



OPEN

## Plasma protein biomarker model for screening Alzheimer disease using multiple reaction monitoring-mass spectrometry

Yeongshin Kim<sup>1,8</sup>, Jaeyeon Kim<sup>1,8</sup>, Minsoo Son<sup>1,8</sup>, Jihyeon Lee<sup>2</sup>, Injoon Yeo<sup>2</sup>,  
Kyu Yeong Choi<sup>3</sup>, Hoowon Kim<sup>3,4</sup>, Byeong C. Kim<sup>5</sup>, Kun Ho Lee<sup>3,6,7</sup>✉ & Youngsoo Kim<sup>1,2</sup>✉

Alzheimer disease (AD) is a leading cause of dementia that has gained prominence in our aging society. Yet, the complexity of diagnosing AD and measuring its invasiveness poses an obstacle. To this end, blood-based biomarkers could mitigate the inconveniences that impede an accurate diagnosis. We developed models to diagnose AD and measure the severity of neurocognitive impairment using blood protein biomarkers. Multiple reaction monitoring-mass spectrometry, a highly selective and sensitive approach for quantifying targeted proteins in samples, was used to analyze blood samples from 4 AD groups: cognitive normal control, asymptomatic AD, prodromal AD, and AD dementia. Multimarker models were developed using 10 protein biomarkers and apolipoprotein E genotypes for amyloid beta and 10 biomarkers with Korean Mini-Mental Status Examination (K-MMSE) score for predicting Alzheimer disease progression. The accuracies for the AD classification model and AD progression monitoring model were 84.9% (95% CI 82.8 to 87.0) and 79.1% (95% CI 77.8 to 80.5), respectively. The models were more accurate in diagnosing AD, compared with single APOE genotypes and the K-MMSE score. Our study demonstrates the possibility of predicting AD with high accuracy by blood biomarker analysis as an alternative method of screening for AD.

Alzheimer disease (AD), characterized by tauopathy, amyloid beta ( $A\beta$ ) plaque formation, and cognitive impairment, is the leading cause of neurodegenerative dementia and the sixth-most frequent cause of death in the US<sup>1,2</sup>. Despite numerous attempts, Food and Drug Administration (FDA)-approved treatments for AD remain undeveloped. Consequently, early diagnosis and monitoring are crucial for improving the prognosis of AD patients<sup>2</sup>.

Diagnosing AD at an early stage is difficult—especially asymptomatic AD, because such patients exhibit a normal cognitive state even after plaque deposition<sup>3</sup>. As damage to neurons accumulates and cognitive impairment worsens, asymptomatic AD progresses to AD with mild cognitive impairment (MCI) (prodromal AD) and AD with dementia. Thus, diagnosing AD in the asymptomatic stage is needed to slow the progression of cognitive impairment<sup>2</sup>.

AD is diagnosed using cognitive assessments and several tests, including genetic tests, biomarker tests, and imaging per clinical criteria<sup>4,5</sup>. In addition, several tests are required to identify the cause of cognitive impairment or dementia, because symptoms can arise from other neurodegenerative diseases. AD biomarkers are also measured to confirm a diagnosis of AD<sup>4,6</sup>. These tests can pose a burden to potential AD patients and are perhaps unsuitable for continuous screening. Among cognitive tests, the Mini-Mental State Examination (MMSE) is widely used<sup>7</sup>, but it has several limitations, such as its vulnerability to illiteracy and association with education level<sup>8</sup>. Other questionnaire-based cognitive tests, such as the Mini-Cog and Montreal Cognitive Assessment

<sup>1</sup>Interdisciplinary Program of Bioengineering, Seoul National University College of Engineering, Seoul, Republic of Korea. <sup>2</sup>Department of Biomedical Engineering, Seoul National University College of Medicine, 28 Yongon-Dong, Chongno-Ku, Seoul 110-799, Republic of Korea. <sup>3</sup>Gwangju Alzheimer's Disease and Related Dementia Cohort Research Center and Department of Biomedical Science, Chosun University, Gwangju 61452, Republic of Korea. <sup>4</sup>Department of Neurology, Chosun University Hospital, Gwangju 61452, Republic of Korea. <sup>5</sup>Department of Neurology, Chonnam National University Medical School, Gwangju 61469, Republic of Korea. <sup>6</sup>Department of Biomedical Science, Chosun University, Gwangju 61452, Republic of Korea. <sup>7</sup>Aging Neuroscience Research Group, Korea Brain Research Institute, Daegu 41062, Republic of Korea. <sup>8</sup>These authors contributed equally: Yeongshin Kim, Jaeyeon Kim and Minsoo Son. ✉email: leekho@chosun.ac.kr; biolab@snu.ac.kr

(MoCA), have different administration times and can lead to nonuniform results<sup>4,7</sup>. Thus, various approaches are encouraged to determine cognitive states that free from such effects.

Cerebrospinal fluid (CSF) is the most widely used type of sample for assessing levels of biomarkers, such as amyloid beta 1–42 (A $\beta$ 42) and tau proteins<sup>9</sup>. However, obtaining CSF is invasive, expensive, and difficult to perform periodically. Because continuous monitoring of its progress is recommended for AD, routine surveillance using blood biomarkers is needed. Because patients with AD have a damaged blood–brain barrier (BBB), their blood can reflect the state of the disease, based on proteins that have passed the disrupted BBB<sup>10,11</sup>.

A $\beta$ 42 and tau proteins are well-known biomarkers of AD that are accurate in diagnosing it<sup>12</sup>. Although these biomarkers have been used in routine clinical procedures, they have several limitations. Because A $\beta$ 42 and phosphorylated tau (p-tau) exist at subnanogram levels in 1 ml of human plasma, antibody enrichment is required to quantify them, introducing variation in testing and increasing the cost<sup>13,14</sup>. Also, quantifying tau proteins is difficult, because they exist as 6 isoforms with varying post-translational modifications<sup>15</sup>. Further, A $\beta$ 42 and t-tau are elevated in other diseases<sup>16,17</sup>, and additional biomarkers are needed to monitor AD owing to the complexity of the causes of AD such as neurodegeneration due to aging, lifestyle, and genetic factors<sup>18</sup>. In this regard, combining biomarkers into a multimarker model would be helpful in diagnosing AD more accurately.

In this study, we analyzed potential biomarkers in plasma from various stages of AD that reflect different severity of neurocognitive stages, and developed multimarker models for monitoring the progression of AD by multiple reaction monitoring-mass spectrometry (MRM-MS). MRM-MS quantifies several target proteins in complex samples simultaneously without antibodies<sup>19,20</sup>, generating results with high reproducibility and consistency between laboratories<sup>21</sup>. Samples from cognitive normal control (CN), asymptomatic Alzheimer disease (AsymAD), prodromal Alzheimer disease (ProdAD), and Alzheimer disease dementia (ADD) patients were analyzed. Our multimarker models, combining proteins with simple AD tests (K-MMSE and apolipoprotein E  $\epsilon$ 4 carrier), showed improved specificity in differentiating AD patients from the control group and evaluating disease progression. Models in this study demonstrate the possibility of diagnosing and monitoring AD more easily with greater discriminatory power.

