In-Depth Site-Specific O-Glycosylation Analysis of Glycoproteins and Endogenous Peptides in Cerebrospinal Fluid (CSF) from Healthy Individuals, Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD) Patients

Zhengwei Chen 1, Danqing Wang 1, Qing Yu 2, Jillian Johnson 2, Richard Shipman 3, Xiaofang Zhong 2, Junfeng Huang 2, Qinying Yu 2, Henrik Zetterberg 4 5 6 7, Sanjay Asthana 8, Cynthia Carlsson 8, Ozioma Okonkwo 8, Lingjun Li 1 2

1Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, United States.

2School of Pharmacy, University of Wisconsin, Madison, Wisconsin 53705, United States.

3Applied Science Program, University of Wisconsin-Stout, Menomonie, Wisconsin 54751, United States.

4Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, 43180 Mölndal, Sweden.

5Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, 43180 Mölndal, Sweden.

6Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, United Kingdom.

7UK Dementia Research Institute at UCL, London WC1E 6BT, United Kingdom.

8School of Medicine and Public Health, University of Wisconsin, Madison, Wisconsin

53726, United States.

†Correspondence: Professor Lingjun Li, School of Pharmacy and Department of Chemistry, University of Wisconsin-Madison, 777 Highland Avenue, Madison, Wisconsin 53705-2222

E-mail: lingjun.li@wisc.edu

Fax: +1-608-262-5345

Phone: +1-608-265-8491

Abstract

Site-specific O-glycoproteome mapping in complex biological systems provides a molecular basis for understanding the structure-function relationships of glycoproteins and their roles in physiological and pathological processes. Previous O-glycoproteome analysis in cerebrospinal fluid (CSF) focused on sialylated glycoforms, while missing information on other glycosylation types. In order to achieve an unbiased O-glycosylation profile, we developed an integrated strategy combining universal boronic acid enrichment, high-pH fractionation, and electron-transfer and higher-energy collision dissociation (EThcD) for enhanced intact O-glycopeptide analysis. We applied this strategy to analyze the O-glycoproteome in CSF, resulting in the identification of 308 O-glycopeptides from 110 O-glycoproteins, covering both sialylated and nonsialylated glycoforms. To our knowledge, this is the largest data set of O-glycoproteins and O-glycosites reported for CSF to date. We also developed a peptidomics workflow that utilized the EThcD and a three-step database searching strategy for comprehensive PTM analysis of endogenous peptides, including N-glycosylation, O-glycosylation, and other common peptide PTMs. Interestingly, among the 1411 endogenous peptides identified, 89 were O-glycosylated, and only one N-glycosylated peptide was found, indicating that CSF endogenous peptides were predominantly O-glycosylated. Analyses of the O-glycoproteome and endogenous peptidome PTMs were also conducted in the CSF of MCI and AD patients to provide a landscape of glycosylation patterns in different disease states. Our results showed a decreasing trend in fucosylation and an increasing trend of endogenous peptide O-

glycosylation, which may play an important role in AD progression.

Introduction

Predominantly produced in choroid plexuses, human cerebrospinal fluid (CSF) circulates within the ventricles of the brain and surrounds the brain and spinal cord.[1] The averaged total volume of CSF is estimated to be about 150 ml in adults, and the daily volume of CSF produced varies between 400 to 600 ml.[2] The functional role of CSF, though remaining to be defined precisely, includes mechanical protection of the central nervous system (CNS), homeostasis of the interstitial fluid in the brain, and regulation of neuronal functioning.[3, 4] Being in direct contact with the CNS, CSF reflects the ongoing physiological or pathological state of CNS most directly.[5, 6] CSF contains metabolites, peptides, proteins, enzymes, and hormones *etc.* that play important roles in many biological processes in CNS.[7] Changes in the compositions of CSF acts as a sign of pathological alterations in the CNS, and thus provides great opportunity for biomarker discovery in neurological diseases.

As one of the most complex forms of protein post-translational modifications (PTMs), glycosylation acts as a key regulatory mechanism controlling protein folding, molecular trafficking, cell adhesion, receptor activation and signal transduction.[8] According to the different amino acids that glycans can attach to, glycosylation is classified into two major classes, N-glycosylation and O-glycosylation. Biosynthesis of N-glycosylation is initiated by transfer of a pre-assembled 14 monosaccharide complex glycan to asparagine residue (Asn) within the consensus motif (Asn-X-Ser/Thr, X≠P)

containing polypeptides followed by sequential addition or removal of certain monosaccharide in a well-defined process.[9] In contrast, O-glycosylation synthesis involves attachment of a single monosaccharide to the serine/threonine (Ser/Thr) residue of a polypeptide without any definable peptide consensus motif and latter transfer of many diverse monosaccharide residues.[10, 11] As a result, a higher extent of occupancy and a larger degree of structure complexity and diversity were found for O-glycosylation.

In general, O-glycosylation can be categorized into mucin and non-mucin types according to the monosaccharide residue directly linked to polypeptides. The attached monosaccharide residue is N-acetylgalactosamine (GalNAc/HexNAc) in mucin type, whereas the attached residue can be N-acetylglucosamine (GlcNAc/HexNAc), fucose (Fuc), galactose (Gal/Hex), mannose, glucose in the non-mucin type.[9] The clustered regions of O-GalNAcylation on mucins have been shown to provide protection from proteolysis as well as unique rheological properties.[12] Mucin-type O-glycosylation on the cell surface and on secreted proteins has also been shown to modulate recognition, adhesion, communication between cells and also their surrounding environments.[13] As a nutrient- and stress-responsive modification, non-mucin type O-GlcNAcylation is extensively involved in the spatiotemporal regulation of diverse cellular processes, including transcription, epigenetic modifications and cell signaling dynamics.[14] Apart from modifying protein, O-glycosylation could also happen on the endogenous peptides, including neurotransmitters and hormones. In fact, it has been well recognized that after initial peptidase cleavages, endogenous peptides could undergo further post-

translational modifications such as amidation, acetylation, phosphorylation, sulfation, and N-/O-glycosylation.[15, 16] As an example, it has been reported that there was an extensive N-/O-glycosylation of gonadotropin, a glycoprotein polypeptide hormones.[17] Recently, our group also discovered that mouse insulin-1B and -2B chains, human insulin-B chain were O-glycosylated, and multiple O-glycoforms were identified.[18] These findings suggest that O-glycosylation may play a crucial role in the folding, trafficking, metabolic clearance and biological activity of neurotransmitters and hormones and it is import to characterize their glycosylation state.

The CSF proteome and endogenous peptidome have been extensively characterized, but there are only few reports on the O-glycosylation study. One study reported 39 O-glycopeptides from 22 CSF O-glycoproteins,[19] and another study identified 43 O-glycopeptides from 28 CSF O-glycoproteins.[20] Utilizing an enrichment approach based on hydrazide chemistry after oxidizing sialic acid (NeuAc), both of the two studies suffered from the weakness that sialic acid information was lost and only a subset of O-glycoproteome (sialylated O-glycopeptides) were characterized. As a result, the current O-glycoproteome depth is far less than satisfactory, hindering the design of studies to explore disease-related O-glycosylation alterations. Regarding the CSF endopeptidome, one study identified 730 endogenous peptides, including 138 peptides with PTMs such as acetylation, amidation, phosphorylation, Gln to pyro-Glu conversion.[21] But none of the identified peptides is reported to be glycosylated. The coverage of the CSF endopeptidome was recently expanded to over 18,000 endogenous peptides, but again glycosylated forms were not clearly defined (ref: PMID: 28044435).

In another study with 563 endogenous peptides identified,[22] Zougman et al. noted the presence of glycan oxonium ions in some spectra, indicating the existence of glycosylated endogenous peptides in CSF. By lowering the HCD collision energy, labile glycan moiety could be partially preserved, and 28 O-glycopeptides were successfully identified. However, only two O-glycan compositions were found. Considering that the two O-glycan modifications set in the search engine were found by merely incidental observation and the large diversity of O-glycans out there, it's highly possible that there are more O-glycosylated endogenous peptides in CSF, even N-glycosylation. Therefore, a systemic approach needs to be developed, which features a glycosylation-centered analytical strategy and searching strategy to take vast categories of N-/O-glycans into account.

In the present study, we first optimized the boronic acid-based enrichment strategy after PNGase F pretreatment to efficiently enrich both sialylated and non-sialylated O-glycopeptides in CSF. Combined with high-pH (HpH) fractionation and intact O-glycopeptide characterization enabled by EThcD, 987 intact O-glycopeptides were identified using 200 μg tryptic peptides from PANC1 cells. The optimized approach was then applied to CSF O-glycoproteome study, 308 intact O-glycopeptides from 182 O-glycosites and 110 O-glycoproteins were confidently identified. About 30% of the O-glycosites identified contained at least two O-glycoforms, revealing the microheterogeneity of O-glycosylation. Up until now, the dataset represents the largest site-specific O-glycoproteome reported for CSF, including 154 novel O-glycosites. For endogenous peptides analysis, we developed a peptidomics workflow that combines

CSF endogenous peptides extraction by 10k molecular weight cut-off (MWCO), EThcD fragmentation, and a three-step database searching strategy for comprehensive PTMs analysis. In this workflow, the use of EThcD enables the preserving of labile PTMs including glycosylation and phosphorylation, facilitating the accurate site localization. The three-step database searching strategy using the Byonic as the search engine allows a comprehensive PTMs analysis of endogenous peptides, including both N-/O-glycosylation, phosphorylation, amidation, acetylation, Gln to pyro-Glu conversion. To our knowledge, the peptidomics workflow developed here is the first workflow that enables a thorough and systemic analysis of N-/O-glycosylation state of endogenous peptides. In total, 1492 enodgenous peptides were identified, and 370 of them underwent post-translational modifications including O-glycosylation, phosphorylation, acetylation, amidation, Gln to pyro-Glu conversion. Among them, 95 endogenous peptides were O-glycosylated, and 15 of them were O-glycosylated neuropeptides. Interestingly, only 1 N-glycosylated peptide was found, indicating endogenous peptides were preferentially O-glycosylated. To provide a landscape picture of the O-glycosylation state of CSF glycoproteins in MCI and AD patients, in-depth O-glycoprofiling experiments were also conducted in parallel, showing a decreased fucosylation in MCI and AD. The PTMs analysis of endogenous CSF peptides in MCI and AD patients showed an increased percentage of PTMs modified peptides, including O-glycosylation, Gln->pyro-Glu and acetylation. Furthermore, the developed strategy here is readily applicable for site-specific O-glycosylation analysis of glycoproteins or endogenous peptides in other complex biological systems.

Experimental Procedures

Chemicals and materials

Dithiothreitol (DTT), PNGase F and sequencing grade trypsin were from Promega (Madison, WI). Optimal LC/MS-grade acetonitrile (ACN), methanol (MeOH) and water were from Fisher Scientific (Pittsburgh, PA). Tris base, urea (UA), sodium chloride and ammonium bicarbonate (ABC) were obtained from Fisher Scientific (Pittsburgh, PA). Formic acid (FA), 10% Sodium dodecyl sulfate solution (SDS), trifluoroacetic acid (TFA), dimethyl sulfoxide (DMSO) were purchased from Sigma-Aldrich (St. Louis, MO). C18 OMIX tips and phenylboronic acid (PBA) solid phase extraction cartridges were obtained from Agilent (Santa Clara, CA). Microcon filters YM-30 (30 kDa) and amicon Ultra-0.5 mL centrifugal filters (10 kDa) were purchased from Merck Millipore (Billerica, MA). PANC-1 pancreatic ductal adenocarcinoma cells were from ATCC (Manassas, VA).

CSF samples

48 enrollees in the Wisconsin Alzheimer's Disease Research Center (ADRC) participated in this study. The subjects comprised of 16 cognitively normal individuals who enrolled in the Wisconsin ADRC at late middle age, 16 individuals with MCI and 16 individuals with AD dementia. Detailed subjects information can be found in **Supplemental Table S1**. All MCI and AD participants were diagnosed via applicable clinical criteria in standardized and multidisciplinary consensus conferences. [23, 24] And cognitive normalcy was determined based on intact cognitive performance by a

comprehensive battery of neuropsychological tests, lack of functional impairment, and absence of neurological or psychiatric conditions that might impair cognition.[25, 26] CSF was collected by lumbar puncture of individuals under written informed consent. The University of Wisconsin Institutional Review Board approved all study procedures. Each enrollee provided a signed informed consent form before participation. CSF aliquots from each of the 16 individuals at each stage were combined into a pool of 1 mL for control, MCI and AD subjects.

