Validation of Plasma Proteomic Biomarkers Relating to Brain Amyloid Burden in the EMIF-Alzheimer's Disease Multimodal Biomarker Discovery Cohort

Sarah Westwood, Alison Baird[1], Sneha Anand[1], Alejo Nevado-Holgado[1], Andrey Kormilitzin[1], Liu Shi[1], Abdul Hye[2], Nicholas Ashton[2], Angharad Morgan[3], Isabelle Bos[4], Stephanie Vos[4], Rik Vandenberghe[5], Noel J Buckley[1], Philip Scheltens[6], Sebastiaan Engelborghs[7,8], Giovanni Frisoni[9], José Luis Molinuevo[9,10], Anders Wallin[11], Alberto Lléo[12], Julius Popp[13,22], Pablo Martinez-Lage[14], Cristina Legido-Quigley[2], Lars Bertram[15], Frederik Barkhof[16], Henrik Zetterberg[17,18,19,20], Paul Morgan[3], Johannes Streffer[21], Pieter Jelle Visser[4,6], Simon Lovestone[1*]

[1]Department of Psychiatry, University of Oxford, UK

[2]Kings College London, London, UK

[3]Cardiff University, Cardiff, UK

[4]Alzheimer Centrum Limburg, Maastricht University, Maastricht, the Netherlands

[5]University Hospital Leuven, Leuven, Belgium

[6]Alzheimer Center, VU University Medical Center, Amsterdam, the Netherlands

[7]Department of Neurology and Memory Clinic, Hospital Network Antwerp (ZNA) Middelheim and Hoge Beuken, Antwerp, Belgium

[8]University of Antwerp, Antwerp, Belgium

[9]University of Geneva, Geneva, Switzerland

[10]IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

[11]Institute of Neuroscience and Physiology, Sahlgrenska Academy at University of Gothenburg, SE-431 41 Moelndal, Sweden

[12]Hospital de la Santa Creu i Sant Pau, Barcelona, Spain

[13]University Hospital of Lausanne, Lausanne, Switzerland

[14]CITA-Alzheimer Foundation, San Sebastian, Spain

[15]University of Lübeck, Lübeck, Germany

[16]Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, The

Netherlands

[17]Department of Psychiatry and Neurochemistry, University of Gothenburg, Mölndal, Sweden

[18]Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden

[19]UK Dementia Research Institute at UCL, London, United Kingdom

[20]Department of Neurodegenerative Disease, UCL Institute of Neurology, London, United Kingdom

[21]Experimental Medicine, Janssen Pharmaceutical Companies, Beerse, Belgium

[22]Geriatric Psychiatry, Department of Mental Health and Psychiatry, Geneva university Hospitals, Geneva, Switzerland

*** Correspondence:**

Professor Simon Lovestone

simon.lovestone@psych.ox.ac.uk

Abstract

We have previously investigated, discovered, and replicated plasma protein biomarkers for use to triage potential trials participants for PET or cerebrospinal fluid measures of Alzheimer's disease (AD) pathology. This study sought to undertake validation of these candidate plasma biomarkers in a large, multi-center sample collection. Targeted plasma analyses of 34 proteins with prior evidence for prediction of in vivo pathology were conducted in up to 1,000 samples from cognitively healthy elderly individuals, people with mild cognitive impairment, and in patients with AD-type dementia, selected from the EMIF-AD catalogue. Proteins were measured using Luminex xMAP, ELISA, and Meso Scale Discovery assays. Seven proteins replicated in their ability to predict in vivo amyloid pathology. These proteins form a biomarker panel that, along with age, could significantly discriminate between individuals with high and low amyloid pathology with an area under the curve of 0.74. The performance of this biomarker panel remained consistent when tested in

apolipoprotein E ε4 non-carrier individuals only. This blood-based panel is biologically relevant, measurable using practical immunocapture arrays, and could significantly reduce the cost incurred to clinical trials through screen failure.

Introduction:

Clinical trials for Alzheimer's disease (AD) modification have recently started to target the earlier prodromal and pre-symptomatic stages with a belief that disease modification efforts are most likely to be effective earlier in the disease process. However, conducting trials in prodromal/preclinical individuals necessitates the use of biomarkers to detect evidence of AD pathology. Currently pathology detection is possible with CSF obtained from lumbar puncture and by PET imaging. Both methods are employed routinely in clinical studies but their limitations impact significantly upon the efficiency of clinical trials. Both CSF and PET can be costly and invasive and are therefore not suitable for large-scale screening or where repeated measures are desirable. The relatively low prevalence of amyloid pathology in people with prodromal, and even more so in preclinical, disease inevitably results in high screen failure rates. The cost of this screen failure by either CSF or PET can be prohibitive. More worryingly, given that such screening is likely to be mandatory as part of clinical implementation of a successful therapeutic, this screen failure rate is likely to constitute an obstacle to clinical translation through a combination of cost factors and the capacity of health systems to enable lumbar puncture or PET imaging at the scale likely to match demands of their populations. One option to overcome this issue is to employ prediction methods such as apolipoprotein E (*APOE*) genotyping prior to trial entry, with *APOE* ε4 carriers most likely to harbour AD pathology. However, individuals with an *APOE* ε4 allele are in the minority and most prodromal/preclinical AD cases are instead *APOE* ε4 non-carriers. Additionally, genotyping only reveals risk and does not indicate current pathological state. Therefore, in order to increase the efficiency of recruitment to clinical trials, a cost effective, minimally invasive method that can be implemented on a large scale to

predict current AD pathology would be enormously valuable. A blood based assay that predicted likely pathological load would become part of a diagnostic screening funnel directing potential trials participants or users of therapy to direct markers of pathology such as CSF measures or PET imaging.

Over a decade ago, we conducted a large agnostic, or untargeted, proteome-wide, discovery study of blood biomarkers in dementia [1]. This study was successful in detecting a signal in the blood that reflected the presence of AD and since then many other studies, by others and by ourselves, have aimed to replicate and improve upon this original signal and have identified blood based protein biomarkers able to distinguish AD 'cases' from cognitively healthy elderly 'controls'. However, the relatively low rate of replication of these biomarkers across studies may be in part due to issues of a study design that compares AD to cognitively healthy elderly controls, many of whom will harbour silent pathology. Other factors limiting replication include technical issues such as assay variance and quality, differences in sampling and storage protocols and the frequently small size of many studies. Nonetheless, such studies demonstrate that there is a signature of disease detectable in blood and the task is now to find a signature that is reproducible.

To attempt this we have since serially refined our study design focussing on an 'endophenotype' approach predicating on outcomes, not of clinical diagnosis, but on a phenotype indicative of disease ('endophenotype') such as brain atrophy measured by structural MRI or Aβ plaque burden measured by PET and CSF. Using this approach we have identified putative plasma markers relating to AD pathology and disease progression using a range of proteomic approaches including mass spectrometry, SOMAscan and immunocapture [2-8] (Westwood et al, *submitted*). In addition, we also performed a series of iterative studies targeting complement and related inflammatory proteins; neuroinflammation itself being an endophenotype associated with disease and amyloid pathology. We previously showed that plasma complement factor H (CFH) was associated with AD [1], a finding replicated by ourselves and others, and several of our other studies show that complement proteins, including C3, C4B and clusterin, are repeatedly associated with amyloid load. Genetic association

studies nominate complement genes as risk factors both in single gene and in pathway analyses and complement signalling clearly plays a role in pathological processes [9,10]. In discovery phase studies, we find clusterin and CFI predict conversion from mild cognitive impairment (MCI) to AD [11] and multiple complement proteins including clusterin are higher in people with an AD polygenic score [12].

