

Discovery of Proteoforms Associated With Alzheimer's Disease Through Quantitative Top-Down Proteomics

Authors

James M. Fulcher, Ashley N. Ives, Shinya Tasaki, Shane S. Kelly, Sarah M. Williams, Thomas L. Fillmore, Mowei Zhou, Ronald J. Moore, Wei-Jun Qian, Ljiljana Paša-Tolić, Lei Yu, Shahram Oveisgharan, David A. Bennett, Philip L. De Jager, and Vladislav A. Petyuk

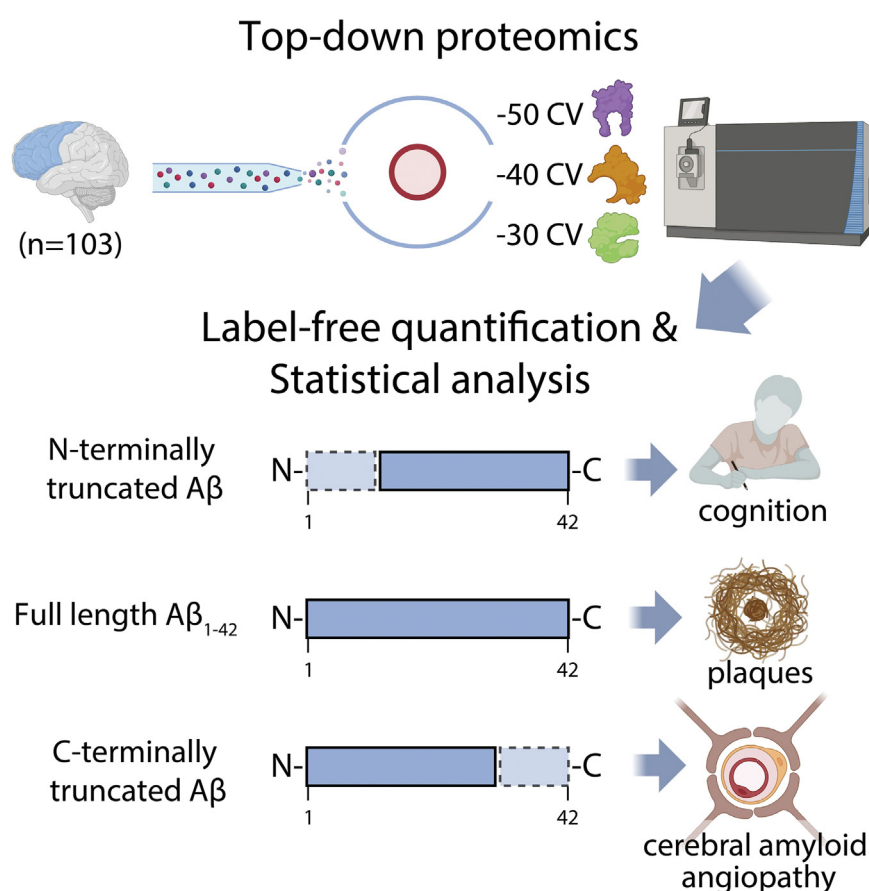
Correspondence

vladislav.petyuk@pnnl.gov

In Brief

Alzheimer's disease involves diverse proteoforms, including amyloid beta and tau, which are challenging to analyze using traditional proteomic techniques. We applied top-down proteomics to 103 frontal cortex samples from the ROS/MAP cohorts, identifying 11,782 proteoforms. Key findings include amyloid beta C-terminal truncation differentiating amyloid plaques from cerebral amyloid angiopathy and N-terminal truncations correlating with cognitive function. Additionally, six VGF neuropeptides were linked to cognition independent of pathology, providing insights into Alzheimer's disease molecular underpinnings.

Graphical Abstract



Highlights

- Top-down proteomics: 11,782 proteoforms, 1,213 proteins in 103 brain samples.
- C-terminal truncated amyloid beta key for amyloid plaque vs. CAA distinction.
- N-truncated Aβ_{x-42} linked to cognitive decline.
- Six VGF neuropeptides tied to cognitive function, independent of pathology.

Discovery of Proteoforms Associated With Alzheimer's Disease Through Quantitative Top-Down Proteomics

James M. Fulcher¹, Ashley N. Ives¹, Shinya Tasaki^{2,3}, Shane S. Kelly⁴, Sarah M. Williams¹, Thomas L. Fillmore⁴, Mowei Zhou⁵, Ronald J. Moore⁴, Wei-Jun Qian⁴, Ljiljana Paša-Tolić¹, Lei Yu^{2,3}, Shahram Oveisgharan^{2,3}, David A. Bennett^{2,3}, Philip L. De Jager⁶, and Vladislav A. Petyuk^{4,*}

The complex nature of Alzheimer's disease (AD) and its heterogeneous clinical presentation has prompted numerous large-scale *-omic* analyses aimed at providing a global understanding of the pathophysiological processes involved. AD involves isoforms, proteolytic products, and posttranslationally modified proteins such as amyloid beta (A β) and microtubule-associated protein tau. Top-down proteomics directly measures these species and thus, offers a comprehensive view of pathologically relevant proteoforms that are difficult to analyze using traditional proteomic techniques. Here, we broadly explored associations between proteoforms and clinicopathological traits of AD by deploying a quantitative top-down proteomics approach across frontal cortex of 103 subjects selected from the ROS and MAP cohorts. The approach identified 1213 proteins and 11,782 proteoforms, of which 154 proteoforms had at least one significant association with a clinicopathological phenotype. One important finding included identifying A β C-terminal truncation state as the key property for differential association between amyloid plaques and cerebral amyloid angiopathy. Furthermore, various N-terminally truncated forms of A β had noticeably stronger association with amyloid plaques and global cognitive function. Additionally, we discovered six VGF neuropeptides that were positively associated with cognitive function independent of pathological burden. The database of brain cortex proteoforms provides a valuable context for functional characterization of the proteins involved in AD and other late-onset brain pathologies.

Alzheimer's disease (AD) is the most prevalent neurodegenerative disorder, and its incidence has more than doubled

over the past three decades in part due to increased lifespan and demographic changes (1). Despite significant investments in dementia research and therapeutic development, a succinct understanding of Alzheimer's remains elusive, and no curative treatments currently exist. Many hypotheses exist on the nature of Alzheimer's etiology, with current leading theories invoking amyloid beta (A β), a fragment derived from amyloid precursor protein, and hyperphosphorylation of microtubule-associated protein tau (2). Although these hallmark protein species are well recognized as being components of senile plaques and neurofibrillary tangles, respectively, AD rarely presents with just these neuropathologies. It is commonly observed in combination with TDP-43 inclusions, hippocampal sclerosis, Lewy bodies/Lewy neurites involving α -synuclein protein, and vascular changes including cerebral amyloid angiopathy (CAA) (3). Furthermore, several clinical subtypes of AD have also been described based on neuroimaging and neuropathology studies (4). This complexity has prompted studies that aim to measure large pools of biological molecules in larger sample groups *via* genomics (5), transcriptomics (6), and proteomics (7–9) to provide new insights into the origins and pathophysiological processes behind AD.

Considering that AD hallmarks involve unusual protein forms, proteomics provides a view that is arguably most directly connected to the underlying pathology. Proteomic techniques most commonly applied require the use of an enzyme such as trypsin to cleave proteins into short peptides that are more amenable for liquid-chromatography mass spectrometric analysis (LC-MS/MS). This is referred to as “bottom-up” proteomics (BUP). Though this approach allows for the quantification of thousands of proteins in a single

From the ¹Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington, USA; ²Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, USA; ³Department of Neurological Sciences, Rush University Medical Center; Chicago, Illinois, USA; ⁴Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington, USA; ⁵Department of Chemistry, Zhejiang University, Hangzhou, Zhejiang, China; ⁶Center for Translational and Computational Neuroimmunology, Department of Neurology & Taub Institute for Research on Alzheimer's disease and the Aging Brain, Columbia University Medical Center; New York, New York, USA

* For correspondence: Vladislav A. Petyuk, vladislav.petyuk@pnnl.gov.

analysis, the digestion step eliminates the ability to capture the entire molecular composition of the protein. Therefore, the detection of properties such as cleavage by endogenous proteases, single nucleotide variants, post-translational modifications (PTMs), and alternative splicing events are drastically diminished. Collectively, these distinct molecular entities are referred to as “proteoforms” and are best measured utilizing top-down proteomics (TDP) which keeps proteins in their intact primary sequence with PTMs by eschewing enzymatic digestion (10). Because the two hallmarks of AD, A β and hyperphosphorylated tau, are noncanonical protein forms, TDP offers many advantages that have yet to be realized in a global TDP analysis of AD. Thus, to broadly explore the association of proteoforms with AD, we deployed a discovery TDP approach to quantify proteoforms across 103 subjects selected from the Religious Orders Study and Rush Memory and Aging Project (ROS/MAP) cohort (11, 12).

TDP, however, is not without its own compromises. It is more limited in proteome coverage compared to the conventional BUP approaches, and tools that enable quantification in a medium/large-scale (>100 samples) TDP studies are at the early inception stage. With regards to low proteome coverage, this can be attributed to several instrument limitations such as rapid ion transient decay in Orbitraps with larger proteins (>30 kDa), signal dilution from having multiple charge states each with isotopologs, and chromatographic peak broadening which leads to overlapping signals during MS acquisition (13). Additionally, A β proteoforms are challenging to isolate and analyze due to their inherent propensity for aggregation. Here, we integrate several innovations to mitigate these limitations (Fig. 1). First, the addition of high-field asymmetric waveform ion mobility spectrometry (FAIMS) provides significantly deeper proteome coverage by fractionating ions using compensation voltages (CVs) in the

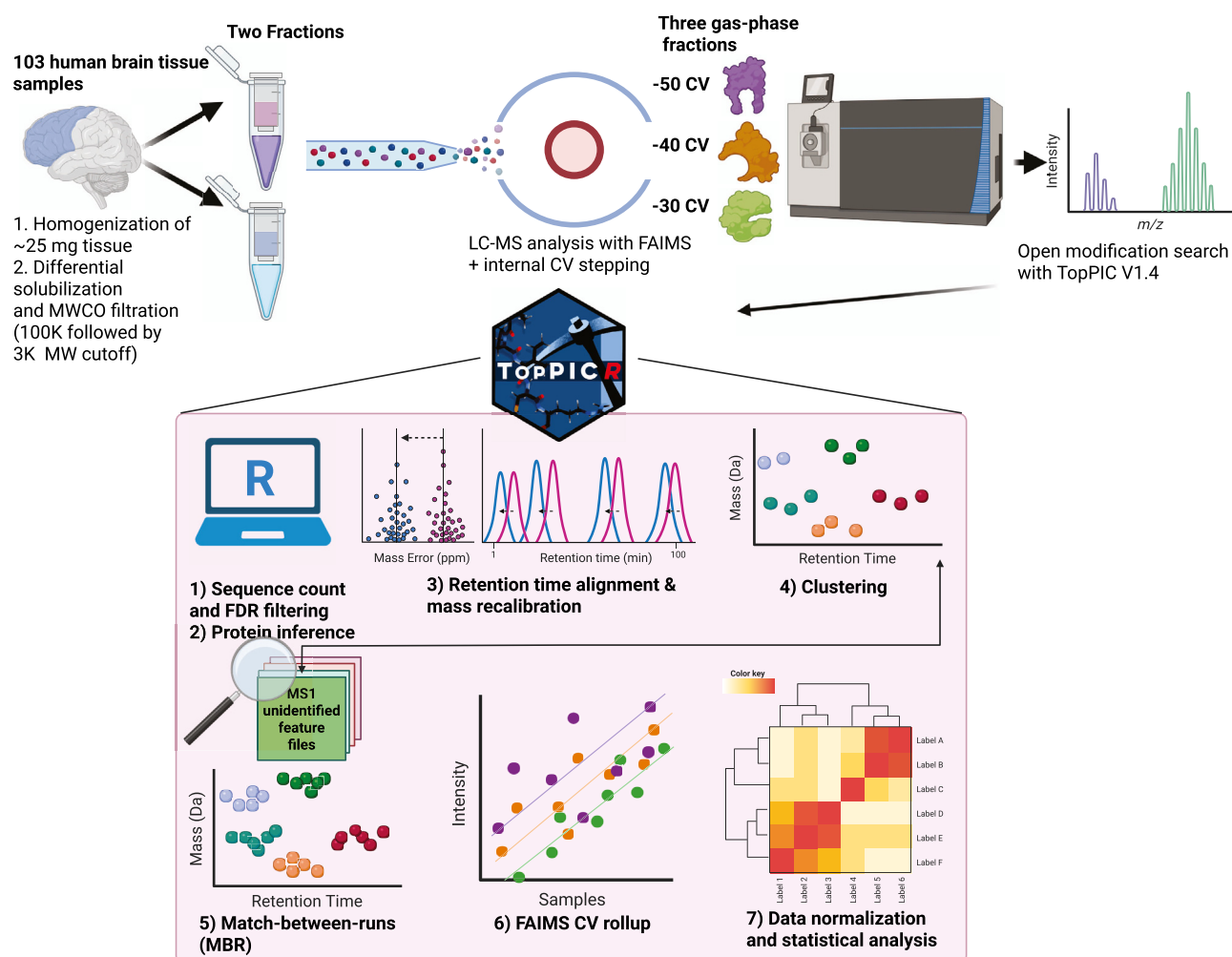


FIG. 1. Workflow for collecting and analyzing top-down proteomics datasets of 103 human brains. Human brain tissue was processed by differential solubilization, separated into two fractions, and then analyzed by LC-MS with gas-phase fractionation using FAIMS. After database searching with TopPIC, proteoform spectrum match (PrSM) result files are imported into R for downstream analysis with TopPICR. Figure was partially created with BioRender.

gas-phase based on their relative size and shape (14, 15). Second, we also implemented a new differential solubilization technique utilizing hexafluoroisopropanol (HFIP). HFIP is a potent hydrogen donor that facilitates dissociation of the oligomeric and fibrillar forms of A β (16). We found that homogenizing human brain tissue with HFIP could extract A β alongside other HFIP-soluble proteoforms into the supernatant, while the pellet could be rehomogenized to provide a fraction with many nonoverlapping proteoforms. Finally, a top-down quantification software (TopPICR (17)) employed several crucial algorithms to reduce data missingness and increase robustness of quantification across the 103 brain samples analyzed. Through the combination of HFIP-differential solubilization, FAIMS, and TopPICR, we demonstrate a novel large-scale TDP approach that enabled identification of 11,782 proteoforms (derived from 1213 proteins) ranging in size from 1 kDa to 28 kDa across 103 subjects and aided one of the deepest TDP studies to date (18, 19).

EXPERIMENTAL PROCEDURES

Study Participants

All antemortem cognitive and postmortem data analyzed in this study were gathered as part of the ROS/MAP (11, 20, 21), two longitudinal cohort studies of the elderly within the United States. Both studies were approved by an Institutional Review Board of Rush University Medical Center (ROS IRB# L91020181, MAP IRB# L86121802). Both studies were conducted according to the principles expressed in the Declaration of Helsinki. Written informed consent, an Anatomic Gift Act, and a repository consent was obtained from all ROS/MAP. Clinical and neuropathologic data collection has been previously reported in (22–26). The entire ROSMAP cohort has also been also extensively characterized with multi-omics approaches (20, 27).

