The Innovation, Volume *5*

Supplemental Information

Alzheimer's disease early diagnostic and staging biomarkers revealed by large-scale cerebrospinal fluid and serum proteomic profiling

Qing-Qing Tao, Xue Cai, Yan-Yan Xue, Weigang Ge, Liang Yue, Xiao-Yan Li, Rong-Rong Lin, Guo-Ping Peng, Wenhao Jiang, Sainan Li, Kun-Mu Zheng, Bin Jiang, Jian-Ping Jia, Tiannan Guo, and Zhi-Ying Wu

Supplemental materials

Detailed Methods

Table S1 The demographic data of the discovery cohort.

Table S2 TMT-based proteome analysis reveals dysregulated proteins in early AD and along with the AD disease progression.

Table S3 The demographic data of the validation cohort.

Table S4 PRM-based targeted proteome analysis verified dysregulated proteins in early AD and along with the AD disease progression.

Table S5 The demographic data of the used subset from the validation cohort.

Figure S1. Quantified CSF proteins identified by TMT-based LC-MS/MS analysis.

Figure S2. Quantified serum proteins identified by TMT-based LC-MS/MS analysis.

Figure S3. The ranking diagrams of CSF and serum proteins.

Figure S4 Validated selected differentially expressed proteins by real-time PCR.

Figure S5. Evaluation of protein biomarker panels to discriminate other neurodegenerative diseases by uniform manifold approximation and projection (UMAP).

Figure S6. Evaluation of serum GFAP and NEFL to discriminate MCI due to AD and other neurodegenerative disease.

Figure S7. Correlations of CSF and serum core dysregulated proteins in MCI with CSF Aβ42, t-tau and p-tau181.

Figure S8. Discovery of CSF protein biomarkers for staging the course of AD.

Figure S9. Network pathway analysis of dysregulated CSF proteins during the different stages of AD reveals several important molecular pathways.

Figure S10. Discovery of serum protein biomarkers for staging the course of AD.

Figure S11. Network pathway analysis of dysregulated serum proteins during the different stages of AD reveals several important molecular pathways.

Figure S12. Quality control of CSF and serum proteomic data acquired by TMT-based LC-MS/MS analysis.

Figure S13. Quality control of PRM proteome data.

Detailed methods

Enzyme-linked immunosorbent assays (ELISAs) for measurement of CSF Aβ42, t-tau, and p-tau181 levels

CSF Aβ42, T-tau, and p-tau181 levels were measured by ELISA Kits (Fujirebio, Ghent, Belgium) according to the manufacturer's instructions as described by our previous study.[1] Briefly, CSF samples, calibrators (CALs), Run Validation Controls (RVC) and all other reagents were thawed and allowed to reach room temperature before use. At the beginning of the tests, conjugate working solution 1 (75μL for Aβ42 and t-tau, and 25μL for p-tau181) and CSF sample/CAL/RVC (25μL for Aβ42 and t-tau, and 75μL for p-tau181) were added to the wells of the antibody-coated plate, and adequately mixed by carefully tapping the stripholder. After incubating in an incubator (60 minutes at 25°C for Aβ42, 16h at 25°C for t-tau, 16h at 4°C for p-tau), all strips were washed 5 times. Add 100μL Conjugate working solution 2 to each well, and incubate in an incubator at 25°C (30 minutes for Aβ42 and t-tau, 60 minutes for p-tau). After washing each well 5 times, add 100μL substrate working solution and incubate for 30 minutes at 25°C in the dark. To stop the reaction, add 50μL Stop Solution to each well and tap the stripholder carefully to ensure optimal mixing. Read the absorbance at 450 nm and calculate the concentration of Aβ42, t-tau and p-tau181.

Blood biomarkers assessment

Serum GFAP and NEFL were quantified by commercial-available Single Molecular Immunity Detection kits (Astrabio, R14060 and R14040). All measurements were performed on the AST-Sc-Lite analyzer (Astrabio) and according to the manufacturer's instructions. Briefly, 25 μL of serum sample was added to an incubation tube, followed by the addition of 25 μL of Reagent 1 which contained 0.1 mg/mL of magnetic beads coated with capture antibodies for GFAP and NEFL. The mixture was then mixed rapidly and incubated at 40°C for 6 minutes. Afterward, Reagent 2, containing detection antibodies labeled with single-molecule imaging fluorophores, was added and mixed, and the mixture was incubated at 40°C for 4 minutes. Subsequently, the reaction mixture was transferred to a flow-cell with a 2*2 mm (width*height) channel for magnetic beads manipulation and imaging. The magnetic beads in the mixture were then absorbed onto the surface of the channel in the flow-cell with the assistance of a permanent magnet. The unlabeled fluorophores were then eliminated by a gentle washing flow of wash buffer and fluorescent images were taken with an integrated fluorescent microscope. The single-molecule signals were analyzed by the machine and protein concentrations were calculated with a standard curve prepared in advance.

Untargeted proteome analysis

Immunodepleting was implemented before digestion to increase the depth of the CSF and serum proteomes. Briefly, 100 μL and 175 μL depletion resin (Thermo Scientific, A36372) were mixed with 100 μL CSF (1:1 CSF/resin volume ratio) and 4 μL serum (4:175 serum/resin volume ratio), respectively, and incubated at room temperature (RT) for 20 min. The mixture was centrifuged at 1000 × g for 2 min to collect the flow-through, which was concentrated by centrifugation at 12000

× g for 30 min in 3k MWCO columns with a molecular weight cutoff (MWCO) of 3 kDa (Thermo Scientific, Cat # 88512). We then added 6 M urea/2 M thiourea (Sigma-Aldrich, Cat # T8656-500G) to the same columns to exchange the buffer system for protein denaturation, and centrifuged at 12000 g until <50 µL solution remained in the chambers. The protein concentrations of depleted CSF and serum samples were determined by a bicinchoninic acid Protein Assay Kit (BCA, Sigma-Aldrich, BCA1 AND B9643) according to the manufacturer's instructions. Concentrated depleted samples were then reduced by 10 mM tris-2(-carboxyethyl)-phosphine (TCEP, Adamas-beta, Cat # 61820E) for 40 min at 32 °C, followed by alkylation with 40 mM Iodoacetamide (IAA, Sigma-Aldrich, Cat # I6125) for 40 min at RT in the dark. The solution of 6 M urea/2 M thiourea, TCEP and IAA were prepared with 100 mM TEAB to make sure the reaction system at a pH of 8.5. Then, we diluted the system with 150 µL 100mM TEAB to make the final concentration of urea/thiourea below 1.5M/0.5M for LysC digestion with 2.5 µg LysC (Hualishi Tech, Cat # HLS LYS002C) for 4 h in 1st step of digestion. Then, we further diluted the reaction system with 50 µL 100mM TEAB and added 2.5 µg trypsin (Hualishi Tech, Cat # HLS TRY001C) for 12 h in 2nd step of digestion. Similarly, serum samples were digested with 0.5 µg LysC and 0.625 µg trypsin in the same way as CSF samples. After being acidified with trifluoroacetic acid (TFA, Thermo Fisher Scientific, Cat # 85183) at a 1% final concentration, peptides were then desalted using 2 mg Solaµ HRP columns (Thermo Scientific, Cat # 60209-001) and the eluate was dried using a SpeedVac.

