

OPEN  
ANALYSIS

# Insights into the changes in the proteome of Alzheimer disease elucidated by a meta-analysis

Hazal Haytural<sup>1</sup>✉, Rui Benfeitas<sup>2</sup>, Sophia Schedin-Weiss<sup>1</sup>, Erika Berezcki<sup>1</sup>, Melinda Rezel<sup>3</sup>, Richard D. Unwin<sup>4</sup>, Xusheng Wang<sup>5</sup>, Eric B. Dammer<sup>6,7</sup>, Erik C. B. Johnson<sup>6,8</sup>, Nicholas T. Seyfried<sup>6,7,8</sup>, Bengt Winblad<sup>1,9</sup>, Betty M. Tijms<sup>10</sup>, Pieter Jelle Visser<sup>1,10,11</sup>, Susanne Frykman<sup>1</sup> & Lars O. Tjernberg<sup>1</sup>✉

Mass spectrometry (MS)-based proteomics is a powerful tool to explore pathogenic changes of a disease in an unbiased manner and has been used extensively in Alzheimer disease (AD) research. Here, by performing a meta-analysis of high-quality proteomic studies, we address which pathological changes are observed consistently and therefore most likely are of great importance for AD pathogenesis. We retrieved datasets, comprising a total of 21,588 distinct proteins identified across 857 postmortem human samples, from ten studies using labeled or label-free MS approaches. Our meta-analysis findings showed significant alterations of 757 and 1,195 proteins in AD in the labeled and label-free datasets, respectively. Only 33 proteins, some of which were associated with synaptic signaling, had the same directional change across the individual studies. However, despite alterations in individual proteins being different between the labeled and the label-free datasets, several pathways related to synaptic signaling, oxidative phosphorylation, immune response and extracellular matrix were commonly dysregulated in AD. These pathways represent robust changes in the human AD brain and warrant further investigation.

## Introduction

Mass spectrometry (MS)-based proteomics is a powerful technique, as it allows a simultaneous identification and quantification of proteins in complex biological samples such as brain tissue. These studies typically use a bottom-up approach in which proteins are first digested, the resulting peptides are then analyzed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS), and lastly the generated mass spectra of the peptide ions are compared against the theoretical spectra from databases for protein identification. Two main approaches, labeled or label-free, are commonly used for quantification of the relative abundance of the identified proteins. In the labeling strategies, such as tandem mass tags (TMTs) or isobaric tags for relative and absolute quantification (iTRAQ), stable tags are chemically attached to free amine on the N-termini of the peptides and lysine side-chains<sup>1,2</sup>. The principle of isobaric labeling strategies relies on the fact that labeled peptides

<sup>1</sup>Department of Neurobiology, Care Sciences and Society, Division of Neurogeriatrics, Center for Alzheimer Research, Karolinska Institutet, Solna, Sweden. <sup>2</sup>National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, S-10691, Stockholm, Sweden. <sup>3</sup>Division of Clinical Protein Science & Imaging, Department of Clinical Sciences (Lund) and Department of Biomedical Engineering, Lund University, Lund, Sweden. <sup>4</sup>Stoller Biomarker Discovery Centre, and Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester Academic Health Sciences Centre, CityLabs 1.0, Nelson Street, Manchester, M13 9NQ, UK. <sup>5</sup>Department of Biology, University of North Dakota, Grand Forks, ND, USA. <sup>6</sup>Goizueta Alzheimer's Disease Research Center, Emory University School of Medicine, Atlanta, GA, USA. <sup>7</sup>Department of Biochemistry, Emory University School of Medicine, Atlanta, GA, USA. <sup>8</sup>Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA. <sup>9</sup>Karolinska University Hospital, Theme of Inflammation and Aging, Huddinge, Sweden. <sup>10</sup>Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, The Netherlands. <sup>11</sup>Alzheimer Center Limburg, School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands. ✉e-mail: [hazal.haytural@ki.se](mailto: hazal.haytural@ki.se); [lars.tjernberg@ki.se](mailto: lars.tjernberg@ki.se)

have the same overall mass due to the isobaric chemical structure of the tags and therefore are chromatographically indistinguishable, but once the peptides are fragmented by collision-induced dissociation via tandem MS (MS2), the reporter ions are released and used for relative protein quantitation<sup>1</sup>. One advantage of this approach is the ability to label different biological samples with different isobaric tags so that they can be analyzed together in the same LC-MS/MS, thus reducing inter-run variability. One drawback with this approach is that the quantification can be affected by interference from isobaric tags derived from coeluting peptides with similar mass, resulting in an underestimation of the fold change – so called ratio compression<sup>3</sup>. In contrast, in the label-free MS approach, each sample is individually analyzed by LC-MS/MS and relative quantitation is often done using the chromatographic precursor ion peak intensity coming from MS1<sup>4</sup>. Thanks to recent advancements allowing more rapid analysis and thus increased proteome coverage, a more reliable protein quantitation can be achieved by LC-MS/MS. This has led to increased application of MS-based proteomics in clinical research to investigate proteins and pathways underlying disease pathophysiology.

One disease that has been extensively studied in the proteomics field is Alzheimer disease (AD), the most common cause of dementia<sup>5</sup>. Many proteomic studies have focused on identifying changes in the proteome of a single brain region in AD cases compared to control subjects<sup>6–10</sup> while others investigated regional vulnerability by comparing multiple brain regions from the same cohort<sup>11–13</sup>. In a few studies, after employing laser microdissection techniques, region-<sup>14,15</sup>, cell-<sup>16</sup> or structure-specific (e.g., amyloid plaques)<sup>17,18</sup> proteomes have also been investigated. Furthermore, the proteomic changes occurring in a specific brain region over the course of Braak stages has been explored<sup>15,19</sup>. Mounting evidence from proteomic studies has shed light onto a number of disrupted cellular mechanisms such as synaptic signaling, mitochondrial bioenergetics, immune response, RNA homeostasis, lipid metabolism and vesicle trafficking – supporting the notion that AD is a multifactorial disease<sup>6,7,14,19</sup>.

Here, we performed a study-level meta-analysis to identify the most robustly affected proteins and pathways in AD brain across cohorts and studies, as well as to investigate potential reasons that could explain the discrepancies observed in the published proteomic studies. Our findings reveal that individual protein alterations observed in AD were clearly dependent on whether a labeled or label-free MS approach was used. However, comprehensive pathway analysis found the involvement of common pathways related to synaptic signaling, oxidative phosphorylation, immune response and extracellular matrix (ECM) in AD brain, irrespective of the MS quantitative method of choice.