Mouse Brain Samples

WT C57BL/6 J mouse brains were acquired directly from The Jackson Laboratory. The study was reviewed and approved by the Institutional Animal Care and Use Committee of Battelle, Pacific Northwest Division (#2021-10).

Experimental Design and Statistical Rationale

For this TDP study, we selected 103 subjects. The rationale behind the subject selection was primarily to ensure that the subjects are well-characterized including genomic, transcriptomic, and epigenomic data. Specifically, we leveraged the same subjects that were used in previous microglia activation studies (28, 29). Selected subjects were representative of the greater study population with respect to age at death and postmortem brain pathology measurements of tangles and A β load (Table 1). Specifically, the differences in the following variables were insignificant: age at death (K-S test, $p = 0.31$), AD diagnosis (χ^2 test, $p = 0.12$), tangles (K-S test, $p = 0.57$), and A β load (K-S test, $p = 0.12$). However, there was a significant difference in sex distribution: 72% female in ROSMAP versus 61% in the selected 103 subjects (χ^2 test, $p = 0.01$). This, however, is closer to the general population, where 55% of individuals over 65 years old are female.

Prior to processing, the 103 patient samples were randomized and assigned to six batches (five batches of 18 and one batch of 13 samples). Each sample was analyzed once, resulting in two LC-MS

TABLE 1

Characteristics and statistics of the 103 participants from the ROSMAP cohort included in this study

Variable	Mean (SD) or n (%)
Demographic	
Age at death (years)	89.4 (5.5)
Female sex (n, %)	63 (61%)
Education (years)	14.9 (2.6)
Postmortem interval (hours)	6.9 (4.6)
APOE ϵ 4 allele (n, %)	26 (25%)
Cognition	
Last visit global cognition (score)	−0.83 (1.0)
Slope of cognitive decline (score/year)	−0.015 (0.097)
Postmortem pathological indices	
β -amyloid load (score)	4.7 (4.8)
PHFtau tangle density (score)	6.3 (7.0)
Cerebral amyloid angiopathy (moderate-severe n, %)	36 (35%)
Macroscopic infarcts (n, %)	35 (34%)
Hippocampal sclerosis (n, %)	9 (9%)
Arteriosclerosis (moderate-severe n, %)	33 (32%)
Neocortical Lewy bodies (n, %)	9 (9%)

datasets: one for HFIP-soluble and one for urea-soluble fractions. To test the significance of the association between proteoforms and clinical or pathological variables, we used the variables as predictors and modeled proteoforms' log₂-transformed relative abundances or spectral counts as outcomes. When justified, we included additional clinicopathologic variables or batch effect covariates as predictors. The details of the data analysis itself are provided in the "Preprocessing and Statistical Analysis" section below.

Sample Preparation

After sample allocation, 15 to 35 mg brain tissue frozen punch biopsies derived from dorsal lateral prefrontal cortex were homogenized on ice with a pellet pestle homogenizer in a 1.5 ml LoBind Eppendorf tube after addition of 8 μ l HFIP per 1 mg of wet tissue. Once tissue was fully disrupted, an equivolume amount of homogenization buffer (HB, consisting of 8 M urea, 10 mM ammonium bicarbonate (ABC), 10 mM tris(2-carboxyethyl)phosphine, 2 mM EDTA) was added and then vortexed for 1 min. Samples were incubated at 15 °C for 5 min with 1200 RPM of shaking on a Thermomixer, diluted with HB to a final concentration of 20% HFIP, and then incubated again at 15 °C for 10 min. Insoluble protein was pelleted with centrifugation at 18,000 RCF (10 °C) for 15 min. The HFIP-soluble fraction containing enriched amyloid beta (supernatant) was transferred to a separate 1.5 ml tube on ice. Forty microliters of HB/mg tissue was then added to the pellet before vortexing for 1 min followed by incubation at 15 °C for 5 min at 1200 RPM to extract proteins that were insoluble with the HFIP cosolvent. Both fractions derived from each brain tissue sample were then centrifuged at 18,000 RCF (10 °C) for 15 min before being transferred to separate 4 ml 100 K MWCO spin filters (Amicon) containing wash buffer (consisting of 8 M urea, 10 mM ABC). After each wash step (three total), the flowthrough was transferred to a separate 3 K MWCO filter where it was concentrated/washed three times. After the final wash, the ~0.1 to 0.2 ml 3 K MWCO filter retentate was transferred to a 1.5 ml LoBind Eppendorf tube where it was acidified to 0.5% formic acid (FA) with a 10% FA stock solution. A bicinchoninic acid assay was then performed using BSA protein standards (0.125–2 mg/ml in 8 M urea, 10 mM ABC, 0.5% FA). Samples were

then normalized by dilution with 8 M urea, 10 mM ABC, 0.5% FA to 0.25 mg/ml (HFIP-soluble fraction), or 0.5 mg/ml (HB-soluble fraction) concentration.

LC-MS Analysis

Samples were analyzed using a Waters NanoACQUITY UPLC system with mobile phases consisting of 0.2% FA in H₂O (mobile phase A) and 0.2% FA in ACN (mobile phase B). Both trapping-precursor (200 μ m i.d., 5-cm length) and analytical column (100 μ m i.d., 50-cm length) were slurry-packed with C2 packing material (5 μ m and 3 μ m for trap/analytical respectively, 300 Å, Separation Methods Technology). Samples were loaded into a 10 μ l loop, corresponding to 2.5 or 5 μ g of loaded material and injected onto the trapping column with an isocratic flow of 1% B at 5 μ l/min over 15 min for desalting. Separation was performed with a 1% to 50% B gradient over 160 min at 300 nl/min. For MS/MS analysis of proteins, the NanoACQUITY system was coupled to a Thermo Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer equipped with the FAIMS Pro interface. Source parameters included electrospray voltage of 2.2 kV, transfer capillary temperature of 275 °C, and ion funnel RF amplitude of 30%. FAIMS was set to standard resolution without supplementary user-controlled carrier gas flow and a dispersion voltage of -5 kV (equivalent to a dispersion field of -33.3 kV/cm), while the CV switched between three voltages (-50, -40, and -30) throughout data collection (referred to as "internal CV stepping") (15).

The Fusion Lumos was set to "Intact Protein" application mode, and data was collected as full profile. MS1 and MS2 data were acquired at a resolution of 120k and 60k across a 500-2000 m/z range, with four and two microscans, and AGC targets of 1E6 and 5E5, respectively. MS1 and MS2 were acquired with a maximum inject time of 250 ms. Data-dependent settings included selection of top six most intense ions, exclusion of ions lower than charge state 5+, exclusion of undetermined charge states, and dynamic exclusion after one observation for 30 s. Ions selected for MS2 were isolated over a $\pm$ 1.5 m/z window and fragmented through collision-induced dissociation with a collision energy of 35%.

Because of the relatively large sample size of this TDP study, several quality control measures were implemented to ensure the quantitative robustness of the approach. To confirm LC-MS instrument stability, we collected quality control datasets every 3 to 8 samples using an intact protein mixture derived from *Shewanella oneidensis*. Instrument operation was continued only if the number of TopPIC proteoform identifications were above a predefined threshold that was determined when the instrument was known to run optimally. Furthermore, the HFIP differential solubilization sample preparation method was empirically evaluated for its reproducibility. We performed the entire workflow, along with a nonfractionated control, on three biological replicate mouse brain samples. Three technical replicates were collected for each mouse brain sample, and these were spread out over a period of months on the same LC-MS instrument used for our analyses.

Database Search Parameters and Acceptance Criteria for Proteoform Identifications

Proteoform identification was performed with TopPIC (30) version 1.5.4. Settings for TopPIC included a precursor window of 3 m/z, mass error tolerance of 15 ppm for precursor and fragment ions, a proteoform cluster error tolerance of 0.8 Da, a mass shift upper bound of 4000 Da and lower bound of -150 Da, and a maximum number of allowed unknown modifications of 1. MS2 spectra were searched against the Swiss-Prot database for *Homo sapiens* containing 20,371 reviewed entries, a variably spliced ("VarSplic") database containing 21,980 splice-isoform entries, and a TrEMBL database containing 57,749 entries (UP000005640- accessed August 29th, 2021). All

databases were scrambled to generate decoys which were concatenated during the search. A list of 14 dynamic modifications (N-terminal methionine excision; N-terminal acetylation, acetylation at lysine; C-terminal amidation; N-terminal carbamylation; carbamylation at lysine; deamidation at glutamine or asparagine; methylation or dimethylation at lysine or arginine; iron adduction at aspartic or glutamic acid; phosphorylation at serine, threonine, or tyrosine; N-terminal pyroglutamate; and dioxidation or oxidation at cysteine, methionine, or tryptophan) were provided during the open modification search to reduce the number of unknown mass shifts; no static modifications were specified. The spectra confidence threshold in the TopPIC searches was set at maximum allowable E-value of 0.05.

Downstream data analysis steps were performed in the R environment for statistical computing and TopPICR (17) package. After combining the datasets, we required that a proteoform sequence should be observed in at least two samples, then the E-values were adjusted such that the false discovery rate does not exceed 1% at gene level for each of the three individual types of annotations of the UniProt database: canonical protein sequences (Swiss-Prot), splice isoforms (VarSplic), and tentative protein sequences (TrEMBL). After these filtering steps followed by parsimonious protein accession inference, we achieved the following final false discovery rates: proteoform-to-spectrum match 0.014%, proteoform 0.02%, sequence (proteoform with stripped modifications) 0.21%, UniProt accession 1.7%, and gene level 1.9%. Manual MS/MS validation was performed using LcMsSpectator 1.1.7158.24217 using a precursor and product ion tolerance of 10 ppm. A minimum S/N threshold of 1.5 and Pearson correlation threshold of 0.7 were used for fragment ion assignment.

Preprocessing and Statistical Analysis

The TopPIC output was converted to MSnSet Bioconductor object using TopPICR package (17) for convenience of the downstream processing and quantitative analysis. Two types of objects were generated: (1) based on MS1 feature intensities and the other (2) using spectral counts as abundance measures. Principal component analysis revealed that processing batch is a significant contributor explaining the variance in the data. This was confirmed by an ANOVA statistical test using the limma package (31). The batch effect was corrected in MS1-based data using the ComBat approach (sva package) (32). Unfortunately, no statistical methods have been developed to correct batch effects in proteomics spectral counting data. Thus, the batch was used as a confounding factor in the statistical modeling of the count data.

As for statistical testing, we used a linear modeling approach for MS1 data (limma package). As for the spectral counting, we used quasi-Poisson-generalized linear modeling approach, known as QuasiTel (33). Besides the sample processing batching factor, we considered other technical (e.g. postmortem interval) and demographic (e.g. age at death and sex) covariates. Out of all the considered variables, only postmortem interval was statistically significant. Therefore, it was added to the models as a covariate.

Proteoform clustering was performed using the weighted gene correlation network analysis WGCNA tool (34) on MS1 intensity-based data. We required that each proteoform have measurement values in more than 50 samples, which resulted in a final set of 6029 proteoforms. To ensure that clustering was not driven by confounding factors, the effect of postmortem interval was regressed out from the log₂ ratios of the relative abundances. Missing values were imputed using the svdImpute method from the pcaMethods package (35). Finally, log₂ ratios were transformed into z-scores by scaling the proteoforms' SDs to ensure that distances reflected the correlation between proteoforms. The WGCNA settings were kept at default, except for the cut-tree distance in the merging step, which was lowered from 0.2 to 0.15.

Code Availability

The analysis pipeline used in this study and the code reproducing most of the figures is publicly available on GitHub at PNNL-Comp-Mass-Spec/ROSMAP103_TopDown repository.

RESULTS

Coverage of the Proteome

The combination of two HFIP fractions, three FAIMS CVs, followed by LC-MS/MS analysis on average identified 6361 ± 784 proteoforms corresponding to 870 ± 71 genes (Fig. 2A) per subject. Combined over 103 samples, we identified 11,782 unique proteoforms corresponding to 1213 genes (Fig. 2B).

We applied two approaches for quantifying proteoforms in this study: MS1 intensity-based and spectral count-based. Conceptually, the approaches are similar to their BUP counterparts (36, 37). While the intensity-based approach can be regarded as more accurate and precise, the advantage of the

semiquantitative spectral counting approach is that count-based statistical modeling does not require a special treatment of the missing values, since they are treated as zero counts. For the intensity-based data, we required the intensity values to be present in at least 50 samples out of 103 samples. Overall, 6124 proteoforms derived from 836 genes were measured using MS1 intensity-based approach with missing values in fewer than 50% of the samples (Fig. 2C) (Supplemental Table S1). In the case of spectral counting, we imposed the requirement that a proteoform must present in at least six samples (3 batches with at least two counts). This resulted in quantifying 7510 proteoforms corresponding to 952 genes (Supplemental Table S2).

The number of identified proteins translates into about 6% of the human proteome (20,361 proteins). However, the current shotgun (*i.e.* nontargeted analysis of an isolated protein) TDP is generally limited to proteins less than 25 kDa (38, 39). Indeed, almost all proteoforms we identified are below 25 kDa (Supplemental Fig. S1). This corresponds to ~20% of the

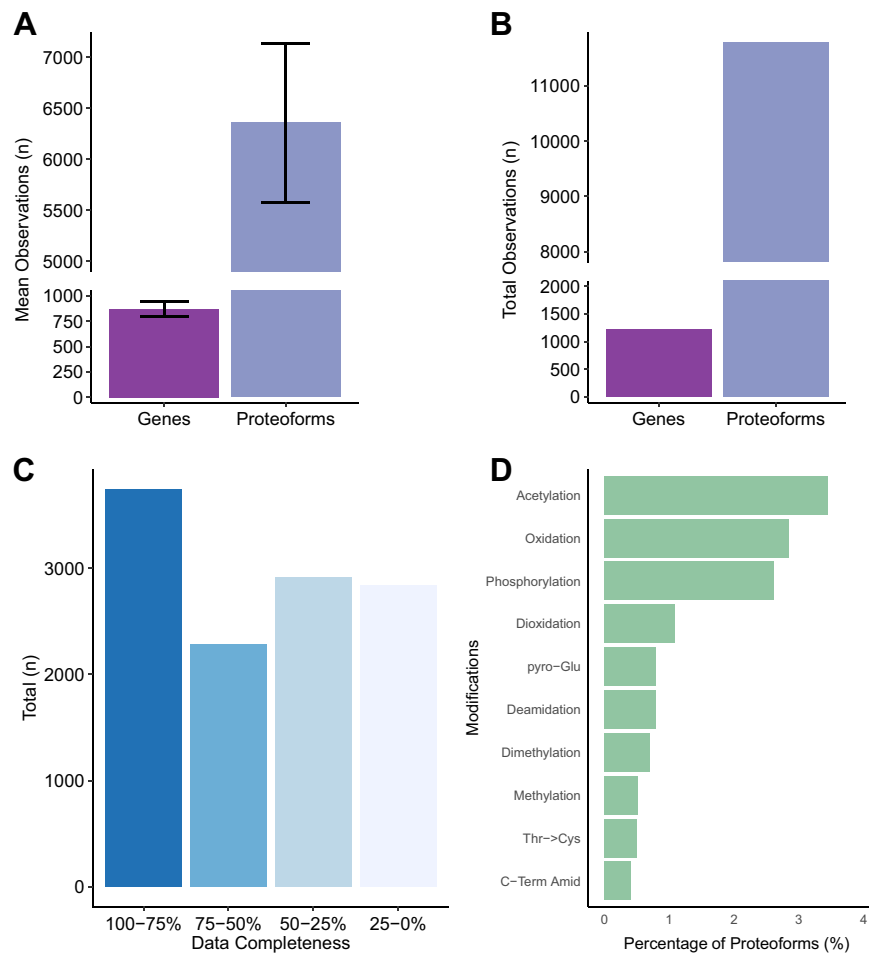


FIG. 2. **Overview of the top-down data.** A, mean number of unique genes and proteoforms observed from each human subject. Error bars represent $\pm$ sd. B, total unique proteoforms and genes found in the entire study. C, percentile bins containing proteoforms with different degrees of data completeness. D, top 10 posttranslational modifications found across the entire study, expressed as a percentage of all proteoforms found with the stated modification.

proteome if measured in its intact form. Notably, the largest full-length protein we could identify in our study was a 28 kDa brain-specific splice variant of tropomyosin (TMB α -3) (40). As proteins are often endogenously cleaved *in vivo*, the proteome coverage also extends to proteins with full-length molecular weights greater than 25 kDa (Supplemental Fig. S1). Although we identify proteoforms derived from proteins far beyond 25 kDa, the bias towards shorter proteins relative to the entire Swiss-Prot database is still present (Supplemental Fig. S1). Proportionally, 89% of our proteoform identifications can be assigned to canonical Swiss-Prot entries, 7% assigned as splice isoforms, and 4% derived from unreviewed TrEMBL entries.