We performed the batch design to equally deposit samples obtained from individuals with different diseases in the same batch for CSF and serum in parallel. In this way, we designed seven batches, and each batch contained 15 samples and one pooled sample. For serum samples, 7 µg of peptide from each sample was labeled with 56 µg of Tandem mass tags (TMT) 16plex reagent (Thermo Fisher Scientific™, San Jose, USA, Cat # A44520) according to the manufacturer's instructions. Due to the lower peptide yield of CSF, 5 µg peptide from each CSF sample was labeled with 40 µg TMT16plex reagent. After 1 h of incubation at RT, the TMT labeling reaction was quenched by hydroxylamine. We utilized TMT16plex-126 to label a pooled serum peptide sample of 49 µg and a pooled CSF peptide sample of 35 µg, which were both produced by mixing equal amounts of peptide from all the serum or CSF samples, followed by equal division into seven batches after labeling quenching, respectively. In this way, 15 labeled samples in the same batch and a labeled pool were combined and desalted using C18 columns (Waters, Sep-Pak Vac tC18 1cc, 50 mg, WAT054960). Fractionation was performed on a Thermo µLtimate Dinex 3000 (Thermo Fisher Scientific™, San Jose, USA) equipped with an XBridge Peptide BEH C18 column (300A, 5 µm × 4.6 mm × 250 mm) (Waters, Milford, MA, USA). Batches were separated using a 60 min gradient from 5% to 35% acetonitrile (ACN) in 10 mM ammonia (pH=10.0) at a flow rate of 1 mL/min to 60 fractions. We further combined the fractions of equal distance (1st and 31st, 2nd and 32nd, …,30th and 60th) to 30 fractions and used speedvac to dry the samples.

Dried peptide powder was re-dissolved in 2% ACN/0.1% formic acid (FA, Thermo Fisher Scientific, Cat # A117-50). Peptide samples were centrifuged at 15000 g for 15 min and then the

supernatants were transferred to sample vials, followed by analysis on LC-MS/MS. The MS raw data were searched by Proteome Discoverer (Version 2.4.1.15, Thermo Fisher Scientific) against a fasta file of Human proteins downloaded from https://www.uniprot.org/ on 7th May 2020, containing 20377 reviewed entries. Precursor ion mass tolerance was set to 10 ppm, and product ion mass tolerance was set to 0.02 Da. Other parameters are kept as default, including the FDR of 1% for PSM level, peptide level and protein level. The target-decoy strategy was setting Automatic, and the software checks whether all searches were validated in same mode in the processing step. The Grouped abundance ratio of 15 samples to pooled sample in the same batch was selected as the intensity of proteins in the protein matrix for the following statistical analysis.

Quality control of TMT-based proteome data

We randomly selected one CSF sample and one serum sample from each type of patient as a biological replicate, and randomly distributed these seven biological replicates into seven batches to control the quality of the proteome discovery workflow. Our data performed a high degree of consistency and reproducibility with the median coefficients of variations (CVs) for 7 biological replicates all below 0.16 (**Figure S11A, B**). Visualization of pooled serum and CSF peptide samples showed minimal batch effects (**Figure S11C, D**).

Targeted proteome analysis

Peptide samples were prepared in the same way as described in the previous proteomic section except that no depletion or TMT labeling was performed. A nanoflow DIONEX UltiMate 3000 RSLCnano System (Thermo Fisher Scientific™, San Jose, USA) coupled with a Q Exactive HF hybrid Quadrupole-Orbitrap (Thermo Fisher Scientific™, San Jose, USA) was applied for the parallel reaction monitoring (PRM) experiment. For each PRM acquisition, 0.5 μg of peptides was injected. The CSF peptide digests were separated at a flow rate of 300 nL/min (precolumn, 3 μm, 100 Å, 20 mm*75 μm internal diameter; analytical column, 1.9 μm, 120 Å, 150 mm*75 μm internal diameter.) with a 60 min effective gradient (from 10% to 30% buffer B). Buffer A was HPLC-grade water containing 2% ACN and 0.1% FA, and buffer B was 98% ACN containing 0.1% FA. The serum peptide digests were separated with 30 min effective gradient (from 10% to 30% buffer B). The resolution values for the full MS and PRM were 60,000 (at *m/z* 200) and 30,000 (at *m/z* 200), respectively. The automatic gain control (AGC) target was set to 2e5, and the maximum IT was set to 80 ms for PRM setting.

The PRM data were manually analyzed with Skyline.[2] The retention time was predicted by the common internal retention time (CiRT) peptides,[3] and the isolation time window was set to 5 min. The mass analyzer for MS1 and MS/MS was set to "Orbitrap", with a resolution power value of 60,000 and 30,000, respectively. We selected the top 6 peptides in terms of abundance for each target protein for detection. Based on the preliminary screening results, we selected 1-2 peptides for final detection for each protein. The final peptides should meet the following empirical criteria mentioned in the previous literature: [4,5] 1) accurate mass (mass error for precursors < 10 ppm, for fragments < 20 ppm), 2) good peak shape of the peak groups, 3) high abundance, 4) retention time within the

predicted range, 5) matching conditions for the abundance ratio of fragment ions with the library. Proteins that do not contain peptides that meet the above conditions are excluded. After selection, a total of 120 peptides including 15 CiRT peptides (**Table S3**) were included in the CSF PRM experiment while 52 peptides including 13 CiRT peptides (**Table S3**) were included in the serum PRM experiment. The value of total area fragment was exported as peptide relative quantitative data, and then converted to protein quantitative data using ProteomeExpert.[6]

