Phosphorylated Tau-Aβ$_{42}$ Ratio as a Continuous Trait for Biomarker Discovery for Early-Stage Alzheimer's Disease in Multiplex Immunoassay Panels of Cerebrospinal Fluid

Oscar Harari, Carlos Cruchaga, John S.K. Kauwe, Benjamin J. Ainscough, Kelly Bales, Eve H. Pickering, Sarah Bertelsen, Anne M. Fagan, David M. Holtzman, John C. Morris, and Alison M. Goate, for the Alzheimer's Disease Neuroimaging Initiative

Background: Identification of the physiologic changes that occur during the early stages of Alzheimer's disease (AD) may provide critical insights for the diagnosis, prognosis, and treatment of disease. Cerebrospinal fluid (CSF) biomarkers are a rich source of information that reflect the brain proteome.

Methods: A novel approach was applied to screen a panel of ~190 CSF analytes quantified by multiplex immunoassay, and common associations were detected in the Knight Alzheimer's Disease Research Center ($N = 311$) and the Alzheimer's Disease Neuroimaging Initiative ($N = 293$) cohorts. Rather than case-control status, the ratio of CSF levels of tau phosphorylated at threonine 181 (ptau$_{181}$) and Aβ$_{42}$ was used as a continuous trait in these analyses.

Results: The ptau$_{181}$-Aβ$_{42}$ ratio has more statistical power than traditional modeling approaches, and the levels of CSF heart-type fatty acid binding protein (FABP) and 12 other correlated analytes increase as AD progresses. These results were validated using the traditional case-control status model. Stratification of the dataset demonstrated that increases in these analytes occur very early in the disease course and were apparent even in nondemented individuals with AD pathology (low ptau$_{181}$, low Aβ$_{42}$) compared with elderly control subjects with no pathology (low ptau$_{181}$, high Aβ$_{42}$). The FABP-Aβ$_{42}$ ratio demonstrates a similar hazard ratio for disease conversion to ptau$_{181}$-Aβ$_{42}$ even though the overlap in classification is incomplete suggesting that FABP contributes independent information as a predictor of AD.

Conclusions: Our results indicate that the approach presented here can be used to identify novel biomarkers for AD correctly and that CSF heart FABP levels start to increase at very early stages of AD.

There is accumulating evidence that the clinical symptoms of Alzheimer's disease (AD) are preceded by a long preclinical phase in which pathologic protein aggregation occurs in the brain (1–4). β-amyloid (Aβ) plaques are estimated to develop ~15–20 years before the onset of cognitive impairment, and neurofibrillary tangles begin to accumulate at least 5 years before symptom onset (2,3). Substantial neurodegeneration is apparent

in mildly symptomatic individuals (1). These observations, together with the failure of clinical trials in symptomatic individuals, illustrate the urgent need for additional biomarkers that characterize the preclinical stage of disease and enable treatment before the brain undergoes irreversible neurologic damage (2–4).

Analytes in cerebrospinal fluid (CSF) reflect the brain proteome and are a rich source of biomarkers. The levels of Aβ$_{42}$, tau, and tau phosphorylated at threonine 181 (ptau$_{181}$) in CSF are related to AD by their association with the presence of Aβ plaques deposition, neurofibrillary tangles, and neuronal cell death (5–8). Similarly, changes in CSF Visinin-like protein-1 (VILIP-1), a neuronal calcium-sensor protein employed as a marker of neuronal injury (2,9), and chitinase 3-like 1 (cartilage glycoprotein-39; YKL-40), an inflammatory biomarker (10), have also been associated with AD. Additionally, both ptau$_{181}$-Aβ$_{42}$ ratio and YKL-40-Aβ$_{42}$ ratio have been shown to be strong predictors of conversion from cognitively normal to very mild to mild cognitive impairment over a 3- to 4-year period (1,10,11). Despite this evidence, there is a need for additional biomarkers to identify and characterize the "preclinical" stage (pathology present with cognition intact) of the disease and to track the effectiveness of therapies. A challenge for clinical trials is to identify a target population of individuals with a high risk for converting from cognitively normal to impaired over a relatively short period of time. At the present time, elevated ptau$_{181}$-Aβ$_{42}$ ratio is used in some clinical trials to select such individuals. Other predictive biomarkers are needed both for the selection of patient cohorts for clinical trials and for monitoring disease progression and the

From the Departments of Psychiatry (OH, CC, SB, AMG), Neurology (AMF, DMH, JCM, AMG), Pathology and Immunology (JCM), Genetics (AMG), and Developmental Biology (DMH), Knight Alzheimer's Disease Research Center (AMF, DMH, JCM, AMG), and Hope Center for Neurological Disorders (CC, AMF, DMH, JCM, AMG), Division of Biology & Biomedical Sciences (BJA), Washington University School of Medicine, St. Louis, Missouri; Department of Biology (JSKK), Brigham Young University, Provo, Utah; and Neuroscience Research Unit (KB, EHP), Worldwide Research and Development, Pfizer, Inc., Groton, Connecticut.

Address correspondence to Alison M. Goate, D.Phil., Department of Psychiatry, Washington University School of Medicine, 425 S. Euclid Avenue, St. Louis, MO 63110; E-mail: goatea@psychiatry.wustl.edu.

Received Jun 19, 2013; revised Oct 7, 2013; accepted Nov 19, 2013.

success of disease-modifying treatment outcomes. Ultimately, biomarkers will enable early intervention in individuals before they present with symptoms, with the goal of delaying the onset or preventing the cognitive decline seen in AD before the brain is irreversibly injured (2).

Two features of AD result in misclassification of subjects and severely reduce the power of traditional clinical measures in the search for novel biomarkers: 1) 30% of cognitively normal individuals show Alzheimer's-type neuropathology by age 75 years (2,3,12,13), and 2) current clinical diagnostic methods are only 83% accurate as defined by neuropathologic confirmation at autopsy (14). To identify new CSF analytes associated with AD, we took a novel approach by developing a quantitative measure of "caseness," employing the CSF $ptau_{181}$-Aβ42 ratio as an endophenotype for AD. This status-independent criterion avoids the misclassification of controls and cases and is a continuous quantitative trait. Both of these characteristics improve the power of this approach over traditional clinical measures.

Methods and Materials

Subjects

In participants enrolled in the Knight Alzheimer's Disease Research Center (Knight-ADRC; $N = 311$), diagnosis was based on criteria from the National Institute of Neurological and Communicative Diseases and Stroke-Alzheimer's Disease and Related Disorders Association (15). The procedure for collection, processing, and assessment of CSF is described elsewhere (11,16). Data for subjects included in the Alzheimer's Disease Neuro-imaging Initiative (ADNI; $N = 292$) were accessed from the ADNI website (see Supplementary Methods and Table S1 in Supplement 1) (8). Processing, aliquoting, and storage were performed according to the ADNI Biomarker Core Laboratory Standard Operating Procedures (http://adni.loni.ucla.edu). Additional details for these two studies are provided elsewhere (10,17). Genotyping of rs7412 and rs429358, which define the apolipoprotein E (*APOE*) ε2, ε3, and ε4 isoforms, was previously described (17,18).

