

RESEARCH

Open Access



# Deep undepleted human serum proteome profiling toward biomarker discovery for Alzheimer's disease

Kaushik Kumar Dey<sup>1†</sup>, Hong Wang<sup>2†</sup>, Mingming Niu<sup>1</sup>, Bing Bai<sup>1,4</sup>, Xusheng Wang<sup>2</sup>, Yuxin Li<sup>2</sup>, Ji-Hoon Cho<sup>2</sup>, Haiyan Tan<sup>2</sup>, Ashutosh Mishra<sup>2</sup>, Anthony A. High<sup>2</sup>, Ping-Chung Chen<sup>1</sup>, Zhiping Wu<sup>1</sup>, Thomas G. Beach<sup>3</sup> and Junmin Peng<sup>1,2\*</sup>

## Abstract

**Background:** Blood-based protein measurement is a routine practice for detecting biomarkers in human disease. Comprehensive profiling of blood/plasma/serum proteome is a challenge due to an extremely large dynamic range, as exemplified by a small subset of highly abundant proteins. Antibody-based depletion of these abundant proteins alleviates the problem but introduces experimental variations. We aimed to establish a method for direct profiling of undepleted human serum and apply the method toward biomarker discovery for Alzheimer's disease (AD), as AD is the most common form of dementia without available blood-based biomarkers in clinic.

**Methods:** We present an ultra-deep analysis of undepleted human serum proteome by combining the latest 11-plex tandem-mass-tag (TMT) labeling, exhaustive two-dimensional liquid chromatography (LC/LC) fractionation (the 1st LC: 3 h for 180 fractions, and the 2nd LC: 3 h gradient per fraction), coupled with high resolution tandem mass spectrometry (MS/MS). AD ( $n = 6$ ) and control ( $n = 5$ ) sera were analyzed in this pilot study. In addition, we implemented a multiplexed targeted LC-MS3 method (TOMAHAQ) for the validation of selected target proteins.

**Results:** The TMT-LC/LC-MS/MS platform is capable of analyzing 4826 protein components (4368 genes), covering at least 6 orders of magnitude in dynamic range, representing one of the deepest serum proteome analysis. We defined intra- and inter- group variability in the AD and control groups. Statistical analysis revealed differentially expressed proteins in AD (26 decreased and 4 increased). Notably, these altered proteins are enriched in the known pathways of mitochondria, fatty acid beta oxidation, and AGE/RAGE. Finally, we set up a TOMAHAQ method to confirm the decrease of PCK2 and AK2 in our AD samples.

**Conclusions:** Our results show an ultra-deep serum discovery study by TMT-LC/LC-MS/MS, and a validation experiment by TOMAHAQ targeted LC-MS3. The MS-based discovery and validation methods are of general use for biomarker discovery from complex biofluids (e.g. serum proteome). This pilot study also identified deregulated proteins, in particular proteins associated with mitochondrial function in the AD serum samples. These proteins may serve as novel AD candidate biomarkers.

**Keywords:** Alzheimer's disease, Biomarker, Human blood, Plasma, Serum, Mass spectrometry, Proteomics, Proteome, Tandem mass tag

\*Correspondence: junmin.peng@stjude.org

<sup>†</sup>Kaushik Kumar Dey and Hong Wang are co-first authors

<sup>1</sup> Departments of Structural Biology and Developmental Neurobiology,

St. Jude Children's Research Hospital, Memphis, TN 38105, USA

Full list of author information is available at the end of the article



## Background

Cellular and biochemical components in blood play a central role in human physiology and their dynamic levels are considered to correlate with an individual's healthy and diseased states [1, 2]. Blood is an exceptionally complex fluid, comprised of cells (i.e. red and white blood cells and platelets) and plasma (the liquid part) from which serum is collected after removing clotting factors with adequate coagulation. Human plasma/serum contains extraordinary diverse proteins, secreted from all types of cells and tissues for normal physiological function, leaked from damaged cells and tissues especially under disease conditions, or released from infectious organisms. Measuring various protein concentrations in plasma/serum is routine in clinical practice. The concentration dynamic range spans at least 10 orders of magnitude, from the most abundant albumin (~50 mg/ml) to cytokines of low abundance (e.g. 4.2, 7.4 and 11.2 pg/ml for interleukin-6, interleukin-17 and TNF- $\alpha$ , respectively) in normal individuals [3, 4]. This extremely high dynamic range raises a significant challenge for profiling the complete plasma/serum by a proteomics platform, commonly based on liquid chromatography-tandem mass spectrometry (LC-MS/MS). Depletion of highly abundant plasma proteins is often used to alleviate the dynamic range challenge, as the top 22 abundant proteins occupy approximately 99% of the total protein mass [1]. The depletion may be achieved by affinity columns immobilized with antibodies against the top abundant proteins [5–7]. However, there are multiple caveats associated with the depletion method: (i) the antibodies are never completely specific and may remove other nonspecific proteins; (ii) the depletion is performed under non-denaturing condition, leading to co-immunoprecipitation and removal of antigen-bound proteins; and (iii) the depletion step generates significant experimental variations [7].

Advances in mass spectrometry (MS)-based proteomics [8, 9], especially in LC separation power and MS resolution and scan rate, enable the profiling of more than 15,000 proteins (>12,000 genes) from mammalian tissue samples [10, 11]. Protein quantification can be achieved through data dependent acquisition (e.g. label free method and stable isotope labeling) [12], as well as data-independent acquisition [13]. Tandem-mass-tag (TMT) is a commonly used stable isotope labeling method, which allows up to 11-plexed analysis [14, 15]. Although the accuracy of TMT measurement is often affected by ion co-elution-induced ratio compression, this issue is largely addressed by the MS3 method [16] or the combination of extensive LC fractionation, MS optimization, and computational correction [17]. With the success of tissue profiling [18, 19], we attempted to apply this latest TMT-LC/LC-MS/MS technology to analyze

blood-based complex biofluids for Alzheimer's disease (AD) biomarker discovery.

Following the discovery of putative biomarkers, it is necessary to validate these candidates in large clinical cohorts, usually by Ab-based approaches or targeted MS methods [2], such as selected single, multiple and parallel reaction monitoring (SRM, MRM, and PRM, respectively) [20, 21]. More recently, Triggered by Offset, Multiplexed, Accurate mass, high resolution, and Absolute Quantitation (TOMAHAQ) has been reported as an isobaric targeted method [22, 23]. For each targeted peptide quantification, TOMAHAQ implements a synthetic, TMT0-labeled peptide, which is used to trigger the quantification of native target peptide by MS3, based on a pre-selected offset mass. During the generation of MS3 spectra, synchronous precursor selection (SPS) can improve quantification accuracy by selecting pre-defined *b*- or *y* ions in MS2.

Alzheimer's disease is the most common form of dementia and the sixth-leading cause of death in the US, affecting more than 5 million Americans with a healthcare cost of \$236 billion in 2016 [24]. By 2050, AD patients are projected to reach 13.8 million in the US [24] and 100 million worldwide [25]. Currently, AD diagnosis is based on patient's symptoms, memory and behavior tests, brain imaging, as well as post-mortem brain pathological assays [26, 27]. Blood-based biomarkers, however, are not available for AD, and most proposed candidates are derived from known disease mechanisms, such as A $\beta$  and tau [28, 29]. Here we present the unbiased, large-scale profiling of human serum specimens, revealing consistent mitochondrial protein changes between control and AD samples.

## Methods

### Patient sample description

Human blood sera were collected from control ( $n=5$ ) and AD patients ( $n=6$ ), provided by the Brain and Body Donation Program at Banner Sun Health Research Institute, with approval for this study. Clinical and pathological diagnoses were based on established criteria [30]. All subjects consented to the study, and informed consent was obtained from each entrant. After clotting and centrifugation, the sera were frozen and stored at  $-80^{\circ}\text{C}$  in aliquots of polyethylene tubes until use.

