenomics and Informatics Center, Washington University, St. Louis, MO, USA. 8 Graduate Program in Neuroscience, Stanford University, Stanford, CA, USA. 9 Laboratory of Behavioral Neuroscience, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA. 10 Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, USA. "Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA. <sup>12</sup>Boston Children's Hospital, F.M. Kirby Neurobiology Center, Boston, MA, USA. <sup>13</sup>Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>14</sup>Quantitative Sciences Unit, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. 15 Department of Neurology, Kuopio University Hospital and Institute of Clinical Medicine – Neurology, University of Eastern Finland, Kuopio, Finland. 16 Department of Pathology, Kuopio University Hospital and Institute of Clinical Medicine – Pathology, University of Eastern Finland, Kuopio, Finland. <sup>17</sup>Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland. <sup>18</sup>Department of Neurosurgery, Kuopio University Hospital and Institute of Clinical Medicine – Neurosurgery, University of Eastern Finland, Kuopio, Finland. 19 Department of Psychology & Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA, <sup>20</sup>Howard Hughes Medical Institute, Boston, MA, USA, <sup>21</sup>Memory and Aging Center, Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA.<sup>22</sup>Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA, USA. 23 Department of Neuroscience, University of California, Berkeley, Berkeley, CA, USA. 24 Stroke Branch, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA.<sup>25</sup>Departments of Population Health and Medicine, New York University Grossman School of Medicine, New York, NY, USA. <sup>26</sup>Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden.<sup>27</sup>Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden.<sup>28</sup>Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK.<sup>29</sup>UK Dementia Research Institute at UCL, London, UK.<sup>30</sup>Hong Kong Center for Neurodegenerative Diseases, Hong Kong, China. <sup>31</sup>Wisconsin Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI, USA. 32 Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. <sup>33</sup>Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. <sup>34</sup>Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, University of Gothenburg, The Sahlgrenska Academy, Gothenburg, Sweden.<sup>35</sup>Memory Clinic, Skåne University Hospital, Malmö, Sweden. e-mail: hamilton.oh@mssm.edu; twc@stanford.edu

## Article

# Methods

#### Participants

**Stanford cohorts.** Plasma and CSF collection, processing and storage for all Stanford cohorts followed a single standard procedure. All studies were approved by the Institutional Review Board (IRB) of Stanford University and written informed consent or assent was obtained from all participants or their legally authorized representatives.

Blood collection and processing followed a rigorous standardized protocol to minimize variation. Briefly, about 10 ml of whole blood was collected in four vacutainer EDTA tubes (Becton Dickinson) and spun at 1,800g for 10 min to separate out plasma, leaving 1 cm of plasma above the buffy coat to avoid contamination. Plasma was aliquoted into polypropylene tubes and stored at -80 °C. Processing took approximately 1 h from draw to freezing and storage. All draws occurred in the morning to minimize circadian effects.

CSF was collected via lumbar puncture using a 20-22 G spinal needle that was inserted in the L4–L5 or L5–S1 interspace. CSF samples were immediately centrifuged at 500g for 10 min, aliquoted in polypropylene tubes and stored at -80 °C.

Plasma from all Stanford cohorts was sent to SomaLogic for proteomics (v.4.1 SomaScan, -7,000 proteins) in the same batch. CSF samples from all Stanford cohorts were sent to SomaLogic for proteomics (v.4.0 SomaScan, -5,000 proteins) in the same batch. The core CSF AD biomarkers  $A\beta_{42}$ ,  $A\beta_{40}$  and  $pTau_{181}$  were measured using the fully automated LUMIPULSE G1200 instrument (Fujirebio) as described previously<sup>54,55</sup>. Descriptions for each cohort are provided below.

A total of 1,160 plasma samples (738 participants, longitudinal sampling) and 371 CSF samples (371 participants, one sample from each) from the Stanford cohorts were included in this study. Per-cohort sample sizes were as follows: ADRC plasma n = 827 (423 participants), CSF n = 113; SAMS plasma n = 222 (215 participants), CSF n = 169; Biomarkers in Parkinson's Disease (BPD) plasma n = 55 (55 participants), CSF n = 68; Stanford Center for Memory Disorders (SCMD) cohort study plasma n = 45 (45 participants), CSF n = 21.

Stanford-ADRC study. Samples were acquired through the NIA-funded Stanford-ADRC, a longitudinal observational study of individuals with clinical dementia and age-matched and sex-matched individuals without dementia. Healthy controls were deemed cognitively unimpaired during a clinical consensus conference that included board-certified neurologists and neuropsychologists. Cognitively impaired individuals underwent CDR and standardized neurological and neuropsychological assessments, including the procedures of the National Alzheimer's Coordinating Center (https://naccdata.org/), to determine cognitive and diagnostic status. Cognitive status was determined during a clinical consensus conference that included neurologists and neuropsychologists. All participants were free from acute infectious diseases and in good physical condition.

**SAMS study.** SAMS is an ongoing longitudinal study of healthy aging. Blood and CSF collection and processing, and neurological and neuropsychological assessments, were performed by the same team and followed the same protocol as in the Stanford-ADRC cohort. All SAMS participants had a CDR = 0 and a neuropsychological test score in the normal range; all SAMS participants were deemed cognitively unimpaired during a clinical consensus conference that included neurologists and neuropsychologists.

**Stanford BPD cohort.** The BPD cohort<sup>56</sup> was a Michael J. Fox Foundation for Parkinson's Research-funded longitudinal study of biomarkers associated with cognitive decline in Parkinson's disease (PD). Research participants were recruited from the Stanford Movement Disorders Center between 2011 and 2015, with a PD diagnosis according to the UK Brain Bank criteria; they required bradykinesia with muscle rigidity or rest tremor. All participants completed baseline cognitive, motor, neuropsychological, imaging and biomarker assessments (plasma and optional CSF), including the Movement Disorders Society-revised Unified Parkinson's Disease Rating Scale. Age-matched healthy controls were also recruited to control for age-associated biomarker changes. After a comprehensive neuropsychological battery, all participants were given a cognitive diagnosis of no CI, MCI or dementia, according to published criteria.

**SCMD study.** The SCMD was an NIA-funded cross-sectional study of people across the cognitive continuum. Participants with mild AD dementia and amnestic MCI were recruited from the Stanford Center for Memory Disorders between 2011 and 2015. Participants were included if they had a diagnosis of probable AD dementia (amnestic presentation) according to the NIA-Alzheimer's Association<sup>57</sup> criteria and a CDR score of 0.5 or 1, or a diagnosis of MCI according to the NIA-Alzheimer's Association criteria<sup>57</sup>, a score of 1.5 s.d. below age-adjusted normative means on at least one test of episodic memory and a CDR score of less than 1. Older healthy controls were recruited from the community, were selected to have a similar average age as the enrolled patients and were required to have normal neuropsychological performance and a CDR of 0. Participants completed cognitive, neuropsychological, imaging and biomarker assessments with plasma.

Knight-ADRC study. The Knight-ADRC is an NIA-funded longitudinal observational study of individuals with clinical dementia and age-matched controls. Research participants undergo longitudinal cognitive, neuropsychological, imaging and biomarker assessments including a CDR. Cases with AD corresponded to those with a diagnosis of dementia of the Alzheimer's type using criteria equivalent to the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS)-Alzheimer's Disease and Related Disorders Association (ADRDA) for probable AD; AD severity was determined using the CDR at the time of the lumbar puncture (for the CSF samples) or blood draw (for the plasma samples). Controls received the same assessment as cases but did not have dementia (CDR = 0).

Blood samples were collected in EDTA tubes (vacutainer purple top, Becton Dickinson) at the time of the visit, immediately centrifuged at 1,500g for 10 min, aliquoted on two-dimensional barcoded Micronic tubes (200  $\mu$ l per aliquot) and stored at -80 °C. Plasma was stored in a monitored -80 °C freezer until it was pulled and sent to SomaLogic (SomaScan 7k) for data generation. Proteomics data from 2,112 plasma samples from each of the 2,122 participants were included in this study.

CSF samples were collected through lumbar puncture from participants after an overnight fast. Samples were processed and stored at -80 °C until they were sent for protein measurement. Proteomics data from 927 CSF samples from each of 927 participants were included in this study. CSF samples from the Knight-ADRC, ADNI and DIAN cohorts were sent for proteomics using the SomaScan platform (SomaScan 7k) in the same batch. CSF A $\beta_{42}$ , A $\beta_{40}$  and pTau<sub>181</sub> were measured using the LUMIPULSE G1200 immunoassay platform according to the manufacturer's specifications.

The IRB of Washington University School of Medicine in St. Louis approved the study and research was performed in accordance with the approved protocols.

**ADNI study.** ADNI is a longitudinal multicenter study designed to develop early biomarkers of AD. All data used in this study were accessed from the ADNI database (https://adni.loni.usc.edu/). ADNI investigators assessed and diagnosed individuals as either cognitively normal (Mini Mental State Examination (MMSE)  $\geq$  24, CDR = 0, nondepressed), MC (MMSE  $\geq$  24, CDR = 0.5, objective memory impairment on education-adjusted Wechsler Memory Scale-II, preserved activities of daily living) or with dementia (MMSE = 20–26, CDR > 0.5, NINCDS/ADRDA criteria for probable AD). Comprehensive details on study design, data acquisition, ethics and policies are described above.

CSF A $\beta_{42}$  and pTau<sub>181</sub> were measured using the xMAP immunoassay platform according to the manufacturer's specifications. Proteomics data from 725 CSF samples from each of 725 participants were included in this study.