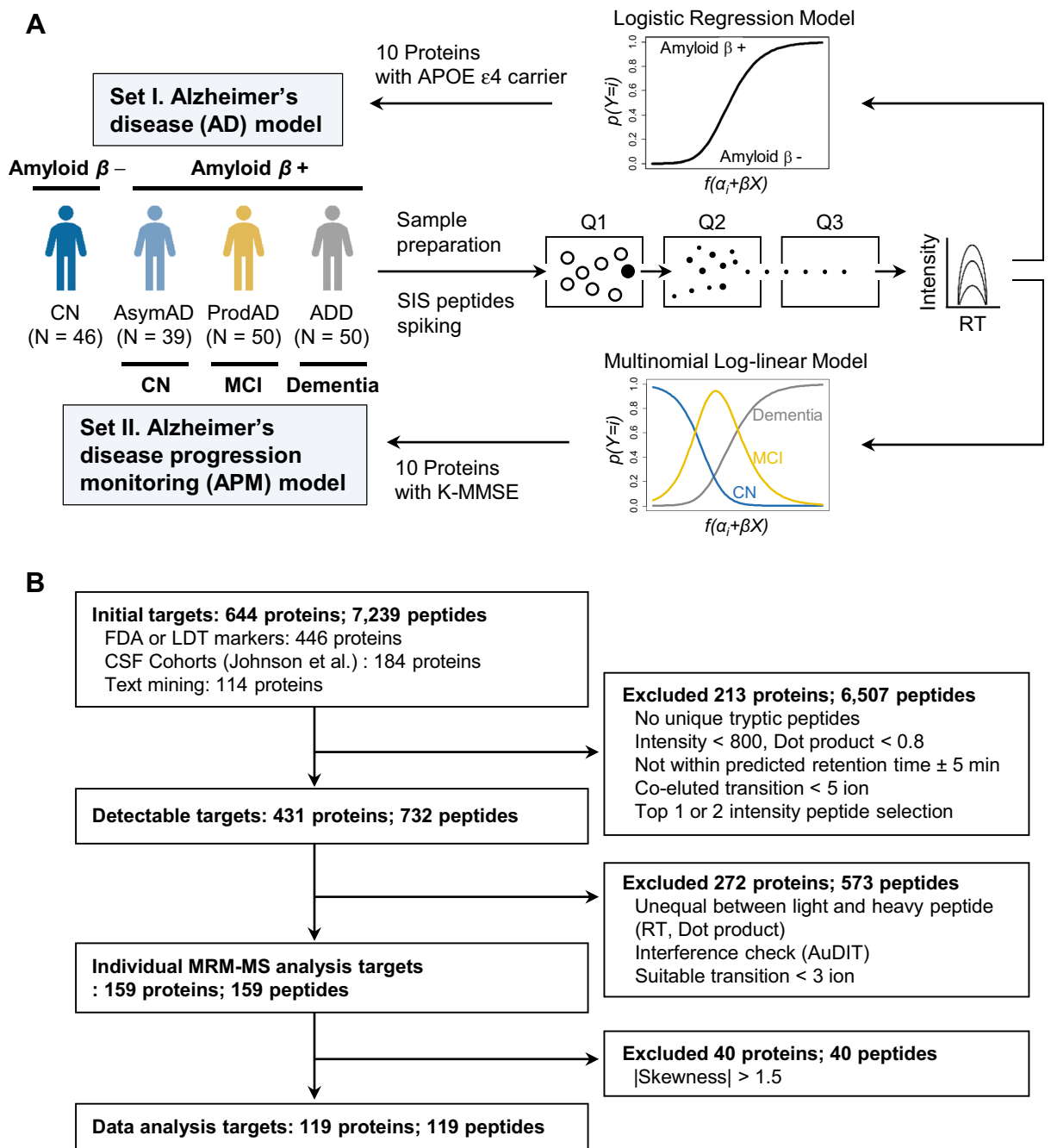
## Results

**Characteristics of the study populations.** A total of 185 samples were selected to measure plasma protein levels for developing a model for classifying AD (Fig. 1A). The patients were categorized into 4 groups, based on a diagnosis of dementia, cognitive impairment status, and amyloid PET: (A) cognitive normal control with negative amyloid beta result (CN; N = 46); (B) asymptomatic AD, defined as cognitively normal but positive for amyloid beta (AsymAD; N = 39); (C) prodromal AD group, which showed mild cognitive impairment and positive amyloid beta results (ProdAD; N = 50); and (D) AD dementia, comprising patients with dementia due to AD with amyloid-positive results (ADD; N = 50).

The demographics and clinical characteristics of the 185 patients are summarized in Table 1. The average age was 71.1 years (95% CI 69.4 to 72.9) for the “CN” group, 73.1 years (95% CI 71.3 to 74.9) for the “AsymAD” group, 74.4 years (95% CI 72.8 to 75.9) in the “ProdAD” group, and 73.1 years (95% CI 71.0 to 75.2) in the “ADD” group. The average education level of the “ADD” group was 7.3 years (95% CI 6.0 to 8.6), which was significantly lower than that in other groups (Bonferroni post hoc test,  $P = 0.028$  for “CN” group,  $P = 5.1 \times 10^{-4}$  for “AsymAD” group, and  $P = 5.3 \times 10^{-4}$  for “ProdAD” group). K-MMSE scores differed significantly between the 4 groups (ANOVA,  $P = 3.4 \times 10^{-27}$ ). The percentage of APOE  $\epsilon$ 4 carriers was 19.6% for the “CN” group, 71.8% for the “AsymAD” group, 62.0% for the “ProdAD” group, and 58.0% in the “ADD” group ( $\chi^2$  test,  $P = 3.0 \times 10^{-6}$ ). A $\beta$  positivity was evaluated using the PET-BAPL scores of 185 subjects. The percentage of A $\beta$  positivity was 100% for the “AsymAD,” “ProdAD,” and “ADD” groups, and there were no A $\beta$ -positive subjects in the “CN” group.

**Target protein selection for quantitation.** Highly quantitative MRM-MS assays for protein biomarkers were developed, as shown in Fig. 1B. To identify proteins with expression patterns that correlated with AD, 644 proteins were selected from the following sources: (1) compiled lists from various sources, including US Food and Drug Administration (FDA)-approved biomarkers, laboratory developed tests (LDT) from the Clinical Laboratory Improvement Amendments (CLIA) database (<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCLIA%20/search.cfm>), Title 21 of the Electronic Code of Federal Regulations (eCFR) (<https://www.ecfr.gov/cgi-bin/ECFR?page=browse>), and Lab Tests Online (<https://labtestsonline.org/tests-index>); (2) targets that were differentially expressed in 2 cerebrospinal fluid AD cohorts from Johnson et al.<sup>22</sup>; and (3) a list of candidates from text mining biomarker studies on AD (Supplementary Table S1).

Targets were considered as detectable in plasma if (1) the peak intensity is greater than 800, (2) the elution profile and ratios of transitions were similar to the spectral library, and (3) peptides were unique tryptic peptides. 431 of 644 proteins were detectable in pooled plasma, and stable isotope labeled-standard (SIS) peptides were synthesized for 732 tryptic peptides coming from 431 detectable proteins. To quantify proteins in plasma samples, we selected the 159 most quantifiable peptides, representing 159 proteins. The best peptide that were interference-free, with highest intensity, and showed equality between endogenous and standard peptides were selected for each protein. 40 of 159 proteins were excluded due to high skewness ( $> 1.5$  or  $< -1.5$ ). Details of the 119 quantitative proteins for quality control are summarized in Supplementary Table S2, Supplementary Fig. S1, and Supplementary Methods. No batch effect for the MRM-MS analysis of 185 samples was observed in the analytical batches (ANOVA,  $P = 0.116$ ), AD groups (ANOVA,  $P = 0.815$ ), or positron emission tomography-brain amyloid plaque load (PET-BAPL) groups (ANOVA,  $P = 0.630$ ), based on their equivalent distribution (Supplementary Fig. S2).

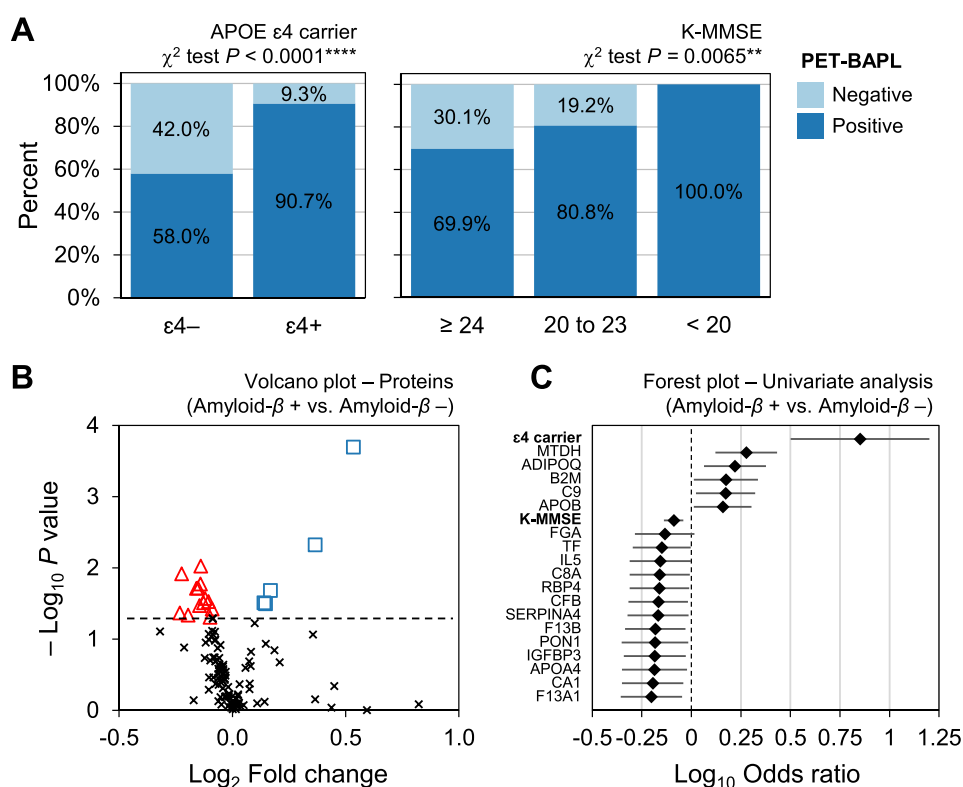


**Figure 1.** Sample and target protein selection. **(a)** Schematic workflow for developing AD classification model using multiple reaction monitoring-mass spectrometry and multinomial log-linear regression. 10 proteins and APOE genotyping results were selected as features for the Alzheimer disease model, and 10 proteins and K-MMSE scores were used for construction of the Alzheimer disease progression monitoring model. **(b)** Target protein selection workflow. A total of 644 proteins were compiled from three types of resources; 431 proteins that were detectable by MRM-MS analysis were selected. Quality of quantitation was assessed by evaluating the interference from SIS peptides. Ultimately, 159 peptides representing 159 proteins were selected for individual sample analysis, of which 119 were selected for data analysis, based on the skewness of the data.

