

Genetic correlations between gut microbiota genera and Alzheimer's Disease

Davis Cammann^{1*}, Yimei Lu^{1*}, Melika J. Cummings¹, Mark L. Zhang², Joan Manual Cue¹, Jenifer Do¹, Jeffrey Ebersole³, Xiangning Chen⁴, Edwin C. Oh^{1,5,6}, Jeffrey L. Cummings⁷, and Jingchun Chen^{1}**

¹Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, Las Vegas, NV 89154, USA.

²Columbia University, West 116 St and Broadway, New York, NY 10027, USA.

³Department of Biomedical Sciences, University of Nevada, Las Vegas, NV 89154, USA.

⁴Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA.

⁵Laboratory of Neurogenetics and Precision Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA.

⁶Department of Internal Medicine, UNLV School of Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA.

⁷Department of Brain Health, School of Integrated Health Sciences, University of Nevada Las Vegas, Las Vegas, NV USA.

*: These two authors contributed equally to this work.

** : corresponding author: Jingchun.chen@unlv.edu; Tel.: 1-702-895-1196

1. ABSTRACT

Background: Dysbiosis of the human gut microbiota has been implicated in the development of Alzheimer’s disease (AD). However, the genetic correlation between gut microbiota and AD is not well addressed.

Methods: Using the largest genome-wide association study (GWAS) of gut microbiota genera from the MiBioGen consortium, we conducted the “best-fit” model from PRSice-2 to determine the genetic correlation between 119 genera and AD in a discovery sample (case/control: 1,278/1,293); we then replicated our findings in an independent sample (case/control: 799/778) and further performed meta-analyses to confirm the correlation. Finally, we conducted a linear regression to assess the correlation between the PRSs for the significant genera and the *APOE* genotype.

Results: In the discovery sample, 20 gut microbiota genera were initially identified genetically associated with AD. Three genera (*Eubacterium fissicatena* as a protective factor, *Collinsella* and *Veillonella* as a risk factor) were validated in the replication sample. Meta-analysis confirmed nine genera to have a significant correlation with AD, three of which were significantly associated with the *APOE* rs429358 risk allele in a direction consistent with their protective/risk designation in AD association. Notably, the proinflammatory genus *Collinsella*, identified as a risk factor for AD, was positively correlated with the *APOE* rs429358 risk allele in both samples.

Conclusion: Host genetic factors influencing the abundance of nine genera are significantly associated with AD, suggesting that these genera may serve as biomarkers and targets for AD treatment and intervention. Our results highlight that proinflammatory gut microbiota might promote AD development through interaction with *APOE*. Larger datasets and functional studies are required to understand their causal relationships.

Keywords: Alzheimer’s disease; microbiome; polygenic risk scores (PRSs); genome-wide association studies (GWASs); neuroinflammation; *APOE*; dysbiosis

2. INTRODUCTION

Alzheimer's disease (AD), the most common form of dementia, is a neurodegenerative disorder characterized by a multitude of pathological and clinical hallmarks such as a progressive decline in cognitive function and the buildup of toxic β -amyloid and tau proteins (1,2). Due to growing elderly populations worldwide, the number of individuals with dementia is projected to reach 150 million globally by the year 2050 (3). Despite this growing burden to world health, the mechanisms underlying the disease pathology are not fully understood, impeding the development of optimally effective treatments (4). Neuroinflammation has emerged as a key feature of AD with mechanistic and treatment implications due to the central role of microglia and inflammation in brain health (5,6). There remains an urgent need to understand the genetic risk factors and pathological basis of neuroinflammation in AD so that individuals with a higher risk can be identified for earlier intervention.

Recently, an association between dysbiosis of the gut microbiome and neuroinflammation has been hypothesized to drive AD. The gut microbiota comprises a complex community of microorganism species that reside in our gastrointestinal ecosystem; alterations in the gut microbiota have been reported to influence not only various gut disorders but also brain disorders such as AD (7,8). The human gut microbiota has been suggested to modulate the brain function and behavior via the microbiota-gut-brain axis (MGBA), a bidirectional communication system connecting neural, immune, endocrine, and metabolic pathways (9). Observational studies across multiple countries show reductions in gut microbiota diversity in AD patients compared to cognitive normal controls (10–12). Current research indicates that bacteria populating the gut microbiota are capable of releasing lipopolysaccharide (LPS) and amyloids, which may induce microglial activation in the brain and contribute to the production of proinflammatory cytokines associated with the pathogenesis of AD (13). The secretion of these biomolecules also harms the integrity of the MGBA and blood-brain-barrier (BBB), which worsens with increasing dysbiosis (8,14). The composition of the human gut microbiota and risk for AD have been suggested as heritable traits

(2,15). Apolipoprotein E ϵ 4 (*APOE* ϵ 4), the most well-established risk gene for AD, has recently been shown to correlate with microbiome composition in humans and mouse models of AD (16–18). However, few studies have explored the correlation between *APOE* alleles and microbiome taxa at the human genomic level. In this study, we aim to determine the genetic correlation between the abundance of gut microbial genera and AD diagnosis. We further investigate if gut microbial genera are correlated with *APOE* ϵ 4 alleles.

One promising approach to exploring this relationship is the use of polygenic risk scores (PRS) analyses. PRS is an overall estimate of an individual's genetic liability for a specific trait. PRSice-2 is designed to calculate the PRS by aggregating and quantifying the effect of many single-nucleotide polymorphisms (SNPs) in the genome weighted by their effect sizes from genome-wide association studies (GWASs) (19). This approach has previously been used to explore the genetic relationship of gut microbial abundance and complex traits like bone mineral density, rheumatoid arthritis, and depression (20–22). In the present study, we used this approach to determine the genetic relationship between 119 microbial genera and AD diagnosis. With the largest GWAS of the human gut microbiota (23), we first conducted PRS analyses in an AD discovery sample to identify the genera genetically correlated with AD. We then verified our results in an independent AD replication sample. Correlation between the top genera and *APOE* genotype was further analyzed by linear regression.

3. METHODS AND MATERIALS

3.1 Study design overview

We obtained GWAS summary data ($n = 18,340$) from the MiBioGen consortium initiative (www.mibiogen.org) (23) and requested genotyping data from dbGaP (<https://www.ncbi.nlm.nih.gov/gap/>). A brief description for these studies is listed in **Table 1**. The overall design of the study is shown in **Figure 1**. Briefly, we used PRSice-2 (19) to calculate PRSs for individuals from our discovery sample. PRSs were calculated based on the summary statistics for 119 microbial genera from the MiBioGen consortium. Significant association between genera

and AD diagnosis was considered when the “Best Fit” PRS model with a $p < 0.00042$ (0.05/119 with Bonferroni correction). We then validated the results in an independent sample. We conducted logistic regression analyses between the PRSs of associated genera and AD diagnosis to generate relative odds ratios (OR) for meta-analyses. Multivariate logistic regression was used to determine if the correlation between the PRSs of the associated genera and AD diagnosis was affected by sex, age, and *APOE* genotype. Furthermore, we conducted a linear regression analysis to evaluate the genetic association between the PRSs of nine significant genera and the *APOE* genotype of individuals in our discovery and replication samples. This study was approved by our institutional review board (IRB) at the University of Nevada Las Vegas (UNLV).

3.2 Data sources

3.2.1 Microbiome GWAS summary statistics

We used the GWAS summary statistics (base data) of human gut microbiota in the MiBioGen consortium (www.mibiogen.org), which is the largest, multi-ethnic genome-wide meta-analysis of the gut microbiome to date (23). Participants ($n = 18,340$) were recruited in the original studies, where 16S rRNA gene sequencing profiles from each individual were utilized to characterize the gut microbiota using SILVA as a reference database (24) with truncation of the taxonomic resolution down to the genus level. In total, 31 loci from host genetic variations were associated with gut microbiota taxa abundance at the genome-wide significant threshold ($p < 5.0 \times 10^{-8}$), most of which ($n = 24$) were at the genus level (23). We chose the effect sizes of the host SNP-microbiota associations (beta) for our analyses. Beta represents how the host genetic loci affect the relative abundance of each microbiome category (mbQTLs) (23). In the present study, we limited our analyses to the summary statistics from the 119 microbial genera, as 16S rRNA sequencing correlates more accurately with the functional role of gut microbiota at lower taxonomic levels (25).

3.2.2 AD genotyping data (target data): discovery and replication samples

From dbGaP (<https://www.ncbi.nlm.nih.gov/gap/>), we requested genotyping data from the NIA/LOAD cohort consents 1 and 2 (ADc12) as our discovery sample (26), and the Multi-Site Collaborative Study for Genotype-Phenotype Associations in Alzheimer's Disease (GenADA) Study as our replication sample (27). In our study, AD cases were considered as any individual with dementia diagnosed with definite, probable or possible AD at any point in their clinical course, according to the Criteria proposed in 1984 by the National Institute of Neurological and Communicative Disorders and Stroke, and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) (28). Included controls were neurologically evaluated individuals who were age-matched cognitively normal. Unspecified dementia, unconfirmed controls, and controls with other neurological diseases from the original study were removed from this study, resulting in 1,278/1,293 cases/controls in the discovery sample ADc12, and 799/778 cases/controls in the replication sample GenADA. Demographic characteristics of the ADc12 and Gen/ADA samples are listed in **Table 2**, along with two major *APOE* SNP genotype information. More detailed descriptions of the data can be found in previous studies (26,27,29). In order to maximize genetic variants, we conducted imputation for both discovery and replication samples at the Michigan Imputation Server (minimac4) (<https://imputationserver.sph.umich.edu>) (30). The 1,000 Genome Phase 3v5 was used as a reference. After the imputation, standard quality control was performed with plink command (--maf 0.01 --hwe 1e-6 --geno 0.01 --mind 0.01) (31,32). The final datasets were composed of 2,571 individuals with 9,997,692 SNPs in the discovery sample, and 1,577 individuals with 8,914,585 SNPs in the replication sample.

