Alzheimer’s disease biomarker discovery using SOMAscan multiplexed protein technology

Martina Sattleckera,b, Steven J. Kiddlea,b, Stephen Newhousea,b, Petroula Proitsia,b, Sally Nelsonc, Stephen Williamsd, Caroline Johnstonea,b, Richard Killicka,b, Andrew Simmonsec,d,b, Eric Westmana,b, Angela Hodgesa,b, Hilkka Kloszewskad, Iwona Kloszewskae, Patrizia Mecoccif, Magda Tsolakig, Bruno Vellasg, Simon Lovestonest, on behalf of the AddNeuroMed Consortium and Richard J. B. Dobsona,b,*

aKing’s College London, Institute of Psychiatry, London, UK
bNIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at South London and Maudsley NHS Foundation Trust, London, UK
cSomaLogic, Boulder, CO, USA
dDepartment of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland
eMedical University of Lodz, Lodz, Poland
fInstitute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy
g3rd Department of Neurology, Aristotle University, Thessaloniki, Greece
hINSERM U 558, University of Toulouse, Toulouse, France

Abstract

Blood proteins and their complexes have become the focus of a great deal of interest in the context of their potential as biomarkers of Alzheimer’s disease (AD). We used a SOMAscan assay for quantifying 1001 proteins in blood samples from 331 AD, 211 controls, and 149 mild cognitive impaired (MCI) subjects. The strongest associations of protein levels with AD outcomes were prostate-specific antigen complexed to a1-antichymotrypsin (AD diagnosis), pancreatic prohormone (AD diagnosis, left entorhinal cortex atrophy, and left hippocampus atrophy), clusterin (rate of cognitive decline), and fetuin B (left entorhinal atrophy). Multivariate analysis found that a subset of 13 proteins predicted AD with an accuracy of area under the curve of 0.70. Our replication of previous findings provides further evidence that levels of these proteins in plasma are truly associated with AD. The newly identified proteins could be potential biomarkers and are worthy of further investigation.

Keywords: Alzheimer’s disease; Blood biomarkers; SOMAscan; Proteins; sMRI; Rate of progression

1. Introduction

Although much progress has been made in understanding the molecular pathology of Alzheimer’s disease (AD), the treatments currently available only temporarily alleviate some symptoms and do not modify pathology. The use of biomarkers to identify individuals with AD before the appearance of clinical symptoms, the so-called predementia phase of the disease, will be essential to the development of drugs for early intervention [1]. Additionally, if sufficiently powered and inexpensive, a biomarker could potentially be used as part of a screening program for at-risk elderly people [2].

Cerebrospinal fluid (CSF) biomarkers such as increased tau and decreased levels of amyloid-β (Aβ1-42) have been found to correlate with postmortem AD pathology [3,4]. These measurements together not only differentiate AD from normal elderly controls with high accuracy but can also predict which subjects with mild cognitive impairment (MCI) are likely to progress to AD within 5 years [5]. However, lumbar puncture to collect CSF is a relatively invasive
procedure, which may not be suitable for use in large-scale trials or for screening populations. Similarly, positron emission tomography imaging of amyloid burden in the brain correlates with clinical diagnosis of AD, Aβ neuropathology at autopsy [3,4,6], and CSF Aβ_{42} levels [7–10] but is equally impractical for use in large groups of frail elderly patients and is restricted to specialist centers.

In contrast, blood is a highly accessible tissue, making it an attractive target for development of a biomarker. Although AD is a disease of the brain, it is increasingly accepted that there is communication between the brain and the periphery, and thus several studies, including those from our own group, have investigated AD and/or MCI-related protein changes in blood plasma using mass spectrometry and antibody capture technologies [11–20].

In this study, we screened plasma from 691 individuals using the SOMAscan (SomaLogic, Inc, Boulder, CO, USA) Multiplexed Proteomic technology for proteins involved in AD pathology. Each of the 1001 measured proteins was tested for association with the following AD outcome measures: clinical diagnosis (case and control), conversion from MCI to AD, rate of cognitive decline, and brain atrophy (left/right hippocampal volume and left/right entorhinal cortex volume). We report results from the largest multiplexed plasma protein study to date for AD markers and as such describe a number of novel protein associations as well as validate a number of previously identified AD biomarkers.

