

## MIT Open Access Articles

*Progressive Shifts in the Gut Microbiome Reflect Prediabetes and Diabetes Development in a Treatment-Naive Mexican Cohort*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**As Published:** 10.3389/FENDO.2020.602326

**Publisher:** Frontiers Media SA

**Persistent URL:** <https://hdl.handle.net/1721.1/133508>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of use:** Creative Commons Attribution 4.0 International license





# Progressive Shifts in the Gut Microbiome Reflect Prediabetes and Diabetes Development in a Treatment-Naive Mexican Cohort

Christian Diener<sup>1,2</sup>, María de Lourdes Reyes-Escogido<sup>3</sup>, Lilia M. Jimenez-Ceja<sup>3</sup>, Mariana Matus<sup>4</sup>, Claudia M. Gomez-Navarro<sup>3</sup>, Nathaniel D. Chu<sup>4</sup>, Vivian Zhong<sup>4</sup>, M. Elizabeth Tejero<sup>1</sup>, Eric Alm<sup>4</sup>, Osbaldo Resendis-Antonio<sup>1,5\*</sup> and Rodolfo Guardado-Mendoza<sup>3,6\*</sup>

## OPEN ACCESS

### Edited by:

Isabel Moreno-Indias,  
University of Málaga, Spain

### Reviewed by:

Tarini Shankar Ghosh,  
University College Cork, Ireland  
Ibiye Owei,  
Texas Tech University Health Sciences  
Center El Paso, United States

### \*Correspondence:

Osbaldo Resendis-Antonio  
oresendis@inmegen.gob.mx  
Rodolfo Guardado-Mendoza  
rguardado@ugto.mx

### Specialty section:

This article was submitted to  
Translational Endocrinology,  
a section of the journal  
Frontiers in Endocrinology

**Received:** 03 September 2020

**Accepted:** 20 November 2020

**Published:** 08 January 2021

### Citation:

Diener C, Reyes-Escogido MdL, Jimenez-Ceja LM, Matus M, Gomez-Navarro CM, Chu ND, Zhong V, Tejero ME, Alm E, Resendis-Antonio O and Guardado-Mendoza R (2021) Progressive Shifts in the Gut Microbiome Reflect Prediabetes and Diabetes Development in a Treatment-Naive Mexican Cohort. *Front. Endocrinol.* 11:602326. doi: 10.3389/fendo.2020.602326

<sup>1</sup> Computational Genomics, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City, Mexico, <sup>2</sup> Gibbons Lab, Institute for Systems Biology, Seattle, WA, United States, <sup>3</sup> Metabolic Research Laboratory, Department of Medicine and Nutrition, University of Guanajuato, León, Mexico, <sup>4</sup> Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, MA, United States, <sup>5</sup> Human Systems Biology Laboratory, Coordinación de la Investigación Científica—Red de Apoyo a la Investigación, Universidad Nacional Autónoma de México (UNAM), Mexico City, Mexico, <sup>6</sup> Research Department, Hospital Regional de Alta Especialidad del Bajío, León, Mexico

Type 2 diabetes (T2D) is a global epidemic that affects more than 8% of the world's population and is a leading cause of death in Mexico. Diet and lifestyle are known to contribute to the onset of T2D. However, the role of the gut microbiome in T2D progression remains uncertain. Associations between microbiome composition and diabetes are confounded by medication use, diet, and obesity.

Here we present data on a treatment-naive cohort of 405 Mexican individuals across varying stages of T2D severity. Associations between gut bacteria and more than 200 clinical variables revealed a defined set of bacterial genera that were consistent biomarkers of T2D prevalence and risk. Specifically, gradual increases in blood glucose levels, beta cell dysfunction, and the accumulation of measured T2D risk factors were correlated with the relative abundances of four bacterial genera. In a cohort of 25 individuals, T2D treatment—predominantly metformin—reliably returned the microbiome to the normoglycemic community state. Deep clinical characterization allowed us to broadly control for confounding variables, indicating that these microbiome patterns were independent of common T2D comorbidities, like obesity or cardiovascular disease. Our work provides the first solid evidence for a direct link between the gut microbiome and T2D in a critically high-risk population. In particular, we show that increased T2D risk is reflected in gradual changes in the gut microbiome. Whether or not these T2D-associated changes in the gut contribute to the etiology of T2D or its comorbidities remains to be seen.

**Keywords:** type 2 diabetes, microbiome, Mexico, metformin, deep phenotyping

## INTRODUCTION

Type 2 diabetes (T2D) is an acquired multifactorial disease that affects more than 8% of the worldwide population and leads to insulin resistance and insufficient insulin production by pancreatic islet cells (1–3). Disease onset is driven or modulated by a variety of factors such as lifestyle, diet, and genetics (4–7). T2D incidence is progressively increasing in the Mexican population and has become a major burden for the national health system and one of the leading causes of death in Mexico (8–10). The particular vulnerability of the Mexican population to this disease is driven by general factors such as a sedentary lifestyle and diet but is also influenced by genetic risk factors that are enriched in the Mexican population (11). For instance, it has been shown that about half of all native Mexicans carry an SLC16A11 variant that increases T2D risk by 20% for each haplotype (12–14). Consequently, there is an urgent need for diagnosis and treatment strategies to limit the progression of T2D in the Mexican high-risk population.

Recently, the gut microbiome has been proposed as an important modulator in the progression of T2D. Several studies have reported a wide array of associations between the gut microbiome and diabetes in European, American, and Chinese cohorts (15, 16). Most of those have suggested that the diabetic microbiome is less efficient in producing short-chain fatty acids (SCFA) due to a loss of butyrate-producing genera (17–19). However, especially when looking across different populations, the bacterial genera associated with diabetes vary (17), which is consistent with findings that the gut microbiome composition varies greatly across populations (20). For example, an increase of Proteobacteria in T2D was reported for Chinese cohorts but was absent in a European cohort (15, 16). Finding robust associations between the microbiome and T2D is further confounded by treatment effects and comorbidities. Metformin, one of the most common medications for T2D, has been shown to modify the gut microbiome which may contribute to its mechanism of action (21, 22). Indeed, studies comparing diabetic treatment-naïve individuals with diabetic metformin-treated individuals showed that most of the associations initially attributed to disease progression were a consequence of the treatment and absent in individuals without a metformin treatment history (23). Apart from medication, changes in lifestyle or diet may also drive changes in the gut microbiome in a disease-independent manner (24, 25). Thus, two major treatment regimens for T2D, metformin treatment, and lifestyle intervention, will likely both trigger their own changes in the gut microbiome and need to be accounted for. Even when isolating the disease from treatment effects, associations may be confounded by comorbidities. The development of T2D is often linked with obesity, a major risk factor in the development of the disease (20, 26). Additionally, T2D increases the risk for cardiovascular disease, which itself has been linked to changes in the gut microbiome (27, 28). Controlling for all of these factors (disease treatment, lifestyle and diet, and comorbidities) might clarify the true associations between the gut microbiome and T2D disease progression. This requires deep phenotyping of the study participants where one measures not only clinically

variables related to the disease of interest but also from other groups such as obesity, cardiovascular health, lifestyle, and diet. Even though this strategy has been shown to be successful in healthy individuals (29), very few studies have done so in the context of T2D.

To address these concerns and explore the relationship between the microbiome and T2D in an understudied population, we present a controlled study in a Mexican cohort from a distinct geographical region that was specifically designed to avoid those shortcomings. Except for a small control group, all participants in the study were treatment-naïve and had never received prior prediabetes or diabetes diagnosis. We also combined a large array of clinical variables related to diabetes with additional phenotype measurements characterizing the lifestyle, diet, obesity prevalence, and cardiovascular health of each individual. This strategy provided a set of more than 200 clinical variables for each individual, allowing us to control for lifestyle and comorbidities and tease out associations specific to different stages of T2D progression. As a result, we identified a set of four bacterial genera that were associated consistently with T2D development. Our work establishes a set of gut microbiome markers for type 2 disease progression in a Mexican population independent of treatment effects or secondary phenotypes.

## MATERIAL AND METHODS

### Study Population

A cross-sectional analysis was performed in patients from Guanajuato, México, from January 2015 to December 2016, as part of the University Cohort Project CARE-In-DEEP Study (Cardiometabolic Risk Evaluation and Interdisciplinary Diabetes Education and Early Prevention). For this particular study, 470 participants who had an anthropometric, nutritional, biochemical, and metabolic evaluation, as well as a stool sample collection, were included; at the end, we had complete data and microbiome composition only for 427. Based on the oral glucose tolerance test, individuals were stratified into normal glucose metabolism (NG, fasting glucose less than 100 mg/dl and 2 h post-OGTT glucose less than 140 mg/dl), isolated impaired fasting glucose (*iIFG*, fasting glucose 100–125 mg/dl and 2 h post-OGTT glucose less than 140 mg/dl), isolated impaired glucose tolerance (*iIGT*, fasting glucose less than 100 mg/dl and 2 h post-OGTT glucose between 140–199 mg/dl), impaired fasting glucose plus impaired glucose tolerance (*IFG + IGT*, fasting glucose between 100–125 mg/dl and 2 h post-OGTT glucose between 140–199 mg/dl), and T2D (T2D, fasting glucose more than 125 mg/dl and/or 2 h post-OGTT glucose higher than 199 mg/dl). A survey was applied to collect general information about the use of medications, family history, risk factors, and previous diseases. The University Research Council evaluated and approved the study protocol. All participants signed informed consent.