Expression studies were carried out using complementary DNA obtained from the parietal lobe from 82 participants with AD (86 $\pm$ 7 years old, 45% male, and 41% *APOE* ε4 carriers) and 39 cognitively normal individuals (85 $\pm$ 9 years old, 41% male, and 23% *APOE* ε4 carriers) obtained through the Knight-ADRC Neuropathology Core. Real-time data were analyzed using the comparative Ct method (19) (see Supplementary Methods in Supplement 1).

Analyte Measurement

The samples from both the Knight-ADRC and the ADNI were evaluated by Rules Based Medicine, Inc. (Austin, Texas) for levels of 190 analytes using the Human Discovery Multi-Analyte Profile (MAP) 1.0 panel and a Luminex 100 platform (Luminex Corporation, Austin, Texas). Only analytes with <10% missing values in each study were analyzed. The protocol used to quantify plasma analytes is described elsewhere (20,21).

For the Knight-ADRC participants, CSF Aβ_{42} and phospho-tau181 ($ptau_{181}$) levels were obtained in duplicate by the Knight-ADRC Biomarker Core by quantitative enzyme-linked immunosorbent assay (Innotest; Innogenetics, Ghent, Belgium) (16). Quantification of CSF VILIP-1 and YKL-40 is described elsewhere (9,10). For the ADNI participants, the CSF Aβ_{42} and

$ptau_{181}$ levels were measured using the multiplex xMAP Luminex platform as described (8).

Statistical Analysis

The analyses were performed in R (The R Foundation for Statistical Computing v2.14.1). Analytes were log-transformed to approximate a normal distribution, and outliers were removed. The associations with the CSF $ptau_{181}$-Aβ_{42} ratio were tested using linear regression models, adjusting for age at lumbar puncture (LP); sex; and *APOE* genotype, encoded as five levels based on the genetic risk (ε22 = 0; ε23 = 1; ε33 = 2; ε24 = 3; ε34 = 4; and ε44 = 5). For analyses using the combined datasets, we included study as a covariate. We used two approaches to correct for the multiple testing associated with examining 64 analytes: Bonferroni correction and simpleM. For 64 tests, the Bonferroni corrected p value corresponding to $p = .05$ is $p = 7.81E-04$. Because this method is most likely too conservative given the correlation structure of the analyte measurements, we also extended and applied the simpleM method (22) (see Supplementary Methods in Supplement 1) and obtained corrected p values ($p = 1.39E-03$ and $p = 1.43E-03$) for the Knight-ADRC and the ADNI studies, respectively. Logistic regression was used to test for association with Clinical Dementia Rating (CDR; CDR = 0 vs. CDR > 0), adjusting for the above-mentioned covariates. Kaplan-Meier survival curves were calculated employing the function survfit, and the Cox proportional hazard regression models were tested using the function coxph (package survival, v2.36-10). The Cox proportional hazard tests were adjusted by age, sex, *APOE* genotype, and study. Hierarchical clustering was calculated using the complete linkage method provided (function hclust), and heatmaps were plotted using the function heatmap.2 (package gplots, v2.10.1). Multivariate stepwise model selection (Table 1A,B) was performed using the function stepAIC (package MASS, v7.3-16), optimizing for the Bayesian information criterion, and included age, sex, and *APOE* genotype as fixed covariates. The minimal multivariate model shown in Table 1C was obtained by selecting the subset of analytes identified as significant in each study (Table 1A,B). We constrained this list to the analytes that were also significant ($p < 5.0E-2$) when the identified models were applied to the other study. The principal component analysis was performed in R running the method prcomp (package stats, v2.14.1). The random forest results were obtained executing the function randomForest (implemented in the package random-Forest, v4.6-6) to grow 1000 trees, and the importance measure was calculated by the mean decrease in accuracy.

Results

CSF $ptau_{181}$-Aβ_{42} Ratio Confers More Statistical Power to Identify CSF Analytes Associated with AD

We first evaluated the statistical power of the CSF $ptau_{181}$-Aβ_{42} ratio measure compared with the CDR at LP, using CSF levels of VILIP-1 (9) and YKL-40 (10), which had previously been measured and shown to be associated with CDR and CDR–sum of boxes in the Knight-ADRC study (11,20,23). The CSF $ptau_{181}$-Aβ_{42} ratio was previously shown to be a strong predictor of both the conversion of cognitively normal subjects to very mild or mild dementia (11) and the rate of decline across time in individuals with very mild dementia (24). The CSF $ptau_{181}$-Aβ_{42} ratio is strongly associated with CDR at LP ($p = 5.05E-16$).

We employed linear regression models to analyze the association of CSF VILIP-1 with the log-transformed levels of CSF

Table 1. Multivariate models

(A)	Inferred Knight-ADRC Model		Replication on ADNI	
	p Value	Effect	p Value	Effect
H-FABP	3.18E-08	.44	1.74E-11	.47
VEGF	1.01E-05	−.63	1.85E-04	−.66
TNF RII	8.56E-04	.63	1.08E-01	.22
Thrombomodulin	2.03E-03	−.37	1.01E-03	−.39
ACE	2.25E-03	−.36	8.66E-01	−.02
Adiponectin	2.29E-03	.17	7.98E-03	.23
HGF	2.69E-03	.51	5.60E-08	.60
MIF	3.30E-03	.33	5.88E-04	.16
SAP	4.26E-03	−.21	9.67E-01	.00
CXCL9	8.17E-03	−.14	9.23E-01	−.01
FasL	1.12E-02	.21	2.95E-01	−.10
Age	1.33E-01	.01	3.17E-01	−.01
Sex	4.84E-01	−.05	3.57E-01	−.07
APOE Genotypes	2.80E-07	.14	3.64E-11	.19
R Squared/BIC	.50/−323.91		.56/−222.42.68	

(B)	Replication on Knight-ADRC		Inferred ADNI Model	
	p Value	Effect	p Value	Effect
H-FABP	2.50E-09	.44	5.06E-16	.50
HGF	2.39E-02	.37	2.19E-11	.68
VEGF	1.65E-04	−.60	4.99E-06	−.68
Complement 3	7.57E-01	−.03	2.62E-05	−.63
ANGPT2	4.04E-01	.12	2.10E-04	.41
APOA	1.13E-02	−.26	2.58E-04	.39
MIF	2.08E-02	.33	6.71E-04	.14
Fibrinogen	2.02E-01	.09	1.73E-03	−.12
Prolactin	6.80E-01	−.04	9.88E-03	−.26
TRAIL.R3	8.13E-01	.04	1.70E-02	−.28
Sex	4.96E-01	.05	8.60E-01	−.01
Age	7.24E-03	.01	9.89E-01	.00
APOE Genotypes	7.90E-08	.16	6.16E-17	.22
R Squared/BIC	.44/−296.91		.60/−370.60	

(C)	Knight-ADRC		ADNI	
	p Value	Effect	p Value	Effect
H-FABP	6.44E-10	.45	1.87E-14	.50
HGF	3.25E-02	.33	5.93E-09	.51
MIF	6.27E-04	.40	3.11E-04	.16
VEGF	3.47E-05	−.59	1.32E-06	−.66
Age	6.78E-03	.01	3.51E-01	.00
Sex	3.27E-01	.06	9.73E-01	.00
APOE Genotypes	9.46E-09	.16	1.44E-14	.21
R Squared/BIC	.42/−336.92		.53/−360.04	

Stepwise optimization in (A) the Knight-ADRC study and (B) the ADNI study. (C) A minimal model that includes the analytes commonly selected in the Knight-ADRC study and the ADNI study.