Efficiency of A β Solubilization

A β proteoforms involved in plaque formation typically require unique extraction and solubilization procedures. A “gold standard” approach for the plaque solubilization utilizes 70 to 100% FA (41–43). However, this approach may not be compatible with TDP as it introduces artifactual formylation unless the sample preparation is performed at temperatures $\leq -20^\circ\text{C}$ (44). Moreover, we desired not to tune the sample preparation method to a specific protein or proteoform in order to capture a greater breadth of proteoforms in our shotgun approach. Therefore, we applied a more generalizable sample preparation method that could also capture A β proteoforms efficiently. Recognizing 8 M urea as a well-established chaotropic agent in protein solubilization, we additionally considered HFIP based on prior studies showing its efficiency for solubilization of A β fibrils (45). Furthermore, the combination of both has been demonstrated to facilitate solubilization of hydrophobic transmembrane peptides (46). Thus, we deployed a hybrid extraction method that relies on both HFIP and 8M urea to provide two fractions of proteoforms for analysis. The first fraction represents proteoforms soluble in $\sim 50\%$ HFIP, 8 M urea, while the second fraction represents proteoforms soluble in just 8 M urea. To evaluate the performance of the HFIP/urea method with respect to extraction of the aqueous-insoluble A β species, we compared the intensity of our observed A β proteoforms to those reported in the prior 5xFAD mouse-based study (47). In this study, the A β species were separated into soluble and insoluble by ultracentrifugation followed by denaturation in 1.5 M or 4 M guanidine-based buffer, respectively. The correlation plots comparing intensities of the A β species between HFIP method and soluble/insoluble fractions of the 5xFAD mouse-based study are shown on (Supplemental Fig. S2). The A β species extracted from the elderly human brains using HFIP method matched best the insoluble species from the aged mice. Thus, the HFIP method is not limited to aqueous-soluble A β species. The correlation between A β species from aged

human brains and insoluble A β species from mice increases with mouse age. This supports the use of aged 5xFAD mouse models for studying A β pathology.

Coverage of Protein Modifications

We next investigated the distribution of protein modifications (cotranslational and posttranslational) across all proteoform identifications. Since the modifications were assigned purely based on the matching of the mass to the UniMod database within 0.1 Da tolerance, which is equivalent to 10 ppm for a 10 kDa proteoform, the results should be interpreted with caution. Some modifications can be mistaken for deisotoping errors, such as deamidation. Others may result from artificial modifications introduced during sample preparation, like carbamylation combined with an isotopic error being confused with acetylation. Additionally, some modifications can be misidentified as nonsynonymous mutations, such as an Ala $\rightarrow$ Ser mutation being confused with oxidation. Nonetheless, as expected, the most frequently identified modification was N-terminal acetylation (26%), typically introduced during the process of translation. Many proteoforms also contain lysine acetylation (4%), oxidation (3%), phosphorylation (2%), and deamidation (1%) (Fig. 2D). The glutaminyl cyclase-mediated N-terminal pyroglutamate (pyro-Glu) ($\sim 1\%$) is abundant and a known modification present on some neuropeptides and found on several A β proteoforms as well (48, 49). We also identified N-terminal myristoylation (~ 210 Da), a type of lipidation, on several proteins known to carry this modification including HPCAL1, MARCKS, and VSNL1. We also observed myristoylation on MMP24OS, which has not previously been annotated with this modification (Supplemental Fig. S3).

The largest detected PTMs were glycosylations ranging in size from 161 to 3937 Da that were assigned to previously documented glycoproteins (Supplemental Table S3) (50–59). These include single hexose modifications of hemoglobin subunits alpha and beta (HBA1, HBB) (53), as well as attachments of more complex sugars to TSC22D1, TSC22D3, HSPE1, CHGA, MAP2, EEF1A1, NDUFB1, UBB, and NAP1L1 (50–59). This list of glycan assignments was manually curated to ensure that these large mass shifts were not misidentified based on protein isoform differences, that is, isoforms that differ by the insertion of several amino acids. Additionally, glycosylated proteins undergo preferential cleavage of the glycosidic linkage during MS/MS *via* collision-induced dissociation (60, 61). This is evident by neutral loss of glycosylation from the assigned fragment ions, as well as neutral loss of glycosylation from the parent ion, which we leveraged as evidence of glycosylation (Supplemental Fig. S4). Other unique modifications included uncommon amino acids and single nucleotide variants. For example, incorporation of selenocysteine was readily detected on selenoprotein W (SPEW1). Furthermore, we could distinguish two isomeric

haptoglobin allele variants (*HP*1F* and *HP*1S*) across different haptoglobin proteoforms. These are differentiated by two missense mutations that lead to an isomeric exchange in amino acids ($^{129}\text{NE}^{130} \rightarrow ^{129}\text{DK}^{130}$, [Supplemental Fig. S5](#)).

Association of Protein Cleavages with Unstructured Regions

Among the types of PTMs that can occur, proteolytic cleavage events represented the most frequently observed modification. Over 75% of the proteoforms identified were found to contain at least one cleavage event. Given the frequency of their observation, we sought to identify any patterns or trends that associated with proteolytic cleavage. Because protease cleavages are typically limited to specific features such as primary sequence motifs or secondary structures, we decided to test if protein local structure correlated with cleavage events. We used the AlphaFold2 predicted local distance difference test (pLDDT) score (62), a per-residue estimate of structural confidence, to quantify structure in a given region. We then curated a list of proteoforms where cleavages occurred after the first three amino acids or before the last three amino acids of a given gene sequence (based on the canonical gene sequence). We then averaged the pLDDT scores within ± 3 amino acids surrounding a given cleavage site and compared these values to the distribution of pLDDT scores for all amino acids of the detected cleaved proteins ([Supplemental Fig. S6](#)). We found a clear shift towards lower pLDDT scores for cleavage sites *versus* the reference data, indicating cleavage sites are enriched in regions predicted to be unstructured or potentially intrinsically disordered (regions with pLDDT <50) (63).

Reproducibility of the Quantification

Repeated Analysis of Mouse Brain Samples—Since label-free top-down proteome quantification is a nascent technique, it is crucial to evaluate its accuracy and reproducibility. To this end, we processed six mouse brain tissue samples using the same HFIP/urea solubilization. This set of resulting 12 fractionated samples was analyzed independently three times in 2 to 3 months intervals. The top-down LC-MS datasets were processed in a similar fashion with TopPIC followed by the companion TopPICR R package. As anticipated, a substantial factor explaining the variance of the data was the batch effect ([Supplemental Fig. S7A](#)). ComBat procedure (32) effectively removed batch effect resulting in the grouping of repeated LC-MS analyses of the same sample together ([Supplemental Fig. S7B](#)). The batch correction procedure also significantly reduced the coefficient of variation or relative SD (we avoid using commonly used CV abbreviation to avoid confusion with FAIMS compensation voltage). After the correction, the median coefficient of variation value dropped from 42.3% down to 31.7%, which is on par within batch values (32.1, 31.8, and 31.9%) and similar to values that are

often achieved with label-free BUP methods ([Supplemental Fig. S7C](#)) (64).

Correlations with Bottom-Up Datasets—One way of validating accuracy of the data acquired with a new method is to compare it, typically by computing Pearson correlation (R) coefficients, with the measurements acquired using a more established approach. There are two types of prior BUP data that have been collected on the same samples: targeted selected reaction monitoring (65) and shotgun tandem mass tag (8). Having two BUP datasets allows us to establish a reference correlation value of two independent measurements of the same protein. The overlap between the selected reaction monitoring and tandem mass tag datasets encompassed 400 samples and 219 proteins. The median correlation value (R) between these bottom-up measurements is 0.2 ([Supplemental Fig. S8](#)). However, the correlation is stronger if a protein has a substantial variance in abundance across the samples, though these abundance fluctuations may not necessarily be relevant to AD or related phenomena. The strongest correlating protein is APP (R = 0.8), peptides of which capture the variation of A β abundance. Other high-correlation genes include C1QA, VGF, C1QC, and GPNMB with R values ranging from 0.68 to 0.71.

Calculating correlation between global protein abundance derived from bottom-up and intact proteoform measurements from top-down is not straightforward, since given our top-down data, there are seven proteoforms per protein on average. Therefore, we evaluated the correlations of all proteoforms with both types of BUP data ([Fig. 3](#)). Overall, the TDP-BUP abundance relationships were positively correlated. For most of the proteins, there is at least one proteoform with the same or higher correlation to the BUP data compared to the reference value based on the two BUP data sets. Thus, despite the complexity of the protein posttranslational processing and bias of the TDP method towards low-MW species, there is evidence that, overall, the data agrees with conventional proteomic approaches.

Proteoforms Linked to Neuropathologies and Cognitive Decline

Next, we tested the association between proteoform relative abundances and clinicopathological variables. For the statistical analysis, we utilized linear and quasi-Poisson modeling approaches for the intensity and spectral counting data, respectively (33). Intensity data was corrected for batch effects prior to statistical testing. In case of spectral counting data, the batch effect was added as a covariate into the statistical models. We also accounted for postmortem interval in both types of measurements, as it was significantly associated with a substantial number of proteoforms: 200 and 164 proteoforms for intensity-based and spectral counting approaches, respectively. The tested variables included tangle density, amyloid load, presence of Lewy bodies in the neocortex, hippocampal sclerosis, cerebral infarcts, cerebral

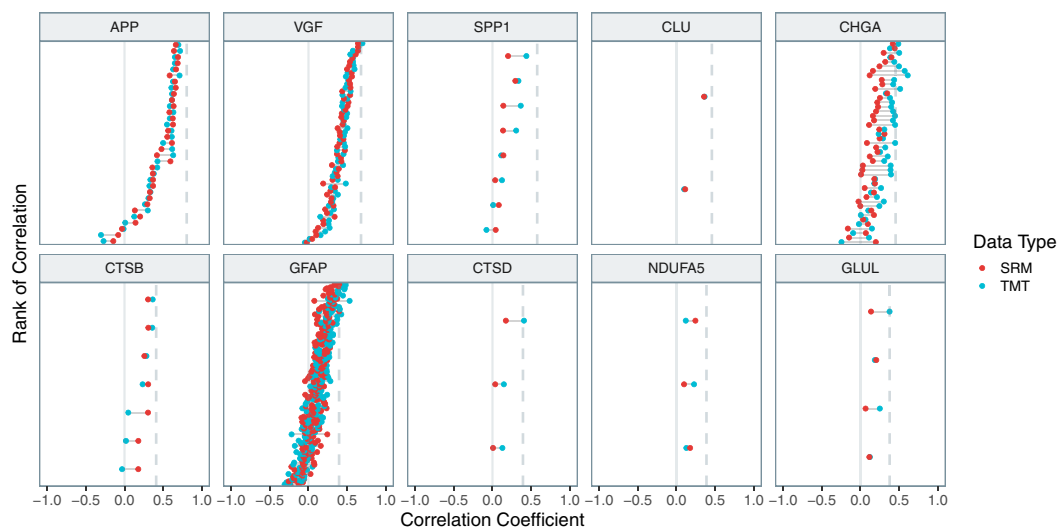


FIG. 3. **Correlation of TDP proteoforms with the parent protein BUP (SRM and TMT) measurements.** Individual points represent proteoforms of a given protein. Plots are shown for proteins that were present in all three sets of data, had high correlation ($R > 0.35$) between the two BUP datasets, and had more than one proteoform in TDP data. Proteoforms are ordered on the Y-axis according to their average correlation to the BUP datasets. The dashed line reflects the cross-BUP (SRM/TMT) correlation value for a given protein.

amyloid angiopathy, *APOE* $\epsilon 4$ allele status, cognition level, slope of cognitive decline, and the slope of cognitive decline adjusted for known Alzheimer's disease and related dementias neuropathologies. The summary heatmap in Figure 4 demonstrates the associations discovered by testing both types of quantitative data. In total, 41 proteoforms in intensity-based quantification and 125 in spectral counting-based quantification showed a statistically significant relationship with at least one variable. The overlap between the two approaches is 18 proteoforms. Many of the significant proteoforms are derived from proteins that have been linked to AD (<https://agora.adknowledgeportal.org/about>) (66). This includes proteins such as growth-associated protein 43 (GAP-43), MAP1B, MAP2 (67), alpha-synuclein (SNCA), ACO2, SYN1, stathmin (STMN1), VGF (68), PACSIN1 (69–71), and VAMP3 (72–74).

Heterogeneity of $A\beta$ Proteoforms

We quantified 37 and 49 $A\beta$ proteoforms using intensity and spectral counting approaches (35 in common), respectively. The most abundant $A\beta$ proteoforms were unmodified $A\beta_{1-42}$, followed by $A\beta_{2-42}$ and $A\beta_{1-40}$ (Fig. 5). The observed species had truncations and extensions of the C-terminus and predominantly truncations from the N-terminus, relative to the canonical $A\beta_{1-40}$ species. The truncations from N- or C-termini were mutually exclusive with only one exception of low abundance $A\beta_{2-38}$ (APP_15) species truncated from both termini.

We observed several PTMs on $A\beta$ as well. The most abundant PTM-bearing $A\beta$ proteoforms included pyroglutamination of Glu3 ($A\beta_{pE3-42}$), acetylation of $A\beta_{1-42}$ at Lys16, and oxidation of $A\beta_{1-42}$ at Met35. We found the $A\beta_{pE3-42}$ (APP_7) abundance to be most associated with amyloid plaque relative

to all proteoforms quantified in this analysis including canonical $A\beta_{1-42}$ (APP_6, Fig. 6). Interestingly, the other main species $A\beta_{1-40}$ (APP_1) abundance is not significantly associated with amyloid plaques or any other clinicopathological variable.