2. Methods

2.1. Subjects

We obtained protein measures for 691 subjects (211 controls, 106 MCI patients, 43 MCI patients converting to AD within a year, and 331 AD patients). Of these subjects, 415 (110 controls, 66 MCI patients, 43 MCI patients converting to AD, and 196 AD patients) were recruited from the European Union–funded AddNeuroMed (ANM) biomarker study [14,21] and 276 subjects (101 controls, 40 MCI, and 135 AD) were recruited from the Maudsley and King’s Healthcare Partners Dementia Case Register (DCR), which incorporates the Alzheimer’s Research UK (ARUK) cohort [13]. Informed consent was obtained for all subjects according to the Declaration of Helsinki (1991), and protocols and procedures were approved by the relevant local ethical committees at each site. All subjects were assessed with a standardized assessment protocol including an informant interview for diagnosis and cognitive assessment such as the Mini-Mental State Examination (MMSE) together with standardized assessment of function, behavior, and dementia severity as previously reported [13,19,20].

2.2. Samples

At the time of assessment, all blood samples were drawn by venipuncture and collected into EDTA glass tubes. Subjects were required to fast for at least 2 hours before collection. ARUK blood samples were centrifuged at 3000 rpm at 4°C for 8 minutes, and ANM and DCR samples were centrifuged with at 2000 g for 10 minutes at 4°C. All samples were centrifuged within approximately 2 hours of collection. Plasma supernatant was collected, divided into aliquots, and frozen at −80°C until further use.

2.3. Protein measures

Proteins were measured using a Slow Off-rate Modified Aptamer (SOMAmer)–based capture array called “SOMAscan” (SomaLogic, Inc). This approach uses chemically modified nucleotides to transform a protein signal to a nucleotide signal that can be quantified using relative florescence on microarrays. Therefore, all gathered SOMAscan measures are relative fluorescence units. This assay has been shown to have a median intra- and interrun coefficient of variation of ~5%. The median lower and upper limits of quantification were ~1 pM and ~1.5 nM, respectively, in buffer and ~2.95 pM and ~1.5 nM, respectively, for a subset of the SOMAmers in plasma; full details are given in the article by Gold et al. [22].

Quality control is performed at the sample and SOMAmer level and involves the use of control SOMAmers on the microarray and calibration samples. At the sample level, hybridization controls on the microarray are used to monitor sample-by-sample variability in hybridization, whereas the median signal over all SOMAmers is used to monitor overall technical variability. The resulting hybridization scale factor and median scale factor are used to normalize data across samples. The acceptance criteria for these values are 0.4 to 2.5, based on historic trends in these values. SOMAmer-by-SOMAmer calibration occurs through the repeated measurement of calibration samples; these samples are of the same matrix as the study samples and are used to monitor repeatability and batch-to-batch variability. Historic values for these calibrator samples for each SOMAmer are used to generate a calibration scale factor. The acceptance criterion for calibrator scale factors is that 95% of SOMAmers must have a calibration scale factor within ±0.4 of the median.

The assay required 8 μL of plasma from each sample. A single assay was used per plasma sample, and thus, no technical replicates were performed. Additionally, the samples were run in two batches ensuring an even mix of diagnosis groups in each batch. All measurements were log2 transformed. Seven outliers, identified using principal component analysis, were removed from the downstream analysis. Principal component analysis also showed that protein measures were affected by the study center, and thus, we either adjusted for the center using linear regression or added the center as a covariate in all downstream analysis. The identified center effect is likely to be caused by differences in sample handling.

The assay measures the level of 1001 human proteins representing different molecular pathways and gene families. Most proteins are involved in the following processes: signal transduction pathways, stress response, immune process,
and phosphorylation, but in addition, proteolysis, cell adhesion, cell differentiation, and intracellular transport proteins are represented.

2.4. Structural magnetic resonance imaging

Volumes of the hippocampi and entorhinal cortices, normalized by intracranial volume [23], were obtained from 273 ANM subjects (93 controls, 62 MCI patients, 19 MCI patients converting to AD within a year, and 99 AD patients) who had undergone structural magnetic resonance imaging (sMRI). Imaging measures were not adjusted for diagnosis, which allowed the identification of proteins associated with brain atrophy caused by AD. The volumetric data were not used to aid in the clinical diagnosis of AD. Detailed information regarding data acquisition, preprocessing, and quality control assessment has been described for this cohort elsewhere [24,25].

2.5. Calculation of the rate of cognitive decline

The rate of cognitive decline (disease progression) in 329 AD patients (214 from ANM, 87 from ARUK, and 28 from DCR) was calculated based on longitudinal MMSE assessments [26]. For the ANM cohort, MMSE scores were gathered at five visits, in which visits were 3 months apart (1-year follow-up). For the ARUK and DCR cohorts, MMSE scores were obtained annually over a period of 2 years (three visits). To estimate the rate of decline, only samples with at least three MMSE measures were included. Linear mixed-effect models with a random intercept and random slope were generated using the package “nlme” in the open source statistical software package R (www.r-project.org). This was done separately for ANM, and for DCR and ARUK combined, because of the differences in assessment windows between the cohorts. Samples and the center were included as random effects in the model. Further covariates including age of onset, disease duration at baseline, gender, apolipoprotein (APOE) ε4 allele presence, living in a nursing home, and years of education were investigated for their effect on the rate of decline.

