

ARTICLE OPEN



Unbiased proteomics and multivariable regularized regression techniques identify SMOC1, NOG, APCS, and NTN1 in an Alzheimer's disease brain proteomic signature

Jackson A. Roberts^{1,2}, Vijay R. Varma¹, Julián Candia³, Toshiko Tanaka³, Luigi Ferrucci³, David A. Bennett⁴ and Madhav Thambisetty¹

Advancements in omics methodologies have generated a wealth of high-dimensional Alzheimer's disease (AD) datasets, creating significant opportunities and challenges for data interpretation. In this study, we utilized multivariable regularized regression techniques to identify a reduced set of proteins that could discriminate between AD and cognitively normal (CN) brain samples. Utilizing *eNetXplorer*, an R package that tests the accuracy and significance of a family of elastic net generalized linear models, we identified 4 proteins (SMOC1, NOG, APCS, NTN1) that accurately discriminated between AD ($n = 31$) and CN ($n = 22$) middle frontal gyrus (MFG) tissue samples from Religious Orders Study participants with 83 percent accuracy. We then validated this signature in MFG samples from Baltimore Longitudinal Study of Aging participants using leave-one-out logistic regression cross-validation, finding that the signature again accurately discriminated AD ($n = 31$) and CN ($n = 19$) participants with a receiver operating characteristic curve area under the curve of 0.863. These proteins were strongly correlated with the burden of neurofibrillary tangle and amyloid pathology in both study cohorts. We additionally tested whether these proteins differed between AD and CN inferior temporal gyrus (ITG) samples and blood serum samples at the time of AD diagnosis in ROS and BLSA, finding that the proteins differed between AD and CN ITG samples but not in blood serum samples. The identified proteins may provide mechanistic insights into the pathophysiology of AD, and the methods utilized in this study may serve as the basis for further work with additional high-dimensional datasets in AD.

npj Aging (2023)9:18; <https://doi.org/10.1038/s41514-023-00112-6>

INTRODUCTION

Alzheimer's disease (AD) remains the most common neurodegenerative disorder and cause of dementia among the elderly, yet causal mechanisms driving sporadic AD remain elusive. Previously, much AD research has focused on hypothesis-driven interrogation of molecular mechanisms, which contributed to the identification of the central roles of amyloid and tau in disease, as well as the genetic association of the *APOE ε4* allele with sporadic AD. However, targeted and hypothesis-oriented research may preclude the discovery of novel molecular mechanisms underlying AD. The development of multiple unbiased omics methods and their application to the large-scale analyses of blood and brain tissue in AD holds promise for the identification of novel molecular mechanisms associated with AD pathogenesis.

Proteomics offers a particularly exciting opportunity for the study of disease mechanisms, as proteins represent the effector molecules of many upstream biological processes and additionally serve as the targets for the majority of known and potential pharmacological therapeutics. Additionally, advances in high throughput data acquisition have transformed proteomics into a field in which rich datasets are readily generated, resulting in an abundance of proteomic data for analysis¹. As data become more abundant, the challenge for researchers is no longer quantification but rather analysis and interpretation of multi-dimensional datasets. In response, systems biology approaches have been

elaborated, including network modeling analyses that seek to cluster variables using multiple methods to establish similarity distances between variables². In proteomics, this may take the form of protein-protein interaction networks that can be generated for different disease states, including AD, which may be followed by pathway enrichment to identify classes of biological mechanisms altered in disease³. In AD, such approaches have contributed to a global understanding of processes associated with disease, proteostasis, immune response, cell signaling, mitochondrial function, metabolism, RNA binding and splicing, and several other biological pathways^{4,5}.

While large-scale, co-expression-based approaches have made significant contributions to understanding altered biologic networks in disease states, studies focused on individual protein contributions to disease remain limited by methodologic heterogeneity⁶. Indeed, studies that examine individual protein differences in the context of disease are largely restricted to univariate analyses that do not take into account the complexity of data structures and inter-variable correlation that characterizes multi-dimensional datasets in small sample size studies. In this context, the elastic net⁷ offers a regularized generalized linear modeling and variable selection method that is flexible and accounts for multi-collinearity, allowing for the selection of a reduced set of features that are highly associated with the outcome. The *eNetXplorer* R package⁸ implemented functionality to sample different predictive elastic net models with a cross-validated

¹Clinical and Translational Neuroscience Section, Laboratory of Behavioral Neuroscience, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA.

²Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA. ³Longitudinal Studies Section, Translational Gerontology Branch, National Institute on Aging, National Institutes of Health, Baltimore, MD 21224, USA. ⁴Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL 60612, USA.

✉email: Jar2344@cumc.columbia.edu; thambisetty@mail.nih.gov

approach that generates model significance and predictive scores. This framework has recently been applied to transcriptomic⁹, brain imaging¹⁰, and proteomic¹¹ data, showing the ability to scan a high-dimensional feature space to find cross-validated signatures predictive of a disease state. Such signatures may, as a result, represent plausible drug targets or provide insight into disease pathomechanisms.

In this study, we undertook an unbiased proteomics approach in brain tissue samples to identify a set of proteins that could accurately discriminate between AD and cognitively normal (CN) individuals in two separate cohorts. We applied brain proteomic analyses in the inferior temporal gyrus (ITG) and middle frontal gyrus (MFG) in the Religious Orders Study (ROS) using regularized generalized elastic net modeling to develop an AD proteomic signature and then tested the discriminatory potential of this signature in the Baltimore Longitudinal Study of Aging (BLSA). We then compared the levels of proteins included in the signature in BLSA and ROS ITG and MFG brain tissue, as well as in blood serum samples collected at time of AD diagnosis, in order to determine if signature proteins are reflected in raw tissue level comparisons both within the signature brain region and additional tissue matrices. Lastly, we examined the association of blood serum and brain tissue levels of these proteins with the severity of AD pathology.

RESULTS

Demographic characteristics

The demographic characteristics of participants who provided brain tissue samples are summarized in Table 1. In the BLSA sample, AD and CN groups did not differ significantly in age at death or sex. The AD group had a higher proportion of white participants (race) than the CN samples. As anticipated, AD and CN groups differed significantly in the severity of neuritic plaques (CERAD scores) and neurofibrillary tangles (Braak scores), with higher levels of pathology in the AD group.

In the ROS sample, all individuals included in the study were of white race. The AD group was older at the age of death and had a higher proportion of female (sex) participants compared to CN. As in BLSA, the AD group displayed higher CERAD and Braak scores, indicative of a higher burden of pathology in this group.

We also compared demographic characteristics across the BLSA and ROS cohorts. Considering the full sample, BLSA and ROS varied significantly in sex and age at death. Comparing BLSA AD participants to ROS AD participants, the BLSA AD group had an earlier age of disease onset, longer duration of disease, had a smaller proportion of female participants, and had higher Braak scores than the ROS AD group.

Additional details regarding the demographic characteristics of individuals with blood serum samples are provided in Supplementary Table 2.

Protein signature evaluation in MFG

We performed eNetXplorer analyses to develop a discriminatory protein signature in ROS MFG samples. In the MFG, the best-performing model ($\alpha = 1$) discriminated between AD and CN brain samples with 83 percent accuracy (Fig. 1a, b). This model identified four proteins (secreted modular calcium-binding protein 1 (SMOC1), noggin (NOG), amyloid P component, serum (APCS), and netrin-1 (NTN1)) based upon the significance of feature frequency, among which SMOC1 displayed a positive feature coefficient—i.e., an increased level of this protein was associated with an increased probability of classification as AD (Supplementary Fig. 1a). Furthermore, these proteins displayed significant feature frequencies at all α levels tested, signifying stability of the signature across model stringencies (Supplementary Fig. 1d). These proteins were reassessed as a signature in a leave-one-out

Table 1. Cohort demographics.

