Featured Article

Association of blood lipids with Alzheimer's disease: A comprehensive lipidomics analysis

Petroula Proitsi[a,*,1], Min Kim[b,1], Luke Whiley[b,**], Andrew Simmons[a,c], Martina Sattlecker[a,c], Latha Velayudhan[a], Michelle K. Lupton[d], Hillka Soininen[e], Iwona Kloszewska[f], Patrizia Mecocci[g], Magda Tsolaki[h], Bruno Vellas[i], Simon Lovestone[j], John F. Powell[a,2], Richard J. B. Dobson[a,c,k,2], Cristina Legido-Quigley[b,2]

[a]King's College London, Institute of Psychiatry, Psychology & Neuroscience, London, UK
[b]King's College London, Institute of Pharmaceutical Science, London, UK
[c]NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at South London and Maudsley NHS Foundation Trust, London UK
[d]QIMR Berghofer Medical Research Institute, Brisbane, Australia
[e]Department of Neurology, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland
[f]Department of Old Age Psychiatry & Psychotic Disorders, Medical University of Lodz, Lodz, Poland
[g]Section of Gerontology and Geriatrics, Department of Medicine, University of Perugia, Perugia, Italy
[h]Memory and Dementia Centre, Aristotle University of Thessaloniki, Thessaloniki, Greece
[i]Department of Internal and Geriatrics Medicine, INSERM U 1027, Gerontopole, Hôpitaux de Toulouse, Toulouse, France
[j]Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK
[k]The Farr Institute of Health Informatics Research, UCL Institute of Health Informatics, UCL, UK

Abstract

Introduction: The aim of this study was to (1) replicate previous associations between six blood lipids and Alzheimer's disease (AD) (Proitsi et al 2015) and (2) identify novel associations between lipids, clinical AD diagnosis, disease progression and brain atrophy (left/right hippocampus/entorhinal cortex).

Methods: We performed untargeted lipidomic analysis on 148 AD and 152 elderly control plasma samples and used univariate and multivariate analysis methods.

Results: We replicated our previous lipids associations and reported novel associations between lipids molecules and all phenotypes. A combination of 24 molecules classified AD patients with $>70\%$ accuracy in a test and a validation data set, and we identified lipid signatures that predicted disease progression ($R^2 = 0.10$, test data set) and brain atrophy ($R^2 \geq 0.14$, all test data sets except left entorhinal cortex). We putatively identified a number of metabolic features including cholesteryl esters/triglycerides and phosphatidylcholines.

Discussion: Blood lipids are promising AD biomarkers that may lead to new treatment strategies.

© 2016 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Alzheimer's disease; Dementia; Brain atrophy; sMRI; Rate of cognitive decline; Lipidomics; Metabolomics; Biomarkers; Machine learning; Multivariate; Classification; Random forest

1. Introduction

Alzheimer's disease (AD) is a devastating illness and one of the major public health challenges of the 21st century. The lack of effective treatments and early diagnosis highlights the importance of the identification of noninvasive biomarkers, for early diagnosis and disease progression. Blood metabolites have recently emerged

as promising AD biomarkers [1–4]. They are small molecules which could theoretically cross the already compromised AD blood-brain barrier [5]; they are easily accessible, and they represent an essential aspect of the phenotype of an organism and a molecular "fingerprint" of disease progression [6,7]. They can therefore aid early diagnosis, recruitment into trials and may help identify new therapeutic targets.

A number of blood metabolomic studies have highlighted the role of lipid compounds, such as phosphatidylcholines (PCs) in AD [1–4]. We previously identified three PCs that were diminished in mild cognitive impairment (MCI) individuals and AD patients [4] and were further associated with poorer memory performance and decreased brain function during aging [8]. We further performed lipidomics analysis and identified 10 metabolites that predicted AD in an unseen test data set with 79% accuracy [9]; six analytes were putatively identified as cholesteryl esters (ChoEs), molecules related to PCs, and were reduced in MCI and AD.

Here, we performed lipidomics analysis in a sample of 142 AD patients and 135 healthy controls aiming to (1) replicate our previous associations [9] and (2) discover new lipids and combinations of lipids associated with clinical AD diagnosis and AD endophenotypes, such as the rate of cognitive decline and brain atrophy measures. This is to our knowledge the most comprehensive blood lipidomics study to date to identify lipid signatures associated with AD and AD endophenotypes, improving our current knowledge of molecules associated with AD.

2. Methods

2.1. Patient sample collection

This study used 148 AD patients and 152 controls from the Dementia Case Register at King's College London and the EU-funded AddNeuroMed study [10]. All individuals with AD patients met criteria for either probable (NINCDS-ADRDA, DSM-IV) or definite (CERAD) AD. All nonpopulation individuals who were controls were screened for dementia using the MMSE or ADAS-cog or were determined to be free from dementia at neuropathologic examination or had a Braak score ≤ 2.5. Diagnosis was confirmed by pathologic examination for a proportion of cases and cognitively normal elderly controls. All AD cases had an age of onset ≥ 60 years, and controls were ≥ 60 years at examination. A total of 102 AD cases and 104 controls had HDL-c, LDL-c, TC, and TG serum levels (mmol/L) available. Nonoverlapping individuals from these cohorts have been previously reported [9]. Each individual was required to fast for 2 hours before sample collection, and 10 mL of blood was collected in tubes coated with sodium ethylenediaminetetraacetic acid to prevent clotting. Whole blood was centrifuged at 2000 g for 10 minutes at 4°C to separate plasma, which was removed and stored at −80°C. All samples were centrifuged within approximately 2 hours of collection.

2.2. Lipidomics

Sample treatment has been described elsewhere [4,9,11] and is explained in detail in Supplementary Methods 1. Briefly, 20 µL of plasma was added to a glass HPLC vial containing a 400-µl glass insert (Chromacol, UK). Ten microliters of high purity water and 40 µL of MS grade methanol were added to each sample, followed by a 2-minute vortex mix to precipitate proteins; 200 µL of Methyl tert-Butyl Ether (MTBE) containing 10 µg/mL of internal standard Tripentadecanoin (TG45:0) was added, and the samples were mixed via vortex at room temperature for 1 hour. After addition of 50 µL of high purity water, a final sample mixing was performed before centrifugation at 3000 g for 10 minutes. The upper, lipid-containing, MTBE phase was then injected onto the LC-MS system directly from the vial by adjustment of the instrument needle height (17.5 mm from bottom).

Lipidomics was performed by a Waters ACQUITY UPLC and XEVO QTOF system. The method has previously been published [4,12] and has been shown to quantitate >4500 metabolite species (Supplementary Methods 1). Samples were analyzed in a randomized order, in four batches, with pooled plasma sampled (QC) at regular intervals throughout the run (n = 30 for both positive and negative ionization). Features were extracted from netCDF files using the R package "XCMS" [13] which performed filtration, peak identification, matching of peaks across samples, and retention time correction. Positive and negative ionization mode data were extracted separately and quantile normalized.

2.3. Structural magnetic resonance imaging

Volumes of whole brain and the hippocampi and entorhinal cortices were obtained using FreeSurfer 5.1.0 from 123 subjects (53 AD patients and 70 Controls) who had undergone sMRI. Regions were normalized by intracranial volume [14]. The volumetric data were not used to aid in the clinical diagnosis of AD. Detailed information regarding data acquisition, pre-processing, and quality control assessment has been described elsewhere [15,16]. Before analyses, sMRI measures were standardized to have a mean of 0 and a standard deviation (SD) of 1.

