Serum metabolic signatures for Alzheimer’s Disease reveal alterations in amino acid composition and energy metabolism – A validation study

Jonas Ellegaard Nielsen
Aalborg University Hospital

Trygve Andreassen
Norwegian University of Science and Technology

Charlotte Held Gotfredsen
Technical University of Denmark

Dorte Aalund Olsen
Vejle Sygehus

Karsten Vestergaard
Aalborg University Hospital

Jonna Skov Madsen
Vejle Sygehus

Søren Risom Kristensen
Aalborg University Hospital

Shona Pedersen (spedersen@qu.edu.qa)
Qatar University, QU Health

Research Article

Keywords: Alzheimer, Metabolites, Biomarker, Blood, Serum, Nuclear Magnetic Resonance, Single Molecule Array

Posted Date: April 5th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2765583/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Background: Alzheimer’s Disease (AD) is complex and novel approaches are urgently needed to characterise disease pathology and to aid in diagnosis. Metabolites are the end-products of upstream molecular alterations, whereby small changes at the genetic or protein level result in substantial changes at the metabolite level. Blood is frequently used as a source for biomarkers; however, its complexity prevents proper detection. The analytical power of metabolomics, coupled with statistical tools, can assist in reducing this complexity. Furthermore, the current bottleneck in biomarker research is reproducibility and appropriate validation. Thus, we sought to validate a previously proposed panel of metabolic blood-based biomarkers for AD and expand our understanding of the pathological mechanisms involved in AD that are reflected in the blood.

Methods: In the validation cohort serum and plasma were collected from 25 AD patients and 25 healthy controls. Serum was analysed for metabolites using nuclear magnetic resonance (NMR) spectroscopy, while plasma was tested for markers of neuronal damage and AD hallmark proteins using single molecule array (SIMOA). A combination of multivariate and univariate statistics were utilized to validate established biomarkers and uncover new disease-related evidence.

Results: The diagnostic performance of the proposed metabolite biomarker panel was confirmed using sparse-partial least squares discriminant analysis (sPLS-DA) with an area under the curve (AUC) of 0.89 (95 % confidence interval: 0.79 – 0.98). Five metabolites (pyruvic acid, valine, leucine, histidine, and isoleucine) were consistently reduced in both the discovery and validation cohorts. Pathway analysis of significantly altered metabolites in the validation set revealed that they are involved in branched-chain amino acids (BCAAs) and energy metabolism (glycolysis and gluconeogenesis). Additionally, a moderate correlation was observed between valine and the proteins neurolament light and glialbrillary acidic protein. By combining the significant protein expression levels measured by SIMOA with the sPLS-DA model, the AUC increased to 0.97 (95 % CI: 0.93 – 1.00).

Conclusions: Our proposed panel of metabolites was successfully validated using a combined approach of NMR and sPLS-DA. It was discovered that cognitive-impairment-related metabolites belong to BCAAs and are involved in energy metabolism.

1. Background

Neurodegenerative diseases, such as Alzheimer's Disease (AD), account for a significant proportion of mortality, morbidity, and healthcare cost globally (1). Clinical examination alone is inadequate for guiding diagnosis, prognosis, and monitoring progress in research, clinical practice, and drug development. Imaging and biomarkers can aid diagnostics by providing an objective indicator of the underlying pathology. In the case of AD, this includes structural, functional, and molecular imaging, as well as measurements of signature proteins in the cerebrospinal fluid (CSF), i.e. amyloid-β (Aβ) and tau isoforms (2, 3). In certain clinical situations, CSF levels of neurofilament light (Nf-L) protein, a marker for neuronal
injury, have been used to diagnose neurodegenerative disorders (4). However, with these current diagnostic methods, several drawbacks have to be accounted for, limiting their applicability as first-line screening tools. Although technological advances could increase the precision of these methods, their expense and lack of patient compliance prevent this from occurring. In addition, advanced scanning methods, including positron emission tomography, are expensive and less accessible for general practitioners (5), while CSF collection through a lumbar puncture is invasive (6). A blood sample may provide a matrix that could outweigh the drawbacks of the currently used biomarkers to diagnose patients with AD. With the benefits of blood being a biofluid in close contact with every organ in the body, its composition could reflect the potential state of the surrounding organ (7). The blood-brain barrier (BBB) separates the central nervous system (CNS) from the periphery, allowing only gaseous exchange, together with small ions, water- and small liposoluble-molecules to pass (8). However, during AD pathogenesis, the BBB becomes permeable (9), potentially enabling the identification of neuronal metabolites in blood samples.