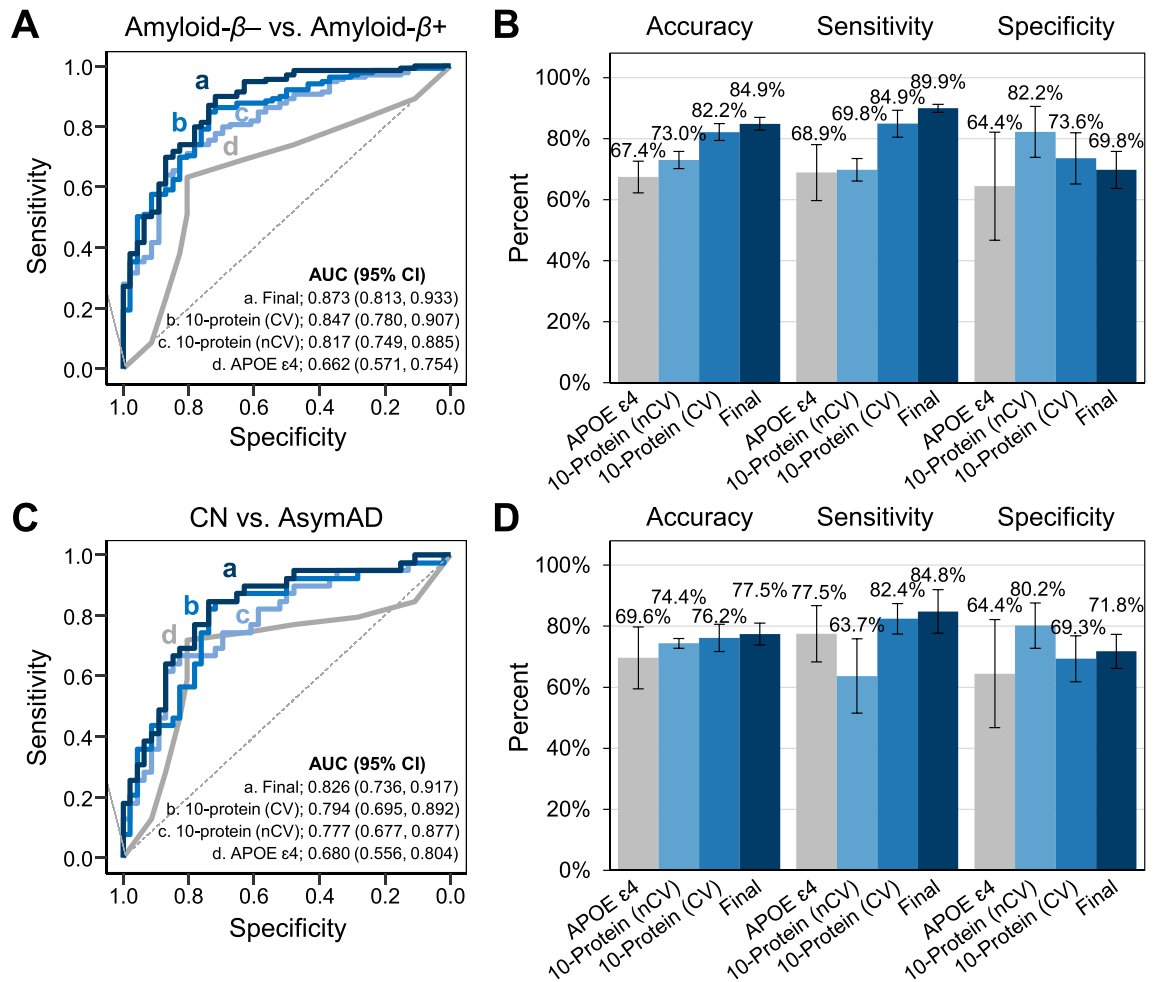
**Development of model for Alzheimer disease classification.** Positron emission tomography-brain amyloid plaque load (PET-BAPL) score, representing the level of deposition of amyloid- $\beta$  ( $A\beta$ ) in patients' brains, correlated well with AD-related factors in our study set (Fig. 2A). Most  $A\beta$ -positive patients were APOE  $\epsilon 4$  carriers ( $\chi^2$  test,  $P < 1.0e-4$ ) and had lower K-MMSE scores ( $\chi^2$  test,  $P = 0.0065$ ). To discover potential AD biomarkers, we compared PET-BAPL-positive ( $A\beta^+$ ) and -negative ( $A\beta^-$ ) groups regarding the expression of 119 proteins, of which 18 showed significant differential expression between groups (Fig. 2B, Supplementary Fig. S3, and Supplementary Table S3).

Descriptions		CN, N = 46		AsymAD, N = 39		ProdAD, N = 50		ADD, N = 50		P value
		Number of patients (Percent)		Number of patients (Percent)		Number of patients (Percent)		Number of patients (Percent)		
Gender	Female	25	54.3%	21	53.8%	16	32.0%	27	54.0%	0.068
	Male	21	45.7%	18	46.2%	34	68.0%	23	46.0%	
		Mean (95% CI)								
Age, yrs		71.1	(69.4–72.9)	73.1	(71.3–74.9)	74.4	(72.8–75.9)	73.1	(71.0–75.2)	0.091
Education years, yrs		10.1	(8.6–11.5)	11.4	(9.9–12.8)	11.1	(9.7–12.5)	7.3	(6.0–8.6)	<0.0001****
K-MMSE <sup>1</sup>		26.8	(26.2–27.5)	27.6	(27.0–28.2)	25.3	(24.3–26.2)	18.0	(16.3–19.8)	<0.0001****
		Number of patients (Percent)								
PET-BAPL	1	46	100.0%	0	0.0%	0	0.0%	0	0.0%	<0.0001****
	2	0	0.0%	18	46.2%	13	26.0%	7	14.0%	
	3	0	0.0%	21	53.8%	37	74.0%	43	86.0%	
APOE genotype	$\epsilon 2/\epsilon 2$	2	4.3%	0	0.0%	0	0.0%	0	0.0%	0.0003****
	$\epsilon 2/\epsilon 3$	3	6.5%	0	0.0%	2	4.0%	0	0.0%	
	$\epsilon 3/\epsilon 3$	32	69.6%	11	28.2%	17	34.0%	21	42.0%	
	$\epsilon 2/\epsilon 4$	1	2.2%	1	2.6%	0	0.0%	3	6.0%	
	$\epsilon 3/\epsilon 4$	7	15.2%	25	64.1%	26	52.0%	24	48.0%	
	$\epsilon 4/\epsilon 4$	1	2.2%	2	5.1%	5	10.0%	2	4.0%	

**Table 1.** Demographics of the three Alzheimer disease groups. <sup>1</sup>Higher score represents better cognitive function; Significance level is represented as “\*\*\*” for  $P < 0.001$  and “\*\*\*\*” for  $P < 0.0001$ ; Cognitive normal, CN; Asymptomatic Alzheimer disease, AsymAD; Prodromal Alzheimer disease, ProdAD; Alzheimer disease dementia, ADD.



**Figure 2.** Identification of differentially expressed proteins between PET-BAPL groups. (a) PET-BAPL score proportion in APOE  $\epsilon 4$  carriers and the K-MMSE group, respectively. (b) Volcano plot of differentially expressed proteins between PET-BAPL positive and negative groups. Eighteen proteins, with  $t$ -test  $P$  values lower than 0.05, are highlighted in red or blue. (c) Forest plot of univariate logistic regression analysis of 18 proteins, APOE genotyping results, and K-MMSE scores.



**Figure 3.** Protein model combined with clinical value for predicting amyloid beta-positive and -negative groups. (a) ROC curves for 10 proteins performed with nested crossvalidation, 10 proteins performed with fivefold crossvalidation, APOE  $\epsilon_4$  carriers, and final AD model (all combined) predicting amyloid  $\beta$ -positive and -negative groups. Area under the curve (AUC) and 95% CI of AUC values for each model are shown. (b) Sensitivity and specificity plots for the 4 models predicting amyloid  $\beta$ -positive and -negative groups. Error bars represent 95% confidence interval. (c) ROC curves of 10 proteins performed with nested crossvalidation, 10 proteins performed with fivefold crossvalidation, APOE  $\epsilon_4$  carriers, and final AD model (all combined) predicting “CN” and “AsymAD” groups. AUC value and 95% CI of AUC values for each model are shown. (d) Sensitivity and specificity plots for the 4 models predicting “CN” and “AsymAD” groups. Error bars represent 95% confidence interval.

Univariate analysis was performed for the 18 proteins, APOE genotypes, and K-MMSE scores (Fig. 2C). Five proteins (MTDH, ADIPOQ, B2M, C9, and APOB) correlated positively with  $A\beta$  positivity, and 13 (FGA, TF, IL5, C8A, RBP4, CFB, SERPINA4, F13B, PON1, IGFBP3, APOA4, CA1, and F13A1) had a negative association. APOE genotype had a stronger association with the  $A\beta$  PET results than other protein biomarkers. In contrast, K-MMSE score contributed less to the classification of AD than the protein biomarkers. Eighteen proteins were used to develop a logistic regression model for classifying the  $A\beta$ -negative (CN;  $N = 46$ ) and  $A\beta$ -positive AD groups (AsymAD, ProdAD, and ADD;  $N = 139$ ).

Applying a recursive feature elimination (RFE) strategy using nested crossvalidation (nCV), 10 of 18 proteins (MTDH, ADIPOQ, APOB, TF, CA1, C9, APOA4, RBP4, F13A1, and FGA) were selected to develop an 10-protein model. Nested CV was performed with 5 outer folds to test the performance of the model, with 5 inner folds from each outer fold for feature selection and model validation in the training model. To validate the performance of the final model, termed the Alzheimer disease (AD) model, and demonstrate improvement in the performance of the multimarker panel compared with the single test, we compared the 10-protein model with the APOE  $\epsilon_4$  carrier model (Fig. 3A), which had AUC values of 0.817 (95% CI 0.749 to 0.855) and 0.662 (95% CI 0.651 to 0.754), respectively. The balanced accuracy values of the 10-protein and APOE  $\epsilon_4$  carrier models were 76.0% (95% CI 88.9 to 83.1) and 66.7% (95% CI 58.9 to 74.5), respectively. The final AD model, combining the 10-protein and APOE  $\epsilon_4$  carrier models, had an AUC value of 0.873 (95% CI 0.813 to 0.933) and a balanced accuracy value of 79.9% (95% CI 73.2 to 86.6) and improved its sensitivity to 89.9% (95% CI 88.6 to 91.3), higher than that of the other models (Fig. 3B).