We found that age of onset, living in a nursing home, and education had significant effects on the rate (P value <.05) and thus were included as fixed effects in the final model. The slope coefficient obtained from the final model for each sample was then used as the rate of cognitive decline, defined as the change in MMSE per day. The cognitive decline slopes were first derived and then tested for associations between protein levels because this allows us to capture a greater subject variance because we had MMSE scores for 329 AD patients, but plasma protein measures only for 239 (173 from ANM, 44 from ARUK, and 22 from DCR).

2.6. Data analysis

2.6.1. Single-analyte analysis

All proteins were analyzed individually for their association with the following AD outcome measures: disease status (AD vs. CTL and MCI stable vs. MCI converter), sMRI imaging measures (volume of left/right entorhinal cortex and left/right hippocampus), and rate of cognitive decline (MMSE change). Logistic regression in R was used for each protein to find associations with the disease status. The association between protein levels and sMRI imaging measures and the association between protein measures and the rate of cognitive decline were investigated using linear regression in R. Subjects’ age at sampling, gender, presence of APOE ε4 alleles, and recruitment center were used as covariates for the imaging models. Imaging models were built using all diagnosis groups, as well as only using control groups. For the rate of cognitive decline models, covariates were not included because they have been adjusted for in the rate of decline calculation. Throughout the single-analyte analysis, we applied false discovery rate (FDR) to correct for multiple testing.

2.6.2. Multivariate analysis

A random forest approach (R package “randomForest”) was used to develop an AD versus control classifier. Default settings were applied (ntree = 500, mtry = square root of the number of variables for classification models, and mtry = number of variables divided by three for regression models). After center adjustment using linear regression, the residuals were split into a test and training set such that the training set comprised cases and controls matched on age, APOE ε4 presence, and gender. The training set consisted of 99 AD samples and 99 control samples, and the independent test set comprised 232 AD and 112 controls. It was possible to match only 99 controls and AD samples, and thus, our training set is smaller. The training set was used to rank the proteins according to their predictive power. This was done by bootstrapping the training data 100 times, such that it was randomly split into a bootstrap training set (75%) and a bootstrap test set (25%). The bootstrap training set was used to build a random forest model and then tested with the bootstrap test set. In a random forest model, each protein typically gets assigned with an importance score, which can be used to rank proteins. Because the bootstrapping procedure was executed 100 times, each of the proteins was assigned a total of 100 ranks. All individually obtained ranks were summed for each protein resulting in a list of proteins sorted according to predictive power. The next step was to select the optimum number of proteins with the aim of keeping numbers as low as possible. This involved conducting a second round of bootstrapping, this time including backward elimination. First, the top 100 proteins were selected, based on the variable importance achieved in the previous optimization step to build 100 bootstrap random forest models, which were then tested with each associated bootstrap test set. This resulted in 100 bootstrap results for these proteins. This was then repeated with the top 90 to 50 proteins in steps of 10, then to 30 proteins in steps of 5, and finally down to 2 proteins in steps of 1. For each set of proteins, the mean bootstrap testing performance was
calculated, and based on the best performance, the optimal number of proteins was identified. A final model was then built in the complete training data using the optimized number of proteins and subsequently tested with the independent test set.

The same model development strategy was applied to build a dichotomous model to predict which MCI patients converted to AD within 1 year of sampling. All data were center adjusted using linear regression. The obtained residuals were then used for the model development including age, APOE ε4 presence, and gender as covariates. Because the data set was highly imbalanced (106 stable MCI patients and 43 MCI patients converting to AD), we randomly selected 75% of the MCI converter samples and then selected the equal number of stable MCI samples (undersampling) to balance the training set. All remaining samples were used in the test set. Additionally, random forest regression models were built to predict the rate of cognitive decline, left/right entorhinal cortex volume, and left/right hippocampal volume. For the brain atrophy regression models, center-adjusted data were used, and age, gender, and APOE ε4 presence were included in the random forest model development. These covariates were omitted in the rate of cognitive decline model because they were already included when calculating the rate of decline. The data set was split randomly into a training set (75%) and a test set (25%) for each of the five regression models.