Even though blood provides a non-invasive biological matrix for investigating disease pathology, its complexity impedes the findings of potential new biomarkers. The omics-era has aided in the realisation of the need to explore such complex samples, with metabolomics being one of the more recent members of the omics family (10). Metabolomics covers the study of all metabolites in a cell, organ, or organism. Metabolites are small molecules < 1,500 Da and comprise amino acids, lipids, peptides, vitamins, etc. (11), and are endpoints of the regulations at the genetic, transcript, and protein levels. Thus, small alterations of upstream molecules could substantially affect the concentration of a metabolite (12). Not only can disease progression cause metabolic perturbations, but environmental factors, treatments, and nutrition also play a role (13). As for clinical applications, metabolic pathways have been shown to be evolutionarily conserved across species, thus bridging the gap between animal studies and human clinical trials (14). Nuclear magnetic resonance (NMR) spectroscopy is among the most informative techniques for studying metabolomics (15), and is able to efficiently analyse and detect hundreds of small molecules in a single measurement in human samples, including plasma and serum (16). Essentially, all metabolites present their own unique and reproducible NMR signature and thus can be used to explore metabolic processes and screen for the presence of known metabolites (17). In addition, NMR is also non-destructive and more informative than other techniques, such as mass spectrometry; however, it lacks sensitivity and requires a larger quantity of sample material (13, 14).

Although biomarker studies contribute to the global search for a solution to the growing problem of the aging population, the literature demonstrates that replication efforts for many promising biomarker findings have failed (18). Thus, the present study aimed to validate suggested metabolite biomarkers presented in our previous discovery study (19) and to provide additional information on metabolic perturbations associated with cognitive impairment. Using NMR-based metabolomics, the serum metabolic signatures from patients with mild to moderate AD were compared to those of cognitively healthy individuals. Furthermore, we validated our initial model using a larger validation cohort by incorporating the commonly identified metabolites and supplementing our findings with additional metabolic perturbations.
2. Methods

2.1. Study demographics

In our previous discovery study, 20 participants were enrolled, with 10 healthy controls and 10 patients with mild to moderate AD. For this validation study, we increased the number of participants to 50, with 25 in each group either diagnosed with mild to moderate AD or as healthy controls. The patients were recruited from the Department of Neurology, Aalborg University Hospital. Recruitment was performed consecutively at the time of diagnosis for the patients and prior to starting their treatment regimen. The diagnosis was based on the following criteria; the International Classification of Diseases and Related Health Problems 10th Edition (ICD\textsubscript{10}) (20), and the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) (21). Paraclinical measurements comprised of Mini-Mental State Examination (MMSE), Addenbrooke's Cognitive Examination (ACE), Functional Activities Questionnaire (FAQ), as well as Aβ, phospho-tau (p-tau), and total-tau (t-tau) measured in CSF. The paraclinical measurements were included when deemed necessary due to diagnostic uncertainty. Age- and gender-related donors were recruited from the blood bank at Aalborg University Hospital to serve as a comparison group for AD patients. In Denmark, blood donors are healthy unpaid volunteers without any apparent illnesses. Inclusion criteria for blood donors were an age > 65 years old and completion of a standard blood donor questionnaire describing physical and mental health, such as experiencing memory impairment, fatigue, or chest pain. Prior to inclusion in the study, all participants signed a written consent form. The study was approved by the local North Denmark Region Committee on Health Research Ethics (N-20150010) and conducted according to the Declaration of Helsinki.

2.2. Sample collection and processing

Blood samples were drawn from study participants and processed as described in a previous study (22). Briefly, blood was collected from the median cubital vein using a 21-gauge needle in 10 mL clot activator tubes (BD Vacutainer, UK) and also 4 mL Ethylenediaminetetraacetic acid (EDTA) tubes (Vacuette, Greiner Bio-One, Austria). After sample collection, the blood was centrifuged twice at 2,500 × g for 15 min at room temperature to obtain serum and plasma. After each centrifugation step, serum and plasma samples were aspirated to approximately 1 cm above the buffy coat or pellet. Finally, serum and plasma samples were aliquoted, snap-frozen using liquid nitrogen, and stored at -80°C until further analyses.

2.3. Routine analyses

Organ markers were routinely measured in serum samples to ensure that none of the study participants had co-morbidities. The clinical biochemistry markers measured were alanine transaminase, albumin, carbamide, cholesterol, creatinine, C-reactive protein, glucose, high and low-density lipoprotein (HDL and LDL, respectively), lactate dehydrogenase (LDH), and triglycerides using the Alinity ci-series (Abbott, Chicago, IL, USA) and haemoglobin using either XN-9000 (Sysmex Europe SE, Germany) or Hb 201 DM (Hemocue AB, Sweden).
2.4. Single Molecule Array

Aβ_{40}, Aβ_{42}, glial fibrillary acidic protein (GFAP), Nf-L, and p-tau181 were measured in EDTA plasma using the respective commercially available kits; Neurology 4-Plex E and P-Tau181 (Quanterix®, Billerica, MA, USA) by Single Molecule Array (SIMOA®) HD-X Analyzer. The analyses were performed according to the manufacturer's instructions. In addition, the manufacturer's commercial controls were applied for quality control.