Overall, $A\beta$ proteoforms can be divided into two major groups and one subgroup based on the starting and ending positions. The longer forms with C-terminal positions greater than 42 are associated with amyloid load, while shorter forms ≤ 40 , with exception of $A\beta_{1-40}$, demonstrate association with CAA. All the CAA-associated proteoforms were truncated either from the N- or C-terminus relative to the 1st and 40th amino acid, respectively. Canonical $A\beta_{1-40}$ proteoforms were not statistically significantly associated with CAA unless they were bearing PTMs such as oxidation of Met35 or acetylation of Lys16. However, in these cases, the association was not as strong as for N- or C-terminus truncated species (e.g. $A\beta_{1-38}$ or $A\beta_{4-40}$). Finally, there is a small subgroup of $A\beta_{x-42}$ species with N-terminal truncations from positions 3 to 9 that demonstrate particularly strong association both with the neuropathology and cognition (e.g. APP_10, Fig. 6).

Association of VGF Proteoforms with Cognitive Decline

VGF is another protein that has been gaining attention with regards to its association with AD (68, 75, 76). Notably, it is expressed as a polypeptide that is proteolytically processed by proprotein convertases (77, 78). The observed VGF fragments in our study align with known cleavage sites and previously annotated neuropeptides (77) (Fig. 7). Mapping of the cleavage sites revealed signal peptide cleavage along with 13 previously described internal cleavage sites and four newly characterized cleavage sites. Notably, cleavages were observed for all dibasic sites.

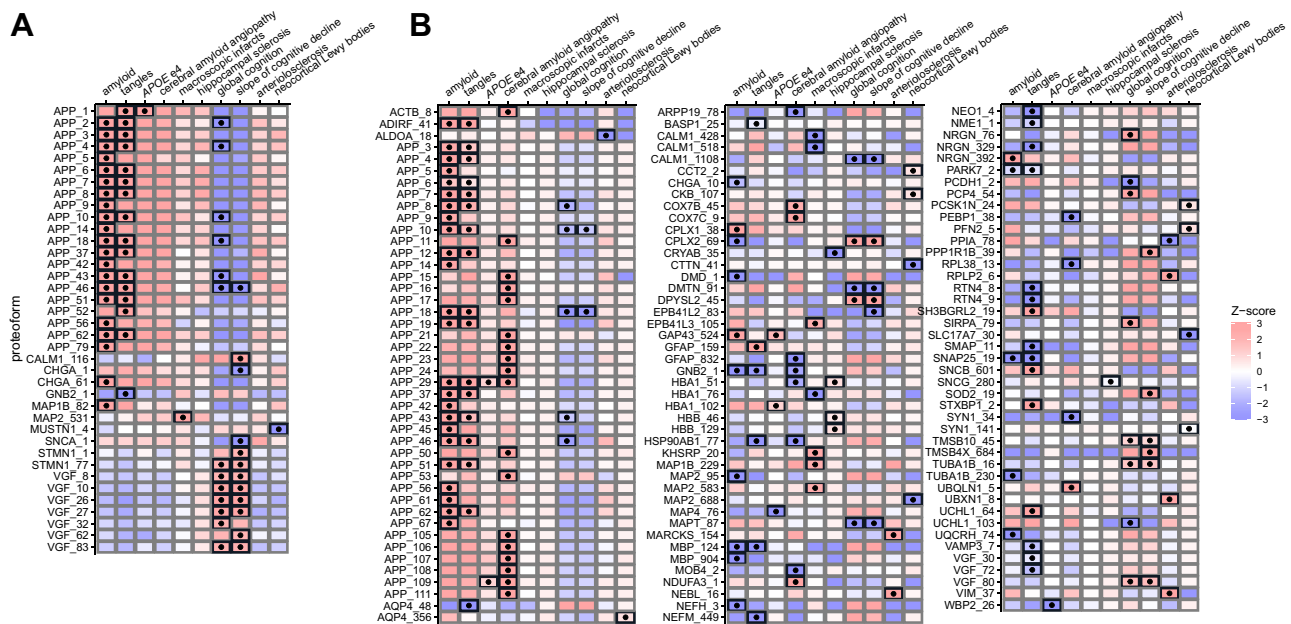


FIG. 4. **Statistical associations.** Punchcard heatmaps of statistical associations between proteoforms and cognitive/pathological variables using (A) intensity-based measurements and (B) spectral counts. For label-free intensities, associations were modeled using linear regression while spectral counts associations were modeled by a quasi-Poisson generalized linear model. In both modeling approaches, abundances were adjusted using postmortem interval as a covariate. Black dots represent statistically significant associations (Benjamini-Hochberg adjusted p -values < 0.05). Proteoforms are labeled with their gene of origin, followed by an arbitrary number assigned based on clustering.

Association of VGF fragments with cognitive decline was not limited to a particular neuropeptide or a region of the VGF sequence. The seven (out of 55 quantified) statistically significant proteoforms are scattered throughout the entire sequence. If no correction for multiplicity of hypothesis testing is applied, then the number of significant proteoforms goes up to 27. However, even in this case, there is no regional dependence of VGF association with cognitive decline (Fig. 7).

Co-abundance Clustering Analysis

Individual Cluster Patterns—To group the proteoforms into co-abundance modules, we applied a widely used WGCNA clustering tool. This resulted in 95 modules, with on average 63 proteoforms per module with sizes ranging from 5 to 3267 proteoforms. The module centroids were correlated with the selected phenotypes. Overall, 10 modules had significant associations with at least one of the traits (Fig. 8 and Supplemental Table S4). For each of the 10 significant modules, we selected the representative gene with the highest number of proteoforms in the module. Some genes, specifically APP and VGF, were representative in more than one module. Furthermore, several modules were represented by proteoforms of predominantly (>85%) a single parent gene: APP-1, APP-2, APP-3, VGF-1, VGF-2, and STMN1. Three additional modules had >50% proteoform representation of a single gene (GFAP, CALM1, and MAP2).

The most significant association is between APP-1 (or “coral”) module and amyloid load (adjusted p -value 4×10^{-19}). It contains $A\beta_{1-42}$ and other $A\beta$ proteoforms ending at the same C-terminal position but recessively shortened at their N-terminal starting positions. The association is exceptionally statistically strong because these $A\beta_{x-42}$ essentially constitute amyloid plaques. APP-2 (or “darkolivegreen4”) is another module that contains similar $A\beta$ proteoforms with the same pattern of associations across the phenotypes as APP-1, but with weaker statistical significance. We did not notice any differences between the APP-1 and APP-2 modules in the size of the proteoforms, their starting N- and ending C-termini positions, or PTMs. However, APP-2 module proteoforms showed noticeably (~30-fold) lower in abundance (Supplemental Fig. S9). The third APP module, APP-3 (or “mediumpurple1”) had a clearly different profile of association with phenotypes. It was not associated with amyloid plaques but showed significant association with CAA. The APP-3 cluster essentially captured the $A\beta$ species truncated at the C-terminus that were described above (Fig. 5).

The split of VGF into two modules is similar to the split between APP-1 and APP-2. The VGF-1 (or “yellowgreen”) is a module composed of less abundant species and overall has stronger association with cognition. The VGF-2 (or “lightcoral”) cluster contains more abundant VGF proteoforms and shows a similar behavior with a weaker association with cognition. No differences were found between the two VGF

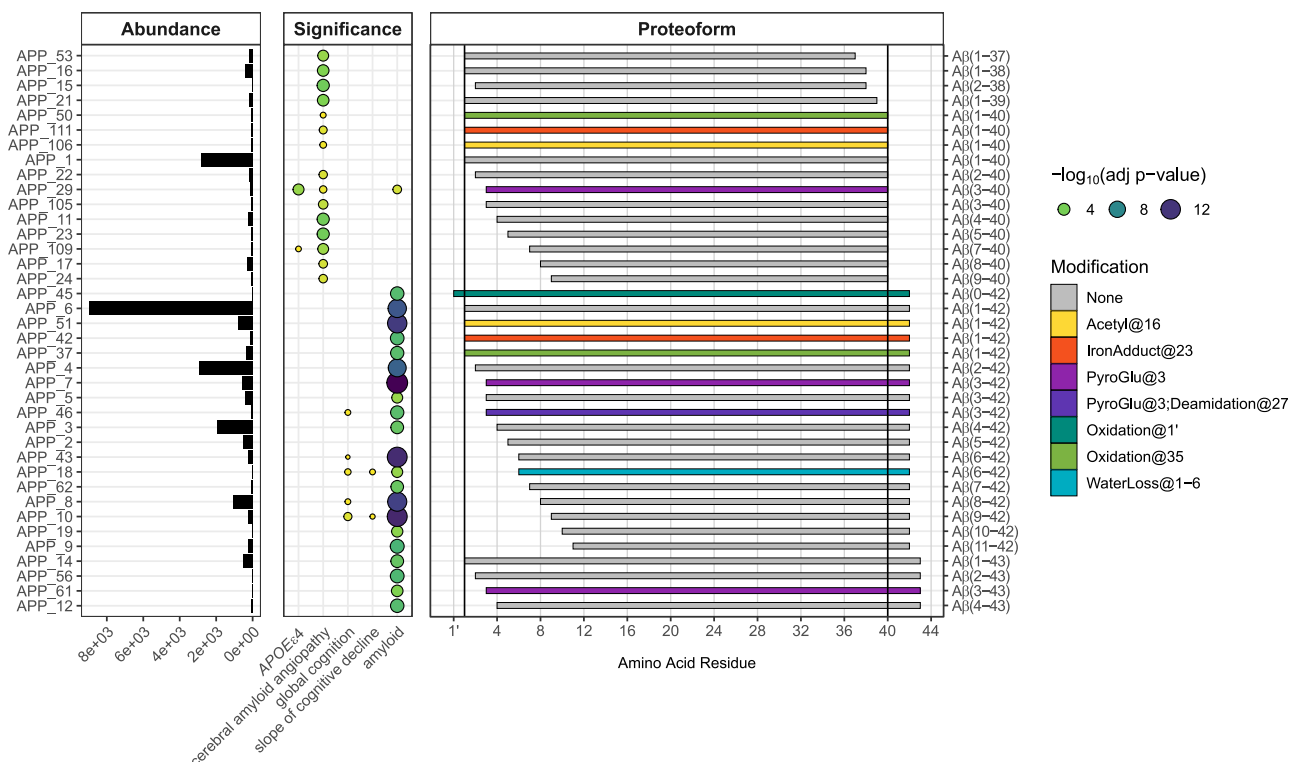


FIG. 5. **Differential association of amyloid beta proteoforms between amyloid plaques and cerebral amyloid angiopathy.** Leftmost panel contains bar plots showing the spectral count abundance for each amyloid beta proteoform. Middle panel displays the significance of the association between each amyloid beta proteoform and several clinicopathological or genetic variables, including APOE ϵ 4 status, cerebral amyloid angiopathy pathology, last visit global cognition score, person-specific rate of cognitive decline adjusted by age, sex, and education, and overall amyloid level (amyloid). The degree of significance is indicated by the circle size and color. The rightmost panel maps every amyloid beta proteoform to an APP reference sequence (dark gray bar) and colors them by the type of PTM identified. Amyloid beta proteoforms are sorted by ascending C-terminal amino acid ending position, followed by ascending N-terminal amino acid starting position.

modules in terms of PTMs, fragments' length, or position within the protein sequence.

Modules "plum4," "darkmagenta," and "darkseagreen3," represented by the genes STMN1, TMSB4X, and CALM1, respectively, show association profiles similar to the VGF modules. The significant associations, like in case of VGF, are primarily contained within the variables reflecting cognitive metrics (Fig. 8). Given that the directionality of the associations is also the same, they may represent the same underlying process related to VGF. Additionally, the "darkmagenta" cluster is enriched with methionine oxidations. Notably, 14 of the 27 proteoforms within the cluster contain mono- or di-oxidation events. These oxidations are distributed across nine out of 16 genes represented in the cluster. By Fisher's exact test, this enrichment is highly significant ($p\text{-value} = 1 \times 10^{-14}$) in comparison to the global data in which oxidation represents <5% of observed modifications (Fig. 2D). While there is currently no established causal relationship between global protein oxidation and progression of AD (79, 80), methionine oxidation does have broad consequences for protein structure and activity (81, 82).

Module "lightblue4," associated with gross chronic infarcts, consists predominantly of MAP2 proteoforms, followed by

DYNLL1, MAP1A, and SYNGAP1. MAP2 itself is a cytoskeletal protein localized in the dendrites, while other members of the module are also annotated as cytoskeletal or dendritic proteins. The "skyblue2" module, associated with arteriosclerosis, is dominated by GFAP proteoforms, followed by proteoforms derived from AQP4, MARCKS, and SIRPA. The latter three proteins are either membrane, trans-membrane, or anchored to the cytoplasmic membrane.

DISCUSSION

Introduction

We quantified 6124 and 7510 proteoforms, mapped to 836 and 952 proteins, using intensity- and spectral counting-based approaches, respectively. Thus, for each protein, we observed approximately 7 to 8 proteoforms on average. This study also provides deep characterization of brain proteoforms, capturing unique PTMs such as phosphorylation, glycosylation, and pyroglutamylation. Our data suggests three quarters of all proteoforms that start at the N-terminus (with or without methionine excision) were N-terminally acetylated (Fig. 2D). This is in agreement with the high frequency of N-terminal acetylation in the human proteome (83).