ACE, angiotensinogen; ADNI, Alzheimer's Disease Neuroimaging Initiative; ANGPT2, angiopoietin 2; APOA, apolipoprotein A; CXCL9, CXC motif chemokine ligand 9'; FasL, Fas ligand; H-FABP, heart, fatty acid binding protein; HGF, hepatocyte growth factor; Knight-ADRC, Knight Alzheimer's Disease Research Center; MIF, macrophage migration inhibitory factor; SAP, serum amyloid P; TNFRII, tumor necrosis factor receptor two; TRAILR3, TRAIL receptor 3; VEGF, vascular endothelial growth factor.

$ptau_{181}$-Aβ_{42} ratio and observed a much stronger association than with CDR ($p = 3.18$E-17 vs. $p = 3.80$E-05). The CSF $ptau_{181}$-Aβ_{42} ratio remained more powerful than CDR even when converted to a dichotomous variable using two partitioning criteria. First, we split the subjects by the median value of CSF $ptau_{181}$-Aβ_{42} ratio; second, we compared the subjects from the upper tercile to the

lower two terciles (intended to resemble the CDR distribution of the Knight-ADRC series). Both logistic models that included the CSF ratio showed much stronger associations ($p = 7.11$E-08 and $p = 4.00$E-09, respectively) than the CDR model ($p = 3.80$E-05).

Similarly, the association of YKL-40 with the CSF $ptau_{181}$-Aβ_{42} ratio was stronger than CDR ($p = 8.99$E-09 vs. $p = 7.76$E-04, respectively). We also tested the normal quantile transformed CSF ratio and the same two dichotomous representations for YKL-40. In every case, the test that included the CSF ratio was more statistically significant than the CDR model ($p = 9.42$E-06 and $p = 3.04$E-05 partitioning by the median tercile and by the upper tercile, respectively). Together, these analyses provide a strong rationale for our subsequent use of the CSF $ptau_{181}$-Aβ_{42} ratio for novel biomarker discovery.

Correlation Structure of CSF Analytes. We applied stringent quality criteria to the CSF analyte measurements, selecting only analytes that were measurable in >90% of the subjects in each series ($N = 311$ for the Knight-ADRC series and $N = 293$ for the ADNI series). This criterion was met by 64 CSF analytes.

Analysis of the correlation matrix for the combined measurements of both datasets showed clear patterns in the analyte levels (Table S2 in Supplement 2). The hierarchical clustering method identified 10 clusters containing 48 of the CSF analytes, with a within-cluster correlation $r^2 > .50$. Six clusters included 2 analytes, and the other four clusters included 3, 4, 9, and 20 analytes (Figure S1 in Supplement 1). The remaining 16 CSF analytes showed only weak correlations to other analytes in the dataset (mean $r^2 = .16$, and maximum $r^2 = .35$).

CSF Analytes Associated with $ptau_{18}$-Aβ_{42} Ratio. Next, we tested for association between each of the CSF analytes and the $ptau_{181}$-Aβ_{42} ratio. There were 13 analytes that were significantly associated with the CSF $ptau_{181}$-Aβ_{42} ratio ($p < 1.0$E-3) and exhibited the same direction of effect in the Knight-ADRC and the ADNI series (Table 2; see Table S3 in Supplement 1 for the complete list of the analytes evaluated). Of these CSF analytes, 11 showed a mean $r^2 = .58$ with each other (Figure 1), whereas the other two, Sortilin (SORT1) and tumor necrosis factor–related apoptosis-inducing ligand R3 (TRAIL-R3), exhibited a slightly lower correlation ($r^2 = .49$ and $r^2 = .48$, respectively) (Figure 1).

In both series, CSF heart-type fatty acid binding protein (H-FABP) was the most significantly associated analyte ($p = 4.77$E-18 for Knight-ADRC and $p = 7.40$E-18 for ADNI) and was associated with higher levels of H-FABP in individuals with CSF $ptau_{181}$-Aβ_{42} ratios indicative of AD (Figure 2A). The regression coefficients for the two studies (Table 2) did not differ significantly (F test $p = .24$).

To investigate whether these associations were apparent in individuals without dementia and individuals with dementia, we combined the subjects from both studies and stratified them by the presence or absence of cognitive impairment. There were 236 subjects with a CDR > 0 and low CSF Aβ_{42} levels, which we employed as a proxy for brain Aβ_{42} deposition (16) (cutoff = 500 pg/mL for the Knight-ADRC study and 192 pg/mL for the ADNI study). To evaluate the impact of Aβ_{42} deposition, we further stratified the 300 cognitively normal subjects (CDR = 0) by CSF Aβ_{42} levels, distinguishing the 192 "clean" controls from the 105 preclinical subjects with lower CSF Aβ_{42} levels. Each of the 13 CSF analytes was significantly associated ($p < 5.0$E-4) in every stratum (Table 3). The analysis of covariance analysis revealed that the effect size for CSF H-FABP was significantly lower in controls compared with the inferred preclinical and clinically assessed cognitively impaired individuals ($p = 4.44$E-04 and $p = 5.87$E-03) and that there was no significant difference in the effect observed in the preclinical and symptomatic cases (Figure 3); this is

Table 2. CSF Analytes Significantly Associated with CSF ptau$_{181}$-Aβ_{42} Ratio ($p <$ 1.0E-03)^a

