**Proteomic Clusters Underlie Heterogeneity in Preclinical Alzheimer Disease Progression**

Julie K. Wisch PhD1 (julie.wisch@wustl.edu), Omar H. Butt, MD, PhD1 (omarhbutt@wustl.edu), Brian A. Gordon PhD2,3,6 (bagordon@wustl.edu), Suzanne E. Schindler, MD, PhD1,6 (schindler.s.e@wustl.edu), Anne M. Fagan, PhD1,6 ([fagana@wustl.edu](mailto:fagana@wustl.edu)), Rachel L. Henson, MS1, Chengran Yang PhD4, 5 (chengranyang@wustl.edu), Anna H. Boerwinkle BS1 (anna.boerwinkle@wustl.edu), Tammie L.S. Benzinger MD, PhD2,6 (benzingert@wustl.edu), David M. Holtzman MD1, 3, 6 (holtzman@wustl.edu), John C. Morris MD1,6 (jcmorris@wustl.edu), Carlos Cruchaga PhD3,6 (cruchagac@wustl.edu), Beau M. Ances MD, PhD1,2,3,6 (bances@wustl.edu)

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](mailto:bances@wustl.edu)

**Running Title:** Proteomic Clustering on AD Progression

**Keywords:** Alzheimer Disease; Biomarkers; Proteome; Heterogeneity; Machine Learning

**Abbreviations:**

AD – Alzheimer Disease

AT(N) – Amyloid Tau Neurodegeneration Framework

CSF – Cerebrospinal Fluid

NfL – Neurofilament Light Chain

MRI – Magnetic Resonance Imaging

PiB – Pittsburgh Compound B

PET – Positron Emission Tomography

AV1451 – 18F – Flortaucipir Tau PET tracer

Knight ADRC – Knight Alzheimer Disease Research Center

WUSTL – Washington University in St. Louis

CDR – Clinical Dementia Rating

APOE – Apolipoprotein E

LP – Lumbar Puncture

ELISA – Enzyme-linked Immunosorbent Assay

FLAIR – Fluid Attenuated Inversion Recovery

WMH – White Matter Hyperintensities

SPM – Statistical Parametric Mapping

SUVR – Standardized Uptake Value Ratio

GAMM - Generalized Additive Mixed Effect Model

QC – Quality Control

AUC – Area Under the Curve

NIAGADS – National Institute on Aging Genetics of Alzheimer’s Disease Data Storage Site

BBB – Blood Brain Barrier

CVD – Cardiovascular Disease

SPARCL1 – SPARC-related Modular Calcium-Binding Protein 1

IGF-1 - Insulin-like Growth Factor-Binding Protein 1

RBP 4 – Retinol-Binding Protein 4

Original Article - Brain

Word Count for Abstract: 324 / 400 words Character Count for Title: 72

Total Word Count: 5970 / 6000 words Number of References: 87

Number of Tables: 1

Number of Figures: 4 Total No. of Display Items: 5 / 8

ABSTRACT

Background: Heterogeneity in progression to Alzheimer Disease poses challenges for both clinical prognosis and clinical trial implementation. Multiple AD-related subtypes have previously been identified, suggesting differences in receptivity to drug interventions. We identified early differences in preclinical Alzheimer Disease biomarkers, assessed patterns for developing preclinical Alzheimer Disease across the Amyloid-Tau-(Neurodegeneration) framework, and considered potential sources of difference by analyzing the cerebrospinal fluid proteome.

Methods: 108 participants enrolled in longitudinal studies at the Knight Alzheimer Disease Research Center completed four or more lumbar punctures. These individuals were cognitively normal at baseline. Cerebrospinal fluid measures of A42, pTau181, and Neurofilament Light chain as well as proteomics values were evaluated. Imaging biomarkers, including positron emission tomography amyloid and tau and structural magnetic resonance imaging were repeatedly obtained when available. Individuals were staged according to the Amyloid-Tau-(Neurodegeneration) framework.

Results: Growth mixture modeling, an unsupervised clustering technique, identified three patterns of biomarker progression as measured by cerebrospinal fluid pTau181 and A42. Two groups (Alzheimer Disease Biomarker Positive and Intermediate Alzheimer Disease Biomarker) had distinct progression from normal biomarker status to having biomarkers consistent with preclinical Alzheimer Disease. A third group (Alzheimer Disease Biomarker Negative) did not develop abnormal Alzheimer Disease biomarkers over time. Participants grouped by CSF trajectories were re-classified using only proteomic profiles (AUCAD Biomarker Positive vs AD Biomarker Negatives = 0.857, AUCAD Biomarker Positive vs. Intermediate AD Biomarkers = 0.525, AUCIntermediate AD Biomarkers vs. AD Biomarker Negative = 0.952).

Conclusions: We highlight heterogeneity in the development of AD biomarkers in cognitively normal individuals. We identified some individuals who became amyloid positive before age 50. A second group, Intermediate AD Biomarkers, developed elevated CSF ptau181 significantly before becoming amyloid positive. 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 non-AT(N) biomarkers for potential therapies, including blood brain barrier-, vascular-, immune-, and neuroinflammatory-related targets.

# Introduction

Alzheimer Disease (AD), a slowly progressive neurodegenerative disorder with an extended prodromal stage, affects nearly 6 million Americans1. It is a disease that progresses from a clinically asymptomatic preclinical phase to a symptomatic clinical phase over many years2. 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 pTau181), 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 pTau181 increases as amyloid accumulates and is a reflection of neurofibrillary tangles 8. Studies have also attempted to forecast symptomatic development of AD using CSF measures 9,10. Of these, the CSF pTau181/ Aβ42 ratio was the most prognostic; however, this ratio was not able to accurately predict time to symptom onset 9. 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.