PANC1 cells

PANC1 pancreatic ductal adenocarcinoma cells were maintained in complete media of DMEM/Ham's F-12 (1:1) (ATCC) supplemented with 10% fetal bovine serum (Hyclone) and 1% antibiotic-antimycotic solution (Cellgro). Cell culture flasks were incubated in the incubator containing 5% CO_2 and 98% humidity. Cells were harvested once 80% confluence was achieved, and cells with a maximum of 15 passages were used. Cell pellets were washed twice with phosphate-buffered saline, flash frozen in dry ice, and stored at -80 °C.

Protein extraction and digestion from PANC1 cells

Protein extraction and trypsin digestion were performed based on previously reported filter-aided sample preparation (FASP) protocol[27] with some modifications. Cell pellets were lysed by sonication in a solution containing digest buffer (4% SDS, 100 mM Tris/Base pH 8.0). To determine the protein concentration, the bicinchoninic acid assay (BCA assay) was conducted. Briefly, the samples were thawed and centrifuged at 16,000g for 5 min. One M DTT (final concentration 0.1 M) was added to 200 µg protein

in digestion buffer. Samples were incubated at 95°C for 3 min to reduce disulfide bonds. Two hundred µL of UA buffer (8 M UA in 100 mM Tris/Base) was added into the vial and transferred onto the 30 kDa filter. The filter was centrifuged at 14,000g for 15 min. Another 200 µL of UA buffer was added to the sample and centrifuged at 14,000g for 15 min. One hundred µL of IAA buffer (0.05 M IAA in UA buffer) was added onto the filter and gently swirled to mix, followed by incubation for 20 min at x degrees/room temperature. The filter was centrifuged at 14,000g for 10 min. One hundred µL of UA buffer was added onto the filter that was centrifuged at 14,000g for 15 min. This step was repeated twice to completely exchange SDS with UA buffer. One hundred µL of ABC buffer (50 mM) was onto the filter that was centrifuged at 14,000g for 15 min. This step was repeated twice. All the centrifugation steps were done at 20°C. Then, 10 µL of trypsin and 40 µL of ABC buffer were added onto the filter followed by incubation at 37 °C in a water bath for 18h. After incubation, the filter was transferred to a fresh collection vial and centrifuged at 14,000g for 10min. Fifty µL 0.5 M NaCl solution was added onto the filter followed by centrifugation at 14,000g for 10min. The procedure was repeated once. TFA was added into the vial to a final concentration of 0.25%. Samples were desalted using a SepPak C18 SPE cartridge (Waters, Milford, MA).

CSF sample processing

CSF samples were separated into peptide fraction and protein fraction using 10 kDa molecular weight cut-off (MWCO) following a previously described protocol.[28, 29] The peptide fraction was injected for LC-MS/MS analysis after desalting with SepPak C18 SPE cartridge. The protein fraction was dissolved in 8 M urea, reduced (5 mM DTT, 1

h at room temperature) and alkylated (15 mM IAA, 30 min at room temperature in the dark). Alkylation was quenched by incubation in 9 mM DTT by adding a second aliquot of DTT at room temperature. Samples were diluted with 50 mM Tris buffer to make urea below 1 M. Trypsin was added in a 1:50 (w/w) ratio and incubated for 18 h at 37 °C. Digestion was quenched by adding TFA to a final concentration of 0.3%. Finally, the samples were desalted on a C18 SepPak cartridge (Waters, Milford, MA) and dried under vacuum. Tryptic peptides were then incubated with PNGase F to remove N-glycans to avoid detection interference with O-glycosylated peptides. Then the samples were subject to C18 SepPak cartridge to get rid of salts and released N-glycans. Samples were dried down under vacuum and stored under -80 °C.

Boronic acid enrichment

Boronic acid enrichment was conducted according to a previously reported protocol with slight modifications.[24] PBA cartridges were first conditioned with 1 mL of anhydrous DMSO for 3 times. Tryptic peptides were dissolved in 35μL DMSO, loaded onto the cartridge and incubated in 37℃ for 2 h with both ends of the cartridge sealed. Unbound peptides were washed away with 1 mL anhydrous can. The washing step was repeated twice. After incubation in 600μL of 0.1% TFA at 37℃ for 1 h, the O-glycopeptides were eluted. The elution step was repeated once. The enriched O-glycopeptides were dried down under vacuum and stored at -80 °C until analysis.

High-pH fractionation

Enriched O-glycopeptides were fractionated using a C18 reverse-phase column (2.1 × 150 mm, 5 μm, 100 Å) operating at 0.3 mL/min in high-pH mode. Samples were first

reconstituted in 100 µL of 10 mM ammonium formate at pH 10 (mobile phase A). Mobile phase B consisted of 90% CAN and 10 mM ammonium formate at pH=10. O-glycopeptides were eluted with a gradient as follows: 1 % A (0–3 min), 1-35% (3-50 min), 35-60% (50-54 min), 60-70% (54-58 min), and 70-100 % (58-59 min). Seven fractions were collected from 4 min to 62 min. Fractions were dried down under vacuum.

LC-MS/MS analysis

After dissolved in 0.1% FA (mobile phase A), samples were analyzed on the Orbitrap Fusion™ Lumos™ Tribrid™ Mass Spectrometer (Thermo Fisher Scientific, San Jose, CA) coupled to a Dionex UPLC system. Mobile phase B consisted of 0.1% formic acid in ACN. Peptides were loaded and separated on a 75 µm x 15 cm homemade column packed with 1.7 µm, 150 Å, BEH C18 material obtained from a Waters UPLC column (part no. 186004661). The LC gradient was set as follows, 3%-30% A (18-98min), 30%-75% A (100-108 min) and 75%-95% A (108-118min). The mass spectrometer was operated in data-dependent (DDA) mode to automatically switch between MS and MS/MS acquisition. An MS1 scan was acquired from 400–1800 (120,000 resolution, $4e^5$ AGC, 100 ms injection time) followed by EThcD MS/MS acquisition of the precursors with the highest charge states in an order of intensity and detection in the Orbitrap (60,000 resolution, $3e^5$ AGC, 100 ms injection time). EThcD was performed with optimized user defined charge dependent reaction time (2+ 50 ms; 3+ 20 ms; 4+ 20 ms; 5+ 20ms; 6 + 9 ms; 7+; 9 ms; 8+ 9ms) supplemented by 33% HCD activation.

Data analysis

All raw data files were searched against UniProt *homo sapien*s reviewed database (08.10.2016, 20, 152 sequences), using PTM-centric search engine Byonic (version 2.9.38, Protein Metrics, San Carlos, CA) incorporated in Proteome Discoverer (PD 2.1). Trypsin was selected as the enzyme and two maximum missed cleavages were allowed. Searches were performed with a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.01 Da. Static modifications consisted of carbamidomethylation of cysteine residues (+57.02146 Da). Dynamic modifications consisted of oxidation of methionine residues (+15.99492 Da), deamidation of asparagine and glutamine (+0.98402 Da). Oxidation and deamidation were set as "rare" modification, and O-glycosylation was set as "common" modification through Byonic node. Two rare modification and one common modification were allowed. Human O-glycan database embedded in Byonic, which contains 70 glycan entities, was used. As for glycopeptide FDR control, Byonic default settings was applied that cuts the protein list after the 20th decoy proteins or at the point in the list at which the protein FDR first reaches 1%, whichever cut gives more proteins. After that, Byonic estimated the spectrum-level FDR of the remaining PSMs to the reported proteins which will typically be in the range 0-5%. Only these O-glycopeptides with PSMs FDR ≤ 1% and Byonic score over 50 and Delta Mod Score over 40 were reported. For endogenous peptides, three consecutive searches were conducted as shown in **Figure S1**. Searches were performed with a precursor mass tolerance of 10 ppm and a fragment mass tolerance of 0.01 Da, and digestion was set to unspecific. The 1st search was using the whole human protein database and total "rare" modification was set to 1, and dynamic modifications includes

oxidation of methionine residues (+15.99492 Da), amidation at peptide C-terminal (-0.984016), acetylation at peptide N-terminal and lysine, serine (+42.010565), and Gln to pyro-Glu conversion (-17.026549). The proteins precursors identified in the first search were combined with the reported protein precursors[21, 22] to construct focused protein database. The 2nd search was using the same dynamic modification as first search, except adding N-glycosylation and O-glycosylation as "common" modification, with total common modification set as 1. The human glycan database used is 182 entities for N-glycan and 70 entities for human O-glycan, and focused protein database was used. The 2nd search yielded a list of identified N-glycans and O-glycans, and these glycans were used to build a focused glycan database. For the 3nd search, dimethylation at peptide N-terminal, lysine, arginine (+28.0313), deamidation at asparagine (+0.984016), methylation at peptide N-terminal, lysine, arginine, glutamine (14.01565), phosphorylation at serine, tyrosine, threonine (+79.966331) were added to dynamic "rare" modifications. N-glycosylation and O-glycosylation as "common" modification, with total common modification set as 1. Focused protein database and focused glycan database were used in the 3rd search. Only these peptides with PSMs FDR $\leqslant$ 1% and Byonic score over 50 were reported. For endogenous peptides with PTMs, a Delta Mod Score over 40 was required. The reported O-glycosylation information was extracted from Uniprot (07.13.2017). The peptide sequence analysis of glycosylation site-containing peptides was conducted using the on-line tool Weblogo 3 (http://weblogo.threeplusone.com).[30]

Results

Optimization of boronic acid enrichment strategy

To avoid the interference of highly abundant nonglycosylated peptides, enrichment is a key step for the success of glycopeptide detection. Different aspects of the properties of O-glycopeptide could be utilized for enrichment purpose, including lectin recognized glycan motif, metabolic labeling by azido analog and sialic acid binding affinity to titanium dioxide (TiO2) beads.[31] The *cis*-diol groups on sialic acid group could also be converted to aldehydes by periodate oxidation, and the sialyalted glycopeptides could further be enriched based on hydrazide chemistry.[19] In addition, the *cis*-diol groups could react with boronic acids to form reversible covalent bond and later be released under acidic conditions without any side effect.[32, 33] The presence of *cis*-diol groups in all glycoproteins and glycopetpides makes it a universal enrichment method for both N-glycopeptide and O-glycopeptide. Although this method has been applied for large-scale N-glycoproteome study[34] and small-scale analysis of O-GlcNAcylation[24], it has not yet been used for the comprehensive O-glycoproteome analysis in complex biological samples.

Here, we used phenylboronic acid (PBA) solid phase extraction cartridge to extract the O-glycopeptides from complex tryptic peptides sample. PNGase F was first used to remove the N-glycans to avoid its interference with the enrichment process and the later MS detection of O-glycopeptides. Starting from 200 μg tryptic peptides from PANC1 cells, a total of 213 intact O-glycopeptides were identified (**Figure 1a**). In the spectra, we found that quite many nonglycosylated peptides were co-enriched which may

suppress the glycopeptide signal. Thus, to reduce sample complexity and minimize the effects of co-eluted nonglycosylated peptides, off-line high-pH fractionation (HpH) was utilized, which has also been shown highly orthogonal to the following LC-MS/MS analysis using low-pH reversed-phase chromatography.[35] Seven HpH fractions were collected and injected for analysis. The results showed that 229 intact O-glycopeptides were identified, with only a slight improvement (**Figure 1a, supplemental table S2)**. A closer look at the distribution of the number of O-glycopeptides identified showed that O-glycopeptides were mainly distributed in the first four fractions **(Figure 1b)**, whereas the number of peptides identified in a typical proteomics study using the same HpH methods resulted in an evenly distributed peptides number in all fractions. This should be no surprise considering that the increased hydrophilicity of O-glycopeptides would shift the elution to an earlier time frame on a C18 column. Then the first fraction was combined with the last three fractions and the saved instrument time by using less fractions will be used for more technical replicates. Another round of experiment using the same amount of starting material employing four fractions and four technical replicates yielded a total number of 987 intact O-glycopeptides, with a 4-fold increase. Such a large increase mainly result from the four replicates employed for each fraction. It is well-known that two or more technical replicates are needed to get the maximum coverage of peptides in a certain sample due to the randomness of data-dependent analysis (DDA) mode in short-gun proteomics.[36] Such effects would be more significant when it comes to glycoproteomics study, as the co-enriched nonglycosylated peptide would compete with glycopeptide for fragmentation and the higher ionization

efficiency would help them gain advantage over glycosylated peptide. Here, four replicates were used to find an optimal number of replicates. As shown in **Figure 1c**, two replicates only yielded less than 60% coverage, and at least three technical replicates were needed to reach a coverage of more than 90%.