**DIAN study.** DIAN, led by Washington University School of Medicine in St. Louis, is a family-based long-term observational study designed to understand the earliest changes of ADAD. All participants undergo clinical and cognitive batteries (that is, global CDR). Comprehensive details on study design, data acquisition, ethics and policies can be found at https://dian.wustl.edu/. The data used in this study are from data freeze 15. CSF A $\beta_{42}$  and pTau<sub>181</sub> were measured using the LUMI-PULSE G1200 platform according to the manufacturer's specifications. Proteomics data from 455 CSF samples from each of 455 participants were included in this study.

**BioFINDER2 study.** BioFINDER2 is a Swedish prospective cohort study (NCT03174938) on age-related neurodegenerative diseases. AD was diagnosed based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition and A $\beta$  positivity (CSF A $\beta_{42}$ :A $\beta_{40}$ ). Proteomics data from a total of n = 829 participants, consisting of n = 480 cognitively unimpaired, n = 213 with MCI and n = 136 with AD dementia were included in this study. Global CDR scores were not measured in BioFINDER2, so for estimation, participants with AD were subdivided according to the MMSE thresholds defined in ref. 58: 21–25 for mild dementia (CDR = 1); 11–20 for moderate dementia (CDR = 2); and 0–10 for severe dementia (CDR = 3). Participants were recruited at Skåne University Hospital and Ängelholm Hospital. The study was approved by the Regional Ethical Committee in Lund, Sweden; all participants gave written informed consent.

CSF samples were collected close in time after baseline clinical examination and handled according to established preanalytical protocols, previously described in detail in ref. 59. All analyses were performed by technicians blinded to all clinical and imaging data. CSF pTau<sub>181</sub>,  $A\beta_{42}$  and  $A\beta_{40}$  were measured using Elecsys assays in accordance with the manufacturer's instructions (Roche Diagnostics). CSF  $A\beta_{42}/A\beta_{40}$  was used to define  $A\beta$  positivity according to a previously established cutoff of <0.08 (ref. 60). CSF samples from the BioFINDER2 cohort were analyzed with liquid chromatography–tandem MS (LC–MS/MS), previously described in detail in ref. 32.

Tau PET was performed using [<sup>18</sup>F]RO948. SUVR images were created for the 70–90-min postinjection interval using the inferior cerebellar cortex as the reference region. A composite corresponding to a Braak I–IV meta-region of interest was used to represent AD-related tau tangle pathology.

**Kuopio University Hospital.** The Kuopio Normal Pressure Hydrocephalus (NPH) and AD Registry and Tissue Bank includes patients from the Eastern Finnish population who were referred to the Kuopio University Hospital neurosurgical unit for suspected NPH. The registry's inclusive criteria encompass a wide range of hydrocephalic conditions and comorbidities: patients must exhibit 1–3 symptoms potentially associated with NPH (such as impaired gait, cognition or urinary continence) along with enlarged brain ventricles (Evans index > 0.3) as seen on computed tomography or magnetic resonance imaging, and no other clear cause that alone explains the observed findings and symptoms. Preoperative comorbidities and conditions were recorded at baseline; patients underwent a systematic differential diagnostic workup followed by a CSF tap test paired with gait evaluation.

A diagnostic right frontal cortical brain biopsy was taken from all patients during shunt surgery from the insertion site of the intraventricular catheter. A neuropathologist (T.R.) analyzed immunoreactivity for A $\beta$  and hyperphosphorylated tau with light microscopy; the results were graded as present or absent. Patients were then further classified into groups according to the presence of the pathology of A $\beta$  or HPT observed in the frontal cortical biopsies<sup>61</sup>. Follow-up was conducted on all operated patients, with optimal shunt function ensured through valve adjustment, brain imaging, shunt valve tapping, lumbar infusion testing and shunt revision if necessary. Global CDR scores were not measured; the Consortium to Establish a Registry for Alzheimer's Disease cognitive score<sup>62</sup>, including the MMSE, was used instead. The latest follow-up MMSE score less than 26 was considered as at least mild dementia<sup>61</sup>.

Lumbar CSF proteomics was performed using high-throughput tandem mass tag-labeling MS, previously described in detail in ref. 63. Data from 90 individuals with CSF proteomics and cognitive scoring performed within 1 year were included in this study.

The study was conducted according to the 1964 Declaration of Helsinki and its later amendments (2013) and all patients provided written informed consent. The Research Ethics Committee of the Northern Savo Hospital District (decision no. 276/13.02.00/2016) approved the study.

ROSMAP studies. All ROSMAP participants enrolled without known dementia and agreed to detailed clinical evaluation and brain donation at death<sup>64</sup>. Both studies were approved by the IRB of Rush University Medical Center (no. L91020181, MAP IRB no. L86121802). Both studies were conducted according to the principles expressed in the Declaration of Helsinki. Each participant signed an informed consent, Anatomical Gift Act and a Rush Alzheimer's Disease Center (RADC) Repository consent (IRB no. L99032481) allowing their data and biospecimens to be repurposed. All MAP participants and a subset of ROS have blood drawn during an annual home visit. For plasma, blood is drawn in a lavender (purple) top EDTA tube. For out of town ROS sites, they were spun, aliquoted into Nunc vials, stored in dry ice and sent to RADC by FedEx, where they were transferred to -80 °C. Samples collected in northeastern Illinois were brought to the RADC laboratory and processed there using the same procedures. A total of 1,046 55-µl samples were shipped to Stanford, then to SomaLogic for proteomics (SomaScan 7k); 973 samples passed quality control.

Clinical and neuropathological data collection has been reported in detail elsewhere<sup>10,65–68</sup>. Regarding clinical diagnosis, an actuarial decision tree designed to mimic expert clinical judgment was implemented by computer to inform several clinical diagnoses, including dementia and AD. It combined data reduction techniques for cognitive performance testing, with a series of discrete clinical judgments made in series by a neuropsychologist and a clinician. Presumptive diagnoses of dementia and AD were calculated that conformed to accepted clinical criteria. The clinician was asked to agree or disagree with the decisions. An algorithm used these decisions to provide diagnoses of MCI and amnestic MCI. Persons with MCI were judged to have CI by the neuropsychologist, and without a diagnosis of dementia by the clinician. Persons without dementia or MCI were categorized as having no CI.

Global CDR scores were not assessed in ROSMAP. Rather, participants underwent a battery of 21 cognitive performance tests, 17 of which were combined into a measure of global cognition which was *z*-scored based on the mean and s.d. of all participants from baseline<sup>69</sup>. For estimation, participants were subdivided according to global cognition *z*-scores: cognitively normal = *z*-score > 0; MCI = -1 < z-score < 0; mild dementia = -2 < z-score <=-1; moderate dementia = -3 < z-score <=-2; and severe dementia *z*-score < = -3. These cutoffs were set based on the distributions of global cognition *z*-scores per clinical diagnosis. Details on cognitive scores, neuropathology and other patient information are described at https://www.radc.rush.edu/documentation.htm. Proteomics data from 973 plasma samples from each of 890 participants were included in this study.

**GNPC.** The GNPC is a major neurodegenerative disease biomarker discovery effort, which hosts the largest collection of SomaScan data (over 40,000 patient samples from over 20 international research groups)

from patient samples across healthy aging, AD, PD, amyotrophic lateral sclerosis and FTD. All cohorts and data are anonymized. Cohorts that had global CDR cognitive scores were included in this study.

**ARIC study.** ARIC is a prospective epidemiological study conducted in four US communities: Forsyth County, NC; Jackson, MS; the northwest suburbs of Minneapolis, MN; and Washington County, MD. The ARIC study enrolled 15,792 mostly White and Black participants aged 45–64 between 1987 and 1989 (ref. 70). After initial enrollment, participants had four additional in-person visits: visit 2 (1990–1992); visit 3 (1993–1995); visit 4 (1996–1999); and visit 5 (2011–2013). Participants were invited back for visit 6 (2016–2017) after 5 years, and visit 7 (2018–2019) immediately thereafter. Additional follow-up visits are ongoing. Blood was drawn for proteomic analysis at visits 2 and 5. For this paper, the late-life (8-year) dementia risk was assessed between visits 5 and 7. Plasma was collected using standardized protocols and frozen at –80 °C until analysis. Proteins were measured using the SomaScan v.4.0 assay; protein quality control steps have been described in detail previously<sup>42</sup>.

Dementia was adjudicated through cognitive assessment tests, telephone screening, informant ratings, hospital records and death record review, as described previously<sup>71</sup>. From visits 2 through 4, a three-instrument cognitive assessment was applied (delayed word recall task, digit symbol substitution from the Wechsler Adult Intelligence Scale-Revised and a letter fluency task). At visits 5 through 7, a surveillance approach was used, where participants received a comprehensive cognitive exam and a functional assessment that included the Clinical Dementia Rating Scale and Functional Activities Questionnaire. With these data, dementia was classified based on the NIA, Alzheimer's Association and Diagnostic and Statistical Manual of Mental Disorder, Fifth Edition criteria. In the time between visits 5 and 6, participants were contacted annually via phone and administered the Six-item Screener (SIS), a brief cognitive assessment. If participants received a low score on the SIS, or if they were unable to participate in the screening via phone, the Ascertain Dementia 8 (AD8) was administered to the participant's informant. For participants who received a dementia diagnosis at visit 6 or 7, SIS, AD8, hospital discharge and death certificate codes were used to define the date of dementia onset. For participants who did not attend visits 6 or 7, SIS, AD8, hospital discharge and death certificate codes were used to define the dementia diagnosis and date of dementia onset.

