

1 **Stratification of the Gut Microbiota Composition Landscape Across the**
2 **Alzheimer's Disease Continuum in a Turkish Cohort**

3
4 **Süleyman Yıldırım*#^{1,2}, Özkan Ufuk Nalbantoglu*^{3,4}, Abdulahad Bayraktar*⁵, Fatma Betül Ercan⁶,**
5 **Aycan Gündoğdu^{4,7}, Halil Aziz Velioğlu^{2,6}, Mehmet Fatih Göl⁷, Ayten Ekinci Soylu⁷, Fatma Koç¹⁸, Ezgi**
6 **Aslan Gürpınar⁷, Kübra Sogukkanlı Kadak⁶, Muzaffer Arıkan², Adil Mardinoğlu^{5,8}, Mehmet Koçak⁹, Emel**
7 **Köseoğlu^{#,10}, Lütfü Hanoglu^{#,11}**

8
9 1. Department of Medical Microbiology, Istanbul Medipol University International School of Medicine,
10 Istanbul, Turkey
11 2. Regenerative and Restorative Medicine Research Center (REMER); Research Institute for Health
12 Sciences and Technologies (SABITA), Istanbul Medipol University, Istanbul, Turkey.
13 3. Department of Computer Engineering, Erciyes University, Kayseri, Turkey
14 4. Genome and Stem Cell Center (GenKok), Erciyes University, Kayseri, Turkey
15 5. Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's
16 College London, London, SE1 9RT, UK
17 6. Graduate Program in Neuroscience, Istanbul Medipol University International School of Medicine,
18 Istanbul, Turkey
19 7. Department of Microbiology and Clinical Microbiology, Erciyes University School of Medicine,
20 Kayseri, Turkey
21 8. Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm, SE-17121, Sweden
22 9. Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis,
23 TN, United States of America
24 10. Department of Neurology, School of Medicine, Erciyes University, Kayseri, Turkey
25 11. Department of Neurology, School of Medicine, Istanbul Medipol University, Istanbul, Turkey.
26
27

28 **Running title:** Gut Microbiota Stratification in Alzheimer's Disease

29 **Key Words:** Gut microbiome, Alzheimer's Diseases, 16S rRNA, Stratification

31 *Equal contribution.

33 #Corresponding authors:

35 Email addresses:

36 Dr. Süleyman Yıldırım
37 suleymanyildirim@medipol.edu.tr

39 Dr. Emel Koseoglu
40 emelk@erciyes.edu.tr

42 Dr. Lutfu Hanoglu
43 lhanooglu@medipol.edu.tr

47 ⁸Current address: APC Microbiome Ireland, University College Cork, Cork T12 YT20, Ireland

50

51 ABSTRACT

52

53 Alzheimer's disease (AD) is a heterogeneous neurodegenerative disorder that spans over a
54 continuum with multiple phases including preclinical, mild cognitive impairment, and dementia.
55 Unlike most other chronic diseases there are limited number of human studies reporting on AD
56 gut microbiota in the literature. These published studies suggest that the gut microbiota of AD
57 continuum patients varies considerably throughout the disease stages, raising expectations for
58 existence of multiple microbiota community types. However, the community types of AD gut
59 microbiota were not systematically investigated before, leaving important research gap for diet-
60 based intervention studies and recently initiated precision nutrition approaches aiming at
61 stratifying patients into distinct dietary subgroups. Here, we comprehensively assessed the
62 community types of gut microbiota across the AD continuum. We analyze 16S rRNA amplicon
63 sequencing of stool samples from 27 mild cognitive patients, 47 AD, and 51 non-demented control
64 subjects using tools compatible with compositional nature of microbiota. To characterize gut
65 microbiota community types, we applied multiple machine learning techniques including
66 partitioning around the medoid clustering, fitting probabilistic Dirichlet mixture model, Latent
67 Dirichlet Allocation model, and performed topological data analysis for population scale
68 microbiome stratification based on Mapper algorithm. These four distinct techniques all converge
69 on Prevotella and Bacteroides partitioning of the gut microbiota across AD continuum while some
70 methods provided fine scale resolution in partitioning the community landscape. The Signature
71 taxa and neuropsychometric parameters together robustly classify the heterogenous groups
72 within the cohort. Our results provide a framework for precision nutrition approaches and diet-
73 based intervention studies targeting AD cohorts.

74

75

76

IMPORTANCE

77

78 The prevalence of AD worldwide is estimated to reach 131 million by 2050. Most disease
79 modifying treatments and drug trials have failed due partly to the heterogeneous and complex
80 nature of the disease. Unlike other neurodegenerative diseases gut microbiota of AD patients is
81 poorly studied. Recently initiated ambitious precision nutrition initiative or other diet-based
82 interventions can potentially be more effective if the heterogeneous disease such as AD is
83 deconstructed into multiple strata allowing for better identification of biomarkers across narrower
84 patient population for improved results. Because gut microbiota is inherently integral part of the
85 nutritional interventions there is unmet need for microbiota-informed stratification of AD clinical
86 cohorts in nutritional studies. Our study fills in this gap and draws attention to the need for
87 microbiota stratification as one of the essential steps for precision nutrition interventions. We
88 demonstrate that while Prevotella and Bacteroides clusters are the consensus partitions the newly
89 developed probabilistic methods can provide fine scale resolution in partitioning the AD gut
90 microbiome landscape.

91

92 **Key words:** Alzheimer's Disease, Gut microbiota, Machine learning, Stratification, Dirichlet,
93 Topological data analysis

94

INTRODUCTION

95

96

97 Alzheimer's Disease (AD) is the most common form of dementia worldwide and its prevalence is
98 estimated to reach 131 million by 2050 [1]. AD spans over a continuum starting with the non-
99 symptomatic pre-clinical stage and advancing through the spectrum of clinical stages. These
100 stages are dashed with distinct pathophysiological states [2], namely the amyloid-tau-
101 neuroinflammation axis. The clinical continuum entails mild memory loss and/or cognitive

102 impairments (mild cognitive impairment, MCI due to AD) and trajectories for function leading to
103 memory problems besides cognitive impairment (dementia phase); and finally complete loss of
104 independent functioning towards the end stage [3]. Moreover, The Alzheimer's dementia phase
105 is further broken down into the stages of mild, moderate and severe, thereby making AD a
106 complex and highly heterogenous disease.

107

108 Traditionally, pathogenesis of AD is attributed to extracellular aggregation of amyloid- β -peptides
109 (A β) in senile plaques and intracellular depositions of hyperphosphorylated tau that forms
110 neurofibrillary tangles [4]. Although numerous clinical trials based on the amyloid postulates have
111 been attempted virtually all of them have failed [5]. The unsettlingly consistent failure of clinical
112 trials targeting single target amyloid pathways prompted researchers to refine the amyloid
113 hypothesis [6] and even extend it to periphery [7]. Recently, a group of AD researchers asserted
114 that infectious agents reach and remain dormant in the central nervous system (CNS) and
115 undergo reactivation during aging, sparking cascades of inflammation, induce A β , and ultimately
116 neuronal degeneration [8]. Chronic inflammation in CNS mediated by microglial toxicity as well
117 as systemic inflammation in the periphery is widely recognized in AD and linked to amyloid
118 cascade hypothesis in animal experiments [9, 10]. None of the drugs available today for
119 Alzheimer's dementia slow or stop the damage and destruction of neurons [11]. Intervention at
120 different points along the Alzheimer's continuum should therefore be multimodal and involve
121 targeting neuropathology in brain, systemic inflammation in the body, and metabolic processes in
122 the periphery that escalate the disease in brain [12]. Non-pharmacologic, targeted, personalized,
123 and multimodal disease modifying interventions in AD, including diet and lifestyle changes to
124 optimize metabolic parameters has recently been under investigation [13-16].

125

126 A growing body of evidence suggest that human gut microbiota is strongly associated with human
127 metabolic processes in all organs including brain [17] and implicated in neuroinflammation via

128 brain-gut axis [18]. Gut microbes across animal models influence CNS by modulation of
129 neuroimmune function, sensory neuronal signaling, and metabolic activity [19]. Several studies
130 using transgenic animal model of AD reported gut microbiota alterations (see [19]) but these
131 animal models poorly mirror human AD. Unexpectedly, only a few human clinical studies on AD
132 were reported in the literature [20-28]. Of these studies, gut microbiota associated metabolites
133 such elevated Trimethylamine N-oxide (TMAO) in CSF [26] and altered bile acids profile [28] were
134 directly implicated in AD dementia. Importantly, dietary pattern of AD patients is at the center of
135 the precision medicine approaches [29]. Also, diet is one of the most important factors modulating
136 gut microbiota-based active metabolites. Disease modifying approaches involving diet should
137 therefore consider microbiota in AD. Indeed, a recent study [23] tested the impact of a modified
138 Mediterranean ketogenic diet on gut microbiome composition and demonstrated that the diet can
139 modulate the gut microbiome and metabolites in association with improved AD biomarkers in
140 CSF. These published studies, however, did not comprehensively investigate AD microbiota
141 subclusters across the disease continuum, leaving important gap in our understanding of human
142 microbiota in a highly heterogenous disease. Recently initiated ambitious precision nutrition
143 approaches [30-33] cannot be applied on a highly heterogenous disease before deconstructing
144 the disease into multiple strata and tailoring therapies accordingly.

145

146 In the present study, we postulated that gut microbiota dysbiosis along the AD continuum should
147 reflect an overlapping yet distinct community types. We show that AD gut microbiota includes
148 distinct community types and the cognitive impairments in AD continuum is associated with
149 unique gut microbiota signatures. Elucidating the diversity and community types of gut microbiota
150 would facilitate identification of stratification biomarkers thereby contributing to precision nutrition
151 approaches in AD.

152

153

154

RESULTS

155

156 **Study Design and Participant Characteristics.**

157

158 The study cohort consisted of 47 AD, 27 MCI (all amnestic), and 51 subjects non-demented
159 controls (N=125). To minimize dietary confounding effect on the microbiome, we included healthy
160 co-habiting spouses of the patients sharing the same diet as controls. The control group therefore
161 largely (n=27) comprised partners of the patients. Participants were recruited in two health centers
162 located in different cities. The cohort groups were statistically not different in terms of sex, but age
163 and education factors were significantly different (Table 1), therefore statistically adjusted in
164 analyses. Expectedly, the groups were also different in cognitive tests including the Mini-Mental
165 State Exam (MMSE), and the Clinical Dementia rating (CDR). Most AD participants had very mild
166 or mild dementia, with clinical dementia rating (CDR) scores ranging from 0.5–3 (median CDR 1
167 for AD; 0.5 for MCI and 0 for the control group). The median MMSE scores were significantly
168 higher in control (MMSE=27) and MCI (MMSE=26) groups than AD (MMSE=16). A subset of AD
169 patients (n=12) was clinically asked to undergo lumbar puncture to ascertain diagnosis using CSF
170 biomarkers including A β 42/A β 40 ratio, phosphorylated tau (p-tau), and the p-tau/A β 42 ratio
171 (Supplementary Table S1). We collected medication information from the patient's registry.

172

173 **Microbiome composition is associated with disease status along the AD continuum**

174

175 The gut microbiota was profiled using the V3-V4 hypervariable region of the 16S rRNA gene; The
176 Nephele automatic pipeline denoised the paired-end sequences and assigned amplicon
177 sequence variants (ASVs) according to DADA2 [34]. The Nephele produced both unrarefied and
178 the rarefied ASV tables. The rarefied table included a total of 3486 ASVs in the table (10769
179 sequences/sample) for downstream analyses.

180 The phylum level taxonomic analysis showed typical human gut microbiota profile in terms of
181 over-abundance of *Firmicutes*, *Bacteroidetes*, and *Proteobacteria* (Figure 1a). Together with
182 *Verrucomicrobia*, and *Actinobacteria* the five phyla comprised 99% of all reads but *Proteobacteria*
183 was overrepresented in AD patient samples. Notably, the genus level relative abundance
184 distributions across samples showed *Prevotella_9* and *Bacteroides* were the most abundant of
185 top30 genera across the samples (Figure 1b). To perform differential abundance analysis
186 between samples we sought concordance analysis among multiple tools. ANCOM-BC or
187 ALDEx2, when used covariates in their models, both agreed that only *Ruminococcus_unclassified*
188 is significantly differentially abundant among the groups (data not shown). Nevertheless, when
189 we employed limma-voom R package (age and sex adjusted, FDR<0.05) we found that
190 *Prevotella_9*, *Bacteroides* and members of *Ruminococcaceae* family were among the top most
191 significant differentially abundant taxa (ASV) between the cohort groups (Supplementary Tables
192 S2-5). A comprehensive comparative statistical assessment of multivariate and compositional
193 methods [35] demonstrated ALDEx2 or alike tools suffer from low power while limma-voom and
194 songbird in their own class were the best performers.

195
196 Alpha diversity indices (Shannon, Inverse Simpson) did not show significant differences after
197 multiple testing corrections (Kruskal-Wallis, Supplementary Figure S1 (a-d), FDR>0.05) but
198 richness index, Chao1, showed significant difference between MCI and the control group
199 (pairwise Wilcoxon rank sum test, p=0.008074).

200
201 We employed both relative abundances based and recently developed compositionally aware
202 tools, namely DEICODE [36] and Songbird [37] to compare the composition and structure of
203 bacterial communities in samples using multiple beta diversity indices (Bray-Curtis, Jaccard, and
204 Aitchison). The principal coordinates analysis showed separation of the three groups by both
205 Bray-Curtis and Jaccard indices (Figure 2a-b). We used adonis2 function in qiime2 plugin (q2-

206 diversity) to perform PERMONAVA analysis with 999 permutations and included interaction terms
207 (Supplementary Table S6) and separation of the groups were highly significant ($P=0.0001$). Age
208 and Sex also significantly contributed to the total variance ($P<0.001$) but the interaction terms
209 were not significant. Furthermore, dispersion between groups tests (PERMDISP) indicated only
210 the dispersion MCI group is significantly heterogeneous (pairwise comparisons $p=0.033$ for AD-
211 MCI; $p=0.024$ for C-MCI; $p=0.672$ for AD-C), which may be attributed to unbalanced design. We
212 added further support for the separation of the three groups from other ordinations. The Canonical
213 Analysis of Principal Coordinates (CAP) analysis unambiguously showed the three groups are
214 distinct (Figure 2c, trace statistic = 0.86855, $p=0.001$, 999 permutations). The final support in beta
215 diversity was provided by the DEICODE analysis (robust Aitchison PCA) (Figure 2d,
216 PERMANOVA $p=0.02$), which indicated that the three groups are distinct, and the community
217 clusters are largely driven by a subset of ASVs with taxonomic assignment *Prevotella_9*,
218 *Bacteroides*, a unclassified genus within *Ruminococcaceae* family
219 (*Ruminococcaceae_unclassified*), and *Escherichia/Shigella*. Moreover, the co-occurrence
220 analysis using SparCC showed that *Prevotella_9* and *Bacteroides* were negatively correlated
221 (Correlation=-0.4445, FDR =0.09355). Moreover, the genus level PCoAs showed partially
222 overlapping clusters of these two taxa while the groups overall were also significantly separated
223 (PERMANOVA, $p <0.0001$, Supplementary Figures S2 (a-c)). We therefore placed particular
224 attention to these two taxa in the rest of the downstream analyses.

225
226 Enrichment analysis by multinomial regression embedded in the songbird tool with regard to
227 covariates (formula: Age+Sex+Edu+MMSE+CDR+Groups(levels=("C", "MCI","AD")) indicated
228 that the natural log ratio of *Prevotella_9* to *Bacteroides* and *Prevotella_9* to *Escherichia/Shigella*
229 significantly separated AD group from the control group (Welch's t-test, FDR adjusted $p=0.04$)
230 but not from the MCI group (Figure 3a-d). Importantly, the songbird excluded 25 samples from
231 this analysis due to zero-rich abundances that do not allow for center-log ratio calculations. We

232 therefore tested the natural log ratio of top 25% allowing to include all samples in the analysis
233 (“Set1” in Supplementary Table S7) to the bottom 25% (“Set2”, Supplementary Table S8) of the
234 ranked ASVs associated with the AD relative to the control group; also, same ratios for MCI
235 relative to the control group (“Set3” and “Set4”, Supplementary Table S8) and the ASVs
236 enriched in each group were visualized with Qurro [38]. Both sets of ranked log ratios revealed
237 significant differences (Graph Pad Prism) between the log ratios of features differentiating groups
238 (Welch’s t-test, FDR adjusted p= 0.0002).