## Results

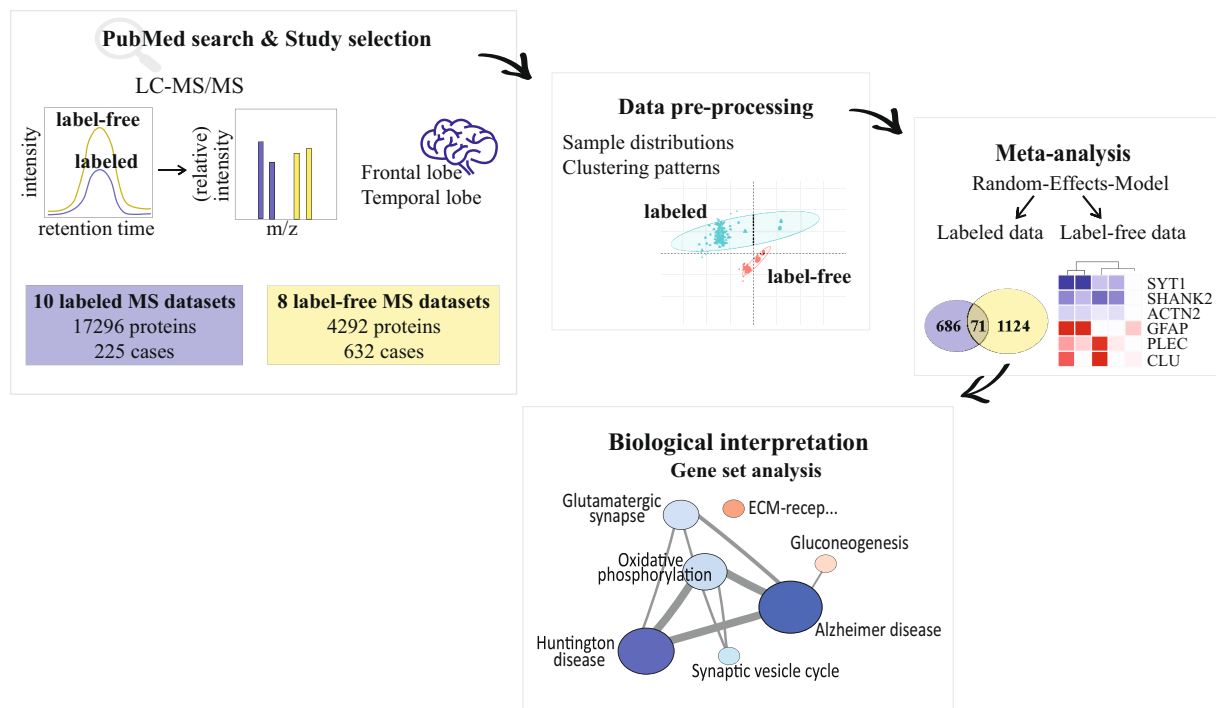
Out of 10 studies using labeled (TMT, iTRAQ and <sup>18</sup>O labeling)<sup>8–12,14</sup> or label-free<sup>6,7,19</sup> MS approaches, 18 datasets that analyzed the proteome of frontal and temporal cortices were retrieved. Data containing isoform-specific information was combined by using UniProtKB accession number as a common identifier between studies. This resulted in a dataset with 21,588 distinct protein IDs, 533 of which were identified in all 857 postmortem human samples. Protein intensities from the retrieved datasets were standardized (i.e., allowing 20% missing values per group, log<sub>2</sub> conversion, median 0 and standard deviation 1). As source of variance could arise from differences in biological samples and methodology (e.g., instrumentation, database search platforms, application of batch correction or other regression models), a random-effects-model was used to calculate the effect size (Fig. 1).

**Samples of labeled vs label-free datasets appears in distinct clusters.** To investigate the largest variability between samples and to determine whether there were clustering patterns that could be explained by some of the methodological differences among the selected studies, principal component analysis (PCA) was performed using the log<sub>2</sub> protein intensities of 533 proteins quantified in all 857 samples (547 AD and 310 control cases). Two main clusters pertaining to labeled or label-free MS approaches (Fig. 2a) suggested that relative protein intensities were specific to the MS method of choice, which may partly be explained by an underestimation of the fold change in the labeled samples due to ratio compression. No clear separation was observed between datasets (Fig. 2b), diagnosis (AD vs control, Fig. 2c), brain regions (frontal vs temporal lobes, Fig. 2d) or lysis buffer (SDS- vs urea-based, Fig. 2e). The top 10 proteins contributing to the first (PC1) and second (PC2) components are also shown in Fig. 2f. Intriguingly, two datasets from Bai *et al.*<sup>9</sup> were clustered further away from all other labeled datasets. While this cannot be explained by the choice of labeling approach, it is possible that differences in methodology (e.g., cell-type correction) could result in this disparity<sup>9</sup>.

**Differences and similarities between the labeled and the label-free datasets.** Based on the distinct sample distributions observed above (Fig. 2a), we performed separate meta-analyses for labeled and label-free datasets. The labeled data consisted of 17,296 distinct proteins quantified across 225 (123 AD and 102 control) cases while the label-free data contained 4,292 distinct proteins quantified across 632 (424 AD and 208 control) cases (Online-only Table 1). When comparing the total protein identifications between the labeled and the label-free data, 3,731 proteins were found to be shared (Fig. 3a). The input data that was used for meta-analyses can be found at the figshare database<sup>20</sup>.

The meta-analysis highlighted that 757 (391 down- and 366 upregulated) proteins were significantly altered in AD in the labeled data (p-value < 0.005 and false discovery rate (FDR) < 10%), compared to 1,195 (634 down- and 561 upregulated) proteins in the label-free data (p-value < 0.03 and FDR < 10%) (Fig. 3b). The findings of both meta-analyses can be found at the figshare database<sup>20</sup>. Notably, among significant alterations, 604 proteins (80%) were originally quantified in two or more labeled datasets and 1038 proteins (87%) were quantified in more than one label-free dataset<sup>20</sup>.

Despite the large number of shared proteins between the labeled and the label-free datasets (Fig. 3a), only 71 proteins (1.9%) were found to be significantly altered in AD in both datasets (Fig. 3b). Of these, 33 proteins (46%) showed consistent alterations in AD, in other words the mean difference between AD and control,



**Fig. 1** The workflow of the study. After literature search, 18 datasets from ten different MS-based proteomic studies, using either labeled or label-free quantification, were retrieved. These studies were done on postmortem human brain tissue, which were categorized into frontal or temporal lobes, severely affected regions by AD pathology, and consisted of AD ( $5 \leq n \leq 252$ ) and control ( $5 \leq n \leq 94$ ) cases. Subsequently, data pre-processing was done so that datasets would be comparable for further statistical analyses. Sample distributions and the presence of any clustering patterns were assessed using principal component analysis (PCA). The datasets generated by labeled and label-free MS approaches were concatenated separately, and meta-analysis using a random-effects-model was performed. Lastly, for better biological interpretation of our findings, gene set analysis was performed.