### Anthropometric Measurements

Weight was measured while participants were barefoot and wearing minimal clothing with a Tanita Scale SC-240 (Tanita

Corporation of America, USA). Height was obtained while the participants were standing barefoot with their shoulders in a normal position with a Tanita stadiometer (Tanita Corporation of America, USA). BMI (kg/m<sup>2</sup>) was obtained from standardized measurements of weight and height and was computed as the ratio of weight (kg) over height squared (m<sup>2</sup>), defining normal weight when BMI was between 18.5–24.9 kg/m<sup>2</sup>, overweight when BMI was between 25–29.9 kg/m<sup>2</sup>, and obesity when BMI was  $\geq 30$  kg/m<sup>2</sup>. Waist circumference was measured at the high point of the iliac crest at the end of normal expiration to the nearest 0.1 cm. Body composition was assessed with electrical bioimpedance through a Tanita Scale SC-240. All measurements were performed by personnel trained to use standardized procedures and reproducibility was evaluated, resulting in concordance coefficients between 0.88 and 0.94.

## Nutritional and Physical Activity Evaluation

A validated semi-quantitative food frequency questionnaire (FFQ) was applied to evaluate dietary intake (30). This questionnaire included data regarding the consumption of 116 food items. For each food, a commonly used portion size (e.g. one slice of bread or one cup of coffee) was specified on the FFQ and participants reported their frequency of consumption of each specific food over the previous year. Energy (kcal/day), carbohydrates (g/day), proteins (g/day), fatty acids (g/day), sucrose (g/day), and fructose (g/day) intake during the last year were obtained from this FFQ. The PA level of participants was assessed using a self-administered questionnaire that was verified when the patient assisted for the metabolic evaluation. The questionnaire has a validated Spanish translation (31), which has been adapted for use in the Mexican population. The questionnaire is self-administered and estimates the minutes devoted to the practice of different recreational physical activities during a typical week in the last year (including walking, running, cycling, aerobics, dancing, and swimming as well as playing football, volleyball, basketball, tennis, fronton, baseball, softball, and squash, among other activities). Each item includes time intervals that allow participants to detail the exact number of minutes or hours they dedicate to each form of recreational PA, as well as the intensity of each PA (light, moderate, vigorous). The total duration of each recreational PA was expressed in minutes per day. We calculated the number of hours per week devoted to each activity, which were then multiplied by the intensity of each activity, defined as multiples of the metabolic equivalent (MET) of sitting quietly. We used the Compendium of Physical Activities to assign METs to each activity (32).

## Metabolic Evaluation and Oral Glucose Tolerance Test (OGTT)

All subjects were admitted to the Metabolic Research Laboratory of the Department of Medicine and Nutrition, Division of Health Sciences at the University of Guanajuato the day of the study between 7 and 8 AM, and a catheter was placed into an antecubital vein for all blood withdrawal. Subjects will not be allowed to eat or drink anything after 10 PM on the night before

until the study is completed. After the intravenous catheter was placed and the first blood sample was drawn, the patients ingested 75 grams of glucose. Plasma samples for glucose measurement were drawn at -15, and 0 min and every 30 min thereafter for 2 h, glucose was measured by colorimetric glucose oxidase. Lipid levels were measured by dry chemistry with a colorimetric method (Vitros 5600; Ortho Clinical Diagnostics). According to the glucose levels at fasting and at 2 h during the OGTT, patients were classified as following: *NG* = fasting glucose <100 mg/dl and 2 h glucose <140 mg/dl, *IFG* = fasting glucose between 100–125 mg/dl and a 2 h glucose <140 mg/dl, *IGT* = fasting glucose <100 mg/dl and 2 h glucose between 140–199 mg/dl, *IFG+ITG* = fasting glucose between 100–125 mg/dl and 2 h glucose between 140–199 mg/dl, *T2D* = fasting glucose >125 mg/dl and/or 2 h glucose >200 mg/dl, and *treated T2D* = previous diagnose of T2D confirmed by the medical record of the patients, consumption of hypoglycemic drugs and fasting glucose >125 mg/dl and/or 2 h glucose >200 mg/dl. HbA1c was measured according to the international guidelines by HPLC in a subset of 182 patients.

Insulin during the OGTT was measured by a solid-phase, enzyme-labeled chemiluminescent immunometric assay (IMMULITE 1000 Siemens Healthcare Diagnostics Products Ltd). The area under the glucose and insulin curve was calculated by the trapezoidal rule.

Insulin resistance was calculated by the homeostasis model assessment (HOMA<sub>IR</sub>) and insulin sensitivity (Matsuda Index) was derived from the insulin and glucose measurements from the OGTT as previously described (33). Insulin secretion was calculated dividing AUC<sub>insulin</sub>\_OGTT by the AUC<sub>glucose</sub>\_OGTT, acute insulin response (AIR) was calculated dividing the insulin change from 0 to 30 min by the glucose change from 0 to 30 min during the OGTT; pancreatic beta cell function was estimated by the disposition index derived from the OGTT (34).

## Faecal Sample Collection

Fecal samples were collected from volunteers in a sterile container, each sample was homogenized and three aliquotes placed in sterile 1 ml screw-cap tubes which were stored at -80°C before DNA extraction.

## DNA Extraction

DNA extraction was performed using MoBio PowerSoil DNA Isolation kit (Mo Bio Laboratories, Inc. Carlsbad, USA) according to the manufacturer's instructions with the following modifications. After adding the C1 solution and mix, 25 µl of proteinase K solution was added and mixed by vortexing. Samples were incubated at 65°C for 10 min, during the incubation tubes were mixed by inversion every 3 min. Tubes were secured horizontally in a vortex adapter tube holder, and vortexed at 3,000 rpm for 15 min. Samples were incubated at 95°C for 10 min, during this time samples were mixed as mentioned above. Total DNA was eluted in 100 µl of sterile water. DNA concentration was quantified spectrophotometrically with a Qubit (Thermo Scientific, USA) and validated by Nanodrop (ND 2000, Thermo Scientific, USA).

## 16S rRNA Gene Amplification and Sequencing

DNA templates were used in a two-step PCR method to sequence the V4 hypervariable region of the bacterial 16S rRNA gene. Fusion primers contained a sequence complementary to the v4 region, as well as Nextera Illumina adapter sequences to allow multiplexing of pooled libraries.

In the initial PCR, we employed primers that were comprised of partial Nextera adapter and the V4 targeting forward or reverse primer sequence in agreement with (35).

NEXT\_16S\_V4\_U515\_F

5'-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGG  
TGCCAGCMGCCGCGGTAA-3'

NEXT\_16S\_V4\_E786\_R

5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGG  
ACTACHVGGGTWTCTAAT-3'

For each sample, we used approximately equal amounts of DNA template (up to 12.5 ng per reaction) and the reactions were carried out with a 3 min denature step at 94°C, followed by 25 cycles of denaturation at 94°C for 45 s, annealing at 50°C for 60 s, and extension at 72°C for 90 s, with a final extension at 72°C for 10 min. In all reactions were used 2x KAPA HiFi HotStart ReadyMix to generate the amplicons.

The amplicons were purified using Agencourt Ampure XP beads (Beckman Coulter) with a proportion of 1.25x (v/v). The PCR products were checked using electrophoresis in 2% (w/v) agarose gels in TAE buffer (Tris- acetate-EDTA) stained with SYBR Gold and visualized under UV light.

For each amplicon, a second PCR was carried out with a 3 min denature step at 95°C, followed by 8 cycles of denaturation at 95°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 30 s, with a final extension at 72°C for 5 min with 5 ul of previous purified DNA template and using primers that attaches dual indices and Illumina sequencing adapters employing the Nextera XT kit. The PCR products were also purified equal to the first PCR reactions and the DNA concentration of each PCR product was determined using a Qubit® 2.0 Broad Range Assay (Life Technologies™). An Agilent TapeStation (Agilent, Santa Clara, CA) with DNA High Sensitivity kit was used to verify the size of the PCR product only to 23 amplicons.

All samples were random distributed in similar proportions in five pools and then mixed in equal amounts (to 10 nM). The final concentration of each pool was again determined using a Qubit® 2.0.

Pools were diluted to a concentration of 9 pM for sequencing using 2x250 bp paired-end sequencing chemistry v2 on an Illumina MiSeq platform. All samples were distributed according to the consecutive number assigned by the experimental laboratory in similar proportions in five pools and then mixed in equal amounts (to 10 nM). The final concentration of each pool was again determined using a Qubit® 2.0. Amplicons were denatured with 0.2 N NaOH and further diluted according to the MiSeq user guide, then combined with denatured PhiX control library. PhiX was spiked into the amplicon pool at 10% relative concentration.

Image processing and base calling was performed on the BaseSpace cloud from Illumina (<http://basespace.illumina.com>).

## Processing of 16S Sequencing Data

Demultiplexed MiSeq FASTQ files were analyzed using the DADA2 workflow (36). High read quality was ensured by filtering and trimming reads before further processing. In brief, the first 5' 10 bp of all reads were trimmed and reads were truncated on 3' to a maximum length of 240 and 200 bp for forward and reverse reads respectively as a dip in sequence quality was observed after that length. Furthermore, all reads with more than two expected errors under the Illumina base model were removed as well. The filtered and trimmed reads were grouped by sequencing run and the error model was fit for each run separately using the DADA2 default parameters. Sequence variants were obtained for each run separately using the previously calculated error models and the dereplicated input sequences. The sequence variants and counts were then joined across all runs in a complete sequence table and *de novo* chimera removal was run on the entire table.

