






OPEN

DATA DESCRIPTOR

Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease

Lingyan Ping^{1,4}, Sean R. Kunding^{1,4}, Duc M. Duong^{1,4}, Luming Yin^{1,4}, Marla Gearing^{2,4}, James J. Lah^{3,4}, Allan I. Levey^{3,4}  & Nicholas T. Seyfried^{1,3,4}  


Alzheimer's disease (AD) is characterized by an early, asymptomatic phase (AsymAD) in which individuals exhibit amyloid-beta (A β) plaque accumulation in the absence of clinically detectable cognitive decline. Here we report an unbiased multiplex quantitative proteomic and phosphoproteomic analysis using tandem mass tag (TMT) isobaric labeling of human post-mortem cortex ($n = 27$) across pathology-free controls, AsymAD and symptomatic AD individuals. With off-line high-pH fractionation and liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) on an Orbitrap Lumos mass spectrometer, we identified 11,378 protein groups across three TMT 11-plex batches. Immobilized metal affinity chromatography (IMAC) was used to enrich for phosphopeptides from the same TMT-labeled cases and 51,736 phosphopeptides were identified. Of these, 48,992 were quantified by TMT reporter ions representing 33,652 unique phosphosites. Two reference standards in each TMT 11-plex were included to assess intra- and inter-batch variance at the protein and peptide level. This comprehensive human brain proteome and phosphoproteome dataset will serve as a valuable resource for the identification of biochemical, cellular and signaling pathways altered during AD progression.

Background & Summary

Alzheimer's disease (AD) is the most common age-related neurodegenerative disease, currently affecting more than 40 million people worldwide^{1,2}. AD is characterized by an early, asymptomatic phase (AsymAD) in which individuals exhibit amyloid-beta (A β) plaque accumulation in the absence of significant tau neurofibrillary tangles (NFT) and cognitive decline^{3,4}. Currently the downstream biochemical and cellular processes that eventually lead to changes in cognition and even dementia are not well understood. Thus, a holistic or systems-level approach that aims to understand these altered processes may yield insight into new drug targets and biomarkers for AD.

Proteins are the proximate mediators of disease, integrating the effects of genetic, epigenetic, and environmental factors. It is well established that distinct mechanisms regulate expression and turnover of RNA and proteins, resulting in weak correlations in their respective levels⁵. This is mainly attributed to the complexity of the human proteome in which 90% of genes yield alternatively-spliced RNA transcripts^{6,7}, and each translated protein isoform is potentially altered by up to 300 or more posttranslational modifications (PTMs)^{8,9}. Protein phosphorylation is one of the key PTMs that govern signaling pathways and pathophysiological mechanisms in AD^{10,11}, inspiring a large body of research that has identified several kinases including GSK-3 β , CDK5, PKC, MAPK, and ROCK2 that have been implicated in the phosphorylation of tau and other key substrates in AD brain¹²⁻¹⁶. Conversely, reduced expression and activity of protein phosphatases like PP2A are also thought to contribute to enhanced phosphorylation of tau and other substrates¹⁷. Thus, quantifying total protein levels and phosphorylated

¹Department of Biochemistry, Emory University School of Medicine, Atlanta, GA, 30322, Georgia. ²Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA, 30322, Georgia.

³Department of Neurology, Emory University School of Medicine, Atlanta, GA, 30322, Georgia. ⁴Center for Neurodegenerative Diseases, Emory University School of Medicine, Atlanta, GA, 30322, Georgia.  e-mail: nseyfri@emory.edu

proteins across AsymAD and AD stages of disease may reveal defects in kinase- or phosphatase-mediated signaling pathways and biomarkers involved in AD progression.

Advances in liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) now facilitate high-throughput detection and quantification of thousands of proteins in a given sample. Data-dependent acquisition (DDA) or shotgun approaches are the traditional methods for proteomics workflows¹⁸. However, a drawback of DDA approaches is that high abundance peptides are biased towards selection for tandem MS/MS and subsequent identification¹⁹. To improve the detection of the low abundance peptides and enhance the depth of the proteome, different off-line fractionation methods have been used to reduce the complexity of peptide mixtures in samples prior to LC-MS/MS analysis. Methods including two-dimensional gel electrophoresis^{20,21}, strong cation exchange (SCX), electrostatic repulsion-hydrophilic interaction chromatography (ERLIC)²², and high-pH reversed-phase chromatography^{23,24} are used to increase peptide identification by separating peptides in an orthogonal dimension²⁵. With the advancement of multiplex isobaric tandem mass tags (TMT), off-line fractionation and high-resolution MS, proteomic datasets are beginning to rival the depth and breadth of transcriptomic datasets^{22,26}. Moreover, integrative proteomic and transcriptomic analyses in an AD post-mortem brain cohort suggests that transcriptome- and proteome-wide analyses can generate both complementary and unique information⁵. Proteomic analyses also offer the important opportunity to identify disease-specific PTMs that may participate in key pathological processes, and potentially serve as novel biomarkers and therapeutic targets.

Here we utilized a multiplex TMT MS-based proteomic approach, using similar protocols established by the Clinical Proteomics Tumor Analysis Consortium (CPTAC)²³, to comprehensively quantify the total proteome and phosphoproteome of human post-mortem cortical cases ($n = 27$) across pathology-free controls, AsymAD and symptomatic AD individuals. This led to the quantification of 11,378 unique protein groups, as well as 48,992 phosphorylated peptides representing 33,652 phosphosites. This dataset can serve as a valuable resource to help researchers elucidate the complexity of AD as it relates to proteomic signatures found in post-mortem human brain.

Methods

Human brain tissue. Human brain post-mortem tissues from the dorsolateral prefrontal cortex (Frontal Cortex, Brodmann Area 9) were obtained from the Emory Alzheimer's Disease Research Center (ADRC) brain bank. All procedures for ADRC are approved by Emory University Institutional Review Board (IRB), and a written informed consent form is obtained before tissue collection. In accordance with Emory University policy, the use of control postmortem tissues was considered exempted research. Postmortem neuropathological evaluation of A β plaque distribution was performed according to the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) criteria²⁷, while the extent of neurofibrillary tangle pathology was assessed in accordance with the Braak staging system²⁸. In total, 27 samples from 3 groups ($n = 10$ control, $n = 8$ AsymAD, and $n = 9$ AD) were used for brain proteome and phosphoproteome analyses. All case metadata, including disease state, age of death, Post-mortem interval (PMI), sex, and apolipoprotein (*ApoE*) genotype are listed in sample traits file²⁹ (Supplementary Tables 1).

Brain tissue homogenization and protein digestion. Procedures for tissue homogenization were performed essentially as previously described²². Approximately 100 mg (wet tissue weight) of brain grey matter tissue was homogenized in 500 μ L of 8 M urea lysis buffer (8 M urea, 100 mM Na₂HPO₄, pH 8.5) with HALT protease and phosphatase inhibitor cocktail (ThermoFisher) using a Bullet Blender (NextAdvance). The samples were homogenized for two full 5 min cycles at 4 °C with ~100 μ L of stainless-steel beads (0.9 to 2.0 mm blend, NextAdvance). The lysates were transferred to new Eppendorf LoBind tubes followed by 3 cycles of sonication consisting of 5 s of active sonication at 30% amplitude with 15 s incubation periods on ice in between sonication pulses. Samples were then centrifuged for 5 min at 15,000 g and the supernatant was transferred to a new tube. Prior to further processing, protein concentration and integrity were tested by bicinchoninic acid (BCA) assay (Pierce) and SDS-PAGE, respectively. For protein digestion, 500 μ g of each sample was aliquoted and volumes were normalized with additional lysis buffer. Samples were reduced with 1 mM dithiothreitol (DTT) for 30 min, followed by 5 mM iodoacetamide (IAA) alkylation in the dark for another 30 min. Samples were diluted 4-fold with 50 mM triethylammonium bicarbonate (TEAB) before incubating with Lysyl endopeptidase (Wako) at 1:100 (w/w) for 12 hr. Trypsin (Promega) was then added at a 1:50 (w/w) ratio and digestion was carried out for another 12 hr after the urea concentration was diluted to 1 M with 50 mM TEAB. The peptide solutions were desalted with a C18 Sep-Pak column (Waters). Briefly, the Sep-Pak columns were activated with 3 \times 1.5 mL of methanol, then equilibrated with 6 \times 1.5 mL 0.1% trifluoroacetic acid (TFA). The samples were loaded after acidification to a final concentration of 1% formic acid (FA) and 0.1% TFA. Each column was washed with 6 \times 1.5 mL 0.1% TFA. Elution was performed with 2 \times 1.5 mL 50% acetonitrile.