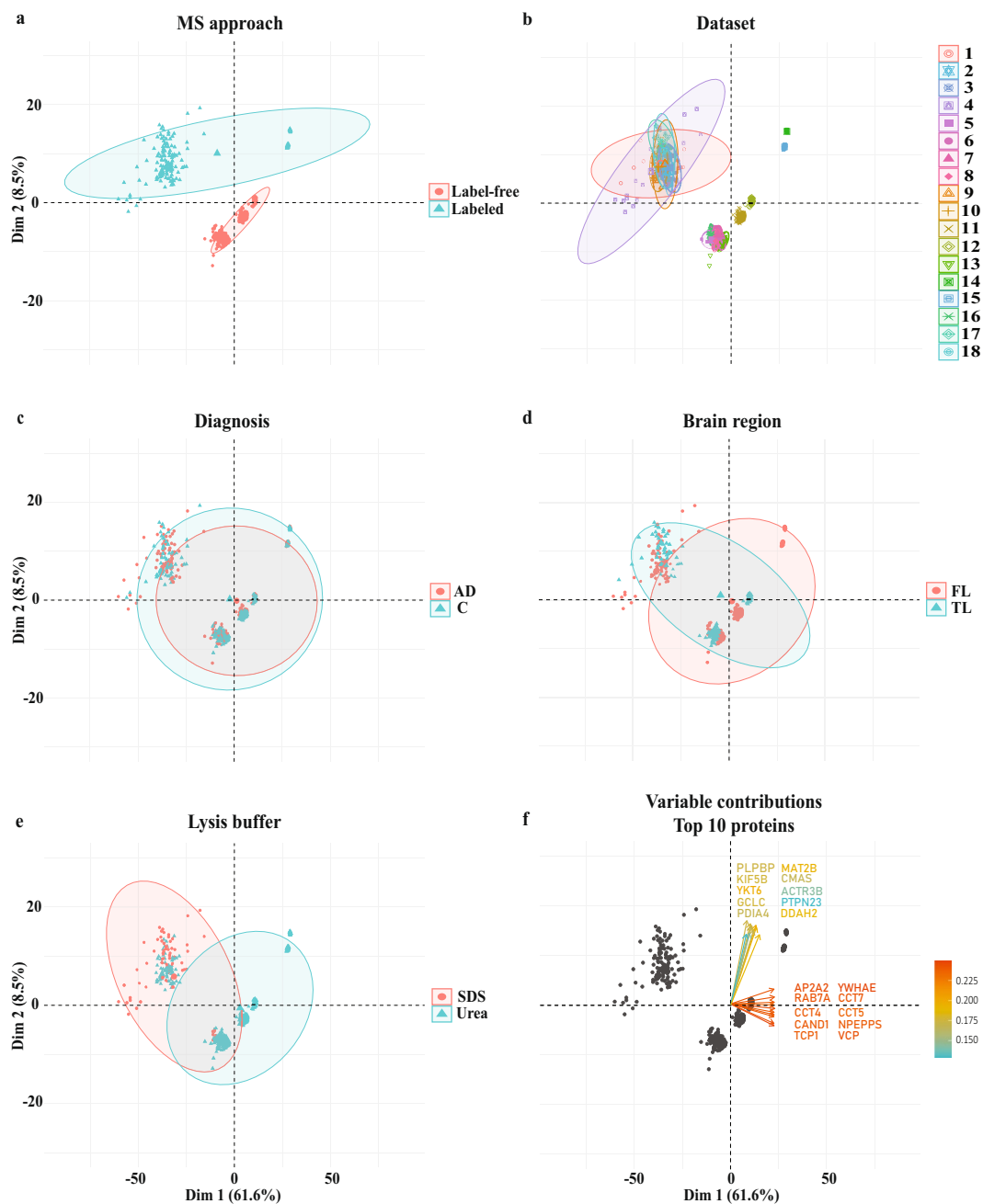
computed by meta-analyses, was in line with the fold changes found in individual studies (Fig. 3c). Several of these proteins were found to be implicated in synaptic signaling pathways, apoptotic signaling and proteasomal protein catabolic processes. Importantly, our analysis led to the identification of novel proteins involved in AD (Fig. 3c), such as GDNF family receptor alpha-2 (GFRA2), voltage-dependent L-type calcium channel subunit beta-3 (CACNB3), utrophin (UTRN), sodium-dependent phosphate transporter 2 (SLC20A2) and synaptotagmin-11 (SYT11).

**Gene set analysis (GSA) identifies common pathways in the meta-analysis of the labeled and the label-free data.** To put single protein alterations into biological context, we performed GSA<sup>21</sup>. All proteins (in gene-centric format) were subjected to GSA (for input data, see<sup>20</sup>), despite that some of them were originally quantified only in one dataset or had opposite directional changes between meta-analyses and individual studies.

By using the KEGG 2019 database, GSA showed that synaptic signaling (e.g., synaptic vesicle cycle), oxidative phosphorylation and pathways related to neurodegenerative disorders, such as AD and Huntington disease (HD), were significantly downregulated in AD in the labeled data (FDR < 5%, Fig. 4a). In addition, complement and coagulation cascade and ECM-receptor interaction pathways were significantly upregulated in AD in the labeled data. Similar changes in biological processes were detected in the label-free data, with the exception that the glycolysis/gluconeogenesis pathway was significantly upregulated in AD (Fig. 4b). Similarly, using the GO biological process database, GSA showed pathways related to synaptic signaling, mitochondrial metabolism, ECM organization and immune response, thereby greatly strengthening the notion that these pathways are involved in disease pathogenesis. All GSA findings are deposited in the figshare database<sup>20</sup>.

## Discussion

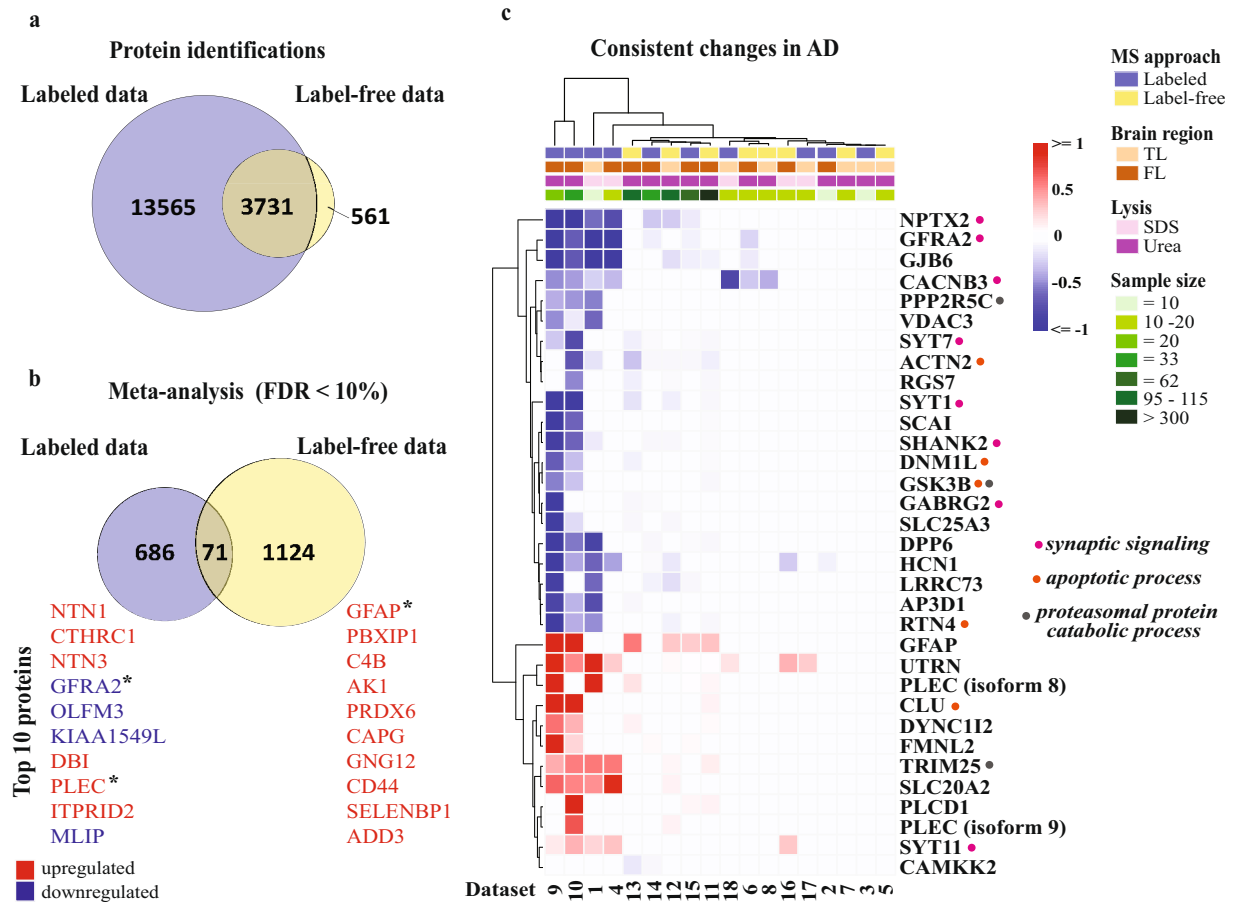
We performed a meta-analysis of proteomic studies with the aim of identifying proteins and pathways that are robustly related to AD pathogenesis as well as elucidating discrepancies between the studies. Three different labeling techniques were employed in the original studies including TMT, <sup>18</sup>O and iTRAQ. Since sample distributions pertaining to these techniques showed a high degree of overlap in the PCA plot, we together call them labeled datasets. Particularly, by comparing the meta-analysis of the labeled and the label-free MS datasets separately, our study provides further knowledge on how the observed proteome might change depending on the MS method of choice. The notion that these two MS approaches rely on different protein quantification methods<sup>22</sup>,



**Fig. 2** Sample distributions visualized by principal component analysis (PCA). This analysis was performed using log<sub>2</sub> intensities of 533 proteins, which were quantified in all 857 (547 AD and 310 control) cases and in all 18 datasets. Sample distributions were grouped by (a) MS approach (labeled vs label-free quantification), (b) datasets, (c) diagnosis (AD vs control), (d) brain region (frontal vs temporal lobe), and lastly (e) lysis buffer (SDS- vs urea-based). PCA showed a clear separation between labeled and label-free MS approaches. (f) The top 10 proteins contributing most to component 1 and component 2 were shown respectively. Ellipses indicate the 95% confidence interval of samples in each of the groups. Data was standardized prior to PCA.