3. Results

Demographic characteristics stratified by cohort are provided in Table 1. Comparisons were undertaken between cohorts with respect to clinical diagnosis using a two-way analysis of variance. Significant distributional differences were found for gender and MMSE score at baseline; no significant differences in the distributions of age and number of APOE ε4 alleles across the three cohorts were observed.

3.1. Disease status

3.1.1. AD versus healthy control subjects

The single-analyte logistic regression for AD cases versus controls showed that 138 proteins were associated with AD at a significance level of P value <.05 and four proteins at a q value <0.05: prostate-specific antigen complexed to the serine protease inhibitor α1-antichymotrypsin (odds ratio [OR], 6.85; β = 1.92; q value = 0.0005), pancreatic prohormone (OR, 2.41; β = 0.88; q value = 0.0009), calcium/calmodulin-dependent protein kinase (OR, 0.01; β = −4.38; q value = 0.0288), and trypsin (OR, 2.24; β = 0.81; q value = 0.0475). Results for all 1001 proteins are provided in Supplementary Table 1.
In the random forest analysis, we found the highest mean training performance was achieved by a model consisting of 80 proteins. However, for a prospective clinical application, it is desirable to keep the number of proteins as low as possible while maintaining a high performance. Thus, we selected a model with 13 proteins because it is the lowest number of proteins before the training performance (sensitivity = 74%; specificity = 72%) dropped drastically (Supplementary Fig. 1). The 13 top proteins (Table 2) and the complete training cohort were used to build the final predictive model, which was then tested with the independent test set. The model predicted 66% of all test samples correctly with a sensitivity of 67%, a specificity of 64%, a positive predictive value of 80%, and a negative predictive value of 48%. A receiver operating curve analysis resulted in an area under the curve (AUC) of 0.70 (Fig. 1).

3.1.2. Conversion of MCI to AD

Logistic regression comparing stable MCI subjects with MCI converters showed that 20 proteins were significantly associated (P value <.05) with conversion to AD. However, none of them passed multiple testing corrections (full results in Supplementary Table 1). We also developed a random forest model to predict the conversion from MCI to AD. Although the training performance in the training set looked promising with a specificity of 85% and a sensitivity of 86%, the model only achieved a specificity of 61%, a sensitivity of 55%, a positive predictive value of 17%, and a negative predictive value of 90% in the independent test set (74 stable MCI samples and 11 converters).

3.2. Brain atrophy

Linear regression was used to compare plasma proteins levels with hippocampi and entorhinal cortices, that is, brain regions known to relate to early AD pathology. Results for all proteins are summarized in Supplementary Table 1.

Table 2

<table>
<thead>
<tr>
<th>Importance rank</th>
<th>Protein</th>
<th>Odds ratio</th>
<th>β</th>
<th>P value</th>
<th>q value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Placenta growth factor</td>
<td>1.86</td>
<td>0.62</td>
<td>.2439</td>
<td>0.6285</td>
</tr>
<tr>
<td>2</td>
<td>Interleukin-17F</td>
<td>1.81</td>
<td>0.59</td>
<td>.1925</td>
<td>0.5802</td>
</tr>
<tr>
<td>3</td>
<td>Fibronectin FN1.4</td>
<td>0.75</td>
<td>−0.29</td>
<td>.0363</td>
<td>0.3519</td>
</tr>
<tr>
<td>4</td>
<td>Fibronectin</td>
<td>0.75</td>
<td>−0.29</td>
<td>.0244</td>
<td>0.3484</td>
</tr>
<tr>
<td>5</td>
<td>Secretory leukocyte protease inhibitor</td>
<td>2.33</td>
<td>0.84</td>
<td>.0450</td>
<td>0.3544</td>
</tr>
<tr>
<td>6</td>
<td>Fibronectin FN1.3</td>
<td>0.74</td>
<td>−0.30</td>
<td>.0353</td>
<td>0.3519</td>
</tr>
<tr>
<td>7</td>
<td>Epithelial cell kinase</td>
<td>1.59</td>
<td>0.46</td>
<td>.1252</td>
<td>0.5001</td>
</tr>
<tr>
<td>8</td>
<td>Prolactin</td>
<td>1.59</td>
<td>0.46</td>
<td>.0799</td>
<td>0.4057</td>
</tr>
<tr>
<td>9</td>
<td>C-C motif chemokine 14</td>
<td>2.58</td>
<td>0.95</td>
<td>.0020</td>
<td>0.1473</td>
</tr>
<tr>
<td>10</td>
<td>Calcium/calmodulin-dependent protein kinase</td>
<td>0.01</td>
<td>−4.38</td>
<td>8.49 × 10⁻⁵</td>
<td>0.0288</td>
</tr>
<tr>
<td></td>
<td>type II subunit α</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Seprase</td>
<td>0.27</td>
<td>−1.31</td>
<td>.0008</td>
<td>0.1080</td>
</tr>
<tr>
<td>12</td>
<td>Pancreatic prohormone</td>
<td>2.41</td>
<td>0.88</td>
<td>1.68 × 10⁻⁶</td>
<td>0.0009</td>
</tr>
<tr>
<td>13</td>
<td>Coagulation factor XI</td>
<td>2.40</td>
<td>0.87</td>
<td>.08540</td>
<td>0.4109</td>
</tr>
</tbody>
</table>

