

# Prognostic Polypeptide Blood Plasma Biomarkers of Alzheimer's Disease Progression

Hongqian Yang<sup>a</sup>, Yaroslav Lyutvinskiy<sup>a</sup>, Sanna-Kaisa Herukka<sup>b</sup>, Hilkkka Soininen<sup>b</sup>, Dorothea Rutishauser<sup>a</sup> and Roman A. Zubarev<sup>a,c,\*</sup>

<sup>a</sup>*Division of Physiological Chemistry I, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden*

<sup>b</sup>*Department of Neurology, School of Medicine, University of Eastern Finland, Kuopio, Finland*

<sup>c</sup>*SciLifeLab, Stockholm, Sweden*

Accepted 13 December 2013

## Abstract.

**Background:** Patients with mild cognitive impairment (MCI) have varying risks of progression to Alzheimer's disease (AD).

**Objective:** To test the utility of the relative abundances of blood plasma polypeptides for predicting the risk of AD progression.

**Methods:** 119 blood plasma samples of patients with MCI with different outcomes (stable MCI and progressive MCI) were analyzed by untargeted, label-free shotgun proteomics. Predictive biomarkers of progressive MCI were selected by multivariate analysis, followed by cross-validation of the predictive model.

**Results:** The best model demonstrated the accuracy of ca. 79% in predicting progressive MCI. Sex differences of the predictive biomarkers were also assessed. We have identified some sex-specific protein biomarkers, e.g., alpha-2-macroglobulin (A2M), which strongly correlates with female AD progression but not with males.

**Conclusion:** Significant sex bias in AD-specific biomarkers underscores the necessity of selecting sex-balanced cohort in AD biomarker studies, or using sex-specific models. Blood protein biomarkers are found to be promising for predicting AD progression in clinical settings.

Keywords: Biomarkers, human blood plasma, label-free quantification, mass spectrometry

## INTRODUCTION

Alzheimer's disease (AD) is the most common cause of senile dementia [1]. It has a long asymptomatic phase, which can last decades before the clinical onset [2]. AD diagnosis relies on medical history, physiological and cognitive tests, and neuroimaging techniques. Currently, there is no cure for AD, which might be

due to the lack of early and accurate diagnosis [3]. To address these issues, the National Institute on Aging redefined three AD stages (dementia due to AD; mild cognitive impairment (MCI) due to AD; and preclinical AD), and recommended implementing biomarkers as a complementary diagnosis tool [4]. For personalized treatment, it is important not only to diagnose AD, but also to identify the MCI patients that will rapidly progress to AD (P-MCI) as opposed to those that are likely to remain stable with MCI (S-MCI). So far the most studied and validated polypeptide biomarkers are those found in cerebrospinal fluid (CSF). Amyloid- $\beta$

\*Correspondence to: Roman A. Zubarev, PhD, Department of Medical Biochemistry & Biophysics, Karolinska Institutet, Scheelesväg 2, SE 17177 Stockholm, Sweden. Tel.: +46 8524 87594; E-mail: roman.zubarev@ki.se.

Table 1  
Descriptive statistics of the study population at baseline

Pooled samples	Stable MCI	Progressive MCI	Individual samples	Stable MCI	Progressive MCI
Patients, <i>n</i> = 139	92	47	Patients, <i>n</i> = 119	76	43
Gender, male/female (%/%)	32/60 (35/65)	13/34 (28/72)	Gender, male/female (%/%)	28/48 (37/63)	15/28 (35/65)
Age at baseline, years ( $\pm$ s.d.)	72 $\pm$ 5	71 $\pm$ 6	Age at baseline, years ( $\pm$ s.d.)	72 $\pm$ 5	71 $\pm$ 6
MMSE ( $\pm$ s.d.)	24.6 $\pm$ 3.0	23.7 $\pm$ 2.7	MMSE ( $\pm$ s.d.)	22.6 $\pm$ 4.1	22.4 $\pm$ 3.3
Follow-up time, months ( $\pm$ s.d.)	28 $\pm$ 16	27 $\pm$ 18	Follow-up time, months ( $\pm$ s.d.)	27 $\pm$ 17	28 $\pm$ 18

peptide (A $\beta$ )<sub>42</sub>, total tau (t-tau), and phosphorylated tau (p-tau) are in use for discriminating AD versus healthy subjects with the sensitivity and specificity around 80–90% [5]. Similarly, somewhat lower levels of accuracy are claimed for the differentiation of P-MCI from S-MCI, although broader validation is needed for these latter claims [6, 7]. However, CSF is an invasive biopsy and not routinely analyzed for MCI patients in most countries.