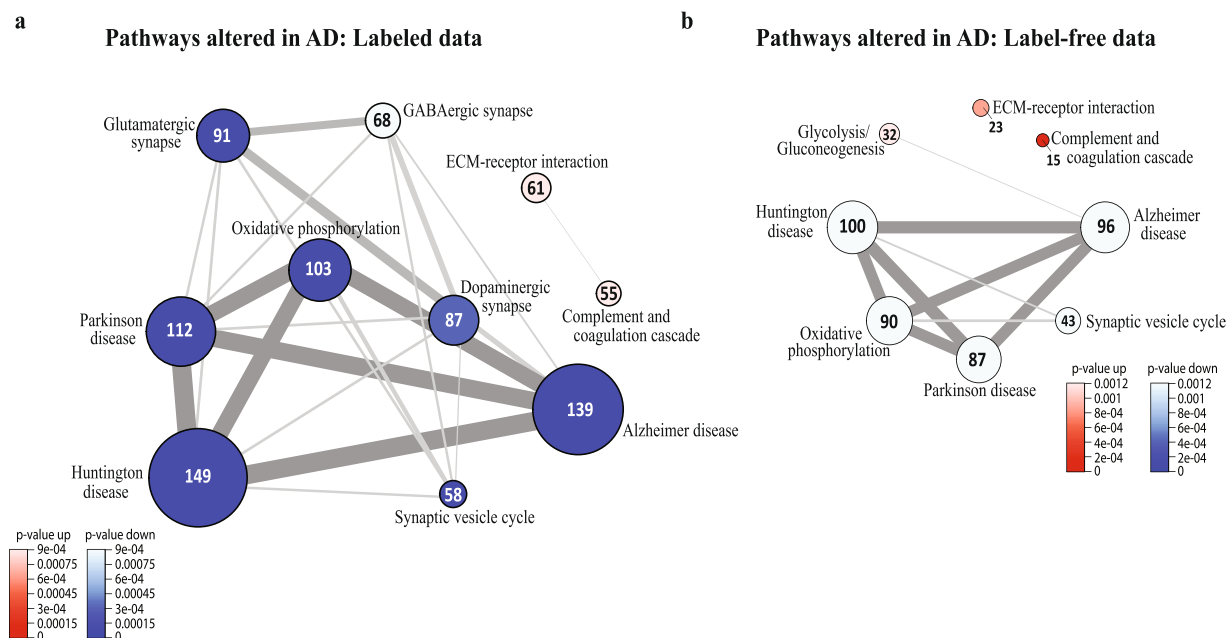
it is not surprising that PCA showed a clear separation, however we cannot rule out that other factors might also contribute to this clustering pattern.

Our meta-analysis findings show that the observed protein alterations were clearly different between the two techniques, since only 71 significantly altered proteins (FDR < 10%) were shared. Several reasons may explain why these 71 proteins stood out from the other significant alterations: they are robustly quantified independent of MS approach or they are not sensitive to discrepancies in sample preparation procedures. To get a better understanding of these 71 significantly altered proteins, we further considered whether the mean difference (i.e., the main outcome of meta-analysis) and the fold changes found in each individual dataset followed the same direction, i.e., the protein was either upregulated or downregulated across datasets. We found 22 consistently



**Fig. 3** Summary of the meta-analysis findings. Venn diagrams showing (a) the total number of identified proteins (with distinct protein IDs) shared between the labeled and the label-free datasets, and (b) the statistically significant proteins with FDR < 10% identified by the meta-analysis of the labeled and the label-free datasets. Despite a large overlap found between the two data subsets (3731 proteins), only 71 proteins were significantly altered in AD in both meta-analyses. The top 10 most significantly altered proteins were shown next to the Venn diagram. \*These proteins were among the shared proteins. (c) Out of 71 significantly altered proteins, 33 were identified as the most robust changes in AD, since the direction of change remained the same across labeled and label-free datasets. Some of these proteins were found to be involved in synaptic signaling, apoptotic and proteasomal protein catabolic processes. Proteins that showed no statistically significant difference between AD and controls as well as the ones that were not identified in the original dataset are indicated by the white boxes.

downregulated and 11 consistently upregulated proteins emerging as the most robust changes occurring in AD brain (Fig. 3b). Importantly, by performing this large dataset analysis, we were able to find novel proteins involved in AD, some of which were involved in synaptic or cell adhesion pathways, which will be discussed below. For instance, we detected upregulation in SLC20A2, which is involved in phosphate transport by absorbing phosphate from interstitial fluid. Interestingly, mutations in this gene are reported to cause primary familial brain calcification, which are often associated with movement disorders<sup>23</sup>. In addition, our findings validate the previously reported data on AD-related proteins. For instance, the astrocytic marker glial fibrillary acidic protein (GFAP) is well-known to be upregulated in AD brain as part of concurrent gliosis. Clusterin (CLU) is identified as a genetic risk factor for sporadic AD<sup>24,25</sup>. Compelling evidence suggests that CLU binds to the amyloid  $\beta$ -peptide (A $\beta$ ) and affects its deposition and clearance (see review by Foster *et al.*<sup>26</sup>). Interestingly, Wojtas *et al.* recently demonstrated that CLU not only affects A $\beta$  pathology<sup>27</sup> but also tau pathology<sup>28</sup>. Another interesting protein from the heatmap is the mitochondrial protein DNM1L (also known as DRP1) which is vital for mitochondrial fission events. Several studies reported a relationship between A $\beta$ -mediated toxicity, DRP1 function and mitochondrial dysfunction<sup>29,30</sup>. In addition, a recent study suggested that DRP1 is required for proper synaptic function in CA1 hippocampal neurons<sup>31</sup>. Several proteins related to synaptic signaling were consistently altered in AD brain across datasets. Among them, synaptotagmin-1 (SYT1) and synaptotagmin-7 (SYT7), acting as calcium sensors triggering neurotransmitter release, are reported to interact with amyloid precursor protein (APP) and presenilin 1 (PS1), a component of  $\gamma$ -secretase that mediates the final cut in APP processing<sup>32-35</sup>. Hyperpolarization-activated cyclic nucleotide-gated channel 1 (HCN1), important for regulating neuronal activity, has been studied extensively in epilepsy research but not much in AD research. However, decreased



**Fig. 4** Summary of the gene set analysis. The network plots show examples of the significantly altered pathways from KEGG database (FDR < 5%) in the (a) labeled and (b) label-free datasets.

levels of HCN1 were in line with our finding observed in AD brain, and loss of HCN1 gene function was reported to increase A $\beta$  levels in mice brain<sup>36</sup>. Recent studies identified dipeptidyl aminopeptidase-like protein 6 (DPP6), which modulates the activity of potassium channels, as a novel genetic factor contributing to dementia<sup>37,38</sup>. Functional studies suggested that DPP6 is important for synaptic structure, hippocampus-dependent learning, and memory<sup>39</sup>. Our meta-analysis confirms that DPP6 is of importance for AD pathogenesis and merits further studies.