	Knight-ADRC			ADNI		
Analyte	*p* Value	Effect	CI	*p* Value	Effect	CI
Fatty Acid Binding Protein (Heart) (H-FABP)	4.77E-18	.54	(.43–.66)	7.40E-18	.50	(.39–.61)
Tumor Necrosis Factor Receptor 2 (TNFR2)	2.64E-13	.91	(.68–1.15)	6.11E-07	.55	(.34–.76)
Tissue Factor (F3)	9.18E-13	.54	(.40–.69)	4.20E-09	.46	(.31–.62)
Hepatocyte Growth Factor (HGF)	4.77E-11	.88	(.62–1.13)	5.13E-14	.66	(.50–.82)
Chromogranin A (CgA)	1.52E-10	1.11	(.78–1.44)	1.27E-07	.95	(.60–1.29)
Stem Cell Factor (KITLG)	1.08E-08	.64	(.42–.85)	2.97E-05	.43	(.23–.63)
Sortilin (SORT1)	3.22E-08	.89	(.58–1.20)	9.10E-04	.50	(.21–.79)
Lectin Like Oxidized LDL Receptor 1 (LOX1)	1.54E-07	.51	(.32–.69)	1.34E-05	.47	(.26–.68)
Beta$_2$ Microglobulin (B2M)	3.36E-07	.70	(.44–.97)	5.07E-05	.53	(.28–.79)
TNF-Related Apoptosis–IL R3 (TRAIL-R3)	3.67E-07	.82	(.51–1.13)	4.73E-04	.38	(.17–.58)
Angiopoietin 2 (ANGPT2)	2.01E-06	.58	(.34–.81)	1.90E-04	.36	(.17–.55)
Apolipoprotein E (*APOE*)	4.83E-05	.45	(.23–.66)	8.26E-04	.40	(.17–.63)
Angiotensin-Converting Enzyme (ACE)	6.18E-05	.39	(.20–.58)	7.86E-04	.34	(.14–.54)

ADNI, Alzheimer's Disease Neuroimaging Initiative; CI, confidence interval; CSF, cerebrospinal fluid; Knight-ADRC, Knight Alzheimer's Disease Research Center.

^{a}Only analytes showing similar effect.

reflected in the different levels of CSF H-FABP among the strata. Although H-FABP levels distinguish cognitively impaired individuals from preclinical AD ($p =$ 5.05E-03), the levels of H-FABP are not significantly different in preclinical individuals compared with clean controls, indicating that early Aβ plaque deposition does not affect the association of CSF H-FABP with the CSF ptau$_{181}$-Aβ_{42} ratio. The association of these analytes with the CSF ptau$_{181}$-Aβ_{42} ratio is robust to the presence of clinically assessed cognitively impaired individuals with uncharacteristic high levels of CSF Aβ_{42} (Table S4 in Supplement 1). In contrast, the plasma levels of these analytes were not significantly associated with the CSF ptau$_{181}$-Aβ_{42} ratio ($p >$ .05) in either the Knight-ADRC series or the ADNI series (consistent with the null correlation between the CSF and plasma levels of each analyte) (Table S5 in Supplement 1).

CSF H-FABP Is Associated with ptau$_{181}$-Aβ_{42} Ratio

To determine whether each of the novel analytes provided independent information to predict the CSF ptau$_{181}$-Aβ_{42} ratio, we combined the subjects from both studies and evaluated joint models that included CSF H-FABP and other CSF analytes from the candidate set, one at a time. Only hepatocyte growth factor (HGF) remained significant after inclusion of CSF H-FABP in the model ($p =$ 8.56E-06) (Table S6 in Supplement 1).

It has been reported previously that CSF *APOE*, which is also associated with the CSF ptau$_{181}$-Aβ_{42} ratio (Table 2), is associated with AD (17). To determine whether the previously reported association of CSF *APOE* levels with CSF ptau$_{181}$-Aβ_{42} ratio is independent of the other analytes, we extended the models to include CSF *APOE*. Only CSF angiopoietin 2 (ANGPT2) and CSF angiotensin-converting enzyme (ACE) showed a decrease in the significance of their association ($p =$ 3.53E-03 and $p =$ 3.34E-02), whereas the other CSF analytes remained significant ($p <$ 1.0E-3) (Table S6 in Supplement 1), indicating they are capturing additional information. Similarly, the joint analysis of CSF H-FABP and VILIP-1 (9) or YKL-40 (10) demonstrated that CSF H-FABP provides nonredundant information to the predictive model for CSF ptau$_{181}$-Aβ_{42} ratio (Supplement 1).

To determine whether CSF H-FABP levels were a strong predictor of future conversion to CDR > 0, we compared the
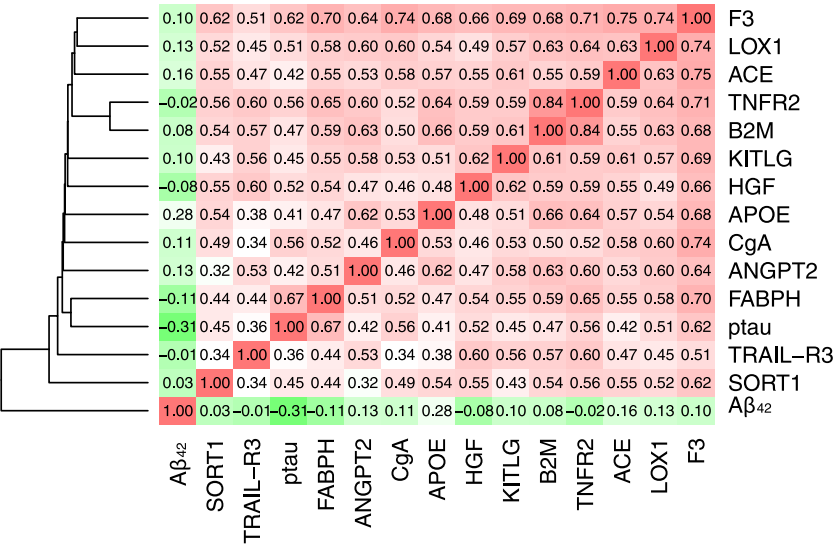
Figure 1. Correlation among the associated analytes and cerebrospinal fluid Aβ_{42} and ptau$_{181}$ in the combined Knight Alzheimer's Disease Research Center and Alzheimer's Disease Neuroimaging Initiative studies. See Table 2 for analytes and their expansions.

Table 3. Stratified Analysis for Cognitively Normal, Preclinical AD, and Symptomatic AD