Compared with CSF, blood analysis is much less invasive and routinely used in clinics in massive screenings, and thus prediction of the AD progress by blood biomarkers in presymptomatic individuals would be highly valuable. Indeed, several studies have investigated blood samples in search for AD biomarkers in the past decade [7]. One of the most influential studies found a panel of 18 signaling plasma proteins that differentiate AD and healthy control with sensitivity and specificity around 90% [8], although later studies similarly based on immunological assays showed worse performance [9]. In the original study [8], the combination of 18 proteins could discriminate S-MCI and P-MCI with the sensitivity and specificity around 80%, which, to our knowledge, is the best performance achieved among such type of studies [10, 11].

This and similar immunology-based studies have frequently focused on proteins linked to AD disease progression, which are often found in blood at low or ultralow concentrations, e.g., cytokines. These proteins are not easily accessible for mass spectrometry analysis, and when they are, the accuracy of the abundance measurements suffers from low signal levels, which reflects in poor values of the coefficient of variability (CV) [12]. In contrast, abundant blood proteins can be measured by label-free analysis with CVs as low as 1% to 3% [13]. We have recently shown that accurately measured levels of ca. 100 most abundant blood proteins reflect important phenotype differences, such as sex. A panel of ca. 20 proteins differentiated males from females with ca. 90% accuracy [13]. Hypothetically, the relative concentrations of highly abundant blood proteins can be predictive of the AD progression. Indeed, the third most abundant protein in blood,

alpha-2-macroglobulin (A2M), is named in literature as one of the AD biomarker candidates [14, 15]. In the present study, we aimed at testing the above hypothesis and determining the predictive power of abundant blood proteins for AD progression, i.e., differentiation between S-MCI and P-MCI. In doing so, we used the same label-free proteomics technique that has previously been employed for sex differentiation from blood plasma samples [13].

## MATERIALS AND METHODS

### Participants

139 blood plasma samples of elderly MCI patients were selected from the Kuopio cohort [16] and pooled into four age-matched groups based on their sexes and disease stages, such as S-MCI and P-MCI. Then 119 samples among them were randomly selected for individual analysis. The participant and pooling information is given in Table 1. Informed written consent was acquired from all the subjects according to the Declaration of Helsinki, and the study was approved by the Ethics Committee of the Kuopio University Hospital (Finland). The plasma samples were collected in the morning and mostly after fasting. The frozen plasma were then stored at  $-80^{\circ}\text{C}$  until further analysis. The follow-up diagnosis performed on average 28 months after the sample collection revealed the S-MCI/P-MCI status of the patients.

### Protein extraction and solubilization

Samples were analyzed as grouped in four pools according to sex and AD progression [17], as well as individually. 0.2  $\mu\text{L}$  plasma proteins from each sample were dissolved in a mixture of 50 mM ammonium bicarbonate (AmBic) in 10% acetonitrile (ACN) with 0.1% Protease MAX<sup>TM</sup> Surfactant Trypsin Enhancer (Promega) to a total volume of 80  $\mu\text{L}$  per sample. The sample mixtures were incubated for 15 min at  $50^{\circ}\text{C}$ , sonicated for 10 min, and centrifuged for 5 min to get rid of the undissolved debris.

### *Trypsin digestion*

Each pooled blood plasma sample was independently digested in triplicate, and individual sample in duplicate. 70  $\mu$ L supernatant from each sample was taken and digested by the MassPrep (Packard) robot. The proteins were reduced by adding 25  $\mu$ L of 20 mM dithiothreitol (DTT) in 50 mM AmBic and incubated at 56°C for 30 min. 25  $\mu$ L of 66 mM iodoacetamide in 50 mM AmBic was further added for alkylation at room temperature for 30 min. Then 25  $\mu$ L of 13 ng/ $\mu$ L sequencing grade modified trypsin (Promega) was added to each sample and incubated at 37°C overnight. The digestion was stopped by adding 7  $\mu$ L of formic acid (FA) and incubating the solution for 20 min at 37°C. Then the samples were desalted by StageTips (Thermo), dried by SpeedVac, and stored at -20°C until further analysis.

### *Mass spectrometry (MS) analysis*

Each digest was resuspended in 0.1% FA, and 0.5  $\mu$ g of protein digest was used in a single analysis. For technical reasons, the analysis of the 238 individual samples (duplicates from 119 patients) was performed in four series of analytical runs, with each group analyzed with an at least two weeks break from the previous one. The analysis sequence was randomized within each group. Group I and II of the runs analyzed 1<sup>st</sup> digestion of S-MCI and P-MCI from both sexes (53 and 66 samples of both sexes, respectively), Group III - 76 samples of the 2<sup>nd</sup> digestion of female samples, and Group IV - 43 samples of the 2<sup>nd</sup> digestion of male samples. Moreover, different chromatographic instrumentation was employed: Group I and Group II analyses were performed using nanoAcquity Ultra Performance LC<sup>®</sup> (Waters), while Group III and IV samples together with pooled samples were analyzed by Easy-nLC system (Thermo Fisher Scientific). In all cases, the LC was coupled online with Velos Orbitrap mass spectrometer (Thermo Fisher Scientific). The fact that the analysis was performed in groups and with different chromatographic equipment complicated the data processing, but made the results more realistic and relevant to clinical settings.