While the observations on an individual protein level are interesting, we sought to test whether and how the meta-analyses of labeled and label-free datasets would show similar patterns in a wider perspective. Notably, our findings indicate that common biological processes were dysregulated in AD brain, even though the individual protein alterations detectable were clearly different between the two MS approaches. For instance, synaptic processes, such as synaptic vesicle cycle (KEGG hsa04721) and chemical synaptic transmission (GO:0007268), were significantly downregulated in AD. This is not surprising considering that synaptic dysfunction occurs early in AD pathogenesis and strongly correlates with cognitive decline<sup>40</sup>. Examples of synaptic proteins involved in these processes included components of the adaptor protein complex 2 (AP2A2 and AP2M1), disk large-associated protein 1 (DLGAP1), glutamate receptors (GRIA2–4 and GRM3), neuronal pentraxin-2 (NPTX2), RAB3A, SNAP25, SYT1 and V-ATPase subunit F (ATP6V1F). As illustrated by the heatmap, the synaptic proteins NPTX2 and SYT1 were consistently decreased in AD brain across studies, representing one of the most robust changes. Interestingly, AP2A2, involved in clathrin-dependent endocytosis, was among the proteins contributing most to the clustering pattern observed on PCA plot. Many studies have reported reduction in synaptic protein expression (e.g., SNAP25, GRIA2) in AD brain and shown that such decrease was correlated with increased rate of cognitive decline<sup>8,41,42</sup>. Besides the above-mentioned synaptic proteins, we also detected robust downregulation in the levels of GFRA2, which is a receptor for the neurotrophic factor called neurturin. A recent genome-wide association study identified the GFRA2 locus as a potential modifier risk of frontotemporal dementia, proposing that GFRA2-related processes may hold a potential as therapeutic targets<sup>43</sup>. Unlike other synaptotagmins, SYT11 does not bind to calcium while mediating vesicle trafficking<sup>44</sup>, and it is also interesting that SYT11 is the only upregulated synaptic proteins in our meta-analyses.

Another prominently downregulated pathway was related to oxidative phosphorylation (KEGG hsa00190) and mitochondrial energy metabolism. Mainly proteins of the electron transport chain (e.g., NDUFS1, NDUFS7, NDUFA7) were associated with these pathways. Altered mitochondrial dynamics and bioenergetic metabolism are commonly observed in neurodegenerative disorders<sup>45</sup>. This is not surprising since activities such as synaptic transmission and synaptic vesicle cycle demand high metabolic energy that is tightly regulated by mitochondria<sup>46</sup>.

It is noteworthy that HD was the pathway with highest number of protein hits. In this regard, it is interesting to note that our previous immunohistochemical studies have shown increased levels of huntingtin in pyramidal neurons both in AD brain<sup>47</sup> and in an AD mouse model (APP-NLF)<sup>48</sup>. Many of the altered proteins were mitochondrial, particularly those involved in the respiratory chain. There were also significant effects on proteins involved in clathrin-mediated endocytosis, and for instance postsynaptic receptors and signaling molecules. These findings support that HD and AD partially share disease pathways involved in mitochondrial function, clathrin-mediated endocytosis, postsynaptic function and signaling.

Several pathways were found to be significantly upregulated in AD, such as complement and coagulation cascade (KEGG hsa04610), cytokine-mediated signaling (GO:0019221), which reflects the presence of neuroinflammation, ECM-receptor interaction (KEGG hsa04512) and ECM organization (GO:0030198), indicating changes in the ECM. The ECM provides structural support that is essential for the cells and regulates many cellular processes such as neurogenesis, axonal outgrowth, synaptic plasticity, and immune response<sup>49</sup>. In line with previous findings<sup>15,50</sup>, we mainly detected upregulation of the ECM components, e.g., collagen, CD44, tenascin, integrin alpha-6, versican core protein and fibronectin-1. In addition, we detected robustly increased levels of UTRN (or dystrophin-related protein 1), which mediates the interaction between the plasma membrane, the cytoskeleton and the ECM.

Our study is the largest meta-analysis study analyzing proteomic data from multiple research centers. Thus far, only a few studies compared proteomic datasets in a systematic manner and identified several proteins as promising targets for maintenance of cognitive resilience<sup>42,51</sup>. Similar to the observations from Wingo *et al.*<sup>42</sup>, we found decreased levels of synaptic (e.g., GRIA2, AP2A2, AP2B1, BAIAP2, DMXL2, DLG4, SYNPO) and mitochondrial proteins (e.g., NDUFS1, PDHA1) but increased levels of proteins involved in myelination and apoptosis processes (e.g., GFAP, GSN, NEFL) in AD brain. Altogether these findings corroborate the involvement of these proteins and processes in disease pathogenesis.

Considering that proteomics data extensively relies on experimental setups such as sample preparation<sup>52</sup>, clinical characteristics of subjects, statistical tests of choice or even cut-offs for statistical significance, a meta-analysis could provide valuable and coherent information that a single dataset analysis could not. However, this study is not without challenges. From a statistical perspective, the random-effects-model could compute the mean difference between AD and control groups, even when a given protein was originally quantified in only one dataset or when proteins had originally opposite directional changes. This is possible because the statistical model gives larger weight to the datasets that contain larger effect size, larger sample size, and lower within-study variability. With that said, from a biological perspective, it is important and relevant to pinpoint proteins with consistent alterations across individual datasets, as they could represent the most robust alterations associated with disease. In this study, we address both perspectives by reporting all significantly altered proteins in the labeled ( $n = 757$ ) and the label-free ( $n = 1,195$ ) datasets as well as the robustly altered ones ( $n = 33$ ). It is important to note that our study does not allow any direct comparison between labeled and label-free MS approaches, since the same biological samples were not studied using both methods. However, the observation of two distinct clusters pertaining to labeled and label-free datasets provided good reason for our stratification wherein we performed two different meta-analyses. At the protein level, our findings suggest that different sets of proteins were significantly altered in the labeled dataset compared to the label-free dataset. It is thus important to emphasize that this could be influenced by how each individual study has dealt with protein inference, which could overestimate the number of distinct protein IDs as well as underestimate commonalities between the datasets. We also noted that the total number of proteins (17,296 vs 4,294 proteins) and sample size (225 vs 632 cases) were rather different between the two methods, which could influence the meta-analysis outcome. Nevertheless, we observed consistent alterations at the pathway level. The analysis of raw MS data in a common pipeline, using the same database search engines for protein identification and quantification, could dissipate some of the methodological differences between the studies, however, this is not without challenges especially when applied to such quantities of data.

Given the complex and multifactorial nature of AD, disentangling pathological mechanisms is of the utmost importance for development of treatment strategies for this debilitating disorder. This study provides a comprehensive analysis of 18 proteomic datasets and offers novel insights into single protein alterations related to AD. Depending on the MS method of choice (labeled vs label-free), different proteins appeared to be significantly dysregulated in AD. Nonetheless, pathway analyses of labeled and label-free MS datasets illustrated that processes related to synaptic signaling, oxidative phosphorylation, immune response and ECM were commonly dysregulated in AD. These observations are largely in line with previous reports, indicating that these mechanisms are central to AD pathogenesis.

