

A deep joint-learning proteomics model for diagnosis of six conditions associated with dementia

Received: 18 April 2025

Accepted: 23 February 2026

Published online: 31 March 2026

 Check for updates

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

Co-pathology is a common feature of neurodegenerative diseases that complicates diagnosis, treatment and clinical management. However, sensitive, specific and scalable biomarkers for in vivo pathological diagnosis are not available for most neurodegenerative neuropathologies. Here we present Proteomics-based Artificial Intelligence for Dementia Diagnosis (ProtAIDe-Dx), a deep joint-learning model on 17,187 patients and controls (age of 70.3 ± 11.5 years, 53.2% female), that uses plasma proteomics to provide simultaneous probabilistic diagnosis across 6 conditions associated with dementia in aging. ProtAIDe-Dx achieves cross-validated balanced classification accuracy of 70–95% and area under the curve of >78% across all conditions. The model's diagnostic probabilities highlighted subgroups of patients with co-pathologies and were associated with pathology-specific biomarkers in an external memory clinic sample, even among individuals without cognitive impairment. Model interpretation revealed a suite of protein networks marking shared and specific biological processes across diseases and identified novel and previously described proteins discriminating each diagnosis. ProtAIDe-Dx significantly improved biomarker-based differential diagnosis in a memory clinic sample, pinpointing proteins leading to diagnostic decisions at an individual level. Together, this work highlights the promise of plasma proteomics to improve patient-level diagnostic workup with a single blood draw.

The past 5 years have seen multiple breakthroughs in the treatment of neurodegenerative diseases. Early disease-modifying therapies have emerged for Alzheimer's disease (AD)^{1,2}, and highly promising drug candidates are currently in clinical trials for AD³, Parkinson's disease (PD)⁴ and amyotrophic lateral sclerosis (ALS)⁵. However, differential diagnosis and disease comorbidity continue to pose considerable challenges in these treatment efforts. Misdiagnosis rates are around 25–30% even in specialized dementia clinics and can exceed 50% in primary care^{6–8}. Meanwhile, comorbidity is common in aging, with 70% of patients 80 years or older harboring multiple neurodegenerative pathologies simultaneously⁹. Misdiagnosis can make it difficult to select the right patients for a drug trial¹⁰, while comorbid

neuropathologies can mask the positive effects of a putative therapy^{11,12}. Once such treatments do become available, misdiagnosis and comorbidity can both lead to treatment mismanagement¹³. With the rapid pace at which promising new drugs are being tested, there is an urgent need for powerful tools for diagnosis and precise identification of underlying comorbid pathologies.

The first step toward mitigating the issues presented by misdiagnosis and comorbidity is the development of biomarkers to identify underlying neurodegenerative pathology with high specificity. Blood-based biomarkers have the potential to be highly accessible, inexpensive and minimally invasive, and the emergence of blood-based biomarkers for AD could facilitate accurate AD diagnosis even in

✉ e-mail: lijun.an@med.lu.se; jacob.vogel@med.lu.se

primary care in the near future⁷. Despite the success of AD biomarkers, scalable, sensitive and specific biomarkers for other neurodegenerative diseases are lacking, and diagnosis can only be made with high confidence at autopsy. Plasma proteomics represent a promising tool toward this aim, allowing robust surveillance of thousands of potential biomarkers and relevant functional effectors with a single blood draw¹⁴. However, despite great promise, plasma proteomics data are not without challenges. Proteomics data are high rank, come with burdensome technological artifacts and probably represent complex nonlinear interactions^{15,16}. In addition, the blood–brain barrier limits the number of brain-expressed proteins relevant in neurological conditions that can be detected in blood¹⁷.

In the present study, we attempt to overcome these limitations by applying artificial intelligence (AI) to the Global Neurodegenerative Proteomics Consortium (GNPC)¹⁸ v1.3MS dataset, the largest neurodegenerative disease plasma proteomics dataset so far. We present a model called Proteomics-based AI for Dementia Diagnosis (ProtAIDe-Dx), a deep multi-task architecture to resolve differential and multiple neurodegenerative diagnoses with a single blood draw (Fig. 1a). We report the performance of ProtAIDe-Dx on multi-diagnosis prediction, evaluate its potential in identifying co-pathology, test its capability in differential diagnosis compared with other clinical markers and explore putative molecular networks contributing to its predictions. Finally, we present a proof of concept using ProtAIDe-Dx for personalized and interpretable diagnostic testing in a clinical scenario. With this approach, we hope to set a benchmark for future plasma-based multi-disease neurodegenerative diagnostic tools.

Results

A subsample of 17,187 participants with SomaLogic 7k proteomics, sampled across 19 contributing sites, was selected from the GNPC v1.3MS dataset for subsequent analysis (Methods). Supplementary Table 1 presents sample and demographics across contributing sites and Supplementary Table 2 presents the frequency of six conditions across each site, namely, AD, PD, frontotemporal dementia (FTD), ALS, previous stroke/transient ischemic attack (TIA) and cognitive unimpaired. Given the high prevalence of vascular dementia (second only to AD¹⁹) but the lack of vascular dementia diagnoses in GNPC, the stroke/TIA group was chosen as a representation of patients with documented cerebrovascular disease.

We applied the ProtAIDe-Dx model to this sample, an architecture capable of generating several features of interest simultaneously: binary diagnosis of each condition, probabilities of each diagnosis and joint embeddings representing low-dimensional, nonlinear protein combinations used by the model for diagnosis (Fig. 1a). We specifically chose a multi-task, joint-learning approach (as opposed to a multi-class classification task) to allow the model to signal disease co-pathology (that is, positive for multiple disorders and probabilities for each disorder).

Joint learning improves multi-diagnostic prediction of neurodegenerative diagnosis from blood in unbalanced samples

We applied ProtAIDe-Dx to the GNPC sample, using tenfold cross-validation stratified for each contributing site. Importantly, we used only proteomic information in the model—no site, demographic, cognitive or diagnostic information was used. We compared the diagnostic performance of ProtAIDe-Dx against multiple machine learning and state-of-the-art deep learning baselines, including Random Forest, XGBoost²⁰ and TabPFN²¹. We also tested an ensemble model combining aspects of both XGBoost and ProtAIDe-Dx (Methods).

ProtAIDe-Dx emerged as the best-performing model overall, while XGBoost was the best non-deep learning model (Fig. 1b). ProtAIDe-Dx achieved a median balanced classification accuracy (BCA) performance above 90% for ALS (95%) and PD (92%) classification, 83% for control, 81% for AD, 72% for FTD and 70% for stroke/TIA (Fig. 1b). ProtAIDe-Dx

significantly outperformed Random Forest across all tasks; XGBoost in AD (false discovery rate (FDR)-corrected P value of 5×10^{-4}), FTD (FDR-corrected P value of 3×10^{-4}) and stroke (FDR-corrected P of 0.004) classification; and significantly outperformed TabPFN in FTD classification (FDR-corrected P value of 0.047). With the exception of Random Forest, all models achieved area under the curve (AUC) >0.8 for all tasks other than stroke/TIA prediction and demonstrated comparable AUCs (Fig. 1b, Supplementary Table 3 and Supplementary Data 1), although AUC alone might convey overoptimistic implications in some imbalanced classification scenarios²² (Supplementary Fig. 1). We found that the ensemble model produced balanced accuracy scores and AUCs that significantly outperformed ProtAIDe-Dx for control and PD diagnosis. The ensemble model also significantly improved BCA scores over XGBoost for all tasks except ALS and PD prediction (Fig. 1b, Supplementary Table 3 and Supplementary Data 1).

As a sanity check, we extracted the probability of AD diagnosis across all individuals and compared them with factors known to be altered in AD. In patients both with and without an AD diagnosis, higher AD probabilities were associated with more copies of the *APOE* $\epsilon 4$ allele and lower AD probabilities with more copies of the $\epsilon 2$ allele (Fig. 1c). In addition, a negative correlation was observed between AD probabilities and Mini-Mental State Examination (MMSE) score, indicating worse cognition associated with higher AD probabilities (Fig. 1d). These analyses suggest that model-derived diagnostic probabilities can serve as continuous proteomic scores associated with indicators of disease progression.

Diagnostic prediction model generalizes to new disease-relevant tasks

Low-dimensional nonlinear proteomic embeddings were extracted from the last layer of the ProtAIDe-Dx model (Fig. 1a). These embeddings should represent a compressed representation of plasma proteomic data, optimized toward tasks related to neurological diseases. To test this hypothesis, we used the embeddings to generalize the ProtAIDe-Dx model to a task that ProtAIDe-Dx was not trained specifically for, namely, prediction of longitudinal clinical progression in healthy controls (Methods). The model differentiated diagnostic progressors (that is, from clinical dementia rating (CDR) 0 to CDR 0.5 or 1; $N = 218$) from non-progressors (remained stably at CDR 0 over time; $N = 1,445$) with a BCA of 70% and an AUC of 74% (Fig. 1e). These results support ProtAIDe-Dx as a flexible and extensible model for neurodegenerative disease-related tasks.

Diagnostic probabilities reveal disease heterogeneity and co-pathology

ProtAIDe-Dx provides probabilities of each condition for each individual. We projected all individuals into a two-dimensional nonlinear embedding on the basis of their disease probabilities (Fig. 2a). As expected, individuals naturally clustered on the basis of their true clinical diagnosis and not by contributing site (Extended Data Fig. 1). Common phenotypic data in the GNPC distributed in expected patterns across the embeddings, with worse cognitive impairment and more *APOE* $\epsilon 4$ carriers in AD regions, fewer $\epsilon 4$ carriers in PD and ALS regions and more hypertension in the stroke/TIA region (Fig. 2b–d).

Next, we used ProtAIDe-Dx to predict etiological diagnoses of patients with ambiguous etiologies that were not used in model training, namely, patients diagnosed with subjective cognitive decline (SCD) or mild cognitive impairment (MCI; $N = 3,116$) and patients characterized as ‘HealthyAD’ (diagnosis of AD but cognitive scores in the healthy range), ‘ComputedDementia’ (no diagnosis but cognitive scores in the dementia range) and ‘Unknown’ (no diagnostic or cognitive information) groups (Methods). When projecting these cases onto the diagnostic embedding (Fig. 2e and Supplementary Fig. 2), the cases were distributed throughout the embedding, with cases falling neatly into regions corresponding to different conditions. This signals the

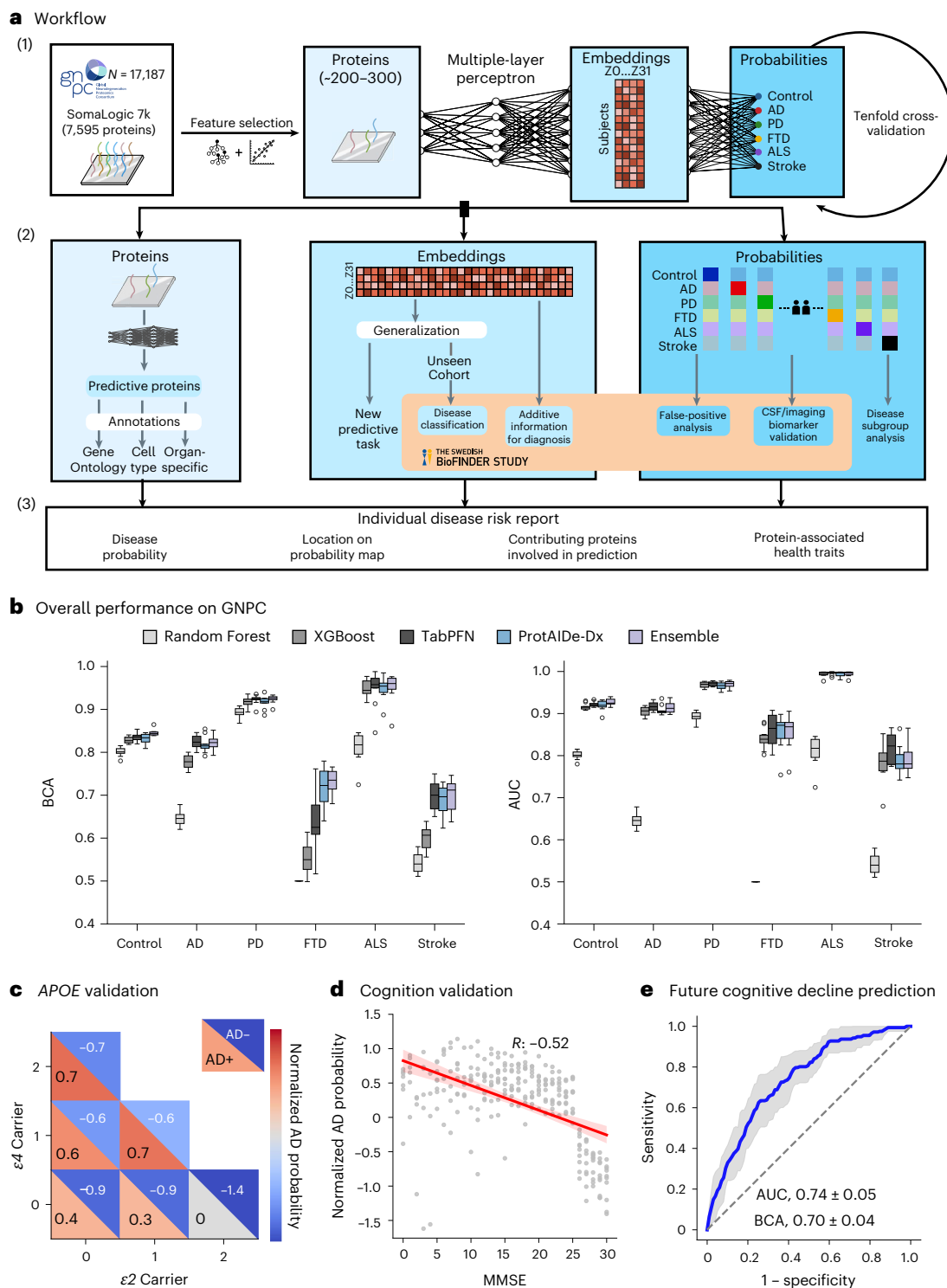


Fig. 1 | Workflow and overall performance of ProtAIDe-Dx on GNPC.

a, Project workflow. (1) Model development on GNPC with tenfold cross-validation procedure. The joint learning framework allows learning of multiple neurodegenerative classification tasks jointly, facilitating shared information during training. (2) Model evaluation on GNPC (blue panels) and BioFINDER-2 cohort (orange panel). Model outputs include not just classifications but also probabilities for each class, and contributing proteins and embeddings can be probed to better understand model choices. (3) Individual neurological disease risk report based on the developed model. **b**, Overall model performance on GNPC. Left: BCA score for each diagnostic task. Right: AUC score for each diagnostic task. Box plots were drawn across 10 cross-validation folds, showing

the median, interquartile range (IQR; 25th to 75th percentiles) and whiskers extending to $1.5 \times$ IQR. Two-sided corrected resample *t*-tests were applied; *P* values were FDR corrected. **c**, Normalized AD probabilities stratified by different *APOE* $\epsilon 2/\epsilon 4$ groups, shown for participants with (lower diagonal) and without (upper diagonal) an AD diagnosis. **d**, Correlation between normalized AD probabilities and MMSE. Error bands represent the 95% confidence interval of the regression line. **e**, Receiver operating characteristic (ROC) curve of model generalization to a new task for predicting longitudinal clinical progression (from no cognitive impairment to future cognitive impairment). Shaded regions represent the mean ROC ± 1 s.d. across ten cross-validation folds. Schematic and logo in **a** created in BioRender; An, L. <https://biorender.com/q2by4y5> (2026).

