

# Using In Vivo Intact Structure for System-wide Quantitative Analysis of Changes in Proteins

John Yates III (✉ [jyates@scripps.edu](mailto:jyates@scripps.edu))

Scripps Research Institute <https://orcid.org/0000-0001-5267-1672>

Ahrum Son

The Scripps Research Institute

Hyunsoo Kim

Chungnam National University

Jolene Diedrich

Scripps Research Institute

Tom Bamberger

The Scripps Research Institute <https://orcid.org/0000-0002-3830-4486>

Daniel McClatchy

<https://orcid.org/0000-0002-0288-5645>

---

## Article

## Keywords:

**Posted Date:** July 3rd, 2023

**DOI:** <https://doi.org/10.21203/rs.3.rs-3118982/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

**Additional Declarations:** There is **NO** Competing Interest.

---

# Using *In Vivo* Intact Structure for System-wide Quantitative Analysis of Changes in Proteins

Ahrum Son<sup>1</sup>, Hyunsoo Kim<sup>1,2</sup>, Jolene K. Diedrich<sup>1</sup>, Casimir Bamberger<sup>1</sup>, Daniel B. McClatchy<sup>1</sup>, and John R. Yates III<sup>1\*</sup>

<sup>1</sup>Department of Molecular Medicine, The Scripps Research Institute, La Jolla, California 92037, United States

<sup>2</sup>Department of Convergent Bioscience and Informatics, Chungnam National University, 99 Daehak-ro, Yuseong-gu, Daejeon, 34134, Republic of Korea

## Abstract

Mass spectrometry-based methods can provide a global expression profile and structural readout of proteins in complex systems. Preserving the *in vivo* conformation of proteins in their innate state is challenging during proteomic experiments. Here, we introduce an approach using perfusion of reagents to create a whole animal *in vivo* protein footprinting method that adds dimethyl labels to exposed lysine residues on intact proteins to maintain information on protein conformations. When this approach was used to measure dynamic structural changes during Alzheimer's disease (AD) progression in a mouse model, we detected 433 proteins that underwent structural changes attributed to AD, independent of aging, across 7 tissues. We identified structural changes of co-expressed proteins and linked the communities of these proteins to their biological functions. Our findings show that structural alterations of proteins precede changes in expression, thereby showing the value of *in vivo* protein conformation measurement. Our method represents a new strategy for untangling mechanisms of proteostasis dysfunction caused by protein misfolding. *In vivo* whole-animal footprinting should have broad applicability for discovering conformational changes in systemic diseases and therapeutic interventions.

## 1 Introduction

The proper functioning of cellular machinery depends on the ability to maintain the functional structures of proteins. Proper folding of proteins is necessary to engage with partners in complexes and to perform catalytic activities. Protein folds or shapes can be measured by powerful, high-resolution *ex vivo* techniques such as X-ray crystallography, NMR, and Cryo-electron microscopy (Cryo-EM).<sup>1, 2, 3, 4</sup> Cryo-EM can be used to analyze large protein complexes if they are extracted from cells or are produced recombinantly prior to deposition on the grid and frozen.<sup>5, 6, 7</sup> Multiplexed Ion Beam imaging (MIBI) and ion beam tomography are capable of imaging cells and tissues, but they are not explicitly used to study the structure of proteins and protein complexes. Modeling algorithms can generate protein structures from *ex-vivo* protein cross-linking data, while *in vivo* cross-linking analyses generate protein-protein interaction data. Thus, because no methods are available to determine the high-resolution structures of proteins *in vivo*, we are still limited in our ability to elucidate the structures of proteins in the cellular milieu.

Protein “footprinting” methods were developed to probe the folding and interactions of proteins (such as epitope sites in antigens) using protease restriction or covalent labeling to identify exposed regions of proteins.<sup>8</sup> The data generated in protein footprinting experiments is often low resolution, but the potential scale of experiments has made it an attractive method. In 2010 West et al. showed proteome-scale footprinting in *S. cerevisiae* to determine off target binding of rapamycin.<sup>9</sup> A variety of protein labeling methods have been developed that provide low resolution *ex vivo* structural information about proteins.<sup>10, 11, 12, 13, 5</sup> Picotti and colleagues developed a limited proteolysis method to map ligand binding and protein folding in cell lysates and biofluids.<sup>14, 15, 5, 16</sup> In 2015 Espino et al. used lasers to activate hydroxy radicals *in vivo* to label proteins, providing the first attempt to footprint an intact cell.<sup>17</sup> Their approach has now been extended to the transparent worm *C. elegans*, which was chosen so the laser beam could penetrate the worm.<sup>18</sup>

Bamberger et al. developed Covalent Protein Painting (CPP), a chemical approach for quantitative protein footprinting to measure *in vivo* changes to protein conformations on a proteome scale.<sup>19</sup> In CPP formaldehyde, a chemical that rapidly permeates through cells and tissues, is used to label proteins by forming a Schiff’s base at solvent exposed lysine residues. These unstable intermediates are converted to dimethyl labels by reduction with cyanoborohydride. After lysis of cells or tissue, denaturation and digestion of proteins, a second labeling with a different “weight” reagent is performed to label inaccessible amino acid residues. By using heavy and light isotope versions of the reagents, a quantitative measure of lysine accessibility can be obtained. Using this method, Bamberger *et al.* probed the conformational changes of a proteome from postmortem brain tissue to reveal structural changes and altered protein-protein interactions in the brain tissue of AD patients.<sup>19</sup> In another study, Bamberger et al. measured the altered conformations of proteins in 60 cancer cell lines (NCI60).<sup>20</sup> Because the CPP protein labeling method begins with the widely used formaldehyde fixation step for *in vivo* dimethyl labeling it should be extensible to whole animal labeling to study models of disease.

Methods to measure alterations of protein conformations *in vivo* are needed to study diseases caused by protein misfolding that create loss or gain of function disruptions to biological processes, including Alzheimer’s disease, a common misfolding disease that is characterized by plaques of amyloid-beta and tangles of tau proteins. As observed by Bamberger et al., late-stage neurodegenerative diseases in humans are characterized by the misfolding of many additional proteins, suggesting that there is a generalized failing of proteostasis.<sup>19</sup> Techniques that allow *in vivo* measurement of protein folding would be a powerful tool for the study of these misfolding diseases.

Here, we used AD as a model to test our hypothesis that the global measurement of structural changes of proteins in tissues can be used to understand changes in their biological functionality during progression of protein misfolding diseases. We reasoned that it is important to capture proteins in their innate states to preserve the complex cellular milieu without the protein degradation that might occur during extraction and homogenization of organs. In this study, we introduce a method to probe the dynamic changes in protein structures *in vivo*, applying CPP for an animal model. This is the first *in vivo* study of structural changes of proteins in progressing AD on a proteome-wide scale in mouse tissue. We identified proteins whose structures were altered in co-expressed protein communities across 7 types of mouse tissue, which helps us understand the role of spatially altered proteins in various biological processes.

## **MATERIALS AND METHODS**

### **Animals/Tissue collection**

Female mice (APP<sup>(NL-F)</sup>)<sup>21</sup> were purchased from RIKEN Brain Science Institute and female C57BL/6 were obtained from The Scripps Research Institute breeding colony. Mice were housed in plastic cages located inside a temperature- and humidity-controlled animal colony and were maintained on a standard cycle (a 12 h day/night cycle). Animal facilities were AAALAC (Association for Assessment and Accreditation of Laboratory Animal Care) approved, and protocols were in accordance with the IACUC (Institutional Animal Care and Use Committee). Mice were sacrificed at 6, 9, 12, and 15 months of age.

### **First dimethyl-labeling and tissue collections**

Mice were anesthetized by inhalation of 1% isoflurane. Chests of the anesthetized mice were opened by cutting the ribcage. The left heart ventricle was punctured with a perfusion needle and a small cut was made in the right atrium to allow outflow of the perfusion solutions. Blood components were washed away with prewarmed pH 7.4 phosphate-buffered saline (PBS) for 10 min. The mice were perfused with 20 mL of fixation solution (1% CD<sub>2</sub>O) at a flow rate of 2.0 mL/min. Immediately afterward, 40 mL of the solution for the first light-dimethylation reaction (0.3 mM NaBH<sub>3</sub>CN, 1% CD<sub>2</sub>O in pH 7.4 PBS) was added at a flow rate of 2.0 mL/min. Organs were quickly excised and cut into 50 mg of tissue blocks. The tissue blocks were incubated in the same labeling solution (0.3 mM NaBH<sub>3</sub>CN, 1% CD<sub>2</sub>O in pH 7.4 PBS) for 10 min, and then the reaction was quenched by immersing the tissue blocks in 50 mM ammonium bicarbonate (ABC) solution for 5 min.

### **Tissue homogenization and protein extraction**

Tissue blocks were placed in 100  $\mu$ L of 20 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) pH 7.4 and were homogenized with a pestle until no chunks were visible. The tissue samples were sonicated for 10 cycles (pulse-on 5 sec, pulse-off 3 sec, amplify 30%) and then homogenates were clarified by centrifugation at 8,000 g at 4 °C for 30 min. Protein precipitation was performed by adding 400  $\mu$ L of 100% methanol, 100  $\mu$ L of 100% chloroform and 300  $\mu$ L of water to the sample. After vortexing vigorously, the samples were centrifuged at 8,000 g at 4°C for 30 min. The large aqueous layer was discarded. The samples were washed by adding 800  $\mu$ L of 100% methanol and vortexing vigorously. After centrifugation at 8,000 g at 4 °C for 30 min, the supernatant was removed. The methanol washing step was repeated 3 times. Methanol was removed and the pellet was air-dried. The pellet was dissolved in 100  $\mu$ L of 1% sodium deoxycholate (SDC) in 20 mM HEPES pH 7.4. The protein concentration was determined with a BCA protein Assay kit following the instructions from the vendor (23225, Thermo Scientific).

### **Proteolysis of labeled proteins with chymotrypsin**

Aliquots of tissue samples that contained 200  $\mu$ g of proteins were adjusted to 80  $\mu$ L with 1% SDC in 20 mM HEPES pH 7.4. The proteins were reduced with 10 mM TCEP (Tris(2-carboxyethyl)phosphine hydrochloride) and 1% SDC in 20 mM HEPES pH 7.4 at 60°C for 60 min on a shaker. Reduced proteins were alkylated with 20 mM IAA (iodoacetamide) for 30 min at 25°C in the dark. Denatured proteins were digested with chymotrypsin (Promega) at 1:100 (enzyme:substrate(w:w)) at 37°C for 16 hr. Samples were acidified with formic acid to a final concentration of 1%. The sample was centrifuged at 8,000 g at 4°C for 30 min and the supernatant was transferred to a new tube. The sample was centrifuged again at 8,000 g at 4°C for 30 min to collect the clean sample and the supernatant was transferred to a new tube.