### Serum protein extraction and quantification

Human serum proteins were extracted in fresh lysis buffer [50 mM HEPES, pH 8.5, 8 M urea, and 0.5% sodium deoxycholate with  $1 \times$  phosphatase inhibitor cocktail (PhosSTOP, Sigma-Aldrich)]. The protein concentration was measured by the BCA assay (Thermo Fisher Scientific) and confirmed by Coomassie-stained

short SDS gel as previously described [31]. The protein lysates were stored at  $-80^{\circ}\text{C}$  in aliquots before use.

#### Protein digestion and TMT labeling

The digestion and labeling were performed based on an optimized protocol [32, 33]. Quantified protein ( $\sim 0.1$  mg in the lysis buffer with 8 M urea) for each TMT channel was directly digested with Lys-C (Wako, 1:100 w/w) at  $21^{\circ}\text{C}$  for 2 h, diluted four-fold to lower urea concentration to 2 M, and further digested with trypsin (Promega, 1:50 w/w) at  $21^{\circ}\text{C}$  overnight. The digestion was terminated by the addition of 1% trifluoroacetic acid (TFA) with centrifugation. The supernatant was desalted with Sep-Pak C18 cartridge (Waters), and then dried by a speedvac vacuum concentrator. Each sample was re-dissolved in 50 mM HEPES, pH 8.5, reacted with TMT reagents, pooled equally, and desalted again before LC/LC-MS/MS.

#### Extensive LC/LC-MS/MS analysis

The pooled TMT labeled peptides were resolved by offline basic pH reverse phase LC, and acidic pH reverse phase LC coupled with MS/MS analysis [34]. The setting of basic pH LC included a XBridge C18 column ( $3.5\ \mu\text{m}$  particle size,  $4.6\ \text{mm} \times 25\ \text{cm}$ , Waters), buffer A (10 mM ammonium formate, pH 8.0), buffer B (95% acetonitrile, 10 mM ammonium formate, pH 8.0) [33], and a 3 h gradient of 15–35% buffer B. Each fraction was collected every minute, ending with a total of 180 fractions. In the acidic pH LC-MS/MS analysis, each previous fraction was analyzed on a column ( $75\ \mu\text{m} \times 25\ \text{cm}$ , heated to  $65^{\circ}\text{C}$  to reduce backpressure) coupled with a Q Exactive HF Orbitrap mass spectrometer (Thermo Fisher Scientific). Peptides were resolved by a 3 h gradient (buffer A: 0.2% formic acid, 5% DMSO; buffer B: buffer A plus 65% acetonitrile). MS settings included MS1 scans (60,000 resolution,  $1 \times 10^6$  AGC and 50 ms maximal ion time) and 20 data-dependent MS2 scans (410–1600  $m/z$ , 60,000 resolution,  $1 \times 10^5$  AGC,  $\sim 150$  ms maximal ion time, HCD, 32% normalized collision energy, and  $\sim 15$  s dynamic exclusion).

#### Identification and quantification of proteins by JUMP software suite

The bioinformatics processing of identification was carried out with our recently developed JUMP search engine, which combines the advantage of pattern- and tag-dependent scoring to improve sensitivity and specificity [35]. A composite target-decoy database was used to estimate false discovery rate (FDR) [36]. The protein database was generated by combining downloaded Swiss-Prot, TrEMBL, and UCSC databases and removing redundancy (human: 83,955 entries). Major

parameters were precursor and product ion mass tolerance ( $\pm 15$  ppm), full trypticity, two maximal missed cleavage, static mass shift for TMT tags ( $+ 229.16293$  on Lys and N-termini) and carbamidomethyl modification ( $57.02146$  on Cys), dynamic mass shift for oxidation ( $+ 15.99491$  on Met), and three maximal modification sites. The resulting PSMs were filtered by mass accuracy, and then grouped by precursor ion charge state followed by the cutoffs of JUMP-based matching scores (J-score and  $\Delta J_n$ ) to reduce FDR below 1% for proteins. When the same peptide is derived from numerous homologous proteins, the peptide was matched to the protein with the top PSM number, according to the rule of parsimony. The quantification was performed as previously described [17].

#### Calculation of abundance index of identified proteins by PSMs

The absolute protein abundance index of serum proteome was calculated based on previously reported methods [37, 38], using the total number of PSMs matched to a particular protein, normalized by theoretically detectable peptides from the protein. It was derived by the formula: (the number of PSMs/the number of theoretically detectable peptides)  $\times$  a scale factor. The scale factor was set to 5000, which generated abundance indexes that were roughly equivalent to protein copy numbers per cell during deep proteomics analyses.

#### Evaluation of sample variations and principal component analysis

The measurement variation was analyzed according to intra- and inter-group replicates. The ratios of all proteins from the samples were modeled with a Gaussian distribution to evaluate standard deviation (SD). Principal component analysis (PCA) was used to visualize the differences among human disease groups. Relative expression of all proteins was used as input of PCA, using a R statistical analysis package (version 3.4.0) [39].

#### Differential expression (DE) analysis, pathway enrichment and protein–protein interaction (PPI) analysis

DE analysis was determined by student  $t$  test in the following steps: (i) calculating  $p$  values and applying a threshold of 0.05; (ii) filtering by at least 1.5 fold of the standard deviation in the analysis; (iii) manually examining all proteins to remove proteins quantified by only one peptide.

Pathway enrichment analysis was used to infer functional groups of proteins enriched in a given pathway. The analysis was performed using Fisher's exact test ( $p$  value) with the BH correction for multiple testing (BH

FDR). Enriched pathways with  $FDR < 0.05$  were considered statistically significant.

DE proteins were matched to a composite PPI database by integrating STRING (v10) [40], BioPlex [41], and InWeb\_IM [42], including 18,515 proteins and 469,993 PPI connections. Modules in each protein cluster were defined as previously reported [18]. Modules were annotated by Gene Ontology, KEGG or Hallmark.

### TOMAHAQ targeted LC–MS3 analysis

The TOMAHAQ assay was based on the initially reported protocol [22]. Selected peptides were synthesized, purified (at least 95% purity), and dissolved in 20% acetonitrile. The peptides were labeled by a TMT0 reagent (Thermo Fisher Scientific), desalted, and spiked into the TMT11-labeled pooled samples. The amount of TMT0-labeled synthetic peptides was adjusted to ensure detection in MS1.

In the LC–MS3 analysis, the TMT0-TMT11 mixed samples were analyzed on a reverse phase LC coupled with MS3 analysis. The setting included a C18 column ( $50 \mu\text{m} \times 15 \text{cm}$ ,  $1.9 \mu\text{m}$  particle size, heated to  $65^\circ\text{C}$  to reduce backpressure), buffer A (0.2% formic acid, 5% DMSO) and buffer B (buffer A plus 65% acetonitrile) in a 1 h gradient of 10–35% buffer B at 250 nl/min, and an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific). The TOMAHAQ workflow comprises a sequence of decisions to prompt quantitative SPS-MS3 in multiple scans. In scan 1, survey MS1 scans (mass range:  $\pm 50 m/z$  of target peptides, 60,000 resolution,  $1 \times 10^6$  AGC and 100 ms maximal ion time) were used to detect one TMT0 labeled, synthetic trigger peptide ( $\pm 15$  ppm). If the intensity threshold ( $1 \times 10^5$ ) was reached, the trigger peptide was fragmented in scan 2 ( $0.4 m/z$  isolation window, and  $\sim 35$  NCE in CID) and detected by Orbitrap (15,000 resolution;  $1 \times 10^5$  AGC; 50 ms maximal ion time). A “Product Ion Trigger” function was used to compare the trigger peptide MS2 spectra to a pre-determined MS2 product ion list ( $\pm 10$  ppm). If at least 6 product ions were matched, it triggered scans 3 and 4 to analyze the corresponding target peptide, using a pre-selected offset (peptide-specific, e.g.  $5.01 m/z$  for  $z=2$  and two TMT tags). In scan 3, target MS2 was collected ( $0.4 m/z$  isolation window,  $\sim 35$  NCE in CID, 15,000 resolution;  $1 \times 10^5$  AGC; 1000 ms maximal ion time). In scan 4, target MS3 was collected based on the previous MS2 and additional MS3 settings: Precursor Ion Exclusion (Low=70, High=5), Isobaric Tag Loss Exclusion (Reagent Tag Type=TMT to exclude “complement” MS2 ions),  $0.4 m/z$  isolation window for 10 pre-defined MS2 product ions on a “Targeted Mass Inclusion List”, 55 NCE in HCD, 60,000 resolution,  $1 \times 10^5$  AGC, and 2,500 ms maximal ion time.