Many of the proteins identified in our discovery-phase 'endophenotype' studies replicate across proteomic platforms and different cohort types. However, unsurprisingly given the wide range of techniques and study designs, not every protein identified replicates in every study. The aim of the present study was to determine the most replicable set of plasma protein markers predicting brain amyloid, by testing 31 of our previously discovered candidate biomarkers in a large, multi-centre cohort of individuals with high and low amyloid burden.

Methods:

Subjects: EMIF Multimodal Biomarker Discovery study (EMIF-AD MBD) cohort

The EMIF-AD MBD is part of the European Medical Information Framework for Alzheimer's disease (EMIF-AD; http://www.emif.eu/), a European wide collaboration to facilitate the re-use of existing healthcare data and the sharing of cohort samples for the benefit of AD research. The EMIF-AD MBD study design, including subject selection criteria, clinical diagnoses, brain amyloid classification and plasma sample collection have all been described previously [13]. In essence though, we sought to assemble a collection of samples from participants in cohort studies from across the full disease spectrum from pre-clinical through prodromal to established disease, in each category seeking to balance those with pathology to those without. To do this we selected, using existing data wherever possible, samples for inclusion from apparently normal, cognitively healthy elderly controls, from participants with diagnosed MCI and from people with established AD. Samples were selected within each category with proven high and low amyloid load wherever possible as previously described [13]. Overall, 1221 participants (494 cognitively healthy controls, 526 MCI and 201 AD) were recruited to

the EMIF-AD MBD study from 11 European cohorts. Each parent cohort was approved by the local medical ethics committee.

The present study selected two sub-cohorts of participants from the EMIF-AD MBD study, all with plasma samples available for analysis. Firstly, 1000 individuals comprising 408 cognitively healthy individuals, 400 individuals with mild cognitive impairments, and 192 AD patients were included for proteomic analysis in the University of Oxford laboratories. Secondly, 866 individuals (93 AD 413 MCI 360 CTL) were included for proteomic analysis in the laboratories at the University of Cardiff. Participants were included in these 'Oxford' and 'Cardiff' sub-cohorts from across three multicentre studies: DESCRIPA [14], EDAR [15], and PharmaCog [16], and eight single centre studies: Amsterdam [17], Antwerp [18], San Sebastian GAP [19], Gothenburg [20], Barcelona IDIBAPS [21], Lausanne [22], Leuven [23] and Barcelona St Pau [24]. Sample number differences between the 'Oxford' and 'Cardiff' cohorts were necessary due to plasma sample availability.

Plasma analyses

Targeted plasma protein analyses were conducted at both Oxford and Cardiff laboratories using Luminex xMAP (Cat#: HNDG1MAG-36K-06, HNDG2MAG-36K-05, HNDG3MAG-36K-07, HND2MAG-39K-02, HKI6MAG-99K-03, HNDG1MAG-36K-01), ELISA (Cat#: CSB-EL008551HU and CSB-E13319H), and MSD assays (in-house optimised using U-plex platform). All assays were performed according to the manufacturer's instructions.

Brain amyloid measurements and group classifications

Measurement and classification of amyloid burden in the EMIF-AD MBD cohort has been described previously [13]. Briefly, where CSF was available, Aβ1-40, and Aβ1-42 were measured using the V-PLEX Plus Aβ Peptide Panel 1 (6E10) Kit from Meso Scale Discovery in a central laboratory (Gothenburg University, Sweden) and the $A\beta_{42/40}$ ratio was established. Where CSF was unavailable then the CSF

$A\beta_{42}$ measurement provided by the parent cohort or the standardised uptake value ratio (SUVR) from an amyloid PET scan was used.

The above measurements were combined into a continuous variable using z-scoring. The Aβ Z-score was calculated using the mean and standard deviation of the control subjects as a reference. In cases where an individual had multiple measures of amyloid (e.g. CSF and PET), all data available were used to generate the mean and standard deviation for each measure. However, the measure included in the final Aβ Z-score was selected from each individual in the following order of priority: CSF $A\beta_{42/40}$ ratio, local CSF $A\beta_{42}$ or the amyloid PET SUVR. PET amyloid Z-scores were multiplied by -1 in order to be combined with CSF derived amyloid Z-scores.

CSF tau measurements

To assess tau pathology, continuous phosphorylated tau (p-tau) and total tau (t-tau), values were obtained from the parent cohorts. As sites were not standardised to each other, the p-tau and t-tau values were Z-scored with controls within each data set as a reference.

MRI acquisition, visual rating and ROI measurements

Full details on the MRI data acquisition, visual rating check and ROI measurements have been previously reported (Ten Kate et al, *submitted*). Briefly, TI-weighted images, acquired according to local protocols, were collected from each site, each image was visually assessed and Freesurfer used to obtain volumetric measurements.

Clinical and cognitive data

Clinical information and neuropsychological test scores were collected from each local site, harmonized, pooled and stored in an online data platform using tranSMART [25]. Full details of the clinical information provided by each site and the harmonization process has been previously described in Bos et al, 2018 [13].

Statistical analyses

All statistical analyses were completed using R (version 3.3.2). Individual participant data was excluded where there was a long interval (>1 year), or missing data on the time interval, between plasma collection and measurement of the outcome variable (amyloid status, n = 69 and 30 excluded from Oxford and Cardiff cohorts respectively; MMSE, n= 73 and 39 excluded from Oxford and Cardiff respectively; brain volume, n=121 and 97 excluded from Oxford and Cardiff respectively). Baseline cohort characteristics between high and low amyloid groups were compared using Mann-Whitney U test. All analyses included age as a covariate when possible (logistic and linear regression). *P*-values and false discovery rate corrected *q*-values are reported.

Univariate statistics

Univariate statistics were performed using identical statistical methods for both the 'Oxford' and 'Cardiff' cohorts and the results for all 31 proteins are presented in this manuscript together. The relationship of each individual protein with group-wise outcome variables was tested using logistic regression. The relationship of proteins with continuous outcome variables was examined using linear regression.

ROC analysis was performed on the values of expression levels of each of the proteins individually, and the outcome was the dichotomous amyloid status. Standard ROC evaluation metrics were computed (sensitivity, specificity, positive and negative predictive values [PPV and NPV, respectively]) along with the area under the curve (AUC). The 95% confidence intervals were estimated using the bootstrap resampling method with n=1000 repetitions [26].

Multivariate amyloid classifier

Logistic regression was used to assess the performance of a multi-protein model for the discrimination between individuals in the high and low brain amyloid groups. The AUC, sensitivity, specificity, PPV and NPV and likelihood ratio (LR) of the model are reported. Other statistical approaches are described in the results

Results

The EMIF-MBD assembled a sample set from multiple cohort studies across Europe using pre-existing data wherever possible to identify people with no apparent cognitive deficit, those with mild cognitive impairment and those with established dementia and in each predementia category seeking to balance those with amyloid burden and those without. In total over 1200 samples were identified of which 1000 were used in the Oxford set and 866 in the Cardiff set with these differences being due to limited sample availability

Clinical characteristics and inter-group differences

The clinical characteristics of the Oxford and Cardiff sets, stratified by amyloid status, are presented in Table 1 along with the Mann-Whitney U inter-group difference significance level. Across the whole cohort, individuals with high amyloid status were older (p<0.001), more frequently *APOE* ε4 carriers (p<0.001) and had lower MMSE scores (p<0.001) compared to those with low amyloid status. Within the Oxford set only, individuals with high amyloid status were more frequently female (p<0.05) compared to those with low amyloid status.

