

## ORIGINAL ARTICLE

# Plasma lipidomics analysis finds long chain cholesteryl esters to be associated with Alzheimer's disease

P Proitsi<sup>1</sup>, M Kim<sup>2</sup>, L Whiley<sup>2</sup>, M Pritchard<sup>1</sup>, R Leung<sup>1</sup>, H Soininen<sup>3</sup>, I Kloszewska<sup>4</sup>, P Mecocci<sup>5</sup>, M Tsolaki<sup>6</sup>, B Vellas<sup>7</sup>, P Sham<sup>8</sup>, S Lovestone<sup>9</sup>, JF Powell<sup>1,11</sup>, RJB Dobson<sup>1,10,11</sup> and C Legido-Quigley<sup>2,11</sup>

There is an urgent need for the identification of Alzheimer's disease (AD) biomarkers. Studies have now suggested the promising use of associations with blood metabolites as functional intermediate phenotypes in biomedical and pharmaceutical research. The aim of this study was to use lipidomics to identify a battery of plasma metabolite molecules that could predict AD patients from controls. We performed a comprehensive untargeted lipidomic analysis, using ultra-performance liquid chromatography/mass spectrometry on plasma samples from 35 AD patients, 40 elderly controls and 48 individuals with mild cognitive impairment (MCI) and used multivariate analysis methods to identify metabolites associated with AD status. A combination of 10 metabolites could discriminate AD patients from controls with 79.2% accuracy (81.8% sensitivity, 76.9% specificity and an area under curve of 0.792) in a novel test set. Six of the metabolites were identified as long chain cholesteryl esters (ChEs) and were reduced in AD (ChE 32:0, odds ratio (OR) = 0.237, 95% confidence interval (CI) = 0.10–0.48,  $P = 4.19E - 04$ ; ChE 34:0, OR = 0.152, 95% CI = 0.05–0.37,  $P = 2.90E - 04$ ; ChE 34:6, OR = 0.126, 95% CI = 0.03–0.35,  $P = 5.40E - 04$ ; ChE 32:4, OR = 0.056, 95% CI = 0.01–0.24,  $P = 6.56E - 04$  and ChE 33:6, OR = 0.205, 95% CI = 0.06–0.50,  $P = 2.21E - 03$ , per (log<sub>2</sub>) metabolite unit). The levels of these metabolites followed the trend control > MCI > AD. We, additionally, found no association between cholesterol, the precursor of ChE and AD. This study identified new ChE molecules, involved in cholesterol metabolism, implicated in AD, which may help identify new therapeutic targets; although, these findings need to be replicated in larger well-phenotyped cohorts.

*Translational Psychiatry* (2015) 5, e494; doi:10.1038/tp.2014.127; published online 13 January 2015

## INTRODUCTION

A better understanding of the biological mechanisms underlying Alzheimer's disease (AD) is required. AD is a devastating illness that currently affects over 496,000 people in the UK ([www.alz.org.uk](http://www.alz.org.uk)). It is one of the major challenges for health care in the 21st century and with estimated longer life expectancy, the number of demented patients worldwide are expected to reach 81.1 million in 2040.<sup>1</sup> The lack of current treatments for AD and the lack of a definite and early diagnosis highlights the absence of a comprehensive understanding of the biological mechanisms underlying the changes that occur during the process of neurodegeneration. This underscores the urgent need for biomarkers that would lead to novel treatment strategies and improve the lives of those affected.<sup>2</sup> Current sulcal cerebrospinal fluid or brain imaging biomarkers are costly, can cause discomfort to the patient and are impractical at large scale. For example, the Medicare Evidence Development and Coverage Advisory Committee concluded that there was too little evidence to show that amyloid positron emission tomography (PET) scans improve AD outcomes or that its benefits outweigh the harms and high cost.<sup>3</sup> A blood-based biomarker could act as a screening tool to identify

at-risk individuals for further investigation or recruitment into clinical trials.

Studies have now demonstrated the potential of using blood metabolites, the repertoire of molecules (size < 1000–1500 Da)<sup>4</sup> present in cells and tissue,<sup>5</sup> as functional intermediate phenotypes in biomedical and pharmaceutical research.<sup>6</sup> Metabolic phenotyping and its subset lipidomics is the rapidly evolving field of the comprehensive measurement, in a non-targeted manner, of ideally all endogenous metabolites in a biological sample.<sup>7</sup> These small-to-medium molecules are the final products of interactions between gene expression, protein expression and the cellular and external environment, and represent an essential aspect of the phenotype of an organism.<sup>8</sup> In addition, it is now known that there is a communication between the brain and the periphery and is increasingly believed that the damage to the blood–brain barrier caused by AD increases theoretically the chance of metabolites crossing to the brain.<sup>2,9,10</sup> Coupled with the fact that blood is relatively easily accessible, plasma metabolites are an ideal source of noninvasive biomarkers and a molecular ‘footprint’ of disease.

Recently, a number of non-targeted blood metabolomic studies (liquid chromatography-mass spectrometry (LC-MS) or direct infusion mass spectrometry (MS)) in AD have emerged,

<sup>1</sup>Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; <sup>2</sup>Institute of Pharmaceutical Science, King's College London, London, UK; <sup>3</sup>Department of Neurology, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland; <sup>4</sup>Department of Old Age Psychiatry & Psychotic Disorders, Medical University of Łódź, Łódź, Poland; <sup>5</sup>Section of Gerontology and Geriatrics, Department of Medicine, University of Perugia, Perugia, Italy; <sup>6</sup>Memory and Dementia Center, 3rd Department of Neurology, Aristotle University of Thessaloniki, Thessaloniki, Greece; <sup>7</sup>Department of Internal and Geriatrics Medicine, INSERM U 1027, Gerontopole, Hôpitaux de Toulouse, Toulouse, France; <sup>8</sup>Department of Psychiatry, State Key Laboratory of Brain and Cognitive Sciences, and Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, the University of Hong Kong, Hong Kong; <sup>9</sup>Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK and <sup>10</sup>National Institute for Health Research Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia at South London and Maudsley NHS Foundation, London, UK. Correspondence: Dr P Proitsi, Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF, UK.

E-mail: petroula.proitsi@kcl.ac.uk

<sup>11</sup>These senior authors contributed equally to this work.

