Molecular Neurodegeneration

RESEARCH ARTICLE Open Access

Quantitative proteomics of cerebrospinal fluid from African Americans and Caucasians reveals shared and divergent changes in Alzheimer's disease

Erica S. Modeste[1], Lingyan Ping[1], Caroline M. Watson[2], Duc M. Duong[1], Eric B. Dammer[1], Erik C. B. Johnson[2], Blaine R. Roberts[1], James J. Lah[2*], Allan I. Levey[2*] and Nicholas T. Seyfried[1,2*]

Abstract

Background Despite being twice as likely to get Alzheimer's disease (AD), African Americans have been grossly underrepresented in AD research. While emerging evidence indicates that African Americans with AD have lower cerebrospinal fluid (CSF) levels of Tau compared to Caucasians, other differences in AD CSF biomarkers have not been fully elucidated. Here, we performed unbiased proteomic profiling of CSF from African Americans and Caucasians with and without AD to identify both common and divergent AD CSF biomarkers.

Methods Multiplex tandem mass tag-based mass spectrometry (TMT-MS) quantified 1,840 proteins from 105 control and 98 AD patients of which 100 identified as Caucasian while 103 identified as African American. We used differential protein expression and co-expression approaches to assess how changes in the CSF proteome are related to race and AD. Co-expression network analysis organized the CSF proteome into 14 modules associated with brain cell-types and biological pathways. A targeted mass spectrometry method, selected reaction monitoring (SRM), with heavy labeled internal standards was used to measure a panel of CSF module proteins across a subset of African Americans and Caucasians with or without AD. A receiver operating characteristic (ROC) curve analysis assessed the performance of each protein biomarker in differentiating controls and AD by race.

Results Consistent with previous findings, the increase of Tau levels in AD was greater in Caucasians than in African Americans by both immunoassay and TMT-MS measurements. CSF modules which included 14–3-3 proteins (YWHAZ and YWHAG) demonstrated equivalent disease-related elevations in both African Americans and Caucasians with AD, whereas other modules demonstrated more profound disease changes within race. Modules enriched with proteins involved with glycolysis and neuronal/cytoskeletal proteins, including Tau, were more increased in Caucasians than in African Americans with AD. In contrast, a module enriched with synaptic proteins including VGF, SCG2, and NPTX2 was significantly lower in African Americans than Caucasians with AD. Following SRM and ROC analysis, VGF, SCG2, and NPTX2 were significantly better at classifying African Americans than Caucasians with AD.

*Correspondence:
James J. Lah
jlah@emory.edu
Allan I. Levey
alevey@emory.edu
Nicholas T. Seyfried
nseyfri@emory.edu
Full list of author information is available at the end of the article

© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions Our findings provide insight into additional protein biomarkers and pathways reflecting underlying brain pathology that are shared or differ by race.

Keywords Biomarkers, Race, Amyloid, Tau, Proteomics, CSF

Background

African Americans are almost twice as likely to have Alzheimer's disease (AD) compared to Caucasians [1–3]. Current evidence suggests that this difference in risk could be explained by a multitude of factors including genetic ancestry and disparities in health, socioeconomic and environmental conditions [4–7]. For example, genome wide association studies (GWAS) show that the *ABCA7* gene has stronger associations with AD risk in individuals with African ancestry than in individuals with European ancestry [6, 7]. *ABCA7* also has a stronger effect size in African Americans than even the strongest genetic risk factor gene for AD, the *APOE* epsilon 4 allele (*APOE* ε4) [6]. Yet, despite the *APOE* ε4 allele being more prevalent amongst African Americans, *APOE* ε4 confers a lower risk for AD compared to Caucasians [8]. Beyond genetic ancestry, chronic health conditions associated with higher risk for dementia, such as cardiovascular disease and diabetes, also disproportionally affect African Americans [4, 5]. Furthermore, societal and environmental disparities that disproportionately affect African Americans, including lower levels and quality of education, higher rates of poverty, and greater exposure to adversity and discrimination, increase risk for both chronic diseases and dementia [4, 5]. This highlights how racial differences in AD risk cannot be exclusively explained by genetics alone [4]. Currently, there is a gap in knowledge of the racial differences underlying pathophysiology related to AD. Therefore, a better understanding of these mechanisms can help move towards a more precise definition of AD across diverse racial, ethnic, and genetic backgrounds.

Amyloid-beta$_{1\text{-}42}$ (Aβ_{42}) and Tau, two core cerebrospinal fluid (CSF) protein biomarkers, comprise the major pathologic hallmarks of senile plaques and neurofibrillary tangles found in AD brain [9, 10]. Thresholds between CSF Aβ_{42} and total Tau (tTau) levels, established from predominantly Caucasian populations, are widely used as diagnostic measures for AD in the clinic [11–15]. Emerging evidence suggests that race is associated with the levels of AD biomarkers in CSF [16, 17]. African Americans with AD have lower levels of CSF tau compared to Caucasians [16, 17], which may explain why African Americans are diagnosed at a later disease stage [18]. Consistently, African Americans have reduced rates of enrollment in clinical trials which utilize tTau/Aβ_{42} CSF ratio as an enrollment criterion [19, 20]. It is thought that elevations in CSF Tau are associated with increasing neuronal damage and cognitive impairment [21], which suggests that African Americans are symptomatic with AD even with lower levels of neuronal loss. This supports a hypothesis that there are other physiological differences contributing to the increased susceptibility and cognitive impairment in African Americans that is reflected in the CSF. An unbiased analysis into the CSF proteome of African Americans could provide insight into additional biomarkers reflecting underlying brain pathology that differ by race in AD.

Proteins are the ideal markers for understanding diseases such as AD because they are most proximal to the phenotypic changes. Unbiased proteomics of human brain coupled with network analysis has emerged as a valuable tool for organizing complex unbiased proteomic data into groups or "modules" of co-expressed proteins that reflect various biological functions linked to AD [22–25]. The direct proximity of CSF to the brain presents a strong rationale to integrate the brain and CSF proteomes to identify biofluid biomarkers associated with brain pathophysiology of AD. Indeed, we recently performed an integrated human AD brain and CSF proteome analysis to reveal that approximately 70% of the CSF proteome overlapped with the brain proteome [26]. While AD brain proteomic networks from large cohorts have been examined [24, 27], AD CSF proteomic networks from large cohorts that include racially diverse subjects have been largely unexplored.

In this study, we used a tandem mass tag mass spectrometry (TMT-MS) approach to generate a deep CSF proteome, without albumin depletion, from over 200 control and AD samples, with over half of the samples derived from African Americans. Over 1,800 CSF proteins were organized into 14 modules based on co-expression network analysis. These modules were associated with brain cell-types and biological pathways known to be altered in AD brain, including synaptic, immune and metabolic processes. Notably, Tau mapped to a module with a magnitude of increase greater in Caucasians than African Americans with AD. In addition, network analysis revealed a core class of CSF markers that demonstrated equivalent disease-related elevations in both African Americans and Caucasians with AD, whereas other modules demonstrated more profound disease changes within race. Namely, a module enriched with post-synaptic neuronal proteins was significantly lower in level

in African Americans than Caucasians with AD. Using a targeted MS approach, selected reaction monitoring (SRM), we measured the neuronal proteins within this module including VGF, SCG2, and NPTX2 which were better classifiers for AD in African Americans than Caucasians. Overall, these results demonstrate the utility of a systems-based approach in the identification of CSF proteins that could serve as markers for AD across a more diverse population. Moreover, these data highlight a need for further investigations into how AD heterogeneity varies across different racial backgrounds.

Materials and methods

CSF samples

All CSF samples were collected as part of ongoing studies at Emory's Goizueta Alzheimer's Disease Research Center (ADRC) including participants in the ADRC Clinical Core, the Emory Healthy Brain Study, and the ADRC-affiliated Emory Cognitive Neurology Clinic. All participants provided informed consent under protocols approved by Emory University's Institutional Review Board. Clinical diagnosis of AD as well as classification as cognitively normal controls was based on review of clinical history, neurological examination, detailed cognitive testing, and diagnostic studies including magnetic resonance imaging and CSF AD biomarker testing. Diagnosis of AD was made by subspecialty certified Cognitive and Behavioral Neurologists with additional input from Neuropsychologists based on current National Institute on Aging and Alzheimer's Association (NIA-AA) criteria [28, 29]. A consensus clinical diagnosis of controls was made without consideration of CSF biomarkers by a panel of experts at the Emory Goizueta ADRC. Criteria for assigning diagnosis are provided in the National Alzheimer Coordination Center coding guidelines, form D1, based on clinician judgment. The basis for this judgment includes many metrics, with controls considered to have normal cognition and normal behavior after reviewing all testing including Montreal Cognitive Assessment (MoCA), Clinical Dementia Rating (CDR) score, and detailed neuropsychological testing. Hence, control participants may have MoCA scores that are lower than traditional cut points for impairment on this screening test. CSF was collected by lumbar puncture and banked according to best practice guidelines outlined by the National Institute on Aging for Alzheimer's Disease Centers (https://alz.washington.edu/BiospecimenTaskForce.html), and identical pre-analytic steps were followed in all groups. Measurements of Amyloid-beta$_{1\text{-}42}$ (Aβ_{42}), total Tau (tTau), and phosphorylated Tau$_{181}$ (pTau$_{181}$) were performed on the Roche Diagnostics Elecsys platform [30–32] using recommended protocols. In total, the cohort was comprised of 105 healthy controls and 98 AD. The racial background of each case was based upon self-identification. Of the 203 cases, 100 identified as Caucasian or White while 103 identified as African American or Black. Case metadata can be found in Supplemental Table 1 along with a summarized version in Table 1.

Protein digestion of CSF

In order to sample the CSF protein in an unbiased manner, and given that we have previously shown that immunodepletion resulted in only a marginal improvement in proteomic coverage, the CSF samples were not immunodepleted prior to digestion [26, 33]. First, 70 µL of CSF was transferred to 1 mL deep well plates for digestion with lysyl endopeptidase (LysC) and trypsin. The samples

Table 1 Cohort characteristics

	CT Cau (N=53)	CT AA (N=52)	p – value[a]	AD Cau (N=47)	AD AA (N=51)	p – value[a]
Characteristics						
Sex	33 F, 20 M	33 F, 19 M	-	29 F, 18 M	32 F, 19 M	-
Age[b]	65 ± 8	64 ± 8	0.8848	68 ± 9	68 ± 9	0.9998
MoCA[c]	26 ± 2	25 ± 3	0.7559	16 ± 6	14 ± 6	0.2055
*Aβ42[d]	1195.2 ± 262.0	1021.4 ± 301.1	**0.0021**	558.2 ± 169.9	483.4 ± 151.8	0.4026
tTau[d]	186.2 ± 61.6	158.5 ± 56.5	0.6573	423.7 ± 189.2	301.5 ± 134.8	**< 0.0001**
pTau181[d]	16.6 ± 5.6	14.1 ± 4.9	0.7787	43.3 ± 20.8	30.1 ± 14.1	**< 0.0001**
tTau/Aβ42 [d]	0.14 ± 0.03	0.14 ± 0.03	0.9994	0.78 ± 0.31	0.66 ± 0.30	0.0850

Abbreviations: *CT* Control, *AD* Alzheimer's disease, *Cau* Caucasian / White, *AA* African American / Black

[a] *p*-values were calculated using one-way ANOVA with Tukey correction, bold indicates $p < 0.05$

[b] Age in years. Values given as average ± standard deviation

[c] Most recent Montreal Cognitive Assessment (MoCA) score. Values given as average ± standard deviation

[d] Aβ42, tTau, pTau181, and tTau/Aβ42 in pg/mL. Values given as average ± standard deviation

[*] Aβ42 levels that reached saturation (1700 pg / mL) were excluded from calculations and analysis

were then reduced and alkylated with 1.4 µL of 0.5 M tris-2(-carboxyethyl)-phosphine (ThermoFisher) and 7 µL of 0.4 M chloroacetamide in a 90 °C water bath for 10 min. The water bath was then turned off and allowed to cool to room temperature along with samples for 5 min. Following this, water bath sonication was performed for 5 min. The samples were then allowed to cool again to room temperature for 5 min prior to adding urea. Then 78 µL of 8 M urea buffer (8 M urea, 10 mM Tris, 100 mM NaH_2PO_4, pH 8.5) and 3.5 µg of LysC (Wako), was added to each sample, resulting in a final urea concentration of 4 M. The samples were then mixed well, gently spun down, and incubated overnight at 25 °C for digestion with LysC. The following day, samples were diluted to 1 M urea with a blend of 468 µL of 50 mM ammonium bicarbonate [34] and 7 µg of Trypsin (ThermoFisher). The samples were subsequently incubated overnight at 25 °C for digestion with trypsin. The next day, the digested peptides were acidified to a final concentration of 1% formic acid and 0.1% trifluoroacetic acid. This was immediately followed by desalting on 30 mg HLB columns (Waters) and then eluted with 1 mL of 50% acetonitrile (ACN) as previously described [35]. To normalize protein quantification across batches [35], 100 µl was taken from all CSF samples and then combined to generate a pooled sample. This pooled sample was then divided into global internal standards (GIS) [36]. All individual samples and the pooled standards were then dried using a speed vacuum (Labconco).