Taxonomy for the final sequence variants was called using DADA2's RDP classifier and using the SILVA database (version 132) (37). Species were identified separately by exact sequence matches where possible (again using SILVA version 132). The final data set was joined with clinical metadata and saved in a phyloseq object for all downstream analyses (38).

In order to identify additional biases or batch effects. We checked whether particular sequence variant read counts were associated with DNA extraction order, DNA extraction date or the scientist that extracted the sample. We could not identify any bias visually and the distribution of correlations between the extraction date and individual sequence variant abundances was similar to one obtained from a random Poisson model. Finally, we also verified that there were no run batch effects by PCoA plots where we observed no particular separation of samples by sequencing run. a notebook for those quality control steps can be found in the study repository as described in "Data availability".

## Association Tests

Association tests were run using DESeq2 with some custom adjustments (39). First, the input count matrix was filtered by a "rule of 10" where we only tested those taxa with an average count of at least 10 reads and which appeared in at least 10% of all samples. This was necessary to avoid bimodal p-value distributions during multiple testing. The count matrix was normalized across samples using the DESeq2 size factors and the "poscounts" correction for zero read counts. This ensured that associations were not driven by library size and was also expected to counteract the compositionality of the data since that normalization scheme is similar to the centered log-ratio transform (40). All continuous clinical variables were standardized (subtraction of mean and division by standard deviation). All tests used sex as a confounding variable. Age did not show major associations with any clinical variables in this study and including it as a confounder did not have any effect. Consequently, we did not include age as a default confounder in our analysis.

Association tests were then run for all combinations between taxa and clinical variables and only for those individuals with non-

missing measurements. Here, significance was evaluated based on a chi-squared likelihood-ratio test testing for a difference of deviance between the model containing only the confounder variables and a model containing the confounder variables and the tested clinical variable (39). All associations discussed in detail in this manuscript were validated manually to confirm the lack of extreme outliers in the scatter plots. P-values were adjusted for false discovery rate using independent hypothesis weighting to avoid biases for tests with low abundance taxa (41).

## RESULTS

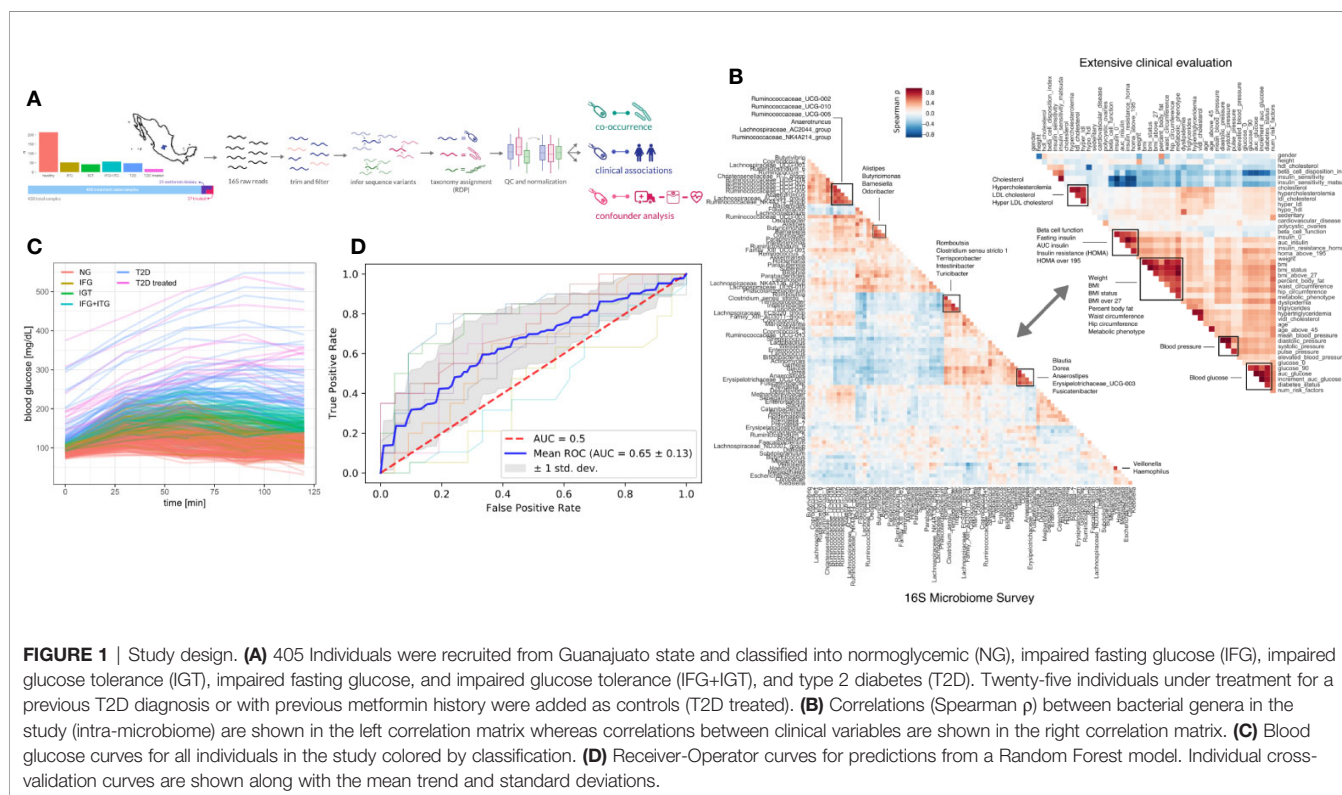
### The Microbiome of Treatment-Naive Individuals Associates With a Wide Range of Clinical Variables

We recruited a cohort of treatment-naive subjects from the Guanajuato region of Mexico as part of the CARE-In-DEEP Study (Cardiometabolic Risk Evaluation and Interdisciplinary Diabetes Education and Early Prevention) of the University of Guanajuato. This cohort consisted of 405 individuals with no previous diabetes diagnosis and a control group of 25 subjects with previously diagnosed T2D or a history of metformin treatment (see **Figure 1A**). Each of the participants in the study underwent extensive clinical characterization consisting of direct measurements as well as a set of validated questionnaires, forming a data set of 226 clinical variables spanning the areas of diabetes, obesity, general health, lifestyle, and diet (**Figure 1B**). Based on an oral-glucose tolerance test,

subjects were stratified into five metabolic groups ranging from normoglycemia and normal glucose tolerance (NG,  $n = 214$ ), to different types of prediabetes (impaired fasting glucose, IFG  $n = 52$ , impaired glucose tolerance IGT  $n = 42$ , and IFG+IGT  $n = 57$ ), and T2D (new T2D  $n = 48$ , and treated previous T2D  $n = 17$ ) (see Methods and **Figure 1C**). As shown in **Table 1**, clinical phenotype varied widely between metabolic groups, with a progressive increase in weight, body fat, glycated hemoglobin (HbA1c), glucose levels, and deteriorating insulin sensitivity and pancreatic beta cell function from the NG group to the T2D group (see **Figures S1A–C**). **Table 2** also shows the frequency of T2D risk factors between the study groups. As shown, only age  $>45$  years, overweight, dyslipidemia, and high blood pressure were significantly different between groups.

To identify links between the microbiome and the progression of T2D, we sequenced the 16S rDNA V4 amplicon from stool samples of the cohort. Sequencing data was analyzed using DADA2 which identified 17,059 exact amplicon sequence variants across all samples (see Methods). These sequence variants mapped to 378 bacterial genera, however only 629 sequence variants and 125 genera were appreciably frequent across samples (found in  $>10\%$  of individuals).

Previous studies have found metformin treatment to lower *Intestinibacter* abundances and to increase *Escherichia* abundances (23). We found similar trends in our data, albeit not significant (Mann-Whitney  $p = 0.05$  and  $0.07$  for *Intestinibacter* and *Escherichia*, see **Figure S1D**). In general, T2D could only be weakly predicted from microbiome composition (Random Forest area under ROC =  $0.69$ , see **Figure 1D**).