239

240 **Discrete multiple subsets of gut microbiota exist along the AD continuum**

241

242 Considering the preceding results, we postulated that gut microbiota profile along the AD
243 continuum does not represent a single state, rather, distinct yet overlapping community types. We
244 addressed this hypothesis using four unique methods: 1- Partitioning around medoid (PAM)-
245 based clustering [39], 2- Fitting Dirichlet multinomial mixture (DMM) models to partition microbial
246 community profiles into a finite number of clusters [40] using the Laplace approximation, 3- Fitting
247 Latent Dirichlet Allocation (LDA) [41, 42] using perplexity measure, and 4- Analyzing topological
248 futures of data density [43] based on the *Mapper* algorithm to capture subtle and non-linear
249 patterns of high-dimensional datasets and population level stratification.

250

251 The PAM-based clustering identified three (k=3) distinct clusters based on Gap statistics
252 (Supplementary Figure S3a). PCoA analysis of the sample abundances in the three clusters
253 indicated significant separation of the clusters (Figure 4a, PERMANOVA, p=0.001) . We
254 confirmed optimum number of clusters using both Jensen-Shannon and Bray-Curtis distance
255 metrics (data not shown). The relative abundance of the genus *Prevotella_9* dominated cluster-
256 1 while the genus *Bacteroides* showed the highest relative abundances in the other two clusters
257 (Figure 4b).

258

259 Next, we employed the Dirichlet multinomial mixtures probabilistic community modeling using the
260 *DirichletMultinomial* R package [40] and fitting genus level absolute abundances. Based on
261 Laplace approximation three clusters (cluster 1, 2, and 3) represented the best model fit
262 (Supplementary Figure 3b), which was congruent with the PAM-based clustering. The PCoA
263 analysis of these clusters and PERMANOVA pairwise tests further supported existence of three
264 distinct clusters within the microbial community (Figure 4c, PERMANOVA, $p=0.01$). The genus
265 *Bacteroides* was the most abundant taxa in the first two clusters and the third cluster was
266 dominated by *Prevotella_9* (Figure 4d). Notably, cluster2 included significantly higher abundance
267 of *Bacteroides* (26.3%) than cluster1 (9.9%) and cluster3 (4.7%). In addition to highly enriched
268 *Bacteroides* in cluster2 the decreasing trend of *Faecalibacterium* abundance and elevated
269 abundance of inflammation associated *Escherchia/Shigella* suggested that cluster2 can be
270 named “*Bacteroides2 (Bact2) enterotype*” as recently described [44, 45]. Reportedly, abundance
271 of *Bacteroides* in Bact2 enterotype can reach as high as 78% in patients with inflammatory bowel
272 disease and is associated with systemic inflammation. These results suggest that cluster2
273 includes patients with aggravated systemic inflammation.

274

275 We also performed SIMPER analysis based on Bray-Curtis distance to identify taxa contributing
276 most to dissimilarities between clusters (data not shown). *Bacteroides*, *Prevotella_9*,
277 *Faecalibacterium*, and taxa within *Ruminococcaceae* family ranked among the top ten taxa
278 contributing most to differences between the three DMM clusters. To examine which factors were
279 associated with the DMM clusters we analyzed distribution of clinical metadata and diversity
280 metrics within the clusters. Alpha diversity indices (Chao1, Shannon, and Inverse Simpson) were
281 statistically different between all three clusters after Benjamini-Hochberg FDR adjustment.
282 However, CDR, MMSE, Age, Sex, and Education were not significant between the clusters

283 (Kruskal Wallis test followed by Dunn's posthoc test, FDR<0.05 and Fisher's Exact test was used
284 for Sex parameter). (Figure 5 a-h)

285
286 We next tested LDA potential to stratify gut microbiota of the cohort participants. This
287 unsupervised machine learning technique is increasingly finding acceptance in the field of
288 microbiome [46-48] for its unique ability to reveal latent or hidden groups within the data cloud.
289 Supplementary Figure S4 shows LDA model's perplexity parameter and log-likelihood values to
290 find optimal number of clusters. Both parameters continued to partition the community without
291 reaching a clear optimum. This finding is unexpectedly consistent with recent publications using
292 LDA in microbial ecology [46-48]. Bacteria probability distributions (ranked by probability $\geq 1\%$ in
293 descending order) across the subgroups are displayed in Figure 6a. Interestingly, of the ten
294 subgroups two subgroups were dominated by *Bacteroides* (topic1 and topic5) and a subgroup
295 (topic2) dominated by *Prevotella_9* with 97% probability. These subgroups therefore resemble
296 subgroups detected by PAM and DMM in terms of prevalence of *Bacteroides* and *Prevotella_9*.
297 Unlike DMM and PAM, however, LDA detected a distinct subgroup (topic10) with top ranking
298 genus was *Escherichia/Shigella*, which also included putatively opportunistic bacteria such as
299 *Enterococcus* and *Klebsiella*. Subgroups 4, 6, and 9 were conspicuous with the genera known to
300 produce butyrate and acetate or is mucophilic. Even though we present first ten subgroups
301 (topics) here we also examined higher order subgroups and observe that the ten subgroups are
302 further partitioned into additional subgroups such as subgroups with top ranking probability of
303 *Lactobacillus* and *Akkermansia* emerge. Finally, we plotted Quetelet index by subgroups to infer
304 associations between subgroups and the cohort groups (Figure 6b). Quetelet index estimates the
305 relative change of the occurrence frequency of a latent subgroup among all the samples compared
306 to that among the samples of the cohort groups. The index showed subgroups 1, 8, 9, 10 are
307 positively associated with AD group. The subgroup 9 is enriched by the members of
308 *Ruminococcaceae* family. The top ranking *Ruminococcaceae_UCG_002* and *Akkermansia* are

309 more abundant in AD group than the control group according to limma-voom analysis.
310 *Akkermansia* overabundance in AD gut microbiota is counterintuitive but was previously reported
311 by others [25] and this genus is more abundant in the gut microbiota of Parkinson's patients, also
312 [49]. The subgroup 10, where *Escherichia/Shigella* is the top ranking genus, is strongly associated
313 with AD group but negatively associated with other groups. Conversely, subgroups 2,4, and 7,
314 which are enriched by short chain fatty acid producers, are positively associated with the control
315 and MCI groups but negatively associated with AD.

316
317 Another and last method we employed to stratify gut microbiota was topological data analysis
318 (TDA), based on the *Mapper* algorithm [50] embedded in recently developed *tmap* tool [43]. The
319 *tmap* tool was developed for network representation for stratification and association study of
320 high-dimensional microbiome data. After constructing TDA microbiome network using Mapper
321 algorithm (ordination, covering, and DBSCAN clustering) the workflow in the second step includes
322 computation of a modified version of the spatial analysis of functional enrichment (SAFE) scores
323 to map both the metadata and microbiome features into the TDA network to generate their vectors
324 of SAFE scores. Vectors of SAFE scores are then used to perform ranking and ordination, and
325 co-enrichment relations to delineate relationship between metadata and microbiome features. To
326 construct TDA network we first applied dimension reduction (filtering) in PCoA using Bray-Curtis
327 distance, followed the above algorithm and also repeated the entire analysis using Jensen-
328 Shannon distance to reveal effect of distance metric, if any. To understand how driver taxa relate
329 to each other and with the clinical metadata we performed Principal Component Analysis (PCA)
330 of SAFE scores. Figure (7a) shows the TDA network and PCA (Bray-Curtis distance) of tax-
331 metadata based on SAFE scores (Supplementary Table S9), respectively. We obtained similar
332 TDA network profile using Jensen-Shannon distance (Figures 7b) and SAFE scores
333 (Supplementary Table S10). Size of each marker is scaled according to the SAFE score and only
334 top30 bacteria species are shown in PCA figures for clarity. A node in the network represents a

335 group of samples sharing similar bacteria genus profiles. Two given nodes are linked when
336 common samples are shared between the two nodes. The TDA analysis using both distance
337 indices resulted in very similar stratification profile with the top ten SAFE scoring genera included
338 *Prevotella_9*, *Bacteroides*, *Ruminococaceae_unclassified*, *species of Lachnospiraceae*, and
339 GCA90006675. Unsurprisingly, a few taxa ranking differed between the two profiles such as
340 *Caprococcus_2*, *Mollicutes_RF39_unclassified*.

341
342 Furthermore, Figures (8a and 8b) show taxa and host covariates based on Bray-Curtis and
343 Jensen-Shannon distances, respectively. Regardless of the distance metric, all three groups were
344 clearly separated. The drivers of microbiome stratification (*Prevotella_9*, *Bacteroides*,
345 *Ruminococcus_unclassified*) are placed near the control, AD and MCI groups, respectively in both
346 PCA figures. Of the clinical metadata, MMSE, sex, and education were grouped with the control
347 group and co-enriched with *Prevotella_9* but also with *Prevotella_2*, and *Haemophilus*, and
348 *Lachnospiraceae_NK4B4_group*. Conversely, CDR, age, and AD group were clustered together
349 and co-enriched with taxa such as *Subdoligranulum*, *Odoribacter*, *Bilophila*, *Alistipes*. The MCI
350 group was co-enriched with *Ruminocoaceae_unclassified*, *Mollicutes_RF39_unclassified*,
351 *Ruminocoaceae_UCG_005*, *Lachnospiraceae_unclassified*. However, some taxa such as
352 *Odoribacter* was placed near the control group in Jensen-Shannon distance PCA Figure (8b),
353 suggesting co-enrichment of certain taxa can be somewhat influenced by the preferred distance
354 metric.

355 **Identification of signature taxa for AD continuum and association with metadata**

356
357 We constructed Random Forest (RF) model on selected features of gut microbiota and
358 psychometric test scores (MMSE and CDR) that are typically used as proxy in clinical diagnosis.
359 Using songbird, we selected 300 ASV (Top 25%) that differentiates between the healthy (control)
360 and the disease groups (MCI and AD). We then plotted the ASVs with the first 20 highest mean

361 decrease Gini values (Figure 9a) and included ASVs with mean decrease Gini values above the
362 breakpoint curve in the RF analysis. We identified the following 9 ASVs above the breakpoint:
363 *Faecalibacterium* (ASV45), *Sutterella*(ASV607), *Coprobacter*(ASV531), *Bacteroides* (ASV81),
364 *Anaerostipes*(ASV364), *Ruminoccaceae unclassified*(ASV203), *Lactobacillus* (ASV65),
365 *Clostridium_sensu_stricto_1* (ASV118), *Ruminococcus_1* (ASV59). Notably, ASVs beyond the
366 breakpoint line are largely the bacterial species responsible for the stratification of gut microbiota
367 in the samples such as *Faecalibacterium*, *Bacteroides*, and *Ruminococcus unclassified*. We next
368 calculated diagnostic accuracy of the RF model by plotting receiver operating characteristics
369 curve (ROC) for the above 9 taxa, MMSE, and CDR separately and in combination for each cohort
370 group (Figure 9b). The ROC value for these selected nine taxa were moderately accurate (AUC
371 63%, Flg 8a) but when we included MMSE and/or CDR, we found that the RF model robustly
372 classify all three groups (groupwise AUC range 0.74-1.0, Figures 9b).

373

374 **Taxa association with clinical parameters**

375

376 We used multivariate association with linear models (MaAsLin2) to assess association between
377 individual taxa and clinical metadata including patients drugs ($q \leq 0.25$). This analysis showed
378 that *Roseburia*, *Lactobacillus*, *Fusicatenibacter* were negatively associated with AD
379 (Supplementary Figure S5). Of the medication categories there are several taxa found to be
380 positively associated with anti-depression and statin. *Blautia*, *Caprococcus*, *Butyricoccus*, *Dorea*,
381 *Lachnospiraceae family members*, some *Ruminoclostridium* and *Ruminococaceae*, known to be
382 butyrate producers are all positively associated with antidepression drugs. Unexpectedly, we
383 found that several taxa were significantly associated with Statin medication and, of these taxa,
384 *Streptococcus* and unclassified member of *Erysipelotrichaceae* were highly significantly
385 associated with statin medication. We also observed the following taxa positively associated with
386 statin medication; unclassified members of *Ruminococaceae* and *Lachnospiraceae*,

387 *Phascolarctobacterium*, *Desulfovibrio*, *Caprobacter*, *Bifidobacterium*, *Butyricoccus*, *Blautia*,
388 *Barnesiella*.

390 DISCUSSION

391 In this study, we demonstrate that gut microbiota across AD continuum not only differentiates
392 between cognitive states but also comprise subgroups delineated by locally dominant co-
393 occurring bacteria. Stratification of the gut microbiota along the AD continuum is major unmet
394 need for diet-based and precision nutrition interventions in AD cohorts and here we present proof-
395 of-concept data that can be insightful for the emerging dietary and precision medicine/nutrition
396 initiatives involving AD patients. A key finding in this study is that these approaches all converge
397 on *Prevotella* and *Bacteroides* stratification, which are also robustly supported by enrichment and
398 ordination analyses that these two species are the drivers of community diversity and
399 composition. Rather than focusing on a single gut microbiota stratification method we have
400 exercised the best practice of implementing multiple methods to compare, contrast, and sought
401 support from alternative analyses. Also, all methods ranked the following taxa among the Top10
402 bacteria contributing to separation of the groups; *Escherichia/Shigella*, *Faecalibacterium*, *Blautia*,
403 *Ruminococcaceae_unclassified*, *Ruminococcaceae_UCG-002*, *Lachnospiraceae_unclassified*,
404 *Parabacteroides*, suggesting these taxa play significant role in the observed community structure
405 of the gut microbiota of the patients in this study.

406
407 PAM clustering and DMM concordantly showed three distinct clusters, one of which is consistent
408 with the recently described Bact2 group [44]. The subjects in this group are likely to have
409 aggravated dysbiosis as manifested from increased abundance of opportunistic pathogens
410 *Escherichia/Shigella* and some species of *Bacteroides* species and lower abundance of
411 *Faecalibacterium* and other SCFA producers. Notably, LDA analysis shuffles similar set of taxa

412 as the number of subgroups increase but *Bacteroides* and *Prevotella_9* are predominantly the
413 most abundant taxa in many of these clusters. Strikingly, *Escherichia/Shigella* dominates one of
414 the subgroups in LDA analysis together with opportunistic *Klebsiella* and *Enterococcus*,
415 suggesting dysbiotic community type may be enriched in this subgroup.

416

417 Topological data analysis (TDA) we used to stratify gut microbiota in this study deserves a
418 particular attention among others. TDA, based on the Mapper algorithm [50], represents the
419 underlying distribution of data in a metric space by dividing the data into overlapping similar
420 subsets according to a filter function, local clustering on each subset and representing the results
421 in an undirected network. A node in the network represents a group of samples with similar
422 microbiome profiles, and if common samples between nodes are shared then the nodes are
423 linked. Next, a modified special analysis of functional enrichment (SAFE) algorithm maps the
424 metadata and taxa into the network. Finally, vectors of SAFE scores can be used in ordination to
425 rank the driver taxa and their relationship with the metadata, all these algorithms are integrated
426 into *tmap* [43]. The SAFE scores we obtain following these algorithms allowed us to identify the
427 driver species that are responsible for community structure and showed their relationship with the
428 metadata. We employed Bray-Curtis and Jensen-Shannon to check the variation resulting from
429 distance metric. *Prevotella_9*, *Bacteroides*, and *Ruminococcus_unclassified* were ranked among
430 the top10 taxa with high SAFE scores, albeit in different order, suggesting TDA is robust and
431 consistent even with different distance metrics. In addition to these three taxa unclassified
432 members of again other taxa within *Ruminococcus* family and *Lachnospiraceae* were congruent
433 with other three methods we tested. Interestingly, this analysis identified GCA-900066575 taxa
434 (Uncultured human intestinal bacterium) as one of the subclusters in contrast with other methods
435 we used. This genus is taxonomically in the family of *Lachnospiraceae*, which includes members
436 of SCFA producers [51], still some other members were associated with metabolic diseases such
437 as obesity [52]. Indeed, another related member of this family GCA-900066225 ranked among

438 the top10 taxa when Bray-Curtis distance was used but enriched around AD. It is therefore
439 important to note that TDA, unlike clustering or probabilistic partitioning methods, provided fine
440 resolution in terms of stratification of the gut microbiota composition. Conversely, TDA did not
441 rank *Escherchia/Shigella* subnetwork among top ten taxa, neither the ordination showed clear
442 association with the disease. Together, bioinformatic tools developed in the field of microbiome
443 have all their strengths and drawbacks and therefore overlaps in bioinformatic analyses should
444 be pursued.

445
446 Several lines of evidence showed human cohorts in microbiome studies can be phenotypically
447 partitioned along *Prevotella* and *Bacteroides* stratification [53-58]. A recent comprehensive report
448 [59] provided evidence that Mediterranean diet-based intervention is associated with specific
449 functional and taxonomic components of the gut microbiome, and its effect is a function of
450 microbial composition. Notably, absence of *Prevotella copri* in the gut microbiomes of a subgroup
451 of participants was associated with the protective health benefits of the dietary intervention,
452 emphasizing the premise that microbiome-informed stratified dietary intervention would be quite
453 effective. Nevertheless, *P. copri* is ambivalently associated with both health and diseases
454 depending on the strain and geography [60], which prompts us to further consider its role in AD.

