Large-scale proteome and metabolome analysis of CSF implicates altered glucose and carbon metabolism and succinylcarnitine in Alzheimer’s disease


1 Department of Population Health Sciences, University of Wisconsin–Madison, Madison, Wisconsin, USA
2 National Center for Quantitative Biology of Complex Systems, University of Wisconsin–Madison, Madison, Wisconsin, USA
3 Department of Biomolecular Chemistry, University of Wisconsin–Madison, Madison, Wisconsin, USA
4 Wisconsin Alzheimer’s Disease Research Center, University of Wisconsin–Madison, Madison, Wisconsin, USA
5 Department of Medicine, University of Wisconsin–Madison, Madison, Wisconsin, USA
6 Wisconsin Alzheimer’s Institute, University of Wisconsin–Madison, Madison, Wisconsin, USA
7 Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA
8 NeuroGenomics and Informatics Center, Washington University School of Medicine, St. Louis, Missouri, USA
9 Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, Missouri, USA
10 Roche Diagnostics GmbH, Penzberg, Germany
11 Roche Diagnostics International Ltd, Rotkreuz, Switzerland
12 William S. Middleton Memorial Veterans Hospital, Madison, Wisconsin, USA
13 Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden
14 Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden
15 Department of Neurodegenerative Disease, UCL Institute of Neurology, London, UK
16 UK Dementia Research Institute at UCL, London, UK
17 Hong Kong Center for Neurodegenerative Diseases, Hong Kong, China
18 Morgridge Institute for Research, Madison, Wisconsin, USA
19 Department of Chemistry, University of Wisconsin–Madison, Madison, Wisconsin, USA

Daniel J. Panyard and Justin McKetney contributed equally to this work.
Joshua J. Coon and Corinne D. Engelman jointly supervised this work.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. Alzheimer’s & Dementia published by Wiley Periodicals LLC on behalf of Alzheimer’s Association.
Abstract

INTRODUCTION: A hallmark of Alzheimer’s disease (AD) is the aggregation of proteins (amyloid beta [A] and hyperphosphorylated tau [T]) in the brain, making cerebrospinal fluid (CSF) proteins of particular interest.

METHODS: We conducted a CSF proteome-wide analysis among participants of varying AT pathology (n = 137 participants; 915 proteins) with nine CSF biomarkers of neurodegeneration and neuroinflammation.

RESULTS: We identified 61 proteins significantly associated with the AT category (P < 5.46 × 10^−5) and 636 significant protein-biomarker associations (P < 6.07 × 10^−6). Proteins from glucose and carbon metabolism pathways were enriched among amyloid- and tau-associated proteins, including malate dehydrogenase and aldolase A, whose associations with tau were replicated in an independent cohort (n = 717). CSF metabolomics identified and replicated an association of succinylcarnitine with phosphorylated tau and other biomarkers.

DISCUSSION: These results implicate glucose and carbon metabolic dysregulation and increased CSF succinylcarnitine levels with amyloid and tau pathology in AD.

KEYWORDS
acylcarnitines, Alzheimer’s disease, amyloid, biomarkers, carbon metabolism, glucose metabolism, metabolism, metabolomics, multiomics, neurodegeneration, neuroinflammation, proteomics, tau

HIGHLIGHTS
∙ Cerebrospinal fluid (CSF) proteome enriched for extracellular, neuronal, immune, and protein processing.
∙ Glucose/carbon metabolic pathways enriched among amyloid/tau-associated proteins.
∙ Key glucose/carbon metabolism protein associations independently replicated.
∙ CSF proteome outperformed other omics data in predicting amyloid/tau positivity.
∙ CSF metabolomics identified and replicated a succinylcarnitine–phosphorylated tau association.
1 | BACKGROUND

Despite much improvement in our understanding of it, Alzheimer’s disease (AD) continues to impose an enormous medical, social, and economic toll on society. An estimated 50 million people have dementia worldwide, and AD is the sixth leading cause of death in the United States, costing an estimated $290 billion annually for health care. Part of the reason for this global impact of AD has been the lack of a cure or effective therapy for the disease, which is driven in part by an incomplete understanding of its causal mechanisms. The core pathological features of AD are well described and center on the accumulation of two proteins, amyloid and tau, into amyloid plaques and neurofibrillary tangles, for which there are validated cerebrospinal fluid (CSF) biomarkers.

Recently, there has been a shift in the research conceptualization of the disease from a focus on clinical signs and symptoms to AD biology measured in vivo. Using CSF assays and neuroimaging related to amyloid deposition and hyperphosphorylation of tau protein, it has become possible to leverage these biomarkers for identifying preclinical AD, mild cognitive impairment (MCI), and AD dementia. In 2018, an explicit research framework for categorizing AD was proposed by the National Institute on Aging and Alzheimer’s Association (NIA-AA). This framework categorized individuals as amyloid positive (A+), tau positive (T+), and/or neurodegeneration positive (N+). This so-called ATN framework—using ATN-based categorizations rather than more traditional clinical diagnoses as outcomes—provided nosological clarity in studying AD and other forms of dementia.

The use of these biomarker-defined categories is most relevant in multomic approaches to studying AD pathophysiology, in which molecular pathways are investigated and clear case definitions are essential. Omics research offers immense promise for understanding complex disease by measuring millions of molecular features spanning the genome, proteome, metabolome, and beyond. Each of these individual omic approaches has already been applied extensively in AD. Genomics research has highlighted numerous loci, from the role of mutations in amyloid beta precursor protein (APP), presenilin 1 (PSEN1), and presenilin 2 (PSEN2) in early-onset familial AD to late-onset AD genetic risk factors like apolipoprotein E (APOE), complement C3b/C4b receptor 1 (Knops blood group) (CR1), and ATP binding cassette subfamily A member 7 (ABCA7). CSF metabolomics studies have identified alterations in cholesterol, sphingolipid, norepinephrine, and other pathways. In the CSF proteome, already known to include the amyloid and tau biomarkers for AD, studies have identified altered proteins related to the immune system and inflammation, carbohydrate metabolism, phospholipids, and the regulation of synapses.

Here, we combined the A and T of the ATN framework of AD with a novel CSF proteomics data set comprising 915 proteins generated for 137 participants, building on our recently published study results in an independent sample. We comprehensively profiled the AD CSF proteome, its relationship to the AT category, and its association with a diverse set of nine AD CSF biomarkers covering measures of amyloid, tau, neurodegeneration, and neuroinflammation (Figure 1). These results were then extensively interrogated for pathway-level and network-based patterns, with top findings replicated in an independent AD proteomics cohort with an alternative proteomics modality. The top implicated biological pathway was then further investigated with a focused metabolomics analysis using the same original participants and samples as well as an independent metabolomics replication cohort of 363 participants. Finally, we combined the proteomics data set with previously generated genome-wide genotypes, 390 CSF metabolome-wide metabolites, and demographic information to examine the relative contributions of different omics data sets to AT-based categories to better understand the multomic landscape of AD. Elucidating the pathophysiology leading to the development of AD pathology and symptoms will help inform the identification of novel drug targets and guide future large-scale multomics research in the field.

2 | METHODS

2.1 | Experimental design

The primary data in this study came from two longitudinal AD cohorts of middle- and older-aged adults: the Wisconsin Registry...
for Alzheimer’s Prevention (WRAP)\textsuperscript{26} and the Wisconsin Alzheimer’s Disease Research Center (WI ADRC),\textsuperscript{27} see Table 1 and Figure 1. Briefly, WRAP includes participants enriched for a parental history of AD dementia who were largely between the ages of 40 and 65 at the time of enrollment, fluent in English, able to perform neuropsychological testing, without a diagnosis or evidence of dementia at baseline, and without any health conditions that might prevent participation in the study. The WI ADRC study includes participants from one of several subgroups: mild late-onset AD, MCI, cognitively unimpaired (CU) middle-aged adults enriched for a parental history of presumed AD dementia, and age-matched healthy older controls (age > 65). Briefly, the WI ADRC participants were > 45 years of age, with decisional capacity, and without a history of certain medical conditions (like congestive heart failure or major neurologic disorders other than dementia) or any contraindication to biomarker procedures. Participants in both the WRAP and WI ADRC cohorts were given diagnoses of AD, MCI, CU, and others that were reviewed by a consensus review committee that included dementia-specialist physicians, neuropsychologists, and nurse practitioners.\textsuperscript{25} The National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRD\textsuperscript{6} and NIA-\textsuperscript{7} criteria were used in defining the clinical diagnoses without reference to the participants’ CSF biomarker status. This study used the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) cohort reporting guidelines\textsuperscript{28} and was performed as part of the Gene Rations Of WRAP (GROW) study, which was approved by the University of Wisconsin Health Sciences Institutional Review Board. Participants in the WRAP and WI ADRC studies provided written informed consent.

\subsection*{2.2 CSF biomarkers}

The CSF samples used for the biomarker analyses were acquired from lumbar punctures (LPs) using a uniform preanalytical protocol between 2010 and 2018 as previously described.\textsuperscript{29} Samples were collected in the morning using a Sprotte 24- or 25-gauge atraumatic spinal needle, and 22 mL of fluid was collected via gentle extraction into polypropylene syringes and combined into a single 30 mL polypropylene tube. After gentle mixing, samples were centrifuged to remove red blood cells and other debris. Then, 0.5 mL CSF was aliquoted into 1.5 mL polypropylene tubes and stored at −80 °C within 30 minutes of collection.

All CSF samples were assayed between March 2019 and January 2020 at the Clinical Neurochemistry Laboratory at the University of Gothenburg. CSF biomarkers were assayed using the NeuroToolKit (NTK; Roche Diagnostics International Ltd, Rotkreuz, Switzerland), a panel of automated Elecsys\textsuperscript{®} and robust prototype immunonasays designed to generate reliable biomarker data that can be compared across cohorts. Measurements with the following immunonasays were performed on a cobas e 601 analyzer (Roche Diagnostics International Ltd, Rotkreuz, Switzerland): Elecsys β-amyloid (1–42) CSF (A\textsubscript{β42}), Elecsys Phospho-Tau (181P) CSF (phosphorylated tau [p-tau]), and Elecsys Total-Tau CSF, β-amyloid (1–40) CSF (A\textsubscript{β40}), and interleukin-6 (IL-6). The remaining NTK panel was assayed on a cobas e 411 analyzer (Roche Diagnostics International Ltd, Rotkreuz, Switzerland), including markers of synaptic damage and neuronal degeneration (neurogranin, neurofilament light protein [NFL], and alpha-synuclein) and markers of glial activation (chitinase-3-like protein 1 [YKL-40] and soluble triggering receptor expressed on myeloid cells 2 [sTREM2]).