The ARIC study protocols were approved by the IRBs at each participating center: University of North Carolina at Chapel Hill; Wake Forest University; Johns Hopkins University; University of Minnesota; and University of Mississippi Medical Center. All ARIC participants gave written informed consent at each study visit; proxies provided consent for participants who were judged to lack capacity.

#### Proteomics

The SomaLogic (https://somalogic.com/) SomaScan assay<sup>72,73</sup>, which uses slow off-rate modified DNA aptamers (SOMAmers) to bind target proteins with high specificity, was used to quantify the relative concentration of thousands of human proteins in plasma and CSF in the Stanford, Knight-ADRC, ADNI, DIAN and ROSMAP cohorts. The v.4.1 (~7,000 proteins) assay was used for all the mentioned cohorts and samples, except for the Stanford CSF, for which the v.4.0 (~5,000 proteins) assay was used. Standard SomaLogic normalization, calibration and quality control were performed on all samples, resulting in protein measurements in relative fluorescence units. Plasma samples were further normalized to a pooled reference using an adaptive maximum likelihood procedure. The resulting values are the data from SomaLogic and are considered 'raw' data. We further performed log<sub>10</sub> normalization as the assay had an expected log-normal distribution. No cohort batch corrections were applied. Acetylcholinesterase was removed before the analyses because it can be confounded with scetylcholinesterase inhibitor treatment. CSF samples from the BioFINDER2 cohort were analyzed with LC-MS/MS, previously described in detail in ref. 32. CSF

samples from the Kuopio cohort were analyzed with high-throughput tandem mass tag-labeling MS, previously described in detail in ref. 63.

#### **CI stage classification**

CI stages reflect global CDR scores. CDR scores of 0, 0.5, 1, 2 and 3 are synonymous with the CI stages of none, MCI, mild dementia, moderate dementia and severe dementia, respectively. The Stanford, Knight-ADRC, ADNI and DIAN cohorts measured global CDR scores. BioFINDER2, ROSMAP and Kuopio did not measure global CDR scores, so we estimated global CDR scores based on cognitive battery tests and clinical diagnoses as described in the sections for each cohort.

#### $A+T_1+$ versus $A-T_1-$ classification

Typically, 'A' positivity is defined by the levels of A $\beta_{42}$  and 'T<sub>1</sub>' positivity by pTau<sub>181</sub>, using a separate Gaussian mixture model for each biomarker to derive value cutoffs<sup>74</sup> (Supplementary Fig. 1a). This leads to four possible groups: A–T<sub>1</sub>–, A+T<sub>1</sub>–, A–T<sub>1</sub>+ and A+T<sub>1</sub>+. However, this classification system does not fit the 'shape' of the data and artificially increases the number of A–T<sub>1</sub>+ individuals<sup>75</sup> (Supplementary Fig. 1a), as the frequency of A–T+ individuals based on PET imaging biomarkers (the gold standard) are extremely rare<sup>75</sup>. To overcome this limitation, we used the CSF pTau<sub>181</sub>:A $\beta_{42}$  ratio, which better fits the shape of the data (Supplementary Fig. 1b), to define A–T<sub>1</sub>– versus A+T<sub>1</sub>+ status (log<sub>10</sub> pTau<sub>181</sub>:A $\beta_{42}$  cutoff = –1, based on Lumipulse or xMAP; different cutoffs were used for Elecsys; Supplementary Fig. 1c). Previous studies showed that pTau<sub>181</sub>:A $\beta_{42}$  appropriately captures A–T<sub>1</sub>– versus A+T<sub>1</sub>+ status<sup>15,16</sup>. A $\beta$  positivity, regardless of T status, was determined using the CSF A $\beta_{42}$ :A $\beta_{40}$  ratio or A $\beta$ PET (gold standards).

#### Statistical analyses

While some cohorts included multiple plasma samples from the same individual (precise numbers are described in the cohort sections), all analyses in this study were performed using proteomics data from only a single time point per individual. Only one CSF sample was collected per individual. For cross-sectional associations with CI, the most recent plasma sample was used to maximize the sample size of cases with dementia, which were fewer than cognitively normal cases. For analyses involving the prediction of future cognitive decline from a cognitively normal or early AD baseline, the earliest plasma sample was used to maximize sample size.

The NumPy<sup>76</sup> and Pandas<sup>77</sup> Python packages were used for data processing and transformation; the Matplotlib<sup>78</sup> and seaborn<sup>79</sup> Python packages were used for plotting.

Linear regression. The stats.pearsonr function from the SciPy<sup>80</sup> Python package was used to assess the Pearson correlations. The ordinary least squares (OLS) function from the statsmodels<sup>81</sup> Python package was used to assess the linear associations between protein levels and CI. For the unbiased proteome-wide association tests in Fig. 1b, we tested the following linear model for each protein:  $CI \sim protein + CSF pTau_{181}$ :  $A\beta_{42}$  + age + sex + APOE4 dose + cohort + PC1. We included the PC1 of the proteome as a covariate because previous studies showed that it represents a large source of non-disease-related variance, potentially related to heterogeneity in CSF production and clearance rates<sup>35,75</sup>. Inclusion of PC1'de-noised' the data and greatly improved the significance of protein associations with Clin every cohort we assessed. Multiple hypothesis testing correction was applied using the Benjamini-Hochberg method, and the significance threshold was set at a 5% false discovery rate (q < 0.05). All other linear regression analyses in the manuscript relied on the same OLS function. Precise covariates used per analysis are displayed in the figures or described in the main text. The proportion of CI variance explained by certain variables was determined using the  $r^2$  values from the OLS models. Bootstrapping (n = 1,000) was applied to derive the 95% confidence intervals and P values for variable (that is, YWHAG:NPTX2 versus pTau<sub>181</sub>:Aβ<sub>42</sub>) comparisons.

**Logistic regression.** The LogisticRegression function from the scikit-learn<sup>82</sup> Python package was used to assess the CI classification based on YWHAG:NPTX2 or pTau<sub>181</sub>:A $\beta_{42}$  (Fig. 1g and Extended Data Fig. 1e). Models were tested to distinguish MCI versus cognitively normal, mild dementia versus MCI, moderate-to-severe versus mild dementia and mild dementia or worse versus cognitively normal. The AUC, accuracy, sensitivity, and specificity were calculated using the confusion\_matrix, accuracy\_score, recall\_score, roc\_curve and roc\_auc functions from scikit-learn<sup>82</sup>. Bootstrapping (n = 1,000) was applied to derive the 95% confidence intervals and P values for YWHAG:NPTX2 versus pTau<sub>181</sub>:A $\beta_{42}$  comparisons.

**Cox proportional hazards regression.** The CoxPHFitter function from the lifelines<sup>83</sup> Python package was used to assess the associations between CSF YWHAG:NPTX2 and future cognitive decline (Figs. 4e-k, 5b,c and 6g-j. An event of cognitive decline was defined as a stage increase in CI (that is, none to MCI, or MCI to mild dementia). An event of conversion from cognitively normal to dementia was defined as a two-stage or more increase in CI from a cognitively normal baseline (none to mild dementia). Additional covariates such as baseline age, sex, *APOE4* dose, CSF pTau<sub>181</sub>:A $\beta_{42}$ , CSF NfL, CSF Ng and CI were included depending on the analysis. The precise covariates used for each analysis are displayed in the figures or in the text. Meta-analyses to compare and aggregate effect sizes and confidence intervals from multiple cohorts were performed in R using the metafor<sup>84</sup> package, with an inverse-variance-weighted fixed effects model.

**Biological pathway enrichment analyses.** gProfiler<sup>85</sup> was used for Gene Ontology (GO) enrichment analyses (Fig. 1c). The GO<sup>86,87</sup> database includes the SynGO<sup>19</sup> database used to subset the synapse proteins; 6,379 unique protein-encoding genes detected by SomaScan were used as background in gProfiler.

**Derivation of CSF YWHAG:NPTX2 ratio.** The LassoCV function from the scikit-learn<sup>\$2</sup> Python package was used to train, in the ADNI cohort, a penalized linear model to predict CI severity based on the levels of 214 synapse proteins that significantly changed with CI in the ADNI and Knight-ADRC cohorts (Fig. 1b). Fivefold cross-validation was implemented to identify the optimal lambda parameter. The RFECV and RFE functions from scikit-learn<sup>\$2</sup> were used to perform RFE on the LassoCV model to further simplify the model to facilitate clinical applications. RFECV showed that two proteins sufficiently captured most of the signal in the model. RFE was used to derive a model with two proteins, which resulted in the normalized ratio between YWHAG and NPTX2. Details on testing the YWHAG:NPTX2 in independent cohorts are provided in the 'Code availability' section of the article.

**Derivation of plasma signature of CI.** Full details are shown in Supplementary Methods. Briefly, the LassoCV function from the scikit-learn<sup>82</sup> Python package was used to train, in the Knight-ADRC and ROSMAP cohorts, a penalized linear model to predict CI severity based on the levels of 745 plasma proteins. Fivefold cross-validation was implemented to identify the optimal lambda parameter. We call this model the 'plasma signature' throughout the paper. Model weights and intercept are provided in Supplementary Table 7 and details on testing the signature in independent cohorts are provided in the 'Code availability' section of the article.