Analyte	Cognitive Normal (N = 195)			Preclinical AD (N = 105)			Symptomatic AD (N = 236)		
	p Value	Effect	CI	*p* Value	Effect	CI	*p* Value	Effect	CI
Fatty Acid Binding Protein (Heart) (H-FABP)	1.67E-10	.30	(.35–.66)	8.42E-14	0.60	(.46–.74)	1.59E-22	.49	(.40–.58)
Tumor Necrosis Factor Receptor 2 (TNFR2)	1.41E-09	.51	(.30–.49)	1.45E-09	1.07	(.75–1.39)	1.64E-14	.70	(.53–.87)
Tissue Factor (F3)	7.30E-14	.40	(.25–.59)	1.49E-13	.72	(.55–.89)	1.10E-23	.63	(.52–.74)
Hepatocyte Growth Factor (HGF)	1.71E-06	.42	(.38–.85)	8.53E-09	1.04	(.71–1.37)	1.86E-16	.66	(.51–.81)
Chromogranin A (CgA)	7.74E-07	.61	(.32–.59)	3.13E-11	1.50	(1.10–1.90)	4.78E-25	1.45	(1.21–1.70)
Stem Cell Factor (KITLG)	1.19E-10	.46	(.13–.57)	4.59E-10	.99	(.71–1.27)	3.40E-16	.71	(.55–.87)
Sortilin (SORT1)	2.30E-03	.35	(.30–.56)	1.49E-06	1.05	(.64–1.45)	1.88E-07	.67	(.42–.91)
Lectin Like Oxidized LDL Receptor 1 (LOX1)	4.08E-10	.43	(.46–.80)	1.30E-10	.85	(.62–1.09)	1.67E-11	.57	(.41–.73)
Beta₂ Microglobulin (B2M)	4.75E-12	.63	(.22–.59)	5.38E-07	.99	(.63–1.36)	1.73E-09	.64	(.44–.85)
TNF-Related Apoptosis–IL R3 (TRAIL-R3)	2.66E-05	.41	(.42–.69)	1.12E-04	.86	(.44–1.29)	1.58E-03	.32	(.12–.52)
Angiopoietin 2 (ANGPT2)	2.66E-14	.55	(.22–.56)	2.86E-05	.75	(.41–1.08)	1.13E-08	.49	(.33–.65)
Apolipoprotein E (*APOE*)	6.82E-06	.39	(.26–.52)	1.29E-06	.76	(.47–1.05)	1.68E-11	.59	(.42–.75)
Angiotensin-Converting Enzyme (ACE)	1.53E-08	.39	(.46–.74)	2.32E-07	.67	(.43–.90)	3.56E-12	.59	(.43–.75)

AD, Alzheimer's disease; CI, confidence interval.

association of CSF H-FABP levels with CDR at LP and CDR after a mean of 4 years following LP. We tested the association of CSF H-FABP with CDR at LP and found that it was significantly associated in the ADNI study but marginally associated in the Knight-ADRC study ($p = 1.55E-04$ and $p = 8.66E-02$, respectively) (Figure 2B). We also evaluated the association with the latest available evaluation of CDR (mean elapsed years, 4.29; SD, 2.30 years) and found it to be significant for the Knight-ADRC study ($p = 3.72E-03$; 36 converted from CDR = 0 to CDR > 0) and for the ADNI study ($p = 6.17E-05$; mean elapsed years, 3.90; SD, 1.92 years). We observed a similar trend when we analyzed the CDR–sum of boxes (25) and the Mini-Mental State Examination (26). In the ADNI study, H-FABP levels were strongly associated with both endophenotypes at LP and the latest available evaluations, whereas in the Knight-ADRC study, H-FABPH levels were significantly associated ($p < .05$) only for the latest evaluations (Table S7 in Supplement 1).

To test whether the messenger RNA expression of H-FABP is also associated with CDR, we measured the messenger RNA levels in parietal lobe tissue from 112 subjects (73 AD cases and 39 controls) (27). The messenger RNA levels of H-FABP in parietal lobe were not associated with CDR at death or at expiration ($p = .76$) or Braak and Braak staging of the pathology ($p = .46$) (28).

CSF H-FABP as a Predictor of Developing Cognitive Impairment. To determine whether higher levels of CSF H-FABP predict

conversion from CDR = 0 to CDR > 0, we performed a survival analysis. The Cox regression analysis showed similar hazard ratios (HRs) for CSF H-FABP (HR = 1.58, 95% confidence interval [CI] = 1.02–2.44, $p = 3.70-02$), CSF ptau₁₈₁ (HR = 2.04, 95% CI = 1.11–3.71, $p = 1.91E-02$), and CSF Aβ₄₂ (HR = 2.93, 95% CI = 1.15–7.42, $p = 2.74E-02$). In addition, we compared the survival curves for CSF ptau₁₈₁-Aβ₄₂ ratio (11) and CSF H-FABP-Aβ₄₂ ratio (29). Both ratios showed similar results (Figure 4) (Cox proportional HR = 2.05 and HR = 1.85, $p = 2.14E-03$ and $p = 3.93E-03$; 95% CI = 1.29–3.27and 95% CI = 1.21–2.79 for CSF ptau₁₈₁/Aβ₄₂ and CSF H-FABP/Aβ₄₂, respectively). Only 62% of the subjects were assigned to the same tercile when stratified by these ratios [$p(Q_{CSF\ ptau/A\beta42} = Q_{CSF\ H-FABP/A\beta42}) = .65$]. Additionally, we observed that the CSF H-FABP-Aβ₄₂ ratio is associated with CDR–sum of boxes (Cox proportional HR = 1.76, $p = 2.83e-05$, 95% CI = 1.35–2.29); it is also associated with the Mini-Mental State Examination (Cox proportional HR = 1.74, $p = 1.47e-04$, 95% CI = 1.30–2.32)

CSF Macrophage Migration Inhibitory Factor and Vascular Endothelial Growth Factor Are Also Associated with CSF ptau₁₈₁-Aβ₄₂ Ratio