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. Important recent work has been performed linking CSF measures of amyloid and tau to the proteome11,12, highlighting observable subtypes in proteomic profile and arguing that perhaps even previously failed drug interventions may have been effective in mitigating AD-related damage in specific sub-populations.

Recent reviews have highlighted the importance of longitudinal research, particularly with respect to heterogeneity 13; 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 14. Longitudinal studies are preferable for studying biomarker dynamics 15. Obtaining repeated samples from the same individuals over time allows for a deeper understanding of potential time courses for disease development.

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 approximately 10 years, in addition to completing other traditional measures of amyloid, tau, and neurodegeneration (AT(N)). For this analysis we compared the longitudinal relationship between CSF Aβ42 and CSF pTau181 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 additional analysis of proteomics.

Evaluating protein expression in preclinical AD and healthy aging provides insight into additional potential biological mechanisms that underlie observed heterogeneity 11,12,16. Beyond simply characterizing our participants using AD biomarkers, proteomics identifies additional 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.

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

# Materials and methods

We included 108 participants (Table 1) enrolled in longitudinal studies at the Knight ADRC, Washington University in St Louis (WUSTL) as previously described 17. 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. 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*). Of the 108 cognitively normal enrollees, 8 developed cognitive impairment (defined by Clinical Dementia Rating (CDR) ≥ 0.5) during participation. 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 CDR scale. The CDR classifies the degree of cognitive impairment through the use of semi-structured interviews 18. 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 18.

**APOE Genotyping**

DNA samples were collected at enrollment and genotyped using either an Illumina 610 or OmniExpress chip. Genotyping methods have been previously described 19.

**Cerebrospinal fluid (CSF) acquisition, collection and processing**

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 20. LP was performed at 8:00 AM following an overnight fast. 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. Centrifugation occurred at 2000xg for 5 minutes. 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 ptau181) 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 21–23. We calculated the average cortical thickness 23.

**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) 24.

**Positron emission tomography (PET) imaging**

PET scans using [11C] PiB were obtained via previously described methods 25. Images were then processed using the PET unified pipeline (PUP, https://github.com/ysu001/PUP) 26,27. Images were smoothed to achieve a spatial resolution of 8 mm. This minimized inter-scanner differences 27,28. A standard image registration technique was used to correct for motion 29,30 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 31. We applied partial volume correction via a geometric transfer matrix approach 32. 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) 33.

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 33. 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 25.

**CSF Proteome Analysis**

All participants had at least one CSF sample processed for proteomics profiling, and nearly all had two. All samples were retained for analysis. Briefly, proteomic data was generated using the SomaScan 1.3k panel (SomaLogic Inc), an aptamer-based platform as previously described 34,35. There were initially 1305 proteins. Quality control (QC) was performed at the sample and aptamer levels using control aptamers (positive and negative controls) and calibrator samples. We removed protein outliers in a four step process: 1) We applied minimum detection filtering, removing samples that were detected at levels less than two standard deviations below the average expression level of the dilution buffer or if the total number of outliers for a given aptamer was more than 15% of the total sample size, we removed the aptamer. 807 aptamers passed this step. 2) We flagged analytes based on the difference of the scale factor. 749 aptamers passed this step. 3) We required that the coefficient of variation of calibrators was less than 0.15. 749 aptamers passed this step. 4) We removed outliers that were more than 1.5-fold of IQR outside of the distribution given the log10-based protein level. 746 aptamers were retained for analysis. 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**

**Latent Cluster Identification**

An unsupervised machine learning technique called growth mixture modeling was used to cluster the longitudinal values of individuals’ CSF pTau181 as a function of CSF Aβ42 36. This approach identifies possible sub-groups within longitudinal data and has previously been employed to study cognitive trajectories in AD 4,37,38 and structural changes 39, but not in preclinical amyloid and tau biomarkers. A single cluster growth mixture model, specified as in Equation 1, functions identically to a linear mixed effect model38, where the objective is to quantify changes over time in a single population, allowing for unique slopes and intercepts by individual. We define it in detail, using notation drawn from structural equation modeling, here. The observed longitudinal data (Y[t]n, in this case, measured CSF pTau181) is a function of the multiple measures of CSF Aβ42 (A0c[t]) and the observed occasion – to – occasion changes in CSF Aβ42 (A1c[t]). g0 and g1 are the latent variables used to express the shape of change, the mean change, and the covariance structure of the data.

Y[t]n = g0n · A0[t] + g1n · A1[t] + e[t]n (1)

Once we expand the growth mixture model to allow for multiple sub-groups within the data (as denoted by the subscript c, for clusters) (Equation 2), it no longer functions in the same manner as a linear mixed effect model. Instead, the model allows for the possibility that individuals are drawn from multiple populations and seeks to represent the probability, πnc, that an individual belongs to a given population (structural equation modeling utilizes c to denote cluster membership). Note that πnc must be a valid probability distribution.

Y[t]n  =∑Cc=1(πnc(g0nc · A0c[t] + g1nc · A1c[t] + e[t]nc)) (2)

In our application of growth mixture modeling, we used individual ID as the subject identifier, using the r function hlme() (lcmm, v1.9.4). We searched for 1, 2, 3, and 4 latent clusters and selected the optimal number of clusters via Bayesian Information Criterion (BIC) minimization. Initial starting values for the latent variables for the multi-clusters models were selected using the optimal values identified for the single cluster model. We searched up to four latent clusters, as standard operating practice in growth mixture modeling is to search for up to one additional cluster beyond what would be anticipated *a priori*36, and prior literature frequently stratifies individuals living with AD into three subtypes based on measured levels of CSF pTau and tTau 40–42. Recent clustering work looking at the proteome in different AD subtypes also identifies three clusters12.