Both LC systems used elution buffer A containing 0.1% FA, and buffer B containing 0.1% FA in ACN with a flow rate of 300 nL/min. The LC elution conditions are given in Supplementary material.

Velos Orbitrap mass spectrometer analyzed the eluted peptides that were ionized with electrospray ionization. The survey mass spectrum was acquired

at a resolution  $R = 60,000$ , with  $m/z$  of ions ranging from 300 to 2,000. Five most abundant ions were selected with a window of 3  $m/z$  units and fragmented by higher-energy collision dissociation (HCD) as well as electron transfer dissociation (ETD) MS/MS. The HCD fragments were detected in the Orbitrap at a resolution  $R = 7,500$ , while ETD fragments were detected in the Velos trap at low resolution.

In summary, there were four pooled samples (S-MCI and P-MCI, for both sexes), each digested in triplicate, with each digest analyzed in two technical replicate LC-MS runs, thus yielding six LC-MS analyses for each pooled sample. For each individual sample, there were two independent digests, with each digest analyzed once. Thus there were two LC-MS analyses for each individual sample.

### *Data processing*

The LC-MS data obtained from each of the four groups of individual samples was processed separately by Quanti software, which performs accurate label-free peptide and protein quantification with correction for instrumental response fluctuations [13]. The data obtained from the pooled samples was also processed in a similar way as described in [13] in details.

### *Proteins and peptides identification*

MS/MS spectra were extracted using a home-written program RAW\_to\_MGF which selected in each MS/MS spectrum up to 200 most abundant ions. The MS/MS data from different LC-MS runs within the same group were clustered together using the program Cluster\_MGF to make a single .MGF file for each group. Cluster\_MGF gathers groups of spectra whose precursors are presumed to be the same peptides. Spectra were included in this group if they shared at least 12 of the 20 most-abundant ions with at least one other MS/MS spectrum in the group. One spectrum from each group with the maximum aggregate intensity is taken as a representative of this group for deposition in the .MGF file. The resultant .MGF files were searched by Mascot search engine (Matrix Science, London, UK) version 2.3.02 against April 2013 version of human reviewed canonical sequence complete proteome database from UniProtKB [18] (contains 33,226 sequences, 16,613 of them are reversed) using HCD and ETD MS/MS data, with precursor mass accuracy of 10 ppm, MS/MS accuracy of 0.6 Da, a maximum of 2 missed tryptic cleavages, carbamidomethylation of cysteine as a fixed modification, asparagine

and glutamine deamidation and methionine oxidation as variable modifications. Peptide assignments were treated as false positives if the best match of the corresponding MS/MS spectrum in the database was to a reversed protein sequence.

#### *Proteins and peptides quantification*

Quantification of the peptides and proteins was performed by Quanti version 2.5.3.1 [13]. This program performs quantification of peptides found by Mascot, or externally supplied, based on peptides' extracted ion chromatograms. Proteins are quantified based on peptide abundances. Quanti uses for quantification only reliably identified (FDR <0.01), first-choice, unmodified, unique-sequence peptides. If two database protein entries have partial sequence overlap, then all the peptides belonging to this overlap are excluded from quantification. As an output, Quanti provided tables of relative abundances for proteins and underlying peptides.

#### *Statistical analysis*

Only proteins found in all four experimental groups (male/female, S-MCI/P-MCI) were submitted for subsequent statistical analysis. Since each sample was analyzed twice, the reproducibility of analysis could be evaluated for each protein in each sample. The overall median CV for relative protein abundances within the same sample measured in different replicates was calculated as 11%, which includes both sample preparation (protein extraction, digestion, desalting) and LC-MS-related variability. Thereafter the abundances of proteins with CV more than 30% (21% of all protein abundance values) were excluded as unreliable measurements (treated as missing data) from the results for the corresponding sample. The protein abundances were log-transformed and then averaged between two replicates. The resulting dataset was submitted to orthogonal projection to latent structures discriminant analysis (OPLS-DA) by SIMCA software (version 13.0.0.0, Umetrics AB, Sweden). OPLS-DA model was built for discrimination of S-MCI and P-MCI groups. To avoid overestimation, sevenfold cross-validated scores were calculated [19]. These scores were used for receiver operator characteristic (ROC) analysis to estimate the predictive accuracy of the model [20]. As a negative control, S-MCI/P-MCI identifier was randomized for all MCI samples, and the above procedure was repeated.

#### *Biological function analysis*

The functional analysis of the potential prognostic biomarkers was performed by the overrepresentation analysis in the pathway analysis of REACTOME [21]. Protein prognostic biomarkers categorized by OPLS-DA as positively/negatively correlating with P-MCI were assessed separately by the overrepresentation analysis using  $p$  value less than  $1.0e-08$  as the significance cutoff empirically.