We investigated whether multivariate models would detect analytes that did not show strong evidence of association in single-analyte models. Our approach included a discovery phase
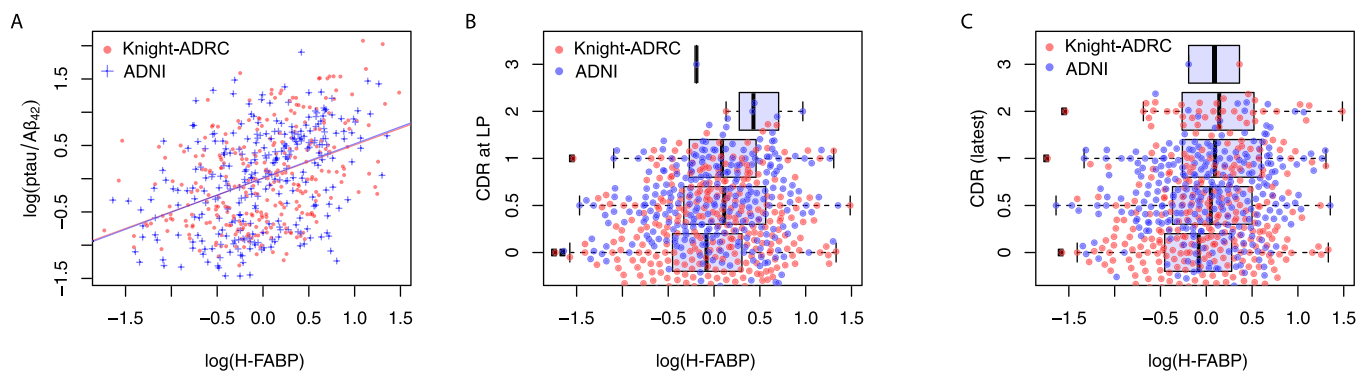
Figure 2. Distribution of cerebrospinal fluid (CSF) heart-type fatty acid binding protein (H-FABP) in the combined Knight Alzheimer's Disease Research Center and Alzheimer's Disease Neuroimaging Initiative (ADNI) studies. **(A)** Linear model of CSF H-FABP compared with CSF ptau/Aβ₄₂. Boxplots for CSF H-FABP stratified by **(B)** Clinical Dementia Rating (CDR) at lumbar puncture (LP) and **(C)** latest ascertained CDR.
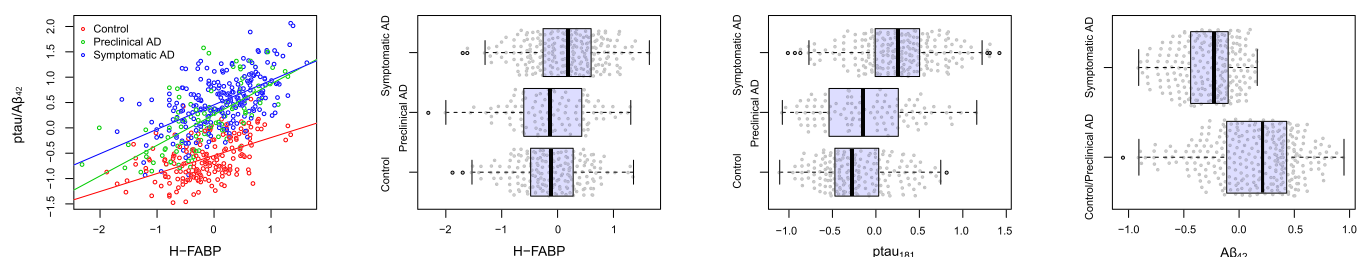
Figure 3. Distribution of cerebrospinal fluid heart-type fatty acid binding protein (H-FABP), ptau$_{181}$, and Aβ_{42} stratified by the cognitive status of the subjects (cognitive normal, preclinical, and cognitive decline) in the combined Knight Alzheimer's Disease Research Center and the Alzheimer's Disease Neuroimaging Initiative studies. AD, Alzheimer's disease.

in which we made use of stepwise regression analysis to optimize a multivariate model for the CSF ptau$_{181}$-Aβ_{42} ratio in one of the datasets, which was followed by a replication phase, in which the model was applied in the second dataset.

The multivariate model optimized for the Knight-ADRC dataset (Table 1A) showed an increased R^2 (.14 for the discovery set and .12 for the ADNI study). It included H-FABP and HGF as well as CSF vascular endothelial growth factor (VEGF), macrophage migration inhibitory factor (MIF), thrombomodulin, and adiponectin (Table 1A). The model optimized for the ADNI dataset (Table 1B) also increased the R^2 for both datasets (.20 and .08 for the ADNI and the Knight-ADRC dataset, respectively). This model also selected H-FABP, HGF, VEGF, and MIF. We constructed a multivariate model with these four analytes and observed that all of them remained significantly associated with the CSF ptau$_{181}$-Aβ_{42} ratio in the two datasets (Table 1C), reflecting an increased goodness of fit for the model (increment of R^2 = .08 for the Knight-ADRC dataset and increment of R^2 = .09 for the ADNI dataset). To evaluate whether the correlation among the analytes produced spurious results, we applied principal component analysis to these selected analytes and evaluated a multivariate

model with the values rotated. We found that all of them remained significant ($p < .05$), indicating that these analytes provide additional information regardless of their correlation.

In the single-analyte analysis, CSF MIF levels show a trend toward association with the CSF ptau$_{181}$-Aβ42 ratio (Table S3 in Supplement 1). Although CSF VEGF is mildly positively associated in the single-analyte analysis (Table S3 in Supplement 1), it has a negative effect in the multivariate model. The stratified analysis revealed that the change of direction occurs only in the cases (Figure S2 in Supplement 1).

The random forest method was also used to analyze these datasets, and it also highlighted the importance of H-FABP, MIF, HGF, and VEGF. These analytes are among the top six for the Knight-ADRC dataset. The analytes H-FABP, HGF, and MIF are among the top eight analytes for the ADNI dataset (Table S8 in Supplement 1).

Discussion

The goal of this study was to identify novel CSF biomarkers for AD using a discovery dataset collected by the Knight-ADRC study and a replication cohort ascertained by the ADNI. Although the
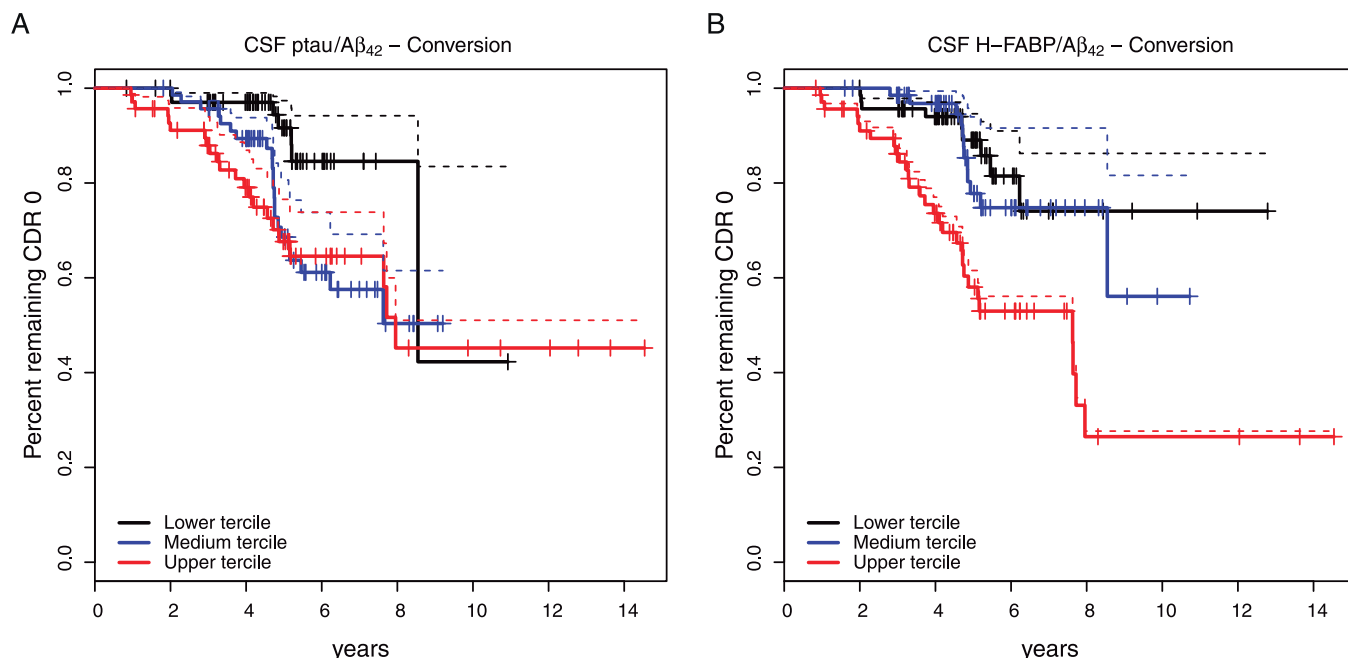
Figure 4. Kaplan-Meier (solid lines) and Cox survival (dashed lines) curves for the conversion (Clinical Dementia Rating [CDR] = 0 to CDR > 0) for cerebrospinal fluid (CSF) ptau$_{181}$/Aβ_{42} and CSF heart-type fatty acid binding protein (H-FABP)/Aβ_{42}. **(A)** Only 7.46% of the subjects within the lowest tercile for CSF ptau-Aβ_{42} ratios converted after 5 years compared with 23.18% and 27.14% of the subjects in the intermediate and higher terciles. **(B)** Similarly, 8.69%, 10.92%, and 34.78% of the subjects within the lower, intermediate, and higher terciles of CSF H-FABP/Aβ_{42} values had converted after 5 years.