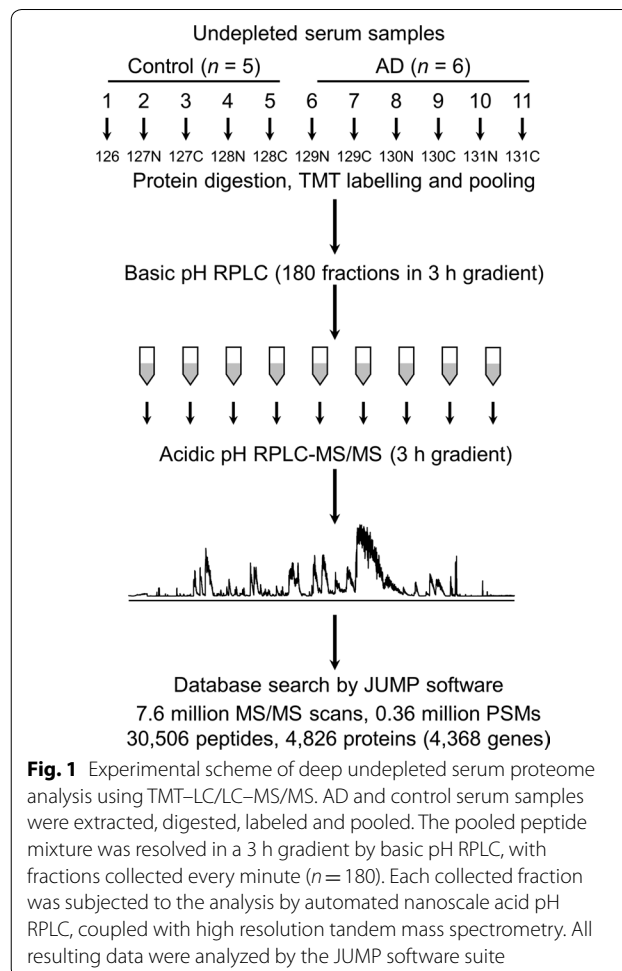
### Availability of data and materials

The mass spectrometry proteomics raw data have been deposited to the Proteome Xchange Consortium (<http://www.proteomexchange.org>) [43] via the PRIDE partner repository with the dataset identifier PXD011482.

## Results and discussion

### Multiplexed quantitative analysis of undepleted human serum proteome

A flowchart of the experiment is presented in Fig. 1, in which we profiled serum proteome without depletion by extensive TMT–LC/LC–MS/MS to maximize sensitivity and proteome coverage. Clinical characteristics of AD patients, together with gender- and age-matched control cases, are summarized in Additional file 1: Table S1. The whole serum protein extracts were denatured and trypsinized into peptides. The resulting peptides were differentially labelled with different TMT tags, and then equally pooled. The pooled peptides were separated by extensive offline basic pH reverse phase (RP) LC, collected into as many as 180 fractions to decrease the complexity in each





fraction. We then used a long, high resolution column and 3 h gradient for acidic pH RPLC (totaling 540 h) coupled to the mass spectrometer. Taking account of maintenance time for LC, we spent approximately one month of instrument time to obtain ultra-deep analysis. In total, we collected 7.6 million MS/MS scans and accepted 0.36 million peptide-spectrum matches (PSMs) after database search, leading to the identification and quantification of 30,506 unique peptides, 4826 proteins with <1% false discovery rate (FDR), corresponding to 4368 genes (Additional file 1: Table S2). The most abundant protein serum albumin (ALB) was identified by 47,006 PSMs (13.1% of all accepted PSMs), and the top 22 proteins occupied 262,008 PSMs (72.2% of the accepted PSMs). In spite of the presence of these exceedingly abundant species, we were still able to identify 3415 (70.8%) proteins with at least two matching peptides, and 3912 (81.1%) proteins with at least two PSMs. To our knowledge, this is one of most comprehensive quantitative analyses of human serum in a single experiment.

#### Estimation of the minimal fraction number to achieve high serum proteome coverage

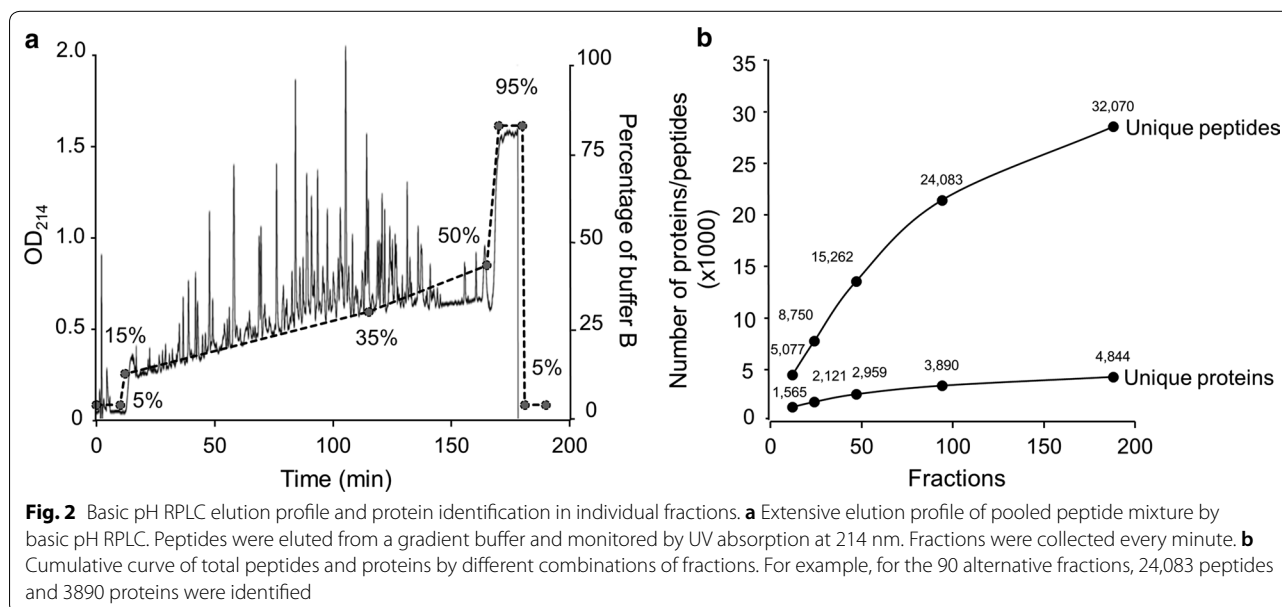
As the serum samples have different protein composition and dynamic range from tissue specimens, we attempted to evaluate the performance of the two dimensional LC and to optimize a strategy for serum proteome analysis. During the 3 h gradient of the 1st dimensional basic pH RPLC (15–35% buffer B, Fig. 2a), the majority of the peptides were eluted between 25 and 165 min. In this pilot analysis, we did not concatenate the 180 fractions because mixing fractions regenerates peptide complexity.