2.5. Nuclear magnetic resonance spectroscopy

Serum samples were initially thawed for one hour and then carefully diluted 1:1 dilution with sodium phosphate buffer (0.075 M, pH 7.4, 20% D_{2}O in H_{2}O, 6 mM NaN_{3}, 4.6 mM 3-(trimethylsilyl)-2,2,3,3-tetradeteropropionic acid (TSP-d4)) and aliquoted into 5 mm NMR tubes. NMR spectra were recorded using a Bruker Avance Neo 600 MHz spectrometer attached to a BBI probe (Bruker Biospin GmbH, Rheinstetten, Germany). IconNMR (Topspin 4.1.1, Bruker Biospin GmbH, Rheinstetten, Germany) and Samplejet autosampler (Bruker Biospin GmbH, Rheinstetten, Germany) were used for sample handling and data acquisition. One-dimensional nuclear Overhauser effect (1D-NOESY) spectra, together with Carr-Purcell-Meiboom-Gill (CPMG), were recorded at 310 K using parameters for acquisition from Dona et al. (23). For the 1D-NOESY spectra, 96k data points were recorded, with 30 ppm spectral width. In contrast, CPMG spectra were recorded with 72k data points and a spectral width of 20 ppm. For both experiments, 32 scans with water suppression (25 Hz) during relaxation delay (4 s) and mixing time (NOESY, 10 ms) were used for recording. Free induction decays were Fourier transformed after artificial zero fillings up to 128k data points and 0.3 Hz line broadening. In accordance with the manufacturer, B.I.Methods (Bruker Biospin GmbH, Rheinstetten, Germany), reference samples were routinely measured and processed in automation for temperature calibration, water suppression determination, and external quantitative referencing. B.I.Quant-PS™ 2.0 (Bruker Biospin GmbH, Rheinstetten, Germany) was used to automatically quantify metabolites.

2.6. Data analysis

For the validation cohort, 40 metabolites were identified using NMR. Information from the discovery cohort can be found in the discovery study (19). Metabolites were filtered for ≥70% valid values in at least one group before statistical analyses were conducted. Prior to validation of the initial metabolic signature for AD diagnosis, data were log-transformed, auto-scaled, and mean-centered.

Three models; Random Forest, Extreme Gradient Boosting (XGBoost), and sparse-partial least squared discriminant analysis (sPLS-DA), were tested and estimated by their performance using the following parameters; Area under the curve (AUC) and 95% confidence interval (CI) were reported to indicate the sensitivity and specificity of the model, together with the accuracy, positive predictive value (PPV), negative predictive value (NPV), and selected number of important metabolites. The Random Forest model performance was estimated by the out-of-bag error rate and optimal number of features was selected using the randomForest v4.71.1 and Boruta v7.0.0 R packages. XGBoosting was performed
using the R package \textit{xgboost v1.6.0.1}, with performance estimated by root mean squared error (RMSE). Optimal number of features were ranked according to importance score, and selected if importance score was > 0.1. As previously described (19), the sPLS-DA model was build with a 5-fold cross-validation (CV) repeated 100 times using the \textit{mixOmics v6.20.0} R package. The optimal number of selected features was estimated using the classification error rate. For visual purpose scores plot for sample groupings, loadings plot for weighted importance of selected metabolites, and receiver operating characteristic (ROC) curve are presented for the most optimal model. Data were assessed for normality and compared between the groups using a Student’s \textit{t}-test, presented as mean ± standard deviations (SD). Correction for multiple comparison was also provided using the Benjamini-Hochberg false-discovery rate (FDR) corrected \textit{p}-value. Nf-L and GFAP were corrected for age using a linear model (24). A significance level of \textit{p} < 0.05 was chosen. Fold changes (FC) between groups were also calculated for the metabolites, using $\text{FC} = (\text{Metabolite}_{\text{AD}} - \text{Metabolite}_{\text{Con}}) / \text{Metabolite}_{\text{Con}}$. Correlations between important metabolites, selected by multivariate data analysis, and clinical data were investigated using Pearson's \textit{r}, with only the significant correlations presented. Data analysis and graphical representations were conducted using R version 4.2.2.

Network analysis was performed using the MetScape (version 3.1.3) App in CytoScape (version 3.9.1). The network was based on KEGG IDs from significantly altered metabolites between AD patients and healthy individuals. In addition, $\log_2(\text{FC})$ was calculated for the metabolites between healthy and diseased individuals to indicate in which direction the metabolites were altered. Raw NMR data for the validation cohort, clinical data, and input data for the network analysis can be found in Additional files 1, 2, and 3, respectively.