Tandem mass tag labeling of CSF peptides

All CSF samples were balanced for diagnosis, race, age, and sex (in that order) across 16 batches using ARTS (automated randomization of multiple traits for study design) [37]. Using a 16-plex Tandem Mass Tag (TMTpro) kit (Thermo Fisher Scientific, A44520, Lot number: VH3111511), 13 channels of each batch were allocated to a CSF sample (127N, 127C, 128N, 128C, 129N, 129C, 130N, 130C, 131N, 131C, 132N, 132C, 133N). The remaining 3 channels were occupied with a GIS pool (126), a standard biomarker negative pool (133C), and a standard biomarker positive pool sample (134N). Information regarding the origination of these pooled samples were reported previously [38]. Supplemental Table 2 provides the sample to batch arrangement. In preparation for labeling, each CSF peptide digest was resuspended in 75 µl of 100 mM triethylammonium bicarbonate (TEAB) buffer meanwhile 5 mg of TMT reagent was dissolved into 200 µl of ACN. Once, both were in suspension, 15 µl of TMT reagent solution was subsequently added to the resuspended CSF peptide digest. After 1 h, the reaction was quenched with 4 µl of 5% hydroxylamine (Thermo Fisher Scientific, 90,115) for 15 min. Then, the peptide solutions were combined according to the batch arrangement. Finally, each TMT batch was desalted with 60 mg HLB columns (Waters) and dried via speed vacuum (Labconco).

High-pH peptide fractionation

Dried samples were re-suspended in high pH loading buffer (0.07% vol/vol NH4OH, 0.045% vol/vol FA, 2% vol/vol ACN) and loaded onto a Water's BEH column (2.1 mm × 150 mm with 1.7 µm particles). A Vanquish UPLC system (ThermoFisher Scientific) was used to carry out the fractionation. Solvent A consisted of 0.0175% (vol/vol) NH4OH, 0.01125% (vol/vol) FA, and 2% (vol/vol) ACN; solvent B consisted of 0.0175% (vol/vol) NH_4OH, 0.01125% (vol/vol) FA, and 90% (vol/vol) ACN. The sample elution was performed over a 25 min gradient with a flow rate of 0.6 mL/min with a gradient from 0 to 50% solvent B. A total of 96 individual equal volume fractions were collected across the gradient. Fractions were concatenated to 48 fractions and dried to completeness using vacuum centrifugation.

Mass spectrometry analysis and data acquisition

All samples (~1 µg for each fraction) were loaded and eluted by an Easy-nLC 1200 (Thermo Scientific) with an in-house packed 15 cm, 150 µm i.d. capillary column with 1.7 µm CSH (Water's) over a 35 min gradient. Mass spectrometry was performed with a high-field asymmetric waveform ion mobility spectrometry (FAIMS) Pro front-end equipped Orbitrap Lumos (Thermo) in positive ion mode using data-dependent acquisition with 1 s top speed cycles for each FAIMS compensative voltage. Each cycle consisted of one full MS scan followed by as many MS/MS events that could fit within the given 1 s cycle time limit. MS scans were collected at a resolution of 120,000 (410–1600 m/z range, 4 × 10^5 AGC, 50 ms maximum ion injection time, FAIMS compensative voltage of -45 and -65). Only precursors with charge states between 2 + and 5 + were selected for MS/MS. All higher energy collision-induced dissociation (HCD) MS/MS spectra were acquired at a resolution of 50,000 (0.7 m/z isolation width, 35% collision energy, 1 × 10^5 AGC target, 86 ms maximum ion time). Dynamic exclusion was set to exclude previously sequenced peaks for 30 s within a 10-ppm isolation window.

Database search and protein quantification

All raw files were analyzed using the Proteome Discoverer Suite (v.2.4.1.15, Thermo Fisher Scientific). MS/MS spectra were searched against the UniProtKB human proteome database (downloaded in 2019 with 20,338 total sequences). The Sequest HT search engine was used to search the RAW files, with search parameters specified

as follows: fully tryptic specificity, maximum of two missed cleavages, minimum peptide length of six, fixed modifications for TMTPro tags on lysine residues and peptide N-termini (+304. 304.2071 Da) and carbamidomethylation of cysteine residues (+57.02146 Da), variable modifications for oxidation of methionine residues (+15.99492 Da), serine, threonine and tyrosine phosphorylation (+79.966 Da) and deamidation of asparagine and glutamine (+0.984 Da), precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.05 Da. Percolator was used to filter peptide spectral matches and peptides to an FDR < 1%. Following spectral assignment, peptides were assembled into proteins and were further filtered based on the combined probabilities of their constituent peptides to a final FDR of 1%. Peptides were grouped into proteins following strict parsimony principles. A complete TMT reporter ion abundance-based table output of assembled protein abundances without adjustments can be found in Supplemental Table 3.

Adjustment for batch and other sources of variance

Only proteins quantified in ≥ 50% of samples were included in subsequent analysis (n = 1,840 proteins). Of the 1,840 proteins, 1,327 proteins were quantified across all samples. As previously reported [24–26], batch correction was performed using a Tunable Approach for Median Polish of Ratio, (https://github.com/edammer/TAMPOR), an iterative median polish algorithm for removing technical variance across batch. Multidimensional scaling (MDS) plots were used to visualize batch contributions to variation before and after batch correction Noticeably, prior to batch correction, cases within the same batch clustered together and batches ran consecutively tended to cluster more closely together (Supplemental Fig. 1A). Following batch correction using a median polish algorithm, the cases were no longer clustering by batch (Supplemental Fig. 1B). The data was then subjected to outlier removal using a robust principal component analysis method, *PcaGrid* [39]. A scree plot graphing the eigenvalue against the principal component (PC) number was utilized to determine the number of PCs to include in the parameters (Supplemental Fig. 1C). Briefly, the parameters used for outlier detection were as follows: the desired number of principal components = 7, method = mean absolute deviation, and criterion for computing cutoff values = 0.99 (Supplemental Fig. 1D). This resulted in the detection and removal of 15 outliers, resulting in a final n = 189 samples. Bootstrap regression was then performed to remove for covariates such as age at collection and sex. Variance partition analysis was performed to confirm appropriate regression of these traits (Supplemental Fig. 1E & F). Since the *variancePartition* package does not allow missing values, proteins with missing quantifications were temporarily imputed using the *impute.knn* function of the impute R package. The final cleaned dataset after regression and log2 transformation can be found in Supplemental Table 4.

Differential expression analysis

One-way ANOVA followed by Tukey's post hoc adjustment for multiple comparisons was performed on four groups (Control-Caucasian, Control-African American, AD-Caucasian, and AD-African American) to identify differentially expressed proteins across diagnosis and within each race. Differentially expressed proteins for comparisons of interest (i.e., Control-Caucasian vs AD-Caucasian and Control-African American vs AD-African American) were then presented as volcano plots using the *ggplot2* package in R v4.1.2. A list of all comparisons computed with corresponding adjusted p-values is provided in Supplemental Table 5.

Weighted gene co-expression network analysis (WGCNA)

As previously published [22, 24–26], the blockwiseModules function from the WGCNA package in R was utilized to derive the weighted protein co-expression network (Supplemental Table 6). Briefly, the following parameters were used: soft threshold power beta = 3, deepSplit = 4, minimum module size = 5, merge cut height = 0.07, and a signed network with partitioning about medoids. Using the *pairwise.wilcox.test* R function with Bonferroni correction, a pairwise Wilcox test was performed to calculate pairwise comparisons between each group with corrections for multiple testing.

Gene ontology (GO) and cell type enrichment analysis

To characterize co-expressed protein module biology, gene ontology (GO) annotations were retrieved from the Bader Lab's monthly updated .GMT formatted ontology lists downloaded July 5, 2022 [40]. A Fisher's exact test for enrichment was performed into each module's protein membership using an in-house script (https://github.com/edammer/GOparallel). Significant ontologies for each module are included in Supplemental Table 7. For cell type enrichment analysis, an in-house marker list was utilized as previously described [24] (Supplemental Table 8). A Fisher's exact test was performed for each module member list against the merged human cell type marker list to determine cell type enrichment. For brain-CSF module overlap, a one-sided Fisher's exact test determined significance of module overlap.

Selected reaction monitoring

Selected reaction monitoring (SRM) assays were performed on 195 of the 203 cases to determine whether a separate targeted proteomic approach could replicate

proteomic changes seen in TMT discovery proteomics. An attempt was made to include all 203 samples from discovery TMT for SRM analysis however, sample GUIDs 52524, 51520, 52055, 48617, 48769, 49537, and 45707 had low remaining sample volume and had to be removed. Sample 62762 was later removed due to irregularities in retention time shifts. Peptide selection, sample preparation, peptide quantification, and data acquisition for the SRM assay was performed as previously described [38]. Briefly, peptides were selected based on their robust detection and significant differential expression in previous CSF discovery proteomic projects for synthesis as heavy standards[25, 26]. More specifically, the peptide had to i) have one or more spectral matches, ii) be significantly differentially abundant when evaluating AD versus control cases, iii) and map to proteins that appeared in brain-based biological panels outlined in Higginbotham et al. 2020[26] that differed in AD. Ultimately, this led to approximately 200 peptides being nominated for synthesis by Thermo Fisher Scientific (Thermo PEPotec SRM Peptide Libraries; Grade 2; crude as synthesized). In addition to the 195 clinical samples from before, two pools of CSF were utilized as AD biomarker positive and AD-biomarker negative quality controls (QC) standards [38]. After all of the CSF samples were blinded and randomized, each sample (50 µL) was reduced, alkylated, denatured and then subjected to digestion as described [38]. After digestion, the heavy labeled standards, 15uL per 50 µL of CSF, were added to each digested sample. Each digested sample was then acidified, desalted and dried under vacuum. Following this, the peptide targets were quantified using TSQ Altis Triple Quadrupole mass spectrometer as previously described [38]. The resulting raw files were uploaded to Skyline-daily software (version 21.2.1.455) for peak integration and quantification by peptide ratios. Peptides were filtered by first assessing retention time reproducibility, then by matching light and heavy transitions, and finally by determining the peptide ratio precision using the coefficient of variation (CV) as described by Watson et al. [38]. The technical CV of each peptide was calculated based on the peptide area ratio for the AD-positive and AD-negative QC pools (Supplemental Table 9). CSF peptide targets with CVs $\leq$ 20% in at least one pooled standard were determined as peptides with high precision and were kept for subsequent analysis, leaving a total of 85 peptides that mapped to 58 proteins. The total area ratio for each targeted peptide in each sample made up the final data matrix. Due to the nature of SRM in that each peptide is explicitly targeted, a value for each peptide is always assigned in each sample (down to and including the limit of detection) as previously published by our group [38]. As a result, the total area ratio for each targeted peptide in each sample made up the final data matrix, leaving a matrix with no blank cells or missing values. In preparation for analysis, this matrix of peptide ratios was $\log_2$ transformed and true zero values were replaced after $\log_2$-transformation with the minimum value for that peptide minus 1 (Supplemental Table 10). Bootstrap regression was then used to regress for age and sex (Supplemental Table 11). Bicor was then utilized to calculate the correlation between SRM peptides and TMT-MS protein measurements (Supplemental Table 12). In cases where multiple peptides mapped to one protein, the most correlated peptide was kept for further analysis (Supplemental Table 13). One-way ANOVA analysis with Tukey adjustment was then utilized once again to examine pairwise interactions (Supplemental Table 14) and receiver operating characteristic (ROC) curve analysis was performed as previously described [38] (Supplemental Fig. 15).