Site-specific O-glycopeptide characterization

Due to the inherent structural complexity of O-glycopeptide, multiple facets exist for site-specific O-glycopeptide analysis, including glycosylation site localization, glycan structure and peptide sequence information. Various analytical technics have been developed to advance the MS-based O-glycopeptide analysis, which has greatly improved the overall performance of large-scale O-glycoproteomics workflow.[31] One focus is the development and optimization of fragmentation techniques. In general, c/z ions produced in ETD-MS/MS provide information on the glycosylation site and the peptide identity, but the abundant unreacted and charge reduced precursors hamper its performance and glycan fragments B/Y ions cannot be obtained in this mode. HCD produces B/Y ions and abundant oxonium ions for glycan identification, as well as b/y ions for peptide sequence information, but it does poorly in glycosylation site localization. In order to obtain comprehensive information, multiple dissociation modes are usually needed to draw a complete picture of O-glycopeptide structure. To this end, the "hybrid" dissociation method electron transfer and higher-energy collision dissociation (EThcD), introduced by Heck and co-workers[37] has shown great potential for O-glycopeptide analysis. Its superior performance in labile PTM analysis was first

demonstrated in phosphopeptide analysis, with improved site localization, richer and more informative peptide backbone spectra compared to ETD and HCD.[38] When applied to O-glycopeptide analysis, EThcD can produce rich fragment ions information for glycan moiety, peptide backbone and glycosylation site identification in one spectrum, enabling enhanced site-specific O-glycopeptide analysis.[39-42]

As an example, EThcD spectrum of the identified O-glycopeptide VHENENIGTTEPGEHQEAK was shown in **Figure 2a**. In the low mass region, the glycan signature oxonium ions, including 138.06 (HexNAc-$2H_2O$-CH_2O), 168.06 (HexNAc-$2H_2O$), 186.08 (HexNAc-H_2O) and 204.09 (HexNAc), indicates the spectrum belongs to a glycopeptide. The glycan fragments 274.05 (NeuAc) and 292.11 (NeuAc-18) suggests the presence of sialic acid. Based on these information, along with fragments HexNAcHex (366.14), HexNAcHexNeuAc (657.23) and accurate precursor mass match (1.2 ppm), the O-glycan composition can be assigned as HexNAcHexNeuAc, a sialylated Tf antigen (**Figure 2a)**. The peptide sequence VHENENIGTTEPGEHQEAK were deduced based on the wealthy backbone fragments b/y ions and c/z ions. Two possible O-glycosylation sites, 9^{th} and 10^{th} threonine, exist on the peptide sequence, and the c2−c10 ions enabled unambiguous localization of the glycan at 9^{th} threonine, which agrees well with previous reports.[19, 20] Besides, abundant glycopeptide peaks with total or partial glycan moiety loss (pep, pep+HexNAc, pep+HexNAcHex) were also detected, further confirming the assignment. In the same way, the O-glycosites of endogenous O-glycopeptide AAVGTSAAPVPSDNH, with 3 potential O-glycosites (5^{th} T, 6^{th} S, 12^{th} S), were also

unambiguously localized at 6th serine residue using the c3-c7 ions, and peptide sequence and O-glycan compositions were deduced using c/z, b/y and glycan fragment ions (**Figure 2b)**. This O-glycopeptide originates from apolipoprotein E, and the O-glycosite at 6th serine has been tentatively identified and reported as a novel O-glycosite in the previous study.[19] In the previous report, although the author narrowed down the O-glycosite to the 5th threonine and 6th serine, the accurate site was not able to be assigned due to a lack of informative fragments generated by electron-capture dissociation (ECD). Furthermore, limited by the enrichment method (sialic acid was periodate-oxidized) used in their study, the sialylation information of the O-glycan was lost. The EThcD spectrum in the current study not only helps us precisely pinpoint this novel O-glycosite, but also explicitly assign the O-glycan composition as HexNAcHexNeuAc(2), fully preserving sialylation information. Similar to O-glycosylation, phosphorylation is also quite labile. Benefiting from EThcD, the detection of c4 ion without phosphorylation group and c5 ion with phosphorylation group help us unambiguously localize the phosphorylation site at the 5th serine residue. Thus, along with other b/y, c/z ions and 98 Da neutral loss, the endogenous phosphorylated peptide VDPKSKEEDKH could be confidently identified as shown in **Figure 2c.**

Site-specific O-glycoproteome analysis in CSF

As discussed previously, the current depth of site-specific O-glycoproteome in CSF is not satisfactory. One of our goal in the present study is to improve the current O-

glycoproteome coverage by developing a more robust analytical strategy. Benefiting from the optimized O-glycopeptide enrichment strategy, along with robust hybrid fragmentation EThcD techniques, we were able to identify 308 intact O-glycopeptides from 181 O-glycosites and 110 O-glycoproteins, which is a large increase compared to previous reports (**Figure 3a**). After combining O-glycopeptides with the same glycosite and glycan composition but different peptide sequence length, there were 292 unique O-glycoforms in total (**Figure 3a**). Overall, the majority of O-glycoproteins (72%) carry only one single O-glycosite, and 25% carry 2-3 O-glycosites, with a slight percentage (5%) more than 4 sites (**Figure 3b**). There are around 30% O-glycosites identified carrying more than 2 O-glycan compositions, indicating the microheterogeneity of O-glycosylation (**Figure 3c**).

Among different kinds of O-glycoforms, mucin-type core 1 (Galβ1→3GalNAc) is known to be the major constituent of O-glycans.[43] In the present study, around 48.3% (141/292) of O-glycoforms had core 1 glycoforms (including $HexNAc_{(1)}Hex_{(1)}$, $HexNAc_{(1)}Hex_{(1)}Fuc_{(1)}$, $HexNAc_{(1)}Hex_{(1)}Fuc_{(1)}NeuAc_{(1)}$, $HexNAc_{(1)}Hex_{(1)}NeuAc_{(1)}$, $HexNAc_{(1)}Hex_{(1)}NeuAc_{(2)}$, $HexNAc_{(1)}Hex_{(1)}NeuAc_{(3)}$, $HexNAc_{(1)}Hex_{(2)}Fuc_{(1)}$). This O-glycosylation feature also agrees quite well with the core 1 O-glycoform percentage in human serum, with 46.4% reported.[44] The sialylated core 1 glycoforms, mainly $HexNAc_{(1)}Hex_{(1)}NeuAc_{(1)}$ and $HexNAc_{(1)}Hex_{(1)}NeuAc_{(2)}$, were the domain compositions (~67%), and the same phenomenon was also observed in human serum. This is probably because more than 80% of CSF proteins originates from the plasma filtrate, preserving a similar glycosylation pattern.[19, 44, 45] Such sialylation signature

information obtained here could never be obtained if O-glycopeptide enrichment was based on sialic acid oxidation, which was the case in previous studies.[19, 20] Overall, the sialylated O-glycoforms accounted for 51% of the total O-glycoforms, suggesting sialylation was a common modification on O-glycans and an enrichment approach that can preserve full sialylation information is highly desirable. Interestingly, acting also as a capping unit to elongate the glycan branch, fucosylated core 1 O-glycoforms only accounted for 7% and total fucosylated O-glycoforms accounted for 29%, a lot smaller percentage compared to sialylated forms. As shown in **Figure 3d**, the number of O-glycoforms with 2 sialic acids largely exceeded the number of O-glycoforms with 2 fucoses, while the number of O-glycoforms with 1 or 3 sialic acids/fucoses were close.

Peptide sequence analysis of CSF O-glycopeptides

Unlike N-glycosylation, there is no consensus motif for O-glycosylation; whereas, some O-glycosylated peptide sequences might be preferred than others. Many preferred structure motifs have been proposed placing significant effects of the amino acids adjacent to the glycosite, including proline, alanine, serine, threonine, glycine, charged residue and amino acids with small side chains.[46-48] Among them, the effects of adjacent proline residues on the O-glycosylation biosynthesis is perhaps the most extensively studied. O'Connell et al[48] has shown the proline from −6 to +4 positions facilitates the process of glycosylating the proteins, agreeing well with another study that proposed proline residue from −4 to +4 positions are important for O-glycosylation.[47] More specifically, another study shows that the proline at -1 and +1 sites enhances O-

glycosylation on recombinant erythropoietin.[49] Using a statistical analysis based on 174 O-glycosites, Wilson et al[50] has shown a high frequency of occurrence at +3 and -1 positions relative to the glycosite. Thus, we performed a proline frequency analysis of the ±10 residues surrounding the 181 identified O-glycosites. Note that such proline frequency analysis results may differ depending on the origins of the O-glycosites dataset, and the proline frequency may vary from different tissues used. Despite of being tissue-specific, we did observe that the proline residue was sequence conserved at the -1 and +3 positions based on the 181 O-glycosites identified in CSF (**Figure 4a, c**), which agrees quite well with previous global statistical analysis[50] and report in CSF.[20] As increasing amount of new O-glycosites information were added to the known database each day, we also conducted the proline frequency analysis of the experimentally verified 435 O-glycosites from Uniprot (2017_12). The results revealed that proline showed the highest frequency from -4 to +7 positions, except at +1 position only next to threonine (**Figure 4b, d**). The conservative proline frequency was most pronounced at -1 and +3 positions in accordance with our findings. Such observation indicates that although a complete picture of conservative sequence may not be fully preserved in a specific tissue carrying a subset of global O-glycosites, the most pronounced conservative sequence was able to be preserved. These conservative sequence features that are unique to CSF may represent CSF-specific features affected by factors such as the origins of the glycoproteins, different biosynthesis routes of these glycoproteins, glycotransferase enzyme actives etc. One of these CSF-specific features is the higher frequency of glutamic acid (E) adjacent to the O-glycosite compared to

the results from global O-glycosites database, especially at +1 and +2 positions. Previous study has shown O-glycosylation was markedly reduced when glutamic acid residue was substituted at position -1 and +3, but glutamic acid replacement at +1 and +2 had no effect,[51] which may explain a higher glutamic acid frequency at +1 and +2 positions rather than other positions.

Selected examples of CSF O-glycoproteins

Selected identified O-glycoprotein examples were discussed below. In the previous CSF O-glycoproteome study, Halim et al. discussed the situation of O-glycosylation identified at Thr residues of the N-glycosylation Asn-X-Ser/Thr consensus sequence.[20] To make a comparison with their results, we will start out discussing these identified O-glycopeptides containing this special O-glycosylation site. As a first example, the GNLTGAPGQR peptide from endothelin B receptor-like protein 2 (ETBR2) was found to be O-glycosylated at 4th Thr residue, and the 2nd asparagine (Asn) underwent deamination **(Table 1)**. As enzyme PNGase F was used in our study to first get rid of N-glycans, any Asn residue with an N-glycan attached would go through deamidation. It should also be noted that chemical deamidation may also occur during the sample preparation process,[52] so further experimental evidence is needed before the deamidated Asn can be assigned to an N-glycosite. Sequence analysis showed that the Asn residue is in the Asn-X-Ser/Thr N-glycosylation consensus motif and it has been annotated as N-glycosylation site in Uniprot. In addition, in our previous CSF N-glycoproteome study (data in a separate manuscript), the 2nd Asn residue has been

confidently identified as an N-glycosite. Thus, it confirmed that the deamidation at 2nd Asn residue was truly due to the release of N-glycans by PNGase F. Such finding is in agreement with Halim *et al*. report.[20] In addition, our study was able to identify the intact O-glycan compositions as $HexNAc_{(2)}Hex_{(2)}$ and $HexNAc_{(2)}Hex_{(2)}Fuc_{(1)}NeuAc_{(1)}$. In the same protein ETBR2, they also identified O-glycopeptide V<u>S</u>GGAPLHLGR from ETBR2, which is different from the proposed signal sequence cleavage at 4th Gly. In our study, the same O-glycopeptide with $HexNAc_{(1)}Hex_{(1)}NeuAc_{(2)}$ attached was identified supporting the previous finding.[20] A second example was the O-glycopeptide LPTTVLNA<u>T</u>AK from protein YIPF3, with O-glycosylation at 9th Thr and deamidation at 7th Asn. Both previous reports and our N-glycoproteome study confirmed 7th Asn as a true N-glycosite, which is also in accordance with Halim et al. report.[20] Two sialylalted O-glycoforms $HexNAc_{(1)}Hex_{(1)}NeuAc_{(1)}$ and $HexNAc_{(1)}Hex_{(1)}NeuAc_{(2)}$ were identified. A third example of O-glycosylation within the N-glycosylation consensus motif that hasn't been reported before are the two O-glycopeptides ELPGVCNE<u>T</u>MMALWEECK and LANL<u>T</u>QGEDQYYLR from apolipoprotein J (also known as clusterin) **(Table 1).** The deamidation at Asn residue in the two peptide sequences, along with our previous N-glycoproteome study results and literature report,[19, 53, 54] confirmed their identity as N-glycosites. The reason that we were able to detect these two sites was related to the different O-glycoforms attached to the two sites on apolipoprotein J. In contrary to the sialylated O-glycoforms detected in the first two examples, the O-glycoforms detected here were nonsialyated forms $HexNAc_{(3)}Fuc_{(1)}$ and $HexNAc_{(6)}Hex_{(5)}Fuc_{(3)}$. The non-sialylated O-glycopeptide cannot be enriched by

the previous enrichment method based on sialic acid oxidation. Whereas, the boronic acid enrichment method used in our study was able to capture the non-sialylated O-glycoforms, which enables the detection of this nonsialylated O-glycopeptide.