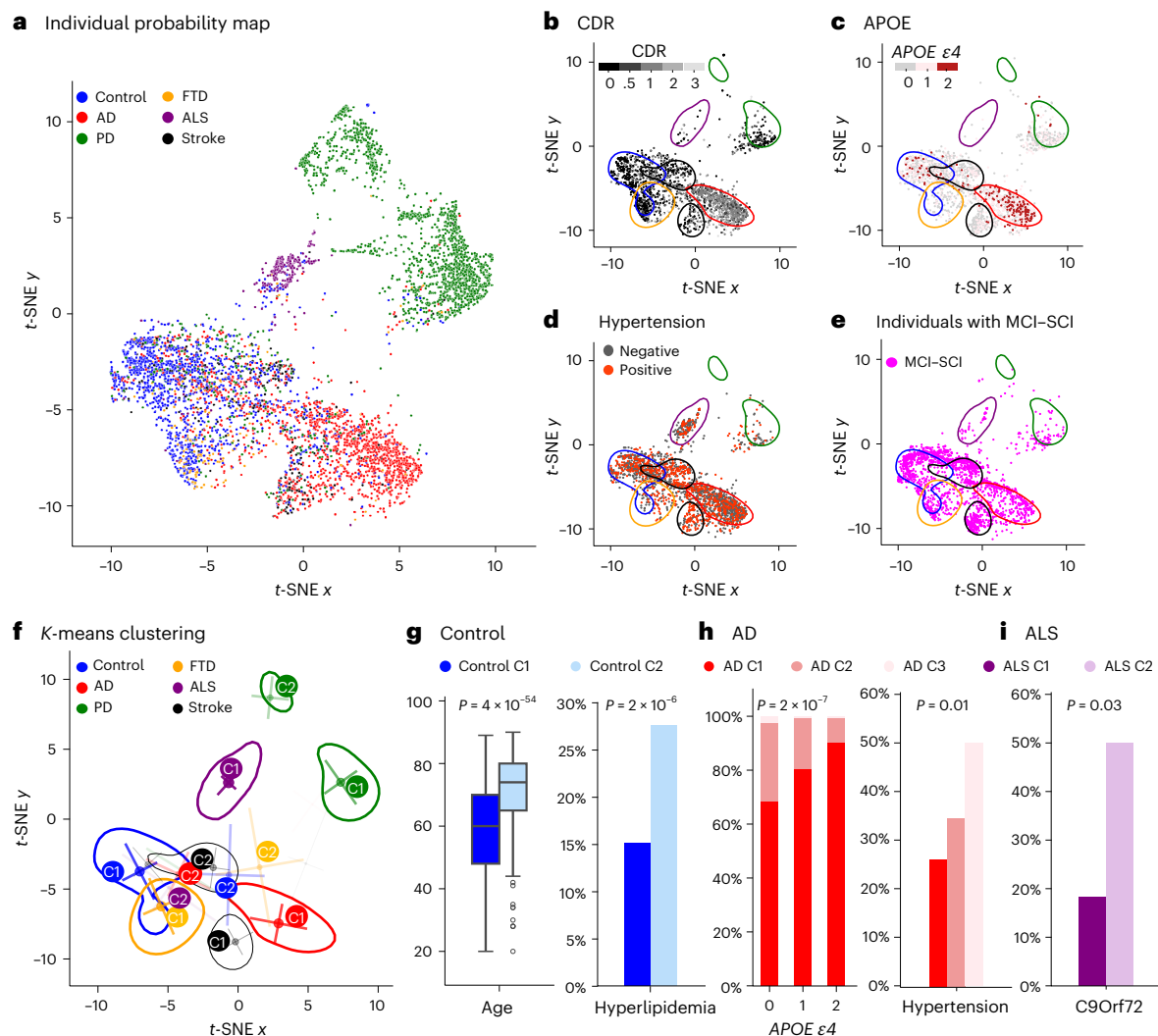


Fig. 2 | Diagnostic probability map derived by ProtAIDe-Dx reveals disease heterogeneity. **a**, Participants from the test set were selected for t-SNE to reduce dimensionality from six (predicted class probabilities) to two (nonlinear disease probability axes), colored by clinical diagnosis. **b–e**, Participants were projected onto the two-dimensional diagnostic probability map, colored by values of CDR (**b**), APOE (**c**), hypertension (**d**) and individuals diagnosed with MCI-SCI (**e**). **f**, K-means clustering was performed within each diagnostic category on the basis of their disease probabilities, where the top two clusters were annotated.

The contour was drawn on the basis of Gaussian kernel density estimation with a threshold of 0.01. **g–i**, Distributions of variables vary across the control (**g**), AD (**h**) and ALS (**i**) clusters (see Extended Data Fig. 2 for further details on disease clusters). Age distributions in **g** are shown for control C1 ($n = 1,132$) and control C2 ($n = 407$) as box plots indicating the median (center), interquartile range (25th to 75th percentiles) and whiskers extending to $1.5 \times$ IQR. P values were FDR corrected.

potential for ProtAIDe-Dx to aid in the diagnosis of patients in early phases of impairment.

There were many cases distributed on the embedding into regions inconsistent with their clinical diagnosis, for example, AD cases distributed into stroke/TIA regions. This observation may result from failed model predictions, incorrect clinical diagnoses or conditions with overlapping molecular etiologies. Figure 2f shows contours onto the embeddings representing the highest density of cases for each diagnosis, as well as subclusters of case densities distributed outside of the primary density (Extended Data Fig. 2). A non-dominant cluster of healthy controls emerged at the intersection of shallow extremes of the AD and stroke/TIA regions and showed older age and higher rates of vascular/metabolic risk factors with worse cognition relative to the dominant cognitively unimpaired (CU) cluster (Fig. 2g and Supplementary Data 2). Two minor AD clusters also emerged, with one colocalizing in the stroke/TIA region and the other in the PD region, each showing distinct clinical profiles (Fig. 2h and Supplementary Data 2). Proteomically, the dominant AD cluster had higher abundance of proteins involved

in cell death, damage response and mitochondrial activity but lower abundance of proteins involved in immune and defense response, whereas both minor AD clusters showed decreased abundance of proteins associated with energetic metabolism (Supplementary Fig. 3a and Supplementary Data 3 and 4). Perhaps most interestingly, a minor ALS cluster emerged closer to the FTD region and showed higher rates of *C9orf72* mutations and MCI, with differential abundance patterns consistent with upregulation of proteins relating to cell death and downregulation of proteins relating to metabolism and immunity (Fig. 2i, Supplementary Figs. 3b,c and 4, and Supplementary Data 2–4). Differential abundance and characteristics for PD, stroke/TIA and FTD are provided in Supplementary Data 3 and 4. Additional details are provided in Supplementary Results 1.

Model interpretation highlights disease-specific networks and key discriminative proteins

While deep learning models are not trivial to interpret, understanding the underlying biological trends driving predictions made by

ProtAIDe-Dx is essential for clinical adoption and biological insight. We used a feature permutation approach at the inference stage²³ to identify the most discriminative proteins used by our model (Fig. 3a and Extended Data Fig. 3). Several expected and previously described proteins emerged from this analysis, such as NEFL for FTD, CPLX2, CLU and SMOC1 for AD, SUMF1 for PD and multiple NPTXR aptamers for multiple neurodegenerative diseases. Several additional proteins emerged as discriminative for different disorders with interesting and highly relevant links to brain pathology, resilience and function, described in detail in Supplementary Results 2. This list included SERPINF2, PRL, C3, GPT2 and HERC1 for PD; CNTFR, TNNT2, PMGNT1 and LRTM1 for ALS; HEY1, SERPINA1, IGF2R, MAEA and STC for both FTD and ALS; and DCPIB, METAP2 and RAN for TIA/stroke (Supplementary Results 2). Certain proteins also emerged with known relationships to drugs commonly prescribed for neurodegenerative diseases. ACHE was unsurprisingly strongly discriminative for AD, consistent with common ACHE-inhibitor treatment²⁴. KCNIP3 showed the strongest ALS signal²⁵, consistent with recent work linking KCNIP3 expression to treatment with riluzole²⁵, a common ALS treatment. Given these findings, we compiled a library of known associations between medications and discriminative proteins identified in this analysis (Supplementary Data 5). We mapped 52 different neurodegenerative or vascular drugs associated with 12 of our discriminative proteins, although 35/52 (67.3%) mapped specifically to ACHE. Perhaps the most interesting set of proteins are those that discriminated healthy controls from all other conditions, as they may inform candidate markers of general brain health and resilience. Several proteins with known relationships to brain function or cognitive reserve emerged from this group, including GLO1, TGFB1, VAT1, STX1A, PDE1A, IGF2 and OMG²⁶ (see Supplementary Results 2 for details).

Next, to better understand the ProtAIDe-Dx model, we probed the proteomic composition of the model's low dimensional embeddings. The embeddings should represent proteins that express unique non-linear relationships in relation to neurodegenerative and/or neurological conditions and therefore may represent isolated disease-relevant molecular networks or processes (Supplementary Data 6 and 7). While we expected these embeddings to represent processes stemming from multiple organs, brain-specific proteins were highly prevalent across all embeddings (Fig. 3b). We therefore tested for enrichment of specific neural cell types (Supplementary Data 9 and 10) and triangulated this with disease discrimination (Fig. 3c) and biomarker associations in an external sample (Fig. 4c and Supplementary Data 8). We found evidence that embedding Z2 may represent neuronal functional decline, reflecting reduced resilience and synaptic dysregulation that contribute to cognitive impairment across aging and neurodegeneration. Meanwhile embedding Z23 may capture glial vulnerability pathways that link aging and sex to increased neurodegenerative disease risk (Supplementary Results 2). Other embeddings emerged with interpretable annotations helping to understand proteomic underpinnings of specific neurodegenerative diagnosis (Fig. 3c, Supplementary Data 6 and 7, and Supplementary Results 2).

Out-of-sample generalization and validation by biomarkers of disease-specific neuropathology

We next wished to test how well ProtAIDe-Dx generalized to new datasets. Out-of-sample generalization is challenging given that within-sample performances tend to be optimistic. In a leave-one-site-out cross-validation approach, ProtAIDe-Dx continued to significantly outperform the Random Forest and XGBoost baselines and showed slight improvements over TabPFN in FTD and stroke prediction (Fig. 4a, Supplementary Table 4 and Supplementary Data 11). The ensemble model did not aid generalization performance. However, all models saw a substantial dip in performance in both balanced accuracy score and AUC compared with whole-sample cross-validation performance, probably driven by high variation in effect sizes of individual proteins across

sites (Extended Data Fig. 4a,b). These performance deficits extended to additional generalization experiments (Supplementary Table 5 and Extended Data Fig. 4b). ProtAIDe-Dx's performance was partially recovered by using finetuning (Methods).

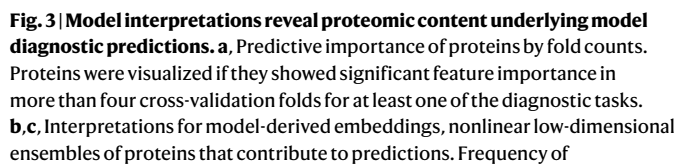
We next applied ProtAIDe-Dx to the BioFINDER-2 dataset ($N=1,786$), a real-life memory clinic dataset with biomarker-supported diagnosis. Note that, while BioFINDER-2 is part of GNPC, this site was excluded from model fitting. Diagnostic performance was close to the median of general leave-one-site-out performance (Fig. 4b). Predicted probabilities across diagnosis revealed expected trends (Extended Data Fig. 5); PD probabilities were elevated in patients with PD but also in dementia with Lewy bodies (DLB) cases, and stroke/TIA probability was elevated in patients with vascular dementia. Subsequently, we tested whether these probabilities correlated with disease-specific biomarkers within disease groups (Fig. 4d, Extended Data Fig. 6 and Supplementary Data 12). Among CU individuals, the model-derived probability of being CU was lower for participants expressing AD, Lewy body or neurovascular pathology (Fig. 4d). This indicates that some 'false positive' results from ProtAIDe-Dx may correctly identify underlying preclinical neuropathology (Supplementary Table 6). Similarly, AD probabilities were higher in non-AD cases with comorbid A β and Tau pathology (Fig. 4d and Supplementary Fig. 6), and higher stroke/TIA probabilities were associated with greater white matter hyperintensity (WMH) burden in both impaired and unimpaired individuals (Fig. 4d). PD probabilities did not show a significant relationship with presence of Lewy body pathology (as measured using CSF α -synuclein seed amplification assays (SAA)) but were correlated with symptom progression (Unified Parkinson's Disease Rating Scale (UPDRS)) in PD cases (Fig. 4d).

Proteomics provide additive information to diagnosis in a memory clinic sample

For models such as ProtAIDe-Dx to be translated to real-life clinical applications, it is important to show evidence that they provide additive value in these settings. Using the same external BioFINDER-2 dataset, we fit a series of models seeking to identify primary etiological diagnosis using a baseline model of only age and sex (model 0), using just ProtAIDe-Dx and demographics (model 1), using accessible clinical markers (model 2: demographics, MMSE, mean cortical thickness of AD-signature meta-region of interest (ROI) (ADSignCT²⁷), plasma p-tau217 and plasma NEFL) and using all of these markers together (model 3). The final model incorporating ProtAIDe-Dx with common clinical biomarkers (model 3) achieved significantly higher BCA than the model using only common clinical markers (model 2), especially adding value in diagnosis of non-AD dementias (Fig. 5a and Supplementary Tables 7 and 8). We also performed an analysis distinguishing subtypes of Lewy body disease (Supplementary Results 3).

Despite being trained exclusively on baseline visits, ProtAIDe-Dx demonstrated the ability to differentiate longitudinal rates of cognitive decline. Baseline clinical diagnoses did not distinguish rates of decline after FDR correction in the GNPC ($P > 0.05$ across all diagnostic groups; Extended Data Fig. 7 and Supplementary Table 9). However, baseline predicted diagnosis from ProtAIDe-Dx did significantly stratify decline trajectories (FDR-corrected P value < 0.05 across all prediction groups; Fig. 5b(1) and Supplementary Table 10), independent of clinical diagnosis. Similarly, in the BioFINDER-2 dataset, patients with MCI predicted as AD by ProtAIDe-Dx declined more rapidly than patients with MCI predicted as control (FDR-corrected P value of 0.0015; Fig. 5b(2) and Supplementary Table 11).

The probability outputs from ProtAIDe-Dx provide clinically interpretable indicators of biomarker status. In the external BioFINDER-2 dataset, we found that when the AD probability exceeded 0.9, most patients were tau positive by Tau-PET (first graph) or CSF p-tau217 (second graph) and showed cortical thickness below the diagnostic threshold (Fig. 5c, third graph). Likewise, when the stroke probability exceeded 0.7, most patients exhibited elevated white matter



1857

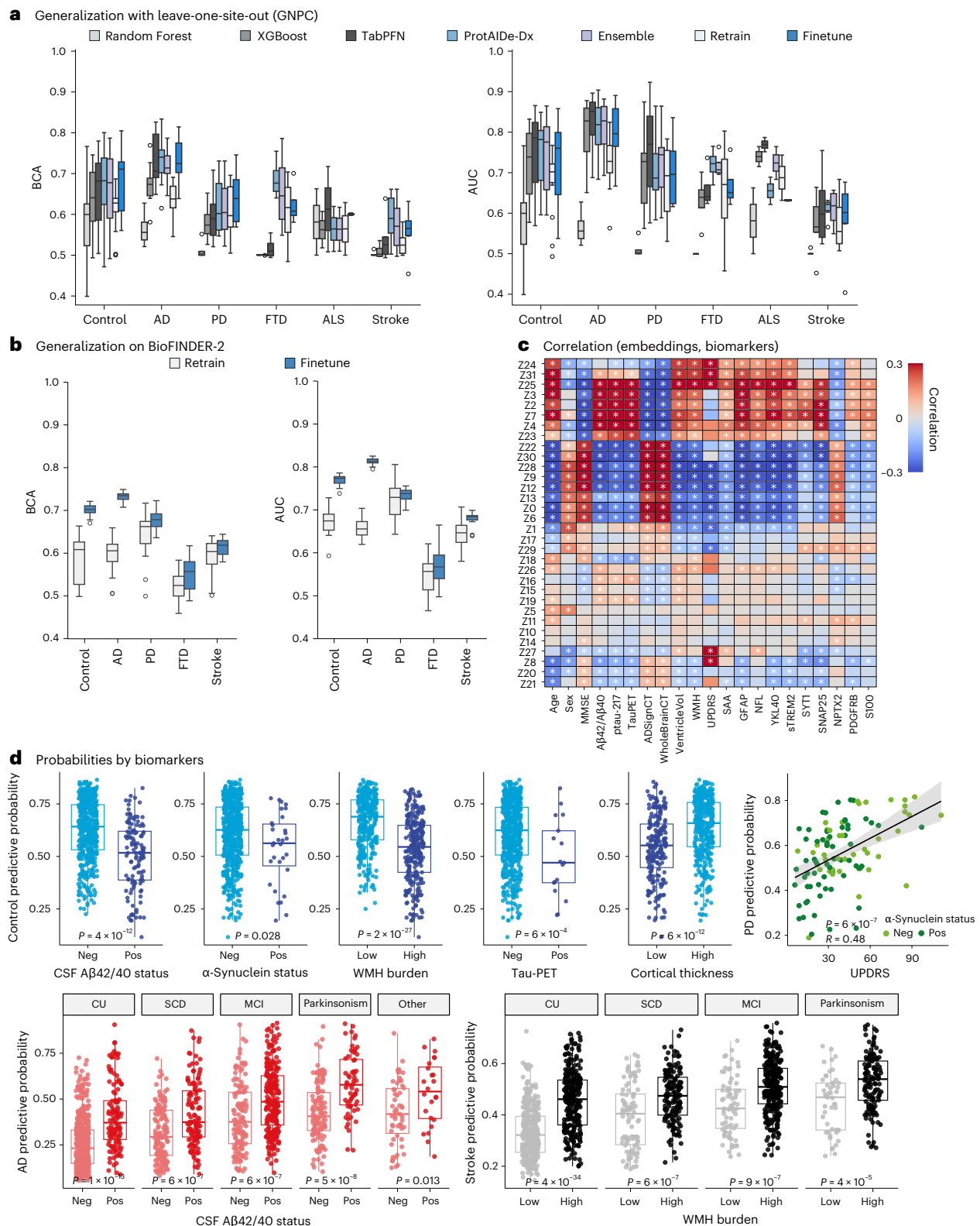


Fig. 4 | Model validation in the external BioFINDER-2 cohort. a, Leave-one-site-out BCAs (left) and AUCs (right) of ProtAIDe-Dx on GNPC across 14 test sites. Retrain: LR model was retrained on $K = 100$ participants' proteins. FineTune: LR model was retrained on $K = 100$ participants' proteomics embeddings. Two-sided corrected resample t -tests were applied; P values were FDR corrected. **b**, Model generalization performance in BioFINDER-2 cohort across 20 random K -shot repeats. Left: BCA. Right: AUC. Two-sided corrected resample t -tests were applied and P values were FDR corrected. **c**, Correlations between embeddings and biomarkers. A white star indicates a significant correlation after FDR

correction across all comparisons. **d**, Diagnostic predictive probabilities for control (blue), AD (red), parkinsonism (PD; green) and stroke/TIA (black) diagnoses, associated with different biomarkers, stratified by biomarker-confirmed clinical diagnosis of 1,786 BioFINDER-2 participants. Neg, negative; pos, positive. The participant number of each clinical group is presented in Supplementary Table 12. Independent t -tests were used to compare group differences, and P values were corrected for multiple comparisons using FDR. All box plots in this figure show the median, interquartile range (25th to 75th percentiles) and whiskers extending to $1.5 \times$ IQR.