NOTE. This table also shows the odds ratio, β coefficient, P value, and q value achieved in the logistic regression AD versus controls.

Fig. 1. Receiver operating curve for the diagnostic model predicting AD in which samples used for training were matched by age, gender, and APOE ε4 presence. The area under the curve is 0.70.

3.2.1. Entorhinal cortex

Sixty-eight proteins correlated with left entorhinal volume (P value <.05), four of which passed multiple testing corrections at a q value threshold of 0.05 (Table 3). The analysis of brain atrophy in control samples only showed that pancreatic prohormone and serine/threonine-protein kinase Chk2 were associated with atrophy in the left entorhinal cortex in controls only at P value <.05. However, they did not pass multiple testing corrections (q value <0.05). Thirty-two proteins correlated with both the left and right entorhinal volumes. Most of these proteins showed a stronger correlation with the left side. Sixty-seven proteins correlated with right entorhinal volume at the uncorrected P value threshold, one of which passed multiple testing corrections at a q value threshold of 0.05 (Table 3). The random forest regression
Chk2 0.18 0.0002** 0.12 9.6

left and right hippocampi in controls only at analysis of brain atrophy in control samples only showed that passed multiple testing corrections (q value right hippocampal volume (right hippocampal volume (training R² = 0.26, R² = 0.20; 212 samples).

3.2.2. Hippocampal volume

All results for the hippocampal volume comparisons are summarized in Supplementary Table 1. Sixty-two proteins were associated with the left hippocampal volume, and one of them, pancreatic prohormone, passed multiple testing corrections (q value <0.05). The analysis of brain atrophy in control samples only showed that pancreatic prohormone was associated with atrophy in the left and right hippocampi in controls only at P value <0.05. However, it did not pass multiple testing corrections (q value <0.05). Among the identified proteins, 37 were correlated with both the left and right hippocampal volumes. Most of these proteins showed a stronger correlation with the left side. The random forest regression model developed for the right and left hippocampal volumes predicted the independent test sets (68 samples) with an R² = 0.06 and R² = 0.14, respectively (training R² = 0.30, R² = 0.31; 212 samples).

3.3. Rate of cognitive decline

The estimated mean loss in MMSE points per year for AD samples used in the protein analysis was ANM 1.5 (standard deviation [STD] = 1.5), ARUK 2.9 (STD = 1.4), and DCR 2.2 (STD = 1.4). One hundred thirty-nine proteins were found to correlate with the rate of cognitive decline (P value <0.05), two of which passed multiple corrections (q value <0.05), namely clusterin (R² = 0.08; β = 2.50; q value = 0.012) and nucleosome assembly protein 2 (R² = 0.06; β = 0.74; q value = 0.044). Both proteins were positively associated with the rate of decline, and therefore, the quantity of these proteins is higher in the plasma of patients with fast cognitive decline. Results for all proteins are summarized in Supplementary Table 1. A random forest regression model optimized for predicting the rate of cognitive decline achieved a training performance of R² = 0.20 (180 samples) and test performance of R² = 0.10 in the independent test set (59 samples).

3.4. Summary single-analyte analysis results

A total of 355 proteins were found to be associated with at least one of the outcome measures at an uncorrected threshold of P value <0.05 (eight passing an FDR q value threshold of <0.05). Scatter plots and box plots for the proteins showing the strongest association with one of the investigated outcome measures are shown in Fig. 2.

We found a high correlation between the four magnetic resonance imaging (MRI) measures, especially between the left and right hippocampal volumes and the left and right entorhinal cortices. In contrast, the rate of cognitive decline in AD subjects is not significantly associated with hippocampal (HC) or entorhinal cortex (EC) atrophy (Supplementary Table 2). Thus, the rate of cognitive decline is the only independent outcome measure. Twenty-four proteins were significantly associated (P value <0.05) with at least three of the following five outcome groups: AD and control, stable MCI and MCI converting to AD, entorhinal cortex (left and/or right), hippocampus (left and/or right), and the rate of cognitive decline (Fig. 3).