Variable	Subjects included in Oxford sample set			Subjects included in Cardiff sample set		
	Low amyloid status	High amyloid status	*P* value	Low amyloid status	High amyloid status	*P* value
N	457	543	/	460	406	/
Aβ Z-score	0.49 ± 0.62	-1.35 ± 0.48	< 0.001*	0.45 ± 0.62	-1.36 ± 0.51	< 0.001*
Age (yrs.)	66.52 ± 8.71	69.81 ± 8.12	< 0.001*	66.61 ± 8.25	70.47 ± 8.22	< 0.001*
Female gender N (%)	223 (49)	301 (55)	<0.05*	253 (55)	226 (56)	0.844
CTL N (%)	289 (63)	119 (22)	/	272 (59)	88 (22)	/
MCI N (%)	147 (32)	253 (47)	/	179 (39)	234 (58)	/
AD N (%)	21 (5)	171 (31)	/	9 (2)	84 (21)	/
***APOE* genotype ε4+ N (%)**	147 (32)	349 (64)	< 0.001*	134 (29)	246 (61)	< 0.001*
MMSE	27.93 ± 2.49	25.11 ± 4.29	< 0.001*	28.09 ± 2.23	25.41 ± 4.26	< 0.001*

Table 1. Demographics of subjects from the EMIF-AD MBD study. Abbreviations: MCI, mild cognitive impairment; AD, Alzheimer's disease, *APOE*, apolipoprotein E; CTL, cognitively healthy control; MMSE, mini-mental state examination. Mean ± standard deviation. *Statistically significant p<0.05

Univariate statistics for amyloid status (high/low group)

Cross-sectional comparisons of protein concentrations between high and low amyloid status groups were performed by logistic regression (Table 2). Seven proteins remained statistically significant after multiple testing corrections (q<0.05; FCN2, B2M, apoE, CC4, cathepsin D, CFI). Five of these proteins have previously been discovered as biomarkers of brain amyloid pathology in one or more of our previous biomarker studies (FCN2, B2M, A1AT, apoE and CC4), and all five replicate the direction of change previously identified. Expression of cathepsin D has previously been found to be decreased in AD fibroblasts, and here we show that this protein is also decreased in plasma with increased AD pathology [27]. Factor I was previously found as a biomarker for conversion from MCI to dementia, with decreased a protein concentration measured in MCI converters [11]. Therefore, the direction of change identified in this study, decreased CFI with increased pathology, agrees with this previous finding.

Logistic regression analysis was also performed separately for each diagnostic group and *APOE* ε4 carrier groups (ε4 non-carrier / ε4 carrier). These results are reported in supplementary tables 1-5.

Table 2. Logistic regression (age as covariate) results for each protein with amyloid status as the outcome variable. *statistically significant <0.05

		Logistic Regression			
Sub-cohort	**protein**	**beta**	***p*-value**	***q*-value**	***N-number***
Oxford	FCN2	0.466	0.000*	0.000*	824
	FGG	-0.071	0.300	0.521	891
	Cystatin C	-0.135	0.049*	0.124	898
	Clusterin	-0.138	0.045*	0.124	906
	B2M	-0.266	0.000*	0.004*	832
	AGP	0.009	0.901	0.963	875
	CP	0.028	0.679	0.780	883
	A2M	-0.017	0.813	0.900	896

	ApoA1	-0.073	0.290	0.521	886
	ApoC3	-0.001	0.991	0.991	900
	ApoE	-0.220	0.002*	0.008*	901
	TTR	-0.039	0.572	0.772	901
	CFH	-0.049	0.475	0.701	907
	CRP	-0.139	0.049*	0.124	854
	A1AT	-0.213	0.004*	0.016*	784
	PEDF	-0.006	0.933	0.964	853
	SAP	-0.052	0.449	0.697	869
	CC4	0.243	0.001*	0.005*	894
	BDNF	-0.036	0.603	0.778	887
	Cathepsin.D	-0.238	0.001*	0.005*	880
	sICAM.1	-0.148	0.033*	0.124	893
	RANTES	-0.100	0.146	0.309	905
	NCAM	0.032	0.645	0.780	891
	sVCAM.1	-0.071	0.302	0.521	891
	PAI.1	-0.134	0.052	0.124	890
Cardiff	CR1	-0.077	0.378	0.617	770
	TCC	0.112	0.149	0.309	770
	CFB	-0.042	0.567	0.772	788
	CFI	-0.284	0.000*	0.004*	754
	Eotaxin	0.176	0.037*	0.124	749
	MCP	0.031	0.680	0.780	769

ROC analysis was performed on each of the seven proteins found to replicate direction of change to determine their individual predictive ability for the discrimination of high / low amyloid status groups, and compared to the discriminant ability of age. Table 3 displays the results of the ROC analysis and Supplementary Figure 1 displays both the AUC and corresponding 95% confidence interval of each protein and age. The AUC for every protein is higher than chance, even when including the lower end of the confidence interval.

For these seven proteins, ROC analysis was also performed separately for each diagnostic group and *APOE* ε4 carrier groups (ε4 non-carrier / ε4 carrier). These results are reported in supplementary tables 6 & 7.

Table 3. AUC statistics per protein, for the classification of high / low brain amyloid status.

Abbreviations: AUC, area under the curve.

Variable	Optimal cutpoint	Sensitivity	Specificity	AUC
FCN2	24607094.990	0.448	0.783	0.640
Age	67.355	0.641	0.560	0.619
CFI	26793.107	0.547	0.623	0.585
CC4	73789.784	0.436	0.713	0.580
B2M	4232.708	0.433	0.698	0.577
Cathepsin D	322.630	0.586	0.547	0.576
ApoE	106.020	0.667	0.445	0.554
A1AT	1675622.528	0.765	0.333	0.552

Multi-protein classifier of amyloid high / low status

It is possible, if the individual protein associations with AD are independent of each other, that a compound marker set of some or all of these proteins would have greater predictive value than any one protein alone. In order to test this, the seven proteins significant by logistic regression after multiple testing corrections (q<0.05; FCN2, B2M, apoE, A1AT, CC4, cathepsin D, CFI) were included in a logistic regression classifier, along with age, to determine their predictive ability for amyloid status when combined. After missing data was removed this 8-feature model was tested on 374 individuals and achieved moderate accuracy (AUC = 0.742 (figure 2, sensitivity = 0.682, specificity = 0.704, PPV = 0.715, NPV = 0.670, LR = 2.3). In comparison, age alone achieved an AUC = 0.617.