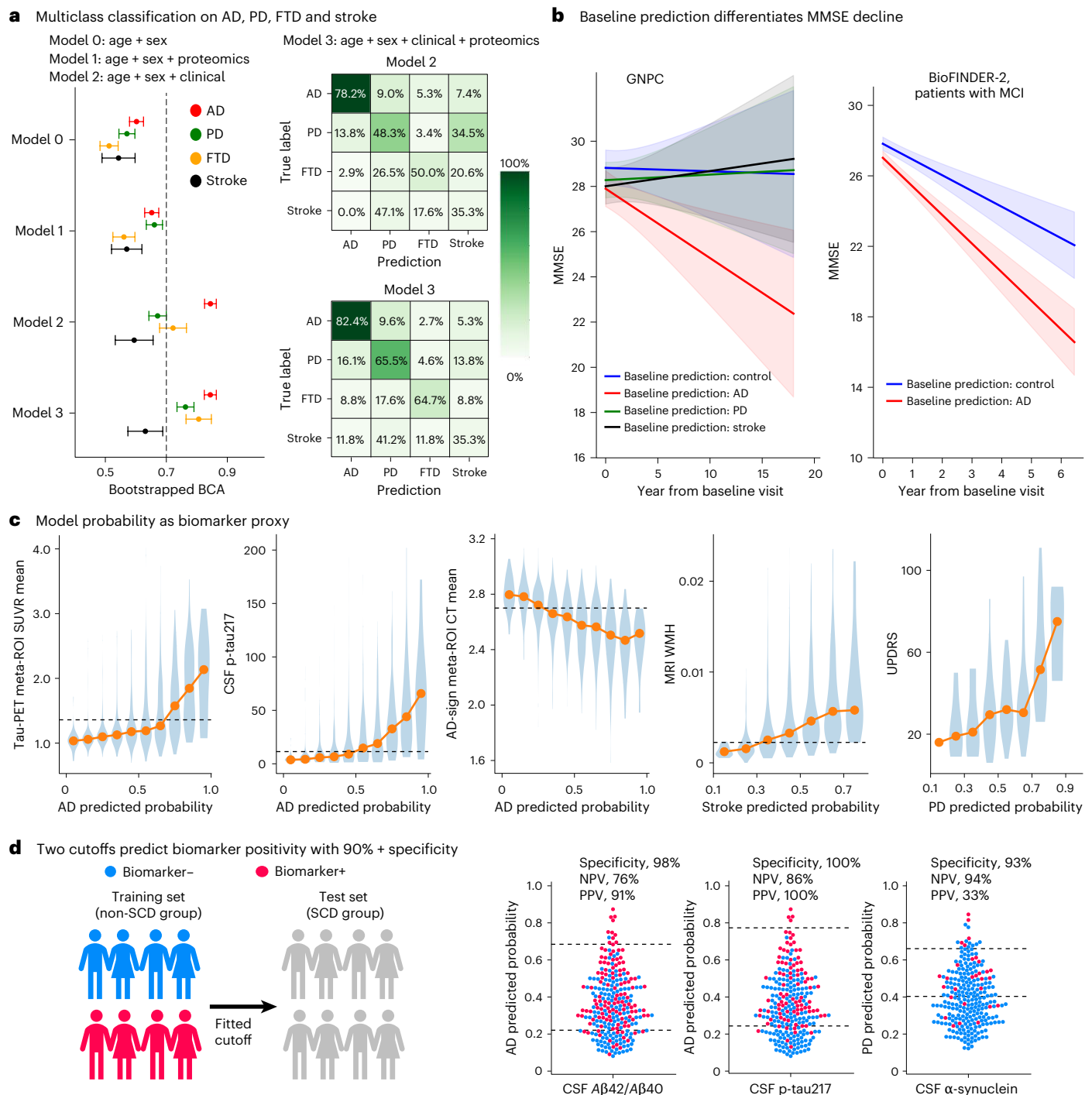


Fig. 5 | Clinical utility of ProtAIDe-Dx. a, Multi-class classification of patients with AD, PD, FTD and stroke using various models inclusive of demographics, ProtAIDe-Dx embeddings and/or common clinical biomarkers. Left: one-versus-rest BCA computed by 1,000 bootstraps on testing patients. Right: confusion matrices on testing patients for model 2 (top) and model 3 (bottom). Error bars represent the s.d. across 1,000 bootstrap resamples. **b**, Left: model-predicted baseline diagnoses differentiated longitudinal MMSE trajectories of GNPC patients irrespective of true baseline clinical diagnosis. Trajectories were modeled using linear mixed-effects model of MMSE with Age, Sex, Site, BaselineDx, BaselinePredictedDx, Year and the BaselinePredictedDx × YearInteraction, with subject-specific random effects for SubjectID.

Right: replication on BioFINDER-2 patients with MCI. Solid lines indicate linear mixed-effects model-predicted mean MMSE trajectories, and shaded regions represent 95% confidence intervals for the linear mixed-effects fixed-effects predictions. **c**, Biomarker distributions across predicted probability bins. CT, cortical thickness. **d**, Two-cutoff strategies for predicting biomarker positivity. Cutoffs were derived from non-SCD BioFINDER-2 participants ($n = 1,524$) to achieve 90% NPVs and PPVs, except for α-synuclein, where the PPV was set at 40% owing to sample validity constraints. These fit cutoffs were then applied to patients with SCD ($n = 263$) to estimate accuracy in this clinically relevant sample. Icons in **d** created in BioRender; An, L. <https://biorender.com/q2by4y5> (2026).

hyperintensities (fourth graph). A two-cutoff strategy derived on participants without SCD achieved >90% specificity and positive predictive value (PPV) in patients with SCD (Fig. 5d). This approach achieved 94%

negative predictive value (NPV) for detecting LBD-related biomarker positivity on the basis of CSF α-synuclein (see Supplementary Fig. 7 for models ensuring 50% coverage or maximizing PPVs and NPVs).

Case A

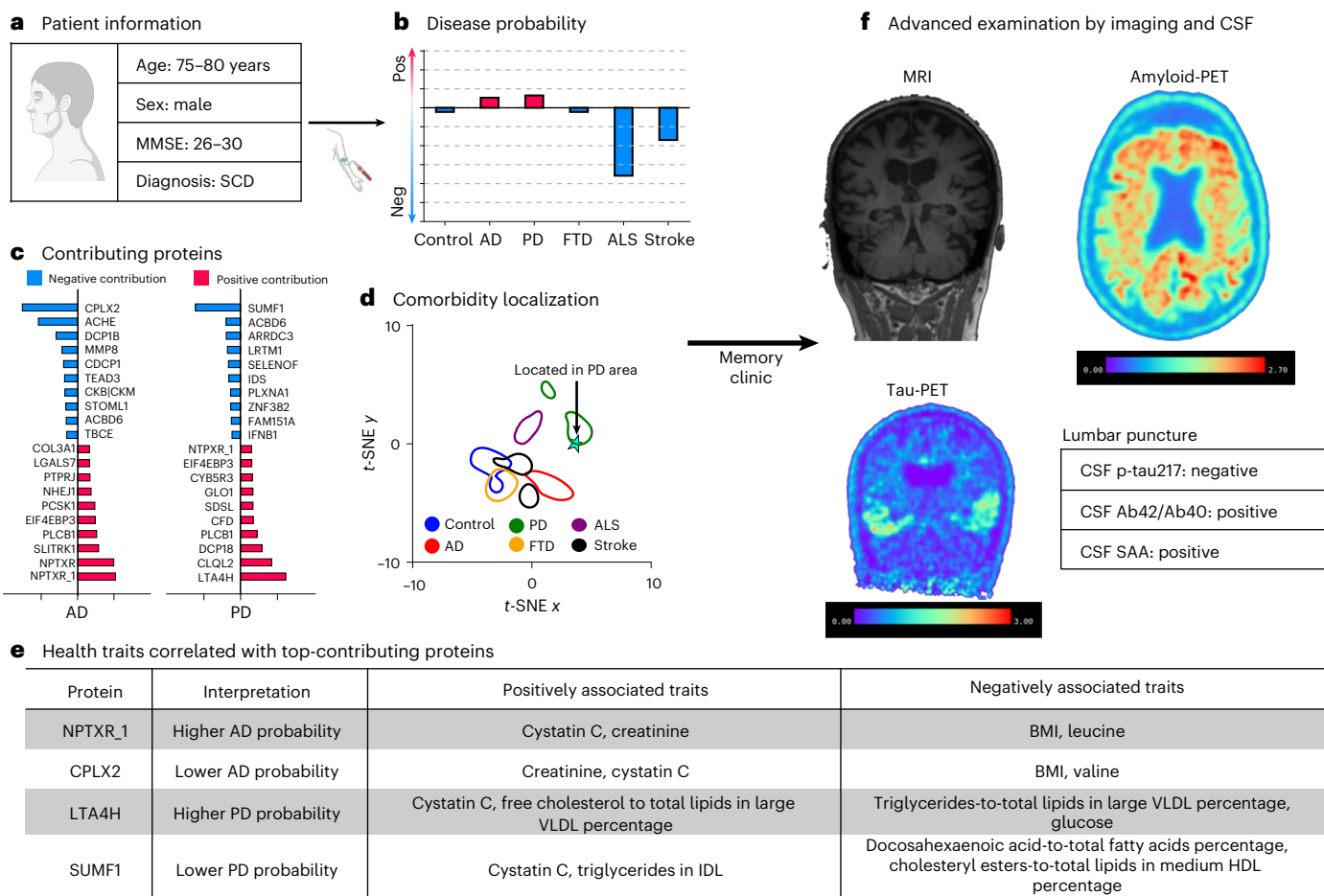


Fig. 6 | Individual neurodegeneration risk report (case A). **a**, Demographics and cognition score of one older participant who was diagnosed with SCD. **b**, ProtAIDe-Dx predicted this participant with higher probabilities of AD and PD. The probabilities were normalized to be centered at zero. **c**, Contributing proteins for the decision were computed on the basis of SHAP values. The top 10 positively and negatively contributing proteins were visualized. **d**, The location of this participant on the diagnostic probability map indicates his proteomic signature was consistent with typical patients with PD. **e**, Health traits correlated with top contributing proteins were listed, informing clinicians to pay attention to these traits. BMI, body mass index; HDL, high-density lipoprotein;

IDL, intermediate-density lipoprotein; VLDL, very low-density lipoprotein. **f**, A signature such as this would be suggestive of follow-up advanced imaging and CSF biomarker examination at a memory clinic to confirm the diagnosis. Here, AD pathology was confirmed by both a positive β -amyloid-PET scan and a positive tau-PET, as well as a positive CSF A β 42/A β 40 ratio, but not positive CSF p-tau217 levels, interestingly. The presence of underlying Lewy body pathology was confirmed through a positive CSF seed-amplification assay for α -synuclein pathology. Illustration in **a** created in BioRender; An, L. <https://biorender.com/q2by4y5> (2026).

Proof-of-concept diagnostic report by ProtAIDe-Dx

The clinical utility analysis shows the potential of ProtAIDe-Dx to provide significant additive information to clinical diagnostic workup. We therefore built a proof of concept for a diagnostic report using ProtAIDe-Dx (Fig. 6 and Extended Data Figs. 8 and 9). The report indicates diagnostic probability across all conditions (Fig. 6b) and includes a localization of the patient to the GNPC disease probability space (Figs. 2a and 6d). Importantly, the report uses model explanation technology to report which proteins contributed to this specific individual's prediction (Fig. 6c), allowing a clear biological explanation of the proteomic diagnosis. Physical traits linked to these patient-specific diagnostic proteins are also provided by programmatically accessing a library of protein–trait associations (Fig. 6e), providing possibility for lifestyle interventions or further explanation. We use three patient case examples to showcase the report. Case A (Fig. 6) was a 75–80-year-old man entering the memory clinic with subjective cognitive complaints but objectively intact cognition. ProtAIDe-Dx predicted underlying comorbid AD and Lewy body pathologies. This was confirmed by positron emission tomography (PET) scans showing cortical A β burden and

temporal tau pathology (AD), as well as positive cerebrospinal fluid (CSF) SAA indicating Lewy body pathology (Fig. 6f). Supplementary Results 4 provides information about the other case studies. While not all ProtAIDe-Dx probabilities were accurate and confirmed by biomarkers (Fig. 4b,d), these examples showcase the potential of ProtAIDe-Dx in adding explainable proteomic data informative toward neurodegenerative etiology to clinical workup.

Discussion

The rapid growth of dementia populations worldwide urges scalable and accessible biomarkers for neurodegenerative diseases. However, while promising biomarkers for certain diseases are on the horizon^{28–30}, many biomarkers under development are either invasive (CSF) or not yet scalable and are usually singular, requiring multiple tests to assess different pathologies. Therefore, a one-shot, multi-disease biomarker panel that is minimally invasive, economical and widely accessible is highly anticipated. This study presents an early attempt toward this goal by applying deep neural networks on a massive ($N = 17,187$) neurodegenerative disease sample and high-rank (7,595 proteins)

plasma proteomics data. Our proposed ProtAIDe-Dx model synthesized novel biological insight and exhibited clinical utility by mining neurodegeneration-related signals from high-rank proteomics data at multiple levels. Moreover, when generalized to an out-of-sample memory clinic dataset, ProtAIDe-Dx improved automated differential diagnosis accuracy beyond the capability of currently accessible clinical biomarkers. This study distinguishes itself from other neurodegenerative disease proteomics studies owing to its focus on making individual-level predictions on new cases, by training a model to recognize multiple copathologies rather than on differential diagnosis, and by rigorously avoiding leakage and overfitting and providing true generalization performance. However, despite providing complementary information capable of enhancing current clinical workup, our data suggest that contemporary high-throughput plasma proteomics assays alone cannot yet replace currently available clinical markers. Altogether, this study sets a benchmark for future proteomics studies by providing a baseline for predictive performance, a resource sorely needed in the biomedicine AI field³¹.

The performance of ProtAIDe-Dx is not sufficient at present to replace currently available clinical markers, although we demonstrate several examples of how it could be integrated into clinical practice as an accessible and cost-effective assistant. Disease-differentiation experiments indicated that ProtAIDe-Dx provides additive value beyond currently available clinical biomarkers, highlighting its clinical potential given the high prevalence of neurodegenerative copathologies in the aging population and the fact that affordable, minimally invasive biomarkers for PD and FTD are still under development^{29,30}. ProtAIDe-Dx's predictions to differentiate longitudinal rates of cognitive decline may offer a useful tool for clinical management, as individuals predicted as having AD tended to exhibit a faster decline irrespective of their current diagnostic status. These findings suggest that diagnostic labels may not fully reflect underlying disease progression, whereas ProtAIDe-Dx tends to capture more informative biological signals that could enhance clinical decision making. Probability outputs from ProtAIDe-Dx also provide clinically interpretable indicators of biomarker status, consistent with established approaches such as the Amyloid Probability Score 2 (APS2) developed by C2N Diagnostics. Higher predicted probabilities generally reflected greater biomarker burden, and the two-cutoffs approach achieved >90% specificity for determining biomarker positivity. Intermediate predicted probabilities may suggest early pathology and should prompt confirmatory assessment with other established biomarkers.

Owing to its diverse site composition and unstandardized data collection, the GNPC creates an excellent dataset for building and testing tools for translation to real-world clinical settings, where data curation and standardization rarely match that of most research datasets. Within the GNPC, ProtAIDe-Dx achieved impressive cross-validated diagnostic accuracy across diseases, especially for ALS and PD. However, this performance decreased in a leave-one-site-out validation setting and a site-to-site generalization setting that more closely simulates a true clinical translation, especially for disease with high imbalanced distribution across sites (for example, patients with ALS mainly come from one site). This indicates that, despite the large and diverse sample, site effects and overfitting remain a challenge for training generalizable models. This cannot necessarily be explained by the overcomplexity of our deep learning architecture, as ProtAIDe-Dx consistently outperformed the less complex Random Forest and XGBoost models. It is likely that advances in generalizable data harmonization³² or standardization³³ will be necessary to overcome these limitations.

In general, several factors must be considered when evaluating the diagnostic performance of ProtAIDe-Dx. One of these factors is the difficulty of the task. For instance, FTD is a highly heterogeneous disorder with multiple underlying pathologies and many different clinical presentations³⁴, and stroke/TIA is a stochastic event and only of many manifestations of neurovascular disease^{35,36}, and itself may often

go undiagnosed. However, one of the most important considerations in evaluating the performance of prediction models in neurological diseases relates to the accuracy of the original clinical diagnosis, which both forms the basis of training and serves as the standard for evaluation. Neurological and neurodegenerative diseases are notoriously difficult to diagnose³⁷, and in our particular case, models were trained and evaluated on the basis of clinical diagnoses that, for the most part, lacked biomarker confirmation. In addition, the clinical diagnosis criteria varied across sites, given that GNPC is a retrospective collection of multiple cohorts. Given that there are multiple criteria for the diagnosis of AD^{38,39} and multiple interpretations of those criteria, this issue will be present in any real-world application. In light of these limitations, many 'false' predictions from ProtAIDe-Dx may not represent incorrect diagnoses, considering the prevalence of asymptomatic disease, co-pathology and misdiagnosis. Along similar lines, we also observed patients with the same diagnosed clinical syndromes but who demonstrate distinct proteomic profiles. These findings are in line with a growing number of studies finding multiple proteomic or transcriptomic subtypes within the disease populations^{40,41}. Together, these observations demonstrate that there is a set of neurodegenerative signatures detectable in blood plasma that may be associated with differing clinical profiles or that may represent stratified responses to general neurological insult. In either case, a patient's underlying biological response to pathology may be just as relevant to treatment and prognosis as their clinical presentation. Future work should leverage longitudinal data to further explore the clinical meaningfulness of these proteomic disease subgroups.