### **Second dimethyl-labeling and desalting**

Pierce C18 spin tips (87784, ThermoFisher) were used for the second dimethyl-labeling step and desalting. A multipipette and a 96-well plate were used to prepare multiple samples in one batch. The C18 tips were activated by aspirating and dispensing 100  $\mu$ L each of 100% methanol and 100% acetonitrile (ACN). After the C18 tips were washed with 100  $\mu$ L of 0.1% formic acid, the samples were loaded onto the C18 tips by aspiration. To clean the samples bound to C18 tips, 100  $\mu$ L of 0.1% formic acid was aspirated and dispensed, and the pH was adjusted by aspirating 20 mM HEPES pH 7.4. Peptides bound to the C18 tips were dimethyl-labeled by aspirating 1% formaldehyde ( $^{13}\text{CD}_3\text{O}$ ), 0.3 mM Sodium cyanoborodeuteride ( $\text{NaBD}_3\text{CN}$ ) and the saturated tips were incubated for 15 min at 25°C. The reaction was quenched by aspirating 50 mM ABC and incubating for 10 min at 25°C. After washing C18 tips with 0.1% formic acid, the labeled peptides were eluted with 100  $\mu$ L of 40% ACN in 0.1% formic acid followed by 100  $\mu$ L of 60% ACN in 0.1% formic acid. The eluted peptides were lyophilized.

### **Strong cation exchange (SCX) fractionation of peptides**

SCX fractionation was conducted with commercial spin columns (90008, ThermoFisher Scientific). The pH of the sample was reduced by adding 800  $\mu$ L of 30% ACN in 0.1% formic acid. The spin column was equilibrated by adding 400  $\mu$ L of 30% ACN in 0.1% formic acid. It was then centrifuged at 1,000 g for 5 min, and the flow-through solution was discarded. The sample was loaded on the spin column and was centrifuged at 1,000 g for 3 min. Flow-through was stored for LC-MS/MS analysis. The peptides were eluted with consecutive 200  $\mu$ L aliquots of elution buffer containing 10 mM, 30 mM, 50 mM, 70 mM 100 mM, 150 mM and 300 mM of ammonium acetate. All elution buffer aliquots contained 0.1% formic acid and 30% ACN.

### **LC-MS/MS analysis**

Samples were loaded onto EvoTips following the manufacturer's protocol. The samples were run on an Evosep One (Evosep) coupled to a timsTOF Pro (Bruker Daltonics). Samples were separated on BEH 1.7  $\mu$ m C18 beads (Waters) packed in a 15 cm  $\times$  150  $\mu$ m inner diameter column with an integrated tip (pulled in-house) using the 30 SPD (samples per day) method. Mobile phases A and B were 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. MS data was acquired in PASEF mode with one MS1 survey TIMS-MS and PASEF MS/MS scans acquired per 1.1 s acquisition cycle. Ion accumulation and ramp time in the dual TIMS analyzer was set to 100 ms each and the ion mobility range spanned from  $1/K_0 = 0.6 \text{ Vs/cm}^2$  to  $1.6 \text{ Vs/cm}^2$ . Precursor ions for MS/MS analysis were isolated with a 2 Th window for  $m/z < 700$  and 3 Th for  $m/z > 700$  with a total  $m/z$  range from 100 to 1700. The collision energy was lowered linearly as a function of increasing mobility starting from 59 eV at  $1/K_0 = 1.6 \text{ VS/cm}^2$  to 20 eV at  $1/K_0 = 0.6 \text{ Vs/cm}^2$ . Singly charged precursor ions were excluded with a polygon filter, and precursors for MS/MS were picked at an intensity threshold of 2,500, target value of 20,000 and with an active exclusion of 24 s.

### **Peptide identification and quantification**

Raw files were searched against mouse proteins from Swiss-Prot-Uniprot database (retrieved 03/13/2022, 51,076 entries) containing canonical and isoform sequences, using MSFragger (version 17.1) in the FragPipe pipeline with mass calibration and parameter optimization enabled.<sup>22</sup> Philosopher was used to filter all peptide-spectrum matches. Quantification analysis was performed with IonQuant. The parameter setting of chymotrypsin allowed for two missed cleavage sites and the minimal required peptide length was set to six amino acids. Dimethyl peptide pairs were identified using variable modifications of light ( $\Delta \text{mass}$ : 32.0564) and heavy labeling ( $\Delta \text{mass}$ : 36.0757) of lysine, oxidation of methionine ( $\Delta \text{mass}$ : 15.9949), fixed modification of heavy dimethylation ( $\Delta \text{mass}$ : 36.0757) on N-terminus and the carbamidomethylation of cysteine ( $\Delta \text{mass}$  57.0214). Precursor tolerance was set to 50 ppm and fragment tolerance was set to 50 ppm. Isotope error was set to 0/1/2. The minimum number of fragment peaks required to include a PSM (peptide-spectrum match) in modeling was set to two, and the minimum number required to report the

match was four. The top 150 most intense peaks were considered, and a minimum of 15 fragment peaks were required to search a spectrum. The data were also searched against a decoy database and protein identifications were accepted at 1% peptide false discovery rate (FDR). All identified peptides were heavy-dimethylated on N-terminus.

### **Determination of accessibility of lysine sites**

Each peptide with a lysine site should be either light- or heavy-dimethylated, depending on the accessibility of lysine site. The difference in intensity of the peptides labeled in the first isobaric labeling step versus the second yields a relative abundance ratio  $R$ .<sup>23</sup> The  $R$  value represents the proportion of the peptide in which a specific lysine site was accessible for dimethylation and is independent of the overall protein amount in the sample.<sup>19</sup> The relative accessibility of a lysine residue for dimethylation is assessed by the value of accessibility; Accessibility (%) =  $R/(1+R) \times 100$ .

### **k-Nearest Neighbor (kNN) Imputation**

Missing values were imputed by kNN machine learning method using the VIM package in R.<sup>24, 25</sup> The kNN method assumes a relationship between spot volume patterns of groups of proteins. The kNN method impute missing values by selecting spots with spot volume patterns similar to the spot of interest.<sup>26</sup> A weighted average of values from the  $k$  most similar spots is used as an estimate for the missing value. The contribution of each spot is weighted by its similarity determined as the Euclidean distance. The optimum number of  $k$ -neighbors must be determined empirically.

### **Weighted correlation network analysis (WGCNA)**

A weighted protein co-expression network was built using the value of protein abundance from blockwiseModules WGCNA function.<sup>27</sup> Construction of weighted gene co-expression networks was conducted independently for each of 7 tissue datasets. The soft thresholding powers were determined with the R function pickSoftThreshold.<sup>28</sup> To pick an appropriate soft-thresholding power for network construction, the value of power was raised to 50. The chosen values were the smallest threshold that resulted in a scale-free  $R^2$  fit of 0.75 and the networks were created by calculating the component-wise minimum values for topologic overlap. Soft threshold powers varied across seven tissues as follows: 22 for brain, 16 for heart, kidney, 12 for liver, muscle and spleen, and 26 for thymus. BlockwiseModule function was run with the following parameters: TOMType = "signed", maxBlockSize = 5000, mergeCutHeight = 0.1, verbose = 3. Module eigenproteins (MEs) were calculated the correlation between the traits of AD (AD vs. NC). Multiple comparisons were accounted for by FDR correction across modules, and the  $P$ -values for the modules were reported.

### **Statistical analysis**

Differentially expressed proteins and altered accessibilities (%) between pairs of different age groups (6, 9, 12, and 15 months) in AD or between different pathological conditions (AD and NC) per age were found using Mann-Whitney tests independently. Kruskal-Wallis was used to simultaneously compare the accessibilities among four age groups (6, 9, 12, and 15 months) or among three age groups (6, 9, and 12 months or 9, 12, and 15 months) in AD. These comparisons were tested with Kruskal-Wallis followed by Bonferroni's comparison *post hoc* test independently. The criterion for significance was a  $P$ -value less than 0.05.

### **Enrichment of Gene ontology (GO) and protein-protein interaction analysis**

GO analysis was performed with ClueGO (a plug-in in Cytoscape) to identify the significant biological functions of the proteins in the WGCNA module.<sup>29</sup> Protein-protein networks were detected using Metascape and the following databases: STRING, BioGrid, InWeb\_IM, OmniPath.<sup>30-32</sup> The resultant network contained the subset of proteins that form physical interactions with at least one other list member, the confidence cutoff of physical interaction was set to medium (0.5) or strong (0.7). Visualization of the protein-protein interaction network was performed on the Cytoscape combining STRING.

## **Complex modeling with AlphaFold2-multimer**

We used AlphaFold2-multimer to predict the protein-protein interaction motif of each complex. AlphaFold2-multimer modeling was performed with ColabFold.<sup>33</sup> Input multiple sequence alignment (MSA) features were generated by local ColabFold using the “MMseqs2 (Uniref + Environmental)” MSA mode. By default, the constructed MSAs contain both unpaired (per-chain) and paired sequences. AlphaFold2-multimer was run with one or several options from the following list: model type = alphafold2\_multimer v3, num recycles = 3, recycle early stop tolerance = 0.5, max msa = auto, num seeds = 1. The models were ranked by confidence score, and rank 1 was selected as the most accurate model. The distance between two lysine residues was calculated using PyMOL2 version 2.5 (Schrödinger, LLC).



## Results

### A comprehensive structural information readout platform for proteome across mouse organs

Our primary goal was to reveal how *in vivo* alterations of protein conformations across tissues were associated with physiological disturbances that characterize Alzheimer's disease (AD). To prevent protein degradation in response to external stimuli, it was necessary to capture the innate folding state *in vivo*. Here, we developed a method to *in vivo* dimethyl label the body-wide proteome of an intact mouse prior to organ harvest. Our method involved sequentially diffusing labeling reagents through blood vessels throughout whole body so that proteins surfaces were light-dimethylated [(CHD<sub>2</sub>)<sub>2</sub>] *in vivo*. The process is fast enough to capture protein structures in a nearly innate state. After harvest and homogenization of each tissue and lysis of the cells, proteins were denatured and proteolyzed with chymotrypsin followed by labeling of the newly exposed lysine sites with heavy-dimethyl [(<sup>13</sup>CD<sub>3</sub>)<sub>2</sub>] tags (Figure 1A). The accessibility of each lysine site was determined from the ratio of the intensity of light-labeled peptide vs. the sum of intensities of the light- and heavy-labeled peptides. We systematically investigated the structural changes in the proteomes of 7 tissues in an AD mouse model ranging in age from 6 to 15 months, as well as in normal control (NC) mice to exclude the effects of aging.