ascertainment and structure of the datasets are quite different, we were able to identify several biomarkers that showed consistent effects across the two datasets. The Rules Based Medicine Discovery Multi-Analyte Profile 1.0 panel includes 64 analytes that are not totally independent of one another. We used the correlation structure of these analytes to approximate the number of independent tests and consequently used this number to adjust our p values for multiple testing. We employed this information to understand the apparent excess of analytes associated with the CSF $ptau_{181}$-$A\beta_{42}$ ratio.

A novel feature of our study is the use of the CSF $ptau_{181}$-$A\beta_{42}$ ratio as the outcome variable in our analyses rather than comparing the levels of the test analytes in cases versus controls. The CSF $ptau_{181}$-$A\beta_{42}$ ratio provides several key advantages over the more traditional approach. It captures the progression of the disease before the onset of clinical symptoms by correcting the misclassification of control subjects with $A\beta$ plaques. In addition, this method correctly assigns individuals with clinical dementia who do not have AD. In a traditional analysis, these individuals would be included as cases, but they are analyzed as non-AD dementias when using the $ptau_{181}$-$A\beta_{42}$ ratio. We believe that each of these factors contributes to the gain in statistical power of the CSF $ptau_{181}$-$A\beta_{42}$ ratio that we observed comparing its performance with the usual case-control model in the evaluation of the CSF levels of VILIP-1 and YKL-40.

The analytes associated with the CSF $ptau_{181}$-$A\beta_{42}$ ratio were correlated with CSF $ptau_{181}$ but not $A\beta_{42}$ (Figure 1), indicating that they do not reflect the very early $A\beta$-related events in the development of the disease, but they do discriminate between preclinical and symptomatic AD. Query of the Database for Annotation, Visualization, and Integrated Discovery (30) failed to identify any pathway that characterized the candidate set of analytes.

The CSF levels of H-FABP, the most significant analyte, were consistently associated with the $ptau_{181}$-$A\beta_{42}$ ratio in both the Knight-ADRC and the ADNI datasets, even when the model included other reported biomarkers, such as CSF VILIP-1 and YKL-40. Two previous studies reported the association of CSF H-FABP with AD in other smaller cohorts (31,32). In the current study, we show that this association is present even at very early stages of the disease, as demonstrated by the analysis of cognitively normal subjects with evidence of $A\beta$ plaques. One limitation of our analysis is that although both studies are longitudinal, the novel biomarkers have been measured only in cross-sectional data. We are unable to address the role of FABP as a novel AD biomarker across the entire course of disease at the present time. Despite this limitation of our study design, we used the available longitudinal data to show how CSF H-FABP can be employed to predict the conversion from cognitive normality to cognitive impairment. It still remains to be determined whether addition of H-FABP in a model including both CSF $ptau_{181}$ and $A\beta_{42}$ levels improves the accuracy of predicting conversion to symptomatic AD. Subjects in different terciles for the CSF H-FABP-$A\beta_{42}$ and the CSF $ptau_{181}$-$A\beta_{42}$ ratios suggest this trend, but the studies included in our analysis did not provide statistical power to test this hypothesis.

In contrast, H-FABP levels do not predict progression from CDR = .5 to CDR > .5, but it is noteworthy that the number of subjects in this analysis is small ($N = 81$) compared with the analysis of conversion from CDR = 0 to CDR > 0 ($N = 219$). Based on these results, we believe that H-FABP may be a very useful biomarker in staging preclinical AD.

The protein H-FABP is a member of a family of proteins characterized as lipid chaperones (33) that transport lipids to specific compartments in the cell or outside the cell (33). It has been demonstrated that in vitro fatty acids induce $A\beta$ assembly and modulate the rate of tau polymer assembly (34). Genetic studies have implicated several genes involved in lipid metabolism as risk factors for AD, including APOE (35) and ABCA7 (36).

The multivariate analysis of the Multi-Analyte Profile 1.0 panel of analytes identified distinct complex models for the Knight-ADRC and the ADNI datasets. Nevertheless, these models included a common subset of analytes. This subset of selected analytes includes CSF H-FABP as well as HGF, which was reported previously to discriminate AD subjects from subjects with other neurodegenerative disorders (32). Similarly, VEGF was reported to discriminate controls from moderately severe AD cases but not from mild AD cases (31). We confirmed the role of these analytes in our dataset by demonstrating an association with CDR at LP.

In conclusion, the use of the CSF $ptau_{181}$-$A\beta_{42}$ ratio as an endophenotype of AD led to the identification of a set of 13 correlated analytes associated with the disease in two independent cohorts. This set of candidate analytes is extended by two additional analytes, which are significantly associated only in the context of multivariate models. Despite the differences in our two datasets (the ADNI study has many symptomatic individuals and fewer nondemented controls, whereas the Knight-ADRC study is largely composed of controls and inferred preclinical AD samples), these analytes were consistently associated with the CSF $ptau_{181}$-$A\beta_{42}$ ratio and AD. Our analysis suggests that these analytes capture distinct information from that identified by the traditional analytes (CSF $A\beta_{42}$ and tau/$ptau_{181}$) for the disease and describe novel facets of disease pathogenesis. Analysis of the role of these analytes as novel biomarkers of AD pathogenesis in additional datasets would help to determine the specificity of these changes to AD. Longitudinal analyses also would enable characterization of the temporal sequence of analyte changes in the pathologic cascade of AD.