As structural components of lipoprotein particles, apolipoproteins play important roles in maintaining their structure and regulating their metabolism and enzyme activities. Studies have shown that apolipoproteins are often modified by O-glycosylation, including apolipoprotein E, A-I, A-II and C-III.[55-57] In our study, four apolipoproteins, apolipoprotein D, E, A-I, and J were found to be O-glycosylated **(Table 1)**. Although N-glycosylation of apolipoprotein D (apoD) has been extensively studied at 65Asn and 98Asn and many potential O-glycosylation sites (8 serine residues and 10 threonine residues) existed in its peptide sequence, O-glycosylation of APOD has not been reported before. The apoD-derived tryptic peptides VLNQELRADGTVNQIEGEATPVNLTEPAK and ADGTVNQIEGEATPVNLTEPAK were found to be O-glycosylated at 86Thr, and deamidation was also found at 96Asn in accordance with the expected outcome of the PNGase F pre-treatment, indirectly supporting the assignment. A total of three O-glycoforms were detected at this site, and all of them were fucosylated ($HexNAc_{(2)}Hex_{(2)}Fuc_{(1)}$, $HexNAc_{(2)}Hex_{(2)}Fuc_{(1)}NeuAc_{(2)}$, $HexNAc_{(4)}Hex_{(4)}Fuc_{(3)}$). For apolipoprotein E (APOE), two O-glycosites at 36^{th} Thr and 212^{th} Thr were identified with 11 O-glycoforms and 2 O-glycoforms detected at each site respectively. Using lectin-based isolation method, Cubedo *et al*. reported the O-glycosylation of serum apolipoprotein A-I, but the exact O-glycosite or O-glycoform information was

lacking.[58] In our study, the O-glycopeptide EQLGPVTQEFWDNLEK from apolipoprotein A-I was identified, with 1 O-glycoform detected ($HexNAc_{(3)}Hex_{(2)}Fuc_{(1)}$). Despite the fact that APOJ has been shown carrying several PTMs such as N-glycosylation, ubiquitination and phosphorylation, O-glycosylation hasn't been reported before.[59-62] In our study, three O-glycosites 105Thr, 210Ser, 376Thr with 5 O-glycoforms were found. Only 210Ser carried a sialylated O-glycoform, and the rest of them carried non-sialylated O-glycoforms.

Comprehensive PTMs analysis of CSF endogenous peptides

Although deemed as a common PTM for endogenous peptides, glycosylation was not considered as a possible dynamic modification in most of current peptidomics workflow. This is largely because unlike other PTMs such as phosphorylation, glycosylation requires its own glycan database that contains hundreds of glycan compositions, which requires special software platforms with the ability to *de novo* glycopeptide spectra.[63] In addition, glycosylation is a quite labile PTM and can be easily lost in the most commonly used CID/HCD-based peptidomics workflow. As a result, to our best knowledge currently there is no peptidomics workflow that takes the glycosylation into account, along with other common PTMs. To solve this two problems, EThcD was utilized to preserve the labile glycosylation and the raw data will be analyzed by PTM-centric Byonic searching engine to enable site-specific glycoprofiling.[64] We first tried to search the raw data directly by setting all the possible PTMs (glycosylation, phosphorylation, acetylation etc.) and using the whole human protein and glycan

database, which leaded to computational explosion. Then a constructed "focused" CSF endogenous peptides database based on literature was used but also failed, indicating that the size of glycan database may be too large. Therefore, we developed a 3-step searching strategy (**Figure S1**) to facilitate the search and details could be found in "Experimental Procedures" section. The key idea is that focused protein and glycan database would be constructed by the first and second wild searches, and would be used to facilitate the final search, where all the possible PTMs (including N-glycosylation and O-glycosylation) could be searched all at once.

For endogenous peptides analysis in CSF, our lab previously developed an optimized 10kDa MWCO-based protocol to achieve an optimal recovery rate.[28, 29] Here, we compared 10kDa MWCO-based protocol developed in our lab with another 30kDa MWCO-based protocol reported previously in terms of their recovery rate.[65] As shown in **Figure 5 (Supplemental Table S3)**, 10k-based protocol outperformed 30k-based protocol in both the total number of identified endogenous peptides, peptides without PTMs and peptides with PTMs. Thus, the 10k-based protocol was used to separate CSF samples into peptide fraction and protein fraction (**Figure S2**). Peptide fraction would be directly injected for LC-MS/MS analysis after desalting, and the acquired data would be subject to search engine Byonic for comprehensive PTMs analysis.

In CSF samples from healthy individual, a total of 1492 endogenous peptides were identified (**Figure 6**), with an over 2-fold increase compared to previous reports.[21, 22] The large increase could possibly result from multiple factors including the use of instrument with higher sensitivity, optimized 10kDa MWCO protocol to increase the

recovery rate, as well as more comprehensive PTMs included. Benefiting from the hybrid fragmentation method EThcD utilized in our study, the labile PTMs including N-/O-glycosylation and phosphorylation could be well preserved, which otherwise could be lost in CID used by previous studies. In total, 370 endogenous peptides with PTMs were identified **(Figure 6**), indicating that endogenous peptides in CSF went through extensive modifications. Among them, 95 and 35 peptides were O-glycosylated and phosphorylated, respectively. It is worth noting that the present study is the first report that CSF endogenous peptides were extensively O-glycosylated, and in fact the number of O-glycosylated endogenous peptides exceeded all other PTMs modified peptides. These O-glycopeptides were derivatized from 32 protein precursors, and 28 O-glycan compositions were identified. Interestingly, only 1 N-glycosylated peptide TNSTFVQALVEHVK from precursor protein prosaposin was found with N-glycosite at previously reported 215th Asn residue.

Of all these identified endogenous peptides, neuropeptides as a subset is of special interest due to their large variety of functions as inter-cellular signaling molecules.[66] By referring to the human neuropeptide database NeuroPep,[67] 282 were identified as neuropeptides **(Figure 6)**. More than half of them (182, 65%) belonged to chromogranin/secretogranin family, with ProSAAS (28, 10%), NPY (19, 7%), VGF (18, 6%), opioid (11, 4%) and 7B2 (7, 2%) following behind (**Figure S3**). As mentioned previously, neuropeptides could undergo further post-translational modifications after initial peptidase cleavages. In the present study, 88 out of the 282 identified neuropeptides were modified by PTMs, including acetylation, amidation, O-

glycosylation, phosphorylation, Gln to pyro-Glu conversion. Among them, 15 neuropeptides were O-glycosylated, with 14 originating from ProSAAS and 1 from secretogranin-1. Broadly expressed in brain and neuroendocrine tissues, ProSAAS is the precursor of a number of neuropeptides.[68] The proSAAS-derived neuropeptides SAAS, PEN, and LEN, which are among the most abundant peptides present in mouse hypothalamus,[69-71] have been implicated in the regulation of food intake and body weight.[72, 73] In the present study, the two O-glycosylation sites (53Asn, 247Asn) detected were in the LEN and SAAS region, and 11 O-glycosylated LEN peptides and 3 O-glycosylated SAAS peptides were identified (**Supplemental Table S4**). Previous study of intact glycoprotein ProSAAS also identified these two sites, but the detailed O-glycoforms information were missing. At site 53Asn, two O-glycoforms with sialylated T-antigen were identified, whereas at site 247Asn three O-glycoforms with T-antigen in both sialylalted form and non-sialylated forms were identified, along with another O-glycoform with non-sialylated Tn-antigen. O-glycosylated neuropeptide EEKGHPQEESEESNV<u>S</u>MA derived from precursor protein secretogranin-1 was also identified with fucosylated O-glycan attached ($HexNAc_{(3)}Hex_{(3)}Fuc_{(1)}$).

CSF O-glycoproteome and endogenous peptidome mapping in MCI and AD

There have been a few studies that reported the N-glycosylation pattern alterations in some of the well-known AD-related proteins, including APP, tau, acetylcholinesterase, and transferrin.[74-77] Some other studies described global glycosylation-related alterations in the brain and CSF of AD patients.[78-83] Most of these global studies relied

on lectin-based staining techniques by recognizing specific glycan motif. Even fewer studies exist for O-glycosylation alteration study due to its larger degree of structure complexity and diversity. As a special case of O-glycosylation, O-GlcNAcylation has been shown to protect tau against aberrant phosphorylation and subsequent aggregation, suggesting O-GlcNAcase could be used as a potential therapeutic target in AD intervention.[77, 84, 85] However, the landscape O-glycosylation mapping in states along the AD progression hasn't been conducted yet. Here, we applied the developed site-specific O-glycoproteomic strategies into CSF samples from MCI and AD to get an idea of the overall picture of O-glycosylation in these disease states. Venn diagram analysis of O-glycoproteins, O-glycosites and O-glycopeptides from control, MCI and AD was shown in **Figure 7a (Supplemental Table S5)**. In order to highlight the more specific features of O-glycosylation pattern belonging to each stage, the percentage of different categories of glycoforms were compared. As discussed previously, the core 1 O-glycoform accounts for the most percentage of all kinds of O-glycoforms, so we compared the core 1 percentage across the three states, along with their sialylated and fucosylated forms. The percentage of global sialylated and fucosylated O-glycoforms were also compared, because altered sialylation and fucosylation were often implicated in various diseases.[86] However, quite comparable percentage of core 1, fucosylated core 1, sialylatd core 1 and global sialylated was obtained between these three states except for the fucosylated O-glycoforms, which showed a decrease trend in MCI and AD (**Figure 7b, Supplemental Table S5**). This is quite interesting because we also found alteration in fucosylation was the most dominant in another separate N-glycosylation

study (in a separate manuscript). Unfortunately, there is few study we can correlate, and further studies are needed to reveal the role of fucosylation in AD mechanism. Endogenous peptides analysis showed that there is a decrease in the number in MCI and AD, and a closer look into the percentage constitute showed shat the decrease is mainly the unmodified peptides (**Figure 8a, Supplemental Table S6**). In terms of the different kind of PTMs detected in control, MCI and AD, there is an gradual increase for O-glycosylation, Gln->pyro-Glu, acetylation, phosphorylation and a decrease for amidation, deamination. (**Figure 8b, Supplemental Table S6**). Such observations above represents a global landscape of the O-glycoproteome and endogenous peptidome from control, MCI and AD, revealing the dominant differences in different states. Further qualitative and quantitative analysis based on individual subject is needed to account for individual variation, pinpointing more specific changes of interesting glycoprotein/glycosite/glycoform.

Conclusion

In summary, a workflow that combined site-specific O-glycoproteome study and comprehensive PTMs analysis of endogenous peptides in CSF was developed. For O-glycoproteome study, an optimized boronic acid enrichment method combined with HpH fractionation was developed for enhanced site-specific O-glycopeptide analysis enabled by EThcD. After applying the optimized approach to CSF O-glycoproteome study, 308 intact O-glycopeptides from 182 O-glycosites and 110 O-glycoproteins were identified, representing the largest dataset for CSF site-specific O-glycoproteome study

so far. For endogenous peptides study, a peptidomics workflow that enables comprehensive PTMs analysis, including glycosylation, phosphorylation etc, of CSF endogenous peptides was developed, which was the first report that incorporates both N-glycosylation and O-glycosylation in the workflow. It was found that 95 CSF endogenous peptides were O-glycosylated, whereas few peptides were N-glycosylated. Among them, 15 O-glycosylated peptides were neuropeptides originating from ProSAAS. Besides, site-specific O-glycoproteome and endogenous peptides analysis were also conducted in CSF samples from MCI and AD patients, which yielded a glycosylation landscape picture to guide any following-up studies.

Acknowledgements

This research was supported in part by the National Institutes of Health (NIH) grants R21AG055377, R01 DK071801, and R56 MH110215. The Orbitrap instruments were purchased through the support of an NIH shared instrument grant (NIH-NCRR S10RR029531) and Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. LL acknowledges a Vilas Distinguished Achievement Professorship with funding provided by the Wisconsin Alumni Research Foundation and University of Wisconsin-Madison School of Pharmacy. HZ is a Wallenberg Academy Fellow. We thank Dr. Marshall Bern from Protein Metrics for providing access to Byonic software package.

