Proteomic Clusters Underlie Heterogeneity in Preclinical AD Progression

Julie K. Wisch PhD[1], Omar H. Butt, MD, PhD[1], Brian A. Gordon PhD[2,3,6], Suzanne E. Schindler, MD, PhD[1,6], Anne M. Fagan, PhD[1,6], Chengran Yang PhD[4, 5], Anna H. Boerwinkle BS[1], Tammie L.S. Benzinger MD, PhD[2,6], David M. Holtzman MD[1, 3, 6], John C. Morris MD[1,6], Carlos Cruchaga PhD[3,6], Beau M. Ances MD, PhD[1,2,3,6]

1. Department of Neurology, Washington University in St. Louis, St. Louis, MO 63110, USA
2. Department of Radiology, Washington University in St. Louis St. Louis, MO 63110, USA
3. Hope Center, Washington University in Saint Louis, St. Louis, MO 63110, USA
4. Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA
5. NeuroGenomics and Informatics Center, Washington University School of Medicine, St Louis, MO, USA
6. Knight Alzheimer Disease Research Center, Washington University School of Medicine, St Louis, MO 63110, USA

Corresponding Author:

Beau M Ances, MD, PhD, MSc

Department of Neurology

Washington University in Saint Louis School of Medicine

Campus Box 8111, 660 South Euclid Avenue, St. Louis, MO 63110 Phone: 314-747-8423 Fax: 314-747 8427 Email: bances@wustl.edu

Original Article

Word Count for Abstract: 350 / 350 words

Character Count for Title: 72

Total Word Count: 5323 words

Number of References: 77

Number of Tables: 1

Number of Figures: 4

SOURCES OF FUNDING

This work was funded by the National Institute of Health (NIH) grants R01NR012907 (BA), R01NR012657 (BA), R01NR014449 (BA), RF1AG053303 (CC), RF1AG058501 (CC), U01AG058922 (CC), K01 AG053474 (BG), P30 AG066444 (JCM), P01AG003991 (JCM), P01AG026276 (JCM), U19 AG032438 (JCM), and U19 AG024904 (JCM). This work was also supported by the generous support of the Barnes-Jewish Hospital; the Washington University Institute of Clinical and Translational Sciences Foundation (UL1 TR000448); the Hope Center for Neurological Disorders; the Paula and Rodger O. Riney Fund; the Daniel J Brennan MD Fund; and Fred Simmons Olga Mohan Fund and the Chuck Zuckerberg Initiative (CZI) .

This work was supported by access to equipment made possible by the Hope Center for Neurological Disorders, the Neurogenomics and Informatics Center (NGI: https://neurogenomics.wustl.edu/) and the Departments of Neurology and Psychiatry at Washington University School of Medicine.

DISCLOSURES:

Julie K. Wisch reports no disclosures

Omar H. Butt reports no disclosures

Brian Gordon reports no disclosures

Carlos Cruchaga reports: Biogen, EISAI, Alector and GSK. The funders of the study had no role in the collection, analysis, or interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. CC is a member of the advisory board of Vivid genetics, Circular Genomics, Halia Therapeutics and ADx Healthcare

Suzanne E. Schindler reports no disclosures

Anne Fagan is a member of the scientific advisory boards for Roche Diagnostics, Genentech and Diadem and also consults for DiamiR and Siemens Healthcare Diagnostics Inc. There are no conflicts.

Chengran Yang reports no disclosures

Anna H. Boerwinkle reports no disclosures

Tammie L. Benzinger has consulted on clinical trials with Biogen, Roche, Jaansen, and Eli Lilly. She receives research support from Eli Lilly and Avid Radiopharmaceuticals. Avid Radiopharmaceuticals provided the AV-1451 used in this study.

David M. Holtzman reports being an inventor on a patent licensed by Washington University to C2N Diagnostics on the therapeutic use of anti-tau antibodies. D.M.H. co-founded and is on the scientific advisory board of C2N Diagnostics. C2N Diagnostics has licensed certain anti-tau antibodies to AbbVie for therapeutic development. D.M.H. is on the scientific advisory board of Denali, Genentech, Cajal Neuroscience, and consults for Eli Lilly and Alector.

John C. Morris reports no disclosures.

Beau M. Ances reports no disclosures.

ABSTRACT

Importance: Heterogeneity in progression to AD poses challenges for both clinical prognosis and clinical trial implementation. In the absence of a well-defined understanding of future disease trajectory, participants may receive unnecessary treatment or true effects of pharmacological intervention may be obscured.

Objectives: We identified early differences in preclinical Alzheimer Disease (AD) biomarkers, assessed patterns for developing preclinical AD across the Amyloid-Tau-(Neurodegeneration) (AT(N)) framework, and considered potential sources of difference by analyzing the CSF proteome.

Design, Setting, and Participants: 108 participants enrolled in longitudinal studies at the Knight Alzheimer Disease Research Center (ADRC) who completed four or more lumbar punctures and were cognitively normal at baseline were included. Cerebrospinal fluid (CSF) measures of Aβ42, $pTau_{181}$, and Neurofilament Light chain (NfL) as well as proteomics values were evaluated. Imaging biomarkers, including positron emission tomography (PET) amyloid and tau and structural magnetic resonance imaging (MRI) were repeatedly obtained when available. This allowed for staging individuals according to the AT(N) framework.

Results: Growth mixture modeling, an unsupervised clustering technique, identified three patterns of biomarker progression as measured by CSF $pTau_{181}$ and CSF Aβ42. Two groups (AD Biomarker Positive and AD Biomarker Intermediate) had distinct progression from normal biomarker status to having biomarkers consistent with preclinical AD. A third group (AD Biomarker Negative) did not develop abnormal AD biomarkers over time. Participants grouped by CSF trajectories were successfully re-classified using only proteomic profiles ($AUC_{\text{AD Biomarker Positive vs AD Biomarker Negatives}} = 0.970$, $AUC_{\text{AD Biomarker Positive vs. Intermediate AD Biomarkers}} = 0.750$, $AUC_{\text{Intermediate AD Biomarkers vs. AD Biomarker Negative}} = 0.698$).

Conclusions and Relevance: We highlight heterogeneity in the development of AD biomarkers in cognitively normal individuals. We identified individuals who became AD Biomarker Positive before age 50. A second group, AD Biomarker Intermediate, developed elevated CSF $ptau_{181}$ in their mid-60's before becoming amyloid positive in their mid-70's. A third group were AD Biomarker Negative over repeated testing. Our results could influence the selection of participants for specific treatments (e.g. amyloid-reducing vs. other agents) in clinical trials. CSF proteome analysis highlighted additional potential opportunities for non-AT(N) focused therapies, including blood brain barrier-, liver-, and neuroinflammatory-related targets.

INTRODUCTION

Alzheimer Disease (AD), a slowly progressive neurodegenerative disorder with an extended prodromal stage, affects nearly 6 million Americans(1). It is a disease that progresses from a clinically asymptomatic preclinical phase to a symptomatic clinical phase over many years(2). The disease continuum is generally thought to progress starting with amyloid accumulation followed by the development of tau pathology concurrent with neurodegenerative changes and finally clinically observable cognitive impairment (AT(N))(3). However, there is significant heterogeneity from the time of development of amyloid to the time of clinical symptoms (4).

Cerebrospinal fluid (CSF) biomarkers have been developed that measure amyloid (CSF Aβ42, CSF Aβ40), phosphorylated tau (e.g. CSF $pTau_{181}$), and neurodegeneration (CSF neurofilament light chain (NfL)). CSF amyloid markers, either low CSF Aβ42 or a low CSF Aβ42/Aβ40 ratio, show high correspondence with amyloid PET imaging (5,6) and postmortem pathology (7). CSF $pTau_{181}$ increases as amyloid accumulates and is not a reflection of tau aggregation but instead neurofibrillary tangles (8). Studies have also attempted to forecast symptomatic development of AD using CSF measures (9,10). Of these, the CSF $pTau_{181}$/ Aβ42 ratio was found to be the most prognostic; however, this ratio was not able to accurately predict time to symptom onset (10). We posit that the limitations of indirect measures of amyloid and tau pathology to forecast AD symptom onset may indicate that other factors, perhaps those expressed in CSF proteomics, can offer an enhanced understanding of the progression from preclinical to symptomatic AD.

In the literature, the observed heterogeneity in preclinical AD is frequently attributed to cognitive resilience (11). Additional factors that have been proposed as possible drivers of resilience, include educational attainment(12,13), cortical thickness(12), personality(14), cardiovascular health(15), and synaptic function(16–18). A substantial proportion of the resilience literature relies on residual methods which allow researchers to define resilience as the source of unexplained variance, rather than identifying clinically meaningful underlying sources (11).

Although difficult, it is vitally important that we understand the etiology of heterogeneity in the preclinical and symptomatic stages of AD. Clinical trials are now targeting drug delivery during either the preclinical or very early symptomatic stages of AD (2). In the absence of a well-defined understanding of future disease trajectory, participants may receive unnecessary treatment or the true effects of pharmacological intervention may be obscured.

In an effort to further interrogate the mismatch in neuropathologic change and progression to clinical symptoms of AD, we evaluated longitudinal CSF samples from a well-characterized cohort of older adults who were cognitively normal at enrollment. More than 100 participants completed at least four lumbar punctures over a period of about 10 years, in addition to completing other traditional measures of amyloid, tau, and neurodegeneration (AT(N)). Recent reviews have highlighted the importance of longitudinal research, particularly with respect to heterogeneity (19); however, relatively few longitudinal studies have been conducted. One recent review identified from more than 1400 studies in AD that had been performed between 1995 and 2015, only 48 included repeat biomarker measurements. From those 48 longitudinal studies, only nine included CSF biomarkers, with almost all of these CSF studies relying on just two

timepoints (20). Longitudinal studies are preferable for studying biomarker dynamics (21); measuring samples from the same individuals over time allows for a deeper understanding of potential time courses for disease development.

We compared the trajectories from the longitudinal CSF samples to well-established neuroimaging markers of PET amyloid (using Pittsburgh compound B (PiB)), PET tau (using AV1451), and neurodegeneration (using magnetic resonance imaging (MRI) measures of cortical thickness and white matter hyperintensities), as well as CSF neurofilament light chain (NfL). Due to the aforementioned limitations of traditional AT(N) pathology to fully explain the heterogeneity in symptomatic progression, we included an analysis of proteomics.

Evaluating protein expression in preclinical AD and healthy aging provides insight into additional potential biological mechanisms that underlie observed heterogeneity (22). Beyond simply characterizing our participants using AD biomarkers, proteomics analysis allows us to consider potential pathways and mechanisms of disease progression. Patterns in protein expression in individuals who progress relatively rapidly to symptomatic AD may point to biological hazards, while proteomic expression in individuals who progress rather slowly may help identify protective factors.

The objectives of this project were as follows: to identify early differences in preclinical AD biomarker development, to assess patterns of development of preclinical AD biomarkers across the AT(N) framework, and to consider potential sources of difference by examining the CSF proteome.

METHODS

We included 108 participants (Table 1) enrolled in longitudinal studies at the Knight ADRC, Washington University in St Louis (WUSTL) as previously described (23). For study inclusion, participants had to be: 1) cognitively normal at time of enrollment; 2) have longitudinal clinical, imaging and CSF measures; and 3) at least one sequenced set of high throughput proteomics. Four CSF data points were required in order to ensure that observed nonlinear dynamics at the individual level were reflective of actual observed nonlinearity rather than noise. Enrollment in the study occurred over a mean period of 11.3 (SD = 2.4) years. A subset of participants also completed a PET PIB scan and/or PET AV1451 scan and/or structural MRI (*Supplemental Figure 1*). This study was approved by the WUSTL Institutional Review Board, and each participant provided signed informed consent.

Data Acquisition

CDR

Participants in the study completed regular clinical assessments and cognitively normal at time of enrollment as defined by the Clinical Dementia Rating® (CDR) scale. The CDR classifies the degree of cognitive impairment through the use of semi-structured interviews (24). Individuals with a CDR of 0 are considered to have no impairment; CDR 0.5 as very mild dementia; CDR 1 as mild dementia; CDR 2 as moderate dementia; and CDR 3 as severe dementia (24).

APOE Genotyping

DNA samples were collected at enrollment and genotyped using either an Illumina 610 or OmniExpress chip. Genotyping methods have been previously described (25).

Cerebrospinal fluid (CSF) acquisition

Each participant enrolled in this study completed at least four lumbar punctures (LP). On average, these occurred approximately 2 years apart. This process has been previously described (26). LP was performed at 8:00 AM following an overnight fast. An atraumatic Sprotte 22-gauge spinal needle was used to collect approximately 25 mL of CSF via gravity drip

Cerebrospinal fluid (CSF) collection and processing

Participants underwent CSF collection by LP at approximately 8 AM following overnight fasting (REF). CSF (20-30 mL) was collected in a 50-mL polypropylene tube via gravity drip using an atraumatic Sprotte 22-G spinal needle. The tube was inverted gently to disrupt potential gradient effects and centrifuged at low speed to pellet any cellular debris. The CSF was then aliquoted into polypropylene tubes and stored at −80°C. Concentrations of CSF Aβ40, CSF Aβ42, CSF tau phosphorylated at 181 (CSF $ptau_{181}$) were measured by chemiluminescent enzyme immunoassay using a fully automated platform (LUMIPULSE G1200, Fujirebio, Malvern, PA, USA). CSF NfL was measured via commercial ELISA kit (UMAN Diagnostics, Umeå, Sweden).

Structural MRI

MRI images were obtained on 3T Siemens scanners. T1-weighted scans were segmented using FreeSurfer 5. 3 (Martinos Center for Biomedical Imaging, Charlestown, Massachusetts, USA), using the Desikan-Killiany atlas. Previous work has identified that cortical thickness decreases with the onset of AD (27–29). We calculated the average cortical thickness (29).

White Matter Hypterintensities

T-2 weighted fluid attenuated inversion recovery (FLAIR) images were also collected. White matter hyperintensities (WMH) were calculated via a legion segmentation toolbox that relies on Statistical Parametric Mapping (SPM) (30).

Positron emission tomography (PET) imaging

PET scans using [11C] PiB were obtained via previously described methods (31). Images were then processed using the PET unified pipeline (PUP, https://github.com/ysu001/PUP) (32,33). Images were smoothed to achieve a spatial resolution of 8 mm. This minimized inter-scanner differences (33,34). A standard image registration technique was used to correct for motion (35,36) using corresponding structural images. The cerebellum was used as the reference region. Regions of interest were defined using the Desikan-Killiany atlas based on the MRI. The standard uptake ratio (SUVR) in each region was evaluated using the 30 – 60 minute post-injection time window (37). We applied partial volume correction via a geometric transfer matrix approach (38). The PET PiB summary value was the arithmetic mean of SUVRs for the following regions: precuneus, prefrontal cortex (FreeSurfer regions: superior frontal and rostral

middle frontal regions), gyrus rectus (FreeSurfer regions: lateral orbitofrontal and medial orbitofrontal regions), and lateral temporal regions (FreeSurfer regions: superior temporal and middle temporal regions) (39).

PET tau imaging utilized [18F]-Flortaucipir (AV-1451), but was otherwise conducted in a similar manner to PET PiB imaging. The SUVR was evaluated using the 80 – 100 minute post-injection time window. The whole cerebellum was used as the reference region (39). The PET Tau summary value, hereafter referred to as "Tauopathy", was the arithmetic mean of SUVRs for the following regions: amygdala, entorhinal cortex, inferior temporal region, and lateral occipital cortex (31).

CSF Proteome Analysis

All participants had at least one CSF sample also processed for proteomics profiling. When multiple CSF samples were available, the most recent sample was retained for analysis. Briefly, proteomic data was generated using the SomaScan 1.3k panel (SomaLogic Inc), an aptamer-based platform as previously described (40,41). Quality control (QC) was performed at the sample and aptamer levels using control aptamers (positive and negative controls) and calibrator samples. At the sample level, hybridization controls on each plate were used to correct for systematic variability in hybridization. The median signal over all aptamers was used to correct for within-run technical variability. This median signal was assigned to different dilution sets within each tissue. For CSF samples, a 20% dilution rate was used.