455
456 Taxonomically, the genus *Prevotella_9* is predicted to belong to *Prevotella copri* complex [61].
457 Comparative genome analysis of the strains of *P. copri* complex, however, show that some strains
458 qualify to be assigned to even a separate species of *Prevotella* due to low genomic similarities
459 [62, 63]. Some *P. copri* strains are associated with disease states such as rheumatoid arthritis
460 [64], while some other strains are associated with habitual diet and life style [54] and
461 underrepresented in Westernized populations. Thus, strain level resolution of *Prevotella_9* is
462 needed to draw inferences. Expectedly, multiple strains of *P. copri* are likely to be part of the
463 bacterial community in the samples. Even though we found *Prevotella_9* to be associated with

464 the control group the enrichment analysis using songbird ranked some ASVs belong to
465 *Prevotella_9* (species level) at the top and few other ASVs at the bottom of the log ratio
466 differentials, suggesting analysis beyond species taxonomic hierarchy would provide better
467 resolution in terms of their associations with human phenotypes. Oligotypes of these two genera
468 in an earlier work were found to be differentially associated with plant based or some others were
469 associated with animal-based diet [55]. A recent report provided evidence that *Bacteroides*
470 *cellulosilyticus* predicted weight gain more precisely than the ratio of *Prevotella* and *Bacteroides*
471 genus. Together, our differential enrichment analysis results are in line with these reports that
472 species or even strain level resolution of these two genera could provide better predictive
473 biomarker power for diet-based intervention studies.

474

475 One limitation of our study was that although we were able control drug induced confounding, we
476 did not control other potential confounders such as diet, BMI, stool consistency. We largely
477 recruited cohabiting spouses as non-demented controls sharing the same diet patterns with the
478 patients and carnivory is rare due to the high cost of meat in the country. We therefore did not
479 predict diet can strongly impact our results.

480

481 In conclusion, we demonstrate in this study that gut microbiota along the Alzheimer's Disease
482 continuum comprises stratified community structure dashed primarily by *Prevotella* and
483 *Bacteroides* but also subnetworks of other taxa exist in the community. The signature taxa when
484 used together with MMSE and CDR robustly classify heterogenous groups hence posing potential
485 biomarker value. The study adds to limited number of clinical studies profiling gut microbiota of
486 AD continuum patients.

487

488

489

490

MATERIALS AND METHODS

491

492 **Subject Recruitment and Study Design:** The Istanbul Medipol University and Erciyes University
493 Ethical Review Boards approved this study (Approval numbers: 186/16.4.2015 and 85/
494 20.02.2015, respectively). All participants were informed of the objectives of this study and signed
495 a written consent form prior to their participation. The diagnosis of dementia and MCI due to AD
496 were based on the criteria of the National Institute on Aging-Alzheimer's Association workgroups
497 on diagnostic guidelines for Alzheimer's disease [65, 66]. Exclusion criteria for this study included
498 history of substance abuse, any significant neurologic disease, major psychiatric disorders
499 including major depression. Also, individuals who used commercial probiotics or antibiotics during
500 the study period or within 1-month prior to providing stool sample, or who major GI tract surgery
501 in past 5 years. Both health centers followed the same protocols in recruiting cohorts and used
502 kits from the same manufacturers to minimize the variations in wet lab procedures.

503

504 **Lumbar puncture, CSF biomarkers assays:** Cerebro Spinal Fluid (CSF) samples were included
505 in the analyses from a subset of AD patients if the patient was requested to donate CSF sample
506 as part of the clinically mandated diagnostic protocol. CSF samples were collected in the morning
507 after overnight fasting using spinal needles (22 gauge) and syringes at the L3/4 or L4/5 interspace.
508 CSF was then aliquoted into 0.5 mL non-adsorbing polypropylene tubes and stored at -80 °C
509 until assay. Biomarker molecules in CSF (A β ₄₂, phosphorylated tau (p-tau), and the p-tau/A β 42
510 ratio) were measured consistent with the Alzheimer's Association flowchart for lumbar puncture
511 and CSF sample processing and the biomarker levels were determined as previously described
512 [67]. Single 96-well ELISA kits (Innogenetics, Ghent, Belgium) were used in quantitation.

513

514 **Sample collection and DNA extraction:** Stool samples from all participants were collected in
515 the neurology clinics of the university training hospitals. The participants were given a collection

516 kit included a sterile tube and provided a brief instruction for collection. Self-collected samples
517 were placed within approximately 30 mins of collection in -80 freezers and kept frozen until DNA
518 extraction.

519
520 16S rRNA gene sequencing and PCR were performed as previously described [68] with minor
521 modifications. Briefly, genomic DNA was extracted from 220 mg fecal samples using QiaAmp
522 DNA Stool Mini Kit (Qiagen, Germany) per manufacturer's instructions with the addition of bead
523 beating (0.1 mm zirconium-beads) and lysozyme and RNase A incubation steps.

524
525 **PCR and amplicon sequencing:** To amplify the variable V3-V4 regions of the 16S rRNA gene,
526 the primers 341 F (5'-CCTACGGNGGCWGCAG-3') and 805 R (5'-
527 GACTACHVGGTATCTAATCC-3') were used. MiSeq sequencing adaptor sequences were
528 added to the 5' ends of forward and reverse primers. Approximately 12.5 ng of purified DNA from
529 each sample was used as a template for PCR amplification in 25 μ l reaction mixture by using 2 \times
530 KAPA HiFi Hot Start Ready Mix (Kapa Biosystems, MA, USA). For PCR amplification, the
531 following conditions were followed: denaturation at 95 °C for 3 min., followed by 25 cycles of
532 denaturation at 95 °C for 30 sec., annealing at 55 °C for 30 sec. and extension at 72 °C for 30
533 sec., with a final extension at 72 °C for 5 min. Amplified PCR products were purified with
534 Agencourt AMPure XP purification system (Beckman Coulter) and Nextera PCR was performed
535 by using sample-specific barcodes. The constructed Nextera libraries were then sequenced by
536 Illumina MiSeq platform using MiSeq Reagent Kit v2 chemistry.

537
538 **Sequence processing and taxonomic assignment:** The pair-end 16S rRNA reads were first
539 used cutadapt v1.9 program [69] for the process of quality filtering, trimming and uploaded on the
540 DADA2 pipeline [34] integrated into the Nephele platform [70] (v.2.0, <http://nephele.niaid.nih.gov>).
541 Chimeric sequences are automatically removed by this pipeline, which generates both rarefied

542 and unrarefied ASV abundance tables. We used Rarefied (10769 reads/sample) ASV table in
543 most downstream analysis due to large differences between some total sample reads except for
544 the scale invariant DEICODE and songbird. We removed any sequences that were classified as
545 either being originated from eukarya, archaea, mitochondria, chloroplasts or unknown kingdoms.

546

547 **Quality control:** We included no sample DNA extractions and no template negative control
548 samples in every sequencing library prepared. Using reads in the negative control samples as
549 reference we identified and removed probable contaminant reads of 13 ASVs from the ASV table,
550 as predicted by Decontam R package [71] using the 'prevalence' method. In this method, the
551 binary coded features across samples are compared to the prevalence in negative controls to
552 identify contaminants. Also, we sequenced the same amplicon of an AD sample three times to
553 check the sequencing variation. Although both centers used same protocols and kits from the
554 same manufacturer in sequencing, we sequenced amplicons amplified from two same genomic
555 DNA templates again from AD samples at both centers to check the center-to-center sequencing
556 concordance. No differences could be identified between the taxonomic compositions of the
557 samples sequenced at both centers nor between the technical replicates (PCoA, PERMANOVA
558 p=0.1).

559

560 **Numerical Ecology and Statistical Analysis:** Most numerical downstream analysis of ASV
561 abundances were performed in R environment [72]. All P values, where appropriate, were
562 adjusted for multiple testing using Benjamini-Hochberg (False Discovery Rate; FDR) method. We
563 measured within samples microbial diversity (alpha diversity) using Observed richness, Chao1,
564 Shannon, and Inverse Simpson in *phyloseq* [73] and tested using Kruskal Wallis. To identify
565 differentially abundant bacterial species we employed animalculus [58] and limma [74] R
566 packages. We assessed microbial diversity between samples (beta diversity) using multiple
567 distance metrics including Bray-Curtis, Jaccard, Canonical Analysis of Principal Components

568 (CAP). CAP analysis and the similarity percentages breakdown (SIMPER) procedure were
569 performed using PRIMER.v7 [75] . Additionally, due to the compositional nature of the data, we
570 also included robust Aitchison PCA, using the Qiime2 DEICODE plugin [36] to calculate beta
571 diversity with feature loadings. The resulting ordination was visualized using Emperor [76]. We
572 tested significance of beta diversity among groups using again Qiime diversity plugin
573 PERMANOVA.

574

575 Next, we used Songbird [37] for multinomial regression to rank species association with disease
576 status with the following parameters: (formula: "MMSE+CDR+Sex+Edu+C(Group, Diff,
577 levels='C','MCI','AD'), -p-epochs 10000 --p-differential-prior 0.5 --p-summary-interval 1 --p-
578 random-seed 3 --min-sample-count 1000 --min-feature-count 0) . Of note, the formula structure
579 follows Patsy formatting (<https://patsy.readthedocs.io/en/latest/>) such that Groups (C, MCI, AD)
580 represent levels=[“healthy”, “mild”, “severe”] states, respectively. A null model was generated
581 using the same parameters. The fitted model demonstrated better fit compared to the null model
582 (pseudo $Q^2 = 0.874027$). Taxa ranks were visualized using Qurro [38]. Significance was
583 determined using a Welch’s t-test between groups, performed by Graph Pad Prism.

584

585 To identify microbial species associated with the clinical metadata including patients’ medication
586 we performed multivariate association with linear models (MaAsLin2) [77]. The control group was
587 excluded from this analysis as they were not normally prescribed these medications. We
588 employed the R package MaAsLin 2.1.0 to perform per-feature tests. We log-transformed relative
589 abundances of microbial species and standardized continuous variables into Z-scores and binary
590 encoded medication information before including them in the MaAsLin models ($q < 0.25$ for
591 significance).

592

593 **Stratification of gut microbiota:** We employed clustering, probabilistic partitioning, and
594 topological data analysis approaches for the stratification of gut microbiota in the samples.
595 Partitioning around the medoid (PAM) approach [39] clusters samples by iteratively updating each
596 cluster's medoid. We assigned samples to community types using the function *pam()* in R
597 package *cluster* based on Bray Curtis and Jensen Shannon distances. The number of clusters
598 was determined by Gap statistic evaluation. Departing from the clustering approach, we next used
599 two distinct probabilistic methods to partition microbiota landscape, namely Dirichlet multinomial
600 mixture models (DMM) [40] and Latent Dirichlet Allocation (LDA) [41, 42]. Genus level
601 abundances were fitted to DMM models to partition microbial community profiles into a finite
602 number of clusters, using the Laplace approximation as previously described [40, 78].

603
604 As a second probabilistic partitioning we performed LDA, is a multi-level hierarchical Bayesian
605 model [41] otherwise used for collections of discrete data such as text corpus analysis in
606 linguistics. LDA is a generalization of Dirichlet multinomial mixture modeling where biological
607 samples are allowed to have fractional membership and distinct microbial communities have
608 different microbial signatures. Thus, for each taxon there is a vector of probabilities across all
609 clusters that it can be assigned to. Each cluster, therefore, has a different probability of containing
610 taxa, indicating chance of microbes in a particular subgroup (strata) co-occurring due to
611 community assembly dynamics. To fit the model we used Gibb's sampling with the R package
612 *MetaTopics* (v.1.0) [79]. The relative abundances of genus collapsed table with abundances more
613 than 0.1% and 5% sample prevalence was input to the model. We plotted perplexity measure and
614 loglikelihood values to estimate model performance and optimal number of topics (subgroups of
615 microbial assemblages) using 5-fold cross-validation. However, we observed that both
616 parameters continued to improve with increasing subgroup number without a clear optimum
617 except the first jump in perplexity was near 10 topics. We therefore picked first 10 topics for the
618 sake of interpretability.

619
620 The final method we applied was topological data analysis (TDA) based on the Mapper algorithm
621 [50] and network representation for stratification and association of study of high dimensional
622 microbiome data, all integrated into *tmap* tool [43]. The framework enables to reveal association
623 of taxa or metadata within the entire network and to identify enrichment subnetworks of different
624 association patterns. Conceptually, the Mapper algorithm transforms a distance matrix and
625 represent the shape of the data cloud in an undirected network. Next, a modified version of special
626 analysis functional enrichment (SAFE) algorithm to map the value of the target feature into the
627 network was employed, followed by ordination of SAFE scores to show taxa-metadata association
628 [43].

629
630 **Signature taxa:** To identify microbial signature of severity of cognitive impairment in AD
631 continuum we implemented a machine learning procedure. We first took advantage of songbird
632 tool to select features including the covariates and healthy (control) and disease states (AD+MCI)
633 in the model formula. We subsequently fit the list of ASV selected this way into Random Forest
634 models. We plotted the area under the receiver operating characteristic curve (AUROC) to show
635 prediction performance of the models. To create the classifiers, a random forest constituted of
636 500 trees were computed using the default settings of the “randomForest” function implemented
637 in the randomForest R package (v4.6-7). Mean decrease Gini values were averaged for each
638 ASV among the 100 random forest replicates. The ASVs with the first 20 highest mean decrease
639 Gini values were plotted. ASVs with mean decrease Gini values above the breakpoint curve were
640 chosen to be part of the classifier. Breakpoints were estimated using the “breakpoints” function
641 included in the strucchange R package [52]. We subsequently fit the list of ASVs selected this
642 way with or without psychometric test values, i.e. MMSE and CDR, into Random Forest models,
643 and bootstrapped for 100 times. We plotted the area under the receiver operating characteristic
644 curve (AUROC) to show prediction performance of the models.

645

646 DATA ACCESSION

647

648 The 16S rRNA generated by this study have been submitted to the NCBI BioProject database,
649 (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA734525.

650

651 AUTHOR CONTRIBUTIONS

652

653 **Conception and Design:** SY, OUN, EK, and LH; **Sample Collection and Processing:** BS, AG,
654 FK, MFG, DK, AES, HAV, EAG, and KS; **Data Analysis:** SY, OUN, AB, MA, and MK; **Data**
655 **Interpretation:** SY, OUN, MK, AM, LH and EK. **Manuscript Writing – Original Draft:** SY; **Writing,**
656 **Review, and Editing:** OUN, MK, AM, LH and EK. All authors read and approved the final
657 manuscript.

658

659 DISCLOSURE DECLARATION

660

661 The authors do not have any conflicts of interest to disclose.

662

663 LEGENDS FOR SUPPLEMENTAL TABLES AND FIGURES

664

665 **Table S1.** Levels of Cerebro-Spinal Fluid Biomarkers of a Subset of AD Patients.

666 **Table S2-S5.** Differentially Abundant ASV and genus level taxa between cohort groups as
667 detected by Limma-Voom Model (Age and Sex Adjusted)

668 **Table S6.** PERMANOVA analysis of covariates

669 **Table S7-S8.** Enrichment analysis by multinomial regression embedded in the songbird (Set1,
670 Set2, Set3, and Set4)

671 **Table S9-S10:** Ranking of SAFE scores calculated using tmap algorithm

672

673

674 **Figure S1. Alpha diversity analysis.** Box plots show (A) Chao1 index, (B) Inverse Simpson, (C)
675 Observed species, (D) Shannon diversity index

676 **Figure S2. Multi-Dimensional Scale (MDS) Analysis of genus relative abundances.** (A) MDS
677 analysis of the samples (B) Gradient of *Prevotella_9* abundances across the samples. (C)
678 Gradient of *Bacteroides* abundances across the samples

679 **Figure S3. Determining the number of clusters in the gut microbiota.** The optimal number of
680 clusters based on (A) Gap statistic with standart error bars for PAM analysis. (B) Laplace method
681 for evaluating model fit for increasing number of Dirichlet mixture components

682 **Figure S4. Latent Dirichlet Allocation Model Performance.** LDA model's perplexity parameter
683 (top) and log-likelihood values (bottom) to find optimal number of clusters.

684 **Figure S5. Associations of the patient drugs with genus-level features.** The heatmap shows
685 per-feature testing in MaAsLin 2 using linear mixed models to identify microbial species
686 associated with drugs used by the patients. Colors of the heatmap reflects the beta coefficient for
687 drugs and age and sex from linear mixed models in MaAsLin 2 with genus-level feature as
688 outcomes.