3.3 Polygenic Risk Score (PRS) analyses via PRSice-2 software

PRSice-2 software was designed to evaluate the genetic correlation between different traits when provided GWAS summary statistics data from a base trait (base data) and genotyping data from a target trait (target data) (19). PRS is a numerical approximation of an individual's genetic liability weighed by the allele number and effect size from a set of genetic variants estimated from GWAS

summary statistics (33). This approach is called polygenic risk score (PRS) analyses, by which investigators can assess an individual's genetic risk to a trait of interest and the genetic correlation between different traits (33). As mentioned above, our base GWAS data were from the 119 microbiome genera in the MiBioGen consortium study (23). Our target samples were discovery sample ADc12 (26) and replication sample GenADA (27). In our study, we first calculated PRSs for the 119 microbiome genera in the discovery sample to determine the genetic correlation between the PRSs and AD diagnosis. PRSs for each genus were calculated from the PRSice-2 program for the "best-fit" PRS model. For this purpose, a range of p value thresholds was set from 5×10^{-8} to 1 with an incremental interval of 0.00005 (--interval 0.00005 --lower 5e-08) with LD clumping (--clump-kb 250kb --clump-p 1.0 --clump-r2 0.1) (19). To validate the significantly associated genera from the discovery sample, we conducted PRS analyses for the significant genera in an independent sample.

To evaluate the overall association from our discovery and replication samples, we rescaled the "best-fit" PRSs of the associated genera with a range of -1 to 1 and performed a simple logistic regression between the rescaled "best-fit" PRSs and AD diagnosis. We then conducted meta-analyses for both samples R packages metafor v3.8-1 (34). A forest plot was generated to visualize the overall AD protective and risk effect across the significant genera using the "forestplot" R package (35). We also created boxplots to compare the PRSs for those significant genera between AD patients and controls in the discovery sample. Box plots were generated by using the R program ggplot2 v.0.4.0 (36). Multivariate logistic regression was conducted by adding sex, age, and *APOE* genotypes (rs429358, rs7412) as covariates.

3.4 Linear regression analyses between *APOE* genotyping and PRSs for the nine significant genera

Two *APOE* SNPs, rs429358 minor allele C and rs7412 major allele C, are well-known risk factors for AD (37,38). We performed linear regression analyses to determine the genetic correlation between the PRSs of the nine significant genera and the two *APOE* SNPs. The association was

further assessed by linear regression adjusted for sex and age. Boxplots were created using the R packages ggplot2 (v3.3.6), ggpubr (v0.4.0), and stats (v0.1.0) (35,36,39).

3.5 Statistical Analyses

The p value threshold for significant association in the discovery sample and meta-analyses was set as $p < 4.20 \times 10^{-4}$ (0.05/119 with Bonferroni correction) (40). For the replication sample, one-side significant level $p < 0.005$ (0.1/20 with Bonferroni correction) was used. For all other statistical analyses, such as linear regression, the ANOVA test, and Wilcoxon signed-rank test, $p < 0.05$ was considered significant. The Wilcoxon signed-rank method was used to test two interrelated samples to see if their sample proportion ranks differed (41). The ANOVA method was utilized to test the association between the PRSs for genera and *APOE* genotype (42).

4. RESULTS

4.1 PRSs for nine microbiome genera were significantly associated with AD diagnosis.

We first calculated the PRSs for the 119 microbiome genera for each individual from the discovery sample ADc12 and conducted the PRS analyses between the PRSs and AD diagnosis using the PRSice-2 program (19). We found that 20 out of the 119 genera were significantly associated with AD diagnosis at the “best-fit” model ($p < 4.20 \times 10^{-4}$) (**Table 3**). Among these 20 significant genera, six were identified as likely risk genera and 14 potentially protective genera for AD diagnosis. Risk genera included *Alistipes* and *Bacteroides* from the Bacteroidetes phylum, *Lachnospira* and *Veillonella* from the Firmicutes phylum, and *Collinsella* and *Sutterella* from the Actinobacteria and Pseudomonadota phyla, respectively. The most significant risk genus was *Bacteroides* ($R^2 = 0.011$, $p = 3.32 \times 10^{-6}$) at the “best-fit” p value of 0.179 with 47,844 SNPs. For protective genera, eleven out of fourteen were from the Firmicutes phylum (*Anaerostipes*, *Candidatus Soleaferrea*, *Catenibacterium*, *Eisenbergiella*, *Eubacterium coprostanoligenes*, *Eubacterium fissicatena*, *Eubacterium nodatum*, *Intestinibacter*, *Lachnospiraceae UCG-008*, *Oscillibacter*, and *Roseburia*), two were from Actinobacteria (*Adlercreutzia* and *Gordonibacter*), and one was from Bacteroidetes

(*Prevotella 9*). The most significant protective genus was *Intestinibacter* ($R^2 = 0.015$, $p = 1.01 \times 10^{-7}$) at the “best-fit” p value of 0.190 with 474,833 SNPs.

To validate our findings for the top 20 genera in the discovery sample, we conducted replication analyses in an independent sample. Three genera, two risk genera (*Collinsella* and *Veillonella*) and one protective genus (*Eubacterium fissicatena*), were successfully validated for their association with AD diagnosis in the replication sample. Ten other genera did not reach the significant p value ($p < 0.005$) but had the same effect direction as in the discovery sample. To evaluate the overall association, we conducted meta-analyses with both the discovery and replication samples. As a result, a total of nine genera, including three genera validated from the replication sample, were significantly associated with AD diagnosis (See **Table S1** and **Figure 2**). Six genera—*Adlercreutzia*, *Candidatus Soleaferrea*, *Eisenbergiella*, *Eubacterium fissicatena*, *Gordonibacter*, and *Prevotella9*—showed as a protective factor for AD, while three genera—*Collinsella*, *Lachnospira*, and *Veillonella*—showed as a risk factor for AD. From the meta-analyses, *Eisenbergiella* was identified as the strongest protective factor for AD with $p < 0.0001$, and OR = 0.568 (95 % CI 0.450 - 0.716), and *Collinsella* was identified as the strongest risk factor for AD $p < 0.0001$ and OR = 1.869 (95 % CI 1.495 - 2.338). Meta-analyses also found four genera to have a suggestive association ($0.00042 < p < 0.05$) with AD diagnosis genera, of which two genera were a potential risk factor (*Bacteroides* and *Intestinibacter*), and one genus was a protective factor (*Roseburia*). In addition, seven genera—*Alistipes*, *Anaerostipes*, *Catenibacterium*, *Eubacterium coprostanoligenes* group, *Lachnospiraceae* UCG-008, *Oscillibacter*, and *Sutterella*—originally identified to be associated with AD, did not show any association due to the opposite effects in the replication sample.

Next, a multivariate logistic regression, including sex, age, and two *APOE* SNPs (rs429358 and rs7412) as covariates, was used to determine any confounding effects on the association between the nine significant genera and AD. As shown in Supplementary **Table S2**, the nine significant genera remained significantly associated with AD diagnosis in the discovery sample ($p < 0.05$),

which suggested that the genetic association between PRSs for those genera and AD diagnosis was independent of age, sex, and *APOE*. As expected, age and *APOE* were strongly associated with AD in the multivariate logistic regression analyses. Specifically, age and rs429358 minor allele C were risk factors because of their positive correlation with AD diagnosis, while rs7412 minor allele T was a protective factor due to its negative correlation with AD diagnosis. However, sex did not show any association with AD in our study.

To better visualize the difference of PRSs between AD cases and controls, we constructed a boxplot along with the Wilcoxon signed-rank test (41) for the nine significant genera in the discovery sample. As compared to cognitively normal controls, **Figure 3A** showed that AD patients had lower PRSs for the six likely protective genera (*Adlercreutzia*, *Candidatus Soleaferrea*, *Eisenbergiella*, *Eubacterium fissicatena*, *Gordonibacter*, and *Prevotella9*). On the other hand, **Figure 3B** showed AD patients had higher PRSs for the three risk genera (*Collinsella*, *Lachnospira*, and *Veillonella*). These results were consistent with the results from our PRS model and logistic regression between PRSs and AD diagnosis.

4.2 Correlation between PRSs for nine significant genera and *APOE* genotyping

APOE is a well-known genetic risk for AD (37,38). Depending on the alleles of two SNPs rs429358 and rs7412, the human *APOE* gene has three major isoforms ($\epsilon 2$, $\epsilon 3$, and $\epsilon 4$) (38). The $\epsilon 4$ isoform is the most influential risk factor beyond age for AD with the presence of a single $\epsilon 4$ allele increasing the risk by 3- to 4-fold compared with $\epsilon 2$ or $\epsilon 3$ allele (37). A few studies have been conducted for the potential links between the *APOE* genotype (rs429358 and rs7412) and the gut microbiota (16–18), but not at the whole genomic level. For this reason, we sought to determine whether there was a genetic link between the PRSs for the nine significant genera and the *APOE* genotype. Linear regression analyses were performed between the rescaled "best-fit" PRSs for the nine significant genera and *APOE* minor alleles at rs429358 and rs7412. *Collinsella* was the only genus that showed significant correlation in both discovery and replication samples. The meta-analyses showed that three out of nine significant genera were correlated with *APOE*

rs429358 risk allele C ($p < 0.05$). Notably, *Collinsella*, a proinflammatory genus, was the only genus that was positively correlated with AD and *APOE* rs429358 risk allele C in both discovery and replication samples ($p < 0.05$) (**Table 4**). To illustrate the correlations between PRSs for *Collinsella* and *APOE* rs429358 risk allele C, we constructed a boxplot along with ANOVA analysis (36,43). As shown in **Figure 4**, a positive correlation between PRSs for *Collinsella* and *APOE* rs429358 risk allele C was found in the discovery sample, which indicated that a genetic factor determining *Collinsella* abundance was more likely to occur in individuals with *APOE* rs429358 allele C (CC and TC) as compared to individuals with two T alleles (TT). Overall, our results showed that PRSs for *Collinsella* were significantly higher in AD cases and individuals with *APOE* risk allele C. Two genera as a protective factor for AD were negatively correlated with *APOE* rs429358 risk allele C (*Adlercreutzia* and *Prevotella9*) (**Table 4**). These associations highlighted that certain members of the microbiome and *APOE* may contribute to modulation in some similar biologic pathways, synergizing in disease risk or protection. The associations between PRSs for the three genera and rs429358 risk allele and were independent of sex and age, as the results remained significant after adjustment for sex and age (Supplementary **Table S2**). For *APOE* rs7412, we did not see any significant correlation.