## Methods

**Study selection and search strategy.** Literature search on MS-based proteomic studies in AD was done using the following string “Alzheimer\* AND proteome AND (proteomic OR mass spectrometry) AND “human brain”” on April 2, 2020 in PubMed (<https://pubmed.ncbi.nlm.nih.gov/>). PRISMA statement was followed during study design<sup>53</sup>. Out of 95 articles, 16 datasets were manually retrieved and curated from high-quality studies performed in seven different research groups<sup>6–12,14,19</sup> (Online-only Table 1), based on the following criteria: (i) sample size of at least five cases per diagnosis group (AD and control), (ii) minimum number of 1000 quantified proteins, (iii) studies in which quantified protein intensities were reported for each individual biological sample, (iv) studies that are publicly available in repositories, and (v) studies in which the proteome of frontal and temporal cortices, severely affected regions by AD-related pathology, was analyzed. In addition, two unpublished datasets (Schedin-Weiss *et al.* in preparation) from our research group were included in this study, as they met with the above-mentioned inclusion criteria (for pre-processed datasets, see<sup>20</sup>). Lastly, studies using pooled biological samples or performing targeted MS approach were excluded from the meta-analysis.

Neuropathological evaluation of neurofibrillary tangles (Braak staging)<sup>54</sup> and neuritic plaques (CERAD criteria)<sup>55</sup> was done in all postmortem human brain tissue included in the above-mentioned studies. While AD cases had often higher Braak stages (IV–VI), control subjects presented little or no pathological alterations beyond normal age-appropriate changes (0–III). Cases with non-AD pathological changes were reported to be excluded in these studies.

**Data pre-processing.** The protein identifiers were manually set to UniProtKB accession numbers in each dataset. Each dataset was standardized in the same manner by allowing 20% of missing values per group, converting protein intensities to log<sub>2</sub> scale, applying median-centered normalization and scaling (median 0 and standard deviation 1) so that they would be comparable for further statistical analyses. Subsequently, sample distributions were visually inspected by PCA for each dataset as well as for the concatenated data, consisting of 533 proteins quantified in all 857 samples and all 18 datasets. All analyses were performed in R (version 4.0.1).

**Meta-analysis by random-effects-model.** The meta-analysis was performed using the meta package (version 4.13) and the metacont function<sup>56</sup>. A random-effects-model, using the DerSimonian and Laird method<sup>57</sup>, was selected, due to methodological differences noted between the selected studies, such as clinical characteristics of postmortem human samples, brain region, sample preparation and LC-MS/MS experiments (Online-only Tables 1 and 2). For each protein, the effect size (i.e., mean difference between AD and control) was computed by taking into account the following parameters: mean of protein intensities per protein per group, standard deviation of protein intensities per protein per group and sample size per group. Particularly, this allowed accounting for group differences even when proteins showed opposite directional changes between datasets or had not been quantified in all datasets. The p-values were corrected for multiple hypothesis testing using Benjamini-Hochberg method and considered significant at  $\leq 10\%$  FDR.

**Gene set analysis.** To gain better insights into the biological processes underlying AD pathogenesis, GSA was performed using the piano R package<sup>21</sup>. Most of the selected proteomic studies were protein-centric, thereby containing the isoform-specific information. Prior to GSA, proteins from the meta-analyses were converted to genes, giving rise to multiple values for the same gene. If isoforms showed consistent mean difference in AD, those with the least significant changes were then excluded from GSA. On the other hand, all isoforms that did not follow the same direction of change were excluded from GSA. Thus, a reduced number of genes were subjected to GSA, compared to the protein list obtained from the meta-analysis<sup>20</sup>. Subsequently, the UniProtKB accession numbers were converted to gene symbol using the UniProt Retrieve/ID mapping tool. All genes together with mean difference and p-value, which were computed by the random-effects-model, were subjected to GSA, thus allowing for a comparison between significant alterations and all identified genes (background) from our meta-analysis data. Two databases, KEGG 2019 and GO biological process, were used as retrieved from Enrichr<sup>58,59</sup>. Minimum number of genes associated with a given geneset was set to five. To get a better understanding of how pathways could be dysregulated in AD, distinct upregulation and distinct downregulation categories were selected. Pathways with FDR < 5% were considered as statistically significant.

### Data availability

The raw data of the following datasets can be found at the ProteomeXchange Consortium via the PRIDE partner repository or the Synapse Web Portal: Dataset 1 (PXD014557)<sup>14,60</sup>, Dataset 4 (PXD006122)<sup>8,61</sup>, Datasets 5–8 (PXD010138)<sup>19,62</sup>, Dataset 9 (PXD007160)<sup>12,63</sup>, Dataset 10 (syn16816734)<sup>10,64</sup>, Dataset 11 (syn21441771)<sup>6,65</sup>, Dataset 12 (syn21441782)<sup>6,66</sup>, Dataset 13 (syn21444768)<sup>6,67</sup>, Datasets 14–15 (PXD007985)<sup>9,68</sup>, Dataset 17 (PXD008739)<sup>11,69</sup>, and Dataset 18 (PXD008806)<sup>11,70</sup>. In addition, Dataset 16 was retrieved from the supplementary data provided in the original publication<sup>7</sup>. Lastly, pre-processed data from Datasets 2 and 3 are available at the figshare database<sup>20</sup>.

To support data sharing and reusability, all fully processed individual datasets are available in github, which can be accessed by [https://github.com/hazhay/Meta-analysis\\_AD](https://github.com/hazhay/Meta-analysis_AD). In addition, input datasets used for meta-analyses and for GSAs as well as their findings are available at the figshare database associated with this article<sup>20</sup>.

### Code availability

The R script used in this research is publicly available and can be found in github ([https://github.com/hazhay/Meta-analysis\\_AD](https://github.com/hazhay/Meta-analysis_AD)).

Received: 4 May 2021; Accepted: 21 October 2021;

Published online: 03 December 2021

### References

1. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical chemistry* **75**, 1895–1904, <https://doi.org/10.1021/ac0262560> (2003).
2. Wiese, S., Reidegeld, K. A., Meyer, H. E. & Warscheid, B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* **7**, 340–350, <https://doi.org/10.1002/pmic.200600422> (2007).
3. Ow, S. Y. *et al.* iTRAQ Underestimation in Simple and Complex Mixtures: “The Good, the Bad and the Ugly”. *Journal of Proteome Research* **8**, 5347–5355, <https://doi.org/10.1021/pr900634c> (2009).
4. Zhu, W., Smith, J. W. & Huang, C.-M. Mass Spectrometry-Based Label-Free Quantitative Proteomics. *Journal of Biomedicine and Biotechnology* **2010**, 840518, <https://doi.org/10.1155/2010/840518> (2010).
5. Winblad, B. *et al.* Defeating Alzheimer’s disease and other dementias: a priority for European science and society. *The Lancet. Neurology* **15**, 455–532, [https://doi.org/10.1016/s1474-4422\(16\)00062-4](https://doi.org/10.1016/s1474-4422(16)00062-4) (2016).
6. Johnson, E. C. B. *et al.* Large-scale proteomic analysis of Alzheimer’s disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nature medicine* **26**, 769–780, <https://doi.org/10.1038/s41591-020-0815-6> (2020).