A total of nine established CSF biomarkers for AD were analyzed in this study: the amyloid beta A\textsubscript{β42}/A\textsubscript{β40} ratio, p-tau, the p-tau/A\textsubscript{β42} ratio, NFL, alpha-synuclein, neurogranin, YKL-40, sTREM2, and IL-6. Because the CSF biomarker measurements were to be used as outcomes, each biomarker was assessed for skewness using the skewness function of the R package moments (version 0.14).\textsuperscript{30} Any biomarker with a skewness ≥ 2 was transformed with a log\textsubscript{10} transformation to better meet the normality assumption of regression. The outcomes that were log\textsubscript{10} transformed were NFL and IL-6.

Samples used in this study were then assigned to pathological categories from the NIA-\textsuperscript{AA} ATN research framework,\textsuperscript{11} using binary cut-offs for CSF amyloid and tau positivity. The development of these research cut-offs is described in detail elsewhere.\textsuperscript{29} Briefly, cut-offs were estimated via receiver operating characteristic (ROC) analysis on a subsample of n = 185 participants (cognitively impaired and unimpaired) who underwent [11C] Pittsburgh compound B positron emission tomography (PiB-PET) imaging within 2 years of an LP. Using the MATLAB perfcurve function\textsuperscript{31} with an equally weighted cost function,\textsuperscript{32} the optimal A\textsubscript{β42}/A\textsubscript{β40} threshold was 0.046 and the optimal p-tau/A\textsubscript{β42} threshold was 0.038. Thresholds for p-tau\textsubscript{181} were determined by establishing a reference group of 223 CSF amyloid (A\textsubscript{β42}/A\textsubscript{β40}) negative, CU younger participants (ages 40–60 years). Biomarker positivity thresholds for these analytes were set at +2 standard deviations (SDs) above the mean of this reference group (p-tau threshold = 24.8 pg/mL). In this study, A+ and T+ were defined based on the CSF A\textsubscript{β42}/A\textsubscript{β40} and p-tau thresholds, respectively. The final pathological categories for this study included amyloid negative and tau negative (A−T−), amyloid positive and tau negative (A+T−), and amyloid positive and tau positive (A+T+). The fourth possible category of amyloid negative and tau positive (A−T+) was not included in the proteomics discovery analyses as these samples are considered to represent non-AD pathological change;\textsuperscript{11} out of 285 participants who were eligible at this point for inclusion in this study, 15 were excluded for being A−T+.

\subsection*{2.3 CSF metabolomics}

All samples for which we generated CSF proteomics data also had CSF metabolomics data available that had been generated in previous work. The details of the CSF sample collection, handling, and metabolomics profiling have been previously described.\textsuperscript{33,34} Briefly, fasting CSF samples were drawn from study participants in the morning through LP and then mixed, centrifuged, aliquoted, and stored at −80 °C. Samples were kept frozen until they were shipped overnight to Metabolon, Inc. (Durham, NC), which similarly kept samples frozen at −80 °C until...
Proteomics sample selection

Protein extraction and digestion

An analysis. Metabolon used ultrahigh-performance liquid chromatography tandem mass spectrometry (UPLC-MS/MS) to conduct an untargeted metabolomics analysis of the CSF samples. The metabolites were then annotated with metabolite identifiers, chemical properties, and pathway information. Metabolite measurements were divided by the median measurement for that metabolite across all samples. Missing values for xenobiotic metabolites were imputed to 0.0001, while missing values for non-xenobiotic metabolites were imputed to half of the minimum value among all other measured samples for that metabolite.

The initial data set contained 412 metabolites from 1172 CSF samples across 687 unique individuals. A total of 13 metabolites that were missing for ≥ 50% of samples were removed. One sample was removed for missing ≥ 40% of metabolite values. A total of nine metabolites with low variance (interquartile range = 0) were then removed. A log₁₀ transformation was applied to all metabolite values. A total of 220 samples were excluded from analysis. The processed data set contained 390 metabolites quantified on 951 CSF samples from 609 unique individuals, including all but one of the CSF samples on which proteomics data were generated for this study.

2.4 Genome-wide genotyping

Genome-wide genotypes were also available for all samples in this study. The genotyping in both the WRAP and WI ADRC had been previously conducted. For the WRAP cohort, DNA from whole blood samples were genotyped with the Illumina Multi-Ethnic Genotyping Array at the University of Wisconsin Biotechnology Center. Pre-imputation quality control (QC) steps included removing samples and variants with a high missingness (> 5%) or inconsistent genetic and self-reported sex. Samples from individuals of European descent were imputed using the Michigan Imputation Server and the Haplotype Reference Consortium (HRC) reference panel. Variants with poor quality (R² < 0.8) or out of Hardy–Weinberg equilibrium (HWE) were removed after imputation, leaving a total of 1198 samples with 10,499,994 single nucleotide polymorphisms (SNPs). In the WI ADRC, whole blood samples were genotyped by the Alzheimer’s Disease Genetics Consortium (ADGC) at the National Alzheimer’s Coordinating Center (NACC) using the Illumina HumanOmniExpress-12v1.A, Infinium HumanOmniExpressExome-8 v1-2a, or Infinium Global Screening Array v1-0 (GSAMD-24v1-0_20011747_A1) BeadChip assay. Initial QC was conducted on each chip’s data separately, removing variants or samples with high missingness (> 1%), out of HWE (P ≤ 1 × 10⁻⁶), or with inconsistent genetic and self-reported sex. The remaining samples were then imputed with the Michigan Imputation Server, phased using Eagle2, and imputed to the HRC reference panel. As before, variants of low quality (R² < 0.8) or out of HWE were removed. The data sets from the different chips were then merged, leaving a data set with 377 samples of European descent and 7,049,703 SNPs. The WRAP and WI ADRC data sets were then harmonized to each other and to the 1000 Genomes Utah residents with Northern and Western European ancestry (CEU) data set, using the GRCh37 genome build. Ambiguous SNPs were removed, and the remaining SNPs were aligned to the same strand and allele orientations as the WI ADRC data set.

The 137 samples from this study were then extracted from this combined genetic data set and further processed using PLINK (v1.90b6.3). To ensure sufficient data were available for use in the prediction models, only SNPs with no missing data and with a minor allele count ≥ 20 among the 137 samples were retained. Linkage disequilibrium (LD) pruning was then applied using a window size of 1000 kb, an R² threshold of 0.1, and the 1000 Genomes CEU samples as the reference data set. The pruning resulted in a data set of 38,652 SNPs.

2.5 APOE genotyping

Each sample was additionally assigned an APOE genotype based on the participant’s combination of the e2, e3, and e4 alleles for APOE from a separate set of genotyping. DNA was extracted from whole blood samples, which was then genotyped for the APOE alleles using competitive allele-specific polymerase chain reaction–based KASP genotyping for rs429358 and rs7412.

2.6 Proteomics sample selection

Based on the results of our pilot study for CSF proteomics, we had estimated that a sample of ~150 would be sufficient for 80% power to detect most of the observed protein–AD diagnosis associations from the original matched case-control analyses in the pilot using the R package pwr (version 1.3-0), though we note that the final study design used here differed from the pilot in that three participant groups were used and age and sex were controlled for in the analyses instead of using a matched design (Figure S1 in supporting information). The process of selecting samples for CSF proteomics generation began by considering all CSF samples from fasted, successful LPs (n = 1440) from 823 unique participants across WRAP and WI ADRC. From there, each CSF sample was matched to its closest set of CSF biomarker data, CSF metabolomics data, and consensus conference diagnosis samples. Samples were excluded if there was insufficient material for proteomics analysis, if they were part of a clinical trial, or if they had been used already in our pilot study. To simplify the downstream analyses, only one sample (the most recent) per participant was considered when there were multiple samples. An approximately equal number of samples per AT-defined subgroup (A–T–, A+T–, A+T+) was selected, prioritizing samples with available genomic data and metabolomic data. A total of 140 samples were selected to have proteomics data generated.

2.7 Protein extraction and digestion

CSF protein concentration was determined by protein bicinchoninic acid (BCA) assay (Thermo Scientific). CSF aliquots were moved to 96-well plates and dried down using a SpeedVac Concentrator.
To increase proteomic depth and protein identifications, offline chromatographic fractionation of a set of pooled representative samples was performed. Pooled samples for each of the three disease groups were created by combining 10 μL of CSF from each sample in that disease group. These three pooled samples were then prepared using the extraction and digestion protocol described above. The three desalted, digested peptide solutions were then fractionated using high-pH reverse-phase liquid chromatography. Separation was performed using an Agilent Infinity 2000 HPLC with a 150 mm C18 reverse-phase column (Waters, XBridge Peptide BEH, particle size 3.5 μm). Mobile phase buffer A was a freshly prepared mixture 10 mM ammonium formate pH 9.5, and mobile phase buffer B was a freshly prepared mixture of 80% MeOH, 10 mM ammonium formate pH 9.5. The gradient method was 20 minutes in length with fractions collected from minute 5 to minute 20, with a flow rate of 800 nL/min across the entire method. The method initiated with a concentration of 5% B before increasing to 35% by minute 2. Percent B increased to 100% by 13 minutes. From 5 to 20 minutes, 32 fractions were collected in before increasing to 35% by minute 2. Percent B then increased to 49% B by 59 minutes before a wash step of 100% B from 62 to 67 minutes. The method finished with an equilibration step from minute 68 to 78 of 0% B.

2.10 Tandem mass spectrometry

Peptides eluting from the column were ionized by electrospray ionization and analyzed using a Thermo Orbitrap Eclipse hybrid mass spectrometer in a data-dependent manner (DDA). Survey scans were collected in the Orbitrap at a resolution of 240,000 with a normalized automatic gain control (AGC) target of 250% (1e5) with Advanced Precursor Determination engaged across the range of 300 to 1400 m/z. Precursors were isolated for tandem MS scans using a window of 0.5 m/z, with a dynamic exclusion duration of 22 seconds and a mass tolerance of 15 ppm. Precursors were dissociated using higher-energy collisional dissociation (HCD) with a normalized collision energy of 25%. Tandem scans were taken over the range 130 to 1350 m/z using the “rapid” setting with a normalized AGC target of 300% (3e4) and a maximum injection time of 18 milliseconds. Data for each participant sample was collected in technical duplicate to account for technical variability. Although we have not directly quantified that variability here, in preliminary experiments for our pilot analysis, we examined preparation replicates and found relatively strong correlation of protein abundances. This finding suggested that technical variability of injection replicates on the mass spectrometer should be quite small for most quantified proteins.