Results

Cohort characteristics

The main objective of this study was to perform an unbiased proteomic study of CSF to define biomarkers and pathways that are similar or divergent between African Americans and Caucasians with AD. Figure 1A provides an overview of our study approach, which included the generation of balanced sets of CSF samples from African American and Caucasian individuals, matched for age and sex with roughly equal numbers of control and AD cases (Table 1). This included 53 Caucasian controls, 52 African American controls, 47 AD Caucasians, and 51 AD African Americans. The majority were female and on average the controls (64.5 years) were slightly younger than AD (68 years). Notably, there were no statistical differences between the ages of the African Americans and the Caucasians within diagnosis (control: $p=0.8848$, AD: $p=0.9998$). As expected, AD cases had lower Montreal Cognitive Assessment (MoCA) scores than controls, but there were no statistically significant differences between MoCA scores across race within controls and AD (control: $p=0.7559$, AD:$p=0.2055$). The AD cases also had lower Amyloid-beta$_{1\text{-}42}$ ($A\beta_{42}$) levels and elevated total Tau (tTau) and phosphorylated Tau_{181} ($pTau_{181}$) levels. Notably, $A\beta_{42}$ levels were significantly lower in African American controls compared to Caucasian controls ($p=0.0021$) but not different between African American AD and Caucasian AD. This may indicate potentially early changes in brain amyloid deposition or processing of APP in African American controls versus Caucasian controls. Notably, the distribution of APOE4 carriers did not differ significantly by race in the control population and so does not account for the pattern observed (Supplemental Table 1). Conversely, tTau and $pTau_{181}$ levels were significantly lower in African Americans with AD

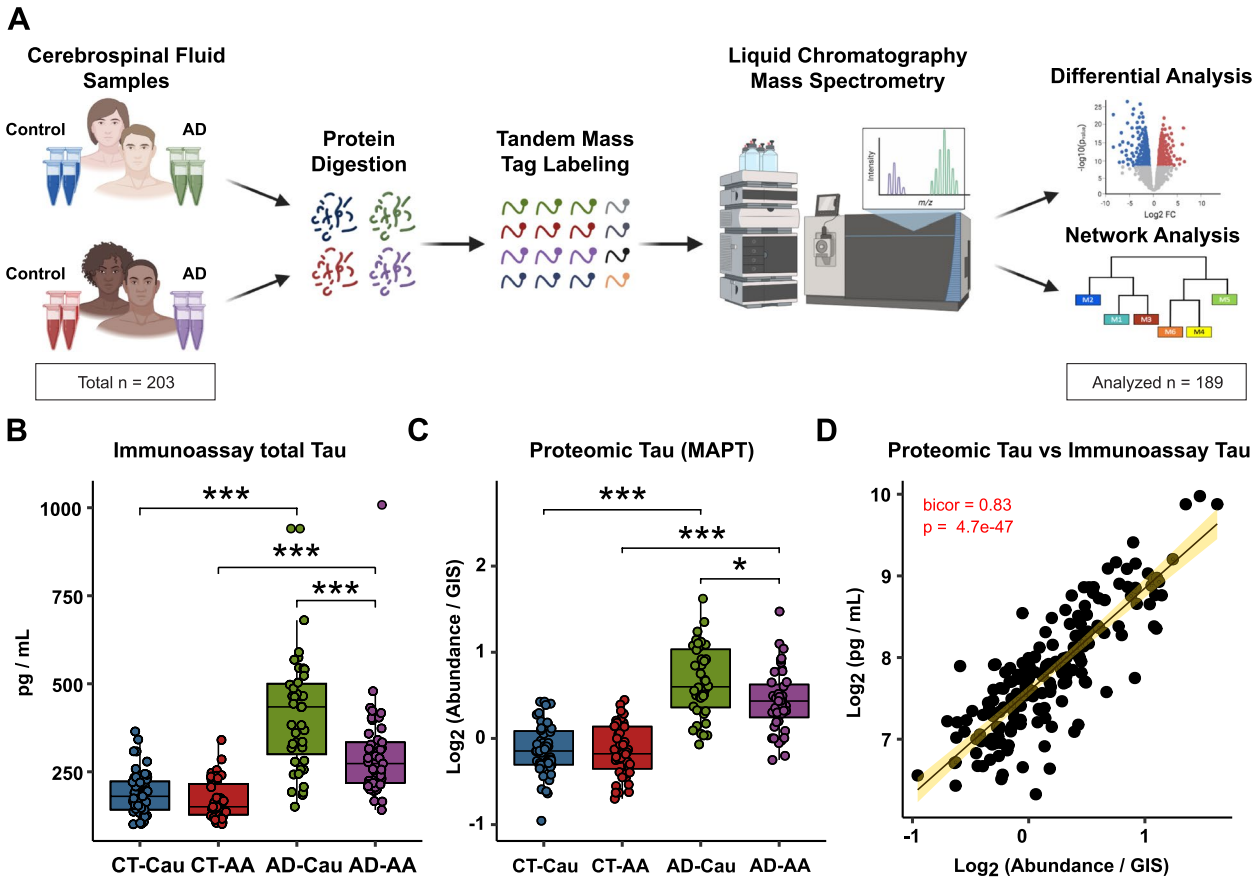

Fig. 1 Schematic of experimental workflow and correlation between proteomic Tau and total Tau immunoassay measurements. **A** Schematic of experimental workflow for quantification of cerebrospinal fluid proteome. **B** Total Tau levels as measured by Roche Elecsys Platform between control (CT) and AD cases and stratified by self-identified race: Caucasian (Cau) or African American (AA) **C** Tau levels measured by mass spectrometry. One-way ANOVA with Tukey post-hoc correction determined pairwise relationships **D** Correlation of Tau levels by TMT-MS (x-axis) to paired immunoassay total Tau levels (y-axis). Biweight midcorrelation coefficient (bicor) with associated p-value is shown. Only 179 cases were included in the linear regression analysis because of sample outlier removal and missing values in the TMT-MS

(tTau:$p < 0.0001$, $pTau_{181}$:$p < 0.0001$) but not different between African American and Caucasian controls. Data on comorbid conditions, including whether or not the person had hypertension, diabetes, dyslipidemia, or cerebrovascular disease, is presented for all cases in Supplemental Table 1. Notably, none of the comorbid conditions was statistically overrepresented in either racial group.

Mass spectrometry and immunoassay measurements of Tau strongly correlate

Following enzymatic digestion, TMT labeling, and offline fractionation, all samples were subjected to liquid chromatography tandem mass spectrometry (LC–MS/MS) (Fig. 1A). In total, TMT-MS proteomic analysis identified 34,330 peptides mapping to 2,941 protein groups across the 203 samples (16 total batches). To account for missing protein measurements across batches, we included only those proteins quantified in at least 50% of samples following outlier removal as previously described [22–26], resulting in the final quantification of 1,840 proteins. Protein abundance was adjusted for batch and age and sex were regressed. Protein levels of Tau (MAPT) by TMT-MS correlated strongly to independently measured tTau levels via immunoassay ($r = 0.83$, $p = 4.7e\text{-}47$). As expected, Tau levels were significantly elevated in both African Americans and Caucasians with AD across both platforms compared to controls (Fig. 1B and C). Consistent with the immunoassay measurements, TMT-MS Tau levels also showed significantly lower levels in African Americans with AD compared to Caucasians with AD (Fig. 1C). Thus, in this study, both platform measures of CSF Tau support a reduction in Tau levels in African Americans with AD, consistent with previous findings [16, 17].

Differential expression analysis of African American and Caucasian CSF proteome reveals unique and shared changes in AD

Differential expression analysis was performed to identify changes in the CSF proteome by race in AD (Supplemental Table 5). Consistent with previous proteomic analyses of AD CSF [25, 26, 41–43], there was a significant increase in Tau (MAPT), 14–3-3 proteins, (YWHAZ, YWHAG, and YWHAE), SMOC1, neurofilaments (NEFM and NEFL) and proteins involved in glucose metabolism in both African Americans and Caucasians with AD compared with race matched controls (Fig. 2A and B). However, Caucasians with AD exhibited a bias towards proteins that were increased in AD, where the number of differentially expressed proteins (DEPs) was nearly double ($n=183$ proteins) the number of decreased DEPs in AD ($n=74$ proteins) (Fig. 2A). In contrast, in African Americans the number of increased and decreased DEPs was more balanced (151 increased proteins vs. 162 decreased proteins). A Venn diagram illustrates the overlap of DEPs from African Americans and Caucasians with AD (Fig. 2C), with the majority of proteins ($n=168$ proteins) differentially expressed in both races. Furthermore, a correlation analysis of both shared and unique DEPs showed overall high agreement in direction of change (bicor = 0.887, p = 2.47e-136, Fig. 2D). However, there were some exceptions including SLIT1 and VSTM2A, which were significantly increased in Caucasians, but decreased in African Americans with AD. Both proteins are predominantly enriched in neuronal-cell types [44, 45]. Thus, despite the differences in the number of significant DEPs in African Americans compared to Caucasians with AD, the direction of change with disease remains largely similar across both races.

Network analysis of the CSF proteome reveals modules related to pathways and brain cell-types

Co-expression network analysis of the AD brain proteome organizes proteins into modules related to molecular pathways, organelles, and cell types impacted by AD pathology [22–25]. Moreover, integration of the human AD brain and CSF proteome revealed that approximately 70% of the CSF proteome overlapped with the brain proteome [26]. While proteomic networks in AD brain have been examined, network changes in the AD CSF proteome, including those associated with race and AD biomarkers are less well understood. Thus, we applied Weighted Gene Co-expression Network Analysis (WGCNA) to define trends in protein co-expression across 1840 CSF proteins in all individuals. These parameters identified 14 modules (M), ranked by size, ranging from the largest M1, with 370 proteins to the smallest, M14, with 16 proteins (Fig. 3A). Many of these modules were significantly enriched for brain-specific cell types (Fig. 3B) as well as established brain gene ontologies (GO), cellular functions and/or organelles (Fig. 3C, Supplemental Table 7). The three largest modules were associated with categories of "Postsynaptic Membrane" (M1), "Complement Activation" (M2), and "Extracellular Matrix" (M3) whereas M5 represented "Lysosome / Catabolism" and M6 "Gluconeogenesis". Other modules included those with GO terms linked to "Cell Morphogenesis" (M4), "Cell Redox / Proteasome" (M7), "Protein Polyubiquitination" (M8), "Angiogenesis / Cell Migration" (M9), "Synapse Assembly" (M10), Myofibril Assembly (M11), "Actin Cytoskeleton" (M12), "Kinase Signaling / Activity" (M13), and "Carbohydrate Metabolism" (M14).