Beyond producing diagnostic probabilities for multiple diseases, ProtAIDe-Dx also identified a compact set of top predictive proteins that significantly influenced model performance on out-of-sample data. While many previous studies have described lists of differentially abundant proteins associated with different diseases^{36,42–44}, predictive modeling is probably more clinically useful for making individual diagnoses⁴⁵. In our study, among the top proteins discriminating different conditions, those proteins that discriminated controls from all neurodegenerative conditions were perhaps the most interesting. GLO1, PDE11A and IGF2 have all been investigated as cognition enhancers in various aging models^{46,47} and may indeed be promising targets given that they also signal healthy aging in our large human cohort. Many of the proteins profiled were brain expressed and played a role in either memory (PDE11A and IGF2), inflammation (TGFB1), neurite regeneration (OMG) or endosomal/lysosomal systems (STX1A). OMG was a highly discriminative protein for controls specifically when differentiating them from not just dementia but also patients with MCI–SCI, suggesting its possible role in maintenance of cognition. Meanwhile, lower expression levels of NPTXR in neurodegenerative groups have been observed in multiple studies^{48,49}, highlighting synapse loss as a common feature of neurodegeneration⁵⁰. These proteins should continue to be investigated as markers of multi-cause neurological malaise, which might be useful in triaging or early diagnosis and monitoring. Further studies should continue to investigate how these proteins change over time in response to other outcomes of brain health.

In all, ProtAIDe-Dx showed great promise in disease diagnosis, probing disease heterogeneity, identifying novel proteins of interest and generating health- and disease-related signatures. However, there are still challenges that we wish to highlight to facilitate future studies exceeding the performance of ProtAIDe-Dx in real-world samples. A major challenge of ProtAIDe-Dx is that its diagnostic accuracy has not yet reached the level required for standalone clinical use. This limitation may be due to several factors. The performance of ProtAIDe-Dx was comparable to that of other studies using UK Biobank data^{51,52}, suggesting that there may be a performance ceiling of high-throughput plasma proteomics as biomarkers. Another factor might be a limited set of specific aptamers that target arbitrary protein conformations, which are also secreted or surface based and are detectable

in blood. This is a particular challenge for brain diseases, as many disease-relevant proteins are probably brain expressed and many of these do not cross the blood–brain barrier. In addition, plasma p-tau biomarkers have achieved much better performance discriminating AD than we found here^{7,53} owing to the identification of specific peptides and post-translational modifications directly related to disease pathology. To achieve similar improvements, mass spectrometry and/or other approaches may be needed for more comprehensive screening of peptides and protein fragments⁵⁴. Another challenge of ProtAIDe-Dx is its relatively poor site generalization performance, which hinders its applicability to new sites. Strong site effects in plasma proteomics⁵⁵ may limit generalization and could be mitigated by harmonization³². As discussed, the diagnostic labels that our models were trained on may not be sufficiently reliable owing to the lack of biomarker support and imperfect harmonization of diagnostic criteria across cohorts. To address this issue, integrating plasma proteomics studies conducted during life with neuropathological assessments at death may provide an eventual solution for enabling model training with ground truth diagnostic labels. Alternatively, future studies are recommended to incorporate more biomarker-confirmed samples to enable models to more accurately capture biologically meaningful signals for disease classification. Another challenge is presented by the potential confounding effects of various factors on protein levels. For example, medication use can significantly affect circulating protein levels, sometimes exceeding normal physiological ranges²⁴, and may therefore dominate model predictions. Hopefully, many of these challenges can be addressed or improved in future releases of GNPC¹⁸.

In summary, ProtAIDe-Dx represents a pioneering attempt toward the development of scalable, minimally invasive and multi-disease diagnostic tools for neurodegenerative diseases. Despite its promise, predictive proteomics as a field faces several challenges, including insufficient diagnostic accuracy for clinical deployment, limited generalizability across sites and potential confounding effects. Future work should address these limitations by incorporating more reliable diagnostic labels to improve accuracy, refining model architecture to enhance generalization and implementing strategies to mitigate confounding biases. Overall, ProtAIDe-Dx establishes a robust benchmark for AI-driven proteomics tools, paving the way for precision medicine in neurodegenerative diseases.

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/s41591-026-04303-y>.

References

- van Dyck, C. H. et al. Lecanemab in early Alzheimer's disease. *N. Engl. J. Med.* **388**, 9–21 (2023).
- Sims, J. R. et al. Donanemab in early symptomatic Alzheimer disease: the TRAILBLAZER-ALZ 2 randomized clinical trial. *J. Am. Med. Assoc.* **330**, 512–527 (2023).
- Mummery, C. J. et al. Tau-targeting antisense oligonucleotide MAPTRx in mild Alzheimer's disease: a phase 1b, randomized, placebo-controlled trial. *Nat. Med.* **29**, 1437–1447 (2023).
- Kirkeby, A. et al. Preclinical quality, safety, and efficacy of a human embryonic stem cell-derived product for the treatment of Parkinson's disease, STEM-PD. *Cell Stem Cell* **30**, 1299–1314 (2023).
- Boros, B. D., Schoch, K. M., Kreple, C. J. & Miller, T. M. Antisense oligonucleotides for the study and treatment of ALS. *Neurotherapeutics* **19**, 1145–1158 (2022).
- Bernstein, A. et al. Dementia assessment and management in primary care settings: a survey of current provider practices in the United States. *BMC Health Serv. Res.* **19**, 919 (2019).
- Palmqvist, S. et al. Blood biomarkers to detect Alzheimer disease in primary care and secondary care. *J. Am. Med. Assoc.* **332**, 1245–1257 (2024).
- Rabinovici, G. D. et al. Association of amyloid positron emission tomography with subsequent change in clinical management among Medicare beneficiaries with mild cognitive impairment or dementia. *J. Am. Med. Assoc.* **321**, 1286–1294 (2019).
- Robinson, J. L. et al. Pathological combinations in neurodegenerative disease are heterogeneous and disease-associated. *Brain* **146**, 2557–2569 (2023).
- Mehta, D., Jackson, R., Paul, G., Shi, J. & Sabbagh, M. Why do trials for Alzheimer's disease drugs keep failing? A discontinued drug perspective for 2010–2015. *Expert Opin. Investig. Drugs* **26**, 735–739 (2017).
- Landau, S. M. et al. Cohort heterogeneity and AD biomarkers in older adults without significant cognitive impairment: a comparison of US POINTER and ADNI. *Alzheimers Dement.* **19**, e081990 (2023).
- Palmqvist, S. et al. Cognitive effects of Lewy body pathology in clinically unimpaired individuals. *Nat. Med.* **29**, 1971–1978 (2023).
- Gaugler, J. E. et al. Characteristics of patients misdiagnosed with Alzheimer's disease and their medication use: an analysis of the NACC-UDS database. *BMC Geriatr.* **13**, 137 (2013).
- Sattlecker, M. et al. Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. *Alzheimers Dement.* **10**, 724–734 (2014).
- Shen, X. et al. Nonlinear dynamics of multi-omics profiles during human aging. *Nat. Aging* **4**, 1619–1634 (2024).
- Candia, J. et al. Variability of 7K and 11K SomaScan plasma proteomics assays. *J. Proteome Res.* **23**, 5531–5539 (2024).
- Zetterberg, H. & Burnham, S. C. Blood-based molecular biomarkers for Alzheimer's disease. *Mol. Brain* **12**, 26 (2019).
- Imam, F. et al. The Global Neurodegeneration Proteomics Consortium: biomarker and drug target discovery for common neurodegenerative diseases and aging. *Nat. Med.* **31**, 2556–2566 (2025).
- Iadecola, C. The pathobiology of vascular dementia. *Neuron* **80**, 844–866 (2013).
- Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (ACM, 2016).
- Hollmann, N. et al. Accurate predictions on small data with a tabular foundation model. *Nature* **637**, 319–326 (2025).
- He, H. & Garcia, E. A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **21**, 1263–1284 (2009).
- Mi, X., Zou, B., Zou, F. & Hu, J. Permutation-based identification of important biomarkers for complex diseases via machine learning models. *Nat. Commun.* **12**, 3008 (2021).
- Bolsewig, K. et al. Increased plasma DOPA decarboxylase levels in Lewy body disorders are driven by dopaminergic treatment. *Nat. Commun.* **16**, 1139 (2025).
- Dergai, O. et al. Skeletal muscle biomarkers of amyotrophic lateral sclerosis: a large-scale, multi-cohort proteomic study. *Ann. Neurol.* **99**, 393–407 (2025).
- Duggan, M. R. et al. OMG! A proteomic determinant of neurodegenerative resiliency. *Mol. Neurodegener.* **21**, 9 (2026).
- Jack, C. R. Jr et al. Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimers Dement.* **13**, 205–216 (2017).
- Kuang, Y. et al. A skin-specific α -synuclein seeding amplification assay for diagnosing Parkinson's disease. *npj Parkinsons Dis.* **10**, 129 (2024).
- Chatterjee, M. et al. Plasma extracellular vesicle tau and TDP-43 as diagnostic biomarkers in FTD and ALS. *Nat. Med.* **30**, 1771–1783 (2024).

30. Siderowf, A. et al. Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using α -synuclein seed amplification: a cross-sectional study. *Lancet Neurol.* **22**, 407–417 (2023).
31. Mahmood, F. A benchmarking crisis in biomedical machine learning. *Nat. Med.* **31**, 1060 (2025).
32. An, L. et al. DeepResBat: deep residual batch harmonization accounting for covariate distribution differences. *Med. Image Anal.* **99**, 103354 (2025).
33. Bethlehem, R. et al. Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
34. Josephs, K. A. et al. Neuropathological background of phenotypical variability in frontotemporal dementia. *Acta Neuropathol.* **122**, 137–153 (2011).
35. Lorenzini, L. et al. Association of vascular risk factors and cerebrovascular pathology with Alzheimer disease pathologic changes in individuals without dementia. *Neurology* **103**, e209801 (2024).
36. Hristovska, I. et al. Identification of distinct and shared biomarker panels in different manifestations of cerebral small-vessel disease through proteomic profiling. *Nat. Aging* <https://doi.org/10.1038/s43587-026-01081-7> (2026).
37. Bradford, A., Kunik, M. E., Schulz, P., Williams, S. P. & Singh, H. Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors. *Alzheimer Dis. Assoc. Disord.* **23**, 306–314 (2009).
38. 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).
39. Dubois, B. et al. Clinical diagnosis of Alzheimer's disease: recommendations of the International Working Group. *Lancet Neurol.* **20**, 484–496 (2021).
40. Tijms, B. M. et al. Cerebrospinal fluid proteomics in patients with Alzheimer's disease reveals five molecular subtypes with distinct genetic risk profiles. *Nat. Aging* **4**, 33–47 (2024).
41. Iturria-Medina, Y. et al. Unified epigenomic, transcriptomic, proteomic, and metabolomic taxonomy of Alzheimer's disease progression and heterogeneity. *Sci. Adv.* **8**, eabo6764 (2022).
42. Deng, Y.-T. et al. Atlas of the plasma proteome in health and disease in 53,026 adults. *Cell* **188**, 253–271 (2025).
43. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
44. Pichet Binette, A. et al. Proteomic changes in Alzheimer's disease associated with progressive A β plaque and tau tangle pathologies. *Nat. Neurosci.* **27**, 1880–1891 (2024).
45. Bzdok, D., Varoquaux, G. & Steyerberg, E. W. Prediction, not association, paves the road to precision medicine. *JAMA Psychiatry* **78**, 127–128 (2021).
46. Pascual-Lucas, M. et al. Insulin-like growth factor 2 reverses memory and synaptic deficits in APP transgenic mice. *EMBO Mol. Med.* **6**, 1246–1262 (2014).
47. Samanta, S. et al. Synaptic mitochondria glycation contributes to mitochondrial stress and cognitive dysfunction. *Brain* **148**, 262–275 (2025).
48. Oh, H. S.-H. et al. A cerebrospinal fluid synaptic protein biomarker for prediction of cognitive resilience versus decline in Alzheimer's disease. *Nat. Med.* <https://doi.org/10.1038/s41591-025-03565-2> (2025).
49. Nilsson, J. et al. Cerebrospinal fluid biomarker panel for synaptic dysfunction in a broad spectrum of neurodegenerative diseases. *Brain* **147**, 2414–2427 (2024).
50. 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).
51. Guo, Y. et al. Plasma proteomic profiles predict future dementia in healthy adults. *Nat. Aging* **4**, 247–260 (2024).
52. Gan, Y.-H. et al. Large-scale proteomic analyses of incident Parkinson's disease reveal new pathophysiological insights and potential biomarkers. *Nat. Aging* **5**, 642–657 (2025).
53. Mielke, M. M. et al. Performance of plasma phosphorylated tau 181 and 217 in the community. *Nat. Med.* **28**, 1398–1405 (2022).
54. Barthélemy, N. R. et al. Highly accurate blood test for Alzheimer's disease is similar or superior to clinical cerebrospinal fluid tests. *Nat. Med.* **30**, 1085–1095 (2024).
55. Halvey, P. et al. Variable blood processing procedures contribute to plasma proteomic variability. *Clin. Proteomics* **18**, 5 (2021).
56. Haufe, S. et al. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* **87**, 96–110 (2014).

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 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

© The Author(s) 2026

Lijun An¹✉, **Alexa Pichet Binette**^{2,3,4}, **Ines Hristovska**², **Gabriele Vilkaite**¹, **Yu Xiao**¹, **Romina Zendehdel**¹, **Zijian Dong**^{5,6}, **Bart Smets**⁷, **Rowan Saloner**⁸, **Shinya Tasaki**⁹, **Ying Xu**^{10,11}, **Varsha Krish**¹², **Farhad Imam**¹², **Shorena Janelidze**¹², **Danielle van Westen**^{13,14}, **The Global Neurodegenerative Proteomics Consortium (GNPC)***, **Erik Stomrud**^{2,15}, **Christopher D. Whelan**^{16,17}, **Sebastian Palmqvist**^{12,15}, **Rik Ossenkoppele**^{2,18,19}, **Niklas Mattsson-Carlgrén**^{2,15}, **Oskar Hansson**¹² & **Jacob W. Vogel**¹✉

¹Department of Clinical Sciences Malmö, SciLifeLab, Lund University, Lund, Sweden. ²Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Lund, Sweden. ³Department of Physiology and Pharmacology, Université de Montréal, Montreal, Quebec, Canada. ⁴Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montreal, Quebec, Canada. ⁵Centre for Sleep and Cognition and Centre for Translational MR Research, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ⁶Department of Electrical and Computer Engineering, National University of Singapore, Singapore, Singapore. ⁷Johnson & Johnson, Beerse, Belgium. ⁸Memory and Aging Center, Department of Neurology, Weill Institute for Neurosciences, University of California, San Francisco, CA, USA. ⁹Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. ¹⁰Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. ¹¹NeuroGenomics and Informatics

Center, Washington University School of Medicine, St. Louis, MO, USA. ¹²Gates Ventures, Seattle, WA, USA. ¹³Department of Diagnostic Radiology, Clinical Sciences, Lund University, Lund, Sweden. ¹⁴Image and Function, Skåne University Hospital, Lund, Sweden. ¹⁵Memory Clinic, Skåne University Hospital, Malmö, Sweden. ¹⁶RCSI University of Medicine and Health Sciences, Royal College of Surgeons in Ireland, Dublin, Ireland. ¹⁷Johnson & Johnson, Cambridge, MA, USA. ¹⁸Amsterdam Neuroscience, Neurodegeneration, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. ¹⁹Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC Location VUmc, Amsterdam, the Netherlands. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: lijun.an@med.lu.se; jacob.vogel@med.lu.se

The Global Neurodegenerative Proteomics Consortium (GNPC)

Lijun An¹, Alexa Pichet Binette^{2,3,4}, Bart Smets⁷, Rowan Saloner⁸, Shinya Tasaki⁹, Ying Xu^{10,11}, Varsha Krish¹², Farhad Imam¹², Niklas Mattsson-Carlsson^{2,15}, Oskar Hansson² & Jacob W. Vogel¹

A full list of members and their affiliations appears in the Supplementary Information.

Methods

Datasets

GNPC. The GNPC (<https://www.neuroproteome.org/>), launched in 2023, is a mega consortium combining multiple dementia and population cohorts, including healthy aging, AD, PD, FTD, ALS and stroke/TIA¹⁸. All cohorts and data were anonymized. Ethics approval for each individual cohort was obtained from their respective institutional review boards. All participating cohorts confirmed that informed consent was obtained from all individuals contributing clinical and generated biosample data before contributing data to the GNPC. The latest GNPC v1.3MS release collected 20,532 participants from 22 contributors (sites), 3,950 of whom had longitudinal visits. We used baseline visit proteomics data for model development.

In this study, we selected 17,187 participants on the basis of SomaLogic 7k proteomics availability. Site U's proteomics data were from serum but we kept site U to maximize the sample size and learn modality-agnostic signatures. For the 9,708 participants that were neither diagnosed as control nor with any of the five above mentioned diseases, we visualized the distribution of these participants (Supplementary Fig. 8). On the basis of MMSE and CDR, we mapped participants with good cognition (MMSE ≥ 26 or CDR of 0) as 'control' and participants with MCI ($20 \leq \text{MMSE} \leq 25$ or CDR of 0.5) as 'MCI-SCI'. Among the remaining 1,606 participants, 1,062 participants with poor cognition (MMSE ≤ 10 or CDR ≥ 1) were labeled as 'ComputedDementia', and 542 participants without valid MMSE or CDR were labeled as 'Unknown'. Supplementary Table 1 presents the demographics and cognition (MMSE) distribution of 17,187 participants by site, and the corresponding distribution of clinical diagnoses is presented in Supplementary Table 2. Race and ethnicity information is shown in Supplementary Fig. 9. It is noted that diagnoses of some sites were confirmed by biomarkers but not for most sites. In the following analysis, the patients with either stroke or TIA were labeled as 'Stroke' in figures for ease of visualizations but are referred to as 'stroke/TIA' in the main text.