The resulting raw data files were searched in MaxQuant43,44 using fast label-free quantification (LFQ) and a full human proteome with isofoms downloaded from UniProt (downloaded June 14, 2017). Oxidation of methionine and acetylation of the N terminus were allowed as variable modifications, and carbamidomethylation of cysteine was set as a fixed modification. Proteins were searched using a false discovery rate (FDR) of 1% with a minimum peptide length of 7 and a 0.5 Da MS/MS match tolerance. Matching between runs was used, applied with a retention time window of 0.7 minutes. Protein abundance data were extracted in the form of LFQ Intensity from the “proteinGroups.txt” output file. Throughout this article, each protein group is referred to by the first listed majority protein from its annotation from MaxQuant. The protein data were annotated with Entrez IDs (via R package org.Hs.eg.db, version 3.11.4), UniProt IDs, and gene information (GENCODE, version 37, and the HUGO Gene Nomenclature Committee, HGNC, database). When the gene annotations conflicted or were absent from one of these databases for a given UniProt ID, the gene identifiers were taken in the order of resources listed. Although the LC-MS analysis was applied to individual samples as well as the three pooled and fractionated samples, only the
individual samples were used for quantitative analysis and statistical investigations described hereafter.

### 2.11 Proteomics quality control

After removing several samples with injection or other technical issues, the proteomics data set included 2040 proteins across 137 samples. These data underwent a strict QC pipeline: proteins that were missing for 33% or more of samples (either overall or within an AT category) were removed; samples missing ≥33% of proteins were removed; and proteins with an interquartile range of 0 were removed (Figure S2, Figure S3 in supporting information). The thresholds of 33% were based on the observed distributions of protein and sample missingness, chosen to balance excluding proteins or samples with a substantial portion of data missing that might have been due to technical issues versus retaining proteins that were missing only a few values that were below the detection limit. A total of 137 samples with 915 proteins remained (Table S1 in supporting information). The LFQ values for each protein were then log2-transformed. The remaining missing values were then randomly imputed based on a normal distribution derived from the lower end of the observed values for that protein (the observed distribution mean was shifted by −1.8 and the SD shrunk by a factor of 0.3), with the expectation that individuals missing a value for one of the remaining proteins in the data set were likely missing the value due to the true value being below the limit of detection (Figure S4 in supporting information). This imputation was performed separately within each AT category. Finally, each protein was Z-score transformed.

### 2.12 Proteomics descriptive analysis

The first step in understanding how the CSF proteome changes in AD is to understand what its contents are and how they compare to the entirety of the human proteome. Thus, our first main objective was to extensively profile the set of proteins quantified in the CSF in this cohort (Table S2 in supporting information). The pairwise correlation of all proteins was calculated (nominally significant results with correlation ρ < 0.05 shown in Table S3 in supporting information) and then visualized with a heatmap with hierarchical clustering to show the underlying patterns of covariation (R package ComplexHeatmap, version 2.4.3; Figure 2A). The structure was further examined with a principal components analysis (PCA), scree plot (Figure S5 in supporting information), and plot of the first two PCs by AT category (R package factoextra, version 1.0.7; Figure S6 in supporting information) to assess the presence of independent signals among the CSF proteome and their relationship to the AT categories. The associations of the top two PCs with age were also examined with a correlation analysis. A pathway analysis was then performed to examine the differences between the set of proteins quantified in the CSF and the overall human proteome. The enrichment of Gene Ontology (GO) terms, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, DisGeNET, and Disease Ontology (DO) gene sets among the CSF proteome against the entire human proteome (with Benjamini–Hochberg adjusted P value < 0.05) was calculated using the R packages clusterProfiler and ReactomePA (version 1.32.0; Figure 2B; Table S4 in supporting information). To visualize the enriched DO terms, the enriched DO terms and any related proteins according to DO were plotted in a network using the R package tidygraph and the Kamada–Kawai algorithm. A community analysis was performed (cluster_fast_greedy method) to group and color the DO terms, with each community labeled with the top-associated DO term within it (according to the Benjamini–Hochberg adjusted P value; Figure 3). To summarize the major constituents of the observed CSF proteome in these cohorts, the presence of clusters was then assessed with Gaussian mixture modeling using the R package mclust (version 5.4.6). The number of clusters (3) was chosen based on the elbow of the plot of the Bayesian information criterion (BIC; Figure S7 in supporting information). Enrichment of gene set ontologies across the clusters was repeated with the GO, KEGG, and DO sets (using the entire human proteome as the comparison and the same Benjamini–Hochberg adjusted P value threshold of 0.05) and plotted (Figure 2C; Table S5 in supporting information).

### 2.13 Protein–AT category associations

The second main objective of this study was the identification of differentially expressed proteins across the three AT categories in order to understand how the CSF proteome changes across the AD trajectory. This analysis was performed using an analysis of covariance (ANCOVA) model comparing each protein across the three groups, controlling for age at LP and sex (Table S6 in supporting information). Additionally, to test the difference in protein level between each pair of AT categories, a logistic regression model was used for each pair of categories (controlling for age at LP and sex; Table S6). A Bonferroni correction for the number of proteins tested (P = 0.05 / 915 = 5.46 × 10⁻⁵) was used for reporting significant results. The distributions of the top-associated proteins across the AT spectrum were plotted (Figure 4A). To assess whether signal enrichment was likely due to an artifact, the ANCOVA analyses were repeated with randomly permuted AT category labels. A quantile–quantile (Q–Q) plot was generated to assess the presence of signal enrichment across the proteome for AT-related differences and to compare the permuted and non-permuted analyses (Figure 4B). Because the APOE gene is known to have a significant effect on AD risk, we examined whether APOE genotype was driving the observed AT–protein associations. The ANCOVA analyses were repeated but with the count of APOE ε4 alleles included as an additional covariate along with age at LP and sex. The same Bonferroni correction was used as before. The resulting AT-associated proteins were compared to the results from the original ANCOVA analyses (Figure S8 in supporting information). The set of associated proteins from the non-permuted analysis was then assessed for enriched GO, KEGG, and DO...
gene sets, using the set of 915 proteins observed in this study as the comparison set of proteins and using a significance threshold of the Benjamini–Hochberg P value < 0.05; Figure 4C, Table S7 in supporting information).

To examine the direction of effect of each protein, a logistic regression was performed with A+T+ (vs. A−T−) as the outcome and a protein as the main predictor, controlling for age at LP and sex and using the same Bonferroni threshold for significance as the ANCOVA analyses. The sample size for the logistic regression was smaller (n = 98) due to the exclusion of the A+T− samples. The overlap between the set of significantly associated proteins and the set of significantly associated proteins from the ANCOVA analysis was displayed in a Venn diagram (R package ggVennDiagram,61 version 1.0.7; Figure S9 in supporting information), and the odds ratios were presented in a volcano plot (Figure S10 in supporting information).

2.14 Protein–CSF biomarker associations

While understanding the relationships of the proteome with amyloid and tau status can provide a high-level view of the changing AD proteome, it is useful to investigate more nuanced connections with markers of neurodegeneration and neuroinflammation to help identify proteins associated with different pathological processes. To this end, our third main objective was conducting a comprehensive set of protein–CSF biomarker analyses using the NTK panel described above. Each protein was tested for association with each of the nine CSF biomarkers (the Aβ42/Aβ40 ratio, p-tau, the p-tau/Aβ42 ratio, NFL, alpha-synuclein, neurogranin, YKL-40, sTREM2, and IL-6). Linear regression models were used to regress each CSF biomarker on each protein, controlling for age at CSF sample and sex and using a Bonferroni correction for the total number of tests (9 × 915 = 8234 tests; P = 0.05 / 8234 = 6.07 × 10⁻⁸⁶; Table S8 in supporting information). For the sake of interpretation, effect estimates where the biomarker outcome was additionally Z-score normalized were also provided (“Effect size [std. biomarker]” columns). The results were summarized with a Q–Q plot showing the signal enrichment for each biomarker along with a sensitivity analysis in which the regression models were repeated with the biomarker values randomly permuted to test the robustness of each biomarker’s signal enrichment (Figure S11 in supporting information). The cross-biomarker relationships among the significantly associated proteins were then visualized as a bipartite graph using the R package tidygraph (version 1.2.0)58 and among the significantly associated proteins were then visualized as a S11 in supporting information). The cross-biomarker relationships provided (“Effect size [std. biomarker]” columns). The results were Bonferroni adjusted P < 0.05; Table S9 in supporting information). An additional enrichment test using the same procedure and threshold was performed using just the set of proteins associated with two or more of p-tau, neurogranin, and alpha-synuclein (Table S10 in supporting information).

2.15 Replication of top pathway results in other AD proteomics cohorts

A replication data set from the Knight ADRC was used to validate findings from the main analyses. The Knight AD data set included samples from CSF (n = 717) and plasma (n = 490). The recruited individuals from the Knight ADRC cohort were evaluated by Clinical Core personnel of Washington University. For individuals with CSF and plasma data, cases received a clinical diagnosis of AD in accordance with standard criteria, and AD severity was determined using the Clinical Dementia Rating (CDR) scale65 at the time of LP (for CSF samples) or blood draw (for plasma samples; in contrast to the WI ADRC, the Knight ADRC does not categorize participants as MCI). Controls received the same assessment as the cases but were non-demented (CDR = 0). As there are myriad pathologies and disease subtypes in clinically diagnosed individuals, our analysis excluded any subjects that had other neurodegenerative diseases and not AD based on the last clinical and biomarker assessment. CSF was collected by LP after overnight fasting, centrifuged, and frozen at −80°C as described previously.66–68 Blood was collected at the time of LP, and serum or plasma was obtained by centrifugation and stored at −80°C. CSF samples were analyzed by immunoassay for β-amyloid 1-42 (Aβ42), total tau, and tau phosphorylated at threonine 181 (p-tau) (INNOTEST, Fujirebio, Ghent, Belgium). The institutional review boards of Washington University School of Medicine in St. Louis approved the study; research was performed in accordance with the approved protocols and participants provided informed consent.

For deep proteomics characterization in the CSF and plasma tissues, the levels of 1305 proteins were quantified using a different methodological approach from that used for the Wisconsin cohorts: the SOMAscan assay, a multiplexed, aptamer-based platform.69 The assay covers a dynamic range of 10⁸ and measures all three major categories: secreted, membrane, and intracellular proteins. The proteins cover a wide range of molecular functions and include proteins known to be relevant to human disease. As previously described by Gold et al.,69 modified single-stranded DNA aptamers are used to bind specific protein targets that are then quantified by a DNA microarray. Protein concentrations are quantified as relative fluorescent units (RFU). Aliquots of 150 μL of tissue were sent to the Genome Technology Access Center at Washington University in St. Louis for protein measurement.