Tandem mass tag (TMT) peptide labeling. A 600 μ L aliquot from each sample was pooled and the mixture was divided into 6 global internal standard (GIS) samples with a total volume of 2400 μ L each, consistent with our previous work²², and peptide solutions were dried by vacuum (Labconco). All 27 samples from 3 groups and 6 GIS were divided and labeled using two sets of 5 mg 11 plex TMT reagents (Thermo Scientific A34808, Lot No for TMT 10-plex: SL258088, and 131 C channel: SJ258847). The batch arrangement is provided in Supplementary Table 2. Briefly, each of the TMT reagents were dissolved in 256 μ L anhydrous acetonitrile and the same channel was combined together from two 5 mg reagents. The samples were reconstituted in 400 μ L of 100 mM TEAB buffer and mixed with 3.2 mg (164 μ L) of the corresponding labeling reagent channel. The reaction was incubated for 1 hr and subsequently quenched with 32 μ L of 5% hydroxylamine (Pierce). For each TMT plex, labeled peptides from all 11 channels were mixed and desalted with a 500 mg Sep-Pak column (Waters). The labeled peptide mixture was eluted in 4.5 mL of 50% acetonitrile and dried by vacuum.

High-pH off-line fractionation. The off-line high pH fractionation method was adapted from the CPTAC protocol²³. Briefly, dried samples were re-suspended in high pH loading buffer (1 mM ammonium formate, 2% (v/v) acetonitrile) and loaded onto ZORBAX Extend 300 C18 columns (4.6 mm × 250 mm with 5 μm beads) from Agilent. An Agilent 1100 HPLC system was used to carry out the fractionation at room temperature with UV absorbance monitored at wavelength 280 nm. Ammonium formate (pH 10), diluted to a concentration of 4.5 mM in 2% (v/v) acetonitrile, was used as basic-pH reverse phase solvent A, whereas 4.5 mM ammonium formate (pH 10) in 90% (v/v) acetonitrile was used as paired solvent B. The peptides were eluted in a 96-min gradient with 0%–16% mobile phase B from 7–13 min, 16%–40% B from 13–73 min, 40%–44% B from 73–77 min, 44%–60% B from 77–82 min and kept at 60% B until the end with a flow rate of 0.8 mL/min (Supplementary Figure 1a). A total of 77 individual fractions were collected across the gradient from 14 to 91 minutes, at a rate of 1 fraction per minute. The 77 individual fractions were then pooled into 24 fractions by combining 3 off-line fractions into one with alternating combinations, except for the 24th fraction, which also contained the last 5 remaining individual fractions (Supplementary Figure 1b). Fractions were acidified to a final concentration of 1% FA, and 5% of sample volumes were dried and reserved for proteome analysis.

IMAC phosphorylated peptide enrichment. For phosphorylated peptide enrichment, 95% of the 24 high-pH fractions were further combined into 12 fractions in an alternating manner (1 and 13, 2 and 14, etc.). Peptide amounts were assumed to be equally distributed in all fractions. The IMAC enrichment method was performed according to CPTAC protocol with some minor modifications²³. Briefly, 1200 μL of slurry, in which the beads/solvent ratio is 1:1 (v/v), was utilized for one batch of TMT fractions. Beads were stripped of nickel with 8 mL of 100 mM EDTA and then equilibrated with 8 mL of 50 mM FeCl₃ both by end-to-end rotation for 30 min. To remove excess Fe³⁺ ion, beads were washed with 3 × 8 mL of water and resuspended in 2.4 mL of 1:1:1 (v/v/v) ratio of acetonitrile/methanol/0.01% (v/v) acetic acid. The beads were re-rinsed with 2.4 mL of 100% acetonitrile/0.1% TFA and divided into 12 tubes. The supernatant was removed before the peptide mixture was added. All 12 dried fractions were reconstituted in 0.4 mL of 50% acetonitrile/0.1% TFA and then diluted 1:1 with 100% acetonitrile/0.1% TFA to obtain a final 75% acetonitrile/0.1% TFA peptide solution at a concentration of 0.5 μg/μL. The peptide mixture was incubated with treated beads for 30 min with end-to-end rotation. Enriched IMAC beads were resuspended in 100 μL of 80% acetonitrile/0.1% TFA before the stage tips were conditioned. Stage tips were equilibrated with 2 × 100 μL methanol washes, 2 × 100 μL 100% acetonitrile/0.1% trifluoroacetic acid washes, followed by 2 × 100 μL of 1% FA washes. The IMAC bead slurry was loaded onto the stage tips and washed with 3 × 100 μL of 80% acetonitrile/0.1% TFA, then 3 × 100 μL of 1% FA. The phosphorylated peptides were released from IMAC beads by 3 × 100 μL 500 mM dibasic sodium phosphate (Na₂HPO₄, Sigma, S9763), pH 7.0, and washed by 3 × 100 μL 1% FA. The phosphorylated peptides were eluted from stage tips by 3 × 100 μL 50% acetonitrile/0.1% FA. The phosphorylated peptide solutions were dried with vacuum.

LC-MS/MS and TMT data acquisition on an Orbitrap Lumos mass spectrometer. Both proteome and phosphoproteome samples were run on a Fusion Lumos equipped with a NanoFlex nano-electrospray source (ThermoFisher). The same volume of loading buffer (19 μL of 0.1% TFA) was added to each of the fractions assuming equal distribution of peptide concentration across all 24 proteomic subfractions. Therefore, an equal 2 μL (1 μg equivalent) of each fraction was loaded for proteomic analysis. It was assumed that phosphorylated peptides were ~ 1% (w/w) of all peptides. The same volume of loading buffer (7 μL, 0.1% TFA) was added to all IMAC elution samples, and of this, 2 μL (1 μg equivalent) was analyzed by mass spectrometry. All proteome and phosphoproteome samples were separated on 25 cm long (75 μm ID) fused silica columns (New Objective, Woburn, MA) packed in-house with 1.9 μm Reprosil-Pur C18-AQ resin (Dr Maisch). All fractions were eluted over a 140 minute gradient using an Easy nLC 1200 (ThermoFisher). The gradient started with 1% buffer B (A: water with 0.1% formic acid and B: 80% acetonitrile in water with 0.1% formic acid) and went to 7% in 3 minutes, then increased from 7% to 30% in 137 minutes, then to 95% within 5 minutes and finally staying at 95% for 25 minutes. The mass spectrometer was operated in top speed mode with 3 second cycles. Both the MS and MS/MS scans were collected in the Orbitrap. The full scan was performed with a range of 375–1500 m/z, a nominal resolution of 120,000 at 200 m/z, automatic gain control (AGC) at 4 × 10⁵, a 50 ms max injection time and a radio frequency (RF) lens setting of 30%. Higher-energy collision dissociation (HCD) MS/MS scans settings were the following: resolution of 50,000, AGC at 1 × 10⁵, isolation width of 0.7 m/z, max injection time of 105 ms, and a collision energy of 38%. Only charge states from 2+ to 7+ were chosen for tandem MS/MS. All resulting raw files (*n* = 108) are provided³⁰.

Protein identification and quantification. Raw data files obtained from the Orbitrap Fusion were processed using Proteome Discoverer™ (version 2.3). MS/MS spectra were searched against the UniProt Knowledgebase (UniProtKB) Human proteome database (downloaded in 2015) with 90,411 total sequences as previously reported²². UniProtKB is a comprehensive³¹ and widely used database³², which contains both reviewed Swiss-Prot and predicted TrEMBL sequences³³. Since the additional depth provided by off-line bRPLC enables the sequencing of rare protein isoforms, this complete database will give us full potential for protein identification. A few peptides were also manually added to the database³⁴, including: one APOE 2/3-specific peptide and one APOE 3/4-specific peptide which allows APOE proteotyping of samples; 4 non-tryptic C-terminal Aβ-peptides specific to Aβ38 (GAIIGLMV), Aβ40 (GAIIGLMVGGVV), Aβ42 (GAIIGLMVGGVVIA), and Aβ43 (GAIIGLMVGGVVIAT). Sequences that map to tau microtubule-binding repeat (MTBR) domains were also set as an additional entry encompassing residues 224–370 (tau 2N4R Isoform, 1–441) in the Uniprot sequence, while all tau isoform sequences were modified by removing MTBR peptides and replicated as new “deltaMTBR” entries^{35,36}. The Aβ sequence (corresponding to residues 1–43) within the canonical APP isoform (P05067) was also excluded. Separation and quantification of these peptide sequences facilitated the investigation of APOE

allele, A β peptide and MTBR peptide-specific regulation of biology in AD datasets³⁵. The respective FASTA database used in this study was deposited on Synapse (syn20820455). The SEQUEST HT search engine was used and parameters were identical for both total and IMAC proteomes and specified as the following: fully-tryptic specificity; maximum of two missed cleavages; minimum peptide length of 6; fixed modifications for TMT tags on lysine residues and peptide N-termini (+ 229.162932 Da) and carbamidomethylation of cysteine residues (+ 57.02146 Da); variable modifications for oxidation of methionine residues (+ 15.99492 Da); deamidation of asparagine and glutamine (+ 0.984 Da); phosphorylation of serine, threonine and tyrosine (+ 79.9663 Da); precursor mass tolerance of 20 ppm; fragment mass tolerance of 0.05 daltons. Percolator was used to filter peptide spectral matches (PSM) and peptides to a false discovery rate (FDR) of less than 1% using target-decoy strategy. The phosphosite localization site threshold was set to 0.75, ensuring < 5% false-localization rate (FLR) of PTM assignments as described³⁷. Following spectral assignment, peptides were assembled into proteins and were further filtered based on the combined probabilities of their constituent peptides to a final FDR of 1%. In cases of redundancy, shared peptides were assigned to the protein sequence in adherence with the principles of parsimony. Reporter ions were quantified from MS2 scans using an integration tolerance of 20 ppm with the most confident centroid setting. The search results and TMT quantification are included³⁸.