A total of 43,014 dimethyl-labeled peptides that mapped to 5,217 proteins across all tissues were identified at a peptide false discovery rate (FDR) of < 1% (Figure 1B and 1C). Among the labeled peptides, 1,219 labeled peptides that mapped to 498 proteins were identified in all 7 tissues, whereas 24,026 labeled peptides that mapped to 4,952 proteins were tissue-specific (Figure 1B). Likewise, the highest proportion of labeled proteins were tissue-specific (37.3% (n = 1,947)), whereas labeled proteins that were identified in all 7 tissues comprised 11.1% (n = 578) of all labeled proteins (Figure 1C).

We next sought to examine the reproducibility of the *in vivo* CPP method for each mouse tissue. We evaluated the correlations of the accessibility between biological replicates for each age, then averaged these correlations. We found the highest R values for NC (0.842) and AD (0.782) in brain, whereas thymus showed the lowest correlation, 0.549 for NC and 0.505 for AD (Figure 1D-G, Table S1). Correlations averaged across all ages in all 7 tissues showed a correlation in accessibility, with R values of 0.668 and 0.649 for NC and AD, respectively. Because we observed strong correlations of the biological replicates regardless of anatomical source or pathological conditions of the tissue, we concluded that perfusion-based dimethyl-labeling was reliable. Also, we noticed that correlations of accessibility were higher within the same tissues, even in the presence of different pathological conditions, compared to those observed between different tissues at the same age.

### Dimethylation pattern encoding tissue-specific protein conformation

With an aim of identifying the tissue-specific perturbations caused by disease, we examined to what extent protein conformation is driven by tissue by comparing which proteins were dimethylated across tissues. Of a total of 23,039 peptides detected in the brain, 39.0% of peptides (n = 8,974) harbored lysine sites, and 8,758 of the 8,974 peptides were labeled with light- or heavy-dimethyl modifications (or both) (Figure 2A). Of a total of 4,295 detected proteins in brain, 2,565 (59.7%) proteins were modified with dimethyl. Among the 8,974 lysine-containing peptides, 8,368 were light-dimethyl labeled through perfusion, indicating a high labeling efficiency (93.2%), while 606 peptides (6.8%) were not labeled with light-dimethyl tags (Figure 2B). Among the 606 peptides that were not light-dimethylated, 390 were solely heavy-dimethylated and 216 were not labeled with either light- or heavy-dimethyl tags. Heart showed the highest efficiency of light-dimethylation, 10,745 peptides (97.5%) of 11,018 K-containing peptides. We presume that the high light-labeling efficiency of heart is due to the extended time that reagent remained in heart before being circulated to other organs. Labeling efficiency was high and consistent across all organs, ranging from 91.5% in muscle to 97.5% in heart, with a standard deviation for all organs of 2.3%. The average of the first labeling efficiency in all 7 tissues was 94.8%. Thus, we confirmed that the first labeling via perfusion was efficient, which provided a valid basis for subsequent comparisons cross tissues.

We then plotted the dimethylation pattern in each tissue. The liver had the second highest number of identified labeled peptides (n = 14,962) but the highest number of tissue-specific labeled peptides (n = 4,321), accounting for 18% of all 24,026 tissue-specific labeled peptides (Figure 2C and 2D). The kidney had the lowest number (11.1%, n = 2,664) of labeled tissue-specific peptides but ranked third lowest in the number of labeled tissue-specific proteins (11.1%, n = 216) (Figure 2F). Muscle-specific proteins represented a small proportion (2.7%, n = 53) of tissue-specific proteins out of a total of 1,947 (Figure 2E

and 2F). However, the presence of highly abundant proteins in muscle may have affected the identification of lower abundance proteins. The top 100 muscle proteins accounted for 78.6% of all identified muscle proteins, while the top 100 proteins in the kidney and heart accounted for 37.8% and 57.5% of the identified proteins, respectively (Figure S1A). Myosin regulatory light chain 11 (MyIpf), the most abundant protein in muscle, contributed 11.3% of the total muscle protein mass, and the top 12 proteins in muscle represented more than half of the mass of the top 100 proteins (Figure S1B). In conclusion, the labeling data provides tissue-specific protein information, which is valuable for understanding the physiological changes associated with disease in each tissue.

### Variability of the conformational changes among 7 tissues

We used labeled peptides that were detected in all seven tissues to quantitatively measure the structural differences in proteins. To minimize the effect of tissue-biased accessibility, we used quantile-normalized values for comparison across tissues (Figure S2). We sought to identify patterns of changes in accessibility that occur specifically for AD in proteins from 6mo to 15mo. To achieve this, we tested for differences over time between NC and AD using spline model from 6 to 15 months for each labeled peptide (Figure 3A). Brain tissue was the most structurally affected by AD, with 686 peptides in AD brain showing significantly different patterns of accessibility compared to NC from 6 to 15 months (Figure 3B). In heart, kidney, and thymus, fewer than 25 peptides showed significantly different patterns of change in AD relative to NC. There were no labeled peptides that exhibited significant different patterns in accessibility across all organs from 6 months to 15 months. However, a total of 10 labeled peptides consistently showed significantly different patterns in accessibility across four different organs each in the AD model. To examine the conformational changes of proteins specifically impacted by AD, we corrected for the confounding effect of age by dividing the individual accessibility of AD by the accessibility of its corresponding sequence of NC, resulting in a metric referred to as "fold-change" in this study. We evaluated the variability in structural changes depending on tissues relative to the brain using 10 peptides that showed distinct patterns of change in accessibility in AD from those in NC in four tissues (Figure 3C-F, Table S2). At the early stage of AD (6 months), the biggest conformational discrepancy due to AD was observed between the liver and brain. The lysine site of TAKGLF (Eno1) was not only more accessible in AD liver than in AD brain, but also more accessible in AD liver than in NC liver. As AD progressed, AD-specific structural changes in muscle and spleen were greater than those in brain when the fold-change between each tissue and the brain was compared. (Figure 3G-H). A functional enrichment analysis of the 10 proteins retrieved KEGG pathways associated with metabolism, glycolysis, and the TCA cycle (Figure 3I). For only 5 (51-67 amino acid (AA) of Atp5f1d, 469-479 AA of Dpysl2, 48-57 AA of Eno1, 1472-1483 AA of Flna, and 277-284 AA of Ppp2cb) out of 1219 peptides detected commonly in the 7 tissues, the accessibility patterns in all 7 tissues during AD progression were not significantly different from those during normal aging, and no significant difference in accessibility was observed across the 7 tissues under each condition (4-ages, disease). This indicates that the regions corresponding to these 5 peptides were not affected by AD in all 7 tissues. As expected, this investigation confirms that the effects of AD are most significantly observed in brain tissue, but by quantifying conformational changes of proteins across all tissues, we found tissue-specific variations associated with AD.

### Conformations of proteins in the brain are changing as AD progresses

Proteomic investigations into AD pathology have primarily relied on the analysis of differential protein expression. Bing et al. profiled the differentially expressed proteins and identified the protein networks that are affected during AD progression.<sup>34</sup> Co-expressed proteins and altered protein expression in human brain tissue of asymptomatic and symptomatic AD patients were reported by Nicholas et al.<sup>35</sup> Savas et al. measured protein expression in several mouse models of AD using quantitative mass spectrometry.<sup>36</sup> Despite extensive proteomic studies on AD and some footprinting studies on AD related proteins there have been no comprehensive *in vivo* studies of protein structures as AD progresses.<sup>19, 37-39</sup>

To uncover the changes in the 3D structure of brain proteins, we focused on 780 proteins, which are known to be expressed in the brain based on the Human Protein Atlas database (<https://www.proteinatlas.org>) and were also found to be dimethyl-labeled in our brain dataset. To identify lysine sites that change significantly in accessibility during the progression of AD and which also differ from NC, we used statistical methods to test the accessibility of 3,456 peptides corresponding to 780 proteins. First, a Kruskal-Wallis test was used to simultaneously compare lysine site accessibility among the four age

groups (6, 9, 12, and 15 months) of AD ( $P$ -value  $\leq 0.05$ ), and then Mann-Whitney was used to compare lysine site accessibility between AD and NC per age ( $P$ -value  $\leq 0.05$ ). Of the 3,456 peptides that were tested, 83 peptides corresponding to 62 proteins showed a significant difference in accessibility in all tests. These results suggest that the accessibility of these peptides changed consistently during the progression of Alzheimer's disease, and that they simultaneously diverged from the NC, thus indicating that the structural changes observed in the AD samples were not induced by aging (Figure 4A, 4B). Four pairs of peptides that shared lysine sites provided validation for the measured changes: APVISAEEKAY and APVISAEEKAYHEQL for Tuba1, HPEQLITGKEDAANNY and ITGKEDAANNY for Tuba1, QVVLVEPKTAW and QYQVVLVEPKTAW for Cnp, and RYLSEVASGENKQTTVSNSQQAY and SEVASGENKQTTVSNSQQAY for Ywhab. Many lysine sites showed a tendency to become inaccessible as aging progressed in both the NC and AD groups. However, the lysine sites in the AD group were found to be more inaccessible than those in the NC group. This suggests that the conformation of proteins in AD may be altered by the physiological changes associated with the disease. For example, 2',3'-cyclic-nucleotide 3'-phosphodiesterase (Cnp), which is associated with neuronal cells and glial cells,<sup>40</sup> was found to be enriched in the brain. When compared to the average expression levels in other tissues, the increase in expression levels of Cnp in the brain ranged from 13.4- to 23.9-fold in the NC and from 17.0- to 27.7-fold in the AD across four age groups (Figure 4C). The accessibility of 2 peptides of Cnp showed a decrease pattern in progressing AD but remained unchanged in the aging NC (Figure 4D and 4E). Of the 62 proteins that showed a significant difference in accessibility, 60 proteins (excluding two brain-specific proteins) exhibited a range of enrichment factor from 0.13-fold (Eef2) to 1,232-fold (Tuba1b) when compared to their expression in other tissues (Figure 4F). Specifically, on average, 51 of these proteins were expressed at higher levels in the brain than in other tissues, while 9 proteins were expressed at lower levels in the brain than in other tissues. Expression of a total 60 proteins was also enriched 21.7-fold in NC and 20-fold in AD on average, but no tendency in expression was observed in either aging or AD status.

These results suggest that the structural information obtained through the accessibility of lysine sites can complement protein expression data by revealing changes that are not fully characterized by protein expression levels alone. This method provides an approach to collect both protein expression and structural information to gain a more comprehensive understanding of the changes that occur in a proteome during the progression of a disease.

### Differential structural changes of tightly regulated proteins

We hypothesized that the changes in protein expression could accompany structural changes to proteins, which may suggest alterations to their physiological function. To evaluate our hypothesis, we initially analyzed the abundance of proteins in tightly co-regulated protein networks by modularizing them into protein communities using a WGCNA algorithm.<sup>28</sup> WGCNA was applied to the dataset of each tissue; 4,295 proteins in brain, 3,074 proteins in heart, 4,186 proteins in kidney, 4,298 proteins in liver, 1,721 proteins in muscle, 4,800 proteins in spleen, 4,159 proteins in thymus were used to build protein co-expression networks. No outliers were detected after all samples were hierarchically clustered using average distance and Pearson's method. For brain, the lowest soft threshold power was 22, with an  $R^2$  of more than 0.75. This network consisted of 17 modules of proteins related by their co-expression across control and disease tissues based on the TOM-based dissimilarity, after merging the modules with dissimilarity (Figure S3A and S3B). The WGCNA analysis also divided the protein data sets into 14, 17, 13, 10, 12 and 10 modules for heart, kidney, liver, muscle, spleen and thymus, respectively (Figure S4A-G).

