**Supplementary Material**

# **Improved Differential Diagnosis of Alzheimer’s Disease by Integrating ELISA and Mass Spectrometry-Based Cerebrospinal Fluid Biomarkers**

# **SAMPLE HANDLING AND ANALYSIS**

## *Sample collection and handling*

 The CSF samples were collected via lumbar puncture into polypropylene tubes. After collection, the CSF was directly centrifuged at 1300 g for 10 min at +4°C to pellet any cell debris. After centrifugation, all samples were visually inspected for blood contamination, frozen and stored at -80°C.

## *Enzyme-linked immunosorbent assay (ELISA)*

 The concentrations of Aβ42, t-tau, and p-tau in CSF were measured at the Clinical Neurochemistry Laboratory, University of Gothenburg, Mölndal, Sweden. The measurements were performed by board-certified laboratory technicians, who were blinded to clinical data, using sandwich ELISAs (INNOTEST, Fujirebio, Ghent, Belgium) and procedures accredited by the Swedish Board of Accreditation and Conformity Assessment.

## *Chemicals and reagents*

 Acetonitrile (ACN), methanol (MeOH), acetic acid (HAc), formic acid (FA), and ammonium bicarbonate (NH4HCO3) were obtained from Merck (Darmstadt, Germany). Acetone, protease inhibitor cocktail, phosphate buffered saline (PBS), trifluoroacetic acid (TFA), sodium citratemonobasic, sodium dodecyl sulphate (SDS), and chicken ovalbumin were purchased from Sigma Aldrich (St. Louis, MO, USA). For tryptic digestion, iodoacetamide (IAA), urea and dithiothreitol (DTT) were obtained from Sigma Aldrich and trypsin/Lys-C mixture (mass spectrometry grade; Promega, Mannheim, Germany). Ultrapure water was prepared by Milli-Q water purification system (Millipore, Bedford, MA, USA).

## *Multiaffinity immunodepletion*

 Prior to nanoLC-MS/MS analysis, the low abundant proteins were enriched through depletion of the seven most highly abundant proteins (albumin, IgG, alpha-1-antitrypsin, IgA, haptoglobin, transferrin, and fibrinogen) using a human Multiple Affinity Removal Spin Cartridge - Hu-7 (Agilent Technologies, Palo Alto, CA, USA). Immunodepletion using the MARS spin cartridge was performed according to the manufacturer’s instruction. Prior to depletion, an aliquot of 310 µL of each CSF sample was filtered through a 0.22 μm cellulose acetate spin filters (Agilent Technologies, Palo Alto, CA, USA) by centrifugation at 15,000 g for 2 min. To minimize any proteolytic degradation during the depletion, the samples were kept at 6oC whenever possible. Next, an aliquot of 300 µL of each CSF sample was loaded onto the spin cartridge and the flow-through (FT) fraction was collected by centrifugation for 2 min at 100 g. Two successive wash steps with 400 µL of MARS-7 Buffer A were carried out to obtain maximum yield. The flow-through and wash (W) fractions were combined. The spin cartridge was washed with 2 mL of MARS-7 Buffer B to remove bound proteins and was then re-equilibrated with Buffer A. The remaining fractions (FT+W) were dried using a SpeedVac (Thermo Scientific, Waltham, MA, USA).