4.3 Association between microbiome abundance and *APOE* genotype

To further investigate *APOE* genotype association with the abundance of all the gut microbiota genera, we retrieved summary statistics for the two *APOE* SNPs rs429358 and rs7412 directly from the 119 genera GWAS summary statistics in the MiBioGen consortium study. As shown in **Table 5**, rs429358 was marginally correlated with the abundance of nine genera, and rs7412 was marginally associated with the abundance of eight genera ($p < 0.05$). Together, these findings indicate that the *APOE* genotype may have some impact on the microbiome abundance at the genus level, and that the association may synergistically contribute to the risk for human diseases such as AD. Our results open the door for future studies to explore the role of the interaction between *APOE* and the gut microbiota and find a new target for treatment in human diseases.

5. DISCUSSION

The microbiota is a complex ecosystem that comprises more than 100 trillion symbiotic microbial cells in the human body, of which 95 % inhabit the human gut (44). The bacteria belonging to phylum Firmicutes and Bacteroidetes form a significant proportion (90 %) of the adult gut microbiota, and Actinobacteria composes the rest (45). Recently, significant evidence has shown that the gut microbiota influences normal systemic physiological homeostasis, and that dysbiosis of gut microbiota may contribute to the pathogenesis of brain diseases, including AD. The gut microbiota interacts with the central nervous system (CNS) across the MGBA via microbial components, metabolic products, and neural stimulation. In this study, we leveraged the extensive GWAS data to study the genetic correlation between gut microbiota genera and AD diagnosis. PRSs for 20 genera were initially found significantly associated with AD in the discovery sample, three of which were validated in the independent replication study. Further meta-analyses from our discovery and replication samples identified a strong genetic association between nine gut microbiota genera and AD diagnosis. Six genera were negatively associated with AD diagnosis and three genera were positively correlated with AD diagnosis. The negative association means that the abundance of these genera is negatively associated with AD diagnosis, thus, PRSs for such genera could be seen as a protective factor for the disease. Similarly, positive association means that the abundance of those genera is positively associated with AD diagnosis, thus, PRSs for these genera would be seen as a risk factor for the disease. Genera identified as a protective factor were primarily from the Firmicutes phylum (*Candidatus Soleaferrea*, *Eisenbergiella*, and *Eubacterium fissicatena*) as well as from Actinobacteria (*Adlercreutzia*, *Gordonibacter*) and Bacteroidetes (*Prevotella*9). Positively correlated, or risk-associated genera were from phyla including Firmicutes (*Lachnospira* and *Veillonella*) and Actinobacteria (*Collinsella*).

In the discovery sample, the correlation of the nine significant genera remained statistically significant after being adjusted for sex, age, and two *APOE* SNPs (rs429358 and rs7412),

suggesting that the genetic correlation between the genera and AD diagnosis was independent of age, sex, or *APOE* genotype. In addition, we found that three of the nine significant genera showed a strong correlation with the *APOE* rs429358 risk allele C via linear regression analyses. Interestingly but not surprisingly, the genera showing positive correlation with *APOE* rs429358 risk allele C tend to have a positive (risk) association with AD, while the genera showing negative correlation with *APOE* rs429358 risk allele C have a negative (protective) association with AD.

In our analyses, *Collinsella* from the phylum Actinobacteria was found to be a risk factor for AD in the discovery samples and then was validated in replication sample. *Collinsella* was also found to be positively correlated with *APOE* rs429358 risk allele C in both samples. *Collinsella* is a proinflammatory genus that is associated with cumulative inflammatory burden in rheumatoid arthritis patients (46). Studies also found that *Collinsella* is abundant in atherosclerosis and type-2 diabetes patients (47,48). Increased abundance of this genus has been observed in AD transgenic mice and AD patients (49,50). As type-2 diabetes is a well-established risk factor for AD (51), our association of genetically-predicted *Collinsella* with AD cases further supports a connection between the two diseases. The strong association between *Collinsella* and AD diagnosis along with its strong correlation with the *APOE* rs429358 risk allele C may provide new insight into how the interaction between the *APOE* ϵ 4 gene and the AD-risk pathogens promotes the risk for AD development. Targeting the risk microbiota might be an effective therapeutic strategy for AD patients with *APOE* ϵ 4 allele. Further studies to evaluate the impact of the *APOE*-associated microbiota on AD-related phenotypes will be necessary to determine if alterations in the gut microbiome represent a novel mechanism and target for the development of potentially novel interventions for AD.

Three genera of the Firmicutes phylum, *Candidatus Soleaferrea*, *Eisenbergiella*, and *Eubacterium fissicatena group*, had a negative association with AD diagnosis. *Eisenbergiella* and *Eubacterium fissicatena group* are known to contain species that metabolize the short-chain fatty acid (SCFA)

butyrate from dietary carbohydrates (52–54). Butyrate is a major SCFA metabolite in the colon and has been shown to be a critical mediator of the colonic inflammatory response. Alongside its anti-inflammatory properties, butyrate is also essential in maintaining the permeability of the gut across the MGBA (55,56). To our knowledge, there are no studies that presently implicate these genera in neurodegenerative diseases. Nevertheless, *Eubacterium fissicatena*, one of the three genera that was validated in the replication sample was found to decrease in mice with periodontitis, a disease linked to a prediabetes condition. Within the limitations of their study, the authors claimed that the gut microbiota may mediate the influence of periodontitis on prediabetes (57). On the other hand, studies found that *Candidatus Soleaferrea* tends to have a higher abundance in controls as compared to a rat prediabetic condition (58). Interestingly, the study found that colon Glucagon-like peptide-2 (GLP-2) was positively correlated with an abundance of *Candidatus Soleaferrea* (58). GLP-2, secreted from the metabolites (e.g., SCFA) of gut microbiota, is typically considered a trophic hormone involved in maintaining intestinal epithelial morphology and function. GLP-2, along with GLP-1, has been shown to possess neuroprotective properties in AD animal models (59,60). It is not clear exactly why *Eubacterium fissicatena* and *Candidatus Soleaferrea* have a protective effect against AD; however, based on the literature, both seem to have a higher abundance in the controls compared to prediabetes-related condition, which may provide some clue for future study on their AD protection. Further study would provide more insights into how to utilize the microbiome and their metabolic molecules to improve AD treatment.

In addition, we identified two Firmicutes genera as a risk factor for AD (*Lachnospira* and *Veillonella*). The association with *Veillonella* was validated in the replication sample. *Veillonella* is a normal bacterium in the intestines and oral mucosa of mammals. Consistently, other studies found that AD patients seem to have more abundance of *Veillonella* in their oral microbiome (61). Lately, another study found that overabundance of *Veillonella parvula* (belonging

to *Veillonella* genus) promotes intestinal inflammation by activating macrophages via the lipopolysaccharide-Toll-like receptor 4 (LPS-TLR4) pathway (62). Increasing intestinal inflammation by *Veillonella* might also promote AD through MGBA. On the other hand, gut *Lachnospira* and *Veillonella* species have also been identified as beneficial or commensal to gut health, such as *Lachnospira* being protective against Crohn's disease, or *Veillonella* interacting with *Streptococcus* species to modulate immune responses in the small intestine (63,64). In an observational study from a Chinese group, patients with AD had decreased *Lachnospira* at the genus level compared with healthy controls (65).

*Prevotella*₉, belonging to Bacteroidetes, was identified as a protective factor against AD in our study. There is an inconsistent relationship between *Prevotella* abundance and intestinal diseases in human studies (66). In humans, *Prevotella* is more common in populations with plant-based and high-carbohydrate diets (67). Conversely, *Bacteroides* is more abundant in those consuming “western” diets high in protein and fat (68). One study found that a higher *Prevotella-Bacteroides* ratio in the gut was predictive of the effectiveness of a 24-week weight-loss regimen in obese patients (69). Another major study showed that *Prevotella* was higher in individuals with greater adherence to Mediterranean diets, which is thought to be protective against neurodegenerative diseases (70–72). The “protective” effects of *Prevotella* abundance may come from the positive dietary effects on the genus. In addition, *Prevotella* was also found to be negatively correlated to *APOE* rs429358 risk allele. It is not clear how the *APOE* gene interacts with the microbiome to provide the protective effect. Nevertheless, more work needs to be done to understand the association between the gut microbiota and *APOE*.

Two protective genera, *Gordonibacter* and *Adlercreutzia*, are from the Actinobacteria phylum. These genera tend to produce metabolites beneficial to mitochondrial function, namely Urolithin-A (UA) and Equol (73,74). UA is an anti-inflammatory compound that enhances mitophagy, the removal of dysfunctional mitochondria in a cell (75). Impaired mitophagy is part of the

pathogenesis of AD, as well as general aging processes, making UA and *Gordonibacter* species promising targets for therapeutics against the disease (76). Equol is an estrogen-like compound that reduces microglial inflammation when stimulated by LPS and downregulates genes in neurons related to apoptosis (77). The positive effects of these bacterial metabolites likely drive the protective association of *Gordonibacter* and *Adlercreutzia* abundance with AD that we found in this study.

The strengths of our study include the use of the largest available GWAS of gut microbiota taxa to date, the identification of multiple genera genetically associated with AD in the discovery, and replication samples after a strict Bonferroni correction. The use of logistic regression alongside our initial PRS analyses allows us to adjust for potential confounders such as sex, age, and *APOE* alleles and further validate that the association was independent of those confounders. Additionally, we are the first to study the genetic correlation between the gut microbiota and *APOE* gene at the human genomic level. We found that three genera were associated with the *APOE* risk allele at rs429358, one of the strongest genetic risk variants for AD. The identified genetic relationship may provide candidate bacteria for subsequent causal relationships and functional studies.