We also tested the APOE  $\epsilon 4$  carrier, 10-protein, and final AD models to classify the “AsymAD” group among normal cognition groups (Fig. 3C). The 10-protein and APOE  $\epsilon 4$  carrier models had AUC values of 0.777 (95% CI 0.677 to 0.877) and 0.680 (95% CI 0.556 to 0.804), respectively; their balanced accuracy values were 71.9% (95% CI 64.4 to 79.4) and 71.0% (95% CI 63.5 to 78.5). The final AD model performed best, based on its AUC of 0.826 (95% CI 0.736 to 0.917), balanced accuracy of 78.3% (95% CI 71.4 to 85.2), and sensitivity of 84.8% (95% CI 77.7 to 92.0) (Fig. 3D). The overall characteristics of the 3 models are summarized in Supplementary Table S5, and the coefficients of the final AD model are listed in Supplementary Table S6. MTDH and FGA contributed most versus the other candidate markers, based the significance of the coefficients (Supplementary Table S6).

**Multiclass model for predicting Alzheimer disease progression.** To develop an AD progression monitoring (APM) model for testing the severity of functional degeneration, we identified 32 differentially expressed proteins (DEPs) between 3 representative groups: the “AsymAD” group for cognitively normal AD, the “ProdAD” group with mild cognitive impairment, and the “ADD” group, representing dementia (student’s *t* test,  $P < 0.05$ ) (Fig. 4A and Supplementary Table S7). Of the 19 DEPs between the “AsymAD” and “ProdAD” groups, 14 were upregulated (HBA1, MTDH, PFN1, DSG3, CFHR3, F13A1, FGA, CST3, F13B, C7, AZGP1, APOB, FGB, and FGG) and 5 (FN1, TF, CFB, COMP, and VTN) declined in the “ProdAD” group (Fig. 4B). Of the 15 DEPs between the “ProdAD” and “ADD” groups, 8 were elevated (CRP, HP, FN1, LAMP2, B2M, ORM1, CALR, and CFI) and 7 were lower (DES, UMOD, APOA4, RBP4, CST3, HRG, and SELL) in the “ADD” group.

Nested CV was performed for the multinomial log-linear regression model with 5 outer folds and 5 inner folds for each outer fold. With RFE selection in the nested CV, the 10-protein model (CALR, F13A1, LAMP2, APOA4, B2M, FN1, FGA, ORM1, DES and MTDH) was deemed to have the best accuracy versus other models with differing numbers of variables (Fig. 4C). Notably, 4 of 10 proteins (APOA4, FGA, F13A1, and MTDH) were also included in the final AD model for classifying  $A\beta$ -positive AD groups. In the comparison of AsymAD and ProdAD, FN1, FGA, and F13A1 contributed more, based on the significance of the coefficients. In addition, APOA4, B2M, and ORM1 contributed more when differentiating between AsymAD and ADD.

Next, we compared the performance of the 10-protein model with other models that used conventional AD factors (APOE  $\epsilon 4$  carrier and K-MMSE) (Fig. 4C and Supplementary Fig. S5). The 10-protein model was more accurate [72.7% (95% CI 70.4% to 75.0%)] than the APOE  $\epsilon 4$  carrier [31.6% (95% CI 28.0% to 35.2%)] and K-MMSE models [60.4% (95% CI 59.3% to 61.6%)]. Further, the final APM model, combining 10 protein variables with K-MMSE, had the highest accuracy [79.1% (95% CI 77.8% to 80.5%)]. It also performed better than other models with conventional factors (Fig. 4D and Supplementary Table S9). The APOE  $\epsilon 4$  carrier model failed to estimate neurocognitive impairments accurately, based on its zero specificity in comparing “AsymAD” with the “ProdAD” plus “ADD” groups and the “ProdAD” group alone, despite a sensitivity of 100%.

Overall, in the comparisons between AD groups, although K-MMSE had balanced sensitivity and specificity in its classification, the final APM model (10-protein with K-MMSE) had diagnostic power with sensitivity and specificity that exceeded 80%. The coefficients of the final APM model are listed in Supplementary Table S10, and box plots of its 11 features (10 proteins and K-MMSE score) are shown in Supplementary Fig. S5.

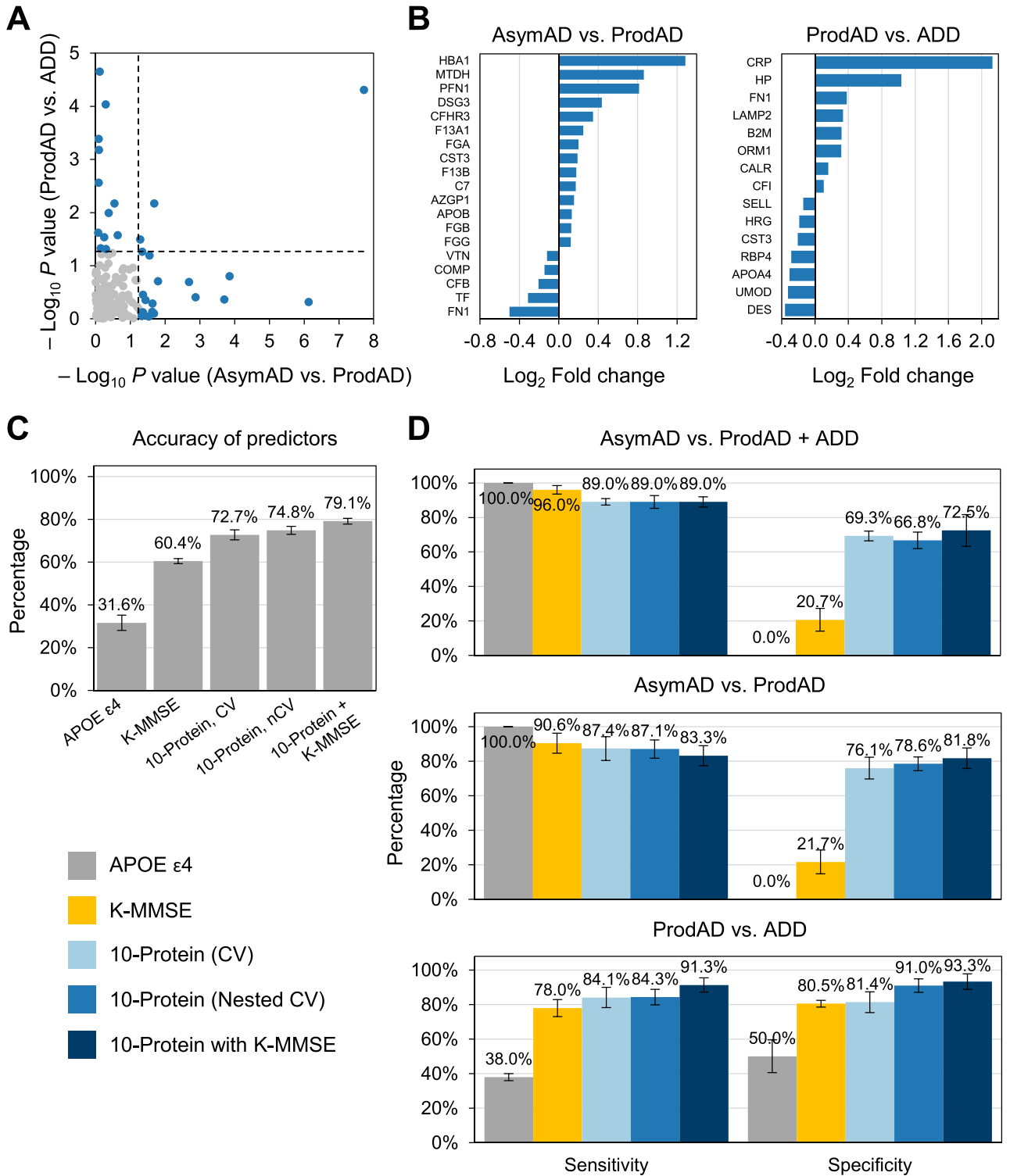
**Association of plasma proteins with CSF proteins.** To examine whether the proteins in the models were derived from the brain or cerebrospinal fluid (CSF) of AD patients, we compared the proteins in the 2 models with those that Johnson et al.<sup>22</sup> identified in their brain and CSF cohorts (Fig. 5A). Nine of 16 proteins (56.3%) overlapped with the module proteins in the brain cohorts (6 proteins; B2M, CA1, F13A1, FGA, ORM1, and TF) or the DEPs in the CSF cohorts (8 proteins; APOA4, B2M, CA1, C9, CALR, FGA, ORM1, and TF). Notably, all 8 CSF proteins correlated significantly with amyloid- $\beta$  or tau levels in the CSF (Supplementary Fig. S6).

The CSF levels of 16 proteins were compared with their plasma levels in 32 CSF samples, matched with plasma samples in our study. Six of 16 proteins (37.5%) correlated positively between plasma and CSF levels (Fig. 5B; Pearson’s correlation,  $P < 0.05$ ). Moreover, APOA4, MTDH, ORM1, and TF differed significantly in the CN and 3 AD groups (Fig. 5C; ANOVA,  $P < 0.05$ ).