3. Results

3.1. Characteristics of study participants

The biochemical parameters, clinical test results for cognitive performances, corresponding clinical parameters, and SIMOA measurements for both study groups have been summarised in Table 1. Briefly, the majority of the biochemical measurements were within the standard reference intervals. A few, but significant differences were also observed between the AD patients and cognitively healthy individuals, including a slightly higher age ($p = 0.00001$), higher LDH levels ($p = 0.03$), and lower glucose levels in the AD patient group ($p = 0.01$). Patients who required additional cognitive testing and paraclinical measurements were identified, where AD patients presented with low MMSE (20.0 ± 4.5) and ACE (58.0 ± 16.5) scores and a high FAQ (11.8 ± 6.2) score, whereas paraclinical tests demonstrated elevated levels of CSF tau, p-tau (81.7 ± 25.0 ng/L) and t-tau (520.4 ± 102.4 ng/L), and decreased levels of CSF Aβ (682.8 ± 216.3 ng/L) for some of the patients, indicating extracellular tau accumulation and intracellular Aβ build-up. Plasma measurements of markers for neuronal injury and AD hallmark proteins were included as additional clinical information. Generally, AD patients had significantly higher plasma levels of Aβ$_{40}$ ($p =$
GFAP ($p = 0.001$), Nf-L ($p = 0.00001$), and p-tau181 ($p = 0.00005$) than healthy individuals; however, $A\beta_{42}$ did not differ between the two groups ($p = 0.5$).
Table 1
Characteristics of study participants.

<table>
<thead>
<tr>
<th>Units</th>
<th>Con (n = 25)</th>
<th>AD (n = 25)</th>
<th>p-value</th>
<th>Reference interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Years</td>
<td>66.6 ± 1.3</td>
<td>75.7 ± 8.2</td>
<td>0.00001</td>
<td>-</td>
</tr>
<tr>
<td>Male/female n</td>
<td>16 / 9</td>
<td>15 / 10</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Biochemical measurements</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALAT U/L</td>
<td>26.3 ± 8.6</td>
<td>22.3 ± 11.6</td>
<td>0.17</td>
<td>10.0–50.0</td>
</tr>
<tr>
<td>Albumin g/L</td>
<td>41.0 ± 1.9</td>
<td>41.5 ± 1.9</td>
<td>0.37</td>
<td>34–45</td>
</tr>
<tr>
<td>Carbamide mmol/L</td>
<td>5.8 ± 1.3</td>
<td>5.7 ± 1.5</td>
<td>0.77</td>
<td>3.1–8.1</td>
</tr>
<tr>
<td>Cholesterol mmol/L</td>
<td>5.4 ± 0.9</td>
<td>5.5 ± 1.1</td>
<td>0.88</td>
<td>4.2–8.5</td>
</tr>
<tr>
<td>Creatinine µmol/L</td>
<td>79.0 ± 10.2</td>
<td>83.4 ± 14.5</td>
<td>0.22</td>
<td>45–105</td>
</tr>
<tr>
<td>CRP mg/L</td>
<td>1.9 ± 1.4</td>
<td>2.2 ± 2.9</td>
<td>0.57</td>
<td>&lt; 8</td>
</tr>
<tr>
<td>Glucose mmol/L</td>
<td>6.4 ± 1.7</td>
<td>5.4 ± 0.9</td>
<td>0.01</td>
<td>4.2–7.8</td>
</tr>
<tr>
<td>Haemoglobin mmol/L</td>
<td>8.8 ± 0.7</td>
<td>8.5 ± 1.0 (n = 15)</td>
<td>0.45</td>
<td>7.3–10.5</td>
</tr>
<tr>
<td>HDL mmol/L</td>
<td>1.5 ± 0.3</td>
<td>1.6 ± 0.4</td>
<td>0.35</td>
<td>0.7–1.9</td>
</tr>
<tr>
<td>LDL mmol/L</td>
<td>3.2 ± 0.8</td>
<td>3.3 ± 0.9</td>
<td>0.71</td>
<td>2.2–5.7</td>
</tr>
<tr>
<td>LDH U/L</td>
<td>170.2 ± 31.2</td>
<td>192.1 ± 38.7</td>
<td>0.03</td>
<td>105–255</td>
</tr>
<tr>
<td>Triglycerides mmol/L</td>
<td>1.5 ± 0.8</td>
<td>1.3 ± 0.8</td>
<td>0.34</td>
<td>0.6–3.9</td>
</tr>
<tr>
<td><strong>Clinical parameters</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ACE</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>FAQ</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CSF Aβ ng/L</td>
<td>-</td>
<td>682.8 ± 216.3 (n = 9)</td>
<td>-</td>
<td>&gt; 500</td>
</tr>
<tr>
<td>CSF p-tau ng/L</td>
<td>-</td>
<td>81.7 ± 25.0 (n = 9)</td>
<td>-</td>
<td>&lt; 61</td>
</tr>
<tr>
<td>CSF t-tau ng/L</td>
<td>-</td>
<td>520.4 ± 102.4 (n = 9)</td>
<td>-</td>
<td>&lt; 450</td>
</tr>
<tr>
<td><strong>SIMOA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aβ40 pg/mL</td>
<td>95.1 ± 10.2</td>
<td>108.7 ± 17.4</td>
<td>0.002</td>
<td>-</td>
</tr>
</tbody>
</table>
Demographics data of study participants together with biochemical measurements, cognitive test results, paraclinical measurements, and SIMOA measurements. Abbreviations; Aβ – Amyloid-β, ACE – Addenbrooke’s Cognitive Examination, AD – Alzheimer’s Disease, ALAT – Alanine transaminase, p-tau – Phosphorylated tau, CRP – C-reactive protein, CSF – Cerebrospinal fluid, FAQ – Functional Activities Questionnaire, GFAP – Glial fibrillary acidic protein, HDL – High-density lipoprotein, LDH – Lactate dehydrogenase, LDL – Low-density protein, MMSE – Mini-Mental State Examination, Nf-L – Neurofilament light, SD – Standard deviation, SIMOA – Single molecule array, t-tau – Total tau.