Received 2 June 2014; revised 28 September 2014; accepted 26 October 2014

highlighting the role of lipid compounds, such as sphingolipids,<sup>11</sup> bile acids,<sup>12</sup> desmosterol<sup>13</sup> and phosphatidylcholines (PCs).<sup>14–16</sup> In a LC-MS/NMR screen followed by subsequent quantification, we previously identified three lipid PC molecules (PC16:0/20:5, PC16:0/22:6 and 18:0/22:6) that were progressively diminished in subjects with mild cognitive impairment (MCI) and AD patients.<sup>16</sup> Interestingly, the latter two were found to be markers in a recent study where a panel of 10 lipids showed 90% area under curve (AUC) for preconversion to amnestic MCI or AD.<sup>14</sup>

Lipid metabolism has been extensively implicated in the pathogenesis of AD through cell biological<sup>17,18</sup> and genetic studies.<sup>19–22</sup> Further support comes from brain tissue metabolic studies, which have shown changes in lipid compounds such as lysobisphosphatidic acid, sphingomyelin, sphingolipids and desmosterol in the brains of AD patients.<sup>23–26</sup>

In this study, we performed untargeted lipidomics utilizing plasma samples from 36 AD patients, 40 healthy controls and 48 individuals with MCI with the aim of discovering new molecules that could diagnose AD patients from controls and MCI and improve our current knowledge of molecules associated with AD and their underlying biology.

## MATERIALS AND METHODS

### Patient sample collection

This study utilized 124 age and gender matched plasma samples (36 AD patients, 48 MCI individuals and 40 controls) from the Dementia Case Register (DCR) at King's College London and the EU funded AddNeuroMed study.<sup>16,27</sup> Normal elderly control subjects were recruited from non-related family members of AD patients, care-givers' relatives, social centres for the elderly or GP surgeries and had no evidence of cognitive impairment. AD and MCI subjects were recruited primarily from local memory clinics, and as such the MCI cohort was expected to be composed largely of subjects with a likely AD end point. Overlapping samples from this study were previously used by Whiley *et al.*<sup>16</sup> to measure the relative amounts of three PCs to a PC internal standard in plasma. Relevant ethics board approved the study and informed consent was obtained for all subjects. Each patient was required to fast for 2 h before sample collection and 10 ml of blood was then collected in tubes coated with sodium ethylenediaminetetraacetic acid to prevent clotting. Whole blood was centrifuged to form a plasma supernatant, which in turn was removed and placed at –80 °C until further use.

### Lipidomics

**Sample treatment.** Sample treatment has been described elsewhere.<sup>4,16</sup> 20 µl of plasma was added to a glass HPLC vial containing a 400 µl glass insert (Chromacol, Welwyn Garden City, UK). Ten microlitres of high purity water and 40 µl of MS-grade methanol were added to each sample, followed by a 2 min vortex mix to precipitate proteins. Then, 200 µl of methyl t-butyl ether (containing 10 µg ml<sup>−1</sup> of internal standard triacylglycerol 45:0) was added, and the samples were mixed via vortex at room temperature for 1 h. After the addition of 50 µl of high purity water, a final sample mixing was performed before centrifugation at 3000 g for 10 min. The upper, lipid-containing, methyl t-butyl ether 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).

The LC-MS-MS method employed has previously been published<sup>2,16</sup> and has shown to measure amounts of >4500 metabolite species, particularly lipids. Instrumentation included a Waters ACQUITY UPLC and XEVO QTOF system (Waters, Milford, MA, USA).

Chromatographic separation was achieved using an Agilent (Palo Alto, CA, USA) Poroshell 120 EC-C8 column (150 mm × 2.1 mm, 2.7 µm), maintained at 60 °C. A gradient was used consisting of 10 mM ammonium formate in water (A) and 10 mM ammonium formate in methanol (B). The solvent was delivered at a flow rate of 0.5 ml min<sup>−1</sup>. The gradient consisted of 0 min (75% B), 23 min (96% B), 36 min (96% B), 36.5 min (100% B), 41.5 min (100% B), 42 min (75% B), 51 min (75% B).

The XEVO QTOF was operated in the positive ion mode with a capillary voltage of 2.5 kV and a cone voltage of 60 V. The desolvation gas flow was 500 l h<sup>−1</sup> and the source temperature was 120 °C. All analyses were acquired using the lock spray setting; leucine enkephalin was used as lock

mass (*m/z* 556.2771 and 278.1141). Data were collected in the centroid mode over the mass range *m/z* 100–1000 with an acquisition time of 1 s per scan.

Samples were analysed in a randomized order, in four batches, with pooled plasma sampled (quality control (QC) samples) at regular intervals throughout the run (*n*=20). LC-MS raw data were aligned and normalized to total mean area, using Waters MarkerLynx software (Waters). Further validation of metabolite concentrations, which were associated with disease status took place using Waters QuanLynx software (Waters) and calculating peak ratios for metabolites that were over the LOQ (limit of quantification) and the area under the peak of the internal standard.

### Statistical analysis

**Data pre-treatment.** Metabolites detected in <80% of each of the diagnostic groups and the four batches were excluded from analyses. Metabolite distributions were inspected using histograms and the Shapiro-Wilks test was used to test for normality of the metabolite distributions. The distribution of a large number of metabolites was skewed and the data were further log2 transformed. The Empirical Bayes method ComBat<sup>28</sup> was used to correct for batch effects in each metabolite. Principal components analysis was then used for the detection of outliers and to check whether the 20 QC samples clustered together. Missing data points were imputed using KNN (*k*-nearest neighbours, *k*=10), separately for each disease phenotype ('impute'). Metabolite correlations were visualized using a heatmap ('ggplot2'). Analysis of variance was used to test for differences in the levels of continuous variables between the three diagnostic groups, followed by Tukey's honest significant differences post hoc test between pairs of groups when the results of analysis of variance were significant at the *P*<0.05 level. Pearson's  $\chi^2$  was used to test for differences in frequency of the categorical variables between the diagnostic groups. All the analyses took place in R.3.01.

**Single analyte analysis.** Analytes were centred around their mean. Logistic regression ('glm') was used to investigate the association of each metabolite individually with disease status (AD versus controls). Logistic regression models were adjusted for the number of APOE ε4 alleles, age, gender and batch. False discovery rate (FDR) correction (0.05) was applied to correct for multiple testing. Secondary models were run to test whether the association of metabolites with disease status was modified by the presence of the APOE ε4 allele by testing for interactions.