Statistical Analysis

An unsupervised machine learning technique called growth mixture modeling was used to cluster the longitudinal trajectories of individuals CSF $pTau_{181}$ / CSF Aβ42 ratios (42). This approach identifies possible sub-groups within longitudinal data and has previously been employed to study cognitive trajectories in AD (4,15,43) and structural changes (44), but not in preclinical amyloid and tau biomarkers. We searched for 1, 2, 3, and 4 latent clusters and selected the optimal number of clusters via Bayesian Information Criterion (BIC) minimization. We compared participant demographics across the identified latent clusters, using the R package table1(45).

To evaluate the time to pathology development, we performed survival analysis (46) to determine age at amyloid positivity (using CSF Aβ42/Aβ40 < 0.0673 pg/mL as the threshold (47)) and age at tau positivity (using the CSF $pTau_{181}$ > 42.5 pg/mL as the threshold (47), and age at symptomatic onset (using the first instance of CDR > 0). Participants were considered pathology positive or to have experienced symptom onset at their visit date where either of their CSF measures surpassed threshold or they had their first CDR > 0 rating at a clinical visit. On average, participant visits where lumbar punctures were performed were spaced about 2 years apart. Clinical visits for CDR rating occur annually. In the survival analysis, we grouped by the latent clusters that we had identified in the previous analysis (relying on the CSF pTau/Aβ42 ratio). For the survival analysis we omitted participants that were amyloid positive (13 / 108) and tau positive (20 / 108) prior to study enrollment for their respective analyses as we could not estimate their age at pathology positivity or change in CDR status.

We then evaluated the trajectory of a variety of amyloid (CSF Aβ42/Aβ40 and PET-PiB summary value), tau (CSF $pTau_{181}$ and PET-AV1451 Tauopathy) and neurodegeneration (cortical thickness, WMH volume, CSF NfL) biomarkers, testing for differences across latent clusters. We applied generalized additive mixed models (GAMMs), selected for their interpretability, capacity for nonlinearity, and ability to handle repeated measures at inconsistent time intervals (48). Age at procedure, identified latent cluster (based on the longitudinal CSF pTau/Aβ42 ratios), and their interaction were included as regressors. The following parameters were utilized as response variables: CSF Aβ42, PET-PiB cortical amyloid summary, CSF pTau, PET-AV1451 tauopathy, cortical thickness, WMH volume, and CSF NfL.

Finally, we applied Pelora, a supervised clustering technique, to classify individuals as members of the previously identified latent clusters using proteomics values(49). A total of 713 CSF proteins passed QC and were utilized as features in the supervised clustering model. For this analysis single timepoint proteome values were compared to clusters derived from longitudinal trajectories.

After applying Pelora to identify ten protein clusters that were predictive of the latent cluster labels, we utilized 10-fold cross validated lasso regression to select clusters relevant for analysis (50). Because this portion of the analysis was for hypothesis generation rather than diagnostic development, we ran Pelora and the subsequent lasso binomial regression on the entire dataset. In order to assess the stability of the classification algorithm, we also completed 10-fold cross validation with an 80% train/20% test split and evaluated the area under the curve (AUC).

To complete our analysis of the proteome, we also calculated predictive power scores for each protein. Further, we performed logistic regression using group membership in the identified latent clusters as the response variable and each protein as the regressor in order to calculate the individual AUC for each protein's ability to classify. We performed a pathway analysis for the proteins identified by Pelora, in order to better understand the function of the proteins associated with latent cluster group membership (51).

All analysis was conducted using R (52). Analysis code has been published at github.com/jwisch/ ProteomicClusters. Proteomic data is available at NIAGADS: https://www.niagads.org/datasets/ng00102 and in the Proteomics Browser: http://ngi.pub:3838/ONTIME_Proteomics/.

RESULTS

Identification of Early Differences in Preclinical AD Pathology

An unsupervised machine learning technique, growth mixture modeling, was used to cluster the longitudinal trajectories for each participant with regards to CSF $pTau_{181}$/ Aβ42 (42). This was a novel application of the algorithm to preclinical amyloid and tau biomarkers. Using a data-driven search for the appropriate number of clusters, three latent growth trajectories were identified (*Figure 1a*). The largest cluster was the "AD Biomarker Negative" group (N = 69) and contained individuals who had relatively low CSF $pTau_{181}$ throughout the time period. Despite CSF Aβ42

progressively decreasing, these individuals continued to have relatively low CSF pTau. The second cluster of individuals was referred to as the "Intermediate AD Biomarkers" group (53) (N = 27). It was comprised of individuals who had higher CSF $pTau_{181}$ and corresponding lower CSF Aβ42 but did not have CSF $pTau_{181}$ as high as the third group. Participants in the third group were referred to as "AD Biomarker Positive" group (N = 12). These individuals had high CSF $pTau_{181}$ and low CSF Aβ42 that were consistent with AD positivity (47). These individuals were at the highest risk for developing AD (54).

In order to understand the different relationships between CSF Aβ42 and CSF $pTau_{181}$ for the three groups, we used a generalized additive mixed effect model (GAMM), fitting cubic splines by group membership. A breakpoint for the Intermediate AD Biomarker cluster occurred just below 1000 pg/mL (Fig 1a). After this threshold, the applied GAMM revealed three distinct slopes: one negative slope, indicating a decreasing relationship between CSF $pTau_{181}$ and declining CSF Aβ42 (AD Biomarker Negative latent cluster), one approximately zero slope, indicating no relationship between CSF $pTau_{181}$ and CSF Aβ42 (Intermediate AD Biomarkers latent cluster), and one positive slope, indicating an increasing relationship between CSF $pTau_{181}$ and declining CSF Aβ42 (AD Biomarker Positive latent cluster).

Given recent work emphasizing the utility of Aβ40 as a means to mitigate abnormally high or low Aβ42 values that could be attributed to individual variation in protein production or ventricular volume(55), we evaluated CSF Aβ42/Aβ40 by CSF $pTau_{181}$ relationship for the three latent clusters. This normalization of CSF Aβ42 by Aβ40 (*Figure 1b*) transforms the apparent three trajectories such that all participants fall on a single monotonically increasing continuum where low CSF Aβ42/Aβ40 is associated with high CSF pTau. Within this continuum, AD Biomarker Negative individuals exist nearly entirely below the thresholds for amyloid and tau positivity, Intermediate AD Biomarkers individuals exist in the transition area, where the relationship between CSF $pTau_{181}$ and CSF Aβ42/Aβ40 goes from a basically flat relationship to a steeply increasing relationship, and AD Biomarker Positive individuals show a steeply increasing relationship between CSF Aβ42/Aβ40 and CSF $pTau_{181}$.

Although Figure 1b seems to show a continuum of pathology, there are important demographic differences across the latent clusters (Table 1). There were differences with regards to age at study enrollment (AD Biomarker Negative individuals were the youngest while Intermediate AD Biomarkers participants were the oldest). These differences in age reveal that although the normalized plot of CSF $pTau_{181}$ by CSF Aβ42/Aβ40 shows a continuum of pathology, this does not represent a continuum across time. There were also differences with regards to APOE status, where the APOE ε4 allele was most frequently found in the AD Biomarker Positive cohort and the APOE ε2 allele was most frequently found in the AD Biomarker Negative cohort. By the conclusion of the study, there was a statistically significant difference in CDR across the three latent clusters. The AD Biomarker Positive cohort had the greatest clinical decline (5 / 12 (42%) converted to CDR ≥ 0.5), while the AD Biomarker Negative group had relatively little change on clinical evaluation (1 / 69 (1%) converted to CDR = 0.5).

Patterns of Development Across the AT(N)

We then performed survival analysis (46) in order to evaluate time to pathology development. We assessed age to amyloid positivity, age to tau positivity, and age to CDR conversion as stratified by the three latent clusters. In addition to survival analysis, we applied GAMMs to biomarkers of AT(N) pathology.

Survival analysis showed a clear separation in age at amyloid positivity (CSF Aβ42/Aβ40 < 0.0673) (47)) across clusters (*Figure 2a*)

Figure *2.* Survival analysis demonstrated that the AD Biomarker Positive group quickly proceeded to both amyloid and tau positivity. The Intermediate AD Biomarkers group lagged the AD Biomarker Positive group by about 10 years for amyloid positivity (as defined by CSF Aβ42/Aβ40 < 0.06753), but no statistically significant difference for tau positivity (as defined by CSF pTau181 > 42.5). A majority of the AD Biomarker Negative group never became amyloid positive and became tau positive only after age 80.. The majority of AD Biomarker Positive participants were amyloid positive before age 65. GAMM modeling shows that the AD Biomarker Positive participants were, on average, CSF Aβ42/Aβ40 positive at time of enrollment (*Figure 3*) (43,53). AD Biomarker Positive were also CSF pTau positive (CSF $pTau_{181}$ > 42.5 ug/mL) around 65 years old (47). Similar results were also seen using PET – AV1451. Relatively few participants converted to CDR > 0; however, the majority of the AD Biomarker Positive cluster developed clinical symptoms by their late 70's (*Figure 2c*).

The majority of Intermediate AD Biomarker participants did not become amyloid positive by CSF Aβ42/Aβ40 until around 75 years old, based on survival analysis. The difference in time to amyloid positivity by cluster is statistically significant (Cox proportional hazard test, $p < 0.001$). GAMM modeling aligns with this observation. Interestingly, positivity as defined by CSF pTau (47) occurred prior to amyloid positivity as defined by CSF Aβ42/Aβ40 for the Intermediate AD Biomarkers cohort (*Figure 2b*). This ordering does not align with the AT(N) hypothesis. For both the Intermediate and AD Biomarker Positive cohorts, clinical symptoms occurred after pathology developed, consistent with the AT(N) hypothesis (56). There were no observable differences in tauopathy as measured by PET-AV1451 between the Intermediate AD Biomarkers and AD Biomarker Negative cohort; however, the data was relatively sparse. There are no observable differences in any of the measures of neurodegeneration (cortical thickness, WMH volume, and CSF NfL) across the three clusters of participants.

Replicating Identified Clusters via the Proteome

We attempted to classify individuals as members of one of the three cohorts using only proteomics data. No additional covariates (e.g. age, sex, APOE genotype) were included in the initial analysis. For each pair of cohorts, we applied Pelora (57) twice: once to the entire set and once using cross validation. We evaluated and reported the proteins identified when we applied Pelora to the entire cohort. We also applied 10-fold cross validation in order to evaluate the generalizability of the classifications; the AUC associated with both model runs subsequently reported. We repeated the analysis after including clinical values of age alone, sex alone, and both age and sex in accordance with the methods outlined (57). Our results remained robust after introduction of these covariates.

The Pelora algorithm(57), which is applied to labeled data (in this case, the labels were AD Biomarker Positive, Intermediate AD Biomarkers, and AD Biomarker Negative) identified groups of proteins that were either upregulated or down regulated. The expression of each of the highlighted proteins, separated by group, is shown in Supplemental Figures 2 – 7. Heatmaps were generated that showed correlation between identified proteins (Supplemental Figures 13 – 15). Proteins that were most important for each group were ranked by binomial log-likelihood. To get a general sense of which proteins played the most significant role in each group

(Intermediate AD Biomarkers vs. AD Biomarker Negative, AD Biomarker Positive vs. AD Biomarker Negative, AD Biomarkers Positive vs. Intermediate AD Biomarkers), the primary function of each of the ten most important proteins (as ranked by binomial log-likelihood) were classified after reviewing available literature. A pathway analysis is also reported in the supplement. Proteins were grouped as primarily associated with the blood brain barrier (BBB), cardiovascular disease (CVD), liver function, amyloid production and/or clearance, inflammation, or neurodegeneration. The relationship between each protein, its' function, and which classification(s) it applied to is shown in *Figure 4.* Detailed comparisons of protein expressions are displayed in Supplemental Figures 2 – 7. Pathway analysis results are displayed in Supplemental Figure 12.

The AD Biomarker Positive group was successfully differentiated from the AD Biomarker Negative group (full dataset classification AUC = 0.970, 10-fold cross-validated classification AUC = 0.873). Three proteins that had the greatest ability to classify individuals with AD Biomarker Positive compared to AD Biomarker Negative via both Pelora, as ranked by binomial log-likelihood, as well as via individual classification were SPARC-related modular calcium-binding protein 1 (SMOC1) ($AUC_{individual}$ = 0.900, PPS = 0.199), 14-3-3 protein family ($AUC_{individual}$ = 0.902, PPS = 0.251), and 14-3-3 protein zeta/delta ($AUC_{individual}$ = 0.909, PPS = 0.292). These proteins that are associated with changes in neurodegeneration (14-3-3 protein family and 14-3-3 protein zeta/delta)(58) and BBB function (SMOC1)(59) . All three of these proteins were previously identified in a study evaluating differences between healthy controls and symptomatic AD individuals (40). A full ranking of all individual proteins demonstrating at least an AUC of 0.750 for classification is shown in Supplemental Figure 9.

The AD Biomarker Positive group was also successfully differentiated from the Intermediate AD Biomarkers group (full dataset classification AUC = 0.750, 10-fold cross-validated classification AUC = 0.780). Nidogen-2 ($AUC_{individual}$ = 0.768, PPS = 0.193), Insulin-like growth factor-binding protein 1 (IGF-1) ($AUC_{individual}$ = 0.648, PPS = 0.199), haptoglobin ($AUC_{individual}$ = 0.657, PPS = 0.199), retinol-binding protein 4 (RBP 4) ($AUC_{individual}$ = 0.762, PPS = 0.199), polymeric immunoglobulin receptor ($AUC_{individual}$ = 0.623, PPS = 0.199), and complement c1r ($AUC_{individual}$ = 0.648, PPS = 0.199) were of primary importance. Functionally, these proteins are associated with changes in BBB function (Nidogen-2(60) and polymeric immunoglobulin receptor(61)), neurodegeneration (IGF-1(62), haptoglobin(63,64), and complement C1R(65)), and liver function (RBP 4(64)). A complete listing of all proteins demonstrating an AUC of at least 0.750 for classification is shown in Supplemental Figure 10.

Classification of the Intermediate AD Biomarkers compared to the AD Biomarker Negative group was at the borderline for acceptable performance (full dataset classification AUC = 0.698, 10-fold cross-validated classification AUC = 0.666). This was the least successful via the Pelora-based clustering approach, as demonstrated by a lower model AUC (0.698 vs. 0.750 or 0.970). The most important proteins for classification via Pelora included proteins associated with neurodegeneration (14-3-3 protein family ($AUC_{individual}$ = 0.829, PPS = 0.251) (58) and neurogenic-locus notch homolog protein3 ($AUC_{individual}$ = 0.650, PPS = 0.199)(66)), amyloid production and regulation (ubiquitin-fold modifier 1(67) ($AUC_{individual}$ = 0.817, PPS = 0.263)),

and BBB function (SPARC-related modular calcium-binding protein 1(59) ($AUC_{individual}$ = 0.764, PPS = 0.199)). Notably, the individual AUC of simply using the 14-3-3 protein family for classification was much higher than the clustering-based approach. Other proteins that demonstrated high efficacy for classifying Intermediate AD Biomarker vs. AD Biomarker Negative are shown in Supplemental Figure 11.

DISCUSSION

This study used a novel application of growth mixture modeling to identify unique pathological trajectories of participants as a function of CSF $pTau_{181}$ / Aβ42. Previously this unsupervised clustering technique has been applied to markers of cognition and neurodegeneration(4,15,43,44). Our objective in doing this was to apply a data driven method to understanding heterogeneity in longitudinal development of AD pathology.