Data Records

All files have been deposited on Synapse³⁹. These include sample traits²⁹, mass spectrometry raw files ($n = 108$) from both total proteome and phosphoproteome³⁰, the FASTA database³⁴, search result³⁸, and the ANOVA analysis input and output result⁴⁰. The mass spectrometry proteomics raw files and data analysis files have also been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository⁴¹.

Technical Validation

Deep dive proteome of human AD brain. We utilized a modified version of the CPTAC protocol to identify the total proteome and phosphoproteome from the same cases across different stages of AD. Control, AsymAD and AD tissues were randomized across the 3 batches (each containing 11 TMT channels) with 9 individual cases per batch (Fig. 1a and Supplementary Tables 1 and 2). Two TMT channels in each batch were dedicated to global reference internal standards (GIS), representing an equivalent amount of pooled peptides from all cases, which allows assessment of the intra- and inter-batch variance²². To reduce sample complexity and increase proteome depth prior to LC-MS/MS, we employed off-line high-pH reversed-phase fractionation essentially as described in the CPTAC protocol²³. A total of 77 individual fractions were collected and combined into 24 fractions for total proteome analysis (Supplementary Figure 1) for each batch. A step-wise concatenation strategy was used for pooling the fractions. A total of 5% of the material by volume was used for total proteome analysis, and the remaining 95% of the sample was used for phosphopeptide enrichment by immobilized affinity chromatography (IMAC) with Fe³⁺-loaded nitrilotriacetic acid (NTA) beads. Each of the 24 fractions were pooled into 12 subfractions prior to IMAC. Both the total proteome ($n = 72$ fractions across 3 batches) and phosphoproteome ($n = 36$ fractions across 3 batches) were analyzed by LC-MS/MS with high-resolution precursor and MS/MS scans on an Orbitrap Fusion Lumos mass spectrometer.

For the total proteome runs, a total of 164,034 unique peptides were identified that mapped to 11,378 protein groups at a 1% FDR on the peptide spectrum match (PSM) level across all batches, which represented 10,373 coding gene products. The total numbers of identified peptides, proteins and PSMs for all batches in the total proteome are listed in Table 1. For each batch, there were approximately 10,000 protein groups identified (Fig. 1b), which was comparable to the depth achieved in the CPTAC protocol using different tissue sources^{42,43}. The confidence of identification for peptide and protein is highly related with the number of PSMs and unique peptides. In the total proteome dataset, more than 77% of the proteins were identified with 2 or more unique peptides (Fig. 2a), while each unique peptide averaged approximately 6 PSMs (Table 1). Approximately 93% of all proteins were identified with at least 2 PSMs (Fig. 2b).

To obtain deep coverage of the phosphoproteome, enrichment strategies are usually applied due to the relatively low abundance of phosphorylation. To assess the quality of our IMAC phosphopeptide enrichment method, we calculated the percent phosphopeptide content (peptide level) in both the total proteome and IMAC phosphoproteome datasets (Tables 1 and 2, and Fig. 1d). The total proteome identified a total of 164,034 unique peptides, and approximately 2% were phosphopeptides (Fig. 1d). Although the IMAC dataset identified less peptides overall ($n = 72,138$), approximately 71% of the IMAC proteome were phosphopeptides (Table 2, and Fig. 1d). The IMAC enrichment method therefore led to an 18-fold increase in phosphopeptide identification using half of the instrument time. We set the threshold of phosphosite identification to 0.75 by SEQUEST, estimating less than 5% false localization rate (FLR) of each assigned site. After filtering, approximately 83% of all phosphosites identified had localization scores greater than or equal to 0.99. There were 51,736 phosphorylated peptides representing 33,652 unique individual phosphosites in total mapping to 8,415 proteins. A total of 34,379 of the phosphorylated peptides were identified in at least two of the three IMAC TMT batches (Fig. 1c). These figures are similar to the depth reported using the same protocol from breast cancer tissue⁴⁴. The numbers of identified peptides in each IMAC batch are listed in Table 2 with calculation of phosphorylation enrichment at both the total peptide and PSM level. Of note, the phosphorylated peptides showed slightly higher level of enrichment at the PSM level (83.55% as average) than peptide level (71.72% as average), which indicates that phosphopeptides were more intense and thus more frequently sequenced by LC-MS/MS than an average unmodified peptide in the phosphoproteome. Indeed, each non-phosphopeptide was identified by an average of 3.5 PSMs, whereas each phosphopeptide was identified by an average of 7 PSMs (Table 2). In total, 74% of all phosphopeptides identified by IMAC enrichment had two or more PSMs, which is consistent with the frequency of PSMs for peptides identified from the total proteome (Fig. 2c,d). Thus, although phosphopeptides were highly enriched in the IMAC