However, there are several limitations to our study. First, the sample size for the microbiome GWAS may still not be large enough to truly cover the effect size of the host genetic variants, even though the MiBioGen study has the largest sample size compared with other microbiome GWASs. The sample size for our target genotyping was also relatively small. Because of this, we may not have enough power to detect some of the associations in our meta-analyses. For example, two genera, *Bacteroides* and *Intestinibacter*, were originally identified to be the strongest risk and protective factors, respectively, from the discovery sample, however, meta-analyses only found a suggestive association with AD. The modest sample size might be one of the reasons. Heterogeneity between the discovery and replication samples could be another reason. Studies

have found that *Bacteroides* is capable of secreting LPS as an endotoxic biomolecule, which has been implicated in pathological endothelial dysfunction of the gut and can induce neuroinflammation in microglia cells (78–80). For *Intestinibacter*, its protective role for AD is unknown; however, the reduction of *Intestinibacter* is a well-known microbial biomarker of Metformin-induced gut dysbiosis in the treatment of type-2 diabetes (81–83). Future studies with larger sample sizes would be more capable of drawing solid conclusions about the genetic connection between the gut microbiota and AD. Second, our genotyping data was specifically drawn from individuals in the United States, which limits the generalization of our conclusions when applied to other nations or ethnic groups. More diverse genotyping and GWAS datasets would enable us to capture the variability in risk for AD across different ethnicities. Third, the 16S rRNA sequencing used to generate genetic associations in the “base” GWAS only provides taxa resolution to the genus level. Understanding the role of bacterial species that may drive the pathology of AD will require more studies of specific species.

Overall, our novel findings provide new insights into understanding AD and point to new therapeutic targets for earlier intervention in the disease. Further investigations are needed to explore the microbial alterations in larger cohorts of AD patients from different ethnic backgrounds and identify the causal effect between the gut microbiome and AD.

6. SUPPLEMENTARY MATERIALS:

Supplementary Tables

Table S1. Logistic regression and meta-analyses between the PRSs for the top 20 gut microbiota genera and AD diagnosis in discovery and replication samples.

Table S2. Association between the PRSs for the nine significant gut microbiota genera and AD adjusted for sex, age, and *APOE* genotype in the discovery sample.

Table S3. Linear regression and meta-analyses between the PRSs for the nine significant gut microbiota genera and *APOE* rs429358 risk allele C adjusted for sex and age.

7. DATA AVAILABILITY STATEMENT:

Full GWAS summary statistics for mbQTLs are available at the www.mibiogen.org website built using the MOLGENIS framework (99).

8. ETHICAL APPROVAL AND CONSENT TO PARTICIPATE

Contributing studies received ethical approval from their respective institutional review boards (IRB). This study was approved by the IRB at the University of Nevada-Las Vegas (IRB #00002305, 10/12/2021).

9. ACKNOWLEDGEMENTS

We thank the patients for their participation in the MiBioGen consortium, NIA/LOAD ADc12 Cohort, and GenADA studies, and the original investigators who conducted these studies and made the data available.

10. AUTHOR CONTRIBUTION

J.C. was responsible for the conception, sample request, data analyses, and writing the manuscript. D.C., Y.L., and M.J.C. were responsible for data analyses and writing the manuscript. M.L.Z., J.M.C, and J.D. were responsible for data analyses. X.C., E.O., J.L.C, and J.E. were responsible for writing, critiquing, and revising the manuscript. All authors have read and agreed to the content of the manuscript before sending to journal for publication.

11. CONFLICTS OF INTEREST

The authors declare no conflict of interest.

12. FUNDING

This research was funded to J.C. in part by NIH grant P20GM121325, NIH grant P20GM121325-02S1 and NIH grant P20GM121325_03S2.

13. REFERENCES

1. Arvanitakis Z, Shah RC, Bennett DA. Diagnosis and Management of Dementia: Review. *JAMA*. 2019 Oct 22;322(16):1589.
2. 2022 Alzheimer's disease facts and figures. *Alzheimers Dement*. 2022 Apr;18(4):700–89.
3. Nichols E, Steinmetz JD, Vollset SE, Fukutaki K, Chalek J, Abd-Allah F, et al. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health*. 2022 Feb;7(2):e105–25.
4. Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther*. 2014;6(4):37.
5. Bachiller S, Jiménez-Ferrer I, Paulus A, Yang Y, Swanberg M, Deierborg T, et al. Microglia in Neurological Diseases: A Road Map to Brain-Disease Dependent-Inflammatory Response. *Front Cell Neurosci*. 2018 Dec 18;12:488.
6. Leng F, Edison P. Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here? *Nat Rev Neurol*. 2021 Mar;17(3):157–72.
7. Chen Y, Zhou J, Wang L. Role and Mechanism of Gut Microbiota in Human Disease. *Front Cell Infect Microbiol*. 2021 Mar 17;11:625913.
8. Goyal D, Ali SA, Singh RK. Emerging role of gut microbiota in modulation of neuroinflammation and neurodegeneration with emphasis on Alzheimer's disease. *Prog Neuropsychopharmacol Biol Psychiatry*. 2021 Mar 2;106:110112.
9. Carabotti M, Scirocco A, Maselli MA, Severi C. The gut-brain axis: interactions between enteric microbiota, central and enteric nervous systems. *Ann Gastroenterol*. :7.
10. Zhuang ZQ, Shen LL, Li WW, Fu X, Zeng F, Gui L, et al. Gut Microbiota is Altered in Patients with Alzheimer's Disease. *J Alzheimers Dis JAD*. 2018;63(4):1337–46.
11. Vogt NM, Kerby RL, Dill-McFarland KA, Harding SJ, Merluzzi AP, Johnson SC, et al. Gut microbiome alterations in Alzheimer's disease. *Sci Rep*. 2017 Oct 19;7(1):13537.
12. Hung CC, Chang CC, Huang CW, Nouchi R, Cheng CH. Gut microbiota in patients with Alzheimer's disease spectrum: a systematic review and meta-analysis. *Aging*. 2022 Jan 14;14(1):477–96.

13. Henry CJ, Huang Y, Wynne AM, Godbout JP. Peripheral lipopolysaccharide (LPS) challenge promotes microglial hyperactivity in aged mice that is associated with exaggerated induction of both pro-inflammatory IL-1 β and anti-inflammatory IL-10 cytokines. *Brain Behav Immun*. 2009 Mar;23(3):309–17.
14. Miller AL, Bessho S, Grando K, Tükel Ç. Microbiome or Infections: Amyloid-Containing Biofilms as a Trigger for Complex Human Diseases. *Front Immunol*. 2021 Feb 26;12:638867.
15. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human Genetics Shape the Gut Microbiome. *Cell*. 2014 Nov;159(4):789–99.
16. Hou M, Xu G, Ran M, Luo W, Wang H. APOE- ϵ 4 Carrier Status and Gut Microbiota Dysbiosis in Patients With Alzheimer Disease. *Front Neurosci*. 2021 Feb 24;15:619051.
17. Zajac DJ, Green SJ, Johnson LA, Estus S. APOE genetics influence murine gut microbiome. *Sci Rep*. 2022 Dec;12(1):1906.
18. Tran TTT, Corsini S, Kellingray L, Hegarty C, Le Gall G, Narbad A, et al. APOE genotype influences the gut microbiome structure and function in humans and mice: relevance for Alzheimer's disease pathophysiology. *FASEB J*. 2019 Jul;33(7):8221–31.
19. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience*. 2019 Jul 1;8(7):giz082.
20. Cheng S, Qi X, Ma M, Zhang L, Cheng B, Liang C, et al. Assessing the Relationship Between Gut Microbiota and Bone Mineral Density. *Front Genet*. 2020 Jan 31;11:6.
21. Wells PM, Adebayo AS, Bowyer RCE, Freidin MB, Finckh A, Strowig T, et al. Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross-sectional study. *Lancet Rheumatol*. 2020 Jul;2(7):e418–27.
22. Yao Y, Qi X, Jia Y, Ye J, Chu X, Wen Y, et al. Evaluating the interactive effects of dietary habits and human gut microbiome on the risks of depression and anxiety. *Psychol Med*. 2022 Jan 25;1–9.
23. Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat Genet*. 2021 Feb;53(2):156–65.
24. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013 Jan;41(Database issue):D590-596.
25. Wakita Y, Shimomura Y, Kitada Y, Yamamoto H, Ohashi Y, Matsumoto M. Taxonomic classification for microbiome analysis, which correlates well with the metabolite milieu of the gut. *BMC Microbiol*. 2018 Dec;18(1):188.
26. Lee JH, Cheng R, Graff-Radford N, Foroud T, Mayeux R, National Institute on Aging Late-Onset Alzheimer's Disease Family Study Group. Analyses of the National Institute on Aging

- Late-Onset Alzheimer's Disease Family Study: implication of additional loci. *Arch Neurol*. 2008 Nov;65(11):1518–26.
27. Li H, Wetten S, Li L, St. Jean PL, Upmanyu R, Surh L, et al. Candidate Single-Nucleotide Polymorphisms From a Genomewide Association Study of Alzheimer Disease. *Arch Neurol*. 2008 Jan 1;65(1):45–53.
 28. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease. *Neurology*. 1984 Jul 1;34(7):939.
 29. European Alzheimer's Disease Initiative (EADI), Genetic and Environmental Risk in Alzheimer's Disease (GERAD), Alzheimer's Disease Genetic Consortium (ADGC), Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE), Lambert JC, Ibrahim-Verbaas CA, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet*. 2013 Dec;45(12):1452–8.
 30. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016 Oct;48(10):1284–7.
 31. Verma SS, de Andrade M, Tromp G, Kuivaniemi H, Pugh E, Namjou-Khales B, et al. Imputation and quality control steps for combining multiple genome-wide datasets. *Front Genet*. 2014;5:370.
 32. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015 Dec 1;4(1):s13742-015-0047–8.
 33. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009 Aug 6;460(7256):748–52.
 34. Michael Dewey, Masayuki Henmi, John Copas. Meta-Analysis Package for R [Internet]. Available from: <https://www.metafor-project.org/doku.php/metafor><https://www.metafor-project.org/doku.php/metafor>
 35. Max Gordon, Thomas Lumley. Advanced Forest Plot Using “grid” Graphics [Internet]. CRAN; 2021. Available from: <https://gforge.se/packages/>
 36. Kassambara A. “ggplot2” Based Publication Ready Plots [Internet]. CRAN; 2020. Available from: <https://rpkgs.datanovia.com/ggpubr/>
 37. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer's Disease in Late Onset Families. *Science*. 1993 Aug 13;261(5123):921–3.
 38. Okuizumi K, Onodera O, Tanaka H, Kobayashi H, Tsuji S, Takahashi H, et al. ApoE-ε4 and early-onset Alzheimer's. *Nat Genet*. 1994 May 1;7(1):10–1.
 39. R Core Team. R: A language and environment for statistical computing. [Internet]. R Foundation for Statistical Computing; 2021. Available from: <https://www.R-project.org/>