We evaluated the correlation of the co-expressed proteins with Alzheimer's disease (AD) by comparing the co-expression of proteins in AD and NC within each module, irrespective of age. Correlation of coefficient  $R > 0.4$  and  $P$ -value  $< 0.05$  were set as the criteria for significant correlation for the AD trait. We found that only a limited number of protein communities were significantly correlated with the AD trait, with one module (M3) in the brain showing a correlation of 0.41, two modules (M11 and M14) in the kidney showing correlations of 0.43 and 0.44, two modules (M2 and M5) in the muscle showing correlations of 0.44 and 0.46, and one module (M4) in the spleen showing a correlation of 0.44. No modules in other tissues showed a significant correlation with the AD trait. (Figure 5A). We also assessed whether the direction and strength of the association between each module and the AD trait remained consistent (positive or negative) following the subdivision of the samples into four age groups (Figure S4A-G). Strong correlations were observed at a certain age in a few modules, while in most modules, the direction of correlation was inconsistent across the four age groups. For example, when samples from all age groups were included, a high correlation was observed for module 3 (M3) of brain ( $R = 0.41$ ) (Figure 5A and 5B).

Following that, the samples were divided by age. M3 proteins from the 9-month samples were found to be highly negatively correlated with a value of  $R = 0.82$ . At 12 and 15 months, samples were also negatively correlated, with  $R$  values of 0.42 and 0.72, respectively. In contrast, M3 proteins from the 6-month samples showed a positive correlation with a value of  $R = 0.4$  (Figure S4). The direction of association between the protein abundance-based module and the AD trait was found to fluctuate during the progression of AD. Therefore, we focused on the proteins in modules displaying significant correlations with the AD trait across all samples to investigate the conformational changes of the co-regulated proteins.

We investigated the conformational changes of the co-regulated proteins from the 4 types of tissues (brain, kidney, muscle, and spleen) that showed the significantly correlated modules based on WGCNA. To examine whether the conformational changes of proteins were specifically impacted by AD, the labeled peptides were clustered based on the fold-change of the accessibility during the progression of AD using the K-means clustering algorithm. The number of clusters was determined via an optimization algorithm (Figure S5). For the brain, the 481 labeled peptides that were mapped onto 174 proteins constituting M3 were clustered into two clusters (Figure 5C). Accessibility of the lysine sites of 268 peptides (123 proteins) in cluster 1 showed a steadily decreasing pattern, with a slight decrease during early AD development (6-9 months), a dramatic decrease from 9 to 12 months, and a slight decrease again during the late stage of AD development from 12 to 15 months (Figure 5D). This suggests that beginning at 9 months, the lysine sites included in cluster 1 were significantly sterically inaccessible due to AD. On the other hand, lysine sites of 213 peptides (113 proteins) in cluster 2 showed relatively stable accessibility during AD development (Figure 5E). Of 174 proteins in M3, 62 have 2 or more peptides included in both cluster 1 and cluster 2, and all labeled peptides of each of 61 and 51 proteins were exclusively included in cluster 1 and cluster 2, respectively (Figure S6). Twenty peptides that mapped to Map1a were identified most often in M3 proteins (Figure S7), and 5 of 20 peptides were hidden as AD progressed, but 15 lysine sites were spatially stable (Figure 5F). Subsequently, 19- and 14-labeled peptides were mapped to Pkg1 and Mdh1, respectively, and the peptides included in cluster 1 exhibited a consistently decreasing pattern of accessibility. The lysine sites of the peptides of Psat1, Mag, and Plp1, which are highly expressed in neuronal cells and were included in 62 brain proteins that constituted M3, became inaccessible during progression of AD (Figure 5G-I). The datasets from kidney, muscle, and spleen were processed separately to reveal the AD-induced spatial changes of proteins (Figure S8). Collectively, our findings suggest that the steric changes of proteins occur concurrently with changes in co-expression of proteins as AD progresses.

### **Biological functions of protein communities whose conformational changes precede expression changes**

We sought to uncover how the network undergoing conformational changes was related to biological functions, particularly those involved in neurodegenerative diseases. To achieve this, we examined the interactions of these proteins and their functional implications. In this analysis, we used proteins that had shown altered patterns of accessibility fold-change via K-means clustering. For instance, the proteins that were included in cluster 1 were used for the brain dataset.

We investigated how changes in the structure and expression of proteins could affect the progression of Alzheimer's disease by examining the pathways and biological functions they are involved in. We performed the analysis for the enrichment network on Metascape (metascape.org), considering the inter-term similarity and intra-term redundancy in the enriched terms.<sup>30</sup> This analysis represented an enriched term as a node connecting other nodes considering Kappa similarities. A total of 113 significantly enriched terms were grouped into 20 clusters based on their similarities and redundancies in the brain dataset (Figure S9 and Table S3). Fifteen proteins (Gnas, Mdh1, Ogdh, Pkg1, Ppp1cb, Slc1a3, Sod2, Taldo1, Sdha, Oxct1, Ndufa8, Epm2aip1, Aldh1l1, Etfa, and Ugp2) were enriched in "generation of precursor metabolites and energy" (node-107) with the most significantly enriched having a  $P$ -value of  $1.78 \times 10^{-9}$ . "Energy derivation by oxidation of organic compounds" (node-1) was also enriched significantly, but was similar to node 107 with kappa score of more than 0.3. It has been demonstrated that an abnormality of carbon and energy metabolism occurs in neurodegenerative disease since neurons require large amounts of energy to maintain their normal activity, and metabolic decline of the brain contributes to cognitive impairment.<sup>41, 42</sup> Interestingly, we discovered that structural changes preceded the expression changes in both protein communities "generation of precursor metabolites and energy" and "carbon metabolism" (node-96) (Figure 6A-B, 6D-E and Figure S10). Protein expressions in node-107 and node-96 increased in AD but remained stable in NC during aging. Expression fold-change increased significantly at

15 months, whereas accessibility fold-change decreased significantly from 12 months. We also noted that “metal ion homeostasis” (node-108) was significantly enriched with 12 proteins (Ank1, Calb1, Calb2, Gnas, Itpr1, Prkcb, Slc12a4, Slc1a3, Sod2, Vapb, Fis1, Immt). The homeostasis of metal ions is also known to be essential to maintain the normal function of brain, and abnormally elevated iron in brain is recognized to induce cell death and be a cause of several neurodegenerative diseases including AD.<sup>43, 44</sup> Zinc also has an essential role in protein binding for enzymatic activity or to modulate synaptic transmission, and abnormal levels of zinc have been reported to be implicated in AD.<sup>45, 46</sup> The enriched proteins in metal ion homeostasis showed an increased but insignificant pattern of expression, while a significantly decreased pattern in accessibility was observed from 9 months (Figure 6C, 6F and Figure S10).

Next, we used the STRING database to identify 45 out of 123 proteins included in cluster 1 that interacted directly with each other through 30 edges. Of these, 39 proteins were associated with either brain-related terms or the enriched terms in which proteins showed conformational changes preceding the expression change (node-97, node-107, and node-108) (Figure 7A, Table S4). We noted that Plp1 and Mag, which were associated with central nervous system and abnormal nervous system and were also known to be located in extracellular space, interacted directly with each other. AlphaFold2-Multimer<sup>47-49</sup> was utilized to predict the complex structure of Mag and Plp1. It showed that the two alpha carbons of the lysine residues FSKNYQDY of Plp1 and YFNSPYPKNYPPVVF of Mag were located within 13.9 Å (Figure 7B). Since protein-protein interactions can occur if proteins exist within 20 Å of each other,<sup>50</sup> it is possible that these adjacent peptides bind with each other. FSKNYQDY showed a greater decrease in accessibility in AD than in NC (Figure 7C), and YFNSPYPKNYPPVVF exhibited a similar accessibility in AD and NC at 6 months, but the lysine sites became inaccessible during AD progression, whereas no difference was observed in NC (Figure 7D).

In addition to the brain, it has been reported that the peripheral system plays a role in amyloid- $\beta$  clearance. Approximately 40%–60% of brain-derived amyloid- $\beta$  is transported across the blood-brain-barrier into the peripheral system for clearance, although the involved periphery and the mechanisms remain unclear.<sup>51</sup> Spleen is composed of a variety of immune cells (with 7-8% of all cells being monocyte/macrophage) and has a role in blood filter and immunological functions. In addition, the spleen monocytes/macrophages are reported to be involved in clearing amyloid beta.<sup>52</sup> Still, the physiological mechanisms underlying the association between the peripheral organs and AD remain unknown. In the ontology network of the spleen dataset, we noted “carbon metabolism” (node-147) and “neutrophil degranulation” (node-151). In these communities of proteins, the lysine sites in AD became exposed during progressing AD, while the accessibility in NC remained stable from 6 months to 15 months (Figure S11A-B). No significant change in the expression of proteins was observed for proteins of carbon metabolism (Figure S11D), but by 15 months the expression of proteins in neutrophil degranulation differed significantly from the expression at 12 months (Figure S11C). The results from the other tissues are shown in the supplementary data (Figure S12-S13 and Table S5-S8). Collectively, the results presented here provide compelling evidence of a relationship between conformational changes and protein expression, thereby highlighting the significant impact of organ-specific alterations of biological function during the progression of AD.

## Discussion

This study elucidated the AD-associated conformational changes in the proteomes of seven tissues in mice. We used the AD mouse model (APP<sup>NL-F</sup>), which expresses APP at wild-type levels while producing elevated pathogenic A $\beta$  through an APP knock-in approach. This model reduces the risk of artificial phenomena that might be observed with an APP overexpressed mouse model and enhances the interpretability of the results.<sup>21</sup> We found co-regulated proteins whose accessibility changed in 4 of 7 tissues, and we connected the structural protein differences in progressing AD compared to normal aging of unaffected mice to possible alterations in their biological functions. Whole animal perfusion was used to deliver reagents to comprehensively dimethyl label mouse organs with minimal intervention such as organ excision, tissue homogenization, and protein extraction that could denature or alter proteins. Our quantitative method for accessibility measurement determines the relative fraction of inaccessible over accessible for each lysine site. Changes in accessibility can be interpreted as a change in protein folding or a change in interaction with another molecule and thus can be a surrogate for protein conformation changes between different conditions.