REFERENCES

1. Ye, L.Q., Gao, P.R., Zhang, Y.B., et al. (2021). Application of Cerebrospinal Fluid AT(N) Framework on the Diagnosis of AD and Related Cognitive Disorders in Chinese Han Population. Clinical interventions in aging ***16***, 311-323, 10.2147/CIA.S294756.
2. MacLean, B., Tomazela, D.M., Shulman, N., et al. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics ***26***, 966-968, 10.1093/bioinformatics/btq054.
3. Parker, S.J., Rost, H., Rosenberger, G., et al. (2015). Identification of a Set of Conserved Eukaryotic Internal Retention Time Standards for Data-independent Acquisition Mass Spectrometry. Molecular & cellular proteomics : MCP ***14***, 2800-2813, 10.1074/mcp.O114.042267.
4. Picotti, P., and Aebersold, R. (2012). Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nature methods ***9***, 555-566, 10.1038/nmeth.2015.
5. Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. Molecular systems biology ***4***, 222, 10.1038/msb.2008.61.
6. Zhu, T., Chen, H., Yan, X., et al. (2021). ProteomeExpert: a Docker image-based web server for exploring, modeling, visualizing and mining quantitative proteomic datasets. Bioinformatics ***37***, 273-275, 10.1093/bioinformatics/btaa1088.

Figure S1. Heatmap of the 3238 CSF proteins quantified by TMT-based LC-MS/MS analysis in the discovery cohort.

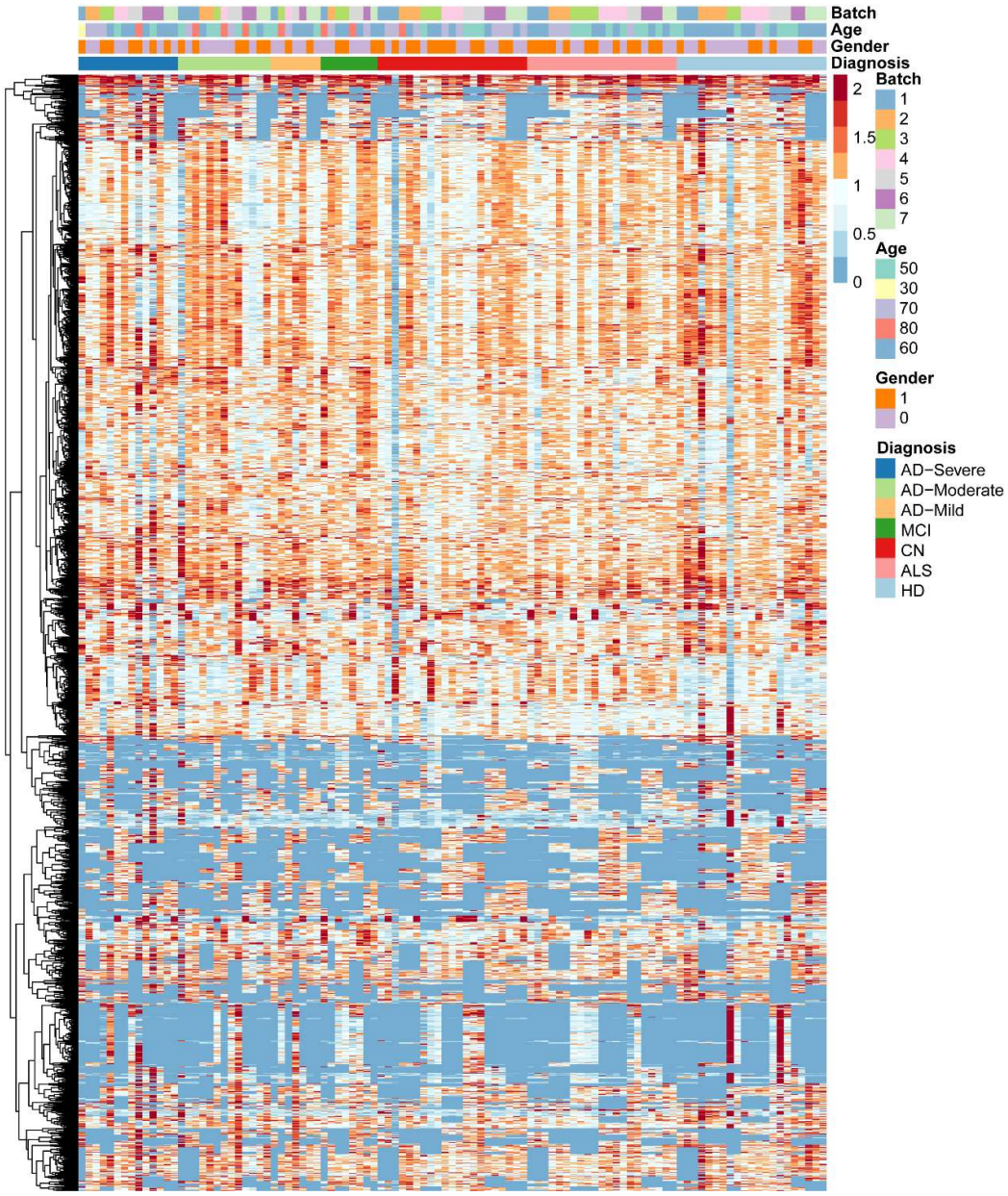

Figure S2. Heatmap of the 1702 quantified serum proteins by TMT-based LC-MS/MS analysis in the discovery cohort.

