

1 **Trajectory of plasma lipidomes associated with the risk of late-onset Alzheimer's disease pathogenesis: a**
2 **longitudinal study in the ADNI cohort**

3 Tingting Wang^{1,2*}, Matthias Arnold^{3,4*}, Kevin Huynh^{1,2,5*}, Patrick Weinsich⁴, Corey Giles^{1,2,5}, Natalie A
4 Mellett¹, Thy Duong¹, Bharadwaj Marella⁴, Kwangsik Nho^{6,7,8}, Alysha De Livera⁹, Xianlin Han¹⁰, Colette
5 Blach¹¹, Andrew J Saykin^{6,8}, Gabi Kastenmüller^{4**}, Peter J Meikle^{1,2,5**}, Rima Kaddurah-Daouk^{3,12,13**}, for
6 the Alzheimer's Disease Neuroimaging Initiative[#]

7 ¹Baker Heart and Diabetes Institute, Melbourne, Australia.

8 ²Baker Department of Cardiometabolic Health, University of Melbourne, Victoria, Australia.

9 ³Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA

10 ⁴Institute of Computational Biology, Helmholtz Zentrum München, German Research Center for
11 Environmental Health, Neuherberg, Germany

12 ⁵Department of Cardiovascular Research Translation and Implementation, La Trobe University, Melbourne,
13 Australia.

14 ⁶Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis IN,
15 USA

16 ⁷Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis,
17 IN, USA

18 ⁸Indiana Alzheimer Disease Center, Indiana University School of Medicine, Indianapolis, IN, USA.

19 ⁹School of Computing, Engineering and Mathematical Sciences, La Trobe University, Melbourne, Australia

20 ¹⁰Barshop Institute for Longevity and Aging Studies, University of Texas Health Science Center at San
21 Antonio, San Antonio, TX, USA

22 ¹¹Duke Molecular Physiology Institute, Duke University, Durham, NC, USA.

23 ¹²Duke Institute of Brain Sciences, Duke University, Durham, NC, USA

24 ¹³Department of Medicine, Duke University, Durham, NC, USA

25 *Equal first author

26 **Equal senior and corresponding author

27 [#]Data used in preparation of this article were generated by the Alzheimer's Disease Metabolomics
28 Consortium (ADMC). As such, the investigators within the ADMC provided data but did not participate in
29 analysis or writing of this report. A complete listing of ADMC investigators can be found
30 at: <https://sites.duke.edu/adnimetab/team/>
31

32 **Abstract**

33 Comprehensive lipidomic studies have demonstrated strong cross-sectional associations between the blood
34 lipidome and late-onset Alzheimer's disease (AD) and its risk factors. However, the longitudinal relationship
35 between the lipidomic variations and progression of AD remains unknown. Here, we employed longitudinal
36 lipidomic profiling on 4,730 plasma samples from 1,517 participants of the Alzheimer's Disease
37 Neuroimaging Initiative (ADNI) cohort to investigate the temporal evolution of lipidomes among diagnostic
38 groups. At baseline, there were 1,393 participants including 437 cognitively normal (CN), 713 with mild
39 cognitive impairment (MCI), and 243 AD cases. During follow up, 329 individuals (29 CN and 300 MCI)
40 developed clinical AD (AD converters). We developed an AD-CN classification model to stratify the non-
41 converting MCI group into AD-like and non AD-like MCI based on their lipidomics profiles at baseline.
42 Longitudinal analysis identified associations between the change in ether lipid species (including
43 alkylphosphatidylcholine, alkenylphosphatidylcholine, lysoalkylphosphatidylcholine, and
44 lysoalkenylphosphatidylcholine) in converters relative to non-converting CN and MCI groups. Further, the
45 AD-CN model efficiently classified MCI into low AD risk and high AD risk, with the high AD risk group having
46 two times higher risk of conversion to AD than the low risk group. These findings suggest that the lipidomic
47 profile can serve as a potential biomarker to identify individuals at higher risk for progressing to AD.

48 **Introduction**

49 Late-onset Alzheimer's disease (AD) is the leading cause of dementia, characterised by the progressive
50 death of neurons and loss of brain structure, usually presenting with memory loss^{1,2}. Many risk factors
51 have been identified to collectively modulate risk for AD, with advanced age (≥ 65 years) being the

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

52 strongest risk factor. Moreover, common genetic risk factors are associated with increased risk³, such as
53 the *APOE* ϵ 4 allele and sex, with females being more likely to develop AD (especially at age \geq 80 years)⁴.

54 Spanning a period of 15-25 years, individuals with AD progress from cognitively normal(CN) through mild
55 cognitive impairment (MCI) to overt dementia². As a transitional state between CN and dementia, MCI has
56 mixed aetiologies with different pathologies, neuropsychological profiles, or biomarker anomalies, often
57 presenting with subtle to mild clinical symptoms^{5,6}. The transition from MCI to dementia can take a varying
58 length of time, with some individuals remaining stable or reverting to CN. The underlying molecular
59 mechanisms contributing to this heterogeneity remain unknown. Accurate stratification of MCI using the
60 molecular level information (such as lipidomic profiling) has the potential to improve the prognostic accuracy
61 at the early stages of disease⁷, which is critical for streamlining clinical trials to shorten drug development
62 cycle and avoid negative results due to this heterogeneity.

63 A growing number of studies have defined an intimate link between the plasma lipidome, measured at a
64 single point in time, and AD⁸⁻¹¹ or AD-related risk factors¹²⁻¹⁷. Besides, baseline lipid metabolic changes in AD
65 patients were also demonstrated to be closely related with cerebrospinal fluid pathology markers, imaging
66 features, and cognitive performance¹⁸. However, the plasma lipidome is highly dynamic, varying in response
67 to environmental exposures (diet, physical activity)¹⁹⁻²¹ and over the longer term with age²²⁻²⁵. Similarly, the
68 progression of cognitive impairment may predispose individuals to lifestyle changes (unbalanced diet or
69 physical inactivity)^{26,27}. These changes will influence peripheral lipid metabolism and may therefore appear
70 to be associated with disease in a cross-sectional analysis (referred to as reverse causation). Longitudinal
71 studies can minimise the impact of reverse causation by defining the relationship between changes in the
72 plasma lipidome prior to AD, during the progression to AD, and in response to AD.

73 In this study, we performed longitudinal analysis of plasma lipidomic profiles in the Alzheimer's Disease
74 Neuroimaging Initiative (ADNI)-1, -GO and -2 cohorts to delineate the relationships between peripheral lipid
75 metabolism and progression to AD. Using this complex data, we also assessed the utility of plasma lipids to
76 identify MCI individuals at high risk of converting to AD.

77 **Results**

78 For this study, we profiled a total of 4,730 longitudinal plasma samples from 1,517 participants of ADNI -1, -
79 GO and -2 cohorts examined from baseline up to the 13th time point (10th years follow up period), with three
80 major time points of baseline, 12 months, and 24 months that include the largest number of individuals
81 (Figure 1). After quality control, the lipidomics profiles on all these measurements consist of 749 lipid species
82 from 48 lipid classes (Supplementary Table 1).

83 ***Longitudinal definition of diagnostic groups***

84 We defined individuals as AD converters if they had a baseline diagnosis of cognitive normal (CN) or mild
85 cognitive impairment (MCI) at the time that they entered the study but progressed to AD at a later time
86 point. In total, there were 363 AD converters progressing from CN or MCI to AD at specific time points. In
87 Supplementary Table 2, we detailed the distributions of AD converters across different time points. At
88 baseline, after removing 25 missingness (detailed in the Method), we had 1,393 individuals with 437 CN, 713
89 MCI, and 243 AD cases. Out of these, there were 29 CN and 300 MCI who converted to AD at later time
90 points (termed as AD converters). We also observed that 71 CN converted to MCI and remained MCI at
91 subsequent time points, which we termed as MCI converters.

92 In Table 1, the difference in the demographics, numbers and distributions of the key risk factors (covariates)
93 among diagnostic classifications were examined on three main time points (Baseline, 12 months, and 24
94 months), using Fisher's exact test for categorical variables or ANOVA for continuous variables.

95 ***Metabotype conservation over time among AD state groups***

96 To globally assess changes in lipidomic profiles over time, we calculated metabotype conservation indices
97 (I_c) for a subset of 755 participants that had lipidomics data available at baseline as well as at follow up visits
98 after 12 and 24 months. The I_c reflects the lipidomics-based self-similarity of an individual over time
99 compared to all other individuals. To avoid potential bias introduced by the strong correlation structure

100 observed in lipidomics data, we clustered highly correlated lipids and calculated eigenlipids for each of the
101 205 obtained clusters. Eighty lipids were not assigned to a cluster and retained as separate variables. Based
102 on these 285 variables, we calculated the I_c . As expected, the I_c after 24 months was overall significantly
103 lower than after 12 months (paired t-test $p = 1.06 \times 10^{-3}$, Wilcoxon test $p = 4.56 \times 10^{-6}$); Figure 2 and
104 Supplementary Table 3). I_c between baseline and 24-month follow-up showed overall high stability of the
105 lipidome, with more than half of individuals reaching the maximum I_c of 1 ($n = 401$). There was no significant
106 difference in the proportion of individuals with an $I_c < 1$ between diagnostic groups (Supplementary Table
107 4). However, comparing distributions of the I_c values smaller than one between diagnostic groups showed
108 the highest conservation in the CN group, with significantly lower levels (Wilcoxon test $p = 0.0039$) observed
109 in cases with AD (Figure 2b).

110 ***Consistent cross-sectional associations of individual lipid species with AD at different time points***

111 At baseline, we observed 192 nominally significant associations between lipid species and AD, relative to CN,
112 after FDR corrections (Figure 3, Supplementary Table 5). The majority of these associations were consistent
113 with previous findings²⁸. Using the 24-month data, there were 181 significant associations, whereas only 52
114 lipid species were significantly associated with the 12-month data (Figure 3). Among 192 associations
115 identified at baseline, we observed that all of these associations were in the same direction as the ones
116 identified at 24 months and only 2 out of 192 were different from the directions at 12 months. Pearson
117 correlations between the beta coefficients (the associations between lipid species and AD state) among
118 three time points were generally quite large (0.78 – 0.84).

119 Across the three time points, 35 lipid species were consistently associated with AD status. These lipid species
120 originated from the ceramide (Cer(d)), dihexosylceramide (Hex2Cer), trihexosylceramide (Hex3Cer), GM3
121 ganglioside (GM3), sphingomyelin (SM), lysophosphatidylcholine (LPC), lysoalkylphosphatidylcholine
122 (LPC(O)), lysoalkenylphosphatidylcholine (LPC(P)), alkylphosphatidylethanolamine (PE(O)),
123 phosphatidylinositol (PI). There were two strongest associations observed between two novel lipid species
124 from the dehydrodesmosterol ester (deDE) class, and AD across all three time points. In detail, they were
125 deDE(18:2) ($p=7.02 \times 10^{-25}$ at baseline; 1.06×10^{-17} at 12 months; 8.80×10^{-12} at 24 months) and deDE(20:4)
126 ($p=7.48 \times 10^{-14}$ at baseline; 2.32×10^{-14} at 12 months; 2.79×10^{-12} at 24 months). As shown in Supplementary
127 Table 6, t-test results showed anticholinesterase medication usage was significantly different between AD
128 and CN groups (0.0 % vs 87.2 %; CN vs AD; p -value $< 1.0 \times 10^{-04}$). In the later analysis, we identified that both
129 deDE lipid species showed the strongest associations with anticholinesterase medication at baseline
130 ($p=7.13 \times 10^{-18}$ for deDE(18:2), $p= 9.68 \times 10^{-12}$ for deDE(20:4); Supplementary Table 7), indicating that they
131 might be driven by the anticholinesterase medication usage (AD-related medication).