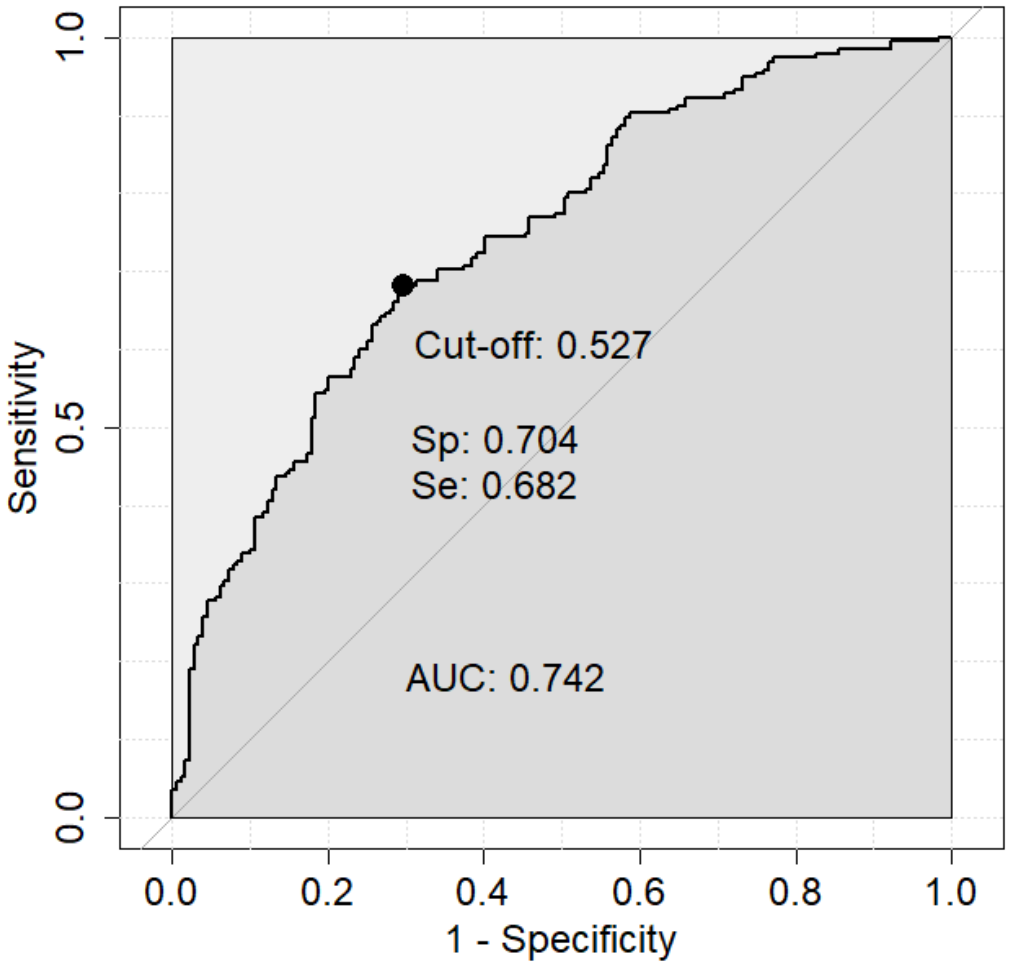


Figure 2. ROC curve obtained for the 8-feature classifier for prediction of amyloid high / low status.

To determine how this 8-feature model performs at different stages of the disease process and also in *APOE* ε4 carriers and non-carriers independently, we tested the classification ability of this model within each separate diagnostic group and *APOE* ε4 carriers and non-carriers. Table 4 displays the performance of this model within each group. The performance within the AD only group could not be accurately determined since removal of subjects with missing data left only 5% of the AD cases as amyloid negative.

Table 4. ROC and AUC statistics for the 8-feature model for the classification of amyloid status within each diagnostic group, and *APOE* ε4 carriers and non-carriers. Abbreviations: MCI, mild cognitive impairment; AD, Alzheimer's disease, *APOE*, apolipoprotein E; CTL, cognitively healthy

control; AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio.

	AUC	Sensitivity	Specificity	PPV	NPV	LR	N-number
Whole cohort (AD, MCI & CTL)	0.742	0.682	0.704	0.715	0.670	2.30	374
MCI	0.743	0.658	0.785	0.815	0.614	3.06	193
CTL	0.768	0.682	0.776	0.577	0.844	3.04	142
MCI & CTL combined	0.724	0.677	0.661	0.641	0.696	2.00	335
APOE ε4 non-carriers	0.736	0.681	0.725	0.604	0.786	2.47	199
APOE ε4 carriers	0.836	0.757	0.767	0.861	0.622	3.25	175

Relationship of classifier proteins with continuous Aβ Z-score

Having determined the relationship between the top 7 proteins and amyloid status (high/low groups) we wanted to determine whether this protein-amyloid relationship remains consistent using the Aβ Z-score (Table 5). All, except one protein (A1AT), were significantly related to Aβ Z-score after testing for multiple testing corrections (q<0.05). A1AT was tending towards significance (q=0.061).

Relationship of classifier proteins with other markers of AD pathology or disease progression

In order to determine whether the classifier proteins were specific to brain amyloid pathology or if they could also perform as a biomarker of the other key hallmark of AD, brain tau pathology, linear regression was used to assess the continuous relationship with both p-tau and t-tau Z-scores (Table 5). None of the proteins were significantly related to either measure after multiple testing correction. We then used a similar approach to examine their relationship to hippocampal volume and MMSE score, and logistic regression to examine their predictive ability for MCI conversion to dementia (Table 5). FCN2 displayed a significant relationship with all three measures (q<0.05). CFI was found to be significantly related to both MMSE and MCI conversion to dementia (q<0.05). A1AT was found to be significantly related to hippocampal volume only (q<0.05).

Table 5. Linear and logistic regression (age as covariate) results for each protein with alternative outcome measures. *statistically significant <0.05

		Aβ Z-score			P-Tau Z-score			T-Tau Z-score			Hippocampal volume			MMSE			MCI conversion to dementia		
		Linear regression			Linear regression			Linear regression			Linear regression			Linear regression			Logistic regression		
Sub-cohort	protein	beta	*p*-value	*q*-value	beta	*p*-value	*q*-value	beta	*p*-value	*q*-value	beta	*p*-value	*q*-value	beta	*p*-value	*q*-value	beta	*p*-value	*q*-value
Oxford	FCN2	-0.19	0.000*	0.000*	-0.125	0.017*	0.098	-0.138	0.019*	0.132	-244.14	0.000*	0.000*	-0.88	0.000*	0.000*	0.398	0.002*	0.006*
	B2M	0.125	0.001*	0.001*	0.019	0.715	0.876	-0.001	0.991	0.991	-14.6	0.758	0.758	0.174	0.183	0.214	0.039	0.729	0.901
	ApoE	0.123	0.000*	0.001*	0.05	0.334	0.584	0.045	0.444	0.778	37.388	0.403	0.47	-0.184	0.146	0.204	-0.015	0.901	0.901
	A1AT	0.07	0.061	0.061	-0.011	0.843	0.876	-0.006	0.924	0.991	172.168	0.000*	0.001*	0.239	0.078	0.181	-0.096	0.52	0.901
	CC4	-0.117	0.001*	0.001*	-0.088	0.09	0.209	-0.045	0.445	0.778	-70.691	0.112	0.261	-0.131	0.306	0.306	0.153	0.207	0.483
	Cathepsin D	0.086	0.013*	0.016*	-0.008	0.876	0.876	0.017	0.776	0.991	46.917	0.306	0.429	0.195	0.118	0.204	0.032	0.792	0.901
Cardiff	CFI	0.117	0.001*	0.002*	0.116	0.028*	0.098	0.107	0.063	0.221	52.396	0.256	0.429	0.442	0.001*	0.002*	-0.932	0.000*	0.000*

Discussion

We have previously used a pathology endophenotype approach to discover plasma proteomic biomarkers designed to be predictive of AD pathology and disease progression. The aim of the present study was to replicate these previously identified candidate biomarkers in a large multi-centre pragmatic sample collection collated from multiple cohorts, as well as to identify a plasma proteomic panel that could classify individuals into high or low brain amyloid groups. Our results show that in around 1000 samples from multiple studies across Europe seven biomarkers replicate, and we confirm a panel of proteins that predict high levels of amyloid pathology with a positive predictive value of 0.72 and a negative predictive value of 0.67. Moreover, the model has predictive value in both *APOE* ε4 carrier and non-carrier individuals generating a biological predictor that could be used to reduce the screen failure rate in clinical trials.