After performing this classification, we examined the demographic characteristics of the three clusters that were identified. The AD Biomarker Positive group had the greatest proportion of APOE ε4+ individuals. This is consistent with studies that have previously identified that the APOE ε4 allele is a risk factor for developing AD (68,69) and at an earlier age for amyloid deposition (70). The AD Biomarker Negative individuals were younger than the other clusters at the time of enrollment. Interestingly, AD Intermediate individuals were the oldest. A priori, we would have anticipated that the oldest group would have been the AD Biomarker Positive cohort. The older age of the Intermediate cohort suggests that this group is developing AD pathology at a slower rate than the AD Biomarker Positive group and may exhibit some resilience in the face of increasing pathology.

When we looked at $pTau_{181}$ as a function of CSF Aβ42/Aβ40, individuals appeared to move along a single continuum rather than three distinct paths. Although it appears in the analysis of the AT(N) that Intermediate individuals attain tau positivity before reaching amyloid positivity, the presented continuum shows steadily decreasing CSF Aβ42/Aβ40 with rapidly increasing CSF $pTau_{181}$ after individuals reach an inflection point. This tipping point is approximately where the Intermediate AD individuals fall along the continuum. Figure 1b shows a relationship between amyloid and tau that is consistent with the prevailing literature (e.g. 2,54,56) however it is important to recall that all participants are not the same age. The Intermediate cohort is oldest, meaning that although we observe a continuum of pathology, it is not aligned temporally.

We further interrogated the proposed AT(N) continuum through the use of survival analysis and application of GAMMs with a variety of biomarkers. Overall, longitudinal changes across the AT(N) aligned with existing literature(2,56,71). Compared to the other groups, the AD Biomarker Positive group had significantly lower CSF Aβ42/Aβ40 at the time of enrollment; that persisted throughout subsequent time points. With the limited PET-PiB data available, we observed an elevation in amyloid starting at enrollment for the AD Biomarker Positive cohort. This has important implications for clinical trials that emphasize early intervention for amyloid detection and potential removal. For the AD Biomarker Positive group, amyloid-related changes occurred in cognitively normal adults before age 50, consistent with previous work (72). Individuals in the Intermediate AD Biomarkers cohort showed a steady decline in CSF

Aβ42/Aβ40 over the study, with many eventually becoming amyloid positive. The Intermediate cohort displayed a very clear increase in amyloid pathology as measured by PET – PiB around age 70.

With regards to CSF pTau, this measure was also elevated at the time of enrollment for the AD Biomarker Positive cohort. These participants were CSF $pTau_{181}$ positive by their mid-50's. Of note, tau positivity as measured by PET tau lagged behind CSF $pTau_{181}$ positivity for this group and is consistent with previously published literature (5). While the AD Biomarker Positive group followed the proposed AT(N) curves (56), the Intermediate cohort developed CSF tau positivity earlier (65 years old) compared to CSF amyloid positivity (75 years old). This development of tau before amyloid is consistent with previous studies suggesting that in some individuals, tau positivity can occur prior to amyloid positivity(73). In the survival analysis, we did not detect a statistically significant difference in time to positivity for amyloid compared to tau for either the AD Biomarker Positive or Intermediate AD Biomarkers cohorts again suggesting that there may be heterogeneity in the progression to symptomatic AD. However, our survival analyses are limited by the number of individuals who were AD biomarker positive upon enrollment.

Throughout enrollment, there were no significant differences in neurodegeneration biomarkers between the three groups. This lack of difference was expected as we focused on cognitively normal individuals who may be at the very earliest stages of AD. Neurodegeneration is proposed to occur during the later stages and our results support the AT(N) hypothesis. At the conclusion of this study, we had one participant with a CDR = 2 and one participant with CDR = 1. Even though amyloid and tau pathology developed in this cohort, participants rarely progressed to symptomatic AD during the duration of the study (~11 years). Of those who did, cognitive decline aligned with our assessment of disease pathology severity (42% of the AD Biomarker Positive cohort had decline on clinical assessments compared to 7% of the Intermediate AD Biomarker cohort).

Perhaps most surprising was our ability to classify individuals as AD Biomarker Positive, Intermediate AD Biomarker, or AD Biomarker Negative – groupings that emerged organically from an unsupervised clustering analysis – using an entirely separate method, namely CSF proteome. Several post mortem studies have previously applied proteomic analysis to identify potential sources of resilience (16–18). In contrast to previous *ex vivo* brain-tissue based proteomics studies of resilience, we utilized *in vivo* CSF samples. Observed differences in the proteome represent a starting point rather than a conclusive identification of discrepancies in early preclinical AD progression. Future work will require quantitative targeted measurements of specific proteins.

Within the AD Biomarker Negative group many proteins associated with neurodegeneration and amyloid production were downregulated. Further studies will be required to determine if the decreased expression of these proteins plays a protective role in slowing the progression of AD pathology. Ubiquitin Modifier 1(74), a protein associated with amyloid production, was downregulated in the AD Biomarker Negative group as compared to the other two cohorts, highlighting the importance of amyloid in the progression to AD. Several BBB associated

proteins (including SMOC, Nidogen-2, and Matrilysin) were also downregulated. Many neurodegeneration associated proteins (e.g. Calcineurin, 14-3-3 protein family, MCL-1, IGF-1, NOTCH03, and Kallikrein-8) were also less commonly expressed in this cohort, consistent with recent findings (40).

Within the Intermediate AD Biomarker group, several proteins were upregulated. This suggests that this cluster may offer insights into slowing of the progression of AD. Angiogenin(75) and Nidogen-2(76), proteins associated with BBB integrity, were elevated in this group. Nidogen-2 on its own was sufficient to distinguish the Intermediate from the AD Biomarker Positive cohort, indicating its overall importance. Retinol Binding Protein 4(77) a protective protein associated with liver health, was identified by Pelora as important and was also upregulated in this group. From this we find evidence consistent with previously published arguments for the importance of BBB integrity (78) and liver health (79) in the prevention of AD.

Limitations

Although this dataset is relatively large in the context of longitudinal CSF studies, we had relatively few datapoints in the context of machine learning. Because of this data sparsity, we were unable to retain a true validation dataset and instead had to rely on cross-validation to evaluate generalizability. We were also were unable to compare results to an external cohort for validation. The requirement of multiple LPs with CSF Aβ42 and CSF $pTau_{181}$ for unsupervised classification in addition to needing fully multiplexed proteomics via CSF makes this a unique dataset. We hope in the future that additional highly characterized longitudinal CSF samples become available for analysis. Future collection of longitudinal PET Tau images could also greatly enhance this dataset and allow for a more complete investigation of tau progression in this cohort.

Conclusions

Our findings on both the timing of amyloid and aggregated tau development, and the ability of the CSF proteome to classify these groupings have important implications for clinical trials. In therapies that focus on the AT(N), here we highlight heterogeneity in amyloid and tau development. The AD Biomarker Positive group developed amyloid and tau pathology before age 50, suggesting very early intervention is necessary for this group. The Intermediate AD Biomarker group developed significant tau pathology before becoming amyloid positive. Perhaps amyloid-reducing agents would demonstrate less efficacy in this group, as they do not seem to require amyloid positivity before developing substantial tau burden. We also identified additional potential non-AT(N) related targets for prospective AD drug development, including BBB integrity, liver function, and neuroinflammation.

REFERENCES

1. Gaugler J, James B, Johnson T, Marin A, Weuve J. 2020 Alzheimer's disease facts and figures. Alzheimer's and Dementia. 2020 Mar 1;16(3):391–460.

2. Aisen PS, Cummings J, Jack CR, Morris JC, Sperling R, Frölich L, et al. On the path to 2025: Understanding the Alzheimer's disease continuum. Vol. 9, Alzheimer's Research and Therapy. BioMed Central Ltd.; 2017.

3. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimer's & Dementia [Internet]. 2018 Apr 1 [cited 2018 Nov 14];14(4):535–62. Available from: https://www.sciencedirect.com/science/article/pii/S1552526018300724

4. Qiu P, Zeng M, Kuang W, Meng SS, Cai Y, Wang H, et al. Heterogeneity in the dynamic change of cognitive function among older Chinese people: A growth mixture model. International Journal of Geriatric Psychiatry. 2020 Oct 1;35(10):1123–33.

5. Boerwinkle AH, Wisch JK, Chen CD, Gordon BA, Butt OH, Schindler SE, et al. Temporal Correlation of CSF and Neuroimaging in the Amyloid-Tau-Neurodegeneration Model of Alzheimer Disease. Neurology. 2021 Jul 6;

6. Schindler SE, Gray JD, Gordon BA, Xiong C, Batrla-Utermann R, Quan M, et al. Cerebrospinal fluid biomarkers measured by Elecsys assays compared to amyloid imaging. Alzheimer's and Dementia. 2018 Nov 1;14(11):1460–9.

7. Grothe MJ, Moscoso A, Ashton NJ, Karikari TK, Lantero-Rodriguez J, Snellman A, et al. Associations of Fully Automated CSF and Novel Plasma Biomarkers With Alzheimer Disease Neuropathology at Autopsy. Neurology [Internet]. 2021 Jul 15;10.1212/WNL.0000000000012513. Available from: http://www.neurology.org/lookup/doi/10.1212/WNL.0000000000012513

8. Toombs J, Zetterberg H. Untangling the tau microtubule-binding region. Vol. 144, Brain. Oxford University Press; 2021. p. 359–62.

9. Moghekar A, Li S, Lu Y, Li M, Wang M-C, Albert M, et al. CSF biomarker changes precede symptom onset of mild cognitive impairment [Internet]. 2013. Available from: www.neurology.org

10. Hansson O, Zetterberg H, Bucchave P, Londos E, Blennow K, Minthon L. Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitiveimpairment: a follow-up study. Lancet Neurology. 2006;5:228–34.

11. Bocancea DI, Catharina van Loenhoud A, Groot C, Barkhof F, van der Flier WM, Ossenkoppele R. Measuring Resilience and Resistance in Aging and Alzheimer Disease Using Residual Methods: A Systematic Review and Meta-analysis. Neurology. 2021 Jul 15;10.1212/WNL.0000000000012499.

12. Ossenkoppele R, Lyoo CH, Jester-Broms J, Sudre CH, Cho H, Ryu YH, et al. Assessment of Demographic, Genetic, and Imaging Variables Associated with Brain Resilience and Cognitive Resilience to Pathological Tau in Patients with Alzheimer Disease. JAMA Neurology. 2020 May 1;77(5):632–42.

13. Stern Y. Cognitive reserve in ageing and Alzheimer's disease. The Lancet Neurology. 2012 Nov;11(11).

14. Terracciano A, Iacono D, O'Brien RJ, Troncoso JC, An Y, Sutin AR, et al. Personality and resilience to Alzheimer's disease neuropathology: A prospective autopsy study. Neurobiology of Aging. 2013 Apr;34(4):1045–50.

15. Haaksma ML, Calderón-Larrañaga A, Olde Rikkert MGM, Melis RJF, Leoutsakos JMS. Cognitive and functional progression in Alzheimer disease: A prediction model of latent classes. International Journal of Geriatric Psychiatry. 2018 Aug 1;33(8):1057–64.

16. Mendonça CF, Kuras M, Nogueira FCS, Plá I, Hortobágyi T, Csiba L, et al. Proteomic signatures of brain regions affected by tau pathology in early and late stages of Alzheimer's disease. Neurobiology of Disease. 2019 Oct 1;130.

17. Arnold SE, Louneva N, Cao K, Wang LS, Han LY, Wolk DA, et al. Cellular, synaptic, and biochemical features of resilient cognition in Alzheimer's disease. Neurobiology of Aging. 2013 Jan;34(1):157–68.

18. Yu L, Petyuk VA, Gaiteri C, Mostafavi S, Young-Pearse T, Shah RC, et al. Targeted brain proteomics uncover multiple pathways to Alzheimer's dementia. Annals of Neurology. 2018 Jul 1;84(1):78–88.

19. McQuail JA, Dunn AR, Stern Y, Barnes CA, Kempermann G, Rapp PR, et al. Cognitive Reserve in Model Systems for Mechanistic Discovery: The Importance of Longitudinal Studies. Vol. 12, Frontiers in Aging Neuroscience. Frontiers Media S.A.; 2021.

20. Lawrence E, Vegvari C, Ower A, Hadjichrysanthou C, de Wolf F, Anderson RM. A systematic review of longitudinal studies which measure Alzheimer's disease biomarkers. Vol. 59, Journal of Alzheimer's Disease. IOS Press; 2017. p. 1359–79.

21. Tarawneh R. Biomarkers: Our Path Towards a Cure for Alzheimer Disease. Vol. 15, Biomarker Insights. SAGE Publications Ltd; 2020.

22. Graves PR, Haystead TAJ. Molecular Biologist's Guide to Proteomics. Microbiology and Molecular Biology Reviews. 2002 Mar;66(1):39–63.

23. Morris JC, Schindler SE, McCue LM, Moulder KL, Benzinger TLS, Cruchaga C, et al. Assessment of Racial Disparities in Biomarkers for Alzheimer Disease. JAMA Neurology. 2019 Mar 1;76(3):264–73.

24. Morris JC. Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type. International Psychogeriatric Association

[Internet]. 1997 [cited 2018 Dec 4];9(1):173–6. Available from: https://doi.org/10.1017/S1041610297004870

25. Cruchaga C, Kauwe J, Harari O, Jin S, Neuron YC-, 2013 U. GWAS of Cerebrospinal Fluid Tau Levels Identifies Risk Variants for Alzheimer's Disease [Internet]. Neuron. 2013 [cited 2019 Oct 31]. p. 256–68. Available from: https://www.sciencedirect.com/science/article/pii/S0896627313001840

26. Fagan AM, Mintun MA, Mach RH, Lee SY, Dence CS, Shah AR, et al. Inverse relation between in vivo amyloid imaging load and cerebrospinal fluid Abeta;42 in humans. Annals of Neurology. 2006 Mar;59(3):512–9.

27. Wang L, Benzinger TL, Hassenstab J, Blazey T, Owen C, Liu J, et al. Spatially distinct atrophy is linked to β-amyloid and tau in preclinical Alzheimer disease. Neurology [Internet]. 2015 [cited 2019 Oct 31];84(12):1254–60. Available from: http://surfer.nmr.mgh.harvard.edu

28. Dickerson BC, Bakkour A, Salat DH, Feczko E, Pacheco J, Greve DN, et al. The cortical signature of Alzheimer's disease: Regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. Cerebral Cortex. 2009 Mar;19(3):497–510.

29. Dincer A, Gordon BA, Hari-Raj A, Keefe SJ, Flores S, McKay NS, et al. Comparing cortical signatures of atrophy between late-onset and autosomal dominant Alzheimer disease. NeuroImage: Clinical. 2020 Jan 1;28.

30. Ithapu V, Singh V, Lindner C, Austin BP, Hinrichs C, Carlsson CM, et al. Extracting and summarizing white matter hyperintensities using supervised segmentation methods in Alzheimer's disease risk and aging studies. Human Brain Mapping. 2014;35(8):4219–35.

31. Mishra S, Gordon BA, Su Y, Christensen J, Friedrichsen K, Jackson K, et al. AV-1451 PET imaging of tau pathology in preclinical Alzheimer disease: Defining a summary measure. NeuroImage. 2017 Nov 1;161:171–8.

32. Su Y, D'Angelo GM, Vlassenko AG, Zhou G, Snyder AZ, Marcus DS, et al. Quantitative analysis of PiB-PET with FreeSurfer ROIs. PLoS ONE. 2013 Nov 6;8(11).

33. Su Y, Blazey TM, Snyder AZ, Raichle ME, Marcus DS, Ances BM, et al. Partial volume correction in quantitative amyloid imaging. NeuroImage. 2015 Feb 5;107:55–64.

34. Joshi A, Koeppe RA, Fessler JA. Reducing between scanner differences in multi-center PET studies. NeuroImage. 2009 May 15;46(1):154–9.