Quality control was performed at the sample and aptamer levels using control aptamers (positive and negative controls) and calibrator samples. At the sample level, hybridization controls on each plate were used to correct for systematic variability in hybridization. The median signal over all aptamers was used to correct for within-run technical
variability. This median signal was assigned to different dilution sets within each tissue. For CSF samples, a 20% dilution rate was used. For plasma samples, three different dilution sets (40%, 1%, and 0.005%) were used.

As described in detail in Yang et al., additional QC was performed by identifying and removing protein and analyte outliers by applying four criteria: (1) Minimum detection filtering. If the analyte for a given sample was less than the limit of detection (LOD), the sample was deemed an outlier. Collectively, if the number of outliers given an analyte was less than 15% of the total sample size, the analyte was kept. (2) Flagging analytes based on the scale factor difference. (3) Coefficient of variation (CV) of calibrators lower than 0.15, in which the CV for each aptamer was calculated as the SD divided by the mean of each calibrator at the raw protein level. (4) Interquartile range (IQR) strategy. Outliers were identified if the subject was located 1.5-fold of the IQR outside of either end of the distribution given the log10-transformation of the protein level. Analytes were kept after passing all the criteria above for the downstream statistical analysis. An orthogonal approach was used to call subject outliers based on IQR. After this second removal of analytes, subject outliers were examined and removed again.

To obtain the proteomic signatures of sporadic AD status, CSF Aβ42, p-tau/Aβ42, and p-tau, differential abundance analysis was performed using linear regression of the log-transformed protein levels. In each tissue, we performed surrogate variable analysis while protecting status and age to correct for unmeasured heterogeneity. Age at death or at measurement, sex, and the resulting surrogate variables were included as covariates. A total of 694 CSF samples with both proteomics and amyloid and tau measures were used for the analysis, including all combinations of amyloid and tau positivity (Table S11 in supporting information).

The Knight ADRC analyses were used as a replication data set for the top findings from the main analyses performed in the WRAP/WI ADRC data set, focusing on the significantly associated proteins from the top implicated biological pathway from the protein–AT category and protein–biomarker analyses. The associations of these proteins were compared to the results from the Knight ADRC association tests conducted in the CSF and plasma to see if their associations and directions of effect were replicated using a significance threshold corrected for the number of tested proteins in the replication (P = 0.05 / 9 = 0.0056; Table 3; Table S12 in supporting information).

To assess replication in studies using a more similar proteomics methodology to what was used for discovery analysis here, the top findings were compared to the published results from Higginbotham et al. and Johnson et al. Both of these data sets used a MS-based approach for proteomics and reported differential protein expression results based on clinical diagnoses. Higginbotham et al. reported differences in CSF protein levels between AD and control participants (n = 20) while Johnson et al. reported differences in brain protein levels among healthy controls, asymptomatic AD, and AD participants (n = 453). A nominal P value threshold (P = 0.05) was used for replication.

### 2.16 Secondary analysis of diagnosis-associated proteins

While this study was designed to analyze the relationship of proteins with amyloid and tau categories, the association of proteins with clinical diagnosis is also of interest. We repeated the ANCOVA analysis to detect proteins associated with clinical diagnosis (AD, MCI, or CUI) instead of AT group, with age at sample and sex again as covariates. One sample was excluded for having MCI not presumed to be AD, leaving 136 samples for analysis. A Bonferroni correction for the number of proteins tested (P = 0.05 / 915 = 5.46 × 10−5) was applied as before to identify significantly associated proteins, with the top-associated proteins visualized by box plot (Table S13, Figure S13 in supporting information). The results were compared to the proteins associated with AT group, though the results were interpreted with caution because the study was not initially designed to answer this question and the distribution of clinical diagnosis was imbalanced across the three diagnoses.

### 2.17 Secondary analysis of insulin-related proteins

Based on the results of the AT category and biomarker associations and the pathway analysis suggesting a relationship with glucose metabolism (described below), the set of proteins excluded during the QC process due to low sample size was examined for proteins related to insulin signaling pathways, including any of the glucose transporter (GLUT) proteins (SLC2A family), insulin (INS), insulin receptor (INSR), insulin-like growth factor 1 (IGF1), IGF-1 receptor (IGF1R), insulin receptor substrate 1 (IRS1), insulin receptor substrate 2 (IRS2), phosphoinositide 3-kinase (PI3K), RAC-alpha serine/threonine-protein kinase (AKT1), mechanistic target of rapamycin (mTOR), and glycogen synthase kinase 3 (GSK3A). Proteins that failed the missingness threshold of 33% but were present for 50% or more of samples were investigated further but without the use of imputed data points. The relationships between the proteins and AT category (Figure S14 in supporting information) and the CSF biomarkers (Figure S15 in supporting information) were plotted, with ANCOVA and linear regression analyses to test for association between the proteins and AT category and the CSF biomarkers performed as previously with age at sample and sex as covariates (Table S14 in supporting information).

### 2.18 Secondary analysis of glycolysis and tricarboxylic acid cycle metabolites

To see whether the implicated pathways observed from the proteomics analyses were also implicated within the metabolomics data, a second form of validation of the main proteomics findings around glucose metabolism was performed using the CSF metabolomics data available in the WRAP and WI ADRC cohorts (described above). Focusing
on the 10 available metabolites from the “glycolysis, gluconeogenesis, and pyruvate metabolism” and “tricarboxylic acid cycle (TCA)” super-pathways (namely, 1.5-anhydroglucitol [1.5-AG], alpha-ketoglutarate, citrate, glucose, glycercate, isocitrate, lactate, malate, pyruvate, and succinylcarnitine [C4-DC]; Table S15 in supporting information), the same AT category and CSF biomarker association analyses for the proteomics data were performed again (using age at sample and sex as covariates) but with these 10 metabolites instead of the proteins. The discovery phase of this analysis was performed using the 136 participant CSF samples from the proteomics analysis that also had CSF metabolomic data available from the same matched CSF samples. The results of the analysis were subjected to a Bonferroni-corrected $P$ value threshold based on the number of metabolites tested ($P = 0.05$ / 10 metabolites = 0.005 for the metabolite–AT category association analyses; $P = 0.05$ / 10 metabolites / 9 biomarkers = 0.00056 for the metabolite–biomarker association analyses). Significantly associated metabolites were visualized with box plots (Figure S16A, Table S16, Table S17 in supporting information).

Because the group of participants in the WRAP and WI ADRC cohorts with CSF proteomics data generated here was only a subset of all of the participants with previously generated CSF metabolomics data in these cohorts, there were an additional 363 unique participants with CSF metabolomics data whose samples were not included in the main proteomics work here (Table S18 in supporting information). These 363 participants’ CSF metabolomics data were used as an independent replication data set for this secondary metabolomics analysis. The same AT category and CSF biomarker association analyses were repeated, using the same covariates and Bonferroni-corrected $P$ value thresholds as in the metabolomics discovery analyses (Figure S16B, Table S16, Table S17).

### 2.19 Multomic prediction of amyloid and tau

With the conceptual shift toward defining AD with amyloid and tau biomarkers in a research context, understanding exactly what other aspects of biology correspond to amyloid and tau, especially in the CSF, is critical, but the relative connections between the different biological omes and amyloid and tau is unclear. Our fourth main objective was to investigate these relationships. We conducted a separate and joint predictive analysis of amyloid and tau categories using CSF proteomics, CSF metabolomics, genomics, and demographic information. The CSF proteomic data set was combined with the CSF metabolomic, genomic, and demographic (age at sample and sex) data sets. After the QC steps described previously for each ome, 136 of the 137 CSF samples had values for all of the multiomic features (915 proteins, 390 metabolites, 38,652 SNPs coded as dosages plus the APOE ε4 allele count, and 2 demographic features [one sample was excluded for lacking CSF metabolomics data]). This multiomic data set was then used to predict different biomarker positivity states: $A_\beta 42/A_\beta 40$-positive, p-tau-positive, and p-tau/$A_\beta 42$-positive. Each ome (CSF proteome, CSF metabolome, genome, and demographics) was used individually along with a fifth multiomics predictor set (comprising all omes) to predict each outcome with an elastic net model (R package glmnet, version 4.0-2; alpha parameter = 0.5). When different CSF-based measurements were used for an individual (e.g., MS-derived CSF protein level, CSF biomarkers from the NTK platform, CSF metabolite levels, etc.), those measurements were all performed on or refer to the same underlying CSF sample; there was no sample date discrepancy between CSF measurements for a given participant.

For each biomarker and predictor pair, the procedure was the same. First, one third of the data were held out as a testing set and the remaining two thirds used as the training set. Within the training set data, 100-iteration, 3-fold cross-validation was used to select the best lambda value (11 possible values ranging from $10^{-5}$ to 1) according to the area under the curve (AUC) using the tidymodels R package (version 0.13). The best-performing model was then run on the entire training data set using the chosen lambda and used to predict the outcome on the held-out testing data set. The performances of the different omic models were then compared with ROC curves and 2D histograms showing the raw biomarker levels against the predicted classifications for each biomarker for each subject (Figure 6). The mean model metrics across each of the 4000 folds were calculated (Table S19 in supporting information).

### 3 RESULTS

#### 3.1 Sample summary

CSF samples from 137 WRAP and WI ADRC participants were selected as described in the Methods, roughly evenly distributed across the three AT categories of interest (Table 1, Figure 1; see Figure S1 for power analysis results based on pilot study). Most (102, 74.5%) of the participants were CU at the time of the sample, with 16 (11.7%) and 19 (13.9%) participants having an MCI or AD dementia diagnosis, respectively. The age and sex distributions across the AT categories varied, with worse AT pathology having a higher average participant age and a greater proportion of males. The amyloid and tau measures reflected the AT categorizations as expected. The remaining CSF biomarkers showed a general increase with increasing AT pathology with the exception of IL-6, which fluctuated across the groups.

#### 3.2 CSF proteomics descriptive analyses

The first major objective was to extensively profile the contents of the CSF proteome. The nLC-MS/MS analysis using DDA, MaxQuant identification, and LFQ quantification generated a total of 2040 protein groups across the participants. After the proteomics quality control steps (Table S1, Figure S2, Figure S3, Figure S4), 915 proteins remained in the WRAP/WI ADRC data set (Table S2). Included in these proteins were YKL-40 (correlation with immunoassay measurement $= 0.352, P = 2.40 \times 10^{-5}$), sTREM2 (correlation with immunoassay measurement of $sTREM2 = 0.490, P = 1.26 \times 10^{-5}$), APOE, APP (correlation with $A_\beta 42 = 0.136, P = 0.114$, amyloid-like protein-1 (APLP1), and...
### TABLE 1 Summary of sample demographics and CSF biomarkers.