The final 7-protein biomarker panel consists of: β2-microglobulin (B2M), cathepsin D (CTSD), ficolin-2 (FCN2), complement component 4 (C4), alpha-1 antitrypsin (A1AT), complement factor I (CFI) and apolipoprotein E (apoE). Although our initial discovery-phase studies often used an unbiased approach to identify proteins that have a relationship with AD and its pathology, it is noteworthy that the resulting biomarker candidates are also biologically relevant to the disease process. B2M shares structural characteristics with Aβ [28] and is one of a number of proteins that form amyoid deposits. Cathepsin D is increased in tangle bearing neurons [29], a process that might be induced by Aβ [30]. Ficolins and mannose-binding lectins are both activators of the lectin complement pathway [31] and CSF MBL levels have been shown to be reduced in AD [32]. The complement proteins (C4 and CFI) are just two members of a pathway repeatedly shown to be associated with AD through genetics and neuropathology as well as from biomarker studies [9]. A1AT is an acute phase chemoreactant that is metabolized by the serpin enzyme complex (SEC) and hence might compete with and affect the activity of another SEC ligand, Aβ [33]. The relationship between the *APOE* ε2/ε3/ε4 polymorphism and AD is well established and complementary to our finding here that apoE protein is clearly a marker

of amyloid load. Overall, complement and inflammatory proteins dominate this protein panel, even beyond CFI that was previously identified in complement-targeted studies.

Recently, Nakamura *et al*. (2018) published an important study demonstrating that Aβ fragments can be reliably detected in blood and perform as well as current CSF biomarkers for predicting brain amyloid [34]. This approach using immunoprecipitation combined with mass spectrometry (IP+MS), was able to predict brain amyloid status with up to 90% accuracy. This finding builds upon other recent work by Ovod *et al*. (2017) who also used an IP+MS technique to identify blood Aβ with high concordance to amyloid PET [35]. Whilst these papers are important in demonstrating the value of blood Aβ as a biomarker tool for brain pathology, replication in larger sample sizes is needed, and the proteomic technology employed will require refinement to enable implementation at scale. Nevertheless, these studies provide considerable further proof of concept for AD blood biomarkers. Whilst the accuracy rates reported in our study are not as high as those reported when measuring blood Aβ levels, we are still able to achieve a level of accuracy to make a substantial impact upon the cost efficiency of clinical trials whilst using immunocapture - a practical and low-cost assay technology already in very wide use in both clinical and research laboratories.

Given that sample collection and storage protocols differed across the 11 cohorts included in the EMIF-AD MBD study, this pragmatic meta-collection reflects the challenges faced by any putative biomarker in the real-world of multi-site, multi-national clinical studies, and even more so in clinical practice, where standardisation of sample collection is sought but rarely achieved. Replication of any putative biomarker set in such a collection of samples has a higher prior probability of effective utility in practice than replication in a single cohort or single site study with a fully standardised sample collection procedure. Nonetheless the limitations of this study are also acknowledged; this study was designed to determine whether candidate biomarkers replicate in their ability to predict amyloid, it was not designed to determine the real-world value of the biomarkers or the biomarker panel they form. Selecting half of the samples from people with high amyloid enables proof of concept but clearly

random prediction will already identify 50% of those with amyloid correctly. Additionally, missing data (4-11% per protein as a result of variable sample volume and quality and occasional assay performance failure) meant that when combining multiple proteins together in a biomarker panel the overall sample size with complete data was significantly reduced. However, even in this reduced sample set with 50% prediction accuracy possible by chance, the likelihood ratio with our biomarker panel is 2.3 in the whole cohort and 3 for preclinical disease, suggesting approximately a 15% and 20% improvement in detection respectively [36]. Applying these figures to the screen failure rate occurred in actual trials demonstrates a significant impact on projected cost of start-up. For example, assuming an approximate screen failure rate of 70% in preclinical disease, a rate similar or lower than found in current clinical trials, and a current screening cost by PET amyloid of approximately $3500 per scan, in order to recruit 5000 individuals successfully the cost of screen failure would be around $41 million. Implementing our blood biomarker panel to reduce the screen failure rate by 20% would save approximately $19 million. Further work is now needed to validate the biomarker panel in samples representing those of people being recruited to clinical trials.

To summarise, our overall goal is to facilitate clinical trials by contributing to rapid and effective selection of research participants most likely to have brain amyloid pathology and hence reducing screen failure rates, reducing cost and time of trial start-up and reducing exposure of potential participants to PET imaging or CSF lumbar puncture. In order to do this, we have previously investigated, discovered and replicated plasma protein biomarkers that could be implemented as a clinical trial entry criterion to triage potential participants for amyloid PET or CSF measures. In the current study 7 of these biomarkers are replicated in a large, multi-centre cohort. These seven proteins form a biomarker panel that is the product of over a decade of research, is biologically relevant and measurable using practical immunocapture arrays and could significantly reduce the cost incurred to clinical trials by screen failure because of absence of amyloid pathology.

Conflict of Interest

SL is named as an inventor on biomarker intellectual property protected by Proteome Sciences and Kings College London and has been an advisor to a biomarker technology company; neither of which are related to the current study. HZ has served at scientific advisory boards for Eli Lilly, Roche Diagnostics, Wave, CogRx and Samumed, has received travel support from Teva, and is a co-founder of Brain Biomarker Solutions in Gothenburg AB (all unrelated to this study).

Author Contributions

SW, AB, SL contributed to study concept and design and all authors contributed to sample selection and/or interpretation of data. SW, AB, SA, AM, HZ were responsible for data acquisition and SW, ANH and AK carried out data analysis and interpretation. SW drafted the manuscript and all authors revised the manuscript.

Acknowledgements

This research was conducted as part of the EMIF-AD project which has received support from the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement n° 115372, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

Ethics Statement

Written informed consent was obtained from all participants before inclusion in the study. The medical ethics committee at each site approved the study (Supplementary Table 8).

Data Availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

References:

1. Hye, A.*, et al.* Proteome-based plasma biomarkers for Alzheimer's disease. *Brain : a journal of neurology* **129**, 3042-3050 (2006).
2. Thambisetty, M.*, et al.* Association of plasma clusterin concentration with severity, pathology, and progression in Alzheimer disease. *Archives of general psychiatry* **67**, 739-748 (2010).
3. Kiddle, S.J.*, et al.* Plasma based markers of [11C] PiB-PET brain amyloid burden. *PloS one* **7**, e44260 (2012).
4. Hye, A.*, et al.* Plasma proteins predict conversion to dementia from prodromal disease. *Alzheimer's & dementia : the journal of the Alzheimer's Association* **10**, 799-807 e792 (2014).
5. Sattlecker, M.*, et al.* Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. *Alzheimer's & dementia : the journal of the Alzheimer's Association* **10**, 724-734 (2014).
6. Ashton, N.J.*, et al.* Blood protein predictors of brain amyloid for enrichment in clinical trials? *Alzheimer's & dementia* **1**, 48-60 (2015).
7. Voyle, N.*, et al.* Blood Protein Markers of Neocortical Amyloid-beta Burden: A Candidate Study Using SOMAscan Technology. *Journal of Alzheimer's disease : JAD* **46**, 947-961 (2015).
8. Westwood, S.*, et al.* Blood-Based Biomarker Candidates of Cerebral Amyloid Using PiB PET in Non-Demented Elderly. *Journal of Alzheimer's disease : JAD* **52**, 561-572 (2016).
9. Aiyaz, M., Lupton, M.K., Proitsi, P., Powell, J.F. & Lovestone, S. Complement activation as a biomarker for Alzheimer's disease. *Immunobiology* **217**, 204-215 (2012).
10. Kolev, M.V., Ruseva, M.M., Harris, C.L., Morgan, B.P. & Donev, R.M. Implication of complement system and its regulators in Alzheimer's disease. *Curr.Neuropharmacol.* **7**, 1-8 (2009).
11. Hakobyan, S.*, et al.* Complement Biomarkers as Predictors of Disease Progression in Alzheimer's Disease. *J Alzheimers Dis* **54**, 707-716 (2016).