## **Reporting summary**

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All data are available upon reasonable request with formal applications submitted to the respective cohort committees to protect patient sensitive data. Stanford data (including the Stanford-ADRC, SAMS, BPD

and SCMD) can be requested at https://web.stanford.edu/group/adrc/ cgi-bin/web-proj/datareg.php. Data from specific Stanford cohorts can be requested from the following cohort leaders: ADRC, T.W.-C. (twc@ stanford.edu); SAMS, E.M. (bmormino@stanford.edu) or A.D.W. (awagner@stanford.edu); and BPD and SCMD, K.L.P. (klposton@stanford. edu). The Knight-ADRC proteomics data were generated by the laboratory of the principal investigator C.C. (cruchagac@wustl.edu) and are available upon reasonable request at https://knightadrc.wustl.edu/ professionals-clinicians/request-center-resources/submit-a-request/. The ADNI data can be requested at the ADNI database (https://adni. loni.usc.edu/). The DIAN data can be requested at https://dian.wustl. edu/our-research/for-investigators/diantu-investigator-resources/ dian-tu-biospecimen-request-form/. Pseudonymized BioFINDER2 data can be shared with qualified academic researchers after request to the cohort leader O.H. (oskar.hansson@med.lu.se) for the purpose of replicating the procedures and results presented in the study. Data transfer must be performed in agreement with EU legislation regarding general data protection regulation and decisions by the Ethical Review Board of Sweden and Region Skåne. The Kuopio data can be requested and accessed via a repository on Terra https://app. terra.bio/#workspaces/marsh-terra-inph/iNPH\_Proteomics\_Workspace. The ROSMAP data can be requested at the RADC Research Resource Sharing Hub (https://www.radc.rush.edu) and https://www. synapse.org/Synapse:syn64957327. The ARIC proteomics data are available through the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (https://biolincc.nhlbi.nih.gov/ studies/aric/). Additional ARIC requests for clinical or proteomic data from individual investigators may be submitted to the ARIC steering committees and will be reviewed to ensure that data can be shared without compromising patient confidentiality or breaching intellectual property restrictions. Participant-level demographic, clinical and proteomic data may be partially restricted based on previously obtained participant consent. Data sharing restrictions may also be applied to ensure consistency with confidentiality or privacy laws and considerations (https://sites.cscc.unc.edu/aric/).g:Profiler can be accessed at https://biit.cs.ut.ee/gprofiler/gost (ref. 85). The Gene Ontology<sup>86,87</sup> database can be accessed at https://geneontology.org/ and the SynGO<sup>19</sup> database can be accessed at https://www.syngoportal.org/.

# **Code availability**

The CSF YWHAG.1:NPTX2 ratio can be derived by log<sub>10</sub>-normalizing the SomaScan (assay v.4.1) protein levels (YWHAG.1 SeqId = 4179-57, NPTX2 SeqId=6521-35), then by z-score normalization using the means and s.d. from our cohorts (YWHAG.1 mean = 3.425, s.d. = 0.183; NPTX2 mean = 4.099, s.d. = 0.171), then taking the difference between normalized YWHAG.1 and NPTX2 values. The thresholds for the YWHAG.1:NPTX2 subgroups are '-1 group' < -1.269 < '0 group' < 0.53 < '1 group' < 1.208 < '2 group' < 2.795 < '3 group'. Importantly, CSF data from other versions of the SomaScan assay (that is, v.4 and v.5) need to be normalized to the v.4.1 assay space if our means, s.d. values and thresholds are to be applied. While scale factors for cross-assay version normalization of plasma data can be requested from SomaLogic, scale factors for CSF data have not been prioritized by SomaLogic at the time of this publication. CSF scale factors can be derived in house by running samples on both the v.4.1 assay and other SomaScan assay versions. CSF scale factors may be derived by SomaLogic in the future. The plasma signature of CI can be derived from independent plasma SomaScan data (assay v.4 and later) by following a tutorial on the synapsesignature GitHub repository (https://github.com/ hamiltonoh/synapsesignature), which uses the protein weights provided in Supplementary Table 7. See Supplementary Methods for details.

## References

54. Wilson, E. N. et al. Performance of a fully-automated Lumipulse plasma phospho-tau181 assay for Alzheimer's disease. *Alzheimers Res. Ther.* **14**, 172 (2022).

- 55. Wilson, E. N. et al. Soluble TREM2 is elevated in Parkinson's disease subgroups with increased CSF tau. *Brain* **143**, 932–943 (2020).
- 56. Plastini, M. J. et al. Multiple biomarkers improve diagnostic accuracy across Lewy body and Alzheimer's disease spectra. *Ann. Clin. Transl. Neurol.* **11**, 1197–1210 (2024).
- 57. McKhann, G. et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939–944 (1984).
- Perneczky, R. et al. Mapping scores onto stages: mini-mental state examination and clinical dementia rating. *Am. J. Geriatr. Psychiatry* 14, 139–144 (2006).
- 59. Palmqvist, S. et al. Discriminative accuracy of plasma phospho-tau217 for Alzheimer disease vs other neurodegenerative disorders. *JAMA* **324**, 772–781 (2020).
- 60. Pichet Binette, A. et al. Amyloid-associated increases in soluble tau relate to tau aggregation rates and cognitive decline in early Alzheimer's disease. *Nat. Commun.* **13**, 6635 (2022).
- 61. Koivisto, A. M. et al. High risk of dementia in ventricular enlargement with normal pressure hydrocephalus related symptoms. *J. Alzheimers Dis.* **52**, 497–507 (2016).
- 62. Nerg, O. et al. The CERAD neuropsychological battery in patients with idiopathic normal pressure hydrocephalus compared with normal population and patients with mild Alzheimer's disease. *J. Alzheimers Dis.* **81**, 1117–1130 (2021).
- 63. Weiner, S. et al. Optimized sample preparation and data analysis for TMT proteomic analysis of cerebrospinal fluid applied to the identification of Alzheimer's disease biomarkers. *Clin. Proteomics* **19**, 13 (2022).
- 64. Bennett, D. A. et al. Religious Orders Study and Rush Memory and Aging Project. J. Alzheimers Dis. **64**, S161–S189 (2018).
- Bennett, D. A. et al. Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* 66, 1837–1844 (2006).
- Bennett, D. A. et al. Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. *Neuroepidemiology* 27, 169–176 (2006).
- 67. Bennett, D. A. et al. Natural history of mild cognitive impairment in older persons. *Neurology* **59**, 198–205 (2002).
- Boyle, P. A. et al. Attributable risk of Alzheimer's dementia attributed to age-related neuropathologies. *Ann. Neurol.* 85, 114–124 (2019).
- Oveisgharan, S. et al. Estrogen receptor genes, cognitive decline, and Alzheimer disease. *Neurology* **100**, e1474–e1487 (2023).
- Wright, J. D. et al. The ARIC (Atherosclerosis Risk In Communities) study: JACC focus seminar 3/8. J. Am. Coll. Cardiol. 77, 2939–2959 (2021).
- Knopman, D. S. et al. Mild cognitive impairment and dementia prevalence: the Atherosclerosis Risk in Communities neurocognitive study. *Alzheimers Dement.* 2, 1–11 (2016).
- 72. Williams, S. A. et al. Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* **25**, 1851–1857 (2019).
- 73. Gold, L. et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5**, e15004 (2010).
- 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).
- Karlsson, L. et al. Cerebrospinal fluid reference proteins increase accuracy and interpretability of biomarkers for brain diseases. *Nat. Commun.* 15, 3676 (2024).
- 76. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).

- 77. McKinney, W. Data structures for statistical computing in Python. In Proc. 9th Python in Science Conference (eds van der Walt, S. & Millman, J.) 56–61 (SCIPY, 2010).
- Hunter, J. D. Matplotlib: a 2D graphics environment. Comput. Sci. Eng. 9, 90–95 (2007).
- 79. Waskom, M. L. seaborn: statistical data visualization. J. Open Source Softw. **6**, 3021 (2021).
- 80. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods.* **17**, 261–272 (2020).
- Seabold, S. & Perktold, J. Statsmodels: econometric and statistical modeling with Python. In Proc. 9th Python in Science Conference (eds van der Walt, S. & Millman, J.) 92–96 (SCIPY, 2010).
- 82. Pedregosa, F. et al. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011).
- 83. Davidson-Pilon, C. Lifelines, survival analysis in Python (v0.27.0). Zenodo https://doi.org/10.5281/zenodo.6359609 (2022).
- 84. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. J. Stat. Softw. **36**, 1–48 (2010).
- Raudvere, U. et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198 (2019).
- 86. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
- Aleksander, S. A. et al. The Gene Ontology knowledgebase in 2023. *Genetics* 224, iyad031 (2023).