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>A-T- (n = 137)</th>
<th>A-T+ (n = 56)</th>
<th>A+T (n = 42)</th>
<th>A+T+ (n = 42)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants (N, %)</td>
<td>137</td>
<td>56 (40.9)</td>
<td>39 (28.5)</td>
<td>42 (30.7)</td>
<td></td>
</tr>
<tr>
<td>Age at CSF sample (mean, SD)</td>
<td>66.1 (8.3)</td>
<td>61.8 (6.7)</td>
<td>67.6 (7.2)</td>
<td>70.6 (8.5)</td>
<td></td>
</tr>
<tr>
<td>Female (N, %)</td>
<td>82 (59.9)</td>
<td>39 (69.6)</td>
<td>23 (59)</td>
<td>20 (47.6)</td>
<td></td>
</tr>
<tr>
<td>AD dementia (N, %)</td>
<td>19 (13.9)</td>
<td>0 (0)</td>
<td>7 (17.9)</td>
<td>12 (28.6)</td>
<td></td>
</tr>
<tr>
<td>MCI (N, %)</td>
<td>16 (11.7)</td>
<td>2 (3.6)</td>
<td>3 (7.7)</td>
<td>11 (26.2)</td>
<td></td>
</tr>
<tr>
<td>Cognitively unimpaired (N, %)</td>
<td>102 (74.5)</td>
<td>54 (96.4)</td>
<td>29 (74.4)</td>
<td>19 (45.2)</td>
<td></td>
</tr>
<tr>
<td>Aβ42/Aβ40 (mean, SD)</td>
<td>0.048 (0.019)</td>
<td>0.068 (0.011)</td>
<td>0.039 (0.006)</td>
<td>0.030 (0.007)</td>
<td></td>
</tr>
<tr>
<td>p-tau (pg/mL) (mean, SD)</td>
<td>0.045 (0.034)</td>
<td>0.018 (0.006)</td>
<td>0.044 (0.014)</td>
<td>0.082 (0.034)</td>
<td></td>
</tr>
<tr>
<td>p-tau (pg/mL) (mean, SD)</td>
<td>23.5 (12.0)</td>
<td>15.9 (4.2)</td>
<td>18.6 (3.6)</td>
<td>38.1 (10.9)</td>
<td></td>
</tr>
<tr>
<td>NFT (pg/mL) (mean, SD)</td>
<td>119.3 (74.2)</td>
<td>80.8 (25.9)</td>
<td>112.9 (57.1)</td>
<td>176.6 (94.4)</td>
<td></td>
</tr>
<tr>
<td>NfL (pg/mL) (mean, SD)</td>
<td>902.4 (357.6)</td>
<td>752.9 (251.9)</td>
<td>711.7 (199.7)</td>
<td>1278.8 (303.3)</td>
<td></td>
</tr>
<tr>
<td>Alpha-synuclein (pg/mL) (mean, SD)</td>
<td>178.0 (77.6)</td>
<td>145.7 (58.1)</td>
<td>143.2 (45.9)</td>
<td>253.4 (71.3)</td>
<td></td>
</tr>
<tr>
<td>YKL-40 (ng/mL) (mean, SD)</td>
<td>164.9 (62.5)</td>
<td>133.0 (37.3)</td>
<td>153.9 (48.5)</td>
<td>217.7 (67.6)</td>
<td></td>
</tr>
<tr>
<td>sTREM2 (ng/mL) (mean, SD)</td>
<td>8.5 (2.5)</td>
<td>7.9 (2.2)</td>
<td>7.6 (2.0)</td>
<td>9.9 (2.8)</td>
<td></td>
</tr>
<tr>
<td>IL-6 (pg/mL) (mean, SD)</td>
<td>5.0 (4.5)</td>
<td>5.1 (4.7)</td>
<td>5.5 (5.5)</td>
<td>4.5 (2.7)</td>
<td></td>
</tr>
</tbody>
</table>

*2 samples were missing IL-6 measurements and thus excluded from analyses of IL-6.

Abbreviations: Aβ, amyloid beta; AD, Alzheimer’s disease; CSF, cerebrospinal fluid; IL-interleukin; MCI, mild cognitive impairment; NFT, neurofilament light chain; p-tau, phosphorylated tau; SD, standard deviation; sTREM2, soluble triggering receptor expressed on myeloid cells 2; YKL-40, chitinase-3-like protein 1.

### Figure 1: Study overview

The core analysis was a robust analysis of the CSF proteome and its relationship to AT category and various CSF biomarkers of neurodegeneration and neuroinflammation (“Proteomics Discovery”). Multiple forms of independent replication of the main findings of these CSF proteomics analyses were pursued, including CSF proteomics in an independent cohort (“Proteomics Replication”) and CSF metabolomics in both the original CSF samples (“Metabolomics Discovery”) and in yet another independent cohort (“Metabolomics Replication”). Various secondary analyses were performed to investigate other questions of high interest (“Proteomics Follow-up”). Note that the previously existing CSF NTK biomarker, metabolomic, and genotyping data in the WRAP and WI ADRC cohorts were generated for the same participants and (for the biomarker and metabolomic data) the same CSF samples as those used to generate the new CSF proteomics data. Of the 137 WRAP and WI ADRC participants used in the proteomics discovery analyses, one participant lacked CSF metabolomics data and was excluded from the comparative multivariate prediction metabolomic discovery analyses, which consequently only had a sample size of 136. AT, amyloid/tau; CSF, cerebrospinal fluid; NTK, NeuroToolKit; WI ADRC, Wisconsin Alzheimer’s Disease Research Center; WRAP, Wisconsin Registry for Alzheimer’s Prevention.
APLP2. The tau protein was not reliably quantified by nLC-MS/MS in the WRAP/WI ADRC samples. Little difference in the percentage of samples missing values for each protein was seen by AT category (Figure S2, Figure S3, Figure S4).

The CSF proteome showed a rich correlation structure with both larger clusters and smaller pockets of highly correlated proteins (Figure 2A, Table S3). Further interrogation with PCA underscored this complexity, with the first four principal components (PCs) collectively explaining only half (49.89%) of the total variance (Figure S5), with the top two PCs not explained by either AT or sex (Figure S6). The top PC (PC1) was weakly correlated with age at sample (correlation = 0.17; P = 0.045), and its top 5 protein contributors (seizure 6-like protein 2 [SEZ6L2], neurofascin [NFASC], neural cell adhesion molecule L1 [L1CAM], protocadherin-1 [PCDH1], and neuronal cell adhesion molecule [NRCAM]) shared a theme of neuronal cell structure and adhesion. The second PC (PC2) was not correlated with age (correlation = −0.019; P = 0.82), and its top 5 protein contributors (neuropeptide-like protein C4orf48, complement decay-accelerating factor [DAF or CD55], multiple epidermal growth factor-like domains protein 10 [MEGF10], epidermal growth factor-containing fibrulin-like extracellular matrix protein 1 [FBLN3], and ciliary neurotrophic factor receptor subunit alpha [CNTFR]) shared a theme of neuropeptides and glial function.

Pathway enrichment analysis comparing the proteins quantified in the CSF to the entire human proteome revealed significant enrichment of terms related to extracellular, neuronal, immune system, platelet pathway, protein processing, and others (Figure 2B, Table S4). There were 167 significantly enriched DO pathways among the observed proteins, with these DO terms encompassing AD, amyloidosis, coronary artery disease, and more (Figure 3).

Given the apparent presence of clusters of proteins based on the correlation structure, the CSF proteome was divided into three clusters based on a Gaussian mixture model (Figure S7). These three clusters were then compared to each other for the differential enrichment of biological pathways (Table S5). The KEgg terms revealed a pattern in which the smallest cluster (2) was enriched for immune system and cholesterol pathways while the other two larger clusters were enriched for extracellular, metabolism, and other pathways (Figure 2C).
FIGURE 3  Enriched Disease Ontology (DO) terms among the observed cerebrospinal fluid (CSF) proteome. A summary network representation of the enriched DO terms among the CSF proteome is shown. The larger colored nodes represent the 167 significantly enriched DO terms among the proteins identified in the CSF compared to the human proteome, and the smaller gray nodes represent the 463 proteins whose genes were connected with one or more of the enriched DO terms. The DO terms and connected proteins are colored according to their group from a community analysis, which are labeled with the top-enriched DO term within each community to provide a high-level overview. The full list of enriched DO terms is included in Table S4 in supporting information.

3.3 Protein–AT associations

The second major objective of this study was to identify the CSF proteins that are associated with AT-defined participant groups. The ANCOVA tests revealed 61 statistically significant associations between proteins and AT category after controlling for age at sample and sex and using a multiple testing correction (\(P < 5.46 \times 10^{-5}\)), with a total of 496 (54.2%) of the proteins nominally associated (\(P < 0.05\); Table S6). The differences in distribution of the top 10 proteins revealed a number of different patterns in relation to amyloid and tau pathology (Figure 4A). Based on both the box plots and logistic regressions, some proteins increased consistently as AT pathology increased (fatty acid-binding protein, heart [FABP3], SPARC-related modular calcium-binding protein 1 [SMOC1]). Other proteins did not change from A– to A+ but did change from T– to T+ (aspartate amino-transferase, cytoplasmic [GOT1], fructose-bisphosphate aldolase A [ALDOA], guanine deaminase [GDA], N(G), N(G)-dimethylarginine dimethylaminohydrolase 1 [DDAH1], ketimine reductase mu-crystallin [CRYM], and activin receptor type-1B [ACVR1B]). Overall, there was an enrichment of statistical signal across the proteome that was not seen in the permutation sensitivity analysis (Figure 4B). Controlling for the APOE ε4 allele count did not substantially change the results, with 53 of the 61 proteins remaining significantly associated when the APOE variable was added to the ANCOVA models (Figure S8). Among the set of significantly associated proteins, seven biological pathways were enriched relative to the entire set of 915 CSF proteins measured in this study (Table S7), including several related to glucose and carbon metabolism (Figure 4C).

When the logistic regression model was used to test for the direction of effect between proteins and A+T+ (vs. A–T–: again controlling for age at sample and sex), only nine proteins were significantly associated with being A+T+, and all of these proteins were also significantly associated in the ANCOVA model (Figure S9). All but one (fibulin-1 [FBLN1]) of the proteins significantly associated with A+T+ were increased in A+T+ relative to A–T– (Figure S10).