12. Morgan, A.R.*, et al.* The Correlation between Inflammatory Biomarkers and Polygenic Risk Score in Alzheimer's Disease. *J Alzheimers Dis* **56**, 25-36 (2017).

13. Bos, I.*, et al.* The EMIF-AD Multimodal Biomarker Discovery study: design, methods and cohort characteristics. *Alzheimers Res Ther* **10**, 64 (2018).

14. Visser, P.J.*, et al.* Development of screening guidelines and clinical criteria for predementia Alzheimer's disease. *Neuroepidemiology* **30**, 254-265 (2008).

15. Reijs, B.L.R.*, et al.* Relation of Odor Identification with Alzheimer's Disease Markers in Cerebrospinal Fluid and Cognition. *J Alzheimers Dis* **60**, 1025-1034 (2017).

16. Galluzzi, S.*, et al.* Clinical and biomarker profiling of prodromal Alzheimer's disease in workpackage 5 of the Innovative Medicines Initiative PharmaCog project: a 'European ADNI study'. *J Intern Med* **279**, 576-591 (2016).

17. van der Flier, W.M.*, et al.* Optimizing Patient Care and Research: The Amsterdam Dementia Cohort. *J Alzheimers Dis* **41**, 313-327 (2014).

18. Somers, C.*, et al.* A Decade of Cerebrospinal Fluid Biomarkers for Alzheimer's Disease in Belgium. *J Alzheimers Dis* **54**, 383-395 (2016).

19. Estanga, A.*, et al.* Beneficial effect of bilingualism on Alzheimer's disease CSF biomarkers and cognition. *Neurobiol Aging* **50**, 144-151 (2017).

20. Wallin, A.*, et al.* Alzheimer's diseasesubcortical vascular disease spectrum in a hospital-based setting: Overview of results from the Gothenburg MCI and dementia studies. *J Cerebr Blood F Met* **36**, 95-113 (2016).

21. Fortea, J.*, et al.* Increased Cortical Thickness and Caudate Volume Precede Atrophy in PSEN1 Mutation Carriers. *J Alzheimers Dis* **22**, 909-922 (2010).

22. Tautvydaite, D.*, et al.* Interaction between personality traits and cerebrospinal fluid biomarkers of Alzheimer's disease pathology modulates cognitive performance. *Alzheimers Res Ther* **9**(2017).

23. Adamczuk, K.*, et al.* Amyloid imaging in cognitively normal older adults: comparison between F-18-flutemetamol and C-11-Pittsburgh compound B. *Eur J Nucl Med Mol I* **43**, 142-151 (2016).

24. Alcolea, D.*, et al.* Relationship Between beta-Secretase, Inflammation and Core Cerebrospinal Fluid Biomarkers for Alzheimer's Disease. *J Alzheimers Dis* **42**, 157-167 (2014).

25. Scheufele, E.*, et al.* tranSMART: An Open Source Knowledge Management and High Content Data Analytics Platform. *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science* **2014**, 96-101 (2014).

26. DiCiccio, T. & Tibshirani, R. Bootstrap Confidence Intervals and Bootstrap Approximations. *Journal of the American Statistical Association* **82**, 163-170 (1987).

27. Urbanelli, L.*, et al.* Cathepsin D expression is decreased in Alzheimer's disease fibroblasts. *Neurobiol Aging* **29**, 12-22 (2008).

28. Sawaya, M.R.*, et al.* Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447**, 453-457 (2007).

29. Ginsberg, S.D., Hemby, S.E., Lee, V.M., Eberwine, J.H. & Trojanowski, J.Q. Expression profile of transcripts in Alzheimer's disease tangle-bearing CA1 neurons. *Ann.Neurol.* **48**, 77-87 (2000).

30. Hoffman, K.B., Bi, X.N., Pham, J.T. & Lynch, G. á-amyloid increases cathepsin D levels in hippocampus. *Neurosci.Lett.* **250**, 75-78 (1998).

31. Fujita, T., Matsushita, M. & Endo, Y. The lectin-complement pathway - its role in innate immunity and evolution. *Immunol Rev* **198**, 185-202 (2004).

32. Lanzrein, A.S.*, et al.* Mannan-binding lectin in human serum, cerebrospinal fluid and brain tissue and its role in Alzheimer's disease. *Neuroreport* **9**, 1491-1495 (1998).

33. Joslin, G.*, et al.* The serpin-enzyme complex (SEC) receptor mediates the neutrophil chemotactic effect of alpha-1 antitrypsin-elastase complexes and amyloid-beta peptide. *J.Clin.Invest.* **90**, 1150-1154 (1992).

34. Nakamura, A.*, et al.* High performance plasma amyloid-β biomarkers for Alzheimer's disease. *Nature* **554**, 249 (2018).

35. Ovod, V.*, et al.* Amyloid β concentrations and stable isotope labeling kinetics of human plasma specific to central nervous system amyloidosis. *Alzheimer's & Dementia* **13**, 841-849 (2017).

36. McGee, S. Simplifying likelihood ratios. *J Gen Intern Med* **17**, 646-649 (2002).

Supplementary Tables

Supplementary Table 1. Logistic regression (age as covariate) results for each protein with amyloid status as the outcome variable in AD subjects only. *statistically significant <0.05

		Logistic Regression		
Sub-cohort	**protein**	**beta**	***p*-value**	***q*-value**
Oxford	FCN2	0.722	0.062	0.635
	FGG	0.140	0.620	0.946
	Cystatin C	-0.307	0.218	0.803
	Clusterin	-0.468	0.083	0.635
	B2M	-0.484	0.035*	0.635
	AGP	0.118	0.665	0.946
	CP	0.043	0.868	0.956
	A2M	-0.261	0.288	0.803
	ApoA1	0.060	0.817	0.956
	ApoC3	-0.247	0.309	0.803
	ApoE	-0.353	0.131	0.675
	TTR	-0.123	0.630	0.946
	CFH	-0.394	0.102	0.635
	CRP	-0.052	0.849	0.956
	A1AT	0.227	0.540	0.946
	PEDF	0.033	0.907	0.956
	SAP	-0.264	0.241	0.803

	CC4	0.658	0.050*	0.635
	BDNF	-0.107	0.671	0.946
	Cathepsin D	-0.232	0.311	0.803
	sICAM1	0.038	0.881	0.956
	RANTES	0.014	0.954	0.956
	NCAM	0.218	0.452	0.877
	sVCAM1	0.029	0.911	0.956
	PAI.1	-0.199	0.381	0.817
Cardiff	CR1	0.023	0.952	0.956
	TCC	-0.312	0.234	0.803
	CFB	0.203	0.650	0.946
	CFI	0.024	0.956	0.956
	Eotaxin	-0.282	0.338	0.805
	MCP	-0.249	0.395	0.817

Supplementary Table 2. Logistic regression (age as covariate) results for each protein with amyloid status as the outcome variable in MCI subjects only. *statistically significant <0.05

		Logistic Regression		
Sub-cohort	**Protein**	**beta**	***p*-value**	***q*-value**
Oxford	FCN2	0.241	0.046*	0.264
	FGG	0.246	0.037*	0.264
	Cystatin C	-0.074	0.499	0.736
	Clusterin	0.065	0.550	0.875
	B2M	-0.213	0.060	0.264
	AGP	0.093	0.403	0.658
	CP	0.170	0.127	0.342
	A2M	0.000	0.998	0.998
	ApoA1	0.163	0.148	0.343
	ApoC3	0.048	0.663	0.859

