RESEARCH ARTICLE

Development and evaluation of a multiplexed mass spectrometry based assay for measuring candidate peptide biomarkers in Alzheimer's Disease Neuroimaging Initiative (ADNI) CSF

Daniel S. Spellman[1], Kristin R. Wildsmith[2], Lee A. Honigberg[2], Marianne Tuefferd[3], David Baker[4], Nandini Raghavan[4], Angus C. Nairn[5], Pascal Croteau[6], Michael Schirm[6], Rene Allard[6], Julie Lamontagne[6], Daniel Chelsky[6], Steven Hoffmann[7], William Z. Potter[8], Alzheimer's Disease Neuroimaging Initiative[9] and the Foundation for NIH (FNIH) Biomarkers Consortium CSF Proteomics Project Team

[1] Department of Pharmacokinetics, Pharmacodynamics and Drug Metabolism, Merck Research Laboratories, Pennsylvania, PA, USA
[2] Department of Pharmacodynamic Biomarkers within Development Sciences, Genentech, Inc (a member of the Roche Group), South San Francisco, CA, USA
[3] Discovery Sciences, Janssen Research & Development LLC, Pharmaceutical Companies of Johnson & Johnson, Beerse, Belgium
[4] Janssen Research & Development LLC, Titusville, NJ, USA
[5] Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA
[6] Caprion Pharmaceuticals, Montreal, QC, Canada
[7] Foundation for the National Institutes of Health, Inc, Bethesda, MD, USA
[8] National Institute of Mental Health, Bethesda, MD, USA
[9] A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Purpose: We describe the outcome of the Biomarkers Consortium CSF Proteomics Project (where CSF is cerebral spinal fluid), a public–private partnership of government, academia, nonprofit, and industry. The goal of this study was to evaluate a multiplexed MS-based approach for the qualification of candidate Alzheimer's disease (AD) biomarkers using CSF samples from the AD Neuroimaging Initiative.

Experimental design: Reproducibility of sample processing, analytic variability, and ability to detect a variety of analytes of interest were thoroughly investigated. Multiple approaches to statistical analyses assessed whether panel analytes were associated with baseline pathology (mild cognitive impairment (MCI), AD) versus healthy controls or associated with progression for MCI patients, and included (i) univariate association analyses, (ii) univariate prediction models, (iii) exploratory multivariate analyses, and (iv) supervised multivariate analysis.

Results: A robust targeted MS-based approach for the qualification of candidate AD biomarkers was developed. The results identified several peptides with potential diagnostic or predictive utility, with the most significant differences observed for the following peptides for differentiating (including peptides from hemoglobin A, hemoglobin B, and superoxide dismutase) or predicting (including peptides from neuronal pentraxin-2, neurosecretory protein VGF (VGF), and secretogranin-2) progression versus nonprogression from MCI to AD.

Correspondence: Dr. Daniel S. Spellman, Department of Pharmacokinetics, Pharmacodynamics and Drug Metabolism, Merck Research Laboratories, WP75-214, 770 Sumneytown Pike, West Point, PA 19486
E-mail: daniel_spellman@merck.com

Abbreviations: Aβ, amyloid beta peptide; **AD**, Alzheimer's disease; **ADNI**, AD Neuroimaging Initiative; **ApoE**, apolipoprotein E; **BDNF**, Brain-derived neurotrophic factor; **CN**, healthy control; **CSF**, cerebral spinal fluid; **Fabps**, fatty acid binding proteins; **FDR**, false discovery rate; **FTD**, frontotemporal dementia; **HBA**, hemoglobin A; **HBB**, hemoglobin B; **ISP**, internal standard peptide; **MCI**, mild cognitive impairment; **NPTX2**, neuronal pentraxin-2; **p-tau**, hyperphosphorylated tau; **PSP**, progressive supranuclear palsy; **SCG-2**, secretogranin-2; **SODE**, superoxide dismutase

Conclusions and clinical relevance: These data provide potential insights into the biology of CSF in AD and MCI progression and provide a novel tool for AD researchers and clinicians working to improve diagnostic accuracy, evaluation of treatment efficacy, and early diagnosis.

Additional supporting information may be found in the online version of this article at the publisher's web-site

1   Introduction

Alzheimer's disease (AD) is the most common cause of dementia in the elderly and affects greater than 10 million people in the United States of America and Europe (Alzheimer's Association, www.alz.org). AD is a progressive and fatal neurodegenerative disorder that leads to memory loss and diminished cognitive function. The personal and financial costs of AD disease are substantial. Patients ultimately lose independence requiring full-time care; the emotional and physical toll on the individual suffering the disease and their families are well documented [1,2]. The system cost for care of AD patients in 2011 in the United States alone was over US$183 billion, with projected annual costs increasing to US$1 trillion by the year 2050 unless effective disease-modifying treatments are developed [3, 4].

Definitive diagnosis of AD is histological and requires post-mortem identification of the presence of two hallmark brain lesions; extracellular deposits of the β-amyloid (Aβ) peptide (amyloid plaques) and intraneuronal accumulations of hyperphosphorylated tau protein (neurofibrillary tangles). Diagnosis of AD during life is based on guidelines established by the National Institute of Neurological Disorders and Stroke-AD and Related Disorders Association (NINCDS-ADRDA) and a patient is considered to have "dementia of the Alzheimer type" [5]. Despite these established guidelines, and experienced clinicians at specialized AD Centers, the accuracy of clinical AD diagnostic methods is still quite low (comprehensively reviewed by Beach et al. [6]). These challenges are greatest in the early stages of the disease (i.e., mild cognitive impairment (MCI)) where therapeutic intervention may have the best chance of being effective but, behaviorally, symptoms are often limited [7, 8].

Extensive effort within the biomedical research community is currently focused on the discovery and development of biomarkers of AD. Most of these efforts focus on clinically assessable and biologically relevant biofluids (i.e., blood and cerebral spinal fluid (CSF)) as well as the use of imaging-based technologies. To date, three CSF-based biomarkers of AD have been established with substantial international acceptance: decreases in Aβ42 and increases in total tau and phosphorylated tau181 (p-tau181, where p-tau is hyperphosphorylated tau) [9–17]. CSF tau and p-tau181 have been shown to predict progression from MCI to AD in some, in particular in combination with apolipoprotein E (ApoE) genotype, but not all studies [18–20]. In addition, the ratio of tau(s) to

Aβ42 has been shown by several groups to be highly predictive of cognitive decline in cognitively normal cohorts as well as individuals with MCI [20–24]. However, given the wide variations in accuracy for established CSF biomarkers across studies, the need for the identification of additional biomarkers that will, either on their own or in combination with more established markers, increase diagnostic accuracy remains a critical need [25]. Thus, additional biomarkers with diagnostic and prognostic value are also needed for AD drug development, especially in the context of clinical trials aiming to treat patients before the onset of dementia, and will enable more efficient trial designs and facilitate better understanding of mechanism of drug action. Novel biomarkers could also help to better define the pathophysiological stages of AD, to identify additional processes involved in AD pathogenesis, to identify individuals at risk of rapid disease progression, and to understand the underlying biology.

While there are several approaches and technologies currently being employed to interrogate large numbers of proteins such as multiplexed immunoassays[26], protein arrays[27], and aptamer arrays [28] to name a few, MS has emerged as the most widely adopted technology platform for proteomics in biomedical research. Discovery proteomics experiments (specifically clinical proteomics) typically involve the characterization of proteomes of normal or diseased tissues or biological fluids (reviewed in [29, 30]). These approaches most commonly involve relative quantitation of peptides, proteins, or PTM differences that associate or correlate with disease state and/or progression. Discovery MS platforms, typically employed for hypothesis generation, have the capability to confidently identify thousands of proteins in complex biological samples in an unbiased fashion (without prespecification of the analytes to be measured). However, broad and unbiased profiling comes at the cost of reduced sensitivity and stochastic sampling. To move along the translational path for biomarkers to verification and hypothesis testing, more sensitive quantitative protein measurements must be made more precisely and reliably. Targeted MS platforms provide an ideal tool for such activities by focusing the resources of the mass spectrometer on a defined subset of analytes [29]. Over the last few decades, targeted MS approaches have been used to increase the speed, sensitivity, and quantitative precision of biomolecule analysis [29, 31, 32].

Mass spectrometers employing modern triple quadrupole mass analyzers (known as triple quads or QQQ) enable rapid

Clinical Relevance

Effective treatment and diagnosis of Alzheimer's disease (AD) is a significant unmet medical need. The number of patients suffering from this most common form of dementia is expected to rise exponentially in the coming years as we continue to live longer lives in the developed world. Diagnosis of AD is confirmed at the time of autopsy based upon the presence of the pathological hallmarks of disease, extracellular plaques composed primarily of amyloid beta (Aβ), and intracellular tangles, composed primarily of tau and hyperphosphorylated tau. Levels of Aβ, tau, and hyperphosphorylated tau have been shown to have diagnostic value when measured in cerebrospinal fluid (CSF). Positron emission tomography imaging techniques using tracers that bind Aβ are also proving to be of use. However, large gaps remain and additional predictive, prognostic, and pharmacodynamic biomarkers are needed to help improve treatment paradigms and accelerate drug-development efforts. To aid in these efforts the Biomarkers Consortium CSF Proteomics Project team, a public–private partnership of government, academia, nonprofit, and industry has developed a targeted proteomic, multiplexed MS-based approach for the qualification of candidate AD biomarkers in the well-characterized AD Neuroimaging Initiative (ADNI) cohort. The data from this study provide important information about the biology of CSF in health and disease, and identify potential biomarkers for follow-up in other studies.

mass selection and fragmentation of specific precursor ions representing analytes of interest and monitoring signals for only a few predefined fragment ions for each analyte. In a contemporary MRM experiment (also commonly referred to as SRM), each fragment ion from an analyte needs to be sampled for only a few milliseconds to obtain interpretable spectra [33]. MRM allows the specific and sensitive quantification of large numbers of peptides and proteins in biological samples in a single run. It is the most sensitive MS-based platform [34] and was demonstrated to be highly reproducible within and across laboratories and instrument platforms [35, 36]. This approach offers an ideal tool to evaluate the host of differentially expressed analytes emerging from "omics" discovery experiments that require verification and qualification using precise, quantitative methods in larger sample numbers and later stages of clinical validation and implementation [30, 37].

We described here the work of the Biomarkers Consortium Project "Use of Targeted Mass Spectrometry Proteomic Strategies to Identify CSF-Based Biomarkers in Alzheimer's Disease." The aim of the project was to determine the ability of a panel of peptides measured with MS to discriminate among disease states. This project is the second part of a multiphased effort seeking to utilize samples collected by AD Neuroimaging Initiative (ADNI) to qualify multiple peptides in CSF to diagnose patients with AD and monitor disease progression. The earlier phase of the program focused on using a multiplexed immuno-based assay (manuscript in preparation) to characterize potential AD biomarkers in CSF.