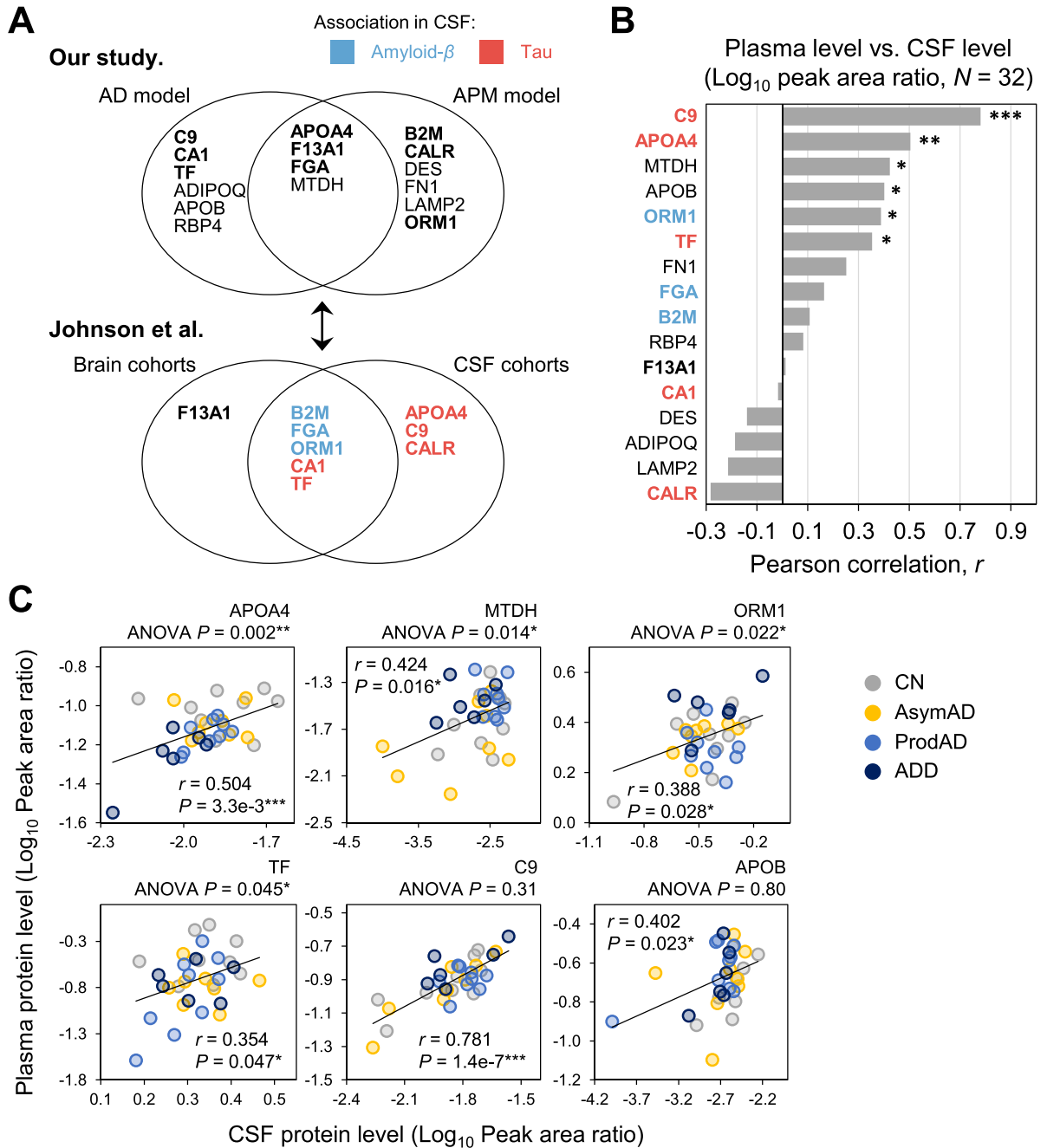
## Discussion

In this study, potential protein biomarkers for AD were analyzed by MRM-MS. MRM-MS can accurately quantify low-abundance AD biomarkers in blood from early-stage AD patients with superior sensitivity. The high reproducibility, accuracy, and sensitivity of MRM-MS render it suitable for the early diagnosis of AD across institutions<sup>23</sup>. The diagnostic performance of the biomarkers was estimated, and a multimarker model was developed by combining biomarkers to improve accuracy of conventional biomarkers, compared with APOE genotyping and MMSE, which need additional tests for accurately diagnosing AD. Amyloid beta plaque deposition is tested primarily using amyloid PET imaging or CSF assay. Although these tests yield valid results on  $A\beta$  deposition status and although their use in measuring biomarkers is mentioned in diagnostic criteria for AD<sup>4</sup>, an  $A\beta$ -positive person by PET might need additional tests to verify whether the amyloidosis actually originates from AD or his advanced age<sup>24</sup>. In addition, lumbar puncture, which is needed to obtain CSF samples from patients, is relatively invasive. Thus, a blood biomarker test would be useful in screening for AD if it can be performed prior to PET imaging or CSF assay.

APOE genotyping and MMSE are also used for diagnosing AD and are typically less strenuous for patients. However, according to a meta-analysis, over 40% of non-APOE  $\epsilon 4$  carriers are at risk of developing AD<sup>25</sup>. MMSE can give varying results, depending on the education level of the subjects or the recall words that examiners choose<sup>8</sup>. In effect, both APOE  $\epsilon 4$  carriers and the K-MMSE model classified AD less accurately in our study—for the  $A\beta$ -positive versus  $A\beta$ -negative groups and the neurocognitive impairment groups.



**Figure 4.** Construction of model for classifying AsymAD, ProdAD, and ADD. (a) Scatterplot of  $P$  values by student's  $t$  test between "AsymAD" versus "ProdAD" groups and "ProdAD" versus "ADD" groups. Proteins with  $P$  values lower than 0.05 are marked as blue dots. (b) Fold-change of DEPs in "AsymAD" versus "ProdAD" groups and "ProdAD" versus "ADD" groups. (c) Accuracy of APOE  $\epsilon 4$  carriers, K-MMSE, 10 proteins performed with nested crossvalidation, 10 proteins performed with fivefold crossvalidation, and 10-protein with K-MMSE models. (d) Bar plots for sensitivity and specificity of APOE  $\epsilon 4$  carriers, K-MMSE, and final model (10-protein with K-MMSE) in model prediction between "AsymAD" versus "ProdAD" plus "ADD" group, "AsymAD" versus "ProdAD" group, and "ProdAD" versus "ADD" group. Error bars represent 95% confidence interval.



**Figure 5.** Association of plasma proteins with CSF proteins. (a) Correlation between Johnson et al. and our study. Between the AD and APM models, APOA4, F13A, FGA, and MTDH were selected by both as features, whereas C9, CA1, TF, APOB, RBP4, and ADIPOQ were unique to the AD model and CALR, B2M, LAMP2, DES, FN1, and ORM1 were unique to the APM model. Among 16 proteins, 9 were mentioned in Johnson et al., in which B2M, FGA, and ORM1 were associated with amyloid and APOA4, C9, CALR, CA1, and TF were associated with tau. (b) Target correlation between CSF and plasma. Correlation of the 16 proteins among the 32 that matched in both CSF and plasma. Seven targets had Pearson’s correlation values above 0.3. (c) Scatter plot of 6 target proteins among the CSF and plasma samples.

Further, the final AD and APM models, combining protein biomarkers with conventional AD factors, made much better predictions. These 2 models could complement conventional diagnostic tests and help identify people with AD with high precision while reducing expenses and simplifying the process. Our models could increase the accuracy of predicting AD and reduce potential hazards, lower medical costs, and increase the quality of life for AD patients<sup>26</sup>. In particular, the APM model might also serve as a surrogate biomarker by determining whether neurodegeneration in AD patients improves through medications in clinical trials. Based on recent data that support treatments for mild and moderate AD, AD patients, especially those with MCI, might benefit from our model by identifying their ailment at an earlier stage and preparing for treatment accordingly<sup>26</sup>. To insure integrity of the method, 16 proteins included in the model were analyzed for limit of detection (LOD),



limit of quantification (LOQ), and stability according to CPTAC guideline (Supplementary Methods). Most of the targets passed the evaluation, whereas CALR and DES failed once under the 24-h condition, DES failed once under the freeze and/ thaw condition, and ADIPOQ failed once in under the 2-time freeze and/ thaw condition, with variability under 30% (Supplementary Table S12, S13). For future validation assays, precision of the analysis using crude SIS peptides can be compared with results of analysis with highly purified peptides in further studies.

Damaged neurons and synapses are a characteristic of AD patient brains. The toxicity of A $\beta$  causes oxidative damage and synaptic alterations, and synaptic dysfunction is one of the main factors that effect cognitive impairments in AD patients<sup>27</sup>. In our study, beta-2-macroglobulin (B2M), which is associated with neuronal function, such as neurogenesis, and neuronal actin expression, was upregulated with the progression of AD<sup>28,29</sup>. Carbonic anhydrase 1 (CA1) is related to the preservation of pH levels in the brain, protecting CSF and neuronal function, and its levels decrease as more A $\beta$  plaque is deposited<sup>30</sup>. Fibrinogen alpha chain (FGA) interacts with A $\beta$  and can be deposited, leading to plaque-induced inflammation<sup>31</sup>. In addition, A $\beta$  is a cause of inflammation, which damages neurons and aggravates the pathogenesis of AD<sup>32</sup>. Coagulation factor XIII A chain (F13A1) is detected in reactive microglia, which is related to neuroinflammation, and F13A1 is involved in transglutaminase-mediated polymerization of amyloid beta proteins<sup>33</sup>. Alpha-1-acid glycoprotein 1 (ORM1) was enriched in pathways that are associated with inflammation, such as IL-6 signaling, in a profiling study on plasma and brain from AD patients<sup>34</sup>. Transferrin (TF) regulates iron levels by binding to iron and controlling oxygen radical formation<sup>35</sup>.

Further, associations between AD and other proteins in our models have been reported in AD studies on inflammation (C9,CALR)<sup>36</sup>, A $\beta$  plaque formation (APOB, FN1, APOA4, LAMP2)<sup>37–39</sup>, and neuronal and brain damage, such as glutamate accumulation in the synapses (MTDH)<sup>40</sup> and synaptic plasticity (ADIPOQ)<sup>41</sup>. C9 is related to complement activation and inflammation. APOB and APOA4 interact with amyloid beta and influence A $\beta$  plaque formation but with opposing effects. APOA4 contributes to the clearance of amyloid beta by binding to amyloid beta and inducing glial cells to take up amyloid beta. APOB also binds to amyloid beta peptide, but its mechanism opposed that of APOA4. APOB has been suggested to have high amyloid-forming potential, and APOB can form amyloid-like structures<sup>42</sup>. FN1 is involved in clotting, which can affect amyloid beta fibrilization<sup>39</sup>. LAMP2 is associated with lysosomal function, and lysosomal dysfunction leads to ROS generation and deposition of wastes that can be neurodegenerative<sup>43</sup>. CALR, which is the receptor for C1q, mediates ROS production by binding to C1q and triggering signaling pathways that are related to ROS<sup>44</sup>. By combining biomarker candidates that represent various pathologies of AD, our AD and APM models can classify the stage of disease with enhanced specificity for AD.