3.4 Protein–CSF biomarker associations

The third major objective of this study was to examine the CSF proteome in AD on a more biological level by examining the association of proteins with CSF biomarkers of neurodegeneration and neuroinflammation. When each of the nine CSF biomarkers (Aβ42/Aβ40, p-tau, p-tau/Aβ42, NfL, alpha-synuclein, neurogranin, YKL-40, sTREM2,
FIGURE 4 Proteins associated with AT category. A, The distributions of the top 10 most significantly associated proteins with AT category are shown. A number of different patterns were seen across increasing AT pathology, including proteins that increased consistently, decreased consistently, or increased only with tau positivity. B, The quantile-quantile (Q–Q) plots of the protein–AT ANCOVA association tests are shown above, with the distribution of P values shown separately for the original (“Not permuted”) and permuted data sets. Substantial signal enrichment was seen across the CSF proteome, with that signal absent in the permuted data set. C, The enriched KEGG pathways among the AT-associated proteins are shown, revealing a general perturbation of glucose and carbon metabolism pathways. ANCOVA, analysis of covariance; AT, amyloid/tau; CSF, cerebrospinal fluid; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.
TABLE 2  Top three significantly associated proteins for each biomarker.

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>UniProt ID</th>
<th>Entrez ID</th>
<th>Protein</th>
<th>Effect size</th>
<th>Effect size (std. biomarker)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aβ42/Aβ40</td>
<td>Q9H4F8</td>
<td>64093</td>
<td>SMOC1</td>
<td>−0.010</td>
<td>−0.54</td>
<td>4.55E-14</td>
</tr>
<tr>
<td>Aβ42/Aβ40</td>
<td>P40925</td>
<td>4190</td>
<td>MDH1</td>
<td>−0.009</td>
<td>−0.46</td>
<td>1.44E-09</td>
</tr>
<tr>
<td>Aβ42/Aβ40</td>
<td>S4R371</td>
<td>2170</td>
<td>FABP3</td>
<td>−0.008</td>
<td>−0.43</td>
<td>2.69E-08</td>
</tr>
<tr>
<td>alpha-synuclein</td>
<td>Q16270</td>
<td>3490</td>
<td>IGFBP7</td>
<td>−0.092</td>
<td>−0.43</td>
<td>4.28E-13</td>
</tr>
<tr>
<td>alpha-synuclein</td>
<td>E7EU1F1</td>
<td>5168</td>
<td>ENPP2</td>
<td>−0.091</td>
<td>−0.43</td>
<td>5.55E-12</td>
</tr>
<tr>
<td>alpha-synuclein</td>
<td>P12259</td>
<td>2153</td>
<td>F5</td>
<td>−0.084</td>
<td>−0.39</td>
<td>5.11E-11</td>
</tr>
<tr>
<td>log10 NfL</td>
<td>Q16270</td>
<td>3490</td>
<td>IGFBP7</td>
<td>−0.092</td>
<td>−0.43</td>
<td>4.28E-13</td>
</tr>
<tr>
<td>log10 NfL</td>
<td>P23142-4</td>
<td>2192</td>
<td>FBLN1</td>
<td>−0.091</td>
<td>−0.43</td>
<td>5.55E-12</td>
</tr>
<tr>
<td>log10 NfL</td>
<td>P12259</td>
<td>2153</td>
<td>F5</td>
<td>−0.084</td>
<td>−0.39</td>
<td>5.11E-11</td>
</tr>
<tr>
<td>neurogranin</td>
<td>E9PG71</td>
<td>2043</td>
<td>EPHA4</td>
<td>266.13</td>
<td>0.74</td>
<td>2.42E-25</td>
</tr>
<tr>
<td>neurogranin</td>
<td>O00451</td>
<td>2675</td>
<td>GFRA2</td>
<td>266.78</td>
<td>0.75</td>
<td>1.12E-23</td>
</tr>
<tr>
<td>neurogranin</td>
<td>Q8TAG5-2</td>
<td>222008</td>
<td>VSTM2A</td>
<td>259.95</td>
<td>0.73</td>
<td>8.53E-23</td>
</tr>
<tr>
<td>p-tau</td>
<td>A0A087WZ50</td>
<td>66004</td>
<td>LYNX1</td>
<td>6.70</td>
<td>0.56</td>
<td>1.05E-14</td>
</tr>
<tr>
<td>p-tau</td>
<td>P40925</td>
<td>4190</td>
<td>MDH1</td>
<td>6.78</td>
<td>0.57</td>
<td>3.29E-14</td>
</tr>
<tr>
<td>p-tau</td>
<td>E9PG71</td>
<td>2043</td>
<td>EPHA4</td>
<td>6.62</td>
<td>0.55</td>
<td>3.81E-14</td>
</tr>
<tr>
<td>p-tau/Aβ42</td>
<td>Q9H4F8</td>
<td>64093</td>
<td>SMOC1</td>
<td>0.019</td>
<td>0.57</td>
<td>1.83E-15</td>
</tr>
<tr>
<td>p-tau/Aβ42</td>
<td>S4R371</td>
<td>2170</td>
<td>FABP3</td>
<td>0.018</td>
<td>0.53</td>
<td>1.68E-12</td>
</tr>
<tr>
<td>YKL-40</td>
<td>P12259</td>
<td>2153</td>
<td>F5</td>
<td>−24.43</td>
<td>−0.39</td>
<td>4.63E-09</td>
</tr>
</tbody>
</table>

Abbreviations: Aβ, amyloid beta; AD, Alzheimer’s disease; CSF, cerebrospinal fluid; IL, interleukin; NfL, neurofilament light chain; p-tau, phosphorylated tau; sTREM2, soluble triggering receptor expressed on myeloid cells 2; YKL-40, chitinase-3-like protein 1

significantly associated proteins (Figure S11). The top three significantly associated proteins per biomarker are summarized in Table 2, with the full list of all statistically significant protein associations available in Table S8. A total of 68 significantly enriched pathways was observed among the biomarker-specific sets of significantly associated proteins compared to the set of CSF proteins measured, with glucose and carbon metabolic pathways noted to be enriched among amyloid and tau biomarkers (Table S9). The only enriched pathways for any biomarker not involving amyloid or tau were the KEGG pathway “cell adhesion molecules” (hsa04514) for neurogranin and the REACTOME pathway “extracellular matrix organization” (R-HAS-1474244) for sTREM2.

The network plot and subsequent community analysis of the protein–biomarker associations revealed three communities (modularity = 0.256) among the network (Figure 5). One community largely comprised the more traditional AD biomarkers of p-tau, p-tau/Aβ42, and Aβ42/Aβ40; a second community centered around the proteins associated with neurogranin; and the third community included the remaining biomarkers of alpha-synuclein, YKL-40, NfL, and sTREM2 (IL-6 had no significant protein associations and was not included). The largest number of shared associations across the biomarkers occurred among neurogranin, p-tau, and alpha-synuclein, which shared 152 protein associations between at least two of those biomarkers (Figure S12, Table S10). Relative to the 915 CSF proteins measured in this study, no GO or KEGG pathways were significantly enriched among the subset of proteins associated with two or more of neurogranin, p-tau, and alpha-synuclein.

3.5 | Replication of top pathway results

Given the historical challenges in replicating CSF proteomics associations in AD, we sought to further replicate our main proteomics associations in an external cohort. The glucose metabolism pathway (REACTOME ID R-HSA-70326) was significantly enriched among the proteins associated with both Aβ42/Aβ40 and p-tau/Aβ42, so the results of the nine proteins from this pathway that were significantly associated with one of the amyloid or tau biomarkers were chosen
Network analysis of protein-biomarker associations. A bipartite graph representation of the proteins significantly associated with the CSF biomarkers after Bonferroni correction is shown. The nodes representing the biomarkers are larger and in green. Proteins are represented as smaller nodes, with an edge to a biomarker representing a significant association between a protein and a biomarker. The colors of the protein nodes and underlying shaded regions correspond to three distinct communities identified with the fast greedy modularity optimization algorithm. Three such communities were identified. Community 1 included proteins associated with the more traditional AD biomarkers, such as p-tau, p-tau/\(\beta\)-42, and \(\beta\)-42/\(\beta\)-40. Community 2 centered around neurogranin, including many proteins uniquely associated with neurogranin. Community 3 included proteins associated with the remaining biomarkers of alpha-synuclein, YKL-40, NFL, and sTREM2. No proteins were significantly associated with IL-6. Many proteins were shared in the center of the network, particularly among neurogranin, p-tau, and alpha-synuclein. A\(\beta\), amyloid beta; AD, Alzheimer’s disease; CSF, cerebrospinal fluid; NFL, neurofilament light chain; p-tau, phosphorylated tau; sTREM2, soluble triggering receptor expressed on myeloid cells 2; YKL-40, chitinase-3-like protein 1.

3.6 | Secondary analysis of diagnosis-associated proteins

While this study was designed to study the relationship of the CSF proteome with amyloid, tau, and other biomarkers, a secondary analysis was performed to examine the relationship between CSF protein levels and clinical diagnosis (AD, MCI, or CU). Because the distribution of clinical diagnosis was imbalanced and the initial study design centered around CSF biomarkers, the results of this analysis were treated as exploratory. We identified 17 proteins significantly associated with clinical diagnosis using the same ANCOVA approach, covariates, and multiple-testing correction used earlier (Table S13, Figure S13). Of these significantly associated proteins, four of them (SMOC1, FBLN1, FABP3, and PGAM1) were also significantly associated with AT group.