Baltimore longitudinal study of aging (BLSA)			
	Total Sample N = 50	AD N = 31	CN N = 19
Age at death, mean (SD)	86.38 (10.70)†	88.46 (8.30)†	82.98 (13.31)
Age of onset, mean (SD)	-	78.81 (9.92)†	-
Disease duration, mean (SD)	-	9.65 (4.47)†	-
Sex, n (% female)	21 (42.00)†	15 (48.39)†	6 (31.58)
Race, n (% white)	48 (96.00)	31 (100)*	17 (89.47)*
CERAD, mean (SD)	1.78 (1.34)	2.77 (0.43)*	0.16 (0.38)*
Braak, mean (SD)	4.14 (1.69)	5.13 (1.12)*†	2.53 (1.12)*
Religious orders study (ROS)			
	Total Sample N = 53	AD N = 31	CN N = 22
Age at death, mean (SD)	90.58 (6.53)†	92.81 (5.45)*†	87.44 (6.75)*
Age of onset, mean (SD)	-	89.14 (5.71)†	-
Disease duration, mean (SD)	-	3.67 (2.94)†	-
Sex, n (% female)	40 (75.47)†	27 (87.10)*†	13 (59.09)*
Race, n (% white)	53 (100.00)	31 (100.00)	22 (100.00)
CERAD, mean (SD)	1.64 (1.26)	2.58 (0.50)*	0.32 (0.65)*
Braak, mean (SD)	3.89 (1.33)	4.65 (0.66)*†	2.82 (1.30)*

* $p < 0.05$ comparing AD to CN within a cohort.
† $p < 0.05$ comparing BLSA to ROS (i.e., AD in BLSA compared to AD in ROS).

cross-validated logistic regression model, which, as expected, was able to accurately discriminate between ROS MFG AD and CN samples with AUC = 0.909 (95% CI: 0.813–1.00) (Fig. 1c). We then tested the ability of these proteins to discriminate between AD and CN MFG samples in the independent BLSA cohort. Remarkably, the protein signature accurately discriminated BLSA MFG samples, with AUC = 0.863 (95% CI: 0.735–0.991), as shown in Fig. 1d. As a result of this independent validation of the ROS MFG signature, these proteins were further evaluated in the remainder of the study.

We additionally tested the ability of the ROS-derived MFG signature to discriminate between AD and CN in BLSA ITG samples. The MFG signature failed to discriminate accurately between AD and CN participants in the BLSA ITG, with an AUC = 0.626 (95% CI: 0.428–0.825).

Protein signature evaluation in ITG

To establish an ITG proteomic signature, we performed eNetXplorer analyses in ROS ITG to determine which multivariable protein model best discriminates between AD and CN participants. In the ITG, the highest cross-validated model performance was obtained in the range $\alpha = 0.8$ –1, which discriminated between AD and CN ROS participants with 91 percent accuracy (Fig. 2a). Based on considerations of parsimony, we selected lasso ($\alpha = 1$), which was able to classify most individuals correctly (Fig. 2b) and identified four significant proteins (Noggin (NOG), interferon lambda 2 (IFNL2), secreted frizzled-related protein 1 (SFRP1), and midkine (MDK)) based on significance of feature frequency. Two of these proteins (NOG and SFRP1) displayed a significant positive feature coefficient—i.e., increased levels of these proteins were associated with an increased probability of classification as AD (Supplementary Fig. 2a). To confirm these findings, we utilized this signature in a leave-one-out cross-validated logistic regression model in the same ROS cohort, which, as expected, was able to discriminate between ROS ITG AD and CN

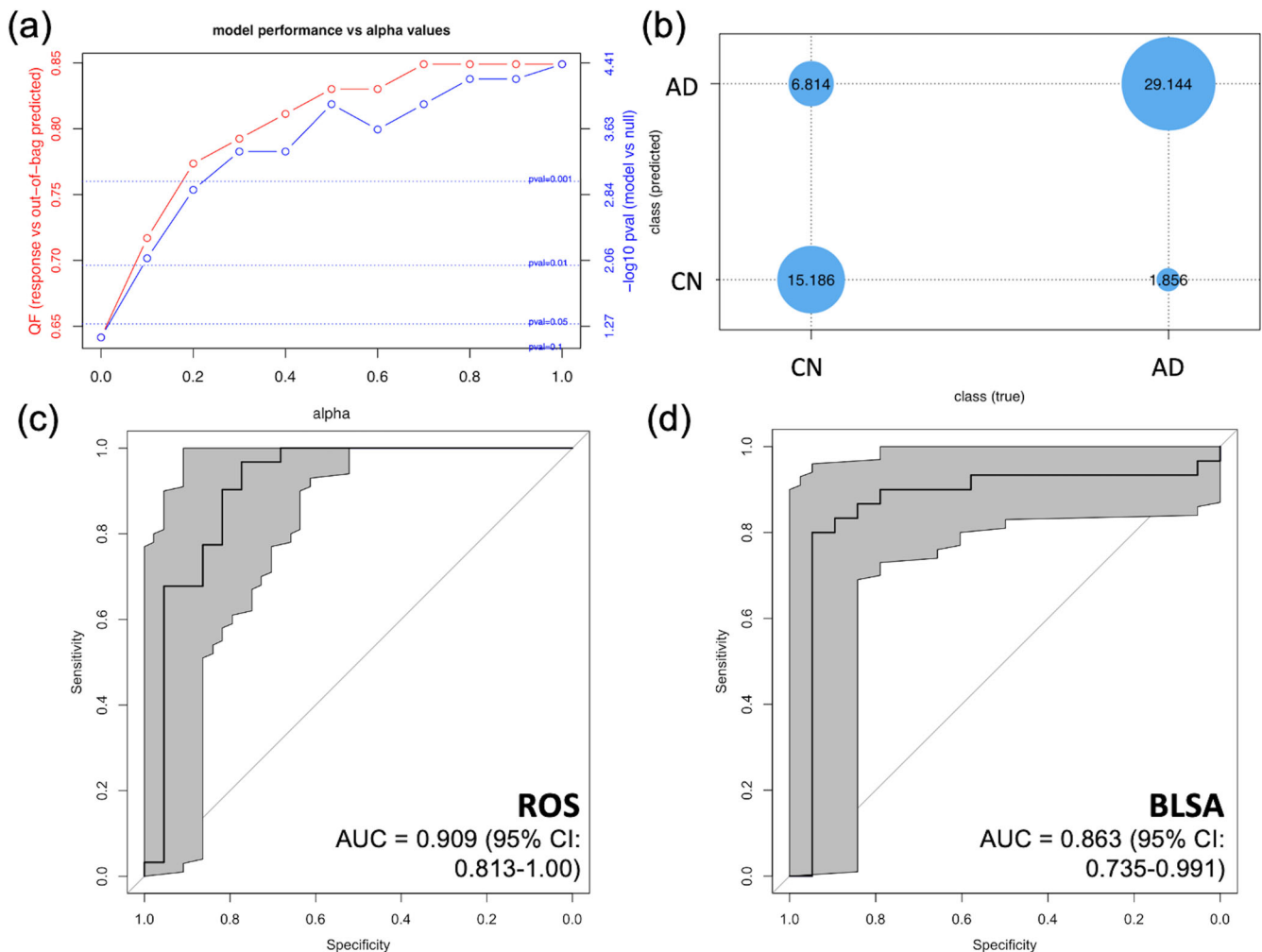


Fig. 1 Establishment and evaluation of MFG proteomic signature. **a** Output of eNetXplorer model selection analyses, performed in ROS MFG. The x-axis indicates the model α value, a parameter that scales from the ridge ($\alpha = 0$) to lasso ($\alpha = 1$) regression. The red line indicates accuracy as quality function (QF), which represents the proportion of out-of-bag class (AD vs. CN) predictions that match the true class of each participant. The blue line indicates the model significance for each α value. **b** Contingency matrix comparing the out-of-bag predicted class and true class for the selected model. **c** Receiver operating characteristic (ROC) curve for leave-one-out cross-validation (LOOCV) of the 4 protein MFG signature in the ROS cohort, derived at the $\alpha = 1$ level. Gray shading indicates the 95 percent confidence interval. **d** ROC curve for LOOCV of MFG samples from an independent cohort of BLSA participants. The area under the curve (AUC) is demonstrated with 95 percent confidence intervals. Gray shading indicates the 95 percent confidence interval.

samples, yielding an AUC = 0.937 (95% CI: 0.846–1.00). The corresponding ROC plot is shown in Fig. 2c. Then, we tested this same signature in an independent cohort, the BLSA, which demonstrated much weaker discriminatory potential, yielding AUC = 0.643 (95% CI: 0.463–0.823) as shown in Fig. 2d.

We additionally tested the ability of the ROS-derived ITG signature to discriminate between AD and CN in BLSA MFG samples. The signature discriminated between AD and CN participants with an AUC of 0.867 (95% CI: 0.746–0.988).

