

Central Lancashire Online Knowledge (CLoK)

Title	Raman spectroscopy to diagnose Alzheimer's disease and dementia with Lewy bodies in blood
Type	Article
URL	https://clock.uclan.ac.uk/23116/
DOI	https://doi.org/10.1021/acschemneuro.8b00198
Date	2018
Citation	Paraskevaidi, Maria, Medeiros-De-morais, Camilo De lelis orcid iconORCID: 0000-0003-2573-787X, Halliwell, Diane E., Mann, David M.A., Allsop, David, Martin-Hirsch, Pierre L. and Martin, Francis L (2018) Raman spectroscopy to diagnose Alzheimer's disease and dementia with Lewy bodies in blood. ACS Chemical Neuroscience, 9 (11). pp. 2786-2794. ISSN 1948-7193
Creators	Paraskevaidi, Maria, Medeiros-De-morais, Camilo De lelis, Halliwell, Diane E., Mann, David M.A., Allsop, David, Martin-Hirsch, Pierre L. and Martin, Francis L

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.1021/acschemneuro.8b00198>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

Raman spectroscopy to diagnose Alzheimer's disease and dementia with Lewy bodies in blood

Maria Paraskevaidi^{a,*}, Camilo L. M. Morais^a, Diane E. Halliwell^a, David M. A. Mann^b, David Allsop^c, Pierre L. Martin-Hirsch^d and Francis L. Martin^{a,*}

^aSchool of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston PR1 2HE, UK

^bDivision of Neuroscience and Experimental Psychology, School of Biological Sciences, University of Manchester, Greater Manchester Neurosciences Centre, Salford Royal Hospital, Salford M6 8HD, UK

^cDivision of Biomedical and Life Sciences, Faculty of Health and Medicine, Lancaster University, Lancaster LA1 4YQ, UK

^dDepartment of Obstetrics and Gynaecology, Central Lancashire Teaching Hospitals NHS Foundation Trust, Preston PR2 9HT, UK

*To whom correspondence should be addressed: mparaskevaidi@uclan.ac.uk or flmartin@uclan.ac.uk

28 Abstract

29

30 Accurate identification of Alzheimer's disease (AD) is still of major clinical importance
31 considering the current lack of non-invasive and low-cost diagnostic approaches. Detection of
32 early-stage AD is particularly desirable as it would allow early intervention and/or recruitment
33 of patients into clinical trials. There is also an unmet need for discrimination of AD from
34 dementia with Lewy bodies (DLB), as many cases of the latter are misdiagnosed as AD.
35 Biomarkers based on a simple blood test would be useful in research and clinical practice.
36 Raman spectroscopy has been implemented to analyse blood plasma of a cohort that consisted
37 of early-stage AD, late-stage AD, DLB and healthy controls. Classification algorithms
38 achieved high accuracy for the different groups: early-stage AD *vs* healthy with 84%
39 sensitivity, 86% specificity; late-stage AD *vs* healthy with 84% sensitivity, 77% specificity;
40 DLB *vs* healthy with 83% sensitivity, 87% specificity; early-stage AD *vs* DLB with 81%
41 sensitivity, 88% specificity; late-stage AD *vs* DLB with 90% sensitivity, 93% specificity; and
42 lastly, early-stage AD *vs* late-stage AD 66% sensitivity and 83% specificity. G-score values
43 were also estimated between 74-91%, demonstrating that the overall performance of the
44 classification model was satisfactory. The wavenumbers responsible for differentiation were
45 assigned to important biomolecules which can serve as a panel of biomarkers. These results
46 suggest a cost-effective, blood-based biomarker for neurodegeneration in dementias.

47

48

49

50 **Keywords:** Alzheimer's disease; Dementia with Lewy bodies; Raman spectroscopy; blood
51 plasma; biomarkers

52 Introduction

53 Alzheimer's disease (AD) and dementia with Lewy bodies (DLB) constitute the two
54 most common causes of dementia. AD and DLB can share common symptoms and clinical
55 characteristics, which can lead to misdiagnosis. A clear distinction between these two causes
56 of dementia is necessary in terms of pharmacological treatment and outcome evaluation ^{1, 2}.
57 The neuropathological hallmarks of AD include senile plaques (containing accumulated
58 amyloid-beta (A β) peptide) and neurofibrillary tangles (composed of hyperphosphorylated tau
59 protein), while in DLB the hallmark pathology is the abnormal aggregation of α -synuclein into
60 Lewy bodies and Lewy neurites ^{3, 4}. The ability to index the presence of these pathological
61 features in very early stages (*i.e.*, prodromal disease), or even before symptoms occur (*i.e.*, pre-
62 clinical disease), would allow an earlier intervention before irreversible neuronal death occurs,
63 as well as facilitating early recruitment into clinical trials.

64 Accurate detection of dementia is essential for improving the lives of those affected.
65 Current diagnostic approaches employ neuroimaging techniques, such as magnetic resonance
66 imaging (MRI) and positron emission tomography (PET) scans (amyloid-PET and more
67 recently tau-PET), or cerebrospinal fluid (CSF) biomarkers, but these methods have many
68 limitations ⁵⁻⁸. A combination of family and clinical history, as well as a series of different
69 memory and psychological tests is often required for diagnosis, but not all pathologically
70 similar cases will present with the same "clinical phenotype"; many studies have shown
71 contradictory results regarding the suitability of these biomarkers for accurate diagnosis.
72 Recently, blood biomarkers have emerged as a potential means to test for neurodegenerative
73 diseases, with some being capable of detecting early-stage disease ⁹⁻¹¹. The rationale behind
74 the use of blood samples is based on the daily release of 500 ml CSF into the bloodstream,
75 which potentially renders blood a rich source of brain biomarkers ¹².

76 Raman is a spectroscopic technique that extracts biological information by applying a
77 monochromatic, laser light onto the sample under interrogation; electrons are thus excited to
78 virtual energy levels. When these electrons return to the original energy level, in the form of a
79 photon, there is no energy shift (known as elastic or Rayleigh scattering), whereas when they
80 return to a lower or a higher energy level there is a gain or loss of energy, respectively (known
81 as inelastic or Raman scattering) ¹³. The shift in the energy allows the generation of a spectrum
82 which is indicative of the chemical bonds present in the sample. The characteristic spectra that
83 are derived from Raman spectroscopy, represent a number of different biomolecules within a
84 sample (*e.g.*, proteins, carbohydrates, lipids, DNA) ¹⁴. Recent studies have employed Raman
85 spectroscopy to study different diseases, such as malaria, oral and colorectal cancer, in
86 biological fluids ¹⁵⁻¹⁷.

87 The aim of the present study was to diagnose patients with Alzheimer's disease, in early
88 and late disease stages, and patients with DLB, as well as to discriminate between AD and
89 DLB. To achieve this, blood plasma was analysed with Raman spectroscopy as a minimally
90 invasive procedure that would also allow repeated measurements for follow-up of individuals.

91 Results

92 We enrolled 56 individuals into this study who were classified into 4 groups; early stage
93 AD (n=11; age range: 50-74 years), late stage AD (n=15; age range: 50-79 years), DLB (n=15;
94 age range: 23-73 years) and healthy controls (n=15; age range: 23-73 years) (Table 1). Early
95 and late-stage AD was defined according to the duration of illness, from designated age at onset
96 up to age at sample collection. P-values were calculated based on age and statistical differences
97 were detected only for the following two subgroup comparisons: Late AD vs Healthy (P=0.004)
98 and DLB vs Healthy (P <0.001). For all the other comparison groups (Early AD vs Healthy,
99 Early AD vs DLB, Late AD vs DLB, Early AD vs Late AD), there was no statistical difference

100 observed due to age ($P > 0.005$) (Supplementary Table 1). Even though there was age difference
101 between the controls and AD individuals, no correlation was observed between age and AD
102 spectra after using partial least squares regression ($R^2 = 0.107$, 2 latent variables with 99.93%
103 cumulative explained variance) and no statistical difference was observed in the spectra of AD
104 patients with age lower and higher than 54 years of age (average control age) with a 95%
105 confidence level ($P > 0.005$). This indicates that age did not affect the spectral distribution
106 within the AD class. Similarly, no statistical differences were observed in the Raman spectra
107 of the different groups due to gender (male vs female) (Supplementary Fig. 8).