Protein-based network analysis in AD brain tissue has shown that the cellular composition represents a major source of biological variance and that many of the network modules are enriched in proteins that are expressed by specific brain cell types [24, 25]. To determine if a similar relationship exists with protein-based networks in CSF, we evaluated the overlap of proteins in each module with brain cell-type specific makers (Fig. 3B, Supplemental Table 8), generated previously from cultured or acute isolated neurons, oligodendrocytes, astrocytes, endothelial, and microglia from brain [44, 45]. The largest module, M1, was enriched with neuron/synaptic proteins (i.e., NPTX1, NPTXR, SCG2, VGF, NRN1, and L1CAM) and to a lesser degree oligodendrocyte proteins (i.e., IGSF8, VCAN, APLP1). Neuronal loss or the active secretion of neuronal proteins into the extracellular space could account for the presence of neuronal proteins in the CSF. The M4 module was also enriched for neuronal protein markers including RTN4R1, LINGO2, OLFM1, and PLXNA2, associated with "Nervous Systems and Cell Morphogenesis". Modules most enriched with microglia markers were M2 (i.e., C2, C3, C1RL, C1QA, C1QB, C1QC, LCP1, etc.) and M5 (i.e., HEXB, CTSZ, HEXA, CTSA, CTSB) consistent with a role in complement activation and lysosome function, respectively. Finally, endothelial markers were mainly overrepresented in modules M3 (i.e., NID2, ECM2, NID1, LTBP4, LAMA5, LAMC1), M9 (IGFBP7, F5, SDCBP, BGN) and M11 (FLNA, ANXA5, S100A11, MYL6) consistent with roles in extracellular matrix, angiogenesis and myofibril assembly, respectively. Thus, as seen in the network analysis of bulk proteome from human brain [24, 25], certain modules of co-expressed proteins in CSF were enriched with markers of specific brain cell-types. To further support this observation, we assessed the protein overlap between modules in CSF and modules from a recent large-scale consensus TMT-MS proteomic network of bulk human AD brain tissue [24]. (Supplemental Fig. 2). Except for M9, M10 and M14, which had

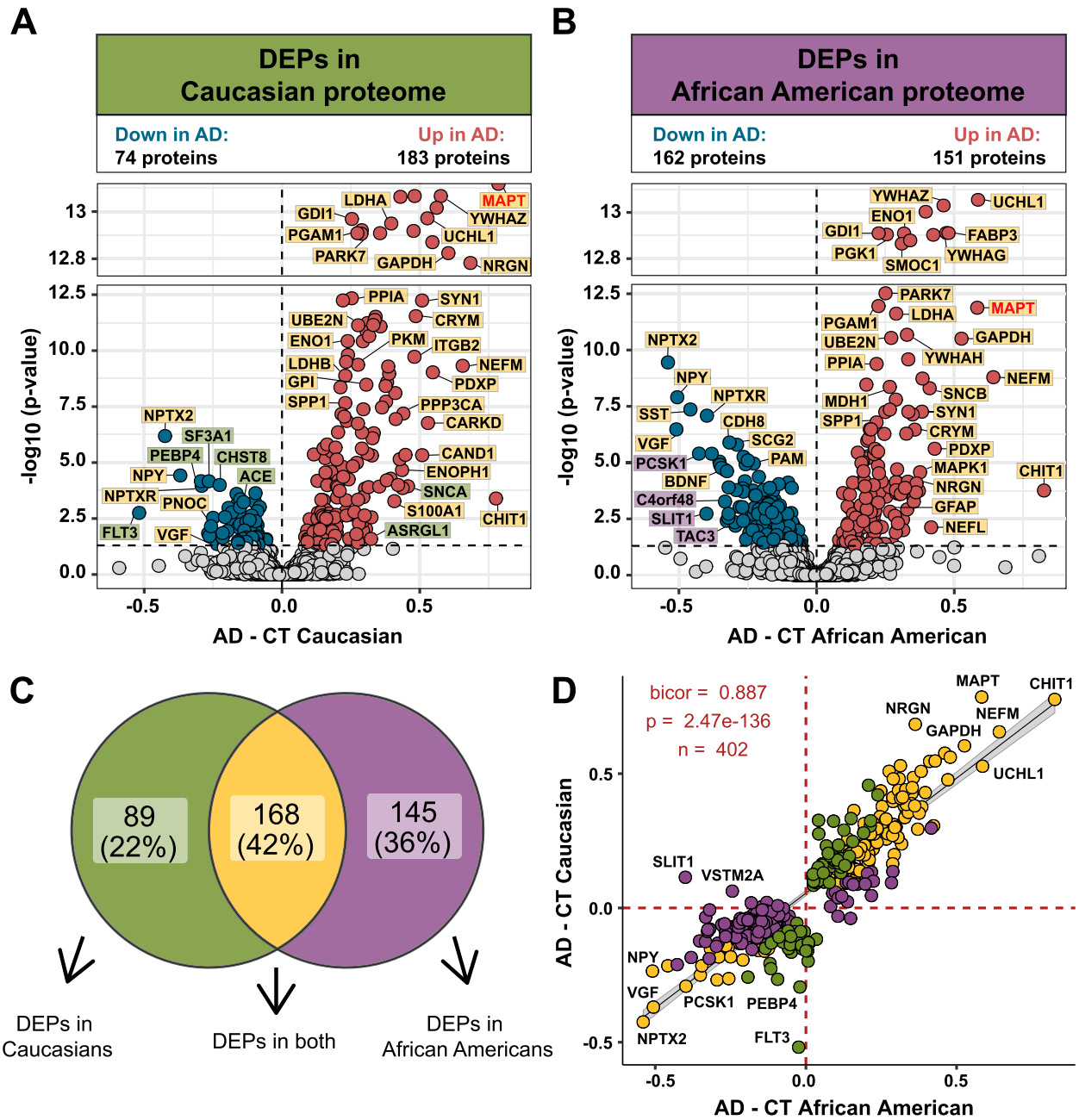

Fig. 2 Differential expression of Caucasian and African American CSF proteomes in AD. Volcano plot displaying the $\log_2$ fold change (FC) (x-axis) against one-way ANOVA with Tukey correction derived -log10 p-value (y-axis) for all proteins ($n = 1840$) comparing AD versus controls for Caucasians **A** and African Americans **B**. Cutoffs were determined by significant differential expression ($p < 0.05$) between control (CT) and AD cases. Proteins with significantly decreased levels in AD are shown in blue while proteins with significantly increased levels in disease were indicated in red. Select proteins were denoted and labeled by whether they were differentially expressed in both proteomes (yellow), in only the Caucasian proteome (green), or in only the African American proteome (purple). **C** Venn diagram illustrating the number of differentially expressed proteins (DEPs) that were uniquely changed in one proteome (green or purple) or changed in both proteomes (yellow) **D** The correlation between the fold change of all DEPs ($n = 402$) across the African American proteome (x-axis) and the Caucasian proteome (y-axis) were strongly correlated (bicor = 0.887, $p = 2.47\text{e-}136$), regardless of whether the DEP was significant in one (green or purple) or both proteomes (yellow)

minimal overlap with the brain, all other modules (79% total) in the CSF network significantly overlapped with at least one of the 44 brain modules (B-M1 to B-M44). For example, there was overlap with CSF proteins in M1 "Postsynaptic Membrane" with several neuronal modules in the consensus brain network (B-M1, B-M4, B-M5,

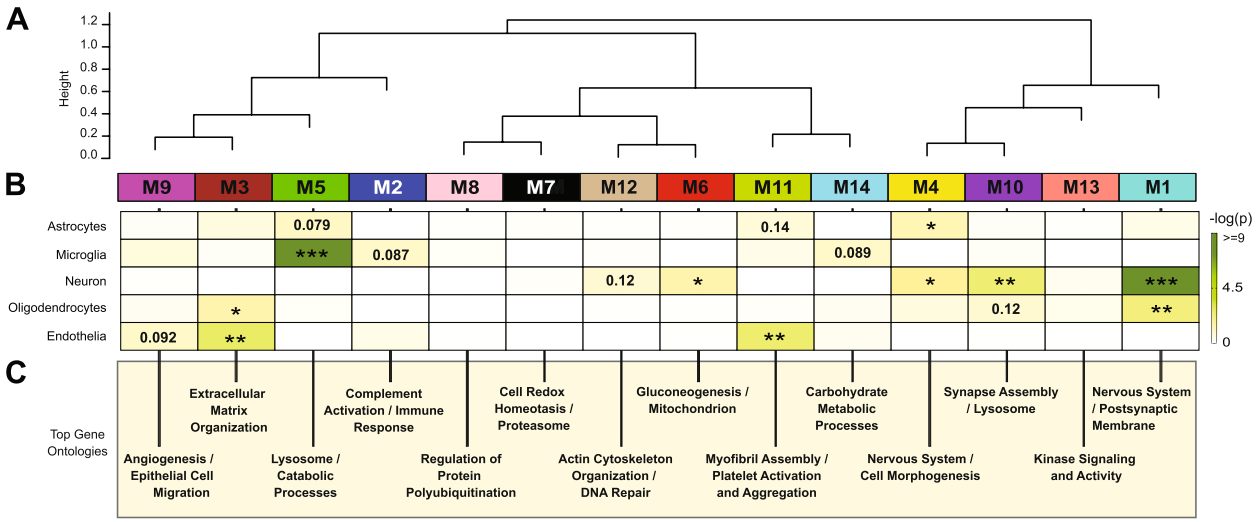

Fig. 3 Network analysis classifies the CSF proteome into modules associated with specific brain cell-types and gene ontologies. **A** Weighted Gene Co-expression Network Analysis cluster dendrogram groups proteins (n = 1840) into 14 distinct protein modules (M1-M14). **B** Cell-type enrichment was assessed by cross referencing module proteins by matching gene symbols using a one-tailed Fisher's exact test against a list of proteins determined to be enriched in neurons, oligodendrocytes, astrocytes, microglia and endothelia. The degree of cell-type enrichment increases from yellow to dark green with asterisks denoting the following statistical significance (*$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq .001$). **C** Top gene ontology (GO) terms were selected from significant GO annotations

B-M10, and B-M15). In addition, M2 "Complement Activation" in CSF overlaps with modules in human brain associated with complement and immune response (B-M26 and B-M40), whereas M3 "Extracellular Matrix" strongly overlapped with B-M27 in brain enriched with endothelial cell markers (Supplemental Fig. 2). Collectively, this supports that the co-expression in protein levels is, in part, shared between CSF and brain tissue, which could reflect changes in activation or phenotypes of specific brain cell types.

CSF protein modules correlate to race and clinicopathological phenotypes of AD

We assessed module correlation to race, cognitive scores (MoCA), and the hallmark AD biomarkers $A\beta_{42}$, tTau, and $pTau_{181}$. The protein network resulted in three main groups/clusters based on module relatedness (Fig. 4A). The first cluster (Group 1) was comprised of four modules (M2 "Complement Activation", M5 "Lysosome / Catabolism", M3 "Extracellular Matrix", and M9 "Angiogenesis / Cell Migration". Of these modules, M3 and M9 exhibited baseline racial differences in abundance levels (Fig. 4B). Notably, the eigenprotein, which corresponds to the first principal component of a given module and serves as a summary expression profile for all proteins within a module, were increased for these two modules in African Americans compared to Caucasians. Of note, these modules were enriched with endothelial cell markers (Fig. 3B) which suggests that genetic ancestry and/or environmental differences influence expression or secretion of these cell-type markers. Similarly, M2 and M5, both of which demonstrated enrichment for microglial markers, trended towards higher levels in both African American controls and AD (Supplemental Fig. 3A),

(See figure on next page.)

Fig. 4 CSF protein modules correlate to race and clinicopathological phenotypes of AD. **A** Modules were clustered based on relatedness defined by correlation of protein co-expression eigenproteins (indicated by position in height). There were three main clusters in the network: Groups 1, 2 and 3. Biweight midcorrelation (bicor) analysis of module eigenprotein levels with diagnostic measures of AD, including MoCA score, immunoassay Amyloid-beta$_{1\text{-}42}$ ($A\beta_{42}$), total Tau (tTau), phosphorylated Tau_{181} ($pTau_{181}$), ratio measures of tTau/$A\beta_{42}$, diagnosis, whether the sample has *APOE* ε4 allele or not, and race. The strength of positive (red) and negative (blue) correlations are shown by a heatmap with annotated bicor correlations and associated *p*-values. **B** Eigenprotein values distributed by race and diagnosis of representative modules for each cluster. **C** Differentially expressed proteins (DEPs) from AD samples compared to controls, by module with Caucasian proteome on the left and African Americans on the right. The height of the bars represents the fraction of module member proteins that were also DEPs compared to controls. The bars are color coded by heatmap for average $\log_2$ difference in abundance, where red represents an increase in abundance in AD, and blue represents a decrease in abundance in AD