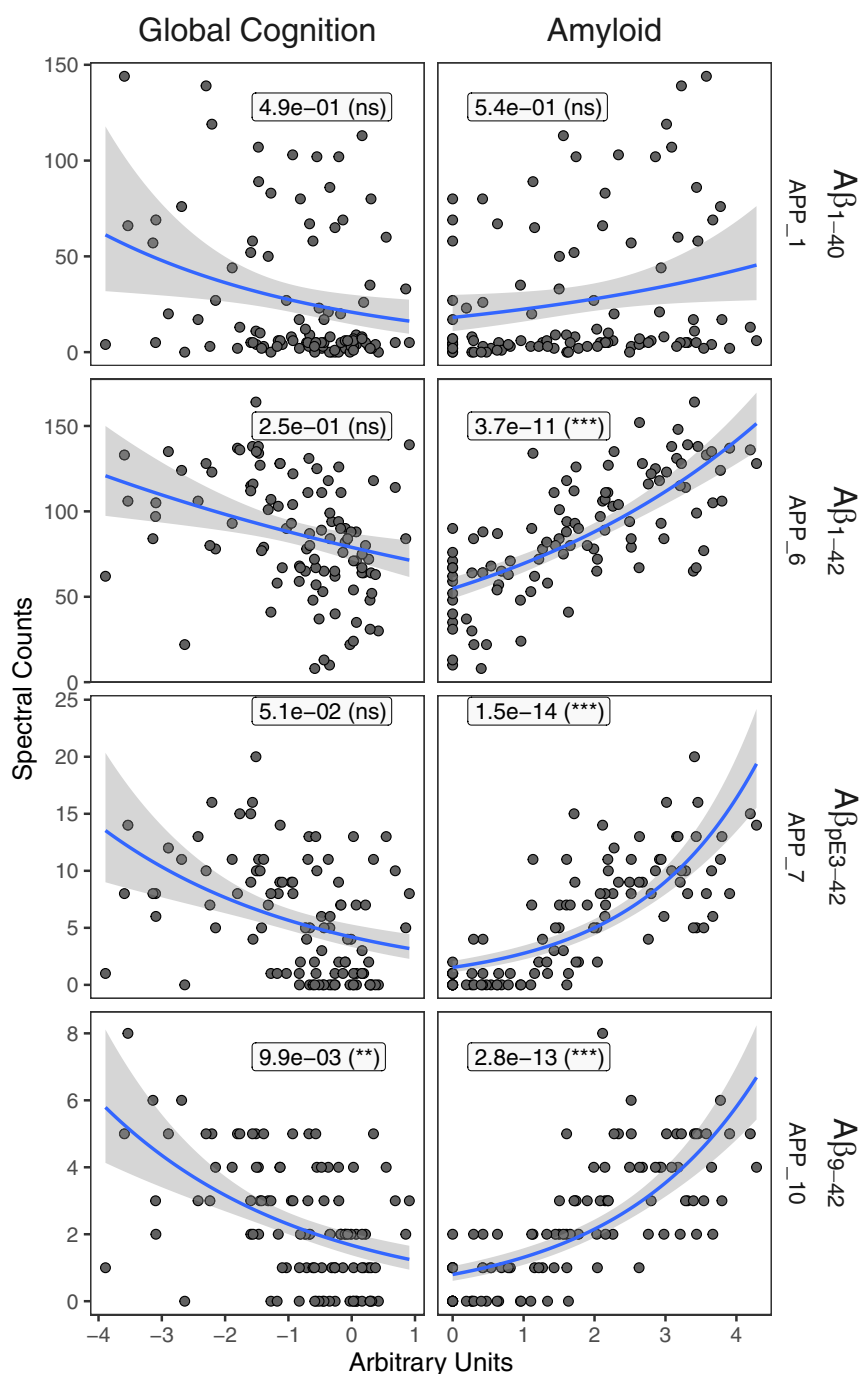


FIG. 6. Association of selected A β proteoform abundances (y axis) with the global cognition level and amyloid plaques load (x axis). Proteoform abundances were quantified using spectral counting and modeled using a quasi-Poisson approach. The estimate of the mean and the corresponding standard error are shown as a blue line and shaded area. For the purpose of visualization, the fitted model does not include the batch effect. The Benjamini-Hochberg adjusted *p*-values (ns > 0.05, * < 0.05, ** < 0.01, and *** < 0.001) are shown for the full model that factors in batches and are the same as in Figure 4.

Both types of quantitative data, proteoform intensity-based abundances and spectral counts, were correlated with clinical and neuropathologic variables relevant to cognitive decline and AD. This approach enabled a detailed look at proteoforms within a given gene with regards to their role in AD. We

identified associations involving canonical AD-related proteins such as amyloid beta, MAPT, and VGF, as well as new proteoform-disease correlations that have not been well characterized, such as with GAP-43 and stathmin. In the case of MAPT (tau), we observed many truncated proteoforms;

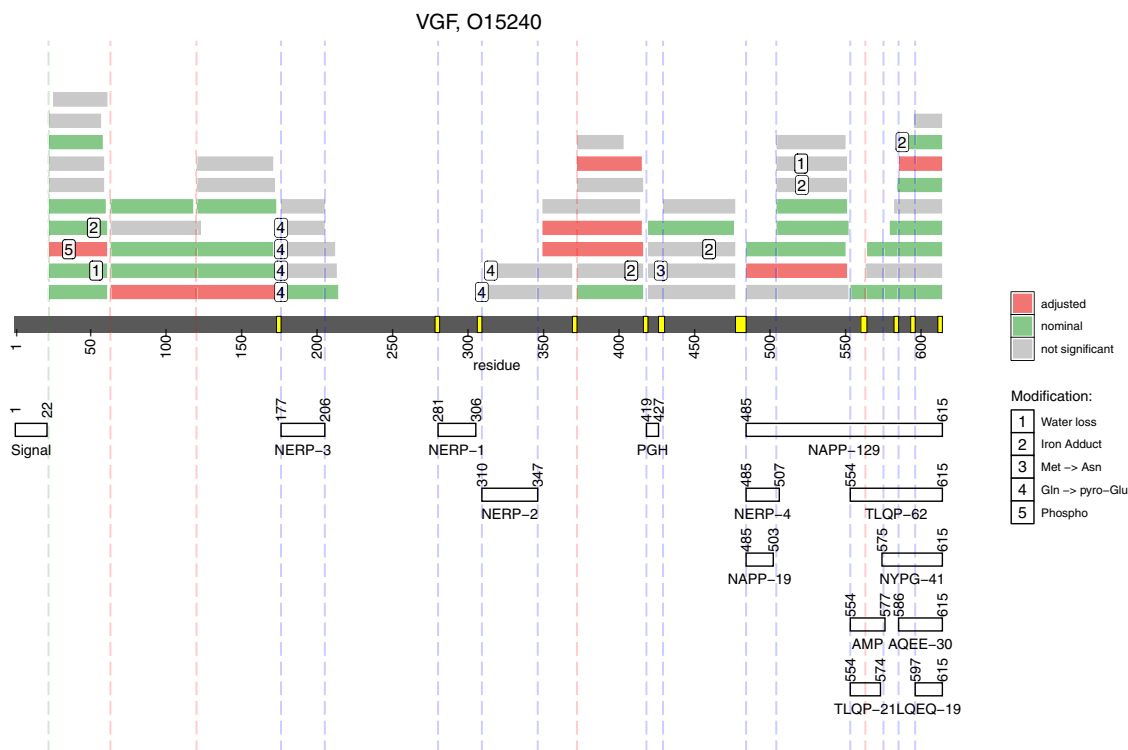


FIG. 7. **Proteolytic cleavage map of VGF.** Middle gray bar and residue numbers indicate the full-length VGF protein. Yellow markings along the protein sequences denote the di-basic sites that are the putative cleavage sites for proprotein convertases. Annotation of the known VGF neuropeptides is shown below this residue labeling (white bars). The VGF proteoforms quantified in this study are shown in the upper half and colored according to significance threshold (nominal significance is shown in green, adjusted significance is shown in red, not significant forms are shown in gray). Observed posttranslational modifications are numbered and annotated at their localized amino acid position. Vertical dashed blue lines indicate known cleavage sites while dashed red lines show the previously unknown cleavage sites. The single dashed green line indicates the cleavage site for the signal peptide.

however, only one was linked to general neuropathology. This does not necessarily indicate a weaker association of tau with neuropathology but instead is a limitation of the method as full-length tau (46 kDa for the 2N4R isoform) is above of the typical molecular weight range for TDP experiments. Thus, capturing pathology-relevant species of tau would require either targeted acquisitions (84, 85), partial digestion or middle-down proteomics approaches (86–89), or alternative data acquisition methods (90).

Considerations for Further Top-Down Quantitative Studies

Unlike bottom-up proteomics, the field of quantitative TDP is still in its early stages. Various approaches have been explored, including spectral counting (19), MS1 intensities, extracted ion chromatograms (47), or even isobaric tag labeling (91). Our initial goal was to use MS1 intensities for quantification. To facilitate this, we developed TopPICR (17), a tool that aligns datasets and performs match-between-runs analysis—a feature first implemented in MaxQuant (92).

Spectral count data is inherently present in any data-dependent acquisition dataset, requiring no additional experimental effort. As a result, both types of quantitative data can be generated. One key challenge in TDP data analysis is the

ambiguity in PTM localization, which can lead to partially characterized proteoforms being assigned different identities across datasets. To address this, we strongly recommend aligning datasets and grouping proteoforms based on their mass and elution time—a process performed by TopPICR—even when spectral counting is used for quantification. Without this step, reports will contain an inflated number of proteoforms and a high degree of missingness or zero counts.

In this study, spectral counting provided broader coverage of proteoforms, likely due to the more stringent data completeness requirements for label-free statistical testing. However, studies involving larger number of samples offer greater statistical power, thus enabling the use of less statistically sensitive spectral counting approaches (e.g., quasi-Poisson). Conversely, for smaller datasets ($N \leq 10$), MS1-based label-free quantification, particularly when combined with moderated statistical tests (e.g., limma), may be a more effective approach.

APP

Introduction—A β proteoforms provided the strongest associations to Alzheimer's clinicopathological variables across both spectral counting and intensity-based analyses. More

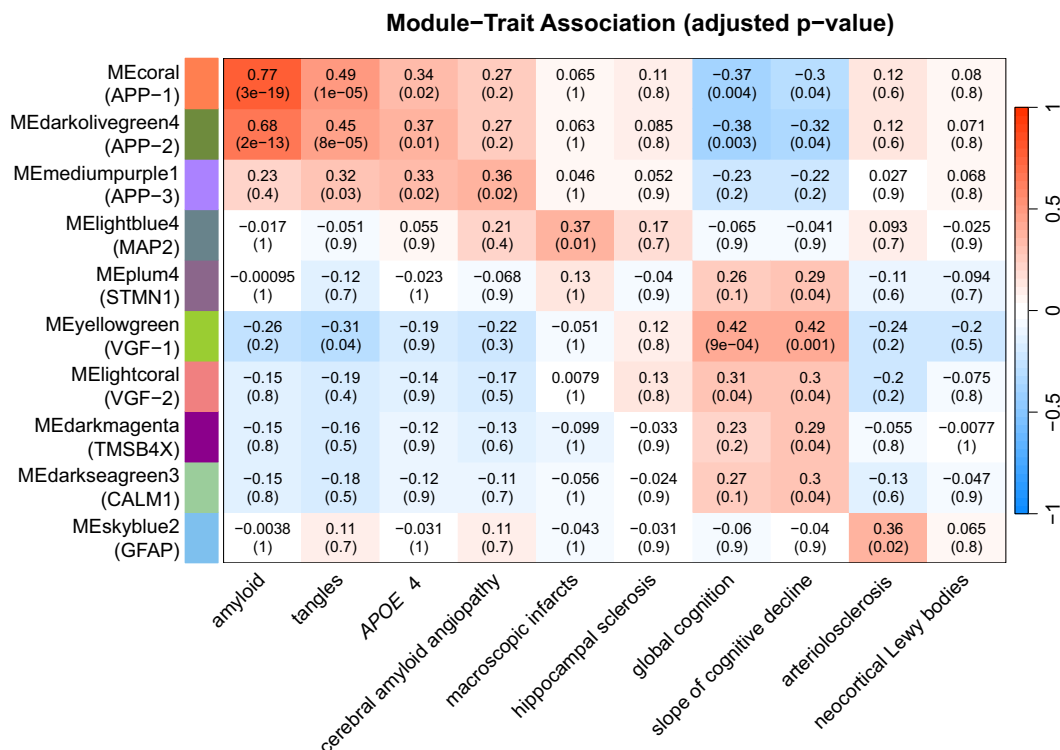


FIG. 8. Heatmap of the significant associations of WGCNA modules' eigengenes with selected clinical and pathological variables. The amplitude and directionality of the correlation in color-coded in blue-white-red gradient and is shown in each cell. The adjusted *p*-values, reflecting the significance of the association, are shown in parentheses.

notably, our study identified the largest number of human brain A β proteoforms to date in a single study, including four extended A β_{x-43} proteoforms that typically cannot be observed in BUP LC-MS/MS studies due to the extreme hydrophobicity of the C-terminal tryptic peptide (93). Part of this depth can be attributed to the HFIP extraction method developed within this study, which can efficiently solubilize A β while at the same time removing potential contaminating proteoforms from the supernatant. Overall, the catalog of A β proteoforms found in this study also agrees with a recent TDP studies using human postmortem samples (94) and the 5xFAD mouse model (47). In particular, all three datasets identify a variety of N- and C-terminally truncated proteoforms in addition to the most abundant proteoforms A β_{1-42} and A β_{1-40} .

HFIP as an A β Solubilizing Agent—A β solubilization is a topic of a broad and persistent interest. The “gold standard” approach for A β solubilization is treatment with 70 to 100% FA (41–43). However, this method is not ideal for TDP, as it can introduce multiple formylation modifications on a proteoform as well as artificial cleavages, complicating downstream analysis (44, 95). To overcome this challenge, we surveyed alternative solubilization strategies with strong chaotropic effects and limited interference with LC-MS instrumental analysis and data interpretation. Alternatives that principally fit into this category were HFIP, guanidinium, and urea. HFIP has

been shown to effectively dissolve synthetic A β_{42} peptides into monomeric forms by disrupting β -sheet structures (96, 97). It has also been used as a mass spectrometry-friendly solvent in a recent top-down study (98). While it remains to be demonstrated that HFIP “fully” solubilizes human brain-derived amyloid plaques, the measured A β proteoforms are significantly associated with expected clinicopathological variables. First, A β_{x-42} proteoforms correlate with plaque content measured *via* histological techniques; even if these species do not originate directly from plaques, this implies that they are at least plaque markers. Second, the abundances of HFIP-extracted A β species align well with those from another study using 4M guanidine to resolubilize pelleted, “insoluble” A β fractions (Supplemental Fig. S2). Third, A β proteoforms were among the most abundant in our dataset, ranking third in spectral count-based abundance estimates, alongside MBP and GFAP. This suggests efficient extraction, with approximately 3.5% of all spectra assigned to A β proteoforms alone.

We acknowledge that the evidence for HFIP's ability to fully solubilize A β plaques is circumstantial at best, and some A β proteoforms may not have been captured. Nonetheless, there is a strong statistical association between HFIP-soluble A β species and both plaques and CAA. Notably, the differential association of A β_{42} species with plaques *versus* A β_{38} and A β_{40} with CAA is well-defined. Thus, we believe, HFIP is a

solubilizing agent worth considering for future proteomics A β species studies.

Differential Association of A β Proteoforms in Human and Mice—The A β proteoforms identified in this study could be divided into two groups based on their association with either amyloid plaques or CAA. The two groups could also be distinguished by the ending C-terminal residue, with CAA-linked proteoforms ending at or before the 40th amino acid and plaque-linked proteoforms ending at the 42nd and 43rd amino acid. Interestingly, the WGCNA modules corresponding to longer A $\beta_{x-(42,43)}$ (“coral” and “darkolivegreen4”) and shorter A $\beta_{x-(37-40)}$ species (“mediumpurple1”) have also been observed in the mouse 5xFAD model as “Group1” and “Group2” proteoforms. However, in the case of the 5xFAD model study, the distinction between the clusters was based on profiles across age groups and soluble/insoluble fractions (Fig. 4 in ref (47)). We can be cautiously optimistic that 5xFAD model reproduces the differences in the behavior between the A $\beta_{x-(42,43)}$ and A $\beta_{x-(37-40)}$ forms. If the group2 A β species have statistically significant association with CAA in the 5xFAD mouse, like we observed in this human study (Fig. 5), that would further validate the model for studying the differences in the A β processing, accumulation, and clearance pathways. Note, that we did not robustly observe A β proteoforms present in the mouse group 3 cluster. There are a few potential reasons for this. One is that we utilized a 3K MWCO filtration step during sample preparation which would remove smaller proteoforms, such as those in the mouse group 3 cluster. Due to LC-MS duty cycle constraints, we also only analyzed three FAIMS CV values between -50 V to -30 V. As FAIMS CVs are correlated with proteoform size (15), observation of shorter A β proteoforms such as A β_{1-25} would likely require analyzing CVs between -50 V and -70 V or lower.