## RESULTS

#### *Protein putative biomarker analysis*

125 proteins were quantified in all four groups with FDR  $\leq 1\%$  and the other requirements described above. OPLS-DA model gave  $R^2 = 0.42$  and  $Q^2 = 0.30$  for the discrimination of S-MCI and P-MCI groups, and showed the predictive accuracy of 79% (Fig. 1). Among the 60 proteins (Supplementary Table 1) significantly contributing to the S-MCI/P-MCI discrimination by the OPLS-DA model, the top twelve (six most positively and six most negatively correlating with P-MCI) proteins are given in Table 2 as putative AD progression biomarkers.

In the negative control, where S-MCI/P-MCI identifier was randomized for all samples, no valid OPLS-DA model ( $R^2 = 0.24$ ,  $Q^2 = -0.08$ ) was obtained, and cross-validation confirmed that the model is statistically indistinguishable from random guessing with the "predictive accuracy" of only 59% (Supplementary Fig. 1).

#### *Sex effects in blood protein biomarkers*

It is known that human blood proteome has sex-specific profiles [13, 17, 22, 23]. To assess such effects among the AD biomarker candidates, we calculated the correlation factors for all the 60 significantly predictive proteins with the disease progression (P-MCI). Opposite regulation directions between the two sexes were found in 13 proteins out of the total 60 proteins (22%) (Supplementary Table 1). Proteins that correlate with P-MCI differently in females and males, e.g., A2M (Fig. 2A-B), are more informative in sex-specific models than in the sex-unified model.

The OPLS-DA models built on the same dataset with sex differentiation showed 86% accuracy for males and 73% accuracy for females (Supplementary Figs. 2 and 3). Since addition of sex specificity should always increase the prediction accuracy, the lower accuracy

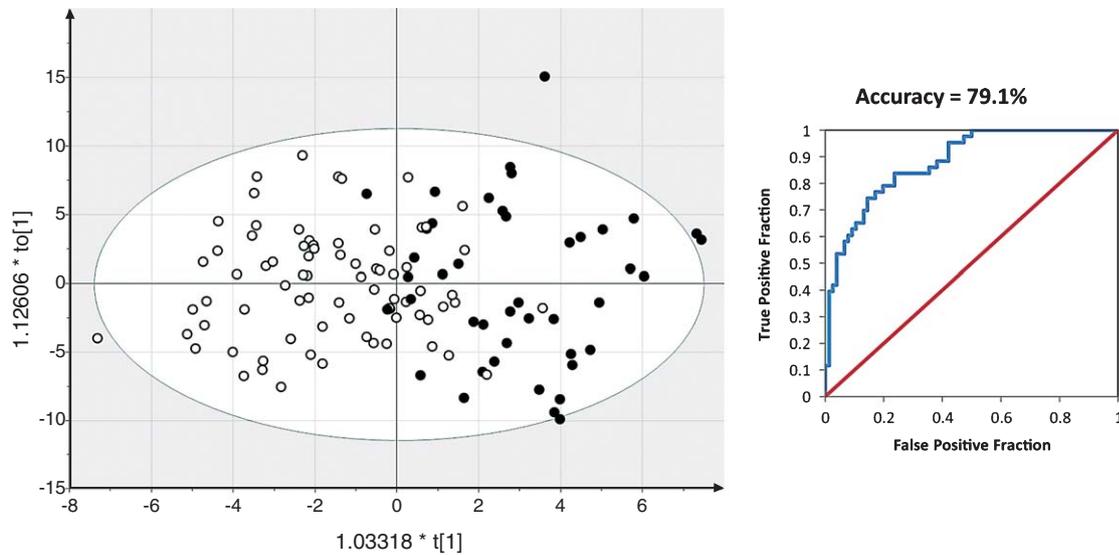


Fig. 1. Left: Orthogonal projection to latent structures discriminant analysis (OPLS-DA) model for discrimination of stable mild cognitive impairment (S-MCI) and progressive MCI (P-MCI) patient samples. Open circles represent S-MCI samples, black dots represent P-MCI samples. Right: ROC-curve based on seven-fold cross validation of the corresponding OPLS-DA model.

Table 2

Blood plasma proteins that, according to OPLS-DA model, most significantly contribute to the discrimination of stable mild cognitive impairment (S-MCI) and progressive MCI (P-MCI) patients. Note that the regulation values here depend upon the variability within the dataset: larger variability leads to smaller regulation values [13]

