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

Süleyman Yıldırım*#[1,2], Özkan Ufuk Nalbantoğlu*[3,4], Abdulahad Bayraktar*[5], Fatma Betül Ercan[6], Aycan Gündoğdu[4,7], Halil Aziz Velioğlu[2,6], Mehmet Fatih Göl[7], Ayten Ekinci Soylu[7], Fatma Koç[1&], Ezgi Aslan Gürpınar[7], Kübra Sogukkanlı Kadak[6], Muzaffer Arıkan[2], Adil Mardinoğlu[5,8], Mehmet Koçak[9], Emel Köseoğlu[#,10], Lütfü Hanoğlu[#,11]

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

Running title: Gut Microbiota Stratification in Alzheimer's Disease

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

*Equal contribution.

#Corresponding authors:

Email addresses:
Dr. Süleyman Yildirim
suleymanyildirim@medipol.edu.tr

Dr. Emel Koseoglu
emelk@erciyes.edu.tr

Dr. Lutfu Hanoglu
lhanoglu@medipol.edu.tr

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

ABSTRACT

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

IMPORTANCE

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

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

INTRODUCTION

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

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

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

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

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

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

RESULTS

Study Design and Participant Characteristics.

The study cohort consisted of 47 AD, 27 MCI (all amnestic), and 51 subjects non-demented controls (N=125). To minimize dietary confounding effect on the microbiome, we included healthy co-habiting spouses of the patients sharing the same diet as controls. The control group therefore largely (n=27) comprised partners of the patients. Participants were recruited in two health centers located in different cities. The cohort groups were statistically not different in terms of sex, but age and education factors were significantly different (Table 1), therefore statistically adjusted in analyses. Expectedly, the groups were also different in cognitive tests including the Mini-Mental State Exam (MMSE), and the Clinical Dementia rating (CDR). Most AD participants had very mild or mild dementia, with clinical dementia rating (CDR) scores ranging from 0.5–3 (median CDR 1 for AD; 0.5 for MCI and 0 for the control group). The median MMSE scores were significantly higher in control (MMSE=27) and MCI (MMSE=26) groups than AD (MMSE=16). A subset of AD patients (n=12) was clinically asked to undergo lumbar puncture to ascertain diagnosis using CSF biomarkers including Aβ42/Aβ40 ratio, phosphorylated tau (p-tau), and the p-tau/Aβ42 ratio (Supplementary Table S1). We collected medication information from the patient's registry.

Microbiome composition is associated with disease status along the AD continuum

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

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

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

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

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

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

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

Discrete multiple subsets of gut microbiota exist along the AD continuum

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

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

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

We also performed SIMPER analysis based on Bray-Curtis distance to identify taxa contributing most to dissimilarities between clusters (data not shown). *Bacteroides, Prevotella_9, Faecalibacterium,* and taxa within *Ruminococcaceae* family ranked among the top ten taxa contributing most to differences between the three DMM clusters. To examine which factors were associated with the DMM clusters we analyzed distribution of clinical metadata and diversity metrics within the clusters. Alpha diversity indices (Chao1, Shannon, and Inverse Simpson) were statistically different between all three clusters after Benjamini-Hochberg FDR adjustment. 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 ≥ 1% in
293 descending order) across the subgroups are displayed in Figure 6a. Interstingly, 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 *Entercoccus* and *Klebsiella*. Subgroups 4, 6, and 9 were conspicious with the genera known to
300 produce butyrate and acetate or is mucinphilic. 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 topranking 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 occurence 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

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

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

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

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

Identification of signature taxa for AD continuum and association with metadata

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

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

Taxa association with clinical parameters

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

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

DISCUSSION

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

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

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

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

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

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

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

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

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

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

MATERIALS AND METHODS

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

Lumbar puncture, CSF biomarkers assays: Cerebro Spinal Fluid (CSF) samples were included in the analyses from a subset of AD patients if the patient was requested to donate CSF sample as part of the clinically mendated diagnostic protocol. CSF samples were collected in the morning after overnight fasting using spinal needles (22 gauge) and syringes at the L3/4 or L4/5 interspace. CSF was then aliquoted into 0.5 mL non-adsorbing polypropylene tubes and stored at −80 °C until assay. Biomarker molecules in CSF ($A\beta_{42}$, phosphorylated tau (p-tau), and the p-tau/Aβ42 ratio) were measured consistent with the Alzheimer's Association flowchart for lumbar puncture and CSF sample processing and the biomarker levels were determined as previously described [67]. Single 96-well ELISA kits (Innogenetics, Ghent, Belgium) were used in quantitation.

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

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

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

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

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

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

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

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

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

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

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

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 loglikelyhood 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.

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

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

DATA ACCESSİON

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

AUTHOR CONTRIBUTIONS

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

DISCLOSURE DECLARATION

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

LEGENDS FOR SUPPLEMENTAL TABLES AND FIGURES

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

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

Table S6. PERMANOVA analysis of covariates

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

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

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

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

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

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

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

ACKNOWLEDGEMENTS

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

References

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. https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf, Accessed Sep 20, 2020*. 2015.
2. Dubois, B., et al., *Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria.* Alzheimers Dement, 2016. **12**(3): p. 292-323.
3. Aisen, P.S., et al., *On the path to 2025: understanding the Alzheimer's disease continuum.* Alzheimers Res Ther, 2017. **9**(1): p. 60.
4. Kumar, A., A. Singh, and Ekavali, *A review on Alzheimer's disease pathophysiology and its management: an update.* Pharmacol Rep, 2015. **67**(2): p. 195-203.
5. Mehta, D., et al., *Why do trials for Alzheimer's disease drugs keep failing? A discontinued drug perspective for 2010-2015.* Expert Opin Investig Drugs, 2017. **26**(6): p. 735-739.
6. Kametani, F. and M. Hasegawa, *Reconsideration of Amyloid Hypothesis and Tau Hypothesis in Alzheimer's Disease.* Front Neurosci, 2018. **12**: p. 25.
7. Wang, J., et al., *A systemic view of Alzheimer disease - insights from amyloid-beta metabolism beyond the brain.* Nat Rev Neurol, 2017. **13**(10): p. 612-623.
8. Itzhaki, R.F., et al., *Microbes and Alzheimer's Disease.* J Alzheimers Dis, 2016. **51**(4): p. 979-84.
9. Guillot-Sestier, M.V., K.R. Doty, and T. Town, *Innate Immunity Fights Alzheimer's Disease.* Trends Neurosci, 2015. **38**(11): p. 674-681.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

50. Singh, G., M´emoli, F., and Carlsson, G. E., *Topo-logical methods for the analysis of high dimensionaldata sets and 3d object recognition.* SPBG, 2007: p. 91-100.

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

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

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

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

55. De Filippis, F., et al., *Unusual sub-genus associations of faecal Prevotella and 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*

workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement, 2011. **7**(3): p. 263-9.

67. Wallin, A.K., et al., *CSF biomarkers predict a more malignant outcome in Alzheimer disease.* Neurology, 2010. **74**(19): p. 1531-7.

68. Demircan, T., et al., *Experimentally induced metamorphosis in highly regenerative axolotl (ambystoma mexicanum) under constant diet restructures microbiota.* Sci Rep, 2018. **8**(1): p. 10974.

69. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.* EMBnet J, 2011. **17:10–12**.