35. Hajnal J v., Saeed N, Soar EJ, Oatridge A, Young IR, Bydder GM. A registration and interpolation procedure for subvoxel matching of serially acquired mr images. Journal of Computer Assisted Tomography. 1995;19(2):289–96.

36. Eisenstein SA, Koller JM, Piccirillo M, Kim A, Antenor-Dorsey JA v., Videen TO, et al. Characterization of extrastriatal D2 in vivo specific binding of [18 F](N-methyl)benperidol using PET. Synapse. 2012 Sep;66(9):770–80.

37. Su Y, Blazey TM, Owen CJ, Christensen JJ, Friedrichsen K, Joseph-Mathurin N, et al. Quantitative Amyloid imaging in autosomal Dominant Alzheimer's disease: Results from the DIAN study group. PLoS ONE. 2016 Mar 1;11(3).

38. Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: Principle and validation. Journal of Nuclear Medicine. 1998;39(5):904–11.

39. Gordon BA, Friedrichsen K, Brier M, Blazey T, Su Y, Christensen J, et al. The relationship between cerebrospinal fluid markers of Alzheimer pathology and positron emission tomography tau imaging. Brain [Internet]. 2016 [cited 2019 Oct 31];139(8):2249–60. Available from: https://academic.oup.com/brain/article-abstract/139/8/2249/1753889

40. Cruchaga C, Ju Sung Y, Yang C, Wang F, Suhy A, Norton J, et al. Multi-tissue proteomics identies molecular signatures for sporadic and genetically dened Alzheimer disease cases.

41. Yang C, Farias FHG, Ibanez L, Suhy A, Sadler B, Fernandez MV, et al. Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. Nature Neuroscience. 2021 Sep 1;24(9):1302–12.

42. Ram N, Grimm KJ. Methods and Measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. Vol. 33, International Journal of Behavioral Development. SAGE Publications Ltd; 2009. p. 565–76.

43. Small BJ, Bäckman L. Longitudinal trajectories of cognitive change in preclinical Alzheimer's disease: A growth mixture modeling analysis. Cortex. 2007;43(7):826–34.

44. Lin W, Donohue MC, Insel P, Schwartzman A, Thompson WK. Bayesian Multivariate Growth Mixture Modeling of Longitudinal Data: An Application to Alzheimer's Disease Study. Available from: https://doi.org/10.1101/2021.03.10.434854

45. Yoshida K, Bohn J. Package "tableone" [Internet]. R. 2019 [cited 2019 Oct 28]. Available from: ftp://cygwin.uib.no/pub/cran/web/packages/tableone/tableone.pdf

46. Therneau T. The survival package. 2019.

47. Volluz KE, Schindler SE, Henson RL, Xiong C, Gordon BA, Benzinger TL., et al. Correspondence of CSF biomarkers measured by Lumipulse assays with amyloid PET. In: 2021 Alzheimer's Association International Conference. 2021.

48. Sørensen Ø, Walhovd KB, Fjell AM. A recipe for accurate estimation of lifespan brain trajectories, distinguishing longitudinal and cohort effects. NeuroImage. 2021;226(July 2020).

49. Dettling M, Bühlmann P. Supervised clustering of genes [Internet]. 2002. Available from: http://genomebiology.com/2002/3/12/research/0069.1

50. Hastie T, Qian J, Tay K. An Introduction to glmnet [Internet]. 2021. Available from: https://cran.us.r-project.org

51. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nature Communications. 2017 Dec 1;8(1).

52. R Core Development Team. A language and environment for statistical computing. 2013 [cited 2019 Oct 28];1. Available from: ftp://ftp.uvigo.es/CRAN/web/packages/dplR/vignettes/intro-dplR.pdf

53. Insel PS, Ossenkoppele R, Gessert D, Jagust W, Landau S, Hansson O, et al. Time to Amyloid Positivity and Preclinical Changes in Brain Metabolism, Atrophy, and Cognition: Evidence for Emerging Amyloid Pathology in Alzheimer's Disease. Frontiers in Neuroscience [Internet]. 2017 May 17 [cited 2019 Apr 3];11:281. Available from: http://journal.frontiersin.org/article/10.3389/fnins.2017.00281/full

54. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Vol. 14, Alzheimer's and Dementia. Elsevier Inc.; 2018. p. 535–62.

55. Graff-Radford J, Jones DT, Wiste HJ, Cogswell PM, Weigand SD, Lowe V, et al. Cerebrospinal Fluid Dynamics and Discordant Amyloid Biomarkers. Neurobiology of Aging [Internet]. 2021 Nov; Available from: https://linkinghub.elsevier.com/retrieve/pii/S0197458021003298

56. Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Vol. 9, The Lancet Neurology. Lancet Publishing Group; 2010. p. 119–28.

57. Dettling M, Bühlmann P. Supervised clustering of genes [Internet]. 2002. Available from: http://genomebiology.com/2002/3/12/research/0069.1

58. Aitken A. 14-3-3 proteins: A historic overview. Vol. 16, Seminars in Cancer Biology. 2006. p. 162–72.

59. Strunz M, Jarrell JT, Cohen DS, Rosin ER, Vanderburg CR, Huang X. Modulation of SPARC/Hevin Proteins in Alzheimer's Disease Brain Injury. Journal of Alzheimer's Disease. 2019;68(2):695–710.

60. Xu L, Nirwane A, Yao Y. Basement membrane and blood-brain barrier. Vol. 4, Stroke and Vascular Neurology. BMJ Publishing Group; 2019. p. 78–82.

61. Dehouck B, Fenart L, Dehouck M-P, Pierce A, Torpier G, Cecchelli R. A New Function for the LDL Receptor: Transcytosis of LDL across the Blood-Brain Barrier [Internet]. Vol. 138, The Journal of Cell Biology. 1997. Available from: http://www.jcb.org

62. Trejo JL, Carro E, Garcia-Galloway E, Torres-Aleman I. Role of insulin-like growth factor I signaling in neurodegenerative diseases. Vol. 82, Journal of Molecular Medicine. 2004. p. 156–62.

63. Song IU, Kim Y do, Chung SW, Cho HJ. Association between serum haptoglobin and the pathogenesis of alzheimer's disease. Internal Medicine. 2015 Mar 1;54(5):453–7.

64. Jung SM, Lee KB, Lee JW, Namkoong H, Kim HK, Kim S, et al. Both plasma retinol-binding protein and haptoglobin precursor allele 1 in CSF: Candidate biomarkers for the progression of normal to mild cognitive impairment to Alzheimer's disease. Neuroscience Letters. 2008 May 9;436(2):153–7.

65. Bonifati DM, Kishore U. Role of complement in neurodegeneration and neuroinflammation. Molecular Immunology. 2007 Feb;44(5):999–1010.

66. Ho DM, Artavanis-Tsakonas S, Louvi A. The Notch pathway in CNS homeostasis and neurodegeneration. Wiley Interdisciplinary Reviews: Developmental Biology. 2020 Jan 1;9(1).

67. Zahra Paylakhi S, Ozgoli S, Paylakhi S. Identification of Alzheimer disease-relevant genes using a novel hybrid method [Internet]. Vol. 6, Progress in Biological Sciences. 2016. Available from: http://www.ncbi.nlm.nih.gov/pubmed

68. Tang MX, Stern Y, Marder K, Bell K, Gurland B, Lantigua R, et al. The APOE-ε4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. Journal of the American Medical Association [Internet]. 1998 [cited 2019 Oct 31];279(10):751–5. Available from: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/187321

69. Li Z, Shue F, Zhao N, Shinohara M, Bu G. APOE2: protective mechanism and therapeutic implications for Alzheimer's disease. Vol. 15, Molecular Neurodegeneration. BioMed Central Ltd; 2020.

70. Morris JC, Roe CM, Xiong C, Fagan AM, Goate AM, Holtzman DM, et al. APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. Annals of Neurology. 2010 Jan;67(1):122–31.

71. Jr CJ, Bennett D, Blennow K, … MC-A&, 2018 U. NIA-AA Research Framework_ Toward a biological definition of Alzheimer's disease _ Elsevier Enhanced Reader [Internet]. Elsevier. [cited 2019 Oct 31]. Available from: https://www.sciencedirect.com/science/article/pii/S1552526018300724

72. Sutphen CL, Jasielec MS, Shah AR, Macy EM, Xiong C, Vlassenko AG, et al. Longitudinal Cerebrospinal Fluid Biomarker Changes in Preclinical Alzheimer Disease During Middle Age. JAMA Neurology. 2015 Sep 1;72(9):1029.

73. Insel PS, Donohue MC, Berron D, Hansson O, Mattsson-Carlgren N. Time between milestone events in the Alzheimer's disease amyloid cascade. NeuroImage. 2021 Feb 15;227.

74. Herrmann J, Lerman LO, Lerman A. Ubiquitin and ubiquitin-like proteins in protein regulation. Vol. 100, Circulation Research. 2007. p. 1276–91.

75. Kim YN, Kim DH. Decreased serum angiogenin level in Alzheimer's disease. Progress in Neuro-Psychopharmacology and Biological Psychiatry. 2012 Aug 7;38(2):116–20.

76. Schneider Thomsen M, Birkelund S, Burkhart A, Stensballe A, Moos T. Synthesis and deposition of basement membrane proteins by primary brain capillary endothelial cells in a murine model of the blood-brain barrier. Journal of Neurochemistry. 2017 Mar;140(5).

77. Jung SM, Lee KB, Lee JW, Namkoong H, Kim HK, Kim S, et al. Both plasma retinol-binding protein and haptoglobin precursor allele 1 in CSF: Candidate biomarkers for the progression of normal to mild cognitive impairment to Alzheimer's disease. Neuroscience Letters. 2008 May 9;436(2):153–7.

78. Sweeney MD, Sagare AP, Zlokovic B v. Blood-brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. Vol. 14, Nature Reviews Neurology. Nature Publishing Group; 2018. p. 133–50.

79. Bassendine MF, Taylor-Robinson SD, Fertleman M, Khan M, Neely D. Is Alzheimer's Disease a Liver Disease of the Brain? Journal of Alzheimer's Disease. 2020 Apr 3;75(1):1–14.

Table 1. Demographics of the three groups of participants identified by longitudinal cerebrospinal fluid (CSF) assessments. Participants are shown stratified by Alzheimer Disease (AD) Biomarker classification. Participants across classifications differed by age and apolioprotein (APOE) ε4 status at enrollment. By the conclusion of the study, participants also differed by clinical dementia rating (CDR).

	Intermediate AD Biomarker	**AD Biomarker Positive**	**AD Biomarker Negative**	**p**
N	27	12	69	
Age at LP - baseline	64.61 (6.01)	61.68 (7.59)	56.83 (7.56)	< 0.001
Age at LP - most recent	75.47 (5.95)	72.83 (6.52)	68.40 (7.56)	< 0.001
Sex				0.386
Male	13 (48.1%)	4 (33.3%)	23 (33.3%)	
Female	14 (51.9%)	8 (66.7%)	46 (66.7%)	
APOE				< 0.001
23	1 (3.7%)	0 (0.0%)	9 (13.2%)	
24	2 (7.4%)	0 (0.0%)	2 (2.9%)	
33	12 (44.4%)	5 (41.7%)	45 (66.2%)	
34	11 (40.7%)	3 (25.0%)	11 (16.2%)	
44	1 (3.7%)	4 (33.3%)	1 (1.5%)	
Race				0.921
Black	3 (11.1%)	1 (8.3%)	5 (7.2%)	
White	**24** (88.9%)	11 (91.7%)	63 (91.3%)	
Unknown	0 (0.0%)	0 (0.0%)	1 (1.4%)	
Education (Years)	**16.22** (2.28)	14.75 (3.60)	16.14 (2.38)	0.186
CDR - baseline				
0	27 (100%)	12 (100%)	69 (100%)	
CDR - most recent				< 0.001
0	25 (92.6%)	7 (58.3%)	68 (98.6%)	
0.5	1 (3.7%)	4 (33.3%)	1 (1.4%)	
1	1 (3.7%)	0 (0.0%)	0 (0.0%)	
2	0 (0.0%)	1 (8.3%)	0 (0.0%)	

Figure 1. (A) Unsupervised clustering identified three distinct trajectories for longitudinal changes in cerebrospinal fluid (CSF) $pTau_{181}$ as a function of CSF Aβ42. AD Biomarker Negative individuals consistently had low levels of CSF p-tau, regardless of CSF Aβ42. Individuals within the Intermediate AD Biomarkers group had borderline positive levels of CSF $pTau_{181}$ with low levels of CSF Aβ42. Individuals within the AD Biomarker Positive group exhibited an increase in CSF $pTau_{181}$ with decreases in CSF Aβ42. (B) When CSF Aβ42 was normalized by CSF Aβ40, individuals appear to progress across a continuum where decreasing CSF Aβ42/Aβ40 was associated with increases in CSF pTau.

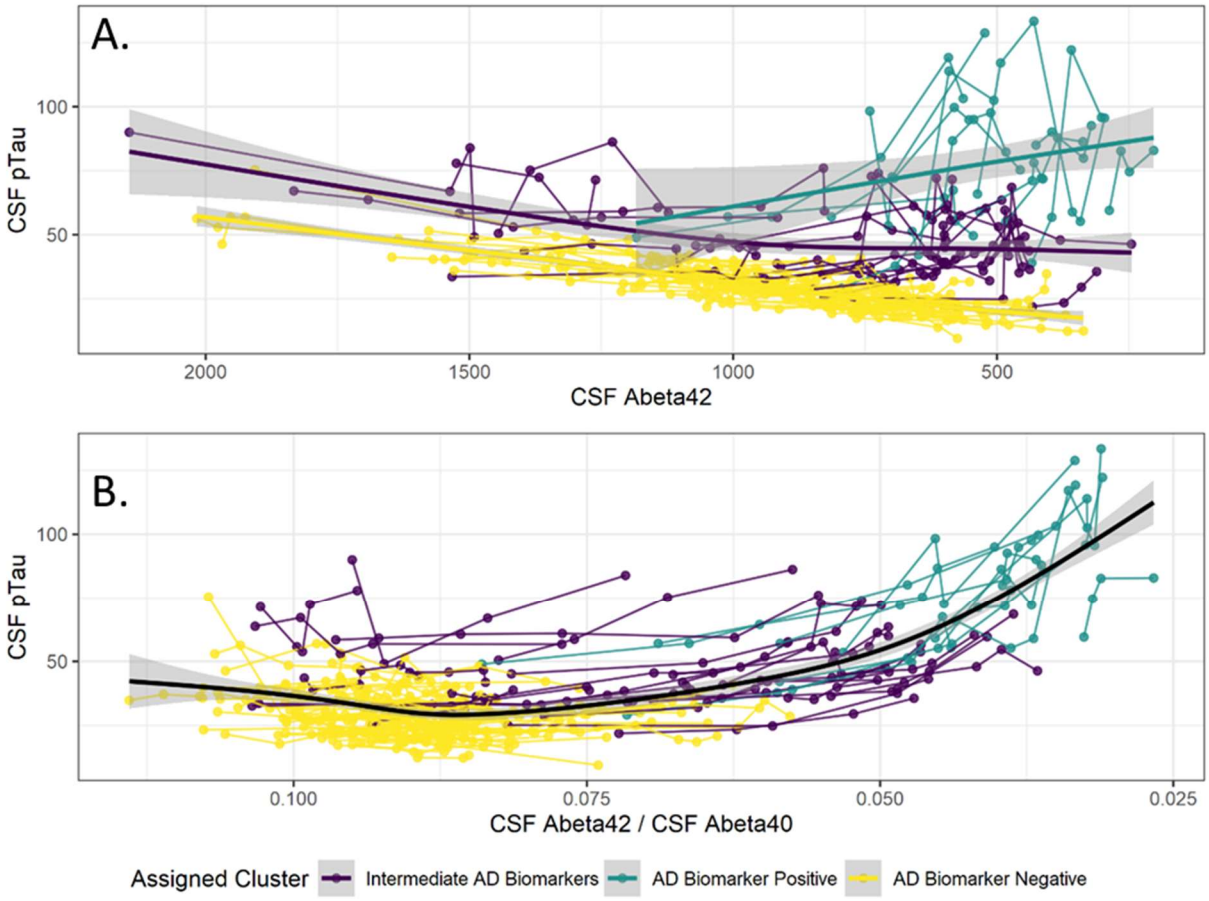


Figure 2. Survival analysis demonstrated that the AD Biomarker Positive group quickly proceeded to both amyloid and tau positivity. The Intermediate AD Biomarkers group lagged the AD Biomarker Positive group by about 10 years for amyloid positivity (as defined by CSF Aβ42/Aβ40 < 0.06753), but no statistically significant difference for tau positivity (as defined by CSF $pTau_{181}$ > 42.5). A majority of the AD Biomarker Negative group never became amyloid positive and became tau positive only after age 80.