**Multivariate analysis.** A Random Forest approach (using 'rf' and 'rfe' in the package 'CARET') was used to develop an AD versus control classifier. Due to the large number of variables and their highly correlated structure, the following approach was used to achieve high performance with a minimal variable set. AD cases and controls were divided into a training (2/3) and an independent test data set (1/3) such that the training set comprised equal numbers of each diagnostic group. The training set was further divided into 100 bootstrap sample sets comprising a bootstrap training set (75%) used to build the Random Forest model and a bootstrap test set (25%) used to test the model. In each model, the default setting for ntree=500 was used and the optimum mtry number after 100 bootstraps was chosen and fitted to the whole training data set. The 'AUC' was used to test for the performance of each classifier.

In each of the 100 bootstrap iterations, each variable was assigned a variable importance score. The ranks were summed for each metabolite over the 100 bootstraps providing an indication of the predictive power of each variable. We then selected the top 10% analytes on the basis of their summed variable importance rankings and ran the Random Forest with recursive feature elimination (rfe), that is, backward elimination on these selected variables using a second round of 100 bootstraps. The bootstrapping was repeated keeping the top 50 to 10 analytes in steps of five, and down to two analytes in steps of one. For each subset of predictors, the mean bootstrap testing performance was calculated, and, on the basis of this, the optimal number of variables was identified using the 'sizeTolerance' function. This takes into account the whole profile during 'rfe' and picks a subset size that is small without sacrificing too much performance. The optimal number of variables was then used to build a final model in the complete training data, which was tested with the independent test set (Model 1).

The final model was also tested on the MCI sample to determine whether it would classify MCI as cases or controls.

We then repeated the steps above including APOE ε4 genotype (0,1 or 2 alleles) during the initial Random Forest bootstrapping step to assess its variable importance and the predictive ability of the metabolites in the presence of APOE.

## RESULTS

The sample demographics are displayed in Table 1 and Supplementary Figure 1 gives a detailed account of the QC steps. In brief, 1878 molecular features were extracted from the 124 samples. Following QC, metabolites detected in <80% of the three diagnostic groups and the four batches were removed and 573 features were left for analysis. The distribution of >90% of the 573 features deviated from the normal, reflecting in some cases different batch distributions. The features therefore underwent log<sub>2</sub> transformation and batch effect correction, resulting in >95% metabolites following a normal distribution. Principal components analysis showed that all 20 QC samples clustered together (Supplementary Figure 2), verifying reproducible results across the batches and highlighting the presence of one outlier (AD case), which was removed. KNN imputation was performed on the 123 samples with the 573 features.

### Single analyte analyses results

Logistic regression analyses for AD versus controls, adjusting for the number of APOE ε4 alleles, age, gender and batch, indicated that 95 analytes were associated with AD at *P*-value < 0.05 and 41 analytes at *q*-value < 0.05. Results for all 573 analytes and all pairwise comparisons are provided in Supplementary Table 1. Secondary models investigating for APOE ε4 specific associations by testing for interactions between analytes and the APOE ε4 allele indicated that the association of 29 analytes with AD seemed to be modified for the APOE ε4 at *P*-value < 0.05; However, none of these associations were significant at *q*-value < 0.05 (results not presented).

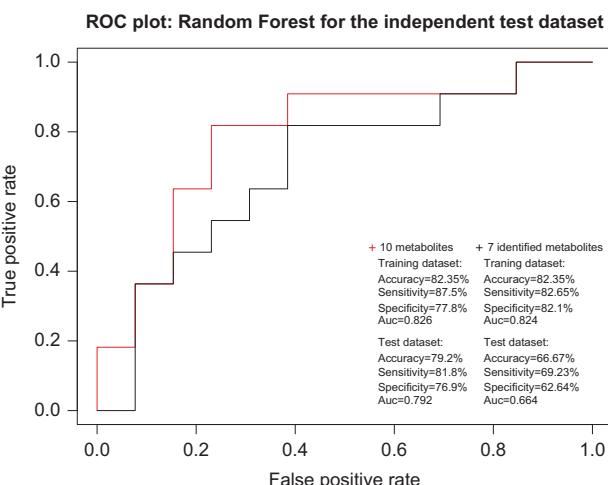
### Multivariate analysis results

Random Forest was performed on the training data set using the 573 features (100 bootstraps). Supplementary Figure 3 displays the variable importance after 100 bootstraps, with the dotted line showing a slight leveling off in the importance measure and which corresponds to the top 10% features (*n* = 57). Random Forest with 'rfe' on the training data set showed that the highest mean training performance was for a model with 25 features. However,

to choose a model with high accuracy while reducing the number of features as low as possible, a model with 10 molecules was chosen (AUC = 0.867) (Supplementary Figure 4). We fitted the Random Forest model with the 10 selected variables on the whole training data set, which predicted the training data set with 82.35% accuracy (sensitivity = 87.5%, specificity = 77.8%, AUC = 0.826) and could classify our test data set with 79.2% accuracy, 81.8% sensitivity, 76.9% specificity, a positive predictive value of 75.0%, a negative predictive value of 83.3% and an AUC of 0.792.

When we repeated the analysis including APOE, APOE had very low variable importance (222 out of 573) and was not selected forward. In addition, APOE was not selected in the final model after forcing it during 'rfe' together with the top 10% of the selected analytes from the 100 Random Forest bootstraps.

Results for the Random Forest model are displayed in Figure 1. Nine of the 10 metabolites of the classifier were associated with AD in single analyte regression analysis at *q* < 0.05 (Supplementary Table 1 and Table 2) and none of them was shown to interact with the APOE ε4 genotype at *P* < 0.05.



**Figure 1.** Receiver operator curve (ROC) for the independent test data set for the selected 10 metabolites after recursive feature elimination for AD versus controls and for the seven identified metabolites, and summaries of the classifier models for the training and test data sets. AD, Alzheimer's disease; AUC, area under curve.