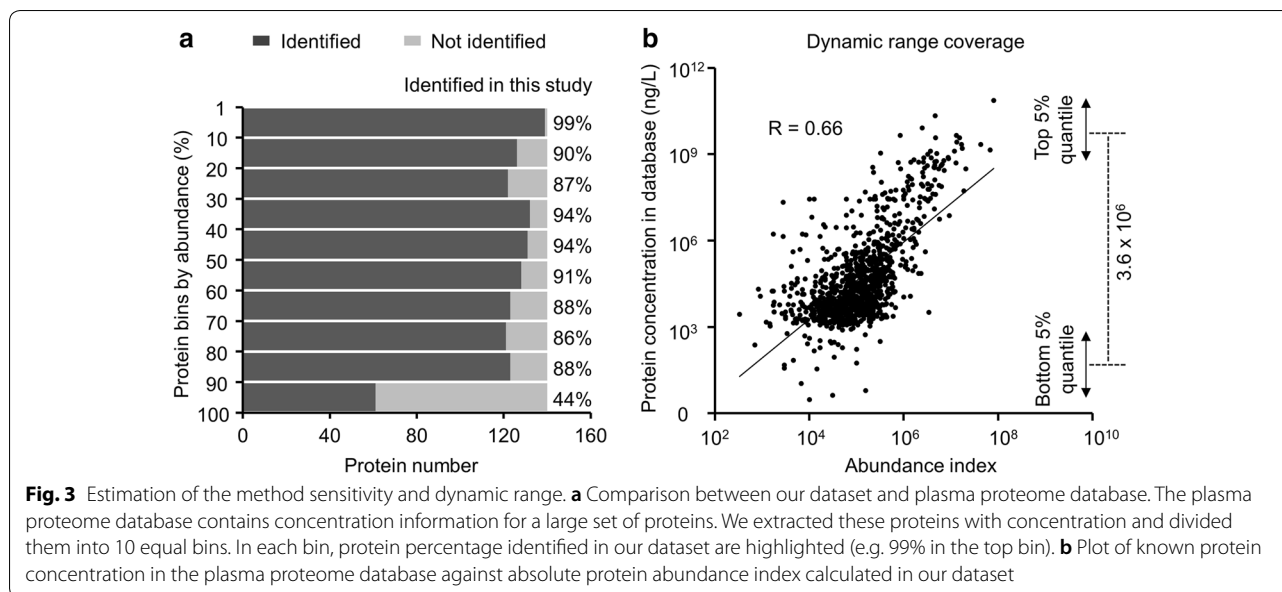
Consistently, in the 2nd dimensional acid pH RPLC–MS/MS, most peptides/proteins were identified between fractions 25–165. However, the number of identified peptides were not evenly distributed, suggesting that concatenation may be a solution to equalize peptide content and improve analytical efficiency.

In shotgun proteomics, longer analytical time is generally rewarded with higher peptide/protein coverage until the saturation point is reached. Indeed, prior to the analysis of the half of fractions ( $n=90$ , every alternative fraction), identified peptides increased with fractions in an approximately linear fashion (Fig. 2b). After 100 fractions, the slope appeared to decrease dramatically, implicating that the analysis was close to saturation. To enhance the throughput of this platform, it is possible to analyze ~4000 proteins with the half of these fractions to balance coverage and MS usage.

#### Evaluation of sensitivity and dynamic range for the identified serum proteome

To assess the sensitivity of our method, we compared our dataset with a public plasma proteome database, and focused on the 1399 proteins with reported plasma concentrations [44]. Out of these proteins in the database, we detected 1206 (86%) proteins in our analysis. While sorting the database proteins into 10 subsets by abundance ( $n \sim 140$  per subset) (Fig. 3a), our analysis identified at least 85% in the top 9 subsets, and the remaining <15% proteins do not have sufficient tryptic peptides compatible with our method. Even in the 10th subset of proteins with the lowest abundance, we still detected 44% of these proteins. Together, this comparison indicates that only





a very small portion of plasma proteins were below our detection limit, demonstrating high sensitivity of our TMT-LC/LC-MS/MS method.

We next computed the abundance index based on PSMs after size normalization (see “Methods”) and evaluated the dynamic range in our dataset. The abundance index is consistent with known protein concentrations in the plasma database ( $R=0.66$ , Fig. 3b, Additional file 1: Table S3). As to the dynamic range we covered, serum albumin has the highest concentration ( $7.6 \times 10^{10}$  ng/L), and cardiac type troponin T2 (TNNT2) has the lowest concentration (3.0 ng/L), spanning a range of more than 10 orders of magnitude. Conservatively, looking at the 5% top and bottom quantile, the estimated dynamic range is  $3.6 \times 10^6$  (Fig. 3b). The results indicate a broad dynamic range is covered by the deep analysis.

#### Quality control analysis and intra- and inter-group variations in AD-control serum proteomes

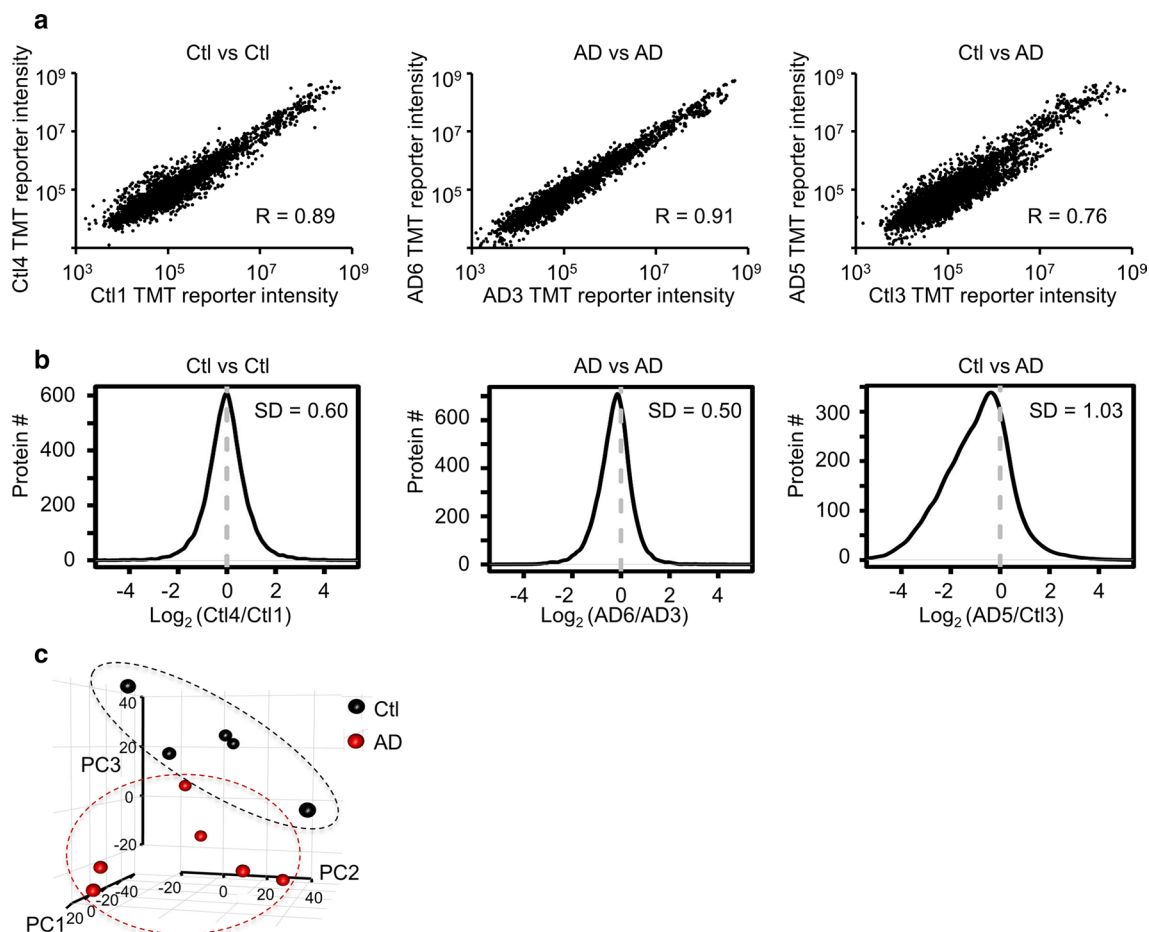
We performed quality control analysis by comparing any of the two samples in the 11 quantified cases, and evaluated intra- and inter-group variations in AD and control cases. To compare the samples, we plotted the TMT reporter intensities for all identified proteins (Fig. 4a). All two-sample comparisons showed a consistent and reproducible pattern ( $R$  of at least 0.7). For example,  $R$  values of the Ctl4/Ctl1, AD6/AD3, and AD5/Ctl3 were 0.89, 0.91 and 0.76. For the above comparisons, we also derived the  $\log_2$  ratio values for all proteins to generate the histograms, which were largely fit into normal distribution to generate standard deviation (Fig. 4b). As anticipated, the intra-group (e.g. Ctl4/Ctl1 and AD6/AD3) and