132 ***The AD-CN model classified AD and stratified the non-converting MCI group into AD-like and non AD-like***

133 An AD-CN model was developed using a ridge regression model within a 5-fold cross-validation framework
134 on the AD and CN sub-cohort at baseline using all the lipid species (except the two deDE lipid species which
135 were affected by the anticholinesterase medication usage) to predict AD status. The AD-CN model was
136 applied to baseline profiles of individuals in the non-converting MCI and AD converter groups (the
137 combinations of individuals converted from both MCI and CN groups; treated as true positives). The model
138 could classify non-converting MCI and AD converters with an AUC of 0.69 (Figure 4). Relative to the basic
139 model using four predictors – age, sex, BMI, and APOE $\epsilon 4$ (AUC=0.65; Supplementary Figure 1), our AD-CN
140 model using the linear combination of the whole lipidome, age, sex, BMI, and APOE $\epsilon 4$ as the predictors
141 showed better predictive power.

142 The predicted values from the model indicate the individual's risk of developing AD. Using a cut-off point of
143 0.37 (illustrated in Supplementary Figure 2), the whole group (the combination of non-converting MCI and
144 AD converters) was divided into low and high AD risk groups. The Kaplan-Meier plot (Figure 5a and 5b)
145 showed the proportion of participants converted to AD in the high AD risk group across different time points
146 was higher than in the low AD risk group. The Fisher's exact test ($p=4.85 \times 10^{-15}$, Odd Ratio=3.14, 95% CI=2.30-
147 4.31) also demonstrated that the model could efficiently stratify the group with most AD converters enriched
148 in the high AD risk subgroup (59%).

149 Among 432 non-converting MCI, we further defined 271 individuals in the low AD risk group as non AD-like
150 MCI and 142 individuals in the high AD risk group as AD-like MCI. We investigated the distribution of selected
151 medications between the two non-converting MCI groups (Supplementary Table 8). Usage of Omega-3 was
152 greater in the non-AD like MCI group (20.4 % vs 36.9 %; $p = 9.0 \times 10^{-04}$), while anticholinesterase usage was
153 greater in the AD-like MCI group (35.9 % vs 17.0 %; $p < 1.0 \times 10^{-04}$).

154 ***The performance of AD-CN model validated across the time points was stable***

155 We further applied the AD-CN model built on the baseline AD and CN data to the whole data set (excluding
156 AD converters) across different time points. To evaluate the prediction performance of the model on each
157 single time point, we re-defined the AD diagnosis status of each individual as a binary variable – AD and non-
158 AD status (combination of CN and MCI). The predictive performance at the three main time points was AUC:
159 0.71 (0.68 – 0.74) at baseline, AUC: 0.68 (0.64 – 0.71) at 12 months, and AUC: 0.72 (0.68 – 0.77) at 24 months
160 (Figure 6). The model demonstrated consistent and robust prediction performance at each time point.

161 ***Trajectories of individual lipid species***

162 The trajectories of individual lipid species were assessed using separate linear mixed-effects models.
163 Significant interactions between AD status and time were observed for 1) the AD converter group versus the
164 non-converter group (combination of non-converting CN and MCI); 2) AD-like versus non-AD-like MCI groups
165 after excluding AD converters; and 3) non AD-like MCI vs CN after excluding AD converters.

166 When comparing AD and CN, only two lipid species (from the sphingosine class) were significant following
167 FDR correction (Figure 7a; Supplementary Table 9).

168 Comparing AD-like and non AD-like MCI groups, more than 121 lipid species from 26 lipid classes showed
169 significantly different trajectories after FDR correction (Figure 7a; Supplementary Table 9). A majority of these
170 were from the SM, GM3, acylcarnitine (AC), alkylphosphatidylcholine (PC(O)), alkenylphosphatidylcholine
171 (PC(P)), LPC, LPC(O), LPC(P), and dehydrocholesterol ester (DE) classes. Of these lipid species, we observed
172 13 lipid species from the lipid classes of GM3, LPC(O), AC, SM, CE and DE also appeared in the top 50
173 predictors in the AD-CN model.

174 The trajectories for 33 lipid species (after FDR correction) were found to significantly differ between AD
175 converters vs non-converters (non-converting CN and MCI) (Figure 7b, Supplementary table 10). These lipids
176 were primarily composed of the LPC(O), LPC(P), and PC(O) classes. The trajectory for all these lipid species
177 were diametrically opposed with those between AD converters and non-converters. As a sensitivity analysis,
178 we grouped MCI converters with AD converters and evaluated the associations of the trajectories of lipid
179 species with the combination of AD and MCI converters (Supplementary Figure 3, Supplementary table 10).
180 Compared to the associations with the AD converters group, we observed slightly more lipid species showing
181 significantly altered trajectories in the combined AD- and MCI-converter group.

182 We selected several lipid species that showed differential trajectories as case studies to examine the
183 trajectory over all time points (Figure 7C). AD and CN showed trajectories in opposite direction for several
184 lipid species, including GM3(d18:1/24:1), Cer(d18:1/18:0), PE(P-16:0/22:6), and PI(18:0_22:6). By contrast,
185 PC(O-16:0/16:0), AC(12:0), deDE(18:2), the oxidised lipid, PC(36:4) [+OH], the dimethyl-cholesteryl ester,
186 dimethyl-CE(18:1), the bile acid dxCA, and CE(24:1) showed trajectories in the same direction. Interestingly,
187 the trends of these lipid species in the converter group transitioned from CN concentrations at baseline to
188 AD group concentrations by the end of the study.

189 ***AD-related medications affect the trajectory of lipid species over time***

190 The aforementioned, t-test results showed that anticholinesterase medication usage was significantly
191 different between AD and CN groups. Further, longitudinal random forest backward selection analysis was
192 performed to examine whether medications affect the trajectory of lipid species. The results showed that
193 the trajectory of four lipid species deDE(18:2), deDE(20:4), GM3(d18:1/22:0), and PE(P)(18:1/22:4) were
194 significantly (variable importance > 30) associated with these anti-dementia drugs (Supplementary Figure 4;
195 Supplementary Table 11. Further, we also found that there were 91 and 160 lipid species (Supplementary

196 Figure 4; Supplementary Table 11), whose trajectories were significantly affected by the medications of
197 omega-3 and statins respectively.

198 **Trajectory of AD-CN scores varied among AD diagnosis groups**

199 After evaluating the AD-CN model across all timepoints, we derived the risk scores for each participant. The
200 distributions of AD-CN scores at each time point is shown in Supplementary Figure 5. We examined the
201 association of the trajectories of the risk scores with AD status (non-converting CN, non AD-like MCI, AD-like
202 MCI, AD converters, and AD) (Table 2). The cross-sectional associations of the lipidomic score with AD-like
203 MCI versus non-converting CN ($p=3.23 \times 10^{-26}$) and AD converters versus the combination of non-converting
204 CN and MCI groups ($p=3.20 \times 10^{-16}$) were significant (Table 2). When we examined the changes of the overall
205 AD-CN lipidomic scores among these groups, the AD-like MCI versus non-converting CN behaved significantly
206 different ($p=3.78 \times 10^{-04}$). The changes of AD converters and the non-converting CN and MCI groups were
207 significant different ($p=3.27 \times 10^{-03}$). The AD group behaved similarly as the CN group.

208 The trajectory of the AD-CN score for all groups is shown in Figure 8. The trajectory of CN and AD groups
209 showed minor trends in the opposite direction. Compared to CN and AD groups, AD-like and non AD-like MCI
210 showed much larger changes in opposite directions with the AD-like group showing a decrease and the non-
211 AD like showing an increase in the AD-CN score. The trajectory of AD converters and non converters was also
212 in opposite directions with the AD converters showing a slight decrease in the AD-CN score and the non-
213 converters showing a slight increase.

214 **Discussion**

215 We performed comprehensive cross-sectional and longitudinal analyses of lipidomic data in the ADNI cohort
216 to identify the associations of the lipid species trajectories with AD diagnosis. Of note, there were many MCI
217 participants in the ADNI study presenting with mixed aetiologies. While some progressed to AD over the
218 follow-up period, others appeared to be stable but displayed heterogeneity in their plasma lipidome, and
219 this was used to stratify these individuals into AD like and non-AD like MCI participants.

220 Previous research^{5,6,29-36} has also recognised this heterogeneity, with some studies implementing machine
221 learning methods on neuroimaging data²⁹⁻³⁴ or plasma metabolites^{35,36} to predict MCI participants at high-
222 risk of converting to AD. These studies showed promising results but were limited by small sample size. In
223 this study, we developed an AD-CN lipidomic model, using ridge regression, within a 5-fold cross validation
224 framework, on the large training dataset (AD and healthy groups; $n=651$). The prediction accuracy of the
225 lipidomic model was assessed within the cross-validation framework of the training dataset and gave an AUC
226 of 0.75 for the separation of CN and AD participants. Further, the AD-CN model efficiently classified MCIs
227 (including AD converters) into low AD risk and high AD risk, with the high AD risk group having two times
228 higher risk of conversion to AD than the low AD risk group. Using the model to stratify the non-converting
229 MCI group showed 142 out of 413 MCI participants were defined as AD-like MCI (participants with high
230 "lipidomic similarity" to AD).

231 The later trajectory analysis of lipid species between AD-like and non AD-like MCI groups delineate the
232 heterogeneity of the MCI groups. We also observed that the AD-like MCI group had a significantly larger
233 proportion of individuals taking anticholinesterase inhibitors, indicating that our AD-CN classification model
234 could efficiently capture individuals in the MCI group who had started to develop AD symptoms.

235 The metabotype conservation index showed that overall, the lipidome was stable across time (the first two
236 years). However, the conservation index in clinical AD was significantly lower than the CN group indicating
237 greater variation in the lipidome over the two years of this analysis. While AD converters and the non-
238 converting MCI group were not significantly different to the CN group.

239 In contrast to the conservation index calculated over two years, the lipid trajectories (calculated over 10
240 years) showed multiple lipid species with trajectories that associated with AD converters relative to the non-
241 converting CN and MCI groups. We identified a greater decrease of (LPC(O), LPC(P)), LPC, and PC(O) species
242 in the converter group. In particular, PC(O-38:5) and PC(O-40:5) decreased during the progression to AD.
243 There was consistent evidence of decreasing ether lipids for participants in the transition to AD²⁸, which may

244 reflect changes in the biosynthetic pathway i.e. a gradual deterioration in peroxisome function, leading to
245 decreased ether lipids in circulation³⁷. AD converters also showed a strong decrease in lysophospholipids
246 species (containing 16:0, 18:0, 18:1, 20:0, 22:0, 22:1 and 24:0 fatty acids) during the progression to AD. In
247 support of this observation, several cross-sectional studies have reported that plasma levels of LPC were
248 decreased in the AD patients compared with healthy group³⁸ and the LPC-to-PC ratio were also inversely
249 associated with AD^{39,40}, suggesting decreasing phospholipase activity as the disease progresses. In humans,
250 LPC(O) and LPC(P) are metabolised by both these metabolic pathways – 1) these lipids are synthesised as
251 ether lipids, originating in the peroxisomes; and 2) these lipids are produced by a plasmalogen specific
252 phospholipase A2 (PLA2), which cleaves the fatty acid from the sn-2 position. Therefore, we hypothesise that
253 both pathways – peroxisome dysfunction³⁷ and decreased PLA2^{41,42} – result in the scarcity of
254 lysophospholipids in participants transitioning to AD. A better understanding of the altered activity of PLA2
255 and ether lipid metabolism could help to identify novel therapeutic targets in AD⁴³.