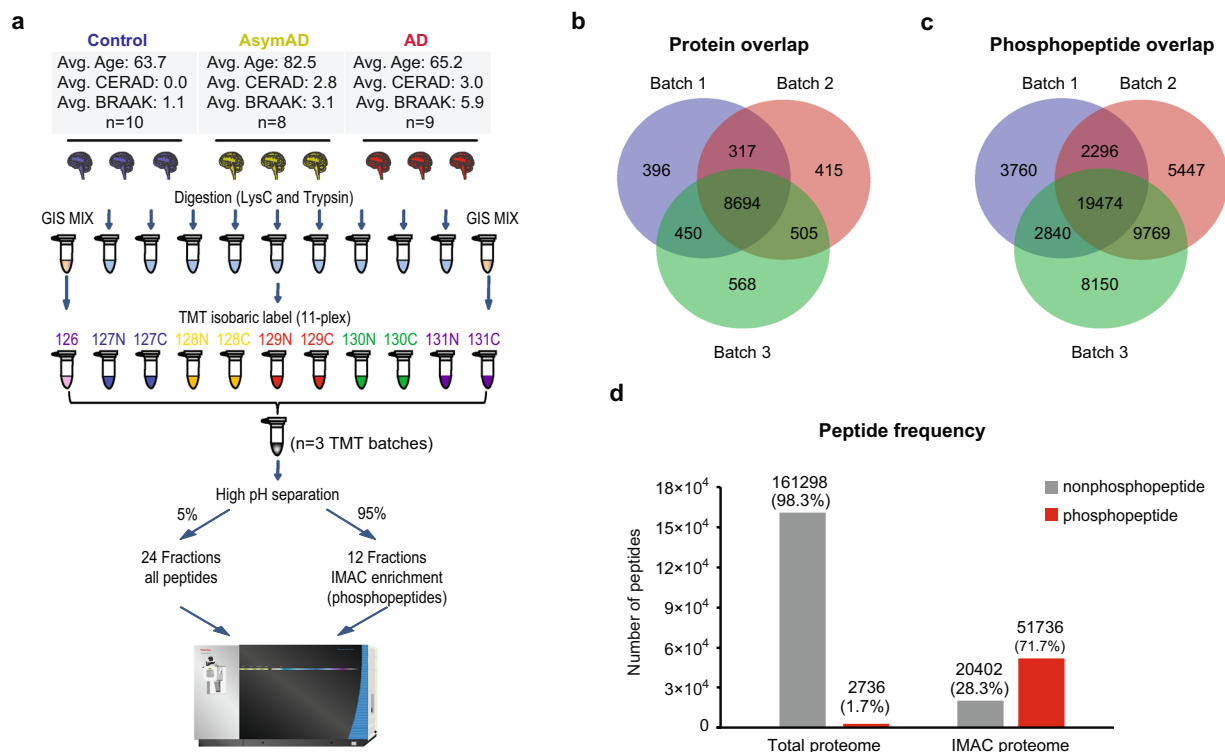


Fig. 1 Deep coverage of human brain proteome and phosphoproteome in Alzheimer's disease using TMT. **(a)** TMT Workflow. There were 27 human post-mortem tissues from the dorsolateral prefrontal cortex, which were digested and labeled with 3 batches of 11 plex TMT reagents. Two pooled global internal standards (GIS) were labeled with channels 126 and 131 C in each batch. There were 9 samples from 3 groups (control, AsymAD and AD) randomized and labeled in the remainder channels. The labeled sample mixture was fractionated by off-line high-pH reversed-phase chromatography. Whole proteome analysis was conducted using 5% of peptides from each fraction. Phosphopeptides were enriched from 95% of the samples by immobilized affinity chromatography (IMAC) with Fe³⁺-loaded nitrilotriacetic acid (NTA) beads. All fractions were analyzed on an Orbitrap Fusion Lumos Tribrid mass spectrometer. **(b)** Venn diagram of total proteome proteins from 3 batches. Approximately 10,000 protein groups were identified from each batch, with 8,694 shared protein groups found in all 3 batches, and 9,966 protein groups found in at least 2 batches. **(c)** Venn diagram of phosphopeptides identified in 3 batches. There were 19,474 shared phosphopeptides found across all 3 batches. **(d)** Peptide composition from both total proteome and IMAC proteome were illustrated. Non-phosphopeptides (grey) and phosphopeptides (red) are shown. Phosphopeptides were greatly enriched in IMAC proteome (71%), compared with the total proteome (1.7%).

Batch Number	Protein Groups	Unique Peptides	PSM	Phosphopeptides (no enrichment)
1	9,857	112,825	289,129	1,719
2	9,931	116,558	315,565	1,785
3	10,217	127,056	351,413	1,809
Total	11,378	164,034	956,107	2,758

Table 1. Protein and peptide identification from Total proteome.

proteome, they were sampled at a rate generally consistent with non-phosphopeptides from the total proteome, allowing greater sequencing depth of the phosphoproteome.

Assessing intra- and inter batch variance utilizing a pooled global internal standard. A major advantage of TMT approaches is the ability to quantify multiple samples in a single run, thereby critically reducing overall MS instrument time. This becomes especially important when the total number of samples increases to dozens or even hundreds^{36,45,46}. Typically, one or more TMT channels are dedicated for global internal standard(s) (GIS) and included in all batches, which can be used to normalize the measurement for protein or peptide signal from all samples across all batches²². In this study, we included two pooled reference standards in each 11-plex TMT batch (channels 126 and 131 C) (Fig. 1a), which allows normalization within and across TMT batches. The two reference standards essentially serve as technical replicates (i.e., a null-experiment), which can

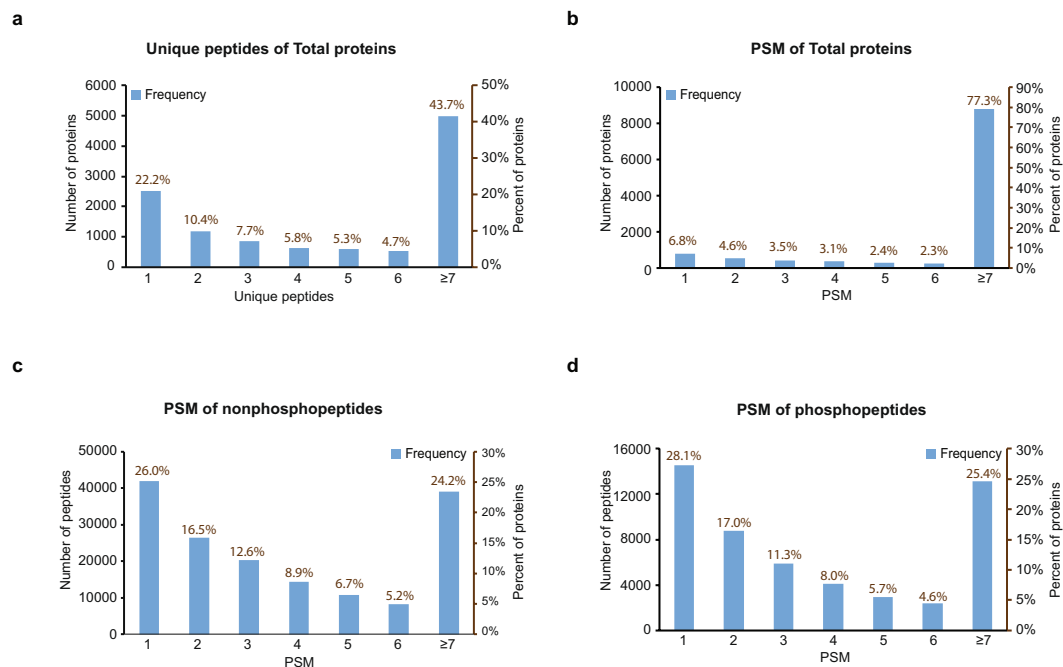


Fig. 2 Assessment of protein and peptide identification by peptide spectral matches for both proteome and phosphoproteome. **(a)** The frequency of protein groups (*blue bars*) per unique peptide bin for the total proteome. The percent of proteins per each unique peptide of all protein groups was also expressed. Approximately 77% of all protein groups identified were matched by at least 2 unique peptides. **(b)** The frequency of protein groups (*blue bars*) per peptide spectral match (PSM) bin for the whole proteome. The percent of proteins of total protein groups identified per each PSM bin was also quantified. More than 93% of the proteins identified were sampled by 2 or more PSMs. **(c)** The frequency of non-phosphopeptides (*blue bars*) matching to different PSM bins from total proteome. The percentage of peptides per each PSM bin out of all non-phosphopeptides identified was also quantified. There were 119,343 non-phosphopeptides identified with 2 or more PSMs, approximately 74% of all non-phosphopeptides identified. **(d)** The frequency of phosphopeptides (*blue bars*) matching to different PSM bins from the IMAC-enriched phosphoproteome. The percentage of peptides per each PSM bin out of all phosphopeptides identified from the IMAC-enriched proteome was also quantified. There were 37,195 phosphopeptides identified with 2 or more PSMs, making up approximately 72% of all IMAC phosphopeptides.

Batch Number	Total Peptide	Phospho-peptide	% of Enrichment (Peptide level)	Total PSM	Phospho-PSM	% of Enrichment (PSM level)
1	34,056	28,370	83.30%	113,894	102,905	90.35%
2	53,349	36,986	69.33%	157,279	118,150	75.12%
3	50,922	40,233	79.01%	163,477	142,074	86.91%
Total	72,138	51,736	71.72%	434,650	363,129	83.55%

Table 2. Peptide identification from IMAC samples.

be used to assess the variance in measurements. Thus, the degree of the signal variation between two internal GIS channels can be used as a threshold to further filter out poor quantitation data. Indeed, the signal of proteins and peptides from 126 and 131 C channels were very consistent and showed very good linear correlation across all three batches (Fig. 3a). We also consistently observed a strong correlation at both the peptide level from the total proteome (Fig. 3b) and phosphoproteome (Fig. 3c). Notably, some peptides exhibited large variation in signal between the two pooled standard channels, especially those peptides with lower total signal abundance as previously described²². According to the central limit theorem, the \log_2 ratio for the two GIS channels (\log_2 TMT channels 126/131 C) should fit a standard Gaussian distribution with the mean at or near zero (Supplemental Figure 2a), which can be used to assess the technical variation of measurements^{47,48}. This allows end-users of the datasets to impose a filtering criterion that can be used to remove peptides or proteins that do not meet variance metrics (> 2 standard deviations (SD) from the mean). Following this filtering criteria, a total of 1,123 peptides were filtered out of the analyses due to large variance, equivalent to $\sim 4\%$ of all quantitated peptides in IMAC Batch 1. In Batch 2 and Batch 3, there were 2,465 and 2,289 peptides filtered out by the > 2 SD standard, respectively, representing 5% of all peptides identified (Supplementary Figure 2b).