689 **ACKNOWLEDGEMENTS**

690

691 This study was supported by funding from The Scientific and Technological Research Council of
692 Turkey (TÜBITAK) to Prof. Dr. Süleyman Yıldırım and Prof. Dr. Emel Köseoglu (Project numbers.
693 2236-115C056 and 215S707, respectively). The funding agency had no role in study design, data
694 collection and interpretation, or the decision to submit the work for publication.

695

696

697 **References**

698

699

700 1. International., A.s.D., *Alzheimer's Disease International. World Alzheimer Report 2015: the Global Impact of Dementia. An Analyses of Prevalence, incidence, Cost and Trends.* 701 <https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf>, Accessed Sep 20, 2020. 702 2015.

703 2. Dubois, B., et al., *Preclinical Alzheimer's disease: Definition, natural history, and* 704 *diagnostic criteria.* Alzheimers Dement, 2016. **12**(3): p. 292-323.

705 3. Aisen, P.S., et al., *On the path to 2025: understanding the Alzheimer's disease* 706 *continuum.* Alzheimers Res Ther, 2017. **9**(1): p. 60.

707 4. Kumar, A., A. Singh, and Ekavali, *A review on Alzheimer's disease pathophysiology and* 708 *its management: an update.* Pharmacol Rep, 2015. **67**(2): p. 195-203.

709 5. Mehta, D., et al., *Why do trials for Alzheimer's disease drugs keep failing? A* 710 *discontinued drug perspective for 2010-2015.* Expert Opin Investig Drugs, 2017. **26**(6): 711 p. 735-739.

712 6. Kametani, F. and M. Hasegawa, *Reconsideration of Amyloid Hypothesis and Tau* 713 *Hypothesis in Alzheimer's Disease.* Front Neurosci, 2018. **12**: p. 25.

714 7. Wang, J., et al., *A systemic view of Alzheimer disease - insights from amyloid-beta* 715 *metabolism beyond the brain.* Nat Rev Neurol, 2017. **13**(10): p. 612-623.

716 8. Itzhaki, R.F., et al., *Microbes and Alzheimer's Disease.* J Alzheimers Dis, 2016. **51**(4): p. 717 979-84.

718 9. Guillot-Sestier, M.V., K.R. Doty, and T. Town, *Innate Immunity Fights Alzheimer's* 719 *Disease.* Trends Neurosci, 2015. **38**(11): p. 674-681.

720

721 10. Rangasamy, S.B., et al., *Selective disruption of TLR2-MyD88 interaction inhibits*
722 *inflammation and attenuates Alzheimer's pathology*. J Clin Invest, 2018. **128**(10): p.
723 4297-4312.

724 11. Association, A.s., *2020 Alzheimer's disease facts and figures*. Alzheimers Dement, 2020.

725 12. Bullain, S. and R. Doody, *What works and what does not work in Alzheimer's disease?*
726 *From interventions on risk factors to anti-amyloid trials*. J Neurochem, 2020.

727 13. Bredesen, D.E., *Reversal of cognitive decline: a novel therapeutic program*. Aging
728 (Albany NY), 2014. **6**(9): p. 707-17.

729 14. Bredesen, D.E., et al., *Reversal of cognitive decline in Alzheimer's disease*. Aging
730 (Albany NY), 2016. **8**(6): p. 1250-8.

731 15. Isaacson, R.S., et al., *Individualized clinical management of patients at risk for*
732 *Alzheimer's dementia*. Alzheimers Dement, 2019. **15**(12): p. 1588-1602.

733 16. Keine, D., et al., *Development, Application, and Results from a Precision-medicine*
734 *Platform that Personalizes Multi-modal Treatment Plans for Mild Alzheimer's Disease*
735 *and At-risk Individuals*. Curr Aging Sci, 2018. **11**(3): p. 173-181.

736 17. Fan, Y. and O. Pedersen, *Gut microbiota in human metabolic health and disease*. Nat
737 Rev Microbiol, 2020.

738 18. Fung, T.C., C.A. Olson, and E.Y. Hsiao, *Interactions between the microbiota, immune*
739 *and nervous systems in health and disease*. Nat Neurosci, 2017. **20**(2): p. 145-155.

740 19. Fang, P., et al., *The Microbiome as a Modifier of Neurodegenerative Disease Risk*. Cell
741 Host Microbe, 2020. **28**(2): p. 201-222.

742 20. Haran, J.P., et al., *Alzheimer's Disease Microbiome Is Associated with Dysregulation of*
743 *the Anti-Inflammatory P-Glycoprotein Pathway*. mBio, 2019. **10**(3).

744 21. Li, B., et al., *Mild cognitive impairment has similar alterations as Alzheimer's disease in*
745 *gut microbiota*. Alzheimers Dement, 2019. **15**(10): p. 1357-1366.

746 22. Liu, P., et al., *Altered microbiomes distinguish Alzheimer's disease from amnestic mild*
747 *cognitive impairment and health in a Chinese cohort*. *Brain Behav Immun*, 2019. **80**: p.
748 633-643.

749 23. Nagpal, R., et al., *Modified Mediterranean-ketogenic diet modulates gut microbiome and*
750 *short-chain fatty acids in association with Alzheimer's disease markers in subjects with*
751 *mild cognitive impairment*. *EBioMedicine*, 2019. **47**: p. 529-542.

752 24. Saji, N., et al., *Analysis of the relationship between the gut microbiome and dementia: a*
753 *cross-sectional study conducted in Japan*. *Sci Rep*, 2019. **9**(1): p. 1008.

754 25. Vogt, N.M., et al., *Gut microbiome alterations in Alzheimer's disease*. *Sci Rep*, 2017.
755 **7**(1): p. 13537.

756 26. Vogt, N.M., et al., *The gut microbiota-derived metabolite trimethylamine N-oxide is*
757 *elevated in Alzheimer's disease*. *Alzheimers Res Ther*, 2018. **10**(1): p. 124.

758 27. Zhuang, Z.Q., et al., *Gut Microbiota is Altered in Patients with Alzheimer's Disease*. *J*
759 *Alzheimers Dis*, 2018. **63**(4): p. 1337-1346.

760 28. MahmoudianDehkordi, S., et al., *Altered bile acid profile associates with cognitive*
761 *impairment in Alzheimer's disease-An emerging role for gut microbiome*. *Alzheimers*
762 *Dement*, 2019. **15**(1): p. 76-92.

763 29. Amini, Y., et al., *The Role of Nutrition in Individualized Alzheimer's Risk Reduction*. *Curr*
764 *Nutr Rep*, 2020. **9**(2): p. 55-63.

765 30. Isaacson, R.S., et al., *The clinical practice of risk reduction for Alzheimer's disease: A*
766 *precision medicine approach*. *Alzheimers Dement*, 2018. **14**(12): p. 1663-1673.

767 31. Kolodziejczyk, A.A., D. Zheng, and E. Elinav, *Diet-microbiota interactions and*
768 *personalized nutrition*. *Nat Rev Microbiol*, 2019. **17**(12): p. 742-753.

769 32. Norwitz, N.G., et al., *Precision Nutrition for Alzheimer's Prevention in ApoE4 Carriers*.
770 *Nutrients*, 2021. **13**(4).

771 33. Schelke, M.W., et al., *Nutritional interventions for Alzheimer's prevention: a clinical*
772 *precision medicine approach*. Ann N Y Acad Sci, 2016. **1367**(1): p. 50-6.

773 34. Callahan, B.J., et al., *DADA2: High-resolution sample inference from Illumina amplicon*
774 *data*. Nat Methods, 2016. **13**(7): p. 581-3.

775 35. Calgaro, M., et al., *Assessment of statistical methods from single cell, bulk RNA-seq,*
776 *and metagenomics applied to microbiome data*. Genome Biol, 2020. **21**(1): p. 191.

777 36. Martino, C., et al., *A Novel Sparse Compositional Technique Reveals Microbial*
778 *Perturbations*. mSystems, 2019. **4**(1).

779 37. Morton, J.T., et al., *Establishing microbial composition measurement standards with*
780 *reference frames*. Nat Commun, 2019. **10**(1): p. 2719.

781 38. Fedarko, M.W., et al., *Visualizing 'omic feature rankings and log-ratios using Qurro*. NAR
782 Genom Bioinform, 2020. **2**(2): p. lqaa023.

783 39. Arumugam, M., et al., *Enterotypes of the human gut microbiome*. Nature, 2011.
784 **473**(7346): p. 174-80.

785 40. Holmes, I., K. Harris, and C. Quince, *Dirichlet multinomial mixtures: generative models*
786 *for microbial metagenomics*. PLoS One, 2012. **7**(2): p. e30126.

787 41. Blei, D., Ng AY, Jordan MI, *Latent Dirichlet allocation*. J Mach Learn Res, 2003. **3**: p.
788 993–1022.

789 42. Sankaran, K. and S.P. Holmes, *Latent variable modeling for the microbiome*.
790 Biostatistics, 2019. **20**(4): p. 599-614.

791 43. Liao, T., et al., *tmap: an integrative framework based on topological data analysis for*
792 *population-scale microbiome stratification and association studies*. Genome Biol, 2019.
793 **20**(1): p. 293.

794 44. Vandeputte, D., et al., *Quantitative microbiome profiling links gut community variation to*
795 *microbial load*. Nature, 2017. **551**(7681): p. 507-511.

796 45. Vieira-Silva, S., et al., *Statin therapy is associated with lower prevalence of gut*
797 *microbiota dysbiosis*. *Nature*, 2020. **581**(7808): p. 310-315.

798 46. Breuninger, T.A., et al., *Associations between habitual diet, metabolic disease, and the*
799 *gut microbiota using latent Dirichlet allocation*. *Microbiome*, 2021. **9**(1): p. 61.

800 47. Hosoda, S., et al., *Revealing the microbial assemblage structure in the human gut*
801 *microbiome using latent Dirichlet allocation*. *Microbiome*, 2020. **8**(1): p. 95.

802 48. Sommeria-Klein, G., et al., *Latent Dirichlet Allocation reveals spatial and taxonomic*
803 *structure in a DNA-based census of soil biodiversity from a tropical forest*. *Mol Ecol*
804 *Resour*, 2020. **20**(2): p. 371-386.

805 49. Romano, S., et al., *Meta-analysis of the Parkinson's disease gut microbiome suggests*
806 *alterations linked to intestinal inflammation*. *NPJ Parkinsons Dis*, 2021. **7**(1): p. 27.

807 50. Singh, G., M'émoli, F., and Carlsson, G. E., *Topo-logical methods for the analysis of*
808 *high dimensional data sets and 3d object recognition*. *SPBG*, 2007: p. 91-100.

809 51. Meehan, C.J. and R.G. Beiko, *A phylogenomic view of ecological specialization in the*
810 *Lachnospiraceae, a family of digestive tract-associated bacteria*. *Genome Biol Evol*,
811 2014. **6**(3): p. 703-13.

812 52. Vacca, M., et al., *The Controversial Role of Human Gut Lachnospiraceae*.
813 *Microorganisms*, 2020. **8**(4).

814 53. Christensen, L., et al., *Microbial enterotypes beyond genus level: Bacteroides species as*
815 *a predictive biomarker for weight change upon controlled intervention with arabinoxylan*
816 *oligosaccharides in overweight subjects*. *Gut Microbes*, 2020. **12**(1): p. 1847627.

817 54. De Filippis, F., et al., *Distinct Genetic and Functional Traits of Human Intestinal*
818 *Prevotella copri Strains Are Associated with Different Habitual Diets*. *Cell Host Microbe*,
819 2019. **25**(3): p. 444-453 e3.

820 55. De Filippis, F., et al., *Unusual sub-genus associations of faecal Prevotella and*
821 *Bacteroides with specific dietary patterns*. *Microbiome*, 2016. **4**(1): p. 57.

822 56. Gorvitovskaia, A., S.P. Holmes, and S.M. Huse, *Interpreting Prevotella and Bacteroides*
823 *as biomarkers of diet and lifestyle*. *Microbiome*, 2016. **4**: p. 15.

824 57. Levy, R., et al., *Longitudinal analysis reveals transition barriers between dominant*
825 *ecological states in the gut microbiome*. *Proc Natl Acad Sci U S A*, 2020. **117**(24): p.
826 13839-13845.

827 58. Wu, G.D., et al., *Linking long-term dietary patterns with gut microbial enterotypes*.
828 *Science*, 2011. **334**(6052): p. 105-8.

829 59. Wang, D.D., et al., *The gut microbiome modulates the protective association between a*
830 *Mediterranean diet and cardiometabolic disease risk*. *Nat Med*, 2021. **27**(2): p. 333-343.

831 60. Ley, R.E., *Gut microbiota in 2015: Prevotella in the gut: choose carefully*. *Nat Rev*
832 *Gastroenterol Hepatol*, 2016. **13**(2): p. 69-70.

833 61. Henderson, G., et al., *Improved taxonomic assignment of rumen bacterial 16S rRNA*
834 *sequences using a revised SILVA taxonomic framework*. *PeerJ*, 2019. **7**: p. e6496.

835 62. Tett, A., et al., *The Prevotella copri Complex Comprises Four Distinct Clades*
836 *Underrepresented in Westernized Populations*. *Cell Host Microbe*, 2019. **26**(5): p. 666-
837 679 e7.

838 63. Tett, A., et al., *Prevotella diversity, niches and interactions with the human host*. *Nat Rev*
839 *Microbiol*, 2021.

840 64. Scher, J.U., et al., *Expansion of intestinal Prevotella copri correlates with enhanced*
841 *susceptibility to arthritis*. *Elife*, 2013. **2**: p. e01202.

842 65. Albert, M.S., et al., *The diagnosis of mild cognitive impairment due to Alzheimer's*
843 *disease: recommendations from the National Institute on Aging-Alzheimer's Association*
844 *workgroups on diagnostic guidelines for Alzheimer's disease*. *Alzheimers Dement*, 2011.
845 7(3): p. 270-9.

846 66. McKhann, G.M., et al., *The diagnosis of dementia due to Alzheimer's disease:*
847 *recommendations from the National Institute on Aging-Alzheimer's Association*

848 *workgroups on diagnostic guidelines for Alzheimer's disease*. Alzheimers Dement, 2011.

849 7(3): p. 263-9.

850 67. Wallin, A.K., et al., *CSF biomarkers predict a more malignant outcome in Alzheimer*

851 *disease*. Neurology, 2010. 74(19): p. 1531-7.

852 68. Demircan, T., et al., *Experimentally induced metamorphosis in highly regenerative*

853 *axolotl (ambystoma mexicanum) under constant diet restructures microbiota*. Sci Rep,

854 2018. 8(1): p. 10974.

855 69. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing*

856 *reads*. EMBnet J, 2011. 17:10–12.

857 70. Weber, N., et al., *Nephele: a cloud platform for simplified, standardized and reproducible*

858 *microbiome data analysis*. Bioinformatics, 2018. 34(8): p. 1411-1413.

859 71. Davis, N.M., et al., *Simple statistical identification and removal of contaminant*

860 *sequences in marker-gene and metagenomics data*. Microbiome, 2018. 6(1): p. 226.

861 72. *R Core Team (2020).R: A language and environment for statistical computing. R*

862 *Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.*

863 2020.

864 73. McMurdie, P.J. and S. Holmes, *phyloseq: an R package for reproducible interactive*

865 *analysis and graphics of microbiome census data*. PLoS One, 2013. 8(4): p. e61217.

866 74. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing*

867 *and microarray studies*. Nucleic Acids Res, 2015. 43(7): p. e47.

868 75. Clarke, K.R.a.G., R. N. , *PRIMER v7: User Manual/Tutorial. PRIMER-E Plymouth*. 1993.

869 76. Vazquez-Baeza, Y., et al., *EMPeror: a tool for visualizing high-throughput microbial*

870 *community data*. Gigascience, 2013. 2(1): p. 16.

871 77. Himmel Mallick, L.J.M., Ali Rahnavard, Siyuan Ma, Yancong Zhang, Long H. Nguyen1,

872 Timothy L. Tickle, George Weingart, Boyu Ren, Emma Schwager, Ayshwarya

873 Subramanian, Yiren Lu, Levi Waldron, Joseph N. Paulson, Eric A. Franzosa, Hector

874 Corrada Bravo, Curtis Huttenhower, *Multivariable Association in Population-scale Meta-*
875 *omics Studies*, in *bioRxiv*. 2021.

876 78. Ding, T. and P.D. Schloss, *Dynamics and associations of microbial community types*
877 *across the human body*. *Nature*, 2014. **509**(7500): p. 357-60.

878 79. Yan, J., et al., *MetaTopics: an integration tool to analyze microbial community profile by*
879 *topic model*. *BMC Genomics*, 2017. **18**(Suppl 1): p. 962.