Cerebrovascular Amyloid—Although several prior studies have observed C-terminally truncated A β proteoforms within cerebrovasculature (99–102), they have been limited in the number of subjects and A β species examined. Another peculiar observation was made during an early TDP characterization effort of A β species (94). The sample preparation method in that study involved separating parenchyma from blood vessels. To the authors surprise, they did not observe any proteoforms ending at the 37th or 38th residue. Given our results, that these short A β forms are more closely associated with CAA, the most likely explanation is that the sample was depleted of these short forms by removing the vasculature. However, the authors did detect N-terminally truncated A β_{x-40} species, several of which were found to have associations with CAA according to our data. Taking these results together, the data imply a regional difference in A β processing. In summary, A β species can be thought of at least as two distinct groups, one consists of C-terminally extended forms (A $\beta_{x-(42,43)}$) associated with dementia-related amyloid plaques, while the other consists of C-terminally truncated forms

(A $\beta_{x-(37-40)}$) associated with CAA-associated deposits restricted to the vasculature.

Effect of N-Terminal Truncations—Analysis of the spectral counting data indicates a subset of A β_{x-42} species with N-terminal truncations that significantly associate with cognitive level and the speed of its decline (Fig. 5). The relevant truncations span from third (pyro-Glu modified) to ninth amino acid. The pattern of A β_{x-42} proteoforms most significantly associated with cognition and amyloid plaques are different depending on the N-terminus, indicating an additional or independent process that involves N-terminal truncations. Unlike gamma-secretase, beta-secretase cleaves specifically between the 671 and 672 residues of APP sequence (103). Thus, the N-terminal truncations are most likely produced by other proteases. Our data suggests that N-terminally truncated species, for example, A β_{9-42} , might be a stronger biomarker of the clinical manifestation of AD compared to the canonical A β_{1-42} form. Thus, it may serve as a better engagement marker for drug clinical trials if captured in blood plasma. Given the rise of assays that are based on A β proteoform ratios (104–107), it may be of interest to investigate the performance of A β_{9-42} , A β_{1-42} , or A β_{1-40} as a correlate or predictor of cognitive decline rather than overall A β plaque pathology.

Pyro-Glu A β —We found pyroglutamate-containing A β_{pE3-42} to be most strongly associated with amyloid plaques, which aligns with prior evidence demonstrating the pyroglutamyl form of A β has a much higher propensity for aggregation relative to A β_{1-x} forms (49). This increased aggregation can be attributed to the loss of two charges, the N-terminal amine and carboxyl from the side-chain of the glutamate, thus reducing overall solubility of the species (108). Cyclization of N-terminal Glu into pyro-Glu can occur as a spontaneous chemical reaction (109) or as an enzymatic reaction catalyzed by the enzyme glutaminyl cyclase (110, 111). Indeed, recent studies have demonstrated that A β_{pE3-x} requires enzymatic conversion by glutaminyl cyclase (112), which has led to consideration of these enzymes as therapeutic targets (113). Our results indicate that, despite its low abundance, A β_{pE3-42} has the strongest association with amyloid plaques.

Context of A β Proteoforms in Clinical Applications—Considering the diversity of A β proteoforms described in this study and their differential pathological associations, the impacts of this heterogeneity on anti-amyloid therapies and immunotherapies have likely been severely underestimated in clinical trials as amyloid burden is often nonspecifically assessed by PET imaging (114, 115). Although recent trials have started to adopt plasma biomarkers such as A β_{1-40} and A β_{1-42} (116, 117), our data suggests that more proteoforms need to be assessed in parallel to fully understand the impacts of AD therapies (114, 118).

For example, what proteoforms are targeted by anti-A β monoclonal antibodies and how these antibodies impact

amyloid plaques, CAA deposits, and potentially other A β proteoforms remains unclear, resulting in unpredictable clinical outcomes (119). One example is ponzemab, a therapeutic antibody that was described as recognizing the C-terminus of A β_{1-40} and required a free carboxy terminus at position 40 for binding (120). Ponzemab was reported to reduce deposition of amyloid in cerebral blood vessels during preclinical research and therefore considered as a treatment for CAA, but was discontinued after showing trends opposite expected outcomes in phase two clinical trials (121, 122). Our results demonstrate that A β_{1-39} , A β_{1-38} , and A β_{1-37} (and several modified A β_{x-40} proteoforms) are more closely associated with CAA than A β_{1-40} . Therefore ponzemab, which lacks binding to A β species truncated from position 40, should not be expected to clear all the proteoforms related to CAA pathology.

Another example is donanemab which targets pyro-Glu A β_{pE3-42} and has shown positive results in clinical trials (123). While we confirmed the importance of the pyro-Glu A β_{pE3-42} based on having the strongest correlation with pathology, which agrees with the rationale behind the development of donanemab (123), the low abundance of pyro-Glu A β_{pE3-x} (relative to other A β proteoforms identified in our study) suggests looking for other anti-amyloid treatments with affinity to more prevalent and more impactful A β proteoforms such as A β_{9-42} .

Finally, a recently developed antibody-based drug aducanumab (aduhelm) (124) recognizes the A β domain 3 to 7 (EFRHD) amino acids (125) and should theoretically bind to a wide range of proteoforms. Its efficacy in humans has been demonstrated using a surrogate endpoint based on PET imaging that does not distinguish parenchymal from vascular A β deposits. Although it was demonstrated that aducanumab more prominently binds parenchymal A β deposits in Tg2576 transgenic mice (124), major safety concerns arose, including hemorrhages in human subjects, specifically amyloid-related imaging abnormalities with vasogenic edema, which clearly point to detrimental changes in the vasculature. This leads to two possible explanations; that aducanumab damages the vasculature by removing vascular A β deposits or conversely exacerbates CAA by transporting parenchymal A β to the vasculature (126, 127). We can envision that the TDP approach, capable of resolving dozens of A β proteoforms, will be an indispensable tool for understanding the effect of aducanumab on A β clearance and brain swelling/hemorrhage complications.

VGF

VGF, also known as secretogranin VII, belongs to the granin family of neuropeptides. It is posttranslationally processed into neuropeptides with various biological activity (reviewed in (77, 78, 128, 129)). Several of them have been implicated in AD and other neurodegenerative or mental disorders. For example, TLQP-21 modulates microglial function (76), TLQP-

62 reduces A β plaque burden and disease-associated microglial activation (68), and other VGF peptides have been noted as decreasing in abundance in the brain, CSF, or blood plasma in AD or Parkinson's diseases (Table 2 in (78)). It is not trivial to reconcile the fact that a variety of VGF-derived neuropeptides have similar associations with AD and yet have different sequences and thus most likely different functions. Accordingly, a recent study targeting these different proteoforms with tryptic peptides on a large number of subjects (N = 1020) found that the association with AD and neuropathologies is not specific to a particular peptide or region of VGF (75).

Out of the 55 VGF fragments quantified in our study, 27 were nominally significant in the association with cognitive decline while seven remained significant after adjustment of the *p*-values for multiplicity of hypothesis testing (Fig. 7). Importantly, statistically significant proteoforms could not be traced to a specific region of the VGF protein. Therefore, our results and prior data indicate the association of VGF with AD is not unique to a single VGF neuropeptide or a specific region of the protein. There may be two explanations that form two, but not mutually exclusive hypotheses. One hypothesis is that, indeed, numerous VGF neuropeptides are necessary for maintaining cognition. An alternative hypothesis is that there is no differential regulation of the VGF-derived peptides. That is, once the full-length VGF protein is synthesized and cleaved by prohormone convertases, there is no differential regulation between the produced peptides, thus they are mutually correlating and have indistinguishable association with cognition.

STMN1

Stathmin, an 18 kDa tubulin-binding protein that inhibits microtubule assembly (130), was observed to have 17 quantifiable proteoforms, two of which were positively correlated with the slope of cognitive decline individually. A larger number of STMN1 proteoforms formed a cluster ("plum4") that had a similar profile to VGF. The associated proteoforms encompassed the full-length protein sequence, shortened fragments, and unlocalized ~15 Da modification which may be indicative of a single-nucleotide variant. A few reports have shown that stathmin is decreased in AD patients and adult patients with Down syndrome, and stathmin-dependent changes in microtubule stability are essential for memory stability (131–133).

GAP-43

GAP-43, also referred to as neuromodulin, is an intracellular growth-associated protein that is associated with nerve growth. An N-terminally truncated, phosphorylated GAP-43 proteoform (GAP43_524, phosphorylated ambiguously at serine 150, 153, or 154) was found to have an inverse relationship with amyloid plaque burden, contrary to previous reports on the full-length protein (134). Phosphorylation of GAP-43 is known to play a role in axon growth and synaptic

plasticity (135), and we also observe this GAP-43 proteoform is associated with reduced cognitive decline.

CONCLUSIONS

This TDP approach has provided a unique perspective to investigate the proteome with respect to AD and associated pathologies. The global approach for the analysis of intact proteins has decisively established the clinicopathological relevance of two major types of A β —one associated with amyloid plaques and the other with similar types of deposits, known as cerebral amyloid angiopathy, that are localized to the brain vasculature. We have also discovered associations between N-terminal truncation state of A β ₄₂ and cognitive decline. Considering these results, we believe the TDP approach will be important for guiding future anti-A β therapies. Furthermore, we have confirmed that VGF does not have a unique neuropeptide fragment associated with cognitive level or its decline. Finally, numerous other proteoforms had significant associations with either cognition or pathology; however, their role in neurodegeneration is less clear and will be a focus of further investigations.

DATA AVAILABILITY

All raw files have been deposited into the MassIVE data repository and can be accessed via accessions MSV000093728 and MSV000093727 for the human and mouse components of the study, respectively.

Supplemental data—This article contains [supplemental data](#) (47).

Acknowledgments—The authors thank In Kwon Choi and Xiaowen Liu from Tulane University for assistance with customization of the TopPIC software. We also thank Matthew Monroe for his assistance in depositing the raw proteomic data onto MassIVE. A portion of the research was performed using EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program.

Funding and additional information—This work was supported by the NIH National Institute of Aging grants U01 AG061356 (P. L. D. J. and D. A. B.) and R01 AG015819, R01 AG17917, P30 AG10161, and P30 AG72975 (D. A. B.). ROS-MAP resources can be requested at <https://www.radc.rush.edu> and www.synapse.org.

Author contributions—J. M. F., A. N. I., S. O., D. A. B., and V. A. P. writing—review and editing; J. M. F. and V. A. P. writing—original draft; J. M. F. and V. A. P. project administration; J. M. F., S. M. W., T. L. F., M. Z., R. J. M., and V. A. P. methodology; J. M. F. investigation; J. M. F., A. N. I., S. T., S. S. K., L. Y., and V. A. P. formal analysis; J. M. F., W.-J. Q., P. L. D. J., and

V. A. P. conceptualization; A. N. I., S. S. K., and L. Y. visualization; S. T. and R. J. M. resources; M. Z., W.-J. Q., L. P.-T., and V. A. P. supervision; D. A. B. and P. L. D. J. funding acquisition.

Conflicts of interests—The authors declare that they have no conflicts of interests with the contents of this article.

Abbreviations—The abbreviations used are: A β , amyloid beta; ABC, ammonium carbonate; AD, Alzheimer's disease; BUP, bottom-up proteomics; CAA, cerebral amyloid angiopathy; CV, compensation voltage; FA, formic acid; FAIMS, field asymmetric waveform ion mobility spectrometry; HB, homogenization buffer; HFIP, hexafluoroisopropanol; LC-MS/MS, liquid-chromatography mass spectrometric analysis; pLDDT, predicted local distance difference test; PTM, post-translational modification; ROS/MAP, Religious Orders Study and Rush Memory and Aging Project; TDP, top-down proteomics.

Received January 28, 2025, and in revised form, April 28, 2025
Published, MCPRO Papers in Press, May 5, 2025, <https://doi.org/10.1016/j.mcpro.2025.100983>

REFERENCES

- GBD 2016 Dementia Collaborators. (2019) 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. *Lancet Neurol.* **18**, 88–106
- Golde, T. E. (2022) Alzheimer's disease - the journey of a healthy brain into organ failure. *Mol. Neurodegener.* **17**, 18
- DeTure, M. A., and Dickson, D. W. (2019) The neuropathological diagnosis of Alzheimer's disease. *Mol. Neurodegener.* **14**, 32
- Ferreira, D., Nordberg, A., and Westman, E. (2020) Biological subtypes of Alzheimer disease: a systematic review and meta-analysis. *Neurology* **94**, 436–448
- Bellenguez, C., Küçükali, F., Jansen, I. E., Kleindam, L., Moreno-Grau, S., Amin, N., et al. (2022) New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436
- Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., et al. (2019) Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337
- Johnson, E. C. B., Carter, E. K., Dammer, E. B., Duong, D. M., Gerasimov, E. S., Liu, Y., et al. (2022) Large-scale deep multi-layer analysis of Alzheimer's disease brain reveals strong proteomic disease-related changes not observed at the RNA level. *Nat. Neurosci.* **25**, 213–225
- Johnson, E. C. B., Dammer, E. B., Duong, D. M., Ping, L., Zhou, M., Yin, L., et al. (2020) Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat. Med.* **26**, 769–780
- Roberts, J. A., Varma, V. R., An, Y., Varma, S., Candia, J., Fantoni, G., et al. (2021) A brain proteomic signature of incipient Alzheimer's disease in young APOE epsilon4 carriers identifies novel drug targets. *Sci. Adv.* **7**, eabi8178
- Toby, T. K., Fornelli, L., and Kelleher, N. L. (2016) Progress in top-down proteomics and the analysis of proteoforms. *Annu. Rev. Anal. Chem.* **9**, 499–519
- Bennett, D. A., Schneider, J. A., Arvanitakis, Z., and Wilson, R. S. (2012) Overview and findings from the religious orders study. *Curr. Alzheimer Res.* **9**, 628–645
- Bennett, D. A., Schneider, J. A., Buchman, A. S., Barnes, L. L., Boyle, P. A., and Wilson, R. S. (2012) Overview and findings from the rush memory and aging project. *Curr. Alzheimer Res.* **9**, 646–663
- Tholey, A., and Becker, A. (2017) Top-down proteomics for the analysis of proteolytic events - methods, applications and perspectives. *Biochim. Biophys. Acta Mol. Cell Res.* **1864**, 2191–2199