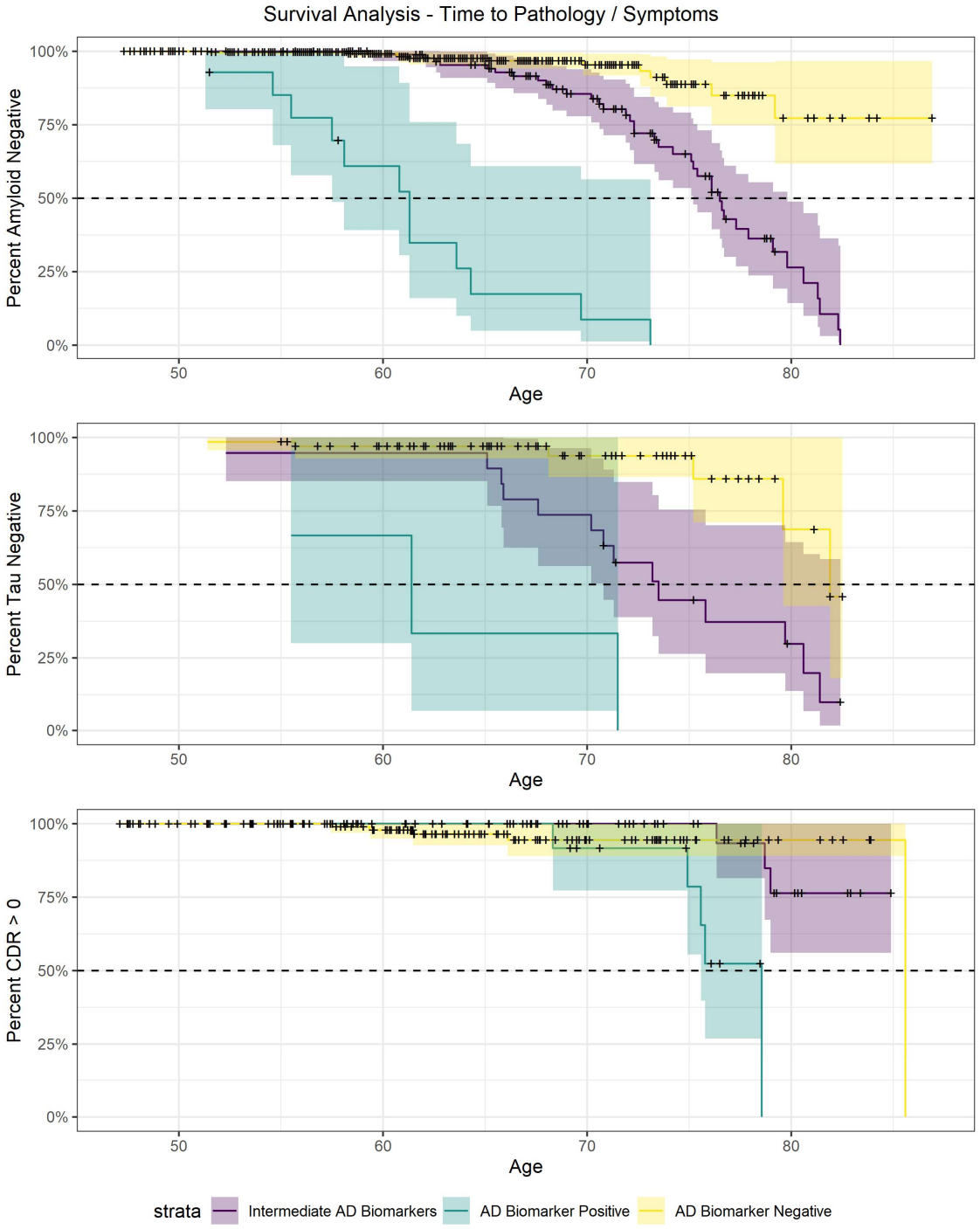


Figure 3. The three clusters exhibit different behaviors across the Amyloid and Tau phases of the AT(N) progression. AD Biomarker Positive individuals have the greatest amyloid accumulation as quantified by both the CSF Aβ42/Aβ40 ratio and PET-PiB imaging. They also have the greatest level of tau accumulation as quantified by both CSF $pTau_{181}$ and PET-AV1451 imaging. The Intermediate cohort develops both amyloid positivity and tau positivity during the period of enrollment of the study, but they become tau positive before they are amyloid positive. There are no differences in the selected neurodegenerative biomarkers across the clusters.

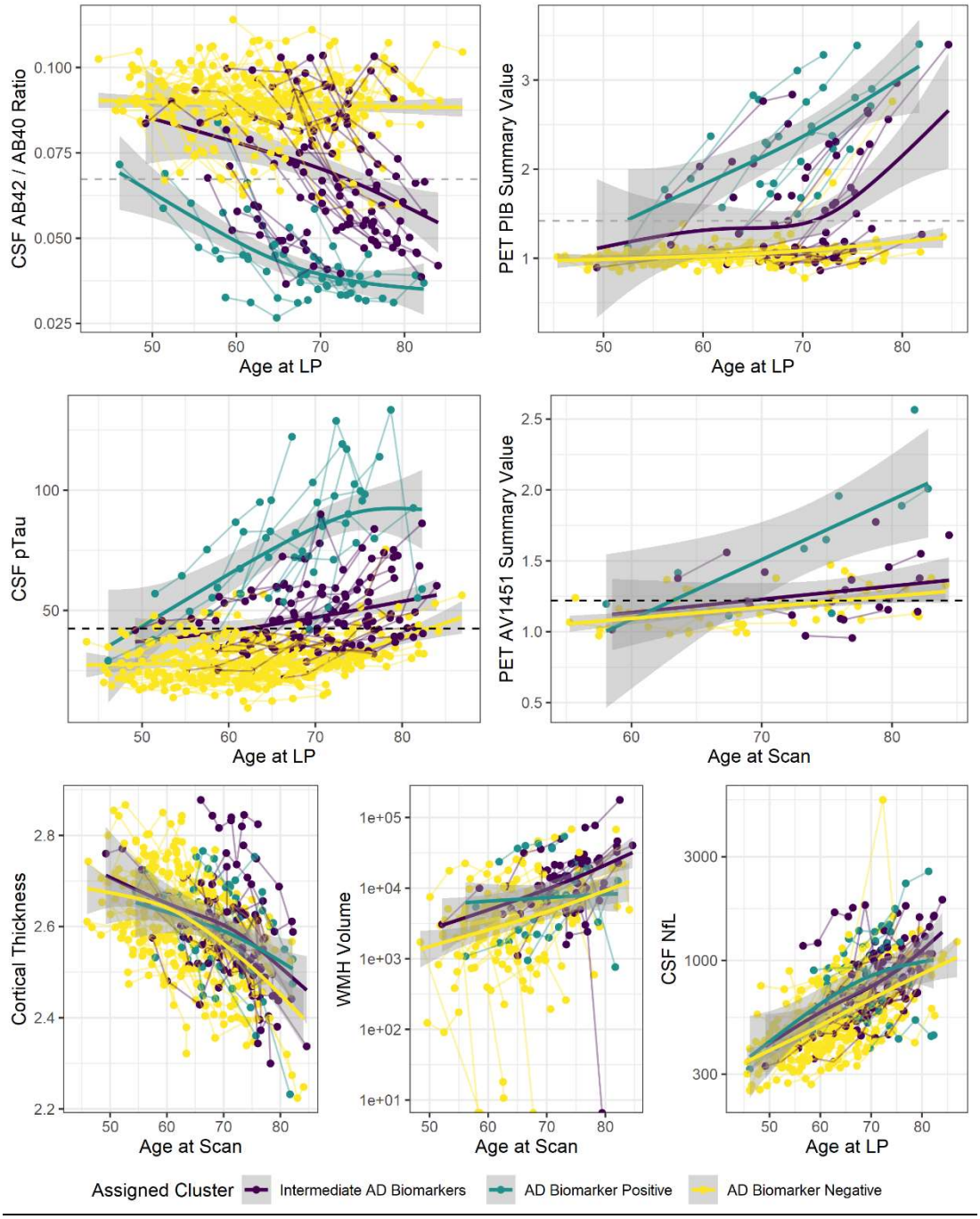
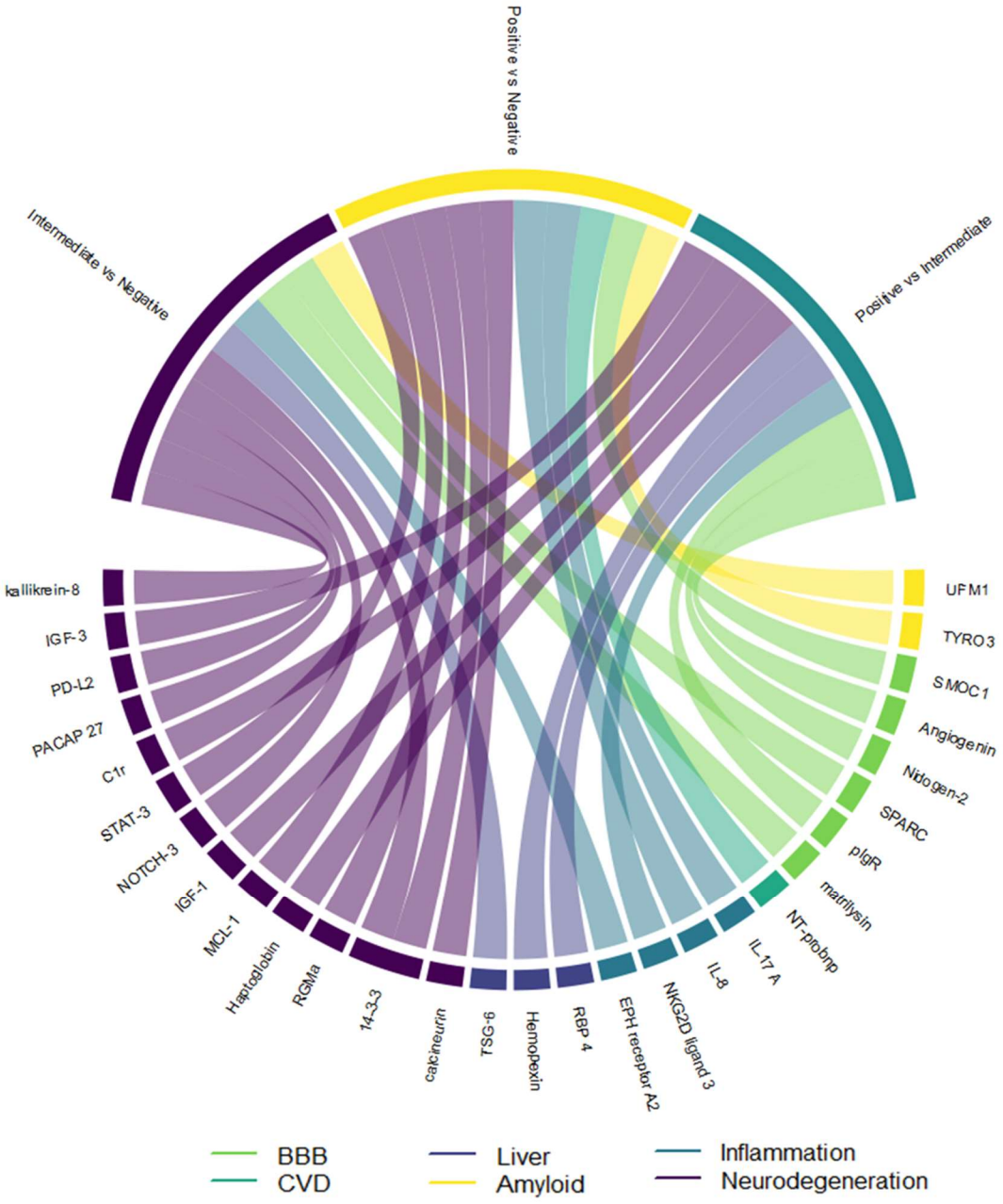
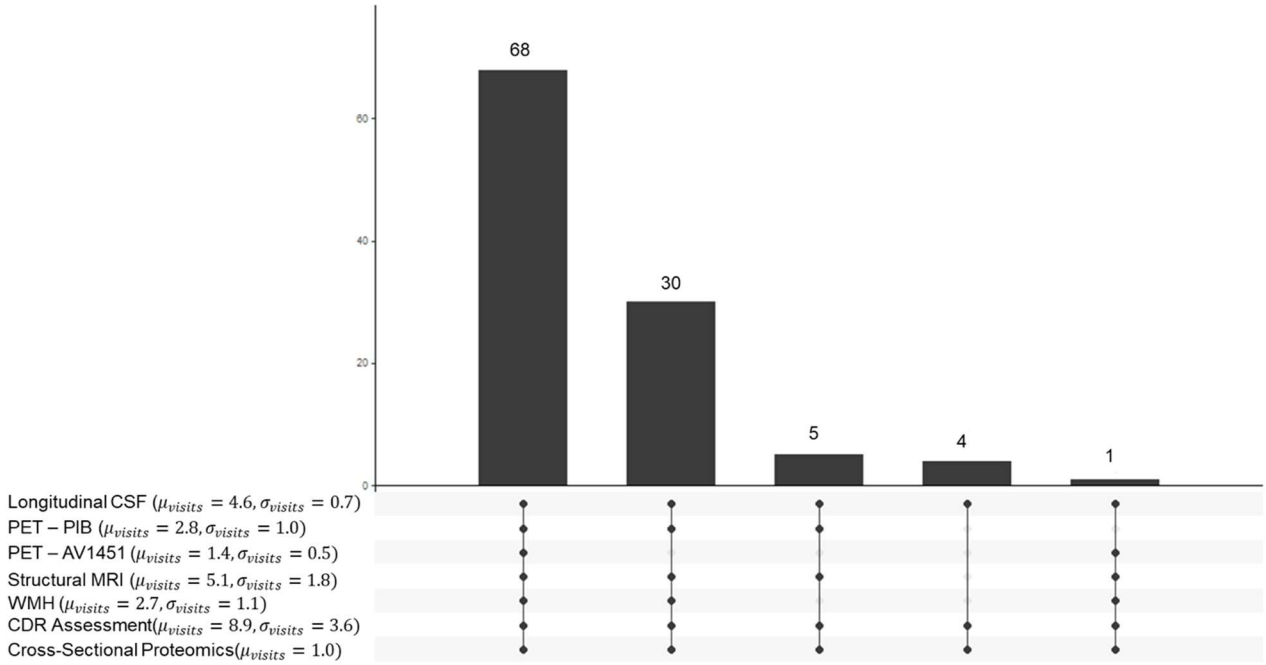