**Table 1.** Sample demographics

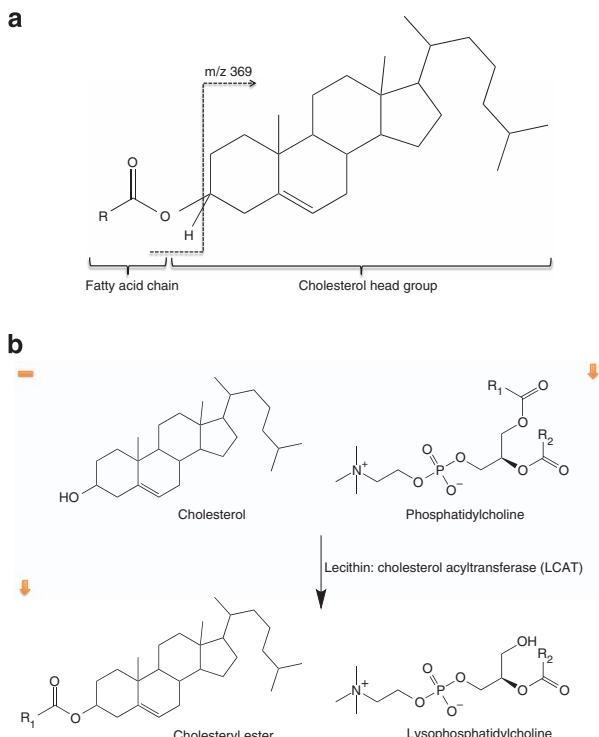
	Comprehensive plasma LC-MS lipidomics (n = 124)			
	Control	MCI	AD	ANOVA F(df1,df2) or x2(df) tests
<i>N</i>	40	48	36	
Age (mean, s.d.)	78.46 (6.7)	78.96 (5.6)	78.14 (7.7)	F = 0.180 (2,121), <i>P</i> = 0.836
MMSE (mean, s.d.)	29.00 (1.1)	26.94 (1.9)	21.49 (4.8)	F = 65.46(2,121), <i>P</i> < 2.0E – 16 <sup>a</sup>
Female/male (N)	17/18	26/22	21/19	$\chi^2$ = 0.259 (2), <i>P</i> = 0.879
APOE ε4 alleles (0/1/2) (N)	33/7/0	28/12/3	15/15/6	$\chi^2$ = 12.812 (2), <i>P</i> < 16.5E – 03 <sup>b</sup>
Batch (1/2/3/4) (N)	9/10/9/12	14/11/10/13	9/8/10/9	$\chi^2$ = 1.256 (2), <i>P</i> = 0.974
Diabetes <sup>c</sup>	1	3	3	$\chi^2$ = 0.749 (2), <i>P</i> = 0.688
Smoking <sup>d</sup>	6	0	9	$\chi^2$ = 2.729 (2), <i>P</i> = 0.255
Statins	10	20	14	$\chi^2$ = 3.552 (2), <i>P</i> = 0.169
Samples (LON)	40	28	36	NA
Samples (EUR)	0	20	0	NA

Abbreviations: AD, Alzheimer's disease; ANOVA, analysis of variance; EUR, samples obtained from the non-London AddNeuroMed European centres; LC-MS, liquid chromatography-mass spectrometry; LDN, samples obtained from London AddNeuroMed- and DCR-based patients; MCI, mild cognitive impairment; MMSE, Mini-Mental State Examination score; NA, not available. <sup>a</sup>Tukey's honest significant differences post hoc tests: AD versus Control *P* < 1.0E – 17; AD versus MCI *P* = 4.02E – 13; MCI versus Control *P* = 3.36E – 03. <sup>b</sup>AD versus Control *P* = 7.5E – 04; AD versus MCI *P* = 0.1364; MCI versus Control *P* = 0.099. <sup>c</sup>Diabetes information was available for 34 AD patients, 29 controls and 43 MCI. <sup>d</sup>Information on smoking was available for 34 AD patients, 29 controls and 6 MCI.

**Table 2.** List of metabolite molecules selected by the Random Forest classifier (AD versus elderly health controls)

Metabolite molecule	Peak height data normalized to whole mean only			Peak ratios normalized to internal standard		
	OR	95% CI	P-value	OR	L95 95% CI	P-value
Mass/z 367	0.124	0.03–0.39	1.31E–03	0.115	0.03–0.36	8.09E–04
ChE 32:0	0.251	0.10–0.52	7.51E–04	0.237	0.10–0.48	4.19E–04
ChE 34:0	0.151	0.05–0.38	3.14E–04	0.152	0.05–0.37	2.90E–04
ChE 34:6	0.231	0.08–0.53	1.56E–03	0.126	0.03–0.35	5.40E–04
ChE 32:4	0.141	0.04–0.43	1.75E–03	0.056	0.01–0.24	6.56E–04
ChE 33:6	0.218	0.07–0.55	3.15E–03	0.205	0.06–0.50	2.21E–03
Mass/z 628	3.669	1.20–13.93	3.49E–02	NA	NA	NA
Mass/z 906	0.210	0.07–0.50	1.40E–03	0.226	0.09–0.48	3.73E–04
Mass/z 315	5.084	1.78–16.88	4.12E–03	NA	NA	NA
ChE 40:4	0.362	0.18–0.67	2.56E–03	0.279	0.12–0.55	6.43E–04
Cholesterol	NA	NA	NA	0.316	0.02–5.01	4.21E–01

Abbreviations: AD, Alzheimer's disease; ChE, cholesteryl ester; CI, confidence interval; NA, not available; OR, odds ratio. Results are presented for peak height and peak ratio (measurement normalized to internal standard) after covariate adjustment. The metabolites are ordered according to their importance during recursive feature elimination. All metabolites except for mass 628 were associated with AD in single analyte regression analysis ( $q < 0.05$ ). Results for semiquantified mass 628 and mass 315 are not presented as they were not consistently above the limit of quantification. Results are also presented for the APOE ε4 allele and for cholesterol peak ratio. The association of APOE ε4 was OR = 5.620, 95% CI = 2.27–16.29,  $P = 5.48E - 04$  per ε4 allele.



**Figure 2.** Cholesteryl ester molecules structure (a) and synthesis from cholesterol catalysed by lecithin: cholesterol acyl transferase (LCAT) (b). ChE is characterized by the presence of cholesterol head group in ESI(+) of 369.