108 **Early stage AD vs healthy individuals.** After pre-processing of the spectral data,
109 principal component analysis followed by linear discriminant analysis (PCA-LDA) was
110 applied to the derived dataset. A one-dimensional (1D) scores plot was generated to account
111 for differences and similarities between early stage AD and healthy subjects (Fig. 1A); after
112 statistical analysis, the two classes showed significant differences ($P < 0.0001$, 95% CI = 0.0503
113 to 0.0622). A loadings plot served as a biomarker extraction method, identifying the top six
114 peaks responsible for differentiation: 1650 cm^{-1} , 1529 cm^{-1} , 1432 cm^{-1} , 1161 cm^{-1} , 996 cm^{-1}
115 and 911 cm^{-1} (Fig. 1B). A statistical test was performed on each peak individually to calculate
116 the P-value and investigate the differences in Raman intensity between the two groups
117 (Supplementary Fig. S2, Supplementary Table 2). Figure 1C summarises the tentative
118 assignments along with the P-values for these peaks (denoted with asterisks). Further analysis
119 was conducted to classify the two classes; support vector machine (SVM) was the classification
120 algorithm that was used, achieving 84% sensitivity and 86% specificity, with the G-score
121 estimated at 85% and Youden's index at 70% (Table 2, Supplementary Fig. 1A).

122 **Late AD vs healthy individuals.** A similar approach was followed for the discrimination
123 between late AD and healthy individuals. Figure 2A represents the scores plot after cross
124 validated PCA-LDA and reveals statistically significant differences between the groups (P

125 <0.0001, 95% CI = 0.0655 to 0.0834). The top six discriminatory peaks that were selected for
126 this comparison group were 1648 cm⁻¹, 1530 cm⁻¹, 1432 cm⁻¹, 1259 cm⁻¹, 1164 cm⁻¹ and 1003
127 cm⁻¹ (Fig. 2, Supplementary Fig. S3 and Supplementary Table 2). After the SVM classification,
128 late AD was discriminated from healthy individuals with 84% sensitivity, 77% specificity and
129 the G-score estimated at 80% and Youden's index at 61% (Table 2, Supplementary Fig. 1B).

130 **DLB vs healthy individuals.** Scores plot was again generated after cross validated
131 PCA-LDA to compare DLB with healthy controls. Statistically significant differences were
132 found between the groups (P <0.0001, 95% CI = 0.0982 to 0.1166) and the wavenumbers that
133 were mostly responsible for this discrimination are shown in the respective loadings plot (Fig.
134 3B): 1647 cm⁻¹, 1604 cm⁻¹, 1418 cm⁻¹, 1384 cm⁻¹, 1002 cm⁻¹ and 933 cm⁻¹. The differences in
135 Raman intensity for each wavenumber are shown in more detail in Supplementary Fig. 4 and
136 Supplementary Table 2. Sensitivity and specificity, after SVM, were 83% and 87%,
137 respectively, while the G-score was calculated at 85% and Youden's index at 70% (Table 2,
138 Supplementary Fig. 1C).

139 **Early stage AD vs DLB.** The scores plot for the comparison between early stage AD and
140 DLB is shown in Fig. 4A. After statistical analysis, the difference between these two cohorts
141 was statistically significant (P <0.0001, 95% CI = -0.0791 to -0.0649). The wavenumbers that
142 were found as the most important, after cross-validated PCA-LDA are shown along with their
143 tentative assignments in Fig. 4 and were the following: 1645 cm⁻¹, 1513 cm⁻¹, 1376 cm⁻¹, 1253
144 cm⁻¹, 1161 cm⁻¹ and 1003 cm⁻¹ (Supplementary Fig. 5 and Supplementary Table 2). The
145 sensitivity and specificity values from this comparison were 81% and 88%, respectively, with
146 the G-score at 84% and Youden's index at 69% (Table 2, Supplementary Fig. 1D).

147 **Late AD vs DLB.** Analyses were conducted to discriminate between late AD and DLB
148 (Fig. 5). Significant differences were found after statistical analysis on the PCA-LDA scores

149 plot ($P < 0.0001$; 95% CI = 0.138 to 0.1596). The following are the top six wavenumbers that
150 were found to be responsible for the observed differentiation: 1646 cm^{-1} , 1614 cm^{-1} , 1437 cm^{-1} ,
151 1216 cm^{-1} , 1164 cm^{-1} and 1003 cm^{-1} . Differences in the Raman intensity at these peaks are
152 given in Supplementary Fig. 6 and Supplementary Table 2. The tentative assignments for these
153 wavenumbers are shown in Fig. 5C. Sensitivity and specificity were 90% and 93%,
154 respectively, with G-score being 91% and Youden's index at 84% (Table 2, Supplementary
155 Fig. 1E).

156 **Early AD vs late AD.** A comparison between early and late-stage AD patients was also
157 performed. After cross validated PCA-LDA, the scores plot revealed statistically significant
158 differences between the two groups ($P < 0.0001$; 95% CI = -0.0943 to -0.0624) (Fig. 6). The
159 loadings plot denoted the following six wavenumbers as the most important: 1650 cm^{-1} , 1476
160 cm^{-1} , 1432 cm^{-1} , 1161 cm^{-1} , 1003 cm^{-1} , 642 cm^{-1} (Supplementary Fig. 7 and Supplementary
161 Table 2). After classification of the two populations, 66% of the early AD spectra were
162 correctly identified with 34% been misclassified as late AD; and 83% of the late AD cases
163 were correctly identified with 17% misclassified as early stage AD (Table 2). G-score was
164 calculated at 74% and Youden's index at 49% (Table 2, Supplementary Fig. 1F).

165 Discussion

166 Amyloid PET imaging has been shown to improve the diagnostic accuracy of AD ¹⁸.
167 However, one of the limitations is that only subjects with advanced dementia and relatively
168 heavy plaque densities will be amyloid PET-positive; thus, individuals may not be identified
169 early enough to be used in prevention studies using anti-amyloid therapeutics ¹⁹. The detection
170 accuracy of neuropathologically defined AD with PET imaging has been estimated at 69-95%
171 sensitivity and 83-89% specificity ²⁰. In the case of DLB patients, PET imaging shows
172 increased A β deposition in >50% of patients with DLB which limits its value in distinguishing
173 between AD and DLB ³. In a recent study, clinical and pathological diagnoses were compared

174 and DLB patients were identified with 73% sensitivity and 93% specificity; such findings
175 suggest that there is still need for improvement in discriminating between these conditions ²¹.
176 When using MRI for AD diagnosis, a decreased volume of hippocampus and other temporal
177 lobe structures is indicative of neurodegeneration; visual rating scales evaluating the degree of
178 atrophy provide ~80-85% sensitivity and specificity when comparing AD to healthy
179 individuals and slightly lower sensitivity and specificity when comparing to amnesic mild
180 cognitive impairment (MCI) ⁶. However, atrophy patterns can be similar in different diseases
181 while at the same time some unusual forms of AD may have atypical patterns ²².