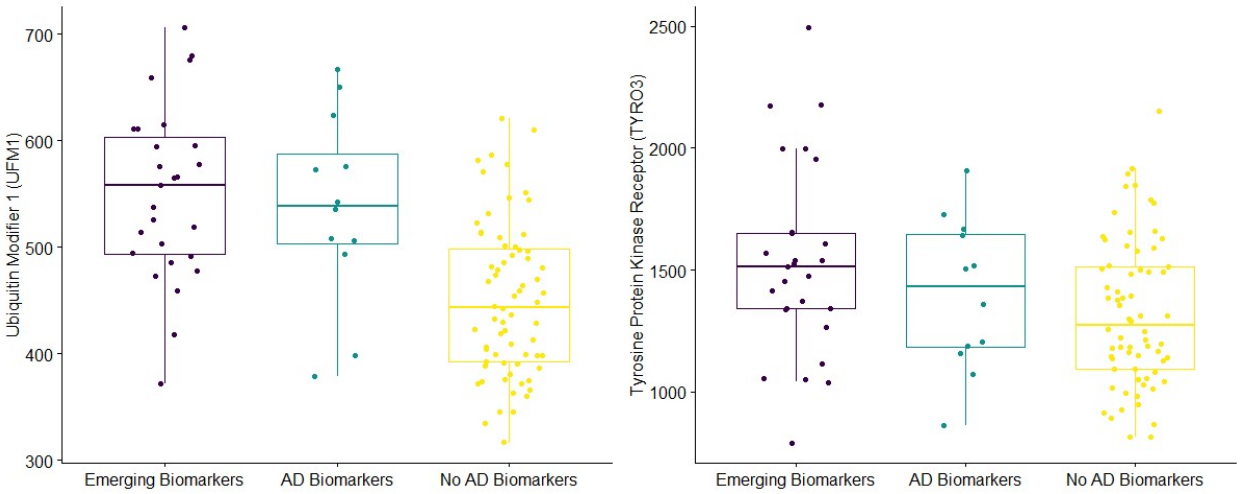
Figure 4. Using the log-loss criterion, we identified ten proteins that were most important for each classification (Intermediate vs. AD Biomarker Negative, AD Biomarker Positive vs. AD Biomarker Negative, AD Biomarker Positive vs. Intermediate AD Biomarkers). We then performed a literature review to classify the primary function of each protein as relating to either blood brain barrier (BBB), cardiovascular disease (CVD), liver function (Liver), amyloid production and/or regulation (Amyloid), inflammation, or neurodegeneration. Following the line from the listed protein to the classification allows the reader to see which proteins associate with which classification model. For example, the largest number of BBB-related proteins helped discriminate AD Biomarker Positive vs. Intermediate AD Biomarkers individuals. This can be detected by following the lines from plgR, Nidogen-2, and Angiogenin connecting the green portion of the protein section to the teal portion of the classification section.



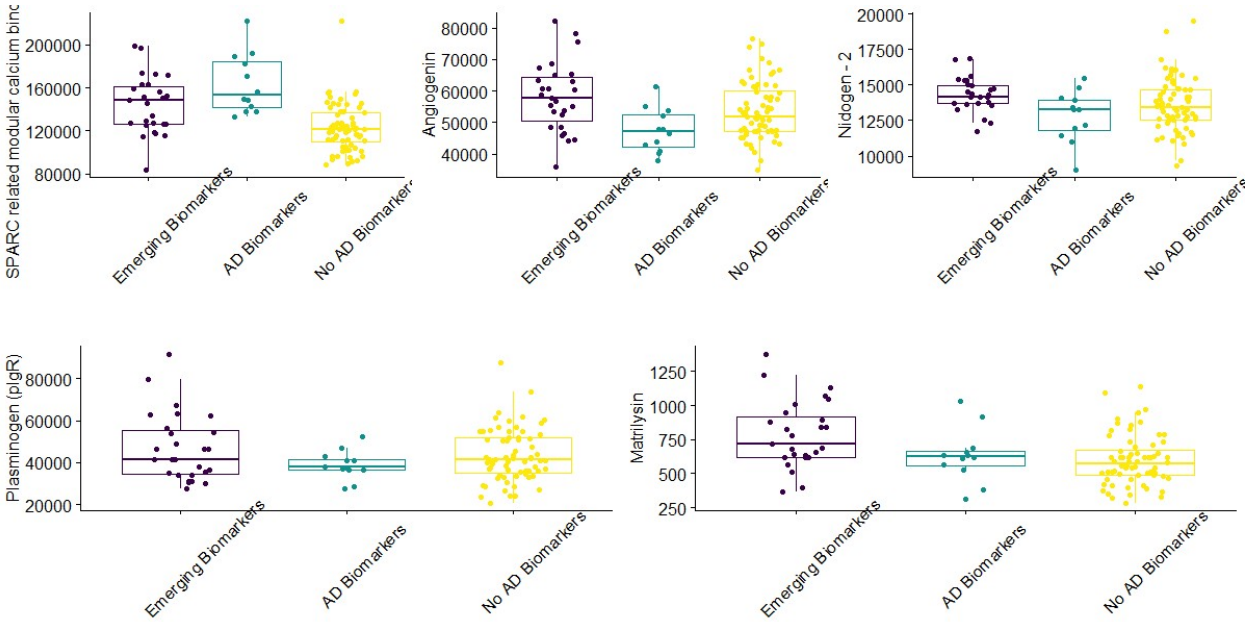
SUPPLEMENTAL MATERIALS



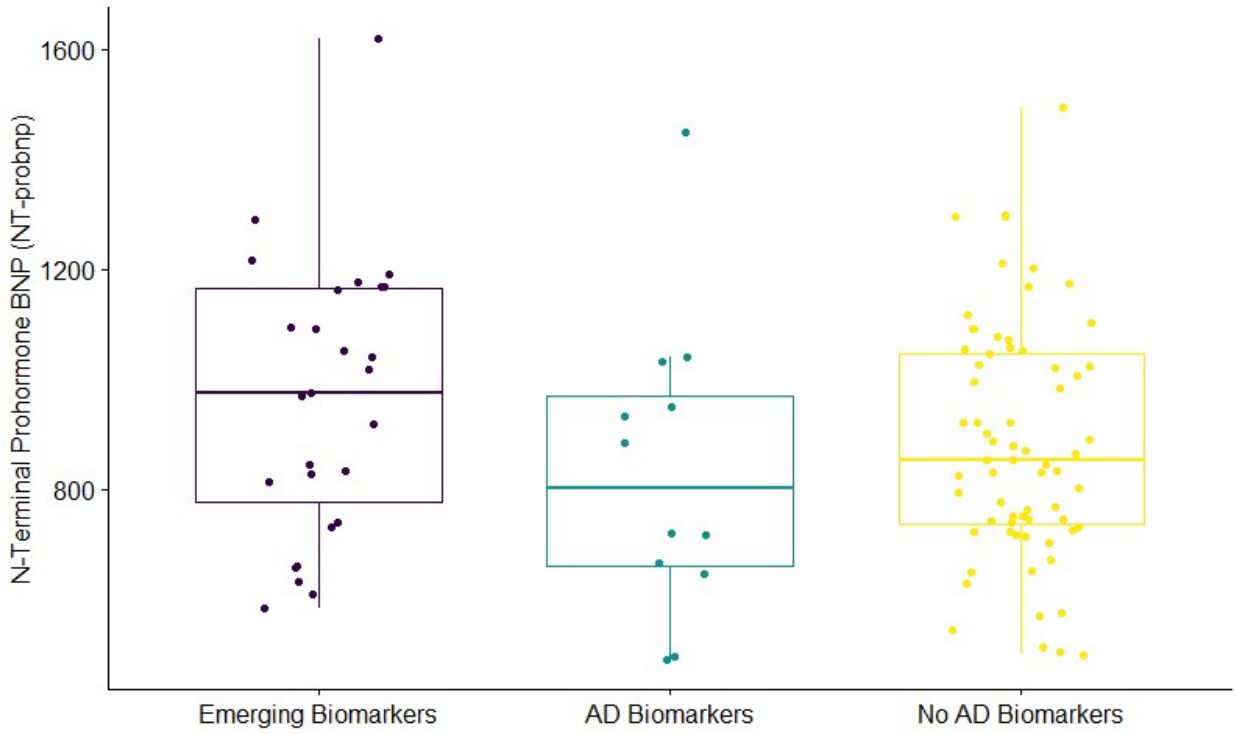
Supplemental Figure 1. Most (N = 68) participants provided all biomarkers and cognitive measures considered in this manuscript. The most frequent missing biomarker was PET – Tau (N = 39). For all measures considered except proteomics, longitudinal data was included where available. The mean number of visits is included in the image.



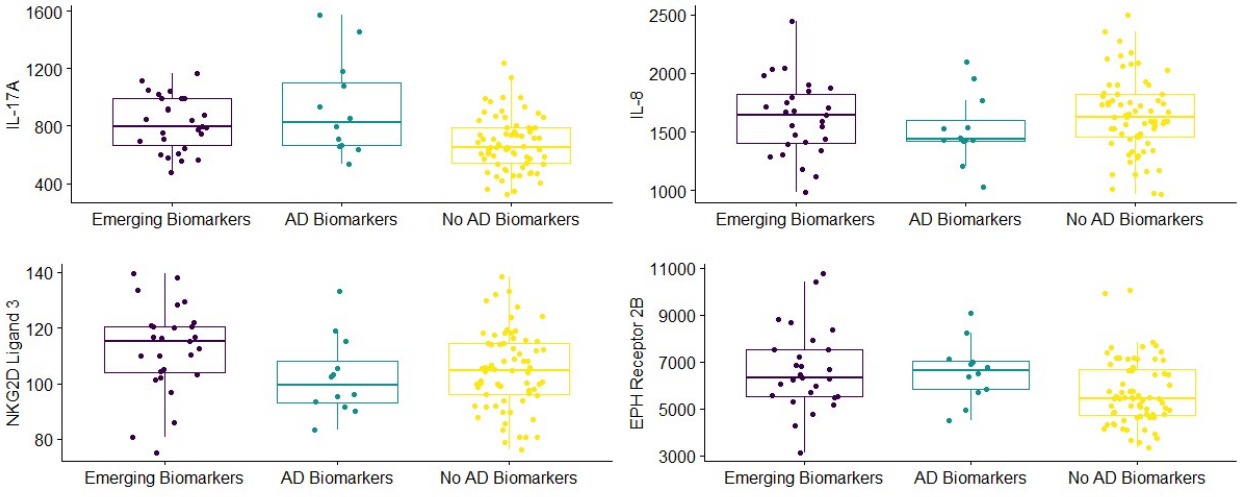
Supplemental Figure 2. Quantification of protein expression for identifying proteins associated with amyloid production and regulation



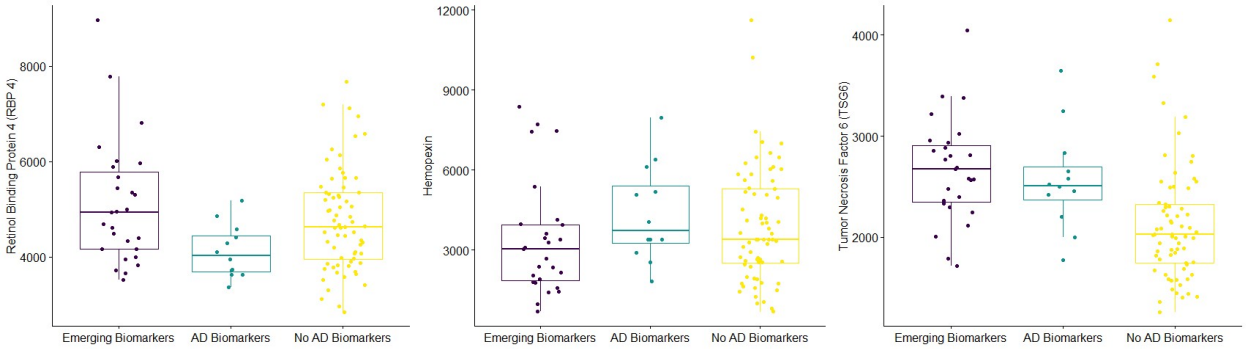
Supplemental Figure 3. Quantification of protein expression for identifying proteins associated with blood brain barrier integrity



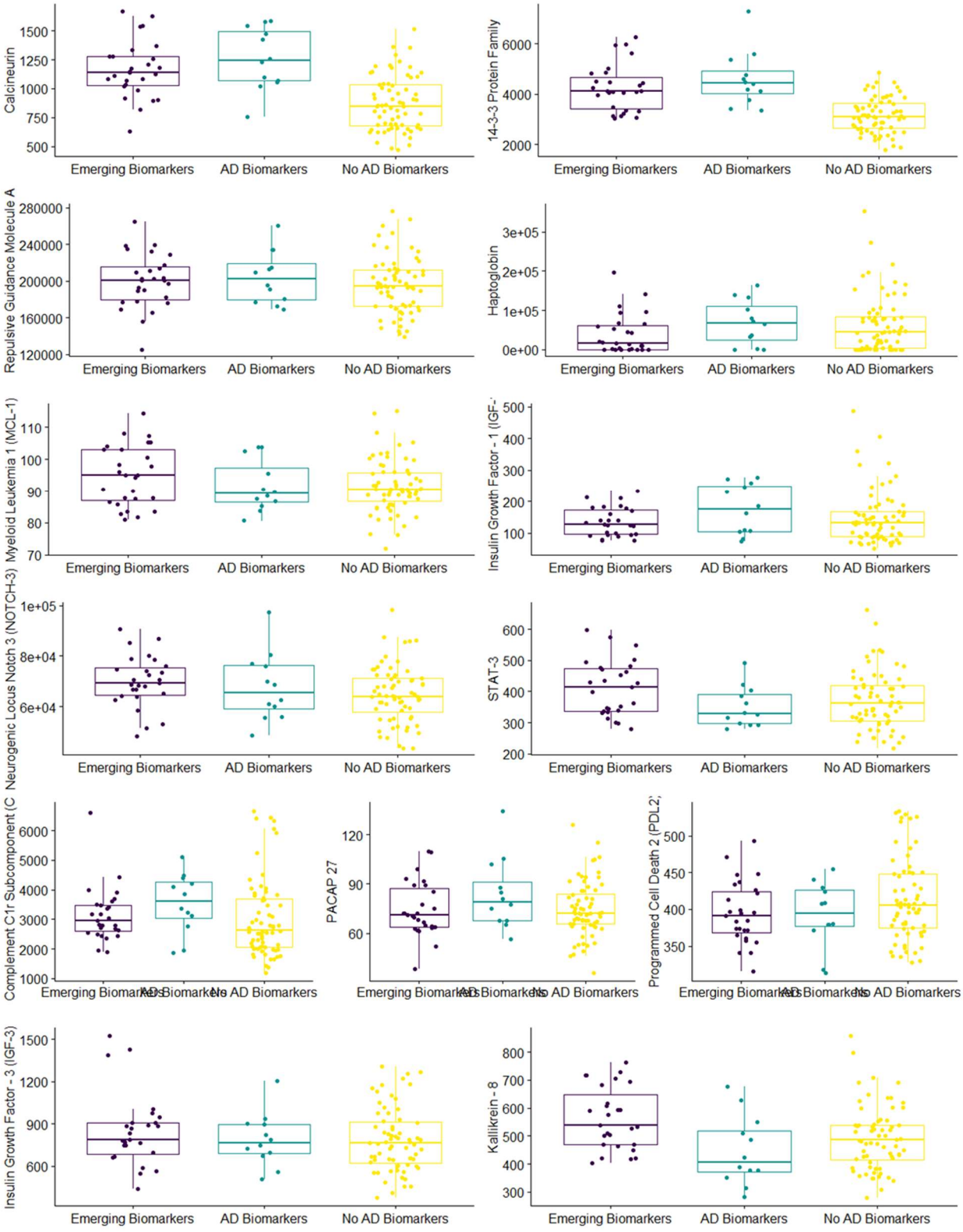
Supplemental Figure 4. Quantification of protein expression for identifying proteins associated with cardiovascular disease.