4. Discussion

We analyzed 1001 proteins in 691 human plasma samples to identify proteins associated with AD clinical diagnosis, conversion from MCI to AD, rate of cognitive decline, and brain atrophy. The single-analyte analysis showed that sets of significant variables differed for the different outcome measures. Although all these outcome measures are linked to AD, they capture different aspects of the pathology. Clinical diagnosis may be noisy and imaging atrophy is continuous and may capture earlier changes. The rate of cognitive decline may be the most independent of the outcome measures. Additionally, although separate proteins are reported for each outcome measure, they may correlate or be part of the same pathways or processes—they may be surrogates for each other. Prostate-specific antigen complexed to the serine protease inhibitor α1-antichymotrypsin (PSA-ACT), pancreatic

Table 3

Proteins significantly correlated with the left entorhinal cortex passing FDR correction q value <0.05 (** are also significantly correlated with the right entorhinal cortex and left/right hippocampal volume at P value <0.05 (*)

<table>
<thead>
<tr>
<th>Protein</th>
<th>Entorhinal cortex</th>
<th>Hippocampus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Left</td>
<td>Right</td>
</tr>
<tr>
<td>R²</td>
<td>β</td>
<td>R²</td>
</tr>
<tr>
<td>Fetuin B</td>
<td>0.20</td>
<td>0.0002**</td>
</tr>
<tr>
<td>Pancreatic prohormone</td>
<td>0.19</td>
<td>-1.0 × 10^{-6**}</td>
</tr>
<tr>
<td>PSA-ACT</td>
<td>0.19</td>
<td>-0.0002**</td>
</tr>
<tr>
<td>Chk2</td>
<td>0.18</td>
<td>0.0002**</td>
</tr>
</tbody>
</table>

Abbreviations: FDR, false discovery rate; PSA-ACT, prostate-specific antigen complexed to the serine protease inhibitor α1-antichymotrypsin; Chk2, serine/threonine-protein kinase Chk2.
Prohormone, clusterin, and fetuin B showed the strongest associations and thus will be discussed in greater detail. Prostate-specific antigen complexed to the serine protease inhibitor α₁-antichymotrypsin (PSA-ACT complex) showed the strongest association with AD diagnosis and was also associated with hippocampal and entorhinal cortex volume atrophy. To our knowledge, this is the first study of plasma PSA-ACT complex in AD. Prostate-specific antigen is a serine protease that in serum is predominantly found complexed to the serine protease inhibitor α₁-antichymotrypsin. α₁-Antichymotrypsin colocalizes with amyloid plaques in AD brain [27] and both prostate-specific antigen [28,29] and α₁-antichymotrypsin [30,31] are detectable in CSF. As such, Mulder et al. [32] found that prostate-specific antigen was present in the brain, but levels of prostate-specific antigen and PSA-ACT complex did not differ across in CSF samples from controls, fronto-temporal lobe dementia, and AD and MCI patients.

Of the 1001 proteins measured, fetuin B showed the strongest correlation with left and right entorhinal cortex volumes. It was also significantly correlated with the left and right hippocampus atrophy. All brain atrophy correlations were positive, and thus, patients with progressing brain atrophy had a reduced level of fetuin B. In a previous study, plasma fetuin A was found to be lower in AD patients than in controls [33], and Denecke et al. [34] have suggested that fetuin B has similar but not identical function to fetuin A. Fetuin A was not present on the SomaLogic panel and so was not measured directly.

Clusterin was the strongest associated protein with the rate of cognitive decline, and thus, we found fast declining AD patients have higher clusterin levels in plasma than slow declining AD patients. This finding replicates pervious findings, which showed the association of clusterin with more severe disease and rapid clinical progression [20,35]. However, 66 of our AD samples (63 ANM and 3 ART; 28% of samples) were previously used by Thambisetty et al. [20]. When removing these 66 samples from the data set and using the remaining 173 samples (72%), clusterin remained significantly associated with the rate of cognitive decline (q value = 0.007), showing independent validation of the association of clusterin with the rate of progression.