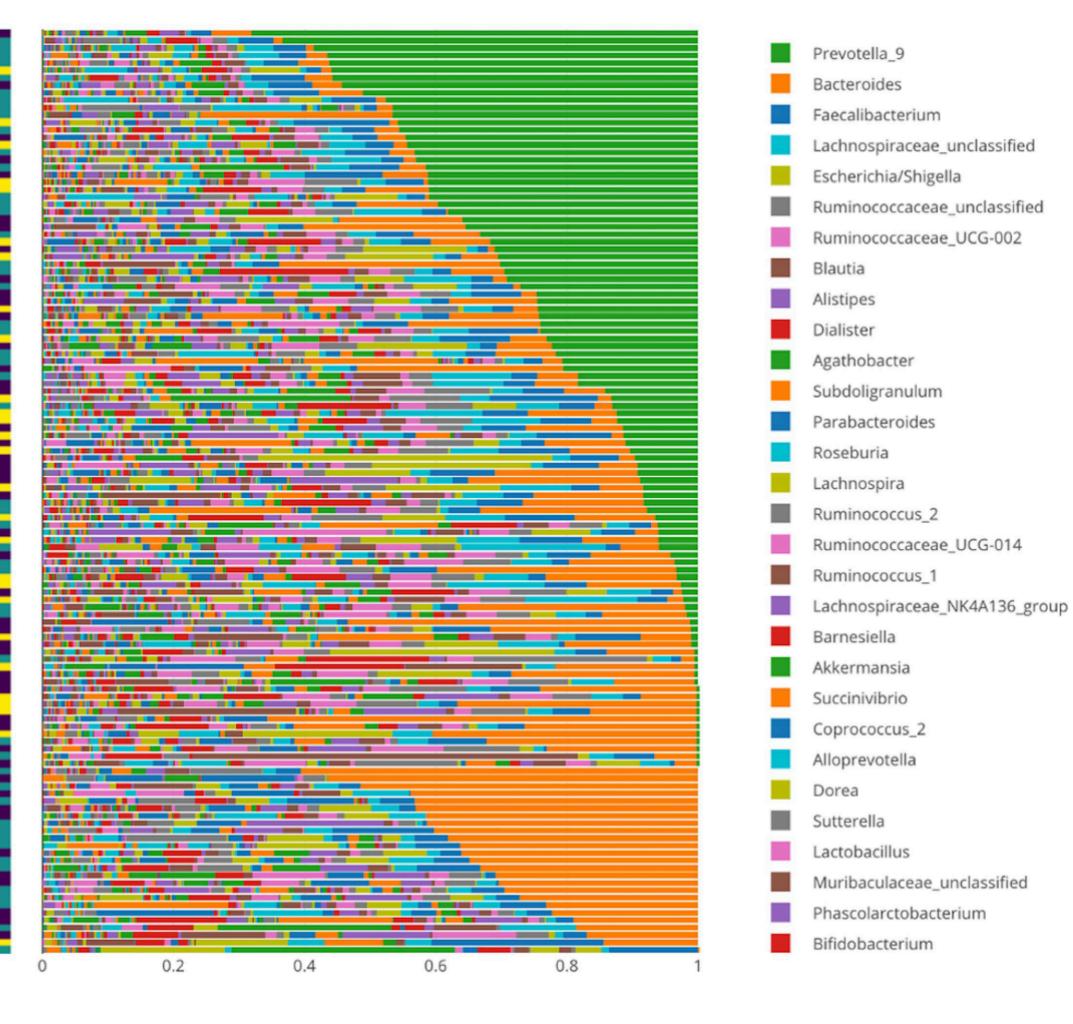
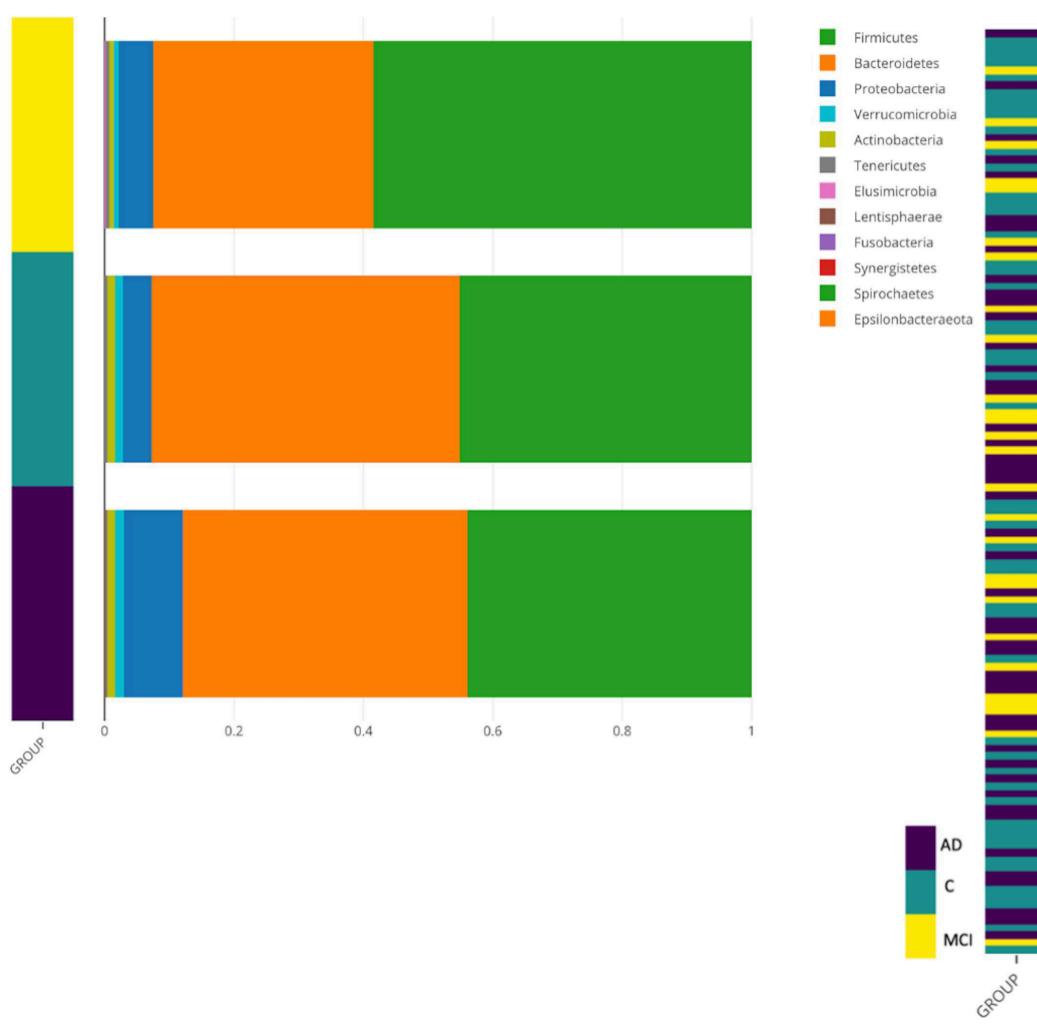
880

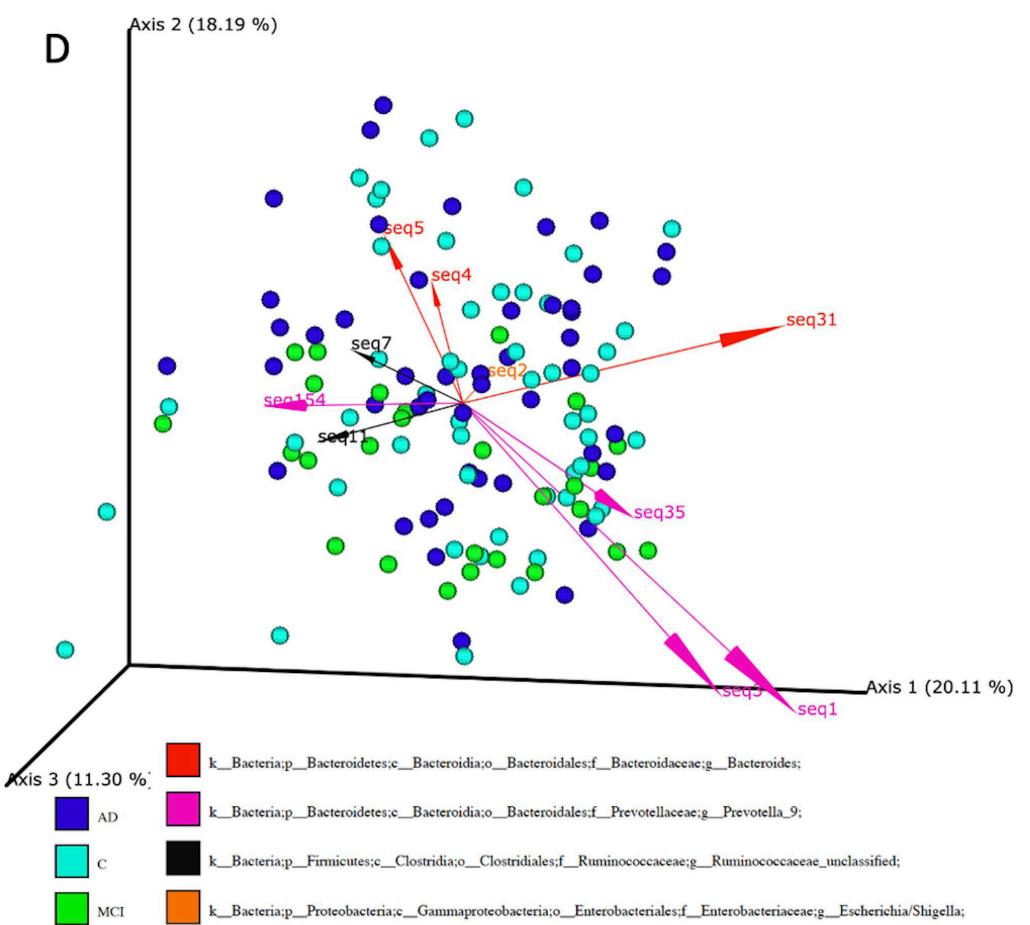
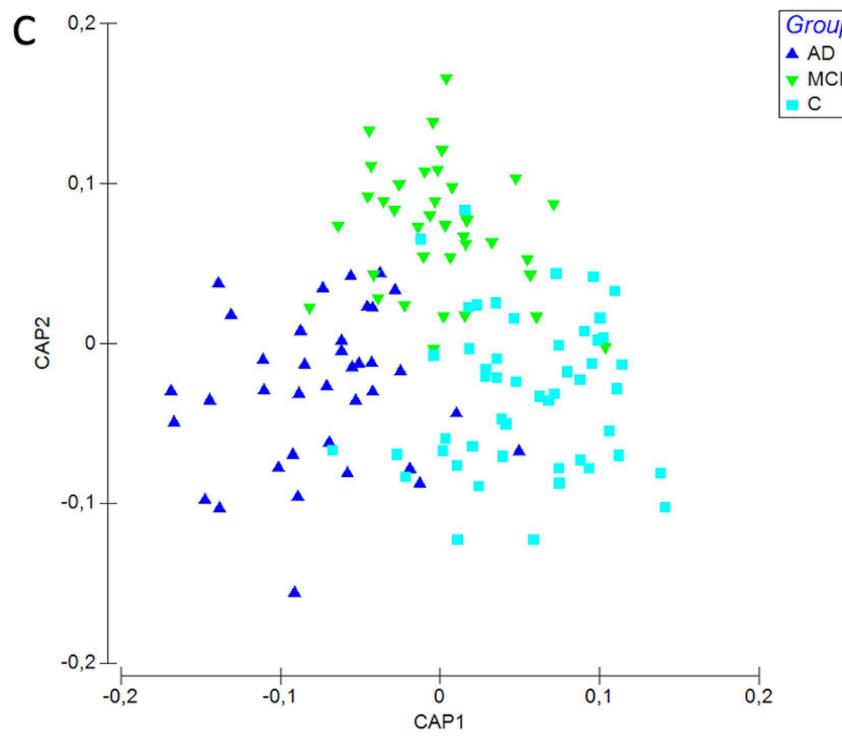
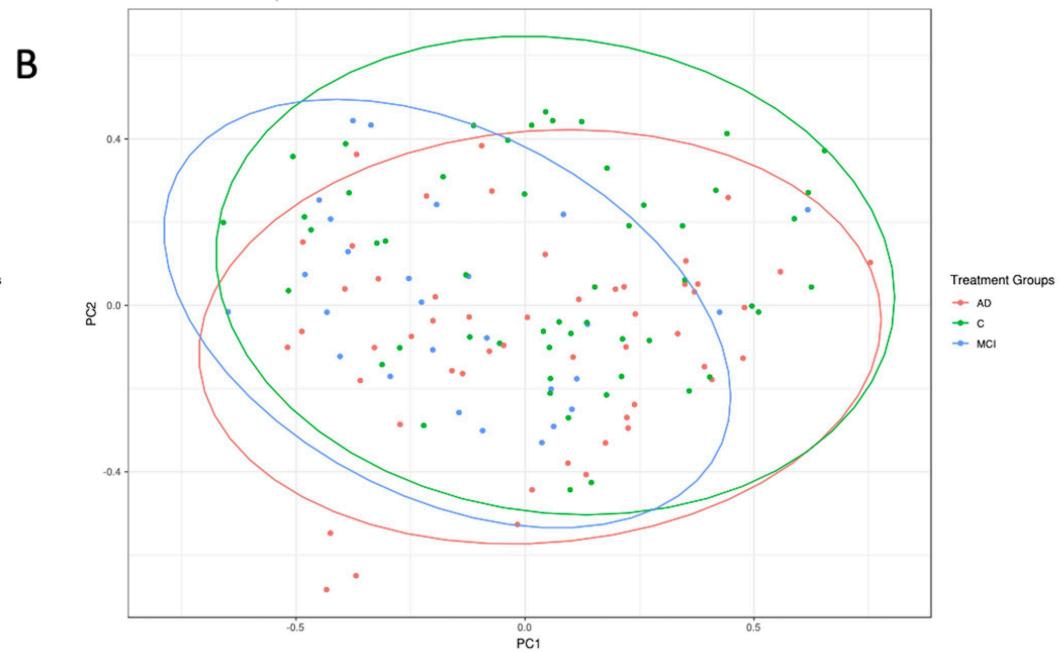
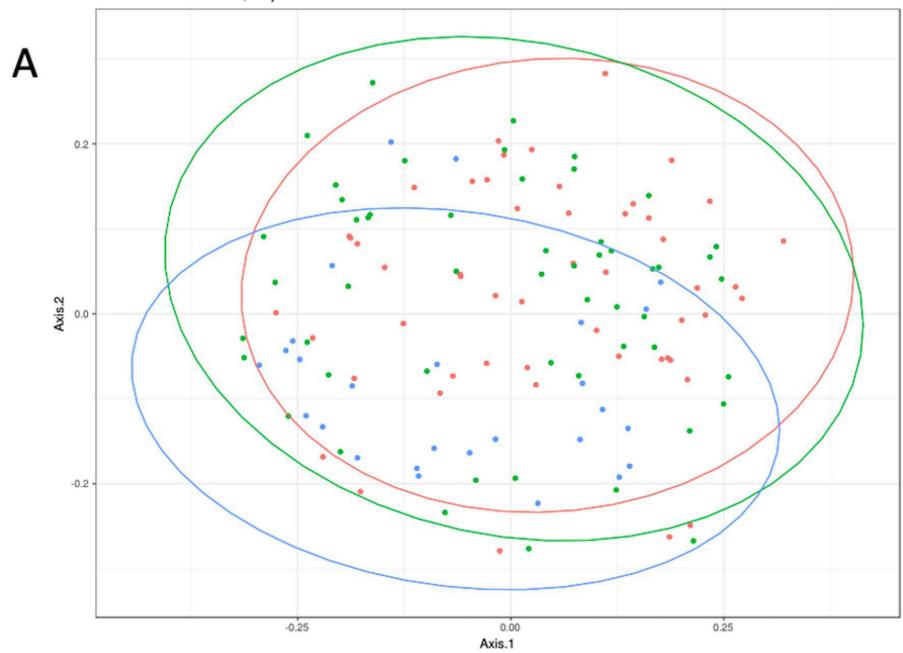
881