**TABLE 1 |** Cohort characteristics.

Variable (n = 430)	NG (n = 214)	IFG (n = 52)	IGT (n = 42)	IFG+IGT (n = 57)	new T2D (n = 48)	treated T2D (n = 17)	p value
Age (y)	38 ± 14	45 ± 14 <sup>a</sup>	45 ± 12	49 ± 12 <sup>a</sup>	51 ± 9 <sup>a</sup>	55 ± 11 <sup>a</sup>	<0.001
Sex (M/F)	49/165	23/29	13/29	18/39	17/31	2/15	0.019
PP (mmHg)	38 ± 9	41 ± 10	42 ± 8	45 ± 11 <sup>a</sup>	45 ± 12 <sup>a</sup>	44 ± 12	<0.001
Weight (kg)	70 ± 15	77 ± 17	76 ± 15	85 ± 17 <sup>a</sup>	85 ± 20 <sup>a</sup>	74 ± 23	<0.001
BMI	26.5 ± 5.4	28.2 ± 5.1	29.3 ± 5.3 <sup>a</sup>	32.4 ± 6.2 <sup>ab</sup>	32.5 ± 7.0 <sup>ab</sup>	31.0 ± 9.0 <sup>a</sup>	<0.001
Body fat%	33 ± 8	32 ± 7	35 ± 8	39 ± 8 <sup>ab</sup>	39 ± 10 <sup>ab</sup>	37 ± 9	<0.001
Visceral fat %	6.6 ± 4.1	9.3 ± 3.9 <sup>a</sup>	9.3 ± 3.9 <sup>a</sup>	11.8 ± 4.3 <sup>a</sup>	12.2 ± 5.6 <sup>a</sup>	10.3 ± 4.2	<0.001
WC (cm)	84.7 ± 13.4	90.3 ± 11.2	92.0 ± 10.3 <sup>a</sup>	98.4 ± 13.3 <sup>a</sup>	97.7 ± 18.5 <sup>a</sup>	95.4 ± 10.5	<0.001
Glucose (mg/dl)	88 ± 7	105 ± 4 <sup>a</sup>	94 ± 4	108 ± 5 <sup>a</sup>	154 ± 60 <sup>abcd</sup>	190 ± 74 <sup>abcde</sup>	<0.001
HbA1c %	5.3 ± 0.3	5.4 ± 0.3	5.4 ± 0.4	5.6 ± 0.4	6.6 ± 1.6 <sup>abcd</sup>	7.5 ± 1.2 <sup>abcd</sup>	<0.001
Total cholesterol (mg/dl)	181 ± 37	189 ± 36	190 ± 32	196 ± 42	188 ± 30	218 ± 30 <sup>a</sup>	<0.001
TG (mg/dl)	131 ± 62	173 ± 93 <sup>a</sup>	171 ± 78 <sup>a</sup>	181 ± 80 <sup>a</sup>	200 ± 85 <sup>a</sup>	202 ± 66 <sup>a</sup>	<0.001

P value column denotes p values of ANOVA with Bonferroni correction. Superscript letters denote the following: (a)  $p < 0.01$  vs NG (b)  $p < 0.01$  vs IFG (c)  $p < 0.01$  vs IGT (d)  $p < 0.01$  vs IFG +IGT (e)  $p < 0.01$  vs treatment-naïve T2D.

**TABLE 2 |** T2D risk factors between the study groups.

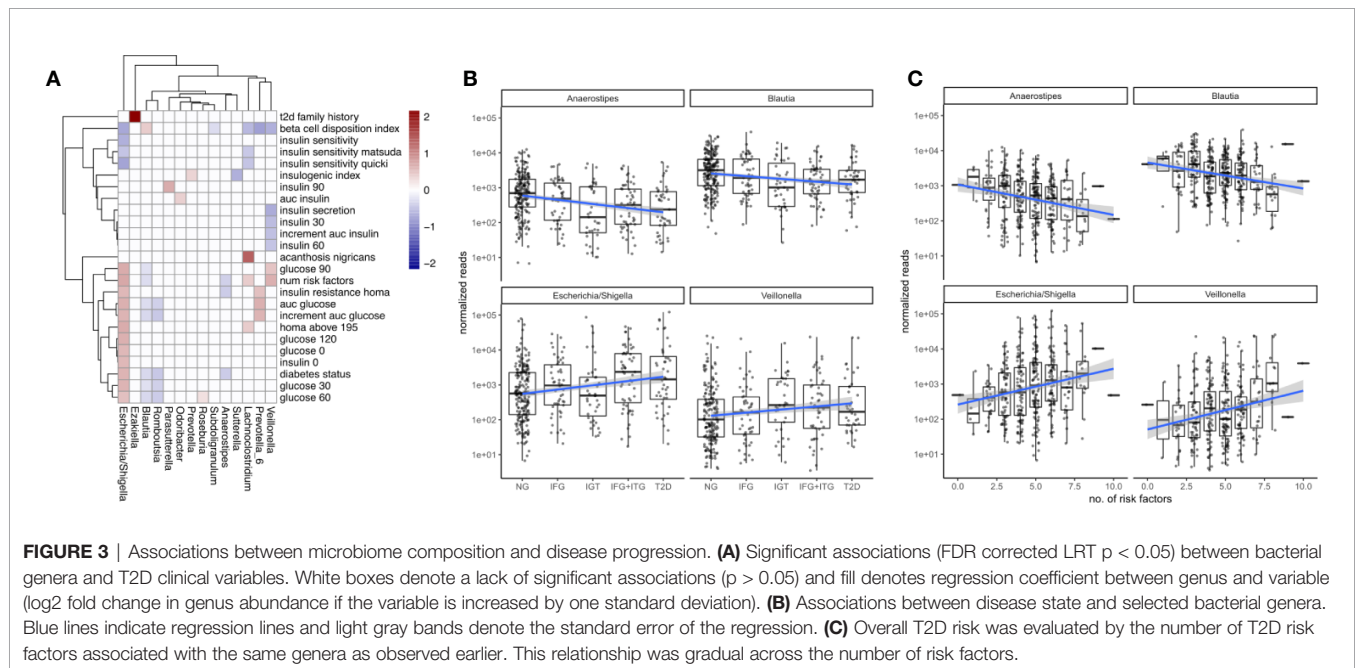
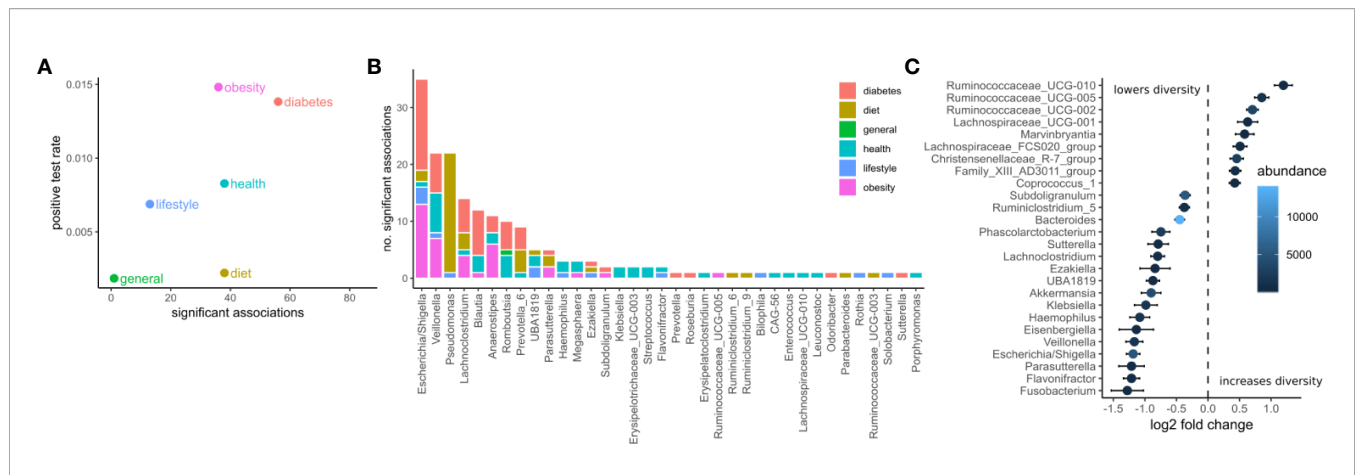
(n = 430)	NG (n = 214)	IFG (n = 52)	IGT (n = 42)	IFG+IGT (n = 57)	new T2D (n = 48)	treated T2D (n = 17)	p value
<b>T2D risk factors (%)</b>							
Physically inactive	61.2	59.6	80.0	71.9	66.0	64.7	0.211
Age >45y	39.2	57.7	47.5	68.4	80.9	76.5	<0.001
T2D Family history	73.7	75.0	77.5	80.7	83.0	100.0	0.160
Overweight	41.8	50.0	64.3	80.7	81.3	64.7	<0.001
Dyslipidemia	69.7	84.6	85.7	89.5	87.2	100.0	<0.001
HBP	25.2	46.2	45.2	54.4	64.6	58.8	<0.001

HBP, high blood pressure.

We identified potential links between the microbiome by exhaustive testing of all combinations between bacterial genera and clinical variables, including alpha diversity (Shannon index). This required careful modeling of the sequencing counts which often do not follow normal distributions. Here we chose negative binomial models as they model the high prevalence of zero read counts and have been shown to represent amplicon sequencing data well (42). Consequently, associations between the microbiome and clinical variables were identified by a robust normalization and testing strategy based on DESeq2 (see Materials and Methods). In summary, read abundances were normalized for library size, regressed against clinical variables with negative binomial models, and significance was judged by a chi-squared likelihood-ratio test on model deviances (LRT) (39). Of the 30,780 tests, 208 were deemed significant under an FDR cutoff of 0.05 (Figures 2A, B). Clinical measurements related to obesity had the most significant associations with microbiome features, while diet-related variables were the least likely to yield a significant association (Figure 2A). The relative paucity of associations between the microbiome and diet may be a consequence of the homogeneous geographical location and the long time-frame covered by the food questionnaire. Additionally, dietary changes usually induce large short-lived shifts in the microbiome which are commonly studied using beta-diversity (43). Those global changes may affect many genera which may conflict with the normalization method used here that required the majority of taxa to be non-differential across individuals.

The genera associated with the most clinical variables was the facultative anaerobe *Escherichia* and the obligate anaerobe *Veillonella*, which had 36 and 23 significant associations respectively (Figure 2B). *Escherichia* was associated mostly with variables related to diabetes and obesity whereas *Veillonella* was associated with variables from many categories. *Ruminococcaceae* genera were the most positively correlated with alpha diversity (Shannon index) whereas *Fusobacterium*, *Flavonifractor*, and *Parasutterella* were the most negatively associated with alpha diversity (Shannon index, Figure 2C).