Among the 16 proteins that were used in the final 2 models, 6 (APOA4, B2M, CALR, F13A1, FGA, and TF) were related to AD in the brain and CSF in a cohort study by Johnson et al.<sup>22</sup>; 4 proteins (B2M, FGA, F13A1, and TF) belonged to modules that were related to blood, myelin or oligodendrocyte, and microglia. By Gene Ontology analysis using all proteins in these modules, enriched terms were associated with major AD pathologies (Supplementary Table S11)<sup>22</sup>. Also, CSF levels of 6 proteins (APOA4, B2M, CALR, CST3, FGA, and TF) correlated with AD-related factors, such as cognitive score (MoCA) and CSF markers (A $\beta$ , total tau, and phospho-tau) (Supplementary Fig. S6). Among the 6 proteins, APOA4 and TF were significantly linked to total tau (t-tau), phospho-tau, and MoCA score, except for CALR linked to t-tau and phospho-tau, but not to MoCA score. Conversely, B2M, CST3, and FGA correlated with A $\beta$  but not MoCA score. Changes in A $\beta$  and tau were involved in a process of preclinical AD in a longitudinal study, and exacerbation of tauopathy during disease progression has been suggested to have a more direct effect on cognitive decline by neurodegeneration due to a damaged cytoskeleton<sup>1,45,46</sup>. In addition, A $\beta$  indirectly affects cognition in a tau-mediated manner, wherein changes in A $\beta$  are followed by those in tau, and neurotoxic A $\beta$  plaques cause neurodegeneration by affecting synaptic function<sup>2,45</sup>. Our APM model, constructed with A $\beta$ - or tau-related protein biomarkers, might be a useful tool for monitoring neurodegeneration in AD patients.

There are several limitations of our study. The samples were obtained from ethnic Koreans, which might limit the applicability of the results to the general populace. Thus, prior to clinical use, the model must be tested in other countries and populations of different ethnicities and etiologies to validate its diagnostic power. Also, the model was developed with limited sample size. Small numbers of samples were included in each group; thus, the models were nested cross-validated to avoid overfitting. To check biasness, comparison between K-fold cross validation and nested cross validation were performed. Nevertheless, the model must be validated with a larger number of samples. MCI patients who were negative for amyloid beta were not included in our study. Because tests for filtering patients with impaired cognitive states due to other diseases are suggested in diagnostic guidelines for AD, further studies with such patients are needed to increase the accuracy of the models. Despite these limitations, the improved diagnostic power of our models demonstrate the possibility of diagnosing AD at an early stage by routine blood test. Our multimarker models thus reduce the burden of tests for diagnosing AD and help discriminate AD patients at an early stage, when they need continuous clinical care to prevent disease progression.

## Methods

**Materials.** Rapigest™ SF was purchased from Waters (Milford, MA, USA). Neat formic acid (FA), dithiothreitol (DTT), iodoacetamide (IAA), ammonium bicarbonate (ABC), and bichinchoninic acid (BCA) solution were obtained from Sigma-Aldrich (St. Louis, MO, USA). Sequencing-grade modified trypsin was acquired from Promega (Madison, WI, USA). High-performance liquid chromatography (HPLC)-grade water, 0.1% FA in HPLC water, 0.1% FA in acetonitrile (ACN), and methanol were purchased from Thermo Fisher Scientific (Bremen, Germany).

**Study participants and sample collection.** A total of 261 patients were prospectively recruited between 2014 and 2019 from the Gwangju Alzheimer's & Related Dementia (GARD) cohort in Gwangju, Korea. All subjects were examined by experienced neurologists and underwent a full dementia screen, which comprised the Seoul Neuropsychological Screening Battery (SNSB) and Clinical Dementia Rating<sup>47,48</sup>. The SNSB is an overall neuropsychological test for 5 domains of cognition and includes the Korean version of the Mini-Mental State Examination (K-MMSE)<sup>49</sup>. Cognitive impairment was defined as a Z-score lower than -1.5 of the standard deviation (normalized for age, sex, and education) on at least 1 of the neuropsychological tests (memory, language, visuospatial function, attention, or frontal/executive). The clinical diagnosis of dementia and mild cognitive impairment (MCI) was performed per the National Institute on Aging and Alzheimer's Association (NIA-AA) and International Working Group 2 (IWG-2) workgroups<sup>4,6,50</sup>. The subjects underwent MRI and Florbetaben positron emission tomography (PET) scans to image the brain. Florbetaben PET images for scoring the brain beta-amyloid plaque load (BAPL) were acquired as detailed<sup>51,52</sup>. PET images were acquired 90 min after intravenous injection of 300 MBq <sup>18</sup>F-Florbetaben using a Discovery ST PET-CT scanner (General Electric Medical Systems, Milwaukee, WI, USA). BAPL scores were measured visually on the PET images: 1 indicated a beta-amyloid-negative PET scan, and 2 and 3 were considered beta-amyloid-positive. Subjects with BAPL scores 2 or 3 were classified as patients having AD.

For genotyping apolipoprotein E (APOE), whole blood was collected in ethylenediaminetetraacetic acid (EDTA) tubes. Peripheral blood leukocytes were isolated from blood, and genomic DNA was extracted from leukocytes. APOE was genotyped using a genome-wide genotyping array (Affymetrix Axion<sup>®</sup> KORV1.0, Santa Clara, CA, USA) at DNALink (Seoul, South Korea). The Center for Genome Science, Korea National Institute of Health, Republic of Korea (4845-301, 3000-3031) designed and optimized the genotyping array<sup>53</sup>, which included 2 APOE-related small nucleotide polymorphisms (rs7412, rs429358).

A total of 76 of 261 subjects were excluded from the study for the following reasons: dementia or MCI that was unrelated to AD, refusal to participate, failure to be administered amyloid PET tracer, and altered status on re-examination. A total of 185 plasma samples were selected. All 185 subjects were classified into the four groups (46 CN subjects, 39 AsymAD patients, 50 ProdAD patients, and 50 ADD patients) according to A $\beta$  positivity by BAPL scores and clinical criteria by NIA-AA and IWG-2 workgroups.

**Ethical approval and consent to participate.** The study was approved by the institutional review board (IRB) of Chosun University Hospital, Korea (IRB approval No. CHOSUN 2013-12-018-070). All research methods were conducted in accordance with the Declaration of Helsinki and relevant guidelines and regulations. Written informed consent was provided by all volunteers and family members or authorized caregivers in the case of cognitively impaired patients.

**Sample preparation.** A total of 185 plasma samples were prepared in block-randomized batches with respect to age, gender, and group, using the psych package (Version 1.9.12) in R (Version 4.0.0). Six high-abundance proteins (albumin, immunoglobulin G, immunoglobulin A, haptoglobin, transferrin, and alpha-1-antitrypsin) were depleted on a high-performance liquid chromatography (HPLC) instrument that was coupled to a Multiple Affinity Removal System Human-6 (MARS Hu-6, 4.6 mm  $\times$  100 mm, Agilent, CA, USA) (Supplementary Methods). Depleted plasma samples were then concentrated at 3470 g for 6 h at 4  $^{\circ}$ C using 3000-Da molecular weight cutoff (MWCO) centrifugal filter units (Amicon Ultra-4 3 K, Millipore, MA, USA). After concentration, the protein in the plasma samples was measured using the Pierce<sup>™</sup> BCA Protein Assay Kit (Thermo Scientific, Rockford, IL, USA). Then, 20  $\mu$ L of solution that contained 0.2% RapiGest, 20 mM dithiothreitol (Merck, Darmstadt, Germany), and 100 mM ABC, pH 8.0 was added to 100  $\mu$ g of protein and incubated at 60  $^{\circ}$ C for 60 min with shaking for denaturation and reduction. Next, 10  $\mu$ L 100 mM iodoacetamide (Sigma, MO, USA) was added and incubated at room temperature for 30 min with shaking in the dark to alkylate the proteins. Subsequently, 40  $\mu$ L 0.1  $\mu$ g/ $\mu$ L trypsin (sequencing-grade, Promega, WI, USA) in 50 mM ABC solution, pH 8.0 was added and incubated at 37  $^{\circ}$ C for 4 h with shaking to digest the proteins. The digestion was quenched with 10  $\mu$ L 10% formic acid at 37  $^{\circ}$ C for 30 min with shaking.

Quenched samples were centrifuged at 16,602 g and 4  $^{\circ}$ C for 60 min, and 90  $\mu$ L of the supernatant was transferred to a clean tube to remove any byproducts of the RapiGest. Digested samples were cleaned using Oasis<sup>®</sup> HLB (hydrophilic-lipophilic balance) 1 cc (30 mg) extraction cartridges (Waters Corp., MA, USA). Columns were activated with 1 mL 100% methanol and equilibrated with 3 mL 100% acetonitrile and 5 mL 0.1% formic acid in distilled water. Each digested plasma samples was loaded onto the column twice and washed with 3 mL 0.1% formic acid in distilled water. Samples were eluted with 0.5 mL 0.1% formic acid in 40% acetonitrile and 0.5 mL 0.1% formic acid in 60% acetonitrile. Eluates were lyophilized on a vacuum centrifuge and resolubilized to 0.25  $\mu$ g/ $\mu$ L in 0.1% formic acid/water.

