Gut Microbiome Multi-Omics and Cognitive Function in the Hispanic Community Health Study/Study of Latinos- Investigation of Neurocognitive Aging

Natalia Palacios1,2,3, Scott Gordon1, Tao Wang4, Robert Burk4, Qibin Qi4, Curtis Huttenhower3,5,6,7, Hector M. Gonzalez8, Robert Knight9,10,11, Charles De Carli12, Marta Daviglus13, Melissa Lamar14, Gregory Telavera15, Wassim Tarraf16, Tomasz Kosciolek17, Jianwen Cai18, Robert C. Kaplan4

1Department of Public Health, University of Massachusetts Lowell, Lowell, MA, 01850.
2Bedford VA Healthcare System, Geriatric Research and Education Clinical Center, Bedford, Massachusetts, USA, 01730.
3Harvard Chan Microbiome in Public Health Center (HCMPH), Boston, MA, 02115.
4Albert Einstein College of Medicine, Bronx, NY, 10461.
5Broad Institute of MIT and Harvard, Cambridge, MA, 02142.
6Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, 02115.
7Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, 02115.
8University of California, San Diego Department of Neurosciences, San Diego, CA, USA 92182.
9Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA, 92182.
10Department of Pediatrics, University of California, San Diego, La Jolla, CA, USA, 92182.
11Center for Microbiome Innovation, University of California, San Diego, La Jolla, CA, USA, 92182.
12University of California, Davis, CA, USA, 95616.
13University of Illinois College of Medicine Institute for Minority Health Research, Chicago, IL, 60612
14Rush University Alzheimer’s Disease Center, Rush University, Chicago, IL, 60612
15South Bay Latino Research Center, San Diego State University, Chula Vista, CA, 92182.
16Wayne State University, Detroit, MI, 48202
17Sano Centre for Computational Precision Medicine, Krakow, Poland
18Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599.

Corresponding author:
Natalia Palacios
Department of Public Health
University of Massachusetts, Lowell
STRUCTURED ABSTRACT

INTRODUCTION: We conducted a study within the Hispanic Community Health Study/Study of Latinos- Investigation of Neurocognitive Aging (HCHS/SOL-INCA) cohort to examine the association between gut microbiome and cognitive function.

METHODS: We analyzed the fecal metagenomes of 2,471 HCHS/SOL-INCA participants to, cross-sectionally, identify microbial taxonomic and functional features associated with global cognitive function. Omnibus (PERMANOVA) and feature-wise analyses (MaAsLin2) were conducted to identify microbiome-cognition associations, and specific microbial species and pathways (Kyoto Encyclopedia of Genes and Genomes (KEGG modules) associated with cognition.

RESULTS: Eubacterium species (E. siraeum and E. eligens), were associated with better cognition. Several KEGG modules, most strongly Ornithine, Serine biosynthesis and Urea Cycle, were associated with worse cognition.

DISCUSSION: In a large Hispanic/Latino cohort, we identified several microbial taxa and KEGG pathways associated with cognition.
BACKGROUND

The human gut microbiome comprises trillions of bacteria, more microorganisms than the number of cells in a human body and 150 times as many genes as the human genome[1]. The microbiome plays key roles in absorption and metabolism of nutrients, breakdown of lipids and polysaccharides, detoxification of xenobiotics, waste particles and pathogenic organisms[2, 3], control of gut motility and homeostasis of the intestinal barrier[4, 5]. It is key to the development and function of the human immune system, with 70% of the body’s lymphocytes located in the gut[6, 7]. The intestinal environment is intimately connected to the central nervous system (CNS) through bi-directional communication pathways regulated, in part, by the gut microbiome via production of bioactive metabolites[8, 9]: the brain-gut-microbiome axis. This communication system is just beginning to be appreciated and includes neural, immune,[10] endocrine, and metabolic pathways[9].

US Hispanics/Latinos are at an increased risk of AD/ADRD and accelerated cognitive decline compared to non-Hispanic/Latino whites[11-14]. Hispanics/Latinos are also one of the fastest growing segments of the US elderly population[15, 16], and are impacted by cognitive decline at younger ages[17-19] compared to non-Hispanic Whites. This elevated risk has been suggested to be due to attributed to higher prevalence of key AD/ADRD risk factors in Hispanics/Latinos, such as poor diet, higher prevalence of diabetes, metabolic syndrome, obesity and cardiovascular disease, and higher blood pressure, which have been documented in HCHS/SOL-INCA[11, 20-27]. A distinct microbial profile in Hispanics/Latinos may also contribute to increased risk[28].