Protein ID	Protein desc.	Corr- female	Corr- male	Corr- all	Regulation: female (%)	Regulation: male (%)	Regulation: all (%)	Min pep	Max pep
CFAB_HUMAN	Complement factor B	-0.23	-0.61	-0.35	-6.5	-15.5	-3.2	16	25
CFAL_HUMAN	Complement factor I	-0.34	-0.45	-0.38	-6.9	-9.0	-2.2	1	4
A1AG1_HUMAN	Alpha-1-acid glycoprotein 1	-0.25	-0.09	-0.20	-15.6	-5.1	-2.0	3	6
IGHG3_HUMAN	Ig gamma-3 chain C region	-0.09	-0.29	-0.16	-8.4	-27.0	-6.8	3	4
CERU_HUMAN	Ceruloplasmin	-0.35	-0.59	-0.41	-9.7	-12.6	-2.6	16	26
CFAH_HUMAN	Complement factor H	-0.24	-0.49	-0.33	-5.9	-14.0	-1.4	26	35
IC1_HUMAN	Plasma protease C1 inhibitor	0.18	0.20	0.18	10.6	14.4	2.3	6	8
FINC_HUMAN	Fibronectin	0.22	-0.08	0.15	23.6	-4.9	11.9	23	41
FIBG_HUMAN	Fibrinogen gamma chain	0.25	0.35	0.28	33.2	36.5	16.3	26	34
FIBB_HUMAN	Fibrinogen beta chain	0.27	0.29	0.27	17.4	15.5	6.3	22	30
FIBA_HUMAN	Fibrinogen alpha chain	0.28	0.23	0.26	18.2	12.7	5.9	29	41
GELS_HUMAN	Gelsolin	0.01	0.25	0.09	2.9	42.7	7.0	10	13

for females-only model compared to sex-unified model was surprising. It was probably due to two reasons: the stochastic nature of the accuracy determination (the estimated uncertainty is  $\pm 10\%$ ), and the detrimental effect of a smaller cohort on statistical analysis.

Note that of the 12 P-MCI predictive proteins in Table 2 selected by OPLS-DA model, all but one protein have the same direction of regulation in all three models. Only fibronectin is found strongly upregulated in females (+25%) and weakly downregulated in males (-5%), with moderate upregulation (+12%) in the joint model. Note also that the regulation factors for the joint model are more modest in value than in sex-specific models. This is a feature of the quan-

tification approach used [13], which takes variability into account when estimating the regulation: stronger variability, as in the joint model, reduces the perceived regulation value. As has been shown earlier, such an approach improves the accuracy of predictive models by suppressing stochastic fluctuations.

*Protein biological function analysis by REACTOME*

Among the 47 proteins negatively correlating with P-MCI patients (Supplementary Table 1), complement cascade was significantly enriched, encompassing 14 proteins ( $p = 6.7e-17$ ). Among the 13 proteins

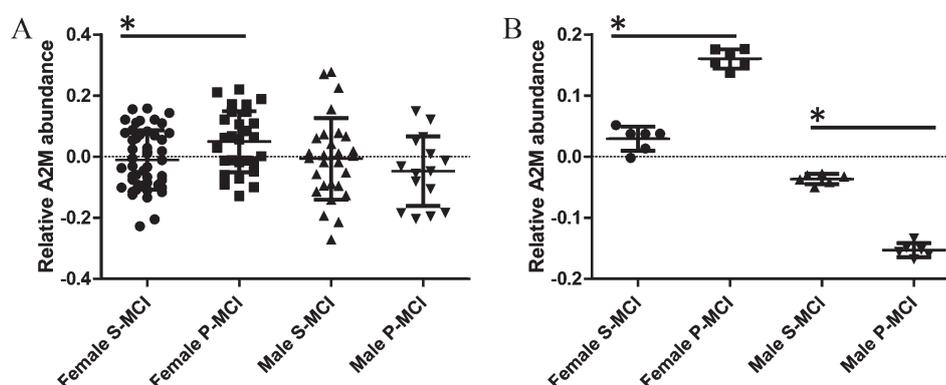


Fig. 2. Box plots of relative protein abundance measured for: A) alpha-2-macroglobulin (A2M) from individual samples, and B) A2M from pooled samples. In both cases, the values correlate differently with AD progression for two sexes. The middle bar indicates mean value, upper and lower bars encompass data deviating by one standard deviation, and \*represents  $p < 0.05$  by two-tailed Student  $t$  test.

positively correlating with P-MCI patients (Supplementary Table 1), hemostasis was enriched, with 9 proteins ( $p = 1.6e-09$ ).

## DISCUSSION

### “Top proteome” blood analysis

Blood proteome has a concentration dynamic range of more than ten orders of magnitude [7]. This makes it a very challenging system to analyze with LC-MS that has a dynamic range of three to four orders of magnitude in direct (“discovery”) analysis and up to six orders in targeted analysis [12]. However, if progressing AD leaves a strong enough signature in the abundances of hundred most abundant blood proteins (“top proteome”), then for predicting this progression, a basic proteomic experiment may suffice that requires minimum sample preparation and lasts an hour, or perhaps even less. This “top proteome hypothesis” has driven us to dedicate our efforts to developing the label-free “top proteome” workflow, focusing on achieving the best possible precision of the abundance measurements as the critical parameter in reducing the time, and thus the cost, of the proteomic analysis [13]. Indeed, based on around 100 abundant blood proteins, we can successfully discriminate S-MCI with P-MCI with ca. 80% accuracy.