3.7 | Secondary analysis of insulin-related proteins

As a follow-up of the theme of glucose and carbon metabolism among the proteomics findings, we looked for insulin-related proteins that had been identified but excluded during the QC process. Of the insulin-related proteins of interest that had not been already included in the QC'd data set, only IGF-1 and AKT1 were identified in the proteomics workflow. AKT1 was only quantified in one subject and was thus not suitable for further analysis, but IGF-1 was quantified in...
TABLE 3 Replication of implicated glucose metabolism-related proteins in the Knight ADRC.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Discovery (WRAP/WI ADRC)</th>
<th>Replication (Knight ADRC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CSF Aβ42/Aβ40</td>
<td>CSF p-tau/Aβ42</td>
</tr>
<tr>
<td>MDH1</td>
<td>-0.0088</td>
<td>1.44E-09</td>
</tr>
<tr>
<td>ALDOA</td>
<td>-0.0070</td>
<td>1.75E-06</td>
</tr>
<tr>
<td>PGK1</td>
<td>-0.0068</td>
<td>5.67E-06</td>
</tr>
<tr>
<td>TPI1</td>
<td>-0.0071</td>
<td>9.01E-07</td>
</tr>
<tr>
<td>PGAM1</td>
<td>-0.0081</td>
<td>4.35E-08</td>
</tr>
<tr>
<td>ALDOC</td>
<td>-0.0048</td>
<td>1.44E-03</td>
</tr>
<tr>
<td>ENO1</td>
<td>-0.0068</td>
<td>4.72E-06</td>
</tr>
<tr>
<td>GOT1</td>
<td>-0.0072</td>
<td>5.67E-07</td>
</tr>
<tr>
<td>PKM</td>
<td>-0.0066</td>
<td>6.93E-06</td>
</tr>
</tbody>
</table>

Note: A * indicates a protein that was not present in the QC'd Knight ADRC data set.
Abbreviations: Aβ, amyloid beta; ADRC, Alzheimer’s Disease Research Center; CSF, cerebrospinal fluid; QC, quality control; WI ADRC, Wisconsin Alzheimer’s Disease Research Center; WRAP, Wisconsin Registry for Alzheimer’s Prevention.

282 samples (59.9%) and analyzed further using only the non-imputed measurements. A trend in missing values by AT category was noted: 44.6% of samples were missing IGF-1 in A−T−, 41.0% in A−T+, 36.6% in A+T−, and 33.3% in A+T+. The ANCOVA analysis (with age at sample and sex as covariates) of IGF-1 did not show a statistically significant difference of the protein across AT categories (P = 0.170), though the distribution of the protein appeared to increase with amyloid positivity (Figure S14). The association analysis between IGF-1 and the CSF biomarkers revealed a nominally significant negative association with Aβ42 (P = 0.011) and positive association with p-tau/Aβ40 (P = 0.009; Table S14, Figure S15).

3.8 Secondary analysis of glycolysis and TCA cycle metabolites

Given the enrichment of glucose and carbon metabolism pathways among the proteomic associations, we were interested to see if direct CSF metabolite measurement of glucose metabolism–related metabolites would similarly be associated with amyloid, tau, or other biomarkers. All but one of the CSF samples for which we generated the CSF proteomics data here also had CSF metabolomics data generated previously, giving us a rich data set with both proteomics and metabolomics available on the same participants. We examined 10 glycolysis and TCA cycle–related metabolites in a discovery analysis: 1,5-anhydroglucitol (1,5-AG), alpha-ketoglutarate, citrate, glucose, glyceraldehyde, isocitrate, lactate, malate, pyruvate, and succinylcarnitine (C4-DC; Table S15). In testing for association with AT category, only one metabolite (succinylcarnitine (C4-DC)) was statistically significantly associated after multiple testing correction and controlling for age at sample and sex (P = 1.57 × 10−6; Table S16; Figure S16). This metabolite was also statistically significantly associated with CSF neurogranin, alpha-synuclein, p-tau, sTREM2, YKL-40, NfL, and p-tau/Aβ42, in each case indicating that succinylcarnitine increases in the CSF along with increasing levels of the biomarker (Table S17).

In an independent replication data set from the WRAP/WI ADRC that had CSF metabolite data but for which we did not generate proteomics data (n = 363; Table S18), succinylcarnitine showed a similar trend to what was seen in the discovery cohort. This association was weaker than that seen in the discovery cohort (P = 0.014), though the replication cohort distribution across the AT categories was strongly skewed to include more A−T− participants than the other two categories. More convincing were the succinylcarnitine–biomarker associations (i.e., with neurogranin, alpha-synuclein, p-tau, sTREM2, YKL-40, and NfL), which were replicated for all biomarkers except for p-tau/Aβ42.

3.9 Multomic prediction models for amyloid and tau

Our fourth and final main objective was to help inform future mutiomics work in AD by comparing the relative information offered by different major omics data types relative to amyloid and tau, a question that was empowered by the rare diversity of biology measured on the same participants and samples in the WRAP and WI ADRC cohorts (genomics, CSF proteomics, CSF metabolomics, and demographics). We built elastic net prediction models using different omic data sets and compared their performance in predicting amyloid and tau positivity in a cross-fold validation framework (with a held-out testing data set). The results of the multomic amyloid and tau prediction models revealed a consistent pattern in which the CSF proteome outperformed the other individual omic data sets in predicting positivity based on the core biomarkers of Aβ42/Aβ40, p-tau/Aβ42, and p-tau (Figure 6, Table S19). The predictive model based on the CSF proteome (number of predictors selected ranged from 75 to 105) achieved a high AUC for all three biomarkers (Aβ42/Aβ40 AUC = 0.839, p-tau/Aβ42
AUC = 0.920, and p-tau AUC = 0.954), performing slightly better than even the integrative model when predicting p-tau or p-tau/\(A\beta_{42}\). For \(A\beta_{42}/A\beta_{40}\) and p-tau positivity, the sensitivity and specificity values further demonstrated the relative superiority of the proteomics model. For \(A\beta_{42}/A\beta_{40}\) positivity, the sensitivity of the CSF proteome model was 0.800 compared to much lower values (0.050–0.450) from the other models. For p-tau positivity, the specificity of the CSF proteome model (0.800) was much higher than the other models (0.000–0.200).

In all cases, the genome-based model performed poorly (AUCs ranged from 0.525–0.679; Figure 6A). The 2D histograms showing the performance of the CSF proteome relative to the raw biomarker values highlighted the effective classification by the proteomic models with effective delineation between positive and negative amyloid statuses (Figure 6B–D).

4 | DISCUSSION

We completed four main CSF proteomics analyses with multiple and diverse sources of independent replication: (1) an extensive characterization of the CSF proteome in an AD-focused cohort (Note S1 in supporting information), (2) a differential proteomics analysis centered on the ATN framework rather than clinical diagnoses, (3) an interrogation of the individual and pathway-level associations of the CSF proteome with a robust set of neurodegeneration and neuroinflammation biomarkers, and (4) a comparative multiomic predictive analysis of amyloid and tau using different omic data. We found that the results of the proteomics discovery and replication analyses highlighted altered (namely increased) levels of glucose- and carbon metabolism–related proteins in the CSF. Further triangulation with CSF metabolomics data
on the same participants from the discovery proteomics cohort and independent replication identified a positive association of succinylcarnitine with p-tau and other biomarkers of neurodegeneration and neuroinflammation. These results provide new insight into the specifics and timing of metabolic dysregulation in AD.

We identified a total of 61 AT-associated CSF proteins after multiple testing correction (Table S6), with 43 of these proteins previously implicated in AD. The protein here with the strongest association with AT category was SMOC1, which increased in the CSF with increasing pathology along the AT categories. The increase of SMOC1 with increasing pathology or disease severity has been noted before,72,73,77 and the protein has also been found to partly colocalize with amyloid plaques.76 FABP3, which was found to increase across AT categories in this study, has also been associated with AD previously.77,79,80 However, several proteins commonly associated with AD were not significantly associated with AT category here, including APOE, clusterin, and secretogranin.80 Although the reason for this lack of association is unclear, it could be due to the present cohort being largely preclinical. We also identified several novel protein associations, including ectonucleotide pyrophosphatase/phosphodiesterase family member 5 (ENPP5, noted to possibly be involved in neuronal cell communications81), heparin cofactor2 (SERPIND1, previously associated with multiple sclerosis82), extracellular matrix protein 2 (ECM2, jointly associated with iron along with APOE83), and glycoprotein endo-alpha-1,2-mannosidase-like protein (MANEAL, where variants in both MANEAL and OSTM1 have been observed in connection with a neurodegenerative disorder84). Beyond the replication in the Knight ADRC described above, we also found replication of our top glucose-metabolism–protein associations in the results of two previously published proteomics studies (Note S2 in supporting information), and we note that numerous associations identified in the exploratory protein–diagnosis secondary analysis replicated prior work as well (Note S3 in supporting information).

Unique to this study was the inclusion of a more comprehensive set of CSF biomarkers relevant to AD, neurodegeneration, and neuroinflammation. We replicated numerous protein associations with p-tau, but found less replication for proteins previously associated with amyloid,77,85 which could be due to differences in the exact amyloid measures used (Note S4 in supporting information). We also saw an imbalanced distribution of protein associations by biomarker, with neurogranin having by far the greatest number of protein associations and IL-6 having none (Note S5 in supporting information).

Among the amyloid and tau biomarkers (but not other biomarkers), the associated proteins shared a common theme of enrichment of glucose metabolism pathways (Table S9), including two proteins (MDH1 and ALDOA) that showed evidence of association with AD diagnosis in the Knight ADRC replication data set and in prior studies.24,72,77,86–89 Glucose metabolism, which is the major source of energy for the brain, has long been known to show signs of dysfunction in AD even before the emergence of symptoms.90–94 Our findings here, in which 74% of participants were CU, support the association of glucose metabolic dysregulation with alterations in amyloid and tau. One potential consequence of altered glucose metabolism is a disruption in autophagy and proteostasis,90 which has been connected to AD.95–98 The general enrichment of association signal across the proteome we observed here along with specific findings for a heat shock protein (HSP90AA1) might be further evidence of disrupted proteostasis, though more work is needed to explore this question in proteomics data (Note S6 in supporting information).

Further underscoring potential abnormalities in energy metabolism in AD is the observation that 18 of the 61 proteins associated with AT here have been previously connected with insulin resistance, including pyruvate kinase (PKM), alpha-enolase (ENO1), and triosephosphate isomerase (TPI).99 The secondary analysis of IGF1, though exploratory, further suggested possible abnormalities in insulin signaling (Note S7 in supporting information).