40. bonferroni correction - Wolfram|Alpha [Internet]. [cited 2022 Sep 16]. Available from: <https://www.wolframalpha.com>
41. Wilcoxon F. Individual Comparisons by Ranking Methods. *Int Biom Soc.* 1945 Dec;1:80–3.
42. Fisher RA. Statistical Methods for Research Workers. In: Kotz S, Johnson NL, editors. *Breakthroughs in Statistics: Methodology and Distribution* [Internet]. New York, NY: Springer New York; 1992. p. 66–70. Available from: https://doi.org/10.1007/978-1-4612-4380-9_6
43. Mather K, Moran PAP, Smith CAB. Commentary on R. A. Fisher's paper on The Correlation Between Relatives on the Supposition of Mendelian Inheritance. *Popul Stud.* 1967 Mar;20(3):372.
44. de J.R. De-Paula V, Forlenza AS, Forlenza OV. Relevance of gutmicrobiota in cognition, behaviour and Alzheimer's disease. *Pharmacol Res.* 2018 Oct 1;136:29–34.
45. O'Toole PW, Jeffery IB. Gut microbiota and aging. *Science.* 2015 Dec 4;350(6265):1214–5.
46. Ruiz-Limón P, Mena-Vázquez N, Moreno-Indias I, Manrique-Arija S, Lisbona-Montañez JM, Cano-García L, et al. *Collinsella* is associated with cumulative inflammatory burden in an established rheumatoid arthritis cohort. *Biomed Pharmacother Biomedecine Pharmacother.* 2022 Sep;153:113518.
47. Karlsson FH, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D, et al. Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat Commun.* 2012 Jan;3(1):1245.
48. Candela M, Biagi E, Soverini M, Consolandi C, Quercia S, Severgnini M, et al. Modulation of gut microbiota dysbioses in type 2 diabetic patients by macrobiotic Ma-Pi 2 diet. *Br J Nutr.* 2016 Jul 14;116(1):80–93.
49. Bello-Medina PC, Hernández-Quiroz F, Pérez-Morales M, González-Franco DA, Cruz-Pauseno G, García-Mena J, et al. Spatial Memory and Gut Microbiota Alterations Are Already Present in Early Adulthood in a Pre-clinical Transgenic Model of Alzheimer's Disease. *Front Neurosci.* 2021 Apr 29;15:595583.
50. Ling Z, Zhu M, Yan X, Cheng Y, Shao L, Liu X, et al. Structural and Functional Dysbiosis of Fecal Microbiota in Chinese Patients With Alzheimer's Disease. *Front Cell Dev Biol.* 2021 Feb 4;8:634069.
51. Barbagallo M. Type 2 diabetes mellitus and Alzheimer's disease. *World J Diabetes.* 2014;5(6):889.
52. Amir I, Bouvet P, Legeay C, Gophna U, Weinberger A. *Eisenbergiella tayi* gen. nov., sp. nov., isolated from human blood. *Int J Syst Evol Microbiol.* 2014 Mar;64(Pt 3):907–14.
53. Togo AH, Khelaifia S, Bittar F, Maraninchi M, Raoult D, Million M. '*Eisenbergiella massiliensis*', a new species isolated from human stool collected after bariatric surgery. *New Microbes New Infect.* 2016 Sep;13:15–6.

54. Oh JK, Vasquez R, Kim SH, Lee JH, Kim EJ, Hong SK, et al. Neogargarooligosaccharides modulate gut microbiota and alleviate body weight gain and metabolic syndrome in high-fat diet-induced obese rats. *J Funct Foods*. 2022 Jan 1;88:104869.
55. Kelly CJ, Zheng L, Campbell EL, Saeedi B, Scholz CC, Bayless AJ, et al. Crosstalk between Microbiota-Derived Short-Chain Fatty Acids and Intestinal Epithelial HIF Augments Tissue Barrier Function. *Cell Host Microbe*. 2015 May;17(5):662–71.
56. Silva YP, Bernardi A, Frozza RL. The Role of Short-Chain Fatty Acids From Gut Microbiota in Gut-Brain Communication. *Front Endocrinol*. 2020 Jan 31;11:25.
57. Li L, Bao J, Chang Y, Wang M, Chen B, Yan F. Gut Microbiota May Mediate the Influence of Periodontitis on Prediabetes. *J Dent Res*. 2021 Nov;100(12):1387–96.
58. Cai J, Zhou L, Song X, Yin M, Liang G, Xu H, et al. Alteration of Intestinal Microbiota in 3-Deoxyglucosone-Induced Prediabetic Rats. *BioMed Res Int*. 2020 Aug 25;2020:1–11.
59. Sasaki-Hamada S, Ikeda M, Oka JI. Glucagon-like peptide-2 rescues memory impairments and neuropathological changes in a mouse model of dementia induced by the intracerebroventricular administration of streptozotocin. *Sci Rep*. 2019 Sep 23;9:13723.
60. Hölscher C. Protective properties of GLP-1 and associated peptide hormones in neurodegenerative disorders. *Br J Pharmacol*. 2022 Feb;179(4):695–714.
61. Guo H, Li B, Yao H, Liu D, Chen R, Zhou S, et al. Profiling the oral microbiomes in patients with Alzheimer’s disease. *Oral Dis*. 2021 Dec 21;
62. Zhan Z, Liu W, Pan L, Bao Y, Yan Z, Hong L. Overabundance of *Veillonella parvula* promotes intestinal inflammation by activating macrophages via LPS-TLR4 pathway. *Cell Death Discov*. 2022 May 6;8(1):1–12.
63. Caparrós E, Wiest R, Scharl M, Rogler G, Gutiérrez Casbas A, Yilmaz B, et al. Dysbiotic microbiota interactions in Crohn’s disease. *Gut Microbes*. 13(1):1949096.
64. Bogert B van den, Meijerink M, Zoetendal EG, Wells JM, Kleerebezem M. Immunomodulatory Properties of *Streptococcus* and *Veillonella* Isolates from the Human Small Intestine Microbiota. *PLOS ONE*. 2014 Dec 5;9(12):e114277.
65. Guo M, Peng J, Huang X, Xiao L, Huang F, Zuo Z. Gut Microbiome Features of Chinese Patients Newly Diagnosed with Alzheimer’s Disease or Mild Cognitive Impairment. *J Alzheimers Dis JAD*. 2021;80(1):299–310.
66. Fox M, Knorr DA, Haptonstall KM. Alzheimer’s disease and symbiotic microbiota: an evolutionary medicine perspective. *Ann N Y Acad Sci*. 2019;1449(1):3–24.
67. Larsen JM. The immune response to *Prevotella* bacteria in chronic inflammatory disease. *Immunology*. 2017 Aug;151(4):363–74.
68. Tomova A, Bukovsky I, Rembert E, Yonas W, Alwarith J, Barnard ND, et al. The Effects of Vegetarian and Vegan Diets on Gut Microbiota. *Front Nutr*. 2019 Apr 17;6:47.

69. Hjorth MF, Blædel T, Bendtsen LQ, Lorenzen JK, Holm JB, Kiillerich P, et al. Prevotella-to-Bacteroides ratio predicts body weight and fat loss success on 24-week diets varying in macronutrient composition and dietary fiber: results from a post-hoc analysis. *Int J Obes.* 2019 Jan;43(1):149–57.
70. Filippis FD, Pellegrini N, Vannini L, Jeffery IB, Storia AL, Laghi L, et al. High-level adherence to a Mediterranean diet beneficially impacts the gut microbiota and associated metabolome. *Gut.* 2016 Nov 1;65(11):1812–21.
71. Scarmeas N, Luchsinger JA, Mayeux R, Stern Y. Mediterranean diet and Alzheimer disease mortality. *Neurology.* 2007 Sep 11;69(11):1084–93.
72. Picchianti Diamanti A, Panebianco C, Salerno G, Di Rosa R, Salemi S, Sorgi ML, et al. Impact of Mediterranean Diet on Disease Activity and Gut Microbiota Composition of Rheumatoid Arthritis Patients. *Microorganisms.* 2020 Dec 14;8(12):1989.
73. Selma MV, Tomás-Barberán FA, Beltrán D, García-Villalba R, Espín JC. *Gordonibacter urolithinifaciens* sp. nov., a urolithin-producing bacterium isolated from the human gut. *Int J Syst Evol Microbiol.* 2014 Jul 1;64(Pt_7):2346–52.
74. Maruo T, Sakamoto M, Ito C, Toda T, Benno Y. *Adlercreutzia equolifaciens* gen. nov., sp. nov., an equol-producing bacterium isolated from human faeces, and emended description of the genus *Eggerthella*. *Int J Syst Evol Microbiol.* 2008 May 1;58(5):1221–7.
75. D’Amico D, Andreux PA, Valdés P, Singh A, Rinsch C, Auwerx J. Impact of the Natural Compound Urolithin A on Health, Disease, and Aging. *Trends Mol Med.* 2021 Jul;27(7):687–99.
76. Pradeepkiran JA, Hindle A, Kshirsagar S, Reddy PH. Are mitophagy enhancers therapeutic targets for Alzheimer’s disease? *Biomed Pharmacother.* 2022 May;149:112918.
77. Subedi L, Ji E, Shin D, Jin J, Yeo JH, Kim SY. Equol, a Dietary Daidzein Gut Metabolite Attenuates Microglial Activation and Potentiates Neuroprotection In Vitro. *Nutrients.* 2017 Mar;9(3):207.
78. Lukiw WJ. *Bacteroides fragilis* Lipopolysaccharide and Inflammatory Signaling in Alzheimer’s Disease. *Front Microbiol.* 2016;7:1544.
79. Dhaliwal G. Alistipes: The influence of a commensal on anxiety and depression. 2019;3(1):9.
80. Lukiw WJ. Gastrointestinal (GI) Tract Microbiome-Derived Neurotoxins—Potent Neuro-Inflammatory Signals From the GI Tract via the Systemic Circulation Into the Brain. *Front Cell Infect Microbiol.* 2020 Feb 12;10:22.
81. Elbere I, Silamikelis I, Dindune II, Kalnina I, Ustinova M, Zaharenko L, et al. Baseline gut microbiome composition predicts metformin therapy short-term efficacy in newly diagnosed type 2 diabetes patients. *PLOS ONE.* 2020 Oct 30;15(10):e0241338.
82. Mueller NT, Differding MK, Zhang M, Maruthur NM, Juraschek SP, Iii ERM, et al. Metformin Affects Gut Microbiome Composition and Function and Circulating Short-. 2021;44:10.