3.2. Validation of metabolic signatures for Alzheimer’s Disease diagnostics

To validate the metabolic signature identified in our discovery study, NMR spectroscopy was applied to measure the concentration of serum metabolites in our validation study cohort. Three prediction models were tested for their performance based on AUC, accuracy, PPV, and NPV. These models included sPLS-DA, random forest, and XGBoost (Table 2). Based on these criteria, sPLS-DA showed the highest performance with five selected metabolites building the model (pyruvic acid, valine, histidine, isoleucine, and glutamine), while random forest performed the second best with four selected metabolites (histidine, valine, pyruvic acid, and glutamine), and lastly the XGBoost with three metabolites (histidine, pyruvic acid, and valine). Thus, sPLS-DA was selected as our data’s most optimal validation model.

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC</th>
<th>95% CI</th>
<th>Accuracy</th>
<th>PPV</th>
<th>NPV</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>sPLS-DA</td>
<td>0.89</td>
<td>0.79–0.98</td>
<td>0.86</td>
<td>0.88</td>
<td>0.85</td>
<td>5</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.87</td>
<td>0.77–0.97</td>
<td>0.76</td>
<td>0.84</td>
<td>0.71</td>
<td>4</td>
</tr>
<tr>
<td>XGBoost</td>
<td>0.84</td>
<td>0.73–0.95</td>
<td>0.74</td>
<td>0.72</td>
<td>0.76</td>
<td>3</td>
</tr>
</tbody>
</table>

Three validation models and their diagnostic performance; sparse-partial least squared discriminant analysis, random forest, and extreme gradient boosting. Abbreviations; AUC – Area under the curve, CI –
Confidence interval, NPV – Negative predictive value, PPV – Positive predictive value, sPLS-DA – Sparse-partial least squared discriminant analysis, XGBoost – Extreme gradient boosting.

The validation model showed a small overlap between the patient and control groups, as seen in the scores plot of the measured serum samples (Fig. 1A). Based on the validated model, five metabolites significantly contribute to sample grouping, accounting for 44% of the group variation (Fig. 1B). Consequently, the model had an AUC performance of 0.89 (95% CI = 0.79–0.98) for discriminating AD patients from cognitively healthy individuals (Fig. 1C). Furthermore, the model had an accuracy = 0.86, PPV = 0.88, and NPV = 0.85, indicating its diagnostic value. Interestingly, when adding the significantly altered proteins (Aβ40, GFAP, Nf-L, and p-tau181) to the validation model, improved its diagnostic performance, resulting in an AUC of 0.97 (95% CI = 0.93–1.00) with an accuracy of 0.94, PPV of 0.96, and NPV of 0.92 (Fig. 1D).

Furthermore, the selected panel of five metabolites was correlated against the clinical data and markers of neuronal damage to determine their possible association with neurodegenerative diseases (Fig. 2). Most metabolites exhibited a negative correlation to Nf-L and p-tau181, with valine showing a moderate Pearson’s correlation of -0.51 and - 0.48, respectively. Pyruvic acid and isoleucine also showed moderate correlations with cognitive scoring tests FAQ (ρ = -0.5) and MMSE (ρ = 0.43).