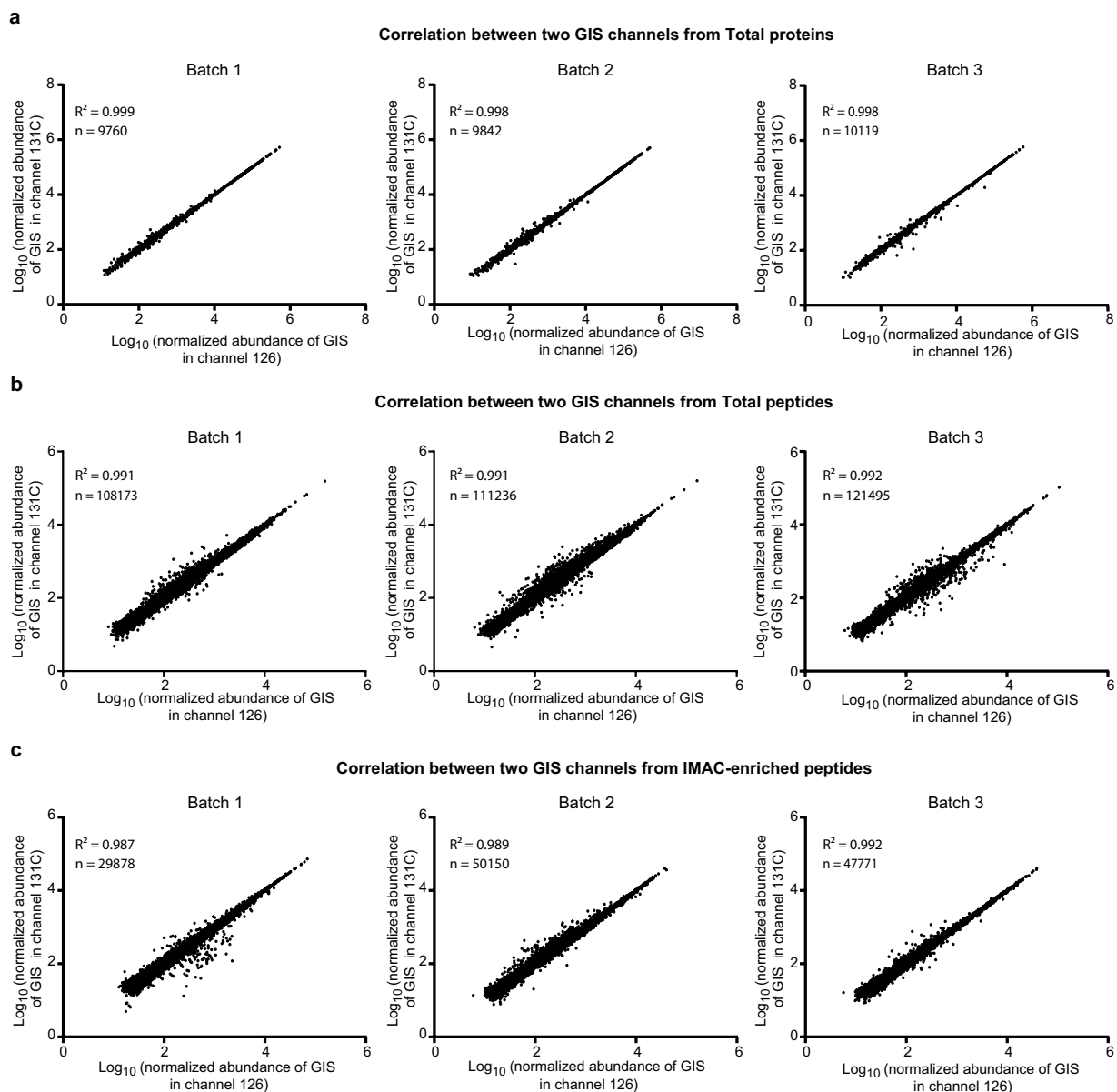


Fig. 3 Correlation of two pooled GIS samples across proteome and phosphoproteome batches. (a) Correlation of two GIS samples, which were in the TMT channels 126 and 131 C, in 3 total proteome batches at the protein level. Strong correlation was observed between 2 GIS channels in all 3 batches. (b) Correlation of 2 GIS samples in 3 whole proteome batches at the peptide level. (c) Correlation of 2 GIS samples in 3 IMAC-enriched phosphoproteome batches at the peptide level. Normalized 126 and 131 C channel intensities illustrate tight correlation.

It is also worth noting that batch effects may be significant when the sample number is large and the variation due to sample preparation cannot be ignored. In this case, post-hoc data normalization strategies should be employed to remove these batch effects^{36,49}. In this project, however, this step was not necessary given the relatively modest sample size ($n = 27$) and since all samples were digested at the same time.

Assessing amyloid levels and tau phosphorylation. A β plaque and hyperphosphorylated tau neurofibrillary tangle (NFT) accumulation in the brain are the core pathological hallmarks of AD^{50,51}. Thus, as a quality control of our measurements, we assessed the levels of A β and tau in our dataset. To confirm increased A β levels in diseased cases, the ion intensities from the first two tryptic peptides of A β were used as a surrogate for amyloid levels in the brain⁵, corresponding to residues 6–16 (Peptide 1) and 17–28 (Peptide 2) of the A β sequence, since the C-terminal non-tryptic peptides were not stably detected in all batches. Indeed, both of these two peptides showed significant increases in AsymAD and AD groups compared with control samples (Fig. 4a). Additionally, measurements of Peptide 1 and Peptide 2 were highly correlated (Fig. 4b). Given this, the sum intensity of the two

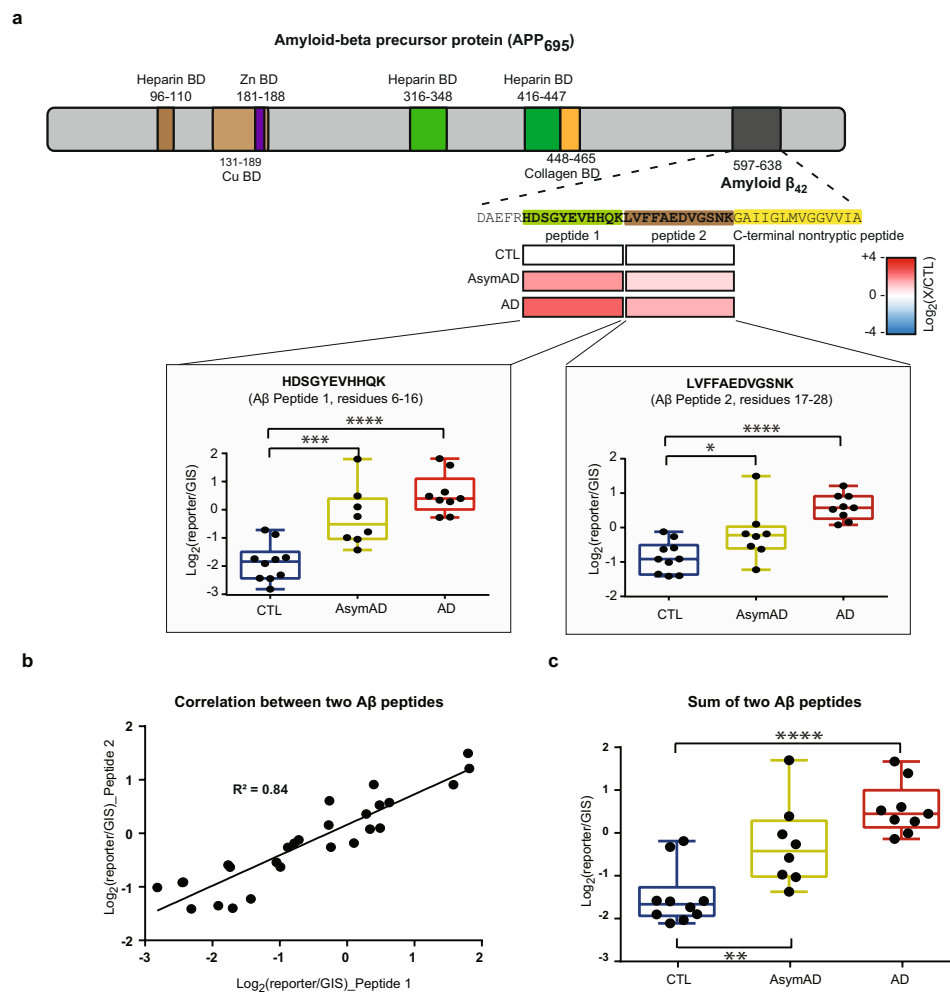


Fig. 4 Quantification of A β peptides. **(a)** Amyloid-beta precursor protein (APP) protein domain map showing the A β_{42} region (dark grey). All binding domain (BD) regions are according to Uniprot. The abundances of tryptic A β peptide 1 (residue 6–16, green) and peptide 2 (residue 17–28, brown) was normalized to global internal reference standards (GIS) and quantified in control (CTL), AsymAD, and AD cases. The \log_2 abundances of diseased cases compared with controls of Peptide 1 and Peptide 2 were illustrated as increased [$\log_2(X/CTL) = +4$, red] or decreased [$\log_2(X/CTL) = -4$, blue]. Both Peptide 1 and Peptide 2 were significantly increased in AD and AsymAD groups. **(b)** Correlation of the abundances of Peptide 1 and Peptide 2 from the same samples. **(c)** The GIS-normalized abundances of the sum of two A β peptides in CTL, AsymAD, and AD cases. The summed level of A β peptides was significantly changed between control and AD groups, with both AD groups (**** $p < 0.0001$ one-way ANOVA) and AsymAD groups (** $p < 0.01$) containing significantly-increased A β .

peptides was used to represent A β levels in each sample, which showed significant increase in both AsymAD and AD samples when compared to the control group (Fig. 4c).