All GNPC blood samples were shipped separately by each contributor and were analyzed by SomaLogic, coordinated by Gates Ventures. SomaLogic applied slow off-rate modified aptamer (SOMAmmer) technology, in which chemically modified nucleotides enable high-specificity and high-affinity binding to target proteins. The resulting proteomic data were standardized, normalized and calibrated, with protein abundances reported in relative fluorescent units. Before integration into the GNPC cohort dataset, aptamers were mapped to UniProt¹⁸. We removed the outlier values for each protein that exceeded six standard deviations (s.d.).

BioFINDER-2 cohort. The Swedish BioFINDER-2 dataset (<https://biofinder.se/two/>, NCT03174938) is a prospective cohort in the south of Sweden spanning the full continuum of AD as well as including patients with non-AD neurodegenerative diseases. Participants are deeply phenotyped, including clinical assessment, CSF/blood sampling, PET and magnetic resonance imaging (MRI) imaging data. All studies were approved by the Institutional Review Board of Lund University and written informed consent or assent was obtained from all participants or their legally authorized representative. BioFINDER-2 is a participating site in GNPC but, in a subanalysis aiming to evaluate the disease probability predicted by the different models in relation to key markers of AD and neurodegenerative diseases, we focused specifically on this cohort. For the BioFINDER-2 subanalysis, we selected 1,786 participants with plasma SomaLogic 7k proteomics data. The demographics and cognition distribution of 1,786 participants are presented in Supplementary Table 12.

We grouped participants into six different groups: CU, SCD, MCI, AD, parkinsonism and other diseases. The BioFINDER-2 dataset inclusion and exclusion criteria have been described in detail previously⁵⁷. In brief, CU participants needed to have an MMSE score of

at least 27 (if <66 years old) or 26 (if ≥ 66 years old) and no signs of cognitive symptoms as assessed by physicians specialized in cognitive disorders. Participants with SCD or MCI were all referred to a memory clinic owing to cognitive symptoms, had an MMSE score between 24 and 30 and did not fulfill criteria for any dementia according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). Participants were classified as MCI if they performed at least 1.5 s.d. below the normative score on at least one cognitive domain from an extensive neuropsychological test battery¹², while participants with SCD performed better than -1.5 s.d. Patients with dementia fulfilled the DSM-5 criteria for dementia, and all patients with AD dementia were A β -positive (based on CSF A β 42/A β 40). Patients with non-AD neurodegenerative diseases were also included in the cohort. Patients with PD, DLB and atypical parkinsonism disorders formed the group 'parkinsonism'. Patients with other diseases were grouped together (labeled 'Other') and included patients with FTD spectrum disorders, vascular dementia and one patient with etiology not otherwise specified. Clinical diagnosis of AD dementia or other neurodegenerative diseases was determined by experienced clinicians.

To benchmark generalization performances on BioFINDER-2 (Fig. 4b) against leave-one-site-out on GNPC, we performed additional diagnostic grouping to match GNPC on the basis of clinical diagnosis and biomarkers of BioFINDER-2. Participants with a clinical diagnosis of normal were labeled as control; patients with an abnormal CSF A β 42/A β 40 ratio and a clinical diagnosis of AD were labeled as AD; patients with a clinical diagnosis of PD, DLB or parkinsonism (not otherwise specified) were labeled as parkinsonism disease (PD); patients with a clinical diagnosis of behavioral variant FTD, semantic variant of primary progressive aphasia, FTD (not otherwise specified) or an SCD with a MAPT mutation were labeled as FTD; and patients with infarcts were labeled as stroke. It is noted that stroke is not exclusive to other diagnosis categories. For example, a patient with AD might also be diagnosed with stroke. In summary, we obtained 609 control participants, 261 patients with AD, 135 patients with PD, 43 patients with FTD and 117 patients with stroke.

Biomarkers of interest in BioFINDER-2. Multiple biomarkers and the MMSE as a global measure of cognition were investigated in BioFINDER-2. The AD biomarkers were A β status (based on CSF A β 42/A β 40) and tau-PET standardized uptake value ratio (SUVR) in a temporal meta-ROI (tracer ¹⁸F-RO948)⁵⁸. The positivity of CSF p-tau217 was determined at a cutoff of 11.42 pg ml⁻¹ (ref. 12). Structural MRI markers of interest were the cortical thickness in an AD signature composed of temporal lobe regions²⁷, whole-brain cortical thickness and ventricular volume (average of lateral ventricle volume in both hemispheres divided by total intracranial volume). T1-weighted MRI was processed with FreeSurfer v6.0. White matter hyperintensity burden (divided by total intracranial volume) was measured on the basis of fluid-attenuated inversion recovery and T1 images processed with Sequence Adaptive Multimodal Segmentation (SAMSEG) tool from FreeSurfer v7.1. α -Synuclein status was available from seed amplification assay performed in the CSF, as described previously¹². For a subset of participants in the parkinsonism group ($n = 100$), the UPDRS was also available.

Model cross-validation and generalization

After diagnostic label mapping, 11,803 participants were labeled as either control or one of the five aforementioned diseases, 662 participants diagnosed with AD but exhibiting healthy cognitive scores (MMSE ≥ 26) were labeled as 'HealthyAD', 3,116 participants were labeled as 'MCI-SCI', 1,064 were labeled as 'ComputedDementia' with poor cognitions (MMSE <19 or CDR ≥ 1) and 542 were labeled as 'Unknown'. The 11,803 participants in the control and disease groups were used as the development set for model training and evaluation,

including tenfold cross-validation and leave-one-site-out procedures. The ‘HealthyAD’, ‘MCI–SCI’, ‘ComputedDementia’ and ‘Unknown’ groups were held out as additional test sets. We also conducted a supplementary experiment by including patients with MCI–SCI in model development (Extended Data Fig. 3).

We conducted a tenfold cross-validation procedure to evaluate the performances of ProtAIDe-Dx models on GNPC. Participants from each site were evenly split into ten folds, so each fold contained data from all sites. A 9–1–1 train–validation–test split was used, with 8 of 10 folds as a training split, one fold as a validation split and the remaining one as a test split. This process was repeated ten times, with each fold serving as the test split once.

To test the site generalization performance of the proposed ProtAIDe-Dx models, a leave-one-contributor-out scheme was used. Data from one site were reserved as the test split, while the data from the remaining contributors served as a training–validation split.

The sites used as test sets were selected on the basis of the following criteria: for any of the six conditions, (1) this site had 200 or more participants with non-missing diagnoses and (2) this site had at least five participants with minor diagnosis categories for this condition. In total, 14 testing sites fit this criteria. Detailed information is presented in Supplementary Table 13. Notably, the BioFINDER-2 cohort was part of GNPC; therefore, we excluded BioFINDER-2 participants when training and tuning the ProtAIDe-Dx model for testing on BioFINDER-2.

Machine learning model development

Feature selection. A feature selection procedure was used to reduce the number of input proteins. On each training and validation split, we conducted both GLM association analysis and XGBoost²⁰ predictive analysis to select informative proteins as input features. As GLM does not require any hyperparameters to set, we merged training and validation participants together to run the GLM association analysis for each of the six conditions and each of the 7,595 SomaLogic 7k proteins following

$$\text{Protein} \sim \text{Condition} + \text{Age} + \text{Sex} + \text{AverageProteinLevel}. \quad (1)$$

AverageProteinLevel is the average protein level across all 7k proteins to control for individual expression differences⁵⁹.

After running 7,595 GLMs for each target, we first selected proteins with fold change of beta Condition >2 or <0.5 and then picked the top 5 proteins with the smallest corrected *P* values of beta Condition.

We also ran XGBoost predictive models on each training and validation split to select predictive proteins. We merged training and validation participants together and resplit these participants into ten subfolds, nine used for training models and one as validation split to tune model hyperparameters. This procedure was repeated ten times to predict each condition, with each subfold serving as the validation split once. After all ten models for each condition were trained, proteins internally selected for building models by all ten XGBoost models were kept.

In this way, the number of input proteins was reduced from 7,595 to 200–300 proteins as input features, which vary across each train–validation–test split. In total, 738 proteins were kept across ten train–validation–test splits. This feature selection procedure was performed for both cross-validation and leave-one-contributor-out experiments.

Classification metrics. Accuracy is an inappropriate and potentially misleading metric given the imbalanced distribution across the six conditions. We therefore chose BCA as classification metric. We also reported AUC scores for reference.

Random Forest as baseline approach. Random Forest models were used as baseline machine learning approaches to classify binary targets separately. We therefore trained six Random Forest models for

classifying six binary conditions. The input proteins to Random Forest models were z-normalized by mean and s.d. computed from the training split. To get optimal validation prediction accuracies, a grid search on hyperparameters, including maximum tree depth, number of features for best split and probability threshold, was performed on the validation split. The Random Forest model trained with optimal hyperparameters was then applied to the test split. This procedure was repeated for each training–validation–test split, including cross-validation and leave-one-site-out.

XGBoost as baseline approach. XGBoost models were used as baseline machine learning approaches to classify binary targets separately. We therefore trained six XGBoost models for classifying six binary conditions. The input proteins to XGBoost models were z-normalized by mean and s.d. computed from the training split. To get optimal validation prediction accuracies, a grid search on hyperparameters, including maximum tree depth, subsampling of training data and probability threshold, was performed on the validation split. The XGBoost model trained with optimal hyperparameters was then applied to the test split. This procedure was repeated for each training–validation–test split, including cross-validation and leave-one-site-out.

TabPFN as baseline approach. TabPFN models were used as baseline deep learning approaches to classify binary targets separately. We therefore fit six TabPFN models on the basis of pretrained TabPFN classifiers for classifying six binary conditions. We performed several additional postprocessing steps before proteomics data were input into deep learning models, which were the same for both TabPFN and ProtAIDe-Dx. First, each participant’s proteomics values were normalized by their averaged protein levels⁵⁹. Second, we fit a 10-nearest neighbor data imputer onto the training split to impute missing protein entries. Third, a Gaussian rank normalizer⁶⁰ was fit on the training split to ensure proteins followed a normal distribution. The optimal validation probability threshold was selected on the basis of best F1 scores on the validation set. The TabPFN model fit with an optimal probability threshold was then applied to the test split. This procedure was repeated for each training–validation–test split, including cross-validation and leave-one-site-out.

ProtAIDe-Dx. The ProtAIDe-Dx models were implemented as multiple layer perceptron-based networks. Input proteins were fed into multiple multi-layer neural networks to classify six binary conditions jointly. The key consideration of choosing a multi-task over multi-class approach for ProtAIDe-Dx is that diagnostic labels are often incomplete in GNPC, as the dataset was aggregated from multiple cohorts with varying research objectives. For example, a participant diagnosed with AD may not have been formally assessed for PD, resulting in missing labels for certain conditions. In a multi-class classification framework using one-hot label vectors across six diagnostic categories, many vectors would contain missing values, necessitating either imputation or exclusion of these samples, both of which would usually be suboptimal. By contrast, a multi-task learning framework maximizes sample utilization, as each task is trained independently on subjects with labels available for that specific condition.

The imbalanced distribution of the six conditions may bias the model toward the majority class and lead to poor generalization if the loss function is not carefully designed. The loss function of the proposed ProtAIDe-Dx model is a weighted combination of binary cross-entropy loss L_{BCE} and multi-class rank loss⁶¹ L_{RL} . Label smoothing mitigates overfitting to potentially noisy clinical annotations by calibrating model confidence, while rank loss enhances robustness to distributional skew by optimizing relative sample rankings rather than absolute probabilities. This combined approach

ensures improved generalization across all conditions despite uneven representation.

$$L_{\text{BCE}} = -\frac{1}{N} \sum_{k=1}^N \sum_{i=1}^6 \left[(y_{k,i}(1-\alpha) + \frac{\alpha}{2}) \log \hat{y}_{k,i} + ((1-y_{k,i})(1-\alpha) + \frac{\alpha}{2}) \log (1-\hat{y}_{k,i}) \right], \quad (2)$$

where N is the number of participants, k is the participant index, i is the index of conditions, $y_{k,i}$ is the true label for participant k and condition i , α is a hyperparameter label smoothing factor to set, and $\hat{y}_{k,i}$ is the model-predicted probability for the participant k and condition i .

The role of rank loss L_{RL} is to constrain the rank of predicted probabilities across conditions, enabling better joint learning of information across targets. For any two conditions i and j , the rank loss $L_{\text{RL}}^{i,j}$ is

$$L_{\text{RL}}^{i,j} = \frac{1}{N} \sum_{k=1}^N \max \left[0, (\hat{y}_{k,i} - \hat{y}_{k,j})(y_{k,j} - y_{k,i}) + \varepsilon \right], \quad (3)$$

where N is the number of participants, k is the participant index, i and j are indexes of two conditions, $\hat{y}_{k,i}$ and $\hat{y}_{k,j}$ are predicted participant k 's probabilities for conditions i and j , and ε is a hyperparameter set as 0.25.

The rank loss across all conditions is

$$L_{\text{RL}} = \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^6 \sum_{j=i+1}^6 L_{\text{RL}}^{i,j}. \quad (4)$$

Therefore, the overall loss function for the proposed ProtAIDe-Dx model is

$$L = L_{\text{BCE}} + \lambda L_{\text{RL}}, \quad (5)$$

where λ is a hyperparameter to control the weight of L_{RL} .

The probability threshold for each target was determined for the highest validation F1 score. We performed a hyperparameter search on the validation split to get the optimal validation BCA using Optuna⁶² with 50 trials. The optimal validation probability threshold was selected on the basis of the best F1 scores on the validation set. The search range of hyperparameters is presented in Supplementary Table 14, with optimal searched hyperparameters presented in Supplementary Table 15 (model evaluated on BioFINDER-2) and Supplementary Data 13 (all cross-validation and leave-one-site-out models). An illustration of ProtAIDe-Dx architecture is shown in Supplementary Fig. 10.

After model fitting, we estimated the model's overfitting by enabling dropout at the inference stage. For each held-out test split, we repeated the prediction 100 times via stochastic forward passes to obtain an empirical distribution of model outputs across 100 repeats. We then summarized this distribution as 'confidence intervals'. We interpreted interval width through the bias-variance tradeoff as a proxy for overfitting. An overfit model may achieve low bias by memorizing the training data but at the cost of high variance because its predictions are sensitive to small fluctuations in the training set. When applied to unseen test data, this instability yields greater dispersion across dropout repeats, producing wider intervals that reflect poor generalization.

We performed several additional postprocessing steps before proteomics data were input into ProtAIDe-Dx models. First, each participant's proteomics values were normalized by their averaged protein levels⁵⁹. Second, we fit a 10-nearest neighbor data imputer onto the training split to impute missing protein entries. Third, a Gaussian rank normalizer⁶⁰ was fit on the training split to ensure normalized proteomics are following a normal distribution.

Proposed ensemble of XGBoost and ProtAIDe-Dx. We conducted an ensemble approach for the developed XGBoost and ProtAIDe-Dx models. The ensemble approach's output probability for each

condition was a weighted sum between XGBoost and ProtAIDe-Dx: $w \times p_{\text{ProtAIDe}} + (1-w) \times p_{\text{XGBoost}}$. On the validation split, we searched for the optimal weight w , ranging from 0 to 1 with the step as 0.01, to get the highest validation balanced accuracy. This procedure was repeated for each training-validation-test split, including cross-validation and leave-one-site-out.

Model generalization to new sites by K -shot transfer learning. We developed a K -shot transfer learning framework⁶³ to help the developed model generalize to new tasks or new sites. Taking generalization to new sites as an example, only K participants were needed to transfer the ProtAIDe-Dx model to this new site, which makes ProtAIDe-Dx easy to scale to real-world clinical settings given model calibration⁶⁴ only needs K participants. In this study, we picked $K = 100$ to balance the available sample size for each condition and generalization performance. First, the ProtAIDe-Dx model was applied to new sites to get proteomic embeddings. Second, a simple logistic regression (LR) model was trained on K participants' proteomics embedding and tested on the remaining participants. The performances were evaluated on the remaining participants. Similarly, the developed ProtAIDe-Dx model could transfer to new neurodegeneration-related tasks by training a simple machine learning model on K participants.

The underlying assumption is that ProtAIDe-Dx's embeddings have captured enough signals to represent neurodegenerative diseases broadly. Therefore, training a simple machine learning model on low-dimensional proteomics embedding with limited subjects will avoid overfitting issues compared with directly training new models on raw proteins.

Estimating predictive importance. To estimate the predictive importance of individual proteins, the PermFIT approach²³ was adopted on each test split. For each input protein, the PermFIT approach randomly permuted this protein across participants and computed the cross-entropy differences compared with unpermuted data. The mean and s.d. of cross-entropy differences for this protein were obtained from 100 permutations, and P values were calculated on the basis of the mean and s.d., assuming a normal distribution.

In this study, we measured the predictive importance of proteins in two ways, including fold count and Z statistics. Once PermFIT was done on all ten splits, the fold count approach counted the frequency of important proteins across those splits using FDR-corrected P values. The Z statistics approach computed the z scores (mean divided by s.d.) on each test split and then averaged z scores across ten splits to get the predictive importance of proteins.