182 Established CSF biomarkers that are currently used in clinical practise to diagnose AD,
183 also known as “core biomarkers”, include decreased levels of A β 42, or decreased A β 42:A β 40
184 ratio, and increased levels of total tau (T-tau) or hyperphosphorylated tau (P-tau) ²³. In a
185 systematic review and meta-analysis, a number of different biomarkers has been associated
186 with AD in both CSF and blood; namely, neurofilament light chain (NfL), neuron-specific
187 enolase (NSE), visinin-like protein 1 (VLP-1), heart fatty acid binding protein (HFABP),
188 chitinase-3-like protein 1 (YKL-40) in CSF, as well as T-tau and P-tau in blood plasma ^{10, 24},
189 ²⁵. More recently, an elevated level of plasma NfL has been suggested as a promising biomarker
190 to distinguish AD and MCI from healthy subjects. The accuracy for the comparison between
191 AD and healthy controls, after testing for NfL, was 87%, which is comparable to accuracies
192 achieved by CSF testing (88% A β 42; 90% T-tau; 87% P-tau; 89% NfL) and plasma tau (78%)
193 ¹¹. Another study, discovered and validated a set of ten lipids in plasma to detect preclinical
194 AD in cognitively normal older adults within a 2-3 year timeframe; this panel achieved 90%
195 accuracy ⁹. Even though it is now established that the α -synuclein gene (*SNCA*) is associated
196 with a few families with Parkinson’s disease (PD) and DLB, CSF α -synuclein is not yet proven
197 as a potential biomarker. CSF and blood biomarkers for the diagnosis of DLB remain elusive,

198 with A β , T-tau and P-tau remaining the most current measurements to predict cognitive decline
199 and determine associated AD pathology ³.

200 In the present study, we included patients with AD, in both early and later stages of the
201 disease, DLB, as well as healthy individuals. The blood-based Raman spectroscopic technique,
202 provided excellent diagnostic accuracy not only between diseased and non-diseased states, but
203 also between the two different types of dementia. Statistically significant age differences were
204 only observed for Late AD vs Healthy (P=0.004) and DLB vs Healthy (P <0.001). The age
205 difference between healthy controls and both Late AD and DLB patients was somehow
206 expected as these diseases manifest mainly in older individuals. A larger dataset containing a
207 wider age range would be necessary for adjusting the model for age. However, the fact that
208 diagnostic accuracies remain exceptionally high for the subgroups with no age differences (*e.g.*,
209 Late AD vs DLB showing 90% sensitivity and 93% specificity), implies that the age factor was
210 not solely responsible for the achieved segregation between the cohorts. Similarly, no statistical
211 differences were observed due to gender after calculating the P values for each spectral
212 wavenumber; therefore, gender differences did not change the spectral profile.

213 Raman spectroscopy can reveal invaluable information about a biological sample as it
214 provides the overall status of a sample, indicating disease. The results from such an approach
215 are comparable to, and in some cases even better than, conventional methods, as they allow for
216 simultaneous investigation of a panel of different biomarkers and therefore may be more
217 suitable for complex diseases. Furthermore, Raman allows for a low-cost, label-free and non-
218 destructive diagnosis in contrast to current imaging techniques and molecular CSF and/or blood
219 tests. Previous studies have estimated the cost of an MRI and PET scan at £163 and £844,
220 respectively, while an enzyme-linked immunosorbent assay (ELISA) measurement (96-well
221 plate) of the core biomarkers (A β 42, T-tau, P-tau) costs £826 per kit ^{26, 27}. In contrast, a blood
222 test employing Raman spectroscopy is negligible in terms of consumables although there

223 would be costs in terms of employee time for samples preparation and analysis; overall cost
224 would fall dramatically as the data infrastructure to allow remote classification of samples
225 became available. Even the upfront cost of Raman instrumentation, often varying from £3,000-
226 £150,000, is low in comparison with other approaches and would again fall with the
227 development of hand-held devices; also the running costs are minimal with electrical power
228 being the only requirement. Over the longer-term, lasers may need to be replaced (~every 6-7
229 years), but daily running costs are close to zero.

230 Discriminatory peaks have also been identified for all of the different comparison
231 groups and could possibly be used as biomarkers for differential diagnosis or screening of high-
232 risk populations. For instance, higher levels of Amide II peaks ($\sim 1530\text{ cm}^{-1}$) were seen in both
233 early ($P < 0.0001$) and late stage AD ($P < 0.0001$) patients and could possibly be represented by
234 an increase in tau proteins or NfL in plasma, which have been suggested previously as
235 promising biomarkers (Supplementary Fig. 2, Supplementary Fig. 3)¹¹. Also, the observed
236 decrease in lipids ($\sim 1432\text{ cm}^{-1}$) could be due to damaged phospholipid membranes caused by
237 oxidative stress. These findings are in line with previous results of a larger-scale study our
238 research team conducted, in which infrared (IR) spectroscopy was employed to diagnose AD
239²⁸. An advantage of Raman spectroscopy over IR is its ability to analyse aqueous samples which
240 would allow the analysis of fresh samples without the need of prior dehydration; this would be
241 particularly beneficial for use in a clinic. Noticeably, in this preceding study, lipid peaks were
242 also decreased ($\sim 1740\text{ cm}^{-1}$, $P < 0.05$; $\sim 1450\text{ cm}^{-1}$, $P < 0.005$) and Amide II was also increased
243 ($\sim 1540\text{ cm}^{-1}$, $P = 0.003$) in AD patients. However, Amide I ($\sim 1650\text{ cm}^{-1}$), which is indicative
244 of A β load, was not found to be statistically different ($P = 0.12$), in contrast to the current study
245 where it was significant in both early ($P = 0.0003$, 95% CI = 0.0008 to 0.0028) and late stage
246 AD ($P < 0.0001$, 95% CI = 0.0016 to 0.0029). Previous studies have noted altered levels of
247 aromatic amino acids in plasma and serum of AD patients²⁹. Some studies have shown an

248 increase in phenylalanine in the brain of AD subjects ³⁰⁻³², while others suggest a decrease ³³,
249 ³⁴. In our study, the level of phenylalanine was increased in DLB cases, whereas in late AD
250 phenylalanine was decreased when compared to healthy subjects. Between AD and DLB
251 patients, the latter cohort showed higher levels of phenylalanine, which could possibly relate
252 to their α -synuclein pathology (Supplementary Fig. 5). Previous studies have shown altered
253 metabolic profiles of PD patients (also related to α -synuclein aggregation) when compared to
254 normal controls, and these differences were related to metabolic pathway variations such as
255 phenylalanine metabolism ³⁵⁻³⁷.

256 We were particularly interested in examining early stage AD cases as it is of crucial
257 importance to identify individuals before brain damage becomes very severe. Evidence of
258 changes here would allow for an on-time intervention, potentially to slow down the disease,
259 psychologically prepare the affected person and their family, as well as provide them with the
260 opportunity to take part in early intervention trials. Surprisingly, the diagnostic accuracy was
261 slightly higher for early AD than for late AD. After comparison of these two groups, 66% of
262 early AD and 83% of late AD were correctly classified. Of the wavenumbers which were shown
263 to contribute the most to the segregation between the classes, a peak assigned to Amide II
264 proteins ($\sim 1476\text{ cm}^{-1}$) and a peak assigned to C-C and C-S vibrations of proteins ($\sim 642\text{ cm}^{-1}$)
265 were found to be statistically significant (Supplementary Fig. 7). A potential explanation for
266 the decreased level of Amide II in early stage AD cases could be the lower density of
267 neurofibrillary tangles in the brain during early stages. Previous studies have suggested that
268 kinase mutations and dysfunction play an important role in the development of disorders such
269 as cancer and neurodegeneration ³⁸. Specifically, cyclin-dependent kinase 5 (cdk5), which is
270 involved in the abnormal hyperphosphorylation of tau, has been suggested to accumulate at a
271 relatively early stage in the neocortex ³⁹; more recent research has also shown that a cellular
272 stress response, caused by accumulation of misfolded proteins, induces the activity of a major

273 tau kinase (GSK-3 β) and occurs at an early stage of neurofibrillary degeneration leading to AD
274 pathogenesis⁴⁰. Therefore, this may potentially explain the increased level of the protein peak
275 at 642 cm⁻¹. Special attention was also given to the accurate diagnosis of DLB and
276 differentiation from AD which is especially important to provide the appropriate treatment;
277 DLB cases respond well to cholinesterase inhibitors but have severe neuroleptic sensitivity
278 reactions, which are associated with significantly increased morbidity and mortality⁴¹.