Five of the metabolites measured in both the discovery and validation data sets were significantly altered when comparing healthy and diseased individuals (Table 3). These five metabolites exhibited identical changes in both the discovery and validation studies.
Table 3
Common significantly altered metabolites.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Discovery study</th>
<th>Validation study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Con [mmol/L]</td>
<td></td>
</tr>
<tr>
<td>Pyruvic acid</td>
<td>0.032 0.007</td>
<td>0.118 0.031</td>
</tr>
<tr>
<td>Valine</td>
<td>0.118 0.019</td>
<td>0.275 0.052</td>
</tr>
<tr>
<td>Leucine</td>
<td>0.149 0.039</td>
<td>0.111 0.035</td>
</tr>
<tr>
<td>Histidine</td>
<td>0.037 0.002</td>
<td>0.120 0.024</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>0.045 0.009</td>
<td>0.065 0.026</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AD [mmol/L]</th>
<th>FC</th>
<th>p-value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Con</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pyruvic acid</td>
<td>0.026</td>
<td>0.004</td>
<td>-0.2</td>
</tr>
<tr>
<td>Valine</td>
<td>0.092</td>
<td>0.011</td>
<td>-0.2</td>
</tr>
<tr>
<td>Leucine</td>
<td>0.119</td>
<td>0.016</td>
<td>-0.2</td>
</tr>
<tr>
<td>Histidine</td>
<td>0.032</td>
<td>0.002</td>
<td>-0.1</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>0.037</td>
<td>0.006</td>
<td>-0.2</td>
</tr>
</tbody>
</table>

| Valine      | 0.079| 0.027 | -0.3  | 0.00002| 0.0001 |
| Leucine     | 0.078| 0.017 | -0.3  | 0.0001 | 0.0005 |
| Histidine   | 0.096| 0.023 | -0.2  | 0.0006| 0.002 |
| Isoleucine  | 0.049| 0.016 | -0.2  | 0.02 | 0.03 |

Five metabolites were dysregulated in serum samples between cognitively affected and healthy individuals, sorted according to \(p\)-value in from the validation cohort. Abbreviations; AD – Alzheimer’s Disease, Con – Healthy controls, FC – Fold change, SD – Standard deviation.

3.3. Metabolic alterations in the validation cohort

To extrapolate novel metabolic information, serum samples from the validation study were examined for significantly altered metabolites between the groups. This brought the total number of significantly different metabolites between the groups to sixteen, with fourteen of these being significant after FDR correction. Eleven metabolites were previously identified in the discovery cohort and the remaining five were novel metabolites not previously identified; 3-hydroxybutyric acid, citric acid, lactic acid, lysine, and succinic acid (Table 4).
Significantly altered metabolites measured in serum samples comparing cognitively affected with healthy individuals, sorted according to the \( p \)-value. Abbreviations; AD – Alzheimer’s Disease, Con – Healthy controls, FC – Fold change, FDR – False-discovery rate, SD – Standard deviation.

A network analysis was performed to investigate biological pathways for the significantly altered metabolites related to cognitive impairment (Fig. 3). Acetoacetate, a substrate for the TCA cycle (\( \log_2 FC = 1.4, p\)-value = 0.008), was the most elevated metabolite in AD. In contrast, pyruvate involved in glycolysis and gluconeogenesis (\( \log_2 FC = -0.6, p\)-value = 0.00002) was the most reduced metabolite in relation to AD. In addition, the metabolic pathways of Biopterin, Glycerophospholipid, Histidine, Lysine, and branch-chained amino acids (BCAAs); valine, leucine, and isoleucine, were also modified. This validation study identified and confirmed changes in histidine and BCAA metabolisms previously found in the discovery study; histidine (\( \log_2 FC = -0.3, p\)-value = 0.00006), isoleucine (\( \log_2 FC = -0.4, p\)-value = 0.02), leucine (\( \log_2 FC = -0.5, p\)-value = 0.00008), and valine (\( \log_2 FC = -0.4, p\)-value = 0.00002).
4. Discussion

In this study, we examined serum-derived metabolites associated with cognitive impairment in patients with mild to moderate AD compared to healthy individuals. The primary objective was to validate the significance of a panel of metabolites previously identified in the discovery study (19). The secondary objective was to add novel information not previously identified.

As previously mentioned, reproducibility is one of the more significant obstacles in biomarker studies (1). The authors of the referred study raised several crucial points to improve the reproducibility of future biomarker studies for neurodegenerative disease. These aspects range from cohort-related factors to independent validation. In the presented study, we have sought to comply with these recommendations, including; 1) consecutively recruitment of study participants to enroll more heterogenous groups, such that multiple factors attenuating the effect of the biomarker to avoid overestimation, 2) avoiding confounding factors, such as the presence of co-morbidities affecting measured biomarker levels, 3) performing validation in a separate validation cohort, as internal CV, i.e. when splitting a single cohort, has shown to fail when replicating the models in independent cohorts (18, 25), possibly due to systematic bias between the groups, 4) reporting not only the overall measure of performance, such as the AUC but also including the parameters PPV, NPV, etc., and lastly 5) metabolomics was performed in a different lab in the validation study compared to the discovery study, thereby also accounting for between-lab variability. (1)