In this study, we focused on changes in proteins of the brain. Amyloid beta is a well-known protein that accumulates as plaques in the brain and serves as a marker of AD, but we were not able to detect structural changes of amyloid beta. We quantified the accessibility of three peptides from APP in only one 15-month mouse. It has been known that the amyloid beta plaque accumulates primarily in the cortex, followed by the hippocampus, basal ganglia, thalamus, and basal forebrain.<sup>53</sup> However, because the purpose of this study was to globally investigate the change of the proteome, we used whole brain tissue sample in this study, not just the amyloid-beta plaque containing cortex region of the brain.

It has been suggested that kidney function is linked to brain activity,<sup>54</sup> and changes in kidney function may play a role in the development and progression of AD. Studies have shown that the MRI images of the brains of AD patients were similar to that of kidney patients<sup>55</sup> and a systematic meta-analysis demonstrated that cognitive impairment is significantly related to malfunction of kidney.<sup>56</sup> Despite persuasive evidence of the link between the kidney and AD, the exact physiological mechanisms underlying this relationship are not fully understood. From our WGCNA analysis, we found two modules (M11 and M14) to be significantly co-expressed in kidney and investigated the enriched functions of the structurally altered proteins using GO enrichment analysis. In both modules, purine-related functions were most significantly enriched. Our findings from kidney are supported by results of metabolomic studies showing that guanosine monophosphate (which is derived from purine guanine and associated with purine metabolism) was dysregulated in the brain of an AD mouse model based on APOE4 allele mutant mice.<sup>57</sup> Therefore, it can be inferred that purine metabolism may play a role in the link between kidney function and AD.

We introduced a feature to the CPP pipeline that can be employed under various experimental conditions. A multidimensional protein identification technology (MudPIT) can be used to generate high sequence coverage of the proteome.<sup>58, 59</sup> MudPIT separates peptides first by charge and then by hydrophobicity to create a two-dimensional separation. In MudPIT, peptides released from a C18 trap column are loaded onto a strong cation exchange column and then released with a buffer that increases in ionic strength through sequential elution steps. Released peptides are bound to the C18 analytical column and are then eluted sequentially to the mass spectrometer based on hydrophobicity. It is not recommended that hydrophobicity be used for the first dimension of the 2D-LC-MS because the hydrophobicity of deuterium is slightly lower than hydrogen, so a partial separation of differentially labeled peptides might occur.<sup>60</sup>

Limitations to this study include the lack of a standard to monitor the distribution of reagent solution to organs in the body during the first labeling step. Although some signs such as body twitching, tail flicking, and head moving in the anesthetized animals was observed, a reliable quantitative standard to assess the extent of labeling in each organ would be useful. We defined labeling efficiency as the ratio of initially labeled peptides over identified lysine-containing peptides per tissue. It is an inevitable limitation of CPP that dimethylation cannot occur on a lysine site that is already modified (i.e., acetylation and ubiquitination); thus, the accessibility of innate partially labeled lysine sites will not be accurate. An additional limitation is the method uses lysine as the conformation reporter which occurs at a frequency of roughly 5-7% in proteins and thus the quality of the analysis scales with sequence coverage of individual proteins. High sequence coverage will generate more MS/MS of lysine containing peptides and increase the completeness of the analysis.

It is likely that changes in protein structure are a result of failing proteostasis and expression

changes are a function of alterations in protein synthesis and/or degradation. Mutations of DNA in somatic cells accumulate as we age<sup>61-64</sup> and can result in changes to protein sequences, necessitating more effort to keep proteins properly folded. Our new method provides a means to measure protein surface accessibility as a surrogate for protein conformation *in vivo* and in animal models to study of the role of protein folding in aging and AD. Our proteomic analysis showed changes in protein structures in multiple tissues during the progression of AD. Even though the patterns of change in non-brain tissue did not correspond exactly with those in the brain tissue (and it is not clear they should), our analysis showed changes in protein structure and expression in other tissues in the AD mouse model. In conclusion, this new method to measure *in vivo* alterations to protein surface accessibility in animal models of disease provides a means to measure a previously unexplored characteristic of proteins to provide insights into how physiological systems are perturbed.

### Data availability

All mass spectrometry raw data in this study have been deposited to MassIVE repository with identifier MSV000091970.

### Acknowledgements

We thank Dr. Claire Delahunty for critical reading the manuscript. This work was supported by NIH/NIA grant RF1AG061846-01 and 5R01AG075862.

### Author contributions

A.S., C.B. and J.R.Y. conceived the project. A.S. and H.K. performed the experiments. J.D. measured the samples on the mass spectrometer. A.S. and H.K. analyzed the data and the results. C.B. and D.M. provided the critical feedback. J.R.Y. supervised the project. A.S. wrote the manuscript and prepared the figures with help from all authors.

### Declaration of interests

The authors declare no competing interests.

### References

1. Jones, S. & Thornton, J.M. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* **93**, 13-20 (1996).
2. Keskin, O., Gursoy, A., Ma, B. & Nussinov, R. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? *Chem Rev* **108**, 1225-1244 (2008).
3. Sakakibara, D. et al. Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* **458**, 102-105 (2009).
4. Oikonomou, C.M. & Jensen, G.J. Cellular Electron Cryotomography: Toward Structural Biology In Situ. *Annu Rev Biochem* **86**, 873-896 (2017).
5. de Souza, N. & Picotti, P. Mass spectrometry analysis of the structural proteome. *Curr Opin Struct Biol* **60**, 57-65 (2020).
6. Benjin, X. & Ling, L. Developments, applications, and prospects of cryo-electron microscopy. *Protein Sci* **29**, 872-882 (2020).
7. Renaud, J.P. et al. Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat Rev Drug Discov* **17**, 471-492 (2018).
8. Sheshberadaran, H. & Payne, L.G. Protein antigen-monoclonal antibody contact sites investigated by limited proteolysis of monoclonal antibody-bound antigen: protein "footprinting". *Proc Natl Acad Sci U S A* **85**, 1-5 (1988).
9. West, G.M. et al. Quantitative proteomics approach for identifying protein-drug interactions in complex mixtures using protein stability measurements. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9078-9082 (2010).
10. Xu, G. & Chance, M.R. Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem Rev* **107**, 3514-3543 (2007).
11. Mendoza, V.L. & Vachet, R.W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom Rev* **28**, 785-815 (2009).
12. Geer, M.A. & Fitzgerald, M.C. Energetics-based methods for protein folding and stability measurements. *Annu Rev Anal Chem (Palo Alto Calif)* **7**, 209-228 (2014).
13. Li, K.S., Shi, L. & Gross, M.L. Mass Spectrometry-Based Fast Photochemical Oxidation of Proteins (FPOP) for Higher Order Structure Characterization. *Acc Chem Res* **51**, 736-744 (2018).
14. Pepelnjak, M., de Souza, N. & Picotti, P. Detecting Protein-Small Molecule Interactions Using Limited Proteolysis-Mass Spectrometry (LiP-MS). *Trends Biochem Sci* **45**, 919-920 (2020).
15. Cappelletti, V. et al. Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. *Cell* **184**, 545-559 e522 (2021).
16. Mackmull, M.T. et al. Global, in situ analysis of the structural proteome in individuals with Parkinson's disease to identify a new class of biomarker. *Nat Struct Mol Biol* **29**, 978-989 (2022).
17. Espino, J.A., Mali, V.S. & Jones, L.M. In Cell Footprinting Coupled with Mass Spectrometry for the Structural Analysis of Proteins in Live Cells. *Anal Chem* **87**, 7971-7978 (2015).