The gut microbiome of the treatment-naïve cohort was associated widely with T2D-related clinical variables. A set of 14 bacterial genera associated at least weakly with 25 of the 31 diabetes-related measures (FDR-corrected LRT p-value < 0.05). However, we observed large differences in how those associations distributed across genera (Figure 3A). Whereas some genera associated with a wide array of T2D measures (for instance *Escherichia*) other associated only with a single measure (e. g. *Ezakiella* with T2D family history), or exclusively with glucose-related measures, but not insulin-related measures (e. g. *Romboutsia*, Figure 3A). In general, we observed more associations with glucose metabolism than insulin levels. *Escherichia* showed by far the most associations with T2D measures and notably associated with all glucose measures included in the study. Given the observed genus-specific patterns of association with T2D, this raised the question of how one could identify a subset of genera that were consistent markers of overall disease progression.



## A Group of Distinct Bacteria Mark the Gradual Progression of Type 2 Diabetes and Modulate Persistent Inflammation

To identify bacterial genera that were strong markers for disease progression, we asked whether some of the 18 genera associating with diabetes measures would do so in a gradual manner across disease progression and risk. Disease progression was quantified by ordering the metabolic groups by severity ranging from normoglycemic (NG) to fully developed T2D. Disease risk was assessed by a set of manually chosen binary indicators (absent/present) for known risk factors and counting their occurrences for each individual (see Materials and **Figure S2**). Thus, an individual

with 8 risk factors would be considered at higher general risk for developing T2D than an individual with only 2 risk factors. Metabolic groups and the number of risk factors did only moderately correlate with each other (Spearman  $\rho = 0.45$ ), confirming that they described different aspects of the disease. Treating the metabolic groups as well as the number of risk factors as continuous descriptors we identified a set of 4 bacterial genera that associated at least weakly with both of them (*Escherichia*, *Veillonella*, *Blautia* and *Anaerostipes*, FDR-corrected LRT  $p < 0.1$ ). It should be noted that those gradual changes do not reflect longitudinal changes within individuals but rather continuous associations with severity across the entire population.

We found that *Escherichia* and *Veillonella* were positively associated with the diabetic state, increasing in abundance with disease progression from normal to T2D (Figure 3B). Conversely, *Blautia* and *Anaerostipes* abundances declined with disease progression (Figure 3B). Whereas *Escherichia* and *Veillonella* are both associated negatively with alpha diversity (Shannon index), *Anaerostipes* and *Blautia* did not (compare Figure 2C). Therefore, the protective association between these genera and T2D cannot be explained by an increased diversity alone. Intriguingly, more than 99% of the *Anaerostipes* sequence variants with unique species assignments belonged to the species *Anaerostipes hadrus*, a known butyrate producer. The four identified genera showed a continuously increasing or decreasing trend with disease progression, with only the prediabetes group (IGT) showing some deviation from this trend (Figure 3B).

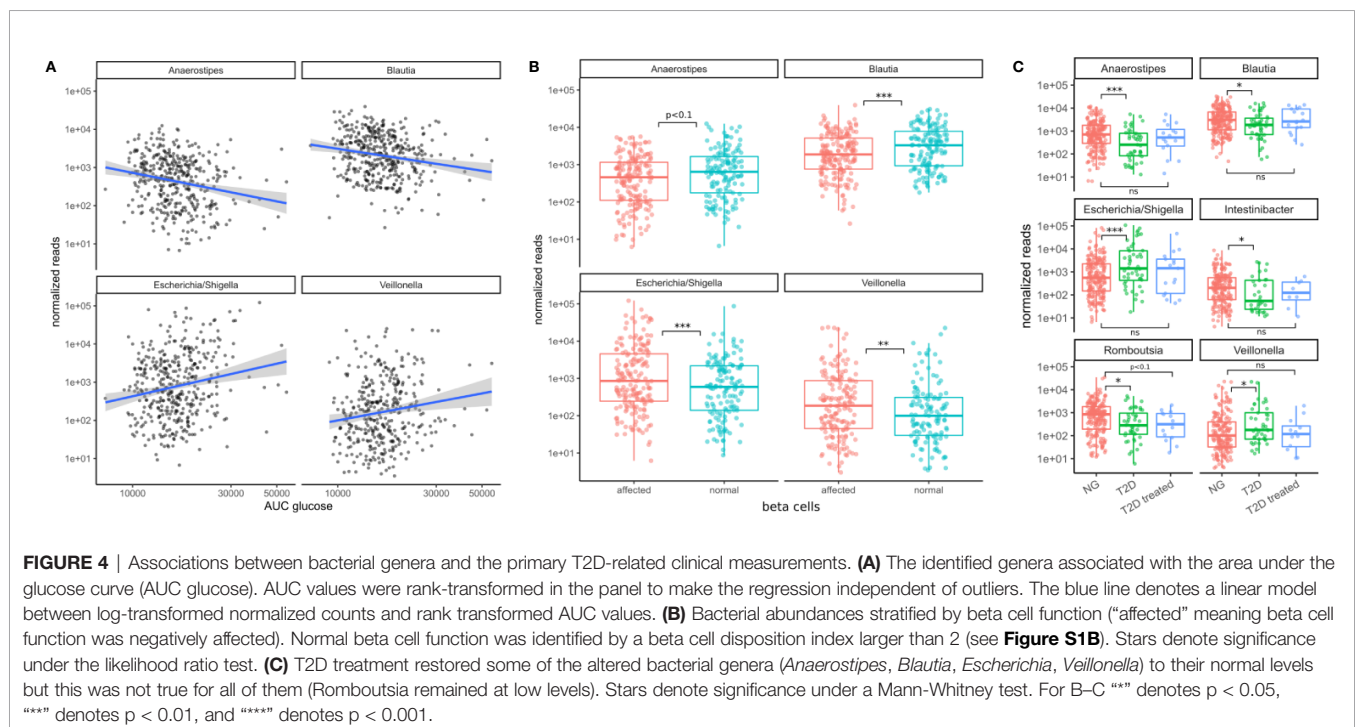
For all of the identified genera, the number of risk factors aligned linearly with the log-transformed counts. Median *Escherichia* levels increased by almost 2 orders of magnitude between individuals with 2 and 8 risk factors respectively and *Anaerostipes* decreased by one order of magnitude (Figure 3C). Notably, individual binary risk factors did show only very few associations with the identified genera (Figure S2). Thus, the accumulation of T2D risk factors across the entire cohort, including healthy individuals, is gradually linked to changes in the microbiome.

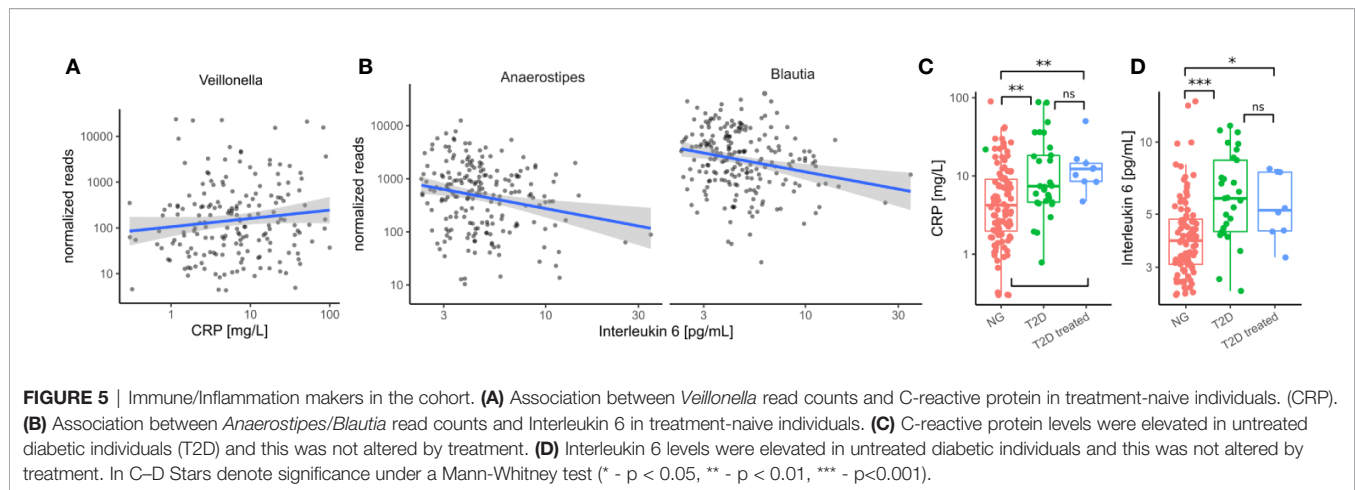
All of the 4 presented genera were also associated with the primary clinical indicators for T2D. Higher levels of *Escherichia* and *Veillonella* accompanied higher area under the glucose curve and diminished beta cell function (FDR adjusted LRT  $p < 0.05$ , see Figure 4). However, *Escherichia* was the only genus that was significantly associated with glycated hemoglobin (log2 fold change 0.5, FDR-adjusted LRT  $p = 0.04$ ), and insulin

sensitivity (Matsuda index, FDR adjusted LRT  $p = 7e-5$ ). Higher levels of *Blautia* and *Anaerostipes* on the other hand were associated with lower areas under the glucose curve and normal beta cell function (FDR adjusted LRT  $p < 0.05$ , see Figure 4). Thus, the associations with markers of metabolic health were consistent with the results of oral glucose tolerance tests.