#### Identification and measurement of putative biomarkers

Six molecules were identified using MS/MS fragmentation patterns (Figure 2a) and found to be cholesteryl esters (ChEs) with very long chain fatty acids,<sup>29,30</sup> synthesized from cholesterol (Figure 2b). These were  $m/z$  866 (ChE 32:0),  $m/z$  894 (ChE 34:0),  $m/z$  882 (ChE 34:6),  $m/z$  856 (ChE 32:4),  $m/z$  868 (ChE 33:6) and  $m/z$  970 (ChE 40:4). In addition, one molecule was proposed as an oxidized form of desmosterol following the expected fragmentation pattern ( $m/z$  367). The six identified ChE molecules and the desmosterol-related molecule predicted the training data set with 82.35% accuracy (sensitivity = 82.61%, specificity = 82.14%, AUC =

0.824) and could classify the test data set with 67% accuracy, 69.23% sensitivity, 62.64% specificity, a positive predictive value of 69.23%, a negative predictive value of 63.64% and an AUC of 0.670.

Relative quantification was produced for eight of the molecules, which were consistently over the limit of quantification (LOQ). Univariate association analysis of the eight peak ratios revealed that the associations of all of them were strengthened. We, additionally, measured cholesterol, the precursor of ChE, finding no association with AD (Table 2).

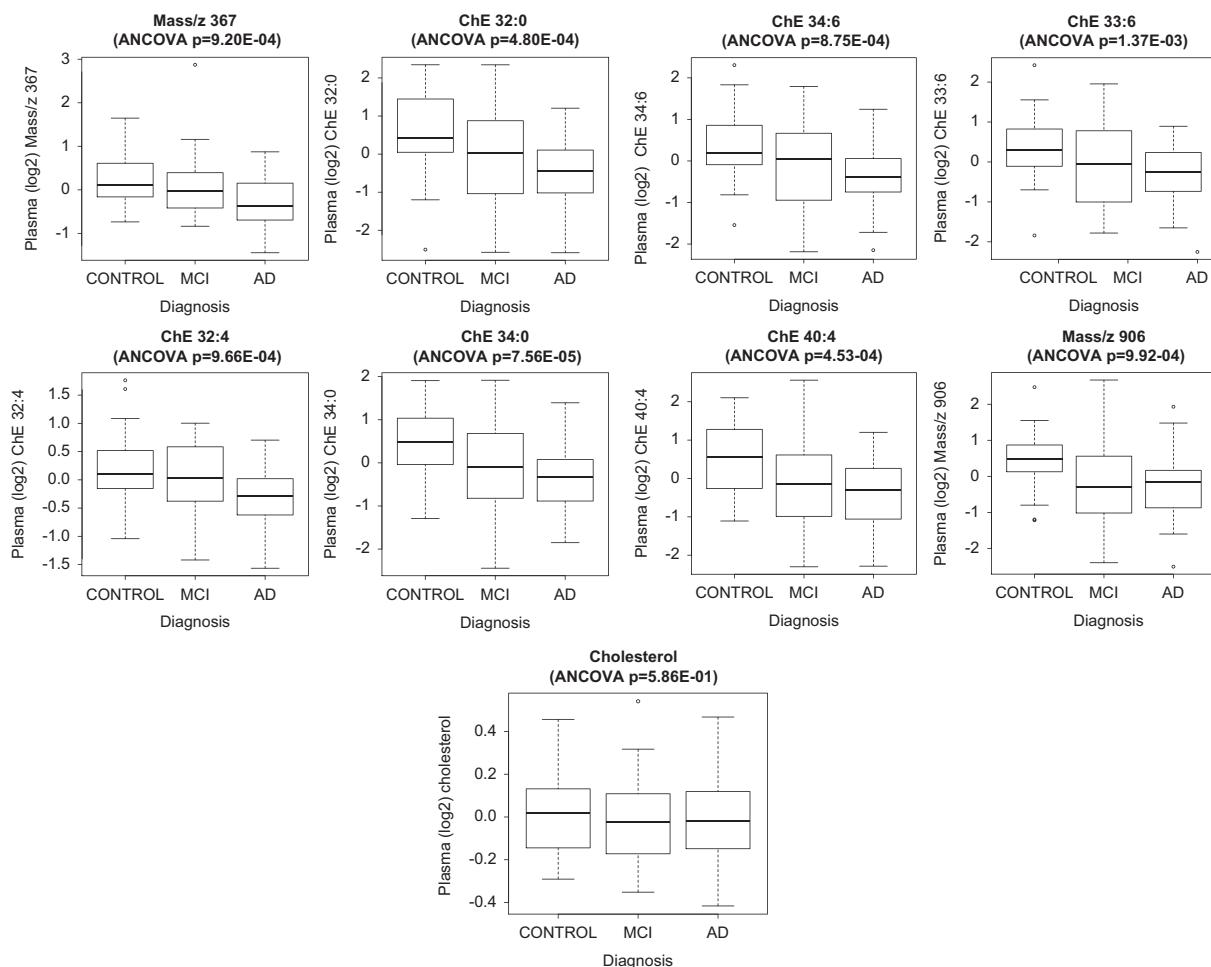
Figure 3 plots the levels of the metabolites and cholesterol in the three diagnostic groups, which highlights the overall decrease pattern in metabolite levels from healthy controls to MCI to AD. To further investigate the association of the metabolites with cognition, we investigated their correlation with MMSE, which showed a weak association (Supplementary Figure 5).

When we tested our classifier on individuals with MCI, 22 were classified as AD and 18 as controls.

Finally, logistic regression analyses including statin use and smoking status as covariates produced identical results (not presented).

#### DISCUSSION

We performed lipidomics on plasma samples from 36 late-onset AD patients, 40 healthy controls and 48 individuals with MCI to identify metabolites that differentiate between AD patients and healthy controls. Univariate analysis identified 41 analytes associated with AD ( $q < 0.05$ ). We then applied Random Forest and a backwards elimination approach to identify a reproducible metabolic signature to classify AD patients and we found a combination of 10 metabolite molecules that classified AD patients in the independent test data set with 79% accuracy. This is one of the largest non-targeted plasma studies to date to use such a systematic analysis pipeline, which included assessment of the model in an independent data set (1/3 of sample), to identify new targets linked to AD and provide improvement to current diagnostic classifiers.<sup>11–16</sup> Although APOE was associated with AD in single analyte analyses, its variable importance in the presence of the metabolites during classification was low and was not included in the classifier probably due to the metabolites already capturing the information that APOE provides. It should be noted that the single metabolite logistic regression results indicated that the metabolites provided information over and above APOE.