256 The lipidomic score provides a global view of the changes in the lipidome over time among AD diagnosis
257 groups. The lipidomic score for non-converting CN and AD groups were stable across time, with AD showing
258 a higher level than the CN group. Consistent with the differences in the trajectories of the lyso and ether
259 lipid species observed between AD converters and non converters (Figure 7b), the overall lipidomic score
260 between the two groups showed significant changes in opposite directions (Figure 8b). Further, we observed
261 differences in the lipidomic score between the AD-like and non AD-like MCI groups and this agreed with the
262 greater number of individual lipid species trajectories showing a significant difference between these groups
263 (Figure 7a). Perhaps surprisingly the AD-like MCI group showed a downward trend in the lipidomic score
264 towards the CN value, suggesting the lipidome may be normalising in this group, while the non AD-like MCI
265 group showed an increase in the lipidomic score crossing the CN value. Clearly these two groups represent
266 different metabolic phenotypes with the AD-like MCI being stable and resilient to progression to AD over the
267 10 year follow up period. This may be associated with a normalisation of their metabolic phenotype as
268 measured by the AD-CN score. In contrast, the non-AD like group which starts with a low AD-CN score
269 appears to progress as evidenced by an increasing AD-CN score but not to the point of conversion (Figure
270 8a). These global and individual lipid species trajectories may provide useful biomarkers to monitor disease
271 progression and better target MCI participants at greatest risk of progression to AD for clinical trials or
272 treatment. However, further studies are required to define the outcomes for the AD-like and non AD-like
273 metabolic phenotypes identified in this study.

274 There were several limitations in this study. Although the follow-up time of the ADNI study extends to 10
275 years, the majority of records are from baseline to 24 months. The metabotype conservation index under
276 two years framework has limited ability to capture the variations of each individual among different AD
277 diagnosis groups. And, the power of the longitudinal analysis on the cohorts within a 2-year period is limited.
278 In addition, a second cohort is needed to externally validate the performance of our AD-CN classification
279 model.

280 In conclusion, we have performed comprehensive lipidomic analyses using the longitudinal ADNI -1, -2 and
281 -GO cohorts. At baseline, we have developed a novel AD-CN classification model to characterise the
282 heterogeneity of MCI participants, providing the potential of using lipidomics to efficiently distinguish high-
283 risk subjects within the MCI group. The subsequent longitudinal analysis using the data set across all the time
284 points highlighted significant changes in the lipidome over time in AD converters relative to CN and AD-like
285 MCI relative to non AD-like MCI groups. These highlight the potential of lipidomic studies to improved our
286 understanding of the relationships between lipid metabolism and progression to AD. Lipidomic biomarkers
287 also show promise to improve clinical risk assessment and management of older individuals at risk of AD.

288 **Methods**

289 ***Participants***

290 Alzheimer's Disease Neuroimaging Initiative (ADNI)-1, -2 and -GO (<http://adni.loni.usc.edu/>) is a longitudinal
291 study, recruiting 1,517 individuals over 55 years old at baseline. At intervals of 6-12 months, blood and
292 clinical data were collected from each individual, up to a maximum of 10 years. Lipidomic profiling was
293 performed on all blood samples, with 4,873 plasma samples examined from baseline up to the 13th time

294 point (at 10 years followed up). After filtering 143 missingness, we ended up with 4,730 samples in total. At
295 baseline, there were originally 1,418 individuals, of which we excluded 25 individuals with missing values (10
296 missing cognitive scores, 3 missing fasting information, and 12 missing BMI). Thereby, there remained 1,393
297 participants at baseline in the study.

298 The definition of probable AD in ADNI followed the NINDS-ADRDA criteria⁴⁴. In brief, individuals with Mini-
299 Mental State Exam (MMSE) scores between 20 and 26 (inclusive) and a Clinical Dementia Rating Scale (CDR)
300 of 0.5 or 1.0 were classified as AD patients⁴⁵. Participants were defined as MCI if they had MMSE scores
301 between 24 and 30, a memory complaint, objective memory loss measured by education-adjusted scores on
302 Wechsler Memory Scale Logical Memory II, a CDR of 0.5, absence of significant levels of impairment in other
303 cognitive domains, and essentially preserved activities of daily living⁴⁶.

304 We further defined the longitudinal status of AD diagnosis group. At baseline, there were 437 CN, 713 MCI,
305 and 243 AD cases. Among all these 1,393 participants, we defined two categories according to their changes
306 of diagnosis status over time: 1) AD converters – 329 participants whose status were CN/MCI at baseline but
307 converted to AD later; 2) MCI converters – 71 CN progressed to MCI and stayed as MCI at the following time
308 points. Additionally, we also observed there were 53 MCI reverted back to CN. And, there were 39 individuals
309 (10 CN, 5 AD, and 24 MCI at baseline) whose status varied across time points. We treated these unstable
310 individuals as undefined (might be affected by the medication usages) and removed them from the
311 longitudinal analysis (though in our cross-section analysis such as AD-CN modelling, we still keep them in the
312 analysis).

313 **Lipidomic profiling**

314 Lipidomic profiling was performed on all plasma samples (n=4,730) using our recently expanded, targeted
315 lipidomic profiling strategy using reverse phase liquid chromatography coupled to an Agilent 6495C QqQ
316 mass spectrometer. The lipid extraction and LC-MS/MS methodology, using scheduled multiple reaction
317 monitoring (MRM), was as previously described¹⁵ with the addition of approximately 200 novel lipid species
318 from 17 lipid classes²⁴. Further details about our latest methodology to measure these lipids are described
319 on our laboratory website (<https://metabolomics.baker.edu.au/method/>).

320 Overall, there were 781 lipid species from 49 lipid classes quantified. Single ion monitoring (SIM) and neutral
321 loss (NL) are two types of measurements for the same TG lipid species, with NL measurements more specific
322 and sensitive. To avoid the redundancy and improve the accuracy in our modelling analysis, we excluded 32
323 TG[SIM] lipids, retaining 749 lipid species from 48 lipid classes in this study. The details for all the lipid species
324 and classes were listed in Supplementary Table 1.

325 **Statistical analysis**

326 In the following analysis, log₁₀ transformation followed by a standard normalisation (zero mean and one-
327 unit standard deviation) was performed on individual lipid species across repeated measurement of all
328 participants. We introduced the covariate set including age, sex, BMI, HDL-C, total cholesterol, triglycerides,
329 fasting status, cohort (a categorical variable indicating ADNI 1, GO, and 2 phases), omega-3, and statin status
330 for the following models.

331 **Development of an AD-CN model.** We sought to use the normalised lipidomic data on AD subjects (n=243)
332 and cognitive normal individuals (CN; n=408) at baseline to build the classification model. Further, we applied
333 the model to stratify non-converting MCI (n=413) from AD converters (n=329). Ridge regression models were
334 created to stratify AD from CN, optimizing C-statistic using the R package ‘glmnet v4.1-4’. Five models were
335 created from an external 5-fold cross-validation framework (Figure 1). All these models were adjusted for
336 age, sex, BMI, *APOE* ϵ 4, HDL-C, total cholesterol, triglycerides, fasting status, cohort (a categorical variable
337 indicating ADNI 1, GO, and 2 phases), omega-3, and statin status. Since we identified that two deDE lipid
338 species (deDE(18:2) and deDE(20:4)) were strongly associated with dementia-related medication, in the
339 models, we used the whole lipidomes except the two deDE lipid species as the main predictors.

340 Beta coefficients from each cross-validation fold were averaged. This model was then applied to the non-
341 converting MCI group to generate the probabilities of the MCI individuals being “AD-like” or “non AD-like”.

342 In addition, the weights were separately applied to the whole dataset across different time points to
343 generate overall AD risk scores for each individual at each time.

344 **The metabotype conservation index.** As changes in metabolic phenotypes (metabotype) over time can be
345 indications for disease onset or progression⁴⁷, we used the metabotype conservation index described by
346 Yousri et al.⁴⁸ to quantify the stability of the metabotype of ADNI participants over time. This analysis was
347 restricted to comparisons between baseline and follow up visits after 12 and 24 months, which had 755
348 common participants with samples across baseline, 12 month and 24 months.

349 As pairwise correlation-based metrics can be distorted by the strong correlation structure observed for lipid
350 measurements, we first aggregated strongly correlated lipids into representative variables (“eigenlipids”).
351 To this end, we first clustered lipid concentrations at baseline using weighted correlation network analysis
352 using the R packages WGCNA (version=1.71)⁴⁹ and dynamicTreeCut (version=1.63.1)⁵⁰. We used a soft
353 threshold of 0.86, which was closest to the recommended threshold of 0.9⁵¹ and corresponded to a power
354 of 10 used for the calculation of the adjacency matrix. We then calculated the topological overlap matrix
355 (TOM) using the adjacency matrix. The resulting distance matrix was used for hierarchical clustering and the
356 dendrogram was cut using the hybrid method of the cutTreeDynamic-function using the following
357 parameters: deepSplit=3, pamDendroRespect=TRUE, and minClusterSize = 2. The latter was chosen to
358 sensitively extract clusters that capture all strong correlations down to at least two lipids. We then calculated
359 one eigenlipid for each individual cluster by extracting its first principal component. Metabolites in the
360 outgroup were retained as separate features. We additionally provide results for several other values of the
361 parameter minClusterSize (5, 10, and the default value of 20). Cluster assignments for each lipid across
362 settings is provided in Supplementary Table 12, with results for higher values of the minimum cluster size (5,
363 10, and 20) in Supplementary Figure 6 and Supplementary Table 3.

364 Projecting the cluster and outgroup assignment from baseline to 12 months and 24 months, we obtained
365 longitudinal measures to calculate the metabotype conservation index $L_c(x)$ for subject x as

366
$$L_c(x) = 1 - \frac{\text{rank}(x) - 1}{N - 1},$$

367 where $\text{rank}(x) = N - z$ and $z = |\{y \in \llbracket 1, N \rrbracket \neq x \mid \rho_p(L_x^{bl}, L_x^{fu}) \geq \rho_p(L_x^{bl}, L_y^{fu})\}|$, with N being the
368 number of subjects, L_i^{bl} being the aggregated lipidomics profile of subject i at baseline, L_i^{fu} being the
369 aggregated lipidomics profile of subject i at follow up, and $\rho_p(L_i^{bl}, L_j^{fu})$ being the Pearson correlation
370 coefficient of the lipidomics profile of subject i at baseline and the lipidomics profile of subject j at follow
371 up for $x, i, j \in \llbracket 1, N \rrbracket$. The resulting $L_c(x)$ ranges between 0 and 1, with 0 meaning no conservation and 1
372 meaning maximal conservation. We compared proportions of individuals with $l_c = 1$ and $l_c < 1$ per diagnostic
373 group using Fisher’s exact test, considering the stable CN group as reference. The resulting distribution of l_c
374 values smaller than 1 (i.e., individuals whose lipid profiles change over the two-year period) were compared
375 across diagnostic groups using the Wilcoxon rank sum test. For assessing changes in the l_c distribution
376 globally between baseline to 12 months and baseline to 24 months, we used both paired t-tests and the
377 Wilcoxon test.

378 **Longitudinal analysis using linear mixed models.** We calculated linear mixed models using repeated
379 measurements across 13 time points to examine the associations between AD diagnosis state and trajectory
380 of lipid species (either species level or overall lipidomic scores) over time using the interaction between time
381 points and AD state. In the model, we treated normalised individual lipid species or overall lipidomic scores
382 as the independent variables, and AD diagnosis state (the category variable) as the main predictor and
383 adjusted the models with a list of covariates of age (at baseline), sex, BMI, HDL-C, total cholesterol,
384 triglycerides, fasting status, cohort, time point (treated as continuous variable), omega-3, and statin status.
385 The interaction between time point and AD diagnosis was introduced into the model, which is the key term
386 for examining the trajectory of lipid species over time among AD states. In the model, we perform two sets
387 of longitudinal analyses on different subsets of the population to: 1) examine the difference in the trajectory
388 of lipid species among AD patients, CN, and non-converting MCI group (AD-like or non AD-like) using the
389 whole population (excluding the converters); 2) use the changes in lipid species to predict the AD converters

390 on the population excluding all prevalent AD cases. The lme4 package in software R 3.6.2 was used to
391 perform the linear mixed models.