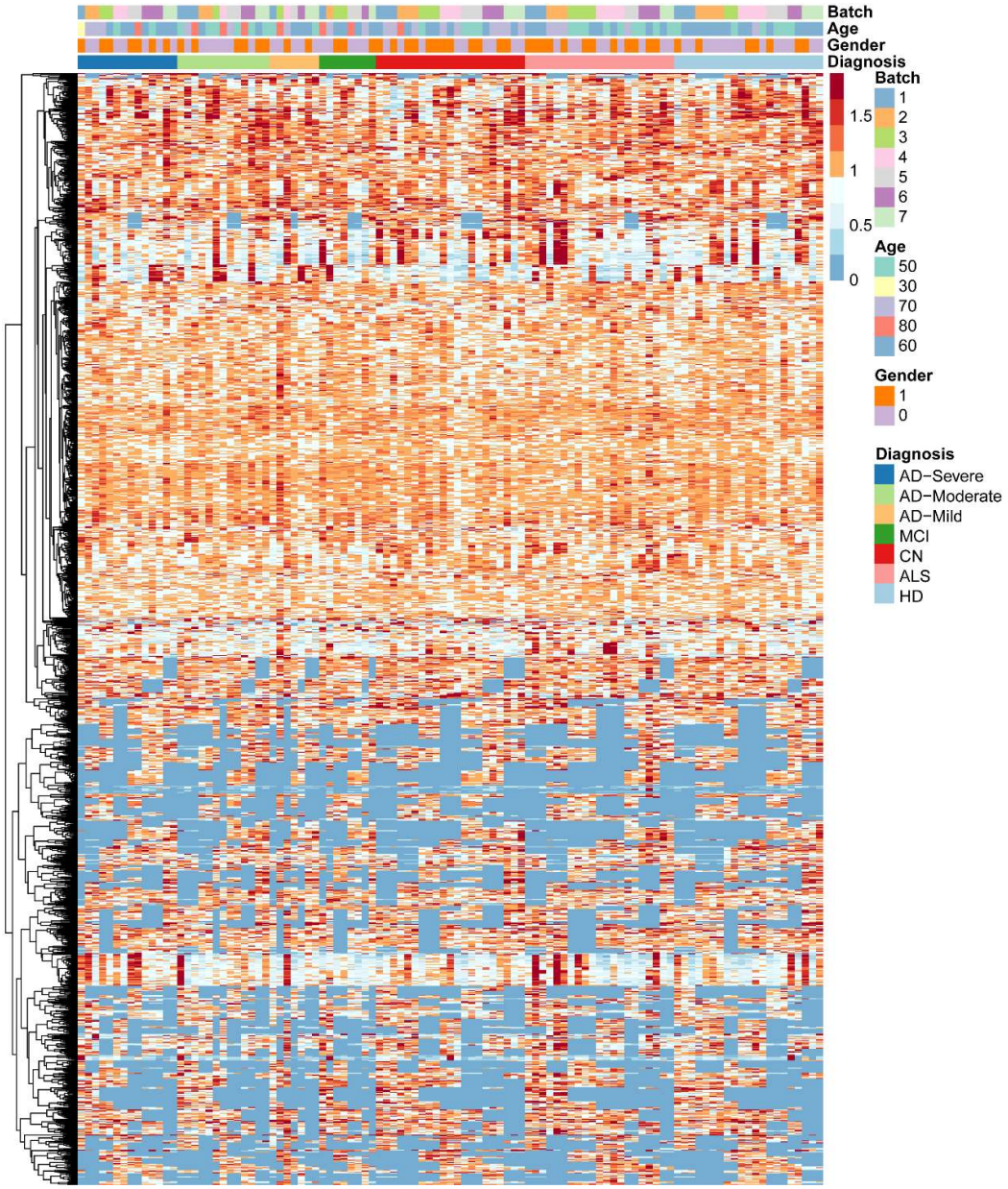

Figure S3. The ranking diagrams of CSF and serum proteins. **(A)** The average intensity of the 3238 CSF proteins identified by TMT-based proteomics was plotted with rank. All identified CSF proteins in this study are in blue for reference. The 19 CSF proteins selected by machine learning model were highlighted in red. **(B)** Concentration range of blood proteins. All identified serum proteins in this study are in blue for reference. The 8 serum proteins selected by machine learning model were highlighted in red.

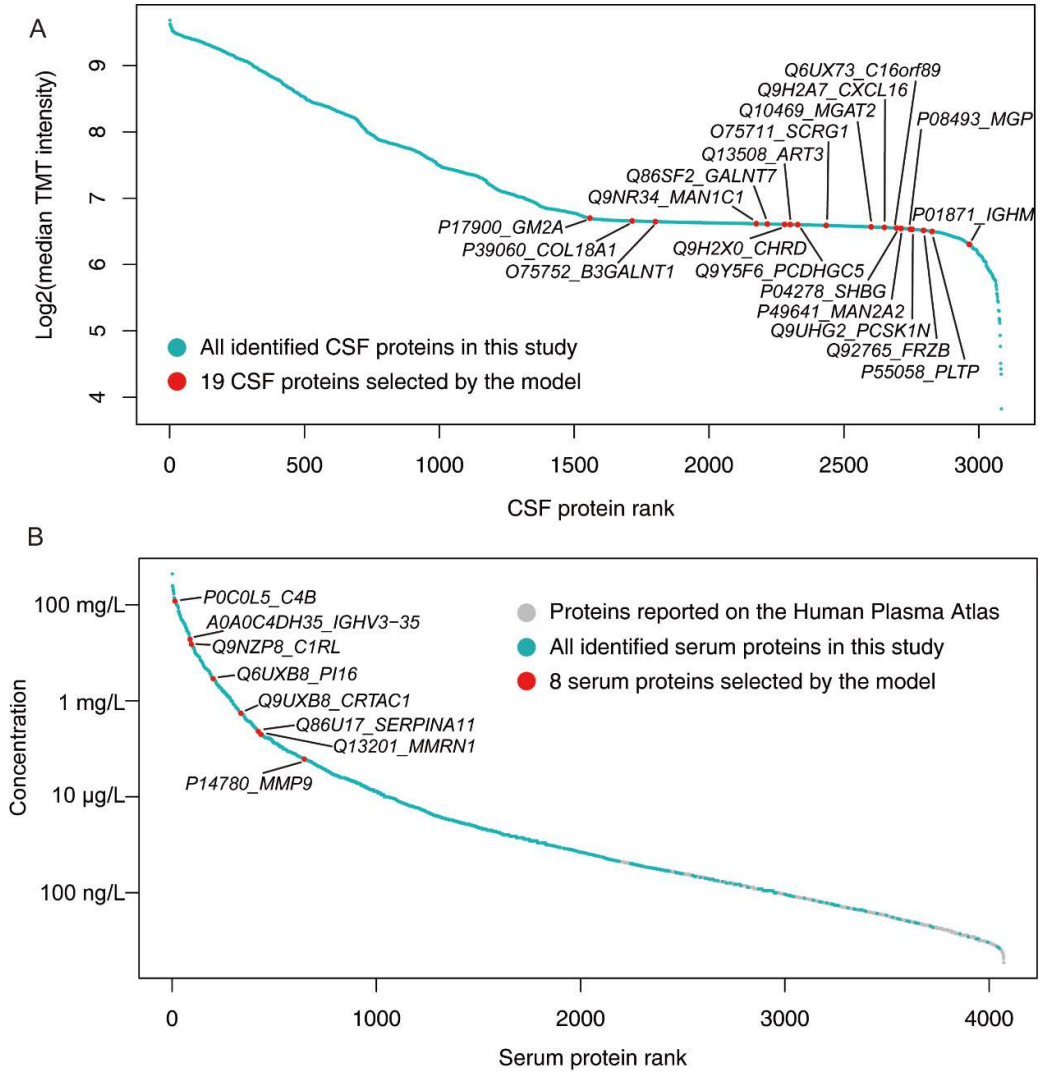

Figure S4. Validated selected differentially expressed proteins by real-time PCR. The transcription levels of the selected differentially expressed proteins (GFAP and GM2A) were evaluated by real-time PCR using the cortex of 5xFAD mice. The transcription levels of GFAP and GM2A were elevated in FAD as compared to wild type (WT) mice, which coincided with the results identified by TMT-based LC-MS/MS analysis. p-value: *, < 0.05; **, < 0.01; ***, < 0.005.