2.4. Calculation of rate of cognitive decline

The ROD was available for 118 AD patients with analyte data and has been described elsewhere [17]. The ROD was based on longitudinal mini mental state examination (MMSE) assessments [18], and only samples with at least three MMSE measures were included in the calculation using linear mixed effect models. After covariate adjustment

[17], the slope coefficient for each sample was used as the ROD defined as the change in MMSE per day.

2.5. Statistical analysis

2.5.1. Quality control

Data QC has been previously described [9] and included filtering of features and individuals, data transformation, batch effect correction, outlier detection, and imputation (Supplementary Methods 2.1 and Supplementary Fig. 1). All analyses took place in R.3.01.

2.5.2. Single-analyte statistical analysis

Logistic regression investigated the association of each metabolite with clinical AD diagnosis and linear regression the association with cognitive decline and sMRI measures. Logistic regression and linear regression models for the sMRI measures were adjusted for age at sampling, gender, presence of the apolipoprotein E (*APOE*) ε4 allele, batch, and study site. For the ROD models, covariate adjustment was only applied for batch as the rest of the covariates were included in the ROD calculation [17]. sMRI measurements were not adjusted for diagnosis to allow identification of features associated with brain atrophy caused by AD. False discovery rate (FDR) correction (0.05) was applied to correct for multiple testing ("fdrtool"). Secondary models investigated whether any associations were modified by the *APOE* ε4 allele or by gender.

Logistic regression results (summary statistics) for the positive ionization metabolites were combined with the results (summary statistics) from the Proitsi et al data set [9] using in-verse variance weighted fixed effect meta-analysis ("metafor"). The published data set [9] was restricted to 576 features extracted using Mass-Lynx, and therefore, the analysis presented here includes a large number of previously unreported molecules extracted using XCMS. All associations are reported as the change per one metabolite standard deviation (SD).

2.5.3. Multivariate statistical analysis

A random forest (RF) classifier approach (using "rf" and "rfe" in "CARET") was used to develop a clinical diagnosis classifier as previously described [9] (Supplementary Methods 2.2). Briefly, AD cases and controls were divided into a training data set (2/3 of the sample) matched for age, gender, and site and an independent data set (rest 1/3 of the sample). An RF model was built on the training data set (100 bootstraps), and in each iteration, each variable was assigned a variable importance (VI) score. The summed VI ranks provided an indication of the predictive power for each variable, and the top 10% molecules were selected for RF with recursive feature elimination (rfe; 100 bootstraps) from 250 down to two features. For each subset of predictors, the mean bootstrap testing performance was calculated, and the optimal number of variables was identified using "sizeTolerance" that picks a subset of variables that is small without sacrificing too much performance. Subsets of variables within 2.5% and 5% of the optimum performance were examined and used to build final models in the complete training data, which were tested on the test set. The final model was also tested in the Proitsi et al data set [9] which was used as a validation data set, after excluding metabolites in the negative ionization mode. The area under curve (AUC) was used to test the performance of each classifier. Receiver operator curves (ROCs)

Table 1
Sample demographics

	AD (N = 142)	Controls (N = 135)	Difference between AD patients and controls*
Age, mean (SD)	77 (6.5)	74 (5.9)	$t = -4.8$ (270), P value $= 2.62 \times 10^{-6}$
Gender (males/females)	48/87	47/95	$\chi^2 = 0.09$ (1), P value $= .761$
APOE ε4 allele (absence/presence)	54/81	99/43	$\chi^2 = 23.53$ (1), P value $= 1.23 \times 10^{-6}$
MMSE, mean (SD; Range)	20.1 (4.6; 10–27)	29.2 (0.9; 27–30)	$t = 22.58$ (143), P value $< 2.0 \times 10^{-16}$
ROD (per year), mean (SD)[†]	−1.46 (1.26)	NA	NA
Entorhinal cortex right, mean (SD)[‡§]	0.00092 (0.0003)	0.0013 (0.0003)	$t = 6.02$ (99), P value $= 2.87 \times 10^{-8}$
Entorhinal cortex left, mean (SD)[‡§]	0.00094 (0.0002)	0.0013 (0.0004)	$t = 5.26$ (86), P value $= 1.58 \times 10^{-6}$
Hippocampus right, mean (SD)[‡§]	0.0019 (0.0004)	0.0025 (0.0003)	$t = 9.06$ (100), P value $= 1.3 \times 10^{-14}$
Hippocampus left, mean (SD)[‡§]	0.0018 (0.0004)	0.0025 (0.003)	$t = 10.57$ (104), P value $< 2.0 \times 10^{-16}$
Mean HDL-c (SD), mmol/L[‖]	1.58 (0.37)	1.55 (0.38)	$\beta = 0.109$ (SE = 0.33), P value $= .068$
Mean LDL-c (SD), mmol/L[‖]	3.42 (1.01)	3.07 (0.82)	$\beta = 0.092$ (SE = 0.15), P value $= .529$
Mean TC (SD), mmol/L[‖]	5.69 (1.17)	5.29 (1.01)	$\beta = 0.209$ (SE = 0.173), P value $= .229$
Mean TG (SD), mmol/L[‖]	1.64 (1.04)	1.52 (0.67)	$\beta = 0.021$ (SE = 0.146), P value $= .885$
Statins (yes/no)	38/97	34/108	$\chi^2 = 0.436$ (1), P value $= .509$

Abbreviations: AD, Alzheimer's disease; MMSE, mini-mental state examination score; ROD, rate of cognitive decline; SD, standard deviation.

*Differences in the means/frequencies of clinical/demographic variables were tested using t test t(df), x2(df) test, or linear regression analyses after adjusting for age, gender, the *APOE* ε4 allele, and study site.

[†]Rate of decline data was available for a subset of AD patients (N = 118).

[‡]sMRI data were available for a subset of study participants (N = 123 [AD = 53, controls = 70]).

[§]Normalized to intracranial volume.