83. Rosario D, Benfeitas R, Bidkhorji G, Zhang C, Uhlen M, Shoaie S, et al. Understanding the Representative Gut Microbiota Dysbiosis in Metformin-Treated Type 2 Diabetes Patients Using Genome-Scale Metabolic Modeling. *Front Physiol.* 2018 Jun 25;9:775.

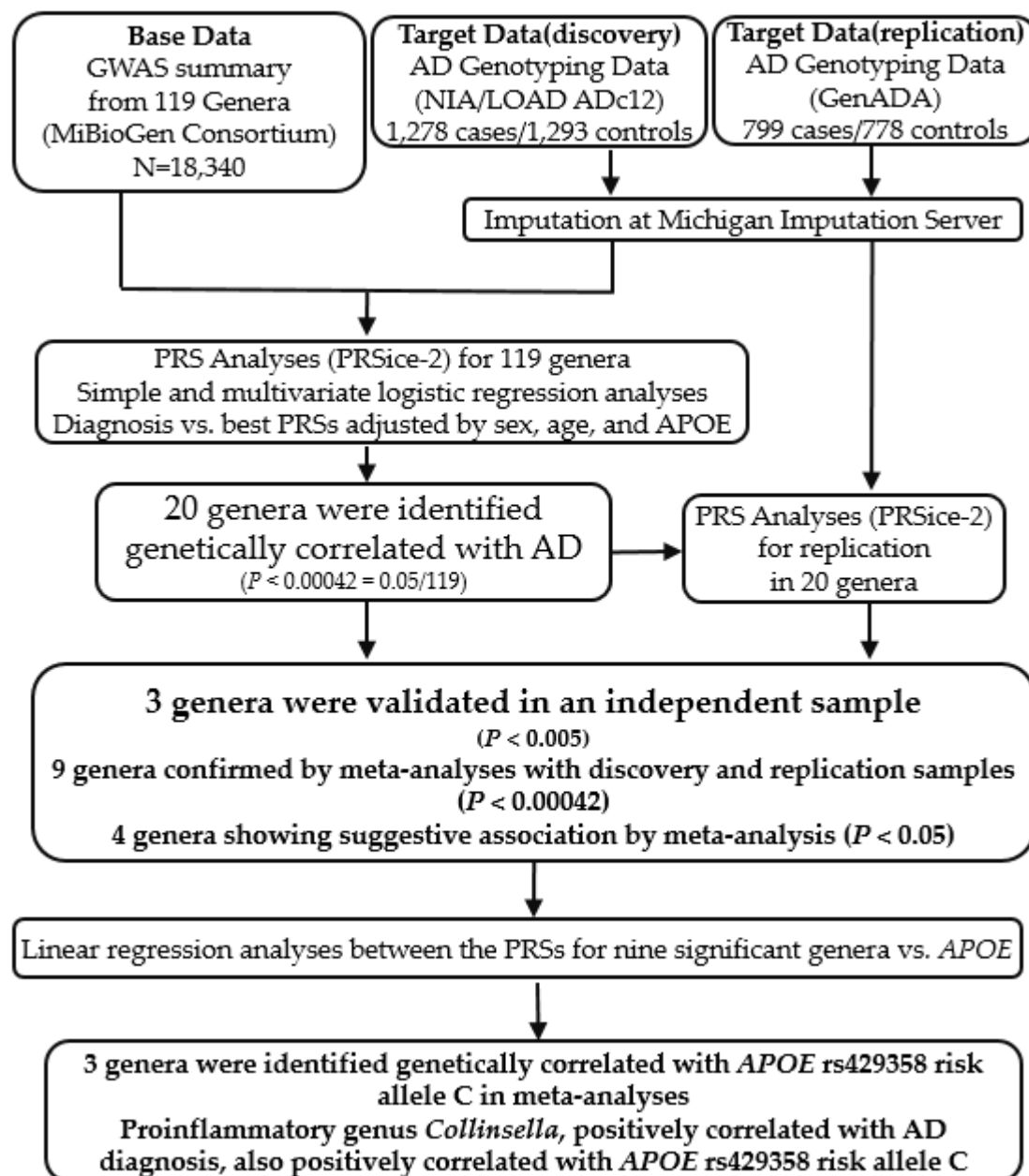


Figure 1. Study design flowchart. In the PRS analysis, “Base” data included summary statistics for 119 microbial genera from the MiBioGen Consortium. “Target” data included AD genotyping data in NIA/LOAD cohort consents 1 and 2 (ADc12) for our discovery sample, and AD genotyping from the GenADA study as our replicate. PRS analyses with PRSice-2, logistic regression analyses, and meta-analyses were used to determine the genetic correlation between the 119 genera and AD diagnosis. 20 genera were found to be significantly genetically associated with AD in the discovery sample, three genera were validated in the replication sample, and nine genera were confirmed by meta-analyses with discovery and replicate samples. Linear regression analyses were used to determine the genetic correlation between the PRSs for nine significant genera and APOE genotyping. Three genera were identified genetically correlated with APOE rs429358 risk allele C.

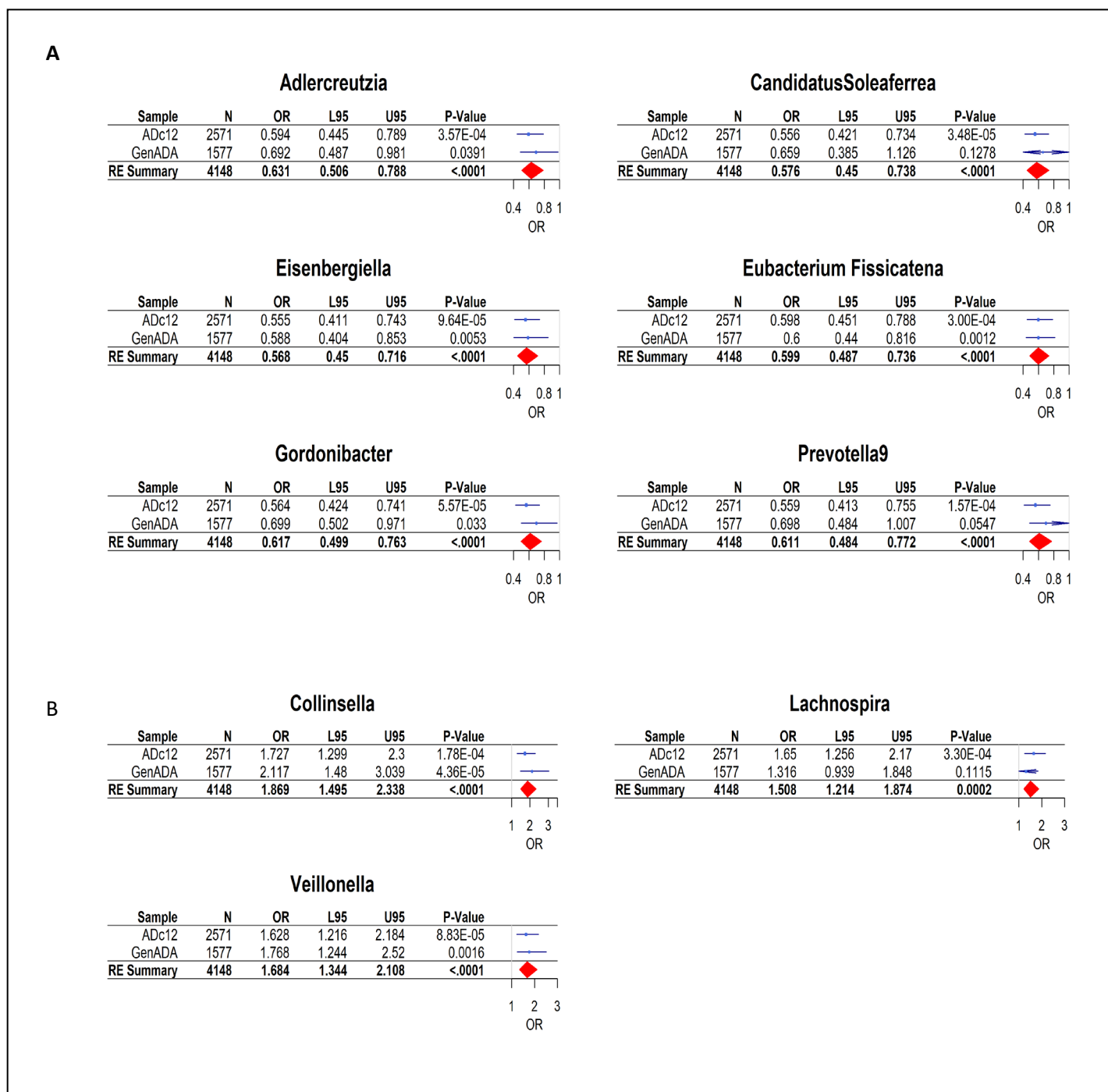


Figure 2: Forest plot of odds ratio and 95% confidence intervals (CI) of rescaled PRSs for nine genera in AD cases compared to controls. The results came from meta-analyses of both discovery (ADc12) and replication samples (GenADA). A. Six genetically predicted genera showed significant association with AD as a protective factor, as $p < 0.00042$ and ORs < 1.0 . B. Three genetically predicted genera showed significant association with AD as a risk factor, as p -value < 0.00042 and the ORs > 1.0 . OR (95%CI): Odds ratio of respective genus with their lower and upper 95 % confidence intervals.