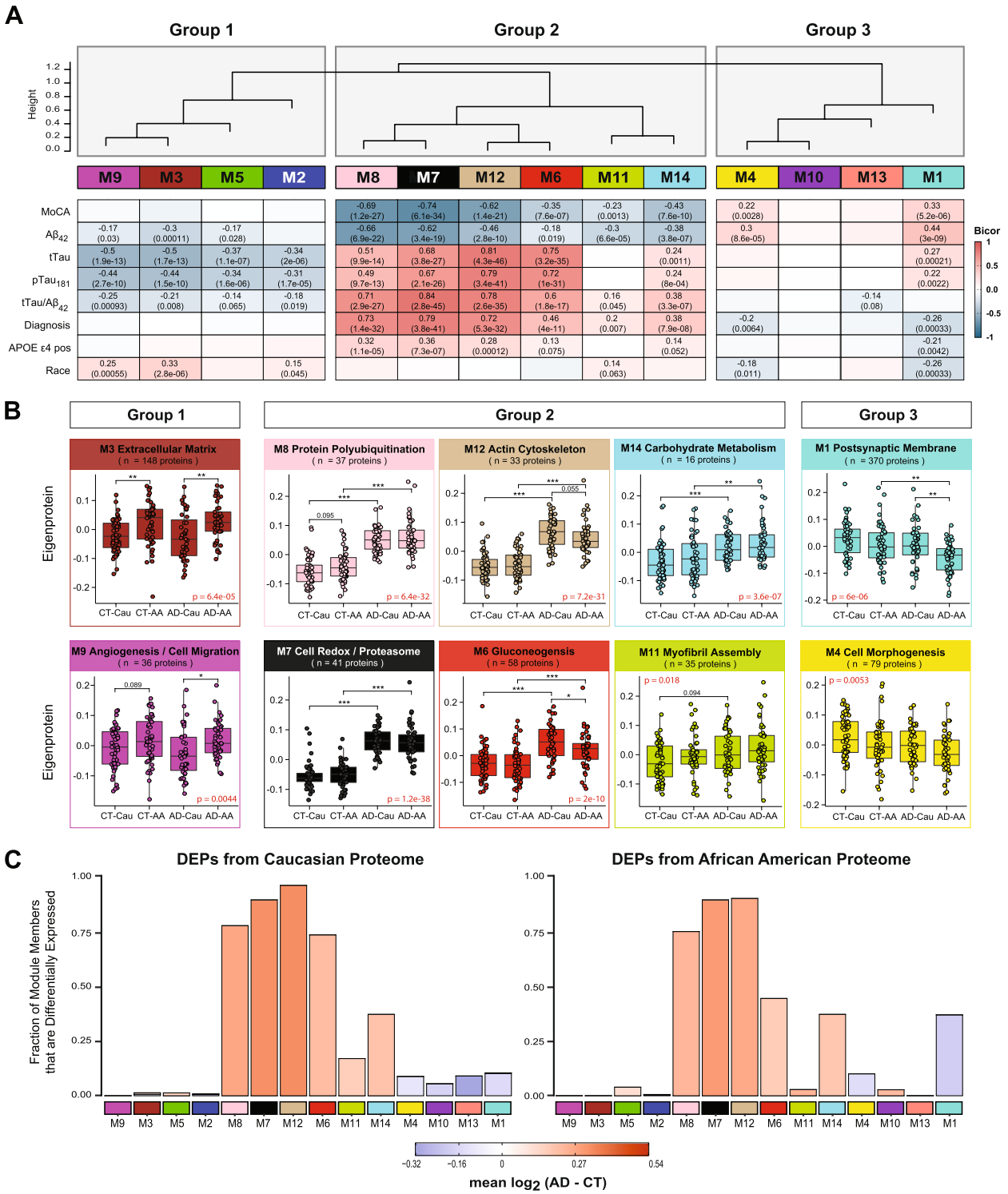

Fig. 4 (See legend on previous page.)

suggesting an accompanying immune response to the vascular alterations seen in modules M3 and M9.

The second cluster of modules (Group 2) was comprised of six modules (M8, M7, M12, M6, M11, and M14) that were all increased in AD (Fig. 4A). These AD modules also demonstrated significant negative correlations to MoCA scores and, conversely, significant positive correlations to tTau/$A\beta_{42}$ ratio. With the exception of

M11, these modules also exhibited positive correlations to APOE ε4 risk (Fig. 4A). Interestingly, a hub protein of the M12 "Actin Cytoskeleton" module was Tau (MAPT). Consistent with CSF levels observed for Tau by immunoassay and TMT-MS (Fig. 1 B and C), the M12 eigenprotein had lower levels in African Americans, compared to Caucasians with AD, albeit not significant ($p=0.055$) (Fig. 4B). Notably, M6 "Gluconeogenesis" was significantly lower in African Americans compared to Caucasians with AD, highlighting another module of CSF proteins that differed by race in AD (Fig. 4 A and B). This also indicated that the increased glycolytic signature of AD previously reported in CSF [25, 26] is higher in Caucasians with AD. Consistently, a greater proportion of increased DEPs in Caucasians with AD mapped to M6 compared to African Americans with AD (Fig. 4C). In contrast, M7 "Cell Redox / Proteasome" and M8 "Protein Polyubiquitination", had the strongest correlations to tTau/Aβ_{42} ratio and cognition (Fig. 4B), and both demonstrated strong, equivalent elevations in African Americans and Caucasians with AD (Fig. 4B). This is consistent with an equivalent fraction of increased DEPs mapping to these modules in African American and Caucasians with AD (Fig. 4C). Therefore, proteins in these modules including 14–3-3 family members (YWHAZ, YWAHB, YWHAG, YWHAE) likely represent the best class of CSF AD biomarkers that are not influenced by race. M14 "Carbohydrate Metabolism" and M11 "Myofibril Assembly" were both elevated in both African Americans and Caucasians with AD (Fig. 4A and B), yet to a lesser degree than M7 and M8.

The final group of modules (Group 3) contained two modules, M1 "Postsynaptic Membrane" and M4 "Cell Morphogenesis", that showed strong correlations to both race and AD diagnosis (Fig. 4A). Both modules were i) decreased in AD compared to controls and ii) and were lower in African Americans compared to Caucasians. In addition, both M1 and M4 were enriched with neuronal markers and positively associated with cognitive MoCA scores (Fig. 4A). Markedly, pairwise statistical analysis of eigenprotein levels for M1 across diagnosis and race revealed significantly lower levels in African Americans with AD (Fig. 4B). To this end, most of the decreased DEPs in African Americans with AD mapped to M1 and to a lesser degree M4, whereas decreased DEPs in Caucasians with AD were equally distributed to M1, M4, M13 and M10 (Fig. 4C). Notably, M10 and M13 within Group 3 did not show any differences with AD or race and did not significantly correlate with traits explored in this study (Fig. 4 and Supplemental Fig. 3). Overall, network analysis effectively organizes the CSF proteome into protein modules that are strongly linked to hallmark AD biomarkers (Aβ_{42}, tTau and pTau$_{181}$) and cognition, which in some cases were also influenced by race.

Validation of shared and divergent CSF protein alterations across AD and race

To further validate these network findings, we used a targeted mass spectrometry method, selected reaction monitoring (SRM), with heavy labeled internal standards to measure CSF proteins across 195 of the 203 cases included in the discovery TMT-MS assays (Fig. 5A). The proteins and corresponding targeted peptides were previously selected based on their robust detection and significant differential expression in previous CSF discovery proteomic datasets [26, 46]. We used pooled CSF samples of control, and AD cases as quality controls replicates ($n=29$ samples total) to assess technical reproducibility. Of the peptides targeted, 85 peptides (mapping to 58 proteins) had a coefficient of variation of < 20% in both the control and AD pools with no missing values (Supplemental Tables 9 and 10). Following adjustments of co-variates (i.e., age and sex), peptide levels were highly correlated with protein levels measured by TMT-MS from the same samples (Supplemental Tables 11 and 12). If a protein was measured by more than one peptide the most correlated peptide to the TMT-MS protein level was selected for further analysis. The final peptide list can be found in Supplemental Table 13. ANOVA analyses determined pairwise significance between the four groups (i.e., Control-Caucasians vs Control-African Americans vs AD-Caucasians vs AD-African Americans, Supplemental Table 14). Figure 5B highlights peptides ($n=24$) that reached significance and that mapped to proteins in CSF modules associated with race and/or AD. Consistent with the TMT-MS protein measurements, proteins measured by SRM within M7 (GAPDH and YWHAG) and M8 (YWAHB and PPIA) had strong elevations ($p<0.001$) in abundance in AD in both races, whereas proteins in M12 (SMOC1, PARK7, and LDHB) had a greater magnitude of change in Caucasians than African Americans with AD (for a list of all M12 members, see Supplemental Table 6). Similarly, a majority of the proteins measured by SRM in M6 (PKM, GDA, TPI1, GOT1, ALDOA and ENO2) were more increased in Caucasians than African Americans with AD (Fig. 5B and C). Proteins in the synaptic M1 module (VGF, SCG2, NPTX2, and NPTXR) were significantly decreased in African Americans with AD compared to Caucasians (Fig. 5B and C), again consistent with TMT-MS protein level abundance. Notably, African Americans with or without APOE ε4 allele in the AD group had reduced levels of these CSF peptide biomarkers compared to Caucasians indicating that race and not APOE status was driving the difference in abundance

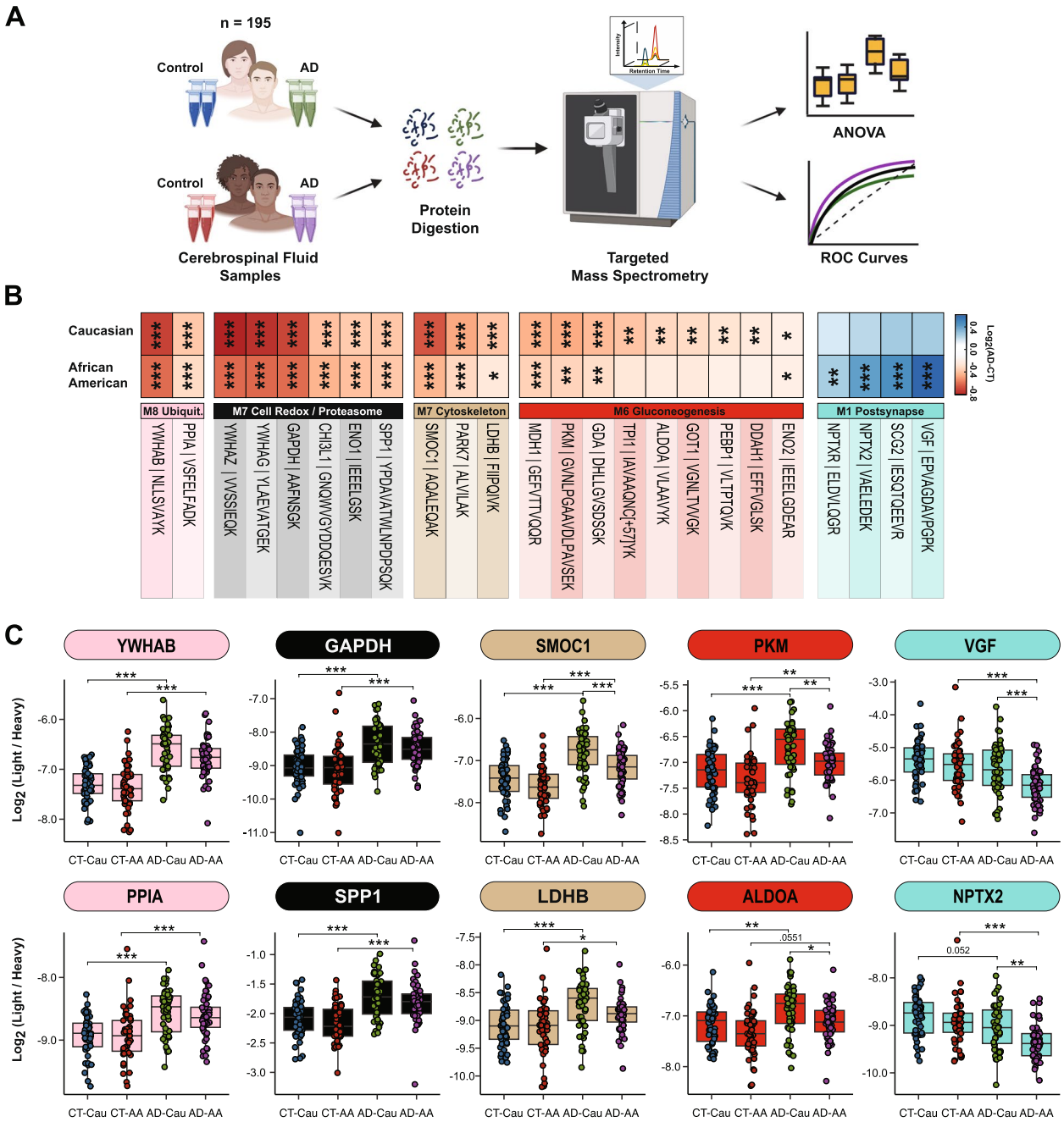

Fig. 5 Validation of shared and divergent CSF protein levels across AD and race. **A** Schematic of experimental workflow for SRM analysis of cerebrospinal fluid proteome **B** Heatmap of peptides that were significantly differentially expressed between control and AD Caucasians or African Americans. Stars are indicative of the level of significant difference (*$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$) seen for each peptide between AD and control within each race. Meanwhile the colors are indicative of the $\log_2$ fold change (FC) of each peptide from control and AD for each race where blue is indicative of the degree of decrease and shade of red is indicative of the degree of increase. **C** Log_2 abundance of peptides that mapped to modules of interest distributed by race and diagnosis. Pairwise significance was calculated using one-way ANOVA with Tukey adjustment

(Supplemental Fig. 4A). Furthermore, these differences across race remained consistent even after removing patients with one or more comorbid condition (i.e., hypertension, diabetes, dyslipidemia, or cerebrovascular disease; Supplemental Fig. 4B and Supplemental Table 1).