[‖]Serum HDL cholesterol, LDL cholesterol, total cholesterol, and triglyceride levels were available for a subset of study participants (N = 208 [AD = 102, Controls = 106]).
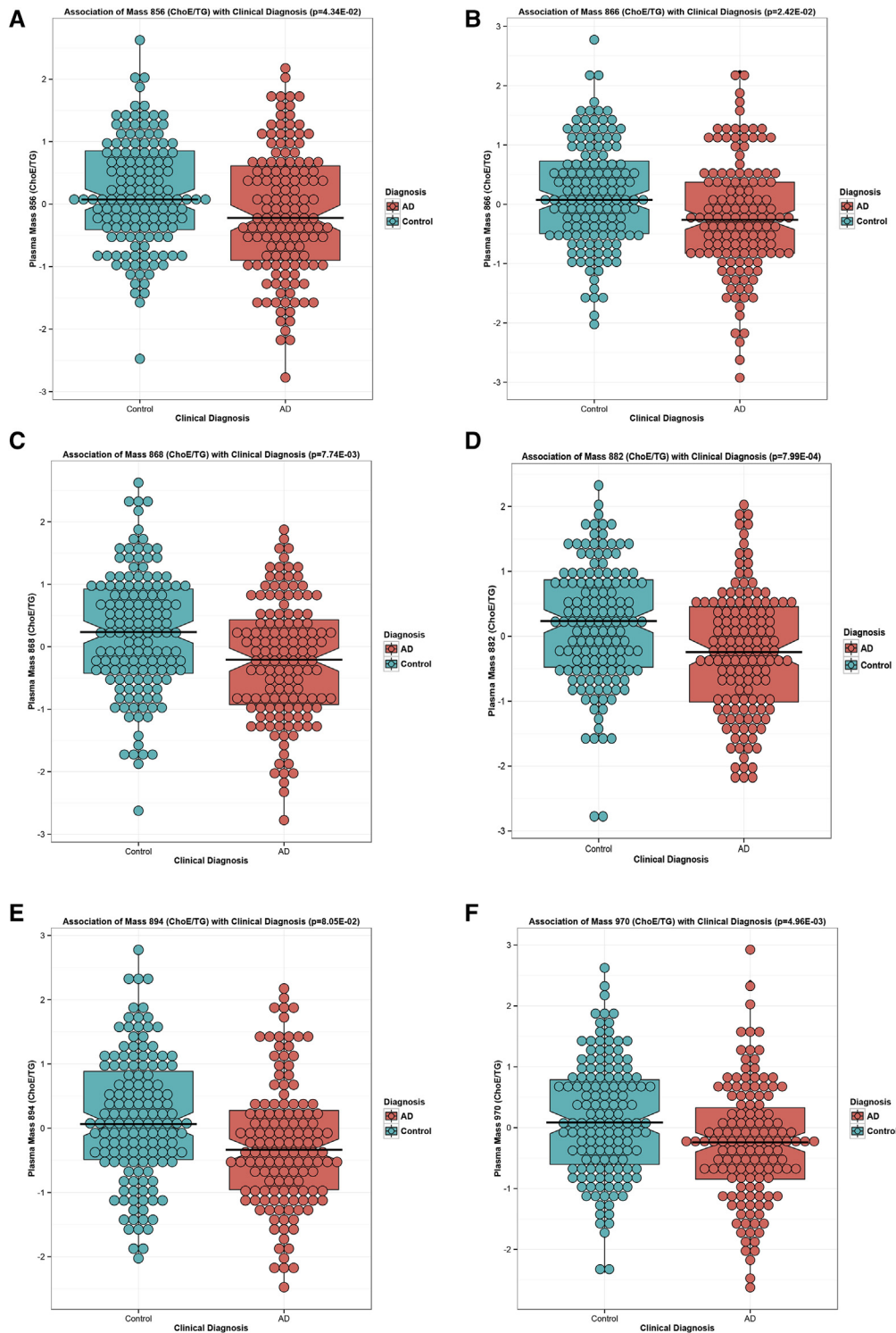
Fig. 1. Associations of previously reported molecules (Proitsi et al 2015) with clinical AD diagnosis in the current data set and associations of putatively annotated molecules, selected through random forest analyses, with the respective phenotype. (A) Association of Mass 856 with clinical AD diagnosis; (B) Association of Mass 866 with clinical AD diagnosis; (C) Association of Mass 868 with clinical AD diagnosis; (D) Association of Mass 882 with clinical AD diagnosis; (E) Association of Mass 894 with clinical AD diagnosis; (F) Association of Mass 970 with clinical AD diagnosis; (G) Association of Mass 882 (−) (PC 40:4) with clinical AD diagnosis; (H) Association of Mass 948 (+) TG (57:1) with clinical AD diagnosis; (I) Association of Mass 919 (+) TG 50:2 with Hippocampus (Right); (J) Association of Mass 943 (+) (ChoE/TG) with Hippocampus Left; (K) Association of Mass 367 (sterol) with Entorhinal Cortex (Right); (L) Association of Mass 816 (+) with Entorhinal Cortex (Left); (M) Association of Mass 771 (+) PC 36:3 with the rate of cognitive decline (ROD). The *P* values displayed are for the univariate regressions after adjusting for covariates. All molecules are scaled to have a mean of 0 and a standard deviation of 1.
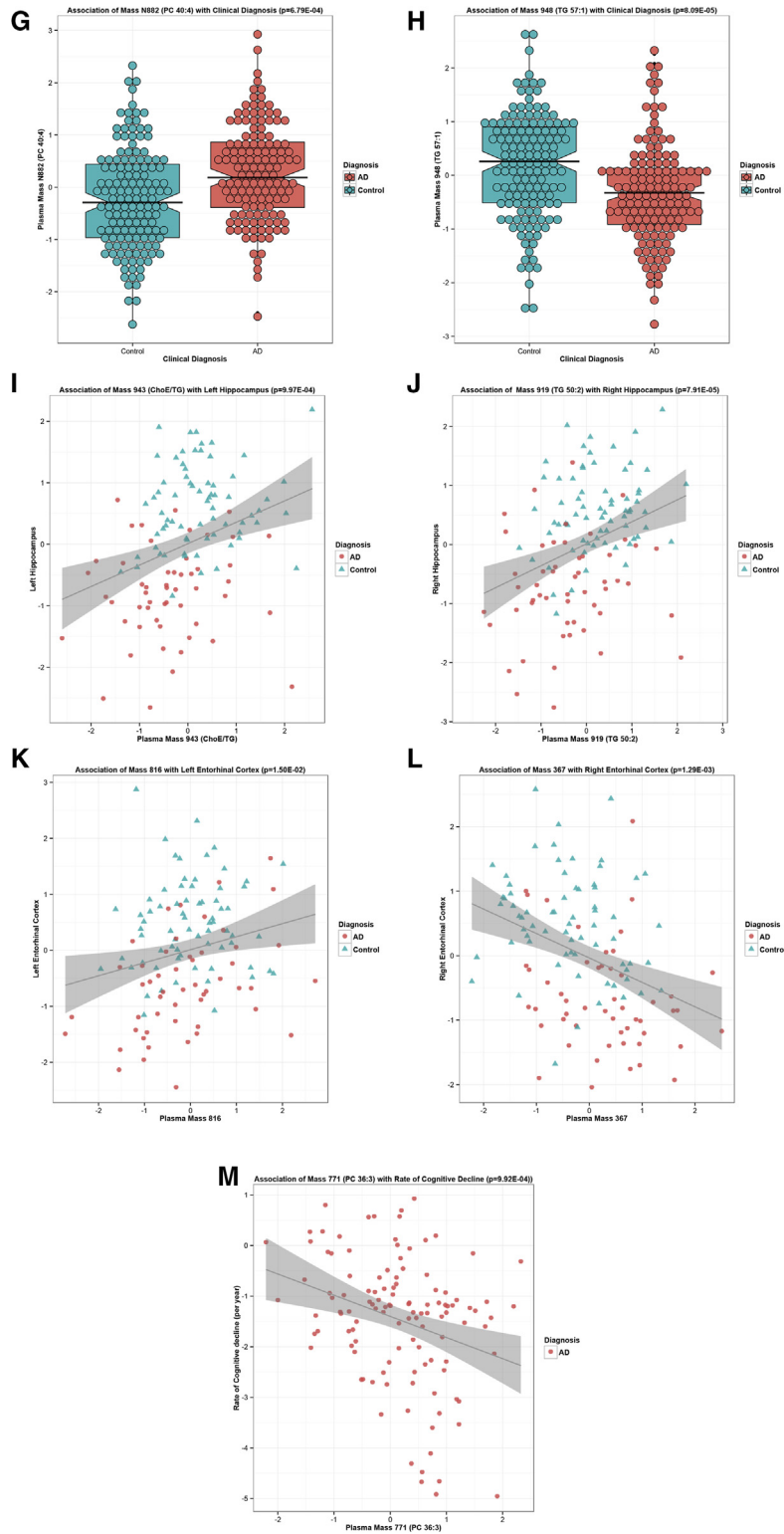
Fig. 1. (Continued)

were plotted using "ROCR". Models including *APOE* ε4 and the six features in Proitsi et al [9] were also tested.