**Figure 3.** Boxplots depicting change in the level of the Random Forest measured molecules which were consistently above the level of quantification (LOQ) and that of cholesterol in the three diagnostic groups. All molecules were decreased in AD compared with controls after putative biomarker measurement. A decrease is also observed in MCI compared with controls. The ANCOVA *P*-value for the difference in metabolite levels in the three groups is displayed above each graph. AD, Alzheimer's disease; ANCOVA, analysis of covariance; MCI, mild cognitive impairment.

Most of the metabolites in the classifier were reduced in AD patients. We observed that their levels in MCI subjects were more variable but overall followed the control > MCI > AD trend with some of them being different at  $P < 0.05$  between AD and MCI or MCI and controls. To investigate this further, we plotted the correlation of metabolites with MMSE, which is a more informative continuous surrogate marker for disease status and observed weak association between the metabolic signature and cognition.

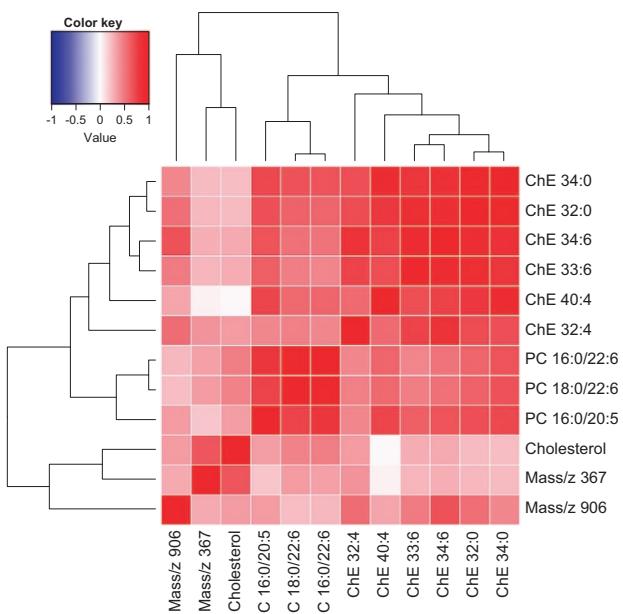
We subsequently extracted single metabolite measures and observed that their associations with AD were strengthened (Table 2). Six molecules were identified as ChEs, molecules which have not been previously associated with AD and one as a potentially oxidized form of desmosterol, which could predict the test data set with 67% accuracy highlighting both their good predictive value and also the importance of the unidentified molecules in AD. We, additionally, measured cholesterol, as it is the precursor of ChE, and wanted to explore the chemical path differences with AD, finding no association. This suggested that the association with AD is specific to metabolites synthesized from cholesterol rather than to cholesterol itself.

A heatmap of the measured molecules together with cholesterol is shown in Figure 4. We have previously reported three PCs to be decreased in AD versus controls<sup>16</sup> using overlapping samples. Although these PCs were not included in our classifier, their raw values were decreased in our data set ( $P < 0.05$ ) with one

of them (mass 780 which is PC16:0/20:5) being associated with AD at  $q < 0.05$ . These molecules were also recently identified as markers of phenoconversion to either amnestic MCI or AD.<sup>14</sup> We have therefore included these three PCs in the heatmap to compare their biological variation with the newly discovered and biochemically related ChEs, which highlights the positive correlation between the metabolites identified here, cholesterol and the three PCs; therefore, it is likely that the information the PCs provide is captured by the analytes selected in the rfe step.

#### Role of cholesterol and its derivatives

Cholesterol esters are largely synthesized in plasma by the transfer of fatty acids to cholesterol from PC, a reaction catalysed by the enzyme LCAT (lecithin: cholesterol acyl transferase). Free cholesterol can be taken up by APOE-containing lipoproteins, such as HDL, but is confined to the outer surface of the particle. The esterification of cholesterol to ChE ensures that more cholesterol is packaged into the interior of lipoproteins and this increases their capacity, allowing more efficient cholesterol transport through the bloodstream. LCAT has a preference for plasma 16:0–18:2 or 18:0–18:2 PCs, therefore, connecting the PCs identified by Whiley et al.<sup>16</sup> and the ChEs identified here through a one-step enzymatic reaction (Figure 2b). Interestingly, LCAT is also expressed in the brain by astrocytes, and, together with APOE and ABCA1, has a key role in the maturation of glial-derived nascent lipoproteins.



**Figure 4.** Heatmap of the eight measured metabolites selected during Random Forest classification, following recursive feature elimination, which were consistently above the level of quantification (LOQ). Cholesterol and the three phosphatidylcholines (PCs) previously published by Whiley *et al.*<sup>16</sup> are also included. AD, Alzheimer's disease; ANCOVA, analysis of covariance; MCI, mild cognitive impairment.

In other tissues, cholesterol is esterified by acyl-coenzyme A: cholesterol acyltransferase 1 and 2 (ACAT1 and ACAT2). ACAT uses acyl-CoA as a source for the acyl chains. Recent evidence suggests a strong link of ACAT1 with amyloid deposition. Pharmacological inhibition of ACAT in AD mice resulted in diminished amyloid plaque burden in their brains and improved cognitive function<sup>25,31</sup> and genetic ablation of Acat1 in AD mice diminished the levels of Aβ42, decreased the amyloid plaque burden and full-length human APP (human amyloid precursor protein), and improved the cognitive function of the mice.<sup>32</sup> Finally, a recent gene therapy study showed that adeno-associated virus targeting Acat1 for gene knockdown delivered to the brains of AD mice decreased the levels of total brain amyloid-β, oligomeric amyloid-β and full-length human APP to levels similar to complete genetic knockdown of Acat1.<sup>33</sup>

Approximately one-third of the ChE is transferred from HDL to APOB-containing lipoproteins, such as VLDLs, in exchange for triglycerides by ChE transfer protein. Thus most of the cholesterol in circulating lipoproteins is ChE produced from HDL lipoproteins by the LCAT-catalysed reaction. This process results in lower HDL cholesterol and indirectly decreases HDL size as frequently observed in type 2 diabetes<sup>34</sup> and a deficiency of ChE transfer protein is associated with increased HDL and decreased LDL levels, a profile typically antiatherogenic.<sup>35,36</sup> Interestingly, an elevation of cholesterol esters (chains 14:20) was observed in the brains of AD patients and of three transgenic familial AD mouse models.<sup>24</sup>

Lower desmosterol levels have been previously found in the plasma and brains of AD patients compared with healthy controls.<sup>13,26</sup> Desmosterol is a precursor of cholesterol and seladin (*DHCR24*), which governs the metabolism of desmosterol to cholesterol in specific brain areas has been shown to counteract the β-secretase cleavage of APP and the formation of amyloid-β.<sup>37,38</sup> An unknown molecule with similar structure and the same mass as desmosterol shows the highest correlation with cholesterol here.