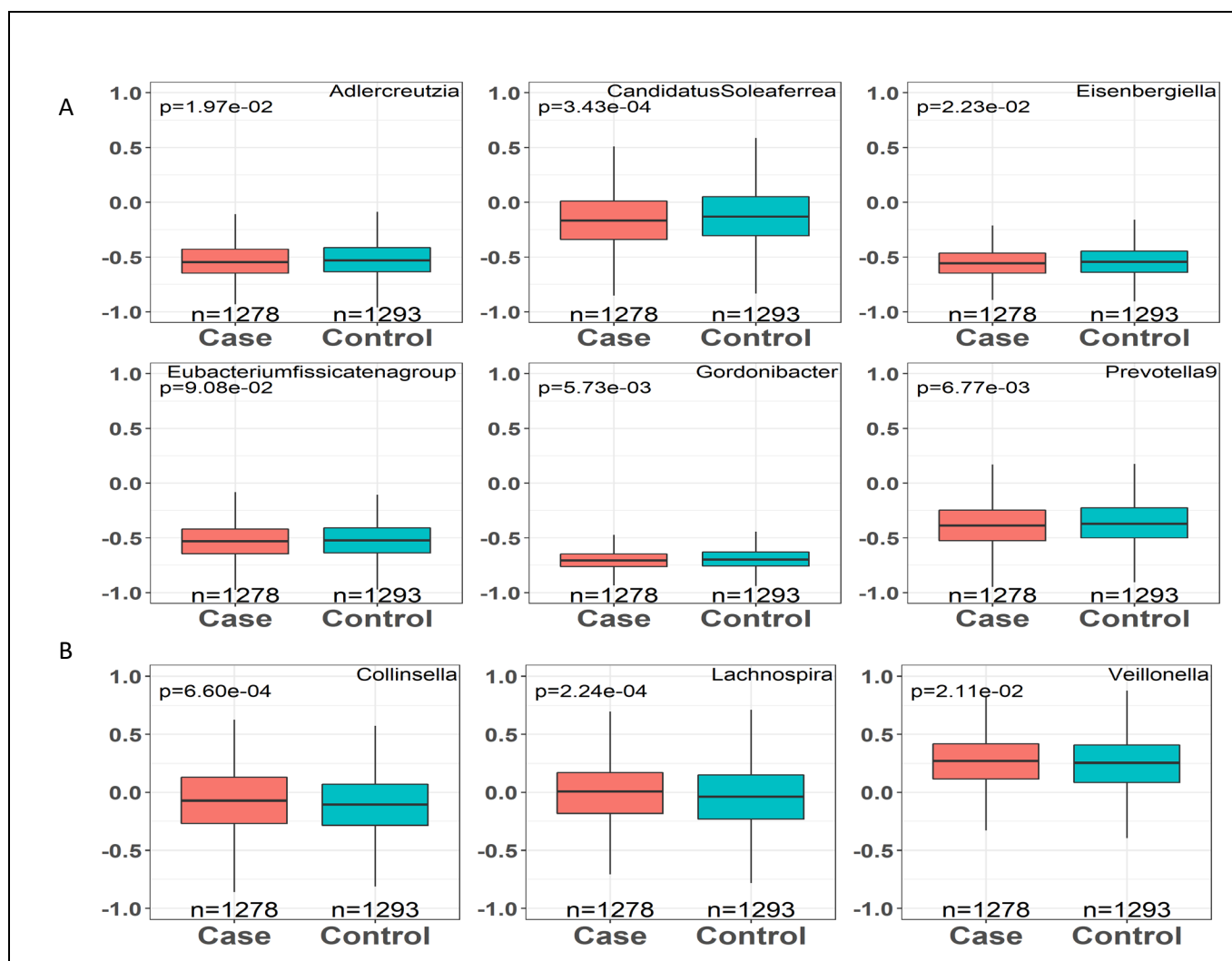


Figure 3. Comparison of rescaled PRSs [-1, 1] for top nine gut genera between AD cases and controls from discovery sample.

- A. PRSs for six genera were relatively lower as compared AD cases with controls ($p < 0.05$), suggesting those genetically predicted genera were likely to be a protective factor for AD.
- B. PRSs for three genera were relatively higher as compared in AD cases vs. controls ($p < 0.05$), suggesting those genetically predicted genera were likely be a risk factor for AD.

Wilcoxon signed-rank test was applied to generate p values. X-axis: Diagnosis (AD cases/controls) with sample count. Y-axis: rescaled PRSs [-1,1] for each of the top nine genera.

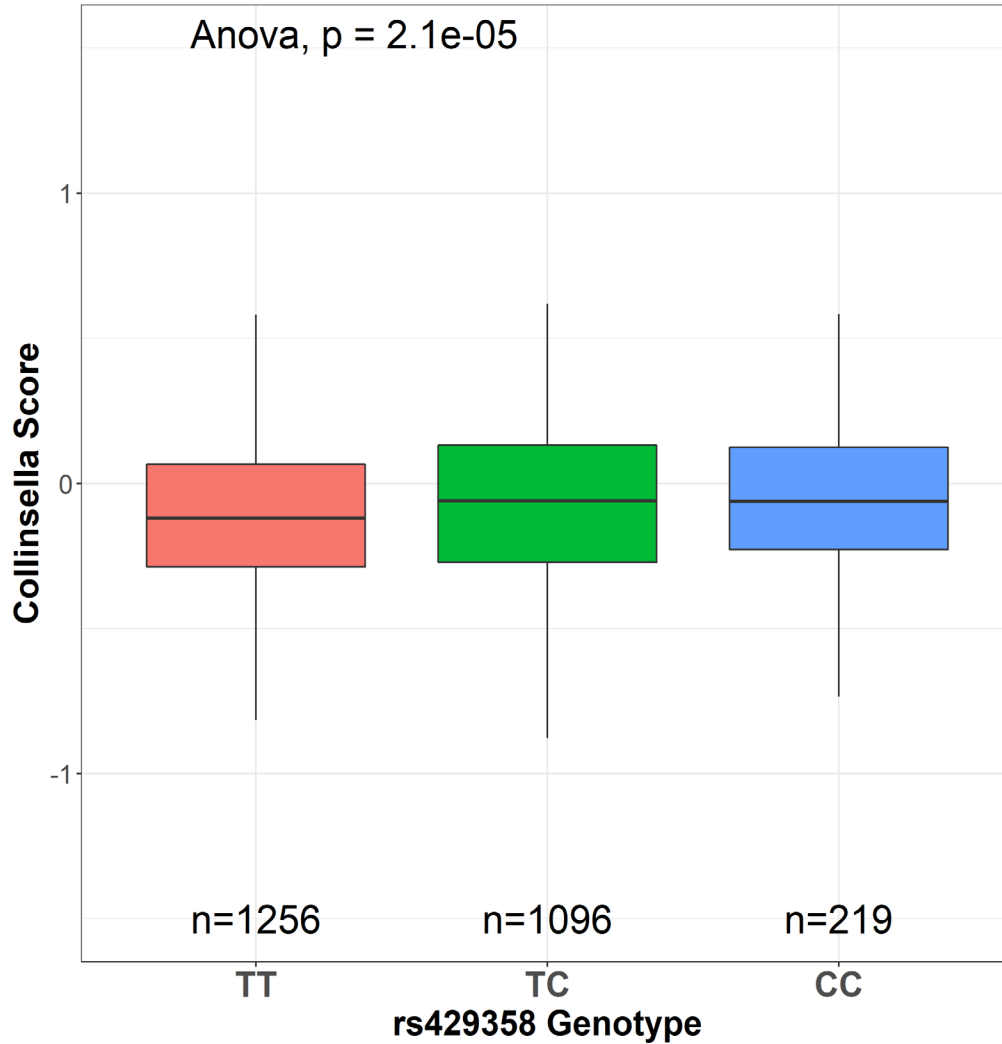


Figure 4. Genetic association between PRSs for *Collinsella* and *APOE* rs429358 minor allele C in the discovery sample. Proinflammatory genus *Collinsella* that was positively associated with AD showed a positive correlation with *APOE* rs429358 minor allele C.

Table 1. Information for studies used in our analyses.

Variable	Consortium or Study	PMID	Year	Sample Size	Ethnic Group
Gut microbiota	MiBioGen	33462485	2021	18340	Multi-ethnic
Alzheimer's Disease	ADc12			2571	
	dbGaP phs000168.v2.p2 NIA/LOAD consent 1 and 2	19001172	2008	1278 cases/1293 controls	Multi-ethnic
	GenADA dbGap phs000219.v1.p1	17998437	2008	1577 799 cases/778 controls	Caucasian

Table 2. Demographic characteristic of the target data (ADc12 and GenADA) with APOE SNP genotyping.

		Discovery Sample (ADc12)			Replication Sample (GenADA)		
		Cases	Controls	Total	Cases	Controls	Total
Age, mean ± SD		1278	1293	2571	799	778	1577
		76.57 ± 6.71*	70.27 ± 10.29		72.24 ± 8.41**	73.40 ± 7.92	
Sex (Male/Female)		443/ 835	471/822	914/1657	339/460	276/502	615/962
APOE SNP Genotype	T/T, n(%)	409 (32.0)	847 (65.5)	1256 (48.9)	296 (37.0)	589 (75.7)	885 (56.1)
	rs429358 T/C, n(%)	682 (53.4)	414 (32.0)	1096 (42.6)	397 (49.7)	177 (22.8)	574 (36.4)
	C/C, n(%)	187 (14.6)	32 (2.5)	219 (8.5)	106 (13.3)	12 (1.5)	118 (7.5)
	C/C, n(%)	1206 (94.4)	1128 (87.2)	2334 (90.8)	739 (92.5)	661 (85.0)	1400 (88.8)
	rs7412 C/T, n(%)	71 (5.6)	159 (12.3)	230 (8.9)	60 (7.5)	113 (14.5)	173 (11.0)
	T/T, n(%)	1 (0.0)	6 (5.1)	7 (0.3)	0 (0.0)	4 (0.5)	4 (0.3)

Note: Both discovery (ADc12) and replication (GenADA) samples were downloaded from dbGaP. Age, mean ± SD. For each case, their “Age” was the age at onset (AAO). For each control, their “Age” was the age at examination (AAE). * $p = 5.97 \times 10^{-71}$ when AAO compared to AAE in the discovery sample. ** $p = 4.94 \times 10^{-3}$ when AAO compared to AAE in the replication sample.

Table 3. Association between top 20 microbiome genera and AD diagnosis at the “best-fit” model from PRSice-2 in the discovery and replication samples.