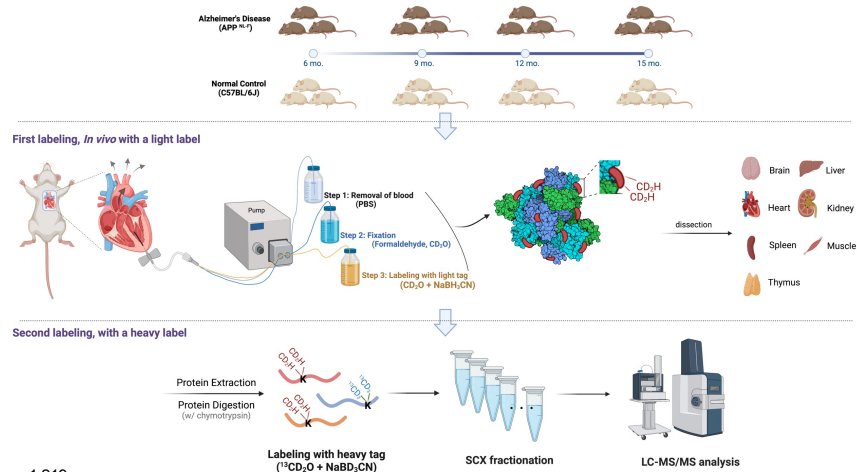


18. Espino, J.A. & Jones, L.M. In Vivo Hydroxyl Radical Protein Footprinting for the Study of Protein Interactions in *Caenorhabditis elegans*. *J Vis Exp* (2020).
19. Bamberger, C. et al. Protein Footprinting via Covalent Protein Painting Reveals Structural Changes of the Proteome in Alzheimer's Disease. *J Proteome Res* **20**, 2762-2771 (2021).
20. Bamberger, C., Diedrich, J., Martinez-Bartholome, S. & Yates, J.R., 3rd Cancer Conformational Landscape Shapes Tumorigenesis. *J Proteome Res* **21**, 1017-1028 (2022).
21. Saito, T. et al. Single App knock-in mouse models of Alzheimer's disease. *Nat Neurosci* **17**, 661-663 (2014).
22. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D. & Nesvizhskii, A.I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat Methods* **14**, 513-520 (2017).
23. Bamberger, C., Pankow, S., Park, S.K. & Yates, J.R., 3rd Interference-free proteome quantification with MS/MS-based isobaric isotopologue detection. *J Proteome Res* **13**, 1494-1501 (2014).
24. Templ, M., Kowarik, A. & Filzmoser, P. Iterative stepwise regression imputation using standard and robust methods. *Comput Stat Data An* **55**, 2793-2806 (2011).
25. Templ, M., Alfons, A. & Filzmoser, P. Exploring incomplete data using visualization techniques. *Adv Data Anal Classif* **6**, 29-47 (2012).
26. Klaus Jung, A.G., Barbara Sitek, Ognjan Apostolov, Alexander Schramm, Helmut E. Meyer, Kai Stühler, Wolfgang Urfer Stastical evaluation of methos for the analysis of dynamic protein expression data from a tumor study. *REVSTAT-STATISTICAL JOURNAL* **4** (2006).
27. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
28. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17 (2005).
29. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-1093 (2009).
30. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* **10**, 1523 (2019).
31. Doncheva, N.T., Morris, J.H., Gorodkin, J. & Jensen, L.J. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res* **18**, 623-632 (2019).
32. Doncheva, N.T. et al. Cytoscape stringApp 2.0: Analysis and Visualization of Heterogeneous Biological Networks. *J Proteome Res* (2022).
33. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682 (2022).
34. Bai, B. et al. Deep Multilayer Brain Proteomics Identifies Molecular Networks in Alzheimer's Disease Progression. *Neuron* **105**, 975-991 e977 (2020).
35. Seyfried, N.T. et al. A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. *Cell Syst* **4**, 60-72 e64 (2017).
36. Savas, J.N. et al. Amyloid Accumulation Drives Proteome-wide Alterations in Mouse Models of Alzheimer's Disease-like Pathology. *Cell Rep* **21**, 2614-2627 (2017).
37. Gau, B., Garai, K., Frieden, C. & Gross, M.L. Mass spectrometry-based protein footprinting characterizes the structures of oligomeric apolipoprotein E2, E3, and E4. *Biochemistry* **50**, 8117-8126 (2011).
38. Klinger, A.L. et al. A synchrotron-based hydroxyl radical footprinting analysis of amyloid fibrils and prefibrillar intermediates with residue-specific resolution. *Biochemistry* **53**, 7724-7734 (2014).
39. Li, K.S., Rempel, D.L. & Gross, M.L. Conformational-Sensitive Fast Photochemical Oxidation of Proteins and Mass Spectrometry Characterize Amyloid Beta 1-42 Aggregation. *J Am Chem Soc* **138**, 12090-12098 (2016).
40. Sprinkle, T.J. 2',3'-cyclic nucleotide 3'-phosphodiesterase, an oligodendrocyte-Schwann cell and myelin-associated enzyme of the nervous system. *Crit Rev Neurobiol* **4**, 235-301 (1989).
41. Knobloch, M. & Paolicelli, R.C. Myeloid Metabolism as a New Target for Rejuvenation?-Comments on Restoring Metabolism of Myeloid Cells Reverses Cognitive Decline in Ageing. *Nature*. 2021 Feb;590(7844):122-128. *Immunometabolism* **3**, e210034 (2021).
42. Navarro, A. & Boveris, A. Mitochondrial nitric oxide synthase, mitochondrial brain dysfunction in aging, and mitochondria-targeted antioxidants. *Adv Drug Deliv Rev* **60**, 1534-1544 (2008).
43. Drayer, B. et al. MRI of brain iron. *AJR Am J Roentgenol* **147**, 103-110 (1986).
44. Levi, S. & Finazzi, D. Neurodegeneration with brain iron accumulation: update on pathogenic mechanisms.

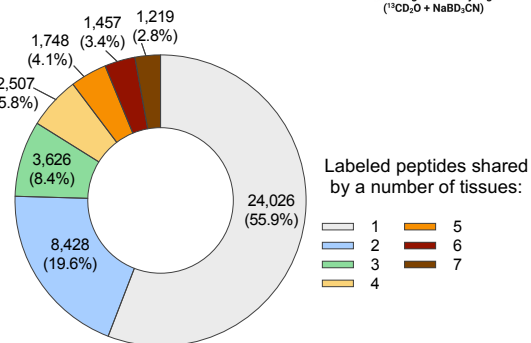
- Front Pharmacol* **5**, 99 (2014).
45. Paoletti, P., Vergnano, A.M., Barbour, B. & Casado, M. Zinc at glutamatergic synapses. *Neuroscience* **158**, 126-136 (2009).
  46. Sensi, S.L., Paoletti, P., Bush, A.I. & Sekler, I. Zinc in the physiology and pathology of the CNS. *Nat Rev Neurosci* **10**, 780-791 (2009).
  47. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).
  48. Tunyasuvunakool, K. et al. Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590-596 (2021).
  49. Evans, R. et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv* (2022).
  50. Vangone, A. & Bonvin, A.M. Contacts-based prediction of binding affinity in protein-protein complexes. *Elife* **4**, e07454 (2015).
  51. Marques, M.A. et al. Peripheral amyloid-beta levels regulate amyloid-beta clearance from the central nervous system. *J Alzheimers Dis* **16**, 325-329 (2009).
  52. Yu, Z.Y. et al. Physiological clearance of Aβ by spleen and splenectomy aggravates Alzheimer-type pathogenesis. *Aging Cell* **21**, e13533 (2022).
  53. Thal, D.R., Rub, U., Orantes, M. & Braak, H. Phases of Aβ-deposition in the human brain and its relevance for the development of AD. *Neurology* **58**, 1791-1800 (2002).
  54. Tang, X. et al. Association of kidney function and brain health: A systematic review and meta-analysis of cohort studies. *Ageing Res Rev* **82**, 101762 (2022).
  55. Guo, Z. et al. Common 1H-MRS Characteristics in Patients With Alzheimer's Disease and Vascular Dementia Diagnosed With Kidney Essence Deficiency Syndrome: A Preliminary Study. *Altern Ther Health Med* **23**, 12-18 (2017).
  56. Etgen, T., Chonchol, M., Forstl, H. & Sander, D. Chronic kidney disease and cognitive impairment: a systematic review and meta-analysis. *Am J Nephrol* **35**, 474-482 (2012).
  57. Nuriel, T. et al. Neuronal hyperactivity due to loss of inhibitory tone in APOE4 mice lacking Alzheimer's disease-like pathology. *Nat Commun* **8**, 1464 (2017).
  58. Link, A.J. et al. Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* **17**, 676-682. (1999).
  59. Washburn, M.P., Wolters, D. & Yates, J.R., 3rd Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* **19**, 242-247 (2001).
  60. Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A.J. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* **4**, 484-494 (2009).
  61. Miller, M.B., Reed, H.C. & Walsh, C.A. Brain Somatic Mutation in Aging and Alzheimer's Disease. *Annu Rev Genomics Hum Genet* **22**, 239-256 (2021).
  62. Choudhury, S. et al. Somatic mutations in single human cardiomyocytes reveal age-associated DNA damage and widespread oxidative genotoxicity. *Nat Aging* **2**, 714-725 (2022).
  63. Kim, J. et al. Prevalence and mechanisms of somatic deletions in single human neurons during normal aging and in DNA repair disorders. *Nat Commun* **13**, 5918 (2022).
  64. Luquette, L.J. et al. Single-cell genome sequencing of human neurons identifies somatic point mutation and indel enrichment in regulatory elements. *Nat Genet* **54**, 1564-1571 (2022).

## Figures

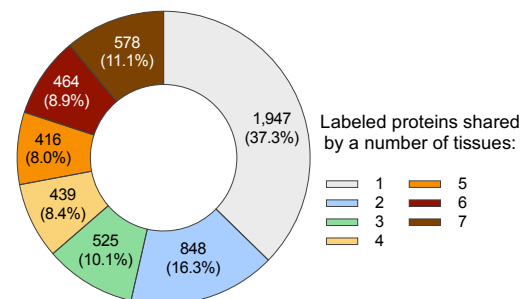
A



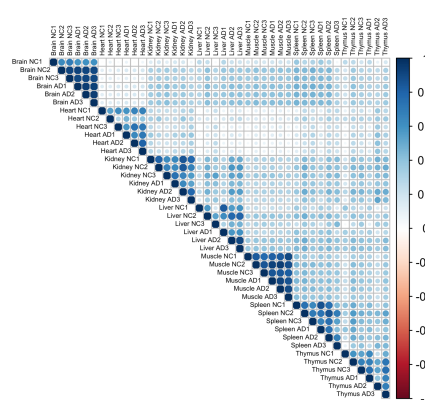
B.