	ApoE	-0.305	0.006*	0.188
	TTR	0.001	0.991	0.998
	CFH	0.035	0.748	0.859
	CRP	-0.158	0.155	0.343
	A1AT	0.009	0.935	0.998
	PEDF	-0.110	0.321	0.562
	SAP	-0.087	0.425	0.658
	CC4	0.217	0.053	0.264
	BDNF	0.021	0.846	0.936
	Cathepsin D	-0.170	0.118	0.342
	sICAM1	-0.106	0.327	0.562
	RANTES	0.036	0.742	0.859
	NCAM	0.177	0.121	0.342
	sVCAM1	-0.038	0.728	0.859
	PAI.1	-0.127	0.235	0.454
Cardiff	CR1	0.234	0.031*	0.264
	TCC	0.176	0.132	0.342
	CFB	-0.137	0.189	0.391
	CFI	-0.263	0.018*	0.264
	Eotaxin	0.198	0.101	0.342
	MCP	0.037	0.725	0.859

Supplementary Table 3. Logistic regression (age as covariate) results for each protein with amyloid status as the outcome variable in cognitively healthy control subjects only. *statistically significant <0.05

		Logistic Regression		
	Protein	**beta**	***p*-value**	***q*-value**
Oxford	FCN2	0.305	0.010*	0.180
	FGG	-0.137	0.262	0.427

	Cystatin C	-0.228	0.066	0.256
	Clusterin	-0.166	0.172	0.368
	B2M	-0.263	0.053	0.256
	AGP	-0.216	0.087	0.298
	CP	-0.254	0.047*	0.256
	A2M	-0.162	0.202	0.368
	ApoA1	-0.245	0.060	0.256
	ApoC3	-0.016	0.891	0.920
	ApoE	-0.175	0.160	0.368
	TTR	-0.180	0.145	0.368
	CFH	-0.178	0.146	0.368
	CRP	0.011	0.924	0.924
	A1AT	-0.359	0.014*	0.180
	PEDF	-0.020	0.872	0.920
	SAP	0.139	0.235	0.404
	CC4	0.150	0.201	0.368
	BDNF	0.094	0.427	0.614
	Cathepsin D	-0.097	0.436	0.614
	sICAM1	-0.288	0.045*	0.256
	RANTES	0.084	0.471	0.634
	NCAM	-0.017	0.888	0.920
	sVCAM1	-0.017	0.884	0.920
	PAI.1	0.156	0.180	0.368

Cardiff	CR1	-0.667	0.017*	0.180
	TCC	-0.044	0.781	0.920
	CFB	0.092	0.497	0.642
	CFI	-0.242	0.137	0.368
	Eotaxin	0.143	0.276	0.428
	MCP	0.040	0.767	0.920

Supplementary Table 4. Logistic regression (age as covariate) results for each protein with amyloid status as the outcome variable in APOE ε4 carrier subjects only. *statistically significant <0.05

		Logistic Regression		
Sub-cohort	**Protein**	**beta**	***p*-value**	***q*-value**
Oxford	FCN2	0.628	0.000*	0.000*
	FGG	-0.258	0.013*	0.120
	Cystatin C	-0.127	0.234	0.409
	Clusterin	-0.136	0.204	0.400
	B2M	-0.111	0.305	0.431
	AGP	0.033	0.761	0.795
	CP	0.031	0.770	0.795
	A2M	0.064	0.575	0.660
	ApoA1	-0.071	0.503	0.606
	ApoC3	-0.074	0.480	0.606
	ApoE	-0.108	0.306	0.431
	TTR	-0.136	0.200	0.400
	CFH	-0.023	0.831	0.831

	CRP	-0.236	0.025*	0.129
	A1AT	-0.221	0.050*	0.192
	PEDF	-0.121	0.251	0.409
	SAP	-0.187	0.081	0.280
	CC4	0.405	0.001*	0.009*
	BDNF	-0.168	0.106	0.329
	Cathepsin D	-0.253	0.017*	0.120
	sICAM1	-0.141	0.179	0.400
	RANTES	-0.241	0.019*	0.120
	NCAM	0.131	0.244	0.409
	sVCAM1	0.071	0.509	0.606
	PAI.1	-0.160	0.122	0.345
Cardiff	CR1	0.047	0.707	0.783
	TCC	0.134	0.297	0.431
	CFB	0.090	0.461	0.606
	CFI	-0.151	0.206	0.400
	Eotaxin	0.325	0.030*	0.133
	MCP	0.202	0.207	0.400

Supplementary Table 5. Logistic regression (age as covariate) results for each protein with amyloid status as the outcome variable in *APOE* ε4 non-carrier subjects only. *statistically significant <0.05

		Logistic Regression		
Sub-cohort	**Protein**	**beta**	***p*-value**	***q*-value**
Oxford	FCN2	0.395	0.000*	0.005*

	FGG	0.083	0.406	0.716
	Cystatin C	-0.206	0.043*	0.222
	Clusterin	-0.241	0.017*	0.125
	B2M	-0.427	0.000*	0.005*
	AGP	-0.062	0.545	0.769
	CP	-0.021	0.836	0.894
	A2M	-0.033	0.751	0.894
	ApoA1	-0.052	0.609	0.821
	ApoC3	0.077	0.439	0.716
	ApoE	-0.022	0.827	0.894
	TTR	0.062	0.539	0.769
	CFH	0.012	0.903	0.908
	CRP	0.082	0.416	0.713
	A1AT	-0.190	0.081	0.288
	PEDF	0.110	0.281	0.669
	SAP	0.065	0.513	0.769
	CC4	0.175	0.075	0.288
	BDNF	0.043	0.668	0.863
	Cathepsin D	-0.249	0.020*	0.125
	sICAM1	-0.177	0.084	0.288
	RANTES	0.037	0.713	0.884
	NCAM	-0.025	0.799	0.894
	sVCAM1	-0.161	0.112	0.348
	PAI.1	-0.090	0.384	0.716
Cardiff	CR1	-0.165	0.268	0.669
	TCC	0.090	0.387	0.716
	CFB	-0.087	0.417	0.716
	CFI	-0.362	0.002*	0.022*
	Eotaxin	0.125	0.234	0.660
	MCP	-0.012	0.908	0.908

Supplementary Table 6. AUC statistics per protein, for the classification of high / low brain amyloid status, within *APOE* ε4 carrier / non-carrier groups separately.

Variable	*APOE* ε4 status	Optimal cutpoint	Sensitivity	Specificity	AUC	CI.low	CI.up
A1AT	carrier	142278.725	0.408	0.706	0.563	0.483	0.650
A1AT	non-carrier	273159.570	0.544	0.571	0.546	0.468	0.623
ApoE	carrier	93.568	0.624	0.456	0.527	0.455	0.609
ApoE	non-carrier	138.613	0.836	0.233	0.501	0.426	0.570
B2M	carrier	2957.260	0.276	0.872	0.554	0.478	0.632
B2M	non-carrier	6191.573	0.658	0.524	0.604	0.531	0.676
Cathepsin D	carrier	270.670	0.395	0.805	0.597	0.526	0.674
Cathepsin D	non-carrier	393.785	0.784	0.363	0.573	0.496	0.642
CC4	carrier	42726.128	0.752	0.450	0.621	0.545	0.698
CC4	non-carrier	73547.781	0.467	0.693	0.556	0.488	0.627
FCN2	carrier	18215589.820	0.630	0.682	0.674	0.596	0.745
FCN2	non-carrier	24620022.060	0.515	0.772	0.626	0.547	0.696
CFI	carrier	25047.264	0.460	0.684	0.552	0.469	0.634
CFI	non-carrier	26770.591	0.529	0.647	0.596	0.524	0.673
Age	carrier	65.010	0.731	0.534	0.652	0.585	0.714
Age	non-carrier	67.355	0.708	0.532	0.643	0.580	0.706

Supplementary Table 7. AUC statistics per protein, for the classification of high / low brain amyloid status, within diagnostic groups separately.