We then asked whether the patterns of these four microbiome markers of the disease might be reversed by treatment. In a control group of subjects that had already received T2D treatment, we noted that type 2 diabetes treatment (mostly metformin alone or in combination with other drugs) led to an approximate return of the 4 genera to normal levels (Mann-Whitney test  $p$  values between 0.4–0.9, see Figure 4C). This behavior was not observed for all genera. For instance, *Romboutsia* levels were not affected as strongly by diabetes treatment (Mann-Whitney test  $p = 0.07$ , Figure 4C). Thus, anti-hyperglycemic treatment for glucose control was sufficient to return the identified genera close to normal levels and this was not the case for other bacterial genera.

Because invasion with Proteobacteria such as *E. coli* is often a sign of intestinal inflammation we also investigated the association with the identified taxa with inflammation markers. We found that *Veillonella* increased with higher concentrations of C-reactive protein (CRP) whereas *Blautia* and *Anaerostipes* decreased with higher concentrations of interleukin 6 (IL-6) in treatment-naïve individuals (all FDR-corrected LRT  $p < 0.1$ , Figures 5A, B). Strikingly, both of the identified inflammation markers were increased in treatment-naïve diabetic individuals compared to healthy individuals and remained elevated in treated individuals (Figures 5C, D). Thus, in contrast to microbial shifts, the increase in inflammation markers is not ameliorated by T2D treatment.





## A Confounder Analysis Across Variable Classes Identifies Diabetes-Specific Associations

As mentioned before, T2D shows comorbidity with many other clinical conditions such as obesity and cardiovascular disease. For instance, we observed correlations of the major glucose metabolism measurements such as the area under the glucose curve and insulin sensitivity with obesity-related variables such as BMI, visceral fat, and waist-to-hip ratio (see **Figure 1D**). Thus, there was a possibility that our observed changes across disease progression were driven by other covariates. For instance, the association between a bacterial genus and glucose metabolism may be a consequence of obesity which itself is associated with higher glucose levels. This is commonly known as confounding and obesity would be the confounder in that case.

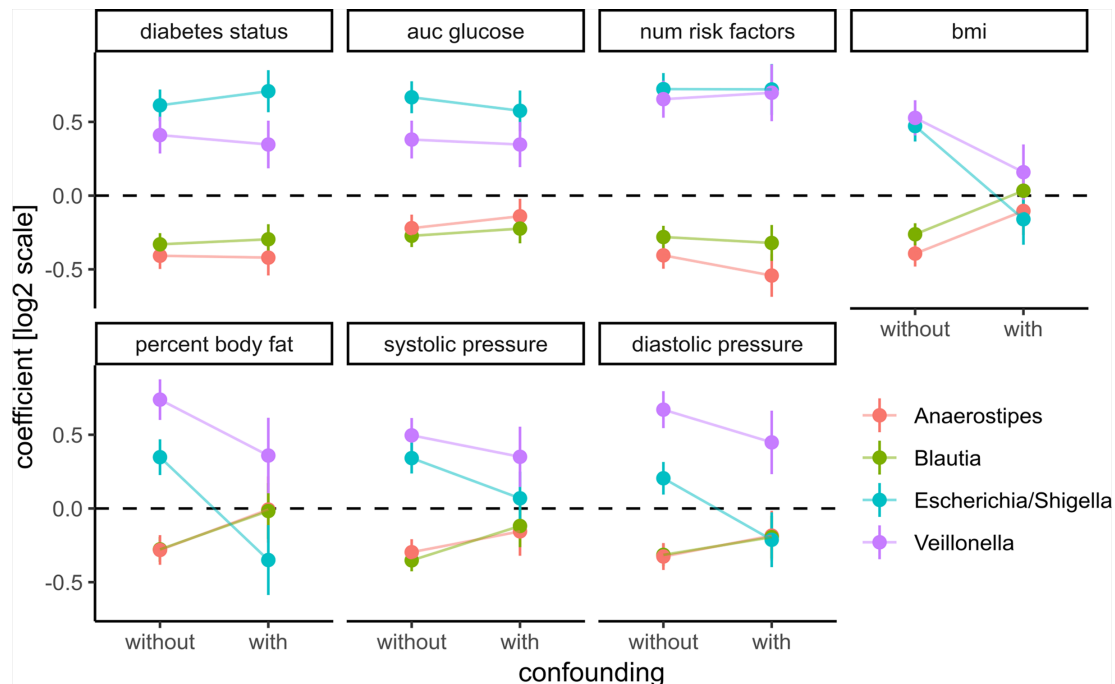
To assess those putative confounding effects, we selected three groups of primary clinical variables that were available for the majority of the samples for T2D, obesity, and cardiovascular health, respectively (see Materials and Methods). Representative clinical variables were chosen by considering only variables measured for the majority of individuals (not all individuals provided information on all measures) and that showed the strongest association with bacterial abundances by themselves. For each of the previously identified bacterial genera and each variable in the three groups, we then ran association tests with either only sex as the confounder (“without”) or with sex and all major variables from the other groups as confounders (“with”). The strength of confounding was evaluated by looking for changes in the regression coefficient for the association between bacterial abundance and the respective clinical variable. If the coefficients were stable across the non-confounded (“without”) and (“with”) group we judged the association robust, whereas a coefficient closer to zero in the confounded setting (“with”) would indicate a diminished association when correcting for additional covariates and, thus, a spurious association.

Coefficients for the diabetes-related clinical variables were not significantly impacted by the introduction of the additional confounders (see **Figure 6**), whereas the coefficients for obesity-related variables were almost completely abolished by

adding the additional confounders. This means that the associations between the four identified bacterial genera and obesity-related clinical variables were essentially lost when correcting for diabetes status. Thus, diabetes measures explained most of the associations between bacterial abundances and obesity but not *vice versa*. Cardiovascular health was also confounded heavily by the T2D-related variables. In particular, we observed that association coefficients between the tested microbial genera and BMI, body fat, or diastolic pressure changed sign when correcting for secondary clinical variables (**Figure 6**). This indicates that non-corrected associations can misinterpret the isolated effect of those clinical variables. Those spurious associations with obesity or cardiovascular disease could be observed with all of the four genera identified in our previous analysis. Here, only *Veillonella* showed residual associations with body fat and blood pressure after correction for some of the clinical variables (body fat and blood pressure) which led us to hypothesize that *Veillonella* seems to associate unspecifically with a variety of “bad health” markers.

## DISCUSSION

One of the challenges in studying the connections between the gut microbiome and T2D is the strong effect of medication on the gut microbiota. Metformin in particular has been shown to induce changes in the microbiome that may themselves alleviate some of the symptoms of T2D either directly or indirectly (22). Consequently, T2D medication with metformin may mask T2D-specific changes in microbial composition. We confirm this in our study and avoided those treatment-specific effects by concentrating on a large treatment-naïve cohort. This allowed us to identify a set of four bacterial genera that are closely connected to T2D disease progression and risk in treatment-naïve individuals of a high-risk population. Notably, all of the four identified genera returned to near-normal levels in treated individuals. Thus, we found that metformin does not only affect more taxa in the



**FIGURE 6** | Adding prominent confounders from other classes of clinical variables did not influence effect size for diabetes-related clinical response variables but did abolish associations in obesity and some cardiovascular responses. Clinical variables are grouped into T2D, obesity, and cardiovascular disease, and association tests between each bacterial genus and variable are either not confounded with additional variables (without confounding) or confounded with all variables from the other groups (with confounding). Points denote the coefficient associated with the response variables under the DESeq2 model (log fold change associated with an increase of one standard deviation in the clinical variable) and error bars denote the standard errors of the model coefficient. Colors denote bacterial genera.

gut microbiome than suggested previously but may also completely disguise microbial changes induced by T2D (**Figure 4A**). It is unclear whether this medication-induced restoration of the gut microbiome is a consequence of alleviated symptoms such as the regulation of blood glucose levels or a direct interaction between drugs and the microbiome. However, our observation that metformin treatment counteracts microbial changes associated with T2D but not with other bacteria seems to suggest that this happens in a disease-dependent manner. Importantly, treatment did not lower the concentrations of the microbiome-associated inflammation markers CRP and IL-6 which indicates that there may be secondary effects of T2D that persist after treatment.

Additionally, the inclusion of a complete characterization of individual phenotypes uncovered the complex pattern of connections between microbial taxa and T2D. Most (25/31) of the diabetes-related covariates included in the study did associate with at least one microbial taxon. However, individual taxa would usually associate with a specific set of clinical measurements. For instance, even though *Escherichia* and *Veillonella* both increased with disease progression, *Escherichia* was preferably associated with measures of blood glucose whereas *Veillonella* was associated with more insulin-related measures (**Figure 3A**). Additionally, we also found that *Blautia* and

*Anaerostipes* did not only decrease with disease progression but also associated with improved beta cell function and insulin efficiency, which is to our knowledge the first time this connection has been described.