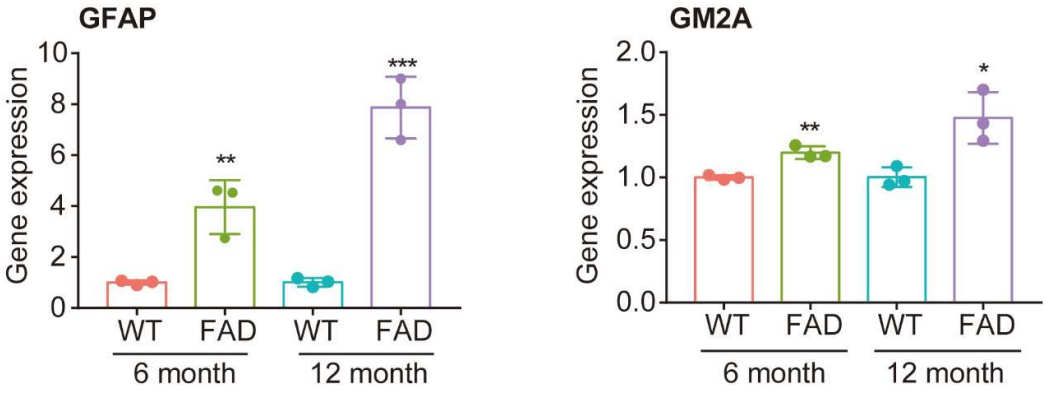

Figure S5. Evaluation of the protein biomarker panels to discriminate other neurodegenerative diseases by uniform manifold approximation and projection (UMAP). **(A)** The 19-protein CSF biomarker panel could well discriminate the MCI from FTD, ALS and HD patients. **(B)** The ability to discriminate the MCI from FTD, ALS and HD was relatively low by the 8-protein serum biomarker panel.

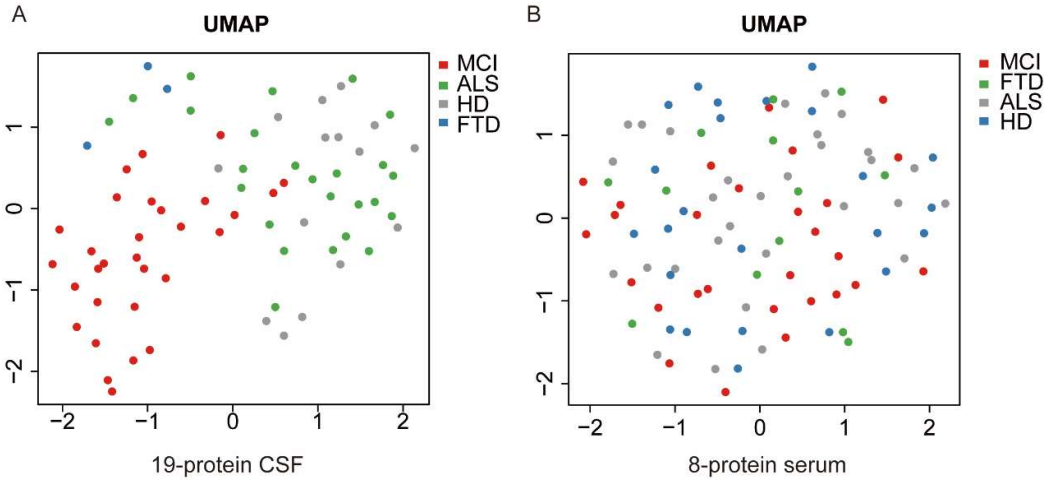

Figure S6. Evaluation of serum GFAP and NEFL to discriminate MCI due to AD and other neurodegenerative disease. (A) The expression level change of serum GFAP and NEFL detected by Single Molecular Immunity Detection kits in the CN controls, MCI due to AD, ALS and FTD. **(B)** Receiver operating characteristic (ROC) analysis was utilized to evaluate the efficacy of serum GFAP and NEFL in distinguishing between MCI due to AD and CN controls. **(C)** Evaluation of the serum GFAP and NEFL to discriminate other neurodegenerative diseases by uniform manifold approximation and projection (UMAP).