Random forest regression (RFR) models were built for cognitive decline and sMRI measures following the same strategy as for clinical diagnosis. The data set was split randomly into a training (2/3 of the data) and test set (1/3 of the data) for each endophenotype such that the training and test data sets were stratified for each endophenotype and contained

equal representation of each site. Age, gender, and *APOE* ε4 presence were included in the model development for the sMRI models, and the root mean squared error (RMSE) was used to evaluate the performance of the models.

3. Results

A total of 2539 positive ionization and 358 negative ionization features were initially extracted from 300 individuals. After QC, 2216 positive and 289 negative ionization features from 277 individuals (142 AD cases and 135 controls) were used in subsequent analyses. Of these, 53 AD patients and 70 controls had sMRI data available, and 118 AD patients had ROD data available. Sample demographics are displayed in Table 1.

3.1. Univariate analyses results

Logistic regression analyses were initially used to investigate the association of each lipid with AD. We then performed fixed-effects meta-analyses between the results of this data set and our previously published data set [9], using fixed-effects meta-analyses. Briefly, 425 features were associated with AD at *P* value <.05 in this data set; of these, 87 features passed correction for multiple testing at Q value <0.05. After meta-analysis, 377 features were associated with AD at *P* value <.05 and 125 at Q value<0.05. All six features from Proitsi et al [9] were associated with AD at Q value <0.05 in meta-analysis (Fig. 1 (A–F) and Table 2).

Linear regression investigated the association of each lipid with brain atrophy and the rate of cognitive decline. A total of 266 features were associated with the ROD at *P* value <.05, but none passed multiple testing correction. A total of 181 features were associated with right hippocampus volume and 224 were associated with left hippocampus volume; only six features were associated with left hippocampus at Q value <0.05. Finally, 156 and 124 features were associated with EC volume (left and right, respectively) at *P* value <.05, but no associations passed correction for multiple testing. Results for all logistic and linear regression analyses are provided in Supplementary Table 1.

Overall, most lipids were reduced in AD compared to controls (54 of the 87 features associated at Q-value <0.05 were reduced in AD). Additionally, we observed substantial overlap between features associated with clinical AD diagnosis and brain atrophy (Supplementary Fig. 2).

We further investigated whether the *APOE* ε4 and gender modified the associations between lipids and clinical AD diagnosis. *APOE* ε4 modified the association of 231 features with AD, and gender modified the association of 191 features with AD (*P* value <.05); none of these associations was significant at Q-value < 0.05. There were only 3 individuals with the ε2/ε4 genotype, and there were therefore no differences in lipids levels between ε4 and non-ε4 carriers after excluding ε2/ε4 individuals.

3.2. Multivariate analysis results

A RF approach was used to identify a panel of molecules associated with clinical AD diagnosis. After an initial RF pre-selection step on the training data set, the top 10% lipids (250 features), in terms of their variable importance, were selected (after 100 Bootstraps). Furthermore, random forest with recursive feature elimination (RF-rfe) on the training data set showed that the best training performance was for a model with 240 features. To choose a model with high accuracy while reducing the number of features as low as possible, a 5% tolerance RF-rfe model (25 features) was fitted on the whole training data set (AUC, 0.87) and classified the test data set with 73% accuracy (Supplementary Fig. 3 and Table 3). The model was then fitted on the training data set, excluding one negative ionization mode analyte and classified the test training data set with 74% accuracy and the Proitsi et al [9] validation data set with 71% accuracy. There was no increase in accuracy when covariates and the features from Proitsi et al [9] were added to the models (Table 3).

Random forest regressions using the same pipeline were applied to the ROD and brain atrophy measures. After RFR-rfe on the training data set, the lowest mean RMSE for ROD was for a model with 40 features. The 5% tolerance model of the lowest RMSE model (10 features) was fitted to the whole training data set ($R^2 = 0.49$) and predicted the test data set with $R^2 = 0.10$ (Table 4).

For right hippocampus, the lowest RMSE was with a model with 70 features that included age. A 5% tolerance model (12 features) was fitted to the training data set ($R^2 = 0.55$) and predicted the test data set with $R^2 = 0.15$ (Table 3). For left hippocampus, the lowest training RMSE was with 100 features that also included age. The 5% tolerance model (12 features) was fitted to the training data set ($R^2 = 0.59$) and predicted the test data set with $R^2 = 0.15$ (Table 3). The performance of the models was almost identical when age was excluded.

For the right EC, the lowest mean training RMSE was with 70 features; a 5% tolerance model (12 features) was fitted to the training data set ($R^2 = 0.54$) and predicted the test data set with $R^2 = 0.14$. Finally, for left EC, a model with 90 features had the lowest RMSE, and a 5% tolerance model (12 features) was fitted to the training data set ($R^2 = 0.42$) and predicted the test data set with $R^2 = 0.01$ (Table 3). Results of all 2.5% models are presented in Supplementary Tables 2 and 3, and the list of molecules included in each classifier is found in Supplementary Table 1. The strength of association between selected features and each model is shown in Fig. 2, and the scaled VI of each lipid after RF-RFE/RFR-RFE for each phenotype is shown in Supplementary Fig. 4.

3.3. Lipid annotation and putative identification

We opted to annotate the top features, in terms of VI from each model and features selected in more than one

Table 2
List of putatively identified metabolite molecules selected by the six random forest models

Logistic regression analysis	m/z (ionization mode)	Putative metabolite molecule	Present study data set			Proitsi et al 2015 data set			Meta-analysis		
			OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Clinical AD diagnosis	882 (−)	PC 40:4	1.996	1.35–3.01	6.79E−04*	NA	NA	NA	NA	NA	NA
	948 (+)	TG 57:1	0.514	0.36–0.71	8.09E−05*	0.522	0.28–0.92	3.01E−02	0.516	0.39–0.69	6.83E−06*
	856 (+)	ChoE/TG[†]	0.711	0.51–0.99	4.34E−02	0.141	0.04–0.43	1.75E−03	0.632	0.46–0.87	4.94E−03*
	866 (+)	ChoE/TG[†]	0.663	0.46–0.94	2.42E−02	0.251	0.10–0.52	7.51E−04*	0.569	0.41–0.79	7.37E−04*
	868 (+)	ChoE/TG[†]	0.65	0.47–0.89	7.74E−03*	0.218	0.07–0.55	3.15E−03	0.591	0.44–0.80	7.16E−04*
	882 (+)	ChoE/TG[†]	0.57	0.41–0.79	7.99E−04*	0.231	0.08–0.53	1.56E−03	0.517	0.38–0.71	3.05E−05*
	894 (+)	ChoE/TG[†]	0.732	0.51–1.04	8.05E−02	0.151	0.05–0.38	3.14E−04*	0.615	0.44–0.86	4.65E−03*
	970 (+)	ChoE/TG[†]	0.643	0.47–0.87	4.96E−03*	0.362	0.18–0.67	2.56E−03	0.58	0.44–0.77	1.27E−04*
Linear regression analysis			Beta	95% CI	P value	NA	NA	NA	NA	NA	NA
Hippocampus (right)	919 (+)	TG 50:2	0.396	0.20–0.59	7.91E−05*	NA	NA	NA	NA	NA	NA
Hippocampus (Left)	943 (+)	ChoE/TG	0.320	0.13–0.51	9.97E−04*	NA	NA	NA	NA	NA	NA
Entorhinal Cortex (Right)	367 (+)	Sterol	−0.201	−0.39 to −0.01	3.88E−02	NA	NA	NA	NA	NA	NA
Entorhinal Cortex (Left)	816 (+)	NA	0.218	0.03–0.41	2.47E−02	NA	NA	NA	NA	NA	NA
ROD	771 (+)	PC 36:3[‡]	−0.412	−0.65 to −0.17	9.92E−04	NA	NA	NA	NA	NA	NA

Abbreviations: AD, Alzheimer's disease; ChoE, cholesteryl ester; CI, confidence interval; m/z, mass-to-charge ratio; OR, odds ratio; PC, phosphatidylcholine; ROD, rate of cognitive decline; TG, Triglyceride.