inter-group standard deviation (e.g. AD5/Ctl3) were 0.60, 0.50 and 1.03, respectively, consistently with the  $R$  values (Fig. 4a).

To fully compare intra- and inter-group variations, we obtained standard deviation values for all two-sample comparisons ( $n=10$  for the control group,  $n=15$  for the AD group, and  $n=30$  for the AD/control group). The averages of standard deviations in the control, AD, and AD/control comparisons were  $0.75 \pm 0.15$ ,  $0.73 \pm 0.12$ , and  $0.78 \pm 0.14$ , respectively. Although the inter-group had slightly larger variations than the intra-group comparisons, there is no statistically significant difference, which may be due to the small cohort size, or large confounding factors, such as gender, age, genetic background, clinical treatment, and other pre- and post-sample collection variance [29, 45]. However, three-dimensional principal-component analysis (PCA) of all quantified proteins displayed the separation of control and AD cases (Fig. 4c), confirming the reproducibility of the analysis.

#### Serum proteomics reveals deregulation of mitochondrial pathways in AD cases

To study serum proteome alterations in AD, we established a computational pipeline by integrating differentially expressed (DE) analysis with pathway enrichment [11, 18] (Fig. 5a). Out of 4826 proteins identified in the serum samples, we initially identified 248 DE proteins ( $p < 0.05$ ), which were filtered by  $\log_2$ (AD/Ctl) changes (1.5 fold of average standard deviation at 0.75, equal to 1.125 on the  $\log_2$  scale, equivalent to 2.2 fold change), resulting in 35 DE proteins. After manual examination



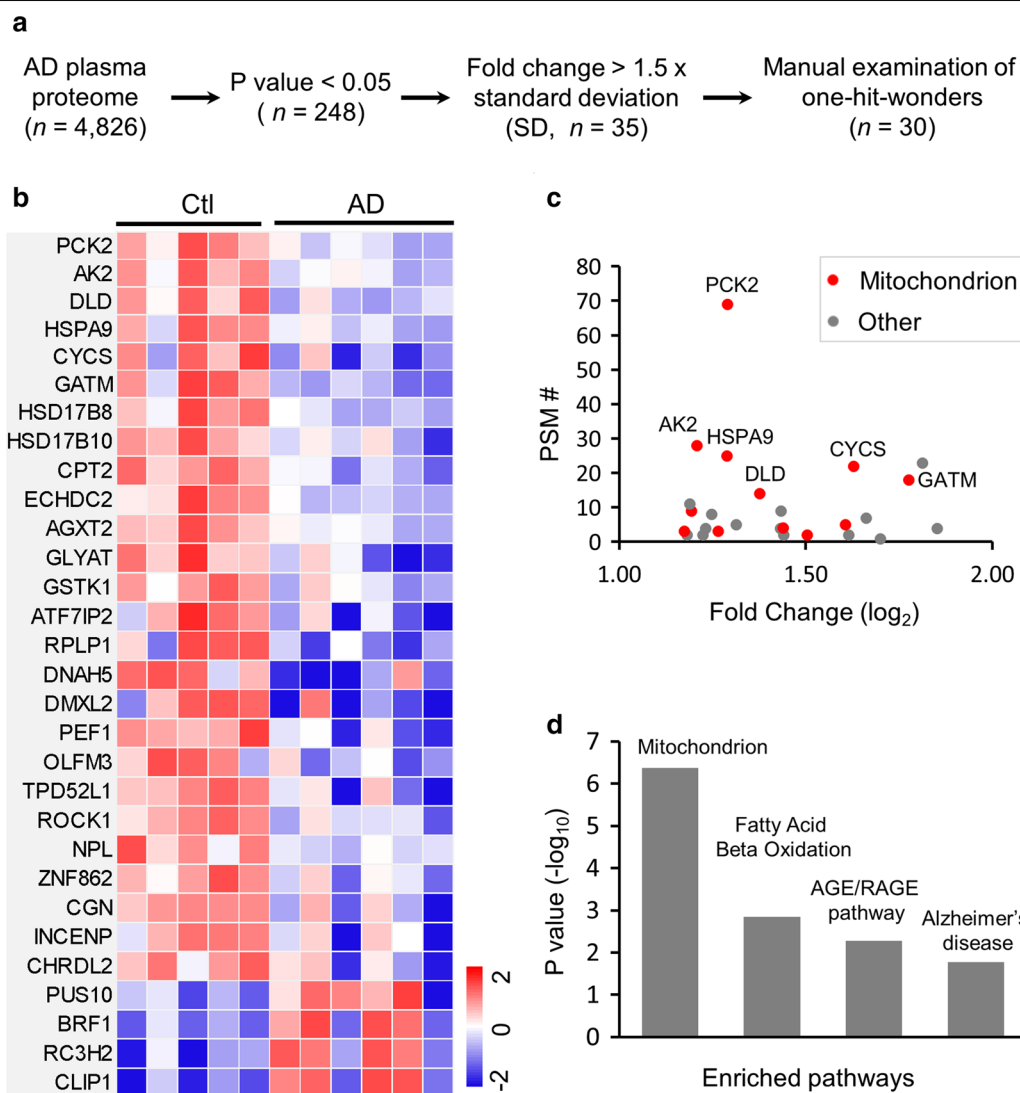
**Fig. 4** Statistical analysis to determine quality and intra/inter-group variations of serum proteome. **a** Representative comparisons of intra- and inter-group variations based on TMT reporter intensities for identified proteins in AD and control cases. **b** Histograms based on protein  $\log_2$  ratios, fitted to normal distribution to derive standard deviation. **c** Principal component analysis of identified proteins

to remove one-hit-wonders, we accepted a final list of 30 proteins (Additional file 1: Table S3), shown in a heat map representing  $\log_2$  ratios between AD and control samples (Fig. 5b). Whereas 4 (13%) proteins (PUS10, BRF1, RC3H2, and CLIP1) showed higher expression in the AD than the control, 26 (87%) proteins had lower expression in AD. Strikingly, out of these downregulated proteins, 12 proteins are localized in mitochondria, including some abundant proteins (e.g. PCK2, AK2, HSPA9, CYCS, DLD, and GATM, with at least 14 PSMs) (Fig. 5c).