70. Weber, N., et al., *Nephele: a cloud platform for simplified, standardized and reproducible microbiome data analysis.* Bioinformatics, 2018. **34**(8): p. 1411-1413.

71. Davis, N.M., et al., *Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data.* Microbiome, 2018. **6**(1): p. 226.

72. *R Core Team (2020).R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.* 2020.

73. McMurdie, P.J. and S. Holmes, *phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data.* PLoS One, 2013. **8**(4): p. e61217.

74. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Res, 2015. **43**(7): p. e47.

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

76. Vazquez-Baeza, Y., et al., *EMPeror: a tool for visualizing high-throughput microbial community data.* Gigascience, 2013. **2**(1): p. 16.

77. Himel Mallick, L.J.M., Ali Rahnavard, Siyuan Ma,Yancong Zhang, Long H. Nguyen1, Timothy L. Tickle, George Weingart, Boyu Ren, Emma Schwager, Ayshwarya Subramanian, Yiren Lu, Levi Waldron, Joseph N. Paulson, Eric A. Franzosa, Hector

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

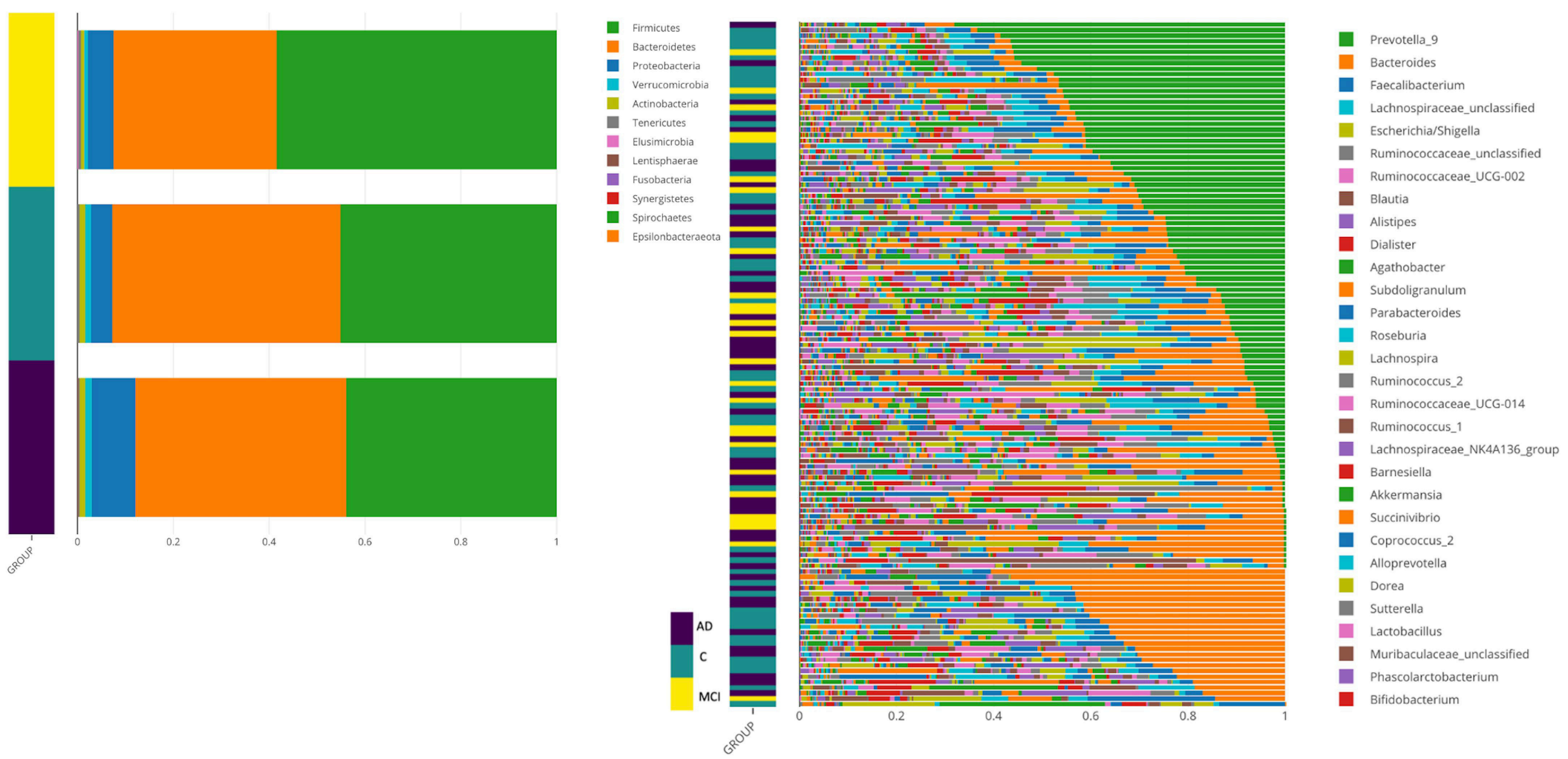
78. Ding, T. and P.D. Schloss, *Dynamics and associations of microbial community types across the human body.* Nature, 2014. **509**(7500): p. 357-60.

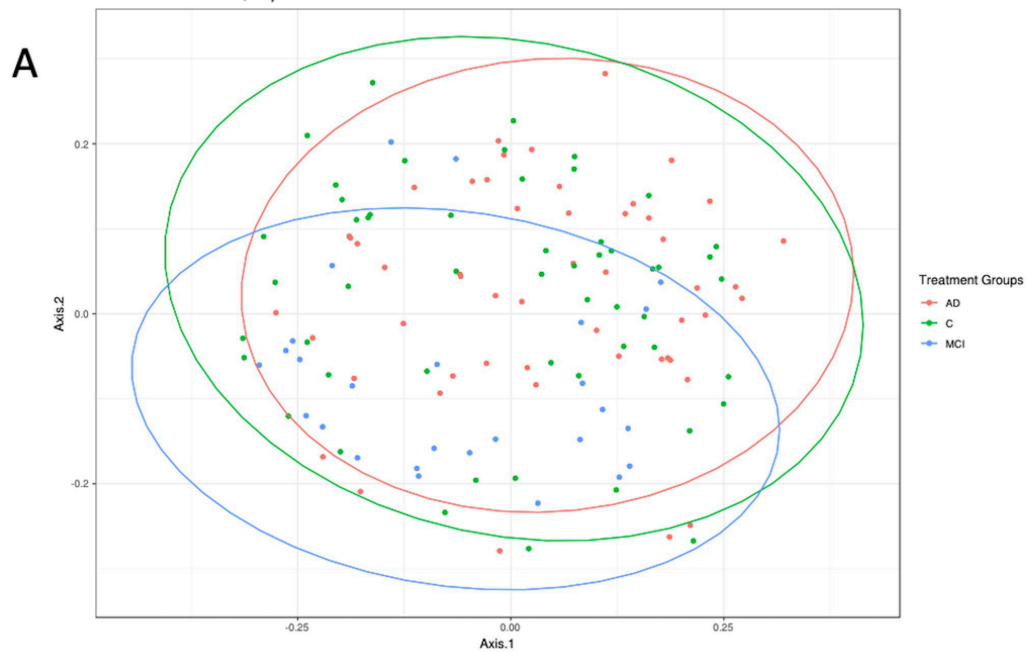
79. Yan, J., et al., *MetaTopics: an integration tool to analyze microbial community profile by topic model.* BMC Genomics, 2017. **18**(Suppl 1): p. 962.