## Acknowledgements

We thank the contributors who collected the samples used in this study, as well as patients and their families, whose help and participation made this work possible. We thank B. Lehallier, J. Rutledge, L. Gold and members of the Wyss-Coray laboratory for feedback and support, and D. Channappa for laboratory management. We are also grateful for the help of M. Parviainen and T. Laaksonen with patient management and cognitive testing. This work was supported by the Stanford-ADRC (NIA grant nos. P50AG047366 and P30AG066515), the National Centralized Repository for Alzheimer's Disease and Related Dementias (no. U24AG021886), the NIA (no. AG072255 to T.W.-C), the Milky Way Research Foundation (T.W.-C.), the Knight Initiative for Brain Resilience (T.W.-C.), the Stanford Graduate Fellowship (H.S.-H.O.), the National Science Foundation Graduate Research Fellowship (H.S.-H.O.), a grant from the KIBR (E.N.W.), a Stanford's Center for Clinical and Translational Education and Research award, under the Biostatistics, Epidemiology and Research Design Program (no. UL1TR003142 to Y.L.G.) and the National Institutes of Health (NIH) Pathway to Independence Award no. 1K99AG088304-01 (to I.H.G.). Samples from the National Centralized Repository for Alzheimer's Disease and Related Dementias, which receives government support under a cooperative agreement grant (no. U24 AG021886) awarded by the NIA, were used in this study. This work was also supported by grants from the NIH (no. R01AG044546 to C.C.; no. P01AG003991 to C.C.; no. RF1AG053303 to C.C.; no. RF1AG058501 to C.C.; no. U01AG058922 to C.C.), the Chan Zuckerberg Initiative, the Michael J. Fox Foundation (to C.C.), the Alzheimer's Association Zenith Fellows Award (no. ZEN-22-848604, awarded to C.C.) and an anonymous foundation. The recruitment and clinical characterization of the research participants at Washington University were supported by NIH grant nos. P30AG066444, P01AG03991 and P01AG026276. This work was supported by access to equipment made possible by the Hope Center for Neurological Disorders, the Neurogenomics and Informatics Center (https://neurogenomics.wustl.edu/) and the Departments of Neurology and Psychiatry at Washington University School of Medicine. The BioFINDER2 study was supported by the European Research Council (no. 101096455), the Alzheimer's Association (nos. ZEN24-1069572 and SG-23-1061717), the Gerald and Henrietta Rauenhorst Foundation,

## Article

the Swedish Research Council (no. 2022-00775), ERA PerMed (no. ERAPERMED2021-184), the Knut and Alice Wallenberg Foundation (no. 2022-0231), the Strategic Research Area MultiPark (multidisciplinary research in Parkinson's disease) at Lund University, the Swedish Alzheimer Foundation (no. AF-980907), the Swedish Brain Foundation (no. FO2021-0293), the Parkinson Foundation of Sweden (no. 1412/22), the Cure Alzheimer's Fund, the Rönström Family Foundation, the King Gustaf V:s and Drottning Victorias Masonic Foundation, the Skåne University Hospital Foundation (no. 2020-0000028), the Regional Research Support (no. 2022-1259) and the Swedish federal government under the ALF agreement (no. 2022-Projekt0080). SAMS is supported by grants from the NIH (nos. R01AG048076 and R21AG058859), the Stanford Wu Tsai Neurosciences Institute and the Stanford Center for Precision Health and Integrated Diagnostics, ROSMAP is supported by grant nos. P30AG10161, P30AG72975, R01AG15819, R01AG17917, U01AG46152 and U01AG61356. The Kuopio study was funded by the Alzheimer's Association, Academy of Finland (grant nos. 338182 and 328287), the KUH VTR Fund, the Sigrid Juselius Foundation, the Strategic Neuroscience Funding of the University of Eastern Finland and the Alzheimer's Association (ADSF-24-1284326-C). H.Z. is a Wallenberg Scholar and a Distinguished Professor at the Swedish Research Council; he is supported by grants from the Swedish Research Council (nos. 2023-00356, 2022-01018 and 2019-02397), the European Union's Horizon Europe Research and Innovation Programme under grant no. 101053962, the Swedish State Support for Clinical Research (no. ALFGBG-71320), the Alzheimer Drug Discovery Foundation USA (no. 201809-2016862), the AD Strategic Fund and the Alzheimer's Association (nos. ADSF-21-831376-C, ADSF-21-831381-C, ADSF-21-831377-C and ADSF-24-1284328-C), the European Partnership on Metrology, cofinanced by the European Union's Horizon Europe Research and Innovation Programme and by Participating States (NEuroBioStand, no. 22HLT07), the Bluefield Project, the Cure Alzheimer's Fund, the Olav Thon Foundation, the Erling Persson Family Foundation, the Rönströms Family Foundation, the Gamla Tjänarinnor Foundation, the Swedish Brain Foundation (no. FO2022-0270), the European Union's Horizon 2020 Research and Innovation Programme under a Marie Skłodowska-Curie grant no. 860197 (MIRIADE), the European Union Joint Programme-Neurodegenerative Disease Research (no. JPND2021-00694), the National Institute for Health and Care Research, the University College London (UCL) Hospitals Biomedical Research Centre and the UK Dementia Research Institute at UCL (no. UKDRI-1003). B.S. was supported by the Alzheimer's Association (grant nos. ADSF-21-836089-C, ADSF-21-836083-C and ADSF-21-836085-C). N.M. was supported by NIH training grant nos. 5T32AG222-30 and 1F32AG079666-01. ARIC is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) grant nos. 75N92022D00001, 75N92022D00002, 75N92022D00003, 75N92022D00004 and 75N92022D00005. Additionally, the ARIC Neurocognitive Study is supported by grant nos. U01HL096812, U01HL096814, U01HL096899, U01HL096902 and U01HL096917 from the NIH (NHLBI, the National Institute of Neurological Disorders and Stroke, the NIA and the National Institute on Deafness and Other Communication Disorders). M.R.D. and K.A.W. were supported in part by the NIA Intramural Research Program of the NIH.

# **Author contributions**

H.S.-H.O. conceptualized the study. H.S.-H.O. led the study design and analyses. D.Y.U. helped with the study design and analyses. L.K. helped with the analyses in the BioFINDER2 cohort. Z.Z. helped with the analyses in the ADNI cohort. Y.S. helped with the analyses in the DIAN cohort. A.F. helped with the analyses in the Stanford-ADRC cohort. M.R.D. and J. Chen helped with the analyses in the ARIC cohort.

N.M. helped with the analyses in the Kuopio cohort. J.T., I.H.G., C.Y., D.W., M.A., Y.L.G. and A.T. provided data, helped with the analyses and provided insights. T.R., S.-K.H., M.H., A.L. and A.J.L. collected the data from the Kuopio cohort. K.L.P. established the Stanford BPD cohort. E.M. and A.D.W. established the SAMS cohort. E.N.W. led the fluid AD biomarker data collection in the Stanford cohorts. D.C. led plasma collection in the Stanford cohorts. V.L., B.S. and H.Z. established the Kuopio cohort. A.J.E. provided key insights on the Alzheimer's field. R.F.G., J. Coresh and K.A.W. lead the ARIC neurocognitive study. D.A.B. established the ROSMAP cohort. N.F. provided key insights on the Alzheimer's field and helped with the analyses in the ADNI cohort. O.H. established the BioFINDER2 cohort. C.C. established the Knight-ADRC cohort. T.W.-C. established the Stanford-ADRC cohort. K.L.P., E.M., A.D.W., E.N.W., V.L., B.S., H.Z., D.A.B., N.F., O.H. and C.C. provided data and insights. H.S.-H.O. produced the figures and wrote the paper. T.W.-C. edited the paper. H.S.-H.O. and T.W.-C. supervised the study. All authors critically revised the paper for intellectual content. All authors read and approved the final version of the paper.

# **Competing interests**

T.W-C. and H.S.-H.O. are cofounders and scientific advisers of Teal Omics and have received equity stakes. C.C. has received research support from GSK and Eisai. C.C. is a member of the scientific advisory board of Circular Genomics and owns stocks. C.C. is a member of the scientific advisory board of ADmit. C.C. and M.A. have an invention disclosure for AT<sub>1</sub> prediction models, including protein IDs, weights, cutoffs and algorithms. O.H. has acquired research support (for the institution) from AVID Radiopharmaceuticals, Biogen, C2N Diagnostics, Eli Lilly, Eisai, Fujirebio, GE Healthcare and Roche. In the past 2 years, he has received consultancy and speaker fees from ALZpath, BioArctic, Biogen, Bristol Myers Squibb, Eisai, Eli Lilly, Fujirebio, Merck, Novartis, Novo Nordisk, Roche, Sanofi and Siemens. H.Z. has served at scientific advisory boards or as a consultant for Abbvie, Acumen, Alector, Alzinova, ALZPath, Amylyx, Annexon, Apellis, Artery Therapeutics, AZTherapies, Cognito Therapeutics, Cognition Therapeutics, Denali, Eisai, Labcorp, Merry Life Biomedical, NervGen Pharma, Novo Nordisk, OptoCeutics, Passage Bio, Pinteon Therapeutics, Prothena, Red Abbey Labs, reMYND, Roche, Samumed, Siemens Healthineers, Triplet Therapeutics and Wave, has given lectures in symposia sponsored by AlzeCure, Biogen, Cellectricon, Fujirebio, Lilly, Novo Nordisk and Roche, and is a cofounder of Brain Biomarker Solutions in Gothenburg, which is a part of the GU Ventures Incubator Program (outside the submitted work). The other authors declare no competing interests.

# **Additional information**

**Extended data** is available for this paper at https://doi.org/10.1038/s41591-025-03565-2.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41591-025-03565-2.

**Correspondence and requests for materials** should be addressed to Hamilton Se-Hwee Oh or Tony Wyss-Coray.

**Peer review information** *Nature Medicine* thanks Sudeshna Das, Tharick Pascoal and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Jerome Staal, in collaboration with the *Nature Medicine* team.