References:

1. McComb, J. G., Recent research into the nature of cerebrospinal fluid formation and absorption. *J. Neurosurg.* **1983,** 59, (3), 369-383.
2. Sakka, L.; Coll, G.; Chazal, J., Anatomy and physiology of cerebrospinal fluid. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* **2011,** 128, (6), 309-316.
3. Stopa, E. G.; Berzin, T. M.; Kim, S.; Song, P.; Kuo-LeBlanc, V.; Rodriguez-Wolf, M.; Baird, A.; Johanson, C. E., Human choroid plexus growth factors: What are the implications for CSF dynamics in Alzheimer's disease? *Exp. Neurol.* **2001,** 167, (1), 40-47.
4. Chodobski, A.; Wojcik, B. E.; Loh, Y. P.; Dodd, K. A.; Szmydynger-Chodobska, J.; Johanson, C. E.; Demers, D. M.; Chun, Z. G.; Limthong, N. P., Vasopressin gene expression in rat choroid plexus. In *Vasopressin and Oxytocin*, Springer: 1998; pp 59-65.
5. Zhang, J., Proteomics of human cerebrospinal fluid–the good, the bad, and the ugly. *PROTEOMICS-Clinical Applications* **2007,** 1, (8), 805-819.
6. Fonteh, A. N.; Harrington, R. J.; Huhmer, A. F.; Biringer, R. G.; Riggins, J. N.; Harrington, M. G., Identification of disease markers in human cerebrospinal fluid using lipidomic and proteomic methods. *Dis. Markers* **2006,** 22, (1-2), 39-64.
7. Yuan, X.; Desiderio, D. M., Proteomics analysis of human cerebrospinal fluid. *J. Chromatogr. B* **2005,** 815, (1), 179-189.
8. Ohtsubo, K.; Marth, J. D., Glycosylation in cellular mechanisms of health and disease. *Cell* **2006,** 126, (5), 855-867.
9. Varki, A.; Cummings, R.; Esko, J.; Freeze, H.; Hart, G.; Marth, J., Essentials of glycobiology. 1999. *Cold Spring Harber Laboratory Press, New York* **1998**.
10. Tabak, L. A. In *The role of mucin-type O-glycans in eukaryotic development*, Semin. Cell Dev. Biol., 2010; Elsevier: 2010; pp 616-621.
11. Ten Hagen, K. G.; Fritz, T. A.; Tabak, L. A., All in the family: the UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferases. *Glycobiology* **2003,** 13, (1), 1R-16R.
12. Tabak, L. A., In defense of the oral cavity: structure, biosynthesis, and function of salivary mucins. *Annu. Rev. Physiol.* **1995,** 57, (1), 547-564.
13. Tran, D. T.; Ten Hagen, K. G., Mucin-type O-glycosylation during development. *J. Biol. Chem.* **2013,** 288, (10), 6921-6929.
14. Yang, X.; Qian, K., Protein O-GlcNAcylation: emerging mechanisms and functions. *Nature Reviews Molecular Cell Biology* **2017**.
15. Bennett, H., Glycosylation, phosphorylation, and sulfation of peptide hormones and their precursors. *Peptide Biosynthesis and Processing* **1991**, 111-140.
16. Bradbury, A.; Smyth, D., Modification of the N-and C-termini of bioactive peptides: amidation and acetylation. *Peptide Biosynthesis and Processing* **1991**, 231-250.
17. Thotakura, N. R.; Blithe, D. L., Glycoprotein hormones: glycobiology of gonadotrophins, thyrotrophin and free α subunit. *Glycobiology* **1995,** 5, (1), 3-10.
18. Yu, Q.; Canales, A.; Glover, M. S.; Das, R.; Shi, X.; Liu, Y.; Keller, M. P.; Attie, A. D.; Li, L., Targeted Mass Spectrometry Approach Enabled Discovery of O-Glycosylated Insulin and Related Signaling Peptides in Mouse and Human Pancreatic Islets. *Anal. Chem.* **2017,** 89, (17), 9184-9191.
19. Nilsson, J.; Rüetschi, U.; Halim, A.; Hesse, C.; Carlsohn, E.; Brinkmalm, G.;

Larson, G., Enrichment of glycopeptides for glycan structure and attachment site identification. *Nat. Methods* **2009,** 6, (11), 809-811.
20. Halim, A.; Rüetschi, U.; Larson, G. r.; Nilsson, J., LC–MS/MS characterization of O-glycosylation sites and glycan structures of human cerebrospinal fluid glycoproteins. *J. Proteome Res.* **2013,** 12, (2), 573-584.
21. Hölttä, M.; Zetterberg, H.; Mirgorodskaya, E.; Mattsson, N.; Blennow, K.; Gobom, J., Peptidome analysis of cerebrospinal fluid by LC-MALDI MS. *PLoS One* **2012,** 7, (8), e42555.
22. Zougman, A.; Pilch, B.; Podtelejnikov, A.; Kiehntopf, M.; Schnabel, C.; Kumar, C.; Mann, M., Integrated analysis of the cerebrospinal fluid peptidome and proteome. *J. Proteome Res.* **2007,** 7, (01), 386-399.
23. Zielinska, D. F.; Gnad, F.; Schropp, K.; Wiśniewski, J. R.; Mann, M., Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. *Mol. Cell* **2012,** 46, (4), 542-548.
24. Wang, X.; Yuan, Z.-F.; Fan, J.; Karch, K. R.; Ball, L. E.; Denu, J. M.; Garcia, B. A., A novel quantitative mass spectrometry platform for determining protein O-GlcNAcylation dynamics. *Mol. Cell. Proteomics* **2016,** 15, (7), 2462-2475.
25. Apweiler, R.; Hermjakob, H.; Sharon, N., On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1999,** 1473, (1), 4-8.
26. Zhang, Y.; Zhang, C.; Jiang, H.; Yang, P.; Lu, H., Fishing the PTM proteome with chemical approaches using functional solid phases. *Chem. Soc. Rev.* **2015,** 44, (22), 8260-8287.
27. Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nature methods* **2009,** 6, (5), 359-362.
28. Cunningham, R.; Wang, J.; Wellner, D.; Li, L., Investigation and reduction of sub‐microgram peptide loss using molecular weight cut‐off fractionation prior to mass spectrometric analysis. *J. Mass Spectrom.* **2012,** 47, (10), 1327-1332.
29. Wang, J.; Cunningham, R.; Zetterberg, H.; Asthana, S.; Carlsson, C.; Okonkwo, O.; Li, L., Label‐free quantitative comparison of cerebrospinal fluid glycoproteins and endogenous peptides in subjects with Alzheimer's disease, mild cognitive impairment, and healthy individuals. *PROTEOMICS-Clinical Applications* **2016,** 10, (12), 1225-1241.
30. Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E., WebLogo: a sequence logo generator. *Genome Res.* **2004,** 14, (6), 1188-1190.
31. Levery, S. B.; Steentoft, C.; Halim, A.; Narimatsu, Y.; Clausen, H.; Vakhrushev, S. Y., Advances in mass spectrometry driven O-glycoproteomics. *Biochimica et Biophysica Acta (BBA)-General Subjects* **2015,** 1850, (1), 33-42.
32. Sparbier, K.; Koch, S.; Kessler, I.; Wenzel, T.; Kostrzewa, M., Selective isolation of glycoproteins and glycopeptides for MALDI-TOF MS detection supported by magnetic particles. *Journal of biomolecular techniques: JBT* **2005,** 16, (4), 407.
33. Xu, Y.; Wu, Z.; Zhang, L.; Lu, H.; Yang, P.; Webley, P. A.; Zhao, D., Highly specific enrichment of glycopeptides using boronic acid-functionalized mesoporous silica. *Anal. Chem.* **2008,** 81, (1), 503-508.

34. Chen, W.; Smeekens, J. M.; Wu, R., A universal chemical enrichment method for mapping the yeast N-glycoproteome by mass spectrometry (MS). *Mol. Cell. Proteomics* **2014,** 13, (6), 1563-1572.
35. Batth, T. S.; Francavilla, C.; Olsen, J. V., Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J. Proteome Res.* **2014,** 13, (12), 6176-6186.
36. Liu, H.; Sadygov, R. G.; Yates, J. R., A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004,** 76, (14), 4193-4201.
37. Frese, C. K.; Altelaar, A. M.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S., Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **2012,** 84, (22), 9668-9673.
38. Frese, C. K.; Zhou, H.; Taus, T.; Altelaar, A. M.; Mechtler, K.; Heck, A. J.; Mohammed, S., Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (EThcD). *J. Proteome Res.* **2013,** 12, (3), 1520-1525.
39. Marino, F.; Bern, M.; Mommen, G. P.; Leney, A. C.; van Gaans-van den Brink, J. A.; Bonvin, A. M.; Becker, C.; van Els, C. c. A.; Heck, A. J., Extended O-GlcNAc on HLA class-I-bound peptides. *J. Am. Chem. Soc.* **2015,** 137, (34), 10922-10925.
40. Zhang, Y.; Xie, X.; Zhao, X.; Tian, F.; Lv, J.; Ying, W.; Qian, X., Systems analysis of singly and multiply O-glycosylated peptides in the human serum glycoproteome via EThcD and HCD mass spectrometry. *J. Proteomics* **2018,** 170, 14-27.
41. Glover, M. S.; Yu, Q.; Chen, Z.; Shi, X.; Kent, K. C.; Li, L., Characterization of intact sialylated glycopeptides and phosphorylated glycopeptides from IMAC enriched samples by EThcD fragmentation: Toward combining phosphoproteomics and glycoproteomics. *Int. J. Mass spectrom.* **2017**.
42. Yu, Q.; Wang, B.; Chen, Z.; Urabe, G.; Glover, M. S.; Shi, X.; Guo, L.-W.; Kent, K. C.; Li, L., Electron-transfer/higher-energy collision dissociation (EThcD)-enabled intact glycopeptide/glycoproteome characterization. *J. Am. Soc. Mass Spectrom.* **2017,** 28, (9), 1751-1764.
43. Mitoma, J.; Petryniak, B.; Hiraoka, N.; Yeh, J.-C.; Lowe, J. B.; Fukuda, M., Extended core 1 and core 2 branched O-glycans differentially modulate sialyl Lewis X-type L-selectin ligand activity. *J. Biol. Chem.* **2003,** 278, (11), 9953-9961.
44. Qin, H.; Cheng, K.; Zhu, J.; Mao, J.; Wang, F.; Dong, M.; Chen, R.; Guo, Z.; Liang, X.; Ye, M., Proteomics Analysis of O-GalNAc Glycosylation in Human Serum by an Integrated Strategy. *Anal. Chem.* **2017,** 89, (3), 1469-1476.
45. Darula, Z.; Sherman, J.; Medzihradszky, K. F., How to dig deeper? Improved enrichment methods for mucin core-1 type glycopeptides. *Mol. Cell. Proteomics* **2012,** 11, (7), O111. 016774.
46. Hansen, J. E.; Lund, O.; Engelbrecht, J.; Bohr, H.; Nielsen, J. O.; Hansen, J. E.; Brunak, S., Prediction of O-glycosylation of mammalian proteins: specificity patterns of UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase. *Biochem. J.* **1995,** 308, (3), 801-813.
47. Elhammer, A. P.; Poorman, R. A.; Brown, E.; Maggiora, L. L.; Hoogerheide, J.;