7. Zhang, Q. *et al.* Integrated proteomics and network analysis identifies protein hubs and network alterations in Alzheimer's disease. *Acta Neuropathologica Communications* **6**, 19, <https://doi.org/10.1186/s40478-018-0524-2> (2018).
8. Berezcki, E. *et al.* Synaptic markers of cognitive decline in neurodegenerative diseases: a proteomic approach. *Brain: a journal of neurology* **141**, 582–595, <https://doi.org/10.1093/brain/awx352> (2018).
9. Bai, B. *et al.* Deep Multilayer Brain Proteomics Identifies Molecular Networks in Alzheimer's Disease Progression. *Neuron* **105**, 975–991.e977, <https://doi.org/10.1016/j.neuron.2019.12.015> (2020).
10. Johnson, E. C. B. *et al.* Deep proteomic network analysis of Alzheimer's disease brain reveals alterations in RNA binding proteins and RNA splicing associated with disease. *Molecular neurodegeneration* **13**, 52, <https://doi.org/10.1186/s13024-018-0282-4> (2018).
11. Xu, J. *et al.* Regional protein expression in human Alzheimer's brain correlates with disease severity. *Communications biology* **2**, 43, <https://doi.org/10.1038/s42003-018-0254-9> (2019).
12. Ping, L. *et al.* Global quantitative analysis of the human brain proteome in Alzheimer's and Parkinson's Disease. *Scientific data* **5**, 180036, <https://doi.org/10.1038/sdata.2018.36> (2018).
13. Seyfried, N. T. *et al.* A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease. *Cell systems* **4**, 60–72.e64, <https://doi.org/10.1016/j.cels.2016.11.006> (2017).
14. Haytural, H. *et al.* The Proteome of the Dentate Terminal Zone of the Perforant Path Indicates Presynaptic Impairment in Alzheimer Disease. *Molecular & cellular proteomics: MCP* **19**, 128–141, <https://doi.org/10.1074/mcp.RA119.001737> (2020).
15. Hondius, D. C. *et al.* Profiling the human hippocampal proteome at all pathologic stages of Alzheimer's disease. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **12**, 654–668, <https://doi.org/10.1016/j.jalz.2015.11.002> (2016).
16. Hashimoto, M. *et al.* Analysis of microdissected neurons by 18O mass spectrometry reveals altered protein expression in Alzheimer's disease. *Journal of cellular and molecular medicine* **16**, 1686–1700, <https://doi.org/10.1111/j.1582-4934.2011.01441.x> (2012).
17. Söderberg, L. *et al.* Analysis of Single Alzheimer Solid Plaque Cores by Laser Capture Microscopy and Nano-electrospray/Tandem Mass Spectrometry. *Biochemistry* **45**, 9849–9856, <https://doi.org/10.1021/bi060331+> (2006).
18. Drummond, E. *et al.* Proteomic differences in amyloid plaques in rapidly progressive and sporadic Alzheimer's disease. *Acta neuropathologica* **133**, 933–954, <https://doi.org/10.1007/s00401-017-1691-0> (2017).
19. Mendonça, C. F. *et al.* Proteomic signatures of brain regions affected by tau pathology in early and late stages of Alzheimer's disease. *Neurobiology of disease* **130**, 104509, <https://doi.org/10.1016/j.nbd.2019.104509> (2019).
20. Haytural, H. *et al.* Insights into the changes in the proteome of Alzheimer disease elucidated by a meta-analysis. *figshare* <https://doi.org/10.6084/m9.figshare.c.5408946> (2021).
21. Våremo, L., Nielsen, J. & Nookaew, I. Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Research* **41**, 4378–4391, <https://doi.org/10.1093/nar/gkt111> (2013).
22. Ahrné, E. *et al.* Evaluation and Improvement of Quantification Accuracy in Isobaric Mass Tag-Based Protein Quantification Experiments. *Journal of Proteome Research* **15**, 2537–2547, <https://doi.org/10.1021/acs.jproteome.6b00066> (2016).
23. Wang, C. *et al.* Mutations in SLC20A2 link familial idiopathic basal ganglia calcification with phosphate homeostasis. *Nature genetics* **44**, 254–256, <https://doi.org/10.1038/ng.1077> (2012).
24. Harold, D. *et al.* Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature genetics* **41**, 1088–1093, <https://doi.org/10.1038/ng.440> (2009).
25. Lambert, J. C. *et al.* Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nature genetics* **41**, 1094–1099, <https://doi.org/10.1038/ng.439> (2009).
26. Foster, E. M., Dangla-Valls, A., Lovestone, S., Ribe, E. M. & Buckley, N. J. Clusterin in Alzheimer's Disease: Mechanisms, Genetics, and Lessons From Other Pathologies. **13**, <https://doi.org/10.3389/fnins.2019.00164> (2019).
27. Wojtas, A. M. *et al.* Astrocyte-derived clusterin suppresses amyloid formation *in vivo*. *Molecular neurodegeneration* **15**, 71, <https://doi.org/10.1186/s13024-020-00416-1> (2020).
28. Wojtas, A. M. *et al.* Clusterin ameliorates tau pathology *in vivo* by inhibiting fibril formation. *Acta Neuropathol Commun* **8**, 210, <https://doi.org/10.1186/s40478-020-01079-1> (2020).
29. Manczak, M., Calkins, M. J. & Reddy, P. H. Impaired mitochondrial dynamics and abnormal interaction of amyloid beta with mitochondrial protein Drp1 in neurons from patients with Alzheimer's disease: implications for neuronal damage. *Human molecular genetics* **20**, 2495–2509, <https://doi.org/10.1093/hmg/ddr139> (2011).
30. Cho, D. H. *et al.* S-nitrosylation of Drp1 mediates beta-amyloid-related mitochondrial fission and neuronal injury. *Science (New York, N.Y.)* **324**, 102–105, <https://doi.org/10.1126/science.1171091> (2009).
31. Shields, L. Y. *et al.* Dynamin-related protein 1 is required for normal mitochondrial bioenergetic and synaptic function in CA1 hippocampal neurons. *Cell death & disease* **6**, e1725, <https://doi.org/10.1038/cddis.2015.94> (2015).
32. Kuzuya, A. *et al.* Identification of the novel activity-driven interaction between synaptotagmin 1 and presenilin 1 links calcium, synapse, and amyloid beta. *BMC biology* **14**, 25, <https://doi.org/10.1186/s12915-016-0248-3> (2016).
33. Zoltowska, K. M. *et al.* Dynamic presenilin 1 and synaptotagmin 1 interaction modulates exocytosis and amyloid  $\beta$  production. *Molecular neurodegeneration* **12**, 15, <https://doi.org/10.1186/s13024-017-0159-y> (2017).
34. Gautam, V., D'Avanzo, C., Berezovska, O., Tanzi, R. E. & Kovacs, D. M. Synaptotagmins interact with APP and promote A $\beta$  generation. *Molecular neurodegeneration* **10**, 31, <https://doi.org/10.1186/s13024-015-0028-5> (2015).
35. Barthelet, G. *et al.* Presenilin-mediated cleavage of APP regulates synaptotagmin-7 and presynaptic plasticity. *Nature communications* **9**, 4780, <https://doi.org/10.1038/s41467-018-06813-x> (2018).
36. Saito, Y. *et al.* Hyperpolarization-activated cyclic nucleotide gated channels: a potential molecular link between epileptic seizures and A $\beta$  generation in Alzheimer's disease. *Molecular neurodegeneration* **7**, 50, <https://doi.org/10.1186/1750-1326-7-50> (2012).
37. Pottier, C. *et al.* Genome-wide analyses as part of the international FTL-D-TDP whole-genome sequencing consortium reveals novel disease risk factors and increases support for immune dysfunction in FTL-D. *Acta neuropathologica* **137**, 879–899, <https://doi.org/10.1007/s00401-019-01962-9> (2019).
38. Cacace, R. *et al.* Loss of DPP6 in neurodegenerative dementia: a genetic player in the dysfunction of neuronal excitability. *Acta neuropathologica* **137**, 901–918, <https://doi.org/10.1007/s00401-019-01976-3> (2019).
39. Lin, L. *et al.* DPP6 Loss Impacts Hippocampal Synaptic Development and Induces Behavioral Impairments in Recognition, Learning and Memory. *Frontiers in cellular neuroscience* **12**, 84, <https://doi.org/10.3389/fncel.2018.00084> (2018).
40. Colom-Cadena, M. *et al.* The clinical promise of biomarkers of synapse damage or loss in Alzheimer's disease. *Alzheimer's research & therapy* **12**, 21, <https://doi.org/10.1186/s13195-020-00588-4> (2020).
41. Berezcki, E. *et al.* Synaptic proteins predict cognitive decline in Alzheimer's disease and Lewy body dementia. *Alzheimer's & dementia: the journal of the Alzheimer's Association* **12**, 1149–1158, <https://doi.org/10.1016/j.jalz.2016.04.005> (2016).
42. Wingo, A. P. *et al.* Large-scale proteomic analysis of human brain identifies proteins associated with cognitive trajectory in advanced age. *Nature communications* **10**, 1619, <https://doi.org/10.1038/s41467-019-09613-z> (2019).
43. Pottier, C. *et al.* Potential genetic modifiers of disease risk and age at onset in patients with frontotemporal lobar degeneration and GRN mutations: a genome-wide association study. *The Lancet. Neurology* **17**, 548–558, [https://doi.org/10.1016/s1474-4422\(18\)30126-1](https://doi.org/10.1016/s1474-4422(18)30126-1) (2018).
44. Shimojo, M. *et al.* Synaptotagmin-11 mediates a vesicle trafficking pathway that is essential for development and synaptic plasticity. *Genes & development* **33**, 365–376, <https://doi.org/10.1101/gad.320077.118> (2019).