**Reprints and permissions information** is available at www.nature.com/reprints.

Article



Extended Data Fig. 1 | CSF YWHAG:NPTX2 ratio versus PTau181:A $\beta$ 42 and cognitive impairment. **a**, Global cognition score versus tau tangle load in A $\beta$  + individuals in the ROSMAP cohort. A $\beta$  and tau do not sufficiently explain cognitive impairment. **b**, Scatterplot showing both change with cognitive impairment independent of PTau181:A $\beta$ 42 (y-axis) and association with PTau181:A $\beta$ 42 (x-axis). Axes show signed –log10 q-values (Benjamini-Hochberg adjusted p-value) from linear regression models. Bold indicates synapse proteins based on SynGO database. **c**, Scatterplot showing PC1-adjusted YWHAG.1 (left) and NPTX2 (right) versus PTau181:A $\beta$ 42, colored by cognitive impairment. Correlation and p-value are shown. **d**, Boxplot showing YWHAG.1:NPTX2 versus cognitive impairment severity across cohorts with SomaScan data per biological sex (male n = 1,024, female n = 1043). Box bounds are Q1, median, and Q3, and whiskers show Q1 – 1.5×1QR and Q3 + 1.5×1QR. Pearson correlations per cohort are shown. **e**, As in **d**, but for PTau181:A $\beta$ 42 (male n = 1,024, female n = 1043). **f**, Results from binary logistic regression models classifying CI stage based on YWHAG.1:NPTX2 or PTau181:AB42. AUC, accuracy, sensitivity, and specificity are shown (total n = 2,067). Bars and error bars represent bootstrapped (n = 1,000) means and 95% confidence intervals. Two-sided p-values were calculated via the empirical distribution of the bootstrapped test statistic. \*p < 0.05, \*\*p < 0.01, \*\*\*\*p < 0.001. **g**, Receiver operating curve (ROC) for classification of mild/ moderate/severe dementia versus cognitively normal is shown.



**Extended Data Fig. 2** | **CSF YWHAG:NPTX2 ratio explains a substantial proportion of variance in cognitive impairment beyond amyloid and tau in AD. a**, Stacked bar plot showing proportions of different cognitive impairment stages among different YWHAG.1:NPTX2 percentile groups, in all and each cohort. **b**, Results from linear models regressing cognitive impairment against the displayed covariates, per cohort (total n = 2,150). Points and error bars represent effect sizes and 95% confidence intervals. **c**, Scatterplot showing YWHAZ versus YWHAG.1, colored by cognitive impairment (top). Scatterplot showing YWHAZ:NPTX2 versus YWHAG.1:NPTX2, colored by cognitive impairment (bottom). **d**, R-squared values from linear models regressing cognitive impairment against covariates displayed on x-axis (n = 2,067). The difference between r-squared values between two models is shown. \*\*\*p < 0.001. e, Boxplot showing YWHAZ:NPTX2 versus cognitive impairment severity based on mass-spectrometry data in BioFINDER2 (n = 829). Box bounds are Q1, median, and Q3, and whiskers show Q1 – 1.5×IQR and Q3 + 1.5×IQR. f, Results from a multivariate linear model regressing cognitive impairment against the displayed covariates in BioFINDER2 (n = 512). Points and error bars represent standardized effect sizes and 95% confidence intervals. g, R-squared values from linear models regressing cognitive impairment against covariates displayed on x-axis in BioFINDER2 (n = 512). The difference between r-squared values between different models are shown. \*\*\*p < 0.001. Colors in a,c,e indicate cognitive impairment stage as shown in Fig. 1e. Bars and error bars in d and g represent bootstrapped (n = 1,000) means and 95% confidence intervals. Two-sided p-values were calculated via the empirical distribution of the bootstrapped test statistic.



Extended Data Fig. 3 | Changes in CSF YWHAG:NPTX2 with normal aging and ADAD. a, Changes in CSF A $\beta$ 42:A $\beta$ 40, PTau181:A $\beta$ 42, and YWHAG.1:NPTX2 with cognitively normal aging in the BioFINDER2 cohort. Lowess regression lines with 95% confidence intervals are shown. b, Changes in CSF PTau181:A $\beta$ 42, and

YWHAG.1:NPTX2 with age in ADAD mutation carriers, binned by estimated age of onset, in the DIAN cohort. Lowess regression (left) and linear regression (right) lines are shown.

## a ADAS13 cognitive score multivariate regression



#### standardized effect size

**Extended Data Fig. 4** | **CSF YWHAG:NPTX2 association with future cognitive impairment independent of tau PET. a**, Results from a multivariate linear model regressing future ADAS13 cognitive score against the displayed covariates in BioFINDER2 (n = 70). PET data are from same time point as the cognitive score. YWHAG.1:NPTX2 and PTau181:A $\beta$ 42 was measured 4-15 years before PET. Points and error bars represent standardized effect sizes and 95% confidence intervals.



**Extended Data Fig. 5** | **Defined YWHAG:NPTX2 groups for prediction of future cognitive decline, meta-analysis. a**, Cox proportional hazard regression was used to associate YWHAG.1:NPTX2 groups with future cognitive decline in  $A + T_1 + MCI$  individuals, while adjusting for PTau181:A $\beta$ 42, CSF NfL, CSF Ng,

*APOE4*, age, and sex. Group 1 was the reference group. Results from a crosscohort fixed-effect meta-analysis are shown (total n = 397). Points and error bars represent hazard ratios and 95% confidence intervals. **b**, As in **a**, but for A +  $T_1$ + cognitively normal individuals (total n = 168). Group 0 was the reference group.

# nature portfolio

Corresponding author(s): Hamilton Se-Hwee Oh and Tony Wyss-Coray

Last updated by author(s): Feb 2, 2025

# **Reporting Summary**

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our <u>Editorial Policies</u> and the <u>Editorial Policy Checklist</u>.

# Statistics

For	all st	atistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.
n/a	Cor	nfirmed
	$\boxtimes$	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
	$\boxtimes$	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
		The statistical test(s) used AND whether they are one- or two-sided Only common tests should be described solely by name; describe more complex techniques in the Methods section.
	$\boxtimes$	A description of all covariates tested
	$\boxtimes$	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
		A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
		For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i> ) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable</i> .
$\boxtimes$		For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
$\boxtimes$		For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
	$\boxtimes$	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
		Our web collection on <u>statistics for biologists</u> contains articles on many of the points above.

# Software and code

Policy information about availability of computer code

Data collection	All analyses involved standard Python or R libraries used for basic data processing, statistics, machine learning, and plotting.
Data analysis	Python 3.9.12 gprofiler-official==1.0.0 lifelines==0.27.3 matplotlib==3.5.1 numpy==1.21.5 pandas==1.4.2 scikit-learn==1.0.2 scipy==1.10.0 statsmodels==0.13.5 R 4.1.2 metafor==4.2.0
	metafor==4.2.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Policy information about <u>availability of data</u>

All manuscripts must include a <u>data availability statement</u>. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data are available upon reasonable request with formal applications submitted to respective cohort committees to protect patient sensitive data. Stanford data (including Stanford-ADRC, SAMS, BPD, SCMD) can be requested at https://web.stanford.edu/group/adrc/cgi-bin/web-proj/datareq.php. Data from specific Stanford cohorts can be requested to the following cohort leaders: ADRC, Tony Wyss-Coray (twc@stanford.edu); SAMS, Beth Mormino (bmormino@stanford.edu) or Anthony Wagner (awagner@stanford.edu); BPD and SCMD, Kathleen Poston (klposton@stanford.edu). Knight-ADRC proteomics data were generated by the laboratory of principal investigator Carlos Cruchaga (cruchagac@wustl.edu) and are available upon reasonable request at https://knightadrc.wustl.edu/ professionals-clinicians/request-center-resources/submit-a-request/. ADNI data can be requested at the ADNI database (https://adni.loni.usc.edu/). DIAN data can be requested at https://dian.wustl.edu/our-research/for-investigators/diantu-investigator-resources/dian-tu-biospecimen-request-form/. Pseudonymized BioFINDER2 data can be shared to qualified academic researchers after request to cohort leader Oskar Hansson (oskar.hansson@med.lu.se) for the purpose of replicating procedures and results presented in the study. Data transfer must be performed in agreement with EU legislation regarding general data protection regulation and decisions by the Ethical Review Board of Sweden and Region Skåne. Kuopio data can be requested and accessed via a repository on Terra https:// app.terra.bio/#workspaces/marsh-terra-inph/iNPH Proteomics Workspace. ROSMAP data can be requested at https://www.radc.rush.edu and www.synpase.org. ARIC proteomics data is available through the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (https://biolincc.nhlbi.nih.gov/studies/ aric/). Additional ARIC requests for clinical or proteomic data from individual investigators may be submitted to the ARIC steering committees and will be reviewed to ensure that data can be shared without compromising patient confidentiality or breaching intellectual property restrictions. Participant-level demographic, clinical and proteomic data may be partially restricted based on previously obtained participant consent. Data sharing restrictions may also be applied to ensure consistency with confidentiality or privacy laws and considerations (https://sites.cscc.unc.edu/aric/).

Here are links and citations to gprofiler39 (https://biit.cs.ut.ee/gprofiler/gost) Gene Ontology40,41 database (https://geneontology.org/) and SynGO42 database (https://www.syngoportal.org/).

# Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Cohorts were generally age-sex matched between AD cases and controls. Sex was included as a covariate for all major regression models. The YWHAG:NPTX2 ratio changes with cognitive impairment, similarly in both sexes.			
Reporting on race, ethnicity, or other socially relevant groupings	Race and other cohort demographics are provided in Supplementary Tables 1 (CSF) and 10 (plasma)			
Population characteristics	PARTICIPANTS Age, sex, race, APOE4 genotype, PTau181:AB42 distributions for each cohort are provided in Supplementary Tables 1 (CSF) and 10 (plasma). Sporadic AD cohorts were made up of participants of ages around 50-90. DIAN, the autosomal dominant AD cohort, was made up of participants around ages 20-60.			
	Stanford (ADRC, SAMS, BPD, SCMD) Plasma and CSF collection, processing, and storage for all Stanford cohorts followed a single standard procedure. All studies were approved by the Institutional Review Board of Stanford University and written informed consent or assent was obtained from all participants or their legally authorized representative.			
	Blood collection and processing followed a rigorous standardized protocol to minimize variation. Briefly, about 10cc of whole blood was collected in 4 vacutainer ethylenediaminetetraacetic acid (EDTA) tubes (Becton Dickinson) and spun at 1800 x g for 10minutes to separate out plasma, leaving 1cm of plasma above the buffy coat to avoid contamination. Plasma was aliquoted into polypropylene tubes and stored at -80oC. Processing took approximately one hour draw to freezing and storage. All draws occured in the morning to minimize circadian effects.			
	CSF was collected via lumbar puncture using a 20-22 G spinal needle that was inserted in the L4-L5 or L5-S1 interspace. CSF samples were immediately centrifuged at 500 x g for 10 minutes, aliquoted in polypropylene tubes, and stored at -80oC.			
	Plasma from all Stanford cohorts were sent to Somalogic for proteomics (v4.1 SOMAscan, ~7,000 proteins) in the same batch. CSF samples from all Stanford cohorts were sent to Somalogic for proteomics (v4.0 SOMAscan, ~5,000 proteins) in the same batch. Core CSF AD biomarkers A $\beta$ 42, A $\beta$ 40, and pTau181 were measured using the fully automated Lumipulse G 1200 instrument (Fujirebio US, Malvern, PA) as previously described1,2. Descriptions for each cohort are provided below.			
	A total of 1160 plasma samples (738 participants, longitudinal sampling) and 371 CSF samples (371 participants, 1 sample from each) from Stanford were included in this study. Per-cohort sample sizes are as follows: ADRC plasma n=827 (423 participants), CSF n=113; SAMS plasma n=222 (215 participants), CSF n=169. BPD plasma n=55 (55 participants), CSF n=68. SCMD plasma n=45 (45 participants), CSF n=21.			
	Stanford- Alzheimer's Disease Research Center (ADRC)			

Samples were acquired through the National Institute on Aging (NIA)-funded Stanford-ADRC, a longitudinal observational study of clinical dementia subjects and age-sex-matched nondemented subjects. Healthy control participants were deemed cognitively unimpaired during a clinical consensus conference that included board-certified neurologists and neuropsychologists. Cognitively impaired subjects underwent Clinical Dementia Rating and standardized neurological and neuropsychological assessments, including procedures of the National Alzheimer's Coordinating Center (https:// naccdata.org/), to determine cognitive and diagnostic status. Cognitive status was determined in a clinical consensus conference that included neurologists and neuropsychologists. All participants were free from acute infectious diseases and in good physical condition.

#### Stanford Aging and Memory Study (SAMS)

SAMS is an ongoing longitudinal study of healthy aging. Blood and CSF collection/processing and neurological/ neuropsychological assessments were performed by the same team and followed the same protocol as in Stanford-ADRC. All SAMS participants had CDR=0 and a neuropsychological test score within the normal range; all SAMS participants were deemed cognitively unimpaired during a clinical consensus conference that included neurologists and neuropsychologists.

#### Stanford Biomarkers in Parkinson's Disease Study (BPD)

The BPD cohort3 was a Michael J Fox Foundation for Parkinson's Research (MJFF) funded longitudinal study of biomarkers associated with cognitive decline in Parkinson's disease (PD). Research participants were recruited from the Stanford Movement Disorders Center between 2011-2015 with a PD diagnosis according to UK Brain Bank criteria and required bradykinesia with muscle rigidity and/or rest tremor. All participants completed baseline cognitive, motor, neuropsychologic, imaging, and biomarkers assessments (plasma and optional CSF) including Movement Disorders Society-revised Unified Parkinson's Disease Rating Scale (MDS-UPDRS). Age-match Healthy Controls (HC) were also recruited to control for age-associated biomarker changes. After comprehensive neuropsychological battery all participants were given a cognitive diagnosis of no cognitive impairment, mild cognitive impairment, or dementia, according to published criteria.

#### Stanford Center for Memory Disorders Cohort Study (SCMD)

The SCMD was an NIA-funded cross-sectional study of people across the cognitive continuum. Participants with mild dementia due to Alzheimer's (AD) and amnestic mild cognitive impairment (aMCI) were recruited from the Stanford Center for Memory Disorders between 2011-2015. Participants were included if they had a diagnosis of probable AD dementia (amnestic presentation) according to the National Institute on Aging–Alzheimer's Association4 (NIA-AA) criteria and a Clinical Dementia Rating (CDR) score of 0.5 or 1, or a diagnosis of MCI according to the NIA-AA criteria4, a score of 1.5 SDs below age-adjusted normative means on at least one test of episodic memory, and a CDR score of less than 1. Healthy older controls (HC) were recruited from the community, were selected to have a similar average age as enrolled patients, and were required to have normal neuropsychological performance and CDR of 0. Participants completed cognitive, neuropsychologic, imaging, and biomarker assessments with plasma.

#### Knight- Alzheimer's Disease Research Center (ADRC)

The Knight-ADRC cohort is an NIA-funded longitudinal observational study of clinical dementia subjects and age-matched controls. Research participants undergo longitudinal cognitive, neuropsychologic, imaging and biomarker assessments including Clinical Dementia Rating (CDR). AD cases corresponded to those with a diagnosis of dementia of the Alzheimer's type (DAT) using criteria equivalent to the National Institute of Neurological and Communication Disorders and Stroke-Alzheimer's Disease and Related Disorders Association for probable AD, and AD severity was determined using the Clinical Dementia Rating (CDR) at the time of lumbar puncture (for CSF samples) or blood draw (for plasma samples). Controls received the same assessment as the cases but were nondemented (CDR=0).

Blood samples were collected in EDTA tubes (Becton Dickinson vacutainer purple top) at the visit time, immediately centrifuged at 1,500g for 10min, aliquoted on two-dimensional barcoded Micronic tubes (200ul per aliquot) and stored at -80°C. The plasma was stored in monitored -80°C freezer until it was pulled and sent to Somalogic (SOMAscan7k) for data generation. Proteomics data from 2,112 plasma samples from each of 2,122 participants were included in this study.

CSF samples were collected through lumbar puncture from participants after an overnight fast. Samples were processed and stored at -80 °C until they were sent for protein measurement. Proteomics data from 927 CSF samples from each of 927 participants were included in this study. CSF samples from Knight-ADRC, ADNI, and DIAN were sent for proteomics using the SOMAscan platform (SOMAscan7k) in the same batch. CSF Aβ42, Aβ40, and pTau181 were measured using the LUMIPULSE G1200 immunoassay platform according to the manufacturer's specifications.

The Institutional Review Board of Washington University School of Medicine in St. Louis approved the study and research was performed in accordance with the approved protocols.

#### Alzheimer's Disease Neuroimaging Initiative (ADNI)

ADNI is a longitudinal multi-center study designed to develop early biomarkers of AD. All data used in this study were accessed from the ADNI database https://adni.loni.usc.edu/. ADNI investigators assessed and diagnosed subjects as either cognitively normal (CN; Mini Mental State Examination [MMSE] $\geq$ 24, Clinical Dementia Rating [CDR]=0, non-depressed), mildly cognitively impaired (MCI; MMSE $\geq$ 24, CDR=0.5, objective memory-impairment on education-adjusted Wechsler Memory Scale II, preserved activities of daily living) or demented (MMSE=20-26, CDR>0.5, NINCDS/ADRDA criteria for probable AD). Comprehensive details on study design, data acquisition, ethics, and policies can be found above. CSF A $\beta$ 42, pTau181 were measured using the xMAP immunoassay platform according to the manufacturer's specifications. Proteomics data from 725 CSF samples from each of 725 participants were included in this study.

#### Dominantly Inherited Alzheimer Network (DIAN)

DIAN, led by Washington University School of Medicine in St. Louis, is a family-based long-term observational study designed to understand the earliest changes of autosomal dominant AD (ADAD). All participants undergo clinical and cognitive batteries (i.e. Clinical Dementia Rating – Global). Comprehensive details on study design, data acquisition, ethics, and policies can be found at https://dian.wustl.edu/. The data used in this study are from data freeze 15 (DF15). CSF Aβ42, pTau181 were measured using the LUMIPULSE G1200 platform according to the manufacturer's specifications. Proteomics data from 455 CSF samples from each of 455 participants were included in this study.

#### BioFINDER2

BioFINDER2 is a Swedish prospective cohort study (NCT03174938) on age-related neurodegenerative diseases. AD was diagnosed based on the Diagnostic and Statistical Manual of Mental Disorders (5th edition) and AB positivity (CSF Aβ42:Aβ40). Proteomics data from a total of n=829 participants, consisting of n=480 cognitively unimpaired, n=213 with mild cognitive impairment and n=136 with AD dementia were included in this study. CDR-Global scores were not measured in BioFINDER2, so for estimation, AD participants were subdivided according to Mini-Mental State Examination (MMSE)thresholds defined in Perneczkey et. al.5: 21–25 for mild dementia (CDR=1), 11–20 for moderate dementia (CDR=2), 0–10 for severe dementia (CDR=3). The participants were recruited at Skåne University Hospital and the Hospital of Ängelholm, Sweden. The study was approved by the Regional Ethical Committee in Lund, Sweden, and all participants gave written informed consent.