Kezdy, F., The specificity of UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase as inferred from a database of in vivo substrates and from the in vitro glycosylation of proteins and peptides. *J. Biol. Chem.* **1993,** 268, (14), 10029-10038.
48. O'Connell, B.; Hagen, F.; Tabak, L., The influence of flanking sequence on the O-glycosylation of threonine in vitro. *J. Biol. Chem.* **1992,** 267, (35), 25010-25018.
49. Elliott, S.; Bartley, T.; Delorme, E.; Derby, P.; Hunt, R.; Lorenzini, T.; Parker, V.; Rohde, M. F.; Stoney, K., Structural requirements for addition of O-linked carbohydrate to recombinant erythropoietin. *Biochemistry* **1994,** 33, (37), 11237-11245.
50. Wilson, I.; Gavel, Y.; Von Heijne, G., Amino acid distributions around O-linked glycosylation sites. *Biochem. J.* **1991,** 275, (2), 529-534.
51. Nehrke, K.; Hagen, F. K.; Tabak, L. A., Charge distribution of flanking amino acids influences O-glycan acquisition in vivo. *J. Biol. Chem.* **1996,** 271, (12), 7061-7065.
52. Palmisano, G.; Melo-Braga, M. N.; Engholm-Keller, K.; Parker, B. L.; Larsen, M. R., Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J. Proteome Res.* **2012,** 11, (3), 1949-1957.
53. Ramachandran, P.; Boontheung, P.; Xie, Y.; Sondej, M.; Wong, D. T.; Loo, J. A., Identification of N-linked glycoproteins in human saliva by glycoprotein capture and mass spectrometry. *J. Proteome Res.* **2006,** 5, (6), 1493-1503.
54. Liu, T.; Qian, W.-J.; Gritsenko, M. A.; Camp, D. G.; Monroe, M. E.; Moore, R. J.; Smith, R. D., Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry. *J. Proteome Res.* **2005,** 4, (6), 2070-2080.
55. Remaley, A.; Wong, A.; Schumacher, U.; Meng, M.; Brewer, H.; Hoeg, J., O-linked glycosylation modifies the association of apolipoprotein A-II to high density lipoproteins. *J. Biol. Chem.* **1993,** 268, (9), 6785-6790.
56. Beisiegel, U.; Weber, W.; Havinga, J. R.; Ihrke, G.; Hui, D. Y.; Wernette-Hammond, M. E.; Turck, C. W.; Innerarity, T. L.; Mahley, R. W., Apolipoprotein E-binding proteins isolated from dog and human liver. *Atertio. Thromb. Vasc. Biol.* **1988,** 8, (3), 288-297.
57. Wada, Y.; Kadoya, M.; Okamoto, N., Mass spectrometry of apolipoprotein C-III, a simple analytical method for mucin-type O-glycosylation and its application to an autosomal recessive cutis laxa type-2 (ARCL2) patient. *Glycobiology* **2012,** 22, (8), 1140-1144.
58. Cubedo, J.; Padró, T.; Badimon, L., Glycoproteome of human apolipoprotein AI: N-and O-glycosylated forms are increased in patients with acute myocardial infarction. *Translational Research* **2014,** 164, (3), 209-222.
59. Woody, S. K.; Zhao, L., Clusterin (APOJ) in Alzheimer's Disease: An Old Molecule with a New Role. In *Update on Dementia*, InTech: 2016.
60. Burkey, B.; Harmony, J., Intracellular processing of apolipoprotein J precursor to the mature heterodimer. *J. Lipid Res.* **1991,** 32, (6), 1039-1048.
61. Liang, H.-C.; Russell, C.; Mitra, V.; Chung, R.; Hye, A.; Bazenet, C.; Lovestone, S.; Pike, I.; Ward, M., Glycosylation of human plasma clusterin yields a novel candidate biomarker of Alzheimer's disease. *J. Proteome Res.* **2015,** 14, (12), 5063-5076.
62. Kapron, J. T.; Hilliard, G. M.; Lakins, J. N.; Tenniswood, M. P.; West, K. A.; Carr, S. A.; Crabb, J. W., Identification and characterization of glycosylation sites in human

serum clusterin. *Protein Sci.* **1997,** 6, (10), 2120-2133.
63. Dallas, D. C.; Guerrero, A.; Parker, E. A.; Robinson, R. C.; Gan, J.; German, J. B.; Barile, D.; Lebrilla, C. B., Current peptidomics: applications, purification, identification, quantification, and functional analysis. *Proteomics* **2015,** 15, (5-6), 1026-1038.
64. Bern, M.; Kil, Y. J.; Becker, C., Byonic: advanced peptide and protein identification software. *Current protocols in bioinformatics* **2012**, 13.20. 1-13.20. 14.
65. Secher, A.; Kelstrup, C. D.; Conde-Frieboes, K. W.; Pyke, C.; Raun, K.; Wulff, B. S.; Olsen, J. V., Analytic framework for peptidomics applied to large-scale neuropeptide identification. *Nature communications* **2016,** 7.
66. Sossin, W. S.; Fisher, J. M.; Scheller, R. H., Cellular and molecular biology of neuropeptide processing and packaging. *Neuron* **1989,** 2, (5), 1407-1417.
67. Wang, Y.; Wang, M.; Yin, S.; Jang, R.; Wang, J.; Xue, Z.; Xu, T., NeuroPep: a comprehensive resource of neuropeptides. *Database* **2015,** 2015.
68. Wardman, J. H.; Berezniuk, I.; Di, S.; Tasker, J. G.; Fricker, L. D., ProSAAS-derived peptides are colocalized with neuropeptide Y and function as neuropeptides in the regulation of food intake. *PLoS One* **2011,** 6, (12), e28152.
69. Fricker, L. D.; McKinzie, A. A.; Sun, J.; Curran, E.; Qian, Y.; Yan, L.; Patterson, S. D.; Courchesne, P. L.; Richards, B.; Levin, N., Identification and characterization of proSAAS, a granin-like neuroendocrine peptide precursor that inhibits prohormone processing. *J. Neurosci.* **2000,** 20, (2), 639-648.
70. Fricker, L. D. In *Neuropeptides and other bioactive peptides: from discovery to function*, Colloquium Series on Neuropeptides, 2012; Morgan & Claypool Life Sciences: 2012; pp 1-122.
71. Mzhavia, N.; Berman, Y.; Che, F.-Y.; Fricker, L. D.; Devi, L. A., ProSAAS processing in mouse brain and pituitary. *J. Biol. Chem.* **2001,** 276, (9), 6207-6213.
72. Gomes, I.; Bobeck, E. N.; Margolis, E. B.; Gupta, A.; Sierra, S.; Fakira, A. K.; Fujita, W.; Müller, T. D.; Müller, A.; Tschöp, M. H., Identification of GPR83 as the receptor for the neuroendocrine peptide PEN. *Science signaling* **2016,** 9, (425), ra43.
73. Gomes, I.; Aryal, D. K.; Wardman, J. H.; Gupta, A.; Gagnidze, K.; Rodriguiz, R. M.; Kumar, S.; Wetsel, W. C.; Pintar, J. E.; Fricker, L. D., GPR171 is a hypothalamic G protein-coupled receptor for BigLEN, a neuropeptide involved in feeding. *Proceedings of the National Academy of Sciences* **2013,** 110, (40), 16211-16216.
74. Van Rensburg, S.; Berman, P.; Potocnik, F.; Taljaard, J., Glycosylation of transferrin in Alzheimer's disease and alcohol-induced dementia. *Metab. Brain Dis.* **2000,** 15, (4), 243-247.
75. Sáez‐Valero, J.; Fodero, L.; Sjögren, M.; Andreasen, N.; Amici, S.; Gallai, V.; Vanderstichele, H.; Vanmechelen, E.; Parnetti, L.; Blennow, K., Glycosylation of acetylcholinesterase and butyrylcholinesterase changes as a function of the duration of Alzheimer's disease. *J. Neurosci. Res.* **2003,** 72, (4), 520-526.
76. Jacobsen, K. T.; Iverfeldt, K., O-GlcNAcylation increases non-amyloidogenic processing of the amyloid-β precursor protein (APP). *Biochem. Biophys. Res. Commun.* **2011,** 404, (3), 882-886.
77. Liu, F.; Iqbal, K.; Grundke-Iqbal, I.; Hart, G. W.; Gong, C.-X., O-GlcNAcylation

regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** 101, (29), 10804-10809.
78. Sihlbom, C.; Davidsson, P.; Sjögren, M.; Wahlund, L.-O.; Nilsson, C. L., Structural and quantitative comparison of cerebrospinal fluid glycoproteins in Alzheimer's disease patients and healthy individuals. *Neurochem. Res.* **2008,** 33, (7), 1332-1340.
79. Palmigiano, A.; Barone, R.; Sturiale, L.; Sanfilippo, C.; Bua, R. O.; Romeo, D. A.; Messina, A.; Capuana, M. L.; Maci, T.; Le Pira, F., CSF N-glycoproteomics for early diagnosis in Alzheimer's disease. *J. Proteomics* **2016,** 131, 29-37.
80. Zhu, Y.; Shan, X.; Yuzwa, S. A.; Vocadlo, D. J., The emerging link between O-GlcNAc and Alzheimer disease. *J. Biol. Chem.* **2014,** 289, (50), 34472-34481.
81. Schedin‐Weiss, S.; Winblad, B.; Tjernberg, L. O., The role of protein glycosylation in Alzheimer disease. *The FEBS journal* **2014,** 281, (1), 46-62.
82. Butterfield, D. A.; Owen, J. B., Lectin‐affinity chromatography brain glycoproteomics and Alzheimer disease: Insights into protein alterations consistent with the pathology and progression of this dementing disorder. *PROTEOMICS-Clinical Applications* **2011,** 5, (1‐2), 50-56.
83. Kanninen, K.; Goldsteins, G.; Auriola, S.; Alafuzoff, I.; Koistinaho, J., Glycosylation changes in Alzheimer's disease as revealed by a proteomic approach. *Neurosci. Lett.* **2004,** 367, (2), 235-240.
84. Yuzwa, S. A.; Shan, X.; Macauley, M. S.; Clark, T.; Skorobogatko, Y.; Vosseller, K.; Vocadlo, D. J., Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat. Chem. Biol.* **2012,** 8, (4), 393-399.
85. Li, X.; Lu, F.; Wang, J. Z.; Gong, C. X., Concurrent alterations of O‐GlcNAcylation and phosphorylation of tau in mouse brains during fasting. *Eur. J. Neurosci.* **2006,** 23, (8), 2078-2086.
86. Pinho, S. S.; Reis, C. A., Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer* **2015,** 15, (9), 540-555.

Figure Legends

Figure 1. Boronic acid enrichment optimization. a) Optimizing the number of fractions and technical replicates used. b) The effects of technical replicates on the overall coverage. c) The distribution of O-glycopeptides identified in different HpH fractions. (F: fractions, R: technical replicates)

Figure 2. EThcD spectra of O-glycopeptide from O-glycoproteome analysis (a), endogenous O-glycopeptide (b), and endogenous phosphorylated peptide (c).

Figure 3. a) The O-glycoproteome mapping in CSF proteins. b) The number of O-glycoforms with 1-3 sialic acids/fucoses. c) The distribution of the number of O-glycosites among O-glycoproteins. d) The distribution of the number of O-glycoforms among O-glycosites.

Figure 4. Weblogo probability plots and proline frequency analysis. a) Weblogo plot for 181 identified O-glycosites in the current study. b) Weblogo plot for 435 O-glycosites from Uniprot. c) Proline frequency analysis of 181 identified O-glycosites in the current study. d) Proline frequency analysis of 435 O-glycosites from Uniprot. ±10 amino acid residues surrounding the O-glycosite were analyzed.

Figure 5. The performance comparison between 10k-based protocol and 30k-based protocol.

Figure 6. The number of endogenous peptides and neuropeptides identified with/without different PTMs.

Figure 7. O-glycoproteome comparison between control, MCI and AD. a) Venn diagram analysis of total O-glycoproteome in control, MCI and AD. b) The comparison of different O-glycoforms in control, MCI and AD.

Figure 8. Endogenous peptidome comparison between control, MCI and AD. a) The number of endogenous peptides identified with/without PTMs. B) The comparison of the percentage of endogenous peptides with different PTMs identified in control, MCI and AD.

Table 1 Representative O-glycopeptides identified from human CSF glycoproteins.

Figure S1. The three-step searching strategy for comprehensive PTMs analysis of CSF endogenous peptides.

Figure S2. The overall workflow for O-glycoproteome and endogenous peptides analysis.

Figure S3. Different families of neuropeptides in CSF.

Supplemental Table S1 Subjects information from healthy control, MCI and AD group.

Supplemental Table S2 The O-glycopeptides identified using different enrichment strategies.

Supplemental Table S3 The endogenous peptides identified using 10k MWCO and 30k MWCO.

Supplemental Table S4 O-glycosylated neuropeptides identified in CSF.

Supplemental Table S5 The O-glycoproteome characterization in control, MCI and AD.

Supplemental Table S6 The endogenous peptides identified in control, MCI and AD.

Figure 1. Chen et al.

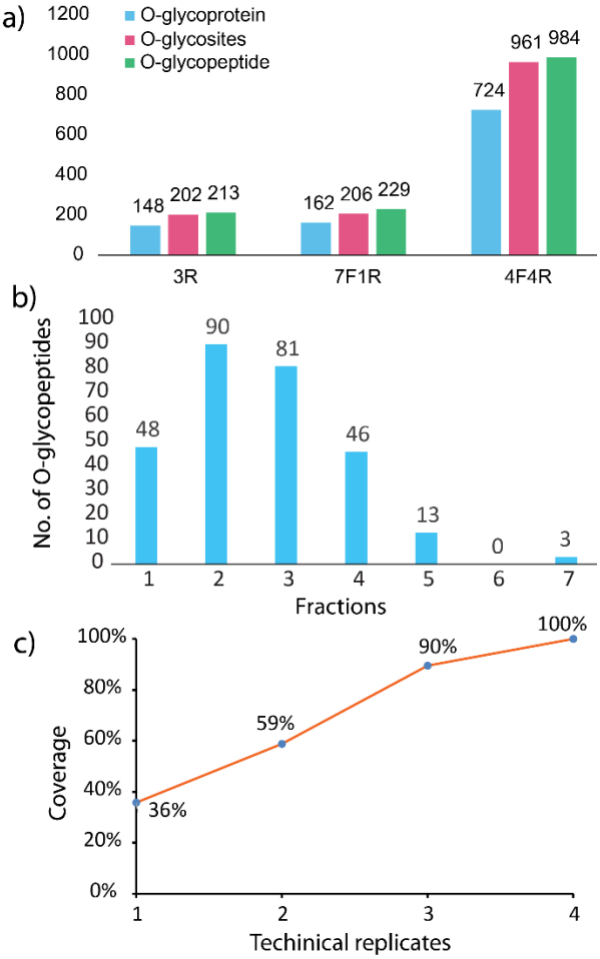


Figure 2. Chen et al.