Brain and blood serum levels of MFG protein signature

To determine differences in levels of SMOC1, NOG, APCS, and NTN1 between AD and CN individuals, we performed proportional odds models separately for each cohort. Comparisons were made for both ROS and BLSA between protein levels in ITG, MFG, and blood serum at the time of sampling concurrent with diagnosis and age-matched controls. Among all 1300 proteins assayed in the MFG, 352 were significantly different ($p < 0.05$) between AD and CN in BLSA participants, and 223 were significantly different between conditions in ROS participants (Supplementary Table 3).

In the MFG brain region in which the proteomic signature was established, all 4 proteins were significantly different between AD and CN individuals in both cohorts at the $p < 0.01$ level (Fig. 3). For all four proteins, levels were higher in AD than in CN individuals in both ROS and BLSA cohorts. In the ITG, the majority of protein comparisons were also significant between AD and CN in both cohorts (Supplementary Fig. 3). Only APCS was not statistically significant ($p = 0.095$) in the ROS ITG, while all other proteins were significantly higher in AD ITG samples compared to CN. In blood serum samples at the time of diagnosis, levels of APCS were higher in AD individuals, and levels of NTN1 were lower compared to CN individuals (Supplementary Fig. 4). These results were significant in the ROS cohort only.

Brain and blood serum levels of ITG protein signature

We also determined differences in levels of NOG, MDK, SFRP1, and IFNL2 between AD and CN brain samples. In the ITG brain region, in which the signature was established, all four proteins were significantly different in ROS, but only one protein (NOG) was significant in BLSA. SFRP1, IFNL2, and NOG were increased in ROS

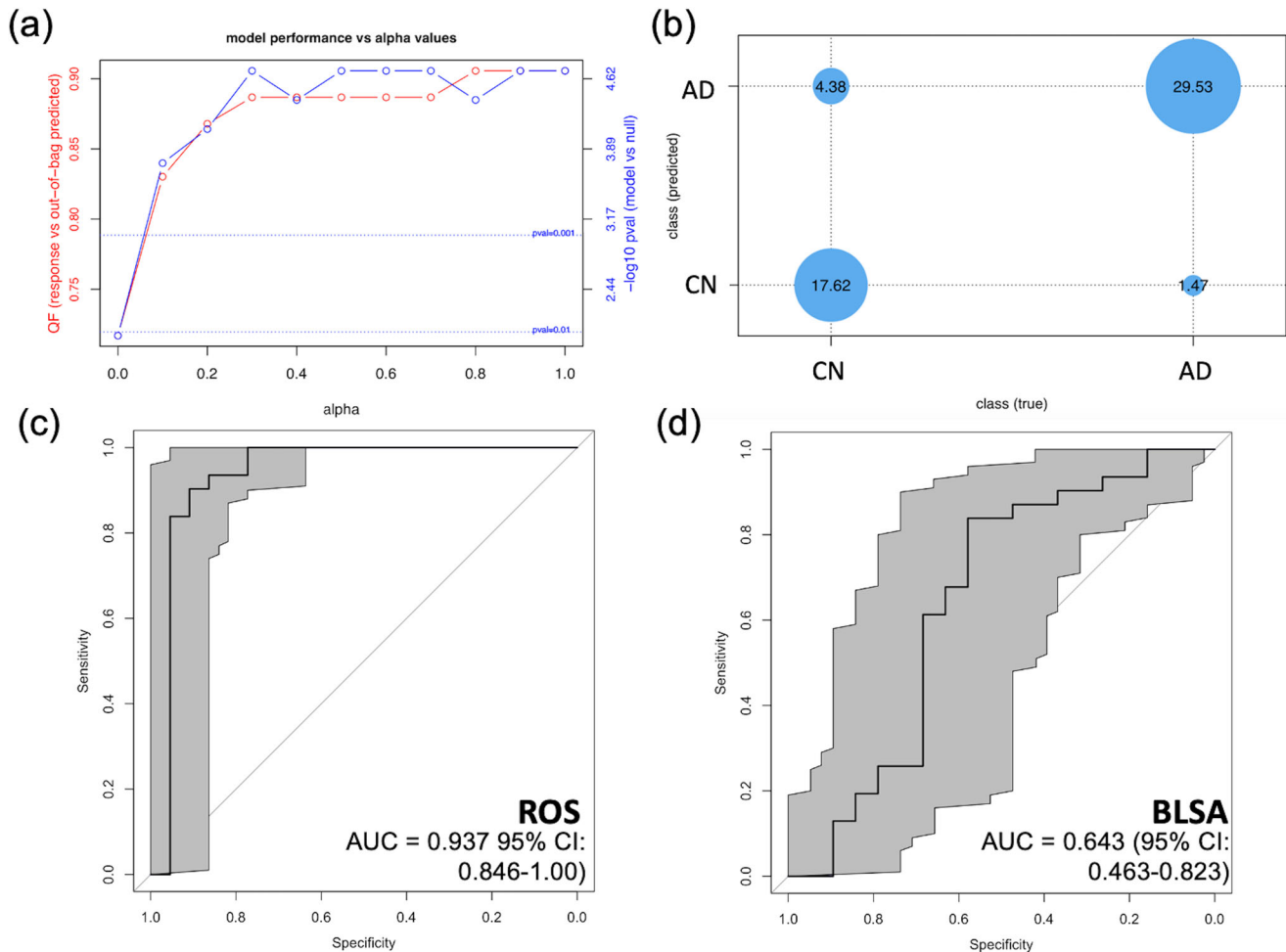


Fig. 2 Establishment and evaluation of ITG proteomic signature. **a** Output of eNetXplorer model selection analyses, performed in ROS ITG. The x-axis indicates the model α value, a mixing parameter that scales continuously from the ridge ($\alpha = 0$) to lasso ($\alpha = 1$) regularized regression models. The red line indicates the out-of-bag prediction performance using accuracy as quality function (QF), which captures the proportion of class (AD vs. CN) predictions that match the true class of each participant. The blue line indicates the estimated model significance for each α value. **b** Contingency matrix comparing the out-of-bag predicted class and true class for the selected model ($\alpha = 1$). **c** Receiver operating characteristic (ROC) curve for leave-one-out cross-validation (LOOCV) of the four protein ITG signature in the ROS cohort, derived at the $\alpha = 1$ level. Gray shading indicates the 95 percent confidence interval. **d** ROC curve for LOOCV of ITG samples from an independent cohort of BLSA participants. The area under the curve (AUC) is demonstrated with 95 percent confidence intervals. Gray shading indicates the 95 percent confidence interval.

AD participants, while IFNL2 was reduced. NOG was increased in BLSA AD participants. In blood serum at the time of diagnosis, none of the four proteins were significant in ROS or BLSA participants. Among all 1300 proteins assayed in the ITG, 150 were significantly different between AD and CN participants in the BLSA cohort, and 353 were significantly different in the ROS cohort (Supplementary Table 3).

Brain and blood serum protein correlations with AD pathology

We also sought to determine if proteins in the MFG signature correlated with Braak and CERAD scores, two measures of AD pathology. We performed partial correlation analyses between ITG and MFG protein levels of each of the four proteins with Braak and CERAD scores separately.

In the MFG, all four proteins included in the signature displayed strong positive correlations with both CERAD and Braak scores (Fig. 4). Similarly, ITG levels of most proteins from the signature displayed strong positive correlations with both scores of AD pathology (Supplementary Fig. 5). The only correlation that did

not achieve statistical significance was the association between ITG levels of SMOC1 and Braak ($p = 0.16$).

As in the ITG and MFG, we performed partial correlation analyses between blood serum levels of proteins in the signature at the time of diagnosis with scores of AD pathology. For the most part, correlations between blood serum proteins and AD pathology scores were not significant. However, NOG displayed a positive correlation with CERAD, and NTN1 displayed a negative association with Braak in the BLSA cohort only (Supplementary Fig. 6).