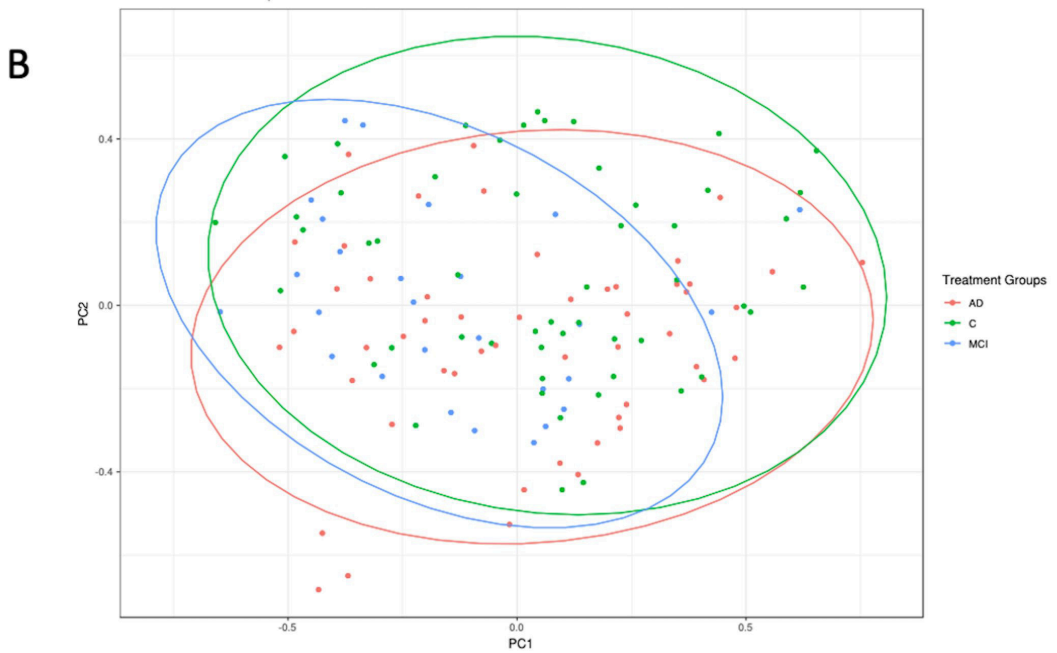
Table 1. Demographic characteristics of the participants in the cohort

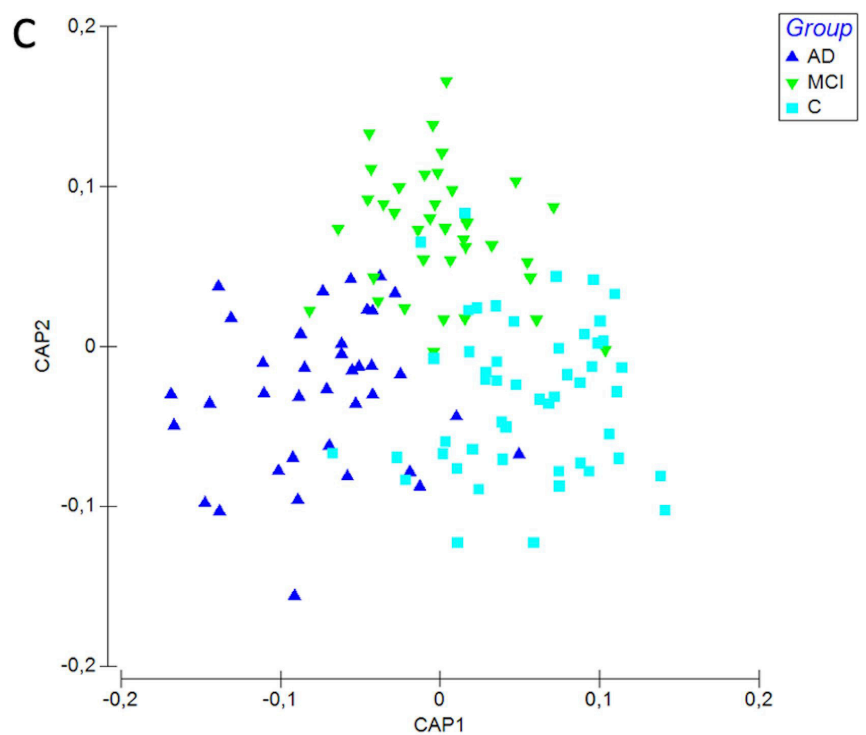
	C	MCI	AD
n, (N=125)	51	27	47
Sex (Female, N%)	45% (23/51)	41% (11/27)	49% (23/47)
Age (years, mean ± SDEV)	67 ± 5.3	69.2 ± 6.4	71.4 ± 5.1
Education (Years)	7.2 ± 4.1	10.4 ± 5.2	4.4 ± 4.1
MMSE	27.1 ± 1.7	25.4 ± 2.7	16.9 ± 5.7
CDR			
0	100%	0%	0
0.5	0	100% (27/27)	29.8% (14/47)
1	0	0	31.9% (15/47)
2	0	0	29.8% (14/47)
3	0	0	8.5% (4/47)
Aβ1–42/P-Tau (pg/mL)	NA	NA	5.97 ± 3.7 (n=14)
Aβ1–42/T-Tau (pg/mL)	NA	NA	0.91 ± 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)