392 **Sensitivity analysis.** We performed sensitivity analysis to examine whether the cross-sectional associations
393 between AD state (AD cases vs CN) and lipid species are consistent across the major time points (baseline,
394 12 months, and 24 months). To perform this, we used linear regression with lipid species as the independent
395 variable, and AD state as the main predictor.

396 To assess whether anticholinesterases were associated with specific lipid trajectories, random forest
397 backward selection⁵² was performed using the Boruta function implemented within the Boruta R package
398 (version 7.0.0). The calculated predictive value (termed as variable importance, calculated across 100
399 permutations of the original dataset) reflects the strength of the association between medications and the
400 changes in lipid species. Predictive values of medication (grouped by ATC codes) on lipids with importance
401 larger than 30 (mean + 4 standard deviations) were chosen as highly relevant⁵².

402 We performed another sensitivity analysis to evaluate whether the trajectories of lipid species within MCI
403 converters behaved similarly as those within AD converters. To do this, linear mixed models was carried out
404 to examine the associations of trajectory of lipid species with the combination of MCI converters and AD
405 converters relative to non-converting CN and MCI groups.

406 **Data Availability**

407 The results published here are in whole or in part based on data obtained from the AD Knowledge Portal.
408 ADNI associated data is also available from the Laboratory of Neuro Imaging Image and Data Archive at
409 <https://ida.loni.usc.edu/login.jsp>.

410

411 **Acknowledgement**

412 Support was provided by the National Health and Medical Research Council of Australia (# 1101320 and
413 1157607). KH is supported by an NHMRC Investigator grant (GNT1197190). This work was also supported in
414 part by the Victorian Government's Operational Infrastructure Support Program.

415 The data available in the AD Knowledge Portal would not be possible without the participation of research
416 volunteers and the contribution of data by collaborating researchers.

417 Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative
418 (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award
419 number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of
420 Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie,
421 Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen;
422 Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and
423 Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE
424 Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson &
425 Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale
426 Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer
427 Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian
428 Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector
429 contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The
430 grantee organization is the Northern California Institute for Research and Education, and the study is
431 coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI
432 data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

433 Metabolomics data is provided by the Alzheimer's Disease Metabolomics Consortium (ADMC) and funded
434 wholly or in part by the following grants and supplements thereto: NIA R01AG046171, RF1AG051550,
435 RF1AG057452, R01AG059093, RF1AG058942, U01AG061359, U19AG063744 and FNIIH: #DAOU16AMPA
436 awarded to Dr. Kaddurah-Daouk at Duke University in partnership with a large number of academic
437 institutions. As such, the investigators within the ADMC, not listed specifically in this publication's author's

438 list, provided data along with its pre-processing and prepared it for analysis, but did not participate in analysis
439 or writing of this manuscript. A complete listing of ADMC investigators can be found at:
440 <https://sites.duke.edu/adnimetab/team/>.

441 Data on lipidome was generated at Baker Heart and Diabetes Institute, a member of ADMC. Details on the
442 lipid profiling technologies are described at: (<https://metabolomics.baker.edu.au/method>).

443 **Author Contributions**

444 Meikle, Kaddurah-Daouk and Kastenmüller led the study design team. Wang, Arnold, and Huynh led the
445 statistical analyses presented in this study. Giles and Livera provided statistical suggestions on linear mixed
446 models. Weinisch and Marella performed the analysis on metabolite conservation index and random forest
447 backward selection. Mellett, Duong, Huynh and Giles supported the acquisition and processing of the
448 lipidomic data for the cohort. Arnold, Kastenmüller, Nho, Saykin, Han and Kaddurah-Daouk were key
449 members of the ADNI team and represent the Alzheimer's Disease Metabolomics Consortium (ADMC): A
450 complete listing of ADMC investigators can be found at <https://sites.duke.edu/adnimetab/who-we-are/>.

451 **Competing interest**

452 Dr. Kaddurah-Daouk is an inventor on a series of patents on use of metabolomics for the diagnosis and
453 treatment of CNS diseases and holds equity in Metabolon Inc., Chymia LLC and PsyProtix. Prof. Meikle leads
454 the provisional patent "METHODS OF ASSESSING ALZHEIMER'S DISEASE" on the development of AD-CN risk
455 scores that has been filed with the Serial No. 63/463,808.

456 Other authors have declared that no conflict of interest exists.

457 **Reference**

- 458
- 459 1. Niccoli, T. & Partridge, L. Ageing as a Risk Factor for Disease. *Current Biology* **22**, R741-R752 (2012).
 - 460 2. Scheltens, P. *et al.* Alzheimer's disease. *The Lancet* **397**, 1577-1590 (2021).
 - 461 3. van der Lee, S.J. *et al.* The effect of APOE and other common genetic variants on the onset of Alzheimer's
462 disease and dementia: a community-based cohort study. *The Lancet Neurology* **17**, 434-444 (2018).
 - 463 4. International., A.s.D. World Alzheimer Report 2018. The state of the art of dementia research: new frontiers.
464 (2018).
 - 465 5. Bowler, J.V. The concept of vascular cognitive impairment. *Journal of the Neurological Sciences* **203**, 11-15
466 (2002).
 - 467 6. Delano-Wood, L. *et al.* Heterogeneity in mild cognitive impairment: differences in neuropsychological profile
468 and associated white matter lesion pathology. *Journal of the International Neuropsychological Society : JINS*
469 **15**, 906-914 (2009).
 - 470 7. Csukly, G. *et al.* The Differentiation of Amnesic Type MCI from the Non-Amnesic Types by Structural MRI.
471 *Frontiers in Aging Neuroscience* **8**(2016).
 - 472 8. Chatterjee, P. *et al.* Plasma Phospholipid and Sphingolipid Alterations in Presenilin1 Mutation Carriers: A
473 Pilot Study. *J Alzheimers Dis* (2016).
 - 474 9. Whiley, L. *et al.* Evidence of altered phosphatidylcholine metabolism in Alzheimer's disease. *Neurobiol Aging*
475 **35**, 271-8 (2014).
 - 476 10. Kling, M.A. *et al.* Circulating ethanolamine plasmalogen indices in Alzheimer's disease: Relation to diagnosis,
477 cognition, and CSF tau. *Alzheimers Dement* **16**, 1234-1247 (2020).
 - 478 11. Barupal, D.K. *et al.* Sets of coregulated serum lipids are associated with Alzheimer's disease pathophysiology.
479 *Alzheimers Dement (Amst)* **11**, 619-627 (2019).
 - 480 12. Smidak, R., Kofeler, H.C., Hoeger, H. & Lubec, G. Comprehensive identification of age-related lipidome
481 changes in rat amygdala during normal aging. *PLoS One* **12**, e0180675 (2017).
 - 482 13. Hancock, S.E., Friedrich, M.G., Mitchell, T.W., Truscott, R.J. & Else, P.L. Decreases in Phospholipids Containing
483 Adrenic and Arachidonic Acids Occur in the Human Hippocampus over the Adult Lifespan. *Lipids* **50**, 861-72
484 (2015).
 - 485 14. Weir, J.M. *et al.* Plasma lipid profiling in a large population-based cohort. *J Lipid Res* **54**, 2898-908 (2013).
 - 486 15. Huynh, K. *et al.* High-Throughput Plasma Lipidomics: Detailed Mapping of the Associations with
487 Cardiometabolic Risk Factors. *Cell chemical biology* (2018).

- 488 16. Lim, W.L.F. *et al.* Relationships Between Plasma Lipids Species, Gender, Risk Factors, and Alzheimer's
489 Disease. *Journal of Alzheimer's Disease* **76**, 303-315 (2020).
- 490 17. Wang, T. *et al.* APOE ϵ 2 resilience for Alzheimer's disease is mediated by plasma lipid species: Analysis of
491 three independent cohort studies. *Alzheimers Dement* (2022).
- 492 18. Toledo, J.B. *et al.* Metabolic network failures in Alzheimer's disease: A biochemical road map. *Alzheimers*
493 *Dement* **13**, 965-984 (2017).
- 494 19. Hyötyläinen, T., Bondia-Pons, I. & Orešič, M. Lipidomics in nutrition and food research. *Mol Nutr Food Res*
495 **57**, 1306-18 (2013).
- 496 20. Morville, T., Sahl, R.E., Moritz, T., Helge, J.W. & Clemmensen, C. Plasma Metabolome Profiling of Resistance
497 Exercise and Endurance Exercise in Humans. *Cell Rep* **33**, 108554 (2020).
- 498 21. Iso-Markku, P. *et al.* Physical activity as a protective factor for dementia and Alzheimer's disease: systematic
499 review, meta-analysis and quality assessment of cohort and case-control studies. *British Journal of Sports*
500 *Medicine* **56**, 701-709 (2022).
- 501 22. Johnson, S.C. *et al.* The Wisconsin Registry for Alzheimer's Prevention: A review of findings and current
502 directions. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **10**, 130-142 (2018).
- 503 23. Darst, B.F., Kosciak, R.L., Hogan, K.J., Johnson, S.C. & Engelman, C.D. Longitudinal plasma metabolomics of
504 aging and sex. *Aging* **11**, 1262-1282 (2019).
- 505 24. Beyene, H.B. *et al.* High-coverage plasma lipidomics reveals novel sex-specific lipidomic fingerprints of age
506 and BMI: Evidence from two large population cohort studies. *PLOS Biology* **18**, e3000870 (2020).
- 507 25. Slade, E. *et al.* Age and sex are associated with the plasma lipidome: findings from the GOLDN study. *Lipids in*
508 *Health and Disease* **20**, 30 (2021).
- 509 26. Ihira, H. *et al.* Association Between Physical Activity and Risk of Disabling Dementia in Japan. *JAMA Network*
510 *Open* **5**, e224590-e224590 (2022).
- 511 27. Floud, S., Balkwill, A., Reeves, G.K., Peto, R. & Beral, V. Cognitive and social activities and long-term dementia
512 risk – Authors' reply. *The Lancet Public Health* **6**, e270 (2021).
- 513 28. Huynh, K. *et al.* Concordant peripheral lipidome signatures in two large clinical studies of Alzheimer's
514 disease. *Nature Communications* **11**, 5698 (2020).
- 515 29. Gray, K.R. *et al.* Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease.
516 *NeuroImage* **60**, 221-229 (2012).
- 517 30. Nozadi, S.H., Kadoury, S. & Initiative, A.s.D.N. Classification of Alzheimer's and MCI patients from
518 semantically parcelled PET images: a comparison between AV45 and FDG-PET. *International journal of*
519 *biomedical imaging* **2018**(2018).
- 520 31. Shi, J. & Liu, B. Stage detection of mild cognitive impairment via fMRI using Hilbert Huang transform based
521 classification framework. *Medical Physics* **47**, 2902-2915 (2020).
- 522 32. Sheng, J. *et al.* A novel joint HCPMMP method for automatically classifying Alzheimer's and different stage
523 MCI patients. *Behavioural brain research* **365**, 210-221 (2019).
- 524 33. Goryawala, M. *et al.* Inclusion of neuropsychological scores in atrophy models improves diagnostic
525 classification of Alzheimer's disease and mild cognitive impairment. *Computational intelligence and*
526 *neuroscience* **2015**(2015).
- 527 34. Jitsuishi, T. & Yamaguchi, A. Searching for optimal machine learning model to classify mild cognitive
528 impairment (MCI) subtypes using multimodal MRI data. *Scientific Reports* **12**, 4284 (2022).
- 529 35. Huang, Y.L. *et al.* Discovery of a Metabolic Signature Predisposing High Risk Patients with Mild Cognitive
530 Impairment to Converting to Alzheimer's Disease. *Int J Mol Sci* **22**(2021).
- 531 36. He, S. *et al.* Blood metabolites predicting mild cognitive impairment in the study of Latinos-investigation of
532 neurocognitive aging (HCHS/SOL). *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **14**,
533 e12259 (2022).
- 534 37. Grimm, M.O.W. *et al.* Plasmalogen synthesis is regulated via alkyl-dihydroxyacetonephosphate-synthase by
535 amyloid precursor protein processing and is affected in Alzheimer's disease. *Journal of Neurochemistry* **116**,
536 916-925 (2011).
- 537 38. Lin, W. *et al.* Studies on diagnostic biomarkers and therapeutic mechanism of Alzheimer's disease through
538 metabolomics and hippocampal proteomics. *European Journal of Pharmaceutical Sciences* **105**, 119-126
539 (2017).
- 540 39. Klavins, K. *et al.* The ratio of phosphatidylcholines to lysophosphatidylcholines in plasma differentiates
541 healthy controls from patients with Alzheimer's disease and mild cognitive impairment. *Alzheimer's &*
542 *Dementia: Diagnosis, Assessment & Disease Monitoring* **1**, 295-302 (2015).
- 543 40. Mulder, C. *et al.* Decreased lysophosphatidylcholine/phosphatidylcholine ratio in cerebrospinal fluid in
544 Alzheimer's disease. *J Neural Transm (Vienna)* **110**, 949-55 (2003).