CSF samples were collected close in time after baseline clinical examination and handled according to established preanalytical protocols, previously described in detail6. All analyses were performed by technicians blinded to all clinical and imaging data. CSF P-tau181, A $\square$ 42 and A $\square$ 40 was measured using Elecsys assays in accordance with the manufacturer's instructions (Roche Diagnostics International Ltd). CSF A $\beta$ 42/A $\beta$ 40 was used to define A $\beta$  positivity according to previously established cutoffs of < 0.087. CSF samples from the BioFINDER2 cohort were analyzed with liquid chromatography-tandem mass spectrometry (LC-MS/MS), previously described in detail8.

Tau-PET was performed using [18F]RO948. Standardized uptake value ratio (SUVR) images were created for the 70-90 min post-injection interval using the inferior cerebellar cortex as reference region. A composite corresponding to a Braak I-IV meta-region of interest was used to represent AD-related tau tangle pathology.

#### Kuopio University Hospital

The Kuopio Normal Pressure Hydrocephalus (NPH) and AD Registry and Tissue Bank includes patients from the Eastern Finnish population who were referred to the KUH neurosurgical unit for suspected NPH. The registry's inclusive criteria encompass a wide range of hydrocephalic conditions and comorbidities: patients must exhibit one to three symptoms potentially associated with NPH (such as impaired gait, cognition, or urinary continence) along with enlarged brain ventricles (Evans' index > 0.3) as seen on computer tomography (CT) or magnetic resonance imaging (MRI), and no other clear cause that alone explains the observed findings and symptoms. Preoperative comorbidities and conditions were recorded at baseline, and patients underwent a systematic differential diagnostic workup followed by a CSF tap test paired with gait evaluation.

Diagnostic right frontal cortical brain biopsy was taken from all the patients during shunt surgery from the insertion site of the intraventricular catheter. A neuropathologist (TR) analyzed immunoreactivity for A $\beta$  and hyperphosphorylated tau with light microscopy and the results were graded as present or absent. The patients were then further classified into groups by the presence of the pathology of A $\beta$  and/or HPt observed in the frontal cortical biopsies9. Follow-up was conducted on all operated patients, with optimal shunt function ensured through valve adjustment, brain imaging, shunt valve tapping, lumbar infusion testing, and shunt revision if necessary. CDR-Global scores were not measured in Kuopio, so the CERAD cognitive score10 including MMSE was used instead. The latest follow-up MMSE score <26 was considered as at least mild dementia9.

Lumbar CSF proteomics was performed using high-throughput tandem mass tag (TMT)-labelling mass spectrometry, previously described in detail11. Data from 90 subjects with CSF proteomics and cognitive scoring performed within 1 year were included in this study.

The study was conducted according to the Declaration of Helsinki and all patients provided informed consent. The Research Ethics Committee of the Northern Savo Hospital District (decision No 276/13.02.00/2016).

#### Religious Order Study and Rush Memory and Aging Project (ROSMAP)

All ROSMAP participants enrolled without known dementia and agreed to detailed clinical evaluation and brain donation at death12. Both studies were approved by an Institutional Review Board of Rush University Medical Center (ROS IRB# L91020181, MAP IRB# L86121802). Both studies were conducted according to the principles expressed in the Declaration of Helsinki. Each participant signed an informed consent, Anatomic Gift Act, and an (Rush Alzheimer's Disease Center) RADC Repository consent (IRB# L99032481) allowing their data and biospecimens to be repurposed. All MAP participants and a subset of ROS have blood drawn during an annual home visit. For plasma, blood is drawn in a lavender (purple) top EDTA tube. For out of town ROS sites, they were spun, aliquoted into nunc vials, stored in dry ice and sent to the RADC by FedEx where they were transferred to -800C. Samples collected in the northeastern Illinois were brought to the RADC laboratory and processed there with the same procedures. A total of 1046 55ul samples were shipped to Stanford, then to Somalogic for proteomics (SOMAscan7k). 973 samples passed quality control.

Clinical and neuropathologic data collection has been reported in detail13–17. Regarding clinical diagnosis, an actuarial decision tree designed to mimic expert clinical judgment was implemented by computer to inform several clinical diagnoses, including dementia and AD. It combines data reduction techniques for the cognitive performance testing with a series of discrete clinical judgments made in series by a neuropsychologist and a clinician. Presumptive diagnoses of dementia and AD are calculated that conform to accepted clinical criteria. The clinician is asked to agree or disagree with the decisions. An algorithm uses these decisions to provide diagnoses of MCI and amnestic MCI. Persons with MCI are judged to have cognitive impairment by the neuropsychologist without a diagnosis of dementia by the clinician. Persons without dementia or MCI are categorized as having no cognitive impairment (NCI).

CDR-Global scores were not assessed in ROSMAP. Rather, participants underwent a batter of 21 cognitive performance tests, 17 of which were combined into a measure of global cognition which was z-scored based on the mean and standard deviation of all participants from baseline18. For estimation participants were subdivided according to Global Cognition z-scores: cognitively normal = z-score>0, mild cognitive impairment = -1<z-score<0, mild dementia = -2<z-score<=-1, moderate dementia -3<z-score<=-2, and severe dementia z-score<=-3. These cutoffs were set based on distributions of Global Cognition z-scores per clinical diagnosis. Details on cognitive scores, neuropathology, and other patient info are

described at https://www.radc.rush.edu/documentation.htm. Proteomics data from 973 plasma samples from each of 890 participants were included in this study.

Global Neurodegeneration Proteomics Consortium (GNPC)

The GNPC19 is a major neurodegenerative disease biomarker discovery effort, which hosts the largest collection of SomaScan data (over 40,000 patient samples from over 20 international research groups) from patient samples across healthy aging, Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis (ALS), and frontotemporal dementia (FTD). All cohorts and data are anonymized. Cohorts that measured CDR-Global cognitive scores were included in this study.

Atherosclerosis Risk in Communities (ARIC) Study

ARIC is a prospective epidemiologic study conducted in four U.S. communities: Forsyth County, NC; Jackson, MS; the northwest suburbs of Minneapolis, MN; and Washington County, MD. The ARIC study enrolled 15,792 mostly white and black participants aged 45-64 between 1987-198920. After initial enrollment, participants had four additional in-person visits: Visit 2 (1990-1992), Visit 3 (1993-1995), Visit 4 (1996-1999), Visit 5 (2011-2013). Participants were invited back for Visit 6 (2016-2017) after five years, and Visit 7 (2018-2019) immediately thereafter. Additional follow-up visits are ongoing. Blood was drawn for proteomic analysis at Visits 2 and 5. For this manuscript, late-life (8-year) dementia risk was assessed between Visit 5 and 7. Plasma was collected using standardized protocols and frozen at -80°C until analysis. Proteins were measured using the SomaScan v4.0 assay; protein QC steps have been described in detail previously21.

Dementia was adjudicated through cognitive assessment tests, telephone screening, informant ratings, hospital records, and death record review, as previously described22. From Visits 2 through 4, a 3-instrument cognitive assessment was applied (delayed word recall task, digit symbol substitution from the Wechsler Adult Intelligence Scale -Revised, and a letter fluency task). At Visits 5 through 7, a surveillance approach was employed, where participants received a comprehensive cognitive exam and a functional assessment that included the Clinical Dementia Rating Scale and Functional Activities Questionnaire. With these data, dementia was classified based on the National Institute on Aging and Alzheimer's Association and the Diagnostic and Statistical Manual of Mental Disorder – Fifth Edition criteria. In the time between Visit 5 and 6, participants were contacted annually via phone and administered the SIS, a brief cognitive assessment. If participants received a low score on the SIS, or if they were unable to participate in the screening via phone, the Ascertain Dementia 8 (AD8) was administered to the participant's informant. For participants who received a dementia diagnosis at visit 6 or 7, SIS, AD8, hospital discharge, and death certificate codes were used to define the date of dementia onset. For participants who did not attend visits 6 or 7, SIS, AD8, hospital discharge, and death certificate codes were used to define the stree used to define dementia diagnosis and date of dementia onset.

ARIC study protocols were approved by IRBs at each participating center: University of North Carolina at Chapel Hill, Chapel Hill, NC; Wake Forest University, Winston-Salem, NC; Johns Hopkins University, Baltimore, MD; University of Minnesota, Minneapolis, MN; and University of Mississippi Medical Center, Jackson, MS. All ARIC participants gave written informed consent at each study visit; proxies provided consent for participants who were judged to lack capacity.

Recruitment

Patient recruitment is detailed per cohort in the methods

Ethics oversight

All cohorts and studies were approved by appropriate ethics boards

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	All samples with CSF/plasma proteomics data and cognitive impairment information were assessed. No other sample size calculations were performed.
Data exclusions	CSF/plasma proteomics data and cognitive impairment information were necessary.
Replication	Multi-cohort study (6 for CSF, 9 for plasma). All attempts at replication were successful.
Randomization	NA, no treatment group
Blinding	NA, no treatment group

# Reporting for specific materials, systems and methods

nature portfolio | reporting summary

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems			Methods	
n/a	Involved in the study	n/a	Involved in the study	
$\ge$	Antibodies	$\boxtimes$	ChIP-seq	
$\times$	Eukaryotic cell lines	$\boxtimes$	Flow cytometry	
$\times$	Palaeontology and archaeology	$\boxtimes$	MRI-based neuroimaging	
$\times$	Animals and other organisms			
$\times$	Clinical data			
$\mathbf{X}$	Dual use research of concern			
$\ge$	Plants			

# Plants

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor
Authentication	was applied. Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.