To estimate the importance of individual model embeddings in predicting specific diagnoses, we used the Haufe approach. Haufe and colleagues⁵⁶ proved that covariance is more reliable than weights (betas) to interpret feature importances in linear models. Given that the embeddings were linearly combined to make predictions, we computed the covariance between embeddings and predicted probabilities for each condition to estimate the embeddings' feature importance.

Model evaluation on GNPC

In addition to reporting classification accuracy, we performed multiple experiments to validate ProtAIDe-Dx models on GNPC data. First, we correlated predicted AD probability with MMSE scores. Given that each train-validation-test split had different probability thresholds, we first divided AD probability by the corresponding AD probability threshold for normalization. Then, we took the logarithm of the mean AD probabilities of participants with the same MMSE score on each testing split. Finally, we computed the Pearson correlation between MMSE scores and the logarithms of the mean AD probabilities. Similarly, we computed the logarithms of mean AD probabilities by *APOE* $\varepsilon 2/\varepsilon 4$ carrier group on each test split and then averaged them across ten splits.

Model generalization to new task with an example of prediction of longitudinal clinical progression. We validated the generalization ability of nonlinear proteomics embeddings to new tasks. More specifically, we took the embeddings from baseline visits and used these to predict whether participants with CDR 0 at baseline visits would progress to CDR 0.5 or 1 in their following visits or not. For 3,942 participants with longitudinal visits, we selected participants with CDR 0 at baseline visits and non-decreasing CDR during the following visits. For example, participants with CDR 0, 1 and 0 in their first, second and third visits would be excluded. In total, we selected 1,445 participants with stable CDR 0 in their following visits and 218 participants with progressive CDR in their following visits. It is noted that the baseline visits of the $1,445 + 218 = 1,663$ participants have been used in the tenfold cross-validation procedure for ProtAlDe-Dx development. We trained LR models to predict whether these 1,663 participants would progress to CDR 0.5 or 1 in their following visits or not, following the former tenfold cross-validation procedure. We adopted default hyperparameters for LR models from the sklearn package (v 1.6.1).

Dimensionality reduction of predicted probabilities with *t*-SNE. We selected 6,332 participants who were either diagnosed as recruited control ($N = 1,540$), AD ($N = 1,637$), PD ($N = 2,287$), FTD ($N = 175$), ALS ($N = 435$) or stroke/TIA ($N = 256$) to project onto a two-dimensional probability map. We excluded individuals with all negatives or multiple positives for the six targets. It is noted that the predicted probabilities of the 6,332 participants were all from test splits, following the former tenfold cross-validation procedure.

The dimensionality reduction algorithm was chosen as *t*-distributed stochastic neighbor embedding (*t*-SNE) from the OpenTSNE library (v 1.0.2). We used OpenTSNE over sklearn because OpenTSNE supports inference on new data with fit *t*-SNE models, allowing us to investigate new participants such as MCI–SCI patients on the same *t*-SNE map. The parameters of *t*-SNE were default values except for setting perplexity to 1,000 to better capture global structures.

Assessing disease heterogeneity. For each target, *K*-means clustering was performed on the two-dimensional *t*-SNE probability maps of participants within each diagnostic category. To pick the optimal number of clusters, we looped it over from 2 to 10 to pick the one with the highest Silhouette score. In summary, we obtained two control clusters, three AD clusters, four PD clusters, two FTD clusters, two ALS clusters and five stroke/TIA clusters.

To figure out the differentially expressed proteins across clusters, we ran GLMs across 7k proteins on positive participants for each diagnostic category. Therefore, for each diagnostic category, we have GLMs following the formula

$$\text{Protein} \sim C_1 + \dots + C_K + \text{Age} + \text{Sex} + \text{Site} + \text{AverageProteinLevel}. \quad (6)$$

$C_1 \dots C_K$ are the binary variables indicating which cluster the participants were from. Site is the categorical variable that indicates which site the participants were from.

Drug-related protein mapping. To systematically evaluate potential medication confounding across the 75 discriminative proteins, we performed a targeted protein–drug lookup in DrugBank⁶⁵ by querying each PermFIT-selected protein using its UniProt identifier. We then filtered and annotated drugs for relevance to neurodegenerative or vascular indications using a rule-based approach that integrated Anatomical Therapeutic Chemical (ATC) codes, DrugBank ‘drug category’ fields and free-text drug descriptions. In brief, ATC codes were tokenized (allowing multiple codes per drug) and matched using either prefix-based rules (for drug classes) or exact-code rules (for specific agents), whereas ‘drug category’ entries and descriptions were screened using curated keyword lists and case-insensitive regular expressions.

Drugs were labeled as AD therapies if they carried ATC codes starting with N06DA or N06DX and/or if DrugBank category/description text contained AD-related terms, including ‘anti-dementia’, ‘Alzheimer’, ‘cholinesterase inhibitor(s)’ or ‘NMDA receptor antagonist(s)’, as well as description mentions of ‘Alzheimer’ or ‘dementia’. Drugs were labeled as PD therapies if they carried ATC codes starting with N04 and/or if category or description text contained PD-related terms including ‘Parkinson’, ‘dopaminergic’ or ‘monoamine oxidase’, as well as description mentions of ‘Parkinson’. ALS drugs were labeled using an exact ATC match to N07XX02 (riluzole) and/or ALS-related description terms including ‘amyotrophic’, ‘ALS’ or ‘motor neuron disease’. Stroke/TIA-related drugs were labeled if they carried ATC codes starting with B01 (antithrombotic agents) and/or if the category or description text contained cerebrovascular terms, including ‘stroke’, ‘TIA’, ‘cerebrovascular’, ‘antiplatelet’, ‘anticoagulant’, ‘thrombolytic’ or ‘plasminogen activator’ or description mentions of ‘transient ischemic’. In addition, we flagged cardiovascular risk factor medications using ATC prefixes C02/C03/C07/C08/C09/C10 (antihypertensives/diuretics/beta blockers/calcium-channel blockers/renin–angiotensin system agents/lipid-modifying agents) and grouped these with stroke/TIA-related drugs.

Assessing site-to-site generalization variability. We conducted a site-to-site generalization experiment in which models were trained exclusively on data from a single site (development site) and evaluated on all remaining sites (test sites). As expected, classification performance was higher on the development site than on the test sites. We selected TabPFN as the predictive model because it does not require hyperparameter tuning. The same sites used in the leave-one-site-out experiments were included here. For each development site, the data were randomly split into two equal halves. Two TabPFN models were trained independently on each half and evaluated on the other half, with development-site accuracy defined as the mean of these two within-site accuracies. The two trained TabPFN models were then applied to the remaining external test sites, and test-site accuracy was defined as the mean performance of both models on the test sites. To quantify site-to-site variability, we defined relative performance as the ratio of test-site accuracy to development-site accuracy, with lower values indicating stronger site effects.

Enrichment analysis

To evaluate the cell-type expression of genes of interest, we utilized two independent single-cell RNA sequencing (scRNA-seq) datasets. The first dataset comprised single-cell transcriptomes from 2.3 million cells obtained from the aged human prefrontal cortex of 427 participants in the Religious Order Study (ROS) and the Rush Memory and Aging Project (MAP)⁶⁶. This dataset includes neuronal, glial and vascular cell types. The second dataset, the Human Brain Vascular Atlas, profiled mainly vascular and perivascular cell types, using 143,793 single-cell transcriptomes from the hippocampus and cortex of eight postmortem samples⁶⁷.

For the two datasets, we downloaded the Seurat objects and used the R package Seurat v4.3.0⁶⁸, applying the AverageExpression function to compute cell-type expression levels. We then calculated the proportion of gene expression across all major neuronal and non-neuronal populations.

For organ enrichment analysis, we used organ-enriched genes as defined in previous studies⁶⁹ on the basis of the Gene Tissue Expression Atlas (GTEx) bulk RNA-seq database. A gene was classified as organ-enriched if its expression was at least fourfold higher in a specific organ or tissue compared with any other, following the Human Protein Atlas definition⁷⁰. We considered organ categories including adipose tissue, artery, brain, esophagus, heart, immune, intestine, kidney, liver, lung, muscle, pancreas, skin, stomach and whole blood. Cell- and organ-enriched genes were then mapped to proteins quantified in the SomaScan assay.

To determine whether a specific set of proteins exhibited preferential expression in certain cell types, we calculated the average expression level of the significant protein list across each cell type. To determine whether this enrichment exceeded random expectation, we generated 10,000 randomly sampled gene lists from the SomaLogic background set, each maintaining the same protein count as the significant list. A probability distribution was then constructed on the basis of the average expression levels of these random lists across cell types. This allowed us to quantify the likelihood that the observed enrichment was greater than expected by chance, while accounting for background expression variability. Cell types with an FDR-corrected P value <0.05 (Benjamini–Hochberg correction) were considered significantly enriched.

We performed overrepresentation analyses of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) Biological Process (BP) terms using `enrichKEGG()` and `enrichGO()` functions in `clusterProfiler` (v4.14.4) and `org.Hs.eg.db` (v3.20.0) in R (v4.4.2). The input gene lists consisted of proteins that were significantly up-regulated, down-regulated or both in disease clusters defined by `logFoldChange` and an FDR-corrected P value <0.05 (Benjamini–Hochberg correction). We used all human proteins as the background set for overrepresentation analysis. We reported only those KEGG pathways and GO BP terms that met an FDR threshold of <0.05 . To reduce redundancy among significantly enriched BP terms, we applied the `clusterProfiler simplify()` function using Wang's semantic similarity measure with a cutoff of 0.7.

Enrichment analyses were performed on proteins differentially expressed across t -SNE clusters. Enrichment analysis was additionally performed on embedding-specific proteins. To annotate the biological process of each nonlinear proteomics embedding, we found 'embedding specific' proteins using regression. For each embedding, we took the 648 proteins from the cross-validation feature selection to train an XGBoost regression model, with grid search on validation split for the lowest mean squared errors. It is noted that the model was trained to generalize to BioFINDER-2 has 32 embedding dimensions. Once these models were trained, we thresholded proteins with 'total_gain' over 50 as 'embedding specific' proteins for enrichment analysis.

Model evaluation on BioFINDER-2

In the BioFINDER-2 dataset, where we correlated each biomarker with the different predicted probabilities separately in each of the six diagnostic groups (CU, SCD, MCI, AD, parkinsonism and other), FDR correction was applied across all comparisons performed. We also calculated the Pearson correlation between embeddings and biomarkers for BioFINDER-2 participants who had corresponding biomarkers available. FDR correction was applied across all comparisons performed. The model applied to the BioFINDER-2 dataset was the leave-one-site-out model trained on the remaining GNPC sites, with BioFINDER-2 entirely excluded. BioFINDER-2 was fully held out from all stages of model development, including feature selection, hyperparameter tuning and model fitting, and was used only for inference and downstream evaluation/clinical utility analyses. Therefore, there was no sample overlap or information leakage between training and testing datasets.

Differential diagnostics of multiple neurodegenerative diseases.

We further investigated whether proteomics embeddings would provide unique or additive value in distinguishing multiple neurodegenerative diseases in the BioFINDER-2 dataset. We selected 231 patients with AD (with abnormal CSF A β 42/A β 40 ratio), 111 patients with PD (with positive CSF α -synuclein status), 39 patients with FTD and 20 patients with stroke (with infarcts), excluding patients with multiple neurodegenerative diseases. We performed a fivefold cross-validation, ensuring the diagnosis distribution was balanced across folds. Three out of five folds were used to train a Support Vector Machine model, one fold was used to tune hyperparameter C from 0.001 to 100 to get

an optimal balanced accuracy score and the remaining one was used for the test. This procedure was repeated five times to ensure each fold was used as a test fold.

Four sets of features were selected to validate the effectiveness of our proteomics embeddings. Model 0 only used age and sex; model 1 used age, sex and the top 5 principal components of proteomic embeddings; model 2 used age and sex plus clinical biomarkers, including MMSE, plasma p-tau217, plasma NEFL and mean cortical thickness of AD-signature meta-ROI; and model 3 used all features in model 0, model 1 and model 2. It is noted that age, MMSE, mean cortical thickness of AD-signature meta-ROI, plasma p-tau217 and plasma NEFL were z-normalized. Both principle component analysis and z-normalization were fit on training splits. Once models were trained, we concatenated participants from five test folds to perform bootstraps 1,000 times to perform statistical tests.

Mixed-effect modeling of longitudinal cognitive decline. To assess whether baseline characteristics (for example, clinical diagnoses and model predictions) could differentiate longitudinal trajectories of decline, we fit linear mixed-effects models with individual variability modeled as random effects. Cognitive performance was measured using the MMSE.

First, we evaluated whether baseline clinical diagnoses could differentiate longitudinal trajectories of decline.

$$\text{MMSE} \sim \text{Age} + \text{Sex} + \text{Site} + \text{BaselineDx} \times \text{Year} + \text{Year}|\text{SubjectID}. \quad (7)$$

BaselineDx is the categorical variable that indicates baseline diagnoses and Year|SubjectID specifies random effects of individuals over time (in units of years).

Then, we evaluated whether baseline predictions could differentiate longitudinal trajectories of decline; we included baseline diagnoses as covariates.

$$\begin{aligned} \text{MMSE} \sim & \text{Age} + \text{Sex} + \text{Site} + \text{BaselineDx} \\ & + \text{BaselinePrediction} \times \text{Year} + \text{Year}|\text{SubjectID}. \end{aligned} \quad (8)$$

BaselinePrediction is the categorical variable that indicates baseline predictions and Year|SubjectID specifies random effects of individuals over time (in units of years).

A two-cutoff approach of predicted probabilities. In the external BioFINDER-2 dataset, we applied a two-cutoff strategy on predicted probabilities to determine biomarker positivity. Participants were divided into non-SCD and SCD groups, with the non-SCD group used to derive the cutoffs and the SCD group serving as the test set. Cutoffs were chosen to yield 90% NPVs and PPVs, except for CSF α -synuclein, where the PPV was set at 40% owing to sample validity constraints.

Individual disease risk report

To assess contributing proteins for individual prediction, Shapley Additive Explanations (SHAP) values by SHAP package (v0.48.0) were computed for each of six conditions, with training participants as the background set. To associate the top contributing proteins with health-related traits, we applied a proteome–phenome atlas⁴² to find the top correlated traits for each protein. The proteome–phenome atlas was computed on the basis of Olink proteomic data. Therefore, there might be no matched results for some of the contributing proteins; we only included available protein–trait associations.

Statistical analysis

We applied a corrected resample t -test to compare model performances in cross-validation and leave-one-site-out experiments. To examine whether the distribution of variables varies across t -SNE clusters or not, we conducted two-proportion z-tests for binary variables

with two clusters, chi-squared tests of independence for both binary variables with three or more groups and multi-category variables, *t*-tests for continuous variables with two clusters and one-way analysis of variance for continuous variables with three or more groups. To compare model performances in comorbidity detection, we applied *t*-tests on 1,000 bootstrapped accuracies. We conducted *t*-tests for comparing probability distributions across biomarker status. It is noted that all *P* values were FDR corrected with α as 0.05 using the Benjamini–Hochberg procedure.

Reporting summary

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

Data availability

GNPC (<https://www.neuroproteome.org/>) is open access. Pseudoanonymized BioFINDER-2 data will be shared by request from a qualified academic investigator for the sole purpose of replicating procedures and results presented in this article and as long as data transfer is in agreement with European Union legislation on the general data protection regulation and decisions by the Swedish Ethical Review Authority and Region Skåne, which should be regulated in a material transfer agreement.

Code availability

Code is publicly available via GitHub at https://github.com/DeMONLab-BioFINDER/An_ProtAIDe-Dx. Y. Xiao reviewed the code before merging it into the GitHub repository to reduce the chance of coding errors.