## *Protein digestion*

 The proteins in the depleted CSF sample were digested using a trypsin/Lys-C mixture. Briefly, the proteins were re-dissolved in 50 μL of digestion buffer (6 M urea, 100 mM NH4HCO3) and a volume of 10 μL chicken ovalbumin solution (0.05 μg/μL) was added to each CSF sample. 15 μL of 45 mM aqueous DTT was then added to all samples and the mixtures were incubated at 37°C for 2 h to reduce the disulfide bridges. The samples were cooled to room temperature (RT) and 15 μL of 100 mM aqueous IAA was added before incubating the mixtures for an additional 40 min at RT in darkness in order to carabamidomethylate the cysteines. Finally, a volume of 50 μL of 100 mM NH4HCO3 was added to all samples followed by the trypsin/Lys-C mixture dissolved in 100 mM TEAB, yielding a final trypsin/protein concentration of 5% (w/w). The tryptic digestion was performed overnight at 37°C. Prior to MS analysis, the peptides were purified and desalted on Isolute C18 solid phase extraction (SPE) columns (1 mL, 50 mg capacity, Biotage, Uppsala, Sweden) by wetting the matrix in 3x500 μL of 100% ACN, followed by equilibration with 3x500 μL 1% HAc. Using five repeated cycles of loading, the tryptic peptides were absorbed into the media. The column was then washed using 3x1 mL of 1% HAc. Finally, the peptides were eluted in 300 μL 50% ACN and 1% HAc. After desalting, the eluent was dried using a vacuum centrifuge and re-dissolved in 60 μL 0.1% TFA prior to nano-LC-MS/MS analysis.

## *NanoLC-MS/MS analysis*

 The nanoLC-MS/MS analyses were performed using a 7 T hybrid LTQ FT mass spectrometer (ThermoFisher Scientific, Bremen, Germany) fitted with a nano-electrospray ionization (ESI) ion source. On-line nanoLC separations were performed using an Agilent 1100 nanoflow system (Agilent Technologies, Waldbronn, Germany). The peptide separations were done on in-house packed 15 cm fused silica emitters (75-μm inner diameter, 375-μm outer diameter). The emitters were packed with a methanol slurry of reversed phase, fully end-capped Reprosil-Pur C18-AQ 3 μm resin (Dr. Maisch GmbH, Ammerbuch Entringen, Germany) using a PC77 pressure injection cell (Next Advance, Averill Park, NY, USA). Five µL of sample was injected into the column (corresponding to 2 µg tryptic peptides) and a flow rate of 200 mL/min was used with mobile phase A being MilliQ water with 0.5% acetic acid and mobile phase B 89.5% acetonitrile, 10% water, and 0.5% acetic acid. A 100 min gradient was used starting from 2% to 50% B followed by a 5 min washing with 98% B. The MS analyses were performed in unattended data-dependent acquisition mode, in which the MS automatically switched between acquiring a high-resolution survey mass spectrum in the FTMS (resolving power 50,000 FWHM) and consecutive low resolution, collision-induced dissociation fragmentation of up to five of the most abundant ions in the ion trap. The samples were run according to a random order and in addition, a pool of CSF was run after every eight biological samples for the purpose of quality control.

## *Data analysis*

### Mass spectrometry identification and quantification

 The raw MS data were converted to an open source format (mzML) using “msconvert” from ProteoWizard [1] and processed using the following workflow on OpenMS platform [2] (the parameters are found in S2 Table). For identification the UniProt/Swiss-Pro human database (release 2017\_02, containing 20168 entries) with ovalbumin chicken protein sequence added to the database combined with a decoy database (the sequences were reversed) was used in the MS-GF+ search engine. The results were imported into Percolator and the peptide matches were used in Fido [3] to score proteins based on peptide-spectrum matches. The proteins with q-value lower than 0.05 were selected for subsequent analysis. For quantification, the mzML files were used in “FeatureFinderCentroid” to find and quantify the features. The resulting features were mapped to identified peptides using “IDMapper” [4]. The retention time shift was corrected for using “MapAlignerIdentification” and features were matched using “FeatureLinkerUnlabeledQT” [4]. The “IDConflictResolver” was used to filter the identifications such that each feature was associated to only one peptide. Finally, peptide abundances were normalized using “ConsensusMapNormalizer” and aggregated to protein abundances using “ProteinQuantifier”, in which the intensity of top three most abundant peptides (with q-value lower than 0.05) were summed and the result was imported to R [5] for further processing. Proteins identified (q-value<0.05) with five or more significant peptides (q-value<0.05) were included in downstream analyses. The proteins with less than 80% missing values in the QC samples and the biological samples and coefficients of variation (CV) <1 in QC samples were selected for modeling. The data was transformed to log2 scale and normalized using cyclic loess normalization on protein level [6]. The levels of MS based proteins as well as Aβ42, t-tau, p-tau, and MMSE were adjusted for age and gender using linear regression (“removeBatchEffect” package in Limma [7]).