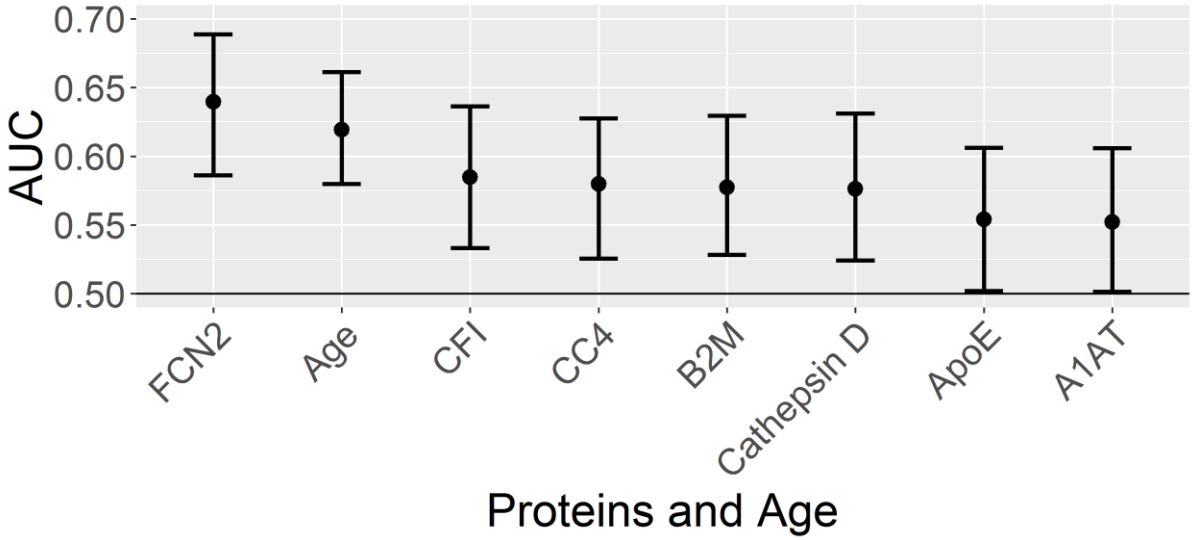
Variable	Diagnosis	Optimal cutpoint	Sensitivity	Specificity	AUC	CI.low	CI.up
A1AT	MCI	253270.314	0.481	0.627	0.526	0.442	0.612
A1AT	AD	54461.556	0.299	0.917	0.500	0.250	0.730
A1AT	CN	3758206.009	0.925	0.240	0.559	0.466	0.655
ApoE	MCI	115.549	0.722	0.449	0.583	0.501	0.669
ApoE	AD	111.074	0.720	0.563	0.626	0.396	0.864
ApoE	CN	104.386	0.693	0.431	0.545	0.462	0.632
B2M	MCI	18721.188	0.948	0.150	0.535	0.448	0.623
B2M	AD	7032.937	0.669	0.643	0.671	0.421	0.899
B2M	CN	7295.702	0.753	0.419	0.589	0.497	0.679
Cathepsin D	MCI	452.920	0.859	0.261	0.551	0.470	0.629
Cathepsin D	AD	319.070	0.600	0.611	0.602	0.418	0.778
Cathepsin D	CN	272.795	0.396	0.741	0.561	0.464	0.655
CC4	MCI	44884.347	0.764	0.387	0.567	0.484	0.643
CC4	AD	77450.398	0.401	0.882	0.655	0.453	0.865
CC4	CN	78252.733	0.354	0.777	0.551	0.459	0.642
FCN2	MCI	22524455.125	0.500	0.712	0.597	0.514	0.681
FCN2	AD	18939335.605	0.723	0.643	0.679	0.446	0.908
FCN2	CN	30220128.285	0.276	0.891	0.571	0.479	0.657
CFI	MCI	26825.973	0.462	0.713	0.587	0.511	0.661
CFI	AD	29249.109	0.875	0.333	0.508	0.033	0.929
CFI	CN	24171.476	0.424	0.752	0.566	0.459	0.678
Age	MCI	67.350	0.709	0.483	0.603	0.536	0.665
Age	AD	68.900	0.571	0.200	0.381	0.206	0.555
Age	CN	66.100	0.593	0.582	0.570	0.484	0.655

Supplementary Table 8. Ethical approval committee of each centre

Centre	Part of multicentre	Country	Approval Committee
Aristotle University, Thessaloniki	DESCRIPA, EDAR, Pharmacog	Greece	Aristotle University of Thessaloniki Medical School Ethics Committee
Central Institute for Mental Health, Mannheim	EDAR	Germany	Ethics Committee of the Medical Faculty Mannheim, University of Heidelberg
GAP, San Sebastian	-	Spain	Ethic and Clinical Research Committee Donostia

Hôpital Timone Adultes, Marseille	Pharmacog	France	Ethics committee Inserm and Aix Marseille University
Hospital Clínic de Barcelona IDIBAPS	Pharmacog	Spain	The Healthcare Ethics Committee of the Hospital Clínic
Hospital de la Santa Creu i Sant Pau, Barcelona	EDAR	Spain	Central Clinical Research and Clinical Trials Unit (UICEC Sant Pau)
INSERM, Toulouse	Pharmacog	France	INSERM Ethical Committee
IRCCS-FBF, Brescia	Pharmacog	Italy	Ethic Committee of the IRCCS San Giovanni di Dio FBF
IRCCS-SDN, Napels	Pharmacog	Italy	Comitato Etico IRCCS Pascale - Napoli
Karolinska Institutet, Stockholm	EDAR	Sweden	Ethics Committee at Karolinska Institutet
Katholieke Universiteit, Leuven	EDAR	Belgium	Ethische commissie onderzoek UZ/KU Leuven
Lausanne University Hospital, Lausanne	-	Switzerland	Research Ethics Committee Lausanne University Hospital
Maastricht University, Maastricht	DESCRIPA, EDAR	Netherlands	Medical ethical committee Maastricht University Medical Center
Rigshospitalet, Copenhagen	EDAR	Denmark	Committee on Health Research Ethics, Region of Denmark
University of Mediterranean, Marseille	Pharmacog	France	Ethics committee of Mediterranean University
University of Lille, Lille	Pharmacog	France	University of Lille Ethics committee
University of Leipzig, Leipzig	Pharmacog	Germany	Ethical Committee at the Medical Faculty, Leipzig University
University of Essen, Essen	Pharmacog	Germany	Ethical Committee at the Medical Faculty, University Hospital Essen
University of Antwerp, Antwerp	-	Belgium	Ethics committee University of Antwerp
University of Genoa, Genoa	Pharmacog	Italy	Ethical Committee of University of Genoa
University of Gothenburg, Gothenburg	-	Sweden	Ethics Committee, University of Gothenburg
University of Perugia, Perugia	Pharmacog	Italy	Human ethics Committee of the University of Perugia
VU Medical Center, Amsterdam	EDAR, Pharmacog	Netherlands	Medical ethics committee VU Medical Center

Supplementary Figures



Supplementary Figure 1. AUC and corresponding 95% confidence intervals plotted per protein, for the classification of high / low brain amyloid status. Abbreviations: AUC, area under the curve.