Gut inflammation has been associated with AD/ADRD risk[29], suggesting that the gut microbiome may contribute to ‘inflamming’ in AD/ADRD. Prior studies focusing on microbiome and AD/ADRD have reported reduction in diversity[30], and lower abundance of beneficial anti-inflammatory taxa, such as *Eubacterium spp.*[31] and *Faecalibacterium spp.*[32, 33] in AD/ADRD. Several species of gut bacteria, such as *Bifidobacterium*, are capable of production of neuroactive metabolites, such as serotonin and GABA.[34-36] Several bacteria including the *Klebsiella, Escherichia, Streptococcus, and Salmonella, Pseudomonas*, species secrete functional amyloid proteins with demonstrated capacity to cross-seed and trigger a cascade of amyloid protein misfolding[37, 38] that can propagate from the gastrointestinal tract to the brain. These amyloid-producing taxa have been reported elevated in AD/ADRD in prior studies[39-41]. Furthermore, secretion of lipopolysaccharide (LPS), a cell wall component of gram negative bacteria[42], can lead to microglial priming, increased production of proinflammatory cytokines, neuroinflammation and neurodegeneration[43]. Pathogens, such as *K. pneumoniae and E. lenta* [44, 45], and pro-inflammatory taxa, such as the sulfate-producing *Desulfovibrio* have been noted in patients with AD/ADRD.

Functional work on the microbiome and AD has implicated both inflammatory and metabolic processes in AD. For example, finding of dysregulation of the P-glycoprotein microbial pathway[46] in AD, suggests a contribution of intestinal inflammation and gastrointestinal infections. Other studies reported dysregulation in pathways related to glucose metabolism and mitochondrial disfunction[47].

Animal studies have supported this epidemiological evidence in humans, reporting altered microbial composition in transgenic AD mouse models[48], and absence of amyloid plaque build-up in germ free mice[49]. Transgenic mice treated with an antibiotic cocktail had fewer insoluble amyloid β plaques and less microglia and astrocyte accumulation around existing amyloid plaques[50]. In a recent report based on a fecal microbiome transfer (FMT) from AD patients to microbiota-depleted healthy young adult rats, FMT resulted in AD symptoms in the AD-colonized rats, as well as changes in the rat cecal and hippocampal metabolomes[51]. The abundance of the pathobiont, pro-inflammatory *Desulfovibrio* was elevated in AD patients compared to healthy controls, inversely correlated to MMSE scores and was the taxa most elevated in the AD-colonized rats after FMT[51]. *Desulfovibrio* has been reported enriched in AD patients by several other studies[46, 52] and has been associated with reduced cecal SCFA levels[53].

To knowledge, no work on the gut microbiome and AD risk has been done in Hispanics/Latinos. We conducted the first study of the gut microbiome in relation to cognitive function in a Hispanics/Latinos cohort.
METHODS

The study was approved by the institutional review boards at Albert Einstein College of Medicine and the University of Massachusetts at Lowell.

Study Population: The study was conducted within Hispanic Community Health Study/Study of Latinos (HCHS/SOL) a large population-based, multisite, prospective cohort study cohort of US Hispanic adults supported by the National Heart Lung and Blood Institute (NHLBI) and other National Institutes of Health (NIH) institutes. Participants (N=16,415) were enrolled during 2008-2011 at four Field Centers located in US cities with large Hispanic/Latino populations: Bronx, NY; Chicago, IL; Miami, FL; and San Diego, CA. Participants are of Cuban, Dominican, Puerto Rican, Mexican, and South American backgrounds. Middle-aged and older Hispanics/Latinos (ages 45-74 years) were oversampled (n=9,652). At study baseline, participants participated in clinic visits with bilingual technicians that included anthropometric measures, blood pressure readings, pulmonary testing, diet assessment as well as a blood draw (with blood frozen and stored) [54, 55].

Neurocognitive assessment: At the second HCHS/SOL visit (V2), neurocognitive testing was conducted among 6,377 participants, at the 2nd HCHS/SOL study visit completed during 2011-2014. All cognitive assessments were performed in-clinic by trained bilingual/bicultural technicians with oversight of a Neurocognitive Reading Center. All three waves included Six-Item Screener[56] (SIS; mental status), (2) Brief-Spanish English Verbal Learning Test (BSEVLT; verbal episodic learning and memory)[57] (3) Verbal Fluency[58] and (4) Digit Symbol Subtest (DSS; processing speed).[59] At v2 and v3, additional testing included the Trail Making Test (TMT; A&B; a test of cognitive function), the NIH Toolbox Picture Vocabulary Test (PVT; test of general cognitive ability/crystalized knowledge)[60] and the 12-item Everyday Cognition (eCog-12) scale of (memory, language, visuospatial, planning, organization, and divided attention)[61]. In this study, our cognitive outcome variable was global cognition scores (GSC), derived from confirmatory factor analyses based on the cognitive measures above[62].

Stool Sample Collection and Storage: During HCHS/SOL V2, participants were provided with a stool collection kit for self-sampling using a disposable paper inverted hat (Protocult collection device, ABC Medical Enterprises, Inc., Rochester, MN). CHS/SOL collected and sequenced over 3,035 stool samples during the V2 phase, placed in RNAlater[28]. Of these 2,470, were part of HCHS/SOL-INCA and had GCS, stool metagenomic and were included in this study.