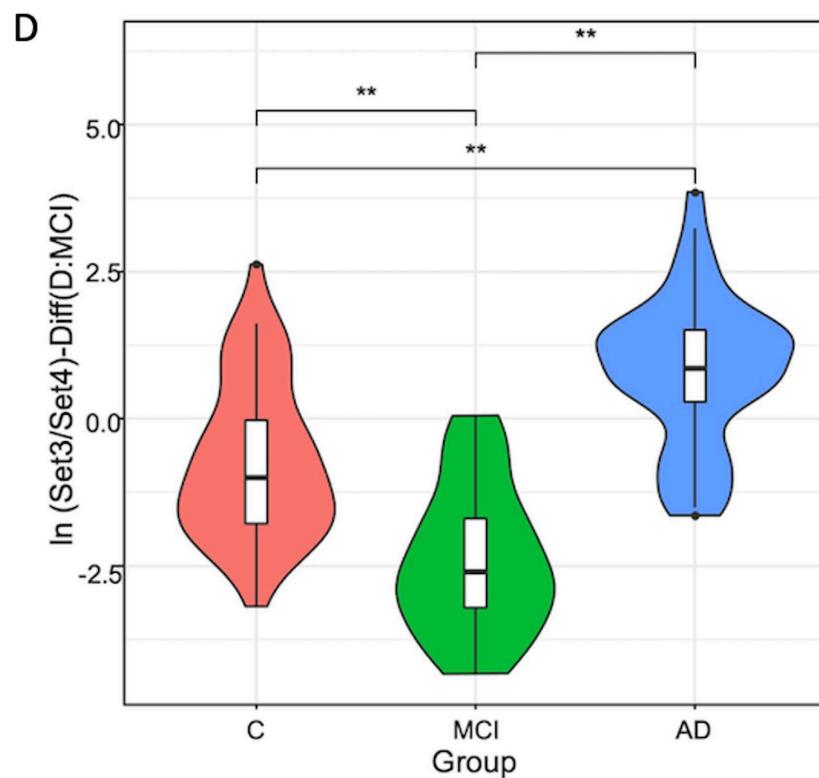
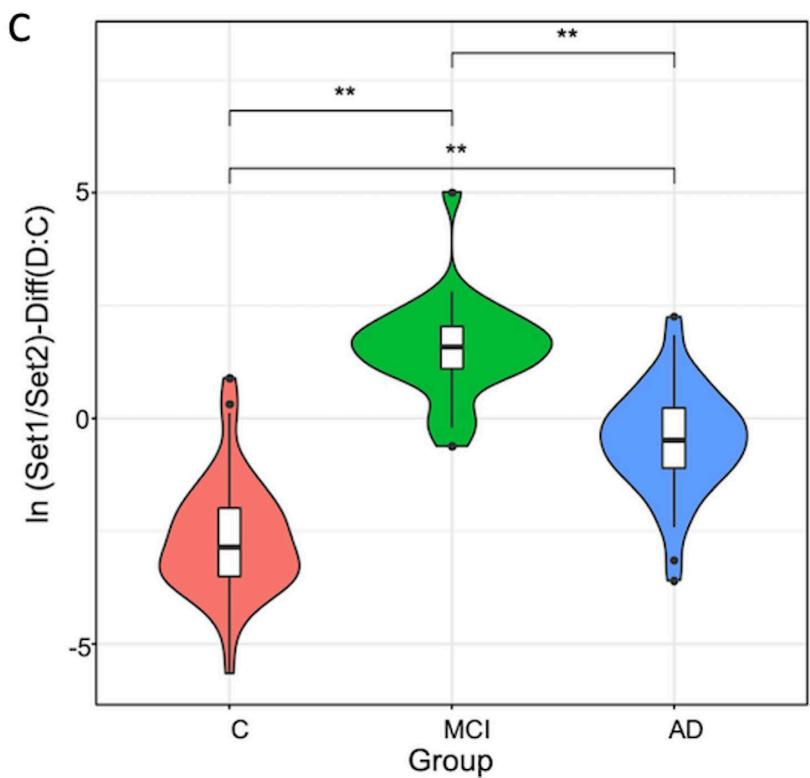
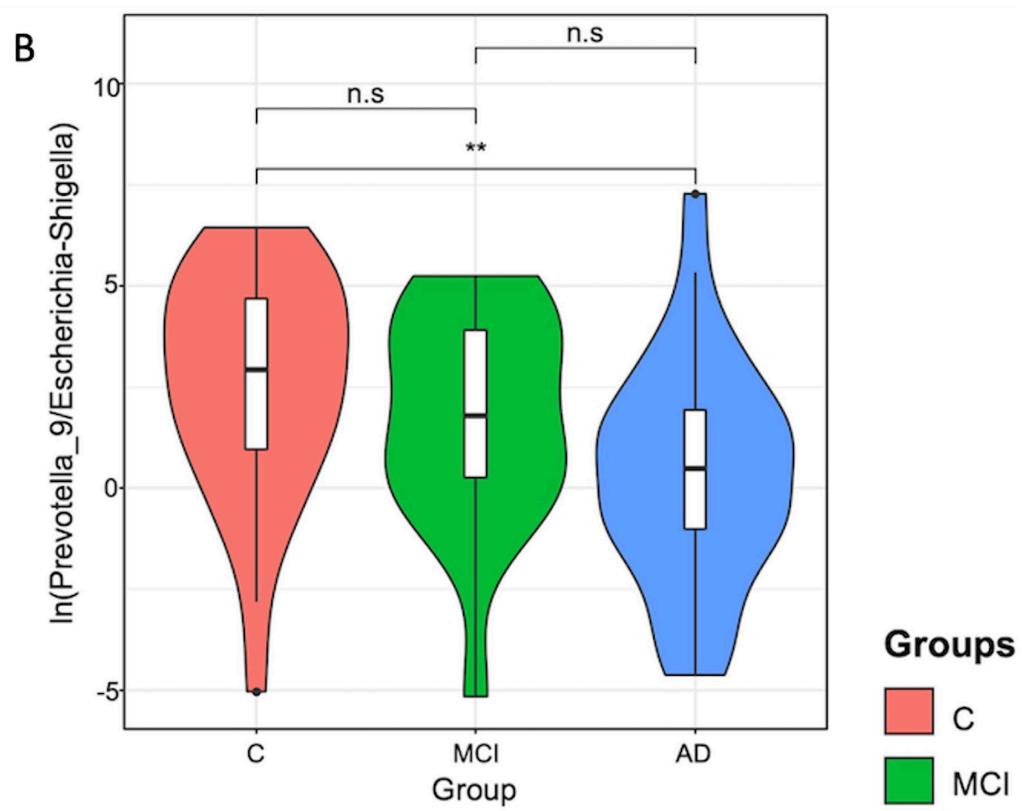
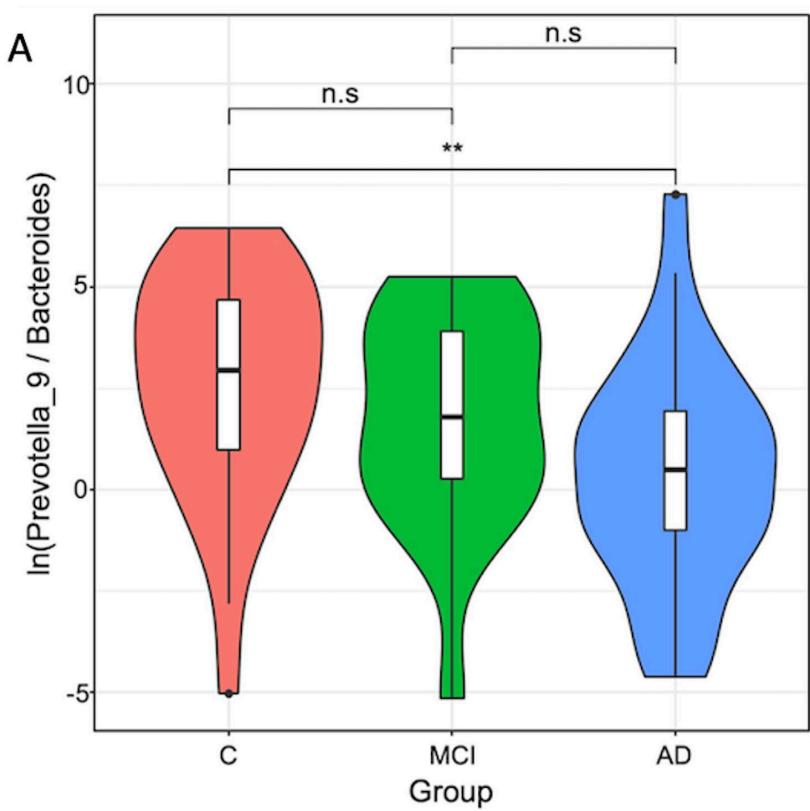
Table 1. Demographic characteristics of the participants in the cohort

	C	MCI	AD
n, (N=125)	51	27	47
Sex (Female, N%)	45% (23/51)	41% (11/27)	49% (23/47)
Age (years, mean \pm SDEV)	67 \pm 5.3	69.2 \pm 6.4	71.4 \pm 5.1
Education (Years)	7.2 \pm 4.1	10.4 \pm 5.2	4.4 \pm 4.1
MMSE	27.1 \pm 1.7	25.4 \pm 2.7	16.9 \pm 5.7
CDR			
0	100%	0%	0
0.5	0	100% (27/27)	29.8% (14/47)
1	0	0	31.9% (15/47)
2	0	0	29.8% (14/47)
3	0	0	8.5% (4/47)
Aβ1-42/P-Tau (pg/mL)	NA	NA	5.97 \pm 3.7 (n=14)
Aβ1-42/T-Tau (pg/mL)	NA	NA	0.91 \pm 0.6 (n=14)
Medications			
AA	NA	37% (10/27)	27.6% (13/47)
ADd	NA	81% (22/27)	87% (41/47)
Adep	NA	66.7% (18/27)	27.6% (13/47)
AE	NA	18.5% (5/27)	8.5% (4/47)
Aht	NA	48% (13/27)	29.8% (14/47)
Apsik	NA	11.1% (3/27)	21.2% (10/47)
Adiab	NA	29.6% (8/27)	19.1% (9/47)
PP	NA	7.4% (2/27)	6.3% (3/47)

C: Control group, MCI: Mild Cognitive Impairment group; AD: Alzheimer's Disease group; MMSE: Mini-Mental State Exam (MMSE); CDR: Clinical Dementia Rating. AA: Antiaggregant; ADd: AD-treatment; Adep: Antidepressant; AE: Antiepileptic; Aht: Antihypertensive; Apsik: Antipsychotic; Adiab: Antidiabetic; PP: Proton-pump inhibitor





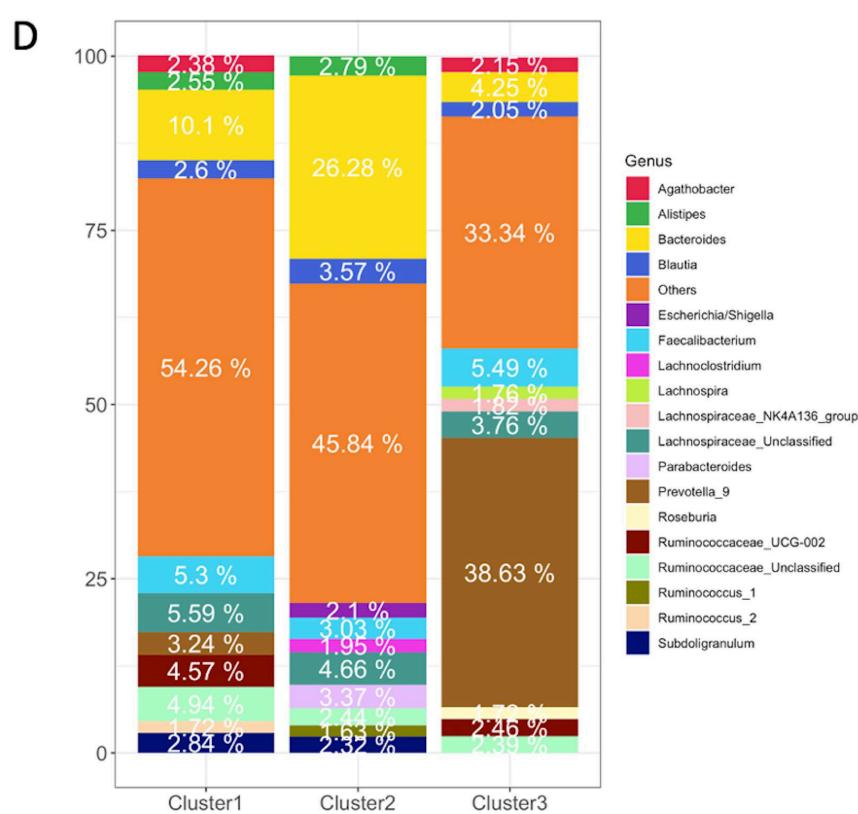
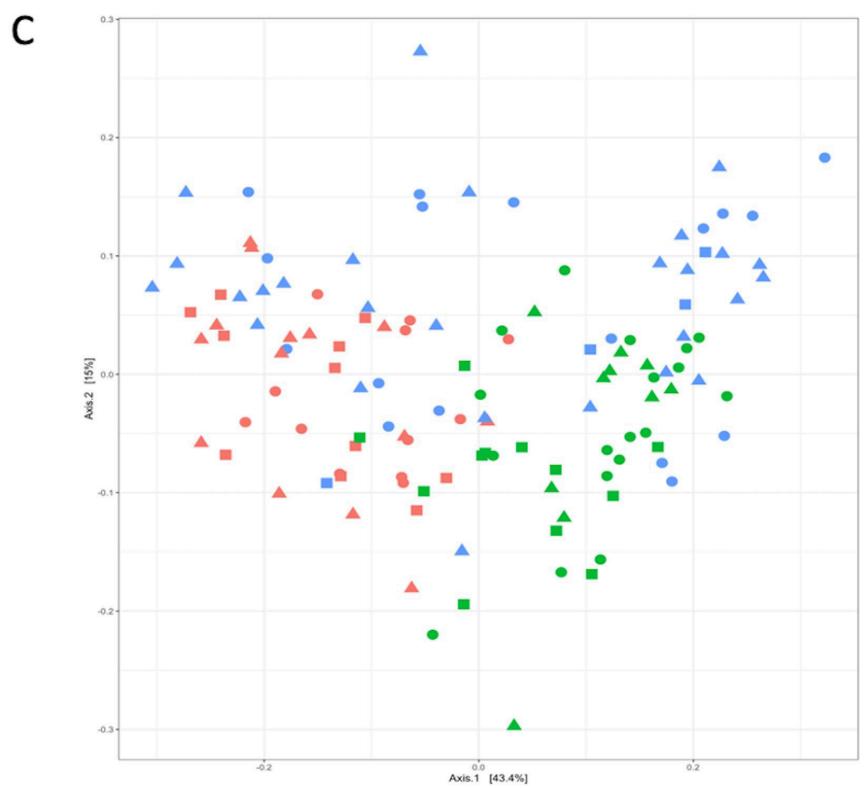
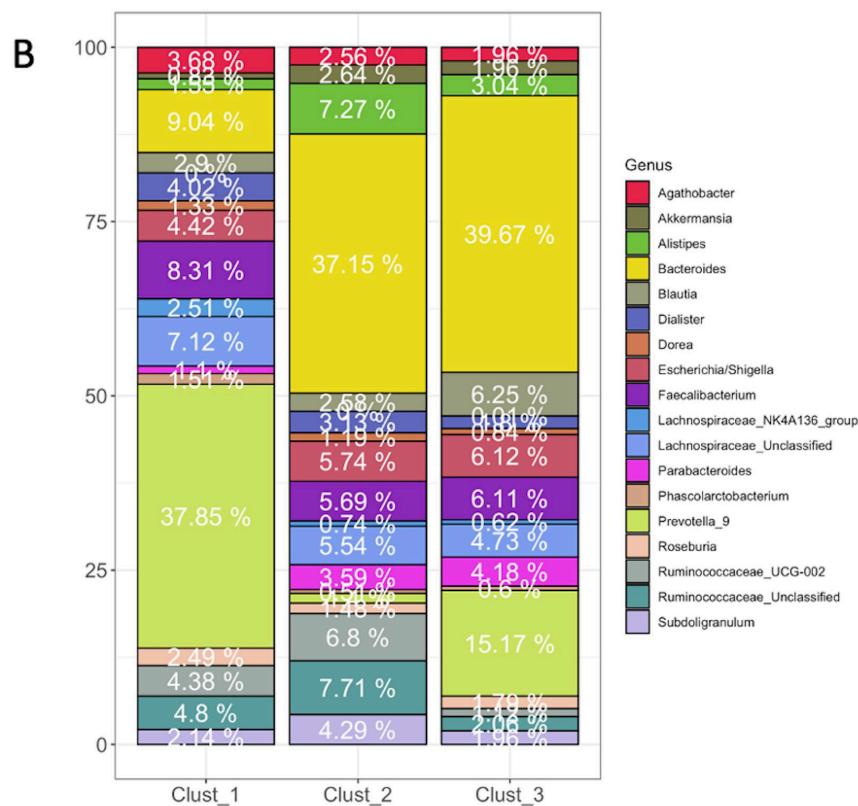
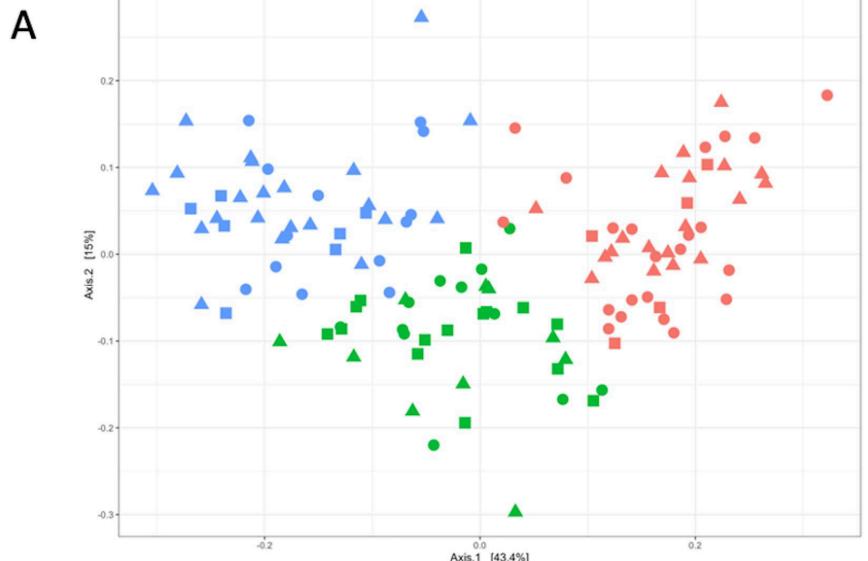