## Multivariate modelling to diagnose AD using Aβ42, t-tau, and p-tau at baseline

 We evaluated if PLS-DA could improve the accuracy of diagnosing AD, MCI, FTD, and non-demented controls by taking combinatorial effects of Aβ42, t-tau, and p-tau into account. A linear model using PLS-DA (mixOmics R package [20]) was trained using the three components for prediction. Importantly, for training, MCI/AD converters and MCI/non-AD converters were regarded as a single group, “MCI”. Therefore, the model was not provided with information on whether the MCI patients were AD converters or not. A leave-one-out cross-validation (“perf” function in mixOmics) was performed to evaluate the accuracy by calculating an area of the ROC curve (AUROC) for AD versus non-AD subjects (stable MCI, FTD and controls), MCI/AD converters versus non-AD subjects (FTD, stable MCI and controls), FTD versus non-FTD subjects (AD, stable MCI and controls), and controls versus all other groups (cognitively declined patients).

## Integrative multivariate statistical analysis

 Using sparse PLS-DA (sPLS-DA), we evaluated if a combination of Aβ42, t-tau, and p-tau levels with levels of proteins evaluated by MS (proteomics data) could improve the diagnostic performance. The training and evaluation of the model, including leave-one-out cross validation, was the same as for the PLS-DA model described above. The variables of importance (VIP) of the four most important variables (proteins) for each of the four components used for the diagnostic prediction were automatically selected and extracted from the model. The analysis was performed using the mixOmics R package [8].

#

# **REFERENCES**

[1] Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, Gatto L, Fischer B, Pratt B, Egertson J, Hoff K, Kessner D, Tasman N, Shulman N, Frewen B, Baker TA, Brusniak MY, Paulse C, Creasy D, Flashner L, Kani K, Moulding C, Seymour SL, Nuwaysir LM, Lefebvre B, Kuhlmann F, Roark J, Rainer P, Detlev S, Hemenway T, Huhmer A, Langridge J, Connolly B, Chadick T, Holly K, Eckels J, Deutsch EW, Moritz RL, Katz JE, Agus DB, MacCoss M, Tabb DL, Mallick P (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**, 918-920.

[2] Rost HL, Sachsenberg T, Aiche S, Bielow C, Weisser H, Aicheler F, Andreotti S, Ehrlich HC, Gutenbrunner P, Kenar E, Liang X, Nahnsen S, Nilse L, Pfeuffer J, Rosenberger G, Rurik M, Schmitt U, Veit J, Walzer M, Wojnar D, Wolski WE, Schilling O, Choudhary JS, Malmstrom L, Aebersold R, Reinert K, Kohlbacher O (2016) OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat Methods* **13**, 741–748.

[3] Serang O, MacCoss MJ, Noble WS (2010) Efficient marginalization to compute protein posterior probabilities from shotgun mass spectrometry data. *J Proteome Res* **9**, 5346-5357.

[4] Weisser H, Nahnsen S, Grossmann J, Nilse L, Quandt A, Brauer H, Sturm M, Kenar E, Kohlbacher O, Aebersold R, Malmstrom L (2013) An automated pipeline for high-throughput label-free quantitative proteomics. *J. Proteome Res* **12**, 1628–1644.

[5] R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

[6] Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18 Suppl 1**, S96–104.

[7] Smyth GK (2005) Limma: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, eds. Springer, New York, pp. 397–420.

[8] Rohart F, Gautier B, Singh A, Lê Cao KA (2017) mixOmics: an R package for ’omics feature selection and multiple data integration. *PLoS Comput Biol* **13**, e1005752.