Consistently, the 30 DE proteins were enriched in mitochondria-related pathway, as well as the signaling of fatty acid beta oxidation and AGE/RAGE (Fig. 5d). Interestingly, several proteins (HSPA9, CYCS, DLD, and GATM) were also enriched in various pathways related to Alzheimer's disease [46–49]. Finally, we superimposed the DE proteins onto PPI network to extract functional modules that are assembled by interacting proteins to form functional units at a systems level. The PPI network

was curated from the most commonly used databases, STRING [40], BioPlex [41], and InWeb\_IM [42]. Computational analysis identified 3 PPI modules, all related to mitochondrial function, including mitochondrial envelope (CYCS and GSTK1), intermembrane space (GATM and AGXT2), and matrix (AK2, DLD, HSPA9, HSD17B10, HSD17B8, and ECHDC2). Mitochondrial failure has been long proposed to play an important role in the development of Alzheimer's disease [50, 51]. The master mitochondrial regulator PGC-1 $\alpha$  [52] was reported to be dysregulated in AD brain during the progression of neuropathology and dementia, leading to the downregulation of mitochondrial genes including PCK2 [53], supporting our proteomic findings. Thus, comprehensive profiling of serum proteome revealed the change of key mitochondrial proteins in AD that may be relevant to disease development.

In this deep proteomics analysis, we also detected tau and APP proteins in the samples. However, these



**Fig. 5** Analysis of whole serum proteome reveals mitochondrial associated signaling pathway. **a** Summary of the computational pipeline for serum proteome analysis. **b** Heatmap of differentially expressed proteins from control and AD samples ( $p < 0.05$ ). **c** Fold change and PSM-based abundance of downregulated proteins in AD. **d** Functional annotations of enriched differentially expressed serum proteome

proteins did not show statistically significant difference between the control and AD samples, partially due to the limited sample size. Recently, Nakamura et al. developed an approach to measure plasma A $\beta$  by immunoprecipitation (IP) and MS, and proposed an AD composite biomarker based on (APP)<sub>669-711</sub>/A $\beta$ <sub>1-42</sub> and A $\beta$ <sub>1-40</sub>/A $\beta$ <sub>1-42</sub> ratios [28]. The composite biomarker displayed high performance for predicting brain A $\beta$  burden, and high correlation with A $\beta$ <sub>1-42</sub> in cerebrospinal fluid. Without the IP enrichment, the detailed ratio analysis could not be performed in our dataset. The IP-MS approach may be used to improve sensitivity for targeted biomarker candidates.

**TOMAHAQ-based multiplexed approach for target validation in AD samples**

Finally, we utilized a TOMAHAQ-based LC-MS3 assay to validate two mitochondrial proteins AK2 and PCK2 which differentially expressed in our LC/LC-MS/MS discovery study. Both candidates were found to be down regulated in AD, with Log<sub>2</sub> (AD/control) values of - 1.05 and - 1.21. In this validation assay, we synthesized two peptides as internal standards. The synthetic peptides were labeled with the TMT0 reagent, and then mixed with endogenous samples labeled with 11-plexed TMT reagents. TOMAHAQ allows simultaneous and accurate quantification of peptides across 11 samples in one



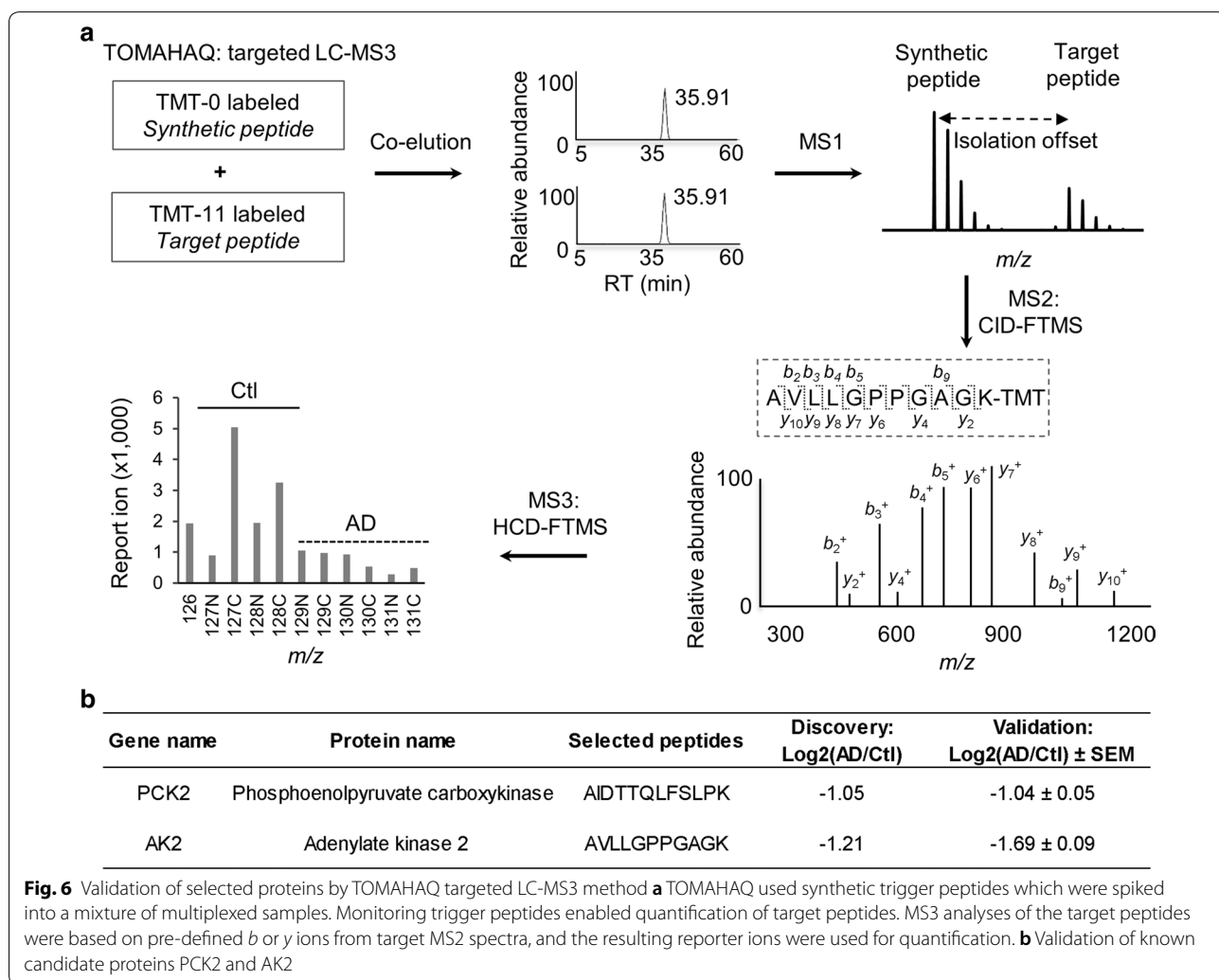
assay. During the LC-MS runs, TMT0-peptides were always detected to trigger the measurement of their corresponding native peptides by MS3 spectra. The reporter ions of native peptides in the MS3 spectra were used for accurate quantification (Fig. 6a, see details in “Methods”). Consistently, the levels of PCK2 and AK2 in the AD cases were significantly lower, with  $\text{Log}_2(\text{AD}/\text{control})$  values of  $-1.04 \pm 0.05$  and  $-1.69 \pm 0.09$ , respectively (Fig. 6b), when compared to the control cases. This multiplexed method may be used for sensitive and accurate quantification of selected targets in large-scale clinical validation in the future.

**Conclusions**

We identified 4826 proteins and demonstrated high proteome coverage, sensitivity and reproducibility, as well as multiplexed targeted assays. Although extensive fractionation and long instrumentation time were employed in this pilot study, we propose to achieve similar results

of ~4000 proteins within a reasonable time frame. This extensive TMT-LC/LC-MS/MS platform will be of general application for the measurement of complex clinical specimens. Remarkably, even in this small cohort, we identified consistent changes of 30 proteins in AD specimens compared to the non-dementia controls, in which 12 proteins were clustered to the mitochondria-related pathway. These novel protein signatures may be related to AD progression and have potential to be followed as biomarkers in a large scale investigation, possibly by the TOMAHAQ-based LC-MS3 assay.