- 545 41. Gattaz, W.F., Maras, A., Cairns, N.J., Levy, R. & Förstl, H. Decreased phospholipase A2 activity in Alzheimer
546 brains. *Biological Psychiatry* **37**, 13-17 (1995).
- 547 42. Ross, B.M., Moszczynska, A., Erlich, J. & Kish, S.J. Phospholipid-metabolizing enzymes in Alzheimer's disease:
548 increased lysophospholipid acyltransferase activity and decreased phospholipase A2 activity. *Journal of*
549 *neurochemistry* **70**, 786-793 (1998).
- 550 43. Llano, D.A. & Devanarayan, V. Serum Phosphatidylethanolamine and Lysophosphatidylethanolamine Levels
551 Differentiate Alzheimer's Disease from Controls and Predict Progression from Mild Cognitive Impairment. *J*
552 *Alzheimers Dis* **80**, 311-319 (2021).
- 553 44. Weiner, M.W. *et al.* Impact of the Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014. *Alzheimer's &*
554 *dementia : the journal of the Alzheimer's Association* **11**, 865-884 (2015).
- 555 45. Arnold, M. *et al.* Sex and APOE ϵ 4 genotype modify the Alzheimer's disease serum metabolome. *Nature*
556 *Communications* **11**, 1148 (2020).
- 557 46. Petersen, R.C. *et al.* Alzheimer's Disease Neuroimaging Initiative (ADNI). *Clinical characterization* **74**, 201-209
558 (2010).
- 559 47. Lacruz, M.E. *et al.* Instability of personal human metabolome is linked to all-cause mortality. *Sci Rep* **8**, 9810
560 (2018).
- 561 48. Yousri, N.A. *et al.* Long term conservation of human metabolic phenotypes and link to heritability.
562 *Metabolomics* **10**, 1005-1017 (2014).
- 563 49. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC*
564 *Bioinformatics* **9**, 559 (2008).
- 565 50. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree
566 Cut package for R. *Bioinformatics* **24**, 719-720 (2007).
- 567 51. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl*
568 *Genet Mol Biol* **4**, Article17 (2005).
- 569 52. Kursa, M.B. & Rudnicki, W.R. Feature Selection with the Boruta Package. *Journal of Statistical Software* **36**, 1
570 - 13 (2010).
- 571
- 572

573
574

Tables

Table 1. The basic characteristic of participants at Baseline, 12m and 24m.

Baseline					
	Stratified by AD disease status				
<i>n</i>	CN	MCI	AD Converters	AD	P values*
	408	413	329	243	
<i>age (mean (SD))</i>	74.14 (5.99)	71.85 (7.67)	74.22 (6.85)	74.93 (7.64)	1.32x10 ⁻⁰⁸
<i>Sex= Male (%)</i>	202 (49.3)	234 (56.7)	195 (59.6)	137 (56.4)	3.04x10 ⁻⁰²
<i>HDL-C (mean (SD))</i>	1.54 (0.38)	1.51 (0.37)	1.55 (0.37)	1.54 (0.36)	5.01x10 ⁻⁰¹
<i>chol (mean (SD))</i>	4.91 (0.94)	4.98 (0.95)	5.02 (0.96)	5.03 (0.99)	3.70x10 ⁻⁰¹
<i>trig (mean (SD))</i>	1.18 (0.56)	1.17 (0.53)	1.17 (0.52)	1.17 (0.48)	9.84x10 ⁻⁰¹
<i>fasting = Yes (%)</i>	393 (95.4)	398 (96.4)	311 (94.8)	230 (95.0)	7.50x10 ⁻⁰¹
<i>BMI (mean (SD))</i>	27.25 (4.85)	27.37 (4.77)	26.59 (4.73)	25.78 (4.12)	9.15x10 ⁻⁰⁵
<i>APOE4 (%)</i>					2.81x10 ⁻³⁶
0	302 (73.3)	240 (58.1)	120 (36.6)	77 (31.8)	
1	99 (24.0)	140 (33.9)	161 (49.1)	115 (47.5)	
2	11 (2.7)	33 (8.0)	47 (14.3)	50 (20.7)	
12 months					
	Stratified by AD disease status				
<i>n</i>	CN	MCI	AD Converters	AD	P values*
	334	397	208	249	
<i>age (mean (SD))</i>	74.86 (6.24)	73.39 (7.54)	74.76 (7.09)	76.12 (7.56)	2.22x10 ⁻⁰⁵
<i>Sex= Male (%)</i>	166 (49.7)	237 (59.7)	112 (53.8)	141 (56.6)	5.28x10 ⁻⁰²
<i>HDL-C (mean (SD))</i>	1.55 (0.37)	1.51 (0.38)	1.61 (0.37)	1.56 (0.39)	7.60x10 ⁻⁰³
<i>chol (mean (SD))</i>	4.90 (0.88)	4.97 (0.98)	5.19 (1.00)	5.06 (1.09)	1.53x10 ⁻⁰²
<i>trig (mean (SD))</i>	1.15 (0.54)	1.23 (0.53)	1.16 (0.54)	1.19 (0.49)	3.90x10 ⁻⁰²
<i>fasting = Yes (%)</i>	312 (93.4)	377 (94.7)	196 (94.7)	235 (93.6)	7.83x10 ⁻⁰¹
<i>BMI (mean (SD))</i>	26.99 (4.57)	27.35 (4.74)	26.56 (5.07)	25.68 (4.18)	1.71x10 ⁻⁰⁴
<i>APOE4 (%)</i>					3.98x10 ⁻³⁰
0	247 (74.0)	234 (58.8)	77 (37.2)	83 (33.1)	
1	79 (23.7)	131 (32.9)	100 (48.3)	117 (46.6)	
2	8 (2.4)	33 (8.3)	30 (14.5)	51 (20.3)	
24 months					
	Stratified by AD disease status				
<i>n</i>	CN	MCI	AD Converters	AD	P values*
	380	341	220	148	
<i>age (mean (SD))</i>	75.58 (6.13)	74.12 (7.63)	75.78 (6.95)	77.13 (7.48)	1.00x10 ⁻⁰⁴
<i>Sex= Male (%)</i>	184 (48.4)	199 (58.4)	131 (59.5)	82 (55.4)	1.85x10 ⁻⁰²
<i>HDL-C (mean (SD))</i>	1.56 (0.37)	1.54 (0.40)	1.55 (0.37)	1.48 (0.36)	1.30x10 ⁻⁰¹
<i>chol (mean (SD))</i>	4.94 (0.92)	4.95 (0.96)	4.98 (0.95)	4.96 (0.98)	9.71x10 ⁻⁰¹
<i>trig (mean (SD))</i>	1.16 (0.57)	1.16 (0.52)	1.24 (0.61)	1.24 (0.51)	6.17x10 ⁻⁰²
<i>fasting = Yes (%)</i>	349 (92.3)	322 (94.4)	206 (94.1)	136 (91.9)	5.91x10 ⁻⁰¹
<i>BMI (mean (SD))</i>	27.16 (4.88)	26.93 (4.68)	26.46 (5.18)	25.68 (3.53)	6.70x10 ⁻⁰³
<i>APOE4 (%)</i>					6.06x10 ⁻³⁴
0	279 (73.4)	204 (59.8)	77 (35.0)	44 (29.7)	
1	92 (24.2)	113 (33.1)	103 (46.8)	73 (49.3)	

2 | 9 (2.4) | 24 (7.0) | 40 (18.2) | 31 (20.9)

575 *P values were obtained using either Fisher’s exact test for categorical variables or ANOVA for continuous
576 variable.

577

578 **Table 2. The association of overall lipidomic scores with AD diagnosis (Intercept and trajectory).**

AD diagnosis groups	Associations in intercept				Associations in trajectory			
	Beta *	95% CI (Lower)	95% CI (Upper)	P value	Beta#	95% CI (Lower)	95% CI (Upper)	P value
non AD-like MCI vs. non-converting CN	-0.192	-0.306	-0.077	1.07x10 ⁻⁰³	0.049	0.026	0.072	2.63x10 ⁻⁰⁵
AD-like MCI vs. non-converting CN	0.759	0.621	0.897	3.23x10 ⁻²⁶	-0.046	-0.071	-0.021	3.78x10 ⁻⁰⁴
AD vs. non-converting CN	0.699	0.575	0.823	2.73x10 ⁻²⁷	-0.003	-0.033	0.026	8.39x10 ⁻⁰¹
AD converters vs. non-converters (CN+MCI)	0.531	0.405	0.657	3.20x10 ⁻¹⁶	-0.026	-0.044	-0.009	3.27x10 ⁻⁰³

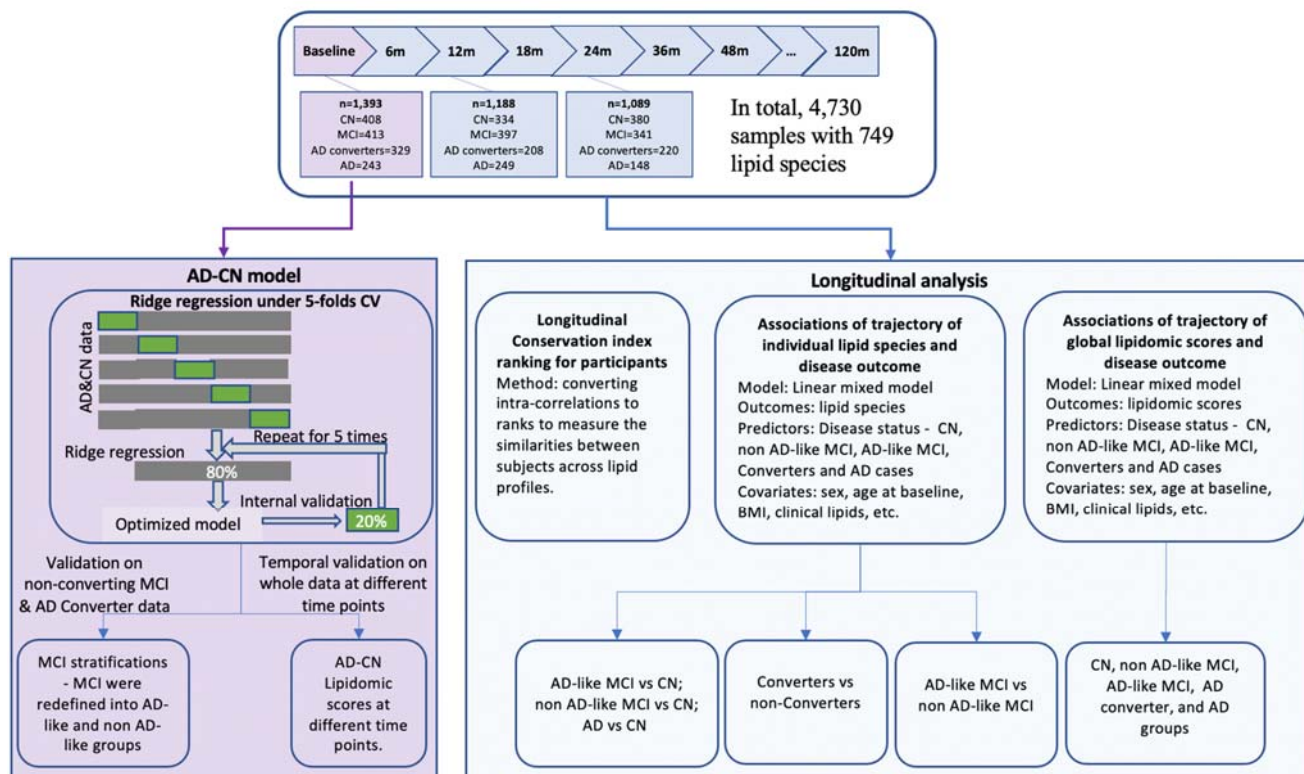
579 *Mean difference in intercept.