279 A critical aspect for every new biomarker, diagnostic or treatment approach is the
280 repetition and validation of the analytical process and in different cohorts. Previously, a few
281 studies also employed Raman spectroscopy to diagnose AD in blood, achieving high
282 classification accuracy. Carmona *et al.* distinguished AD (n=35) from normal (n=12) with 89%
283 sensitivity and 92% specificity⁴². Ryzhikova *et al.* included serum samples from 20 AD
284 patients, 18 patients with other neurodegenerative dementias (OD) (5 with DLB, 10 with
285 Parkinson's disease dementia and 3 with frontotemporal dementia) and 10 healthy individuals
286 and achieved 95% sensitivity and specificity⁴³. However, the fact that a range of different
287 dementias were all taken in the same group, may obscure the actual classification capability
288 between AD and DLB. Moreover, no spectroscopic approach has been employed so far to
289 investigate DLB in more detail.

290 A limitation of the current study is the small number of participants, which can affect
291 sensitivity and specificity estimates. However, G-score values were estimated at 74-91%,
292 denoting that the models were not overfitted. G-score does not account the size of classes, thus
293 providing robust information about the classification ability even in smaller cohorts⁴⁴.
294 Youden's index values ranged between 49% (early AD vs late AD) and 84% (late AD vs DLB).
295 This parameter is a probability indicator of the model's ability to avoid failure. Youden's
296 indexes above 70% for early AD vs healthy, DLB vs healthy and late AD vs DLB indicate that
297 these models have low probability of misclassification in the future. Another limitation of this

298 study, as well as similar previous ones, is the lack of serial samples from the same individuals
299 which would validate the results and demonstrate repeatability.

300 In summary, diagnosis of early stage AD, late stage AD, DLB as well as differentiation
301 between the two dementias was achieved, opening a new road for potential applications in a
302 clinical setting. Some of the future uses of spectroscopy could be the detection of
303 prodromal/pre-demented cases; the differential diagnosis of different dementias that would
304 allow the appropriate treatment and/or recruitment into clinical trials; and the further
305 monitoring of patients that do finally take part in clinical trials.

306 **Methods**

307 **Patient information.** We enrolled 56 individuals into this study who were classified
308 into four groups; early stage AD (n=11), late stage AD (n=15), DLB (n=15) and healthy
309 controls, usually spouses (n=15). Early and late-stage AD was defined according to the duration
310 of illness, from designated age at onset up to age at sample collection. Early stage was defined
311 as up to two years from designated age at onset, whereas late stage AD was defined as any
312 duration beyond this time point. Clinical and demographic data is summarised in Table 1.
313 Information on apolipoprotein $\epsilon 4$ (*APOE4*) status and gender was not available for two subjects
314 from the healthy control group. Patients were recruited at Salford Royal Hospital (Salford, UK)
315 with informed consent prior to enrolment in accordance with Local Ethical Approval
316 (05/Q1405/24 conferred by North West 10 Research Ethics Committee Greater Manchester
317 North). Patients were diagnosed according to battery of psychological testing (Manchester
318 Neuropsychology Inventory) performed at a Specialist referral Centre (Cerebral Function Unit,
319 Greater Manchester Neurosciences Centre, Salford Royal Hospital). All methods were
320 performed in accordance with the relevant guidelines and all other applicable laws and
321 regulations. At time of diagnosis patients were not receiving any medications, such as

322 anticholinesterase treatments. Most patients had received MRI scans but these were used only
323 to support the neuropsychological outcomes.

324 **Sample preparation and *APOE* genotyping.** Whole blood samples were collected into
325 EDTA tubes, centrifuged at 2000 rpm at 4°C for 10 min to separate erythrocytes from plasma.
326 Plasma was collected in 0.5 mL clean, plastic tubes, stored at -80°C and thawed at room
327 temperature prior to spectroscopic interrogation. After the samples were thawed, 50 µL were
328 deposited on glass slides covered with aluminium foil, which has been shown to be featureless
329 in Raman ¹⁴, and were then left to air-dry overnight. DNA was extracted by routine methods
330 from blood samples of patients and control subjects; *APOE* alleles were determined by PCR
331 ⁴⁵.

332 **Raman spectroscopy.** Raman spectra were collected with an InVia Renishaw Raman
333 spectrometer coupled with a charge-coupled device (CCD) detector and a Leica microscope.
334 A 200 mW laser diode was used at a wavelength of 785 nm with a grating of 1200 l/mm, and
335 the system was calibrated to 520.5 cm⁻¹ with a silicon source, before every run. After trial-and-
336 error measurements to optimise the experimental parameters, we concluded to a 10 second
337 exposure time, 5% laser power and 2 accumulations at a spectral range 2000-400 cm⁻¹ to
338 achieve optimum spectral quality. Twenty-five point spectra were taken per sample using a
339 50× objective to focus the laser beam on the sample.

340 **Pre-processing of spectral data and multivariate analysis.** Spectra were initially
341 corrected for cosmic rays using the Renishaw WiRe software. An in-house developed IRootLab
342 toolbox (<http://trevisanj.github.io/irootlab/>) was then implemented within MATLAB
343 (MathWorks, Natick, USA) for further pre-processing and computational analysis of the data.
344 All spectra were cut at 1750-500 cm⁻¹, first order differentiated with Savitzky-Golay (SG)
345 (window of 9 points; 2nd polynomial filter) to smooth out the noise and vector normalised to

346 account for non-biological differences, such as varying concentration or thickness of the
347 sample; the resulting dataset was then mean-centered before implementation of cross-validated
348 (k-fold=10, leave-one-out) principal component analysis followed by linear discriminant
349 analysis (PCA-LDA). The leave-one-out cross-validation was implemented to avoid
350 overfitting. This ensures that one sample is removed from the training set and predicted as
351 external sample during model construction in an interactive process until all samples are
352 predicted; this provides more realistic classification results. All classification models were
353 validated using 10% of the samples in a test set. PCA is an unsupervised method that reduces
354 the spectral dataset to only a few important principal components (PCs) which are responsible
355 for the majority of the variation; using a Pareto function, a number of 10 PCs was found as
356 optimal. LDA is a supervised technique, often coupled with PCA, to maximise the between-
357 class distance and minimise the within-class distance. Scores plots and loadings plots were
358 generated after PCA-LDA to visualise the differences and similarities between the groups as
359 well as to identify specific spectral peaks responsible for this differentiation; these peaks were
360 tentatively assigned to different biomolecules which can potentially serve as biomarkers ^{46, 47}.
361 After the six peaks were identified from the loadings plot, they were then extracted from
362 polynomial corrected, vector normalised spectra in order to avoid the spectral transformation
363 that first order differentiation can cause. Classification of the different comparison groups was
364 conducted using support vector machine (SVM) which is a machine-learning technique to
365 classify spectral data. For SVM implementation, the pre-processed dataset (*i.e.*, cut, SG
366 differentiated, vector normalised) was normalised to the [0, 1] range and then the optimal (C,
367 γ) combination was found using grid search. Sensitivities and specificities were therefore
368 calculated for each comparison group ⁴⁸. In order to overcome the limitation of using a small
369 cohort in this study, G-score values were also calculated to assess the overall performance of
370 the classification model ⁴⁴. The G-score is calculated as the square root of sensitivity times

371 specificity. Youden's index was calculated to assess the classifier's ability to avoid failure.
372 This parameter is estimated as sensitivity minus (1 – specificity).

373 **Statistical analysis.** The values generated after cross-validated PCA-LDA, were imported
374 into GraphPad Prism 7 to conduct the statistical analyses and calculate the P-values for each
375 comparison. Differences between two groups were assessed using a Student's t-test (two-tailed,
376 non-parametric, Mann-Whitney test, 95% confidence interval). The data were expressed as the
377 mean \pm standard deviation (SD). A P-value of 0.05 or less was considered significant in all
378 statistical tests.