Sex differences in human blood proteome are well known [13, 17, 22, 23]. The level of A2M, the third most abundant protein in plasma, is typically 15%–20% higher in adult females than in adult males [24]. In literature, A2M is often listed as an AD biomarker [14, 15]. The 5′ splice-site deletion in exon 18 of A2M was genetically linked to AD with the

same degree of certainty as the APOE- $\epsilon 4$  allele [25]. Since then, several studies have implicated A2M in AD, although its function in AD pathology still remains unclear. It has also been reported that the level of A2M increases in AD patients compared with healthy controls by 20% (quantified by western blot) [14]. Also, A2M levels have been found to correlate with AD progression [26]. These findings are only partially consistent with our results here. As shown in Fig. 3A, A2M levels are increased in P-MCI females, but decreased (although without statistical significance) in P-MCI males. The picture is even clearer for pooled data, where the same individual samples were pooled according to their sexes and disease progression [13]. A2M level in P-MCI females is higher than in S-MCI females by 13% and in males the corresponding level is lower by 12% (Fig. 2B). The strong dependence upon sex precluded the use of A2M as predictive marker for AD in Table 2. Several other proteins also show sex differences (Supplementary Table 1). Fortunately, it was possible to select proteins that demonstrate more disease specificity than sex specificity. Among all the putative biomarker candidates in Table 2, only fibronectin shows “opposite” behavior in male and female models.

Yet it is important to remember that such a strong phenotype difference as sex inevitably affects the blood proteome [13]. We have previously shown that significant sex differences exist in blood at the level of abundant proteins [13], glycosylation of IgG [27], and isoaspartate content [17]. These results put under question the previous AD biomarker findings made without regard to patient sex. For instance, strong dominance (>75%) of females in the AD cohort due to higher AD prevalence can skew the results of a sex-unified model

[7]. This conclusion is valid not only for diagnostic and prognostic AD biomarkers, but also for blood serum biomarkers of basic activities of daily living in AD patients, among which strong sex differences have recently been reported [27]. Since in the cohort used in the current study females dominate males as 2 : 1, we were extra cautious to avoid sex bias in the predictive model.

Biological roles of the proteins in Table 2 can hint on the mechanism of disease progression. The stark feature of AD progression is the elevated levels in P-MCI of the three fibrinogen chain proteins (alpha, FIBA, beta FIBB, and gamma FIBG), fibronectin (FINC) as well as plasma protease C1 inhibitor (IC1). Moreover, REACTOME pathway overrepresentation analysis has also identified hemostasis as the enriched process in all the P-MCI elevated blood proteins. This feature strongly supports the recently uncovered role of the hemostatic system and the clotting process in AD [29]. Fibrinogen that is normally circulating in blood, deposits in AD together with A $\beta$  in the brain parenchyma and cerebral blood vessels. The interaction of A $\beta$  and fibrin(ogen) leads to increased fibrinogen aggregation, A $\beta$  fibrillization, and the formation in brain of degradation-resistant fibrin clots [29].

In Table 2, decreased levels of immunoglobulin, like Ig gamma-3 chain C region (IGHG3), as well as complement factors (complement factor B (CFAB), complement factor I (CFAI)) in P-MCI patients suggest the involvement of immune system from this group of biomarkers. Among all the 47 negatively correlating proteins, complement cascade was found over-represented. Complement activation is a key component of neuroinflammation in AD, which is potently induced by aggregated A $\beta$  [30]. The complement system may play a neuroprotective role in eliminating aggregated proteins, but an exaggerated/insufficient activation may also be neurodestructive by generating proinflammatory cytokines and oxidative species [31]. Although the role of complement activation as a positive or negative factor in AD is not clear, the involvement of the complement cascade proteins in AD has been documented in multiple independent studies [32], as well as our previous work related to IgG glycosylation [27].

#### *Performance of AD prognostic model*

As mentioned above, the current "gold standard" for discriminating P-MCI versus S-MCI by immunoassay of blood cytokines exhibits the specificity and sen-

sitivity of around 80% [5]. Here, based on a set of abundant proteins analyzed by proteomics, we could achieve a similar accuracy in predicting AD onset (P-MCI). Among all the putative biomarkers proteins in Table 2, complement factor I [33], ceruloplasmin [34], plasma protease C1 inhibitor [26], and fibrinogen [29] have been reported as increased in AD blood samples compared with healthy controls. These reports support the validity of our predictive model.

Regarding the importance of taking sex into account in biomarker discovery, it is clear that the cohorts for mixed-sex discovery have to be more or less sex-balanced. We are also of the opinion that one should also employ sex-specific models to validate sex-unified models. Larger cohorts are likely to clearly demonstrate superior performance of sex-specific models, as would be expected from the theoretical point of view.