580 #Mean change in slope over time.

581

582

583 **Figures.**
584

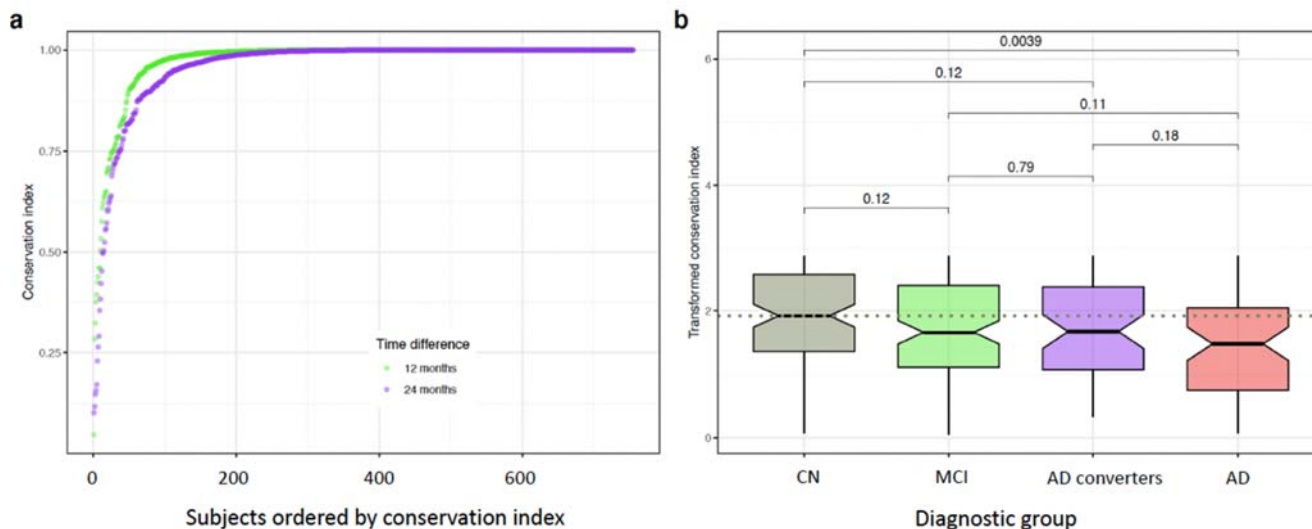


585
586 **Figure 1. Study design.**

587 This study had three parts. Part 1 involved the development of an AD-CN risk model, using baseline data, to
588 characterise the heterogeneity of the non-converting MCI group and to calculate lipidomic risk scores for
589 individuals across different time points. A ridge regression model, built within a five-fold cross-validation
590 framework was used to stratify the non-converting MCI group into AD-like and non AD-like sub-groups. In
591 the development of the model, we treated AD-CN status as outcome with the predictors including all the
592 lipid species, age, sex, BMI, and *APOE ε4*. Part 2 involved the calculation of a metabotype conservation index
593 to quantify the stability of the lipidome over time. Part 3 was the longitudinal analysis on the repeated
594 measurements across 13 time points to examine the associations of changes in lipid species and lipidomic
595 risk scores with AD status. Associations of the trajectories of individual lipid species and disease outcomes
596 were examined using linear mixed models to uncover the difference of trajectories of lipid species
597 between different groups. The covariates included age, sex, BMI, HDL-C, triglycerides, cholesterol, fasting
598 status, omega-3, and statin status. Thereafter, global lipidomic scores combining all lipid species derived from
599 the AD-CN model were fitted into a linear mixed model to define the associations with disease outcomes.
600 Similarly, the covariate set included age, sex, BMI, clinical lipids, fasting status, omega-3, and statin status.

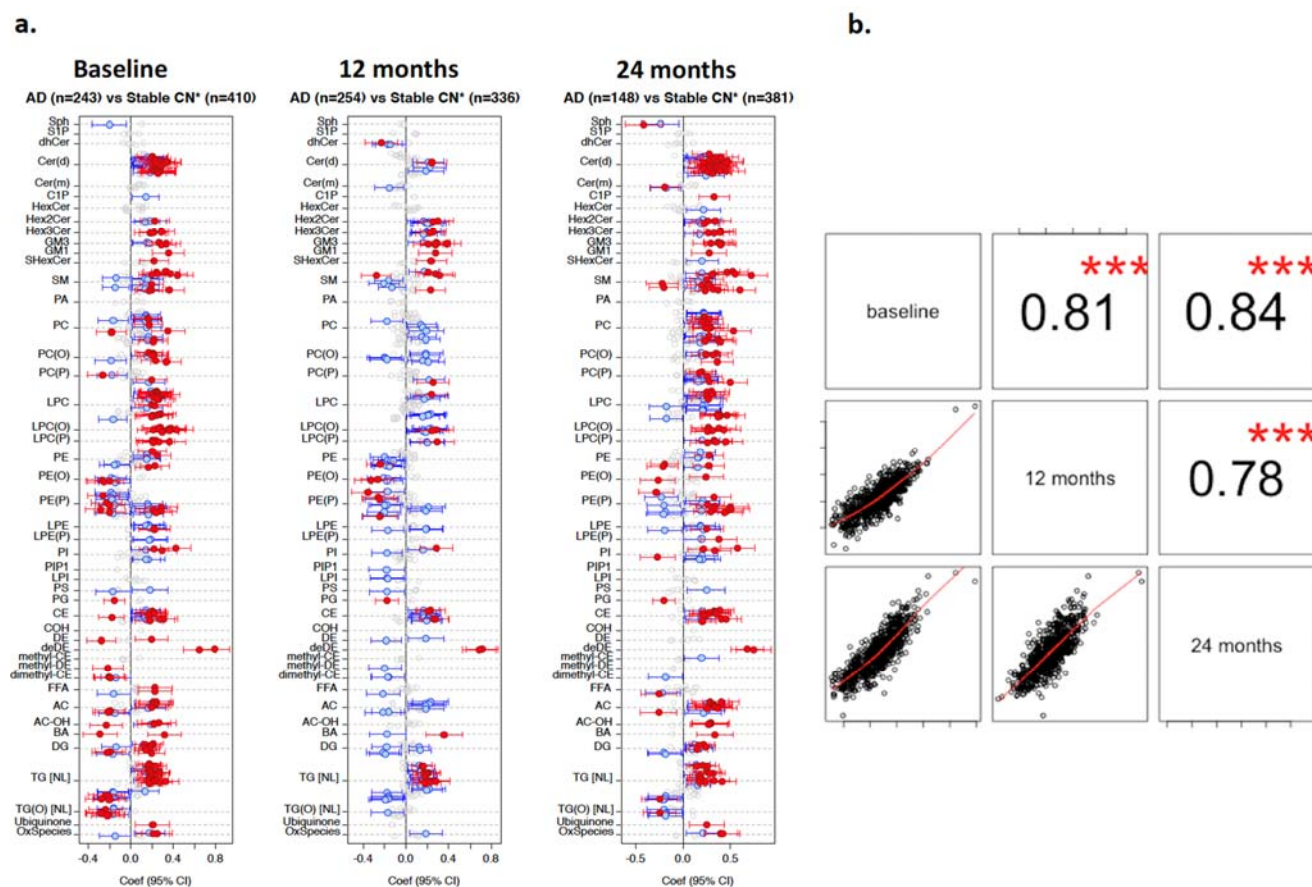
601
602
603
604
605
606
607

608



609
610
611
612
613
614
615

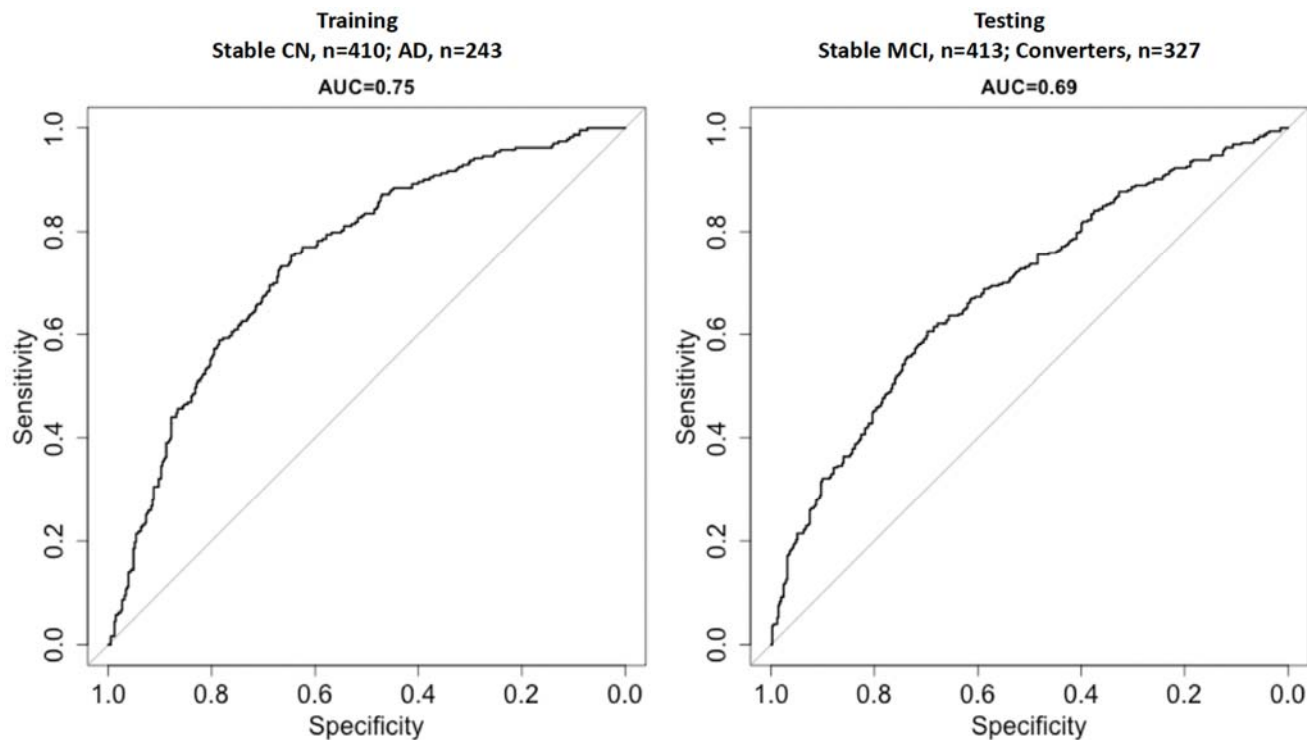
Figure 2. Metabotype conservation index. **a.** Comparison of metabotype conservation index (I_c) across two time intervals (baseline to 12 and 24 months, respectively). **b.** Boxplot comparing the distribution of I_c values < 1 between diagnostic groups after 24 months. P-values are calculated using the Wilcoxon rank sum test and show a significantly lower I_c in the AD group compared to the CN group. For visualization purposes, we transformed the index using $-\log_{10}(1 - I_c)$.