To our knowledge, this study (30,506 peptides from 4826 proteins) represents one of the deepest, undepleted serum proteome profiling experiments from human biofluid. Previous studies usually attempted to increase the serum/plasma proteome coverage by immunodepletion of abundance proteins and extensive separation [4]. In 2006, the combination of immunodepletion, chemical fractionation (isolating cysteinyl- peptides and



**Fig. 6** Validation of selected proteins by TOMAHAQ targeted LC-MS3 method **a** TOMAHAQ used synthetic trigger peptides which were spiked into a mixture of multiplexed samples. Monitoring trigger peptides enabled quantification of target peptides. MS3 analyses of the target peptides were based on pre-defined *b* or *y* ions from target MS2 spectra, and the resulting reporter ions were used for quantification. **b** Validation of known candidate proteins PCK2 and AK2

glycol-peptides) and LC/LC–MS/MS, allowed the identification of 22,267 peptides from 3654 different proteins. In 2011, human plasma proteome datasets were compiled to produce a non-redundant list of 1929 proteins (20,433 peptides) of high confidence [54]. In 2015, with the advance of better fractions and instrumentation, about 4600 proteins were analyzed in human plasma by immunodepletion, isobaric labeling and LC/LC–MS/MS. In 2017, the human plasma proteome draft included 3509 proteins identified at least two peptides, and about 1300 additional ambiguous proteins [55]. The drawbacks of immunodepletion are the removal of non-targeted proteins, associated quantitative variability, and the cost of the antibody cartridge [7]. Our study demonstrates the possibility to achieve deep analysis without the step of immunodepletion. However, all of these deep plasma/serum profiling experiments were time consuming due to a large number of fractions, which are not well suited for large clinical studies. Alternatively, a single-run, label-free protocol was introduced for rapid analysis of hundreds of plasma proteomes, and with additional pre-fractionation, interpretation of 1000 proteins became possible [56]. Other approaches, such as SWATH, was used to quantify more than 300 plasma proteins in 232 plasma samples [57]. Furthermore, the throughput of profiling of biofluids can be increased by sample multiplexing, such as iTRAQ/TMT labeling [58]. Here, we adapted the TMT-derived TOMAHAQ method for targeted protein analysis. The integration of deep proteome coverage by extensive TMT–LC/LC–MS/MS in the discovery phase, and targeted measurement by TOMAHAQ in the validation phase, will represent a balance between comprehensive profiling and analytical time.

## Additional file

**Additional file 1: Table S1.** Summary of human cases used in this study. **Table S2.** Serum proteome profiling of AD and control cases by TMT–LC/LC–MS/MS. **Table S3.** Dynamic range analysis based on plasma proteome database concentration and MS-derived abundance index. **Table S4.** Differentially expressed serum proteins in AD and control cases by TMT–LC/LC–MS/MS.

## Abbreviations

AD: Alzheimer's disease; MS: mass spectrometry; TMT: tandem mass tags; LC/LC–MS/MS: two dimensional liquid chromatography coupled with tandem mass spectrometry; TOMAHAQ: triggered by offset multiplexed accurate mass high resolution absolute quantification.

## Authors' contributions

JP, BB, PCC, KKD, and TGB contributed to the conception and design of the project. T.G.B. provided the human specimens. KKD, MN, HT, ZP, AM and AAH performed the proteomics experiments. HW, KKD, BB, XW, YL, JHC, and JP analyzed and interpreted the data. JP, KKD, and HW wrote the manuscript. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> Departments of Structural Biology and Developmental Neurobiology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA. <sup>2</sup> Center for Proteomics and Metabolomics, St. Jude Children's Research Hospital, Memphis, TN 38105, USA. <sup>3</sup> Banner Sun Health Research Institute, Sun City, AZ 85351, USA. <sup>4</sup> Present Address: Department of Laboratory Medicine, Nanjing Drum Tower Hospital, Nanjing University Medical School, Nanjing 210008, Jiangsu, China.

## Acknowledgements

The authors thank all other lab and facility members for helpful discussion, and Brian K. Erickson and Steven P. Gygi for TOMAHAQ consultation. The MS analysis was performed in the Center for Proteomics and Metabolomics; and the peptide synthesis was carried out in the Hartwell Center, both at St. Jude Children's Research Hospital.

## Completing interests

The authors declare that they have no competing interests.

## Availability of data and materials

The mass spectrometry proteomics data have been deposited to the Proteome Xchange Consortium via the PRIDE partner repository with the dataset identifier PXD011482. Data are available via Proteome Xchange with identifier PXD011482. Reviewer account details: Username: reviewer00296@ebi.ac.uk, Password: Q9Yjbjnr.

## Consent for publication

The authors declare that they have no competing interests.

## Ethics approval and consent to participate

All the participants have given the written informed consent. This research was approved by the Institutional Review Committee. All experiments were performed in accordance with the relevant guidelines and regulations.

## Funding

This work was partially supported by National Institutes of Health grants R01AG047928 (J.P.), R01AG053987 (J.P.), U24NS072026 (T.G.B.), P30AG19610 (T.G.B.), Arizona Department of Health Services (contract 211002) (T.G.B.), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001) (T.G.B.), ALSAC (American Lebanese Syrian Associated Charities), and St. Jude Children's Research Hospital, partially supported by NIH Cancer Center Support Grant (P30CA021765).

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 29 October 2018 Accepted: 10 April 2019

Published online: 17 April 2019

## References

- Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics*. 2002;1:845–67.
- Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat Biotechnol*. 2006;24:971–83.
- Arican O, Aral M, Sasmaz S, Ciragil P. Serum levels of TNF-alpha, IFN-gamma, IL-6, IL-8, IL-12, IL-17, and IL-18 in patients with active psoriasis and correlation with disease severity. *Mediators Inflamm*. 2005;2005:273–9.
- Geyer PE, Holdt LM, Teupser D, Mann M. Revisiting biomarker discovery by plasma proteomics. *Mol Syst Biol*. 2017;13:942.
- Pieper R, Su Q, Gatlin CL, Huang ST, Anderson NL, Steiner S. Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. *Proteomics*. 2003;3:422–32.
- Qian WJ, Kaleta DT, Petritis BO, Jiang H, Liu T, Zhang X, et al. Enhanced detection of low abundance human plasma proteins using a tandem