Fecal sample processing and sequencing: Shotgun sequencing was performed in the laboratory of Dr. Knight, using previously described procedures[63]. Briefly, DNA was extracted from fecal samples on FTA cards, following the Earth Microbiome Project protocol[64, 65]. Input DNA was quantified, and thereafter DNA fragmentation is performed, followed by end-repair and A-tailing. Adapters and barcode indices are added following the iTru adapter protocol[66]. Each plate of 384 libraries is generated without repeating barcodes, eliminating the problem of sequence misassignment due to barcode swapping[67]. The libraries were purified, quantified and normalized for sequencing on Illumina NovaSeq. Sequencing depth ranged from 500K-8,945K reads/sample, with average 955K reads/sample.

Bioinformatic profiling of fecal samples: FASTQ sequence reads were demultiplexed, filtered to remove reads mapping to the human genome, and trimmed to remove low quality bases. The reads are then aligned against the NCBI RefSeq representative prokaryotic genome collection (release 82)[68] using Bowtie2,[69] and per strain coverage calculated using default SHOGUN[63] settings. The α-diversity indices (Shannon) and β-diversity (Bray-Curtis dissimilarity) were calculated using R vegan packages[70, 71]. Functional profiles were obtained using SHOGUN via sequence alignment to a nucleotide gene database derived from NCBI RefSeq (release 82) and annotated with Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology[72, 73].

Metabolomic Profiling: Metabolomics data, capturing >600 metabolites within eight defined classes, was obtained using fasting plasma specimens collected at HCHS/SOL V2. Metabolomics profiling was conducted...
via Metabolon HD4Discovery platform (Metabolon Inc., Durham, NC) (Morrisville, NC) using liquid chromatography-MS/MS methods with positive ion and negative ion modes (Waters ACQUITY ultra-performance liquid chromatography, as previously described[74]. Of the 2,471 participants with metagenomic and cognitive data at V2, 480 had concurrent serum metabolome data. Data on these 2 participants were used in multi-omic analyses below.

Assessment of Covariates: Comprehensive covariate data were collected during in-person examinations, as well as annual telephone interviews and ongoing ascertainment of major incident health events. Our model was adjusted for factors previously associated with AD, such as age, biological sex, education, AHEI2010 dietary index, metformin use, Hispanic background (Central American, South American, Dominican, Mexican, Cuban, Puerto Rican), age moved to US, fiber intake and diabetes status. Age, education, metformin use[75], Hispanic background and years living in the US[28] were collected as part of in-person interview at V2.

Definition of diabetes and prediabetes are based on contemporaneous ADA criteria by a combination of diagnosis and measured levels of fasting and 2hr glucose and hemoglobin A1c tests.[76] Long-term diet is one of the primary factors driving human microbiome composition and function[77], including among immigrants[78, 79]. In HCHS/SOL, all participants completed, at baseline, two 24h dietary recalls at ~6 week intervals, to capture usual intake of foods and nutrients[80], from which fiber use and Alternative Healthy Eating/AHEI score were derived. We adjusted our analyses for use of Metformin use, as it is used by ~12% of our participants, is the most common diabetes medication, has been shown to lead to alterations in gut microbiome composition and function.[81] We adjusted for Hispanic background and, in sensitivity analyses, for years living in the US as these factors were strongly related to the microbiome of HCHS/SOL participants in our prior studies.

Statistical Analysis: All analyses were conducted on 2,471 HCHS/SOL-INCA participants who had concurrent GCS and stool metagenomic data (Figure 1A). Analyses involving the metabolome, were conducted in the 480 participants who also had serum metabolomic data. Our initial dataset contained 3,314 microbial species after quality control and removal of host ‘contaminant’ sequences, and of these, 198 were retained after filtering (as detailed above) and used in our statistical analyses. Likewise, of 445 KEGG Modules, 216 were retained after filtering (Figure 1B).

Global Cognitive Function Variable and Categories: GCS was treated as a continuous outcome variable in MaAsLin2 and univariate PERMANOVA analyses. For analyses requiring categorical outcomes, including pCoA visualization, GCS was re-categorized into an ordinal variable with the following values: better (>=1SD above mean GCS; N = 413), medium (between 1SD above and 1SD below mean GCS; N = 1621) and worse (>1SD below mean GCS; N = 429).

Overall community patterns of microbial variation: The Bray-Curtis dissimilarity metric was used for all beta-diversity analyses, both of taxonomic composition and functional potential. We performed ordination via the Principal Coordinates Analyses (PCoA), to visualize Beta-diversity relationships across the three GCS categories (better, medium, worse cognition). We performed omnibus testing with permutational multivariate analysis of variance (PERMANOVA) of Bray-Curtis dissimilarities to quantify the percent variance explained by GCS (continuous) and study covariates.