45. Devine, M. J. & Kittler, J. T. Mitochondria at the neuronal presynapse in health and disease. *Nature Reviews Neuroscience* **19**, 63–80, <https://doi.org/10.1038/nrn.2017.170> (2018).
46. Rangaraju, V., Calloway, N. & Ryan, T. A. Activity-driven local ATP synthesis is required for synaptic function. *Cell* **156**, 825–835, <https://doi.org/10.1016/j.cell.2013.12.042> (2014).
47. Axenhus, M., Winblad, B., Tjernberg, L. O. & Schedin-Weiss, S. Huntingtin Levels are Elevated in Hippocampal Post-Mortem Samples of Alzheimer's Disease Brain. *Current Alzheimer research* **17**, 858–867, <https://doi.org/10.2174/1567205017666201203125622> (2020).
48. Schedin-Weiss, S. *et al.* Proteomics Time-Course Study of App Knock-In Mice Reveals Novel Presymptomatic A $\beta$ 42-Induced Pathways to Alzheimer's Disease Pathology. *Journal of Alzheimer's disease: JAD* **75**, 321–335, <https://doi.org/10.3233/jad-200028> (2020).
49. Frantz, C., Stewart, K. M. & Weaver, V. M. The extracellular matrix at a glance. *Journal of cell science* **123**, 4195–4200, <https://doi.org/10.1242/jcs.023820> (2010).
50. Ma, J. *et al.* Extracellular Matrix Proteins Involved in Alzheimer's Disease. *Chemistry (Weinheim an der Bergstrasse, Germany)* **26**, 12101–12110, <https://doi.org/10.1002/chem.202000782> (2020).
51. Yu, L. *et al.* Cortical Proteins Associated With Cognitive Resilience in Community-Dwelling Older Persons. *JAMA Psychiatry*, <https://doi.org/10.1001/jamapsychiatry.2020.1807%J> *JAMA Psychiatry* (2020).
52. Piehowski, P. D. *et al.* Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. *J Proteome Res* **12**, 2128–2137, <https://doi.org/10.1021/pr301146m> (2013).
53. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. **339**, b2535, <https://doi.org/10.1136/bmj.b2535%JBMJ> (2009).
54. Braak, H. & Braak, E. Neuropathological staging of Alzheimer-related changes. *Acta neuropathologica* **82**, 239–259, <https://doi.org/10.1007/BF00308809> (1991).
55. Mirra, S. S. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* **41**, 479–486, <https://doi.org/10.1212/wnl.41.4.479> (1991).
56. Harrer, M., Cuijpers, P. & Ebert, D. Doing Meta-Analysis in R. *Zenodo* <https://doi.org/10.5281/zenodo.2551803> (2019).
57. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Controlled Clinical Trials* **7**, 177–188, [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2) (1986).
58. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90–97, <https://doi.org/10.1093/nar/gkw377> (2016).
59. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics* **14**, 128, <https://doi.org/10.1186/1471-2105-14-128> (2013).
60. Mermelekas, G. Microdissected brain tissue Alzheimer's. *PRIDE Archive* <https://identifiers.org/pride.project:PXD014557> (2020).
61. Branca, R. Proteomics on post-mortem human brain tissue of patients with Alzheimer's, Parkinson's and Lewy body dementias. *PRIDE Archive* <https://identifiers.org/pride.project:PXD006122> (2017).
62. Rezeli, M. Human brain Alzheimer's. *PRIDE Archive* <https://identifiers.org/pride.project:PXD010138> (2019).
63. Duong, D. *Global quantitative analysis of the human brain proteome in Alzheimer's and Parkinson's Disease*. <https://identifiers.org/pride.project:PXD007160> (2018).
64. Dammer, E. AD Splicing Changes - Johnson Mol Neurodegen 2018. *Synapse* <https://doi.org/10.7303/syn16816734> (2018).
65. Dammer, E. The Consensus Brain Protein Coexpression (ConsensusProteinCoexpression) Study - Main Arm (Consensus Network). *Synapse* <https://www.synapse.org/#!Synapse:syn21441771> (2019).
66. Dammer, E. The Consensus Brain Protein Coexpression (ConsensusProteinCoexpression) Study - Alternative Brain Region Analysis - Mayo Temporal Cortex. *Synapse* <https://www.synapse.org/#!Synapse:syn21441782> (2019).
67. Dammer, E. The Consensus Brain Protein Coexpression (ConsensusProteinCoexpression) Study - Univ. Pennsylvania Multi-disease Specificity Cohort - Discovery LFQ Proteomics. *Synapse* <https://www.synapse.org/#!Synapse:syn21444768> (2019).
68. Wang, X. Deep Multilayer Tissue Proteomics Identifies Molecular Networks in Alzheimer's Disease Progression, human data. *PRIDE Archive* <https://identifiers.org/pride.project:PXD007985> (2019).
69. Unwin, R. D. Localised protein expression changes in different regions of human Alzheimer's brain correlate with disease severity - Hippocampus. *PRIDE Archive* <https://identifiers.org/pride.project:PXD008739> (2019).
70. Unwin, R. D. Localised protein expression changes in different regions of human Alzheimer's brain correlate with disease severity - Entorhinal cortex. *PRIDE Archive* <https://identifiers.org/pride.project:PXD008806> (2019).

## Acknowledgements

The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Project SNIC 2020/15-164. RDU would like to thank the families and patients who supported this research by donation of brains to the New Zealand Neurological Foundation Human Brain Bank. This project has received funding from Swedish Research Council, Alzheimerfonden, Gun och Bertil Stohnes Stiftelse, Stiftelsen för Gamla Tjänarinnor, Demensfonden, Hjärnfonden, Margaretha af Ugglas' foundation and Fondation Recherche sur Alzheimer. RDU was supported by Alzheimer's Research UK (ARUK-PPG2014B-7), Manchester NIHR Biomedical Research Centre and the Greater Manchester Comprehensive Local Research Network and the Stoller Biomarker Discovery Centre which was established with an award from the Medical Research Council (MR/M008959/1). Support for this research was provided by funding from the National Institute on Aging (R01AG053960, R01AG061800, RF1AG057471, RF1AG057470, R01AG057339, RF1AG062181), the Accelerating Medicine Partnership for AD (U01AG046161 and U01AG061357).

## Author contributions

H.H., S.S.W., B.W., B.M.T., P.J.V., S.F. and L.O.T. designed the study. S.S.W., E.B., M.R., R.D.U., X.W., E.B.D., E.C.B.J., N.T.S. and L.O.T. previously collected data, which were used in this study. R.B. provided statistical support. H.H. and R.B. analysed the data, and together with L.O.T. interpreted the data. HH wrote the manuscript. All authors contributed to the interpretation of results and writing or revision of the manuscript.

## Funding

Open access funding provided by Karolinska Institute.

## Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to H.H. or L.O.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021