NOTE. The six random forest models were for the clinical AD diagnosis, ROD, hippocampus (R/L), and entorhinal cortex (R/L) phenotypes. The association of each molecule is presented with the respective phenotype (i.e., primary phenotype of association). The association of the six molecules previously reported by Proitsi et al 2015 with AD is also presented.

*Q value <0.05.

[†]Features identified by Proitsi et al 2015 and for the Proitsi et al., data set semiquantified values are presented.

[‡]PC 36:3 has m/z 770, and m/z 771 is its C13 isotope. ChoE/TG indicates co-elution of ChoE and TG molecules.

model, using our in-house lipid database and MS/MS fragmentation patterns [4,11,12]. These features were annotated as mainly long-chain triglycerides (LCTs) and ChoEs, some were PCs and a sterol. Fig. 1 (G–M) and Table 2 present the univariate associations of these molecules with the respective phenotypes. The association of the annotated molecules with all phenotypes is shown in Supplementary Fig. 5. The raw intensity counts for each AD associated lipid across AD and controls, along with the coefficients of variation (relative standard deviation [RSD]) of the pooled samples (QCs) are shown in Supplementary Table 4.

Table 3
Random forest classifier model results (clinical AD diagnosis) for the training data set and predictions on the test data set and the Proitsi et al data set

Model (5% tolerance)	Training data set (N = 179)			Test data set (N = 98)						Validation data set (N = 75)					
	Sens.	Spec.	AUC	Acc.	Sens.	Spec.	AUC	PPV	NPV	Acc.	Sens.	Spec.	AUC	PPV	NPV
Covariates only*	0.72	0.71	0.77	0.56	0.54	0.58	0.56	0.5	0.62	0.6	0.57	0.63	0.6	0.57	0.63
25 features[†]	0.82	0.82	0.87	0.73	0.77	0.69	0.73	0.66	0.79	NA	NA	NA	NA	NA	NA
24 features[‡]	0.81	0.82	0.86	0.74	0.74	0.73	0.74	0.68	0.78	0.71	0.69	0.73	0.71	0.69	0.73
25 features[†] + covariates*	0.83	0.83	0.88	0.75	0.79	0.71	0.75	0.68	0.81	NA	NA	NA	NA	NA	NA
24 features[‡] + covariates*	0.84	0.82	0.88	0.75	0.79	0.71	0.75	0.68	0.81	0.71	0.69	0.73	0.71	0.69	0.73
25 features[†] + 6 ChoE/TG[§]	0.82	0.82	0.87	0.71	0.74	0.67	0.71	0.64	0.77	NA	NA	NA	NA	NA	NA
24 features[‡] + 6 ChoE/TG[§]	0.82	0.81	0.87	0.71	0.79	0.66	0.72	0.65	0.80	0.72	0.71	0.73	0.72	0.69	0.74
25 features[‡] + covariates* + 6 ChoE/TG[§]	0.83	0.83	0.88	0.74	0.77	0.71	0.74	0.68	0.80	NA	NA	NA	NA	NA	NA
24 features[‡] + covariates* + 6 ChoE/TG[§]	0.83	0.82	0.88	0.74	0.79	0.69	0.74	0.67	0.81	0.71	0.69	0.73	0.71	0.69	0.73

Abbreviations: Acc, accuracy; AUC, area under the curve; ChoE, cholesteryl ester; NPV, negative predictive value; PPV, positive predictive value; Sens, sensitivity; Spec, specificity; TG, triglyceride.

*Age, sex, ε4.

[†]5% tolerance model including negative ionization molecule.

[‡]5% tolerance model excluding negative ionization molecule.

[§]Six features identified by Proitsi et al 2015.

Table 4
Random forest regression model results for the training data set and predictions on the test data set for each AD endophenotype

Phenotype	Model (5% tolerance)	Train data set (n = 93)		Test data set (n = 28)	
		RMSE	R^2	RMSE	R^2
Hippocampus (right)	Covariates only*	0.92	0.28	1.09	0.02
	12 features[†]	0.58	0.55	0.9	0.15
Hippocampus (Left)	Covariates only*	0.89	0.34	1.21	0.01
	12 features[†]	0.64	0.59	0.99	0.15
Entorhinal cortex (right)	Covariates only*	0.95	0.22	1.14	<0.01
	12 features	0.66	0.54	0.92	0.14
Entorhinal cortex (left)	Covariates only*	1.00	0.22	1.17	<0.01
	12 features[†]	0.77	0.42	1.07	0.01
ROD	10 features[‡]	0.93	0.49	1.09	0.10

Abbreviations: RMSE, root mean squared error; ROD, rate of cognitive decline.

*Age, sex, ε4.

[†]Age was included in the final model. There was no difference in either train or test data set performance when age was excluded.

[‡]Covariates were already included in the calculation for the ROD.

4. Discussion

This is to our knowledge, the largest nontargeted blood lipidomics study in AD to date. Here, we expanded our recent work [9], and using univariate and multivariate approaches, we replicated the associations between six previously reported blood lipids and AD [9] and reported their association with brain atrophy. We further identified combinations of lipids that classified AD patients with relatively good accuracy when tested in both a test and a validation data set (>70%), and combinations of molecules that predicted changes in disease progression (R^2 = 0.10 for test data set) and brain atrophy ($R^2 \geq 0.14$ for all test data sets except for left EC). Overall, we observed substantial overlap between features associated with clinical AD diagnosis and brain atrophy. The associations of all molecules included in each model with all phenotypes is shown in Fig. 2. Although these signatures cannot be used for diagnostic purposes yet, they suggest important biological mechanisms associated with AD.

4.1. Identification and role of lipids in AD

We putatively identified two PC molecules; additionally, ChoEs and triglycerides (TGs) were tentatively annotated due to chromatographic coelution, and finally, we putatively annotated a molecule as a sterol. The higher MS-MS sensitivity achieved here enabled the detection of a number of additional lipids that co-eluted with ChoE; these were annotated as TGs (Table 2).

The association of PCs with AD and cognition has been extensively described [4,8]. Here, one of the molecules

most strongly associated with AD is a putative PC (PC 40:4), and the top lipid in the ROD model is also a putative PC (PC 36:3). In contrast to the same species of molecules, we have previously identified, both PCs are increased in AD, and PC 36:3 is associated with faster ROD. Although most studies to date have reported a reduction of PC levels in AD, an increase in CSF PCs has been observed in AD compared to control brains [19] and recently in "AD-like" patients based on their CSF Amyloid-beta42, Tau, and Phospho-Tau-181 levels [20]. A recent study also reported a parallel increase of PCs containing saturated and short-chain fatty acids in serum from AD patients [21]. These suggest deregulation in the biosynthesis, turnover, and acyl chain remodeling of phospholipids, in accordance with increased phospholipid breakdown due to PLA2 [21] overactivation.