References

57. Palmqvist, S. et al. Discriminative accuracy of plasma phospho-tau217 for Alzheimer disease vs other neurodegenerative disorders. *J. Am. Med. Assoc.* **324**, 772–781 (2020).
58. Leuz, A. et al. Comparison of group-level and individualized brain regions for measuring change in longitudinal tau positron emission tomography in Alzheimer disease. *JAMA Neurol.* **80**, 614–623 (2023).
59. Karlsson, L. et al. Cerebrospinal fluid reference proteins increase accuracy and interpretability of biomarkers for brain diseases. *Nat. Commun.* **15**, 3676 (2024).
60. Zajac, Z. Preparing continuous features for neural networks with GaussRank. *FastML* <https://fastml.com/preparing-continuous-features-for-neural-networks-with-rankgauss/> (2018).
61. Xue, C. et al. AI-based differential diagnosis of dementia etiologies on multimodal data. *Nat. Med.* **30**, 2977–2989 (2024).
62. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: a next-generation hyperparameter optimization framework. In *Proc. 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 2623–2631 (Association for Computing Machinery, 2019).
63. He, T. et al. Meta-matching as a simple framework to translate phenotypic predictive models from big to small data. *Nat. Neurosci.* **25**, 795–804 (2022).
64. Alba, A. C. et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. *J. Am. Med. Assoc.* **318**, 1377–1384 (2017).
65. Knox, C. et al. DrugBank 6.0: the DrugBank knowledgebase for 2024. *Nucleic Acids Res.* **52**, D1265–D1275 (2024).
66. 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).
67. Yang, A. C. et al. A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* **603**, 885–892 (2022).
68. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
69. Zhao, R. et al. Plasma proteomics-based organ-specific aging for all-cause mortality and cause-specific mortality: a prospective cohort study. *Geroscience* **47**, 1411–1423 (2025).
70. Uhlen, M. et al. Proteomics. tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

Acknowledgements

We acknowledge J. Rittmo and L. Chauveau for aiding production of neuroimages. This work was supported by the SciLifeLab and Wallenberg Data Driven Life Science Program (grant no. KAW 2020.0239 to J.W.V.), the Crafoord Foundation (grant no. 20230790 to J.W.V.), the Swedish Research Council (grant no. 2024-03642 to J.W.V.) and the US National Institutes of Health (grant no. U01 AG079847-02 to J.W.V.). The BioFINDER study was supported by the Alzheimer's Association (grant nos ZEN 24-I 069572 to O.H., iLEADS-24-1277370 to O.H. and SG-231061717 to S.P.), European Research Council (grant no. 101096455 to O.H.), Knut and Alice Wallenberg Foundation (grant no. KAW 2022.0231), Mats Paulsson Foundation (grant no. 2025-0076 to O.H.), Michael J. Fox Foundation (grant nos MJFF-025741 to O.H. and MJFF-025507 to N.M.-C.), Region Skåne (grant nos 2025-2026-2024-2426 to O.H. and 2025-2026-2024-2028 to N.M.-C.), Swedish Brain Foundation (grant nos FO2024-0133-HK-46 to O.H., FO 2025-0055 and FO2025-0055 HK267 to N.M.-C., and FO 2024-0284 to S.P.), Swedish Parkinson Foundation (grant nos 1589/24 to O.H., 1685/25 to N.M.-C. and 1698/25 to S.P.), Swedish Research Council (grant nos 2023-06428 to O.H., and 2021-02219 and 2025-02319 to N.M.-C.), the Swedish federal government through the ALF agreement (grant nos 2022-Project0080 to O.H. and 2022-Project0107 to N.M.-C.), Alzheimer's Association and GHR Foundation (grant no. ALZSI-26-1523522 to N.M.-C.), Cure Alzheimer's Fund (to N.M.-C.), the Family Rönström Foundation (grant nos FRS-0013 and AF-1011799 to N.M.-C. and AF-1011949 to S.P.), GHR Foundation (grant nos 14358 to N.M.-C. and 13943 to S.P.), Global Research Platform, LLC (to N.M.-C.), Greta och Johan Kocks Stiftelser (grant no. F2024/228 to N.M.-C.), Skåne University Hospital's Foundations and Donations (to N.M.-C. and S.P.), Swedish Alzheimer Foundation (grant nos A-1032795 to N.M.-C. and AF-1032524 to S.P.), WASP and DDLS (grant no. WASP/DDLS22-066 to N.M.-C.), Bundy Academy (to S.P.), EU Commission: ERA PerMed (grant no. ERAPERMED2021-184 to S.P.), Greta och Johan Kocks stiftelser (to S.P.), Innovative Health Initiative (grant no. 101132933 to S.P.), the Kamprad Family Foundation for Entrepreneurship, Research and Charity (grant no. 20243058 to S.P.) and National Institute of Aging (grant no. R01AG083740-02 to S.P.). The precursor of ¹⁸F-flutemetamol was sponsored by GE Healthcare. The precursor of ¹⁸F-RO948 was provided by Roche. Our computational work was supported by workspace mms-proteomics-kb from the Alzheimer's Disease Data Initiative (<https://www.alzheimersdata.org>), project sens2023026 by the National Academic Infrastructure for Supercomputing in Sweden (NAISS; <https://www.naiss.se>) at UPPMAX, project NAISS 2024/22-457 by Chalmers e-Commons at Chalmers and the Berzelius resource (Berzelius-2025-231) funded by the Knut and Alice Wallenberg Foundation at the National Supercomputer Centre. NAISS is partially funded by the Swedish Research Council through grant agreement no. 2022-06725. The funding sources had no role in the design and conduct of the study; in the collection, analysis and interpretation of the data; or in the preparation, review or approval of the paper.

Author contributions

L.A. and J.W.V. conceptualized the study and designed the methodology. L.A. developed the ProtAIDe-Dx model. L.A., A.P.B., I.H. and G.V. performed the primary analyses. Y. Xiao reviewed the study code.

R.Z., Z.D., B.S., S.T., Y. Xu and C.D.W. provided input on predictive modeling design. V.K. and F.I. coordinated GNPC data harmonization, access, management and computational resources. S.J., D.v.W., E.S., S.P., R.O., N.M.-C. and O.H. contributed to BioFINDER-2 data collection, processing and management. L.A., A.P.B., I.H., G.V., R.S., O.H. and J.W.V. interpreted the results. L.A., A.P.B., I.H., G.V. and J.W.V. drafted the original paper, and all authors reviewed and edited the final version. J.W.V. provided funding and supervised the project.

Funding

Open access funding provided by Lund University.

Competing interests

Y. Xu is the cofounder of Andia Health. E.S. has acquired research support (for the institution) from Beckman Coulter, Bristol Myers Squibb, C2N Diagnostics, Eisai, Fujirebio, GE Healthcare and Roche Diagnostics. S.P. has acquired research support (for the institution) from Avid Radiopharmaceuticals and ki elements through ADDF. In the past 3 years, he has received consultancy/speaker fees from BioArtic, Danaher, Eisai, Eli Lilly, Novo Nordisk, Roche and Sanofi. R.O. is currently a full-time employee of Eli Lilly and Company. His contribution to the work presented in this paper was performed as an employee of Amsterdam University Medical Centers and Lund University. R.O. has received research funding/support from European Research Council, ZonMw, NWO, National Institute of Health, Alzheimer Association, Alzheimer Nederland, Stichting Dioraphte, Cure Alzheimer's fund, Health Holland, ERA PerMed, Alzheimerfonden, Hjärnfonden, Avid Radiopharmaceuticals, Janssen Research &

Development, Roche, Quanterix and Optina Diagnostics, has given lectures in symposia sponsored by GE Healthcare, received speaker fees from Springer and was an advisory board/steering committee member for Asceneuron, Biogen, Johnson & Johnson and Bristol Myers Squibb. All the aforementioned has been paid to Amsterdam University Medical Centers and Lund University. N.M.-C. has received consultancy/speaker fees from Biogen, Eli Lilly, Owkin and Merck. O.H. is an employee of Lund University and Eli Lilly. J.W.V. has received advisory fees from Manifest Technologies within the past 2 years. The other authors declare no competing interests.

Additional information

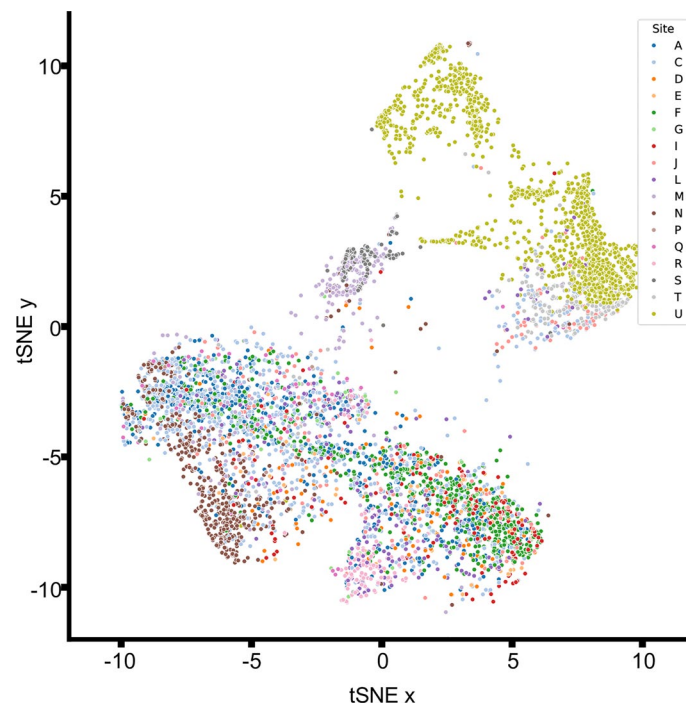
Extended data is available for this paper at <https://doi.org/10.1038/s41591-026-04303-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-026-04303-y>.

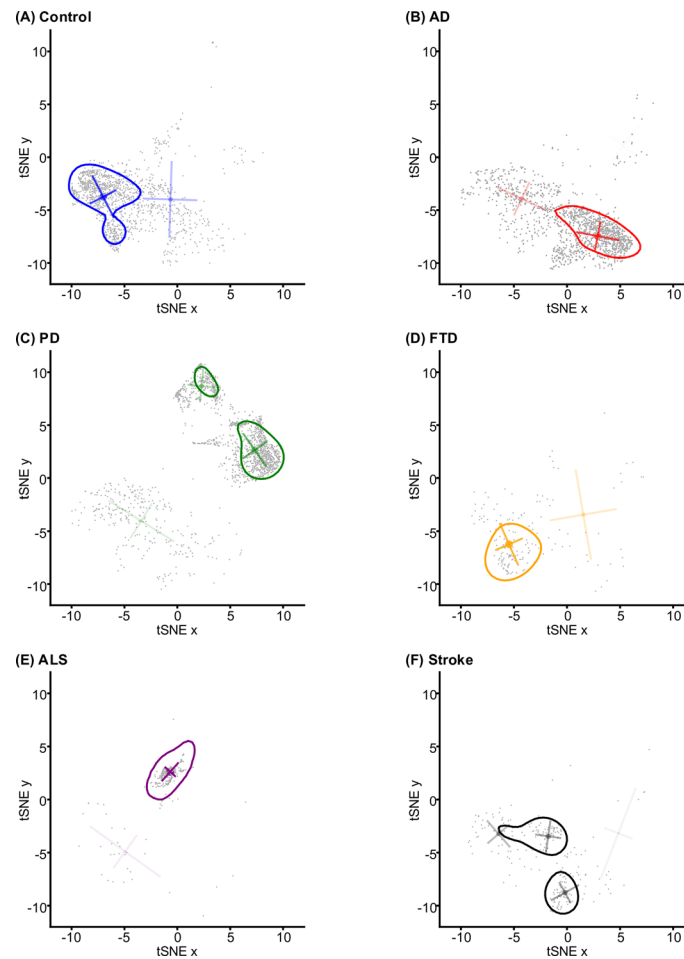
Correspondence and requests for materials should be addressed to Lijun An or Jacob W. Vogel.

Peer review information *Nature Medicine* thanks Ahmet Tarik Baykal and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Liam Messin, in collaboration with the *Nature Medicine* team.

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

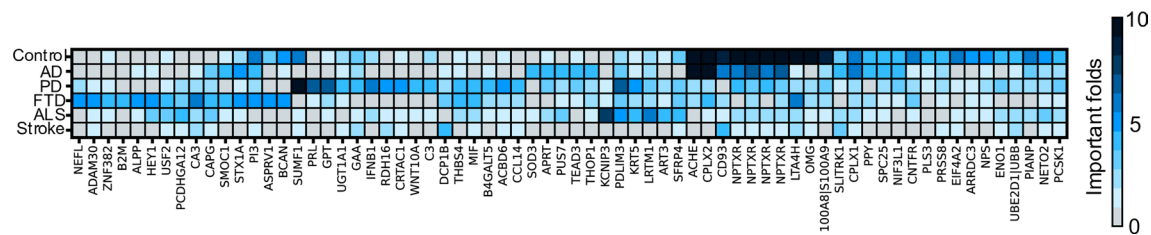


Extended Data Fig. 1 | Site distribution of 6,332 test participants on the t-SNE diagnostic probability map. The test participants and t-SNE embedding were the same as in Fig. 2, while the participants were re-colored by sites.



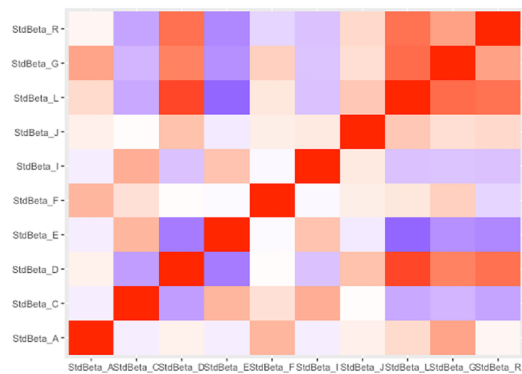
Extended Data Fig. 2 | K-means clustering of participants within diagnosis on the t-SNE diagnostic probability map. The contour was drawn based on Gaussian kernel density estimation with a threshold of 0.01. **(A)** Control

participants have two clusters. **(B)** AD patients have three clusters. **(C)** PD patients have four clusters. **(D)** FTD patients have two clusters. **(E)** ALS participants have two clusters. **(F)** Stroke/TIA participants have five clusters.

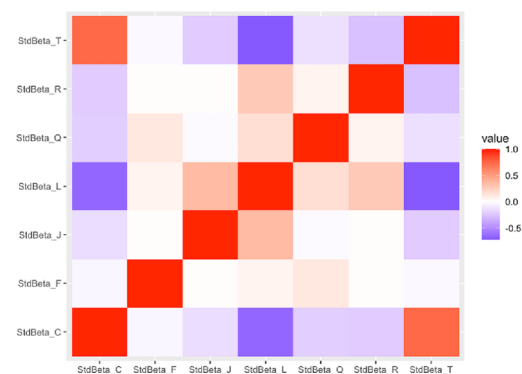


Extended Data Fig. 3 | Predictive importance of proteins by cross-validation fold counts with MCI-SCL patients in the model development. Proteins were visualized if they showed significant feature importance in more than 4 folds for at least one of the diagnostic tasks.

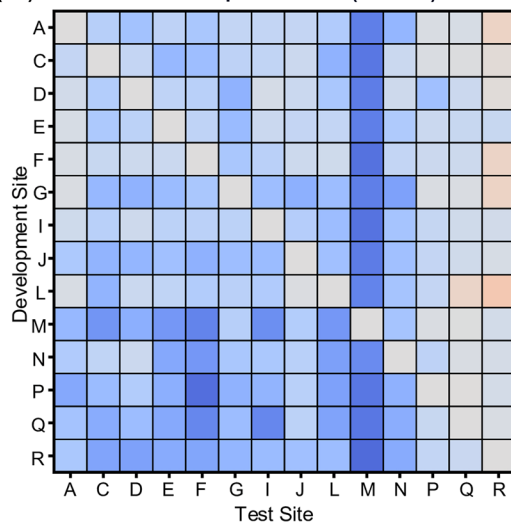
(A1) AD versus control



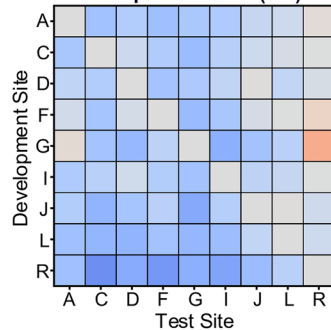
(A2) PD versus control



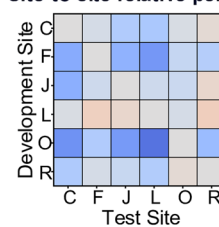
(B1) Site-to-site relative performance (Control)



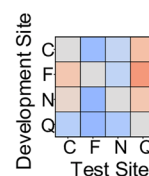
(B2) Site-to-site relative performance (AD)



(B3) Site-to-site relative performance (PD)



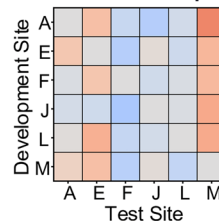
(B4) Site-to-site relative performance (FTD)



(B5) Site-to-site relative performance (ALS)



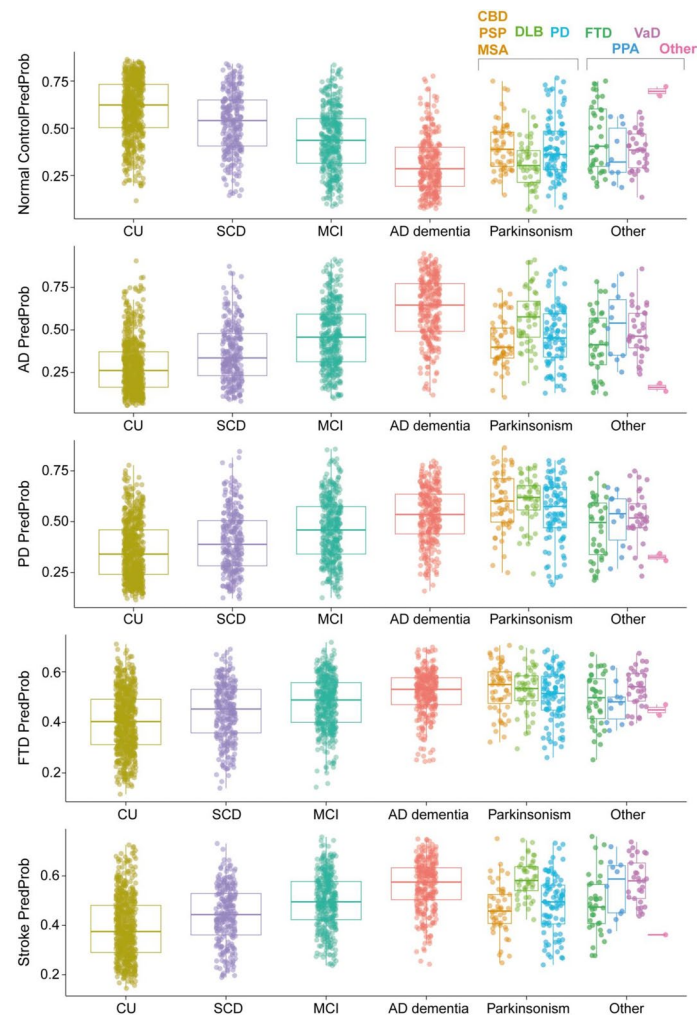
(B6) Site-to-site relative performance (Stroke/TIA)



Relative Performance (Test / Development)

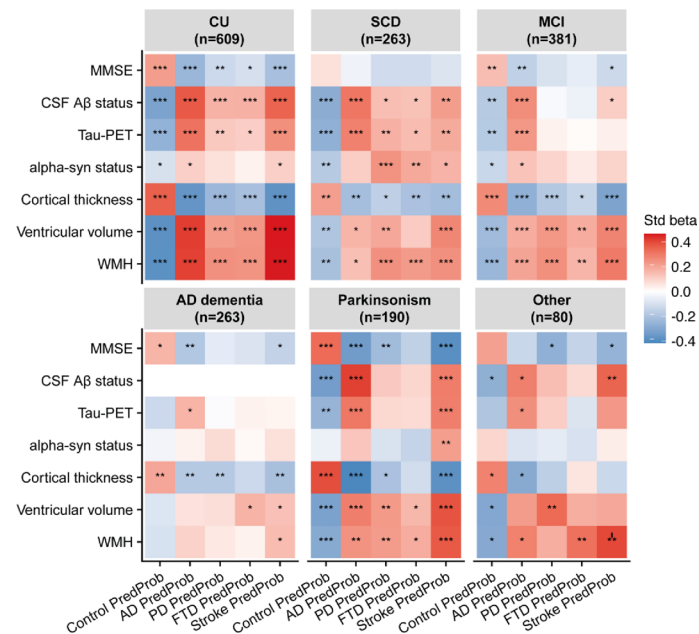
Extended Data Fig. 4 | Site effects of proteomics quantified by univariate logistic regression and deep learning model TabPFN. (A) At each site, a univariate logistic regression was run for every protein to examine the effect of disease status (disease vs. control), with age, sex, and mean standardized protein level included as covariates; reported effect sizes correspond to the regression coefficients (β). For each pair of sites, we computed the similarity between sites by correlating their per-protein β vectors. The heatmaps show pairwise Pearson

correlation values. (A1) AD versus control. (A2) PD versus control. (B) Site-to-site generalization performance using TabPFN. Heatmaps show the change in model performance when a TabPFN model trained on a development site is evaluated on a separate test site. Blue tones indicate poorer generalization; red tones indicate improved performance on the test site. (B1)-(B6) performances for Control, AD, PD, FTD, ALS, stroke/TIA, respectively.