Following this analysis, we compared participant demographics across the identified latent clusters, using the R package tableone43. Presented p values are derived from chi-squared tests of categorical variables and one-way ANOVA of continuous variables. This post-hoc analysis of identified clusters is the recommended approach to understanding previously unobserved sub-populations36. We further applied generalized additive mixed models (GAMM) to a visualization of this stratification by subpopulation. CSF pTau181 was the response variable. We fitted a cubic regression spline to CSF Aβ42, and also allowed CSF Aβ42 to vary by the identified latent cluster, fitting it with a factor smooth interaction (“fs”), which is effectively the equivalent of a random slope in a linear mixed effects model.

**Comparison of Latent Clusters to A-T(N) Measures**

To evaluate the time to pathology development, we performed survival analysis 44 to determine age at amyloid positivity (using CSF Aβ42/Aβ40 < 0.0673 pg/mL as the threshold 45) and age at tau positivity (using the CSF pTau181 > 42.5 pg/mL as the threshold 45, 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 approximately 2 years apart. Clinical visits for CDR rating occurred roughly annually, although this varied by cluster (see Table 1). In the survival analysis, we grouped by the latent clusters that were identified in the previous analysis (relying on the relationship of CSF pTau to Aβ42 ratio). Some participants were amyloid positive (13 / 108) and/or tau positive (20 / 108) prior to study enrollment, necessitating the application of interval syntax, which allows for both left and right censored data in survival analysis. We did not correct for any additional covariates in this analysis.

We then evaluated the trajectory of a variety of amyloid (CSF Aβ42/Aβ40 and PET-PiB summary value), tau (CSF pTau181 and PET-AV1451 Tauopathy) and neurodegeneration (cortical thickness, WMH volume, CSF NfL) biomarkers, testing for differences across latent clusters. We again applied GAMMs, following previously described methodology 46. We fitted a cubic spline to age at the time of the procedure and allowed age to vary by identified latent cluster (based on the longitudinal CSF pTau/Aβ42 ratios), fitting the splinewith a factor smooth interaction. We limited the age by latent cluster interaction to 3 knots (because there were 3 latent clusters we could maximally vary the age by cluster interaction by 3 knots).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.

**Proteome Classification and Analysis**

Finally, we applied Pelora, a supervised clustering technique, to classify individuals as members of the previously identified latent clusters using proteomics values47. A total of 713 CSF proteins passed QC and were utilized as features in the supervised clustering model. These values were scaled and centered. For this analysis all available proteome values were compared to clusters derived from longitudinal trajectories.

Most (N = 105) participants had two sequenced proteomes. For the application of Pelora, we set aside 20% of the participant IDs as a hold out set. In stratifying our data into training, testing, and hold out sets, we did not allow participants to be split across training/testing/hold out sets, as this could lead to target leakage. We then applied Pelora for classification on the training dataset. We trained the algorithm using 1 – 10 protein clusters, then evaluated performance on the testing dataset. We selected the optimal number of protein clusters based on the AUC performance on the testing dataset. After the number of clusters was chosen, we applied Pelora to the hold out dataset. We calculated bootstrapped confidence intervals for the AUC, and these values are reported in the results.

In the supplement, we present a comparison between the models presented in the main text and a simple baseline model. We performed binomial lasso regression (Pelora is built on similar L1 norm concepts) using only sex and age as covariates, attempting to classify individual membership in the previously identified latent clusters. We also applied Pelora while including age and sex as covariates.

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 also present the spearman correlations between all proteins identified as useful for classification by Pelora in the supplement.49

**Data Availability**

Analysis code has been published at github.com/jwisch/ ProteomicClusters. All analyses, excluding proteome pathway analysis, was conducted using R 50. Proteomic data is available at NIAGADS: https://www.niagads.org/datasets/ng00102 and in the Proteomics Browser: <http://ngi.pub:3838/ONTIME_Proteomics/>. Imaging data is available via OASIS <https://www.oasis-brains.org/>. Imaging data, as well as biomarker data, is also available via request to the Knight ADRC https://knightadrc.wustl.edu/data-request-form/.

**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 pTau181 as a function of CSF Aβ42 36. 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 pTau181 throughout enrollment. Even for visits where lower CSF Aβ42 was measured, these individuals had relatively low CSF pTau. The second cluster of individuals was referred to as the “Intermediate AD Biomarkers” group 51 (N = 27). It was comprised of individuals who had higher mean CSF pTau181 and corresponding lower mean CSF Aβ42 but did not have mean CSF pTau181 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 mean CSF pTau181 and low mean CSF Aβ42 that were consistent with AD positivity 45. These individuals were at the highest risk for developing AD 52.

In order to understand the different relationships between CSF Aβ42 and CSF pTau181 for the three groups, we used a 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 pTau181 and declining CSF Aβ42 (AD Biomarker Negative latent cluster), one approximately zero slope, indicating no relationship between CSF pTau181 and CSF Aβ42 (Intermediate AD Biomarkers latent cluster), and one positive slope, indicating an increasing relationship between CSF pTau181 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 volume53, we evaluated CSF pTau181 as a function of CSF Aβ42/Aβ40 across 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 pTau181 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 pTau181.

Although Figure 1b seems to show a continuum of pathology, there are important demographic differences across the latent clusters (Table 1). Differences were observed with regards to age at study enrollment. In particular, 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 pTau181 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 4 allele was most frequently found in the AD Biomarker Positive cohort and the APOE 2 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 44 in order to evaluate time to pathology development. Age to amyloid positivity, age to tau positivity, and age to CDR conversion were stratified for the three latent clusters. In addition to survival analysis, we applied GAMMs to biomarkers of AT(N) pathology.