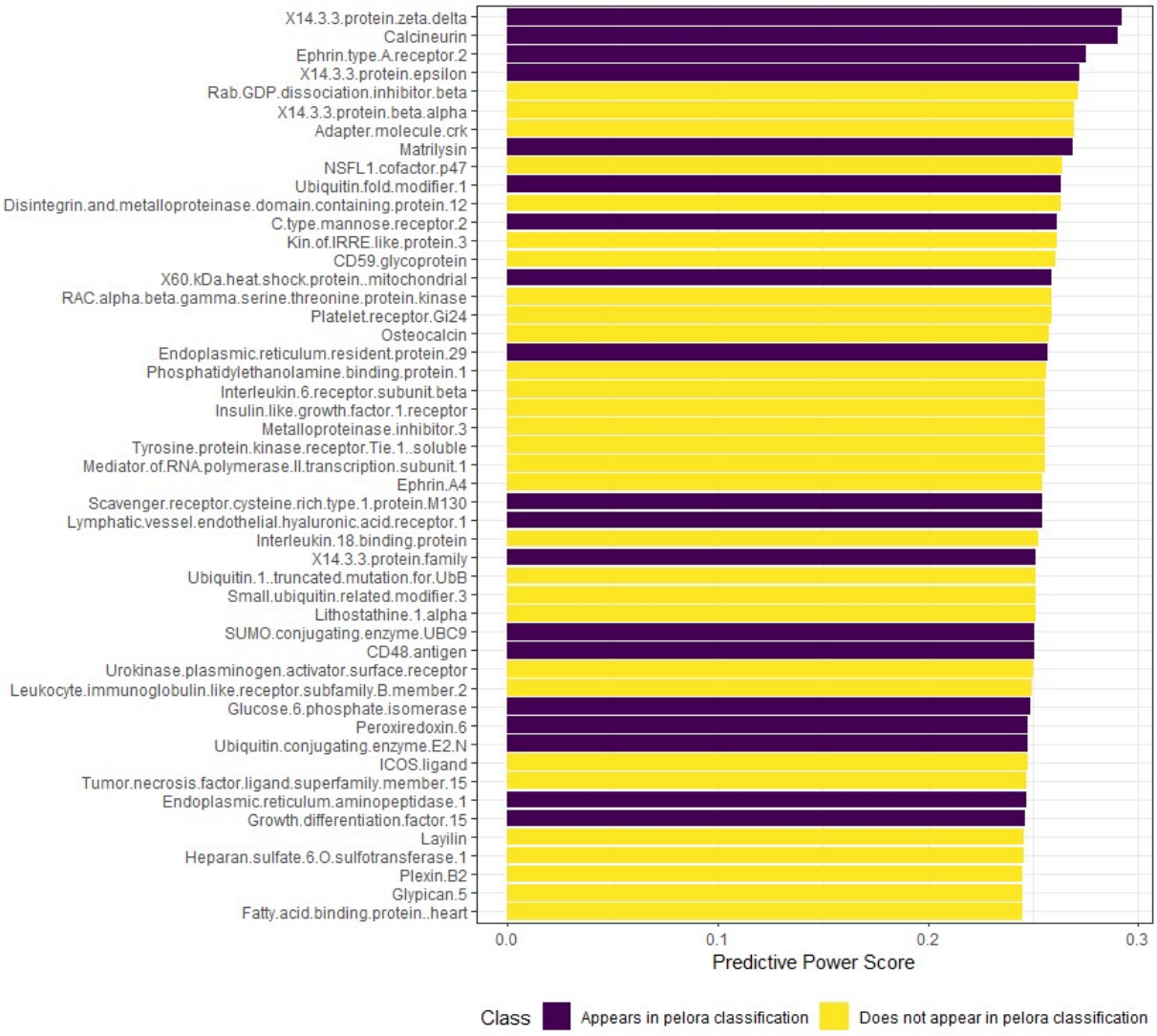
Supplemental Figure 5. Quantification of protein expression for identifying proteins associated with inflammation.



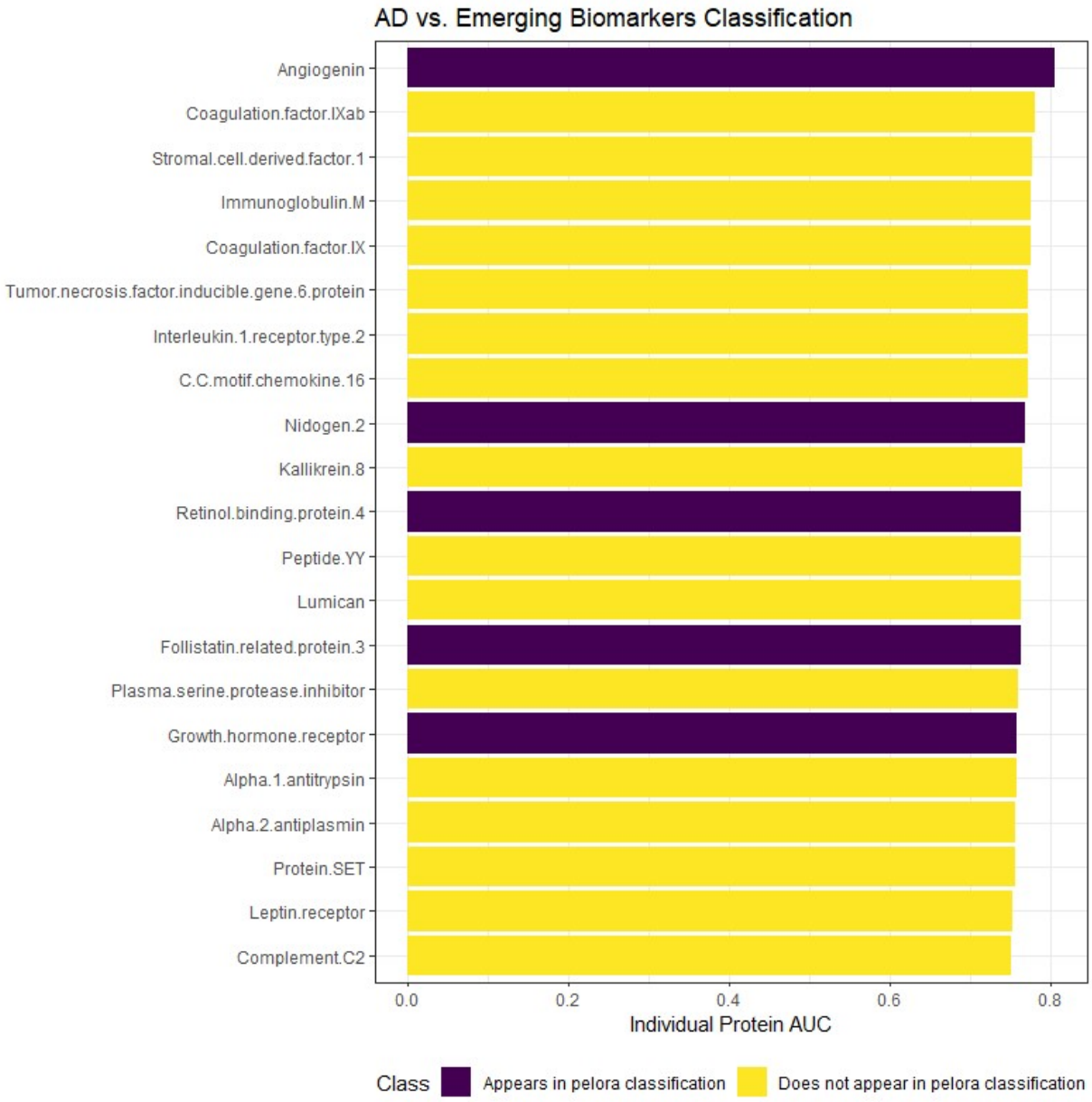
Supplemental Figure 6. Quantification of protein expression for identifying proteins associated with liver function



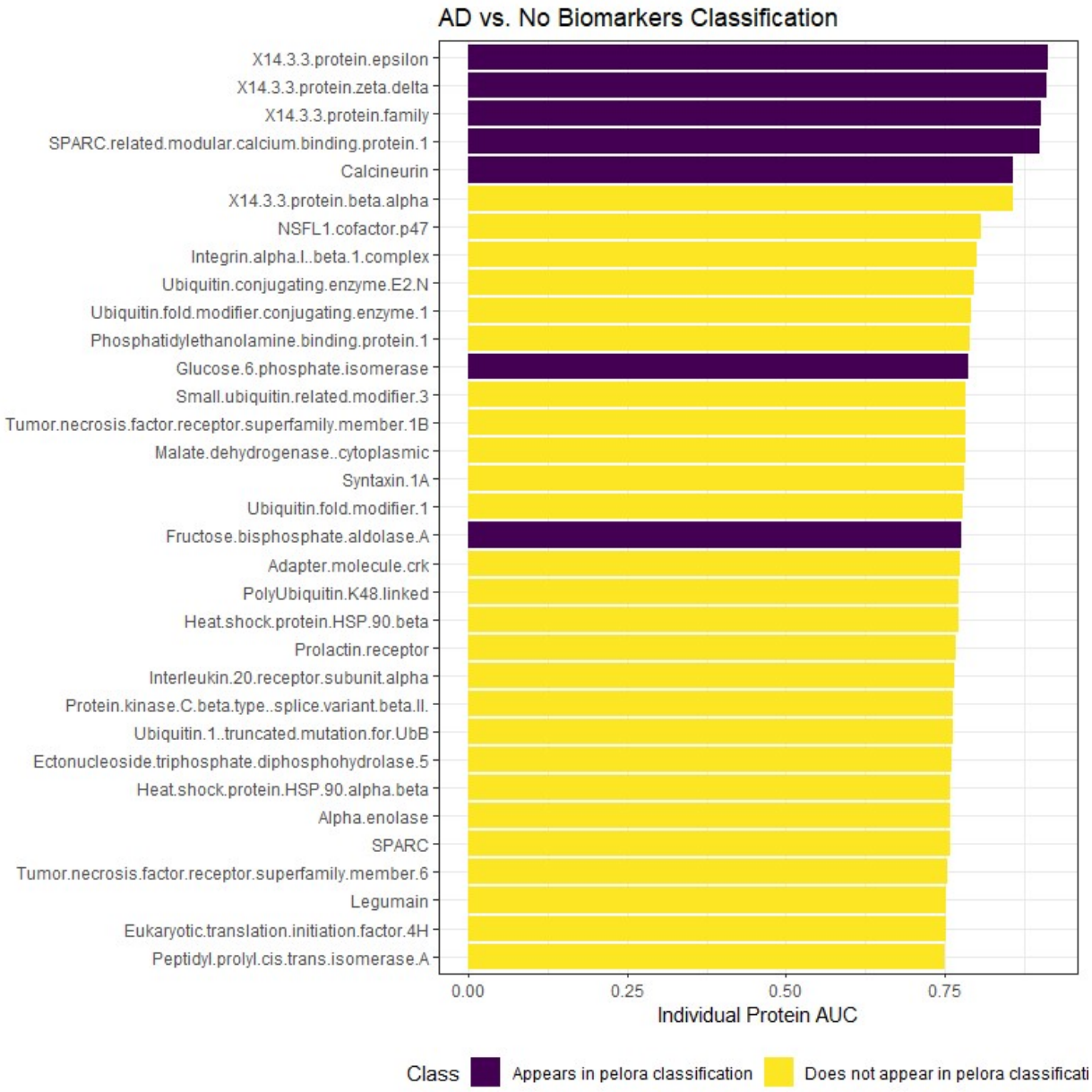
Supplemental Figure 7. Quantification of protein expression for identifying proteins associated with neurodegeneration



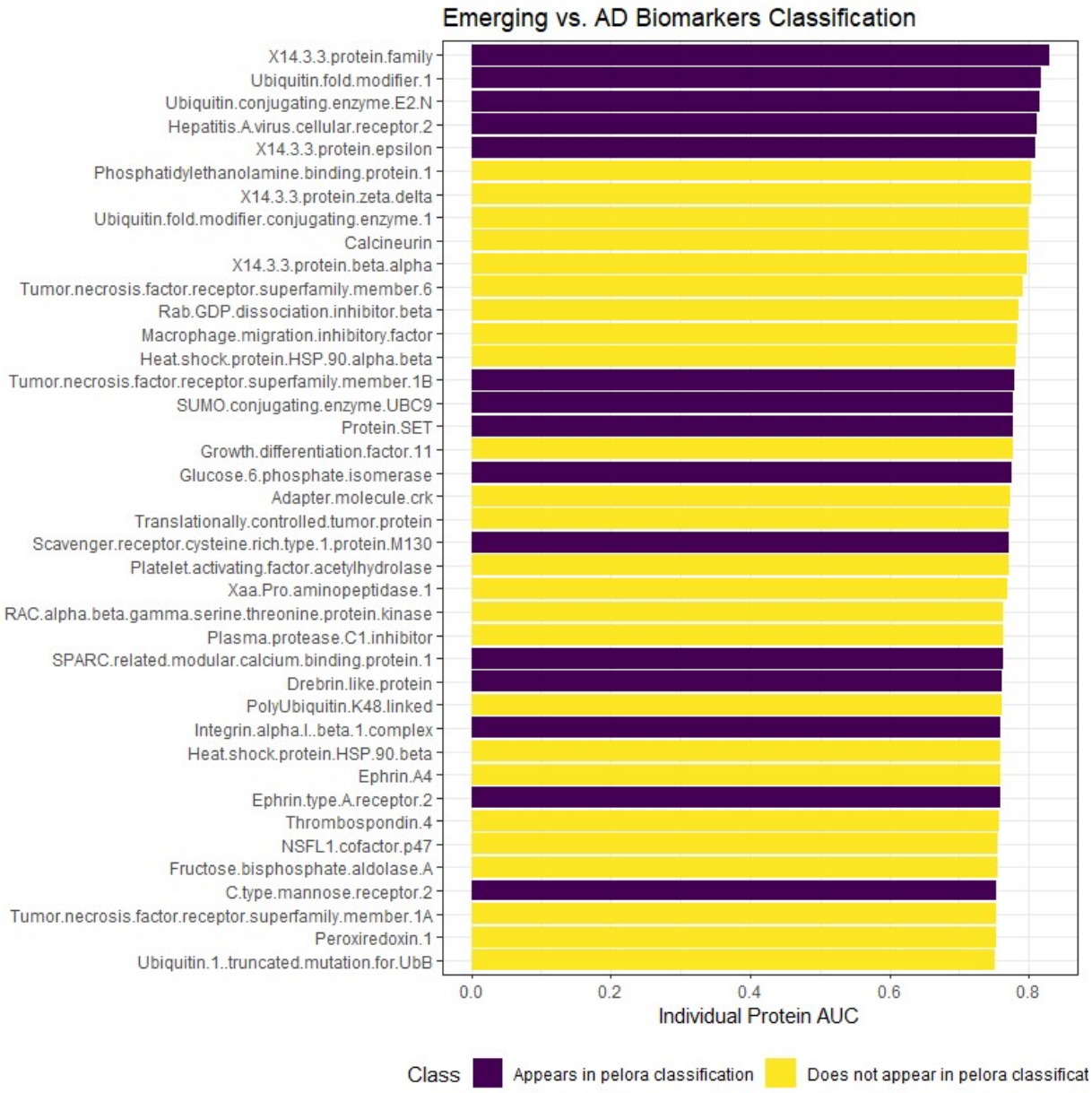
Supplemental Figure 8. The individual predictive power score (PPS) of each protein for classifying individuals for group membership is shown above. PPS is a value that ranges between 0 and 1. The three proteins with the highest individual predictive power (14-3-3 protein zeta-delta, calcineurin and ephrin type A receptor 2) were also identified by the pelora-based classification scheme. This indicates that the individual proteins are strongly tied to group membership, rather than being important by virtue of their interactions with other proteins.