We have also reported associations with low-chain and very-low-chain triglycerides (LCTs/VLCTs; fatty acid chain length >16 carbons). One of the most interesting findings was that due to the higher MS-MS sensitivity achieved in this study, we were able to observe putative VLCTs that were coeluting with ChoEs (Table 2). We have previously reported on the synthesis of ChoEs [9]; briefly, it takes place by transfer of fatty acids from PC to cholesterol, a reaction catalyzed by lecithin cholesterol acyl transferase in plasma and by acyl-coenzyme A: cholesterol acyl transferase 1 and 2 (ACAT1 and ACAT2) in other tissues, including the brain. The association of LCTs/VLCTs with AD is noteworthy. Although overall TGs are seen as risk factors for many disorders including cardiovascular disease (CVD) and type 2 diabetes (T2D), numerous investigations point to the diverse role of TGs with different chain lengths. It is known for example that medium-chain triglycerides (MCTs) and LCTs have different metabolic pathways in digestion and absorption [22]. Moreover, although LCTs of lower carbon number and double bond content have been associated with increased CVD [23] and T2D risk [24], LCTs with higher carbon number and double bond content, like the ones here, have been associated with decreased risk of T2D [24], whereas no associations between T2D and total triglyceride levels were observed in the same individuals [24]. Furthermore, decreased concentration of LCTs and an increased concentration of VLCTs have been associated with longevity [25]. These findings are particularly interesting as most vegetable oils are comprised of long-chain fatty acids; however, only MTCs have to our knowledge been implicated in AD, although findings are controversial [26]. When we previously investigated the association of total cholesterol and TGs with AD in overlapping individuals using Mendelian randomization, we found no evidence for an association with AD [27]. Additionally, we observed no difference in serum triglycerides, total cholesterol, LDL cholesterol, and HDL cholesterol between AD patients and controls for a random subset of study participants in this study (AD = 102, controls = 106) that had serum lipid measures available for the same visit, as well as no difference in the
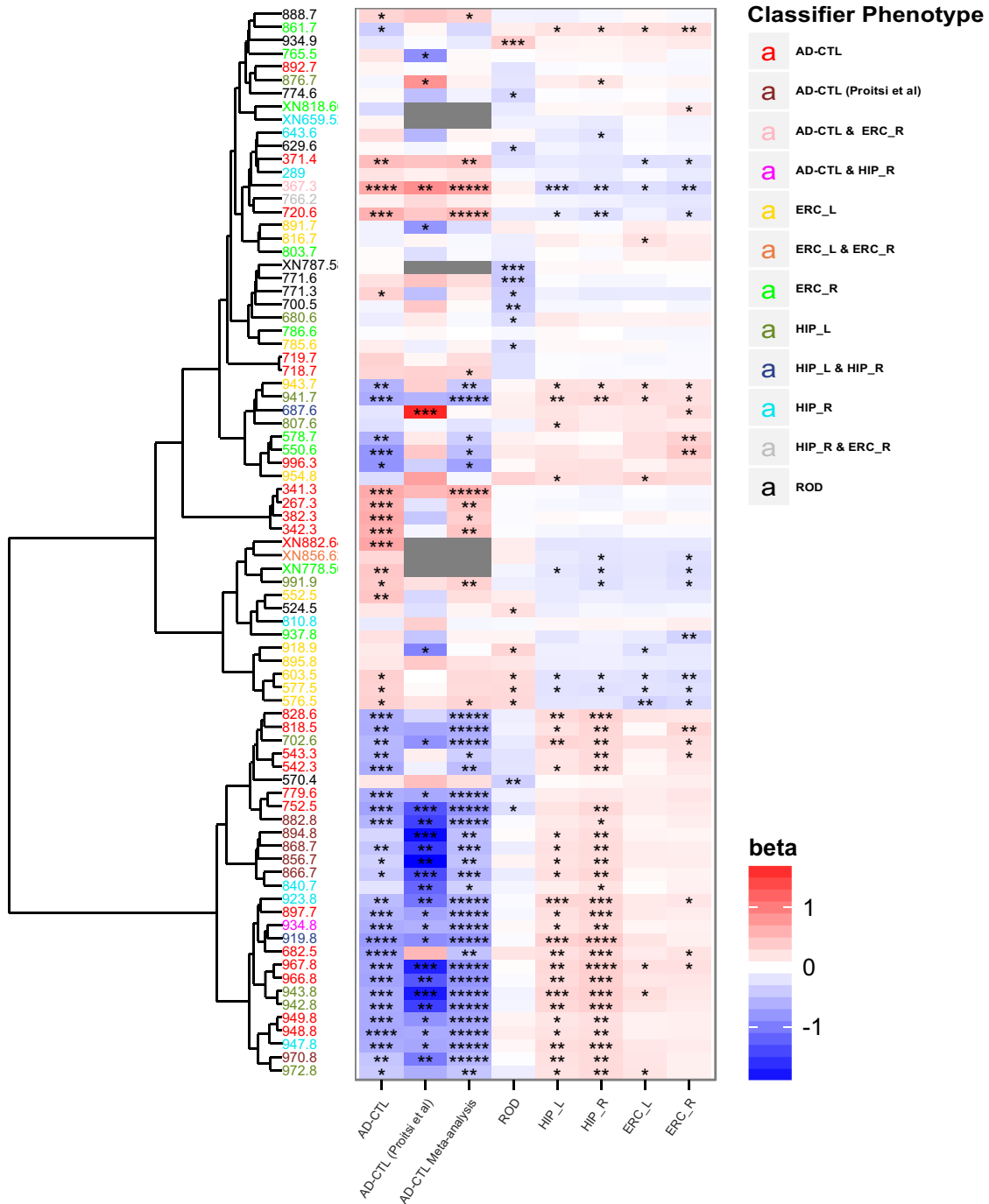
Fig. 2. Heatmap of the univariate associations between features selected during random forest analyses for each phenotype. The color of each box represents the univariate logistic regression beta coefficient (log[OR]) for clinical diagnosis or the univariate scaled linear regression beta coefficients for the rate of cognitive decline and brain atrophy, after adjusting for covariates. The stars on each box represent the strength of the association: *P value <.05; **P value <.01; ***P value <.001; ****P value <.0001; *****P value <.00001. The order of the metabolite molecules on the y-axis is based on a hierarchical clustering using the metabolites pairwise correlations. XN denotes negative ionization mode feature. Abbreviations: AD-CTL, Clinical AD diagnosis; ROD, rate of cognitive decline; HIP_L, left hippocampus; HIP_R, right hippocampus; ERC_L, left entorhinal cortex; ERC_R, right entorhinal cortex.

frequency of AD patients and controls who were taking statins. On the other hand, a recent study reported an overlap between genes involved in elevated plasma lipid levels and inflammation and the risk for AD [28]. All these highlight the relevance of investigating smaller lipid fractions as they

highlight specific steps in their biosynthesis and metabolism that may be associated with AD.

Finally, we observed an association between most phenotypes and a feature of m/z 367. We previously described a molecule with the same mass and similar retention time to

be reduced in AD [9]. The molecule discovered here was associated with an increased risk for AD in both data sets and with reduced brain volume and was included in the ERC and clinical diagnosis models. We believe this feature is a fragment and a sterol, specifically an isomer of desmosterol. Desmosterol is a precursor of cholesterol and seladin (*DHCR24*), which governs the metabolism of desmosterol to cholesterol in specific brain areas. Desmosterol has been shown to inhibit β-secretase cleavage of APP, and the formation of amyloid-β and lower desmosterol levels has been found in the plasma and brains of AD patients compared to controls [29–32].