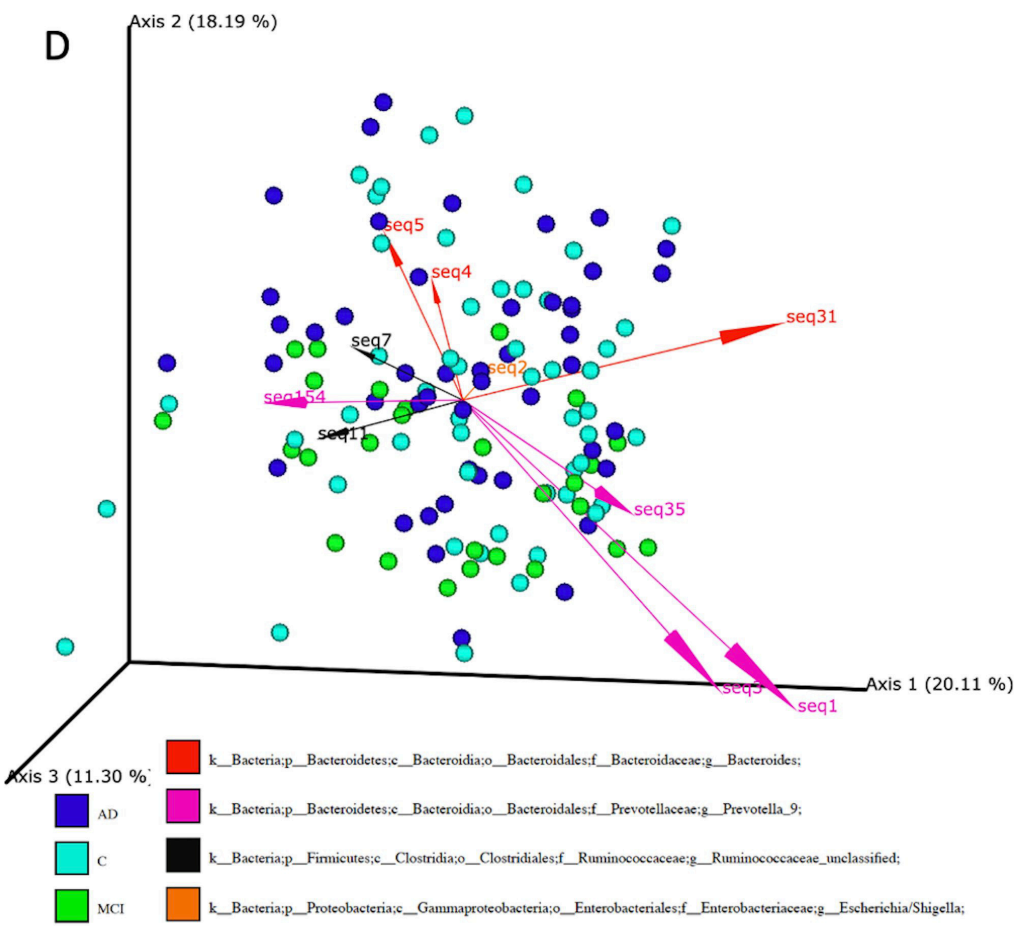
C: Control group, MCI: Mild Cognitive Impairment group; AD: Alzheimer's Disease group; MMSE: Mini-Mental State Exam (MMSE); CDR:Clinical Dementia Rating. AA:Antiaggregant; ADd:AD-treatment; Adep: Antidepressant; AE:Antiepileptic; Aht:Antihypertansive; Apsik: Antipychotic; Adiab:Antidiabetic; PP: Proton-pump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

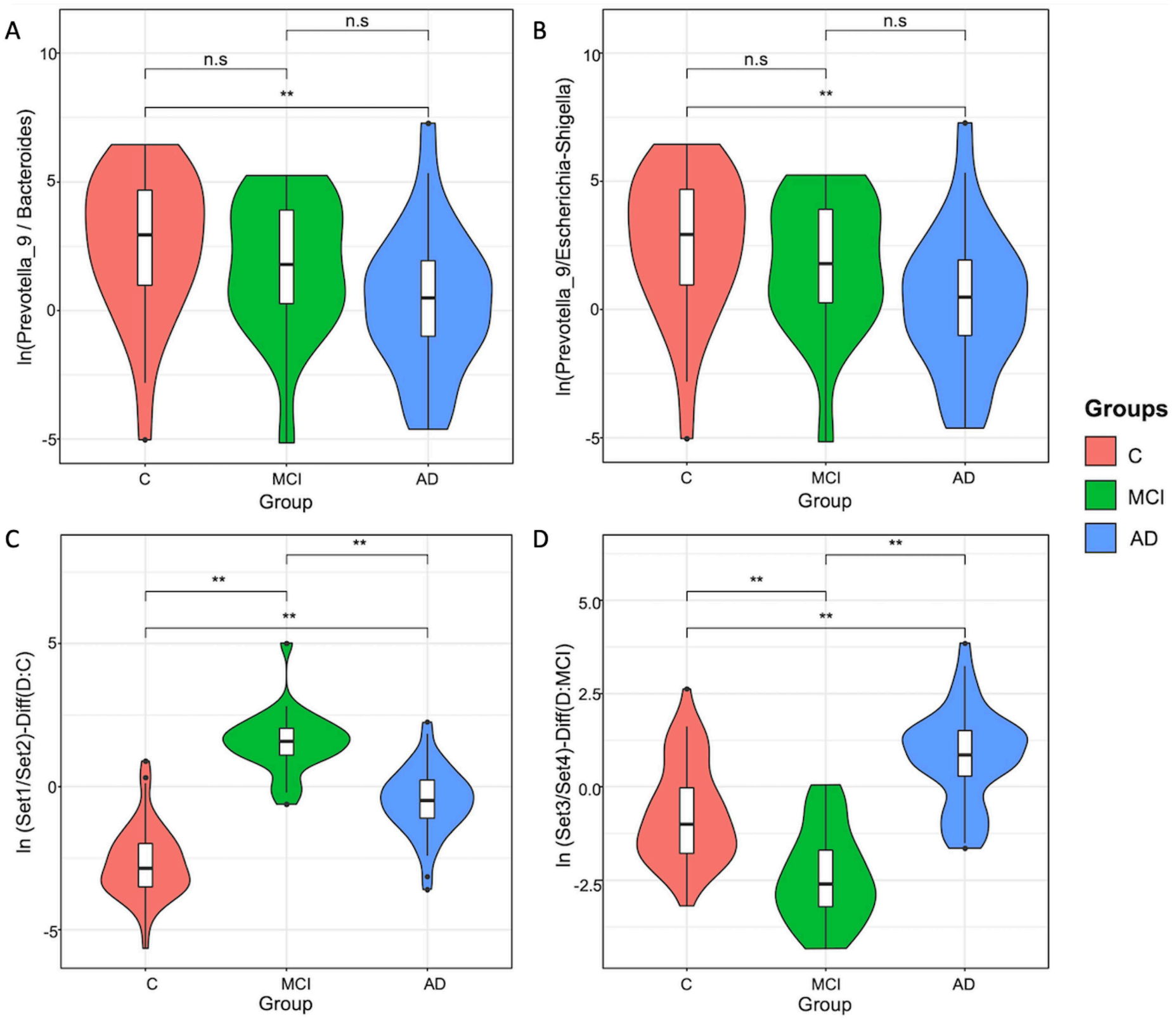

Firmicutes
Bacteroidetes
Proteobacteria
Verrucomicrobia
Actinobacteria
Tenericutes
Elusimicrobia
Lentisphaerae
Fusobacteria
Synergistetes
Spirochaetes
Epsilonbacteraeota
AD
C
MCI
GROUP
0
0.2
0.4
0.6
0.8
1
Prevotella_9
Bacteroides
Faecalibacterium
Lachnospiraceae_unclassified
Escherichia/Shigella
Ruminococcaceae_unclassified
Ruminococcaceae_UCG-002
Blautia
Alistipes
Dialister
Agathobacter
Subdoligranulum
Parabacteroides
Roseburia
Lachnospira
Ruminococcus_2
Ruminococcaceae_UCG-014
Ruminococcus_1
Lachnospiraceae_NK4A136_group
Barnesiella
Akkermansia
Succinivibrio
Coprococcus_2
Alloprevotella
Dorea
Sutterella
Lactobacillus
Muribaculaceae_unclassified
Phascolarctobacterium
Bifidobacterium