Thirteen of 93 participants converted to amyloid positive during study enrollment (a further 15 participants were amyloid positive prior to enrollment). Survival analysis showed a clear separation in age at amyloid positivity (CSF Aβ42/Aβ40 < 0.0673) 45) across clusters (*Figure 2a*). The majority of AD Biomarker Positive participants were amyloid positive before age 65 years old. GAMM modeling shows that the AD Biomarker Positive participants were, on average, CSF Aβ42/Aβ40 positive at time of enrollment (*Figure 3*) 38,51. AD Biomarker Positive individuals were also CSF pTau positive (CSF pTau181 > 42.5 ug/mL) very early (Age 48.0, 95% Confidence Interval (CI): 43.6, 53.3) 45. Twenty of 89 participants converted to CSF pTau positive during study enrollment, with 19 additional participants evaluated as CSF pTau positive prior to enrollment. Similar results were also seen when 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 45 (Age 63.1, 95% CI: 46.3, 68.4) occurred prior to amyloid positivity as defined by CSF Aβ42/Aβ40 (Age 70.1, 95% CI: 72.3, 75.4) for the Intermediate AD Biomarkers cohort (*Figure 2b*). This ordering does not align with the proposed AT(N) hypothesis. For both the Intermediate and AD Biomarker Positive cohorts, clinical symptoms occurred after pathology developed, consistent with the AT(N) hypothesis 54. 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 cortical thickness or WMH volumeacross the three clusters of participants; however, AD Biomarker Positive and Intermediate AD Biomarkers cohorts have significantly higher NfL levels than the AD Biomarker Negative cohort (AD Biomarker Positive NfL levels are significantly higher than the AD Biomarker Negative cohort for ages 45.4 – 86.0; AD Intermediate Biomarkers NfL levels are significantly higher than the AD Biomarker Negative cohort for ages 52.4 – 81.6).

**Predicting Identified Clusters via the Proteome**

We attempted to classify individuals as members of one of the three latent clusters using only proteomics data. No additional covariates (e.g. age, sex, APOE genotype) were included in the initial analysis. For each pair of latent clusters, we applied Pelora 55. We repeated the analysis after including clinical values of both age and sex in accordance with the methods outlined 55. The results relying on the proteome alone outperformed the models that included age and sex, although not significantly (Supplemental Table 1).

The Pelora algorithm55, 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. We were able to distinguish between both AD Biomarker Positive and AD Biomarker Negative (AUC: 0.952, (0.850, 1.00) as well as Intermediate AD Biomarkers and AD Biomarker Negative cohorts (AUC: 0.857 (0.729, 0.986), but not AD Biomarker Positive vs. Intermediate AD Biomarkers (AUC: 0.525 (0.146, 0.903). The expression of each of the highlighted proteins in a useful classification, separated by group, is shown in Supplemental Figures 2 – 5. A Heatmap was generated that showed correlation between identified proteins (Supplemental Figure 9). 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. Proteins were grouped as primarily associated with the blood brain barrier/vascular function (BBB),immune function , liver function, inflammation, or neurodegeneration. The relationship between each protein, its’ function, and which classification(s) it applied to is shown in *Figure 4*Individual predictive power scores for proteins are shown in Supplemental Figure 6.

The Intermediate AD Biomarkers cohort was classified distinctly from the AD Biomarker Negative group (Figure 4A). The most used protein in this classification was 14-3-3 protein zeta/delta (AUCindividual = 0.909, PPS = 0.292). The most heavily weighted protein, by binomial log-likelihood was the 14-3-3 protein family (AUCindividual = 0.902, PPS = 0.251). SPARC-related modular calcium-binding protein 1 (AUCindividual = 0.900, PPS = 0.199), Neuronal Growth Regulator 1 (AUCindividual = 0.590, PPS = 0.199), Interleukin 20 receptor subunit alpha (AUCindividual = 0.726, PPS = 0.199), and Antileukoproteinase (AUCindividual = 0.694, PPS = 0.218) were also highly weighted for classification utility.

We were not able to successfully distinguish between the AD Biomarker Positive and Intermediate AD Biomarkers cohorts (Figure 4B). The AD Biomarker Positive group was successfully differentiated from the AD Biomarker Negative group (Figure 4C). The 14-3-3 protein family (AUCindividual = 0.829, PPS = 0.251) was the most important protein for classification, followed by Thrombospondin 4 (AUCindividual = 0.708, PPS = 0.228), Immunoglobulin G (AUCindividual = 0.667, PPS = 0.199), Growth Differentiation Factor 15 (AUCindividual = 0.614, PPS = 0.246), Endostatin (AUCindividual = 0.662, PPS = 0.199), and SPARC-related modular calcium-binding protein 1 (AUCindividual = 0.900, PPS = 0.199).

# Discussion

This study used a novel application of growth mixture modeling to identify unique pathological patterns of participants as a function of CSF pTau181 / A42. Previously this unsupervised clustering technique has been applied to markers of cognition and neurodegeneration4,37–39. Our objective in doing this was to apply a data driven method to understand heterogeneity in the longitudinal development of AD pathology.

After performing this classification, we examined the demographic characteristics for the three identified clusters. The AD Biomarker Positive group had the greatest proportion of APOE 4+ individuals (53.3%), followed by the AD Intermediate group (44.4%). This is consistent with studies that have previously identified that the APOE 4 allele is a risk factor for developing AD 65,66 with amyloid deposition occurring at an earlier age 67. 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 evaluatingf pTau181 as a function of CSF A42/A40, individuals appeared to move along a single continuum rather than three distinct paths (Figure 1B). 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 A42/A40 with rapidly increasing CSF pTau181 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 literature2,3,54. Compared to the other groups, the AD Biomarker Positive group had significantly lower CSF A42/A40 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 years old, consistent with previous work 68. Individuals in the Intermediate AD Biomarkers cohort showed a steady decline in CSF A42/A40 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 years old (Mean Age 71.0, 95% CI: (45.4, 73.4)).