2 Materials and methods

Data used in preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. Further details about ADNI are given in the Acknowledgments section.

2.1 CSF from ADNI-1 subjects

306 ADNI-1 baseline CSF samples were used in the study, including 16 blinded technical replicates. Thus 289 unique ADNI-1 baseline subjects were represented: 85 healthy control (CN) subjects, 66 AD subjects, and 134 MCI subjects. The cohort demographics are described in Table 1. Written informed consent was obtained from all participants and the study was conducted with prior institutional ethics approval. Further information about ADNI can be obtained from www.adni-info.org.

2.2 CSF sample processing

CSF sample aliquots of 0.5 mL were stored at −80°C until use. After thawing, 100 μL of each sample was depleted of high abundance proteins using a MARS-14 immunoaffinity resin (4.6 × 100 mm column, Agilent) on an Agilent 1200 HPLC. The depletion gradient used is provided in Supporting Information Table 1). Samples were run in batches of 12, 20, or 21 over 15 days, using two separate MARS-14 columns. Details of run order and column usage are included in Supporting Information Table 2). Three in-run QC samples (HGS-CSF, human gold standard CSF (Bioreclamation, lot BRH631340)) were included per depletion day (beginning, middle, and end). These QC samples were processed at the same time and the same manner as the study samples and were used to assess the reproducibility of the sample processing and MS analysis. The depleted samples, containing the remaining lower abundance proteins, were stored at −80°C. After all samples were depleted, the frozen samples were lyophilized over 72 h. The lyophilized samples were digested

Table 1. Baseline characteristics of the ADNI proteomics cohort

	Controls	LMCI	Progressors	AD
n	85	134	67	66
Age, mean (SD)	75.6 (5.6)	74.7 (7.5)	74.9 (7.3)	75 (7.6)
Gender, percentage of females	41 (48)	44 (33)	25 (37)	37 (56)
Education, mean (SD)	15.61 (2.97)	16 (2.98)	15.82 (3.03)	15.1 (2.95)
MMSE, mean (SD)	29.04 (1.01)	26.9 (1.735)	26.49 (1.73)	23.52 (1.85)
Aβ1-42, mean (SD)	207.8 (56.26)	161.37[a] (52.9)	146[a] (41.95)	141.12 (37.39)
p-tau181P, mean (SD)	24.19 (12.02)	35.06 (15.02)	38.47 (14.51)	41.95 (20.6)
APOE4 (% allele)	64 (75)	64 (48)	26 (39)	19 (29)

a) One sample failed Aβ measurement.

overnight with trypsin (Promega) at 1:10 protease-to-protein ratio, based on the protein amount determined by BCA (bicinchoninic acid assay kit, Pierce). The digested samples were lyophilized and desalted using an Empore C18 96-well plate (3M). Two sets of replicate MS plates were prepared for each sample. The plates were dried by vacuum evaporation and stored at −20°C prior to MS analysis. A flow chart providing an overview of the main steps of sample processing is included in Supporting Information Fig. 1).

2.3 CSF multiplex MRM panel

The MRM panel consisted of 567 peptides representing 222 proteins and, for each peptide, two mass transitions were monitored. Non-HPLC purified synthetic reference peptide standards (unlabeled) were obtained at a cost of approximately $12 per peptide sequence, and used for method development, verification of instrument reproducibility prior to and after sample analysis, and as a reference standard for data analysis. The full list of transitions, peptide sequences, and corresponding proteins (Table 2) are listed in Supporting Information Data. The 640 detectable transitions (320 peptides from 142 proteins) are listed first, followed by the transitions monitored for the internal standards, followed by the remaining 494 transitions that were not detectable and not carried forward in subsequent analysis.

A number of steps were performed in order to QC and combine (or "roll-up") transitions into a peptide quantitation and peptides into a protein quantitation. These results are reported in arbitrary signal intensity units on a natural log scale. The raw data and all the intermediate steps leading up to the final dataset are available online at adni.loni.ucla.edu under the "Biomarkers Consortium CSF Proteomics MRM data set" in the document "CSFMRM Consolidated Data.xlsx" or in Supporting Information Data.

Before analyzing the study samples, a system suitability test of the LC/MRM-MS system was performed. The reconstitution solution, which includes five internal standard peptides (ISPs) (peptide sequences and transitions can be found in Supporting Information Data, annotated as IS-1 through IS-5 in the "Transitions" tab) at 100 ng/mL was injected in replicates of five. For the system to pass the test, the CV of the

Table 2. Proteins selected for inclusion in the LC/MRM-MS panel

1433Z	CMGA	IFNB	NELL2	SCG3
A1AT	CNDP1	IGSF8	NEO1	SDCB1
A1AT	CNTF	IL10	NEUS	SE6L1
A1BG	CNTN1	IL12B	NFH	SHSA7
A2GL	CNTN2	IL17	NFL	SIAE
A2MG	CO2	IL1A	NFM	SLIK1
A4	CO3	IL27A	NGF	SMOC1
AACT	CO4A	IL6	NICA	SODC
AATM	CO5	IL6RA	NLGN3	SODE
AFAM	CO6	ITIH1	NPTX1	SORC1
ALDOA	CO8B	ITIH5	NPTX2	SORC2
AMBP	COCH	ITM2B	NPTXR	SORC3
AMD	CRP	JAK1	NPY	SPON1
APLP2	CSTN1	KAIN	NRCAM	SPRL1
APOA	CSTN3	KCC2B	NRX1A	STX12
APOA1	CUTA	KI67	NRX2A	SV2A
APOB	CYTC	KLK10	NRX3A	SYNJ1
APOC1	DAG1	KLK11	NSG1	SYT11
APOD	DIAC	KLK12	OSTP	TADBP
APOE	ENOG	KLK3	PCD17	TAU
B2MG	ENPP2	KLK6	PCMD1	TCRG1
B3GN1	EXTL2	KLK9	PCSK1	TEN3
BACE1	FABP5	KLKB1	PDIA3	TGFB1
BASP1	FABP6	KNG1	PDYN	TGFB2
BDNF	FABP7	KPCZ	PEDF	TGFB3
BTD	FABPH	KPYM	PGRP2	TGON2
C1QA	FABPI	L1CAM	PIMT	THRB
C1QB	FAM3C	LAMB2	PLDX1	TIMP1
C3AR	FBLN1	LFTY2	PLMN	TNF14
CA2D1	FBLN3	LPHN1	PPN	TNFA
CAD13	FETUA	LRC4B	PRDX1	TNR1B
CADM3	FMOD	LTBP2	PRDX2	TNR21
CAH1	GFAP	MIME	PRDX3	TNR6
CATA	GLNA	MMP2	PRDX4	TRBM
CATD	GOGB1	MMP9	PRDX5	TRFE
CATL1	GOLM1	MMRN2	PRDX6	TRFM
CCKN	GRIA4	MOG	PTGDS	TTHY
CCL25	HBA	MTHR	PTPRD	UBB
CD14	HBB	MUC18	PTPRN	UCHL1
CD59	HEMO	NBL1	PVRL1	VASN
CERU	HERC4	NCAM1	RIMS3	VGF
CFAB	I18BP	NCAM2	SAP	VTDB
CH3L1	IBP2	NCAN	SCG1	X3CL1
CLUS	IBP6	NEGR1	SCG2	

median peak area of the five ISP must have been below 7.5%, and the CV of the median retention time of the five ISP must have been below 0.25%. In addition, a synthetic peptide mix containing all synthesized crude peptides (JPT Peptide Technologies), each at a concentration 200 pmol/mL, was injected in replicate of five. It was required that the targeted peptides elute within 30 s of their predicted retention times and that the median peak area CV over all measured transitions was below 10%.

Sample analysis was initiated after a successful system suitability test. The processed samples were resolubilized with 11 μL of a reconstitution solution containing five ISP each at 100 ng/mL. In addition, one of the HGS-CSF samples from the backup plate was resolubilized with 10 μL of the reconstitution solution plus 1 μL of the synthetic peptide mix at 200 pmol/mL. This sample was used later for the retention time alignment in Elucidator. Eight microliters of material were injected per sample onto a NanoAcquity UPLC (Waters) coupled to a 5500 QTRAP mass spectrometer (AB Sciex). Peptide separation was achieved using a 320 μm × 150 mm, 5 μm particle size, Thermo Biobasic C18 column. Peptides were eluted with a linear LC gradient of 7.5–25% buffer B (0.2% formic acid in ACN) over 23 min followed by 1 min at 60% buffer B and 6 min at 92.5% buffer A (0.2% formic acid in water). The flow rate was 10 μL/min.

2.4 Data processing, normalization, and peptide quantitation

The raw mass spectrometer data files (WIFF) were converted to mzXML format and loaded into the Elucidator software (version 3.3.0.1 SP4.25, Rosetta Biosciences [38]) and processed using the "PeakTeller" processing pipeline for chromatogram alignment, noise filtering, data smoothing, peak detection, and quantitation. The peak alignment was then manually reviewed. If more than 20% of the peaks of a sample were not well aligned with the others, the sample was excluded. The following set of five additional peptide verification criteria were implemented to using Perl scripts developed in house and orchestrated using Elucidator's "visual scripts" plug-in interface:

(i) *Detection threshold*: A transition was included if it was observed in at least 10% of the total samples analyzed (HGS-CSF and study samples) with a peak area of 7500 or more.

(ii) *Wrong intensity ratio*: The expected intensity ratios of the two transitions for each peptide were calculated from ten injections of synthetic peptide standards spiked in buffer (ratio of peak areas summed across ten samples). The observed transition intensity ratio was then calculated across all CSF samples having peak areas >10 000 on both transitions. For each peptide, the distribution of the observed ratios was considered and the 25th, 50th, and 75th percentiles of the distribution were extracted. If the expected ratio was not contained within the 25th and 75th percentiles and was not within 1.5-fold of the

median ratio, then the peptide was flagged (across all samples) because of suspected matrix interference. Interference was then manually confirmed by visual inspection of chromatograms.

(iii) *Low intensity correlation of transition pair*: A peptide was flagged if the squared Pearson correlation coefficient of the transition pair was lower than 0.25.

(iv) *Imperfect coelution*: A peptide was flagged if the retention time difference of the centroid of the two transitions was greater than 0.05 min (3 s).