Pancreatic prohormone is a promising candidate, which we found to be significantly associated with three outcome measures, namely AD clinical diagnosis, left hippocampal atrophy, and left entorhinal cortex volume. It was also found to be associated with changes in brain volume in the left/ right hippocampus and the left entorhinal cortex in controls.
Fig. 3. Univariate heatmap showing proteins, which were significantly associated with at least three of the seven studied Alzheimer’s disease (AD) outcome measures: case versus control (AD), mild cognitive impairment (MCI) to AD conversion (MCIc), rate of cognitive decline (ROD), left/right entorhinal cortex atrophy (EC.left/EC.right), and left/right hippocampal atrophy (HC.left/HC.right). Proteins signed with * were found to be significant at a P value < 0.05 and proteins signed with ** were found to be significant at a P value < 0.01. Scaled β values were used to generate the heatmap and thus red indicated positive association and blue negative association. LRIG3, leucine-rich repeats and immunoglobulin-like domains protein 3; PSA-ACT, prostate-specific antigen complexed to the serine protease inhibitor a1-antichymotrypsin; IGFBP-2, insulin-like growth factor–binding protein 2; BPI, bactericidal permeability-increasing protein; PGRP-S, peptidoglycan recognition protein 1; NKG2D, NKG2-D type II integral membrane protein; HIBADH, 3-hydroxyisobutyrate dehydrogenase, mitochondrial.
Only. This finding might indicate that pancreatic prohormone is associated with preclinical changes. However, this would require follow-up of the control samples in order to be verified. Furthermore, it was among the 13 proteins, which were present in the final AD diagnostic model. Previous studies have shown that pancreatic prohormone levels are associated with AD diagnosis [15,36–38] and amyloid burden [39]. Also, the use of pancreatic prohormone in a protein signature for AD prediction was reported [36]. Our study is the first non-Luminex xMAP (Luminex Corporation, Austin, TX, USA) study to identify pancreatic prohormone as a potential marker for AD, and we also show that it associates with other outcome measures such as brain atrophy.

Other proteins generally linked to AD pathology in the literature that we also found to be altered in the periphery include glypican 3, higher levels of which were found in AD patients, individuals with high left and right hippocampal atrophy, and MCI patients converting to AD within 1 year; glypican 1 was previously found to be present in both diffuse and classic cerebral senile plaques and tangles [40]; neurexophilin-1, a flanking single-nucleotide polymorphism (rs6463843) of which was reported as a top hit in an Alzheimer’s Disease Neuroimaging Initiative (ADNI) genome-wide association study of brain-wide MRI phenotypic measures of gray matter density, volume, and cortical thickness [41], had increased levels in AD patients whose cognitive decline was rapid; visfatin, which in accumulations may lead to damage of the blood-brain barrier, amyloid formation, and brain damage [42], was found to be significantly increased in patients with fast cognitive decline. We found that levels of metalloproteinase inhibitor 3, which has previously been shown to be increased in AD brains [43], was significantly increased in the plasma of AD patients with a fast rate of cognitive decline.

We were able to replicate a number of specific findings from previous AD biomarker studies of the periphery. z1-Antitrypsin has been found to be associated with AD [36,44–46] and brain amyloid burden [39]. Other proteins that we were able to replicate include apolipoprotein E [19,37–39,47], granulocyte colony–stimulating factor [15,18], matrix metallopeptidase 9 [36,39], fibronectin [48], complement component C6 [49], and immunoglobulin E [39].

Additionally, we developed and optimized multivariate classifiers for predicting clinical diagnosis, conversion from MCI to AD, right/left hippocampal/entorhinal volume, and rate of cognitive decline. The only model that performed well was the AD diagnostic model, which consisted of 13 proteins using samples matched by age, gender, and APOE ε4 presence in the training set. The other models, although performing well in training did not perform well in the independent test suggesting the training models were overfitted, as a possible side effect of the small sample size. Ray et al. [18] reported a model comprising 18 plasma proteins that achieved an overall accuracy of 89% for AD diagnosis in a data set consisting of 85 AD patients and 79 controls. They did not report adjustment for covariates, such as age, gender, and APOE ε4 genotype. Björkqvist et al. [50] quantified the same 18 proteins in plasma from 174 controls, 142 patients with AD, and 88 patients with other dementias and found that these 18 proteins could classify patients with AD from controls only with low diagnostic precision (AUC = 0.63). Recently, Doecke et al. [36] also identified a separate 18–protein biomarker signature in blood plasma using samples from 207 AD patients and 754 healthy controls recruited for the Australian Imaging Biomarker and Lifestyle study. They validated the signature in the ADNI cohort (80% accuracy; 108 AD patients and 57 healthy controls) but included age, gender, and APOE ε4 genotype as predictors in the model. Interestingly, age, gender, and APOE ε4 alone could predict AD with an accuracy of 77% [36]. We chose to develop our diagnostic classifier by matching cases and controls in the training set. The resulting model achieved a sensitivity of 67%, a specificity of 64%, and an AUC of 0.70 when tested in an independent test set. The performance is less than that reported in some previous studies, but it is also less likely to be confounded by age, gender, and APOE ε4 presence. The performance of this classifier is possibly too low for screening of a wider population; however, it might contribute to gain greater understanding of AD pathophysiology, and it is worth noting that two proteins in our diagnostic classifier have been reported in previous classification studies, namely pancreatic prohormone [36] and prolactin [15]. The remaining 11 proteins were unique to our classifier; however, we do not know if this is because they have not been studied before or they were not found to be discriminatory in other studies. Nonetheless, 8 of the 13 proteins used in the classifier were also found to be significantly correlated with AD status in our univariate analysis: fibronectin, fibronectin FN1.3, fibronectin FN1.4, secretory leukocyte protease inhibitor, C-C motif chemokine 14, calcium/calmodulin-dependent protein kinase type II subunit α, pancreatic prohormone, and seprase. The remaining five proteins may have been selected during the multivariate optimization procedure because of interaction effects between the proteins, which were not tested for in the univariate analysis. An alternative approach to the used one would be to select proteins with previously reported AD association, which might lead to improved performance.