Feature-wise analyses: To identify microbial features and functions associated with GCS, we used Multivariate Association with Linear Models version 2 (MaAsLin2), Version 2.1.16.0, (https://huttenhower.sph.harvard.edu/maaslin2)[82]) a modified general linear model for feature-wise multivariate testing in microbial community profiles. MaAsLin2 was run in R Version 4.3.1. Model specification and covariates were as follows:

```r
  microbial feature ~ global cognitive score (continuous) + age + biological sex + education + AHEI2010 diet index + metformin use + Hispanic background + age moved to US + fiber intake + diabetes status
```
We conducted sensitivity analyses MaAsLin2 including 1) univariate models, 2) additional adjustment for pack years smoking, chronic stress, depression (CESD10), hypertension, waist-hip ratio, alcohol intake, whether participant was US born, and the Multi-Ethnic Study of Atherosclerosis MESA acculturation score[83], a score representing extent of acculturation to the US.

For features (taxa and KEGG modules) most strongly associated with cognition (FDR <= 0.2), we examined the Spearman correlation of these cognition-associated taxa with cognition-associated KEGG modules. To understand the relationship between the identified taxa and KEGG modules to key serum microbial metabolites in serum, we examined the Spearman correlation of these cognition-associated taxa and KEGG modules with serum metabolites from the HCHS/SOL V2 collection. We performed these correlation analyses for serum metabolites within the Short Chain Fatty Acid (SCFA), Branched Chain Amino Acid (BCAA) and tryptophan metabolism networks.

RESULTS

Table 1 outlines the characteristics of the study participants with cognitive measures and metagenomic profiles at HCHS/SOL V2 who were included in this study. The average age of study participants was 60.83 years, 10% were US born. The highest proportion of participants (38.46%) were of Mexican heritage, followed by Puerto Rican (18.30%), Cuban (13.85%), Dominican (10.28%), Central American (9.88%), South American (7.04%) and mixed (2.02%) heritage. Most participants were either had pre-diabetes (47%) or diabetes (34%), 12% reported metformin use, and 54.04 had hypertension.

Overall patterns of microbial community variation in relation to GCS: Applying univariate PERMANOVA of Bray-Curtis dissimilarities, we observed variables such as the MESA Acculturation Scale, Sex, Puerto Rican descent and age moved to the US explained the greatest amount of variation in gut microbiome composition (Figure 2A, Taxa), while metformin use, age and diabetes explained the largest amount of variation in functional potential (Figure 2A, KEGG Modules), consistent with our prior publications[28]. Diet (AHEI dietary score), fiber intake, diabetes, and several other lifestyle variables, also explained significant amount of variation in microbiome composition (Figure 2A, Taxa) and function (Figure 2A, KEGG Modules). In multivariate PERMANOVA analyses, global cognitive function did not explain a significant amount of variation in microbial taxa (R² = 0.07%; p = 0.22), or functions (R² = 0.07%; p = 0.24) (Figure 2B).

Microbial species associated with GCS: The most abundant species in stool included Bacteroides vulgatus, Prevotella copri, Bacteroides uniformis and Eubacterium rectale (Figure 2C, Taxa). The most abundant KEGG modules included Ribosome bacteria M00178, Ribosome archaea M00179, Aminoacyl-tRNA biosynthesis prokaryotes M00360, Aminoacyl-tRNA biosynthesis eukaryotes M00359 and ABC-2 type transport system (BR:ko02000) M00254 (Figure 2C, KEGG Modules).

In feature-wise, MaAsLin2 [84] models, after adjustment for age, biological sex, education, AHEI2010 diet index, metformin use, Hispanic background, age moved to US, fiber intake and diabetes status, we observed an inverse association between several bacterial species, including Prevotella sp P4.76, Blifobacterium longum Prevotella bryantii, Desulfovibrio piger Ruminococcus faecis, and GCS. Abundance of several bacterial species, most strongly Clostridium phoenicesis, Eubacterium eligens, and also Eubacterium siraeum, Holdemania massiliensis, Intestimonas massiliensis, Bacteroides barnesiae, and others was positively associated with global cognitive function (Figure 3A). In sensitivity analyses, after additional adjustment for pack years smoking, chronic stress, depression (CESD10), hypertension, waist-hip ratio, alcohol intake, whether participant was US born, and the MESA acculturation score, abundance of Prevotella sp P4.76 and Ruminococcus faecis remained inversely and abundance of Intestimonas massiliensis and Eubacterium eligens remained positively associated with GCS (Supplemental Figure 1).