- IgY12-SuperMix immunoaffinity separation strategy. *Mol Cell Proteomics*. 2008;7:1963–73.
7. Tu C, Rudnick PA, Martinez MY, Cheek KL, Stein SE, Slebos RJ, et al. Depletion of abundant plasma proteins and limitations of plasma proteomics. *J Proteome Res*. 2010;9:4982–91.
  8. Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR 3rd. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev*. 2013;113:2343–94.
  9. Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. *Nature*. 2016;537:347–55.
  10. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*. 2016;534:55–62.
  11. Stewart E, McEvoy J, Wang H, Chen X, Honnell V, Ocarz M, et al. Identification of therapeutic targets in rhabdomyosarcoma through integrated genomic, epigenomic, and proteomic analyses. *Cancer Cell*. 2018;34(4):11–26. e19.
  12. Altaalar AF, Munoz J, Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet*. 2013;14:35–48.
  13. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol Syst Biol*. 2018;14:e8126.
  14. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, Schmidt G, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem*. 2003;75:1895–904.
  15. Rauniyar N, Yates JR 3rd. Isobaric labeling-based relative quantification in shotgun proteomics. *J Proteome Res*. 2014;13:5293–309.
  16. Ting L, Rad R, Gygi SP, Haas WS. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nat Methods*. 2011;8:937–40.
  17. Niu M, Cho JH, Kodali K, Pagala V, High AA, Wang H, et al. Extensive peptide fractionation and y1 ion-based interference detection method for enabling accurate quantification by isobaric labeling and mass spectrometry. *Anal Chem*. 2017;89:2956–63.
  18. Tan H, Yang K, Li Y, Shaw TI, Wang Y, Blanco DB, et al. Integrative proteomics and phosphoproteomics profiling reveals dynamic signaling networks and bioenergetics pathways underlying T cell activation. *Immunity*. 2017;46:488–503.
  19. Du X, Wen J, Wang Y, Karmaus PWF, Khatamian A, Tan H, et al. Hippo/Mst signalling couples metabolic state and immune function of CD8alpha(+) dendritic cells. *Nature*. 2018;558:141–5.
  20. Gillette MA, Carr SA. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat Methods*. 2013;10:28–34.
  21. Kusebauch U, Campbell DS, Deutsch EW, Chu CS, Spicer DA, Brusniak MY, et al. Human SRMAtlas: a resource of targeted assays to quantify the complete human proteome. *Cell*. 2016;166:766–78.
  22. Erickson BK, Rose CM, Braun CR, Erickson AR, Knott J, McAlister GC, et al. A strategy to combine sample multiplexing with targeted proteomics assays for high-throughput protein signature characterization. *Mol Cell*. 2017;65:361–70.
  23. Rose CM, Erickson BK, Schweppe DK, Viner R, Choi J, Rogers J, et al. Tomahaqcompanion: a tool for the creation and analysis of isobaric label based multiplexed targeted assays. *J Proteome Res*. 2019;18:594–605.
  24. Alzheimer's A. 2016 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2016;12:459–509.
  25. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement*. 2007;3:186–91.
  26. Hyman BT, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Carrillo MC, et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement*. 2012;8:1–13.
  27. Risacher SL, Saykin AJ. Neuroimaging and other biomarkers for Alzheimer's disease: the changing landscape of early detection. *Annu Rev Clin Psychol*. 2013;9:621–48.
  28. Nakamura A, Kaneko N, Villemagne VL, Kato T, Doecke J, Dore V, et al. High performance plasma amyloid-beta biomarkers for Alzheimer's disease. *Nature*. 2018;554:249–54.
  29. Olsson B, Lautner R, Andreasson U, Ohrfelt A, Portelius E, Bjerke M, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol*. 2016;15:673–84.
  30. Beach TG, Adler CH, Sue LI, Serrano G, Shill HA, Walker DG, et al. Arizona study of aging and neurodegenerative disorders and brain and body donation program. *Neuropathology*. 2015;35:354–89.
  31. Xu P, Duong DM, Peng JM. Systematical optimization of reverse-phase chromatography for shotgun proteomics. *J Proteome Res*. 2009;8:3944–50.
  32. Pagala VR, High AA, Wang X, Tan H, Kodali K, Mishra A, et al. Quantitative protein analysis by mass spectrometry. *Methods Mol Biol*. 2015;1278:281–305.
  33. Bai B, Tan H, Pagala VR, High AA, Ichhaporis VP, Hendershot L, et al. Deep profiling of proteome and phosphoproteome by isobaric labeling, extensive liquid chromatography, and mass spectrometry. *Methods Enzymol*. 2017;585:377–95.
  34. Wang H, Yang Y, Li Y, Bai B, Wang X, Tan H, et al. Systematic optimization of long gradient chromatography mass spectrometry for deep analysis of brain proteome. *J Proteome Res*. 2015;14:829–38.
  35. Wang X, Li Y, Wu Z, Wang H, Tan H, Peng J. JUMP: a tag-based database search tool for peptide identification with high sensitivity and accuracy. *Mol Cell Proteomics*. 2014;13:3663–73.
  36. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J Proteome Res*. 2003;2:43–50.
  37. Peng J, Kim MJ, Cheng D, Duong DM, Gygi SP, Sheng M. Semiquantitative proteomic analysis of rat forebrain postsynaptic density fractions by mass spectrometry. *J Biol Chem*. 2004;279:21003–11.
  38. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics*. 2005;4:1265–72.
  39. Ihaka R, Gentleman R. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*. 1996;5:299–314.
  40. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43:D447–52.
  41. Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, et al. Architecture of the human interactome defines protein communities and disease networks. *Nature*. 2017;545:505–9.
  42. Li T, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkowitz G, et al. A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat Methods*. 2017;14:61–4.
  43. Deutsch EW, Csordas A, Sun Z, Jarnuczak A, Perez-Riverol Y, Ternent T, et al. The ProteomeXchange consortium in 2017: supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res*. 2017;45:D1100–6.
  44. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, Sharma R, et al. Plasma proteome database as a resource for proteomics research: 2014 update. *Nucleic Acids Res*. 2014;42:D959–65.
  45. Zhou JY, Hanfelt J, Peng J. Clinical proteomics in neurodegenerative diseases. *Proteomics Clin Appl*. 2007;1:1342–50.
  46. Hoshino T, Muraio N, Namba T, Takehara M, Adachi H, Katsuno M, et al. Suppression of Alzheimer's disease-related phenotypes by expression of heat shock protein 70 in mice. *J Neurosci*. 2011;31:5225–34.
  47. Fukui H, Diaz F, Garcia S, Moraes CT. Cytochrome c oxidase deficiency in neurons decreases both oxidative stress and amyloid formation in a mouse model of Alzheimer's disease. *Proc Natl Acad Sci USA*. 2007;104:14163–8.
  48. Brown AM, Gordon D, Lee H, Wavrant-De Vrieze F, Cellini E, Bagnoli S, et al. Testing for linkage and association across the dihydrolipoyl dehydrogenase gene region with Alzheimer's disease in three sample populations. *Neurochem Res*. 2007;32:857–69.
  49. Walker DG, Link J, Lue LF, Dalsing-Hernandez JE, Boyes BE. Gene expression changes by amyloid beta peptide-stimulated human postmortem brain microglia identify activation of multiple inflammatory processes. *J Leukoc Biol*. 2006;79:596–610.
  50. Swerdlow RH, Burns JM, Khan SM. The Alzheimer's disease mitochondrial cascade hypothesis: progress and perspectives. *Biochim Biophys Acta*. 2014;1842:1219–31.
  51. Moreira PI, Carvalho C, Zhu X, Smith MA, Perry G. Mitochondrial dysfunction is a trigger of Alzheimer's disease pathophysiology. *Biochim Biophys Acta*. 2010;1802:2–10.
  52. Finck BN, Kelly DP. PGC-1 coactivators: inducible regulators of energy metabolism in health and disease. *J Clin Invest*. 2006;116:615–22.

53. Qin W, Haroutunian V, Katsel P, Cardozo CP, Ho L, Buxbaum JD, et al. PGC-1alpha expression decreases in the Alzheimer disease brain as a function of dementia. *Arch Neurol*. 2009;66:352–61.
54. Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, et al. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol Cell Proteomics*. 2011;10(M1110):006353.
55. Schwenk JM, Omenn GS, Sun Z, Campbell DS, Baker MS, Overall CM, et al. The human plasma proteome draft of 2017: building on the human plasma peptideatlas from mass spectrometry and complementary assays. *J Proteome Res*. 2017;16:4299–310.
56. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. Plasma proteome profiling to assess human health and disease. *Cell Syst*. 2016;2:185–95.
57. Liu Y, Buil A, Collins BC, Gillet LC, Blum LC, Cheng LY, et al. Quantitative variability of 342 plasma proteins in a human twin population. *Mol Syst Biol*. 2015;11:786.
58. Sathe G, Na CH, Renuse S, Madugundu AK, Albert M, Moghekar A, et al. Quantitative proteomic profiling of cerebrospinal fluid to identify candidate biomarkers for Alzheimer's disease. *Proteomics Clin Appl*. 2018 e1800105.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