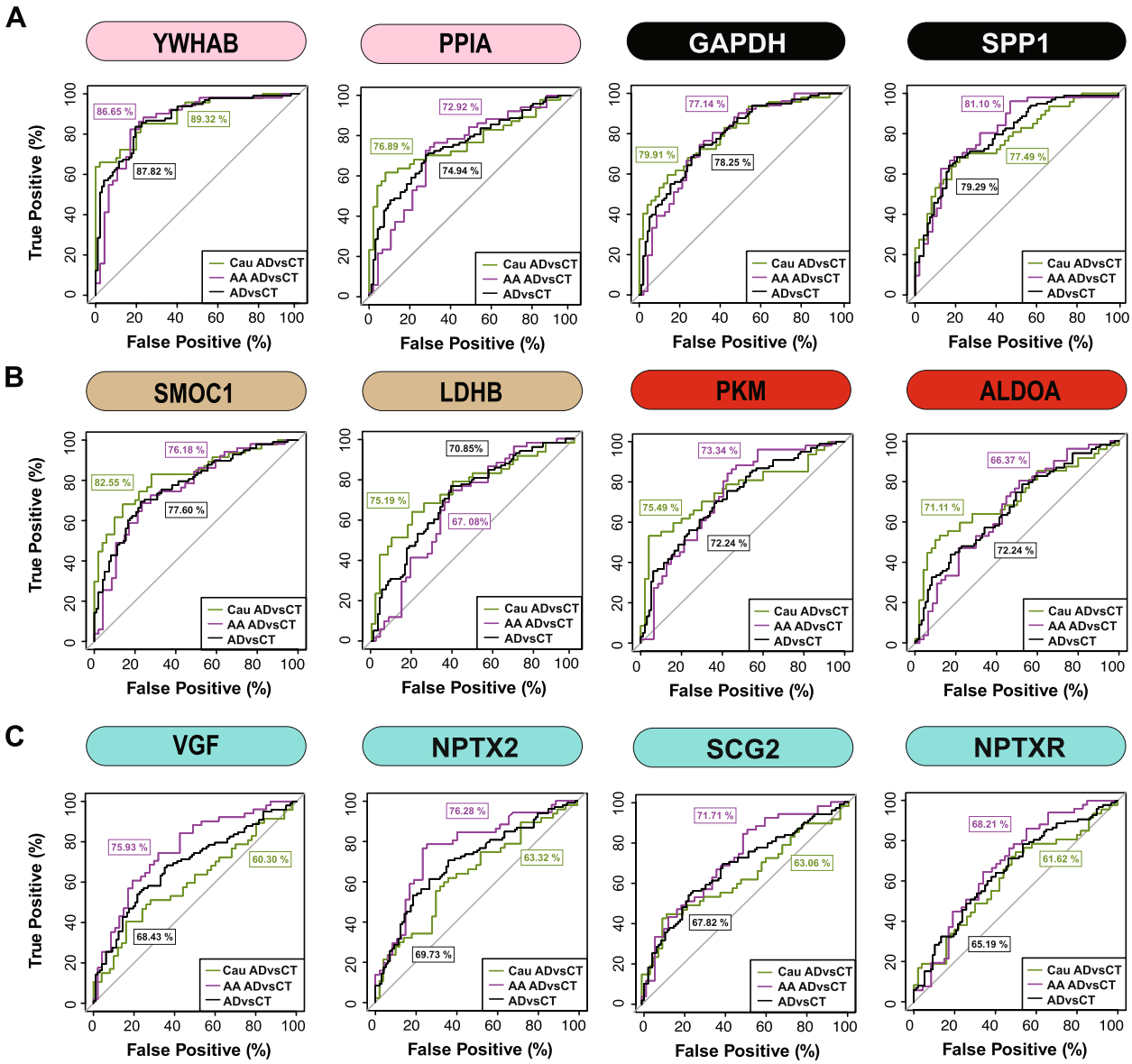

Fig. 6 ROC analysis to evaluate CSF protein classification of AD by race. **A** YWHAB, PPIA, GAPDH, and SPP1 had similar performance in classifying Caucasians and African Americans with AD **B** SMOC1, LDHB, PKM and ALDOA showed modest improvement in the AUC for Caucasians with AD compared to African Americans with AD **C** VGF, SCG2, and NPTX2 were better classifiers for AD in African Americans compared to Caucasians, whereas NPTXR showed modest improvement in classification of AD in African Americans. All protein AUCs with p-values and confidence internals (CI) are provided in Supplemental Table 15

Finally, a receiver operating characteristic (ROC) curve analysis was performed to assess the performance of each peptide biomarker in differentiating controls and AD by race (Fig. 6 and Supplemental Table 15). We generated an area under the curve (AUC) for AD in African American and Caucasian individuals for each protein biomarker (considered separately in each race). As expected, proteins mapping to M8 and M7 including 14–3-3 proteins (YWHAB, YWHAG and YWHAZ) were equally able to discriminate AD from control irrespective of racial background. Notably, despite having lower levels in African Americans with AD compared to Caucasians with AD, only a modest improvement in the AUC for SMOC1 was observed for classifying AD in Caucasians AUC = 0.8255 (p = 1.71e-08, CI = 0.7421–0.9090) compared to African Americans AUC = 0.7618 (p = 4.12e-06, CI = 0.6660–0.8576). Similar findings were observed for another M12 protein, LDHB, as well as M6 proteins PKM and ALDOA.

However, the M1 protein VGF was only nominally significant at classifying AD in Caucasian AUC = 0.6030 ($p = 0.0406$, CI = 0.4887–0.7173), yet highly significant in African Americans AUC = 0.7593 ($p = 5.03e-06$, CI = 0.6634–0.8552). Similar results were observed for other synaptic M1 proteins, NPTX2 and SCG2, whereas NPTXR showed only a modest improvement in the AUC between African Americans compared to Caucasians with AD (Fig. 6 and Supplemental Table 15). Collectively this supports a hypothesis that African Americans with AD have lower levels of a subset of neuronal biomarkers compared to Caucasians with AD.

Discussion

Here we performed an unbiased quantitative analysis of the AD CSF proteome to identify protein biomarkers reflective of underlying brain physiology that are shared or unique across race. Using a network analysis, we organized the CSF proteome into 14 modules of highly correlated proteins. Notably, these modules were associated with cell-types and biological pathways in brain and largely overlapped with modules in a consensus human AD brain proteomic network [24]. Consistent with previous findings [16, 17], we also show that Tau levels are lower in African Americans with AD compared to Caucasians. Notably, Tau mapped to a CSF module enriched with other related neuronal/cytoskeletal proteins with a magnitude of increase greater in Caucasians than in African Americans with AD. In contrast, CSF modules which included 14–3-3 proteins, were elevated equivalently in both African Americans and Caucasians with AD. A module enriched with neuronal/synaptic proteins including VGF, SCG2, and NPTX2 was significantly lower in African Americans than Caucasians with AD. VGF, SCG2, and NPTX2 levels in CSF measured by SRM were also significantly better at classifying African Americans with AD than Caucasians. Thus, our findings suggest that there are likely distinct mechanisms underlying the abundance and/or secretion of CSF neuronal markers including Tau and VGF that differ by race. Collectively, these data underscore the need for further investigations into how AD biomarkers and underlying physiology vary across different racial backgrounds.

In a previous study we performed unbiased TMT-MS from a small discovery cohort of control and AD CSF samples ($n = 40$) and mapped these proteins onto a human AD brain co-expression network, which revealed that approximately 70% of the CSF proteome overlapped with the brain network [26]. The increased sample size in this study afforded the opportunity to build an independent co-expression network of the CSF proteome and then assess overlap with a consensus brain network. We observed a strong overlap between CSF and brain modules, indicating that there is conservation of protein co-expression across brain and CSF. This correlation is, in part, driven by the cell-types changes, as we observed modules that were enriched with neuronal, glial and endothelial specific markers. In terms of the CSF network biology that differed by race, it is noteworthy that modules significantly enriched in endothelial proteins (M3 and M9) were increased in African Americans across both control and AD individuals. This suggests that fundamental differences exist in the levels and/or activation state of cells residing in the vasculature between African Americans and Caucasians. Whether this biological difference between the two races is observed in brain tissues or relates to a higher incidence of vascular health disparities between African Americans and Caucasians [47] requires further investigation. Additionally, further studies of the proteomic differences between races across disease subgroups, will allow construction of robust and widely applicable biomarker panels.

The current biological framework for the pre-symptomatic stages of AD is based on the presence of β-amyloidosis (A), tauopathy (T), and neurodegeneration (N) also termed the A/T/N framework [13]. CSF remains the gold standard for A/T/N biomarkers of neurodegenerative disease as it maintains direct contact with the brain and reflects biochemical changes in amyloid, tau and neurodegeneration. A strength of our study was the balanced nature of African American samples, which offered the ability to examine racial differences in both cognitively normal controls with biomarker positive (A + /T +) individuals diagnosed with AD. Our mass spectrometry measurements of Tau strongly correlated with immunoassay levels measured on the Roche Elecsys platform reinforcing measurements made by TMT-MS. Increased Tau in CSF is considered to result from neurodegeneration, however, it can be increased in early pre-symptomatic disease stages when neurodegeneration is limited [13, 48]. Recently, Tau CSF levels have been linked to enhanced synaptic plasticity [49]. A large proportion of synaptic proteins in this study mapped to M1 and M4 which were positively associated with cognition and were decreased in AD. However, other modules, M6 and M12, also contained neuronal proteins including Tau (MAPT), SNCB, SYN1, BASP1, GAP43, SYT1, NRGN among others that were negatively associated with cognition and increased in AD. Thus, there are likely distinct mechanisms that result in the activation or abundance of neurons that contribute to this discordance in levels in AD CSF. We observed that African Americans in this study had on average lower levels of neuronal markers mapping to M1 and M4 in the network, which are reduced in AD. Paradoxically, African Americans also have lower levels of neuronal proteins in M6 and M12, which all increase

in AD. Thus, increased CSF Tau, thought to be a marker of neurodegeneration, does not equate to decreased levels of VGF and other postsynaptic markers in CSF. Consistent with this observation, in a recent CSF proteomic study in a asymptomatic Caucasian European population stratified by Tau CSF levels, individuals deemed to have high Tau levels maintained levels of M1 post-synaptic proteins (CADM3, NEO1, NPTX1, CHGB, PCSK1, NEGR, L1CAM, PTPRN, CACNA2D, PAM, VEGFA, NBL1 etc.) compared to individuals with lower Tau levels [49]. This observation is analogous to differences we see between African Americans and Caucasians with AD. M1 members VGF and NPTX2, strongly correlate to antemortem cognitive measures [50–52] and VGF and NPTX2 has been nominated as biomarkers of neurodegeneration (N) as CSF levels enhance prediction of MCI to AD [52–54]. Collectively, this would suggest that a specific sub-group of individuals with AD, including African Americans, have a higher burden of neurodegeneration (N) despite low CSF Tau levels. Future studies that analyze these CSF changes in neuronal proteins and other module members longitudinally will be important to better resolve these changes by race throughout disease progression.