Although the association of aberrant lipid metabolism in AD pathogenesis is undisputed [33–35]; at this stage, the mechanisms by which these changes in lipids might occur in AD are unclear. One possibility involves AD selective alterations to circulating lipid metabolism. However, another possibility relates to cellular lipid production. A number of phospholipids are synthesized within a specialized region of the endoplasmic reticulum (ER) that is closely associated with mitochondria, the mitochondria-associated ER membranes (MAM). The close association of MAM to mitochondria facilitates Ca^{2+} and phospholipid exchange between the two organelles [36–38]. Recent studies have shown that MAM contacts are damaged in AD [39–42]. Because ER-mitochondria contacts are required for the synthesis of certain lipids [36–38], such changes may affect lipid metabolism and lead to some of the changes described here. Indeed, different *APOE* alleles have been shown to influence MAM [43].

4.2. Strengths and limitations

Here, we have used a large well-characterized AD cohort and a careful and systematic analysis pipeline. Through bootstrapping, we have reduced over-fitting, and subsequently, we validated our results in an unseen data set for each phenotype and an additional validation data set for clinical AD diagnosis. Our AD diagnostic classifier achieved 88% accuracy on the training data set (summary of 100 bootstraps) and predicted the test and validation data sets with >70% accuracy. Our training data set comprised of individuals matched on age and gender. On the other hand, the test data set consisted primarily by females, and AD patients were significantly older than controls. Additionally, the validation data set [9] included AD patients and controls of UK origin only, older than the individuals of the current data set. These findings highlight robustness in the model. For the AD endophenotypes, the Random Forest Regression models had a very good performance on all training data sets. Although the performance dropped significantly in the test data sets, we observed $R^2 > 0.10$ for all phenotypes except for Left EC ($R^2 = 0.01$). The drop in performance can be attributed to over-fitting of the training data sets and the smaller number of individuals with ROD/brain atrophy measures. The poor performance of the Left EC is in agreement with our univariate analyses that highlighted weaker associations with the EC for the whole sample; however, it is in contrast to the overall right-to-left asymmetry in AD [17].

A limitation of this study is that we were not able to decipher the exact fatty acid chain structure of some features. Owing to the higher MS-MS sensitivity, we observed a number of putative ChoEs and TGs co-eluting, which is commonly observed in lipidomics studies due to hundreds of lipids detected in one analysis; to minimize co-elution problems, our chromatographic run is 2 hours long using ultra pressure chromatography [11,44].

Additionally, although this is the largest AD lipidomics study to date, we acknowledge that the sample size is still modest and further replication is required, especially for the ROD and brain atrophy phenotypes. Moreover, although we had information on the ROD, this calculation was based on the MMSE, which is a crude measure of measurement of cognition. Furthermore, the present study did not contain an MCI cohort or information on conversion to MCI/AD, and therefore, we do not know whether these features are associated with initiation of AD. This study additionally suffers from limitations inherent to AD case-control studies, such as the large number of co-morbidities in old age, the possibility that some of the elderly controls may already carry pathology, and that some of the clinically diagnosed AD may be pathologically non-AD dementias. Finally, this study lacks information on BMI and body fat distribution that could potentially explain some of the differences between AD patients and controls.

However, through the longitudinal nature of these cohorts, we know that all the AD patients used for our analysis maintained the diagnosis of AD as did all controls for at least 3 years from their baseline visit. Additionally, our information on disease progression and brain atrophy provide us with more precise phenotypes that capture different stages of disease pathology including the early preclinical stages. Given the good performance of these models, we believe that enrichment with additional individuals and pathology information would increase their performance. Finally, although we did not have BMI information for our cohort, we observed no difference in statin use or serum lipids between AD cases and controls.

5. Conclusion

In conclusion, the findings of this study deepen our knowledge of AD disease mechanisms and emphasize the importance of investigating in detail different lipid fractions in dementia research. As it is not known whether the observed changes in lipid levels are causally related to or are just a marker of changes in lipoprotein dynamics and composition, studies that address causality are essential, as the success of targeting specific molecules and

identifying potentially causal pathways amenable to intervention is predicated on these molecules being on the causal pathway. Finally, integrating additional types of biological modalities such as protein, gene expression, and genotype information may increase the fit of these models and help us to understand more about the biological context in which these molecules operate.

Acknowledgments

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jalz.2016.08.003.

RESEARCH IN CONTEXT

1. Systematic review: The authors reviewed the literature using PubMed and reported key publications. There is a pressing need to identify noninvasive Alzheimer's disease (AD) biomarkers, and blood metabolites are promising biomarkers that could aid early diagnosis and ultimately lead to the development of more effective interventions. Recent blood metabolomic studies have highlighted the role of lipid compounds in AD. However, most studies are small and relatively heterogeneous.

2. Interpretation: This study replicated previous associations between blood lipids and AD and reported novel associations between blood lipids and clinical AD diagnosis, the rate of cognitive and brain atrophy. These findings deepen our knowledge of AD disease mechanisms and suggest novel targets for future work.

3. Future directions: Results of this study could be complemented with protein and genetic data. Future studies should address whether these changes are causally related to AD or are just a marker of changes in lipoprotein dynamics and composition.

References

[1] Mapstone M, Cheema AK, Fiandaca MS, Zhong X, Mhyre TR, MacArthur LH, et al. Plasma phospholipids identify antecedent memory impairment in older adults. Nat Med 2014;20:415–8.

[2] Oresic M, Hyotylainen T, Herukka SK, Sysi-Aho M, Mattila I, Seppanan-Laakso T, et al. Metabolome in progression to Alzheimer's disease. Transl Psychiatry 2011;1:e57.

[3] Trushina E, Dutta T, Persson XM, Mielke MM, Petersen RC. Identification of altered metabolic pathways in plasma and CSF in mild cognitive impairment and Alzheimer's disease using metabolomics. PLoS One 2013;8:e63644.

[4] Whiley L, Sen A, Heaton J, Proitsi P, Garcia-Gomez D, Leung R, et al. Evidence of altered phosphatidylcholine metabolism in Alzheimer's disease. Neurobiol Aging 2014;35:271–8.

[5] Zipser BD, Johanson CE, Gonzalez L, Berzin TM, Tavares R, Hulette CM, et al. Microvascular injury and blood-brain barrier leakage in Alzheimer's disease. Neurobiol Aging 2007;28:977–86.

[6] Lindon JC, Holmes E, Nicholson JK. Metabonomics in pharmaceutical R&D. FEBS J 2007;274:1140–51.

[7] Suhre K, Shin SY, Petersen AK, Mohney RP, Meredith D, Wagele B, et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature 2011;477:54–60.

[8] Simpson BN, Kim M, Chuang YF, Beason-Held L, Kitner-Triolo M, Kraut M, et al. Blood metabolite markers of cognitive performance and brain function in aging. J Cereb Blood Flow Metab 2016; 36:1212–23.

[9] Proitsi P, Kim M, Whiley L, Pritchard M, Leung R, Soininen H, et al. Plasma lipidomics analysis finds long chain cholesteryl esters to be associated with Alzheimer's disease. Transl Psychiatry 2015;5:e494.

[10] Lovestone S, Francis P, Kloszewska I, Mecocci P, Simmons A, Soininen H, et al. AddNeuroMed–the European collaboration for the discovery of novel biomarkers for Alzheimer's disease. Ann N Y Acad Sci 2009;1180:36–46.

[11] Whiley L, Godzien J, Ruperez FJ, Legido-Quigley C, Barbas C. In-vial dual extraction for direct LC-MS analysis of plasma for comprehensive and highly reproducible metabolic fingerprinting. Anal Chem 2012;84:5992–9.