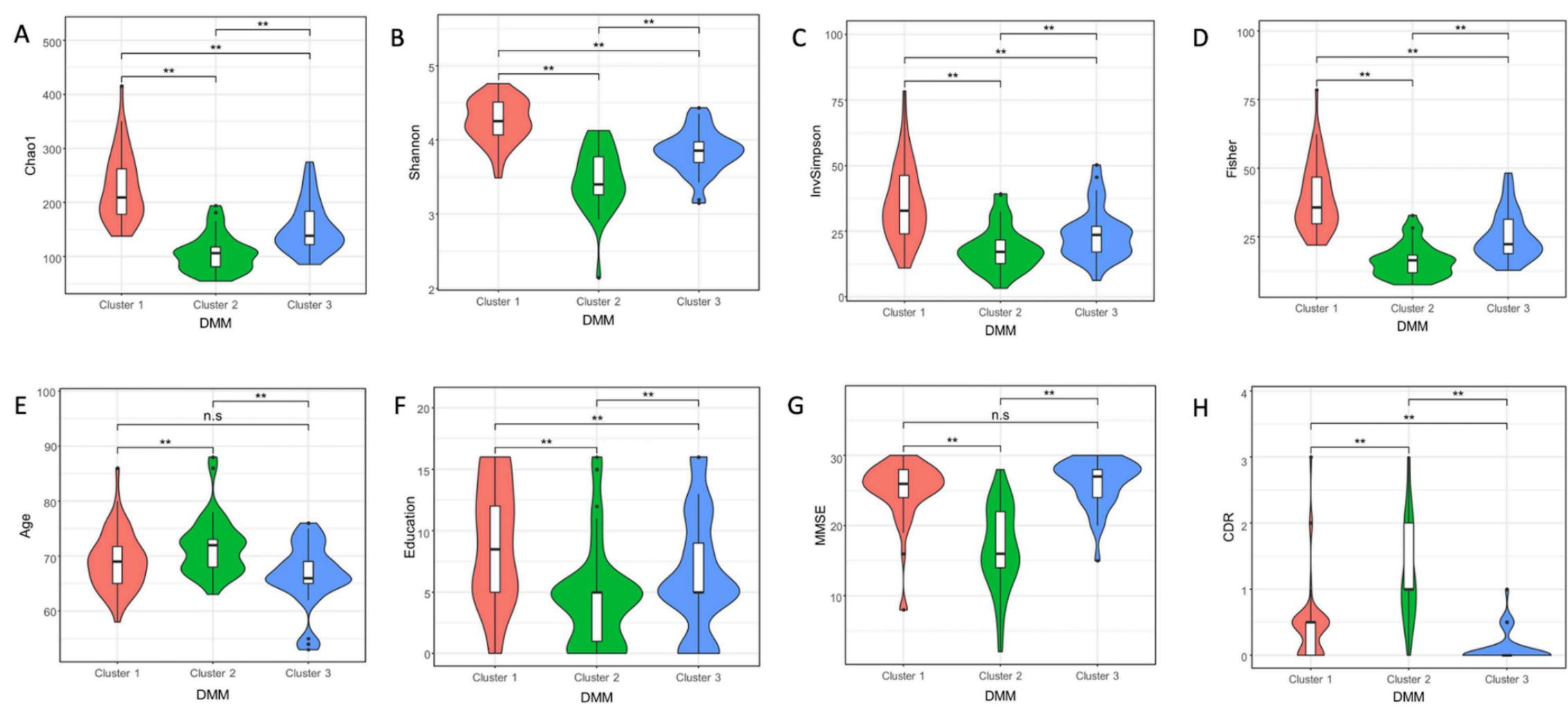
Groups

C

MCI

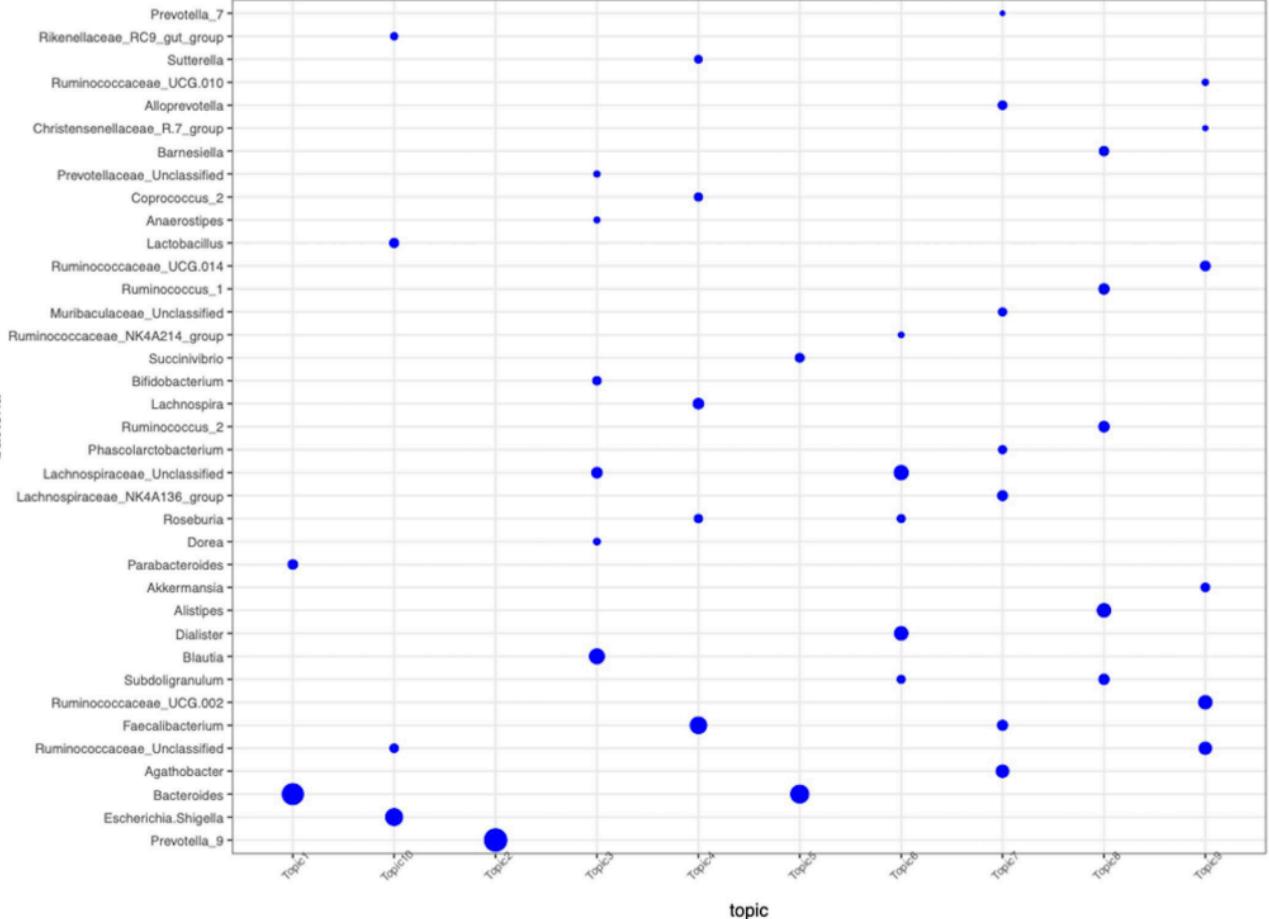
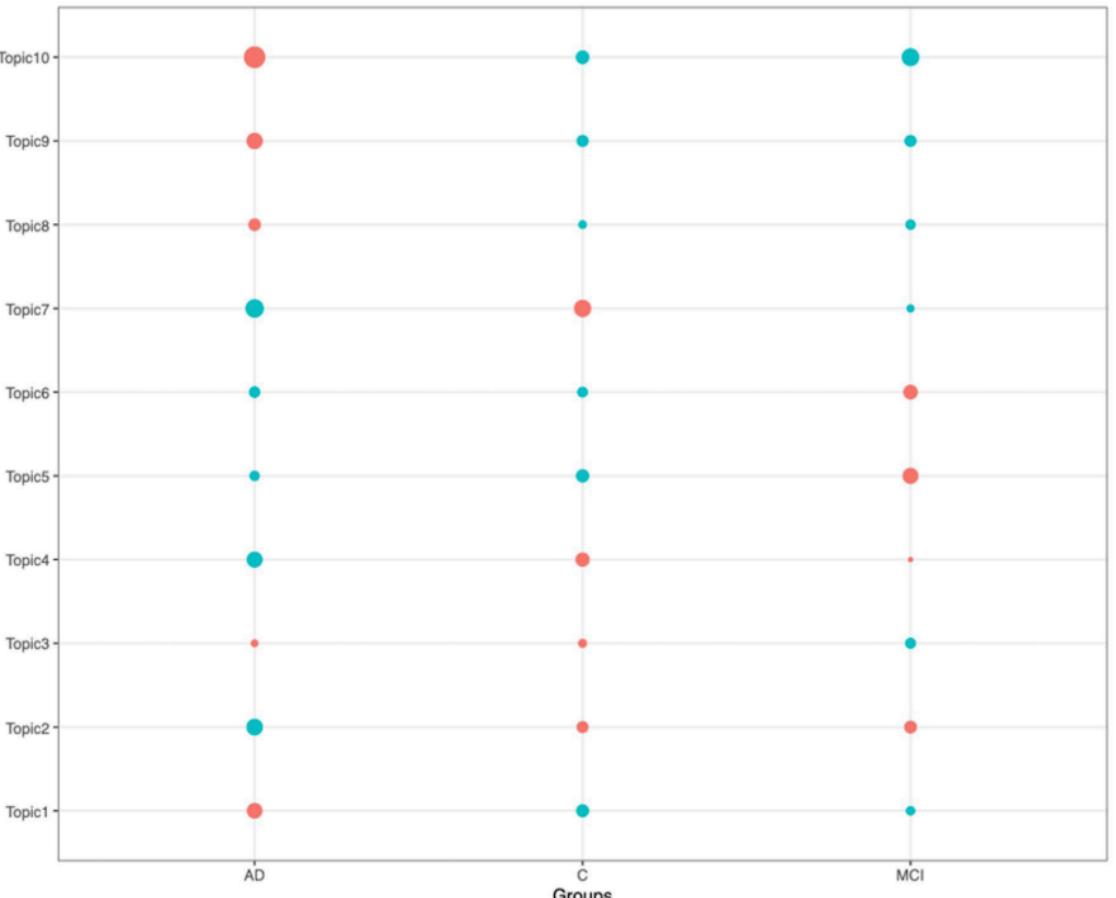
AD

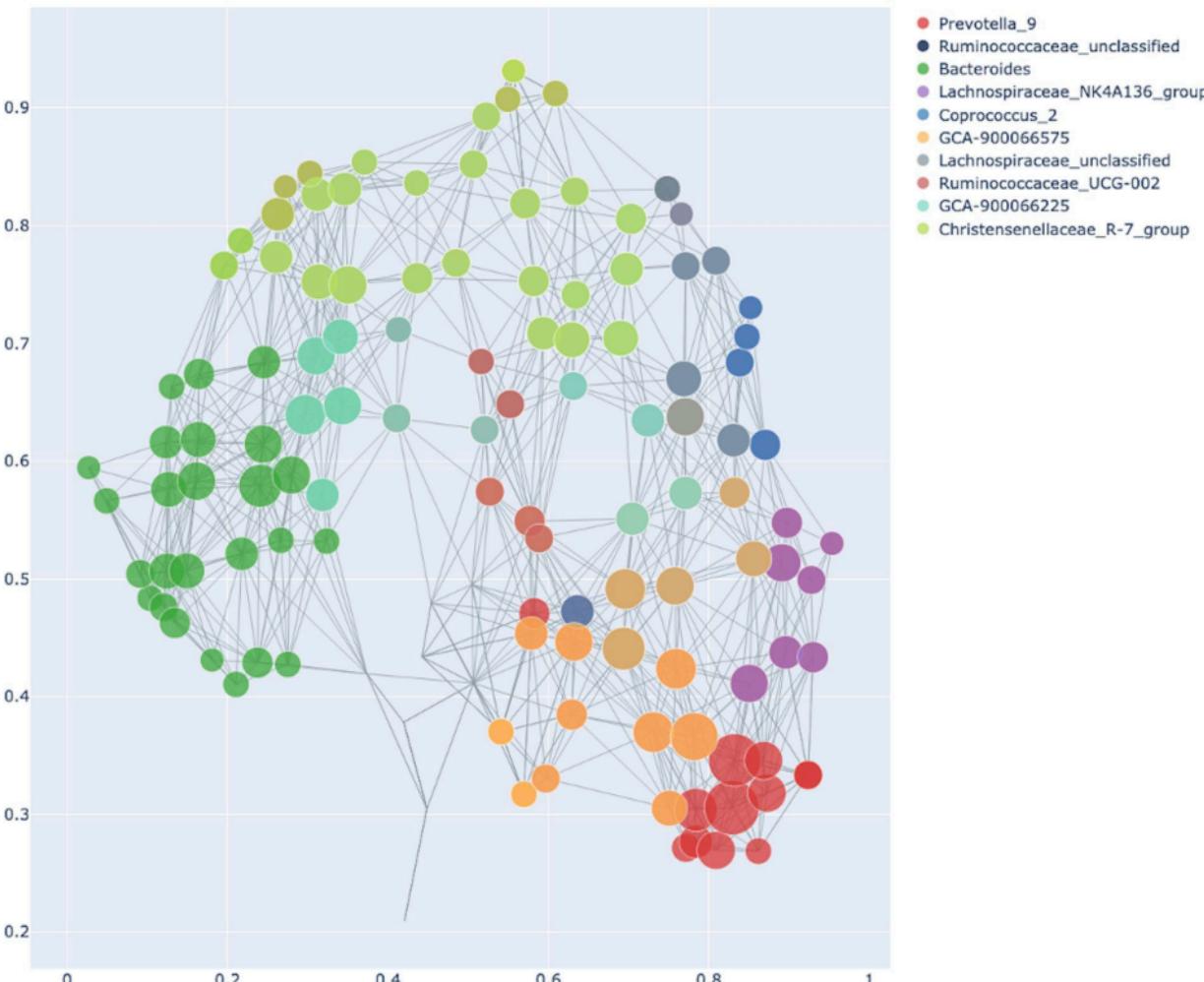
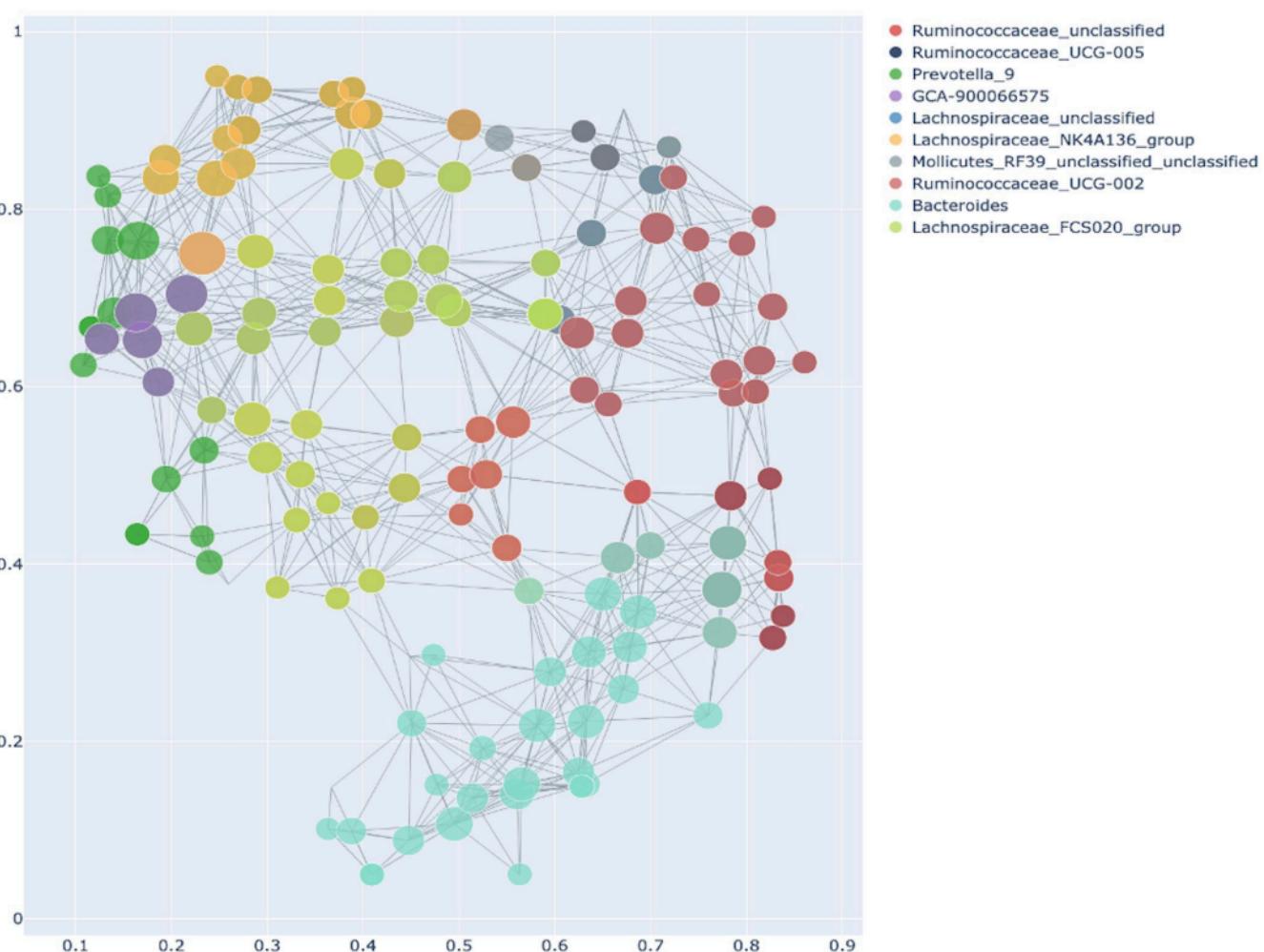