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 pTau181 positive by age 48 (95% CI: (43.6, 53.3). Of note, tau positivity as measured by PET tau (Mean Age 59.4, 95% CI: (55.2, 70.7) occurred statistically significantly after CSF pTau181 positivity, which is consistent with previously published literature 5. While the AD Biomarker Positive group followed the proposed AT(N) curves 54, the Intermediate cohort developed CSF tau positivity significantly earlier (Mean Age 63.1, 95% CI: (46.3, 68.4)) than CSF amyloid positivity (Mean Age 72.8, 95% CI: (70.1, 75.4)). This development of tau before amyloid is consistent with previous studies suggesting that in some individuals, tau positivity can occur prior to amyloid positivity69.

Throughout enrollment, there were no significant differences in white matter hyperintensity or cortical thickness 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. However, the AD Biomarker Negative cohort did have significantly lower NfL levels than the other two cohorts, suggesting that some early neurodegenerative changes could be detected. In the analysis of the proteome we further see the importance of nonspecific neurodegenerative markers (namely 14-3-3 and SPARCL1) in distinguishing the AD Biomarker Positive and Intermediate AD Biomarkers cohorts from the AD Biomarker negative cohort. At the conclusion of this study, one participant had attained a CDR = 2 and one participant with a 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 70–72. To date, we are only aware of one study that has sought to identify proteomic subtypes associated with CSF-based measures of amyloid and tau11,12; this study did not link their results to corresponding neuroimaging data. As noted by Tijms et al, observed differences in the proteome represent a starting point rather than a conclusive identification of discrepancies in early preclinical AD progression11. Future work will require quantitative targeted measurements of specific proteins.

Two of the most useful proteins (14-3-3 and SPARCL1) in this classification effort have previously been identified as nonspecific markers of neurodegeneration 56,57,73. These proteins are not specific to AD and are reflective of general neuropathology.

It was somewhat surprising to us that markers of inflammation were more prevalent in classifying AD Biomarker Positive vs. AD Biomarker Negative individuals, given that the Intermediate AD Biomarkers cohort was the oldest group. We would have expected the oldest group to display the highest levels of inflammation; however, it was the group that demonstrated the classic A-T(N) progression of disease pathology that had a distinct pattern of inflammation. This supports existing evidence that links AD pathology development with inflammation74, and suggests that inflammatory response is more closely associated with AD pathology rather than aging.

The cohort that had a T-first progression (the Intermediate AD Biomarkers group) was instead distinguished by its markers of neurodegeneration and BBB/Vascular function. Tau proteins are thought to play a key role in the regulation of the neurovascular unit which is comprised of the BBB as well as circulating immune and peripheral tissue cells.75 While we do not know why this cohort develops pTau positivity out of sequence with the traditional AT(N) framework (or even the proposed updated APT(N) framework76), the prevalence of neurovascular unit related proteins used for classification provides additional evidence linking tau development to the neurovascular system. When considering potential drug targets for AD, amyloid reducing agents may demonstrate less efficacy in this group. Instead, tau targets and BBB/Vascular related drugs should be considered.

**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 used multiple datapoints from the same individuals. Although we were careful to segregate individuals into only training, testing, or holdout datasets, this potentially inflates individual-specific random effects. We were also were unable to compare results to an external cohort for validation. The requirement of multiple LPs with CSF A42 and CSF pTau181 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. At this stage of analysis, our results are largely descriptive and hypothesis generating rather than conclusively establishing links between specific proteome expression and amyloid and tau pathology. This is an important first step, but does not identify optimal drug targets or definitively establish the mechanisms associated with AT(N) pathological development.

**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. As previously suggested in other works, specific AD subtypes may be responsive to different types of therapies12. 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 years old, suggesting very early intervention is necessary for this group. The Intermediate AD Biomarkers 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, immune function, and neuroinflammation.

# Acknowledgements

We acknowledge Dr. Duber Gomez-Fonseca for his assistance in searching for potential validation datasets for this project.

# 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.

# Competing interests

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.

# Supplementary material

Supplementary material is available at *Brain* online.

# References

1. Gaugler J, James B, Johnson T, Marin A, Weuve J. 2020 Alzheimer’s disease facts and figures. *Alzheimer’s and Dementia*. 2020;16(3):391-460. doi:10.1002/alz.12068

2. Aisen PS, Cummings J, Jack CR, et al. On the path to 2025: Understanding the Alzheimer’s disease continuum. *Alzheimers Res Ther*. 2017;9(1). doi:10.1186/s13195-017-0283-5

3. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer’s disease. *Alzheimer’s & Dementia*. 2018;14(4):535-562. doi:10.1016/J.JALZ.2018.02.018

4. Qiu P, Zeng M, Kuang W, et al. Heterogeneity in the dynamic change of cognitive function among older Chinese people: A growth mixture model. *Int J Geriatr Psychiatry*. 2020;35(10):1123-1133. doi:10.1002/gps.5334

5. Boerwinkle AH, Wisch JK, Chen CD, et al. Temporal Correlation of CSF and Neuroimaging in the Amyloid-Tau-Neurodegeneration Model of Alzheimer Disease. *Neurology*. Published online July 6, 2021. doi:10.1212/wnl.0000000000012123

6. Schindler SE, Gray JD, Gordon BA, et al. Cerebrospinal fluid biomarkers measured by Elecsys assays compared to amyloid imaging. *Alzheimer’s and Dementia*. 2018;14(11):1460-1469. doi:10.1016/j.jalz.2018.01.013

7. Grothe MJ, Moscoso A, Ashton NJ, et al. Associations of Fully Automated CSF and Novel Plasma Biomarkers With Alzheimer Disease Neuropathology at Autopsy. *Neurology*. Published online July 15, 2021:10.1212/WNL.0000000000012513. doi:10.1212/WNL.0000000000012513

8. Toombs J, Zetterberg H. Untangling the tau microtubule-binding region. *Brain*. 2021;144(2):359-362. doi:10.1093/brain/awaa468

9. Hansson O, Zetterberg H, Buchhave P, Londos E, Blennow K, Minthon L. Association between CSF biomarkers and incipient Alzheimer’s disease in patients with mild cognitive impairment: a follow-up study. *Lancet Neurol*. 2006;5(3):228-234. doi:10.1016/S1474-4422(06)70355-6