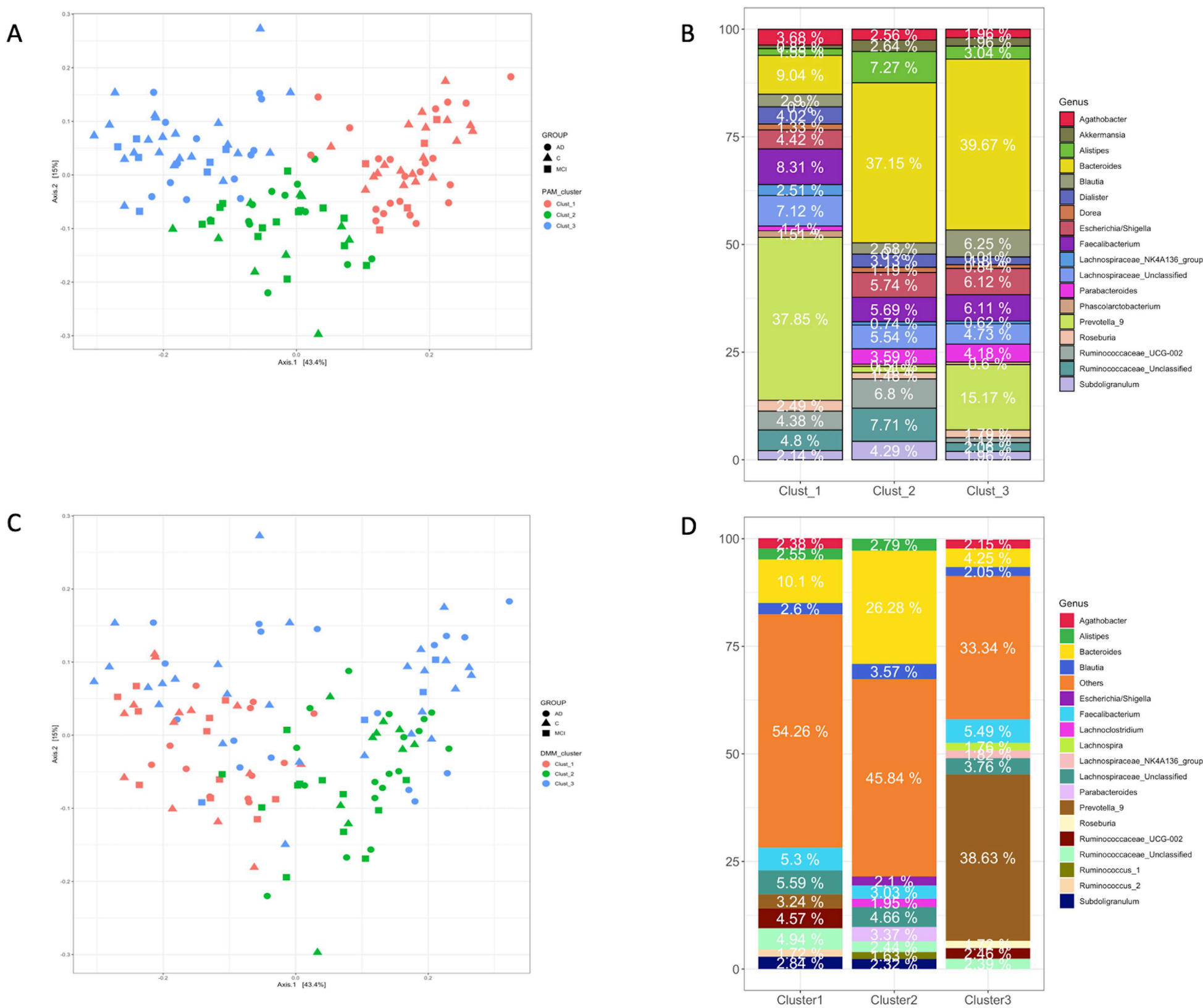

A
Axis.2
Axis.1
0.2
0.0
-0.2
-0.25
0.00
0.25
Treatment Groups
AD
C
MCI

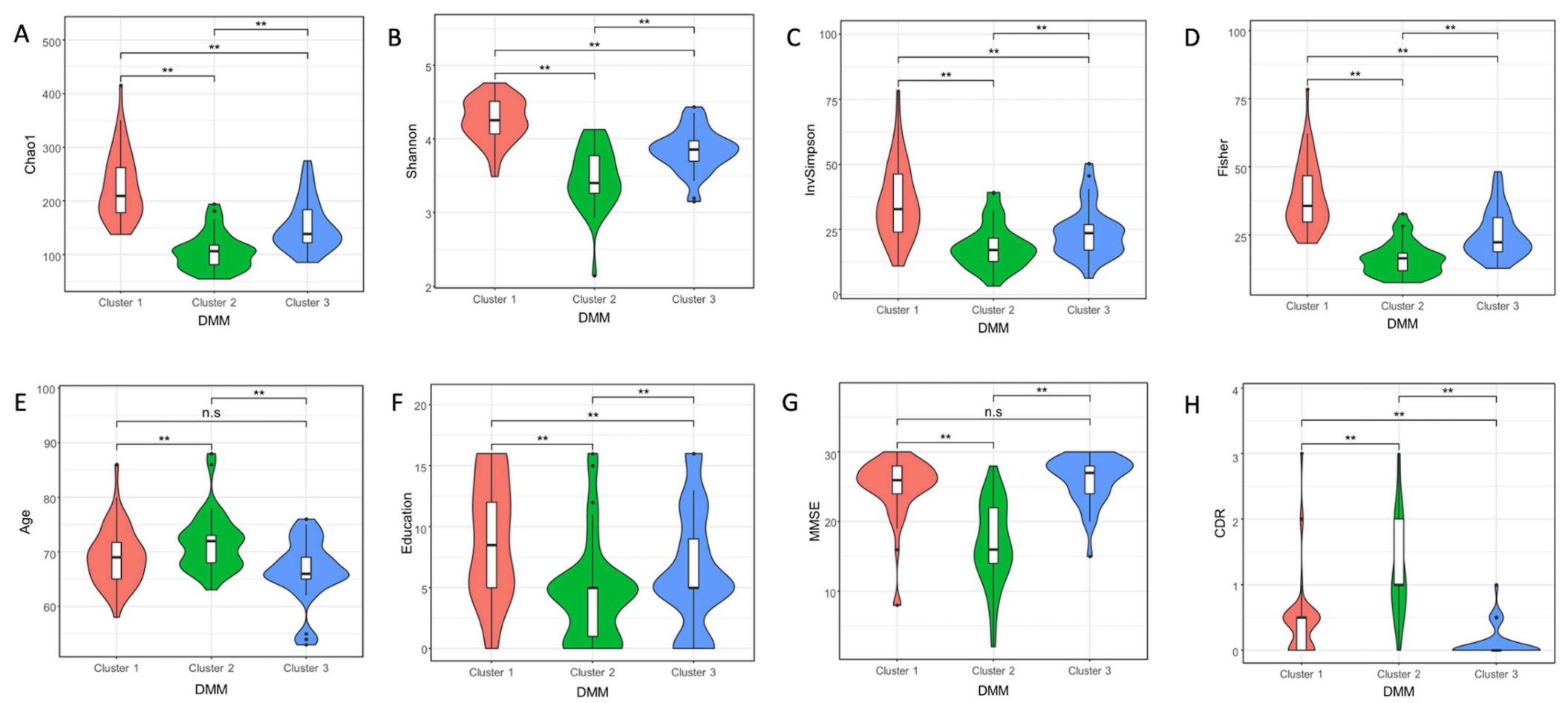

B
PC2
PC1
0.4
0.0
-0.4
-0.5
0.0
0.5
Treatment Groups
AD
C
MCI

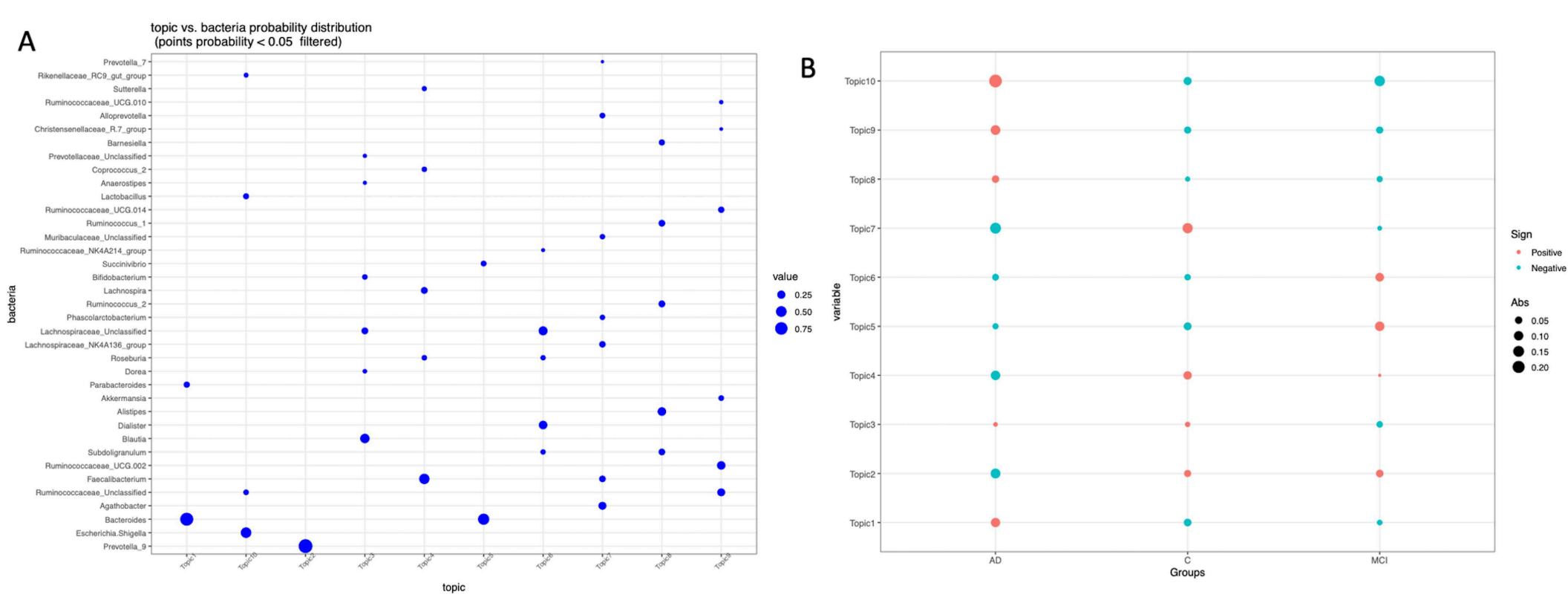

C
Group
AD
MCI
C
CAP2
CAP1
0,2
0,1
0
-0,1
-0,2
-0,2
-0,1
0
0,1
0,2

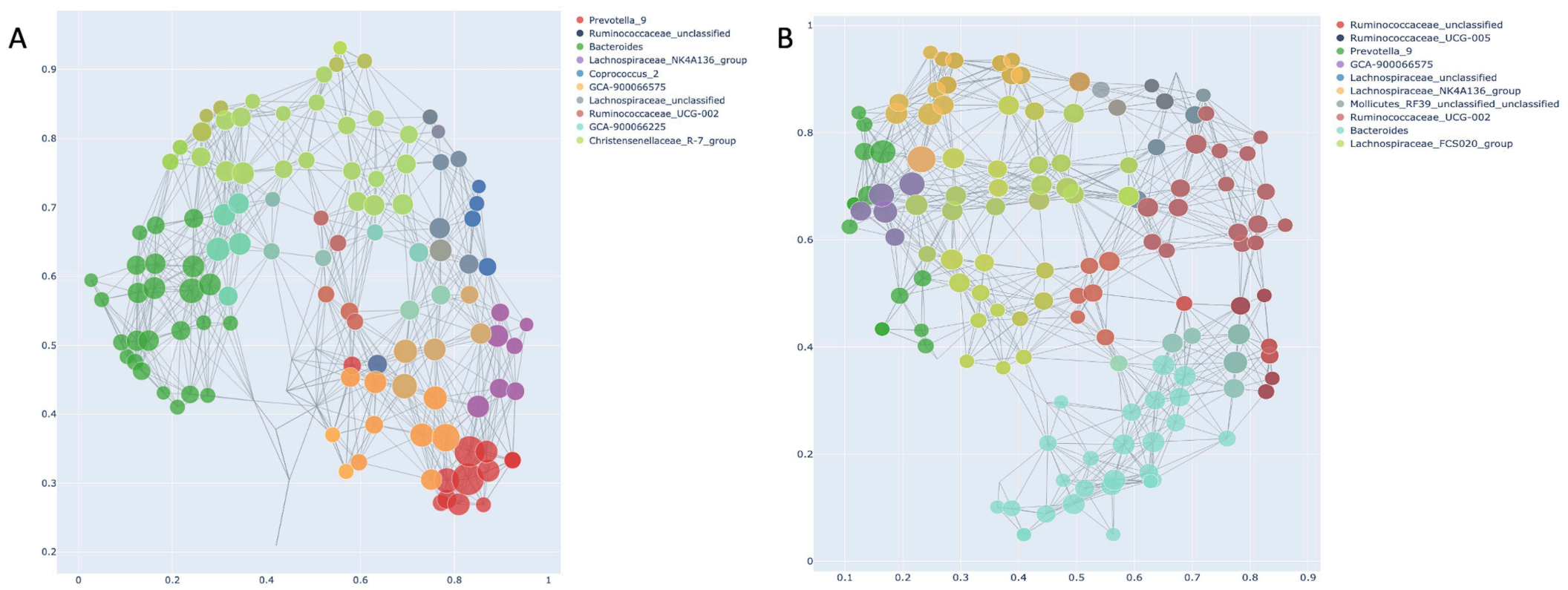

D
Axis 2 (18.19 %)
Axis 1 (20.11 %)
Axis 3 (11.30 %)
seq5
seq4
seq31
seq7
seq2
seq154
seq11
seq35
seq1
AD
C
MCI
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Bacteroidaceae;g__Bacteroides;
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella_9;
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales;f__Ruminococcaceae;g__Ruminococcaceae_unclassified;
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Escherichia/Shigella;