**Liquid chromatography-tandem mass spectrometry.** The plasma samples were analyzed on an Agilent 6490 triple quadrupole (QQQ) mass spectrometer (Agilent, CA, USA) with a Jetstream electrospray source, coupled to a 1260 Infinity HPLC system (Agilent, CA, USA). Buffers that consisted of 0.1% formic acid/water (v/v) and 0.1% formic acid/acetonitrile (v/v) were used as mobile phases A and B during the mass analysis. A reversed-phase analytical column (150 mm  $\times$  0.5 mm id, Agilent Zorbax SB-C18, 3.5- $\mu$ m particle size) was used with mobile phases A and B to separate the peptides. In every analysis between batches, LC quality control and a suitability test were performed using 6  $\times$  5 LC-MS/MS Peptide Reference Mix from Promega (Madison, WI, USA).

The multiple reaction monitoring-mass spectrometry (MRM-MS) analyses were performed in positive mode. A total of 10  $\mu$ L of sample was injected to the mass spectrometer for each analysis. The total run time of LC was

70 min, and the specific LC buffer system was as follows: The flow rate of the mobile phases was 40  $\mu\text{L}/\text{min}$ . The LC gradient started at 97% mobile phase A and 3% mobile phase B. The gradient of mobile phase B was increased linearly from 3 to 40% and flowed for 52 min for separation of elution of the peptides. The buffer with 60% mobile phase B flowed for 3 min, and the column was equilibrated for the next run for 15 min with the buffer with 3% mobile phase B. The ion spray capillary voltage was 2500 V, and the nozzle voltage was set to 2000 V. The cell accelerator voltage was 5 V, the delta electron multiplier voltage (EMV) was adjusted to 200 V, and the fragment voltage was set to 380 V. The temperature of the drying gas was set to 250  $^{\circ}\text{C}$  at 15 L/min, and the sheath gas was adjusted to 350  $^{\circ}\text{C}$  at a flow rate of 12 L/min.

**Proteomic data processing and statistical analysis.** Raw MRM-MS data were processed in Skyline (MacCoss Lab, University of Washington, Seattle, WA) to compute the peak areas of the transitions. The peak areas of endogenous peptides were normalized to those of their corresponding stable isotope-labeled standard (SIS) peptides to compare the relative abundance of candidate peptides between samples. SIS peptides were synthesized with lysine or arginine that were heavy isotope-labeled ( $^{13}\text{C}_6^{15}\text{N}_2$  or  $^{13}\text{C}_6^{15}\text{N}_4$ ) at the C-termini. Targets with interference signals were excluded, per the Automated Detection of Inaccurate and Imprecise Transitions (AuDIT) algorithm<sup>54</sup>. The  $\log_{10}$ -transformed, protein-wise centered and scaled peak area ratios were used for data processing. Interference-free check and quality control for individual analysis are detailed in Supplementary Methods.

Differences between patient groups were determined by student's *t* test for bivariate and analysis of variance (ANOVA) for multiple groups using SPSS (versions 23.0 and 25.0; IBM, Chicago, IL), and unadjusted *P* values were used to save potential candidate biomarkers and minimize type II errors<sup>55</sup>. Pearson's correlation and  $\chi^2$  test were performed to verify the association between protein level and clinical information. The skewness of the log-transformed protein level was assessed using the e1071 package (version 1.7–3) in R. Logistic regression was performed to develop a binary classification model using the 'glm' function in the caret package (version 6.0–86). To develop a multiclass classification model, multinomial log-linear regression was performed using the 'multinom' function in the nnet (version 7.3–14) and caret packages (version 6.0–86) in R. Both models were validated using nested crossvalidation (5 outer folds and 5 inner folds for each outer fold) and fivefold cross validation<sup>56,57</sup>. The associations of the 16 proteins in the AD model for detecting AD and the APM model for monitoring the progression of AD with biological functions that were related to the progression of AD were characterized using protein lists of postmortem brain network modules and published proteomic data on CSF cohorts from Johnson et al.<sup>22</sup>. Functional annotations of proteins that were enriched in each risk subgroup were identified using the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources (Version 6.8)<sup>58</sup>.

## Data availability

The raw MRM-MS files for all 185 plasma samples were deposited into PeptideAtlas with the quantitation target lists (Dataset identifier: PASS01631; Password: SH6233b). The quantitation information were also deposited to Panorama Public repository (<https://panoramaweb.org/CXISFG.url>). Clinical information and cleaned expression data from Johnson et al. that were used to characterize the associations of the 17 proteins in the AD and APM models were downloaded from the "Synapse" data portal (<https://www.synapse.org/consensus>; Synapse ID: syn20933797).

Received: 5 March 2021; Accepted: 11 January 2022

Published online: 24 January 2022

## References

- Braak, H. & Braak, E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259. <https://doi.org/10.1007/BF00308809> (1991).
- Alzheimer's disease facts and figures. *Alzheimers Dement.* <https://doi.org/10.1002/alz.12068> (2020).
- Driscoll, I. & Troncoso, J. Asymptomatic Alzheimer's disease: A prodrome or a state of resilience?. *Curr. Alzheimer Res.* **8**, 330–335 (2011).
- McKhann, G. M. et al. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 263–269. <https://doi.org/10.1016/j.jalz.2011.03.005> (2011).
- McKhann, G. et al. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group\* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939–939 (1984).
- Albert, M. S. et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 270–279. <https://doi.org/10.1016/j.jalz.2011.03.008> (2011).
- Tsoi, K. K., Chan, J. Y., Hirai, H. W., Wong, S. Y. & Kwok, T. C. Cognitive tests to detect dementia: A systematic review and meta-analysis. *JAMA Int. Med.* **175**, 1450–1458. <https://doi.org/10.1001/jamainternmed.2015.2152> (2015).
- Carnero-Pardo, C. Should the Mini-Mental State Examination be retired?. *Neurología (English Edition)* **29**, 473–481. <https://doi.org/10.1016/j.nrleng.2013.07.005> (2014).
- Blennow, K. & Hampel, H. CSF markers for incipient Alzheimer's disease. *Lancet Neurol.* **2**, 605–613. [https://doi.org/10.1016/S1474-4422\(03\)00530-1](https://doi.org/10.1016/S1474-4422(03)00530-1) (2003).
- Rehman, S. H. et al. Proteomics as a reliable approach for discovery of blood-based Alzheimer's disease biomarkers: A systematic review and meta-analysis. *Ageing Res. Rev.* **60**, 101066. <https://doi.org/10.1016/j.arr.2020.101066> (2020).
- Sweeney, M. D., Sagare, A. P. & Zlokovic, B. V. Blood-brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nat. Rev. Neurol.* **14**, 133–150. <https://doi.org/10.1038/nrneuro.2017.188> (2018).
- Mulder, C. et al. Amyloid-beta(1–42), total tau, and phosphorylated tau as cerebrospinal fluid biomarkers for the diagnosis of Alzheimer disease. *Clin. Chem.* **56**, 248–253. <https://doi.org/10.1373/clinchem.2009.130518> (2010).