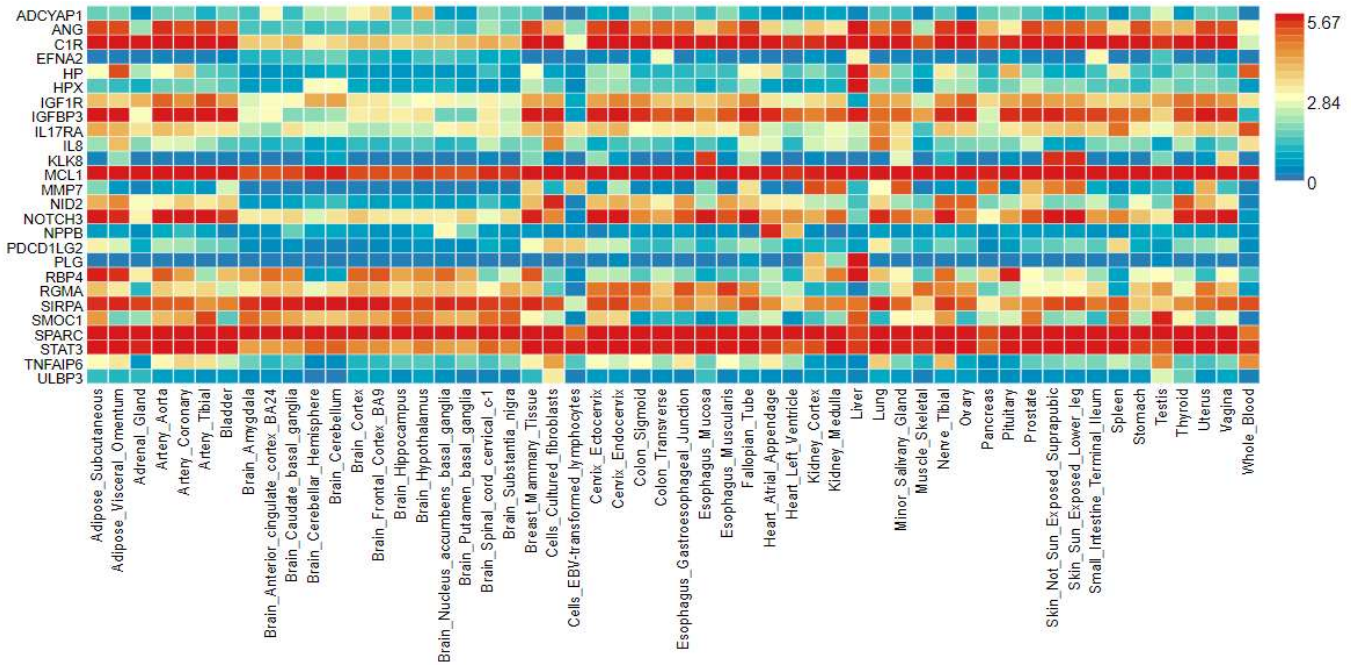
Supplemental Figure 9. All proteins with an individual AUC of greater than 0.75 when classifying individuals as belonging to either the AD Biomarker or Intermediate Biomarker cohort. The protein with the highest individual AUC, angiogenin, is also important for the pelora-based classification algorithm. Recall that the AD Biomarker vs. Intermediate Biomarker classification AUC was 0.75. This means that each of these individual proteins would actually be more useful for classification than the identified groupings of up and down-regulated proteins.



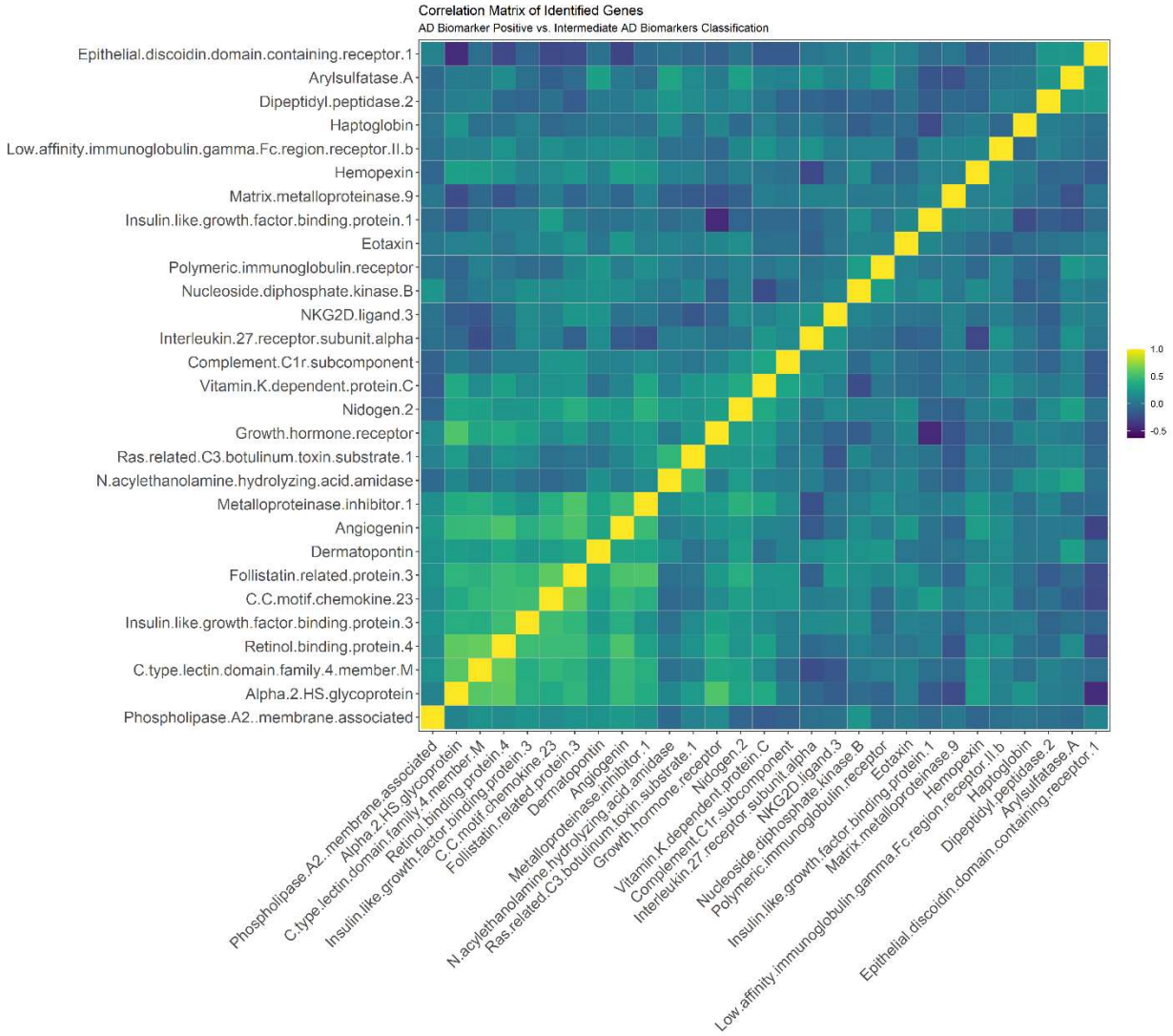
Supplemental Figure 10. All proteins with an individual AUC of greater than 0.75 when classifying individuals as belonging to either the AD Biomarker or AD Biomarker Negative cohort. Recall that the AD Biomarker vs. AD Biomarker Negative classification AUC was 0.98. This means that the combination of proteins identified as useful for classification (including proteins from the 14-3-3 family, SMOC1 and calcineurin) was substantially more powerful than any individual protein for classification.



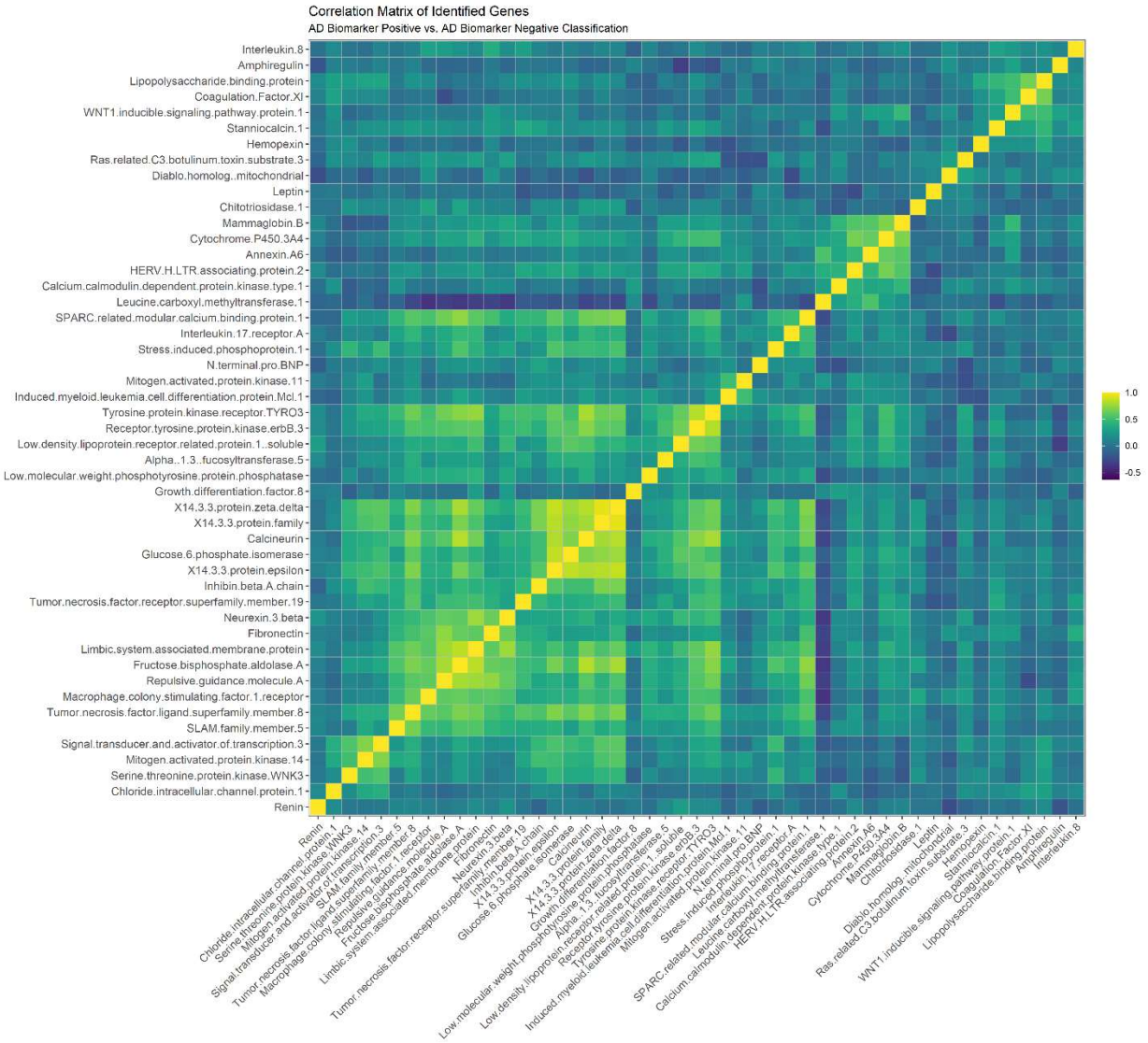
Supplemental Figure 11. All proteins with an individual AUC of greater than 0.75 when classifying individuals as belonging to either the Intermediate Biomarker or AD Biomarker Negative cohort. Recall that the Intermediate Biomarker vs. AD Biomarker Negative classification AUC was 0.68. This means that each of these individual proteins would actually be more useful for classification than the identified groupings of up and down-regulated proteins.



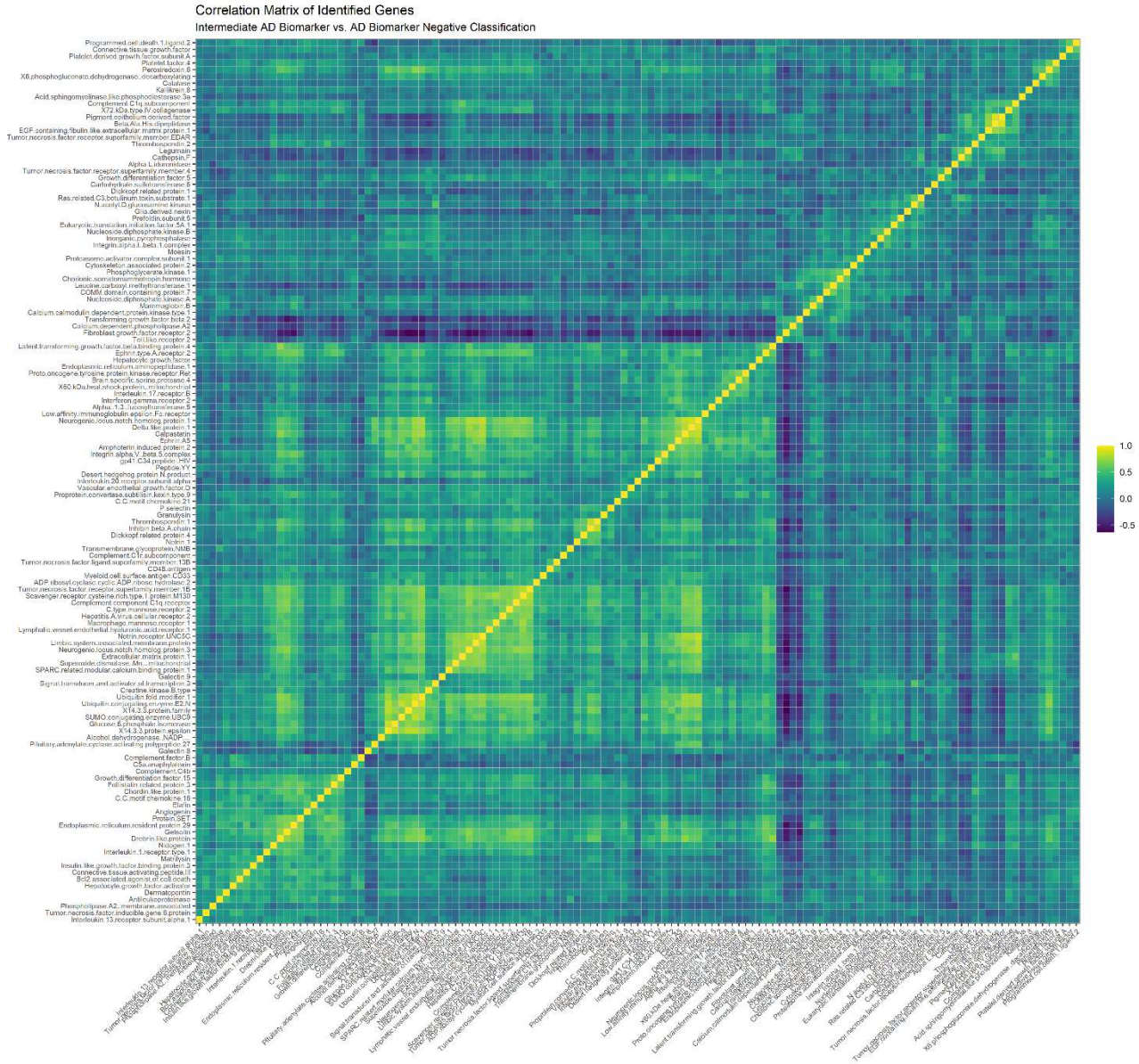
Supplemental Figure 12. Heatmap showing the average expression per protein (log2 transformed). This is the product of the pathway analysis conducted on the proteins identified as important for classification via pelora.



Supplemental Figure 13. Correlations between all proteins identified by Pelora as important for the AD Biomarker Positive – Intermediate AD Biomarker classification.



Supplemental Figure 14. Correlations between all proteins identified by Pelora as important for the AD Biomarker Positive – AD Biomarker Negative classification.



Supplemental Figure 15. Correlations between all proteins identified by Pelora as important for the Intermediate AD Biomarker – AD Biomarker Negative classification.