(v) *Departure from expected RT*: A peptide was flagged if the retention time of the pair of transitions was significantly offset (more than 10 s) compared to its expected retention time.

Peptides with one or more flags were manually reviewed and were either kept or discarded, depending on the overall peak shape, the quality of the alignment, and the presence of a neighboring interference. Once the final set of transitions was validated, the peak area data were then exported to a tab-delimited file for further processing (Supporting Information Data reported in the spreadsheet "Raw Intensity"). All intensity values were transformed on the natural log scale, in order to bring their distribution closer to a Gaussian one. To account for the zero values, the exact transformation is $\ln(\text{intensity} + 1)$ and is referred to as log intensity. This transformation was used for all following steps, except when mentioned otherwise. A two-step normalization procedure was applied to the raw peak area data (final normalized data is available in Supporting Information Data and is reported in the spreadsheet "Normalized Intensity"). A detailed description of the normalization procedure in provided in the Supporting Information Methods and Supporting Information Fig. 4 (and can be found online at adni.loni.ucla.edu. under the "Biomarkers Consortium CSF Proteomics MRM data set" in the "Data Primer" document). The intensities for the validated transitions as described above were used to project "rolled-up" peptide and protein values as the score of the first principal component (PCA), which optimally captures the covariation in the data due to concentration differences between the samples (described in detail in the Supporting Information Methods)

A mix of synthetic standard peptides in buffer, each at 200 pmol/mL, was analyzed before ($n = 5$) and after ($n = 5$) the study samples to assess instrument reproducibility and sensitivity. The non-normalized transition intensity values for these ten replicates are included in Supporting Information Data in the spreadsheet "External Standards." At the transition level, an estimate of protein concentration in the study samples can be calculated as follows:

Protein concentration in nanograms per milliliter
= ((Transition intensity in study sample × inj. volume
× synthetic peptide conc. × MW protein)
/(average transition intensity in peptide mix × 1000 × CSF volume injected))

These estimates are included in Supporting Information Data in the spreadsheet "Estimated Concentrations." Because the transition intensities from the standard mixes were not included in the roll up into peptide and protein intensities, no concentration estimates at the peptide or protein level were calculated. The reported protein concentrations at the transition level are considered as an illustrative estimate and have not been demonstrated to be accurate or precise. Subsequent studies using heavy isotope labeled ISPs are required to generate robust concentration measurements.

2.5 Unblinded statistical analysis

Analyses were conducted following data processing as described above, and the quality control process described below in detail, after the dataset was made publically available (Posted April 9, 2014 on adni.loni.ucla.edu) and unblinded. Data analysis was performed at the peptide level from the final peptide quantitation data (Supporting Information Data in the spreadsheet "CSV-Export"). The metadata used for the unblinded analyses consisted of subject-specific covariates including age, gender, diagnosis at baseline, diagnosis at month 36 to identify progressors, APOE4 status, CSF Aβ, CSF total, and p-tau. All metadata used in this analysis is publicly available and were obtained from the ADNI database (adni.loni.ucla.edu).

The demographic and diagnostic data were obtained from the ADNIMERGE R-package [02.2014]. Two hundred eighty-five samples, one sample/individual in the ADNI 1 cohort, picked randomly from the test–retest dataset, of which 66 had a baseline diagnosis of AD, 134 had a baseline diagnosis of MCI, and 85 were CN. Of the 134 MCIs, 67 had progressed to AD by the month 36 visit, and 67 had not. One sample was discarded because of incomplete CSF data. There were 142 proteins (320 peptides) profiled including three blood contamination markers, hemoglobin A (HBB), hemoglobin A (HBA), and APOB.

Four sets of analyses were conducted: (i) univariate association analyses, (ii) univariate prediction models, (iii) exploratory multivariate analyses, and (iv) supervised multivariate analysis. To determine whether specific panel analytes were associated with baseline pathology (MCI, AD) versus CNs, univariate association analyses were conducted. Differentially expressed analytes were identified comparing (i) AD and controls, (ii) MCI and controls, and (iii) MCI converters and MCI nonconverters, adjusting for age, gender, and APOE4 status. The Benjamini–Hochberg false discovery rate (FDR) was applied for multiple testing corrections. Univariate prediction models were used to determine whether specific panel analyte values at baseline were predictive of progression from MCI to AD by 3 years, over and above that given by age, gender, APOE4 status, and log(Tau/Ab142). The latter has been established as a marker of progression [39]. Progressors were defined as MCI subjects who had converted to AD by the month 36 visit in the ADNI1 cohort. Within the MCI popula-

tion, the reference model (Ref-Model) applied was a logistic regression model for MCI progressors versus MCI nonprogressors, with age, gender, APOE, and log(Tau/Aβ1-42) as covariates in the model. The test model (Test-Model) was the same as the reference model, with each protein/peptide of interest as an additional covariate. Additionally, unsupervised, exploratory multivariate analysis was employed to look for patterns of separation in the peptides and proteins, as well as subgroups of subjects, using spectral maps based on centered PCA [40]. The analysis dataset consisted of the demographic and diagnostic data, as well as the protein and peptide data.

Supervised multivariate analysis was performed to determine if a specific panel of analytes at baseline could be predictive of progression to AD within the MCI group. Panel selection was based on penalized logistic regression. The idea is to fit a logistic regression model with all peptides as predictors and then keep the best contributors to the model as the selected proteins. While L1 penalization would have provided natural selection, L2 penalization was preferred for better predictive accuracy. Regression weights were computed for each peptide (indexed by l) as $(\beta_l^2 / \sum_i \beta_i^2)$, where the βs are the regression coefficient of the peptides. Predictive accuracy was estimated by the AUC score in a cross-validation frame. The cross-validation frame was a fivefold stratified random split, with stratification based on MCI status (converters, nonconverters), dubbed layer 1. Each fold of layer 1 was further split into 100 random training and testing sets (proportion of 80–20%, stratified by MCI status). This is layer 2 of the cross-validation frame. Regression weights were estimated in each training set of the second layer and averaged to get the layer 1 regression weights. Final weights were the average across all five layer 1 folds. Transitions with a weight above 3.5% were selected to be part of the regression model. AUC was estimated by fitting the regression model, with selected transitions, on each layer 1 training set and evaluating accuracy on the corresponding testing set. Reported AUC is the average of the five AUCs.

For panels that do not need selection, as for the panels of common peptides between univariate and multivariate analyses, predictive accuracy was estimated by the AUC score in a cross-validation frame. The cross-validation frame was a five-fold stratified random split, with stratification based on MCI status (converters, nonconverters). Regression models were fit in each training set and AUC calculated on the corresponding testing set. Reported AUC is the average of the five AUC. Peptides were represented by one of their transition. Only transitions with a detection rate (i.e., ratio of intensity above 10 000) above 50% were considered. For the eligible transitions, the within-group (converters, nonconverters) variance and average of squared lag below ln(10 001) was summed (a log intensity over ln(10 001) contributes 0 to the average) and the worse score between converters and nonconverters was kept as a transition score. The transition with the lowest score was selected as the peptide representative.

Statistical analysis was performed using R version 3.0.1 (http://www.R-project.org/). The limma package [41], available from the Bioconductor [42] package version 2.12, was used for the association study, a generalized linear model was applied for the prediction study, and penalized regression was implemented using R package glmnet version 1.9-8. The mpm package, also available from Bioconductor, was used for the spectral map analysis. Spotfire (version 5.5.1) was used for visualizations.

3 Results and discussion

3.1 Building a CSF multiplex MRM panel

Proteins and peptides were selected based upon their previous detection in CSF, relevance to AD, and previous results from the rules based medicine (RBM) multiplex immunoassay analysis of ADNI CSF (manuscript in preparation). CSF samples were depleted of abundant plasma proteins, digested, and analyzed by LC-MS/MS operating in MRM mode. In a pilot study, 25 ADNI CSF samples were analyzed to determine which proteins could be detected in CSF using the MRM platform, and to identify the best representative peptides for monitoring each protein in the final MRM assay. For each protein that had not been detected in previous in-house LC-MS/MS analysis of CSF (unpublished data), approximately five peptides were tested and two were typically selected for subsequent use. In the pilot study, an initial list of candidate markers constructed from a wide ranging search of published and unpublished AD biomarker studies conducted in CSF, brain, and cell lines was evaluated. A total of 198/510 peptides (representing 121/267 proteins) were detectable in ADNI CSF in this pilot. Peptides not detected are likely present at concentrations below the LOD of the platform. As the platform was capable of monitoring ~500 peptides, for the final MRM panel, we supplemented the detectable peptides from the pilot study with a large number of additional peptides, not all of which were known to be detectable in CSF. These additional peptides included a series of peptides representing inflammatory markers and peptides representing particular proteins of interest identified in the RBM study. The final MRM panel consisted of 567 peptides representing 222 proteins and, for each peptide, two mass transitions were monitored. As described above, a number of steps were performed in order to QC and combine (or "roll-up") transitions into a peptide quantitation and peptides into a protein quantitation. These results are reported in arbitrary signal intensity units on a natural log scale. Since alternate splicing or posttranslational processing could result in biologically significant differences in the levels of two peptides from the same protein, we focused on the use of the "log peptide intensity" for further analysis. The raw data and all the intermediate steps leading up to the final dataset are included in Supporting Information Data. The final results indicate that 320 of 567 peptides in the final MRM panel were detectable in >10% of ADNI samples and are included in

the final results file Supporting Information Data in the spreadsheet "CSV-Export."

3.2 Data quality control and peak detection

All data quality control and test/retest analysis was performed before the sample ID's were unblinded. Reproducibility and quality of abundant protein immunodepletion carried out via HPLC was evaluated by monitoring the CV of the mean AUC for total protein in the flow-through fraction of the in-run QCs per depletion day. In this study, the CV of the mean flow-through AUC value was ≤2.2% per depletion day, ≤4.7% per depletion column, and 4.9% across all QC CSF samples.

Before analyzing the study samples, a system suitability test of the LC/MRM-MS system was performed. The reconstitution solution, which includes five ISPs at 100 ng/mL was injected in replicates of five. The CV of the median peak area of the five ISP in this study was 2.5%. The CV of the median retention time of the five ISP in the current study was 0.07%. It was verified that the targeted peptides elute within 30 s of their predicted retention times (based on a synthetic peptide mix containing all synthesized crude peptides at 200 pmol/mL injected in replicate of five) and that the median peak area CV over all measured transitions was below 10% (4.8% in this study).