#### ACKNOWLEDGMENTS

The proteomics data were acquired within the EU FP7 project PredictAD, which supported their preliminary analysis. The final data analysis work reported here was supported by the Knut and Alice Wallenberg Foundation, VINNOVA Foundation, Alzheimerfonden as well as the Swedish Research council. Carina Palmberg and Marie Stahlberg are gratefully acknowledged for their invaluable contribution in sample preparation and LC/MS experiment running.

Authors' disclosures available online (<http://www.j-alz.com/disclosures/view.php?id=2062>).

#### SUPPLEMENTARY MATERIAL

Supplementary material is available in the electronic version of this article: <http://dx.doi.org/10.3233/JAD-132102>.

#### REFERENCES

- [1] Humpel C (2011) Identifying and validating biomarkers for Alzheimer's disease. *Trends Biotechnol* **29**, 26-32.
- [2] Troncoso JC, Cataldo AM, Nixon RA, Barnett JL, Lee MK, Checler F, Fowler DR, Smialek JE, Crain B, Martin LJ, Kawas CH (1998) Neuropathology of preclinical and clinical late-onset Alzheimer's disease. *Ann Neurol* **43**, 673-676.
- [3] Perrin RJ, Fagan AM, Holtzman DM (2009) Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. *Nature* **461**, 916-922.
- [4] Hyman BT, Phelps CH, Beach TG, Bigio EH, Cairns NJ, Carrillo MC, Dickson DW, Duyckaerts C, Frosch MP, Masliah E, Mirra SS, Nelson PT, Schneider JA, Thal DR, Thies B, Trojanowski JQ, Vinters HV, Montine TJ (2012) National Institute on Aging-Alzheimer's Association guidelines for