We also studied the relationship between the identified bacterial genera and T2D risk based on several established T2D risk factors. Here, we found a clear pattern of microbial shift associated with the accumulation of risk factors. This complements previous studies that have described a connection between the microbiome and the coincidence of T2D diabetes but not on T2D risk itself (23). We observed that this association was stable even in a subpopulation with a low number of risk factors. This is consistent with the pathophysiology of T2D and shows that T2D-specific changes in the microbiome may precede observable symptoms (44, 45). At the present point, one cannot say whether those associations observed across our cohort are indeed present in single individuals during the disease trajectory. Thus, those results rather present co-occurrence than causality. Longitudinal studies may capture properties of T2D progression that are missed by cross-sectional studies (46). Thus, we anticipate that future longitudinal studies will shed more light on the causality between diabetes and the microbiota.

Nevertheless, deep clinical phenotyping allowed us to control for many of the known comorbidities of T2D and confirm the robustness of our findings. For instance, we show that the

strongest associations between the microbiome and obesity-related clinical indicators (BMI and visceral fat) are almost entirely confounded by diabetes covariates and cannot be maintained when controlling for diabetes status. The implications of this observation go beyond this study and demonstrate the potential for extensive confounding in microbiome-obesity studies. As we have shown, this can be avoided by extensive phenotyping of the study subjects and can help to identify effects that are specific to the studied condition and not a secondary effect of another phenotype. In particular, we feel that the combination of correcting for additional phenotypes combined with studying microbial changes that are reversed by treatment is a feasible strategy to constrain the number of associations and identify connections between disease and the microbiome that are good candidates for causal relationships.

On a coarse level, our study is in agreement with previous T2D microbiome studies which mostly report a depletion of butyrate producers. On a fine level, however, we find that the identified genera in our study differ from what has been found in previous studies. For instance, we do not find a depletion of the butyrate-producing *Roseburia*, *Faecalibacterium*, or *Eubacterium* (18) but rather observe a decrease in *Anaerostipes hadrus*, another known butyrate-producer (47). Some studies have also reported an increase of *E. coli* (15, 16), however, we do not observe an increase in *Lactobacillus* or *Streptococcus* as reported there. Consistent with previous findings in treatment-naïve subpopulations, we found that T2D could only be weakly predicted from microbiome composition when correcting for metformin treatment (23). Hyperglycemia itself has been shown to increase the risk for enteric infection by driving intestinal barrier permeability which is consistent with the tight association we observe between *Escherichia* abundance and blood glucose levels (48). Functionally, many of the observed associations point towards gut inflammation. Blooms of proteobacteria, like *E. coli*, have been associated previously with an inflamed gut and are often observed in irritable bowel disease (49, 50). Loss of *Blautia* has also been associated with an inflamed gut in Crohn's disease and other clinical conditions (51, 52). Additionally, alterations in solute carrier expression, as present in the Mexican population (12), have been observed in the development of irritable bowel disease and have been linked to inflammation (53, 54).

Though there is some evidence that gut inflammation may be modulated by the microbiome, it is still unclear whether one could potentially target T2D via altering the gut microbiome (55, 56). We did identify a microbiome-inflammation axis formed by three of the four identified taxa and observed that inflammation markers remained elevated even in treated individuals. *E. coli* was not directly associated with CRP or IL-6 (FDR-adjusted LRT  $p = 0.7$ ) which indicates that there may be additional factors driving the colonization with Proteobacteria. Thus, the observed compositional changes consistent with inflammation might be useful as markers for long-term effects of diabetes-induced phenotypes. For instance, the gut microbiome may help to identify diabetes patients with a high risk for irritable bowel disease or colorectal cancer which have a higher incidence in

T2D patients (57–59). In the end, additional studies will be required to elucidate the causal connections between the gut microbiome and T2D.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. Raw sequencing data along with metadata is provided on the sequence read archive (SRA) under the Bioproject PRJNA541332 at <https://www.ncbi.nlm.nih.gov/>. All additional primary input files as well as intermediate files and R notebooks to reproduce the analysis and figures are provided at <https://github.com/resendislab/mext2d>. More complex functions that could be potentially useful in the analysis of other data sets are furthermore provided along with documentation in a dedicated R package at <https://github.com/resendislab/mbtools>.

## ETHICS STATEMENT

The study protocol was reviewed and approved by the Research Council of the University of Guanajuato. All participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

OR-A, RG-M, and EA designed the study and wrote the manuscript. CD analyzed the data and wrote the manuscript. MR-E, LJ-C, MM, and CG-N extracted the samples and obtained the clinical data. NC and VZ performed the co-occurrence and machine learning analyses and wrote the manuscript. MT analyzed the diet data. All authors contributed to the article and approved the submitted version.

## FUNDING

The authors thank the financial support from CONACYT (Grant Ciencia de Frontera 2019, FORDECYT-PRONACES/425859/2020), an internal grant of the National Institute of Genomic Medicine (INMEGEN, México), a grant from the University of Guanajuato (RG-M, Grants 010/2014, 018/2015, and 1098/2016), and the MIT International Science and Technology Initiatives (MISTI-2015-2016). CG-N received a Ph.D. scholarship from the CONACYT.

## ACKNOWLEDGMENTS

We would like to thank Naeha Subramanian for the valuable discussions. The authors thank the technical support from Alfredo Mendoza and the genome sequencing facility at INMEGEN, México.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fendo.2020.602326/full#supplementary-material>

**SUPPLEMENTARY FIGURE 1 | (A)** Increment of the area under the glucose curve stratified by metabolic group. "nT2D" denotes new (treatment-naïve) T2D and pT2D denotes previous (treated) T2D. Bars denote the standard error of the mean. **(B)** Insulin sensitivity over the metabolic groups. Bars denote the standard error of the mean. **(C)** Beta cell disposition index stratified by metabolic group. The dashed line denotes the cutoff value 2 that was used to separate functional from

non-functional beta cells. **(D)** Abundances of *Escherichia* and *Intestinibacter* stratified by metformin history. Superscripts denote the following comparisons in a t-test: \*p < 0.01 vs. NG group, \*\*p < 0.01 vs. NG and IFG groups, <sup>†</sup>p < 0.01 vs all groups.

**SUPPLEMENTARY FIGURE 2 |** Binary risk factors used to calculate the overall risk of developing T2D. Shown are bacterial abundances stratified by individual risk factors. Absence is denoted by zero and presence by 1. The following 3 risk factors were summarized into a single one (any of the 3 present) when calculating overall risk: polycystic ovary syndrome, cardiovascular disease, having a baby born with more than 4 kg of weight (macrosomia). Reads were normalized across samples as described in Methods.

## REFERENCES

- Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* (2017) 14:88–98. doi: 10.1038/nrendo.2017.151
- Folli F, La Rosa S, Finzi G, Davalli AM, Galli A, Dick EJ Jr, et al. Pancreatic islet of Langerhans' cytoarchitecture and ultrastructure in normal glucose tolerance and in type 2 diabetes mellitus. *Diabetes Obes Metab* (2018) 20 (Suppl 2):137–44. doi: 10.1111/dom.13380
- Guardado-Mendoza R, Davalli AM, Chavez AO, Hubbard GB, Dick EJ, Majluf-Cruz A, et al. Pancreatic islet amyloidosis, -cell apoptosis, and -cell proliferation are determinants of islet remodeling in type-2 diabetic baboons. *Proc Natl Acad Sci* (2009) 106:13992–7. doi: 10.1073/pnas.09064711106
- Diabetes Prevention Program Research Group. Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin. *N Engl J Med* (2002) 346:393–403. doi: 10.1056/NEJMoa012512
- Ley SH, Hamdy O, Mohan V, Hu FB. Prevention and management of type 2 diabetes: dietary components and nutritional strategies. *Lancet* (2014) 383:1999–2007. doi: 10.1016/S0140-6736(14)60613-9
- Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, et al. Replication of Genome-Wide Association Signals in UK Samples Reveals Risk Loci for Type 2 Diabetes. *Science* (2007) 316:1336–41. doi: 10.1126/science.1142364
- Wellcome Trust Case Control Consortium, Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* (2008) 40:638–45. doi: 10.1016/S0084-3741(08)79224-2
- Barquera S, Campos-Nonato I, Aguilar-Salinas C, Lopez-Ridaura R, Arredondo A, Rivera-Dommarco J. Diabetes in Mexico: cost and management of diabetes and its complications and challenges for health policy. *Global Health* (2013) 9:3. doi: 10.1186/1744-8603-9-3
- Alegre-Diaz J, Herrington W, López-Cervantes M, Gnatiuc L, Ramirez R, Hill M, et al. Diabetes and Cause-Specific Mortality in Mexico City. *N Engl J Med* (2016) 375:1961–71. doi: 10.1056/NEJMoa1605368
- Barquera S, Tovar-Guzmán V, Campos-Nonato I, González-Villalpando C, Rivera-Dommarco J. Geography of diabetes mellitus mortality in Mexico: an epidemiologic transition analysis. *Arch Med Res* (2003) 34:407–14. doi: 10.1016/S0188-4409(03)00075-4
- Moreno-Estrada A, Gignoux CR, Fernandez-Lopez JC, Zakharia F, Sikora M, Contreras AV, et al. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* (2014) 344:1280–5. doi: 10.1126/science.1251688
- The SIGMA Type 2 Diabetes Consortium. Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature* (2013) 506:97–101. doi: 10.1038/nature12828
- Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. *Nature* (2016) 536:41–7. doi: 10.1038/nature18642
- Villarreal-Molina MT, Flores-Dorantes MT, Arellano-Campos O, Villalobos-Companan M, Rodriguez-Cruz M, Miliar-Garcia A, et al. Association of the ATP-Binding Cassette Transporter A1 R230C Variant With Early-Onset Type 2 Diabetes in a Mexican Population. *Diabetes* (2008) 57:509–13. doi: 10.2337/db07-0484
- Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* (2013) 498:99–103. doi: 10.1038/nature12198
- Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* (2012) 490:55–60. doi: 10.1038/nature11450
- de Vos WM, Nieuwdorp M. A gut prediction: Genomics. *Nature* (2013) 498:48–9. doi: 10.1038/nature12251
- Hartstra AV, Bouter KEC, Bäckhed F, Nieuwdorp M. Insights Into the Role of the Microbiome in Obesity and Type 2 Diabetes. *Diabetes Care* (2015) 38:159–65. doi: 10.2337/dc14-0769
- Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* (2017) 8:601. doi: 10.1038/s41467-017-01973-8
- Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature* (2012) 486:222–7. doi: 10.1038/nature11053
- Sun L, Xie C, Wang G, Wu Y, Wu Q, Wang X, et al. Gut microbiota and intestinal FXR mediate the clinical benefits of metformin. *Nat Med* (2018) 108:1167. doi: 10.1038/s41591-018-0222-4
- Wu H, Esteve E, Tremaroli V, Khan MT, Caesar R, Mannerås-Holm L, et al. Metformin alters the gut microbiome of individuals with treatment-naïve type 2 diabetes, contributing to the therapeutic effects of the drug. *Nat Med* (2017) 23:850–8. doi: 10.1038/nm.4345
- Forslund K, Meta HIT consortium, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* (2015) 528:262–6. doi: 10.1038/nature15766
- David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* (2014) 505:559–63. doi: 10.1038/nature12820
- Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* (2018) 555:210–5. doi: 10.1038/nature25973
- Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* (2006) 444:840–6. doi: 10.1038/nature05482
- Jie Z, Xia H, Zhong S-L, Feng Q, Li S, Liang S, et al. The gut microbiome in atherosclerotic cardiovascular disease. *Nat Commun* (2017) 8:169. doi: 10.1038/s41467-017-00900-1
- Ryan PM, London LEE, Björndahl TC, Mandal R, Murphy K, Fitzgerald GF, et al. Microbiome and metabolome modifying effects of several cardiovascular disease interventions in apo-E–/– mice. *Microbiome* (2017) 5:576. doi: 10.1186/s40168-017-0246-x
- Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, et al. A wellness study of 108 individuals using personal, dense, dynamic data clouds. *Nat Biotechnol* (2017) 35:747–56. doi: 10.1038/nbt.3870
- Hernández-Avila M, Romieu I, Parra S, Hernández-Avila J, Madrigal H, Willett W. Validity and reproducibility of a food frequency questionnaire to assess dietary intake of women living in Mexico City. *Salud Públ México* (1998) 40:133–40. doi: 10.1590/s0036-36341998000200005
- Martínez-González MA, López-Fontana C, Varo JJ, Sánchez-Villegas A, Martínez JA. Validation of the Spanish version of the physical activity questionnaire used in