616
617
618
619
620
621

Figure 3. The associations of lipid species with AD state (AD vs CN). **a.** The associations of lipid species with AD versus CN at different time points; **b.** Pearson correlations of beta coefficients among baseline, 12 months, and 24 months. Linear regression models of lipid species against AD adjusted for age, sex, BMI, clinical lipids, fasting status, cohort, omega-3, and statin status were performed on each time point (baseline, 12 month and 24 months). The coefficients were exacted from plot

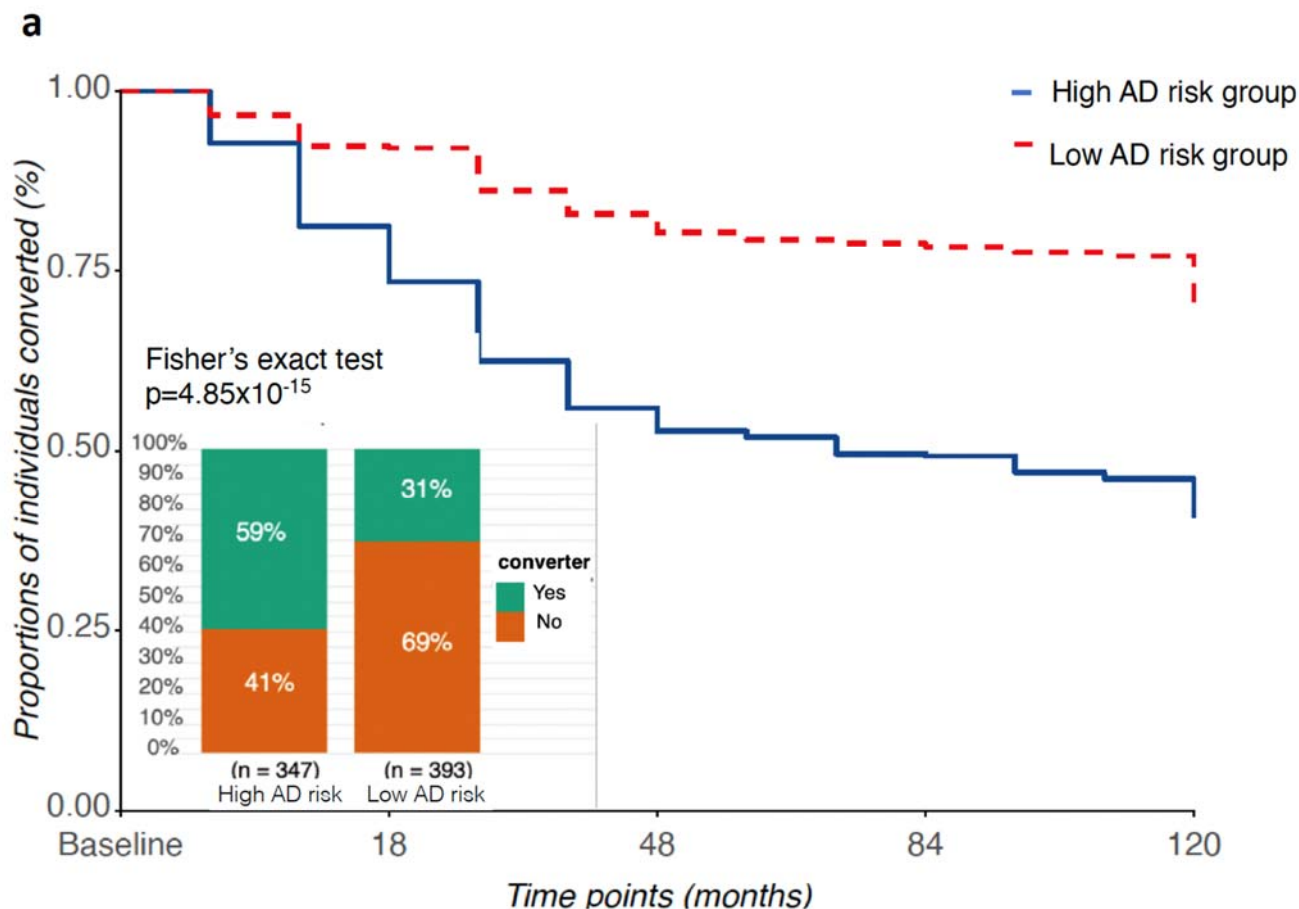
622 **a** and compared in the scatter plot **b**. Grey dots = not significant; blue dots = uncorrected $p < 0.05$; red dots
623 = BH corrected $p < 0.05$, whiskers showed 95% confidence intervals.
624



625
626 **Figure 4. Prediction performance on training set (left) and testing set (right).**

627 Ridge regression model under 5-folds CV framework was performed using the disease status (training set -
628 AD and CN group) as the binary outcome with a list of covariates including age, sex, BMI, fasting status,
629 clinical lipids, cohort, omega-3, and statin status. The models were validated on non-converting MCI and
630 converter groups with converters treated as true positives.

631
632



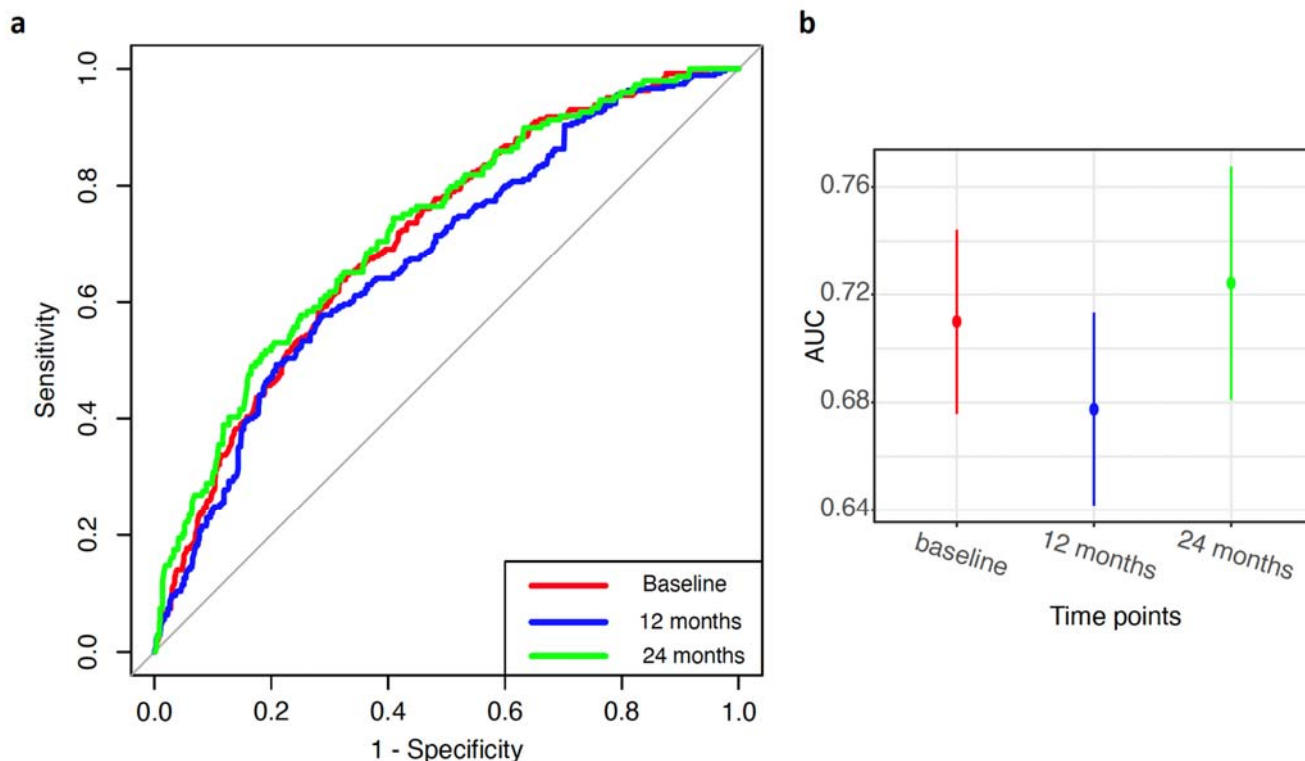
b

	BL	6m	12m	18m	24m	36m	48m	60m	72m	84m	96m	108m	120m
High AD risk group	347 (0)	347 (25)	322 (40)	282 (27)	255 (38)	217 (23)	194 (11)	183 (3)	180 (8)	172 (1)	171 (8)	163 (3)	160 (19)
Low AD risk group	393 (0)	393 (13)	380 (17)	363 (1)	362 (23)	339 (13)	326 (10)	316 (4)	312 (2)	310 (2)	308 (3)	305 (2)	303 (31)

633
634
635
636
637
638
639

Figure 5. The proportion of converters in stratified non-converting MCI and converter groups.

a. The proportions of individuals converted to AD at different time points between high AD and low AD risk groups (combination of non-converting and AD converters) were plotted in the Kaplan-Meier plot. The fisher's exact test showed the distribution of converters in the high and low AD risk groups. **b.** The exact numbers of converters out of the total number of individuals at each time point were detailed in brackets.