Genera	Samples	Threshold	R ²	P	Coeff.	SE	SNP#	SampleSize
Adlercreutzia*	ADc12	0.3705	0.0068	3.57E-04	-1870.66	524.00	101798	2571
	GenADA	0.0020	0.0036	3.91E-02	-224.26	108.71	1684	1577
Alistipes	ADc12	0.1676	0.0080	1.04E-04	2581.83	665.28	67574	2571
	GenADA	0.3246	0.0023	1.03E-01	-2953.44	1810.61	99785	1577
Anaerostipes	ADc12	0.0032	0.0078	1.18E-04	-469.15	121.86	2834	2571
	GenADA	0.2989	0.0054	1.18E-02	4134.46	1641.01	96076	1577
Bacteroides*	ADc12	0.1792	0.0114	3.32E-06	3930.96	845.35	71984	2571
	GenADA	0.0009	0.0037	3.68E-02	221.55	106.10	922	1577
Candidatus Soleaferrea*	ADc12	0.0002	0.0090	3.48E-05	-71.29	17.22	177	2571
	GenADA	0.0199	0.0020	1.28E-01	-404.97	265.96	12431	1577
Catenibacterium	ADc12	0.3839	0.0082	8.54E-05	-1276.68	324.95	71434	2571
	GenADA	0.0831	0.0080	2.11E-03	1068.34	347.56	25728	1577
Collinsella	ADc12	0.0002	0.0073	1.78E-04	125.63	33.51	230	2571
	GenADA	0.0003	0.0143	4.36E-05	229.15	56.06	335	1577
Eisenbergiella*	ADc12	0.1049	0.0082	9.64E-05	-858.39	220.13	42727	2571
	GenADA	0.0439	0.0066	5.28E-03	-1131.73	405.70	22002	1577
Eubacterium coprostanoligenes	ADc12	0.1020	0.0086	7.19E-05	-1085.22	273.37	48130	2571
	GenADA	0.0101	0.0038	3.36E-02	707.65	333.01	7483	1577
Eubacterium fissicatena	ADc12	0.0404	0.0070	3.00E-04	-418.31	115.71	19567	2571
	GenADA	0.0008	0.0090	1.17E-03	-142.85	44.01	663	1577
Eubacterium nodatum*	ADc12	0.4353	0.0072	2.52E-04	-1240.65	338.95	95579	2571
	GenADA	0.1259	0.0025	8.82E-02	-805.51	472.45	42817	1577
Gordonibacter*	ADc12	0.0330	0.0089	5.57E-05	-254.45	63.13	16821	2571
	GenADA	0.0116	0.0039	3.30E-02	-351.65	164.98	7135	1577
Intestinibacter*	ADc12	0.1903	0.0154	1.01E-07	-3014.37	566.17	70292	2571
	GenADA	0.0024	0.0024	9.22E-02	-246.05	146.10	2152	1577
Lachnospira*	ADc12	0.0034	0.0067	3.30E-04	530.79	147.83	2997	2571
	GenADA	0.0004	0.0021	1.11E-01	113.71	71.44	439	1577
LachnospiraceaeUCG008	ADc12	0.0784	0.0081	1.08E-04	-776.50	200.53	35344	2571
	GenADA	0.0041	0.0017	1.57E-01	197.29	139.40	3224	1577
Oscillibacter	ADc12	0.0124	0.0068	3.08E-04	-630.36	174.67	8305	2571
	GenADA	0.0051	0.0027	7.55E-02	302.62	170.24	3884	1577
Prevotella9*	ADc12	0.0084	0.0075	1.57E-04	-556.34	147.21	6609	2571
	GenADA	0.0008	0.0031	5.47E-02	-147.32	76.68	819	1577
Roseburia*	ADc12	0.2061	0.0100	1.35E-05	-3564.78	819.18	78446	2571
	GenADA	0.0037	0.0027	7.64E-02	-366.88	207.02	3135	1577
Sutterella	ADc12	0.4928	0.0071	2.53E-04	3594.16	982.20	124985	2571
	GenADA	0.0002	0.0009	3.08E-01	-34.75	34.10	156	1577
Veillonella	ADc12	0.0070	0.0081	8.83E-05	539.94	137.72	5406	2571

GenADA	0.0021	0.0085	1.56E-03	369.43	116.78	1843	1577
--------	--------	--------	-----------------	--------	--------	------	------

Note: The association from the “best-fit” threshold was generated from PRSice-2 with a range of p -value thresholds from 5×10^{-8} to 1 and incremental interval of 5×10^{-5} ; R^2 : Variance explained by the PRS model; P : p -value of model fit for the association; Coeff: Coefficient of the model; SE: standard error; # of SNPs: Number of SNPs included in the model at the specified threshold. Genera in bold are three genera identified to have genetically significant association with AD in both discovery and replicate samples. *: indicated ten genera that have the same direction in both discovery and replicate samples. Seven genera, originally identified to be significantly associated with AD in the discovery sample, did not survive the replicate analysis due to the opposite direction in the replicate sample.

Table 4. Association between PRSs for nine significant gut microbiota genera and *APOE* rs429358.

Trait	Samples	Coeff.	SE	z-value	p	OR (95%CI)
Adlercreutzia	ADc12	-0.014	0.009	-1.643	0.1006	0.99(0.97 - 1.00)
	GenADA	-0.026	0.011	-2.309	0.0211	0.97(0.95 - 1.00)
	meta-analysis	-0.019	0.007	-2.709	0.0068	0.98(0.97 - 0.97)
Candidatus Soleaferrea	ADc12	-0.020	0.009	-2.242	0.0250	0.98(0.96 - 1.00)
	GenADA	-0.004	0.007	-0.486	0.6271	1.00(0.98 - 1.01)
	meta-analysis	-0.011	0.008	-1.366	0.1719	0.99(0.97 - 0.97)
Collinsella	ADc12	0.037	0.008	4.413	1.06E-05	1.04(1.02 - 1.06)
	GenADA	0.033	0.011	2.930	0.0034	1.03(1.01 - 1.06)
	meta-analysis	0.036	0.007	5.278	<.0001	1.04(1.02 - 1.02)
Eisenbergiella	ADc12	-0.029	0.008	-3.382	0.0007	0.97(0.96 - 0.99)
	GenADA	0.002	0.011	0.180	0.8575	1.00(0.98 - 1.02)
	meta-analysis	-0.014	0.015	-0.922	0.3565	0.99(0.96 - 0.96)
Eubacterium fissicatena	ADc12	-0.032	0.009	-3.557	0.0004	0.97(0.95 - 0.99)
	GenADA	-0.003	0.013	-0.252	0.8011	1.00(0.97 - 1.02)
	meta-analysis	-0.019	0.014	-1.353	0.1762	0.98(0.95 - 0.95)
Gordonibacter	ADc12	-0.032	0.009	-3.460	0.0005	0.97(0.95 - 0.99)
	GenADA	-0.004	0.012	-0.352	0.7248	1.00(0.97 - 1.02)
	meta-analysis	-0.019	0.014	-1.398	0.1621	0.98(0.95 - 0.95)
Lachnospira	ADc12	0.014	0.009	1.524	0.1277	1.01(1.00 - 1.03)
	GenADA	0.014	0.012	1.169	0.2425	1.01(0.99 - 1.04)
	meta-analysis	-0.017	0.018	-0.969	0.3324	0.98(0.95 - 0.95)
Prevotella9	ADc12	-0.012	0.008	-1.486	0.1373	0.99(0.97 - 1.00)
	GenADA	-0.020	0.011	-1.836	0.0666	0.98(0.96 - 1.00)
	meta-analysis	-0.015	0.007	-2.291	0.0220	0.99(0.97 - 0.97)
Veillonella	ADc12	0.024	0.008	2.946	0.0032	1.02(1.01 - 1.04)
	GenADA	-0.008	0.011	-0.664	0.5071	0.99(0.97 - 1.01)
	meta-analysis	0.0094	0.016	0.5905	0.5549	1.01(0.98 - 1.04)

Note: Genera in bold are three genera identified to have genetically significant correlation with *APOE* rs429358 minor allele C in the meta-analysis ($p < 0.05$). *Collinsella* was the only genus that showed significant correlation in both discovery and replication samples.

Table 5. List of gut microbiome genera that were nominally associated with *APOE* SNPs rs429358 and rs7412.

rsID (CHR:BP:Effect Allele)	Microbiome Genera	Beta	SE	SZ	P	N	# cohorts
rs429358 (19:45411941:C)	Bacteroides	-0.044	0.015	-2.881	0.0040	18173	23
	Butyricimonas	-0.044	0.020	-2.241	0.0250	10657	23
	Dorea	0.030	0.015	2.127	0.0334	17494	23
	Eubacterium coprostanoligenes	0.033	0.015	2.120	0.0340	17261	23
	Faecalibacterium	-0.035	0.015	-2.188	0.0287	17960	23
	Olsenella	0.064	0.033	1.993	0.0462	3721	13
	Parasutterella	0.040	0.019	2.216	0.0267	11291	23
	Senegalimassilia	0.050	0.024	2.207	0.0273	6888	21
	Veillonella	0.044	0.021	2.032	0.0421	9194	23
rs7412 (19:45412079:T)	Butyricoccus	0.048	0.021	2.468	0.0136	15637	20
	Collinsella	0.052	0.023	2.214	0.0268	12811	19
	Coprococcus3	0.050	0.022	2.344	0.0191	14323	20
	Eubacterium hallii	0.053	0.021	2.453	0.0142	14846	20
	Lachnospiraceae UCG001	-0.071	0.027	-2.534	0.0113	9357	20
	Olsenella	-0.103	0.043	-2.346	0.0190	3721	13
	Ruminococcaceae UCG004	0.080	0.028	2.901	0.0037	9049	20
	Senegalimassilia	0.082	0.032	2.584	0.0098	6508	18

Note: The data were extracted from the largest GWAS of human gut microbiome conducted among 24 multi-ethnic cohorts in the MiBioGen consortium (www.mibiogen.org) (24). rsID (CHR:BP:Effect Allele): the rs-number of single nucleotide polymorphisms, chromosome number (CHR), base pair (BP), with the effect allele (genome assembly GRCh37/hg19). Microbiome: only contains the genus level with significant abundance associated with *APOE* two SNPs. Beta: Beta coefficient. SE: Standard Error. SZ: Weighted sum of z-scores. P: p-value (p<0.05 was considered as nominal association between the microbiome genera and *APOE*). N: sample count. # Cohorts: Number of cohorts involved.