Microbial functions associated with GCS: MaAsLin2 identified 14 KEGG Modules significantly associated with GCS (Figure 3B). Of these, Ornithine biosynthesis modules (M00536), Urea cycle (M00029), Serine biosynthesis glyceral 3P serine (M00020), Triacylglycerol biosynthesis (M00089), and Putative ABC
As expected, taxa associated with better cognition were, for the most part, correlated with each other (Supplemental Figure 2A). This was also true for taxa associated with worse cognition, with the exception of *B. longum*, the association of which with worse cognition is surprising as it is generally recognized as a beneficial taxon. Thus it's consistent with prior work that *B. longum* might correlate with beneficial taxa. Similarly, KEGG modules associated with better cognition were positively correlated with each other and negatively correlated with modules associated with rose cognition (Supplemental Figure 2B). We observed a modest correlation between, the top taxa and KEGG modules positively associated with cognition, and analogously a correlation between taxa and pathways negatively associated with cognition (Figure 4A). *Alistipes shahii* was most strongly correlated to various cognition-associated KEGG modules in this study (Figure 6), with strong correlations with the chondroitin degradation (*r* = 0.88, FDR P < 0.05), sulfate deradation (*r* = 0.88, FDR P < 0.05) and the pectin degradation (*r* = 0.87, FDR P < 0.05) KEGG modules among others. Other strong correlations included the positive correlation between abundance of *Bacteroides plebeius* and HydH HydG metal tolerance two component regulatory system (*r* = 0.45, FDR P < 0.05) and the Formaldehyde assimilation ribulose monophosphate module (*r* = 0.35, FDR P < 0.05).

Several cognition-associated taxa identified in this study were correlated with concurrently collected (HCHS/SOL V2) serum metabolites within SCFA, BCAA and tryptophan metabolic networks. Consistent with results previously published by our group[85], *B. longum*, a beneficial taxa surprisingly associated with worse cognition in our study, was negatively correlated to most metabolites in the BCAA pathway, and strongly positively correlated with indolepropionate (*r* = 0.19, FDR P < 0.05). *B. longum* was also inversely correlated with most BCAA metabolites. We also observed, a positive correlation between indolepropionate and many of taxa positively associated with cognition, including *E. eligens* (*r* = 0.28, FDR P < 0.05) and *E. siraeum* (*r* = 0.1, FDR P < 0.05), *I. massiliensis* (*r* = 0.1, FDR P < 0.05), *A. obesi* (*r* = 0.12, FDR P < 0.05) and *A. shahii* (*r* = 0.13, FDR P < 0.05) (Figure 4B). Indolepropionate was also correlated with several modules associated with cognition, such as eg. Phenylalanine biosynthesis chorismate phenylalanine (*r* = 0.17, p<0.05), (Figure 4C).

**DISCUSSION**

To our knowledge, this is the first, largest, and most comprehensive multi-omic study of the gut microbiome in a Hispanic/Latino cohort to date. Overall, our findings are consistent with prior reports of a pro-inflammatory, pathogen-enriched microbiome in cognitive decline and AD/ADRD[32, 33, 41, 45, 86]. In our MaAsLin2 regression models, abundance of anti-inflammatory, strictly anaerobic taxa, such as *Eubacterium*, e.g.: *E.eligens*, *E. siraeum*, and *E.ventriosum* was associated with better cognition. We noted an inverse association between the abundance of the sulfate-producing pathobiont *Desulfovibrio* and cognitive function, this association has been seen in prior animal and human studies[51]. Notable also, is the relationship between the identified taxa, and KEGG modules to serum metabolites, particularly BCAA and tryptophan.

In this study, higher abundance of *Alistipes obesi* and *Alistipes shahii* were associated with better cognitive function. Several *Alistipes* spp., including *Alistipes shahii* can hydrolyze tryptophan to indole[87]. Tryptophan is an essential amino acid, metabolized via host/kynurenine pathway (kynurenate, xanthurenic and
A Bacteroides cellulosilyticus, Bacteroides plebeius and Bacteroides barnesiae were associated with better cognitive function in this study. Several Bacteroides species are involved in microbiota-derived γ-
aminobutyric acid GABA[92], due to carriage a key enzyme a pyridoxal-5′-phosphate- dependent GAD (gadB) for converting glutamate to GABA in Bacteroides genomes. Prior work reported mixed associations between Bacteroides abundance and AD, with some studies reporting increased abundance in AD[45] and MCI[93], and others decreased abundance[94]. Bacteroides are gram-negative bacteria, that produce Lipopolysaccharide (LPS) as part of their cell wall[42]. LPS has been shown to lead to microglial priming, increased production of proinflammatory cytokines, neuroinflammation and neurodegeneration[43]. Bacteroides have been reported as elevated in persons with diabetes[95]. Thus, this study adds to the mixed evidence regarding Bacteroides and cognition that should be examined in future work. Two Prevotella species, Prevotella sp P476 and Prevotella bryantii were associated with worse cognitive function in our study. Prior work on Prevotella in AD have been mixed. Lower Prevotella abundance has been reported in patients with MCI[96] in one study, while another study reported enrichment of Prevotella in AD[97]. Prevotella abundance has been associated with risk of Parkinson’s disease, another neurodegenerative disease[98]. Interestingly, in HCHS/SOL, a higher Prevotella to Bacteroides ratio was associated with obesity [28], while the opposite association had been reported in other populations[79].