10. Moghekar A, Li S, Lu Y, et al. CSF biomarker changes precede symptom onset of mild cognitive impairment. *Neurology*. 2013;81(20):1753-1758. doi:10.1212/01.wnl.0000435558.98447.17

11. Tijms B, Gobom J, Teunissen C, et al. CSF Proteomic Alzheimer’s Disease-Predictive Subtypes in Cognitively Intact Amyloid Negative Individuals. *Proteomes*. 2021;9(3):36. doi:10.3390/proteomes9030036

12. Tijms BM, Gobom J, Reus L, et al. Pathophysiological subtypes of Alzheimer’s disease based on cerebrospinal fluid proteomics. *Brain*. 2020;143(12):3776-3792. doi:10.1093/brain/awaa325

13. McQuail JA, Dunn AR, Stern Y, et al. Cognitive Reserve in Model Systems for Mechanistic Discovery: The Importance of Longitudinal Studies. *Front Aging Neurosci*. 2021;12. doi:10.3389/fnagi.2020.607685

14. 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. *Journal of Alzheimer’s Disease*. 2017;59(4):1359-1379. doi:10.3233/JAD-170261

15. Tarawneh R. Biomarkers: Our Path Towards a Cure for Alzheimer Disease. *Biomark Insights*. 2020;15. doi:10.1177/1177271920976367

16. Graves PR, Haystead TAJ. Molecular Biologist’s Guide to Proteomics. *Microbiology and Molecular Biology Reviews*. 2002;66(1):39-63. doi:10.1128/mmbr.66.1.39-63.2002

17. Morris JC, Schindler SE, McCue LM, et al. Assessment of Racial Disparities in Biomarkers for Alzheimer Disease. *JAMA Neurol*. 2019;76(3):264-273. doi:10.1001/jamaneurol.2018.4249

18. Morris JC. Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type. *International Psychogeriatric Association*. 1997;9(1):173-176. doi:10.1017/S1041610297004870

19. 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. Neuron. doi:10.1016/j.neuron.2013.02.026

20. Fagan AM, Mintun MA, Mach RH, et al. Inverse relation between in vivo amyloid imaging load and cerebrospinal fluid Abeta;42 in humans. *Ann Neurol*. 2006;59(3):512-519. doi:10.1002/ana.20730

21. Wang L, Benzinger TL, Hassenstab J, et al. Spatially distinct atrophy is linked to β-amyloid and tau in preclinical Alzheimer disease. *Neurology*. 2015;84(12):1254-1260. doi:10.1212/WNL.0000000000001401

22. Dickerson BC, Bakkour A, Salat DH, 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;19(3):497-510. doi:10.1093/cercor/bhn113

23. Dincer A, Gordon BA, Hari-Raj A, et al. Comparing cortical signatures of atrophy between late-onset and autosomal dominant Alzheimer disease. *Neuroimage Clin*. 2020;28. doi:10.1016/j.nicl.2020.102491

24. Ithapu V, Singh V, Lindner C, et al. Extracting and summarizing white matter hyperintensities using supervised segmentation methods in Alzheimer’s disease risk and aging studies. *Hum Brain Mapp*. 2014;35(8):4219-4235. doi:10.1002/hbm.22472

25. Mishra S, Gordon BA, Su Y, et al. AV-1451 PET imaging of tau pathology in preclinical Alzheimer disease: Defining a summary measure. *Neuroimage*. 2017;161:171-178. doi:10.1016/j.neuroimage.2017.07.050

26. Su Y, D’Angelo GM, Vlassenko AG, et al. Quantitative analysis of PiB-PET with FreeSurfer ROIs. *PLoS One*. 2013;8(11). doi:10.1371/journal.pone.0073377

27. Su Y, Blazey TM, Snyder AZ, et al. Partial volume correction in quantitative amyloid imaging. *Neuroimage*. 2015;107:55-64. doi:10.1016/j.neuroimage.2014.11.058

28. Joshi A, Koeppe RA, Fessler JA. Reducing between scanner differences in multi-center PET studies. *Neuroimage*. 2009;46(1):154-159. doi:10.1016/j.neuroimage.2009.01.057

29. 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. *J Comput Assist Tomogr*. 1995;19(2):289-296. doi:10.1097/00004728-199503000-00022

30. Eisenstein SA, Koller JM, Piccirillo M, et al. Characterization of extrastriatal D2 in vivo specific binding of [ 18 F](N-methyl)benperidol using PET. *Synapse*. 2012;66(9):770-780. doi:10.1002/syn.21566

31. Su Y, Blazey TM, Owen CJ, et al. Quantitative Amyloid imaging in autosomal Dominant Alzheimer’s disease: Results from the DIAN study group. *PLoS One*. 2016;11(3). doi:10.1371/journal.pone.0152082

32. Rousset OG, Ma Y, Evans AC. Correction for partial volume effects in PET: Principle and validation. *Journal of Nuclear Medicine*. 1998;39(5):904-911.