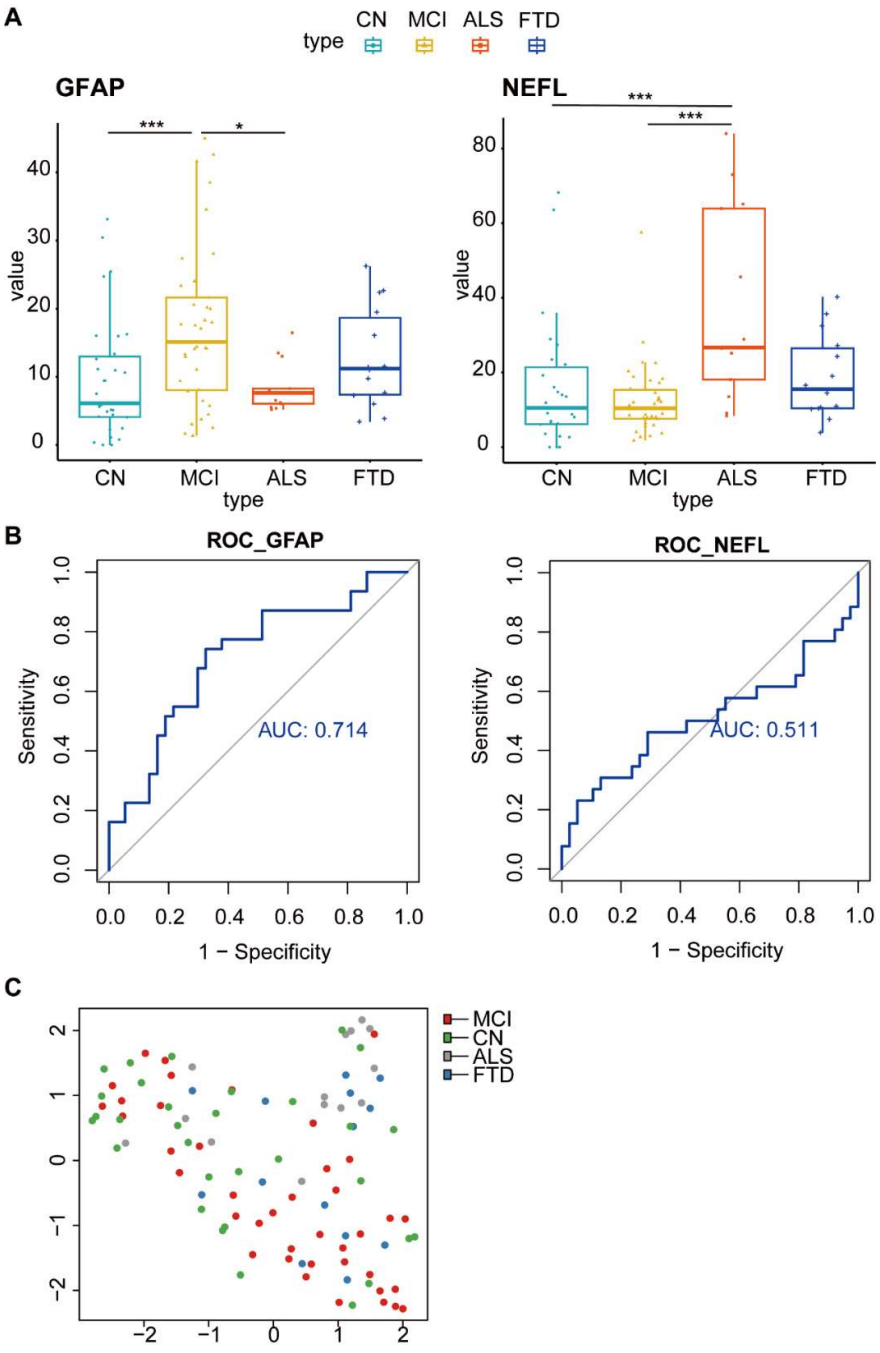

Figure S7. Correlations of the levels of CSF and serum core dysregulated proteins in MCI with the levels of CSF $A\beta_{42}$, t-tau and p-tau_{181}.

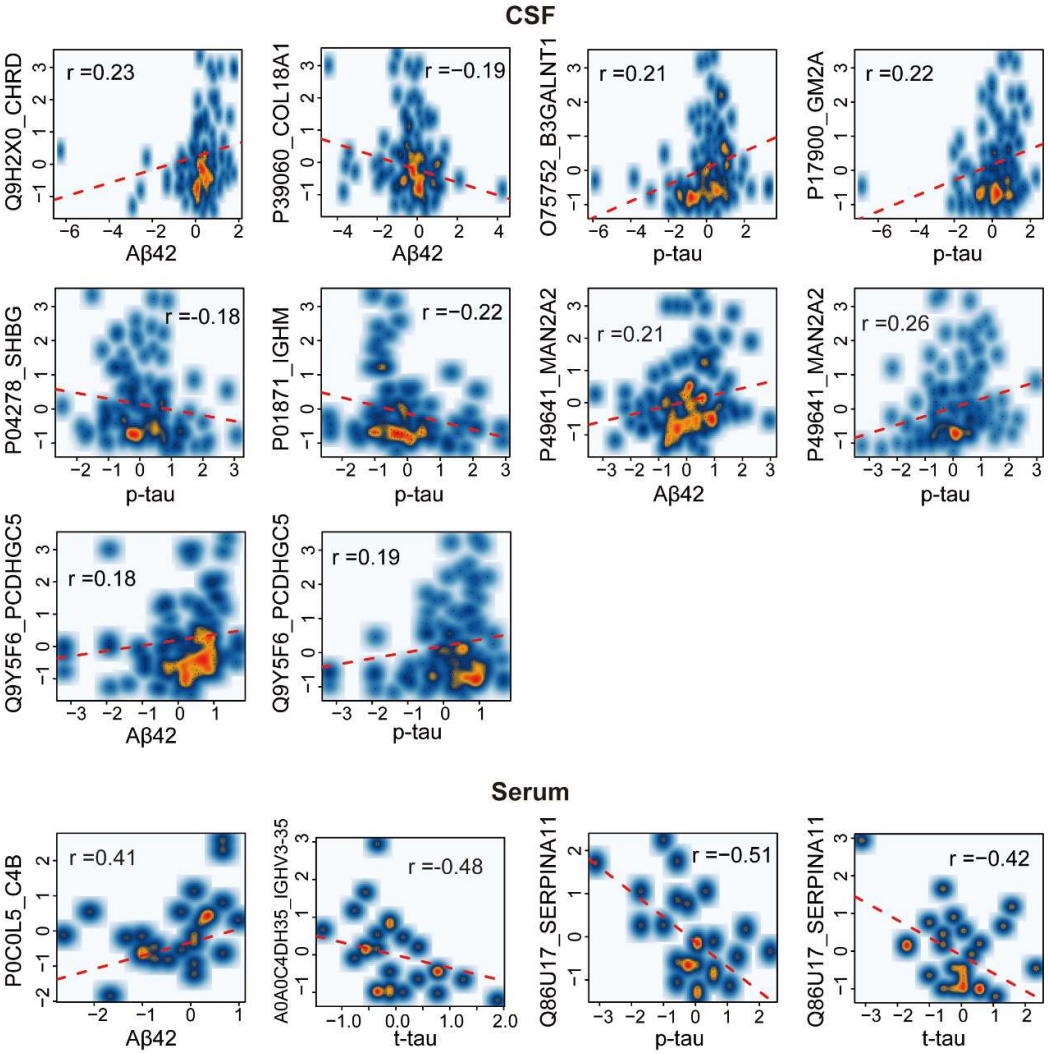

Figure S8. Eight clusters of 2461 CSF proteins identified with the Mfuzz analysis. Among them, proteins in cluster 7 and cluster 8 showed the same regulatory trend with disease progression.