We examined the serum metabolome of our study participants through NMR spectroscopy in combination with multivariate data analysis. Three models were tested using the discovery and validation data sets. Overall, these models primarily selected the same metabolites, indicating the relevance of these metabolites for distinguishing cognitively impaired persons from healthy individuals. The commonly selected metabolites included pyruvic acid, valine, and histidine. Pyruvic acid, or pyruvate, is the end-product of glycolysis and the substrate for mitochondrial adenosine triphosphate (ATP) synthesis. The nervous system is vulnerable to alterations in pyruvate metabolism due to the high ATP demand (26), which is used to maintain neuronal activity and homeostasis of the extracellular space and to defend against oxidative stress (27, 28). In contrast, to our findings, increased levels of pyruvate have been observed in CSF of AD patients (29), but similar alterations were identified in blood samples from patients with Parkinson's disease (30). As stated in the discovery study, valine and histidine are well-studied amino acids with respect to AD pathology (19). Consequently, the results of the present study further validate their importance related to AD. In accordance with previous findings identifying decreased levels of valine in serum (31) and CSF and positive correlations between CSF-valine and MMSE score, researchers have continued investigating this particular amino acid in relation to Alzheimer's disease (32). Valine was identified as a potential marker for predicting the transition from mild cognitive impairment to AD (31). Histidine is an essential amino acid (33), and a known scavenger of hydroxyl radicals, part of the reactive oxygen species (34). Using a cell model for anti-aging effects, increased proliferation and neurogenesis, as well as up-regulation of anti-oxidant enzymes, have been demonstrated as positive effects of histidine (33).
As an additional clinical characteristic parameter, we included measurements of established non-disease specific markers of neurological damage (GFAP and Nf-L), as well as hallmark targets of AD (Aβ\textsubscript{40}, Aβ\textsubscript{42}, and p-tau181) measured in blood, as the literature strongly implicates their diagnostic performance and significance in relation to neurological disease. (35, 36). Overall, the present study confirms previous findings; however, we found significantly elevated levels of Aβ\textsubscript{40} and no difference in Aβ\textsubscript{42} concentrations. This may be due to underlying cardiovascular conditions, such as hypertension, ischemic heart disease, and cardio-protective medications, which have been shown to influence plasma concentrations of Aβ. (37). In the biochemical measurements, glucose and LDH levels were significantly different between the control group and the patients, but both groups had levels within the normal range. Interestingly, reduced glucose utilization has been shown in AD brains (38), and this could also explain the observed significantly lower pyruvate concentration. With limited glucose availability, brain mitochondria use ketone bodies as energy substrates, with the two main ketone bodies being 3-hydroxybutyric acid and acetoacetic acid (39), which also were found elevated in the AD patient group. In addition, the ketones have been proposed as neuroprotective factors, reducing amyloid neurotoxicity and improving memory ability (39).

5. Limitations

Even though our results confirmed important findings and contributed to the search for valid blood-based biomarkers to aid in diagnosing Alzheimer’s disease, it is essential to note the limitations of our study. First, although the patient group was clinically confirmed to have AD, not all patients underwent neuropsychological testing or had CSF proteins measured because their physician deemed it unnecessary for the patient’s diagnosis. Secondly, to thoroughly verify our metabolite panel as a diagnostic tool for AD, it would be necessary to test its accuracy against other neurodegenerative diseases and different stages of AD. Thirdly, we discovered a significant age difference between our study groups, with AD patients being, on average, older than the control group. Unfortunately, it is not possible to recruit older blood donors, and we cannot exclude that this age difference could affect the results even though their age ranges overlapped. Fourthly, adding CSF samples to a study of the metabolome in relation to cognitive impairment can strengthen biomarker panels identified in the peripheral system. Lastly, both p-value and FDR corrected values were reported to minimize type II errors, however, multivariate statistics were also applied to encompass the overall information in the data, including covariance and correlation between metabolites.

Our findings validated the significance of the identified metabolite biomarker panel from the discovery study in distinguishing cognitively healthy individuals from patients with cognitive impairment. In addition, we evaluated various models to validate the performance of our panel, finding sPLS-DA to be the best fit. Lastly, new insights into disruptions in energy metabolism were uncovered.

6. Conclusions
In the current study, we validated our blood-based biomarker model derived from a discovery study consisting of five metabolites; pyruvic acid, valine, histidine, isoleucine, and glutamine. The novel information provided by the validation cohort confirmed the involvement of significantly altered metabolites in the BCAAs and histidine metabolisms. Moreover, metabolites from the energy metabolism were changed. Combining the proposed metabolite biomarker panel with neurodegenerative markers in plasma increased the diagnostic value. Although our validation yielded very intriguing results, the diagnostic performance of this panel of metabolic markers must also be assessed against other types of neurodegenerative diseases and various stages of AD progression.