Sample analysis was initiated after a successful system suitability test. The five ISP were used to monitor the instrument's performance during sample analysis. In the current study, the median CV was 14.8% across all samples. The number and percentage of detected transitions, peptides, and proteins is reported in Supporting Information Fig. 3. The median CV of the synthetic peptide mix in buffer was 4.9% in the prerun samples, 5.9% in the postrun samples, and 12.1% overall. Supporting Information Table 2 lists the median CV of the ISP in the study samples and the median CV of the ISP and all detected non-ISPs in the HGS-CSF for each study day, each column, and for the entire analysis. Sample ZGJ0297 was excluded because peaks shifted outside the MRM detection window. Sample ZGJ0038 was excluded because it aligned poorly with the other samples during the Elucidator peak alignment. This was likely caused by the relative high amount of hemoglobin present in these two samples (samples were pink).

A transition was considered detected if it was observed in at least 10% of the total samples analyzed (HGS-CSF and study samples) with a peak area of 7500 or more. A peptide was considered detected if both of its transitions were detected. A protein was considered detected if at least one of its peptides was detected. Quality and reproducibility of sample processing and MS analysis was also evaluated using the 45 HGS-CSF samples. The median peak area CV over all transitions detected was ≤22.7%. The median peak area CV over all transitions detected across all QC CSF samples was 22.6%. Sixty-eight transitions, representing 39 different peptides, with CV >35% over the 45 HGS-CSF samples were identified (see column CV(HGS) Supporting

Information Data in the spreadsheet "Normalized Intensity"). Supporting Information Fig. 3 displays individual transition CV in HGS-CSF versus average intensity. The average intensity and CV of individual transitions across all HGS-CSF samples are included in the columns ln(AVE(HGS)) and CV(HGS) in Supporting Information Data in the spreadsheet "Normalized Intensity." Higher CV across the HGS-CSF samples is associated with lower signal intensity. Rather than select an arbitrary cutoff for exclusion from the roll-up process, all 640 detected transitions were included in the roll up to peptide quantitation.

Prior to quantitation, we determined the proportion of samples with a prenormalized intensity value of at least 10 000 the intensity threshold established for quantitation as compared to the lower value of 7500 accepted for detection (see column "%raw transition sample values >10 000" in Supporting Information Data in the spreadsheet "Normalized Intensity"). The transitions with detection in fewer than 50% of samples are listed in Supporting Information Table 3. These transitions include peptides that were intentionally included in the panel as markers of blood contamination, and so might be expected to be present in a small number of samples. Thus no filter was applied based on the number of CSF samples in which a transition was detected because peptides that are present in only a small number of samples may still be informative and should be available for subsequent analysis.

3.3 Outlier and pattern detection

The distributions of the number of detected transitions by sample with a peak area greater than 0 (Supporting Information Fig. 2a) or greater than 10 000 (Supporting Information Fig. 2b) were examined. The distribution of the sample average intensity was also evaluated (Supporting Information Fig. 2c). In the case of the averages, the peak areas were first transformed on the log scale ln(peak area + 1). Two samples with average intensity greater than three standard deviations from their group mean were flagged as outliers (ZGJ0043 (Batch 13) and ZGJ0005 (Batch 14)) but these samples were still included in subsequent peptide and protein quantitation steps. A standard PCA was applied to the peak area data in order to identify patterns unrelated to biological variability (described in detail in Supporting Information Methods). Supporting Information Fig. 2d displays the sample distribution based on the first two principal components of the PCA performed on raw intensity data, using all transitions. No samples were flagged as outliers based on this analysis.

3.4 Technical reproducibility evaluation by test/retest analysis

Three hundred six ADNI-1 baseline CSF samples were analyzed, including 16 blinded technical replicates. The 16 blinded technical replicates were distributed throughout the MS analysis runs and used to assess assay reproducibility. As

described in more detail below, there was very high concordance between these "test/retest" replicates, indicating overall robustness in the detection and data-processing pipeline. Data quality was assessed using the ADNI CSF test/retest samples using the peptide-level intensity data for all 320 peptides contained in Supporting Information Data in the spreadsheet in the spreadsheet "log(Peptide Intensity)." Reproducibility between each of the 16 replicate pairs was assessed graphically and by computing Spearman correlation and concordance coefficients between each pair (see Supporting Information Data test–retest). Overall very high reproducibility of intensity was observed, with correlation between technical replicates above 0.959 for all the 16 pairs. Most of the peptides with larger variability had relatively lower signal intensities.

We developed a statistic to flag less-reliable peptides. In Step 1, the most variable peptides were empirically identified in the test–retest dataset, based on differences in expression level between technical duplicates. Supporting Information Fig. 5a shows the distribution of the maximum difference in any of the 16 pairs for each peptide, a bimodal distribution of differences was observed. An arbitrary cutoff of 5 was then applied, identifying 24 peptides (red dots represent technical outliers). Column "Step 1" in Supporting Information Data in the spreadsheet "Test-Retest Flagged Peptides" lists these 24 peptides. These peptides showed a difference of greater than 5 on the log scale within at least one of the 16 technical replicate pairs. The column "log(Peptide Intensity)" can be found in Supporting Information Data in the spreadsheet "Test-Retest Flagged Peptides" to indicate these 24 test/retest flagged peptides. However, we noted that several of these peptides were included in the panel as potential markers of blood contamination (HBA, HBB, and APOB). If blood contamination was minimal in the 16 CSF samples represented in the test/retest set, these peptides would be expected to have low and noisy signal but they might still have utility in detecting rare blood contaminated samples in the full sample set. Most of the other flagged peptides had low signal in the test/retest samples. Given the small number of flagged peptides and the potential for these peptides to still be informative in a small number of samples in full sample set, we did *not* exclude the 24 flagged peptides in the final dataset, Supporting Information Data in the spreadsheet "CSV-Export."

To further understand the relationship of test/retest performance to signal intensity, we explored various metrics on the test–retest dataset to identify outliers. The metrics and corresponding rules are summarized in Supporting Information Fig. 5b. These included the standard deviation, maximum, median absolute deviation, and minimum for each peptide computed across the 32 samples in the test–retest dataset. These statistics are shown plotted against the mean in Supporting Information Fig. 5b. Of the four statistics, the minimum appeared to be the most consistent in identifying the same peptides identified in Supporting Information Fig. 5a. We use a threshold of 0 for the minimum log(roll up) value in at least one of the 32 samples to flag peptides.

Nineteen peptides were identified and, of these, 17 were among the 24 peptides flagged as technical outliers in Step 1 and two additional peptides seen as black dots below the red line in Supporting Information Fig. 5b. Column "Step 2" shows the peptides identified. In a final step to explore these relationships, we applied this rule of identifying all negative log(roll up) values to the entire dataset of 304 samples. Supporting Information Fig. 5c shows the minimum versus mean plot for the complete dataset. This rule applied to the entire dataset identified 27 peptides, including 19 of the 24 technical outliers as defined in Step 1. All peptides identified in this step are reported in column "Step 3."

In summary, the test–retest dataset allowed us to estimate the technical reproducibility of the MRM technology on the ADNI CSF dataset. All 16 pairs (technical replicates) showed a comparable and very high correlation and concordance. The most variable peptides were associated with low expression in both the test–retest dataset (32 samples) and the entire dataset (304 samples). Defining a threshold of a minimum log(roll up) value in any of the samples below zero allows us to capture 79% of the technically variable peptides flagged in the test/retest evaluation. However, given that a number of the peptides that were flagged using these metrics were in fact specifically included in the panel as blood contamination markers and might have utility in a small subset of samples, all 320 peptides in Supporting Information Data in the spreadsheet "log(Peptide Intensity)" were accepted as "QC Pass" in the test/retest analysis and subsequently included in the CSV export. Flagging information is included for informational purposes.

3.5 Unblinded data analysis

Upon completion of data processing and quality control analysis, the dataset was made publically available (posted April 9, 2014 on adni.loni.ucla.edu) and unblinded. Given the potential for differences between the levels of peptides from a single protein and the unknown biological relevance of any such differences, data analysis was performed at the peptide level from the final peptide quantitation data (Supporting Information Data, spreadsheet "CSV-Export"). Data analysis aimed to identify peptides that were differentially expressed relative to diagnostic category, after controlling for covariates, as well as peptides that differentiate between progressors and nonprogressors. Four sets of analyses were conducted: (i) univariate association analyses, (ii) univariate prediction models, (iii) exploratory multivariate analyses, and (iv) supervised multivariate analysis.

The goal of the association analysis was to determine whether specific peptides were associated with baseline pathology (MCI, AD) versus CNs, after controlling for age, gender, and APOE4. Table 3 shows the top-ranked peptides and Fig. 1 shows a plot of the raw *p*-values for each peptide against the log fold-change for AD versus CN, MCI versus

CN, and MCI converters versus MCI nonconverters. Each peptide is colored by its adjusted *p*-value range. Interestingly, two peptides from FABPH topped the list of peptides for both AD versus CN and MCI versus CN (Table 3, Supporting Information Fig. 6A), although they did not show statistical significance upon FDR correction. Peptides from HBA, HBB, and superoxide dismutase (Table 3, Supporting Information Fig. 6B–E) demonstrated significance for differentiation between progressors and nonprogressors within the MCI group.

The goal of the prediction models was to determine whether specific panel analytes at baseline were predictive of progression by 3 years, over and above that given by age, gender, APOE4 status, and log(Tau/Aβ1-42). The latter has been established as a marker of progression [39]. Table 1 shows the basic characteristics of the samples at baseline, including summary statistics of age, gender, and baseline MMSE, p-tau, Aβ1-42, and APOE4 status. Within the MCI population, the reference model (Ref-Model) applied was a logistic regression model for MCI progressors versus MCI nonprogressors, with age, gender, APOE, and log(Tau/Aβ1-42) as covariates in the model. The test model (Test-Model) was the same as the reference one, with each peptide of interest as an additional covariate. Figure 2 shows the plots of the misclassification rate (*x*-axis) from the Test-Model versus the *p*-value (*y*-axis) for the specific protein in the Test-Model. This plot helps understand how useful the specific peptide was for the prediction accuracy of the model, relative to the reduction in misclassification rate achieved. The misclassification rate for the Ref-Model is given by the gray line at 0.36 on the *x*-axis. For example, peptide VLEYLNQEK from secretogranin-2 (SCG2) achieves a reduction in misclassification (down to 0.28 from 0.36 for the Ref-Model), and it is evident from the *p*-value of 0.005 on the *y*-axis that this protein is contributing significantly to the model. The most significant peptides from this analysis (indicated by green squares in Fig. 2) were LESLEHQLR and TESTLNALLQR from NPTX2 (where NPTX2 is neuronal pentraxin-2), THLGEALAPLSK, AYQGVAAPFPK, and NSEPQDEGELEFQGVDPR from VGF, and VLEYLNQEK from SCG2 (Supporting Information Fig. 6F and G).