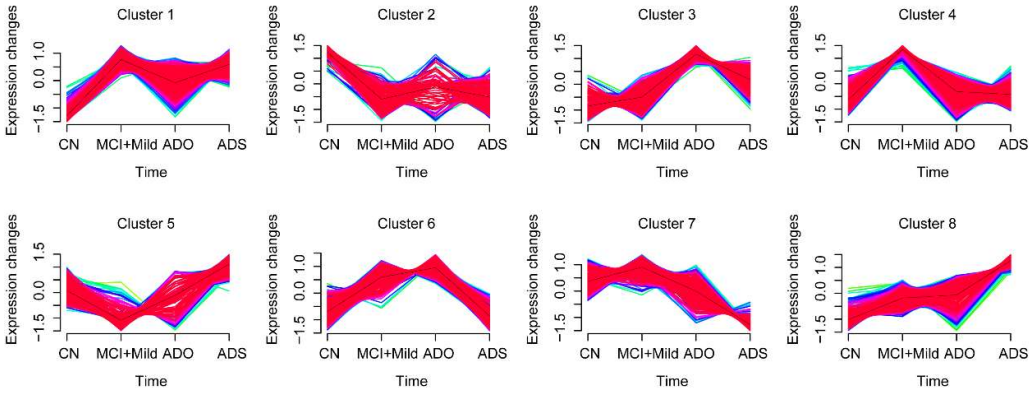

Figure S9. Network pathway analysis of dysregulated CSF proteins during the different stages of AD reveals several important molecular pathways.

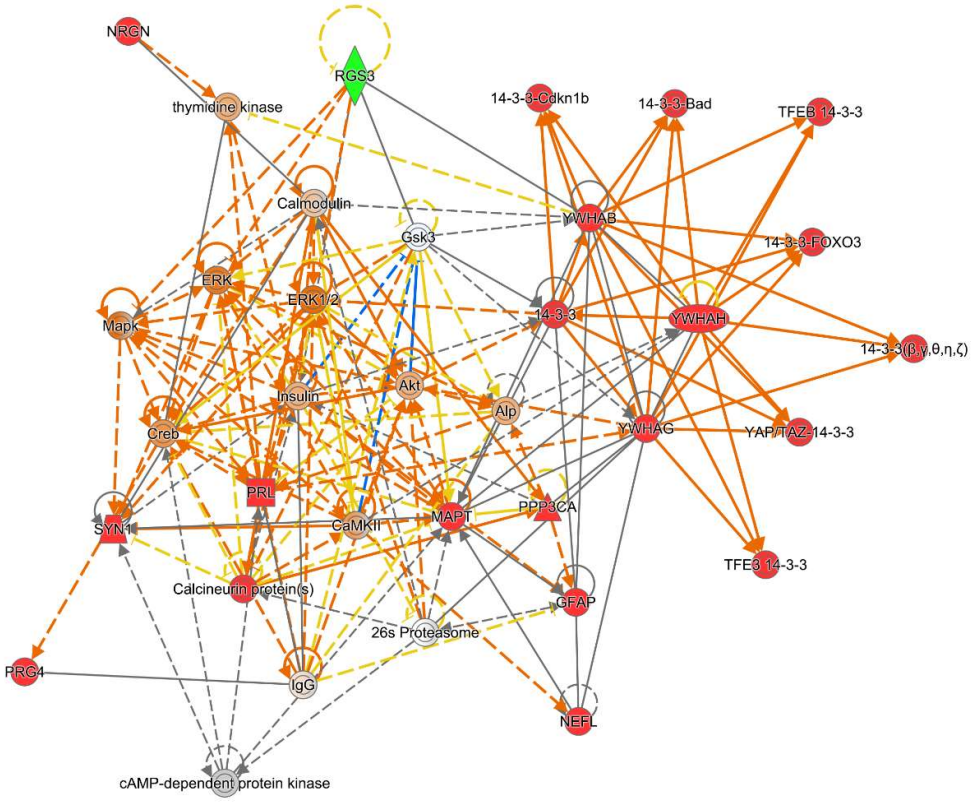

Figure S10. Eight clusters of 1330 serum proteins identified with the Mfuzz analysis. Among them, proteins in cluster 5 and cluster 6 showed the same regulatory trend with disease progression.

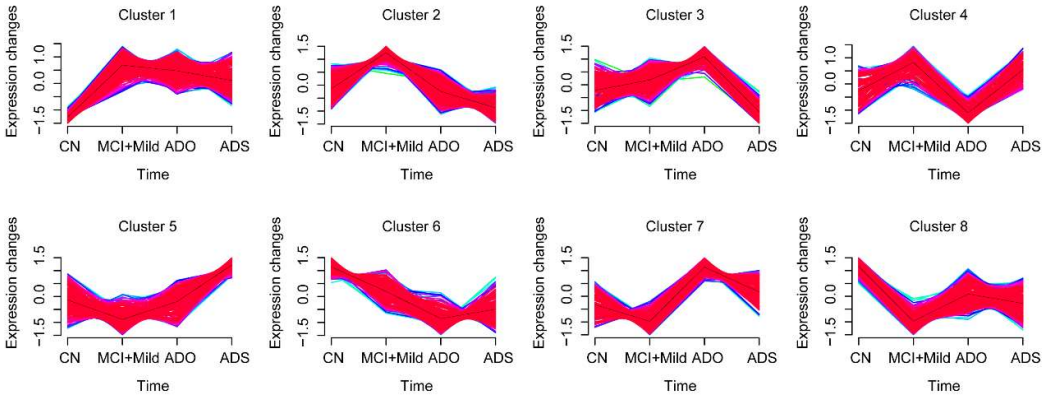

Figure S11. Network pathway analysis of dysregulated serum proteins during the different stages of AD reveals several important molecular pathways.