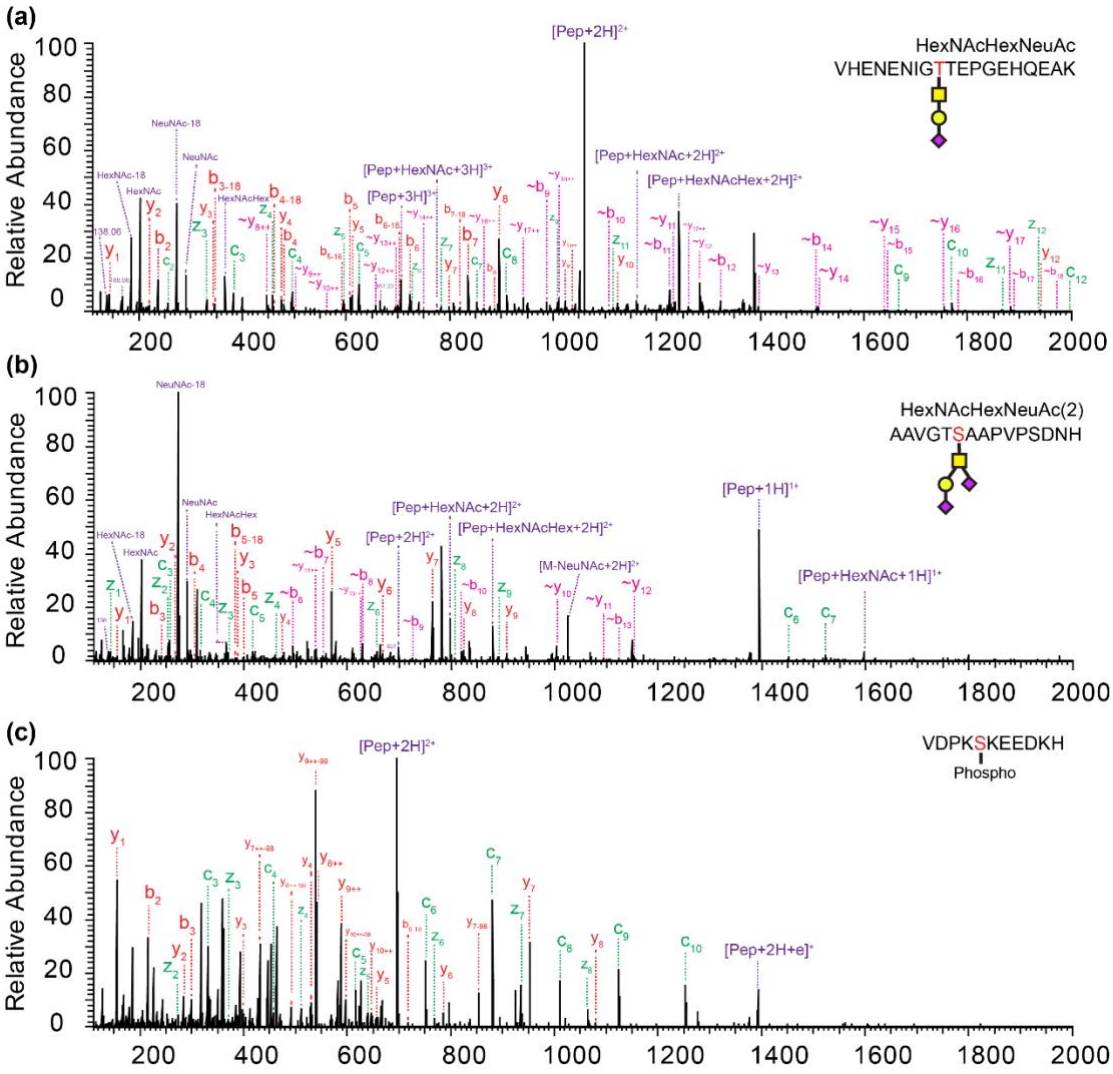


Figure 3. Chen et al.

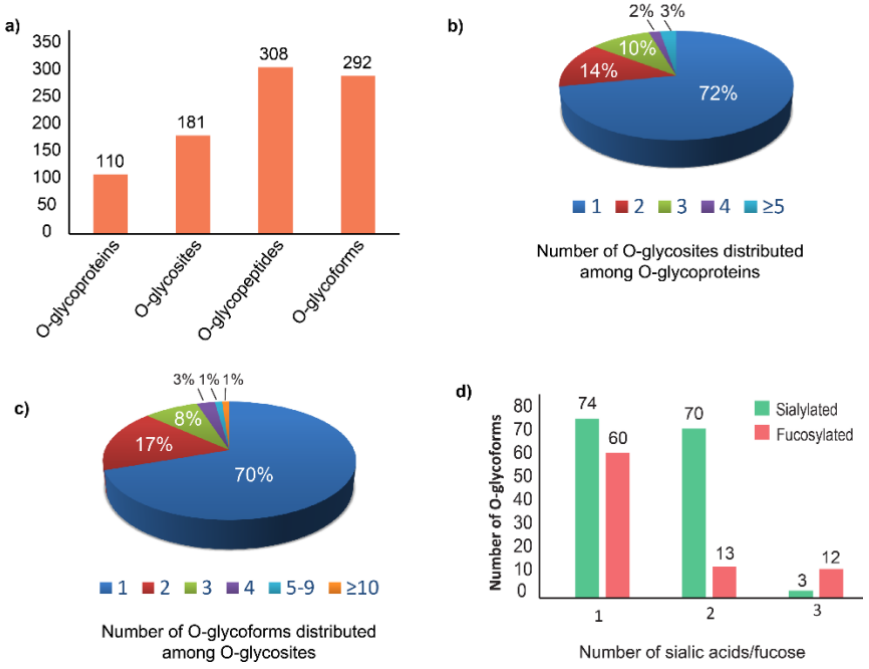


Figure 4. Chen et al.

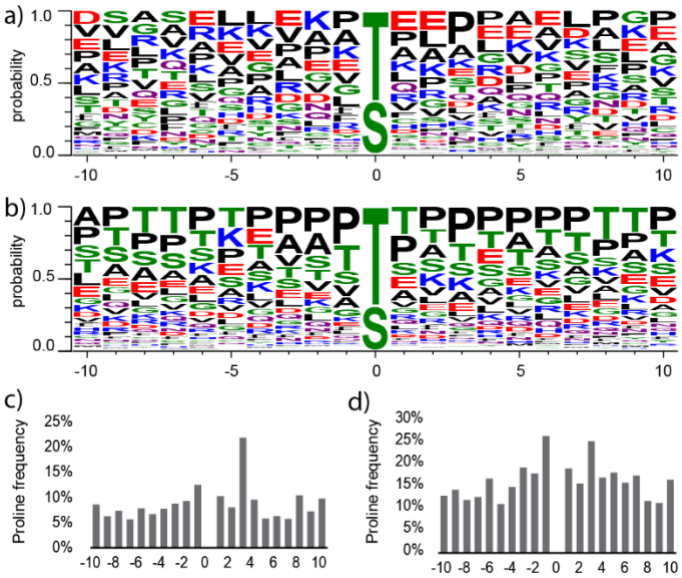


Figure 5. Chen et al.

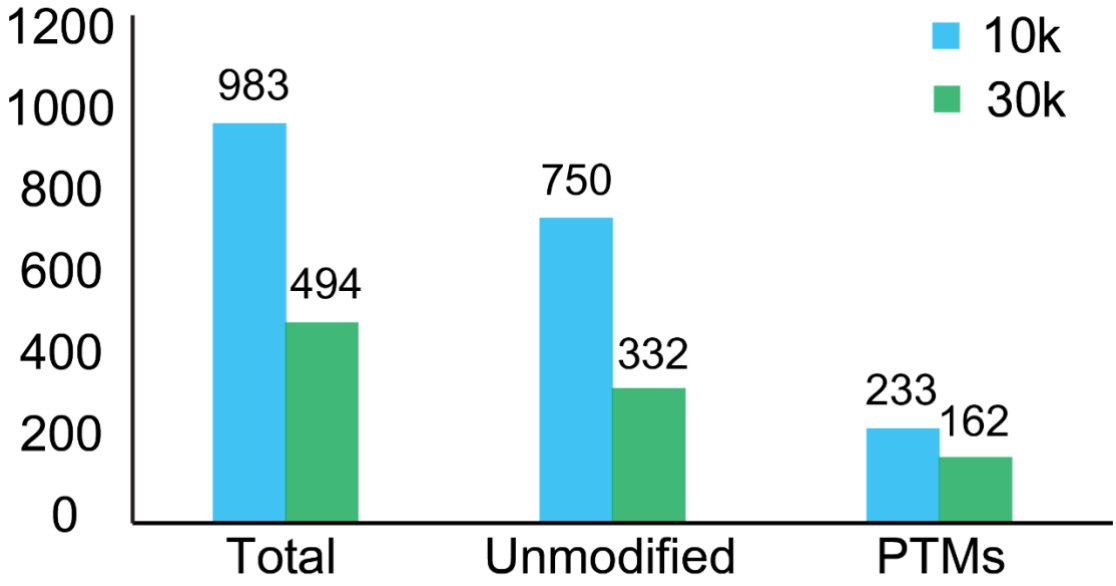


Figure 6. Chen et al.

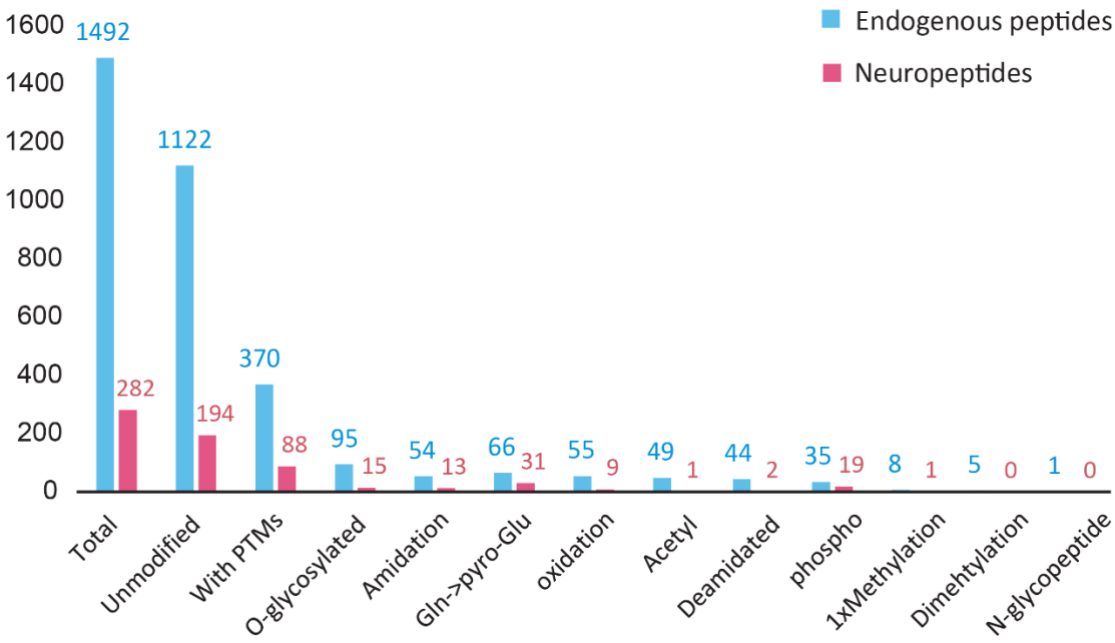


Figure 7. Chen et al.