14. Kaulich, P. T., Cassidy, L., Winkels, K., and Tholey, A. (2022) Improved identification of proteoforms in top-down proteomics using FAIMS with internal CV stepping. *Anal. Chem.* **94**, 3600–3607
15. Fulcher, J. M., Makaju, A., Moore, R. J., Zhou, M., Bennett, D. A., De Jager, P. L., et al. (2021) Enhancing top-down proteomics of brain tissue with FAIMS. *J. Proteome Res.* **20**, 2780–2795
16. Stine, W. B., Dahlgren, K. N., Krafft, G. A., and LaDu, M. J. (2003) *In Vitro* characterization of conditions for amyloid- β peptide oligomerization and fibrillogenesis. *J. Biol. Chem.* **278**, 11612–11622
17. Martin, E. A., Fulcher, J. M., Zhou, M., Monroe, M. E., and Petyuk, V. A. (2023) TopPICR: a companion R package for top-down proteomics data analysis. *J. Proteome Res.* **22**, 399–409
18. Catherman, A. D., Durbin, K. R., Ahlf, D. R., Early, B. P., Fellers, R. T., Tran, J. C., et al. (2013) Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. *Mol. Cell Proteomics* **12**, 3465–3473
19. Melani, R. D., Gerbasi, V. R., Anderson, L. C., Sikora, J. W., Toby, T. K., Hutton, J. E., et al. (2022) The Blood Proteoform Atlas: a reference map of proteoforms in human hematopoietic cells. *Science* **375**, 411–418
20. De Jager, P. L., Ma, Y., McCabe, C., Xu, J., Vardarajan, B. N., Felsky, D., et al. (2018) A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data* **5**, 180142
21. Bennett, D. A., Buchman, A. S., Boyle, P. A., Barnes, L. L., Wilson, R. S., and Schneider, J. A. (2018) Religious orders study and rush memory and aging project. *J. Alzheimers Dis.* **64**, S161–S189
22. Bennett, D. A., Schneider, J. A., Arvanitakis, Z., Kelly, J. F., Aggarwal, N. T., Shah, R. C., et al. (2006) Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* **66**, 1837–1844
23. Bennett, D. A., Wilson, R. S., Schneider, J. A., Evans, D. A., Beckett, L. A., Aggarwal, N. T., et al. (2002) Natural history of mild cognitive impairment in older persons. *Neurology* **59**, 198–205
24. Bennett, D. A., Schneider, J. A., Aggarwal, N. T., Arvanitakis, Z., Shah, R. C., Kelly, J. F., et al. (2006) Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. *Neuro-epidemiology* **27**, 169–176
25. Boyle, P. A., Yu, L., Leurgans, S. E., Wilson, R. S., Brookmeyer, R., Schneider, J. A., and Bennett, D. A. (2019) Attributable risk of Alzheimer's dementia attributed to age-related neuropathologies. *Ann. Neurol.* **85**, 114–124
26. Boyle, P. A., Wang, T., Yu, L., Wilson, R. S., Dawe, R., Arfanakis, K., et al. (2021) To what degree is late life cognitive decline driven by age-related neuropathologies? *Brain* **144**, 2166–2175
27. Iturria-Medina, Y., Adewale, Q., Khan, A. F., Ducharme, S., Rosa-Neto, P., O'Donnell, K., et al. (2022) Unified epigenomic, transcriptomic, proteomic, and metabolomic taxonomy of Alzheimer's disease progression and heterogeneity. *Sci. Adv.* **8**, eabo6764
28. Felsky, D., Patrick, E., Schneider, J. A., Mostafavi, S., Gaiteri, C., Patso-poulos, N., et al. (2018) Polygenic analysis of inflammatory disease variants and effects on microglia in the aging brain. *Mol. Neurodegener.* **13**, 38
29. Felsky, D., Roostaei, T., Nho, K., Risacher, S. L., Bradshaw, E. M., Petyuk, V., et al. (2019) Neuropathological correlates and genetic architecture of microglial activation in elderly human brain. *Nat. Commun.* **10**, 409
30. Kou, Q., Xun, L., and Liu, X. (2016) TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **32**, 3495–3497
31. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47
32. Johnson, W. E., Li, C., and Rabinovic, A. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127
33. Li, M., Gray, W., Zhang, H., Chung, C. H., Billheimer, D., Yarbrough, W. G., et al. (2010) Comparative shotgun proteomics using spectral count data and quasi-likelihood modeling. *J. Proteome Res.* **9**, 4295–4305
34. Langfelder, P., and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559
35. Stacklies, W., Redestig, H., Scholz, M., Walther, D., and Selbig, J. (2007) pcaMethods—a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* **23**, 1164–1167
36. Zhu, W., Smith, J. W., and Huang, C. M. (2010) Mass spectrometry-based label-free quantitative proteomics. *J. Biomed. Biotechnol.* 840518. <https://doi.org/10.1155/2010/840518>
37. Lundgren, D. H., Hwang, S. I., Wu, L., and Han, D. K. (2010) Role of spectral counting in quantitative proteomics. *Expert Rev. Proteomics* **7**, 39–53
38. Compton, P. D., Zamdborg, L., Thomas, P. M., and Kelleher, N. L. (2011) On the scalability and requirements of whole protein mass spectrometry. *Anal. Chem.* **83**, 6868–6874
39. Anderson, L. C., DeHart, C. J., Kaiser, N. K., Fellers, R. T., Smith, D. F., Greer, J. B., et al. (2017) Identification and characterization of human proteoforms by top-down LC-21 tesla FT-ICR mass spectrometry. *J. Proteome Res.* **16**, 1087–1096
40. Stamm, S., Casper, D., Lees-Miller, J. P., and Helfman, D. M. (1993) Brain-specific tropomyosins TMBR-1 and TMBR-3 have distinct patterns of expression during development and in adult brain. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 9857–9861
41. Kawarabayashi, T., Younkin, L. H., Saido, T. C., Shoji, M., Ashe, K. H., and Younkin, S. G. (2001) Age-dependent changes in brain, CSF, and plasma amyloid (beta) protein in the Tg2576 transgenic mouse model of Alzheimer's disease. *J. Neurosci.* **21**, 372–381
42. Mori, H., Takio, K., Ogawara, M., and Selkoe, D. J. (1992) Mass spectrometry of purified amyloid beta protein in Alzheimer's disease. *J. Biol. Chem.* **267**, 17082–17086
43. Moore, B. D., Chakrabarty, P., Levites, Y., Kukar, T. L., Baine, A. M., Moroni, T., et al. (2012) Overlapping profiles of Abeta peptides in the Alzheimer's disease and pathological aging brains. *Alzheimers Res. Ther.* **4**, 18
44. Zheng, S., and Doucette, A. A. (2016) Preventing N- and O-formylation of proteins when incubated in concentrated formic acid. *Proteomics* **16**, 1059–1068
45. Stine, W. B., Jr., Dahlgren, K. N., Krafft, G. A., and LaDu, M. J. (2003) *In vitro* characterization of conditions for amyloid-beta peptide oligomerization and fibrillogenesis. *J. Biol. Chem.* **278**, 11612–11622
46. Shen, F., Tang, S., and Liu, L. (2011) Hexafluoro-2-propanol as a potent cosolvent for chemical ligation of membrane proteins. *Sci. China Chem.* **54**, 110–116
47. Kandi, S., Cline, E. N., Rivera, B. M., Viola, K. L., Zhu, J., Condello, C., et al. (2023) Amyloid beta proteoforms elucidated by quantitative LC/MS in the 5xFAD mouse model of Alzheimer's disease. *J. Proteome Res.* **22**, 3475–3488
48. Huang, K. F., Liu, Y. L., Cheng, W. J., Ko, T. P., and Wang, A. H. (2005) Crystal structures of human glutamyl cyclase, an enzyme responsible for protein N-terminal pyroglutamate formation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 13117–13122
49. Jawhar, S., Wirths, O., and Bayer, T. A. (2011) Pyroglutamate amyloid-beta (A β): a hatchet man in Alzheimer disease. *J. Biol. Chem.* **286**, 38825–38832
50. Wang, S., Yang, F., Petyuk, V. A., Shukla, A. K., Monroe, M. E., Gritsenko, M. A., et al. (2017) Quantitative proteomics identifies altered O-GlcNAcylation of structural, synaptic and memory-associated proteins in Alzheimer's disease. *J. Pathol.* **243**, 78–88
51. Zhang, W., Liu, T., Dong, H., Bai, H., Tian, F., Shi, Z., et al. (2017) Synthesis of a highly azide-reactive and thermosensitive biofunctional reagent for efficient enrichment and large-scale identification of O-GlcNAc proteins by mass spectrometry. *Anal. Chem.* **89**, 5810–5817
52. He, J., Fan, Z., Tian, Y., Yang, W., Zhou, Y., Zhu, Q., et al. (2022) Spatiotemporal activation of protein O-GlcNAcylation in living cells. *J. Am. Chem. Soc.* **144**, 4289–4293
53. Noguchi-Shinohara, M., Yuki-Nozaki, S., Abe, C., Mori, A., Horimoto, M., Yokogawa, M., et al. (2022) Diabetes mellitus, elevated hemoglobin A1c, and glycated albumin are associated with the presence of all-cause dementia and Alzheimer's disease: the JPSC-AD study. *J. Alzheimer's Dis.* **85**, 235–247
54. Nandi, A., Sprung, R., Barma, D. K., Zhao, Y., Kim, S. C., Falck, J. R., et al. (2006) Global identification of O-GlcNAc-Modified proteins. *Anal. Chem.* **78**, 452–458

55. Lin, C.-H., Liao, C. C., Wang, S. Y., Peng, C. Y., Yeh, Y. C., Chen, M. Y., *et al.* (2022) Comparative O-GlcNAc proteomic analysis reveals a role of O-GlcNAcylated SAM68 in lung cancer aggressiveness. *Cancers* **14**, 243
56. Teo, C. F., Ingale, S., Wolfert, M. A., Elsayed, G. A., Nöt, L. G., Chatham, J. C., *et al.* (2010) Glycopeptide-specific monoclonal antibodies suggest new roles for O-GlcNAc. *Nat. Chem. Biol.* **6**, 338–343
57. Phoomak, C., Park, D., Silsirivanit, A., Sawanyawisuth, K., Vae-teewoottacharn, K., Detarya, M., *et al.* (2019) O-GlcNAc-induced nuclear translocation of hnRNP-K is associated with progression and metastasis of cholangiocarcinoma. *Mol. Oncol.* **13**, 338–357
58. Song, H., Ma, J., Bian, Z., Chen, S., Zhu, J., Wang, J., *et al.* (2019) Global profiling of O-GlcNAcylated and/or phosphorylated proteins in hepatoblastoma. *Signal. Transduct. Targeted Ther.* **4**, 40
59. Levine, Z. G., Fan, C., Melicher, M. S., Orman, M., Benjamin, T., and Walker, S. (2018) O-GlcNAc transferase recognizes protein substrates using an asparagine ladder in the tetratricopeptide repeat (TPR) superhelix. *J. Am. Chem. Soc.* **140**, 3510–3513
60. Hanisch, F.-G., Green, B. N., Bateman, R., and Peter-Katalinic, J. (1998) Localization of o-glycosylation sites of MUC1 tandem repeats by QTOF ESI mass spectrometry. *J. Mass Spectrom.* **33**, 358–362
61. Fridriksson, E. K., Beavil, A., Holowka, D., Gould, H. J., Baird, B., and McLafferty, F. W. (2000) Heterogeneous glycosylation of immunoglobulin E constructs characterized by top-down high-resolution 2-D mass spectrometry. *Biochemistry* **39**, 3369–3376
62. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589
63. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Židek, A., *et al.* (2021) Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596
64. Piehowski, P. D., Petyuk, V. A., Orton, D. J., Xie, F., Moore, R. J., Ramirez-Restrepo, M., *et al.* (2013) Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. *J. Proteome Res.* **12**, 2128–2137
65. Yu, L., Petyuk, V. A., Gaiteri, C., Mostafavi, S., Young-Pearse, T., Shah, R. C., *et al.* (2018) Targeted brain proteomics uncover multiple pathways to Alzheimer's dementia. *Ann. Neurol.* **84**, 78–88
66. Askenazi, M., Kavanagh, T., Pires, G., Ueberheide, B., Wisniewski, T., and Drummond, E. (2023) Compilation of reported protein changes in the brain in Alzheimer's disease. *Nat. Commun.* **14**, 4466
67. Fife, A., Spönnle, I., Koziel, V., Kriem, B., Yen Potin, F. T., Bihain, B. E., *et al.* (2006) Microtubule-associated protein MAP1A, MAP1B, and MAP2 proteolysis during soluble amyloid beta-peptide-induced neuronal apoptosis. Synergistic involvement of calpain and caspase-3. *J. Biol. Chem.* **281**, 229–240
68. Beckmann, N. D., Lin, W. J., Wang, M., Cohain, A. T., Charney, A. W., Wang, P., *et al.* (2020) Multiscale causal networks identify VGF as a key regulator of Alzheimer's disease. *Nat. Commun.* **11**, 3942
69. Regan, P., Mitchell, S. J., Kim, S. C., Lee, Y., Yi, J. H., Barbat, S. A., *et al.* (2021) Regulation of synapse weakening through interactions of the microtubule associated protein tau with PACSIN1. *J. Neurosci.* **41**, 7162–7170
70. Greenwood, A. K., Montgomery, K. S., Kauer, N., Woo, K. H., Leanza, Z. J., Poehlman, W. L., *et al.* (2020) The AD knowledge portal: a repository for multi-omic data on Alzheimer's disease and aging. *Curr. Protoc. Hum. Genet.* **108**, e105
71. Greenwood, A. K., Gockley, J., Daily, K., Aluthgamage, D., Leanza, Z., Sieberts, S. K., *et al.* (2020) Agora: an open platform for exploration of Alzheimer's disease evidence. *Alzheimer's Dement.* **16**, e046129. <https://doi.org/10.1002/alz.046129>
72. Pechstein, A., Bacetic, J., Vahedi-Faridi, A., Gromova, K., Sundborger, A., Tomlin, N., *et al.* (2014) Regulation of synaptic vesicle recycling by complex formation between intersectin 1 and the clathrin adaptor complex AP2. *P Natl. Acad. Sci. U. S. A.* **111**, 4206–4211
73. McMahon, H. T., Ushkaryov, Y. A., Edelmann, L., Link, E., Binz, T., Niemann, H., *et al.* (1993) Cellubrevin is a ubiquitous tetanus-toxin substrate homologous to a putative synaptic vesicle fusion protein. *Nature* **364**, 346–349
74. Guo, J. T., Yu, J., Grass, D., de Beer, F. C., and Kindy, M. S. (2002) Inflammation-dependent cerebral deposition of serum amyloid A protein in a mouse model of amyloidosis. *J. Neurosci.* **22**, 5900–5909
75. Yu, L., Petyuk, V. A., Lopes, K. P., Tasaki, S., Menon, V., Wang, Y., *et al.* (2023) Associations of VGF with neuropathologies and cognitive health in older adults. *Ann. Neurol.* **94**, 232–244
76. El Gaamouch, F., Audrain, M., Lin, W. J., Beckmann, N., Jiang, C., Hariharan, S., *et al.* (2020) VGF-derived peptide TLQP-21 modulates microglial function through C3aR1 signaling pathways and reduces neuropathology in 5xFAD mice. *Mol. Neurodegener.* **15**. <https://doi.org/10.1186/s13024-020-0357-x>
77. Lewis, J. E., Brameld, J. M., and Jethwa, P. H. (2015) Neuroendocrine role for VGF. *Front. Endocrinol. (Lausanne)* **6**, 3
78. Quinn, J. P., Kandigian, S. E., Trombetta, B. A., Arnold, S. E., and Carlyle, B. C. (2021) VGF as a biomarker and therapeutic target in neurodegenerative and psychiatric diseases. *Brain Commun.* **3**, fcab261
79. Granold, M., Moosmann, B., Staib-Laszczik, I., Arendt, T., Del Rey, A., Engelhard, K., *et al.* (2015) High membrane protein oxidation in the human cerebral cortex. *Redox Biol.* **4**, 200–207
80. Aksenov, M. Y., Aksenova, M. V., Butterfield, D. A., Geddes, J. W., and Markesbery, W. R. (2001) Protein oxidation in the brain in Alzheimer's disease. *Neuroscience* **103**, 373–383
81. Chao, C. C., Ma, Y. S., and Stadtman, E. R. (1997) Modification of protein surface hydrophobicity and methionine oxidation by oxidative systems. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 2969–2974
82. Cadenas-Garrido, P., Schonvandt-Alarcos, A., Herrera-Quintana, L., Vázquez-Lorente, H., Santamaría-Quiles, A., Ruiz de Francisco, J., *et al.* (2024) Using redox proteomics to gain new insights into neurodegenerative disease and protein modification. *Antioxidants (Basel)* **13**. <https://doi.org/10.3390/antiox13010127>
83. Arnesen, T., Van Damme, P., Polevoda, B., Helsens, K., Evjenth, R., Colaert, N., *et al.* (2009) Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8157–8162
84. Mair, W., Muntel, J., Tepper, K., Tang, S., Biernat, J., Seeley, W. W., *et al.* (2016) FLEXITau: quantifying post-translational modifications of tau protein in vitro and in human disease. *Anal. Chem.* **88**, 3704–3714
85. Wesseling, H., Mair, W., Kumar, M., Schlaffner, C. N., Tang, S., Beerepoot, P., *et al.* (2020) Tau PTM profiles identify patient heterogeneity and stages of Alzheimer's disease. *Cell* **183**, 1699–1713.e1613
86. Molden, R. C., and Garcia, B. A. (2014) Middle-down and top-down mass spectrometric analysis of Co-occurring histone modifications. *Curr. Protoc. Protein Sci.* **77**. <https://doi.org/10.1002/0471140864.ps2307s77>
87. Sidoli, S., Lin, S., Karch, K. R., and Garcia, B. A. (2015) Bottom-up and middle-down proteomics have comparable accuracies in defining histone post-translational modification relative abundance and stoichiometry. *Anal. Chem.* **87**, 3129–3133
88. Ives, A. N., Dunn, H. A., Afsari, H. S., Seckler, H. D. S., Foroutan, M. J., Chavez, E., *et al.* (2022) Middle-down mass spectrometry reveals activity-modifying phosphorylation barcode in a class C G protein-coupled receptor. *J. Am. Chem. Soc.* **144**, 23104–23114
89. Fornelli, L., Ayoub, D., Aizikov, K., Beck, A., and Tsybin, Y. O. (2014) Middle-down analysis of monoclonal antibodies with electron transfer dissociation orbitrap fourier transform mass spectrometry. *Anal. Chem.* **86**, 3005–3012
90. Fornelli, L., Durbin, K. R., Fellers, R. T., Early, B. P., Greer, J. B., LeDuc, R. D., *et al.* (2017) Advancing top-down analysis of the human proteome using a benchtop quadrupole-orbitrap mass spectrometer. *J. Proteome Res.* **16**, 609–618
91. Yu, D., Wang, Z., Cupp-Sutton, K. A., Guo, Y., Kou, Q., Smith, K., *et al.* (2021) Quantitative top-down proteomics in complex samples using protein-level tandem mass tag labeling. *J. Am. Soc. Mass Spectrom.* **32**, 1336–1344
92. Tyanova, S., Temu, T., and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319
93. Grasso, G. (2011) The use of mass spectrometry to study amyloid-beta peptides. *Mass Spectrom. Rev.* **30**, 347–365
94. Wildburger, N. C., Esparza, T. J., LeDuc, R. D., Fellers, R. T., Thomas, P. M., Cairns, N. J., *et al.* (2017) Diversity of amyloid-beta proteoforms in the Alzheimer's disease brain. *Sci. Rep.* **7**, 9520