13. Mehta, P. D. *et al.* Plasma and cerebrospinal fluid levels of amyloid  $\beta$  proteins 1–40 and 1–42 in Alzheimer disease. *Arch. Neurol.* **57**, 100–105. <https://doi.org/10.1001/archneur.57.1.100> (2000).
14. Tatebe, H. *et al.* Quantification of plasma phosphorylated tau to use as a biomarker for brain Alzheimer pathology: Pilot case-control studies including patients with Alzheimer's disease and down syndrome. *Mol Neurodegener* **12**, 63. <https://doi.org/10.1186/s13024-017-0206-8> (2017).
15. Hernandez, F. & Avila, J. Tauopathies. *Cell. Mol. Life Sci.* **64**, 2219–2233. <https://doi.org/10.1007/s00018-007-7220-x> (2007).
16. Blennow, K. Cerebrospinal fluid protein biomarkers for Alzheimer's disease. *NeuroRx* **1**, 213–225. <https://doi.org/10.1602/neuroRx.1.2.213> (2004).
17. Chen, G. F. *et al.* Amyloid beta: Structure, biology and structure-based therapeutic development. *Acta Pharmacol. Sin.* **38**, 1205–1235. <https://doi.org/10.1038/aps.2017.28> (2017).
18. Boyle, P. A. *et al.* Person-specific contribution of neuropathologies to cognitive loss in old age. *Ann. Neurol.* **83**, 74–83. <https://doi.org/10.1002/ana.25123> (2018).
19. Gillette, M. A. & Carr, S. A. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat. Methods* **10**, 28–34. <https://doi.org/10.1038/nmeth.2309> (2013).
20. Whiteaker, J. R. *et al.* A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nat. Biotechnol.* **29**, 625–634. <https://doi.org/10.1038/nbt.1900> (2011).
21. Kennedy, J. J. *et al.* Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. *Nat. Methods* **11**, 149–155. <https://doi.org/10.1038/nmeth.2763> (2014).
22. Johnson, E. C. B. *et al.* Large-scale proteomic analysis of Alzheimer's disease brain and cerebrospinal fluid reveals early changes in energy metabolism associated with microglia and astrocyte activation. *Nat. Med.* **26**, 769–780. <https://doi.org/10.1038/s41591-020-0815-6> (2020).
23. Domanski, D. *et al.* MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. *Proteomics* **12**, 1222–1243. <https://doi.org/10.1002/pmic.201100568> (2012).
24. Scheltens, P. *et al.* Alzheimer's disease. *Lancet* **388**, 505–517. [https://doi.org/10.1016/s0140-6736\(15\)01124-1](https://doi.org/10.1016/s0140-6736(15)01124-1) (2016).
25. Farrer, L. A. *et al.* Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: A meta-analysis. *JAMA* **278**, 1349–1356 (1997).
26. Leifer, B. P. Early diagnosis of Alzheimer's disease: Clinical and economic benefits. *J. Am. Geriatr. Soc.* **51**, 5281–5288. <https://doi.org/10.1046/j.1532-5415.5153.x> (2003).
27. Bayer, T. A. & Wirths, O. Intracellular accumulation of amyloid-Beta—a predictor for synaptic dysfunction and neuron loss in Alzheimer's disease. *Front. Aging Neurosci.* **2**, 8. <https://doi.org/10.3389/fnagi.2010.00008> (2010).
28. Smith, L. K. *et al.* beta2-microglobulin is a systemic pro-aging factor that impairs cognitive function and neurogenesis. *Nat. Med.* **21**, 932–937. <https://doi.org/10.1038/nm.3898> (2015).
29. Forner, S. *et al.* Intra- and extracellular beta-amyloid overexpression via adeno-associated virus-mediated gene transfer impairs memory and synaptic plasticity in the hippocampus. *Sci. Rep.* **9**, 15936. <https://doi.org/10.1038/s41598-019-52324-0> (2019).
30. Provensi, G. *et al.* A new kid on the block? Carbonic anhydrases as possible new targets in Alzheimer's disease. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms20194724> (2019).
31. Ahn, H. J., Chen, Z. L., Zamolodchikov, D., Norris, E. H. & Strickland, S. Interactions of beta-amyloid peptide with fibrinogen and coagulation factor XII may contribute to Alzheimer's disease. *Curr. Opin. Hematol.* **24**, 427–431. <https://doi.org/10.1097/MOH.0000000000000368> (2017).
32. Akiyama, H. *et al.* Inflammation and Alzheimer's disease. *Neurobiol. Aging* **21**, 383–421 (2000).
33. Gerardino, L. *et al.* Coagulation factor XIII Val34Leu gene polymorphism and Alzheimer's disease. *Neurol. Res.* **28**, 807–809. <https://doi.org/10.1179/016164106X110454> (2006).
34. Chen, M. & Xia, W. Proteomic profiling of plasma and brain tissue from Alzheimer's disease patients reveals candidate network of plasma biomarkers. *J. Alzheimers Dis.* **76**, 349–368. <https://doi.org/10.3233/JAD-200110> (2020).
35. Fischer, P., Götz, M. E., Danielczyk, W., Gsell, W. & Riederer P. Blood transferrin and ferritin in Alzheimer's disease. *Life Sciences* **60**, 2273–2278 (1997).
36. 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).
37. Cui, Y., Huang, M., He, Y., Zhang, S. & Luo, Y. Genetic ablation of apolipoprotein A-IV accelerates Alzheimer's disease pathogenesis in a mouse model. *Am J Pathol* **178**, 1298–1308. <https://doi.org/10.1016/j.ajpath.2010.11.057> (2011).
38. Namba, Y., Tsuchiya, H. & Ikeda, K. Apolipoprotein B immunoreactivity in senile plaque and vascular amyloids and neurofibrillary tangles in the brains of patients with Alzheimer's disease. *Neurosci. Lett.* **134**, 264–266 (1992).
39. Yang, H. *et al.* Prognostic polypeptide blood plasma biomarkers of Alzheimer's disease progression. *J. Alzheimers Dis.* **40**, 659–666 (2014).
40. Yoo, B. K. *et al.* Astrocyte elevated gene-1 (AEG-1): A multifunctional regulator of normal and abnormal physiology. *Pharmacol. Ther.* **130**, 1–8. <https://doi.org/10.1016/j.pharmthera.2011.01.008> (2011).
41. Kim, J. Y., Barua, S., Jeong, Y. J. & Lee, J. E. Adiponectin: The potential regulator and therapeutic target of obesity and Alzheimer's disease. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms21176419> (2020).
42. Das, M. & Gursky, O. Amyloid-forming properties of human apolipoproteins: Sequence analyses and structural insights. *Adv. Exp. Med. Biol.* **855**, 175–211. [https://doi.org/10.1007/978-3-319-17344-3\\_8](https://doi.org/10.1007/978-3-319-17344-3_8) (2015).
43. Koh, J.-Y., Kim, H. N., Hwang, J. J., Kim, Y.-H. & Park, S. E. Lysosomal dysfunction in proteinopathic neurodegenerative disorders: Possible therapeutic roles of cAMP and zinc. *Mol. Brain* **12**, 1–11 (2019).
44. Luo, X., Weber, G. A., Zheng, J., Gendelman, H. E. & Ikezu, T. C1q-calreticulin induced oxidative neurotoxicity: Relevance for the neuropathogenesis of Alzheimer's disease. *J. Neuroimmunol.* **135**, 62–71. [https://doi.org/10.1016/s0165-5728\(02\)00444-7](https://doi.org/10.1016/s0165-5728(02)00444-7) (2003).
45. Hanseeuw, B. J. *et al.* Association of amyloid and tau with cognition in preclinical Alzheimer disease: A longitudinal study. *JAMA Neurol.* <https://doi.org/10.1001/jamaneurol.2019.1424> (2019).
46. Ballatore, C., Lee, V. M. & Trojanowski, J. Q. Tau-mediated neurodegeneration in Alzheimer's disease and related disorders. *Nat Rev Neurosci* **8**, 663–672. <https://doi.org/10.1038/nrn2194> (2007).
47. Kang, Y., Jang, S. & Na, D. L. *Seoul Neuropsychological Screening Battery (SNSB)* (Human Brain Research and Consulting Co., Seoul, 2012).
48. Morris, J. C. The clinical dementia rating (CDR): Current version and scoring rules. *Neurology* **43**, 2412–2414. <https://doi.org/10.1212/wnl.43.11.2412-a> (1993).
49. Kang, Y., Na, D. L. & Hahn, S. A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. *J. Korean Neurol. Assoc.* **15**, 300–308 (1997).
50. Dubois, B. *et al.* Revising the definition of Alzheimer's disease: A new lexicon. *Lancet Neurol.* **9**, 1118–1127. [https://doi.org/10.1016/s1474-4422\(10\)70223-4](https://doi.org/10.1016/s1474-4422(10)70223-4) (2010).
51. Barthel, H. *et al.* Individualized quantification of brain beta-amyloid burden: Results of a proof of mechanism phase 0 florbetaben PET trial in patients with Alzheimer's disease and healthy controls. *Eur J Nucl Med Mol Imaging* **38**, 1702–1714. <https://doi.org/10.1007/s00259-011-1821-1> (2011).
52. Patt, M. *et al.* Metabolite analysis of [<sup>18</sup>F]Florbetaben (BAY 94–9172) in human subjects: A substudy within a proof of mechanism clinical trial. *J. Radioanal. Nucl. Chem.* **284**, 557–562. <https://doi.org/10.1007/s10967-010-0514-8> (2010).



53. Jee, Y. H., Lee, S. J., Jung, K. J. & Jee, S. H. Alcohol intake and serum glucose levels from the perspective of a mendelian randomization design: The KCPS-II Biobank. *PLoS ONE* **11**, e0162930. <https://doi.org/10.1371/journal.pone.0162930> (2016).
54. Abbatiello, S. E., Mani, D. R., Keshishian, H. & Carr, S. A. Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. *Clin. Chem.* **56**, 291–305. <https://doi.org/10.1373/clinchem.2009.138420> (2010).
55. Pascovici, D., Handler, D. C., Wu, J. X. & Haynes, P. A. Multiple testing corrections in quantitative proteomics: A useful but blunt tool. *Proteomics* **16**, 2448–2453. <https://doi.org/10.1002/pmic.201600044> (2016).
56. Vabalas, A., Gowen, E., Poliakoff, E. & Casson, A. J. Machine learning algorithm validation with a limited sample size. *PLoS ONE* **14**, e0224365. <https://doi.org/10.1371/journal.pone.0224365> (2019).
57. Parvande, S., Yeh, H. W., Paulus, M. P. & McKinney, B. A. Consensus features nested cross-validation. *Bioinformatics* **36**, 3093–3098. <https://doi.org/10.1093/bioinformatics/btaa046> (2020).
58. da Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57. <https://doi.org/10.1038/nprot.2008.211> (2009).

## Acknowledgements

This work was supported by the Industrial Strategic Technology Development Program (No. 20000134) and the Korea Health Industry Development Institute (No. HL19C0020 & No. HI19C1132). This study was supported by the KBRI Basic Research Program through the Korea Brain Research Institute, funded by the Ministry of Science and ICT (20-BR-03-02), and the Original Technology Research Program for Brain Science of the National Research Foundation (NRF), funded by the Korean government, MSIT (NRF-2014M3C7A1046041). I.Y. was supported by BK21 FOUR, funded by the National Research Foundation of Korea. Plasma samples and demographic data from all subjects in this study were provided from Chonnam National University Hospital and Chosun University Hospital, including the National Research Center for Dementia in Gwangju, Korea. We thank all subjects who donated blood for this study. The results that are published here, especially Supplementary Fig. S6 and Supplementary Table S11, are based in part on data from the AD Knowledge Portal (<https://adknowledgeportal.synapse.org/>). For Supplementary Fig. S6, these data were generated from cerebrospinal fluid (CSF) that was collected through the Emory Goizueta Alzheimer's Disease Research Center. Samples from this study are also part of the M2OVE-AD Consortium, supported by NIH/NIA (PI: Ihab Hajjar, MD, MS grant numbers: AG051633 and AG057470). The investigators acknowledge the contribution of the study participants who donated their time and effort for this study.

## Author contributions

Y. K. and K. H. L. contributed to the study concept and design. Y. K., J. K., M. S., I. Y., K. Y. C., H. K., and B. C. K. contributed to the acquisition of data, data processing, statistical analysis, and reagents. Y. K., J. K., M. S., J. L., K. Y. C., and Y. K. drafted the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-05384-8>.

**Correspondence** and requests for materials should be addressed to K.H.L. or Y.K.

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

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



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

© The Author(s) 2022