DISCUSSION

In this study, we utilized a machine learning approach to perform feature reduction with the goal of identifying a parsimonious set of proteins whose expression levels in the brain could accurately discriminate between AD and CN individuals. We utilized two independent cohorts of AD individuals with autopsy brain tissue samples, performing model training and feature selection in ROS and then testing the selected model in BLSA. This methodology

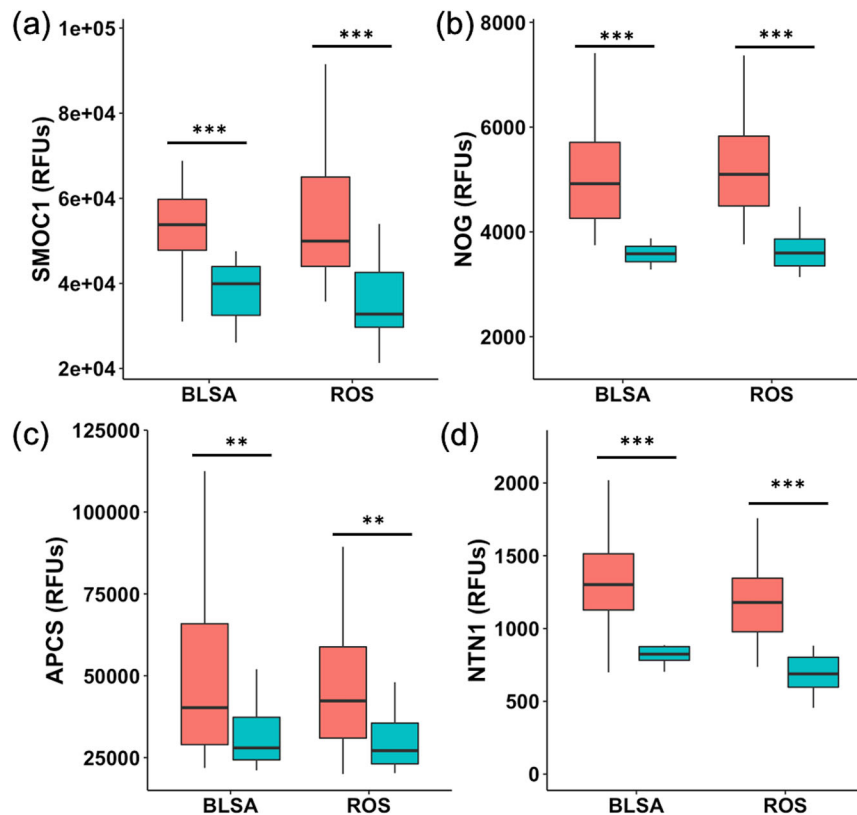


Fig. 3 MFG levels of proteins present in the proteomic signature. Comparison of MFG levels of **a** SMOC1, **b** NOG, **c** APCS, and **d** NTN1 between AD (red) and CN (blue) individuals in BLSA and ROS. Protein levels are on the y-axis in relative fluorescence units (RFUs). Statistical significance was calculated using sex and age-adjusted proportional odds models. Horizontal lines within each box represent the mean, and error bars represent the standard deviation. ** $p < 0.01$, *** $p < 0.001$.

allowed us to establish a four-protein signature (SMOC1, APCS, NOG, and NTN1) that accurately discriminated AD and CN MFG brain tissue samples in both ROS and BLSA cohorts. We then compared levels of these four proteins in MFG and ITG brain tissue samples at autopsy, as well as blood serum samples at the time of AD diagnosis. In both brain regions, these proteins consistently differed between AD and CN individuals. Notably, serum concentrations of APCS and NTN1 additionally differed between AD and CN individuals in the ROS cohort at the time of AD diagnosis, suggesting these brain proteomic alterations may extend to the peripheral system. Lastly, we sought to determine the correlation between proteins in this signature and the severity of AD pathology, identifying strong positive correlations between ITG and MFG levels of all four proteins and CERAD and Braak in ROS and BLSA. While relationships between blood serum levels and AD pathology were less consistent, NOG and NTN1 did correlate with AD pathology in BLSA samples.

The use of stringent feature reduction techniques to identify proteins that are highly associated with AD status allows for a more granular exploration of the potential mechanisms of the relationship between proteomic alterations and AD pathogenesis. APCS, or serum amyloid P component, is a pentraxin that has consistently been identified in CSF, as well as bound to amyloid deposits and neurofibrillary tangles, in individuals with AD^{12,13}. APCS has been shown to stabilize amyloid plaques, prevent proteolysis, and promotes deposition of amyloid *in vitro*^{14,15}. In animal models and cell culture experiments, APCS has further been shown to induce neuronal apoptosis^{16,17} and plays a role in the immunological response via its effects on macrophage polarization¹⁸. More systemically, APCS functions alongside CRP as a key mediator of the innate immune response, binding to pattern-recognition sequences on microbes, as well as recognizing

damaged tissue^{19,20}. Both APCS and CRP activate the complement activation cascade, with APCS binding C1q and C4b proteins²¹. APCS, and other pentraxins, may therefore represent a possible mediator between long-standing inflammation and AD, further contributing to hallmark features of pathology. Indeed, studies have shown that APCS deposition on neurofibrillary tangles occurs prior to the activation of complement, thereby contributing to local inflammation and resultant neuronal loss²².

SMOC1 is a member of the SPARC matricellular protein family with a role in the regulation of the interaction between cell matrices via the binding of cell surface receptors²³. SMOC1 has been discussed as a possible mediator of inflammation, as nitric oxide inhibits SMOC1 function, which in turn results in reduced TGF- β signaling²⁴. SMOC1 appears to be consistently identified as a key protein in AD, as multiple proteomics studies have found elevations of SMOC1 in brain tissue and CSF, as well as associations with amyloid pathology^{25–27}. Such consistent findings suggest that SMOC1 likely represents a plausible biomarker for AD that is sensitive as a marker of AD pathology. SMOC1 also plays a role as a pro-angiogenic factor²⁸, which has frequently been proposed as a biological process perturbed in AD²⁹.

NOG is a protein with a role in the development of multiple tissues, including neural cells, and is a member of the TGF- β protein family, acting as an antagonist of bone morphogenetic protein-4 (BMP-4). Prior studies have identified the role of NOG in the regeneration of neurons following injury³⁰. Others have found that NOG signaling through BMP changes throughout the lifecycle, finding that age-related increases in BMP signaling are associated with impaired neurogenesis that may be buffered by NOG³¹. Animal studies additionally suggest that NOG can rescue hippocampal neurogenesis and behavioral deficits³². However, in this study, we identified an increase in NOG in AD