95. Kaulich, P. T., Jeong, K., Kohlbacher, O., and Tholey, A. (2024) Influence of different sample preparation approaches on proteoform identification by top-down proteomics. *Nat. Methods* **21**, 2397–2407
96. Pachahara, S. K., Chaudhary, N., Subbalakshmi, C., and Nagaraj, R. (2012) Hexafluoroisopropanol induces self-assembly of beta-amyloid peptides into highly ordered nanostructures. *J. Pept. Sci.* **18**, 233–241
97. Zyla, A., Martel, A., Jurczak, P., Moliński, A., Szymańska, A., and Kozak, M. (2023) Human cystatin C induces the disaggregation process of selected amyloid beta peptides: a structural and kinetic view. *Sci. Rep.* **13**, 20833
98. Melby, J. A., Brown, K. A., Gregorich, Z. R., Roberts, D. S., Chapman, E. A., Ehlers, L. E., et al. (2023) High sensitivity top-down proteomics captures single muscle cell heterogeneity in large proteoforms. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2222081120
99. Kakuda, N., Miyasaka, T., Iwasaki, N., Nirasawa, T., Wada-Kakuda, S., Takahashi-Fujigasaki, J., et al. (2017) Distinct deposition of amyloid-beta species in brains with Alzheimer's disease pathology visualized with MALDI imaging mass spectrometry. *Acta Neuropathol. Commun.* **5**, 73
100. Moro, M. L., Giaccone, G., Lombardi, R., Indaco, A., Uggetti, A., Morbin, M., et al. (2012) APP mutations in the Abeta coding region are associated with abundant cerebral deposition of Abeta38. *Acta Neuropathol.* **124**, 809–821
101. Prelli, F., Castano, E., Glenner, G. G., and Frangione, B. (1988) Differences between vascular and plaque core amyloid in Alzheimer's disease. *J. Neurochem.* **51**, 648–651
102. Suzuki, N., Iwatsubo, T., Odaka, A., Ishibashi, Y., Kitada, C., and Ihara, Y. (1994) High tissue content of soluble beta 1-40 is linked to cerebral amyloid angiopathy. *Am. J. Pathol.* **145**, 452–460
103. Vassar, R., Kovacs, D. M., Yan, R., and Wong, P. C. (2009) The beta-secretase enzyme BACE in health and Alzheimer's disease: regulation, cell biology, function, and therapeutic potential. *J. Neurosci.* **29**, 12787–12794
104. Perez-Grijalva, V., Romero, J., Pesini, P., Sarasa, L., Monleón, I., San-José, I., et al. (2019) Plasma Abeta42/40 ratio detects early stages of Alzheimer's disease and correlates with CSF and neuroimaging biomarkers in the AB255 study. *J. Prev. Alzheimers Dis.* **6**, 34–41
105. Li, Y., Schindler, S. E., Bollinger, J. G., Ovod, V., Mawuenyega, K. G., Weiner, M. W., et al. (2022) Validation of plasma amyloid-beta 42/40 for detecting Alzheimer disease amyloid plaques. *Neurology* **98**, e688–e699
106. Delaby, C., Estellès, T., Zhu, N., Arranz, J., Barroeta, I., Carmona-Iragui, M., et al. (2022) The Abeta1-42/Abeta1-40 ratio in CSF is more strongly associated to tau markers and clinical progression than Abeta1-42 alone. *Alzheimers Res. Ther.* **14**, 20
107. Weber, D. M., Taylor, S. W., Lagier, R. J., Kim, J. C., Goldman, S. M., Clarke, N. J., et al. (2024) Clinical utility of plasma Abeta42/40 ratio by LC-MS/MS in Alzheimer's disease assessment. *Front. Neurol.* **15**, 1364658
108. Schlenzig, D., Manhart, S., Cinar, Y., Kleinschmidt, M., Hause, G., Willbold, D., et al. (2009) Pyroglutamate formation influences solubility and amyloidogenicity of amyloid peptides. *Biochemistry* **48**, 7072–7078
109. Yu, L., Vize, A., Huff, M. B., Young, M., Remmele, R. L., and He, B. (2006) Investigation of N-terminal glutamate cyclization of recombinant monoclonal antibody in formulation development. *J. Pharm. Biomed. Anal.* **42**, 455–463
110. Fischer, W. H., and Spiess, J. (1987) Identification of a mammalian glutaminyl cyclase converting glutaminyl into pyroglutamyl peptides. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 3628–3632
111. Busby, W. H., Jr., Quackenbush, G. E., Humm, J., Youngblood, W. W., and Kizer, J. S. (1987) An enzyme(s) that converts glutaminyl-peptides into pyroglutamyl-peptides. Presence in pituitary, brain, adrenal medulla, and lymphocytes. *J. Biol. Chem.* **262**, 8532–8536
112. Schilling, S., Appl, T., Hoffmann, T., Cynis, H., Schulz, K., Jagla, W., et al. (2008) Inhibition of glutaminyl cyclase prevents pGlu-Abeta formation after intracortical/hippocampal microinjection *in vivo/in situ*. *J. Neurochem.* **106**, 1225–1236
113. Schilling, S., Zeitschel, U., Hoffmann, T., Heiser, U., Francke, M., Kehlen, A., et al. (2008) Glutaminyl cyclase inhibition attenuates pyroglutamate Abeta and Alzheimer's disease-like pathology. *Nat. Med.* **14**, 1106–1111
114. Weller, J., and Budson, A. (2018) Current understanding of Alzheimer's disease diagnosis and treatment. *F1000Res.* **7**. <https://doi.org/10.12688/f1000research.14506.1>
115. Tolar, M., Abushakra, S., Hey, J. A., Porsteinsson, A., and Sabbagh, M. (2020) Aducanumab, gantenerumab, BAN2401, and ALZ-801—the first wave of amyloid-targeting drugs for Alzheimer's disease with potential for near term approval. *Alzheimer's Res. Ther.* **12**, 95
116. Bateman, R. J., Blennow, K., Doody, R., Hendrix, S., Lovestone, S., Salloway, S., et al. (2019) Plasma biomarkers of AD emerging as essential tools for drug development: an EU/US CTAD task force report. *J. Prev. Alzheimer's Dis.* **6**, 169–173
117. Jack, C. R., and Holtzman, D. M. (2013) Biomarker modeling of Alzheimer's disease. *Neuron* **80**, 1347–1358
118. Klyucherev, T. O., Olszewski, P., Shalimova, A. A., Chubarev, V. N., Tarasov, V. V., Attwood, M. M., et al. (2022) Advances in the development of new biomarkers for Alzheimer's disease. *Transl. Neurodegener.* **11**, 25
119. Avgerinos, K. I., Ferrucci, L., and Kapogiannis, D. (2021) Effects of monoclonal antibodies against amyloid-beta on clinical and biomarker outcomes and adverse event risks: a systematic review and meta-analysis of phase III RCTs in Alzheimer's disease. *Ageing Res. Rev.* **68**, 101339
120. La Porte, S. L., Bollini, S. S., Lanz, T. A., Abdiche, Y. N., Rusnak, A. S., Ho, W. H., et al. (2012) Structural basis of C-terminal beta-amyloid peptide binding by the antibody ponezumab for the treatment of Alzheimer's disease. *J. Mol. Biol.* **421**, 525–536
121. Leurent, C., Goodman, J. A., Zhang, Y., He, P., Polimeni, J. R., Gurol, M. E., et al. (2019) Immunotherapy with ponezumab for probable cerebral amyloid angiopathy. *Ann. Clin. Transl. Neurol.* **6**, 795–806
122. Greenberg, S. M., Cordonnier, C., Schneider, J. A., Smith, E. E., van Buchem, M. A., van Veluw, S. J., et al. (2021) Off-label use of aducanumab for cerebral amyloid angiopathy. *Lancet Neurol.* **20**, 596–597
123. Mintun, M. A., Lo, A. C., Duggan Evans, C., Wessels, A. M., Ardayfio, P. A., Andersen, S. W., et al. (2021) Donanemab in early Alzheimer's disease. *N. Engl. J. Med.* **384**, 1691–1704
124. Sevigny, J., Chiao, P., Bussière, T., Weinreb, P. H., Williams, L., Maier, M., et al. (2016) The antibody aducanumab reduces Abeta plaques in Alzheimer's disease. *Nature* **537**, 50–56
125. Arndt, J. W., Qian, F., Smith, B. A., Quan, C., Kilambi, K. P., Bush, M. W., et al. (2018) Structural and kinetic basis for the selectivity of aducanumab for aggregated forms of amyloid-beta. *Sci. Rep.* **8**, 6412
126. Sveikata, L., Charidimou, A., and Viswanathan, A. (2022) Vessels sing their ARIAs: the role of vascular amyloid in the age of aducanumab. *Stroke* **53**, 298–302
127. Greenberg, S. M., Bacskai, B. J., Hernandez-Guillamon, M., Pruzin, J., Sperling, R., and van Veluw, S. J. (2020) Cerebral amyloid angiopathy and Alzheimer disease - one peptide, two pathways. *Nat. Rev. Neurol.* **16**, 30–42
128. Ferri, G. L., Noli, B., Brancia, C., D'Amato, F., and Cocco, C. (2011) VGF: an inducible gene product, precursor of a diverse array of neuroendocrine peptides and tissue-specific disease biomarkers. *J. Chem. Neuroanat.* **42**, 249–261
129. Lin, W. J., Jiang, C., Sadahiro, M., Bozdagi, O., Vulchanova, L., Alberini, C. M., and Salton, S. R. (2015) VGF and its C-terminal peptide TLQP-62 regulate memory formation in Hippocampus via a BDNF-TrkB-dependent mechanism. *J. Neurosci.* **35**, 10343–10356
130. Cassimeris, L. (2002) The oncoprotein 18/stathmin family of microtubule destabilizers. *Curr. Opin. Cell Biol.* **14**, 18–24
131. Uchida, S., Martel, G., Pavlowsky, A., Takizawa, S., Hevi, C., Watanabe, Y., et al. (2014) Learning-induced and stathmin-dependent changes in microtubule stability are critical for memory and disrupted in ageing. *Nat. Commun.* **5**, 4389
132. Jin, L. W., Masliah, E., Iimoto, D., Deteresa, R., Mallory, M., Sundsmo, M., et al. (1996) Neurofibrillary tangle-associated alteration of stathmin in Alzheimer's disease. *Neurobiol. Aging* **17**, 331–341
133. Cheon, M. S., et al. (2001) Gert Lubec. In: *Protein Expression in Down Syndrome Brain*, Springer, Vienna, Austria: 281–288
134. Martzen, M. R., Nagy, A., Coleman, P. D., and Zwiars, H. (1993) Altered phosphorylation of growth-associated protein B50/GAP-43 in Alzheimer disease with high neurofibrillary tangle density. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11187–11191
135. Spencer, S. A., Schuh, S. M., Liu, W. S., and Willard, M. B. (1992) GAP-43, a protein associated with axon growth, is phosphorylated at three sites in cultured neurons and rat brain. *J. Biol. Chem.* **267**, 9059–9064