33. Gordon BA, Friedrichsen K, Brier M, et al. The relationship between cerebrospinal fluid markers of Alzheimer pathology and positron emission tomography tau imaging. *Brain*. 2016;139(8):2249-2260. doi:10.1093/brain/aww139

34. Cruchaga C, Ju Sung Y, Yang C, et al. Multi-tissue proteomics identifies molecular signatures for sporadic and genetically defined Alzheimer disease cases. *Preprint*. doi:10.21203/rs.3.rs-923492/v1

35. Yang C, Farias FHG, Ibanez L, et al. Genomic atlas of the proteome from brain, CSF and plasma prioritizes proteins implicated in neurological disorders. *Nat Neurosci*. 2021;24(9):1302-1312. doi:10.1038/s41593-021-00886-6

36. Ram N, Grimm KJ. Methods and Measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *Int J Behav Dev*. 2009;33(6):565-576. doi:10.1177/0165025409343765

37. 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. *Int J Geriatr Psychiatry*. 2018;33(8):1057-1064. doi:10.1002/gps.4893

38. Small BJ, Bäckman L. Longitudinal trajectories of cognitive change in preclinical Alzheimer’s disease: A growth mixture modeling analysis. *Cortex*. 2007;43(7):826-834. doi:10.1016/S0010-9452(08)70682-8

39. 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. doi:10.1101/2021.03.10.434854

40. Wallin AK, Blennow K, Zetterberg H, Londos E, Minthon L, Hansson O. CSF biomarkers predict a more malignant outcome in Alzheimer disease. *Neurology*. 2010;74(19):1531-1537. doi:10.1212/WNL.0b013e3181dd4dd8

41. van der Vlies AE, Verwey NA, Bouwman FH, et al. CSF biomarkers in relationship to cognitive profiles in Alzheimer disease. *Neurology*. 2009;72(12):1056-1061. doi:10.1212/01.wnl.0000345014.48839.71

42. Iqbal K, Flory M, Khatoon S, et al. Subgroups of Alzheimer’s disease based on cerebrospinal fluid molecular markers. *Ann Neurol*. 2005;58(5):748-757. doi:10.1002/ana.20639

43. Yoshida K, Bohn J. Package “tableone.” *R*. Published online 2019. Accessed October 28, 2019. ftp://cygwin.uib.no/pub/cran/web/packages/tableone/tableone.pdf

44. Therneau TM. A Package for Survival Analysis in R. Published online 2022.

45. Volluz KE, Schindler SE, Henson RL, et al. Correspondence of CSF biomarkers measured by Lumipulse assays with amyloid PET. In: *2021 Alzheimer’s Association International Conference*. ; 2021.

46. 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). doi:10.1016/j.neuroimage.2020.117596

47. Dettling M, Bühlmann P. *Supervised Clustering of Genes*.; 2002. http://genomebiology.com/2002/3/12/research/0069.1

48. Hastie T, Qian J, Tay K. *An Introduction to Glmnet*.; 2021. https://cran.us.r-project.org

49. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1). doi:10.1038/s41467-017-01261-5

50. R Core Development Team. A language and environment for statistical computing. 2013;1. Accessed October 28, 2019. ftp://ftp.uvigo.es/CRAN/web/packages/dplR/vignettes/intro-dplR.pdf

51. Insel PS, Ossenkoppele R, Gessert D, et al. Time to Amyloid Positivity and Preclinical Changes in Brain Metabolism, Atrophy, and Cognition: Evidence for Emerging Amyloid Pathology in Alzheimer’s Disease. *Front Neurosci*. 2017;11:281. doi:10.3389/fnins.2017.00281

52. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer’s disease. *Alzheimer’s and Dementia*. 2018;14(4):535-562. doi:10.1016/j.jalz.2018.02.018

53. Graff-Radford J, Jones DT, Wiste HJ, et al. Cerebrospinal Fluid Dynamics and Discordant Amyloid Biomarkers. *Neurobiol Aging*. Published online November 2021. doi:10.1016/j.neurobiolaging.2021.10.017

54. Jack CR, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer’s pathological cascade. *Lancet Neurol*. 2010;9(1):119-128. doi:10.1016/S1474-4422(09)70299-6

55. Dettling M, Bühlmann P. *Supervised Clustering of Genes*.; 2002. http://genomebiology.com/2002/3/12/research/0069.1

56. Aitken A. 14-3-3 proteins: A historic overview. *Semin Cancer Biol*. 2006;16(3):162-172. doi:10.1016/j.semcancer.2006.03.005

57. 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. doi:10.3233/JAD-181032

58. Xu L, Nirwane A, Yao Y. Basement membrane and blood-brain barrier. *Stroke Vasc Neurol*. 2019;4(2):78-82. doi:10.1136/svn-2018-000198

59. Dehouck B, Fenart L, Dehouck MP, Pierce A, Torpier G, Cecchelli R. *A New Function for the LDL Receptor: Transcytosis of LDL across the Blood-Brain Barrier*. Vol 138.; 1997. http://www.jcb.org

60. Trejo JL, Carro E, Garcia-Galloway E, Torres-Aleman I. Role of insulin-like growth factor I signaling in neurodegenerative diseases. *J Mol Med*. 2004;82(3):156-162. doi:10.1007/s00109-003-0499-7

61. Song IU, Kim Y do, Chung SW, Cho HJ. Association between serum haptoglobin and the pathogenesis of alzheimer’s disease. *Internal Medicine*. 2015;54(5):453-457. doi:10.2169/internalmedicine.54.2876

62. Jung SM, Lee KB, Lee JW, 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. *Neurosci Lett*. 2008;436(2):153-157. doi:10.1016/j.neulet.2008.03.010

63. Bonifati DM, Kishore U. Role of complement in neurodegeneration and neuroinflammation. *Mol Immunol*. 2007;44(5):999-1010. doi:10.1016/j.molimm.2006.03.007

64. Zahra Paylakhi S, Ozgoli S, Paylakhi S. Identification of Alzheimer disease-relevant genes using a novel hybrid method. *Prog Biol Sci*. 2016;6:37-46. doi:10.22059/PBS.2016.59006

65. Tang MX, Stern Y, Marder K, et al. The APOE-ε4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. *J Am Med Assoc*. 1998;279(10):751-755. doi:10.1001/jama.279.10.751

66. Li Z, Shue F, Zhao N, Shinohara M, Bu G. APOE2: protective mechanism and therapeutic implications for Alzheimer’s disease. *Mol Neurodegener*. 2020;15(1). doi:10.1186/s13024-020-00413-4