1. Price JL, McKeel DW Jr, Buckles VD, Roe CM, Xiong C, Grundman M, et al. (2009): Neuropathology of nondemented aging: Presumptive evidence for preclinical Alzheimer disease. *Neurobiol Aging* 30: 1026–1036.
2. Holtzman DM, Goate A, Kelly J, Sperling R (2011): Mapping the road forward in Alzheimer's disease. *Sci Transl Med* 3:114ps48.
3. Holtzman DM, Morris JC, Goate AM (2011): Alzheimer's disease: The challenge of the second century. *Sci Transl Med* 3:77sr1.
4. Bateman RJ, Xiong C, Benzinger TLS, Fagan AM, Goate A, Fox NC, et al. (2012): Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med* 367:795–804.
5. Buerger K, Ewers M, Pirttilä T, Zinkowski R, Alafuzoff I, Teipel SJ, et al. (2006): CSF phosphorylated tau protein correlates with neocortical neurofibrillary pathology in Alzheimer's disease. *Brain* 129:3035–3041.
6. Hardy JJ, Selkoe DJD (2002): The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science* 297:353–356.
7. de Leon MJM, DeSanti SS, Zinkowski RR, Mehta PDP, Pratico DD, Segal SS, et al. (2004): MRI and CSF studies in the early diagnosis of Alzheimer's disease. *J Intern Med* 256:205–223.
8. Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, et al. (2009): Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol* 65: 403–413.
9. Tarawneh R, Lee J-M, Ladenson JH, Morris JC, Holtzman DM (2012): CSF VILIP-1 predicts rates of cognitive decline in early Alzheimer disease. *Neurology* 78:709–719.
10. Craig-Schapiro R, Perrin RJ, Roe CM, Xiong C, Carter D, Cairns NJ, et al. (2010): YKL-40: A novel prognostic fluid biomarker for preclinical Alzheimer's disease. *Biol Psychiatry* 68:903–912.
11. Fagan AM, Roe CM, Xiong C, Mintun MA, Morris JC, Holtzman DM (2007): Cerebrospinal fluid tau/beta-amyloid(42) ratio as a prediction of cognitive decline in nondemented older adults. *Arch Neurol* 64: 343–349.
12. Price JL, Davis PB, Morris JC, White DL (1991): The distribution of tangles, plaques and related immunohistochemical markers in healthy aging and Alzheimer's disease. *Neurobiol Aging* 12:295–312.
13. Price JL, Morris JC (1999): Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. *Ann Neurol* 45:358–368.
14. Toledo JB, Brettschneider J, Grossman M, Arnold SE, Hu WT, Xie SX, et al. (2012): CSF biomarkers cutoffs: The importance of coincident neuropathological diseases. *Acta Neuropathol* 124:23–35.
15. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984): 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* 34:939–944.
16. Fagan AM, Mintun MA, Mach RH, Lee S-Y, Dence CS, Shah AR, et al. (2006): Inverse relation between in vivo amyloid imaging load and cerebrospinal fluid Abeta42 in humans. *Ann Neurol* 59:512–519.
17. Cruchaga C, Kauwe JSK, Nowotny P, Bales K, Pickering EH, Mayo K, et al. (2012): Cerebrospinal fluid APOE levels: An endophenotype for genetic studies for Alzheimer's disease. *Hum Mol Genet* 21: 4558–4571.
18. Soares HD, Potter WZ, Pickering E, Kuhn M, Immermann FW, Shera DM, et al. (2012): Plasma biomarkers associated with the apolipoprotein E genotype and Alzheimer disease. *Arch Neurol* 69: 1310–1317.
19. Muller PY, Janovjak H, Miserez AR, Dobbie Z (2002): Processing of gene expression data generated by quantitative real-time RT-PCR. *BioTechniques* 32:1372–1374, 1376, 1378–1379.
20. Craig-Schapiro R, Kuhn M, Xiong C, Pickering EH, Liu J, Misko TP, et al. (2011): Multiplexed immunoassay panel identifies novel CSF biomarkers for Alzheimer's disease diagnosis and prognosis. *PLoS ONE* 6: e18850.
21. Hu WT, Holtzman DM, Fagan AM, Shaw LM, Perrin R, Arnold SE, et al. (2012): Plasma multianalyte profiling in mild cognitive impairment and Alzheimer disease. *Neurology* 79:897–905.
22. Gao X, Starmer J, Martin ER (2008): A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol* 32:361–369.
23. Johnstone D, Milward EA, Berretta R, Moscato P, Alzheimer's Disease Neuroimaging Initiative (2012): Multivariate protein signatures of preclinical Alzheimer's disease in the Alzheimer's disease neuroimaging initiative (ADNI) plasma proteome dataset. *PLoS One* 7:e34341.
24. Snider BJ, Fagan AM, Roe C, Shah AR, Grant EA, Xiong C, et al. (2009): Cerebrospinal fluid biomarkers and rate of cognitive decline in very mild dementia of the Alzheimer type. *Arch Neurol* 66:638–645.
25. Morris JC (1993): The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* 43:2412–2414.
26. Jefferson AL, Cosentino SA, Ball SK, Bogdanoff B, Leopold N, Kaplan E, Libon DJ (2002): Errors produced on the mini-mental state examination and neuropsychological test performance in Alzheimer's disease, ischemic vascular dementia, and Parkinson's disease. *J Neuropsychiatry Clin Neurosci* 14:311–320.
27. Karch CM, Jeng AT, Nowotny P, Cady J, Cruchaga C, Goate AM (2012): Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. *PLoS One* 7:e50976.
28. Braak H, Braak E (1991): Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* 82:239–259.
29. Chiasserini D, Parnetti L, Andreasson U, Zetterberg H, Giannandrea D, Calabresi P, Blennow K (2010): CSF levels of heart fatty acid binding protein are altered during early phases of Alzheimer's disease. *J Alzheimers Dis* 22:1281–1288.
30. Huang DW, Sherman BT, Lempicki RA (2009): Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4:44–57.

31. Ohrfelt A, Andreasson U, Simon A, Zetterberg H, Edman A, Potter W, *et al.* (2011): Screening for new biomarkers for subcortical vascular dementia and Alzheimer's disease. *Dement Geriatr Cogn Dis Extra* 1: 31–42.
32. Hu WT, Chen-Plotkin A, Arnold SE, Grossman M, Clark CM, Shaw LM, *et al.* (2010): Novel CSF biomarkers for Alzheimer's disease and mild cognitive impairment. *Acta Neuropathol* 119:669–678.
33. Furuhashi M, Hotamisligil GS (2008): Fatty acid-binding proteins: Role in metabolic diseases and potential as drug targets. *Nat Rev Drug Discov* 7:489–503.
34. Wilson DM, Binder LI (1997): Free fatty acids stimulate the polymerization of tau and amyloid beta peptides. In vitro evidence for a common effector of pathogenesis in Alzheimer's disease. *Am J Pathol* 150:2181–2195.
35. Castellano JM, Kim J, Stewart FR, Jiang H, DeMattos RB, Patterson BW, *et al.* (2011): Human apoE isoforms differentially regulate brain amyloid-β peptide clearance. *Sci Transl Med* 3:89ra57.
36. Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, *et al.* (2011): Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 43:429–435.