## Strengths and limitations

To our knowledge, ours is the first study to implicate ChEs in AD. We should highlight, however, that these molecules are biochemically related to metabolites previously associated with AD initiation and progression and are, therefore, biologically relevant. Although lipid molecules have been previously implicated in AD, results are not always consistent. It is worth noting that even if previously published studies go under the umbrella of plasma lipidomics and/or metabolomics, differences are considerable in terms of patient cohort selection (such as differences in patient groups selection, in age and in dementia severity), in terms of technical treatment (such as sampling/storage conditions and fasting state before sample collection) and in clinical data availability. As an example, sample groups can differ from a simple case versus control design<sup>11</sup> to including MCI<sup>12–16</sup> and in one case including time points with pre- and post-conversion patients.<sup>14</sup> Sample groups will influence the selectivity of results, as well as patient numbers, which ranged from <50 in early studies<sup>11</sup> to the low hundreds in more recent studies.<sup>13–16</sup> In addition, LC-MS fingerprinting methods can be qualitative or/and quantitative, hence the information can be relative amounts<sup>11,12,15</sup> or a combination of relative and absolute amounts for chosen metabolites.<sup>13,14,16</sup> Another factor for differences observed between studies would be the different data processing tools applied for data mining and statistical methods used to produce AD biomarker predictive models. Some studies showed predictive models of AD based on one to three molecules<sup>13,16</sup> or a battery ranging from three to 10 markers and have used a wide range of statistical analysis methods ranging from simple single metabolite approaches<sup>11</sup> to methods using multivariate regression approaches and resampling techniques,<sup>15</sup> and validation of the panels in independent data sets.<sup>14</sup> Here, we have used a well-characterized AD cohort matched for age and gender and performed a careful and systematic analysis pipeline to extract metabolites associated with AD, by performing bootstrapping to avoid over-fitting and validating our results in an unseen data set. After the molecules associated with AD were identified, an additional method of metabolite quantification in all the samples was performed, which was possible due to internal standards and a run separating lipids, which acquired data for 2 h for each sample.

Our study also suffers from potential limitations. Although ours is one of the largest AD metabolomics studies to date, the sample size is still modest and replication is required in larger cohorts. In addition, this study suffers from limitations inherent to AD case-control studies, such as the possibility that some of the healthy elderly controls may already carry pathology, and that some of the clinically diagnosed AD may be pathologically non-AD dementias. Through follow-up data on the individuals used in this study, however, we know that all AD patients used for our analysis maintained the diagnosis of AD as did all controls for at least 3 years from their baseline visit. We also believe that achieving such high performance in both training and test data sets for an AD case-control data set highlights the efficacy of the classifier, as AD diagnostic classifiers rarely achieve such high performance; we, therefore, believe that having additional pathology information would only increase its performance. Furthermore, we need to acknowledge that individuals with MCI are more heterogeneous in pathology; however, the MCI subjects used in this study were recruited primarily from local memory clinics and were, therefore, expected to be composed largely of subjects with likely an AD end point.

Finally, due to the large number of comorbidities in old age, we have to acknowledge that our metabolite signal may not be AD specific but it could be associated with overall ill health and other comorbid conditions, making it potentially a not good biomarker for recruitment in clinical trials. Investigating and

comparing the metabolic profiles of other disorders would increase the specificity of our panel.

## CONCLUSIONS

In this study, we used a Random Forest approach and identified a combination of 10 metabolites, which predicted AD with near 80% accuracy. We subsequently identified six of the metabolites to be ChEs, molecules not previously implicated in AD, which are connected to PCs through a one-step enzymatic reaction. The newly identified molecules were reduced in AD patients compared with controls. All these, combined with the lack of association between cholesterol and AD, suggest that it is the dysregulation of specific steps in cholesterol metabolism, rather than cholesterol itself, that is responsible for these observations and suggest novel targets for future work. These findings need to be replicated in larger well-phenotyped cohorts, which will be possible in the near future through the large biomarker consortia being set up. In addition, information on pathology status, such as amyloid marker cerebrospinal fluid, PET or brain atrophy measurements, through magnetic resonance imaging (MRI), would provide more precise phenotypes for biomarker discovery and would capture different stages of disease pathology, and, the comparison of metabolite levels between MCI patients who converted to AD and those who remained stable would provide us with metabolites, which are associated with disease initiation. Finally, integrating additional types of biological modalities, such as protein, gene expression and genotype information, will help investigate the origin of ChE dysregulation in AD.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

We thank the individuals and families who took part in this research. We acknowledge the use of the computational Linux cluster and the Biomedical Research Centre Nucleus Informatics Team supported by National Institute for Health Research (NIHR) Mental Health Biomedical Research Centre and Dementia Unit at South London and Maudsley NHS Foundation Trust and (Institute of Psychiatry, Psychology and Neuroscience) King's College London. This work was also supported by the 7th Framework Programme of the European Union (ADAMS project, HEALTH-F4-2009-242257). AddNeuroMed was funded through the EU FP6 programme. PP is an Alzheimer's Society post-doctoral fellow. The funding bodies had no role in the design, collection, analysis, interpretation of the data, writing of the manuscript and the decision to submit the manuscript.