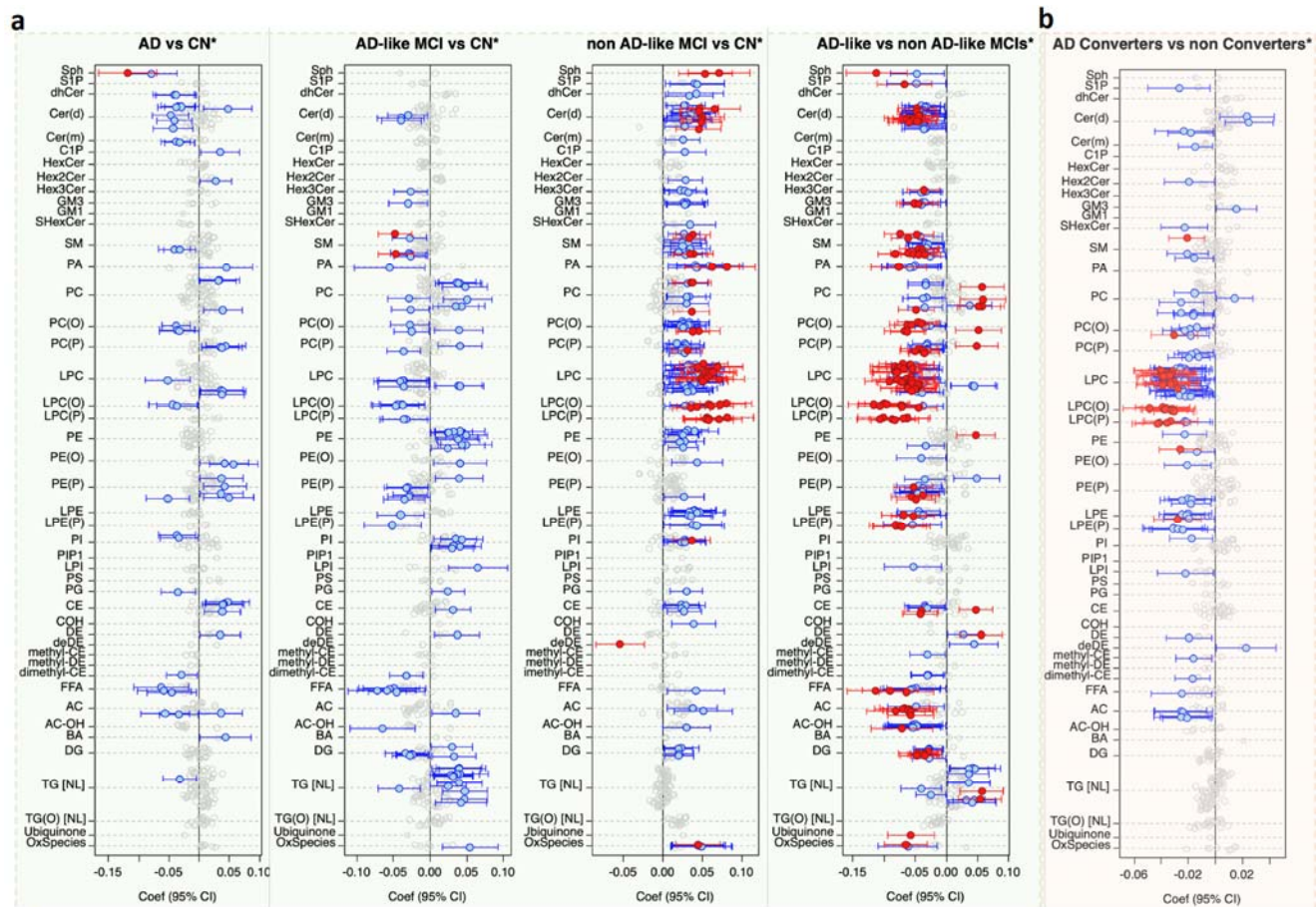


640
641
642
643
644
645
646
647
648
649
650

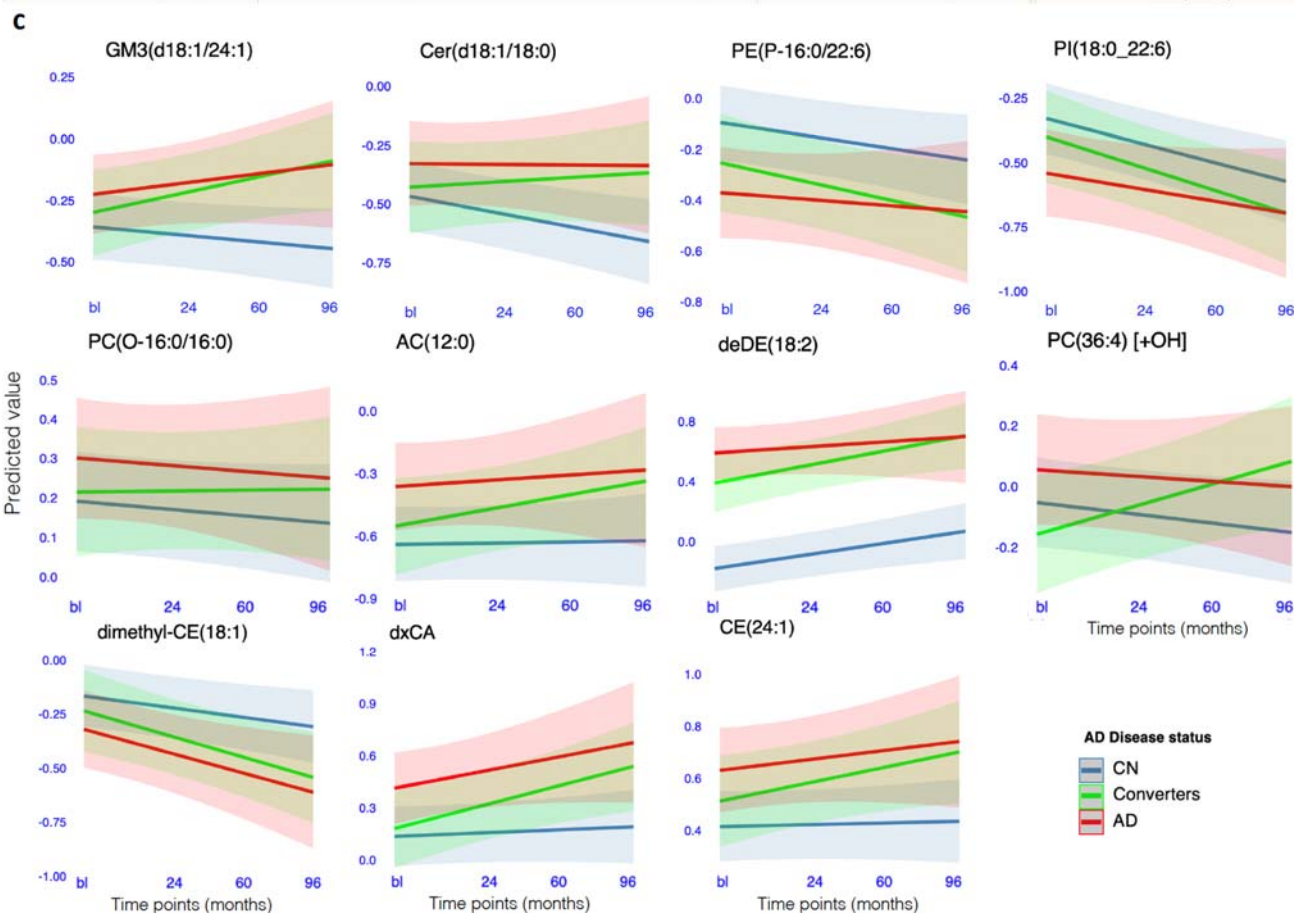
Figure 6. The performance of Lipidomic scores from AD-CN model across baseline, 12 months), and 24 months.

a ROC AUC curves across baseline (red), 12 months (blue), and 24 months (green); b The AUC and 95% confident interval across baseline (red), 12 months (blue), and 24 months (green). The weights were first generated from ridge regression models on AD-CN data set at baseline using five-folds cross-validation. Then, the weights were separately applied to baseline, 12 months, and 24 months to generate the lipidomic scores at different time points. Further, the AUC was calculated on the lipidomic scores and the AD diagnosis status of each individuals (binary variable - AD and non-AD status that was the combination of CN and MCI).

651



652

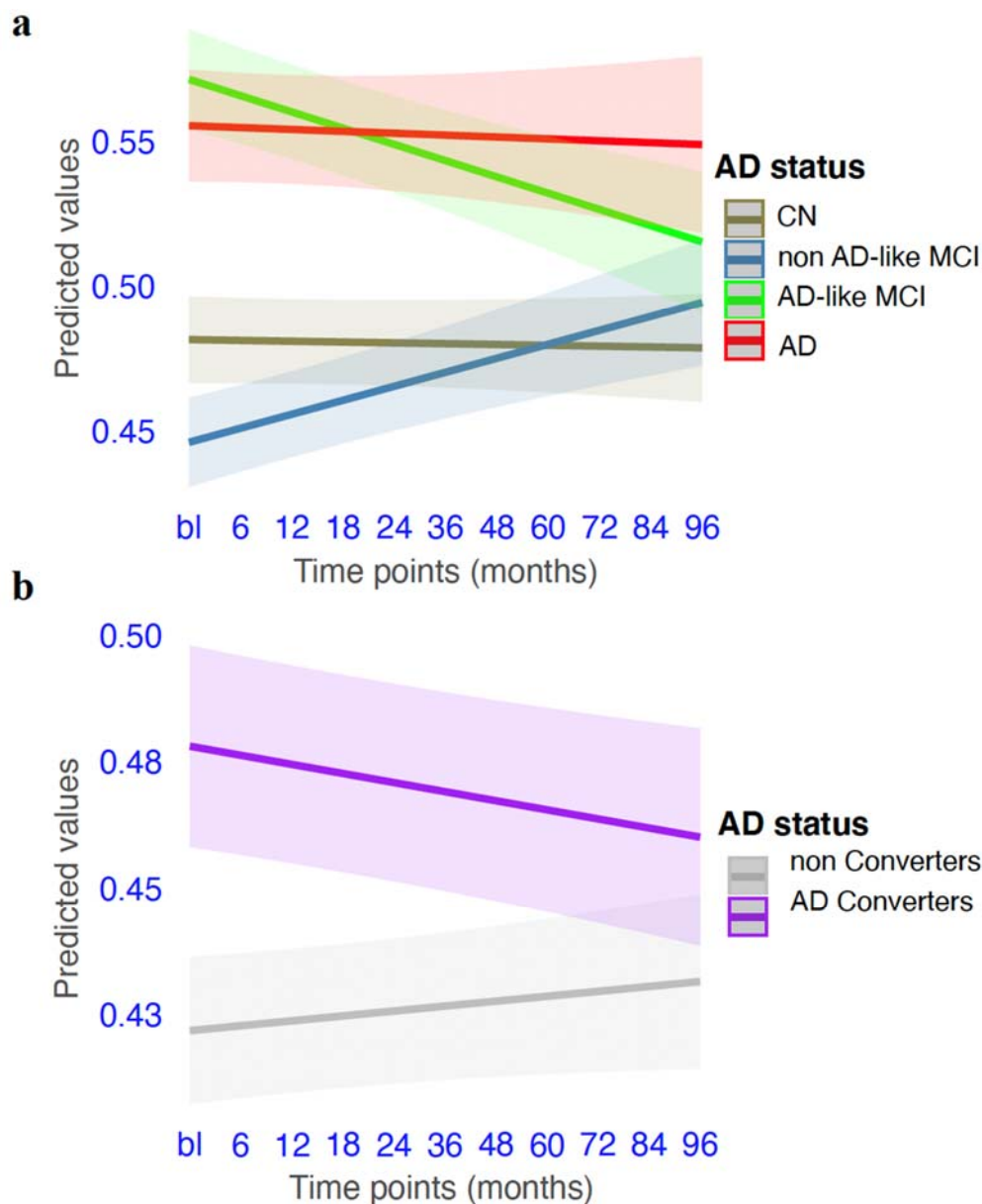


653

654

Figure 7. Trajectory of lipid species between different AD diagnosis groups.

655 The linear mixed model was performed to examine the association of the changes of individual lipid species
656 with AD diagnosis state: **a.** After excluding the AD converters, we compared trajectory of lipid species
657 between the AD (652) vs non-converting CN (1,049), AD-like MCI (433) vs non-converting CN (1,049), and
658 non AD-like MCI (781) vs non-converting CN (1,037); **b.** Using the similar model in the data sets excluding AD
659 cases, we compared the converters (1,353) vs non-converter groups (the combination of non-converting CN
660 and two non-converting MCI groups; 2,283). The covariates included age, sex, BMI, fasting status, HDL-C,
661 total cholesterol, triglycerides, cohort, omega-3, and statin status; **c.** The trajectory of selected individual
662 lipid species was further examined among the diagnosis groups of non-converting CN, AD converters, and
663 AD using the above linear mixed model on the data set excluding MCI.
664
665



666 **Figure 8. Trajectory of AD-CN scores between different AD diagnosis groups.**
667 **a.** The change of scores among non-converting CN, non AD-like MCI, AD-like MCI, and AD; **b.** The change of
668 scores between AD converters and non converters (the combination of non-converting CN and MCIs).
669 Linear mixed model was used to examine the changes of AD-CN scores among AD diagnosis groups
670 adjusted by age, sex, BMI, fasting status, AD diagnosis status at baseline, total cholesterol, HDL-C,
671 triglycerides, cohort, omega-3, and statin status.
672
673