Bifidobacterium longum was associated with worse cognitive function in our study. This result is somewhat surprising, given that Bifidobacteria are generally considered to be anti-inflammatory and beneficial to health, promoting gut epithelial integrity[99] and that B. longum was negatively correlated to most metabolites in the BCAA pathway, and strongly positively correlated with indolepropionate, a beneficial microbial metabolite, that has also been associated with reduced diabetes risk[85]. Other taxa associated with worse cognition in SOL, exhibited the opposite pattern vis. a vis these metabolites. Most, but not all prior studies have noted reduced Bifidobacterium abundance in AD.[45] However, increased abundance of Bifidobacterium[100-108] has been observed in Parkinson’s disease, including in recent work by our group[109]. Furthermore, Bifidobacteria abundance has been associated with aging and have been shown to be elevated in centenarians[110]. Interestingly, and more consistent with a protective role for Bifidobacterium, another Bifidobacterium species, B. dentium was the top species contributing to the discrimination between participants with better vs. worst cognition, with a positive (beneficial) variable importance. Bifidobacteria are in involved in GABA, and acetate production, and B. dentium is thought to be most prolific GABA producer among Bifidobacteria[111]. These mixed results highlight the need to further examine the role of Bifidobacteriina in AD/ADRD. Similarly, abundance of R. faecalis was associated with worse cognitive function in our study. Ruminococcus are generally beneficial bacteria[112], so the inverse association with cognition is surprising.

We noted dysregulation in KEGG pathways involving metabolism of neuroactive molecules (eg. GABA, ornithine), sugars (eg. ribose and glucose) and BCAA with global cognition. The GMB has a known capacity to produce as well as metabolize neurotransmitters, influencing neurotransmitter levels in the brain.[113] Several bacterial species, including members of Bifidobacterium Bacteroides and Lactobacillus are capable of GABA production.[114] The GABAergic system has been proposed as a potential therapeutic target for AD.[115, 116] Ornithine, derived from L-arginine and glutamate, has been implicated in the pathogenesis of AD.[117, 118] Shikimate, a tryptophan metabolism pathway, was identified as one of the pathways associated with cognition.
in our study, and has been linked with neurodegeneration and Alzheimer’s disease. This KEGG module correlated positively with tryptophan metabolite concentration in our sample. Serum tryptophan was correlated with many of the KEGG modules identified in this study, positively with modules associated with better cognition and negatively with KEGG modules associated with worse cognition. We observed correlations between other metabolites involved in tryptophan metabolism, such as the *B. longum* and indolepropionate as well as *B. cellulosityticus* and kyturenate and 3- indoxyl sulfate, suggests a potential role for microbial tryptophan metabolism in AD/ADRD.

Our study has several notable strengths. To our knowledge this is the largest and the first multi-omic study of the microbiome in Hispanics/Latinos to date. HCHS/SOL contained a culturally appropriate, comprehensive assessment of cognitive function, incorporating multiple cognitive domains. Simultaneously collected data on GMB composition, function (KEGG modules) and serum metabolites in HCHS/SOL, allowed us to conduct our multi-omic analyses, examining the inter-relationships between these features. The study also has several limitations. Because the microbiome and cognition were, at the time of writing, were only concurrently available at visit 2 in HCHS/SOL, our study was limited to a cross sectional design. Future work will consider the longitudinal relationship between GMB and cognitive decline and AD/ADRD risk. While the stool samples in this study were self-collected by the participants and based on a single collection, extensive work has shown that that self-collection is effective the microbiome is very stable with between-person variability greatly exceeding within-person differences over time. HCHS/SOL-INCA participants are relatively young (mean age = 60.83), however they also have a high burden of AD-related comorbidities, such as a prevalence of diabetes or prediabetes over 80%, increasing their AD/ADRD risk beyond that appropriate for their calendar age.

In summary, our work in this large, well characterized Hispanic/Latino cohort points to a subtle, but noticeable dysregulation of the gut microbiome associated with cognitive function. We confirmed prior reports of a positive association between anti-inflammatory taxa, and negative association with pathogenic taxa and cognitive function. We also identified several KEGG pathways, with global cognitive function, which correlated strongly with serum tryptophan levels.
REFERENCES


**FIGURE LEGENDS:**

**Figure 1.** Microbiome and cognition in HCHC/SOL-INCA study design. **Figure 1A:** Sequencing and profiling of HCHC/SOL-INCA Visit 2 samples with microbiome and metabolomics data. **Figure 1B:** STORMS flowchart outlining enrollment, data collection, and processing for our study of human gut metagenomes and metabolomes in relation to global cognitive function.

**Figure 2.** Overall microbiome structure in relation to GCS and other study covariates. **Figure 2A:** Univariate R2 (Bray-Curtis dissimilarities, PERMANOVA) explained by key covariates. **Figure 2B:** Distributions of taxonomic Bray-Curtis dissimilarities by phenotype group show that overall microbiome structure does not vary significantly with global cognitive function. Multivariate PERMANOVA adjusted for age, gender, education, diet (AHEI), metformin use, Hispanic background (list). **Figure 2C:** Most abundant Taxa and KEGG Modules in HCHC/SOL-INCA according to GCS.