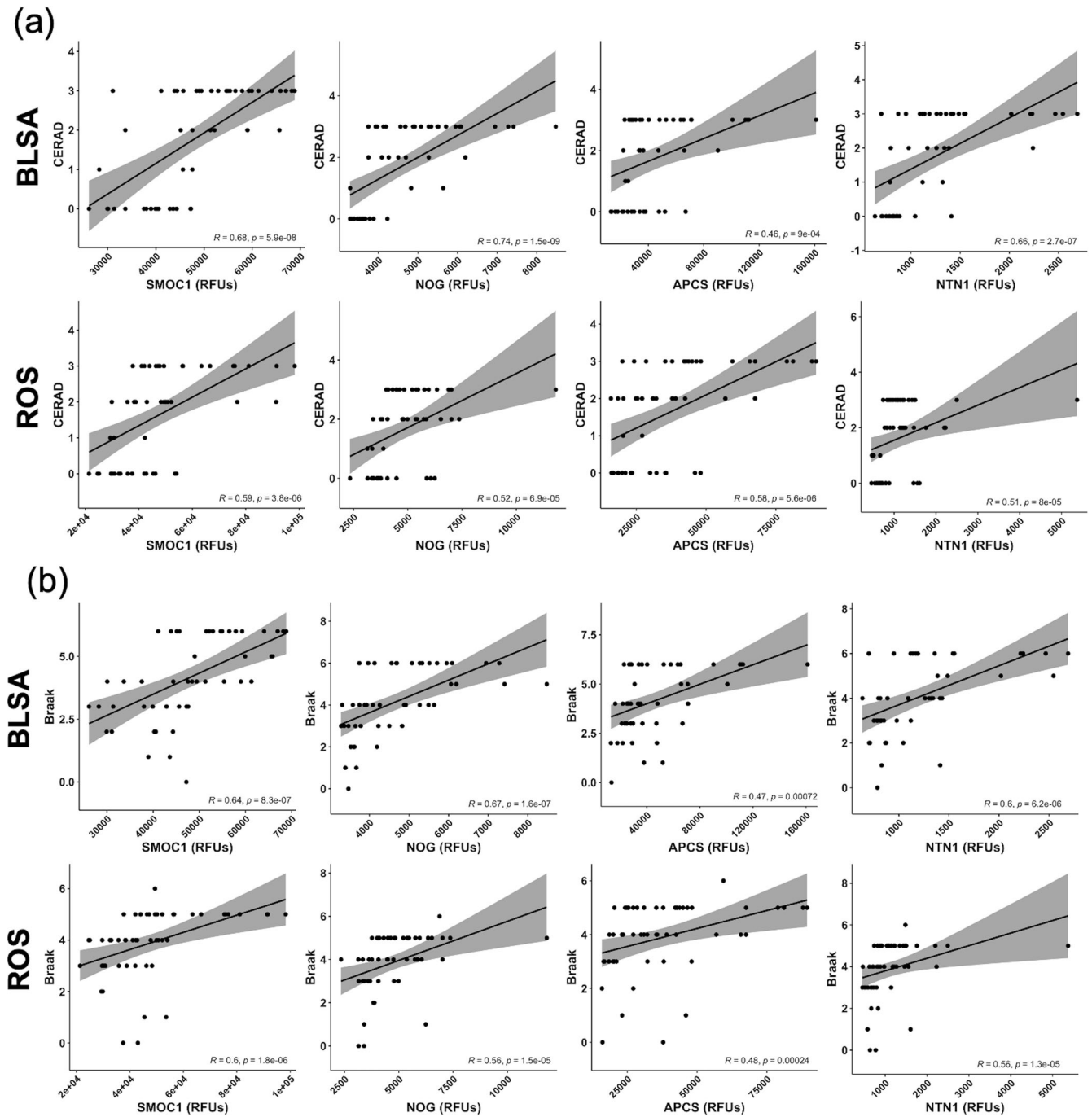


Fig. 4 Correlation of MFG protein levels with AD pathology. Partial Spearman correlation, adjusted for sex and age at sampling, between MFG protein levels and **a** CERAD and **b** Braak scores in BLSA and ROS. Protein levels are given in relative fluorescence units (RFUs) on the x-axis. Shading indicates the 95 percent confidence interval of the estimate.

and a positive correlation with pathology, which may suggest that NOG elevation may represent an attempt to promote neurogenesis in response to neuronal loss in AD. Previously, NOG has been found to be hypermethylated in AD, suggesting that genetic modification leading to an imbalance of NOG and BMP-4 may be a feature of disease³³.

NTN1 belongs to a family of proteins called netrins that play a role in neuronal axon guidance through modulation of microtubular function and additionally function to prevent apoptosis initiation. NTN1 has previously been proposed as a therapeutic for AD after the finding that administration of NTN1 to transgenic mice reduced amyloid beta and improved working memory³⁴.

Additionally, NTN1 gene expression in blood has previously been found to be reduced in mild cognitive impairment (MCI)³⁵. Interestingly, we found NTN1 to be reduced in blood serum in ROS individuals with AD at the time of diagnosis yet increased in BLSA and ROS postmortem samples and positively correlated with AD pathology. This may suggest that early in the disease course, reductions in NTN1 could contribute to AD pathogenesis, while later increases in expression could indicate a response to the disease.

In summary, we have undertaken a unique multivariable regression approach to identify a reduced set of proteins that could discriminate between AD and CN brain tissue samples. We

achieved this by deriving a proteomic signature in one cohort, ROS, before testing the discriminatory signature in an independent cohort, BLSA. This signature was identified despite significant demographic differences across cohorts, suggesting that the identified protein signature may be associated with disease independent of the study sample. This process identified four proteins in the MFG brain region that discriminated accurately between AD and CN samples in both cohorts that may represent key features of AD pathogenesis. Despite strong replication in brain tissue, however, the MFG signature failed to replicate in blood serum obtained from participants at a baseline visit. This suggests the signature is unlikely to have significant utility as a serum screening biomarker in pre-clinical populations. It may specifically reflect mechanisms of disease pathophysiology in brain tissue, though exploration of these proteins in cerebrospinal fluid may further evaluate their utility as AD biomarkers.

In contrast, the ITG protein signature identified in ROS failed to discriminate between AD and CN in BLSA. This may be due to a number of demographic differences between the two cohorts, as ROS participants were on average older, more often female, and with shorter disease duration as compared to BLSA. In addition, while CERAD scores, reflective of amyloid pathology, were similar across cohorts, AD participants in the BLSA had, on average, higher Braak scores, reflective of greater accumulated neurofibrillary tangle pathology. The ITG was specifically selected in this study as a brain region vulnerable to tau-related neurofibrillary tangle pathology, while the MFG was selected as a brain region vulnerable to amyloid accumulation^{36,37}. Numerous studies have suggested that amyloid beta precedes and may induce tau aggregation, and distinct neurodegenerative alterations occur when tau and amyloid colocalize^{38–40}. Our findings of a strong MFG-specific proteomic signature are, therefore, most likely to represent a signature of AD that is closely related to amyloid pathology, while the failure of the ITG signature to replicate may reflect varying severity of neurofibrillary tangle pathology across cohorts. Indeed, SMOC1 and NTN1 were recently identified among the most highly-enriched proteins in amyloid plaques in early-onset AD and individuals with Down syndrome and AD⁴¹. Similarly, both proteins were also among the most highly-enriched proteins in amyloid plaques in late-onset AD⁴², strongly suggesting the MFG signature is reflective of proteomic alterations related to amyloid plaque formation. In addition, MDK, which was present in the ITG signature that also replicated in the MFG, was identified as a highly-enriched protein in amyloid plaques in both of these prior studies^{41,42}.

In sum, we have applied a rigorous statistical methodology to derive an MFG protein signature discriminating between AD and CN participants in two independent cohorts that may have significant specificity to amyloid pathology. This study confirms prior work identifying SMOC1 and NTN1 enrichment within amyloid plaques and extends results to suggest that NOG and APCS are amyloid plaque-related proteins warranting further evaluation. The approach taken here may provide further utility in evaluating larger datasets of blood and CSF samples throughout the course of AD development in order to identify improved biomarkers or earlier targets for therapeutics. Additionally, this approach may be adapted as even greater multi-dimensional datasets emerge, allowing researchers to parse through such datasets for rapid identification of proteins that may most substantially contribute to disease outcomes.

METHODS

Participants: Baltimore Longitudinal Study of Aging (BLSA)

The National Institute on Aging's (NIA) BLSA is among the longest-running scientific studies of aging in the United States⁴³. This observational study began in 1958 and includes longitudinal,

radiological, clinical, and laboratory evaluations of community-dwelling volunteer participants. The individuals included in this study were participants in the autopsy sub-study of the BLSA, which has been described previously⁴⁴. Postmortem brains were inspected by an expert neuropathologist to assess AD pathology. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) and Braak criteria were used to assess the severity of AD pathology based on neuritic plaques⁴⁵ and neurofibrillary tangles⁴⁶, respectively, as described previously in ref. ⁴⁷. Clinical diagnosis of dementia and AD have previously been described⁴⁸ and were based on the Diagnostic and Statistical Manual (DSM)-III-R⁴⁹ and the National Institute of Neurological and Communication Disorders and Stroke–Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria, respectively⁵⁰.

Diagnosis and cognitive status were determined at consensus diagnosis conferences using procedures described in detail previously⁴⁸. Briefly, autopsy participants were classified as either AD or CN according to the following criteria: AD participants had a clinical diagnosis of AD or mild cognitive impairment (MCI) within 1 year of death in addition to a postmortem CERAD pathology score >1 (i.e., CERAD B or C); cognitively normal (CN) participants had normal cognition within 1 year of death and a CERAD pathology score ≤1 (i.e., CERAD 0 or A). Demographic characteristics of the BLSA cohort are included in Table 1. The BLSA study protocol has ongoing approval from the Institutional Review Board of the National Institute of Environmental Health Science, the National Institutes of Health, and informed consent was obtained from all participants.

Participants: Religious Orders Study (ROS)

The Religious Orders Study (ROS) has enrolled Catholic nuns, priests, and brothers from a multitude of communities across the United States since 1994⁵¹. This longitudinal observational study collects information from clinical, neuroimaging, laboratory, and self-report evaluations of employed and retired community-dwelling individuals. At the time of enrollment, participants are without a diagnosis of known dementia. All participants agreed to organ donation and annual clinical evaluation.