379 [Additional Information](#)

380 **Availability of data and material**

381 All data (raw and pre-processed spectra) along with appropriate code identifiers will be
382 uploaded onto the publicly accessible data repository Figshare.

383 **Conflict of Interest Disclosure**

384 The authors declare that they have no competing interests.

385 **Funding**

386 This study was supported by Rosemere Cancer Foundation (RCF).

387 **Authors' contributions**

388 M.P performed the experiments, analysed the data and wrote the manuscript. The manuscript
389 was written with contributions from C.L.M.M, D.E.H, D.M.A.M, D.A, P.L.M-H and F.L.M.
390 All authors have read and approved the final manuscript.

391 **Acknowledgements**

392 MP would like to thank the Rosemere Cancer Foundation (RCF) for funding. CLMM would
393 like to thank CAPES-Brazil (grant 88881.128982/2016-01) for financial support.

394 **Supplementary Information**

395 Sensitivity and specificity rates after classification with support vector machine (SVM);
396 differences in the intensity levels of important biomolecules after the comparison between
397 early-stage AD and healthy individuals; differences in the intensity levels of important
398 biomolecules after the comparison between late-stage AD and healthy individuals; differences
399 in the intensity levels of important biomolecules after the comparison between DLB and
400 healthy individuals; differences in the intensity levels of important biomolecules after the
401 comparison between early-stage AD and DLB; differences in the intensity levels of important
402 biomolecules after the comparison between late-stage AD and DLB; differences in the intensity
403 levels of important biomolecules after the comparison between early-stage AD and late-stage
404 AD; mean values, standard deviations (SD), 95% confidence intervals (CI) and P-values after
405 statistical analysis for the different comparison groups; P-values (two-tail, 95% confidence
406 level) for the different ages in each comparison group.

407

408 References

- 409 [1] Minoshima, S., Foster, N. L., Sima, A. A., Frey, K. A., Albin, R. L., and Kuhl, D. E. (2001) Alzheimer's
410 disease versus dementia with Lewy bodies: cerebral metabolic distinction with autopsy
411 confirmation, *Ann Neurol* 50, 358-365.
- 412 [2] Mueller, C., Ballard, C., Corbett, A., and Aarsland, D. (2017) The prognosis of dementia with Lewy
413 bodies, *Lancet Neurol*.
- 414 [3] McKeith, I. G., Boeve, B. F., Dickson, D. W., Halliday, G., Taylor, J.-P., Weintraub, D., Aarsland, D.,
415 Galvin, J., Attems, J., Ballard, C. G., Bayston, A., Beach, T. G., Blanc, F., Bohnen, N., Bonanni, L.,
416 Bras, J., Brundin, P., Burn, D., Chen-Plotkin, A., Duda, J. E., El-Agnaf, O., Feldman, H., Ferman,
417 T. J., ffytche, D., Fujishiro, H., Galasko, D., Goldman, J. G., Gomperts, S. N., Graff-Radford, N.
418 R., Honig, L. S., Iranzo, A., Kantarci, K., Kaufer, D., Kukull, W., Lee, V. M. Y., Leverenz, J. B.,
419 Lewis, S., Lippa, C., Lunde, A., Masellis, M., Masliah, E., McLean, P., Mollenhauer, B., Montine,
420 T. J., Moreno, E., Mori, E., Murray, M., O'Brien, J. T., Orimo, S., Postuma, R. B., Ramaswamy,
421 S., Ross, O. A., Salmon, D. P., Singleton, A., Taylor, A., Thomas, A., Tiraboschi, P., Toledo, J. B.,
422 Trojanowski, J. Q., Tsuang, D., Walker, Z., Yamada, M., and Kosaka, K. (2017) Diagnosis and
423 management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium,
424 *Neurology* 89, 88-100.
- 425 [4] Walker, Z., Possin, K. L., Boeve, B. F., and Aarsland, D. (2015) Lewy body dementias, *Lancet* 386,
426 1683-1697.
- 427 [5] McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E.,
428 Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens,
429 P., Carrillo, M. C., Thies, B., Weintraub, S., and Phelps, C. H. (2011) The diagnosis of dementia
430 due to Alzheimer's disease: Recommendations from the National Institute on Aging-
431 Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease,
432 *Alzheimers Dement* 7, 263-269.
- 433 [6] Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., and Thompson, P. M. (2010) The clinical use of
434 structural MRI in Alzheimer disease, *Nat Rev Neurol* 6, 67-77.
- 435 [7] Saint-Aubert, L., Lemoine, L., Chiotis, K., Leuzy, A., Rodriguez-Vieitez, E., and Nordberg, A. (2017)
436 Tau PET imaging: present and future directions, *Mol Neurodegener* 12, 19.
- 437 [8] Ossenkoppele, R., Schonhaut, D. R., Schöll, M., Lockhart, S. N., Ayakta, N., Baker, S. L., O'Neil, J. P.,
438 Janabi, M., Lazaris, A., and Cantwell, A. (2016) Tau PET patterns mirror clinical and
439 neuroanatomical variability in Alzheimer's disease, *Brain* 139, 1551-1567.
- 440 [9] Mapstone, M., Cheema, A. K., Fiandaca, M. S., Zhong, X., Mhyre, T. R., MacArthur, L. H., Hall, W. J.,
441 Fisher, S. G., Peterson, D. R., Haley, J. M., Nazar, M. D., Rich, S. A., Berlau, D. J., Peltz, C. B.,

442 Tan, M. T., Kawas, C. H., and Federoff, H. J. (2014) Plasma phospholipids identify antecedent
443 memory impairment in older adults, *Nat Med* 20, 415-418.

444 [10] Olsson, B., Lautner, R., Andreasson, U., Öhrfelt, A., Portelius, E., Bjerke, M., Hölttä, M., Rosén, C.,
445 Olsson, C., and Strobel, G. (2016) CSF and blood biomarkers for the diagnosis of Alzheimer's
446 disease: a systematic review and meta-analysis, *Lancet Neurol* 15, 673-684.

447 [11] Mattsson, N., Andreasson, U., Zetterberg, H., Blennow, K., and for the Alzheimer's Disease
448 Neuroimaging, I. (2017) Association of plasma neurofilament light with neurodegeneration in
449 patients with alzheimer disease, *JAMA Neurol* 74, 557-566.

450 [12] Hye, A., Lynham, S., Thambisetty, M., Causevic, M., Campbell, J., Byers, H. L., Hooper, C., Rijdsdijk,
451 F., Tabrizi, S. J., Banner, S., Shaw, C. E., Foy, C., Poppe, M., Archer, N., Hamilton, G., Powell, J.,
452 Brown, R. G., Sham, P., Ward, M., and Lovestone, S. (2006) Proteome-based plasma
453 biomarkers for Alzheimer's disease, *Brain* 129, 3042-3050.

454 [13] Raman, C. V., and Krishnan, K. S. (1928) A new type of secondary radiation, *Nature* 121, 501-502.

455 [14] Butler, H. J., Ashton, L., Bird, B., Cinque, G., Curtis, K., Dorney, J., Esmonde-White, K., Fullwood,
456 N. J., Gardner, B., Martin-Hirsch, P. L., Walsh, M. J., McAinsh, M. R., Stone, N., and Martin, F.
457 L. (2016) Using Raman spectroscopy to characterize biological materials, *Nat Protoc* 11, 664-
458 687.

459 [15] Sahu, A., Nandakumar, N., Sawant, S., and Krishna, C. M. (2015) Recurrence prediction in oral
460 cancers: a serum Raman spectroscopy study, *Analyst* 140, 2294-2301.

461 [16] Li, X., Yang, T., and Li, S. (2012) Discrimination of serum Raman spectroscopy between normal
462 and colorectal cancer using selected parameters and regression-discriminant analysis, *Phys
463 Chem Chem Phys* 51, 5038-5043.

464 [17] Garrett, N. L., Sekine, R., Dixon, M. W., Tilley, L., Bamberg, K. R., and Wood, B. R. (2015) Bio-
465 sensing with butterfly wings: naturally occurring nano-structures for SERS-based malaria
466 parasite detection, *Phys Chem Chem Phys* 17, 21164-21168.