**Figure 3.** Taxa and KEGG modules significantly associated with GCS. Volcano plots show beta coefficient (x-axis) and FDR-corrected log10(q value) (y-axis). The most significantly enriched and depleted taxa (Figure 3A) and KEGG modules (Figure 3B) associated with phenotype group identified via feature-wise analyses (MaAsLin 2). All analyses consider global cognitive function as a continuous predictor and are adjusted for age, gender, education, diet (AHEI), metformin use, Hispanic background. All p-values are presented with FDR correction for multiple comparisons.

**Figure 4.** Multi-omic correlations, taxa, KEGG modules and metabolites. Correlation coefficients are partial Spearman correlations adjusted for age, gender, education, diet (AHEI), metformin use, Hispanic background. **Figure 4A:** We observed modest, but consistent correlation between top cognition-associated taxa and KEGG modules (FDR <0.2). Cognition-positive taxa were generally correlated with cognition-positive pathways and, likewise, cognition-negative taxa were overall associated with cognition-negative pathways. **Figure 4B:** Correlation between top cognition-associated taxa (FDR <0.2) with and metabolites in the SCFA, BCAA and tryptophan pathways in HCHS/SOL. **Figure 4C:** Correlation between top cognition-associated KEGG modules (FDR <0.2) with and metabolites in the SCFA, BCAA and tryptophan pathways in HCHS/SOL-INCA.
SUPPLEMENTAL FIGURE LEGENDS:

Supplemental Figure 1. Correlations within a. Taxa and b. KEGG modules associated with cognition in HCHC/SOL-INCA. Spearman (unadjusted correlations) within the top Supplemental Figure 1A: Taxa and Supplemental Figure 1B: KEGG modules associated with cognition are. Numbers in cells denote correlation coefficients, '*' represents correlation p<0.05.

Supplemental Figure 2. Volcano Plot for MaAsLin2 model for taxa (Supplemental Figure 2A) and KEGG modules (Supplemental Figure 2B) and with additional adjustment for pack years smoking, chronic stress, depression (CESD10), hypertension, waist-hip ratio, alcohol intake, whether participant was US born, and the Multi-Ethnic Study of Atherosclerosis MESA acculturation score[83].

ACKNOWLEDGEMENTS: The authors thank the Hispanic Community Health Study/Study of Latinos- Participants for their participation in the study.

CONFLICTS: Curtis Huttenhower serves on the Scientific Advisory Board for Seres Therapeutics and Empress Therapeutics. Tomasz Kosciolek serves as a scientific adviser to Human Biome Institute. Other co-authors have nothing to disclose.

FUNDING: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The Hispanic Community Health Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN268201300001I / N01-HC-65233), University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago (HHSN268201300003I / N01-HC-65236 Northwestern Univ), and San Diego State University (HHSN268201300005I / N01-HC-65237). The following Institutes/ Centers/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements.

This work was additionally supported by RF1AG075922, R01NS097723, 1R01MD011389, RF1AG054548, 1R01DK119268, 1R01AG048642, 1R01DK134672, R01 AG075758.

DATA AVAILABILITY: HCHS/SOL data are archived in the dbGap and BIOLINCC managed by the National Institutes of Health. Members of the scientific community can apply to access participant data and materials at HCHS/SOL via an established process, such requests are reviewed by the project’s Steering Committee, described at https://sites.cscn.unc.edu/hchs/ (accessioned February 20, 2024). Reasonable requests for data and specimen access can be sent to the corresponding author, which will be referred to the Steering Committee of HCHS/SOL.

CONSENT STATEMENT: all human subjects provided informed consent.

KEY WORDS: Hispanic, Latino, Cognition, Microbiome, Metabolomics
Enrollment and Data Collection

A

B

STOOL MICROBIAL METAGENOMICS
2471 SOL participants at V2 with GCS and metagenomics

Sequencing: Knight Laboratory
Illumina NovaSeq
955 read/sample average coverage

Quality control:
- removal of host 'contaminant' sequences
- trimming to remove low quality bases

Taxonomic Profiling
3314 microbial species

Functional Profiling
445 KEGG Modules

Serum UNTARGETED METABOLICOMICS
480 participants with metagenomic, Metabolic, and GCS data

Untargeted Metabolomic Profiling:
Metabolon HD/Discovery platform

Serum Metabolites
677 metabolites

Figure 1
Click here to access/download: Figure1.tif
A

**TAXA:**
- Central American
- South American
- Cigarettes
- Hypertension
- Dominican
- Chronic Stress
- Alcohol Use
- Global Cognition
- Education
- CESD
- Diabetes
- Metformin
- Mexican
- Waist Hip Ratio
- Fiber
- AHEI
- Cuban
- Acculturation Score
- Age Moved to US
- Female
- Puerto Rican
- Hispanic
- Metformin

**Univariate R-squared (%)**

B

**TAXA:**

**R Squared:** 0.0007 (p = 0.22)

**Global Cognitive Function:**
- high: >=1SD above mean N = 413
- medium: within 1SD of mean N = 1629

C

**TAXA**
- worst
- GCS
- best

**Relative Abundance**

**KEGG MODULES**
- worst
- GCS
- best

**Relative Abundance**
Table 1. Characteristics of 2471 HCHS/SOL-INCA participants with cognitive and microbiome measures at HCHS/SOL-INCA V2.