Another hallmark of AD is hyperphosphorylated tau⁵², which is the core component of neurofibrillary tangles (NFTs) in diseased neurons. Remarkably, 22 phosphorylated tau peptides were detected in the total proteome even without IMAC enrichment, which highlights the robust phosphorylation of this protein in AD brain (Fig. 5a). After IMAC enrichment, a total of 112 tau phosphopeptides were identified. Of note, there were 47 peptides containing two or more phosphosites, which was approximately 42% of all phosphopeptides mapped to tau (Fig. 5a). Since the MTBR domains form the core of neurofibrillary tangles and is required to seed tau aggregation^{53,54}, it was included as an additional protein entry as MAPT-MTBR within the database, while all other tau isoforms were replaced as new “deltaMTBR” entries after the MTBR sequence was removed from the original sequences. As shown in Fig. 5b,c, both MAPTdeltaMTBR (MAPT Δ MTBR) and MAPT-MTBR show differences between AD and control groups. However, as expected, the effect size (\log_2 fold change) for the tau MTBR is larger than tau without the MTBR (Δ MTBR) in AD. The tau MTBR sequenced from the phosphoproteome, which contained stronger phosphopeptide signal, yielded even better separation between AD and control groups compared with MTBR sequenced from the total proteome.

A one-way ANOVA of peptide levels across three groups (CTL, AsymAD and AD) was also performed⁴⁰, and peptide volcano plots were calculated, showing \log_2 fold changes and \log_{10} -transformed p-values of peptides

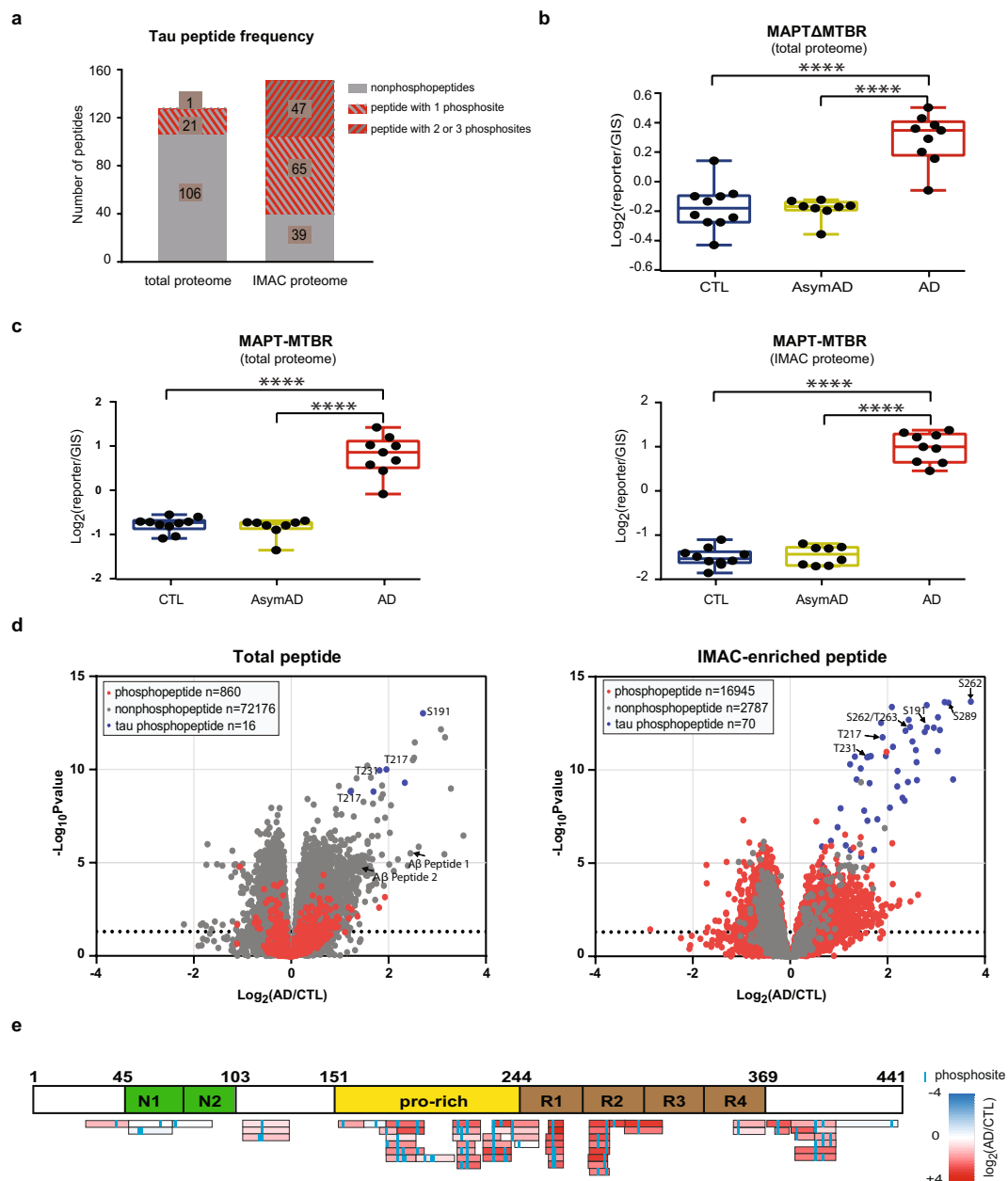


Fig. 5 Tau phosphopeptides are highly enriched in AD. **(a)** Tau peptide composition from both total proteome and IMAC proteome. In the total proteome, 17% of peptides mapping to Tau were phosphopeptides, while 74% of all Tau peptides in the IMAC proteome were phosphopeptides. Approximately 42% of the Tau phosphopeptides in the IMAC proteome contained 2 or more phosphosites. **(b)** The GIS-normalized abundances of MAPTΔMTBR in CTL, AsymAD, and AD cases. The MTBR was removed from MAPT and set as an additional sequence in the FASTA. AD group was significantly increased compared to control and AsymAD groups. **(c)** The GIS-normalized abundances of MAPT-MTBR in CTL, AsymAD, and AD cases. AD group was significantly increased compared to control and AsymAD groups. Compared to MAPTΔMTBR, MTBR domains showed better separation from both the total proteome and phosphoproteome data. **(d)** Volcano plots showing \log_2 abundance fold changes (AD/CTL) and \log_{10} -transformed p-values of non-phosphopeptides (grey) and phosphopeptides (red) after one-way ANOVA across three groups (CTL/AsymAD/AD) in the total proteome dataset (left) and IMAC-enriched phosphoproteome dataset (right). In order to achieve better accuracy, only the peptides quantified across all 3 batches were plotted. Several phosphorylated MAPT peptides (blue) identified from both total proteome and IMAC proteome were significantly increased in AD compared with CTL and labeled. Both tryptic A β peptides increased in the AD group and labeled. **(e)** The \log_2 fold changes in AD over CTL brains of tau phosphopeptides were largely increased in the Proline-rich (Pro-rich) domain (yellow) and MTBR domains (brown). Unchanged phosphopeptides also mapped to the N-terminal acidic domains (green). The \log_2 abundances of Tau peptides quantified in at least two batches were illustrated ranging from increased [$\text{Log}_2(\text{X}/\text{CTL}) = +4$, red] to decreased [$\text{Log}_2(\text{X}/\text{CTL}) = -4$, blue].

from both the total proteome and phosphoproteome datasets (Fig. 5d). In total proteome peptide data, we observed both Peptide 1 and Peptide 2 from A β to be significantly increased in AD when compared to controls. Additionally, the tau phosphopeptides were among the most changed peptides between AD and control, with significantly increased peptide abundances (Fig. 5d). In agreement with this, tau phosphopeptides were among the most significantly-changed peptides in the IMAC proteome as well. To illustrate this, all tau phosphopeptides quantified in more than two batches were colored according to the degree of fold change between AD and control from IMAC proteome (Fig. 5e). Importantly, the IMAC enrichment allowed deep sequencing and quantification of phosphopeptides mapping to the Proline-rich (Pro-rich) domain (residues 103–244) and MTBR (residues 244–368) domain⁵⁵. Both these regions showed the most consistently increasing in abundance in AD compared with other regions of the tau protein.