Our sample consisted of a subset of participants from the larger ROS cohort study. All ROS participants provided written informed consent and the study was approved by an Institutional Review Board of Rush University Medical Center. Participants signed an Anatomical Gift Act for organ donation and a repository consent to allow their data and biospecimens to be shared. At each study visit, dementia status was determined by trained clinicians using all cognitive and clinical data blinded to prior years based on NINCDS-ADRDA criteria. A final consensus clinical diagnosis was determined at death, blinded to all neuropathologic data. Autopsies were performed based on standard methods reported using 4% paraformaldehyde-fixed 1-cm sections of brain tissue for neuropathology analyses⁵². Postmortem brains were examined by an expert neuropathologist or trained technician to assess AD pathology. CERAD and Braak criteria were used to assess the severity of AD pathology, as described previously⁵³.

Participants were classified into two groups. AD participants had a final clinical diagnosis of AD and an NIA-Reagan score of an intermediate or high likelihood of AD. NIA-Reagan criteria are based on both neuritic plaques (CERAD score) and neurofibrillary tangles (Braak score)⁵⁴. CN participants had a clinical diagnosis of no cognitive impairment (NCI) and an NIA-Reagan score of a low likelihood of AD or no AD. Diagnosis and cognitive status were determined based on a three-stage process described previously⁵¹, which included a medical history, neurological examination, neuropsychiatric testing, and review of neuroimaging data when present. At death, clinical data were reviewed in order to make a likely clinical diagnosis.

Brain tissue collection and homogenization

BLSA (AD = 31, CN = 19) and ROS (AD = 31, CN = 22) brain tissue samples were selected from two a priori specified regions: the inferior temporal gyrus (ITG) and middle frontal gyrus (MFG). The ITG was selected to represent a brain region vulnerable to neurofibrillary tangle deposition and the MFG was selected to represent a region vulnerable to β -amyloid accumulation^{36,37}. Approximately 4-mm-diameter tissue punches were extracted from the cortical surface of the brain tissue regions using a sterile technique and stored at -80°C prior to proteomic assays.

To ~10 mg of brain tissue, 110 μl of T-PER (tissue protein extraction reagent) (Thermo Fisher Scientific, USA) with 2 μl of Halt™ Protease and Phosphatase Inhibitor cocktail (Thermo Fisher Scientific, USA) was added and placed in a CKMix grinding tube, containing soft tissue homogenizing lysis beads (a mix of 1.4 mm and 2.8 mm ceramic (zirconium oxide) beads) (Bertin Technologies, San Quentin, France). The tubes were placed in the Precellys Evolution tissue homogenizer (Bertin Technologies, San Quentin, France) and homogenized for two 30 s cycles of 6500 rpm and a 30 s rest in between. The homogenate was removed and placed in an Eppendorf tube and centrifuged at 16,000 $\times g$ for 5 min at 4 $^{\circ}\text{C}$. The supernatant was removed and centrifuged a second time for 10 min at 16,000 $\times g$ at 4 $^{\circ}\text{C}$. The supernatant was collected at 4 $^{\circ}\text{C}$, 2.5 μl was aliquoted, and protein quantitation was carried out using a MicroBCA Protein Assay Kit (Thermo Scientific, USA). The total protein concentration was determined, and the samples were diluted to a final volume of 200 $\mu\text{g}/\text{ml}$ with PBS 1X and stored at -80°C until analysis.

Blood serum collection

Blood serum samples were collected from BLSA participants (AD = 26, CN = 20) at time of AD diagnosis, 43 of whom were included in the autopsy sample, at the NIA Clinical Research Unit in Harbor Hospital, Baltimore. Details on collection and processing have been published previously⁵⁵. Briefly, venous blood samples were collected between 6 and 7 a.m. following an overnight fast. Serum samples were aliquoted into 0.5-mL volume in Nunc cryogenic tubes and stored at -80°C until further use. Samples were not subject to any freeze–thaw cycles prior to proteomic assays. Additional details on sample selection have been published previously⁵⁵. ROS participants (AD = 29, CN = 22) provided blood serum at the time of AD diagnosis, all of whom were included in the autopsy sample, as described previously⁵⁶.

SOMAscan proteomic quantification

Sample total protein was adjusted to 16 $\mu\text{g}/\text{mL}$ in SB17T buffer (40 mM HEPES, 125 mM NaCl, 5 mM KCl, 5 mM MgCl_2 , 1 mM EDTA, 0.05% Tween-20 at pH 7.5). Proteomic profiles for 1322 SOMAmers were assessed using the 1.3 K SOMAscan assay (SomaLogic, Inc., Boulder, CO, USA) at the Trans-NIH Center for Human Immunology and Autoimmunity, and Inflammation (CHI), National Institute of Health (Bethesda, MD, USA). The SOMAscan assay platform includes 1322 SOMAmer Reagents, of which 12 are hybridization controls, five are viral proteins, and five are non-specifically-targeted SOMAmers. As a result, analyses spanned 1300 SOMAmer Reagents. The proteins targeted by SOMAmers included in this assay are shown in Supplementary Table 1. The experimental procedure for proteomic assessment and normalization has been previously reported⁵⁷. Briefly, the SOMAscan assay uses SOMAmers to translate protein concentrations into measurable DNA signals, which can be quantified using standard DNA detection procedures. This is achieved by affinity binding and biotin capture on streptavidin beads. The DNA concentrations obtained from this method are reported as relative fluorescence units (RFUs), resulting from fluorescent SOMAmer hybridized to its

complementary probe on an Agilent array, and are directly proportional to the reported relative abundance of SOMAmer Reagents. The data normalization process includes hybridization, control normalization, median signal normalization, and calibration normalization, as previously described in refs. ^{57,58}.

Study/cohort-specific samples were run in the same batch. Within the study/cohort, sample ordering was randomized by disease (AD and CN), for brain tissue by brain region (ITG and MFG), sex, and age. SomaScan reports protein abundances measured in RFU with no missing or below limit of detection (LOD) values. Following standard SomaScan normalization procedures⁵⁸, RFU data were processed using hybridization normalization, median signal normalization, plate-scale normalization, and inter-plate calibration, which were designed to remove nuisance variability due to differences in loading volume, leaks, washing conditions, microarray hybridization, and other sources of intra- and inter-plate technical effects. Additional quality checks based on principal component analysis were performed at each normalization step; the fully normalized data used for downstream analysis showed no evidence of significant technical bias. Outliers were not excluded.

Statistical analysis: demographic characteristics

Demographic characteristics are summarized in Table 1. Comparisons between the BLSA and ROS cohorts were performed using two-sample *t*-tests for continuous variables and chi-square tests for categorical variables.

Statistical analysis: brain proteomic signature identification

To establish a brain proteomic signature discriminating AD and CN brain samples, we utilized a two-step process in which we first identified a signature in ROS and then validated the signature independently in BLSA. We performed the signature identification step separately in ROS ITG and MFG brain tissue samples utilizing binomial classification in eNetXplorer⁸, an R package, which tests the accuracy and significance of a family of elastic net generalized linear models ranging from ridge regression ($\alpha = 0$) to lasso regression ($\alpha = 1$)⁷. For analysis in eNetXplorer, the outcome was a binary group variable (AD vs. CN) with predictors of \log_{10} -transformed aptamer relative fluorescence units (RFU) in the brain and covariates of sex and age at death. The elastic net mixing parameter α was scanned from 0 to 1 in 0.2 intervals. The α model best discriminating between AD and CN in ROS samples was selected based on overall performance assessed as a function of the fivefold cross-validated quality function (accuracy) and the empirical *p* value estimated from comparing the model against a statistical ensemble of null models generated by random permutations of the response (i.e., AD and CN class labels). This procedure required the evaluation of a large number of elastic net realizations arising from 500 randomly generated fivefold cross-validation runs, each of them compared against 250 null model permutations, in which participant class labels (AD or CN) are randomly assigned. Model performance is quantified by a quality function (i.e., classification accuracy), and the comparison between the model performance and null model performance distribution is quantified with an empirical *p* value. To identify proteins defining a discriminative proteomic signature, we selected proteins with a significant (*p* value <0.05) feature frequency (frequency with which a protein was selected by the model) or feature coefficient in the selected α model. Proteins identified in this manner by eNetXplorer are those that discriminate most accurately between AD and CN brain samples in ROS. False-discovery rate corrections were not applied to selected protein features as such methodologies assume statistical independence, whereas eNetXplorer preserves the data covariance structure across feature space.