<table>
<thead>
<tr>
<th></th>
<th>All (N=2471)</th>
<th>Women (N=1625)</th>
<th>Men (N=846)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (mean, SD)</td>
<td>60.83 (7.11)</td>
<td>60.77 (7.06)</td>
<td>60.94 (7.19)</td>
</tr>
<tr>
<td>Hispanic/Latino Heritage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominican</td>
<td>10.28 (254)</td>
<td>11.51 (187)</td>
<td>7.92 (67)</td>
</tr>
<tr>
<td>Central American</td>
<td>9.88 (244)</td>
<td>9.98 (162)</td>
<td>9.69 (82)</td>
</tr>
<tr>
<td>Cuban</td>
<td>13.85 (342)</td>
<td>12.25 (199)</td>
<td>16.90 (143)</td>
</tr>
<tr>
<td>Mexican</td>
<td>38.46 (950)</td>
<td>40.02 (650)</td>
<td>35.46 (300)</td>
</tr>
<tr>
<td>Puerto Rican</td>
<td>18.30 (452)</td>
<td>17.36 (282)</td>
<td>20.09 (170)</td>
</tr>
<tr>
<td>South American</td>
<td>7.04 (174)</td>
<td>68.97 (112)</td>
<td>7.33 (62)</td>
</tr>
<tr>
<td>2+ Heritage</td>
<td>2.02 (50)</td>
<td>1.72 (28)</td>
<td>2.60 (22)</td>
</tr>
<tr>
<td>Diabetes Status</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without Diabetes</td>
<td>19%</td>
<td>20%</td>
<td>18%</td>
</tr>
<tr>
<td>With Pre-Diabetes</td>
<td>47%</td>
<td>47%</td>
<td>45%</td>
</tr>
<tr>
<td>With Diabetes</td>
<td>34%</td>
<td>33%</td>
<td>37%</td>
</tr>
<tr>
<td>MCI at V2</td>
<td>9%</td>
<td>10%</td>
<td>8%</td>
</tr>
<tr>
<td>Education:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;8 grade</td>
<td>33%</td>
<td>39%</td>
<td>34%</td>
</tr>
<tr>
<td>8 grade - HS</td>
<td>20%</td>
<td>19%</td>
<td>20%</td>
</tr>
<tr>
<td>&gt;HS</td>
<td>38%</td>
<td>36%</td>
<td>41%</td>
</tr>
<tr>
<td>Born in the US</td>
<td>10%</td>
<td>9%</td>
<td>12%</td>
</tr>
<tr>
<td>Depression (CESD score)</td>
<td>7%</td>
<td>7%</td>
<td>6%</td>
</tr>
<tr>
<td>Hypertension</td>
<td>54.05% (1335)</td>
<td>53.45% (868)</td>
<td>55.20% (467)</td>
</tr>
<tr>
<td>Waist/Hip Ratio</td>
<td>0.94 (0.085)</td>
<td>0.92 (0.07)</td>
<td>0.99 (0.069)</td>
</tr>
<tr>
<td>Diet/AHEI Score</td>
<td>51.1</td>
<td>50.6</td>
<td>52.3</td>
</tr>
<tr>
<td>Metformin Use</td>
<td>12%</td>
<td>13%</td>
<td>12%</td>
</tr>
<tr>
<td>Fiber Intake (g/day)</td>
<td>19</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>CESD Depression</td>
<td>6.70 (6.09)</td>
<td>7.32 (6.30)</td>
<td>5.53 (5.49)</td>
</tr>
</tbody>
</table>
**Highlights**

- Largest metagenomic study in a Hispanic/Latino cohort to date.
- *Eubacterium* species (*E. siraeum* and *E. eligens*), were associated with better cognition.
- Several KEGG modules, most strongly Ornithine, Serine biosynthesis and Urea Cycle, were associated with worse cognition.
Research in Context:

1. **Systematic review:** The authors reviewed the literature using PubMed sources. According to the literature, US Hispanics/Latinos are at an increased risk of AD/ADRD and accelerated cognitive decline compared to non-Hispanic/Latino whites and have a distinct microbial profile. While the role of the microbiome in cognitive decline and AD/ADRD has not been fully studied, and no studies have been conducted in Hispanics/Latinos to date. The relevant citations are appropriately cited in the manuscript.

2. **Interpretation:** Our study identified several species and KEGG pathways associated with cognitive function, in the largest study on Hispanics/Latinos to date. We confirmed prior reports of a positive association between anti-inflammatory taxa, and negative association with pathogenic taxa and cognitive function. We identified several KEGG modules associated with global cognitive function. Many of the cognition-associated species and functions correlated with serum tryptophan levels.

3. **Future directions:** Future, prospective work is needed to identify microbiome alterations associated with cognitive decline and dementia. More work is needed in underserve populations, such as Hispanics/Latinos.