Usage Notes

Ultimately, these deep human brain proteomic and phosphoproteomic datasets serve as a valuable resource for a variety of research endeavors including, but not limited to, the following applications:

Use case 1: Protein abundance at steady state. This dataset provides a reference for relative protein abundance in brain, especially if an investigator wants to determine whether their protein of interest is abundantly expressed in human brain³⁸.

Use case 2: AD stage-specific differential protein expression. There were three separate clinical and pathological groups of human post-mortem tissues representing three stages of AD. One can compare the expression differences between different stages at the protein, peptide or phosphopeptide level. The volcano plots shown in Fig. 5d displays the substantive changes in peptide levels between AD and control groups. The same analysis between AsymAD and control can also be applied. The ANOVA analysis between the three groups (i.e., proteins, peptides and phosphopeptides) is included⁴⁰. The average levels from each group are also included in the output, which can be used to assess stage-specific trends across the groups.

This analysis also includes the quantification of peptides with and without phosphorylation sites³⁸ in the same peptide within the same sample, which can greatly benefit the investigators working to fully describe the phosphorylation stoichiometry of certain proteins.

Use case 3: Protein co-expression network analysis. In this dataset, there were more than 10,000 proteins quantified which is more than enough to conduct systems-level analysis. WeiGhted Co-expression Network Analysis (WGCNA) and related algorithms can be utilized for systems-based network analyses, which generate modules of proteins clustered by correlated expression patterns^{5,36,46}. The protein clusters can then be correlated to molecular functions and pathways. These programs can also be used to correlate expression clusters to various biological traits. Furthermore, the cell-type specificity of individual proteins may be investigated according to the module membership of a protein and the brain cell-type enrichment data for that particular module. The systems-level co-expression analysis also includes the average abundance of proteins or phosphopeptides across disease stages that form early, intermediate and late change clusters⁴⁶.

Use case 4: Identification and quantification of signaling pathways. Pathway analysis is routine with software⁵⁶ or web services^{57,58} to analyze different high-throughput omics data, like genomics, transcriptomics, proteomics, lipidomics and metabolomics. Pathway analyses help to organize a list of proteins into a cohesive list of pathway maps to interpret proteomics results. These analyses have proved to be a very powerful interpretation tool in biological research, facilitating novel insights in disparate fields including development⁵⁹, apoptosis⁶⁰, cancer^{61,62}, and other diseases^{63,64}. Several biological pathways have been linked to AD using similar methods^{36,65–67}. Given the excellent coverage of the AD proteome and phosphoproteome from the same samples described here, this dataset may therefore serve as a useful resource for pathway analysis.

Use case 5: Domain or motif analysis. A protein domain or motif is a part of a given protein sequence that serves as a substrate for kinases or other enzymes to recognize and chemically modify, and is replicated in other sequences in the proteome, playing conserved roles in protein function^{68,69}. Recent advances in genomics and proteomics sequencing following the development of bioinformatics^{70,71} make large-scale domain or motif analyses possible. As kinases reliably phosphorylate motif sequences specific to that enzyme, the altered phosphorylation of certain motifs may reflect impaired kinase dynamics in AD. Given the enhanced coverage of the AD proteome and phosphoproteome, this dataset can be an excellent tool for AD-related domain or motif analysis.

Use case 6: Targeted proteomics. Due to the multiplexing nature of the TMT method, proteomic sample processing has become increasingly high throughput and a more popular mode of research. As innovative technical advances in instrumentation, computing and processing have steadily improved, TMT-labeled peptide analyses have begun to be applied to targeted proteomic methods, such as TOMAHAQ⁷². In this dataset, we have identified 164,034 peptides and 51,736 phosphopeptides through TMT isobaric labeling. Importantly, this dataset includes peptide-specific characteristics such as intensity, charge and modification state³⁸, which can serve as a resource to reference for targeted proteomics analyses in the future.

Code availability

The authors do not have code specific to this work to disclose.

Received: 3 June 2020; Accepted: 18 August 2020;

Published online: 28 September 2020

References

- Global, regional, and national burden of Alzheimer's disease and other dementias, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet. Neurology* **18**, 88–106 (2019).
- Guo, T. *et al.* Molecular and cellular mechanisms underlying the pathogenesis of Alzheimer's disease. *Molecular neurodegeneration* **15**, 40 (2020).
- Sperling, R. A. *et al.* Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging–Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **7**, 280–292 (2011).
- Sperling, R. A. *et al.* The impact of amyloid-beta and tau on prospective cognitive decline in older individuals. *Annals of neurology* **85**, 181–193 (2019).
- Seyfried, N. T. *et al.* A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. *Cell systems* **4**, 60–72.e64 (2017).
- Bush, S. J., Chen, L., Tovar-Corona, J. M. & Urrutia, A. O. Alternative splicing and the evolution of phenotypic novelty. *Philos Trans R Soc Lond B Biol Sci* **372**, 1–7 (2017).
- Li, H. D., Menon, R., Omenn, G. S. & Guan, Y. The emerging era of genomic data integration for analyzing splice isoform function. *Trends Genet* **30**, 340–347 (2014).
- Cox, J. & Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annual review of biochemistry* **80**, 273–299 (2011).
- Maiolica, A., Junger, M. A., Ezkurdia, I. & Aebersold, R. Targeted proteome investigation via selected reaction monitoring mass spectrometry. *Journal of proteomics* **75**, 3495–3513 (2012).
- Oliveira, J., Costa, M., de Almeida, M. S. C., da Cruz, E. S. O. A. B. & Henriques, A. G. Protein Phosphorylation is a Key Mechanism in Alzheimer's Disease. *Journal of Alzheimer's disease: JAD* **58**, 953–978 (2017).
- Sathe, G. *et al.* Multiplexed Phosphoproteomic Study of Brain in Patients with Alzheimer's Disease and Age-Matched Cognitively Healthy Controls. *Omics: a journal of integrative biology* **24**, 216–227 (2020).
- Herskowitz, J. H. *et al.* Pharmacologic inhibition of ROCK2 suppresses amyloid-beta production in an Alzheimer's disease mouse model. *The Journal of neuroscience: the official journal of the Society for Neuroscience* **33**, 19086–19098 (2013).
- Hanger, D. P., Hughes, K., Woodgett, J. R., Brion, J.-P. & Anderton, B. H. Glycogen synthase kinase-3 induces Alzheimer's disease-like phosphorylation of tau: Generation of paired helical filament epitopes and neuronal localisation of the kinase. *Neuroscience Letters* **147**, 58–62 (1992).
- Noble, W. *et al.* Cdk5 Is a Key Factor in Tau Aggregation and Tangle Formation *In Vivo*. *Neuron* **38**, 555–565 (2003).
- Cole, G., Dobkins, K. R., Hansen, L. A., Terry, R. D. & Saitoh, T. Decreased levels of protein kinase C in Alzheimer brain. *Brain Research* **452**, 165–174 (1988).
- Chin, J. Y. *et al.* Microtubule-Affinity Regulating Kinase (MARK) Is Tightly Associated with Neurofibrillary Tangles in Alzheimer Brain: A Fluorescence Resonance Energy Transfer Study. *Journal of Neuropathology & Experimental Neurology* **59**, 966–971 (2000).
- Gong, C. X., Singh, T. J., Grundke-Iqbal, I. & Iqbal, K. Phosphoprotein phosphatase activities in Alzheimer disease brain. *J Neurochem* **61**, 921–927 (1993).
- Michalski, A., Cox, J. & Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *Journal of proteome research* **10**, 1785–1793 (2011).
- O'Connell, J. D., Paulo, J. A., O'Brien, J. J. & Gygi, S. P. Proteome-Wide Evaluation of Two Common Protein Quantification Methods. *Journal of proteome research* **17**, 1934–1942 (2018).
- Lim, H. *et al.* Identification of 2D-gel proteins: a comparison of MALDI/TOF peptide mass mapping to mu LC-ESI tandem mass spectrometry. *Journal of the American Society for Mass Spectrometry* **14**, 957–970 (2003).
- Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y. & Aebersold, R. Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 9390–9395 (2000).
- Ping, L. *et al.* Global quantitative analysis of the human brain proteome in Alzheimer's and Parkinson's Disease. *Scientific data* **5**, 180036 (2018).
- Mertins, P. *et al.* Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. *Nature protocols* **13**, 1632–1661 (2018).
- Tan, A. & Fanaras, J. C. Use of high-pH (basic/alkaline) mobile phases for LC-MS or LC-MS/MS bioanalysis. *Biomedical chromatography: BMC* **33**, e4409 (2019).
- Horvatovich, P., Hoekman, B., Govorukhina, N. & Bischoff, R. Multidimensional chromatography coupled to mass spectrometry in analysing complex proteomics samples. *Journal of separation science* **33**, 1421–1437 (2010).
- Bai, B. *et al.* Deep Profiling of Proteome and Phosphoproteome by Isobaric Labeling, Extensive Liquid Chromatography, and Mass Spectrometry. *Methods in enzymology* **585**, 377–395 (2017).
- Mirra, S. S. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathological assessment of Alzheimer's disease. *Neurology* **41**, 479–486 (1991).
- Braak, H. & Braak, E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* **82**, 239–259 (1991).
- Ping, L., Duong, D. M. & Seyfried, N. T. Sample traits. *Synapse* <https://doi.org/10.7303/syn20820456> (2020).
- Ping, L., Duong, D. M. & Seyfried, N. T. RAW files. *Synapse* <https://doi.org/10.7303/syn20820454> (2020).
- UniProt: a worldwide hub of protein knowledge. *Nucleic acids research* **47**, D506–d515 (2019).
- Müller, J. B. *et al.* The proteome landscape of the kingdoms of life. *Nature* **582**, 592–596 (2020).
- Breuzza, L. *et al.* The UniProtKB guide to the human proteome. *Database: the journal of biological databases and curation* **2016** (2016).
- Ping, L., Duong, D. M. & Seyfried, N. T. Database. *Synapse* <https://doi.org/10.7303/syn20820455> (2020).
- Dai, J. *et al.* Effects of APOE Genotype on Brain Proteomic Network and Cell Type Changes in Alzheimer's Disease. *Frontiers in molecular neuroscience* **11**, 454 (2018).
- Johnson, E. C. B. *et al.* Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nature medicine* **26**, 769–780 (2020).
- Wiese, H. *et al.* Comparison of alternative MS/MS and bioinformatics approaches for confident phosphorylation site localization. *J Proteome Res* **13**, 1128–1137 (2014).
- Ping, L., Duong, D. M. & Seyfried, N. T. Search result. *Synapse* <https://doi.org/10.7303/syn20820472> (2020).
- Ping, L. *et al.* Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *Synapse* <https://doi.org/10.7303/syn20820053> (2020).
- Ping, L., Duong, D. M. & Seyfried, N. T. Anova analysis. *Synapse* <https://doi.org/10.7303/syn21986491> (2020).
- Ping, L. *et al.* Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *PRIDE Archive* <http://identifiers.org/pride.project:PX020296> (2020).
- Zhang, H. *et al.* Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian. *Cancer. Cell* **166**, 755–765 (2016).
- Clark, D. J. *et al.* Evaluation of NCI-7 Cell Line Panel as a Reference Material for Clinical Proteomics. *Journal of proteome research* **17**, 2205–2215 (2018).
- Mertins, P. *et al.* Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 (2016).