The generally lower test performance in comparison with the training performance might be due to slight overfitting to the training set. Sample size could also be affecting performance; for the MCI conversion classifier, we had only 43 MCI samples converting to AD. Other possibilities that could have negatively influenced the classifier performance are sample heterogeneity and short follow-up for MCI samples. Cohort differences might have also played a role, although we adjusted the data for center effects. The performance model predicting cognitive decline might be improved by dichotomizing the individuals, into fast and slow decliners.
In our study, a larger number of proteins were found to be associated with the left hippocampus/entorhinal cortex than their right equivalent, which may relate to changes in the natural right-to-left asymmetry during the progression of AD. Changes to the hippocampal right-to-left asymmetry were alluded by previous studies [51–53].

To the best of our knowledge, this study reports the largest number of tested proteins in plasma for their suitability as AD biomarkers using array technology. Rather than only comparing protein quantities with the clinical disease status, we further compared protein levels with AD endophenotypes of brain atrophy (left/right entorhinal and hippocampal volumes) and also the rate of cognitive decline (rate of MMSE change). We found 53 proteins to be associated with three or more AD outcome measures, some novel and some previously identified. Some of the novel proteins may not have been studied before in the context of AD and could be good candidates for further validation.

Acknowledgments

This work was supported by InnoMed (Innovative Medicines in Europe), an integrated project funded by the European Union of the Sixth Framework program priority (FP6-2004-LIFESCIHEALTH-5); the Alzheimer’s European Union of the Sixth Framework program priority (FP6-2004-LIFESCIHEALTH-5); the Alzheimer’s Alzheimer’s disease: a clinical follow-up study of 4.7 years. J Alzheimers dis 2010;21:1119–28.

To the best of our knowledge, this study reports the largest number of tested proteins in plasma for their suitability as AD biomarkers using array technology. Rather than only comparing protein quantities with the clinical disease status, we further compared protein levels with AD endophenotypes of brain atrophy (left/right entorhinal and hippocampal volumes) and also the rate of cognitive decline (rate of MMSE change). We found 53 proteins to be associated with three or more AD outcome measures, some novel and some previously identified. Some of the novel proteins may not have been studied before in the context of AD and could be good candidates for further validation.

Acknowledgments

This work was supported by InnoMed (Innovative Medicines in Europe), an integrated project funded by the European Union of the Sixth Framework program priority (FP6-2004-LIFESCIHEALTH-5); the Alzheimer’s Research Trust UK; the John and Lucille van Geest Foundation; and the NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at the South London, Maudsley NHS Foundation Trust and Kings College London, and a joint infrastructure grant from Guy’s and St Thomas’ Charity and the Maudsley Charity; Kuopio University Hospital (HS) and funding from UEF-BRAIN (HS). SOMAscan and SOMAmer are trademarks of SomaLogic, Inc. The authors acknowledge Abhishek Dixit for his help with data management and access. Disclosure: Intellectual property has been registered on the use of plasma proteins for the use as biomarkers for AD by King’s College London and Proteome Sciences, with Simon Lovestone named as an inventor. Sally Nelson and Stephen Williams are employees of SomaLogic, Inc, and the proteomic assay reported in this manuscript was performed using reagents supplied by SomaLogic, Inc.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jalz.2013.09.016.

References


[35] Schrijvers EM, Koudstaal PJ, Hofman A, Breteler MM. Plasma clus-