**Abbreviations**

Aβ  
Amyloid-β  
ACE  
Addenbrooke’s cognitive examination  
AD  
Alzheimer’s Disease  
ATP  
Adenosine triphosphate  
AUC  
Area under the curve  
BBB  
Blood-brain barrier  
BCAA  
Branched-chain amino acid  
CNS  
Central nervous system  
CSF  
Cerebrospinal fluid  
CV  
Cross-validation  
EDTA  
Ethylenediaminetetraacetic acid  
FAQ  
Functional activities questionnaire  
FC  
Fold change  
FDR  
False-discovery rate  
GFAP
Glial fibrillary acidic protein
LDH
Lactate dehydrogenase
LDL
Low-density lipoprotein
MMSE
Mini-mental state examination
Nf-L
Neurofilament light
NMR
Nuclear Magnetic Resonance
NPV
Negative predictive value
p-tau
Phospho-tau
PPV
Positive predictive value
ROC
Receiver operating characteristics
SIMOA
Single Molecule Array
sPLS-DA
Sparse partial least squared discriminant analysis
t-tau
Total-tau
XGBoost
Extreme gradient boosting

Declarations

Ethics approval and consent to participate

The study was approved by the local North Denmark Region Committee on Health Research Ethics (N-20150010) and conducted according to the Declaration of Helsinki.

Consent for publication

Not applicable.

Availability of data and materials
All data generated or analysed during this study are included in this published article (and its supplementary information files).

Competing interests

The authors declare that they have no competing interests.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author’s contributions

JEN processed and prepared the samples, analysed and interpreted the data, performed the statistical analysis, and wrote the original draft for the manuscript. TA performed the NMR spectroscopy on the validation cohort. CHG performed the NMR spectroscopy on the discovery cohort. DAO performed the SIMOA analysis on the validation cohort. KV enrolled patients in the study and provided clinical data. JSM provided review and editing of the manuscript. SRK conceptualized and supervised the study, and was a major contributor in writing the manuscript. SP conceptualized and supervised the study, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to acknowledge the assistance provided by Mette Ullits Thomsen, Helle Dalsgaard Holst, Helle Hylander, and Mette Jespersgaard for their invaluable help in enrollment and blood sample collection from AD patients and blood donors. The NMR Center at DTU and the Villum Foundation are also acknowledged for providing access to the 800 MHz NMR spectrometer used for the discovery cohort. The NMR experiments on the validation cohort were performed at the MR Core Facility, Norwegian University of Science and Technology (NTNU). MR core facility is funded by the Faculty of Medicine and Health sciences at NTNU and Central Norway Regional Health Authority.

References


**Figures**
Figure 1

Validation of metabolic signature for cognitive impairment. (A) Scores plot for sparse-partial least squared discriminant analysis (sPLS-DA), with each score representing a sample. (B) Loadings plot for selected metabolites representing their mean importance for sample grouping reflected in the scores plot. The colour-coding of the bars indicates their importance for the corresponding group. (C) Receiver operating characteristics (ROC) curve indicates the ability of the model to distinguish the groups. Together with the ROC curve is the area under the curve (AUC) with the presented 95 % CI, accuracy, PPV, and NPV. (D) ROC curve of selected metabolites combined with significantly altered markers of neurodegeneration ($A\beta_{40}$, GFAP, Nf-L, $p$-tau181) showing an improved diagnostic efficacy, also presented with AUC and the 95 % CI, accuracy, PPV, and NPV. Abbreviations; AD – Alzheimer’s Disease, AUC – Area under the curve, CI – Confidence interval, Con – Healthy controls, LV – Latent variable, NPV – Negative predictive value, PPV – Positive predictive value.
Correlogram of metabolites of interest and clinical parameters. Only correlations shown to be significant are included. The colour of the square indicates if the correlation is positive (red) or negative (blue), and the intensity of the colour corresponds to the level of the correlation. Abbreviations; ACE – Addenbrooke’s cognitive examination, CSF – Cerebrospinal fluid, FAQ – Functional activities questionnaire, GFAP – Glial fibrillary acidic protein, MMSE – Mini-mental state examination, Nf-L – Neurofilament light, p-tau – Phospho-tau, t-tau – Total-tau.
Figure 3

Network analysis of dysregulated pathways related to cognitive impairment. Square nodes represent altered metabolites identified in the validation study, and circular nodes represent metabolites as part of the pathways not identified in the study. Colour of the square nodes represents the log2 FC value of the corresponding metabolite, according to alterations between the groups, with blue indicating a downregulation, and red indicating an upregulation. Metabolites are mapped according to their KEGG IDs.
Supplementary Files

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

- Additionalfile1RawNMRdata.xlsx
- Additionalfile2Clinicaldata.xlsx
- Additionalfile3InputforNetworkanalysis.xlsx