After identifying proteins accurately discriminating between AD and CN brain samples in ROS with eNetXplorer, we validated the

discriminative ability of these proteins by applying leave-one-out cross-validation (LOOCV) with logistic regression models in an independent cohort of BLSA brain samples. In LOOCV, one sample is held out from the data as a test set and all other samples are used in the training set. This procedure is repeated until all samples in the dataset have been used as the test set⁵⁹. LOOCV was performed separately within each brain region: i.e., the ROS ITG brain proteomic signature was applied to BLSA ITG tissue samples, and the ROS MFG signature was applied to BLSA MFG samples. Prior to model evaluation, we performed three inter-quartile range (IQR) winsorization, in which protein values outside the 3 IQR range were “winsorized” and set equal to the three IQR limit⁶⁰. Logistic regression models included the group outcome (AD vs. CN), and predictors included levels of proteins in the brain proteomic signature, as well as covariates—sex and age. Results of LOOCV were visualized as receiver operating characteristic (ROC) curves, from which we calculated area under the curve (AUC) with 95 percent confidence intervals.

Statistical analysis: protein abundance comparisons

To identify differences in levels of proteins identified by eNetXplorer between AD and CN participants, we used proportional odds models⁶¹, a generalization of the non-parametric Wilcoxon and Kruskal-Wallis tests. All proportional odds models included ranked protein levels (outcome), the group predictor (AD vs. CN), and covariates—the age at sampling and sex. Models were run separately for ROS and BLSA cohorts. For each cohort, three sets of cross-sectional proportional odds models were performed, one for each protein sampling matrix: ITG brain tissue, MFG brain tissue, and blood serum at the time of diagnosis. The significance for between-group comparisons was set as $p < 0.05$.

Statistical analysis: associations with AD pathology

We also sought to determine if blood serum and brain levels of proteins in the brain proteomic signature were associated with the severity of AD pathology in the ITG and MFG. As in prior work⁶², we examined partial Spearman correlations of CERAD and Braak scores with ranked aptamer values, controlling for covariates—sex and age at sampling. Correlations were performed separately in ROS and BLSA for ITG, MFG, and serum protein levels at the time of samples concurrent (<1 year) with AD diagnosis and age-matched controls. A significant (at a p value threshold of 0.05) positive correlation indicated that a higher concentration of the protein was associated with higher AD pathology (i.e., higher CERAD or Braak scores) and, conversely, a significant negative correlation indicated that a lower concentration of the protein was associated with higher AD pathology.

Reporting summary

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

DATA AVAILABILITY

NIH and BLSA policy prohibits the deposition of the proteomic data associated with this manuscript in a public repository. However, unprocessed SomaLogic proteomic data from the Baltimore Longitudinal Study of Aging (BLSA) are available to researchers and can be requested at <https://blsa.nih.gov/researchers>, in accordance with BLSA and NIH policy. SomaLogic proteomic data from the Religious Orders Study (ROS) are available to researchers at www.radc.rush.edu.

CODE AVAILABILITY

All analyses were performed in R Studio version 1.3.1073 using publicly-available packages as described in the Methods. Requests for code utilized in this manuscript may be made to the corresponding author.

REFERENCES

- Roter, A. H. Large-scale integrated databases supporting drug discovery. *Curr. Opin. Drug Discov. Dev.* **8**, 309–315 (2005).
- Altaf-UI-Amin, M., Afendi, F. M., Kiboi, S. K. & Kanaya, S. Systems biology in the context of big data and networks. *Biomed. Res. Int.* **2014**, 428570 (2014).
- Karbalaei, R., Allahyari, M., Rezaei-Tavirani, M., Asadzadeh-Aghdaei, H. & Zali, M. R. Protein-protein interaction analysis of Alzheimer's disease and NAFLD based on systems biology methods unhide common ancestor pathways. *Gastroenterol. Hepatol. Bed Bench* **11**, 27–33 (2018).
- Zhang, Q. et al. Integrated proteomics and network analysis identifies protein hubs and network alterations in Alzheimer's disease. *Acta Neuropathol. Commun.* **6**, 19 (2018).
- Johnson, E. C. B. et al. Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat. Med.* **26**, 769–780 (2020).
- Araujo, R. P., Liotta, L. A. & Petricoin, E. F. Proteins, drug targets and the mechanisms they control: the simple truth about complex networks. *Nat. Rev. Drug Discov.* **6**, 871–880 (2007).
- Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Series B Stat. Methodol.* **67**, 301–320 (2005).
- Candia, J. & Tsang, J. S. eNetXplorer: an R package for the quantitative exploration of elastic net families for generalized linear models. *BMC Bioinformatics* **20**, 189 (2019).
- Candia, J. et al. The genomic landscape of Mongolian hepatocellular carcinoma. *Nat. Commun.* **11**, 4383 (2020).
- Pat, N., Wang, Y., Bartonicek, A., Candia, J. & Stringaris, A. Explainable machine learning approach to predict and explain the relationship between task-based fMRI and individual differences in cognition. *Cereb. Cortex* **33**, 2682–2703 (2022).
- Minas, T. Z. et al. Serum proteomics links suppression of tumor immunity to ancestry and lethal prostate cancer. *Nat. Commun.* **13**, 1759 (2022).
- Hawkins, P. N. et al. Concentration of serum amyloid P component in the CSF as a possible marker of cerebral amyloid deposits in Alzheimer's disease. *Biochem. Biophys. Res. Commun.* **201**, 722–726 (1994).
- Kalaria, R. N., Galloway, P. G. & Perry, G. Widespread serum amyloid P immunoreactivity in cortical amyloid deposits and the neurofibrillary pathology of Alzheimer's disease and other degenerative disorders. *Neuropathol. Appl. Neurobiol.* **17**, 189–201 (1991).
- Hamazaki, H. Amyloid P component promotes aggregation of Alzheimer's beta-amyloid peptide. *Biochem. Biophys. Res. Commun.* **211**, 349–353 (1995).
- Tennent, G. A., Lovat, L. B. & Pepys, M. B. Serum amyloid P component prevents proteolysis of the amyloid fibrils of Alzheimer disease and systemic amyloidosis. *Proc. Natl Acad. Sci. USA* **92**, 4299–4303 (1995).
- Urbanyi, Z., Lakics, V. & Erdo, S. L. Serum amyloid P component-induced cell death in primary cultures of rat cerebral cortex. *Eur. J. Pharmacol.* **270**, 375–378 (1994).
- Pisalyaput, K. & Tenner, A. J. Complement component C1q inhibits beta-amyloid- and serum amyloid P-induced neurotoxicity via caspase- and calpain-independent mechanisms. *J. Neurochem.* **104**, 696–707 (2008).
- Killock, D. Connective tissue diseases: SAP-induced macrophage polarization: a potential therapeutic option for SLE? *Nat. Rev. Rheumatol.* **7**, 497 (2011).
- Poon, I. K., Hulett, M. D. & Parish, C. R. Molecular mechanisms of late apoptotic/necrotic cell clearance. *Cell Death Differ* **17**, 381–397 (2010).
- Du Clos, T. W. & Mold, C. Pentraxins (CRP, SAP) in the process of complement activation and clearance of apoptotic bodies through Fc gamma receptors. *Curr. Opin. Organ Transplant.* **16**, 15–20 (2011).
- Sjoberg, A. P., Trouw, L. A., McGrath, F. D., Hack, C. E. & Blom, A. M. Regulation of complement activation by C-reactive protein: targeting of the inhibitory activity of C4b-binding protein. *J. Immunol.* **176**, 7612–7620 (2006).
- Veerhuis, R. Histological and direct evidence for the role of complement in the neuroinflammation of AD. *Curr. Alzheimer Res.* **8**, 34–58 (2011).
- Vannahme, C. et al. Characterization of SMOG-1, a novel modular calcium-binding protein in basement membranes. *J. Biol. Chem.* **277**, 37977–37986 (2002).
- Westenfeld, R. et al. Selective inhibition of inducible nitric oxide synthase enhances intraglomerular coagulation in chronic anti-Thy 1 nephritis. *Kidney Int.* **61**, 834–838 (2002).
- Johnson, E. C. B. et al. Large-scale deep multi-layer analysis of Alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level. *Nat. Neurosci.* **25**, 213–225 (2022).
- Bai, B. et al. Deep multilayer brain proteomics identifies molecular networks in Alzheimer's disease progression. *Neuron* **106**, 700 (2020).
- Zhou, M. et al. Targeted mass spectrometry to quantify brain-derived cerebrospinal fluid biomarkers in Alzheimer's disease. *Clin. Proteomics* **17**, 19 (2020).