Unsupervised, exploratory, multivariate analysis was conducted with the goal of identifying patterns of separation in the peptides, as well as subgroups of subjects, using spectral maps based on centered PCA [40]. Figure 3 shows a spectral map of the MCI subjects in the data, with MCI converters in orange and MCI nonconverters in gray. From this plot, we can see an emerging separation of certain subjects corresponding to elevate/suppressed values of peptides EFTPPVQAAYQK, VNVDEVGGEALGR, and SAVTALWGK from HBB; VGAHAGEYGAEALER, FLASVSTVLTSK, and TYFPHFDLSHGSAQVK from HAA; and VLDALQAIK and YSSLAEAASK from CAH1(Supporting Information Fig. 6B–D).

A supervised multivariate analysis was performed, combining peptide-level intensity data with established clinical

Table 3. Summary of peptide level univariate association analyses

Peptide	Protein	logFC	Average expression	t	p-Value	Adj. p-value
Top 20 peptides on the AD vs. CN comparison						
SIVTLDGGK	FABPH	0.386	14.666	3.096	0.002	0.694
SLGVGFATR	FABPH	0.316	15.706	2.832	0.005	0.796
VEIDTK	TTHY	0.248	15.603	2.514	0.013	0.891
LVEAFGGATK	NPTXR	−0.385	18.425	−2.504	0.013	0.891
LESLEHQLR	NPTX2	−0.395	9.338	−2.303	0.022	0.891
ELDVLQGR	NPTXR	−0.315	21.740	−2.285	0.023	0.891
GATLALTQVTPQDER	MUC18	−0.201	16.624	−2.197	0.029	0.891
GEAAGAVQELAR	PCSK1	−0.330	21.425	−2.114	0.035	0.891
VPGLYYFTYHASSR	C1QB	−0.312	17.603	−2.113	0.036	0.891
ALAHLLEAER	PCSK1	−0.351	19.072	−2.105	0.036	0.891
SVLGQLGITK	A1AT	0.445	9.833	2.035	0.043	0.891
SSIIFAGGDK	NEGR1	−0.229	20.941	−1.936	0.054	0.891
SDLYIGGVAK	NRX1A	−0.247	16.574	−1.929	0.055	0.891
FVVTDGGITR	CA2D1	−0.287	20.252	−1.923	0.056	0.891
ITTQITAGAR	NRX1A	−0.275	16.234	−1.921	0.056	0.891
LSALTLSTVK	NRX2A	−0.263	19.328	−1.891	0.060	0.891
TESTLNALLQR	NPTX2	−0.394	10.571	−1.882	0.061	0.891
NSDPALGLDDDPDAPAAQLAR	PCSK1	−0.294	13.999	−1.870	0.063	0.891
LENLEQYSR	NPTX1	−0.238	15.057	−1.857	0.064	0.891
Top 20 peptides on the MCI vs. CN comparison						
SIVTLDGGK	FABPH	0.315	14.666	2.622	0.009	0.998
SLGVGFATR	FABPH	0.249	15.706	2.311	0.022	0.998
LSINTHPSQKPLSITVR	CO3	1.926	7.393	2.074	0.039	0.998
LEQGENVFLQATDK	C1QB	−0.193	17.127	−1.868	0.063	0.998
VEIDTK	TTHY	0.160	15.603	1.677	0.095	0.998
VPGLYYFTYHASSR	C1QB	−0.228	17.603	−1.598	0.111	0.998
TITLEVEPSDTIENVK	UBB	0.155	22.402	1.579	0.115	0.998
WADLSGITK	KAIN	0.285	17.639	1.501	0.135	0.998
LGAEVYHTLK	ENOG	0.496	6.795	1.488	0.138	0.998
ESSQEQSSVVR	KLK6	0.182	10.108	1.480	0.140	0.998
SVLGQLGITK	A1AT	0.312	9.833	1.475	0.141	0.998
VGSALFLSHNLK	KAIN	0.266	13.665	1.444	0.150	0.998
SWLAELQQWLKPGLK	CD14	−0.138	20.664	−1.427	0.155	0.998
SYPEILTLK	ENPP2	−0.082	22.329	−1.403	0.162	0.998
ESTLHLVLR	UBB	0.173	22.771	1.396	0.164	0.998
FYYLIASETPGK	KAIN	0.247	16.784	1.391	0.165	0.998
TGEVLDTK	FAM3C	0.189	12.125	1.348	0.179	0.998
ATYIQNYR	DIAC	0.110	15.474	1.347	0.179	0.998
ALQASALK	ALDOA	0.160	19.168	1.345	0.180	0.998
IIEVEEEQEDPYLNDR	FBLN1	0.142	9.947	1.343	0.180	0.998
Top 20 peptides on the progressors vs. nonprogressors comparison						
VGAHAGEYGAEALER	HBA	2.938	19.149	4.023	0.000	0.028
FLASVSTVLTSK	HBA	2.488	15.117	3.870	0.000	0.028
VNVDEVGGEALGR	HBB	2.938	17.911	3.694	0.000	0.028
SAVTALWGK	HBB	3.199	20.449	3.628	0.000	0.028
TYFPHFDLSHGSAQVK	HBA	2.327	16.299	3.615	0.000	0.028
VTGVVLFR	SODE	−0.243	25.187	−3.436	0.001	0.043
EFTPPVQAAYQK	HBB	3.380	12.291	3.293	0.001	0.059
AGLAASLAGPHSIVGR	SODE	−0.232	15.298	−3.086	0.003	0.100
IGKPAPDFK	PRDX2	0.641	12.962	3.027	0.003	0.107
GLFIIDGK	PRDX2	0.656	19.629	2.944	0.004	0.124
ALQLPYR	KLK10	−0.440	8.566	−2.877	0.005	0.138
ITQVTWQK	PVRL1	−0.279	14.046	−2.779	0.006	0.168
IHWESASLLR	CO3	−0.886	14.047	−2.696	0.008	0.197
ESDTSYVSLK	CRP	−0.797	15.218	−2.646	0.009	0.210
YSSLAEAASK	CAH1	1.834	15.993	2.306	0.023	0.456
TESTLNALLQR	NPTX2	−0.415	10.599	−2.291	0.024	0.456
NGQWTLIGR	AMD	−0.288	16.270	−2.275	0.025	0.456
VLDALQAIK	CAH1	1.698	18.069	2.231	0.027	0.456
IILEALR	SCG2	−0.241	20.846	−2.195	0.030	0.456
INENTGSVSVTR	CAD13	−0.275	11.873	−2.170	0.032	0.456

logFC, log2 fold-change of peptide expression for each comparison of interest; average expression, average peptide expression for each comparison of interest; t, moderated t statistic for each peptide and each comparison of interest; p-value, raw p-value of the moderated t statistic; adj. p-value, adjusted p-value for multiple testing (applying Benjamini–Hochberg false discovery rate correction).
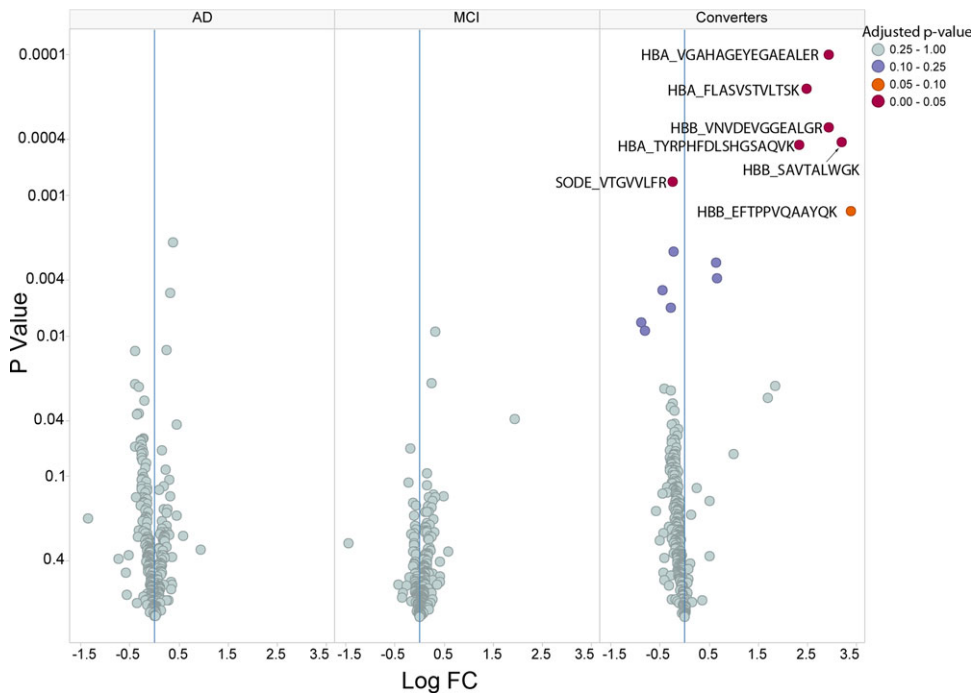
Figure 1. Volcano plots of peptides association analysis. Each filled circle represents a peptide, −log10(raw *p*-value) versus log(fold change). Each peptide is colored by its corresponding adjusted *p*-value (after Benjamini–Hochberg FDR correction). FDR <5% in red, 5–10% in orange, 10–25% in purple, >25% in gray.