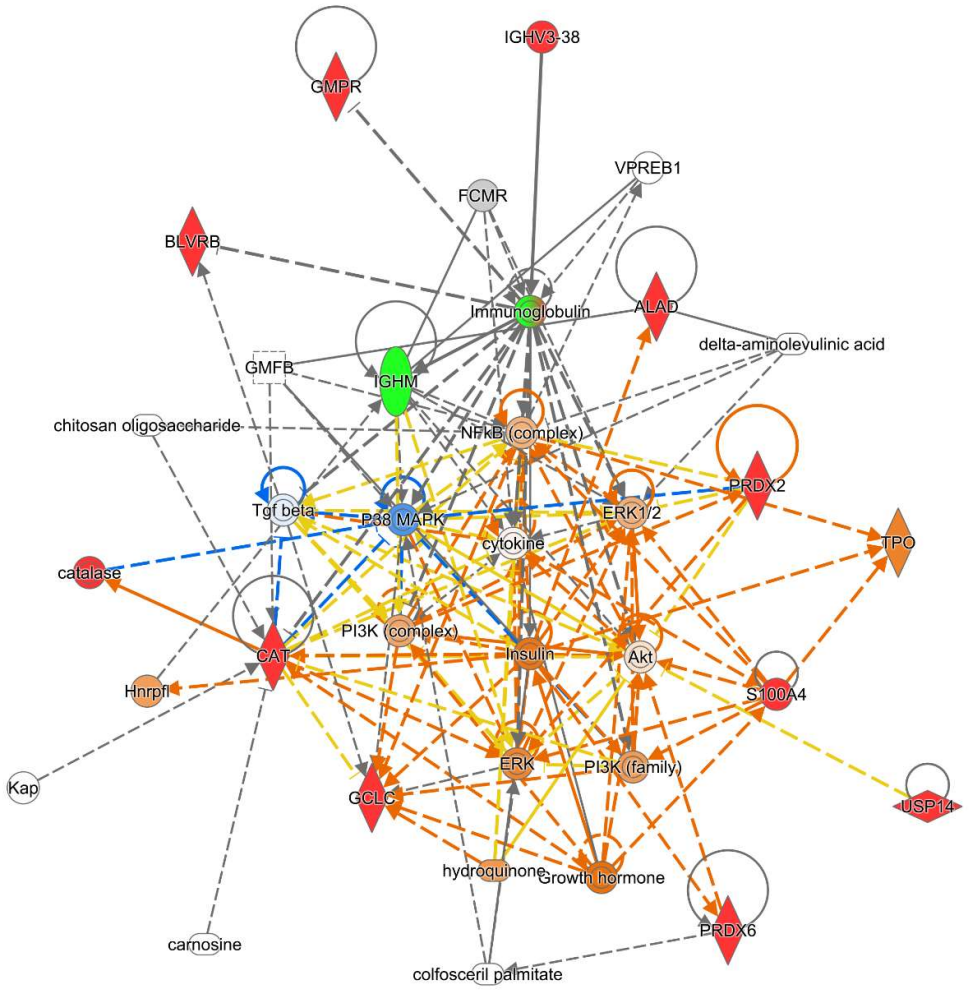

Figure S12. Quality control of CSF and serum proteomic data acquired by TMT-based LC-MS/MS analysis. The coefficient of variation (CV) between the pairing of 7 replicates for A) CSF and B) serum samples. Evaluation of batch effects using the pooled C) CSF and D) serum samples.

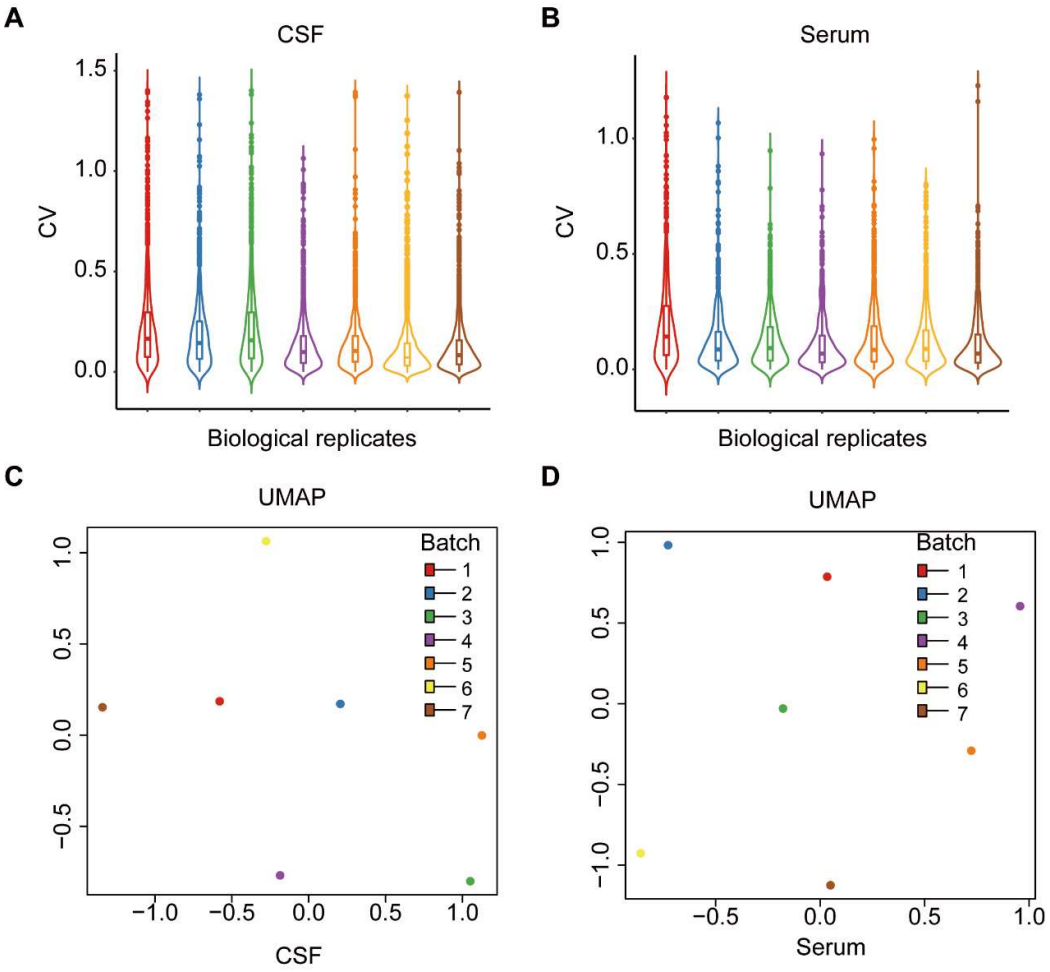

Figure S13. **Quality control of PRM proteome data**. The CVs of 11 and 16 technical replicates in **(A)** CSF and **(B)** serum samples, respectively. The Pearson correlation coefficient (r) for **(C)** 8 CSF pooled samples and **(D)** 10 serum pooled samples. Evaluation of batch effect using the pooled **(E)** CSF and **(F)** serum samples.