467 [18] Foster, N. L., Heidebrink, J. L., Clark, C. M., Jagust, W. J., Arnold, S. E., Barbas, N. R., DeCarli, C. S.,
468 Scott Turner, R., Koeppe, R. A., Higdon, R., and Minoshima, S. (2007) FDG-PET improves
469 accuracy in distinguishing frontotemporal dementia and Alzheimer's disease, *Brain* 130, 2616-
470 2635.

471 [19] Beach, T. G. (2017) A Review of Biomarkers for Neurodegenerative Disease: Will They Swing Us
472 Across the Valley?, *Neurol Ther* 6, 5-13.

473 [20] Beach, T. G., Schneider, J. A., Sue, L. I., Serrano, G., Dugger, B. N., Monsell, S. E., and Kukull, W.
474 (2014) Theoretical impact of Florbetapir (18F) amyloid imaging on diagnosis of alzheimer
475 dementia and detection of preclinical cortical amyloid, *J Neuropathol Exp Neurol* 73, 948-953.

476 [21] Skogseth, R. E., Hortobágyi, T., Soennesyn, H., Chwiszczuk, L., Rongve, A., Ballard, C., and Aarsland,
477 D. (2017) Accuracy of Clinical Diagnosis of Dementia with Lewy Bodies versus Neuropathology,
478 *J Alzheimers Dis*, 1-14.

479 [22] Johnson, K. A., Fox, N. C., Sperling, R. A., and Klunk, W. E. (2012) Brain Imaging in Alzheimer
480 Disease, *Cold Spring Harb Perspect Med* 2, a006213.

481 [23] Frisoni, G. B., Boccardi, M., Barkhof, F., Blennow, K., Cappa, S., Chiotis, K., Démonet, J.-F.,
482 Garibotto, V., Giannakopoulos, P., and Gietl, A. (2017) Strategic roadmap for an early diagnosis
483 of Alzheimer's disease based on biomarkers, *Lancet Neurol* 16, 661-676.

484 [24] Zetterberg, H., Wilson, D., Andreasson, U., Minthon, L., Blennow, K., Randall, J., and Hansson, O.
485 (2013) Plasma tau levels in Alzheimer's disease, *Alzheimers Res Ther* 5, 9.

486 [25] Tatebe, H., Kasai, T., Ohmichi, T., Kishi, Y., Kakeya, T., Waragai, M., Kondo, M., Allsop, D., and
487 Tokuda, T. (2017) Quantification of plasma phosphorylated tau to use as a biomarker for brain
488 Alzheimer pathology: pilot case-control studies including patients with Alzheimer's disease
489 and down syndrome, *Mol Neurodegener* 12, 63.

490 [26] Humpel, C. (2011) Identifying and validating biomarkers for Alzheimer's disease, *Trends
491 Biotechnol* 29, 26-32.

- 492 [27] Escudero, J., Ifeachor, E., Zajicek, J. P., Green, C., Shearer, J., and Pearson, S. (2013) Machine
493 learning-based method for personalized and cost-effective detection of Alzheimer's disease,
494 *IEEE Trans Biomed Eng* 60, 164-168.
- 495 [28] Paraskevaidi, M., Morais, C. L., Lima, K. M., Snowden, J. S., Saxon, J. A., Richardson, A. M., Jones,
496 M., Mann, D. M., Allsop, D., and Martin-Hirsch, P. L. (2017) Differential diagnosis of
497 Alzheimer's disease using spectrochemical analysis of blood, *Proc Natl Acad Sci USA*,
498 201701517.
- 499 [29] Griffin, J. W., and Bradshaw, P. C. (2017) Amino Acid Catabolism in Alzheimer's Disease Brain:
500 Friend or Foe?, *Oxid Med Cell Longev* 2017.
- 501 [30] Xu, J., Begley, P., Church, S. J., Patassini, S., Hollywood, K. A., Jullig, M., Curtis, M. A., Waldvogel,
502 H. J., Faull, R. L., Unwin, R. D., and Cooper, G. J. (2016) Graded perturbations of metabolism in
503 multiple regions of human brain in Alzheimer's disease: Snapshot of a pervasive metabolic
504 disorder, *Biochim Biophys Acta* 1862, 1084-1092.
- 505 [31] Nilsen, L. H., Witter, M. P., and Sonnewald, U. (2014) Neuronal and astrocytic metabolism in a
506 transgenic rat model of Alzheimer's disease, *J Cereb Blood Flow Metab* 34, 906-914.
- 507 [32] Wissmann, P., Geisler, S., Leblhuber, F., and Fuchs, D. (2013) Immune activation in patients with
508 Alzheimer's disease is associated with high serum phenylalanine concentrations, *J Neurol Sci*
509 329, 29-33.
- 510 [33] Gonzalez-Dominguez, R., Garcia-Barrera, T., and Gomez-Ariza, J. L. (2015) Metabolite profiling for
511 the identification of altered metabolic pathways in Alzheimer's disease, *J Pharm Biomed Anal*
512 107, 75-81.
- 513 [34] Trushina, E., Dutta, T., Persson, X.-M. T., Mielke, M. M., and Petersen, R. C. (2013) Identification
514 of altered metabolic pathways in plasma and CSF in mild cognitive impairment and
515 Alzheimer's disease using metabolomics, *PLoS one* 8, e63644.
- 516 [35] Luan, H., Liu, L. F., Meng, N., Tang, Z., Chua, K. K., Chen, L. L., Song, J. X., Mok, V. C., Xie, L. X., Li,
517 M., and Cai, Z. (2015) LC-MS-based urinary metabolite signatures in idiopathic Parkinson's
518 disease, *J Proteome Res* 14, 467-478.
- 519 [36] Luan, H., Liu, L.-F., Tang, Z., Zhang, M., Chua, K.-K., Song, J.-X., Mok, V. C., Li, M., and Cai, Z. (2015)
520 Comprehensive urinary metabolomic profiling and identification of potential noninvasive
521 marker for idiopathic Parkinson's disease, *Sci Rep* 5.
- 522 [37] Havelund, J. F., Heegaard, N. H., Færgeman, N. J., and Gramsbergen, J. B. (2017) Biomarker
523 Research in Parkinson's Disease Using Metabolite Profiling, *Metabolites* 7, 42.
- 524 [38] Lahiry, P., Torkamani, A., Schork, N. J., and Hegele, R. A. (2010) Kinase mutations in human
525 disease: interpreting genotype-phenotype relationships, *Nat Rev Genet* 11, 60.
- 526 [39] Pei, J.-J., Grundke-Iqbal, I., Iqbal, K., Bogdanovic, N., Winblad, B., and Cowburn, R. F. (1998)
527 Accumulation of cyclin-dependent kinase 5 (cdk5) in neurons with early stages of Alzheimer's
528 disease neurofibrillary degeneration, *Brain Res* 797, 267-277.
- 529 [40] Hoozemans, J. J., van Haastert, E. S., Nijholt, D. A., Rozemuller, A. J., Eikelenboom, P., and Scheper,
530 W. (2009) The unfolded protein response is activated in pretangle neurons in Alzheimer's
531 disease hippocampus, *Am J Pathol* 174, 1241-1251.
- 532 [41] McKeith, I. (2004) Dementia with Lewy bodies, *Dialogues Clin Neurosci* 6, 333-341.
- 533 [42] Carmona, P., Molina, M., Calero, M., Bermejo-Pareja, F., Martinez-Martin, P., and Toledano, A.
534 (2013) Discrimination analysis of blood plasma associated with Alzheimer's disease using
535 vibrational spectroscopy, *J Alzheimers Dis* 34, 911-920.
- 536 [43] Ryzhikova, E., Kazakov, O., Halamkova, L., Celmins, D., Malone, P., Molho, E., Zimmerman, E. A.,
537 and Lednev, I. K. (2015) Raman spectroscopy of blood serum for Alzheimer's disease
538 diagnostics: specificity relative to other types of dementia, *J Biophotonics* 8, 584-596.
- 539 [44] Parikh, K. S., and Shah, T. P. (2016) Support vector machine—a large margin classifier to diagnose
540 skin illnesses, *Procedia Technology* 23, 369-375.