Although a strength of our study was the large number of African Americans included there are several limitations that should be noted. First, we acknowledge that many of the protein changes we observe in the CSF across race could be due to ancestral or genetic differences [55, 56]. There was no genetics *a priori* performed on these study participants to confirm enrichment of African *vis a vis* European ancestry [57] as we stratified race solely by self-identification. Future studies, which include the integration of genetics and protein abundance to define protein quantitative trait loci (pQTL) will be necessary to resolve which proteins are under genetic control across race [58–60]. It is noteworthy that the expression level of most modules which differed between racial groups were decreased in African Americans relative to Caucasians. Upon integration with whole genome profiling of larger cohorts, these patterns may help in the future to identify pQTLs or other mechanisms influencing synthesis and turnover of proteins that differ by race that differ by race. Only a few studies to date have investigated proteomic difference by race in AD [61, 62], which have predominately focused on brain tissues and not on the scale of this current study. However, a major initiative of the Accelerating Medicine Partnership (AMP)-AD partnership [63] is to increase the number of diverse tissues in the multi-omic analyses performed, which will complement data generated from these previous studies. Second, despite the well documented differences in the quality of education, higher rates of poverty, and greater exposure to adversity and discrimination that increase risk for dementia [4, 5], these metrics were not captured on the participants in this study. Integrating CSF protein levels with vascular risk factors, and other environmental metrics in larger cohorts may help better resolve some of the underlying racial differences in the CSF proteome. Finally, in this study we adjusted for co-factors such as age and sex to pinpoint changes that are most likely to be associated with race and AD. Sex and age have an impact on the abundance of CSF Tau and other protein levels [64]. Therefore, future studies that assess the interactions between, age, sex and race will be informative. Nevertheless, this study reveals an impressive view of protein co-expression in AD CSF across race, which provides new insights into the pathways underlying cell-type changes and further evidence that race may mediate these in AD. These data also provide a resource for further investigations into how AD biomarkers vary across different racial backgrounds.

Conclusions

We developed a TMT-MS based quantification pipeline for proteomic analysis of human CSF in both Caucasians and African Americans with AD that led to additional insights into proteins and pathways reflecting underlying brain pathology that are shared or differ by race. Our findings also demonstrate the utility of a systems-based approach in the identification of CSF proteins that could serve as markers for AD across a more diverse population. Ultimately, these results highlight a need for further investigations into how AD heterogeneity varies across different racial backgrounds and affirm the urgency and importance of increased inclusion of under-represented groups to ensure the completeness and generalizability of research findings.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13024-023-00638-z.

Additional file 1: Supplemental Table 1. Cohort characteristics (Aβ42 levels that reached saturation (1700 pg / mL) were excluded from calculations and analysis). **Supplemental Table 2.** TMT batch arrangement. **Supplemental Table 3.** TMT output before adjustments. **Supplemental Table 4.** Post-regression protein table. **Supplemental Table 5.** TMT MS anova table. **Supplemental Table 6.** WGCNA KME table. **Supplemental Table 7.** Module go terms and *P*-values. **Supplemental Table 8.** Cell type marker list. **Supplemental Table 9.** SRM PEPTIDE coefficients of variation. **Supplemental Table 10.** SRM ratio abundances pre-regression. **Supplemental Table 11.** SRM ratio abundances post-regression. **Supplemental Table 12.** Correlation values between SRM AND TMT-MS. **Supplemental Table 13.** SRM culled list. **Supplemental Table 14.** SRM anova table. **Supplemental Table 15.** ROC-AUC analysis table.

Additional file 2: Supplemental Figure 1. Batch correction, outlier removal and bootstrap regression. Multidimensional scaling (MDS) plots were used to illustrate batch contributions to variance before and after batch correction. In MDS plots, the distance a case is from one another is

reflective of how similar or dissimilar a case is from the other. (**A**) Prior to batch correction, the samples clustered by batch (**B**) After batch correction, the samples no longer cluster by batch. (**C**) After batch correction, a principal component (PC)-based outlier removal method was utilized to detect outliers. By graphing the eigenvalue of each component against the PC number, the elbow or bend in the graph, which in this case was 7, was indicative of the ideal number of components to include within the parameters. (**D**) With a criterion for computing cutoff values set to 0.99, the cutoffs for the detection of outliers for the orthogonal distance and score were 16.79257 and 4.654674 respectively. This resulted in the detection of 15 outliers (b1.128N, b11.130C, b13.130N, b14.127N, b14.133N, b15.128C, b15.131N, b2.127C, b2.128C, b2.133N, b5.131C, b7.127C, b7.130N, b8.129N, b9.127C). B14.133N was such an extreme outlier because it was an empty channel. (**E**) After outlier removal, the matrix underwent bootstrap regression to remove variations in the dataset that were due to age and sex. Variance partition plots were employed to illustrate the percent contribution of diagnosis, race, age, and sex to the variance of each protein. (**F**) Following bootstrap regression, variations explained by age and sex were removed.

Additional file 3: Supplemental Figure 2. (A) Protein module enrichment across the CSF and brain was assessed by matching gene symbols of proteins in each module from the CSF network against gene symbols for protein in each module from a human AD consensus brain network using a one-tailed Fisher's exact test. The degree of enrichment increases from pink to light purple to dark purple with asterisks denoting the following statistical significance (**$p \leq 0.01$ and ***$p \leq .001$). (**B**) Similar to CSF, cell-type enrichment was assessed by cross referencing brain module proteins against a list of proteins determined to be enriched in neurons, oligodendrocytes, astrocytes, and microglia using a one-tailed Fisher's exact test. The degree of cell-type enrichment increases from yellow to green-yellow to dark green with asterisks denoting the following statistical significance(**$p \leq 0.01$ and ***$p \leq .001$).

Additional file 4: Supplemental Figure 3. Additional CSF network protein modules. (**A**) Eigenprotein levels were distributed by race and diagnosis for remaining modules not shown in main Figure 4. This includes M5, M2, M10, and M13.

Additional file 5: Supplemental Figure 4. Stratification of SRM CSF protein measurements in by APOE genotype and comorbidity. (**A**) Within each race, protein levels for were not affected by APOE ε4 genotype for YWHAB, GAPDH, SMOC1, PKM, VGF, PPIA, SPP1, LDHB, ALDOA, and NPTX2. (**B**) Within each race, protein levels for were not affected by patient co-morbidities (hypertension, diabetes, dyslipidemia, or cerebrovascular disease) for YWHAB, GAPDH, SMOC1, PKM, VGF, PPIA, SPP1, LDHB, ALDOA, and NPTX2. Only cases with co-morbidities (**Supplemental Table 1**) were included in these boxplots.

Acknowledgements
We acknowledge Dr. Lenora Higginbotham and Dr. Eddie Fox and our colleagues at Emory for providing critical feedback.

Authors' contributions
Conceptualization, E.S.M., J.J.L., and A.I.L., and N.T.S.; Methodology, E.S.M, L.P., D.M.D., C.M.W., B.R.R., J.J.L and N.T.S; Investigation, E.S.M., L.P., C.M.W, E.B.D., J.J.L., A.I.L., and N.T.S.; Formal Analysis, E.S.M., C.M.W., E.B.D.; Writing – Original Draft, E.S.M., D.M.D., and N.T.S.; Writing – Review & Editing, E.S.M., C.M.W., E.B.D, E.C.B.J., J.J.L., A.I.L., and N.T.S.; Funding Acquisition, J.J.L, A.I.L. and N.T.S.; Resources, J.J.L. A.I.L. and N.T.S., and.; Supervision, A.I.L., J.J.L., and N.T.S. All authors read and approved the final manuscript.

Funding
This study was supported by the following National Institutes of Health funding mechanisms: U01AG061357 (A.I.L and N.T.S), R01AG070937 (J.J.L) and P30AG066511 (A.I.L) and the Foundations for the National Institute of Health AMP-AD 2.0 grant.

Availability of data and materials
Raw mass spectrometry data and pre- and post-processed protein expression data and case traits related to this manuscript are available from CSF proteomics can be found at https://www.synapse.org/EmoryDiversityCSF. The results published here are in whole or in part based on data obtained from the AMP-AD Knowledge Portal (https://adknowledgeportal.synapse.org). The AMP-AD Knowledge Portal is a platform for accessing data, analyses and tools generated by the AMP-AD Target Discovery Program and other programs supported by the National Institute on Aging to enable open-science practices and accelerate translational learning. The data, analyses and tools are shared early in the research cycle without a publication embargo on secondary use. Data are available for general research use according to the following requirements for data access and data attribution (https://adknowledgeportal.synapse.org/#/DataAccess/Instructions).

Declarations

Ethics approval and consent to participate
Local institutional review boards approved the procedures for this study and written informed consent was obtained.

Consent for publication
Not applicable.

Competing interests
A.I.L, N.T.S. and D.M.D. are co-founders of Emtherapro Inc.

Author details
[1]School of Medicine, Department of Biochemistry, Emory University, Atlanta, GA, USA. [2]School of Medicine, Department of Neurology, Emory University, Atlanta, GA, USA.

Received: 14 December 2022 Accepted: 21 June 2023
Published online: 19 July 2023

References

1. Rajan KB, Weuve J, Barnes LL, Wilson RS, Evans DA. Prevalence and incidence of clinically diagnosed Alzheimer's disease dementia from 1994 to 2012 in a population study. Alzheimers Dement. 2019;15:1–7.
2. Potter GG, Plassman BL, Burke JR, Kabeto MU, Langa KM, Llewellyn DJ, et al. Cognitive performance and informant reports in the diagnosis of cognitive impairment and dementia in African Americans and whites. Alzheimers Dement. 2009;5:445–53. https://doi.org/10.1016/j.jalz.2009.04.1234.
3. Gurland BJ, Wilder DE, Lantigua R, Stern Y, Chen J, Killeffer EH, et al. Rates of dementia in three ethnoracial groups. Int J Geriatr Psychiatry. 1999;14:481–93.
4. Glymour MM, Manly JJ. Lifecourse social conditions and racial and ethnic patterns of cognitive aging. Neuropsychol Rev. 2008;18:223–54.
5. Lines L, Sherif N, Wiener J. Racial and ethnic disparities among individuals with Alzheimer's disease in the United States: A literature review RTI Press. 2014. https://doi.org/10.3768/rtipress.2014.RR.0024.1412.
6. Reitz C, Jun G, Naj A, Rajbhandary R, Vardarajan BN, Wang LS, et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E 4, and the risk of late-onset Alzheimer disease in African Americans. JAMA. 2013;309:1483–92.
7. Logue MW, Schu M, Vardarajan BN, Buros J, Green RC, Go RC, et al. A comprehensive genetic association study of Alzheimer disease in African Americans. Arch Neurol. 2011;68:1569–79. https://doi.org/10.1001/archneurol.2011.646.
8. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. a meta-analysis. APOE and Alzheimer disease meta analysis consortium. JAMA. 1997;278:1349–56.
9. Ballatore C, Lee VM, Trojanowski JQ. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. Nat Rev Neurosci. 2007;8:663–72.
10. Selkoe DJ. Amyloid beta protein precursor and the pathogenesis of Alzheimer's disease. Cell. 1989;58:611–2.