- the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement* **8**, 1-13.
- [5] Blennow K (2004) CSF biomarkers for mild cognitive impairment. *J Intern Med* **256**, 224-234.
- [6] Andreasen N, Blennow K (2005) CSF biomarkers for mild cognitive impairment and early Alzheimer's disease. *Clin Neurol Neurosurg* **107**, 165-173.
- [7] Lista S, Faltraco F, Prvulovic D, Hampel H (2013) Blood and plasma-based proteomic biomarker research in Alzheimer's disease. *Prog Neurobiol* **101-102**, 1-17.
- [8] Ray S, Britschgi M, Herbert C, Takeda-Uchimura Y, Boxer A, Blennow K, Friedman LF, Galasko DR, Jutel M, Karydas A, Kaye JA, Leszek J, Miller BL, Minthon L, Quinn JF, Rabinovici GD, Robinson WH, Sabbagh MN, So YT, Sparks DL, Tabaton M, Tinklenberg J, Yesavage JA, Tibshirani R, Wyss-Coray T (2007) Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nat Med* **13**, 1359-1362.
- [9] Marksteiner J, Kemmler G, Weiss EM, Knaus G, Ullrich C, Mechtcheriakov S, Oberbauer H, Auffinger S, Hinterhölzl J, Hinterhuber H, Humpel C (2011) Five out of 16 plasma signaling proteins are enhanced in plasma of patients with mild cognitive impairment and Alzheimer's disease. *Neurobiol Aging* **32**, 539-540.
- [10] Furney SJ, Kronenberg D, Simmons A, Güntert A, Dobson RJ, Proitsi P, Wahlund LO, Kloszewska I, Mecocci P, Soininen H, Tsolaki M, Vellas B, Spenger C, Lovestone S (2011) Combinatorial markers of mild cognitive impairment conversion to Alzheimer's disease - cytokines and MRI measures together predict disease progression. *J Alzheimers Dis* **26**, 395-405.
- [11] Bazenot C, Lovestone S (2012) Plasma biomarkers for Alzheimer's disease: Much needed but tough to find. *Biomark Med* **6**, 441-454.
- [12] Surinova S, Schiess R, Hüttenhain R, Cerciello F, Wollscheid B, Aebersold R (2010) On the development of plasma protein biomarkers. *J Proteome Res* **10**, 5-16.
- [13] Lyutvinskiy Y, Yang H, Rutishauser D, Zubarev RA (2013) In silico instrumental response correction improves precision of label-free proteomics and accuracy of proteomics-based predictive models. *Mol Cell Proteomics* **12**, 2324-2331.
- [14] Hye A, Lynham S, Thambisetty M, Causevic M, Campbell J, Byers HL, Hooper C, Rijdsdijk F, Tabrizi SJ, Banner S, Shaw CE, Foy C, Poppe M, Archer N, Hamilton G, Powell J, Brown RG, Sham P, Ward M, Lovestone S (2006) Proteome-based plasma biomarkers for Alzheimer's disease. *Brain* **129**, 3042-3050.
- [15] Thambisetty M, Lovestone S (2010) Blood-based biomarkers of Alzheimer's disease: Challenging but feasible. *Biomark Med* **4**, 65-79.
- [16] Orešič M, Hyötyläinen T, Herukka SK, Sysi-Aho M, Mattila I, Seppänen-Laakso T, Julkunen V, Gopalacharyulu PV, Hallikainen M, Koikkalainen J, Kivipelto M, Helisalmi S, Lötjönen J, Soininen H (2011) Metabolome in progression to Alzheimer's disease. *Transl Psychiatry* **1**, e57.
- [17] Yang H, Lyutvinskiy Y, Soininen H, Zubarev RA (2011) Alzheimer's disease and mild cognitive impairment are associated with elevated levels of isoaspartyl residues in blood plasma proteins. *J Alzheimers Dis* **27**, 113-118.
- [18] Griss J, Martin M, O'Donovan C, Apweiler R, Hermjakob H, Vizcaino JA (2011) Consequences of the discontinuation of the International Protein Index (IPI) database and its substitution by the UniProtKB "complete proteome" sets. *Proteomics* **11**, 4434-4438.
- [19] Broadhurst DI, Kell DB (2006) Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2**, 171-196.
- [20] Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* **27**, 861-874.
- [21] Haw R, Hermjakob H, D'Eustachio P, Stein L (2011) Reactome pathway analysis to enrich biological discovery in proteomics data sets. *Proteomics* **11**, 3598-3613.
- [22] Miike K, Aoki M, Yamashita R, Takegawa Y, Saya H, Miike T (2010) Proteome profiling reveals gender differences in the composition of human serum. *Proteomics* **10**, 2678-2691.
- [23] Corzett TH, Fodor IK, Choi MW, Walsworth VL, Turteltaub KW, McCutchen-Maloney SL, Chromy BA (2010) Statistical analysis of variation in the human plasma proteome. *J Biomed Biotechnol* **2010**, 258494.
- [24] Ganrot PO, Scherstén B (1967) Serum  $\alpha$ 2-macroglobulin concentration and its variation with age and sex. *Clinica Chimica Acta* **15**, 113-120.
- [25] Blacker D, Wilcox MA, Laird NM, Rodes L, Horvath SM, Go RC, Perry R, Watson B Jr, Bassett SS, McInnis MG, Albert MS, Hyman BT, Tanzi RE (1998) Alpha-2 macroglobulin is genetically associated with Alzheimer disease. *Nat Genet* **19**, 357-360.
- [26] Akuffo EL, Davis JB, Fox SM, Gloger IS, Hosford D, Kinsey EE, Jones NA, Nock CM, Roses AD, Saunders AM, Skehel JM, Smith MA, Cutler P (2008) The discovery and early validation of novel plasma biomarkers in mild-to-moderate Alzheimer's disease patients responding to treatment with rosiglitazone. *Biomarkers* **13**, 618-636.
- [27] Lundström SL, Yang H, Lyutvinskiy Y, Rutishauser D, Herukka SK, Soininen H, Zubarev RA (2013) Blood plasma IgG Fc glycans are significantly altered in Alzheimer's disease and progressive mild cognitive impairment. *J Alzheimers Dis* **38**, 567-579.
- [28] Hall JR, Johnson LA, Barber RC, Vo HT, Winter AS, O'Bryant SE (2012) Biomarkers of basic activities of daily living in Alzheimer's disease. *J Alzheimers Dis* **31**, 429-437.
- [29] Cortes-Canteli M, Zamolodchikov D, Ahn HJ, Strickland S, Norris EH (2012) Fibrinogen and altered hemostasis in Alzheimer's disease. *J Alzheimers Dis* **32**, 599-608.
- [30] Wang Y, Hancock AM, Bradner J, Chung KA, Quinn JF, Peskind ER, Galasko D, Jankovic J, Zabetian CP, Kim HM, Leverenz JB, Montine TJ, Ghingina C, Edwards KL, Snaipinn KW, Goldstein DS, Shi M, Zhang J (2011) Complement 3 and factor h in human cerebrospinal fluid in Parkinson's disease, Alzheimer's disease, and multiple-system atrophy. *Am J Pathol* **178**, 1509-1516.
- [31] Bonifati DM, Kishore U (2007) Role of complement in neurodegeneration and neuroinflammation. *Mol Immunol* **44**, 999-1010.
- [32] Aiyaz M, Lupton MK, Proitsi P, Powell JF, Lovestone S (2012) Complement activation as a biomarker for Alzheimer's disease. *Immunobiology* **217**, 204-215.
- [33] Thambisetty M, Simmons A, Hye A, Campbell J, Westman E, Zhang Y, Wahlund LO, Kinsey A, Causevic M, Killick R, Kloszewska I, Mecocci P, Soininen H, Tsolaki M, Vellas B, Spenger C, Lovestone S (2011) Plasma biomarkers of brain atrophy in Alzheimer's disease. *PLoS One* **6**, e28527.
- [34] Squitti R, Salustri C, Siotto M, Ventriglia M, Vernieri F, Lupoi D, Cassetta E, Rossini PM (2010) Ceruloplasmin/transferrin ratio changes in Alzheimer's disease. *Int J Alzheimers Dis* **2011**, 231595.