risk factors (APOE4 genotype, age, gender, and Aβ/p-tau ratio), to determine if a multipeptide panel would be predictive of conversion to AD within the MCI group. A selection methodology was developed to identify peptides that served as the best contributors to the performance of a multipeptide panel (Table 4). Cross-validation was used to estimate the accuracy of the panel. Panels were also identified with various combinations of peptide-level intensity data, APOE4 genotype, and clinical data. The addition of peptide-level intensity data to established clinical factors increases appreciably the predictive accuracy of the model. The model APOE genotype + Aβ/p-tau + age + gender, AUC of 0.67, improved to an AUC of 0.79 with the addition of the peptide-level intensity data (for the peptides listed in Table 4). Some of
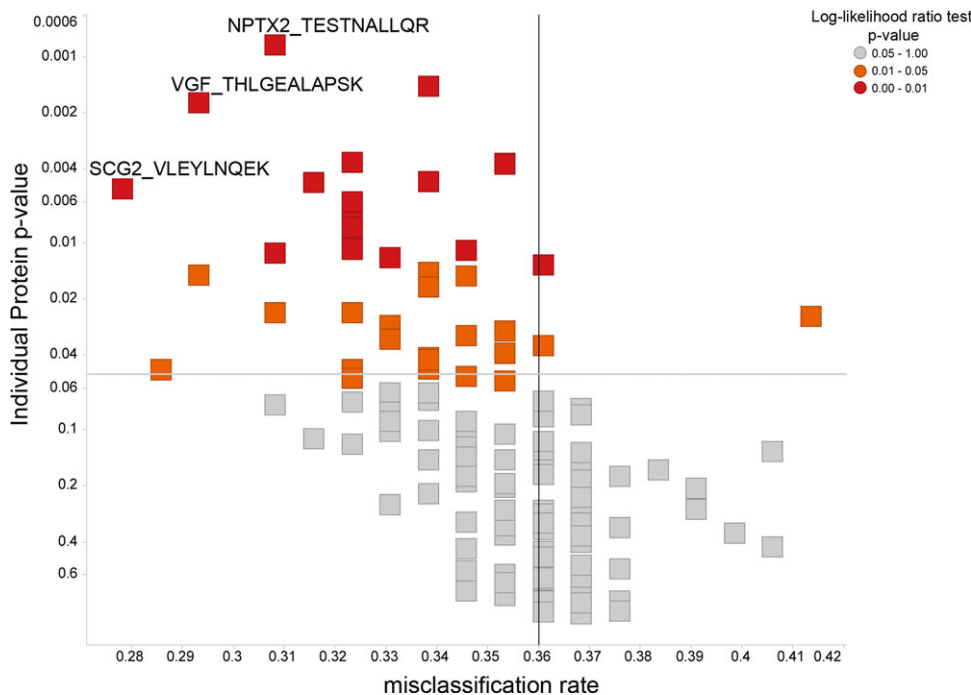
Figure 2. Univariate prediction analysis of progression from MCI to AD at 36 months. Each square represent a protein in the test versus reference model comparison (protein-specific *p*-value vs. misclassification rate from the test-model). The color scheme applied for each protein is based on log-likelihood ratio test significance comparing the test-model and the reference-model (*p*-value <0.01 in red, 0.01–0.05 in orange, >0.05 in gray). The top three peptides are labeled.
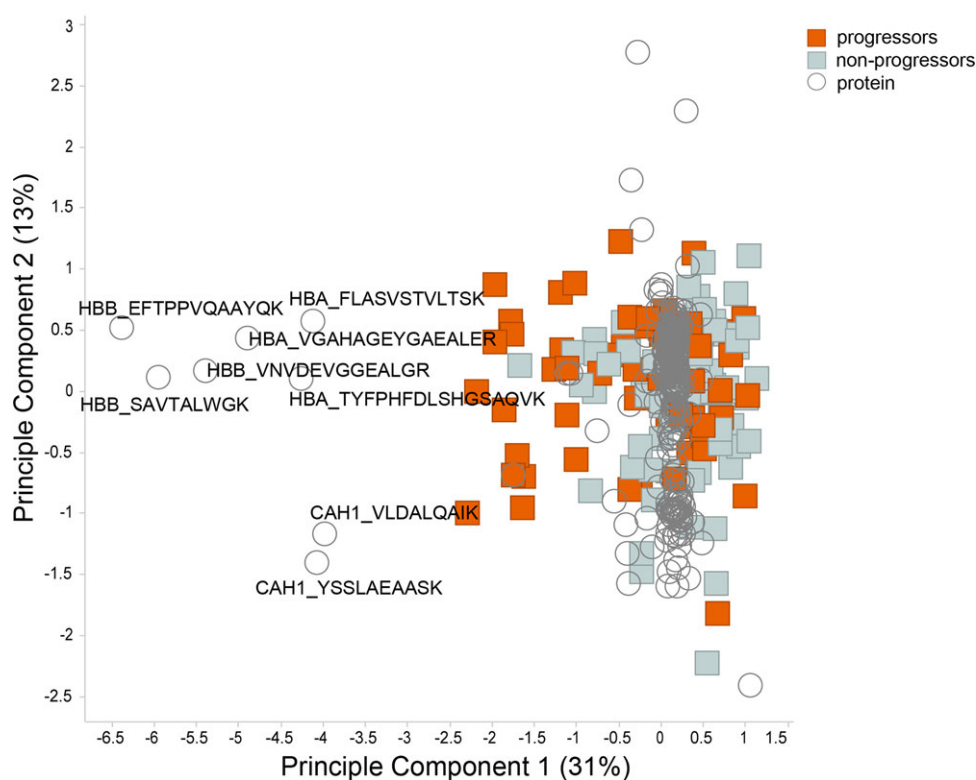
Figure 3. Centered PCA. Two first components of a centered PCA (spectral map) representing the 134 baseline samples identified in the MCI population. In this biplot, each square represents a sample colored by its progression status at 36 months (in orange, progressors; in gray, nonprogressors). The first component explains 31% of the variability between the samples, whereas the second component explains 13% of the variability between the samples. The peptides driving most of the variability between samples are represented as dots (top eight peptide names are represented).

the peptides selected in the multivariate analysis also showed significant differential expression in the univariate analyses. Table 5 shows the estimated AUC of panels involving only those peptides common to the two approaches. It appears from this analysis that it may be possible to achieve similar accuracy with a more modest panel size.

3.6 Biological insights from MRM assay findings

An important goal of the analysis conducted here, in addition to the qualification of candidate biomarkers in a well-characterized and curated clinical cohort, was to gain potential insights into the biology of CSF in AD and MCI progression. Many of the findings in this study confirm or replicate orthogonal measures (from either immunoassay-based or alternative MS-based approaches) in the ADNI cohort as well as other AD CSF collections. Each peptide identified as significant within our multiple approaches to data analysis and visualization potentially represents an interesting biological insight. Here we focus on a subset of findings that have gained particular interest and traction in the literature including, fatty acid binding proteins (Fabps), blood proteins in CSF, the granin family of neurosecretory proteins, and ApoE.

FABPH topped the list of peptides for both AD versus CN and MCI versus CN (Table 3) in univariate analysis, although they did not show statistical significance upon FDR

correction. It is interesting to note that elevated FABPH (also known as Fabp3) concentrations as measured with immunoassay-based platforms have been observed previously both in the current ADNI cohort as well as other US- and EU-based groups of patients diagnosed with AD ([43] and references included therein). The consistency and relative robustness of this finding across studies, even without the added confidence in the finding provided by the MRM results reported here (Supporting Information Fig. 6A), have generated the speculation that "central nervous system dyshomeostasis" may play a role in AD [43] providing some support for an earlier hypothesized role for altered CNS lipid biology [44]. Fabps function as cytoplasmic shuttle molecules that transport the long and short chain fatty acids that have more direct biological functions in the brain such as docosahexanoic and arachidonic acids (reviewed in [45]). At least ten genes encoding Fabps have been identified [46] with three Fabp family molecules, Fabp3, Fabp5, and Fabp7 localized in neural stem/progenitor cells, neurons, and/or glia. It is speculated (reviewed by [45]) that one or more of the three Fabps in the brain may be involved in hippocampal neurogenesis with any defect in their function contributing to hippocampal loss in AD. Indeed, Fabp3 concentration at baseline had been found to be associated with the subsequent rate of decline of entorhinal cortex size, which is correlated with hippocampal loss [43]. The correlation between Fabp3 concentrations reported here and the subsequent hippocampal rate of decline in each individual remains to be explored. The

Table 4. Peptides selected for panel to classify converters from MCI to AD

Peptide	Protein
ALQASALK	ALDOA
SVSLPSLDPASAK	APOB
VLNQELR	APOD
INENTGSVSVTR	CAD13
VSEADSSNADWVTK	CFAB
SFTLASSETGVGAPISGPGIPGR	CH3L1
VTIDSSYDIAK	CH3L1
ILGQQVPYATK	CH3L1
VIASNILGTGEPSGPSSK	CNTN2
IHWESASLLR	CO3
ESDTSYVSLK	CRP
SIVTLDGGK	FABPH
SLGVGFATR	FABPH
YLPFVPSR	FMOD
ESSQEQSSVVR	KLK6
TESTLNALLQR	NPTX2
ELDVLQGR	NPTXR
AIPVAQDLNAPSDWDSR	OSTP
SSFVAPLEK	PEDF
DISLSDYK	PRDX1
LVQAFQFTDK	PRDX1
GLFIIDGK	PRDX2
IGKPAPDFK	PRDX2
LSILYPATTGR	PRDX6
ITQVTWQK	PVRL1
GEAGAPGEEDIQGPTK	SCG1
AYQGVAAPFPK	VGF
THLGEALAPLSK	VGF
NSEPQDEGELFQGVDPR	VGF

Table 5. Estimated AUC of multivariate panels restricted to top univariate analytes

Panel	AUC
INENTGSVSVTR (CAD13), IHWESASLLR (CO3), ESDTSYVSLK (CRP), TESTLNALLQR (NPTX2), GLFIIDGK (PRDX2), IGKPAPDFK (PRDX2) and ITQVTWQK (PVRL1) APOE genotype	0.74
INENTGSVSVTR (CAD13), IHWESASLLR (CO3), ESDTSYVSLK (CRP), TESTLNALLQR (NPTX2), GLFIIDGK (PRDX2), IGKPAPDFK (PRDX2) and ITQVTWQK (PVRL1) APOE genotype + Aβ/p-tau + age + gender	0.76

interpretation of any such relationship, however, will at a minimum require longitudinal data on alterations of Fabp3 over time versus alterations in hippocampal volume for each individual studied. A recently described polymorphism in Fabp3 produces a frame-shift protein, FABP3 E132fs, which forms cellular aggregates and is unstable when expressed in cultured cells. Work from the same group includes the observation that Fabp3 knockout mice show decreased social memory and novelty seeking among the investigated behaviors [47]. Far too little is known, however, about what influences concentrations of CSF Fabp3 to know if changes

in its concentrations truly reflect alterations in functionally relevant lipid homeostasis in human brain.

It has been hypothesized for some time that the blood–brain barrier is leaky in AD, leading to entry into the brain of blood proteins and other molecules (reviewed in [48]). Interestingly, our data demonstrate that changes for abundant blood proteins HBA and HBB were only apparent in differentiating converters from nonconverters within the MCI patient set, and that we do not see significant differences between AD and CN or MCI and CN for these peptides (Supp frame-shift Fig. 6B and C). If the elevated hemoglobins reflected increased permeability of the blood–brain barrier and if this was a characteristic of a reasonable proportion of subjects with AD we should have observed an AD versus CN difference. During the CSF sample collection, erythrocyte counts were measured as a surrogate for blood contamination (defined as more than 500 red blood cells/μL [49]). In our cohort, 13 CSF samples demonstrated some blood contamination, including six samples (two progressors) of the 134 samples from the MCI cohort. It is unlikely that our findings relate to blood contamination as there was no clear correlation between HBA and HBB levels and erythrocyte counts (Supporting Information Fig. 7).