- 541 [45] Davidson, Y., Gibbons, L., Purandare, N., Byrne, J., Hardicre, J., Wren, J., Payton, A., Pendleton, N.,
542 Horan, M., Burns, A., and Mann, D. M. A. (2006) Apolipoprotein E ϵ 4 Allele Frequency in
543 Vascular Dementia, *Dement Geriatr Cogn Dis* 22, 15-19.
- 544 [46] Movasaghi, Z., Rehman, S., and Rehman, I. U. (2007) Raman spectroscopy of biological tissues,
545 *Appl Spectrosc Rev* 42, 493-541.
- 546 [47] Pichardo-Molina, J. L., Frausto-Reyes, C., Barbosa-García, O., Huerta-Franco, R., González-Trujillo,
547 J. L., Ramírez-Alvarado, C. A., Gutiérrez-Juárez, G., and Medina-Gutiérrez, C. (2007) Raman
548 spectroscopy and multivariate analysis of serum samples from breast cancer patients, *Lasers*
549 *Med Sci* 22, 229-236.
- 550 [48] Trevisan, J., Angelov, P. P., Carmichael, P. L., Scott, A. D., and Martin, F. L. (2012) Extracting
551 biological information with computational analysis of Fourier-transform infrared (FTIR)
552 biospectroscopy datasets: current practices to future perspectives, *Analyst* 137, 3202-3215.

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

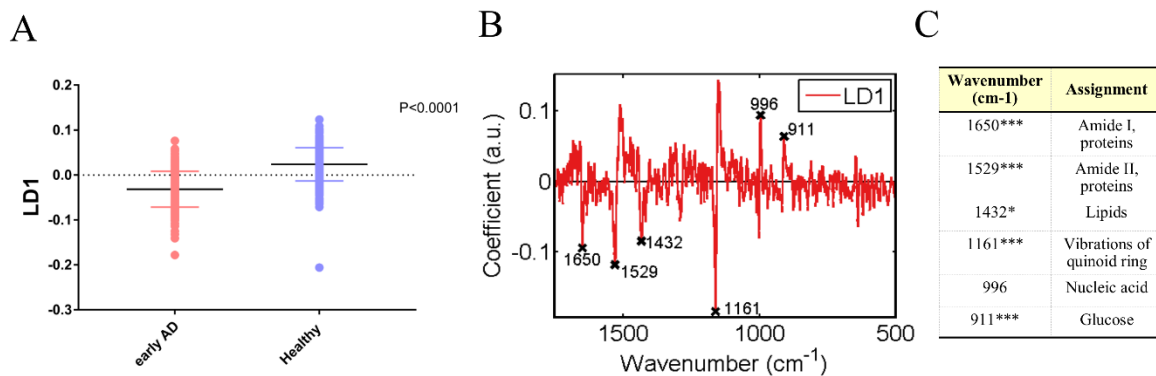
571

572

573

574

575



578

579 **Figure 1. Early stage Alzheimer's disease (AD) versus healthy individuals.** One-
 580 dimensional (1D) scores plot after cross-validated PCA-LDA ($P < 0.0001$, 95% CI = 0.0503 to
 581 0.0622) (A); loadings plot showing the top six discriminatory peaks between the two classes
 582 (B); important peaks along with their tentative assignments^{45,46} (C). Data are expressed as the
 583 mean \pm standard deviation (SD). A P-value of 0.05 or less was considered significant; $P < 0.05$
 584 (*), $P < 0.005$ (**), or $P < 0.0005$ (***)

585

586

587

588

589

590

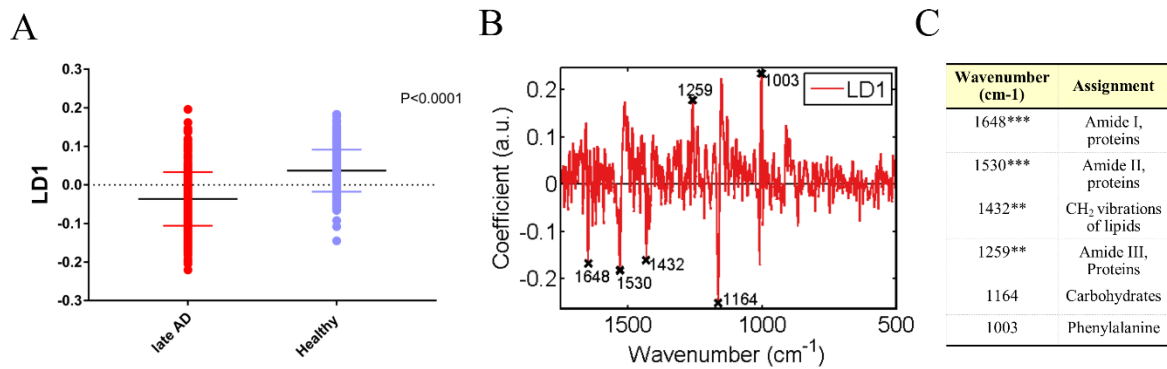
591

592

593

594

595



596

597 **Figure 2. Late stage Alzheimer's disease (AD) versus healthy individuals.** One-dimensional
 598 (1D) scores plot after cross-validated PCA-LDA ($P < 0.0001$, 95% CI = 0.0655 to 0.0834) (A);
 599 loadings plot showing the top six discriminatory peaks between the two classes (B); important
 600 peaks along with their tentative assignments^{45,46} (C). Data are expressed as the mean \pm standard
 601 deviation (SD). A P-value of 0.05 or less was considered significant; $P < 0.05$ (*) or $P < 0.005$
 602 (**) or $P < 0.0005$ (***)).

603

604

605

606

607

608

609

610

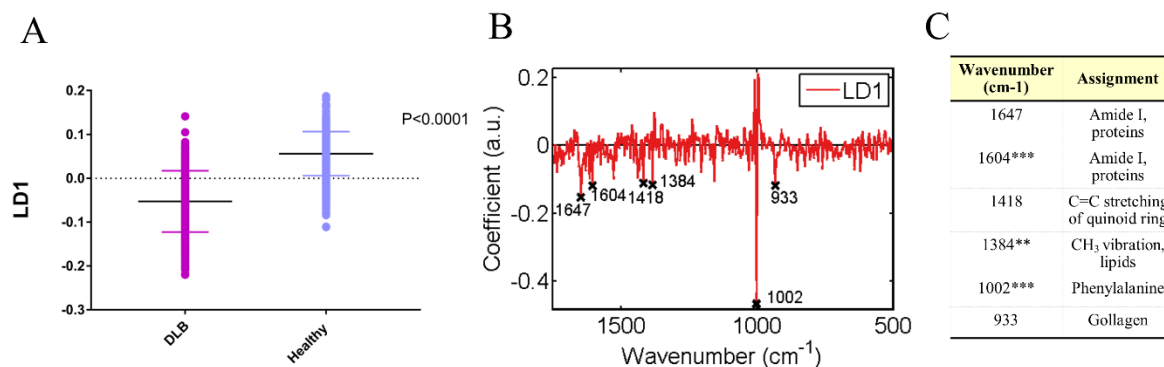
611

612

613

614

615



616

617 **Figure 3. Dementia with Lewy bodies (DLB) versus healthy individuals.** One-dimensional
 618 (1D) scores plot after cross-validated PCA-LDA ($P < 0.0001$, 95% CI = 0.0982 to 0.1166) (A);
 619 loadings plot showing the top six discriminatory peaks (B); important peaks along with their
 620 tentative assignments^{45,46} (C). Data are expressed as the mean \pm standard deviation (SD). A P-
 621 value of 0.05 or less was considered significant; $P < 0.05$ (*) or $P < 0.005$ (**) or $P < 0.0005$
 622 (***).