45. Wingo, A. P. *et al.* Large-scale proteomic analysis of human brain identifies proteins associated with cognitive trajectory in advanced age. *Nature communications* **10**, 1619 (2019).
46. Higginbotham, L. *et al.* Integrated Proteomics Reveals Brain-Based Cerebrospinal Fluid Biomarkers in Asymptomatic and Symptomatic Alzheimer's Disease. bioRxiv, 806752 (2019).
47. Seyfried, N. T. *et al.* Multiplex SILAC analysis of a cellular TDP-43 proteinopathy model reveals protein inclusions associated with SUMOylation and diverse polyubiquitin chains. *Molecular & cellular proteomics: MCP* **9**, 705–718 (2010).
48. Dammer, E. B. *et al.* Quantitative phosphoproteomics of Alzheimer's disease reveals cross-talk between kinases and small heat shock proteins. *Proteomics* **15**, 508–519 (2015).
49. Brenes, A., Hukelmann, J., Bensaddek, D. & Lamond, A. I. Multibatch TMT Reveals False Positives, Batch Effects and Missing Values. *Molecular & cellular proteomics: MCP* **18**, 1967–1980 (2019).
50. Glenner, G. G. & Wong, C. W. Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochemical and biophysical research communications* **120**, 885–890 (1984).
51. Grundke-Iqbal, I. *et al.* Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proc Natl Acad Sci USA* **83**, 4913–4917 (1986).
52. Hampel, H. *et al.* Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimer's Dement* **4**, 38–48 (2008).
53. Kolarova, M., Garcia-Sierra, F., Bartos, A., Ricny, J. & Ripova, D. Structure and pathology of tau protein in Alzheimer disease. *International journal of Alzheimer's disease* **2012**, 731526 (2012).
54. Falcon, B. *et al.* Conformation determines the seeding potencies of native and recombinant Tau aggregates. *The Journal of biological chemistry* **290**, 1049–1065 (2015).
55. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185–190 (2017).
56. Zhou, X. & Su, Z. EasyGO: Gene Ontology-based annotation and functional enrichment analysis tool for agronomical species. *BMC genomics* **8**, 246 (2007).
57. Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**, 1–13 (2009).
58. Franceschini, A. *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research* **41**, D808–815 (2013).
59. Pouloupoulos, A. *et al.* Subcellular transcriptomes and proteomes of developing axon projections in the cerebral cortex. *Nature* **565**, 356–360 (2019).
60. Dix, M. M. *et al.* Functional interplay between caspase cleavage and phosphorylation sculpts the apoptotic proteome. *Cell* **150**, 426–440 (2012).
61. Kolch, W. & Pitt, A. Functional proteomics to dissect tyrosine kinase signalling pathways in cancer. *Nature reviews. Cancer* **10**, 618–629 (2010).
62. Koomen, J. M. *et al.* Proteomic contributions to personalized cancer care. *Molecular & cellular proteomics: MCP* **7**, 1780–1794 (2008).
63. Blagoev, B., Ong, S. E., Kratchmarova, I. & Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nature biotechnology* **22**, 1139–1145 (2004).
64. Lopez-Villar, E. *et al.* A proteomic approach to obesity and type 2 diabetes. *Journal of cellular and molecular medicine* **19**, 1455–1470 (2015).
65. Inestrosa, N. C. & Toledo, E. M. The role of Wnt signaling in neuronal dysfunction in Alzheimer's Disease. *Molecular neurodegeneration* **3**, 9 (2008).
66. Bai, B. *et al.* Deep Multilayer Brain Proteomics Identifies Molecular Networks in Alzheimer's Disease Progression. *Neuron* **105**, 975–991.e977 (2020).
67. Zhu, X., Lee, H. G., Raina, A. K., Perry, G. & Smith, M. A. The role of mitogen-activated protein kinase pathways in Alzheimer's disease. *Neuro-Signals* **11**, 270–281 (2002).
68. Tompa, P., Davey, N. E., Gibson, T. J. & Babu, M. M. A million peptide motifs for the molecular biologist. *Molecular cell* **55**, 161–169 (2014).
69. Makalowski, W., Zhang, J. & Boguski, M. S. Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. *Genome research* **6**, 846–857 (1996).
70. Krystkowiak, I., Manguy, J. & Davey, N. E. PSSMSearch: a server for modeling, visualization, proteome-wide discovery and annotation of protein motif specificity determinants. *Nucleic acids research* **46**, W235–w241 (2018).
71. Bidin, S., Vujaklija, I., Paradzik, T., Bielen, A. & Vujaklija, D. Leitmotif: protein motif scanning 2.0. Bioinformatics (Oxford, England) (2020).
72. Rose, C. M. *et al.* TomahaqCompanion: A Tool for the Creation and Analysis of Isobaric Label Based Multiplexed Targeted Assays. *Journal of proteome research* **18**, 594–605 (2019).

Acknowledgements

We are grateful to the patients and families that donate tissue samples to the Emory University brain bank and for their contributions to this study. Support for this research was provided by funding from the National Institute on Aging (R01AG053960, R01AG061800, RF1AG057471, RF1AG057470, R01AG057339, RF1AG062181), the Accelerating Medicine Partnership for AD (U01AG046161 and U01AG061357) and the Emory Alzheimer's Disease Research Center (P50AG025688).

Author contributions

Conceptualization, L.P., D.M.D., J.J.L., A.I.L., N.T.S.; Methodology, L.P., D.M.D., L.Y., M.G., N.T.S.; Investigation and Formal Analysis, L.P., D.M.D., S.R.K., N.T.S.; Writing – Original Draft, L.P., S.R.K., N.T.S.; Writing – Review and Editing, L.P., S.R.K., D.M.D., J.J.L., A.I.L., N.T.S.; Funding Acquisition, A.I.L. and N.T.S.; Supervision, N.T.S.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41597-020-00650-8>.

Correspondence and requests for materials should be addressed to N.T.S.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2020