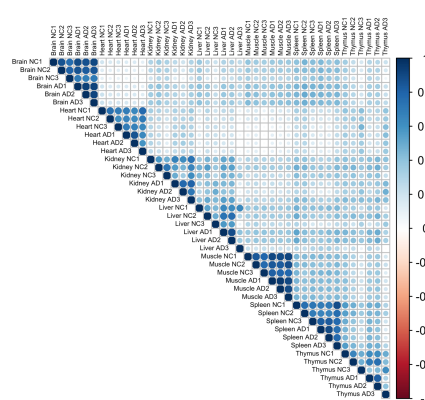
C



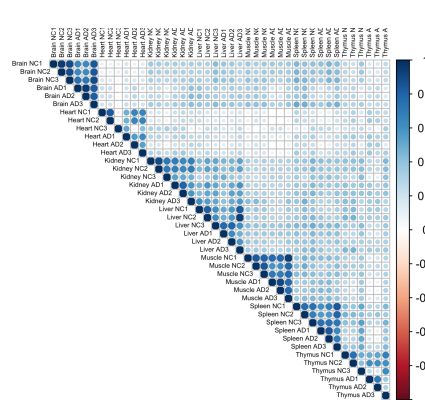
D



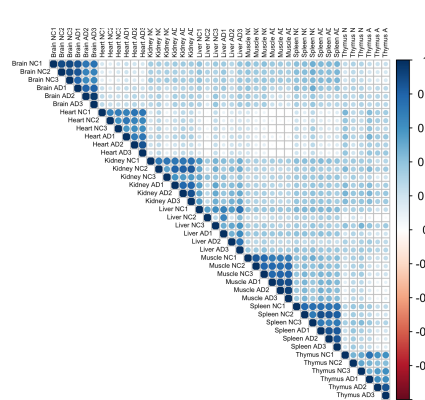
E



F



G



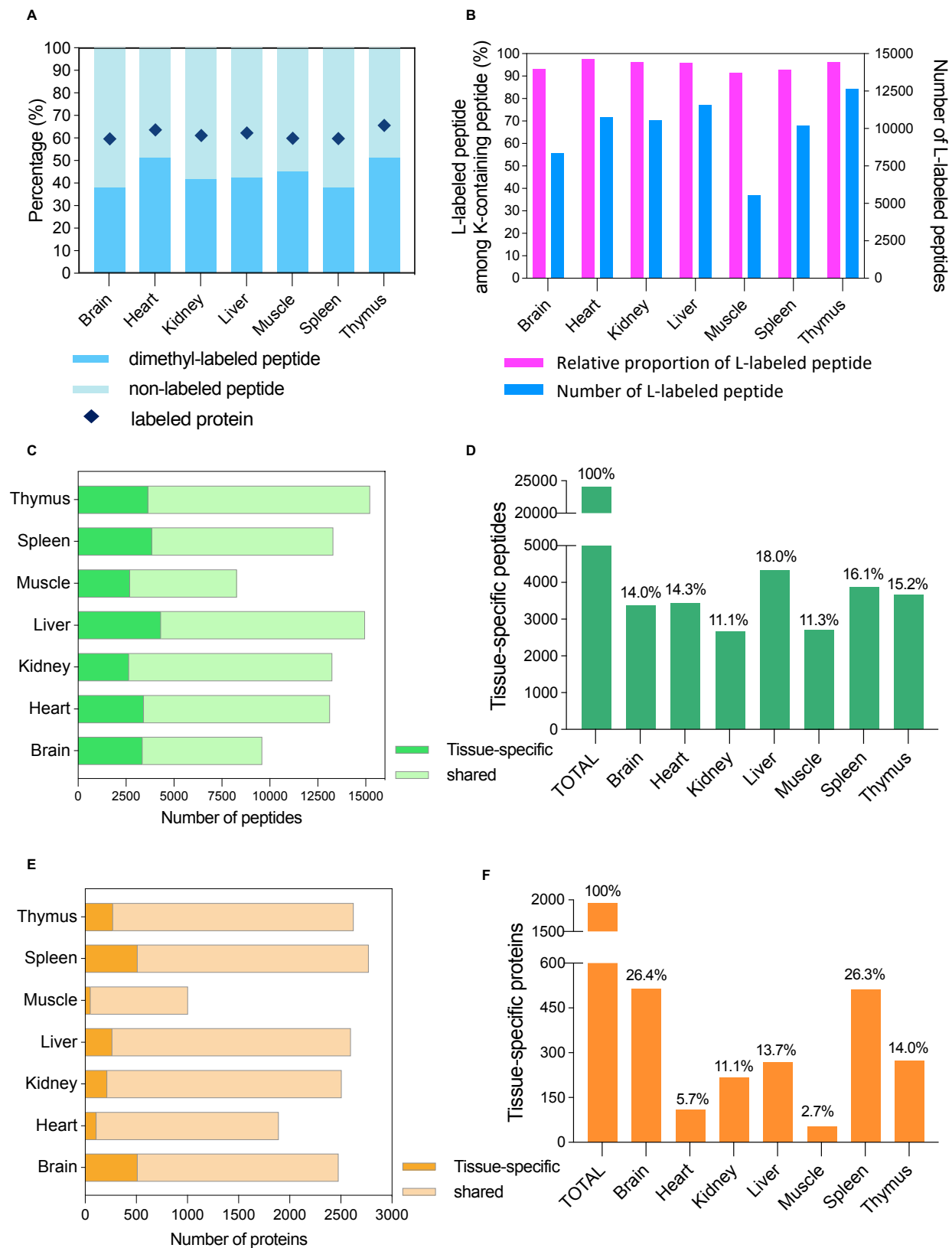
**Figure 1. Strategy for identifying dimethyl-labeled peptides**

**A.** Three mice per each age group (ranging 6 months to 15 months) were used for AD (APP<sup>NL-F</sup>) and NC (C57BL6/J). The first step of the CPP workflow consists of three sub-steps that were conducted via perfusion: i) blood was washed by PBS, ii) tissue was fixed by formaldehyde, and iii) exposed lysine sites of the native proteins were labeled with light-dimethylation ([CD<sub>2</sub>H]<sub>2</sub>). Proteins from each of the seven organs were extracted and digested separately with chymotrypsin, after which the newly exposed lysine sites were labeled with heavy dimethylation ([C<sup>13</sup>D<sub>3</sub>]<sub>2</sub>).

**B.** More than half of the total labeled peptides were tissue-specific. Less than 3% of a total of labeled peptides were peptides common to all seven tissues.

**C.** The proportions of labeled proteins were determined by assigning proteins to the labeled peptides. Unlabeled proteins were not counted. The largest portion of labeled proteins was tissue-specific proteins, and the portion of proteins common to all 7 tissues was the third largest portion.

**D-G** Biological triplicates were correlated across 7 tissues at 6 months (D), 9 months (E), 12 months (F) and 15 months (G).



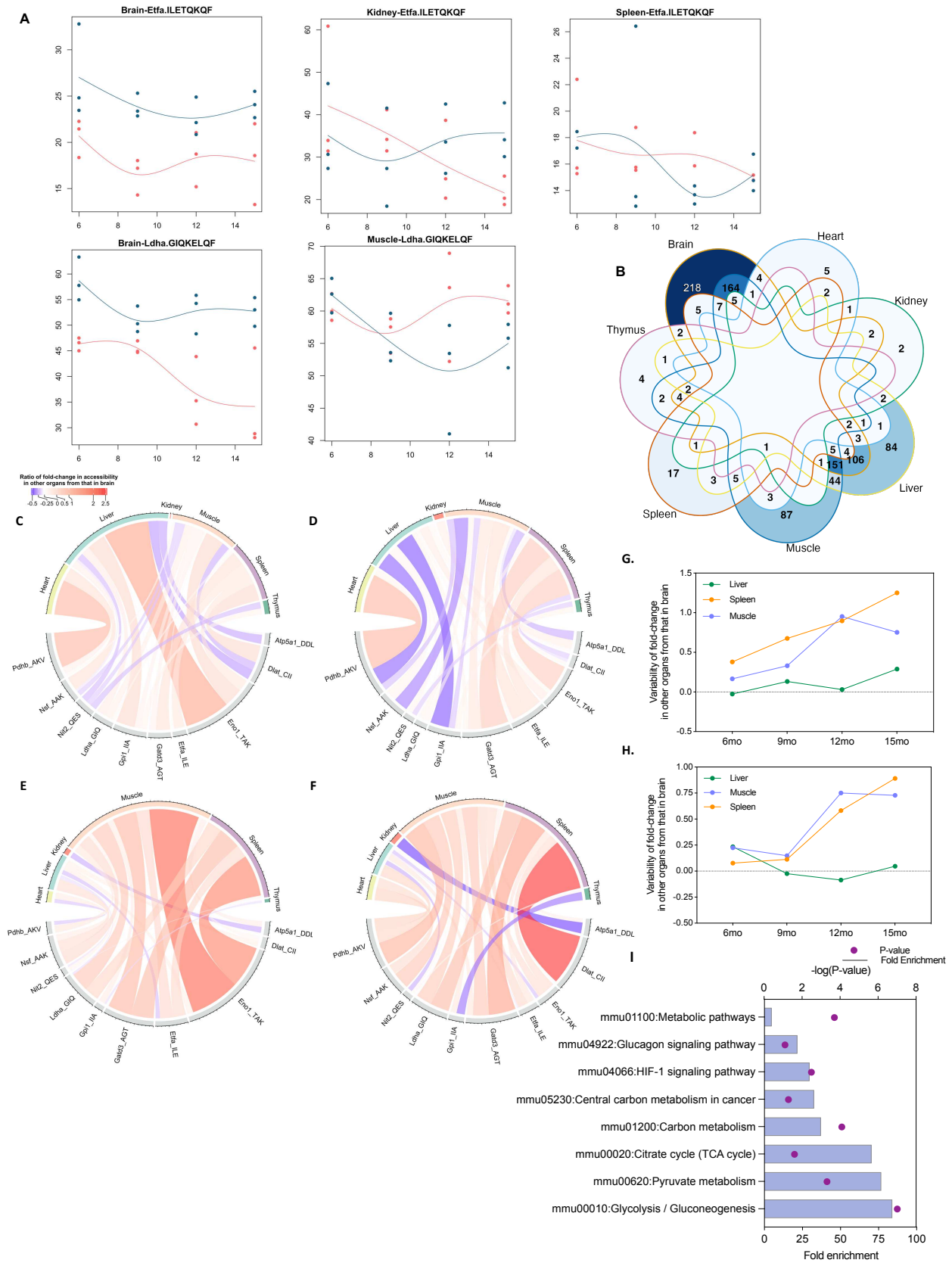
15 **Figure 2. Labeling efficiency of the 1<sup>st</sup> labeling via perfusion**

**A.** All the labeled peptides (light, heavy or both light and heavy) were counted as identified peptides for each tissue sample. The proportions of identified peptides that were labeled ranged from 38.01% (n = 8,758) in brain to 51.20% (n = 10,779) in heart (blue). Diamonds indicate the relative proportion of proteins that were labeled.

**B.** The proportions of lysine-containing peptides that were light-labeled were determined for each tissue. The proportions ranged from 91.5% (n = 5,583) in muscle to 97.5% (n = 10,745) in heart (pink). All tissue samples show a labeling efficiency of more than 90% for labeling via the perfusion method, while the number of light-labeled peptides were variable (blue).

**C, E.** The tissue-specific labeled peptides (C)/proteins (E) (green/orange) and shared peptides/proteins (pale green/pale orange) were determined for each tissue sample.

**D, F.** These bar graphs represent the contribution of each tissue to the total number of tissue-specific labeled peptides (D)/proteins (F).



28 **Figure 3. Variability of the conformational changes depending on the tissues**

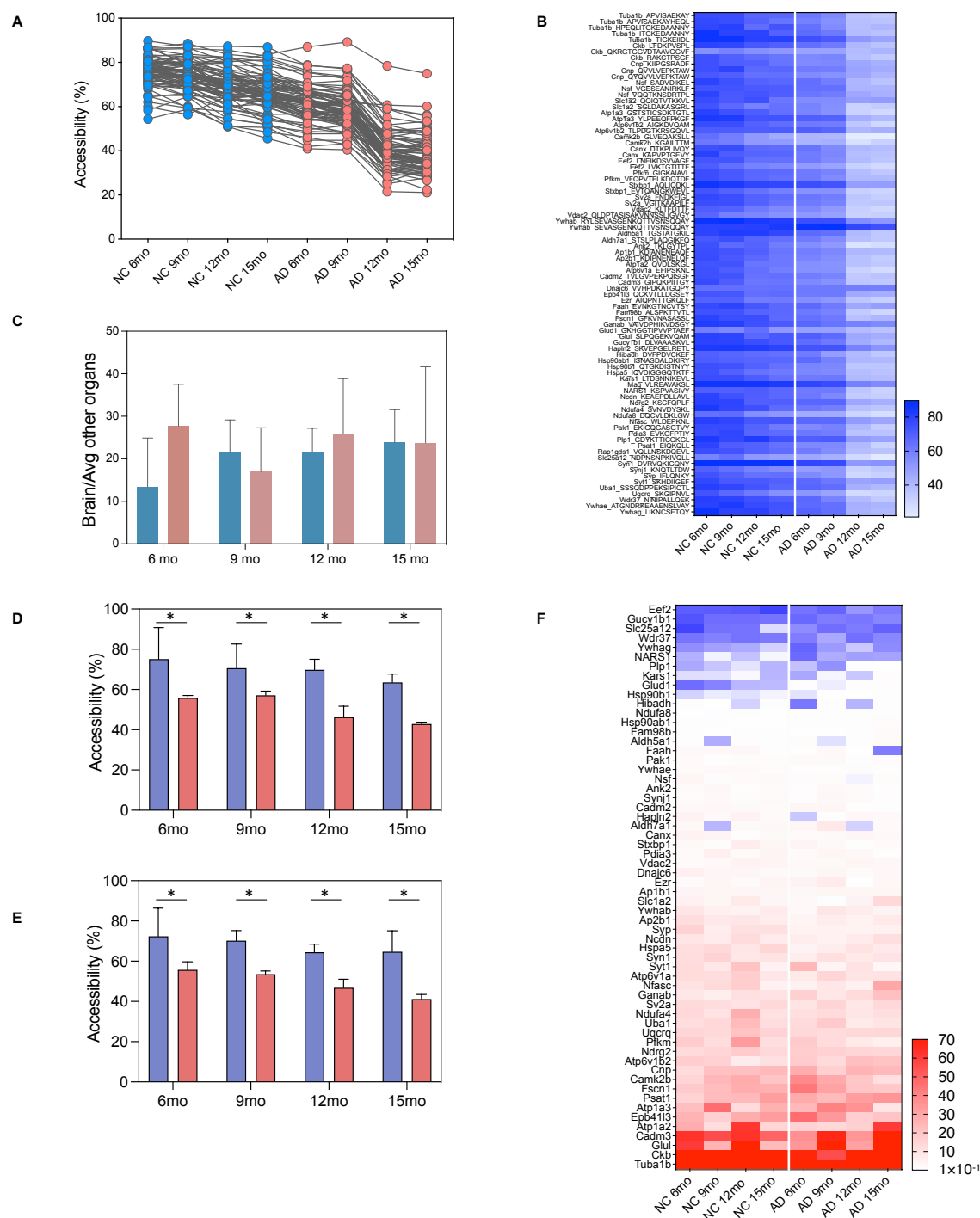
29 **A.** Two representative peptides (ILETQKQF and GIQKELQF) were shown representatively. The normalized  
30 values were utilized to fit spline models. The accessibility of each peptide for both NC (blue) and AD (red)  
31 exhibited significantly distinct patterns from 6 to 15 months (P-value < 0.05).