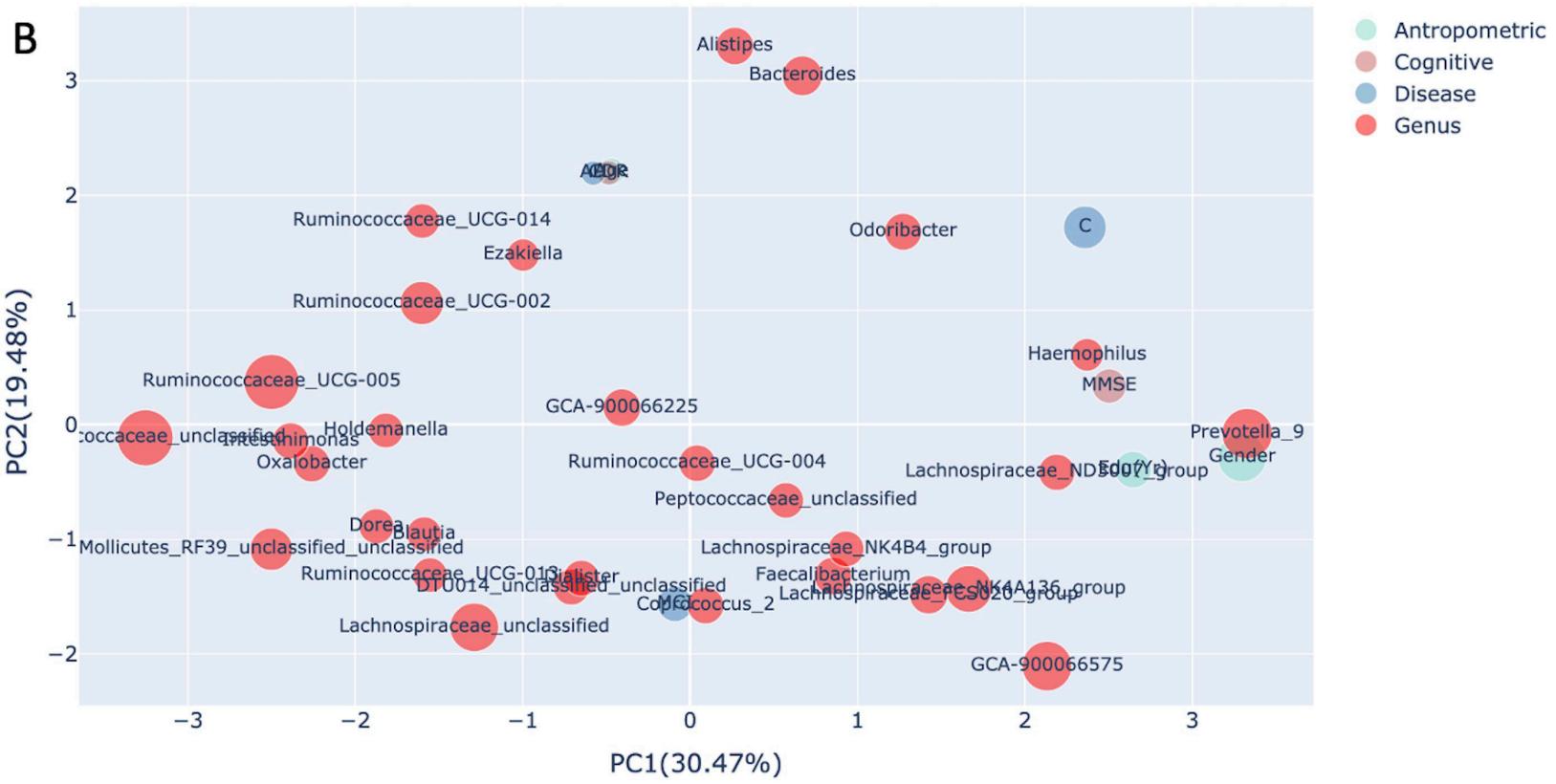
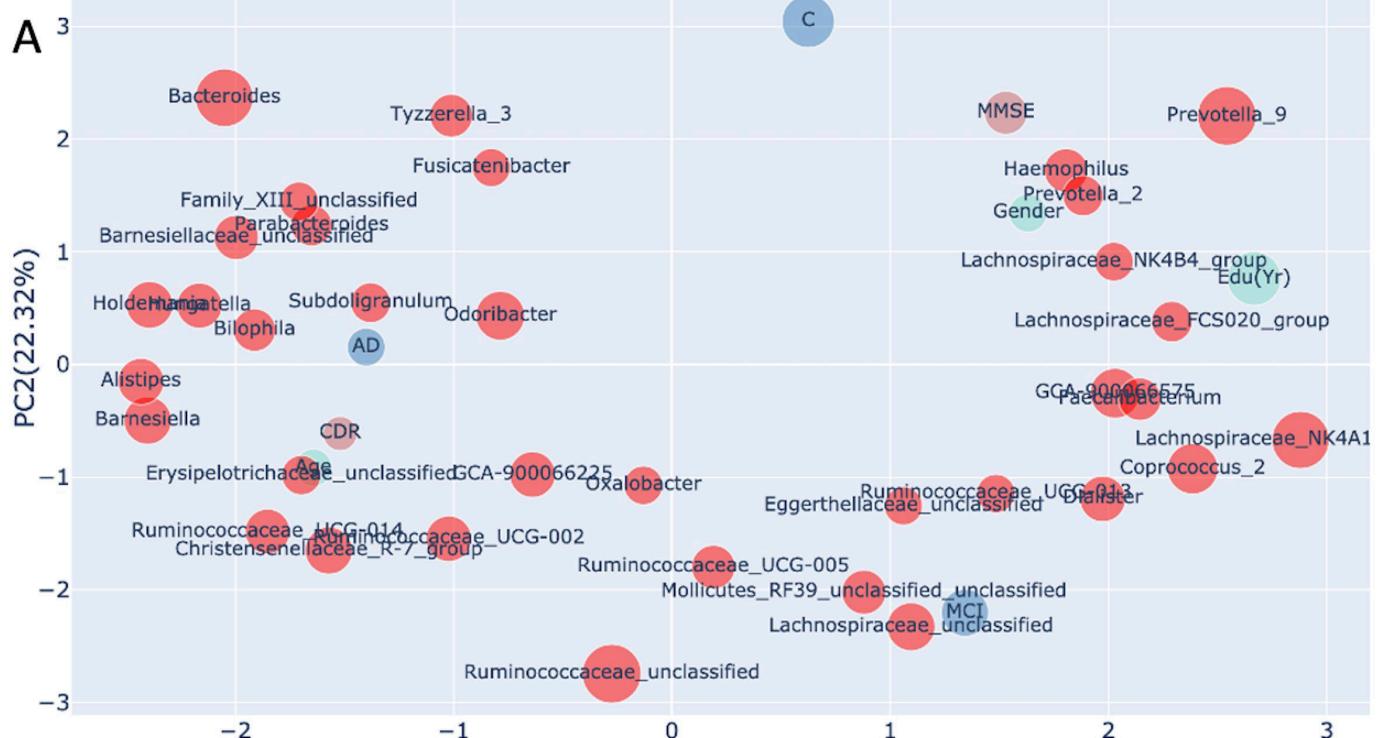


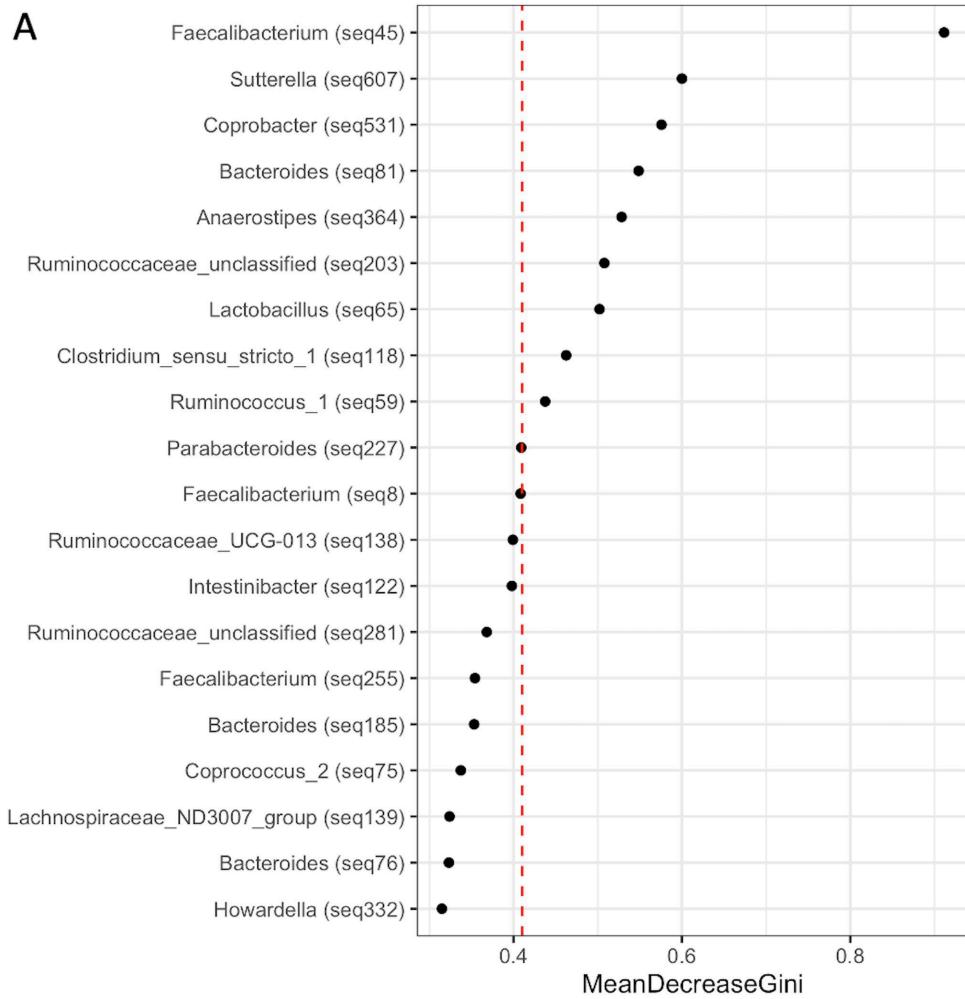
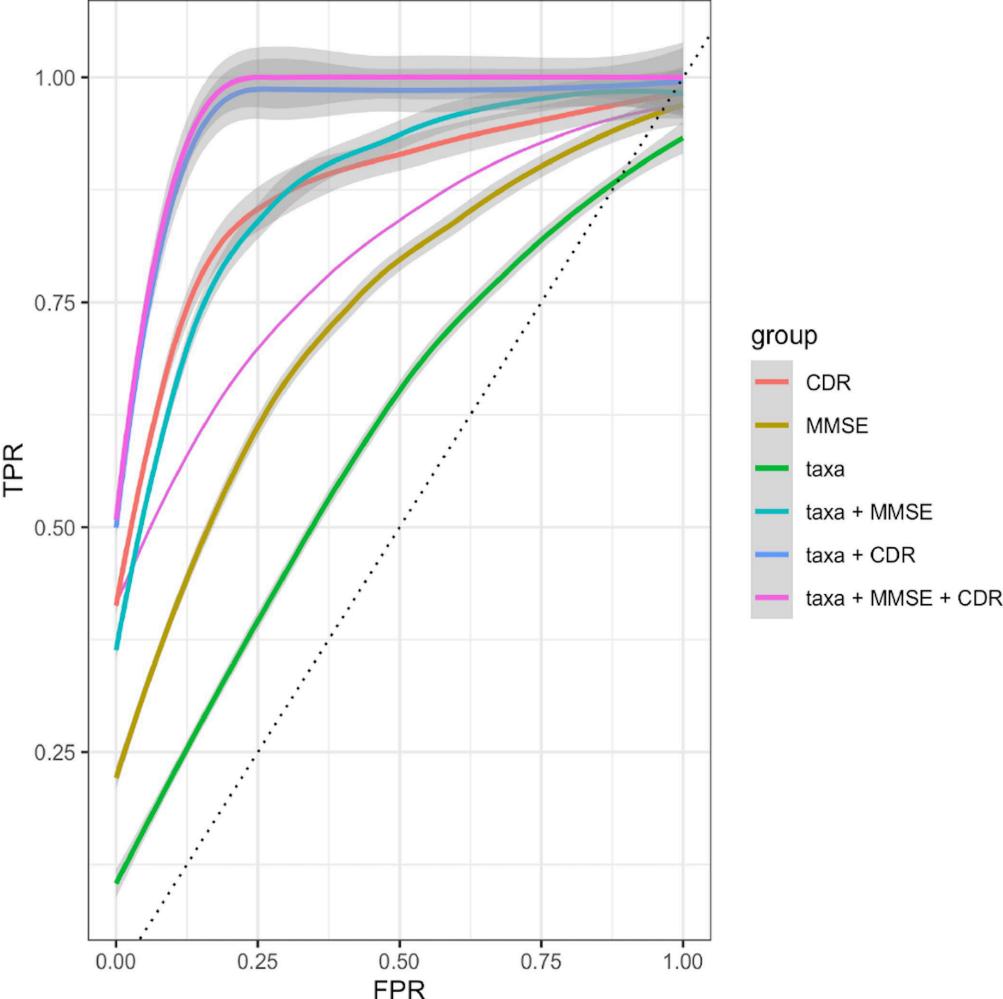
A

topic vs. bacteria probability distribution
(points probability < 0.05 filtered)

**B**

A**B**



A**AUC for ROC****B****group**

- CDR
- MMSE
- taxa
- taxa + MMSE
- taxa + CDR
- taxa + MMSE + CDR