Whatever the source of the hemoglobin findings, it was observed in both univariate analysis of the data, where peptides from the hemoglobins are significantly associated with the diagnosis for converters versus nonconverters (Table 3), and unsupervised multivariate analysis where separation of converting versus nonconverting subjects corresponded to elevated values of the same peptides (Fig. 2). The current study cannot address whether these changes are specific to HBB and HBA, or reflect general change in proteins of blood origin in CSF, given that other abundant blood proteins were depleted from our samples (MARS-14 immunoaffinity resin removes albumin, haptoglobin, transferrin, IgG, IgA, a1-antitrypsin, a2-macroglobulin, a1-acid glycoprotein, apolipoprotein AI, apolipoprotein aII, complement C3, IgM, transthyretin, and fibrinogen). Additional analysis, and possibly experiments will be needed to understand whether this reflects artifacts of the experiment or some underlying biological phenomenon. Longitudinal CSF samples in the MCI group may prove especially helpful in understanding these unexpected hemoglobin peptide findings.

A growing body of evidence, with several recent publications from a number of groups, have suggested the neurosecretory granins and other dense core protein family members, including VGF and SCG2, in both their intact and proteolytic forms, serve as putative markers of synaptic loss and neuronal injury/degeneration [50–54]. Our results demonstrate significant decreases for multiple peptides from these two proteins between converting and nonconverting MCI patients. These proteins are typically transported in synaptic vesicles and in an activity-dependent fashion [55]. Decreases in CSF concentrations for these analytes may point toward alterations in vesicle maturation and transport during AD progression. Indeed, a reduction has been
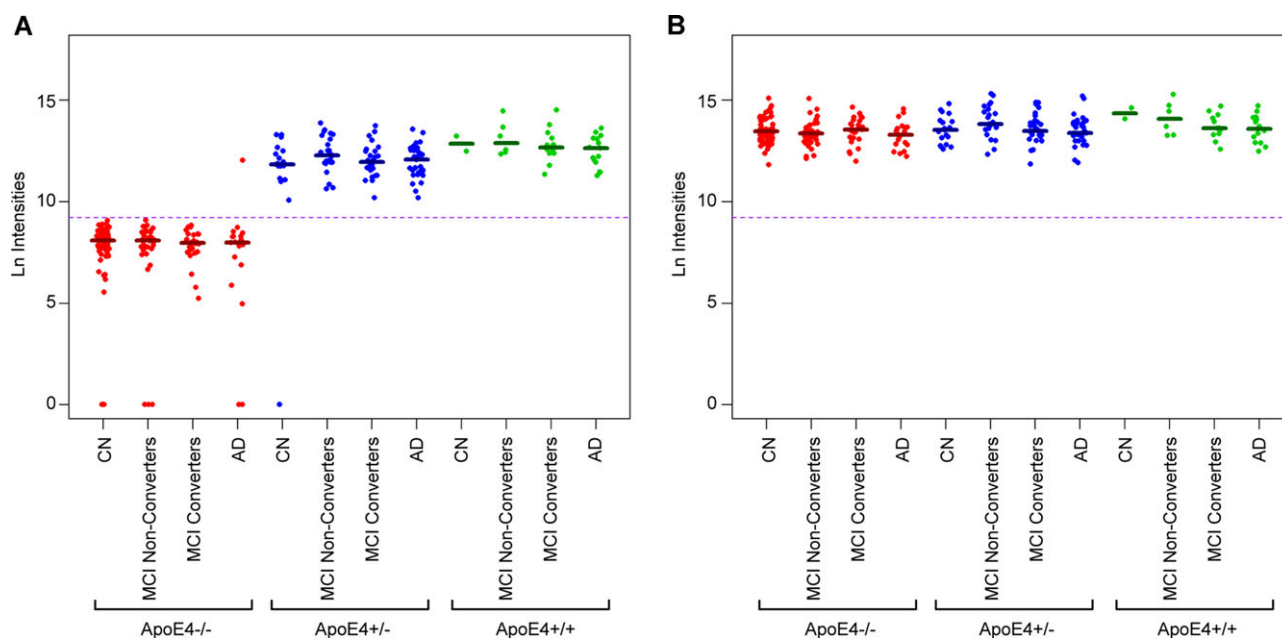
Figure 4. ApoE peptide levels do not differ by disease state in CSF. (A) Levels of ApoE4 peptide (LGADMEDVR) are below the LOQ (dashed line) in noncarriers. (B) Total levels of common ApoE peptide (LGPLVEQGR) are the same independent of genotype (color by ApoE4 allele: red, ApoE4$^{-/-}$; blue, ApoE4$^{+/-}$; green, ApoE4$^{+/+}$; group by disease state: CN, normal; MCI, mild cognitive impairment; AD, Alzheimer's disease; solid black line represents the group median).

observed in VGF peptides studied in the parietal cortex of AD patients postmortem suggesting involvement in the physiological or pathophysiological mechanisms occurring in patients with AD [56]. By longitudinal magnetic resonance imaging studies in subjects who progressed to AD within 4–5 years, the parietal cortex exhibited a high rate of atrophy that correlated with changes in clinical severity and decline in cognition [57]. Furthermore, in the parietal cortex of patients with AD, a significant deficit in proBDNF protein was also found [58], whereas an overlap in VGF and *trk* mRNA expression has been reported in the rat CNS [59]. Hence, it can be hypothesized that a decrease in VGF may be correlated with a loss of neuronal products such as BDNF, resulting in the failure of neuronal protective functions.

Due to the strong risk associated with ApoE *ε*4 allele and AD, multiple ApoE peptides were included in the assay. This included an ApoE4-specific peptide (LGADMEDVR), which was below the LOQ in ApoE4-negative patients, as expected (Fig. 4A). There was not a significant difference between levels of ApoE between normal, MCI, and AD patients in CSF, consistent with what others have observed [60–62]. This observation was consistent for the five ApoE peptides that were quantitated, and two of the peptides are shown in Fig. 4. However, unlike Cruchaga et al. who observed a difference in total ApoE levels depending on ApoE4 status (0 > 1 > 2 alleles), our results using a peptide common to all the isoforms indicate that there is no relationship between ApoE4 status and CSF ApoE concentration (Fig. 4B). This result is supported by another group who used the isoform-specific

peptides for quantitation with MS [62]. The differences in the literature are most likely due to the difference in the method of quantitation. The observed relationship between genotype and ApoE levels in CSF as measured by an immunoassay could simply be due to a bias of the antibody for one isoform over another.

4   Concluding remarks

We describe the outcome of the Biomarkers Consortium CSF Proteomics Project, a public–private partnership of government, academia, nonprofit, and industry. This project aimed to begin to address the large gap that remains in the qualification of candidate predictive, prognostic, and pharmacodynamic biomarkers for AD. To this end, we carried out an analysis of a large number of candidate markers on what is widely recognized as the most well-characterized cohort of patient samples in AD. The level of rigor applied to the data processing and analysis, in our opinion, has rarely, if ever, been applied to this type of dataset in the area of Biomarker research. The data from this study provide important information about the biology of CSF in health and disease, and identify potential biomarkers for follow-up in other studies. In addition, the assay constructed and described here provides researchers with a tool from further analysis of independent clinical cohorts. It should be mentioned that several outstanding issues remain to be addressed, including but not limited to, the need for follow-up analysis in longitudinal samples as

many candidate markers present here may not differentiate a baseline but could when assessing their change over time, additional development including heavy isotope peptide standards for most promising analytes to provide absolute abundance measures as opposed to relative, additional correlation analysis of orthogonal datasets available for these samples (i.e., correlation with RBM, BACE activity, and many other datasets available publically through the ADNI website), additional analysis of disease-relevant demographic data such as racial background or body mass index, and the application of the panel to other neurological and neurodegenerative diseases such as Parkinson's, FTD, PSP, schizophrenia, and depression to determine the specificity of the markers evaluated.

5 References

[1] D'Aoust, R. F., Brewster, G., Rowe, M. A., Depression in informal caregivers of persons with dementia. *Int. J. Older People Nurs.* 2015, *10*, 14–26.

[2] Shuter, P., Beattie, E., Edwards, H., An exploratory study of grief and health-related quality of life for caregivers of people with dementia. *Am. J. Alzheimers Dis. Other Demen.* 2013, *29*, 379–385.

[3] Brookmeyer, R., Evans, D. A., Hebert, L., Langa, K. M. et al., National estimates of the prevalence of Alzheimer's disease in the United States. *Alzheimers Dement.* 2011, *7*, 61–73.

[4] Fagan, A. M., Perrin, R. J., Upcoming candidate cerebrospinal fluid biomarkers of Alzheimer's disease. *Biomark. Med.* 2012, *6*, 455–476.

[5] McKhann, G., Drachman, D., Folstein, M., Katzman, R. 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* 1984, *34*, 939–944.

[6] Beach, T. G., Monsell, S. E., Phillips, L. E., Kukull, W., Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *J. Neuropathol. Exp. Neurol.* 2012, *71*, 266–273.

[7] Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J. et al., Mild cognitive impairment: clinical characterization and outcome. *Arch. Neurol.* 1999, *56*, 303–308.

[8] Price, J. L., Ko, A. I., Wade, M. J., Tsou, S. K. et al., Neuron number in the entorhinal cortex and CA1 in preclinical Alzheimer disease. *Arch. Neurol.* 2001, *58*, 1395–1402.

[9] Blennow, K., Hampel, H., Weiner, M., Zetterberg, H., Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nat. Rev. Neurol.* 2010, *6*, 131–144.

[10] Sunderland, T., Linker, G., Mirza, N., Putnam, K. T. et al., Decreased beta-amyloid1-42 and increased tau levels in cerebrospinal fluid of patients with Alzheimer disease. *JAMA* 2003, *289*, 2094–2103.

[11] Clark, C. M., Xie, S., Chittams, J., Ewbank, D. et al., Cerebrospinal fluid tau and beta-amyloid: how well do these biomarkers reflect autopsy-confirmed dementia diagnoses? *Arch. Neurol.* 2003, *60*, 1696–1702.

[12] Fagan, A. M., Mintun, M. A., Mach, R. H., Lee, S. Y. et al., Inverse relation between in vivo amyloid imaging load and cerebrospinal fluid Abeta42 in humans. *Ann. Neurol.* 2006, *59*, 512–519.

[13] Forsberg, A., Engler, H., Almkvist, O., Blomquist, G. et al., PET imaging of amyloid deposition in patients with mild cognitive impairment. *Neurobiol. Aging* 2008, *29*, 1456–1465.