67. Morris JC, Roe CM, Xiong C, et al. APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. *Ann Neurol*. 2010;67(1):122-131. doi:10.1002/ana.21843

68. Sutphen CL, Jasielec MS, Shah AR, et al. Longitudinal Cerebrospinal Fluid Biomarker Changes in Preclinical Alzheimer Disease During Middle Age. *JAMA Neurol*. 2015;72(9):1029. doi:10.1001/jamaneurol.2015.1285

69. Insel PS, Donohue MC, Berron D, Hansson O, Mattsson-Carlgren N. Time between milestone events in the Alzheimer’s disease amyloid cascade. *Neuroimage*. 2021;227. doi:10.1016/j.neuroimage.2020.117676

70. Mendonça CF, Kuras M, Nogueira FCS, et al. Proteomic signatures of brain regions affected by tau pathology in early and late stages of Alzheimer’s disease. *Neurobiol Dis*. 2019;130. doi:10.1016/j.nbd.2019.104509

71. Arnold SE, Louneva N, Cao K, et al. Cellular, synaptic, and biochemical features of resilient cognition in Alzheimer’s disease. *Neurobiol Aging*. 2013;34(1):157-168. doi:10.1016/j.neurobiolaging.2012.03.004

72. Yu L, Petyuk VA, Gaiteri C, et al. Targeted brain proteomics uncover multiple pathways to Alzheimer’s dementia. *Ann Neurol*. 2018;84(1):78-88. doi:10.1002/ana.25266

73. Gu Q, Cuevas E, Raymick J, Kanungo J, Sarkar S. Downregulation of 14-3-3 Proteins in Alzheimer’s Disease. *Mol Neurobiol*. 2020;57(1):32-40. doi:10.1007/s12035-019-01754-y

74. Sarkar D, Fisher PB. Molecular mechanisms of aging-associated inflammation. *Cancer Lett*. 2006;236(1):13-23. doi:10.1016/j.canlet.2005.04.009

75. Michalicova A, Majerova P, Kovac A. Tau Protein and Its Role in Blood–Brain Barrier Dysfunction. *Front Mol Neurosci*. 2020;13. doi:10.3389/fnmol.2020.570045

76. Groot C, Smith R, Stomrud E, et al. Phospho-tau with subthreshold tau-PET predicts increased tau. *Brain*. Published online 2022. doi:10.1093/brain/awac329/6695020

77. Herrmann J, Lerman LO, Lerman A. Ubiquitin and ubiquitin-like proteins in protein regulation. *Circ Res*. 2007;100(9):1276-1291. doi:10.1161/01.RES.0000264500.11888.f0

78. Kim YN, Kim DH. Decreased serum angiogenin level in Alzheimer’s disease. *Prog Neuropsychopharmacol Biol Psychiatry*. 2012;38(2):116-120. doi:10.1016/j.pnpbp.2012.02.010

79. 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. *J Neurochem*. 2017;140(5). doi:10.1111/jnc.13789

80. Jung SM, Lee KB, Lee JW, 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. *Neurosci Lett*. 2008;436(2):153-157. doi:10.1016/j.neulet.2008.03.010

81. Sweeney MD, Sagare AP, Zlokovic B v. Blood-brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nat Rev Neurol*. 2018;14(3):133-150. doi:10.1038/nrneurol.2017.188

82. 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;75(1):1-14. doi:10.3233/jad-190848

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) 4 status at enrollment. By the conclusion of the study, participants also differed by clinical dementia rating (CDR).

Figure 1. (A) Unsupervised clustering identified three distinct trajectories for longitudinal changes in cerebrospinal fluid (CSF) pTau181 as a function of CSF A42. AD Biomarker Negative individuals consistently had low levels of CSF p-tau, regardless of CSF A42. Individuals within the Intermediate AD Biomarkers group had borderline positive levels of CSF pTau181 with low levels of CSF A42. Individuals within the AD Biomarker Positive group exhibited an increase in CSF pTau181 with decreases in CSF A42. (B) When CSF A42 was normalized by CSF A40, individuals appear to progress across a continuum where decreasing CSF A42/A40 was associated with increases in CSF pTau. Published cutoffs for amyloid and tau positivity are included as dashed lines6.Figure 2. Survival analysis demonstrated that the AD Biomarker Positive group quickly proceeded to both amyloid (A) and tau positivity (B). The Intermediate AD Biomarkers group lagged the AD Biomarker Positive group by about 10 years for amyloid positivity (as defined by CSF A42/A40 < 0.06753), but were not statistically significant different 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 years old. There were no differences in progression to clinically evident cognitive impairment (C).

Figure 3. The three clusters exhibit different behaviors across the Amyloid (A & B) and Tau (C & D) phases of the AT(N) progression. AD Biomarker Positive individuals have the greatest amyloid accumulation as quantified by both the CSF A42/A40 ratio and PET-PiB imaging. They also have the greatest level of tau accumulation as quantified by both CSF pTau181 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. Both the AD Biomarker Positive and Intermediate AD Biomarker groups have elevated NfL levels (G) as compared to the AD Biomarker negative group.

Figure 4. Using the log-loss criterion, we identified ten proteins that were most important for each successful classification (Intermediate vs. AD Biomarker Negative (A), AD Biomarker Positive vs. AD Biomarker Negative (C)). We then performed a literature review to classify the primary function of each protein as relating to either blood brain barrier or vascular function (BBB/Vascular), immune function, inflammation, or neurodegeneration. The Intermediate AD Biomarkers cohort was distinguished from the AD Biomarker Negative cohort using primarily markers of neurodegeneration, BBB/Vascular function and immune function (D); The AD Biomarker Positive cohort was distinguished from the AD Biomarker Negative cohort using primarily markers of neurodegeneration, BBB/vascular function, inflammation and immune function (E).

SUPPLEMENTAL MATERIALS

Supplemental Materials are provided in a separate PDF.