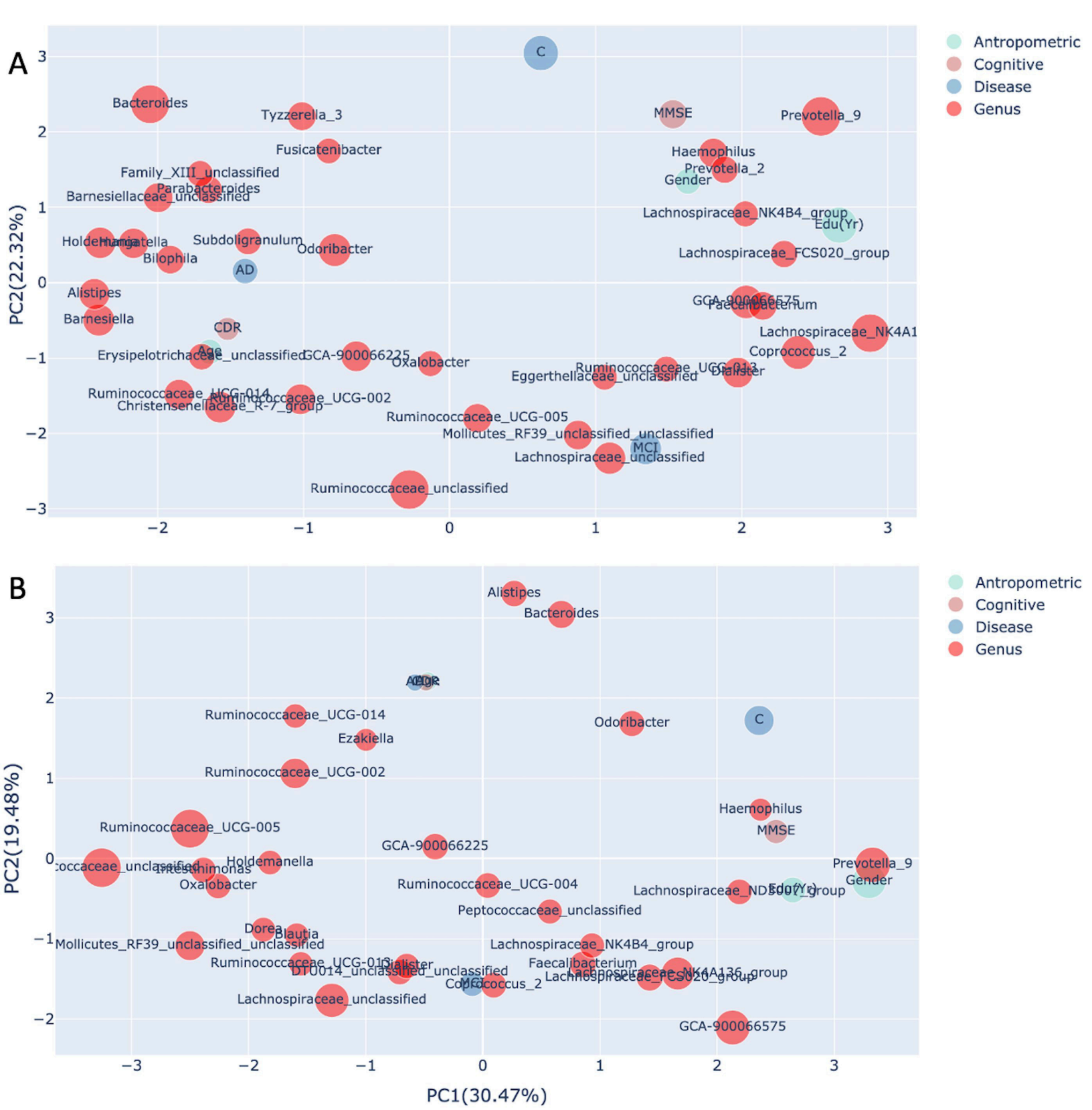

A
B
C
D
n.s
**
ln(Prevotella_9 / Bacteroides)
ln(Prevotella_9/Escherichia-Shigella)
ln (Set1/Set2)-Diff(D:C)
ln (Set3/Set4)-Diff(D:MCI)
C
MCI
AD
Group
Groups