[12] Whiley L, Legido-Quigley C. Current strategies in the discovery of small-molecule biomarkers for Alzheimer's disease. Bioanalysis 2011;3:1121–42.

[13] Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 2006;78:779–87.

[14] Westman E, Aguilar C, Muehlboeck JS, Simmons A. Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment. Brain Topogr 2013; 26:9–23.

[15] Simmons A, Westman E, Muehlboeck S, Mecocci P, Vellas B, Tsolaki M, et al. MRI measures of Alzheimer's disease and the AddNeuroMed study. Ann N Y Acad Sci 2009;1180:47–55.

[16] Simmons A, Westman E, Muehlboeck S, Mecocci P, Vellas B, Tsolaki M, et al. The AddNeuroMed framework for multi-centre MRI assessment of Alzheimer's disease : experience from the first 24 months. Int J Geriatr Psychiatry 2011;26:75–82.

[17] Sattlecker M, Kiddle SJ, Newhouse S, Proitsi P, Nelson S, Williams S, et al. Alzheimer's disease biomarker discovery using SOMAscan multiplexed protein technology. Alzheimers Dement 2014;10:724–34.

[18] Wilson RS, Arnold SE, Schneider JA, Kelly JF, Tang Y, Bennett DA. Chronic psychological distress and risk of Alzheimer's disease in old age. Neuroepidemiology 2006;27:143–53.

[19] Walter A, Korth U, Hilgert M, Hartmann J, Weichel O, Hilgert M, et al. Glycerophosphocholine is elevated in cerebrospinal fluid of Alzheimer patients. Neurobiol Aging 2004;25:1299–303.

[20] Koal T, Klavins K, Seppi D, Kemmler G, Humpel C. Sphingomyelin SM(d18:1/18:0) is significantly enhanced in cerebrospinal fluid samples dichotomized by pathological amyloid-beta42, tau, and phospho-tau-181 levels. J Alzheimers Dis 2015;44:1193–201.

[21] Gonzalez-Dominguez R, Garcia-Barrera T, Gomez-Ariza JL. Combination of metabolomic and phospholipid-profiling approaches for the study of Alzheimer's disease. J Proteomics 2014;104:37–47.

[22] St-Onge MP, Mayrsohn B, O'Keeffe M, Kissileff HR, Choudhury AR, Laferrere B. Impact of medium and long chain triglycerides consumption on appetite and food intake in overweight men. Eur J Clin Nutr 2014;68:1134–40.

[23] Stegemann C, Pechlaner R, Willeit P, Langley SR, Mangino M, Mayr U, et al. Lipidomics profiling and risk of cardiovascular disease in the prospective population-based Bruneck study. Circulation 2014; 129:1821–31.

[24] Rhee EP, Cheng S, Larson MG, Walford GA, Lewis GD, McCabe E, et al. Lipid profiling identifies a triacylglycerol signature of insulin resistance and improves diabetes prediction in humans. J Clin Invest 2011;121:1402–11.

[25] Montoliu I, Scherer M, Beguelin F, DaSilva L, Mari D, Salvioli S, et al. Serum profiling of healthy aging identifies phospho- and sphingolipid species as markers of human longevity. Aging (Albany NY) 2014; 6:9–25.

[26] Henderson ST, Vogel JL, Barr LJ, Garvin F, Jones JJ, Costantini LC. Study of the ketogenic agent AC-1202 in mild to moderate Alzheimer's disease: a randomized, double-blind, placebo-controlled, multicenter trial. Nutr Metab (Lond) 2009;6:31.

[27] Proitsi P, Lupton MK, Velayudhan L, Newhouse S, Fogh I, Tsolaki M, et al. Genetic predisposition to increased blood cholesterol and triglyceride lipid levels and risk of Alzheimer disease: a mendelian randomization analysis. PLoS Med 2014;11:e1001713.

[28] Desikan RS, Schork AJ, Wang Y, Thompson WK, Dehghan A, Ridker PM, et al. Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. Circulation 2015;131:2061–9.

[29] Crameri A, Biondi E, Kuehnle K, Lutjohann D, Thelen KM, Perga S, et al. The role of seladin-1/DHCR24 in cholesterol biosynthesis, APP processing and Abeta generation in vivo. EMBO J 2006;25:432–43.

[30] Greeve I, Kretzschmar D, Tschape JA, Beyn A, Brellinger C, Schweizer M, et al. Age-dependent neurodegeneration and Alzheimer-amyloid plaque formation in transgenic Drosophila. J Neurosci 2004; 24:3899–906.

[31] Sato Y, Suzuki I, Nakamura T, Bernier F, Aoshima K, Oda Y. Identification of a new plasma biomarker of Alzheimer's disease using metabolomics technology. J Lipid Res 2012;53:567–76.

[32] Wisniewski T, Newman K, Javitt NB. Alzheimer's disease: brain desmosterol levels. J Alzheimers Dis 2013;33:881–8.

[33] Reitz C, Mayeux R. Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. Biochem Pharmacol 2014; 88:640–51.

[34] Reitz C. Dyslipidemia and the risk of Alzheimer's disease. Curr Atheroscler Rep 2013;15:307.

[35] Reitz C. Dyslipidemia and dementia: current epidemiology, genetic evidence, and mechanisms behind the associations. J Alzheimers Dis 2012;30:S127–45.

[36] Rowland AA, Voeltz GK. Endoplasmic reticulum-mitochondria contacts: function of the junction. Nat Rev Mol Cell Biol 2012;13:607–25.

[37] Helle SC, Kanfer G, Kolar K, Lang A, Michel AH, Kornmann B. Organization and function of membrane contact sites. Biochim Biophys Acta 1833;2013:2526–41.

[38] van Vliet A, Verfaillie T, Agostinis P. New functions of mitochondria associated membranes in cellular signalling. Biochim Biophys Acta 2014;1843:2253–62.

[39] Area-Gomez E, Del Carmen Lara Castillo M, Tambini MD, Guardia-Laguarta C, de Groof AJ, Madra M, et al. Upregulated function of mitochondria-associated ER membranes in Alzheimer disease. EMBO J 2012;31:4106–23.

[40] Zampese E, Fasolato C, Kipanyula MJ, Bortolozzi M, Pozzan T, Pizzo P. Presenilin 2 modulates endoplasmic reticulum (ER)-mitochondria interactions and Ca2+ cross-talk. Proc Natl Acad Sci U S A 2011;108:2777–82.

[41] Sepulveda-Falla D, Barrera-Ocampo A, Hagel C, Korwitz A, Vinueza-Veloz MF, Zhou K, et al. Familial Alzheimer's disease-associated presenilin-1 alters cerebellar activity and calcium homeostasis. J Clin Invest 2014;124:1552–67.

[42] Hedskog L, Pinho CM, Filadi R, Ronnback A, Hertwig L, Wiehager B, et al. Modulation of the endoplasmic reticulum-mitochondria interface in Alzheimer's disease and related models. Proc Natl Acad Sci U S A 2013;110:7916–21.

[43] Tambini MD, Pera M, Kanter E, Yang H, Guardia-Laguarta C, Holtzman D, et al. ApoE4 upregulates the activity of mitochondria-associated ER membranes. EMBO Rep 2016;17:27–36.

[44] Sen A, Wang Y, Chiu K, Whiley L, Cowan D, Chang RC, et al. Metabolic phenotype of the healthy rodent model using in-vial extraction of dried serum, urine, and cerebrospinal fluid spots. Anal Chem 2013; 85:7257–63.