32 **B.** Venn diagram shows the number of peptides exhibiting significant differences in the trend of accessibility  
33 changes between NC and AD. There were no peptides from AD that showed a significant difference in  
34 accessibility changes compared to NC in all 7 tissues during the period from 6 to 15 months. The value of  
35 zero was not indicated.

36 **C-H** During AD progression, 10 common peptides exhibited distinct patterns in accessibility between NC  
37 and AD in different four tissues. The variabilities for the structural changes in each tissue were calculated  
38 based on the value of brain using the formula: (fold-change of other tissue - fold-change of brain) / fold-  
39 change of brain at 6 mo (C), 9 mo (D), 12 mo (E), and 15 mo (F). Only the first three amino acids were  
40 shown. AGTAEAIKAL of Gatd3 (G) and GIQKELQF of Ldha (H) showed a difference in the magnitude of  
41 accessibility change in muscle and spleen compared to that in the brain as AD progressed.

42 **I.** Enriched KEGG pathways with 10 proteins.





**Figure 4. Changes in the structures of proteins that are known to be associated with brain.**

**A, B.** Accessibility of 83 labeled peptides that mapped to 62 proteins were significantly changed. The accessibility values in AD groups decreased more steeply than those in NC groups (A). The proteins and peptides corresponding to each change in the accessibility are indicated (B).

**C, D, E.** Cnp was more highly expressed in brain than in the other tissues (C). Two of the three peptides of

48 Cnp share one lysine site, and variations in accessibility for these two peptides are represented.  
49 KIIPGSRADF (D) is located at 87-96 amino acid of Cnp and QYQVVLVEPKTAW (E) is located at 141-153  
50 amino acid of Cnp. While they exhibited a decreasing trend in AD, the accessibility values were lower in AD  
51 compared to NC, and the trend in AD was steeper than in NC. Blue indicates NC groups, pink indicates AD  
52 groups. Asterisk (\*) denotes significance.

53 **F.** Expression of 60 brain proteins were compared to averaged expression in six other tissues (F). The  
54 minimum enrichment factor was 0.13 for Eef2 in NC at 15 mo and the maximum enrichment factor was  
55 1,232 for Tuba1b in NC at 12 mo. Expression of 6 proteins (Eef2, Gucy1b1, NARS1, Slc25a12, Wdr37,  
56 Ywhag) was lower in brain than the expression of the corresponding proteins in other tissues at all ages in  
57 NC and AD. The bar indicates the enrichment factor, with red indicating an enrichment factor more than 1,  
58 and blue indicating an enrichment factor less than 1. Proteins enriched more than 70-fold are marked in  
59 red.



62 **A.** Four of the seven tissues (brain, kidney, muscle and spleen) showed significant correlated modules ( $R^2$   
63  $> 0.4$ , P-value  $< 0.05$ ).

64 **B.** In module 3 (M3) of brain, the eigenprotein level between AD and NC was assessed, with dot colors  
65 indicating the age of the mice.

66 **C-E.** The labeled peptides of M3 proteins were clustered based on the fold-change of the accessibility (C).  
67 Distribution of the fold-change of the accessibility of cluster 1 (D) and cluster 2 (E). Of 481 labeled peptides  
68 that mapped onto 174 proteins in M3, fold-change of accessibility of 268 peptides showed a consistent  
69 decrease in progressing AD in cluster 1 and fold-change of accessibility of 213 peptides in cluster 2 did not  
70 significantly change in progressing AD. \* $p < 0.05$ , \*\* $p < 0.005$  and \*\*\*\* $p < 0.0001$

71 **F-I.** The fold changes of the accessibility for Map1a (F), Psat (G), Mag (H), and Plp1 (I) are presented. The  
72 peptides in the bold box were included in cluster 1. Peptides in the bold box clearly decreased.

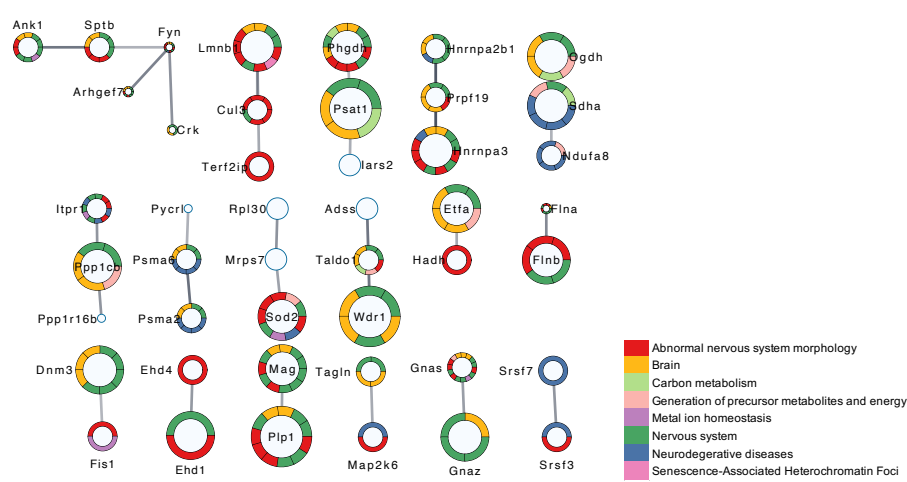


74 **A-C.** Structural changes of proteins enriched in generation of precursor metabolite and energy (A), carbon  
75 metabolism (B), and metal ion homeostasis (C) are shown. The heatmap (right) showed variations in the  
76 fold-change of the accessibility based on peptide sequence. The scatter plots (left) were plotted irrespective  
77 of peptide sequences.

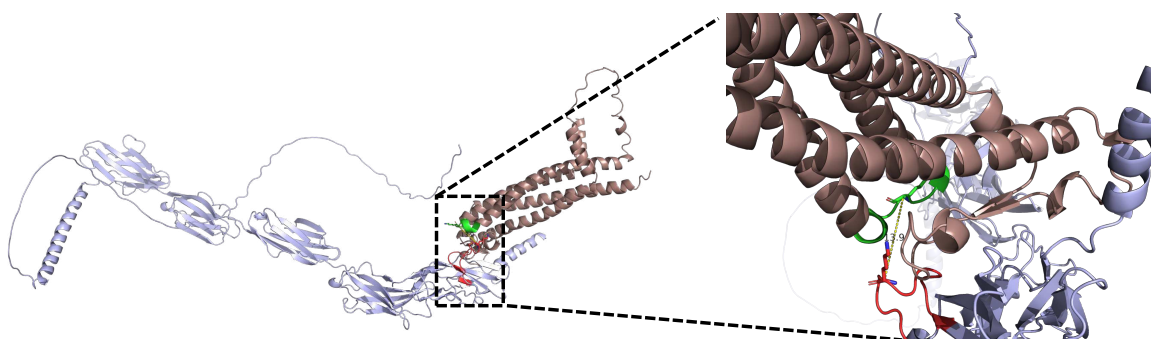
78 **D-F.** Expression change of proteins enriched in generation of precursor metabolite and energy (D), carbon  
79 metabolism (E), and metal ion homeostasis (F). The fold-changes of the expression level are presented.

80 \* $p < 0.05$ , \*\* $p < 0.005$  and \*\*\*\* $p < 0.0001$

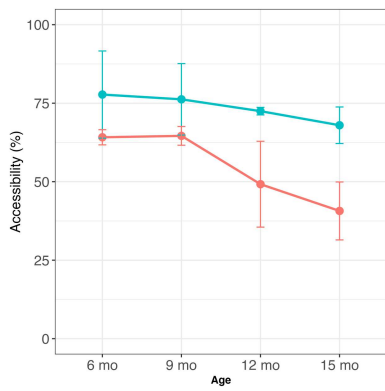
A



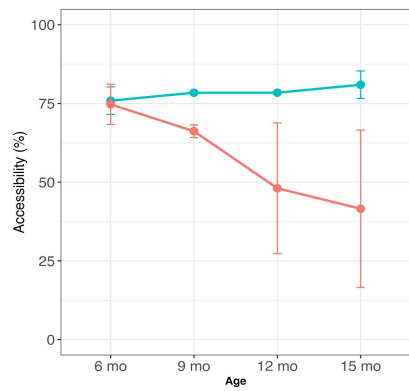
B



C.



D.



81 **Figure 7. The physical interactions of the proteins in M3 of brain**

**A.** Forty-five proteins in the brain dataset were physically interacted. Each node indicates a protein. The ring color of the node indicates the terms that the protein is associated with. The size of node represents the number of significantly changed lysine sites, with very small nodes indicating no significantly changed peptides, small nodes indicating one significantly changed peptide, medium nodes indicating 2-4 significantly changed peptides, and large nodes indicating more than 4 significantly changed peptides.

**B.** The structure of the Plp1-Mag complex was predicted using AlphaFold-Multimer. The structure in dark pink is Plp1 and the structure in light purple is Mag. FSKNYQDY of Plp1 and YFNSPYPKNYPPVVF of Mag were presented in green and red, respectively. The right panel is an enlarged view of the complex on the left. The distance between alpha-carbons of two lysine sites was 13.9 Å.

**C, D.** Structural changes in adjacent peptide regions with potential binding, with variation of the accessibility of site in AD (pink) and NC (green) for FSKNYQDY (C) and YFNSPYPKNYPPVVF (D).



# Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementalinformation.pdf](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)
- [TableS8.xlsx](#)