623

624

625

626

627

628

629

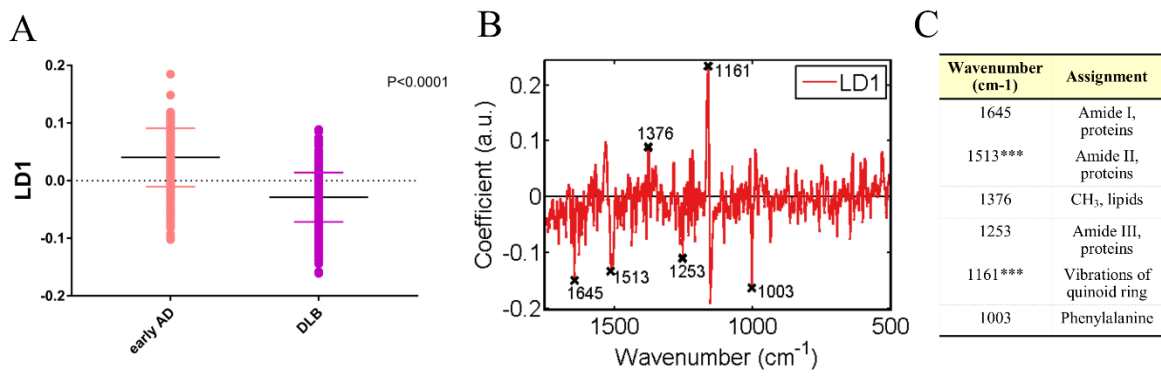
630

631

632

633

634



635

636 **Figure 4. Early stage Alzheimer's disease (AD) versus dementia with Lewy bodies (DLB).**

637 One-dimensional (1D) scores plot after cross-validated PCA-LDA ($P < 0.0001$, 95% CI = -

638 0.0791 to -0.0649) (A); loadings plot showing the top six discriminatory peaks (early AD was

639 used as reference class) between the two classes (B); important peaks along with their tentative

640 assignments^{45,46} (C). Data are expressed as the mean \pm standard deviation (SD). A P-value of

641 0.05 or less was considered significant; $P < 0.05$ (*) or $P < 0.005$ (**) or $P < 0.0005$ (***)

642

643

644

645

646

647

648

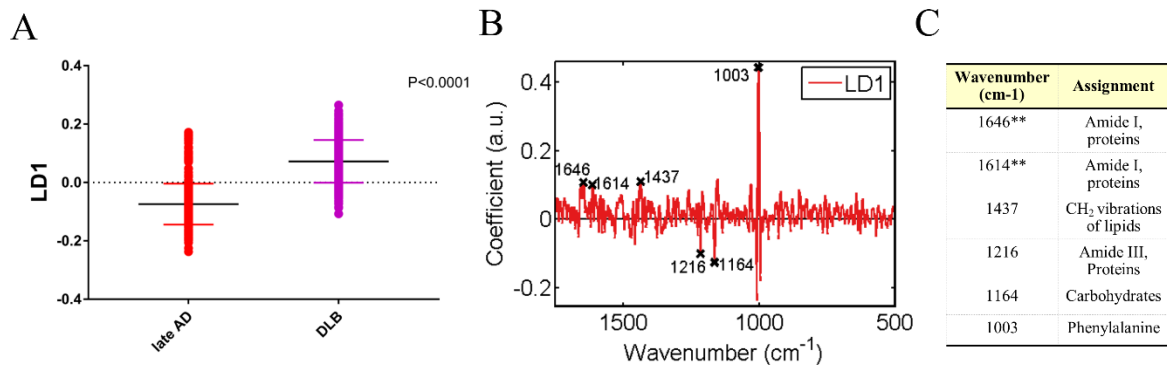
649

650

651

652

653



654

655 **Figure 5. Late stage Alzheimer's disease (AD) versus dementia with Lewy bodies (DLB).**
 656 One-dimensional (1D) scores plot after cross-validated PCA-LDA ($P < 0.0001$, 95% CI = 0.138
 657 to 0.1596) (A); loadings plot showing the top six discriminatory peaks (late AD was used as
 658 reference class) (B); important peaks along with their tentative assignments^{45,46} (C). Data are
 659 expressed as the mean \pm standard deviation (SD). A P-value of 0.05 or less was considered
 660 significant; $P < 0.05$ (*) or $P < 0.005$ (**) or $P < 0.0005$ (***)).

661

662

663

664

665

666

667

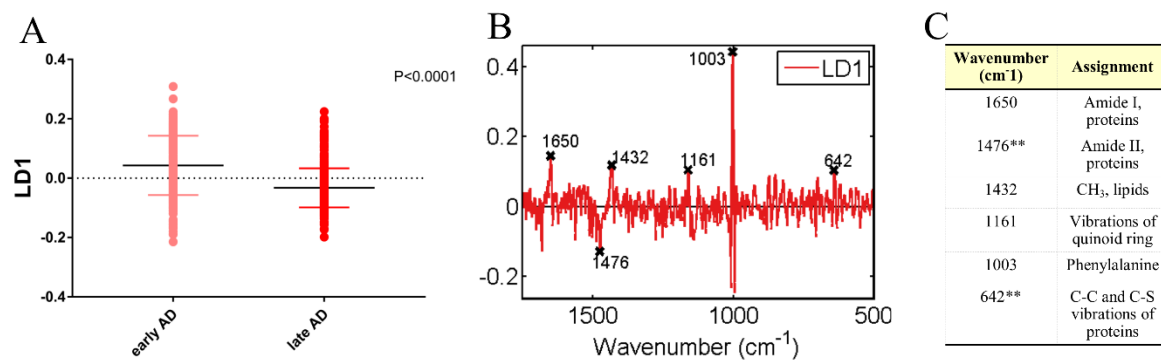
668

669

670

671

672



673

674 **Figure 6. Early stage Alzheimer's disease (AD) versus late AD.** One-dimensional (1D)
 675 scores plot after cross-validated PCA-LDA ($P < 0.0001$, 95% CI = -0.0943 to -0.0624) (A);
 676 loadings plot showing the top six discriminatory peaks (early AD was used as reference class)
 677 (B); important peaks along with their tentative assignments^{45,46}. (C). Data are expressed as the
 678 mean \pm standard deviation (SD). A P-value of 0.05 or less was considered significant; $P < 0.05$
 679 (*), $P < 0.005$ (**), or $P < 0.0005$ (***)).

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699 **Table 1. Patient characteristics.**

	Early stage AD	Late stage AD	DLB	Healthy^a
Sample size, n	11	15	15	15
Age, years, mean (SD, range)	62 (10, 50-74)	64 (8, 50-79)	71 (6, 61-80)	54 (18, 23-73)
<i>APOE4</i> carriers, n (%)	6 (55)	11 (73)	6 (40)	6 (40)
Female, n (%)	5 (45)	3 (20)	3 (20)	9 (60)
Duration, years, mean (\pm SD)	1.28 (\pm 0.5)	4.56 (\pm 3)	2.46 (\pm 1)	n/a

700

701 AD: Alzheimer’s disease DLB: dementia with Lewy bodies; *APOE4*: apolipoprotein E4; n/a: not
 702 applicable

703 ^a Two individuals from the ‘Healthy’ group had no information on *APOE4* load (13%) and gender
 704 (13%).

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721 **Table 2. Sensitivity, specificity and G-score and Youden’s index for the different comparison**
 722 **groups after classification with support vector machine (SVM).**

Comparison Group	Sensitivity (%)	Specificity (%)	G-Score (%)	Youden’s index (%)
Early AD vs Healthy	84	86	85	70
Late AD vs Healthy	84	77	80	61
DLB vs Healthy	83	87	85	70
Early AD vs DLB	81	88	84	69
Late AD vs DLB	90	93	91	84
Early AD vs Late AD	66	83	74	49

723 AD: Alzheimer’s disease; DLB: dementia with Lewy bodies