The analysis of specific metabolites in the glycolysis and TCA cycle pathways identified an increase of CSF succinylcarnitine (C4-DC) levels in the A+T+ group relative to both the A−T− and A+T− groups. Succinylcarnitine is a type of acylcarnitine, which typically plays a role in metabolism by transporting fatty acids from the cytoplasm into the mitochondria for beta-oxidation100 or by transferring succinyl groups between the mitochondria and the cytosol through the carnitine shuttle.101 With regard to the TCA cycle, succinyl-CoA (one of the TCA cycle intermediates) may be esterified into succinylcarnitine. Our results here provide evidence in humans that implicates succinylcarnitine in AD in conjunction with changing tau levels, complementing previous AD genomics findings relating a nearby enzyme (succinyl-CoA ligase, SUCLG2) with CSF amyloid levels102 and mouse models of aging and AD (Note S8 in supporting information). Moreover, recent AD metabolomics work in the blood and brain identified general acylcarnitine dysregulation and specifically found decreased levels of short-chain acylcarnitines103 (the same kind of acylcarnitine as succinylcarnitine), which potentially indicates that either the CSF or succinylcarnitine is an exception to this general trend. More broadly, our proteomics and follow-up metabolomics analyses reveal a situation in which the protein machinery of glucose and carbon metabolism is increased as amyloid levels change, while succinylcarnitine levels increase with changes in tau, hypothetically revealing problems with metabolic pathways trying to keep up with the demands imposed by AD. Given the known increases in acylcarnitines more generally in aging,104,105 it is possible that increasing succinylcarnitine could be connected with oxidative stress in the brain as well. Additional joint multomics work will be needed to pinpoint the timing, context, and implications of these changes in succinylcarnitine levels and to investigate the lack of signal for other metabolites.

A few limitations deserve mention. First, although our discovery sample size of 137 was comparable to other CSF proteomics work in AD and we had multiple sources of replication, analysis in a larger discovery sample would provide more precision, particularly for the novel joint multomic analyses presented here. Our study was also limited to individuals of European ancestry, which restricts its generalizability. Another limitation of our study was the lack of the N (neurodegeneration) category in our main analyses. Expanding the categories to all relevant combinations of ATN would allow for a more complete picture of the changing AD proteome. We also acknowledge that there are potential gray areas in categorizing individuals
as positive or negative for a given biomarker, but those gray areas were part of the reason we also studied protein associations with continuous measures of amyloid and tau. The similarity of the enriched pathways between dichotomous and continuous measures of amyloid and tau reduces the concern that imprecision in the AT categorizations substantially affected our results. Also, having comparison groups for other non-AD causes of dementia would allow for better triangulation of AD-specific proteomic changes; more in-depth analysis of A–T participants (which were only included in the Knight ADRC data used here) could be a reasonable place to start. We note as well that the Knight ADRC replication data set used a different proteomics platform than the discovery data set (aptamer-based vs. MS-based) and that the correlation between the MS-based and immunoassay-based measurements of YKL-40 and sTREM2 were only moderate. One the one hand, technical differences between the platforms could have introduced different biases or issues in protein identification that might have impacted the observed results. On the other hand, one way to deal with potential differences in proteomics platforms across cohorts is to seek replication through different technologies, as we did here. The presence of concordant replication of association for three out of five overlapping proteins despite the difference in platform shows the robustness of these CSF signals.

Nevertheless, our study provides a thorough investigation of the CSF proteome and its relationship to AD, AD biomarkers, and other omic data types. Among the numerous proteins associated with amyloid, tau, and other CSF biomarkers of neurodegeneration and neuroinflammation, we demonstrated that the CSF proteome associated with amyloid and tau was enriched for glucose and carbon metabolic pathways in contrast to the other biomarkers. Follow-up independent proteomic replication analyses and CSF metabolomics interrogation provided robust support for the increase in the CSF of these proteins and highlighted succinylcarnitine as a relevant metabolite, corroborating several previous lines of AD experimental and genomic evidence. In total, these results showcase the power of multomic analyses and provide a new look at the CSF proteome in AD in relation to amyloid and tau.

AUTHOR CONTRIBUTIONS

Authors DJP, JM, JJC, and CDE conceived the study design and conducted the initial pilot study. Author DJP conducted the quality control, data integration, and analyses of the various Wisconsin data sets; prepared the figures and tables; and led the writing of the manuscript. Authors JM and JJC generated the Wisconsin proteomics data. Authors YKD, ARM, GEE, and BBB contributed to the secondary analyses and interpretations of the main proteomics findings. Author EMJ contributed to the statistical design of the study. Author CAVH contributed to the use and development of the biomarker-based categorizations. Authors CC, CY, YS, and MA conducted the replication analyses in the Knight ADRC data. Authors HZ, KB, GK, IS, and AB contributed to the generation of the NTK CSF biomarker data set. Authors CDE, CMC, SCJ, BBB, SA, JJC, HZ, and KB contributed resources or funding. Authors CMC, SCJ, BBB, SA, and CDE contributed to cohort sample or data collection. Authors CMC, SCJ, and SA oversaw the Wisconsin Alzheimer’s research cohort studies. All authors contributed to and critically reviewed the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank WRAP and WI ADRC participants and the Wisconsin Alzheimer’s Institute (WAI) and WI ADRC staff for their contributions to the WRAP and WI ADRC studies. Without their efforts this research would not be possible. This research is supported by National Institutes of Health (NIH) grants R01AG27161 (Wisconsin Registry for Alzheimer Prevention: Biomarkers of Preclinical AD), R01AG054047 (Genomic and Metabolomic Data Integration in a Longitudinal Cohort at Risk for Alzheimer’s Disease), P41GM108538 (National Center for Quantitative Biology of Complex Systems), R01AG037639 (White Matter Degeneration: Biomarkers in Preclinical Alzheimer’s Disease), R01AG021155 (The Longitudinal Course of Imaging Biomarkers in People at Risk of AD), and P50AG033514 and P30AG062715 (Wisconsin Alzheimer’s Disease Research Center Grant), the Clinical and Translational Science Award (CTSA) program through the NIH National Center for Advancing Translational Sciences (NCATS) grant UL1TR000427, and the University of Wisconsin-Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. Computational resources were supported by a core grant to the Center for Demography and Ecology at the University of Wisconsin-Madison (P2CHD047873). We also acknowledge use of the facilities of the Center for Demography and Health at the University of Wisconsin–Madison, funded by the National Institute on Aging Center grant P30AG017266. Author DJP was supported by NLM training grants to the Bio-Data Science Training Program (T32LM012413) and the Interdisciplinary Training Program in Cardiovascular and Pulmonary Biostatistics (5T32HL83806). Author YKD was supported by a training grant from the National Institute on Aging (T32AG000213). Author GEE was supported by an Alzheimer’s Association Research Fellowship (2019-AARF-643973). Author HZ is a Wallenberg Scholar supported by grants from the Swedish Research Council (#2018-02532), the European Research Council (#681712), Swedish State Support for Clinical Research (#ALFGBG-720931), the Alzheimer Drug Discovery Foundation (ADDF), USA (#201809-2016862), the AD Strategic Fund and the Alzheimer’s Association (#ADSF-21-831376-C, #ADSF-21-831381-C and #ADSF-21-831377-C), the Olav Thon Foundation, the Erling-Persson Family Foundation, Stiftelsen för Gamla Tjänarinnor, Hjärnfonden, Sweden (#FO2019-0228), the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860197 (MIRIADE), and the UK Dementia Research Institute at UCL. Author KB was supported by the Swedish Research Council (#2017-00915), the Alzheimer Drug Discovery Foundation (ADDF), USA (#201809-2016862), the AD Strategic Fund and the Alzheimer’s Association (#ADSF-21-831376-C, #ADSF-21-831381-C and #ADSF-21-831377-C), the Olav Thon Foundation, the Erling-Persson Family Foundation, Stiftelsen för Gamla Tjänarinnor, Hjärnfonden, Sweden (#FO2019-0228), the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860197 (MIRIADE), and the UK Dementia Research Institute at UCL. Author KB was supported by the Swedish Research Council (#2017-00915), the Alzheimer Drug Discovery Foundation (ADDF), USA (#201809-2016862), the AD Strategic Fund and the Alzheimer’s Association (#ADSF-21-831376-C, #ADSF-21-831381-C and #ADSF-21-831377-C), the Olav Thon Foundation, the Erling-Persson Family Foundation, Stiftelsen för Gamla Tjänarinnor, Hjärnfonden, Sweden (#FO2019-0228), the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860197 (MIRIADE), and the UK Dementia Research Institute at UCL.
data collection and analysis, decision to publish, or preparation of the manuscript. Author CC receives support from the National Institutes of Health (R01AG044546, R01AG064877, RF1AG053303, R01AG058501, U01AG058922, R01AG064614, 1RF1AG074007), and the Chuck Zuckerberg Initiative (CZI). The recruitment and clinical characterization of research participants at Washington University were supported by NIH P30AG066444, and P01AG003991. 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. ELECSYS, COBAS, and COBAS E are trademarks of Roche. The Roche NeuroToolKit robust prototype assays are for investigational purposes only and are not approved for clinical use. We thank the University of Wisconsin Madison Biotechnology Center Gene Expression Center for providing Illumina Infinium genotyping services. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

CONFLICT OF INTEREST STATEMENT

Author CC receives research support from Biogen, EISAI, Alector, GSK, and Parabon; these 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. Author CC is a member of the advisory board and owns stock of Vivid Genomics and Circular Genomics. Author HZ has served on scientific advisory boards and/or as a consultant for Alector, Elsai, Denali, Roche Diagnostics, Wave, Samumed, Siemens Healthineers, Pionente Therapeutics, Nervgen, AZTherapies, CogRx, and Red Abbey Labs; has given lectures in symposia sponsored by Cellectricon, Fujirebio, Alzecure, and Biogen; and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program. Author KB has served as a consultant, on advisory boards, or on data monitoring committees for Abcam, Axon, Biogen, JOOMDD/Shimadzu, Julius Clinical, Lilly, MagQu, Novartis, Roche Diagnostics, and Siemens Healthineers, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (BBS), which is a part of the GU Ventures Incubator Program. Author GK is a full-time employee of Roche Diagnostics GmbH. Author IS is a full-time employee and shareholder of Roche Diagnostics International Ltd. Author AB is a full-time employee and shareholder of Roche Diagnostics GmbH. Author SCJ serves as a consultant to Roche Diagnostics and receives research funding from Siemens Healthineers. Other authors have no competing interests to declare. Author disclosures are available in the supporting information.

CONSENT STATEMENT

For the data from the WRAP and WI ADRC: this study was performed as part of the Generations of WRAP (GROW) study, which was approved by the University of Wisconsin Health Sciences Institutional Review Board; participants in the WRAP and WI ADRC studies provided written informed consent. For the data from the Knight ADRC: the institutional review boards of Washington University School of Medicine in St. Louis approved the study; research was performed in accordance with the approved protocols and participants provided informed consent.

DATA AVAILABILITY STATEMENT

The data sets generated and analyzed in this study from the WI ADRC may be requested at https://www.adrc.wisc.edu/apply-resources. The Knight-ADRC proteomic data are available at NIAGADS: NG00102 collection and can be interactively explored at http://ngi.pub:3838/ONTIME_Proteomics/. Version information of the software used in this study is provided in the Methods.

REFERENCES


PANYARD ET AL.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.