- the Nurses' Health Study and the Health Professionals' Follow-up Study. *Public Health Nutr* (2005) 8:920–7. doi: 10.1079/PHN2005745
32. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. *Med Sci Sports Exerc* (2000) 32:S498–504. doi: 10.1097/00005768-200009001-00009
  33. Matsuda M, DeFronzo RA. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care* (1999) 22:1462–70. doi: 10.2337/diacare.22.9.1462
  34. Ahren B, Pacini G. Importance of quantifying insulin secretion in relation to insulin sensitivity to accurately assess beta cell function in clinical studies. *Eur J Endocrinol* (2004) 150:97–104. doi: 10.1530/eje.0.1500097
  35. Caporaso JG, Gregory Caporaso J, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* (2012) 6:1621–4. doi: 10.1038/ismej.2012.8
  36. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* (2016) 13:581–3. doi: 10.1038/nmeth.3869
  37. 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) 41:D590–6. doi: 10.1093/nar/gks1219
  38. McMurdie PJ, Holmes S. Shiny-phyloseq: Web application for interactive microbiome analysis with provenance tracking. *Bioinformatics* (2015) 31:282–3. doi: 10.1093/bioinformatics/btu616
  39. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* (2014) 15:550. doi: 10.1101/002832
  40. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol* (2017) 8:2224. doi: 10.3389/fmicb.2017.02224
  41. Ignatiadis N, Klaus B, Zaugg JB, Huber W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat Methods* (2016) 13:577–80. doi: 10.1038/nmeth.3885
  42. Calgaro M, Romualdi C, Waldron L, Risso D, Vitulo N. Assessment of single cell RNA-seq statistical methods on microbiome data. *Genome Biol* (2020) 21:191. doi: 10.1186/s13059-020-02104-1
  43. Johnson AJ, Vangay P, Al-Ghalith GA, Hillmann BM, Ward TL, Shields-Cutler RR, et al. Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. *Cell Host Microbe* (2019) 25:789–802.e5. doi: 10.1016/j.chom.2019.05.005
  44. DeFronzo RA, Ferrannini E, Groop L, Henry RR, Herman WH, Holst JJ, et al. Type 2 diabetes mellitus. *Nat Rev Dis Primers* (2015) 37:15019. doi: 10.1038/nrdp.2015.19
  45. Gastaldelli A, Gaggini M, DeFronzo RA. Role of Adipose Tissue Insulin Resistance in the Natural History of Type 2 Diabetes: Results From the San Antonio Metabolism Study. *Diabetes* (2017) 66:815–22. doi: 10.2337/db16-1167
  46. Reitmeier S, Kiessling S, Clavel T, List M, Almeida EL, Ghosh TS, et al. Arrhythmic Gut Microbiome Signatures Predict Risk of Type 2 Diabetes. *Cell Host Microbe* (2020) 28:258–72.e6. doi: 10.1016/j.chom.2020.06.004
  47. Kant R, Rasinkangas P, Satokari R, Pietilä TE, Palva A. Genome Sequence of the Butyrate - Producing Anaerobic Bacterium *Anaerostipes hadrus* PEL 85. *Genome Announc* (2015) 3:238. doi: 10.1128/genomeA.00224-15
  48. Thaïs CA, Levy M, Grosheva I, Zheng D, Soffer E, Blacher E, et al. Hyperglycemia drives intestinal barrier dysfunction and risk for enteric infection. *Science* (2018) 359:1376–83. doi: 10.1126/science.aar3318
  49. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* (2012) 13:R79. doi: 10.1186/gb-2012-13-9-r79
  50. Graessler J, Qin Y, Zhong H, Zhang J, Licinio J, Wong M-L, et al. Metagenomic sequencing of the human gut microbiome before and after bariatric surgery in obese patients with type 2 diabetes: correlation with inflammatory and metabolic parameters. *Pharmacogenom J* (2013) 13:514–22. doi: 10.1038/tpj.2012.43
  51. Gevers D, Kugathasan S, Denson LA, Vázquez-Baeza Y, Van Treuren W, Ren B, et al. The Treatment-Naïve Microbiome in New-Onset Crohn's Disease. *Cell Host Microbe* (2014) 15:382–92. doi: 10.1016/j.chom.2014.02.005
  52. Hall AB, Tolonen AC, Xavier RJ. Human genetic variation and the gut microbiome in disease. *Nat Rev Genet* (2017) 18:690–9. doi: 10.1038/nrg.2017.63
  53. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature* (2011) 474:307–17. doi: 10.1038/nature10209
  54. Coburn LA, Singh K, Asim M, Barry DP, Allaman MM, Al-Greene NT, et al. Loss of solute carrier family 7 member 2 exacerbates inflammation-associated colon tumorigenesis. *Oncogene* (2018) 38:1067–79. doi: 10.1038/s41388-018-0492-9
  55. Litvak Y, Byndloss MX, Bäuml AJ. Colonocyte metabolism shapes the gut microbiota. *Science* (2018) 362:eaat9076. doi: 10.1126/science.aat9076
  56. Levy M, Thaïs CA, Zeevi D, Dohnalová L, Zilberman-Schapira G, Mahdi JA, et al. Microbiota-Modulated Metabolites Shape the Intestinal Microenvironment by Regulating NLRP6 Inflammasome Signaling. *Cell* (2015) 163:1428–43. doi: 10.1016/j.cell.2015.10.048
  57. Jurjus A, Eid A, Al Kattar S, Zeenny MN, Gerges-Geagea A, Haydar H, et al. Inflammatory bowel disease, colorectal cancer and type 2 diabetes mellitus: The links. *BBA Clin* (2016) 5:16–24. doi: 10.1016/j.bbacli.2015.11.002
  58. Tilg H, Adolph TE, Gerner RR, Moschen AR. The Intestinal Microbiota in Colorectal Cancer. *Cancer Cell* (2018) 33:954–64. doi: 10.1016/j.ccell.2018.03.004
  59. Amshoff Y, Maskarinec G, Shvetsov YB, Raquinio PH, Grandinetti A, Setiawan VW, et al. Type 2 diabetes and colorectal cancer survival: The multiethnic cohort. *Int J Cancer* (2018) 143:263–8. doi: 10.1002/ijc.31311

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Diener, Reyes-Escogido, Jimenez-Ceja, Matus, Gomez-Navarro, Chu, Zhong, Tejero, Alm, Resendis-Antonio and Guardado-Mendoza. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.