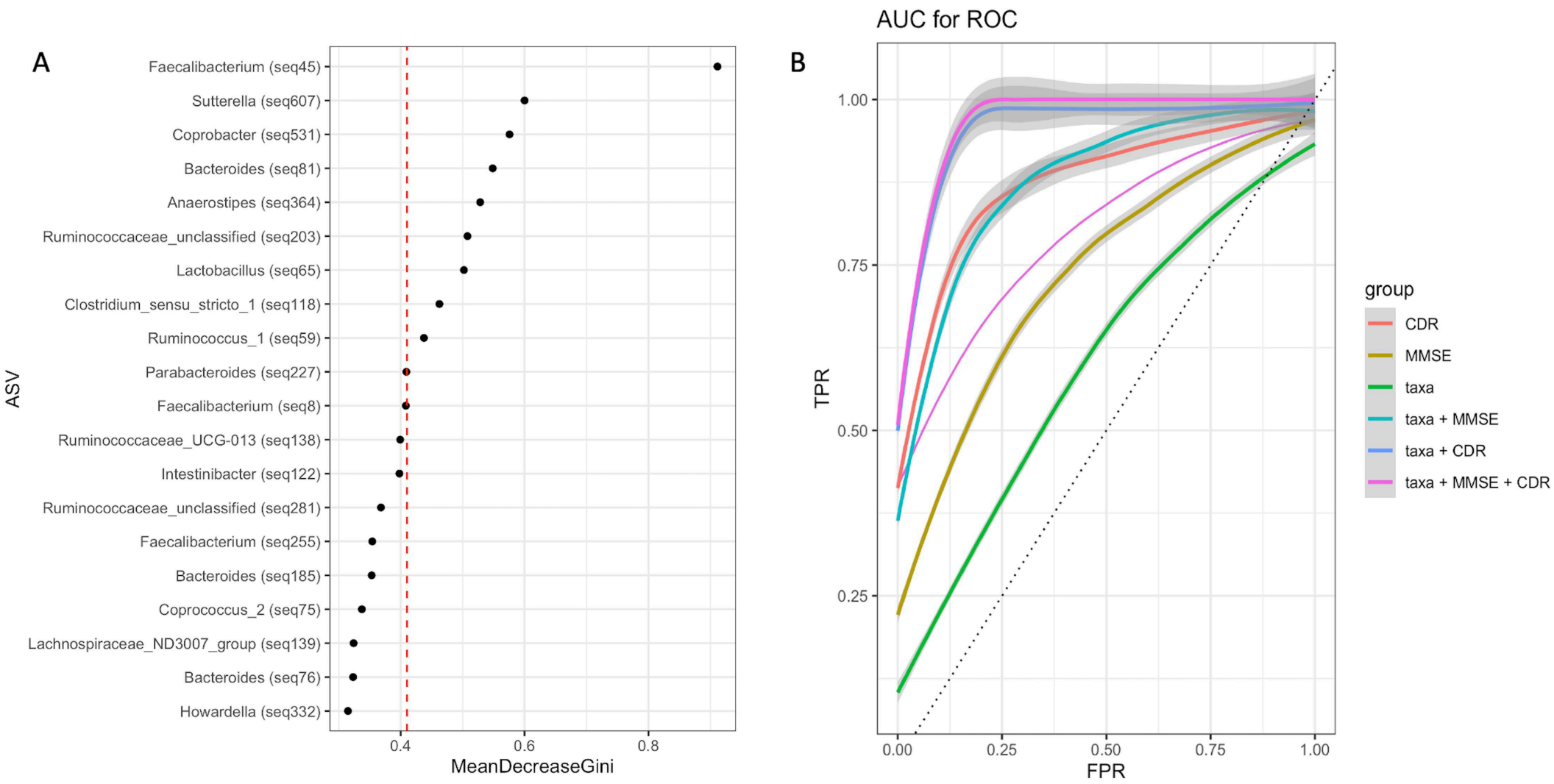

A
Axis.2 [15%]
Axis.1 [43.4%]
GROUP
AD
C
MCI
PAM_cluster
Clust_1
Clust_2
Clust_3
B
Clust_1
Clust_2
Clust_3
Genus
Agathobacter
Akkermansia
Alistipes
Bacteroides
Blautia
Dialister
Dorea
Escherichia/Shigella
Faecalibacterium
Lachnospiraceae_NK4A136_group
Lachnospiraceae_Unclassified
Parabacteroides
Phascolarctobacterium
Prevotella_9
Roseburia
Ruminococcaceae_UCG-002
Ruminococcaceae_Unclassified
Subdoligranulum
C
Axis.2 [15%]
Axis.1 [43.4%]
GROUP
AD
C
MCI
DMM_cluster
Clust_1
Clust_2
Clust_3
D
Cluster1
Cluster2
Cluster3
Genus
Agathobacter
Alistipes
Bacteroides
Blautia
Others
Escherichia/Shigella
Faecalibacterium
Lachnoclostridium
Lachnospira
Lachnospiraceae_NK4A136_group
Lachnospiraceae_Unclassified
Parabacteroides
Prevotella_9
Roseburia
Ruminococcaceae_UCG-002
Ruminococcaceae_Unclassified
Ruminococcus_1
Ruminococcus_2
Subdoligranulum


A
Chao1
500
400
300
200
100
**
**
**
Cluster 1
Cluster 2
Cluster 3
DMM
B
Shannon
5
4
3
2
**
**
**
Cluster 1
Cluster 2
Cluster 3
DMM
C
InvSimpson
100
75
50
25
0
**
**
**
Cluster 1
Cluster 2
Cluster 3
DMM
D
Fisher
100
75
50
25
**
**
**
Cluster 1
Cluster 2
Cluster 3
DMM
E
Age
100
90
80
70
60
n.s
**
**
Cluster 1
Cluster 2
Cluster 3
DMM
F
Education
20
15
10
5
0
**
**
**
Cluster 1
Cluster 2
Cluster 3
DMM
G
MMSE
30
20
10
n.s
**
**
Cluster 1
Cluster 2
Cluster 3
DMM
H
CDR
4
3
2
1
0
**
**
**
Cluster 1
Cluster 2
Cluster 3
DMM


A
topic vs. bacteria probability distribution
(points probability < 0.05 filtered)
Prevotella_7
Rikenellaceae_RC9_gut_group
Sutterella
Ruminococcaceae_UCG.010
Alloprevotella
Christensenellaceae_R.7_group
Barnesiella
Prevotellaceae_Unclassified
Coprococcus_2
Anaerostipes
Lactobacillus
Ruminococcaceae_UCG.014
Ruminococcus_1
Muribaculaceae_Unclassified
Ruminococcaceae_NK4A214_group
Succinivibrio
Bifidobacterium
Lachnospira
Ruminococcus_2
Phascolarctobacterium
Lachnospiraceae_Unclassified
Lachnospiraceae_NK4A136_group
Roseburia
Dorea
Parabacteroides
Akkermansia
Alistipes
Dialister
Blautia
Subdoligranulum
Ruminococcaceae_UCG.002
Faecalibacterium
Ruminococcaceae_Unclassified
Agathobacter
Bacteroides
Escherichia.Shigella
Prevotella_9
bacteria
Topic1
Topic10
Topic2
Topic3
Topic4
Topic5
Topic6
Topic7
Topic8
Topic9
topic
value
0.25
0.50
0.75
B
Topic10
Topic9
Topic8
Topic7
Topic6
Topic5
Topic4
Topic3
Topic2
Topic1
variable
AD
C
MCI
Groups
Sign
Positive
Negative
Abs
0.05
0.10
0.15
0.20


A
Prevotella_9
Ruminococcaceae_unclassified
Bacteroides
Lachnospiraceae_NK4A136_group
Coprococcus_2
GCA-900066575
Lachnospiraceae_unclassified
Ruminococcaceae_UCG-002
GCA-900066225
Christensenellaceae_R-7_group
B
Ruminococcaceae_unclassified
Ruminococcaceae_UCG-005
Prevotella_9
GCA-900066575
Lachnospiraceae_unclassified
Lachnospiraceae_NK4A136_group
Mollicutes_RF39_unclassified_unclassified
Ruminococcaceae_UCG-002
Bacteroides
Lachnospiraceae_FCS020_group


A
Antropometric
Cognitive
Disease
Genus
PC2(22.32%)
C
Bacteroides
Tyzzerella_3
MMSE
Prevotella_9
Fusicatenibacter
Haemophilus
Prevotella_2
Family_XIII_unclassified
Gender
Parabacteroides
Barnesiellaceae_unclassified
Lachnospiraceae_NK4B4_group
Edu(Yr)
Subdoligranulum
Odoribacter
Lachnospiraceae_FCS020_group
Bilophila
AD
Alistipes
Barnesiella
CDR
Lachnospiraceae_NK4A1
Erysipelotrichaceae_unclassified
GCA-900066225
Oxalobacter
Coprococcus_2
Eggerthellaceae_unclassified
Dialister
Ruminococcaceae_UCG-014
Ruminococcaceae_UCG-002
Christensenellaceae_R-7_group
Ruminococcaceae_UCG-005
Mollicutes_RF39_unclassified_unclassified
MCI
Lachnospiraceae_unclassified
Ruminococcaceae_unclassified
B
Antropometric
Cognitive
Disease
Genus
PC2(19.48%)
PC1(30.47%)
Alistipes
Bacteroides
Ruminococcaceae_UCG-014
Odoribacter
C
Ezakiella
Ruminococcaceae_UCG-002
Haemophilus
MMSE
Ruminococcaceae_UCG-005
GCA-900066225
Prevotella_9
Gender
Holdemanella
Oxalobacter
Ruminococcaceae_UCG-004
Peptococcaceae_unclassified
Dorea
Blautia
Mollicutes_RF39_unclassified_unclassified
Lachnospiraceae_NK4B4_group
Faecalibacterium
Dialister
Coprococcus_2
Lachnospiraceae_unclassified
GCA-900066575


A
Faecalibacterium (seq45)
Sutterella (seq607)
Coprobacter (seq531)
Bacteroides (seq81)
Anaerostipes (seq364)
Ruminococcaceae_unclassified (seq203)
Lactobacillus (seq65)
Clostridium_sensu_stricto_1 (seq118)
Ruminococcus_1 (seq59)
Parabacteroides (seq227)
Faecalibacterium (seq8)
Ruminococcaceae_UCG-013 (seq138)
Intestinibacter (seq122)
Ruminococcaceae_unclassified (seq281)
Faecalibacterium (seq255)
Bacteroides (seq185)
Coprococcus_2 (seq75)
Lachnospiraceae_ND3007_group (seq139)
Bacteroides (seq76)
Howardella (seq332)
ASV
0.4
0.6
0.8
MeanDecreaseGini
B
AUC for ROC
1.00
0.75
0.50
0.25
TPR
0.00
0.25
0.50
0.75
1.00
FPR
group
CDR
MMSE
taxa
taxa + MMSE
taxa + CDR
taxa + MMSE + CDR