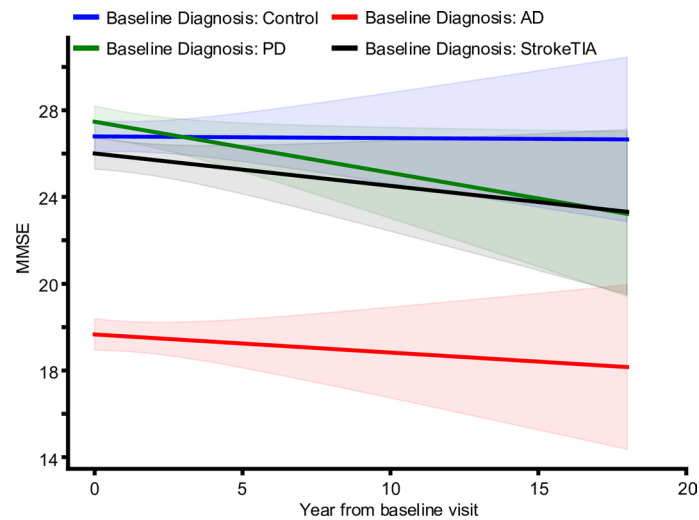


Extended Data Fig. 5 | Probabilities by each diagnostic group in the BioFINDER-2 cohort. Box plots were stratified by clinical diagnostic groups among 1,786 BioFINDER-2 participants. CU (n = 609), SCD (n = 263), MCI (n = 381),

AD dementia (n = 263), CBS/PSP/MSA (n = 55), DLB (n = 48), PD (n = 78), FTD (n = 33), PPA (n = 10), VaD (n = 32), Other (n = 5). Box plots show the median, interquartile range (25th–75th percentiles), and whiskers extending to 1.5×IQR.



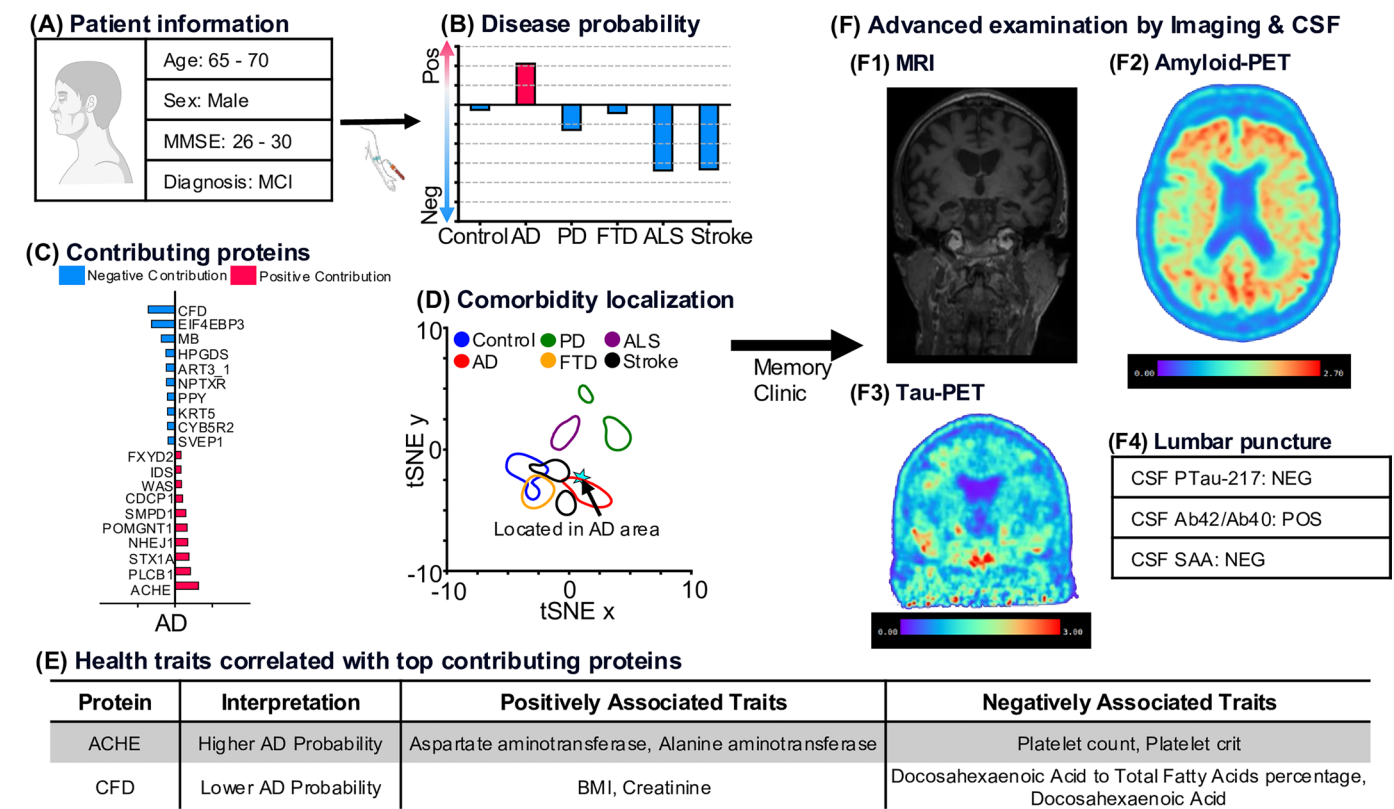
Extended Data Fig. 6 | Correlations between predicted probabilities and biomarkers. Significant results from two-sided t-tests are indicated by asterisks after FDR correction across all comparisons: * FDR-adjusted P = 0.01–0.05; ** FDR-adjusted P = 0.001–0.01; *** FDR-adjusted P < 0.001. Detailed correlations and FDR-corrected p values can be found in SuppData 12.



Extended Data Fig. 7 | MMSE longitudinal trajectories by baseline clinical diagnosis in the GNPC. Longitudinal trajectories were assessed using the linear mixed-effects model: $\text{MMSE} = \text{Age} + \text{Sex} + \text{Site} + \text{BaselineDx} \times \text{Year} + (\text{Year} | \text{SubjectID})$. Fixed effects: Age, Sex, Site, and the interaction between baseline diagnosis and time (Year); random intercepts and slopes for Year

were included for each subject. The $\text{BaselineDx} \times \text{Year}$ term tests whether the annual change in MMSE differs by baseline diagnosis. Solid lines indicate linear mixed-effects model predicted mean MMSE trajectories, shaded regions represent 95% confidence intervals for the LME fixed-effects predictions.

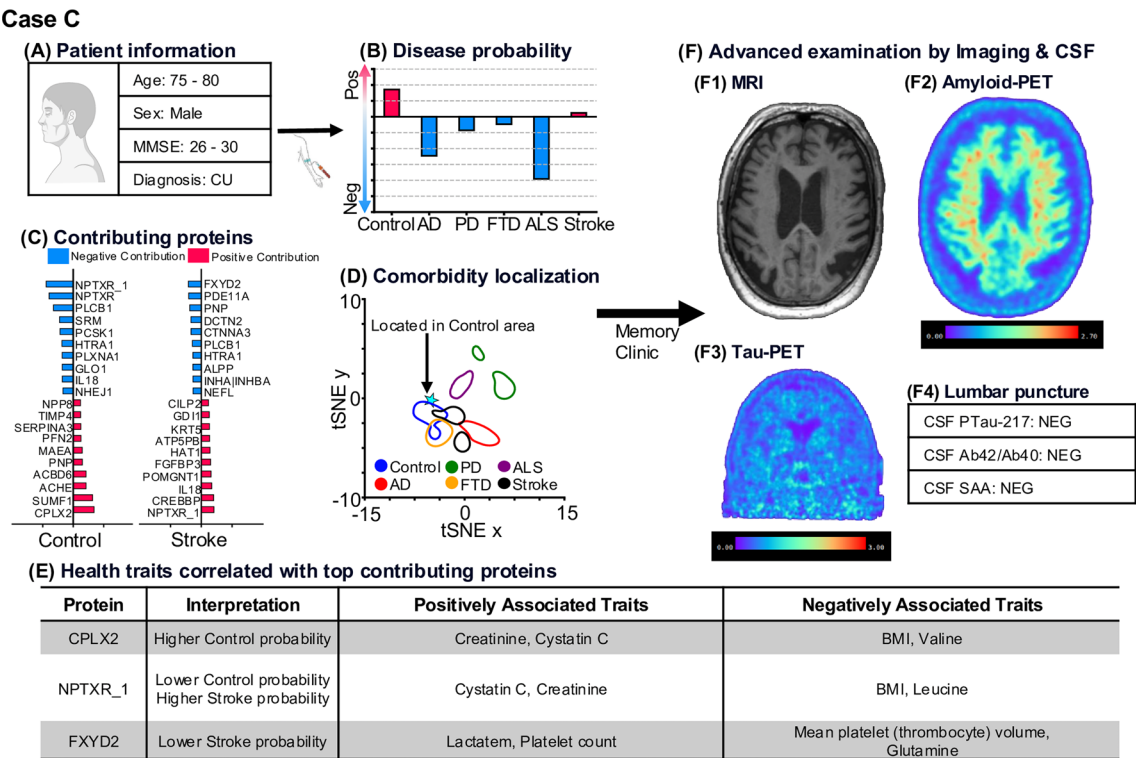
Case B



Extended Data Fig. 8 | Individual neurodegeneration risk report (Case B).

(A) Demographics and cognition score of one older participant, who was diagnosed with MCI. (B) ProtAIDe-Dx predicted this participant with higher probabilities of AD. The probabilities were normalized to be centered at zero. (C) Contributing proteins for making the decision were computed based on SHAP values. The top 10 positive and top 10 negative proteins were visualized.

(D) The location of this participant on the diagnostic probability map indicates he was close to typical AD patients. (E) Health traits correlated with top contributing proteins were listed, informing clinicians to pay attention to these traits. (F) Advanced imaging and CSF biomarker examination at the memory clinic to confirm AD pathology. Illustration in a created in BioRender; An, L. <https://biorender.com/q2by4y5> (2026).



Extended Data Fig. 9 | Individual neurodegeneration risk report (Case C). (A) Demographics and cognition score of one older participant, who was diagnosed with MCI. (B) ProtaIDe-Dx predicted this participant with higher probabilities of AD. The probabilities were normalized to be centered at zero. (C) Contributing proteins for making the decision were computed based on SHAP values. The top 10 positive and top 10 negative proteins were visualized.

(D) The location of this participant on the diagnostic probability map indicates he was close to typical AD patients. (E) Health traits correlated with top contributing proteins were listed, informing clinicians to pay attention to these traits. (F) Advanced imaging and CSF biomarker examination at the memory clinic to confirm AD pathology. Illustration in a created in BioRender; An, L. <https://biorender.com/q2by4y5> (2026).

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 [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| ☐ | ☒ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | 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
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ | A description of all covariates tested |
| ☐ | ☒ | 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. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☐ | ☒ | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection The data is directly downloaded from GNPC database. No data collection software/code applicable.

Data analysis All code are publicly available at: https://github.com/DeMONLab-BioFINDER/An_ProtAIDe-Dx; We mainly used Python (v3.9.21), torch (v2.5.0), numpy (v2.0.2), scikit-learn (v1.6.1), scipy (v1.13.1), XGBoost (v2.1.4), and tabpfn (v2.1.0) for our analysis. For detailed information, please check our configuration file at: https://github.com/DeMONLab-BioFINDER/An_ProtAIDe-Dx/blob/main/replication/config/DeMONLab_ProtAIDe_env.yaml

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.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). 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](#)

GNPC (<https://www.neuroproteome.org/>) is open-access. Pseudonymized BioFINDER-2 data will be shared by request from a qualified academic investigator for the

sole purpose of replicating procedures and results presented in the article and as long as data transfer is in agreement with EU legislation on the general data protection regulation and decisions by the Swedish Ethical Review Authority and Region Skåne, which should be regulated in a material transfer agreement.

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	Sex was self-reported. All statistical analysis included sex as covariate.
Reporting on race, ethnicity, or other socially relevant groupings	Race and ethnicity were not included as confounding factors in this manuscript.
Population characteristics	Detailed information is given in Tables S1, S2 and S12.
Recruitment	The participants were either health aging people, patients with mild cognitive impairment, or with five of the most common incapacitating age-related neurological conditions in old age – AD, PD, neurovascular disease (represented here as stroke), ALS and frontotemporal dementia (FTD).
Ethics oversight	Ethical approvals were given by the IRBs of all participating cohorts in GNPC. All studies were approved by the Institutional Review Board (IRB) of Lund University and written informed consent or assent was obtained from all participants or their legally authorized representative.

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	We selected all GNPC (N=17,187) and BioFIDNER-2 (N=1,786) participants with SomaLogic 7k plasma proteomics data.
Data exclusions	We excluded subjects without SomaLogic 7k plasma proteomics data.
Replication	We replicated our model trained on GNPC on external clinical cohort BioFIDNER-2.
Randomization	We performed 10-fold cross validation with random seeds for randomization.
Blinding	Proteomic measurements were performed blinded to any demographics or clinical characteristics.

Reporting for specific materials, systems and methods

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

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Antibodies

Antibodies used	Details about antibody-based proteomic technology is provided by SomaLogic with all details found here: https://somallogic.com/wp-content/uploads/2022/04/SL00000712_Rev-1_2022-03_SomaScan-Infographic-1.pdf
Validation	Extensive validation has been conducted by SomaLogic with all details of the SomaLogic 7k Assay available here: https://somallogic.com/poster/somascan-platform-confirmation-and-performance-validation/

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	BioFINDER-2: NCT03174938
Study protocol	Please see www.biofinder.se for BioFINDER studies
Data collection	BioFINDER-2 participants include a mix of population-based and memory clinic-based studies in Lund and Malmö, in Sweden and all imaging data was acquired at Skane University Hospital between April 2017 and December 2022.
Outcomes	The primary outcomes was that ProtAIDe-Dx driven proteomics signatures provided comparable performance to an automated diagnosis combining multiple neuroimaging and clinical biomarkers currently used in memory clinics.

Plants

Seed stocks	N/A. We could not unselect this tab.
Novel plant genotypes	N/A. We could not unselect this tab.
Authentication	N/A. We could not unselect this tab.

Magnetic resonance imaging

Experimental design

Design type	Structural MRI imaging.
Design specifications	Whole brain structural MRI scan.
Behavioral performance measures	No behavior performances measures during structural imaging.

Acquisition

Imaging type(s)	structural
Field strength	3
Sequence & imaging parameters	T1-weighted images were acquired on a Siemens Prisma scanner (Siemens Medical Solutions) with a 64-channel head coil using an MPRAGE sequence (in-plane resolution=1×1 mm ² , slice thickness=1 mm, repetition time=1900 ms, echo time=2.54ms, flip-angle=9°).
Area of acquisition	Whole brain scan.
Diffusion MRI	☐ Used ☒ Not used

Preprocessing

Preprocessing software	FreeSurfer version 6.0
------------------------	------------------------

Normalization	Volume of white matter lesion (SAMSEG) was normalied by intracrainal volume.
Normalization template	Data extracted from single subject space.
Noise and artifact removal	N.A.
Volume censoring	N.A.

Statistical modeling & inference

Model type and settings	Predictive
Effect(s) tested	Prediction of diagnosis group of patients using structural MRI.
Specify type of analysis:	☐ Whole brain ☒ ROI-based ☐ Both
Anatomical location(s)	AD-signature meta-ROI, which is the mean cortical thickness in the following individual ROIs: entorhinal, inferior temporal, middle temporal, and fusiform.
Statistic type for inference	N.A.
(See Eklund et al. 2016)	
Correction	N.A.

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis
Multivariate modeling and predictive analysis	We used SVM as machine learning to predict diagnosis group of patients.