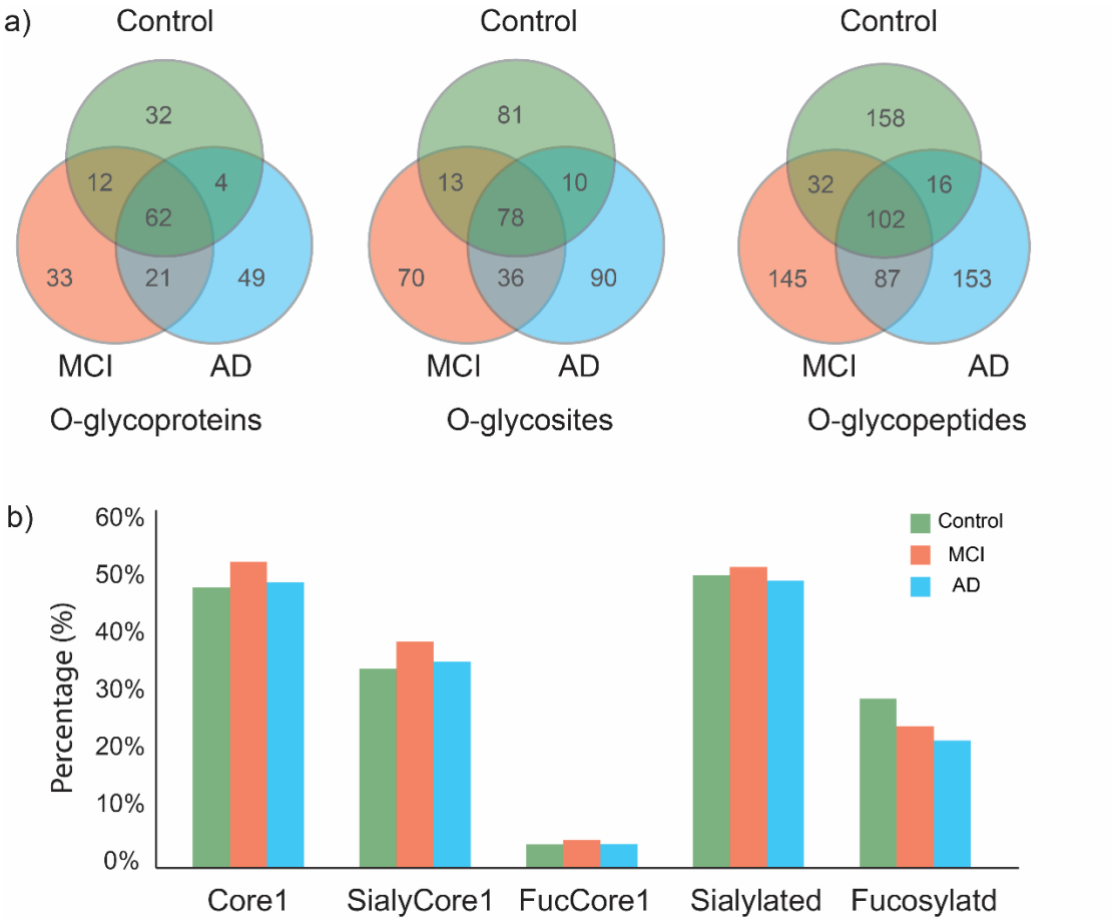


Figure 8. Chen et al.

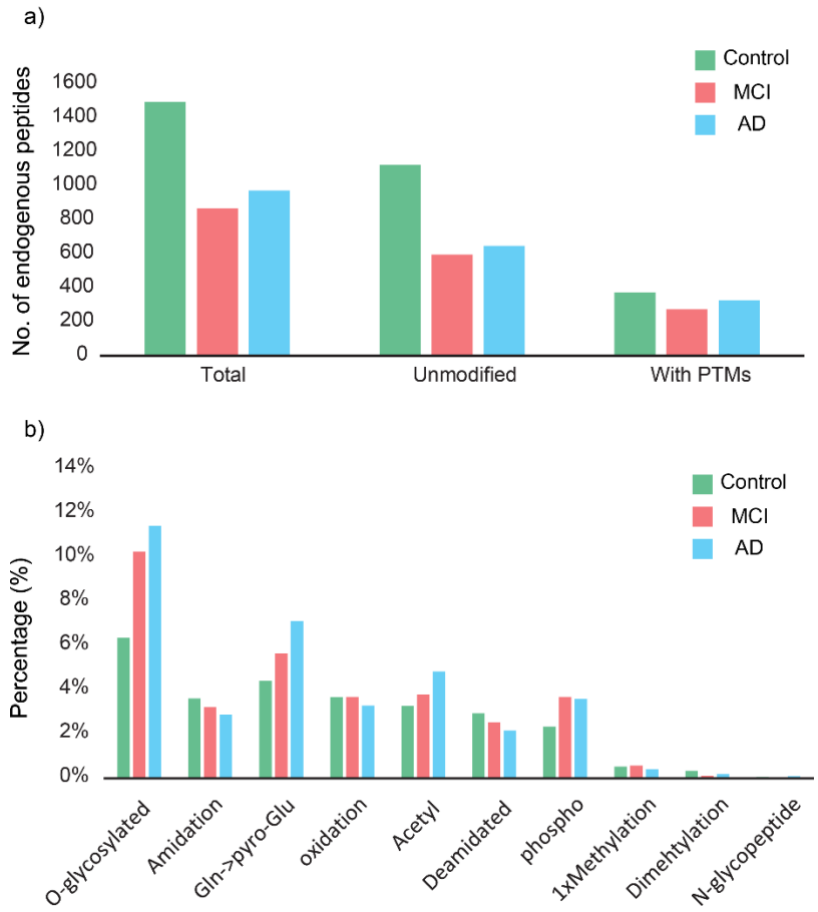


Table 1

O-glycoprotein ID	O-glycosite	O-glycopeptide sequence	O-glycan compositions
Prosaposin receptor GPR37L1	107	[R].GNLTGAPGQR.[L]	HexNAc(2)Hex(2)
	107	[R].GNLTGAPGQR.[L]	HexNAc(2)Hex(2)Fuc(1)NeuAc(1)
Protein YIPF3	339	[R].LPTTVLNATAK.[A]	HexNAc(1)Hex(1)NeuAc(1)
	339	[R].LPTTVLNATAK.[A]	HexNAc(1)Hex(1)NeuAc(2)
Apolipoprotein D	86	[K].VLNQELRADGTVNQIEGEATPVNLTEPAK.[L]	HexNAc(2)Hex(2)Fuc(1)
	86	[K].VLNQELRADGTVNQIEGEATPVNLTEPAK.[L]	HexNAc(2)Hex(2)Fuc(1)NeuAc(2)
	86	[R].ADGTVNQIEGEATPVNLTEPAK.[L]	HexNAc(4)Hex(4)Fuc(3)
Apolipoprotein E	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(1)Hex(1)NeuAc(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(1)Hex(1)NeuAc(2)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(2)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(2)Fuc(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(2)Fuc(1)NeuAc(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(2)NeuAc(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(2)Hex(2)NeuAc(2)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(3)Hex(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(3)Hex(1)Fuc(1)
	212	[R].AATVGSLAGQPLQER.[A]	HexNAc(3)Hex(1)Fuc(1)NeuAc(1)
	36	[R].QQTEWQSGQR.[W]	HexNAc(1)Hex(1)NeuAc(1)
	36	[R].QQTEWQSGQR.[W]	HexNAc(1)Hex(1)NeuAc(2)
Apolipoprotein A-I	92	[R].EQLGPVTQEFWDNLEK.[E]	HexNAc(3)Hex(2)Fuc(1)
Apolipoprotein J	105	[K].ELPGVCNETMMALWEECK.[P]	HexNAc(3)Fuc(1)
	210	[R].EPQDTYHYLPFSLPHR.[R]	HexNAc(1)Hex(1)
	210	[R].EPQDTYHYLPFSLPHR.[R]	HexNAc(1)Hex(1)NeuAc(2)
	376	[R].LANLTQGEDQYYLR.[V]	HexNAc(5)Hex(5)Fuc(3)
	376	[R].LANLTQGEDQYYLR.[V]	HexNAc(6)Hex(5)Fuc(3)

Figure S1. Chen et al.

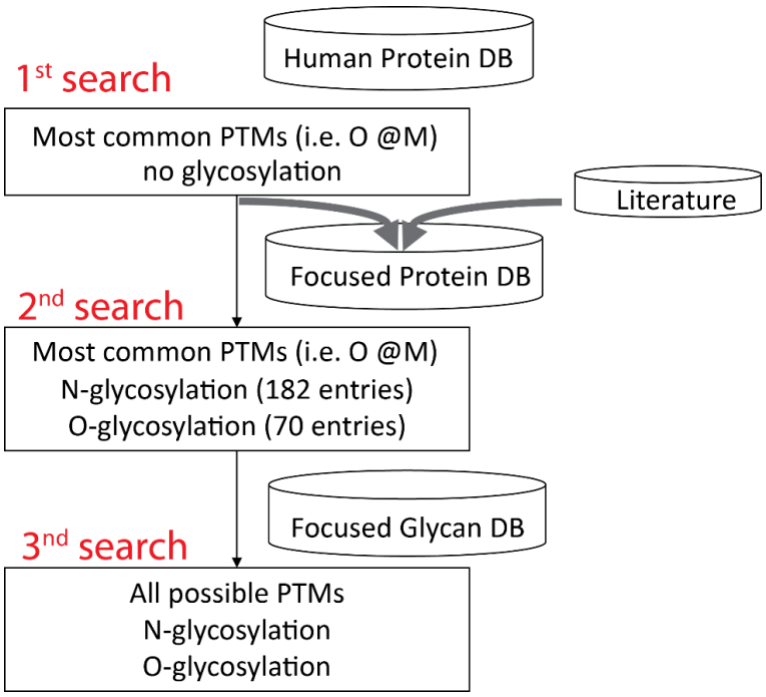


Figure S2. Chen et al.

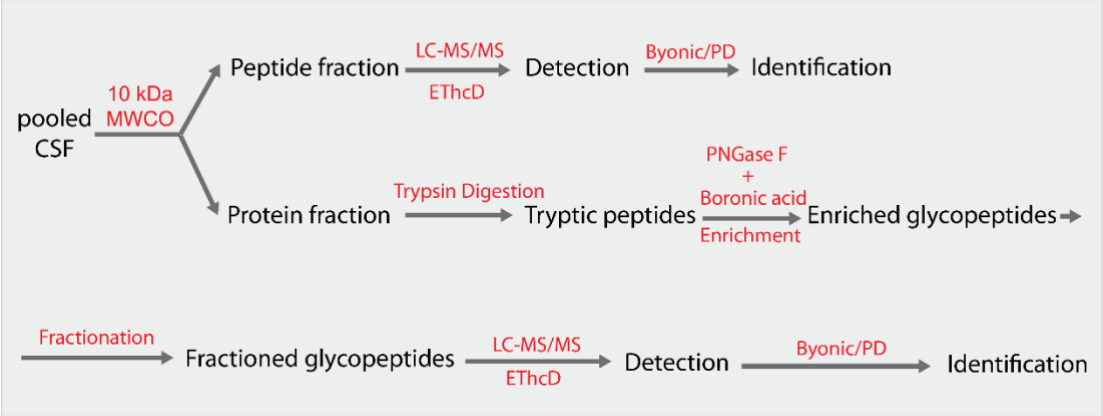
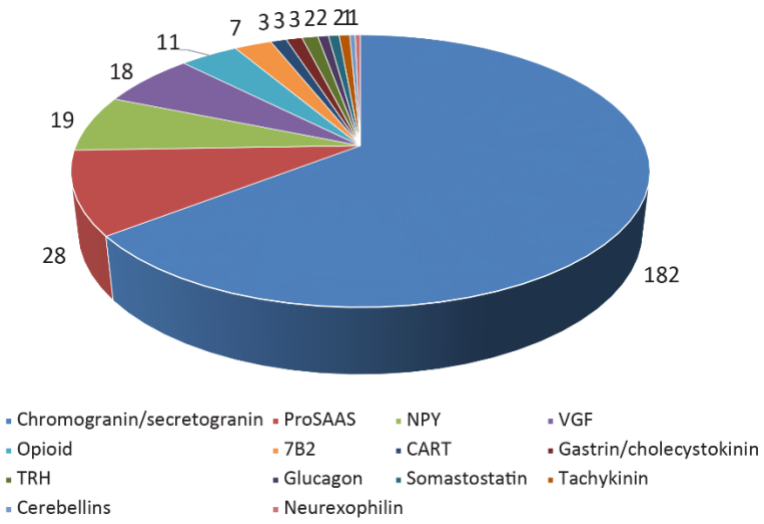


Figure S3. Chen et al.


182
28
19
18
11
7
3
3
3
2
2
2
1
1
Chromogranin/secretogranin
ProSAAS
NPY
VGF
Opioid
7B2
CART
Gastrin/cholecystokinin
TRH
Glucagon
Somastostatin
Tachykinin
Cerebellins
Neurexophilin