11. Blennow K, Hampel H, Weiner M, Zetterberg H. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Nat Rev Neurol. 2010;6:131–44. https://doi.org/10.1038/nrneurol.2010.4.
12. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7:280–92. https://doi.org/10.1016/j.jalz.2011.03.003.
13. Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018;14:535–62. https://doi.org/10.1016/j.jalz.2018.02.018.
14. Olsson B, Lautner R, Andreasson U, Ohrfelt A, Portelius E, Bjerke M, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. Lancet Neurol. 2016;15:673–84. https://doi.org/10.1016/S1474-4422(16)00070-3.
15. Ashton NJ, Scholl M, Heurling K, Gkanatsiou E, Portelius E, Hoglund K, et al. Update on biomarkers for amyloid pathology in Alzheimer's disease. Biomark Med. 2018;12:799–812. https://doi.org/10.2217/bmm-2017-0433.
16. Howell JC, Watts KD, Parker MW, Wu J, Kollhoff A, Wingo TS, et al. Race modifies the relationship between cognition and Alzheimer's disease cerebrospinal fluid biomarkers. Alzheimers Res Ther. 2017;9:88. https://doi.org/10.1186/s13195-017-0315-1.
17. Morris JC, Schindler SE, McCue LM, Moulder KL, Benzinger TLS, Cruchaga C, et al. Assessment of racial disparities in biomarkers for Alzheimer disease. JAMA Neurol. 2019;76(3):264–73. https://doi.org/10.1001/jamaneurol.2018.4249.
18. Clark PC, Kutner NG, Goldstein FC, Peterson-Hazen S, Garner V, Zhang R, et al. Impediments to timely diagnosis of Alzheimer's disease in African Americans. J Am Geriatr Soc. 2005;53:2012–7. https://doi.org/10.1111/j.1532-5415.2005.53569.x.
19. Hernandez S, McClendon MJ, Zhou XH, Sachs M, Lerner AJ. Pharmacological treatment of Alzheimer's disease: effect of race and demographic variables. J Alzheimers Dis. 2010;19:665–72.
20. Garrett SL, McDaniel D, Obideen M, Trammell AR, Shaw LM, Goldstein FC, et al. Racial disparity in cerebrospinal fluid amyloid and tau biomarkers and associated cutoffs for mild cognitive impairment. JAMA Netw Open. 2019;2:e1917363.
21. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol. 2009;65:403–13. https://doi.org/10.1002/ana.21610.
22. Seyfried NT, Dammer EB, Swarup V, Nandakumar D, Duong DM, Yin L, et al. A multi-network approach identifies protein-specific co-expression in asymptomatic and symptomatic Alzheimer's disease. Cell Syst. 2017;4(60–72):e64.
23. Johnson ECB, Dammer EB, Duong DM, Yin L, Thambisetty M, Troncoso JC, et al. Deep proteomic network analysis of Alzheimer's disease brain reveals alterations in RNA binding proteins and RNA splicing associated with disease. Mol Neurodegener. 2018;13:52. https://doi.org/10.1186/s13024-018-0282-4.
24. Johnson ECB, Carter EK, Dammer EB, Duong DM, Gerasimov ES, Liu Y, et al. Large-scale deep multi-layer analysis of Alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level. Nat Neurosci. 2022;25:213–25. https://doi.org/10.1038/s41593-021-00999-y.
25. Johnson ECB, Dammer EB, Duong DM, Ping L, Zhou M, Yin L, et al. Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. Nat Med. 2020;26:769–80. https://doi.org/10.1038/s41591-020-0815-6.
26. Higginbotham L, Ping L, Dammer EB, Duong DM, Zhou M, Gearing M, et al. Integrated proteomics reveals brain-based cerebrospinal fluid biomarkers in asymptomatic and symptomatic Alzheimer's disease. Sci Adv. 2020;6:eaaz9360.
27. Bai B, Wang X, Li Y, Chen P-C, Yu K, Dey KK, et al. Deep multilayer brain proteomics identifies molecular networks in Alzheimer's disease progression. Neuron. 2020;105:975-991.e977. https://doi.org/10.1016/j.neuron.2019.12.015.
28. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7:263–9. https://doi.org/10.1016/j.jalz.2011.03.005.
29. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7:270–9. https://doi.org/10.1016/j.jalz.2011.03.008.
30. Bittner T, Zetterberg H, Teunissen CE, Ostlund RE Jr, Militello M, Andreasson U, et al. Technical performance of a novel, fully automated electrochemiluminescence immunoassay for the quantitation of beta-amyloid (1–42) in human cerebrospinal fluid. Alzheimers Dement. 2016;12:517–26.
31. 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. Alzheimers Dement. 2018;14:1460–9. https://doi.org/10.1016/j.jalz.2018.01.013.
32. Hansson O, Seibyl J, Stomrud E, Zetterberg H, Trojanowski JQ, Bittner T, et al. CSF biomarkers of Alzheimer's disease concord with amyloid-beta PET and predict clinical progression: a study of fully automated immunoassays in BioFINDER and ADNI cohorts. Alzheimers Dement. 2018;14:1470–81. https://doi.org/10.1016/j.jalz.2018.01.010.
33. Dammer EB, Ping L, Duong DM, Modeste ES, Seyfried NT, Lah JJ, et al. Multi-platform proteomic analysis of Alzheimer's disease cerebrospinal fluid and plasma reveals network biomarkers associated with proteostasis and the matrisome. Alzheimers Res Ther. 2022;14:174. https://doi.org/10.1186/s13195-022-01113-5.
34. Winiarska A, Zareba L, Krolczyk G, Czyzewicz G, Zabczyk M, andUndas, A. decreased levels of histidine-rich glycoprotein in advanced lung cancer: association with prothrombotic alterations. Dis Markers. 2019;2019:8170759.
35. Ping L, Duong DM, Yin L, Gearing M, Lah JJ, Levey AI, et al. Global quantitative analysis of the human brain proteome in Alzheimer's and Parkinson's Disease. Sci Data. 2018;5:180036.
36. Magistretti PJ, Allaman I. A cellular perspective on brain energy metabolism and functional imaging. Neuron. 2015;86:883–901.
37. Maienschein-Cline M, Lei Z, Gardeux V, Abbasi T, Machado RF, Gordeuk V, et al. ARTS: automated randomization of multiple traits for study design. Bioinformatics. 2014;30:1637–9.
38. Watson CM, Dammer EB, Ping L, Duong DM, Modeste E, Carter EK, et al. Quantitative mass spectrometry analysis of cerebrospinal fluid protein biomarkers in Alzheimer's Disease. Sci Data. 2023;10:261.
39. Chen X, Zhang B, Wang T, Bonni A, andZhao, G. Robust principal component analysis for accurate outlier sample detection in RNA-Seq data. BMC Bioinformatics. 2020;21:269.
40. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler. GSEA, Cytoscape and EnrichmentMap. Nat Protoc. 2019;14:482–517. https://doi.org/10.1038/s41596-018-0103-9.
41. Bader JM, Geyer PE, Muller JB, Strauss MT, Koch M, Leypoldt F, et al. Proteome profiling in cerebrospinal fluid reveals novel biomarkers of Alzheimer's disease. Mol Syst Biol. 2020;16:e9356.
42. Dayon L, Núñez Galindo A, Wojcik J, Cominetti O, Corthésy J, Oikonomidi A, et al. Alzheimer disease pathology and the cerebrospinal fluid proteome. Alzheimers Res Ther. 2018;10:66. https://doi.org/10.1186/s13195-018-0397-4.
43. Tijms BM, Gobom J, Reus L, Jansen I, Hong S, Dobricic V, et al. Pathophysiological subtypes of Alzheimer's disease based on cerebrospinal fluid proteomics. Brain. 2020;143:3776–92. https://doi.org/10.1093/brain/awaa325.
44. Sharma K, Schmitt S, Bergner CG, Tyanova S, Kannaiyan N, Manrique-Hoyos N, et al. Cell type- and brain region-resolved mouse brain proteome. Nat Neurosci. 2015;18:1819–31. https://doi.org/10.1038/nn.4160.
45. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci. 2014;34:11929–47.
46. Zhou M, Haque RU, Dammer EB, Duong DM, Ping L, Johnson ECB, et al. Targeted mass spectrometry to quantify brain-derived cerebrospinal fluid

biomarkers in Alzheimer's disease Clin. Proteomics. 2020;17:19. https://doi.org/10.1186/s12014-020-09285-8.
47. Carnethon MR, Pu J, Howard G, Albert MA, Anderson CAM, Bertoni AG, et al. Cardiovascular health in African Americans: a scientific statement from the American Heart association. Circulation. 2017;136:e393–423. https://doi.org/10.1161/CIR.0000000000000534.
48. Hanseeuw BJ, Betensky RA, Jacobs HIL, Schultz AP, Sepulcre J, Becker JA, et al. Association of amyloid and tau with cognition in preclinical Alzheimer disease: a longitudinal study JAMA. Neurology. 2019;76:915–24. https://doi.org/10.1001/jamaneurol.2019.1424.
49. Visser PJ, Reus LM, Gobom J, Jansen I, Dicks E, van der Lee SJ, et al. Cerebrospinal fluid tau levels are associated with abnormal neuronal plasticity markers in Alzheimer's disease. Mol Neurodegener. 2022;17:27. https://doi.org/10.1186/s13024-022-00521-3.
50. Quinn JP, Kandigian SE, Trombetta BA, Arnold SE, Carlyle BC. VGF as a biomarker and therapeutic target in neurodegenerative and psychiatric diseases. Brain Commun. 2021;3(4):fcab261.
51. Wingo AP, Dammer EB, Breen MS, Logsdon BA, Duong DM, Troncosco JC, et al. Large-scale proteomic analysis of human brain identifies proteins associated with cognitive trajectory in advanced age. Nat Commun. 2019;10:1619. https://doi.org/10.1038/s41467-019-09613-z.
52. Libiger O, Shaw LM, Watson MH, Nairn AC, Umaña KL, Biarnes MC, et al. Longitudinal CSF proteomics identifies NPTX2 as a prognostic biomarker of Alzheimer's disease. Alzheimers Dement. 2021;17:1976–87.
53. Llano DA, Devanarayan P, Devanarayan V. CSF peptides from VGF and other markers enhance prediction of MCI to AD progression using the ATN framework. Neurobiol Aging. 2023;121:15–27. https://doi.org/10.1016/j.neurobiolaging.2022.07.015.
54. Xiao MF, Xu D, Craig MT, Pelkey KA, Chien CC, Shi Y, et al. NPTX2 and cognitive dysfunction in Alzheimer's Disease. eLife. 2017;6:e23798.
55. Sjaarda J, Gerstein HC, Kutalik Z, Mohammadi-Shemirani P, Pigeyre M, Hess S, et al. Influence of Genetic Ancestry on Human Serum Proteome. Am J Hum Genet. 2020;106:303–14. https://doi.org/10.1016/j.ajhg.2020.01.016.
56. Ghosh S, Nehme R, Barrett LE. Greater genetic diversity is needed in human pluripotent stem cell models. Nat Commun. 2022;13:7301.
57. Peterson RE, Kuchenbaecker K, Walters RK, Chen CY, Popejoy AB, Periyasamy S, et al. Genome-wide association studies in ancestrally diverse populations: opportunities. Methods, Pitfalls, and Recommendations. Cell. 2019;179:589–603. https://doi.org/10.1016/j.cell.2019.08.051.
58. Zhang J, Dutta D, Köttgen A, Tin A, Schlosser P, Grams ME, et al. Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. Nat Genet. 2022;54:593–602. https://doi.org/10.1038/s41588-022-01051-w.
59. Kachuri L, Mak ACY, Hu D, Eng C, Huntsman S, Elhawary JR, et al. Gene expression in African Americans, Puerto Ricans and Mexican Americans reveals ancestry-specific patterns of genetic architecture. Nat Genet. 2023. https://doi.org/10.1038/s41588-023-01377-z.
60. Robins C, Liu Y, Fan W, Duong DM, Meigs J, Harerimana NV, et al. Genetic control of the human brain proteome. Am J Hum Genet. 2021;108:400–10. https://doi.org/10.1016/j.ajhg.2021.01.012.
61. Stepler KE, Mahoney ER, Kofler J, Hohman TJ, Lopez OL, Robinson RAS. Inclusion of African American/Black adults in a pilot brain proteomics study of Alzheimer's disease. Neurobiol Dis. 2020;146:105129.
62. Desaire H, Stepler KE, Robinson R AS. Exposing the brain proteomic signatures of Alzheimer's disease in diverse racial groups: leveraging multiple data sets and machine learning. J Proteome Res. 2022;21:1095–104.
63. Hodes RJ, Buckholtz N. Accelerating medicines partnership: Alzheimer's Disease (AMP-AD) knowledge portal aids Alzheimer's drug discovery through open data sharing. Expert Opin Ther Targets. 2016;20:389–91.
64. Wesenhagen KEJ, Gobom J, Bos I, Vos SJB, Martinez-Lage P, Popp J, et al. Effects of age, amyloid, sex, and APOE ε4 on the CSF proteome in normal cognition. Alzheimers Dement (Amst). 2022;14:e12286.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

BMC