## REFERENCES

- 1 Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M et al. Global prevalence of dementia: a Delphi consensus study. *Lancet* 2005; **366**: 2112–2117.
- 2 Whiley L, Legido-Quigley C. Current strategies in the discovery of small-molecule biomarkers for Alzheimer's disease. *Bioanalysis* 2011; **3**: 1121–1142.
- 3 Samson K. MEDICARE panel: Too little evidence to show amyloid PET scans improve Alzheimer outcomes. *Neurol Today* 2013; **13**: 36–42.
- 4 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**: 5992–5999.
- 5 Lindon JC, Holmes E, Nicholson JK. Metabonomics in pharmaceutical R&D. *FEBS J* 2007; **274**: 1140–1151.
- 6 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.
- 7 Suhre K, Meisinger C, Doring A, Altmaier E, Belcredi P, Gieger C et al. Metabolic footprint of diabetes: a multiplatform metabolomics study in an epidemiological setting. *PLoS One* 2010; **5**: e13953.
- 8 Barba I, Fernandez-Montesinos R, Garcia-Dorado D, Pozo D. Alzheimer's disease beyond the genomic era: nuclear magnetic resonance (NMR) spectroscopy-based metabolomics. *J Cell Mol Med* 2008; **12**: 1477–1485.
- 9 Lunnon K, Ibrahim Z, Proitsi P, Lourdusamy A, Newhouse S, Sattlecker M et al. Mitochondrial dysfunction and immune activation are detectable in early Alzheimer's disease blood. *J Alzheimers Dis* 2012; **30**: 685–710.
- 10 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**: 977–986.
- 11 Han X, Rozen S, Boyle SH, Hellegers C, Cheng H, Burke JR et al. Metabolomics in early Alzheimer's disease: identification of altered plasma sphingolipidome using shotgun lipidomics. *PLoS One* 2011; **6**: e21643.
- 12 Greenberg N, Grassano A, Thambisetty M, Lovestone S, Legido-Quigley C. A proposed metabolic strategy for monitoring disease progression in Alzheimer's disease. *Electrophoresis* 2009; **30**: 1235–1239.
- 13 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**: 567–576.
- 14 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* 2010; **20**: 415–4182014.
- 15 Oresic M, Hytönen T, Herukka SK, Sysi-Aho M, Mattila I, Seppanen-Laakso T et al. Metabolome in progression to Alzheimer's disease. *Transl Psychiatry* 2011; **1**: e57.
- 16 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**: 271–278.
- 17 Reitz C. Dyslipidemia and the risk of Alzheimer's disease. *Curr Atheroscler Rep* 2013; **15**: 307.
- 18 Reitz C, Mayeur R. Alzheimer disease: epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochem Pharmacol* 2014; **88**: 640–651.
- 19 Harold D, Abraham R, Hollingworth P, Sims R, Gerrish A, Hamshere ML et al. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 2009; **41**: 1088–1093.
- 20 Hollingworth P, Harold D, Sims R, Gerrish A, Lambert JC, Carrasquillo MM et al. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 2011; **43**: 429–435.
- 21 Lambert JC, Heath S, Even G, Campion D, Sleegers K, Hiltunen M et al. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nat Genet* 2009; **41**: 1094–1099.
- 22 Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**: 1452–1458.
- 23 Astarita G, Jung KM, Vasilevko V, Dipatrizio NV, Martin SK, Cribbs DH et al. Elevated stearoyl-CoA desaturase in brains of patients with Alzheimer's disease. *PLoS One* 2011; **6**: e24777.
- 24 Chan RB, Oliveira TG, Cortes EP, Honig LS, Duff KE, Small SA et al. Comparative lipidomic analysis of mouse and human brain with Alzheimer disease. *J Biol Chem* 2012; **287**: 2678–2688.
- 25 Hejazi L, Wong JW, Cheng D, Proschoglo N, Ebrahimi D, Garner B et al. Mass and relative elution time profiling: two-dimensional analysis of sphingolipids in Alzheimer's disease brains. *Biochem J* 2011; **438**: 165–175.
- 26 Wisniewski T, Newman K, Javitt NB. Alzheimer's disease: brain desmosterol levels. *J Alzheimers Dis* 2013; **33**: 881–888.
- 27 Lovestone S, Francis P, Strandgaard K. Biomarkers for disease modification trials—the innovative medicines initiative and AddNeuroMed. *J Nutr Health Aging* 2007; **11**: 359–361.
- 28 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; **8**: 118–127.
- 29 Butovich IA. Fatty acid composition of cholesterol esters of human meibomian gland secretions. *Steroids* 2010; **75**: 726–733.
- 30 Butovich IA, Uchiyama E, McCulley JP. Lipids of human meibum: mass-spectrometric analysis and structural elucidation. *J Lipid Res* 2007; **48**: 2220–2235.
- 31 Hutter-Paier B, Huttunen HJ, Puglielli L, Eckman CB, Kim DY, Hofmeister A et al. The ACAT inhibitor CP-113,818 markedly reduces amyloid pathology in a mouse model of Alzheimer's disease. *Neuron* 2004; **44**: 227–238.
- 32 Bryleva EY, Rogers MA, Chang CC, Buen F, Harris BT, Rousselet E et al. ACAT1 gene ablation increases 24(S)-hydroxycholesterol content in the brain and ameliorates amyloid pathology in mice with AD. *Proc Natl Acad Sci USA* 2010; **107**: 3081–3086.
- 33 Murphy SR, Chang CC, Dogbevia G, Bryleva EY, Bowen Z, Hasan MT et al. Acat1 knockdown gene therapy decreases amyloid-beta in a mouse model of Alzheimer's disease. *Mol Ther* 2013; **21**: 1497–1506.
- 34 de Vries R, Borggreve SE, Dullaart RP. Role of lipases, lecithin:cholesterol acyltransferase and cholesterol ester transfer protein in abnormal high density lipoprotein metabolism in insulin resistance and type 2 diabetes mellitus. *Clin Lab* 2003; **49**: 601–613.

- 35 Barter P, Rye KA. Cholesteryl ester transfer protein inhibition to reduce cardiovascular risk: where are we now? *Trends Pharmacol Sci* 2011; **32**: 694–699.
- 36 Barter PJ, Brewer HB Jr, Chapman MJ, Hennekens CH, Rader DJ, Tall AR. Cholesteryl ester transfer protein: a novel target for raising HDL and inhibiting atherosclerosis. *Arterioscler Thromb Vasc Biol* 2003; **23**: 160–167.
- 37 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**: 3899–3906.
- 38 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**: 432–443.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Supplementary Information accompanies the paper on the Translational Psychiatry website (<http://www.nature.com/tp>)