28. Awwad, K. et al. Role of secreted modular calcium-binding protein 1 (SMOC1) in transforming growth factor beta signalling and angiogenesis. *Cardiovasc. Res.* **106**, 284–294 (2015).
29. Vagnucci, A. H. Jr. & Li, W. W. Alzheimer's disease and angiogenesis. *Lancet* **361**, 605–608 (2003).
30. Cho, Y. E. et al. Older age results in differential gene expression after mild traumatic brain injury and is linked to imaging differences at acute follow-up. *Front. Aging Neurosci.* **8**, 168 (2016).
31. Yousef, H. et al. Age-associated increase in BMP signaling inhibits hippocampal neurogenesis. *Stem Cells* **33**, 1577–1588 (2015).
32. Diaz-Moreno, M. et al. Noggin rescues age-related stem cell loss in the brain of senescent mice with neurodegenerative pathology. *Proc. Natl Acad. Sci. USA* **115**, 11625–11630 (2018).
33. Ren, J., Zhang, B., Wei, D. & Zhang, Z. Identification of methylated gene biomarkers in patients with Alzheimer's disease based on machine learning. *Biomed. Res. Int.* **2020**, 8348147 (2020).
34. Spilman, P. R. et al. Netrin-1 interrupts amyloid-beta amplification, increases sABPPa in vitro and in vivo, and improves cognition in a mouse model of Alzheimer's disease. *J. Alzheimers Dis.* **52**, 223–242 (2016).
35. Park, M. K. et al. NLRP1 and NTN1, deregulated blood differentially methylated regions in mild cognitive impairment patients. *J. Mol. Neurosci.* **66**, 561–571 (2018).
36. Li, Y. et al. Regional analysis of FDG and PIB-PET images in normal aging, mild cognitive impairment, and Alzheimer's disease. *Eur. J. Nucl. Med. Mol. Imaging* **35**, 2169–2181 (2008).
37. Buckley, R. F. et al. Region-specific association of subjective cognitive decline with tauopathy independent of global beta-amyloid burden. *JAMA Neurol.* **74**, 1455–1463 (2017).
38. Stancu, I. C., Vasconcelos, B., Terwel, D. & Dewachter, I. Models of beta-amyloid induced Tau-pathology: the long and "folded" road to understand the mechanism. *Mol. Neurodegener.* **9**, 51 (2014).
39. Wennstrom, M. et al. Cellular localization of p-tau217 in brain and its association with p-tau217 plasma levels. *Acta Neuropathol. Commun.* **10**, 3 (2022).
40. Busche, M. A. & Hyman, B. T. Synergy between amyloid-beta and tau in Alzheimer's disease. *Nat. Neurosci.* **23**, 1183–1193 (2020).
41. Drummond, E. et al. The amyloid plaque proteome in early onset Alzheimer's disease and Down syndrome. *Acta Neuropathol. Commun.* **10**, 53 (2022).
42. Xiong, F., Ge, W. & Ma, C. Quantitative proteomics reveals distinct composition of amyloid plaques in Alzheimer's disease. *Alzheimers Dement.* **15**, 429–440 (2019).
43. Ferrucci, L. The Baltimore longitudinal study of aging (BLSA): a 50-year-long journey and plans for the future. *J. Gerontol. A Biol. Sci. Med. Sci.* **63**, 1416–1419 (2008).
44. O'Brien, R. J. et al. Neuropathologic studies of the Baltimore longitudinal study of aging (BLSA). *J. Alzheimers Dis.* **18**, 665–675 (2009).
45. Mirra, S. S. et al. The consortium to establish a registry for Alzheimer's disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479–486 (1991).
46. Braak, H. & Braak, E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259 (1991).
47. Troncoso, J. C. et al. Effect of infarcts on dementia in the Baltimore longitudinal study of aging. *Ann. Neurol.* **64**, 168–176 (2008).
48. Kawas, C., Gray, S., Brookmeyer, R., Fozard, J. & Zonderman, A. Age-specific incidence rates of Alzheimer's disease: the Baltimore longitudinal study of aging. *Neurology* **54**, 2072–2077 (2000).
49. APA. *Diagnostic and Statistical Manual of Mental Disorders: DSM-III-R.* (American Psychiatric Association, 1987).
50. McKhann, G. et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939–944 (1984).
51. Bennett, D. A. et al. Religious orders study and rush memory and aging project. *J. Alzheimers Dis.* **64**, S161–S189 (2018).
52. Schneider, J. A. et al. Relation of cerebral infarctions to dementia and cognitive function in older persons. *Neurology* **60**, 1082–1088 (2003).
53. Bennett, D. A. et al. Apolipoprotein E epsilon4 allele, AD pathology, and the clinical expression of Alzheimer's disease. *Neurology* **60**, 246–252 (2003).
54. Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. The National Institute on Aging, and Reagan Institute Working Group on diagnostic criteria for the neuropathological assessment of Alzheimer's disease. *Neurobiol. Aging* **18**, S1–S2 (1997).
55. Casanova, R. et al. Blood metabolite markers of preclinical Alzheimer's disease in two longitudinally followed cohorts of older individuals. *Alzheimers Dement.* **12**, 815–822 (2016).
56. Tanaka, T. et al. Plasma proteomic signatures predict dementia and cognitive impairment. *Alzheimers Dement.* **6**, e12018 (2020).
57. Candia, J. et al. Assessment of variability in the SOMAScan assay. *Sci. Rep.* **7**, 14248 (2017).
58. Candia, J., Daya, G. N., Tanaka, T., Ferrucci, L. & Walker, K. A. Assessment of variability in the plasma 7k SomaScan proteomics assay. *Sci. Rep.* **12**, 17147 (2022).
59. Baek, S., Tsai, C. A. & Chen, J. J. Development of biomarker classifiers from high-dimensional data. *Brief. Bioinform.* **10**, 537–546 (2009).
60. Thomas, J. W. & Ward, K. Economic profiling of physician specialists: use of outlier treatment and episode attribution rules. *Inquiry* **43**, 271–282 (2006).
61. Roberts, J. A. et al. Blood metabolite signature of metabolic syndrome implicates alterations in amino acid metabolism: findings from the Baltimore longitudinal study of aging (BLSA) and the Tsuruoka metabolomics cohort study (TMCS). *Int. J. Mol. Sci.* **21**, 1249 (2020).
62. Roberts, J. A. et al. A brain proteomic signature of incipient Alzheimer's disease in young APOE epsilon4 carriers identifies novel drug targets. *Sci. Adv.* **7**, eabi8178 (2021).

ACKNOWLEDGEMENTS

We are grateful to BLSA and ROS participants for their invaluable contributions. The funders had no role in study design, data collection and analysis, the decision to publish, or preparation of the manuscript. This work was supported by NIH grants P30AG10161 (D.A.B.), R01AG15819 (D.A.B.); Andrew and Lillian A. Posey Foundation (M.T.); and Intramural Research Program of the NIH, National Institute on Aging (M.T.).

AUTHOR CONTRIBUTIONS

J.A.R.: visualization, writing—original draft, formal analysis, investigation, and conceptualization; V.R.V.: writing—original draft, investigation, methodology, data curation, and conceptualization; J.C.: visualization, conceptualization, methodology, and writing—original draft; T.T.: methodology, data curation, and writing—review and editing; L.F.: writing—review and editing and project administration; D.A.B.: data curation, methodology, project administration, writing—review and editing, and funding acquisition; M.T.: project administration, methodology, writing—review and editing, funding acquisition, and conceptualization.

FUNDING

Open Access funding provided by the National Institutes of Health (NIH).

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41514-023-00112-6>.

Correspondence and requests for materials should be addressed to Jackson A. Roberts or Madhav Thambisetty.

Reprints and permission information is available at <http://www.nature.com/reprints>

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023