[14] Tolboom, N., van der Flier, W. M., Yaqub, M., Boellaard, R. et al., Relationship of cerebrospinal fluid markers to 11C-PiB and 18F-FDDNP binding. *J. Nucl. Med.* 2009, *50*, 1464–1470.

[15] Grimmer, T., Riemenschneider, M., Forstl, H., Henriksen, G. et al., Beta amyloid in Alzheimer's disease: increased deposition in brain is reflected in reduced concentration in cerebrospinal fluid. *Biol. Psychiatry* 2009, *65*, 927–934.

[16] Jagust, W. J., Landau, S. M., Shaw, L. M., Trojanowski, J. Q. et al., Relationships between biomarkers in aging and dementia. *Neurology* 2009, *73*, 1193–1199.

[17] Buerger, K., Ewers, M., Pirttila, T., Zinkowski, R. et al., CSF phosphorylated tau protein correlates with neocortical neurofibrillary pathology in Alzheimer's disease. *Brain* 2006, *129*(Pt 11), 3035–3041.

[18] Blom, E. S., Giedraitis, V., Zetterberg, H., Fukumoto, H. et al., Rapid progression from mild cognitive impairment to Alzheimer's disease in subjects with elevated levels of tau in cerebrospinal fluid and the APOE epsilon4/epsilon4 genotype. *Dement. Geriatr. Cogn. Disord.* 2009, *27*, 458–464.

[19] Hansson, O., Zetterberg, H., Buchhave, P., Londos, E. et al., Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. *Lancet Neurol.* 2006, *5*, 228–234.

[20] Fagan, A. M., Roe, C. M., Xiong, C., Mintun, M. A. et al., Cerebrospinal fluid tau/beta-amyloid(42) ratio as a prediction of cognitive decline in nondemented older adults. *Arch. Neurol.* 2007, *64*, 343–349.

[21] Li, G., Sokal, I., Quinn, J. F., Leverenz, J. B. et al., CSF tau/Abeta42 ratio for increased risk of mild cognitive impairment: a follow-up study. *Neurology* 2007, *69*, 631–639.

[22] Hansson, O., Zetterberg, H., Buchhave, P., Londos, E. et al., Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study. *Lancet Neurol.* 2006, *5*, 228–234.

[23] Snider, B. J., Fagan, A. M., Roe, C., Shah, A. R. et al., Cerebrospinal fluid biomarkers and rate of cognitive decline in very mild dementia of the Alzheimer type. *Arch. Neurol.* 2009, *66*, 638–645.

[24] Landau, S. M., Harvey, D., Madison, C. M., Reiman, E. M. et al., Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* 2010, *75*, 230–238.

[25] Fagan, A. M., Perrin, R. J., Upcoming candidate cerebrospinal fluid biomarkers of Alzheimer's disease. *Biomark. Med.* 2012, *6*, 455–476.

[26] Marquette, C. A., Corgier, B. P., Blum, L. J., Recent advances in multiplex immunoassays. *Bioanalysis* 2012, *4*, 927–936.

[27] Chandra, H., Reddy, P. J., Srivastava, S., Protein microarrays and novel detection platforms. *Expert Rev. Proteomics* 2011, *8*, 61–79.

[28] Gold, L., Ayers, D., Bertino, J., Bock, C. et al., Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One* 2010, *5*, e15004.

[29] Gillette, M. A., Carr, S. A., Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat. Methods* 2013, *10*, 28–34.

[30] Rifai, N., Gillette, M. A., Carr, S. A., Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* 2006, *24*, 971–983.

[31] Lawson, A. M., The scope of mass spectrometry in clinical chemistry. *Clin. Chem.* 1975, *21*, 803–824.

[32] Grebe, S. K., Singh, R. J., LC-MS/MS in the clinical laboratory—where to from here? *Clin. Biochem. Rev.* 2011, *32*, 5–31.

[33] Yost, R. A., Enke, C. G., Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Anal. Chem.* 1979, *51*, 1251–1264.

[34] Lange, V., Picotti, P., Domon, B., Aebersold, R., Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 2008, *4*, 222.

[35] Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J. et al., Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* 2009, *27*, 633–641.

[36] Kennedy, J. J., Abbatiello, S. E., Kim, K., Yan, P. et al., Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. *Nat. Methods* 2014, *11*, 149–155.

[37] Percy, A. J., Chambers, A. G., Yang, J., Domanski, D., Borchers, C. H., Comparison of standard- and nano-flow liquid chromatography platforms for MRM-based quantitation of putative plasma biomarker proteins. *Anal. Bioanal. Chem.* 2012, *404*, 1089–1101.

[38] Paweletz, C. P., Wiener, M. C., Bondarenko, A. Y., Yates, N. A. et al., Application of an end-to-end biomarker discovery platform to identify target engagement markers in cerebrospinal fluid by high resolution differential mass spectrometry. *J. Proteome Res.* 2010, *9*, 1392–1401.

[39] Samtani, M. N., Raghavan, N., Shi, Y., Novak, G. et al., Disease progression model in subjects with mild cognitive impairment from the Alzheimer's disease neuroimaging initiative: CSF biomarkers predict population subtypes. *Br. J. Clin. Pharmacol.* 2013, *75*, 146–161.

[40] Lewi, P. J., Spectral mapping, a technique for classifying biological activity profiles of chemical compounds. *Arzneimittelforschung* 1976, *26*, 1295–1300.

[41] Smyth, G. K., Yang, Y. H., Speed, T., Statistical issues in cDNA microarray data analysis. *Methods Mol. Biol.* 2003, *224*, 111–136.

[42] Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B. et al., Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004, *5*, R80.

[43] Desikan, R. S., Thompson, W. K., Holland, D., Hess, C. P. et al., Heart fatty acid binding protein and Abeta-associated

Alzheimer's neurodegeneration. *Mol. Neurodegener.* 2013, *8*, 39.

[44] Di Paolo, G., Kim, T. W., Linking lipids to Alzheimer's disease: cholesterol and beyond. *Nat. Rev. Neurosci.* 2011, *12*, 284–296.

[45] Matsumata, M., Inada, H., Osumi, N., Fatty acid binding proteins and the nervous system: their impact on mental conditions. *Neurosci. Res.* 2014, pii: S0168-0102(14)00188-6.

[46] Smathers, R. L., Petersen, D. R., The human fatty acid-binding protein family: evolutionary divergences and functions. *Hum. Genomics* 2011, *5*, 170–191.

[47] Shimamoto, C., Ohnishi, T., Maekawa, M., Watanabe, A. et al., Functional characterization of FABP3, 5 and 7 gene variants identified in schizophrenia and autism spectrum disorder and mouse behavioral studies. *Hum. Mol. Genet.* 2014, *23*, 6495–6511.

[48] Erickson, M. A., Banks, W. A., Blood-brain barrier dysfunction as a cause and consequence of Alzheimer's disease. *J. Cereb. Blood Flow Metab.* 2013, *33*, 1500–1513.

[49] Teunissen, C. E., Tumani, H., Engelborghs, S., Mollenhauer, B., Biobanking of CSF: international standardization to optimize biomarker development. *Clin. Biochem.* 2014, *47*, 288–292.

[50] Perrin, R. J., Craig-Schapiro, R., Malone, J. P., Shah, A. R. et al., Identification and validation of novel cerebrospinal fluid biomarkers for staging early Alzheimer's disease. *PLoS One* 2011, *6*, e16032.

[51] Simonsen, A. H., McGuire, J., Hansson, O., Zetterberg, H. et al., Novel panel of cerebrospinal fluid biomarkers for the prediction of progression to Alzheimer dementia in patients with mild cognitive impairment. *Arch. Neurol.* 2007, *64*, 366–370.

[52] Carrette, O., Demalte, I., Scherl, A., Yalkinoglu, O. et al., A panel of cerebrospinal fluid potential biomarkers for the diagnosis of Alzheimer's disease. *Proteomics* 2003, *3*, 1486–1494.

[53] Jahn, H., Wittke, S., Zurbig, P., Raedler, T. J. et al., Peptide fingerprinting of Alzheimer's disease in cerebrospinal fluid: identification and prospective evaluation of new synaptic biomarkers. *PLoS One* 2011, *6*, e26540.

[54] Wijte, D., McDonnell, L. A., Balog, C. I., Bossers, K. et al., A novel peptidomics approach to detect markers of Alzheimer's disease in cerebrospinal fluid. *Methods* 2012, *56*, 500–507.

[55] Levi, A., Ferri, G. L., Watson, E., Possenti, R., Salton, S. R., Processing, distribution, and function of VGF, a neuronal and endocrine peptide precursor. *Cell Mol. Neurobiol.* 2004, *24*, 517–533.

[56] Cocco, C., D'Amato, F., Noli, B., Ledda, A. et al., Distribution of VGF peptides in the human cortex and their selective changes in Parkinson's and Alzheimer's diseases. *J. Anat.* 2010, *217*, 683–693.

[57] Foundas, A. L., Leonard, C. M., Mahoney, S. M., Agee, O. F., Heilman, K. M., Atrophy of the hippocampus, parietal cortex, and insula in Alzheimer's disease: a volumetric magnetic resonance imaging study. *Neuropsychiatry Neuropsychol. Behav. Neurol.* 1997, *10*, 81–89.

[58] Fahnestock, M., Garzon, D., Holsinger, R. M., Michalski, B., Neurotrophic factors and Alzheimer's disease: are we focusing on the wrong molecule? *J. Neural Transm. Suppl.* 2002, *62*, 241–252.

[59] Snyder, S. E., Li, J., Salton, S. R., Comparison of VGF and trk mRNA distributions in the developing and adult rat nervous systems. *Brain Res. Mol. Brain Res.* 1997, *49*, 307–311.

[60] Wahrle, S. E., Shah, A. R., Fagan, A. M., Smemo, S. et al., Apolipoprotein E levels in cerebrospinal fluid and the effects of ABCA1 polymorphisms. *Mol. Neurodegener.* 2007, *2*, 7.

[61] Cruchaga, C., Kauwe, J. S., Nowotny, P., Bales, K. et al., Cerebrospinal fluid APOE levels: an endophenotype for genetic studies for Alzheimer's disease. *Hum. Mol. Genet.* 2012, *21*, 4558–4571.

[62] Martinez-Morillo, E., Hansson, O., Atagi, Y., Bu, G. et al., Total apolipoprotein E levels and specific isoform composition in cerebrospinal fluid and plasma from Alzheimer's disease patients and controls. *Acta Neuropathol.* 2014, *127*, 633–643.