HEPATOLOGY

HEPATOLOGY, VOL. 65, NO. 2, 2017

Gut Microbiota Profiling of Pediatric Nonalcoholic Fatty Liver Disease and Obese Patients Unveiled by an Integrated Meta-omics-Based Approach

Federica Del Chierico,[1*] Valerio Nobili,[2,3*] Pamela Vernocchi,[1] Alessandra Russo,[1] Cristiano De Stefanis,[3] Daniela Gnani,[3] Cesare Furlanello,[4] Alessandro Zandonà,[4] Paola Paci,[5,6] Giorgio Capuani,[7] Bruno Dallapiccola,[8] Alfredo Miccheli,[7] Anna Alisi,[3] and Lorenza Putignani[1,9]

There is evidence that nonalcoholic fatty liver disease (NAFLD) is affected by gut microbiota. Therefore, we investigated its modifications in pediatric NAFLD patients using targeted metagenomics and metabolomics. Stools were collected from 61 consecutive patients diagnosed with nonalcoholic fatty liver (NAFL), nonalcoholic steatohepatitis (NASH), or obesity and 54 healthy controls (CTRLs), matched in a case-control fashion. Operational taxonomic units were pyrosequenced targeting 16S ribosomal RNA and volatile organic compounds determined by solid-phase microextraction gas chromatography-mass spectrometry. The α-diversity was highest in CTRLs, followed by obese, NASH, and NAFL patients; and β-diversity distinguished between patients and CTRLs but not NAFL and NASH. Compared to CTRLs, in NAFLD patients Actinobacteria were significantly increased and Bacteroidetes reduced. There were no significant differences among the NAFL, NASH, and obese groups. Overall NAFLD patients had increased levels of *Bradyrhizobium*, *Anaerococcus*, *Peptoniphilus*, *Propionibacterium acnes*, *Dorea*, and *Ruminococcus* and reduced proportions of *Oscillospira* and Rikenellaceae compared to CTRLs. After reducing metagenomics and metabolomics data dimensionality, multivariate analyses indicated a decrease of *Oscillospira* in NAFL and NASH groups and increases of *Ruminococcus*, *Blautia*, and *Dorea* in NASH patients compared to CTRLs. Of the 292 volatile organic compounds, 26 were up-regulated and 2 down-regulated in NAFLD patients. Multivariate analyses found that combination of *Oscillospira*, Rickenellaceae, *Parabacteroides*, *Bacteroides fragilis*, *Sutterella*, Lachnospiraceae, 4-methyl-2-pentanone, 1-butanol, and 2-butanone could discriminate NAFLD patients from CTRLs. Univariate analyses found significantly lower levels of *Oscillospira* and higher levels of 1-pentanol and 2-butanone in NAFL patients compared to CTRLs. In NASH, lower levels of *Oscillospira* were associated with higher abundance of *Dorea* and *Ruminococcus* and higher levels of 2-butanone and 4-methyl-2-pentanone compared to CTRLs. *Conclusion:* An *Oscillospira* decrease coupled to a 2-butanone up-regulation and increases in *Ruminococcus* and *Dorea* were identified as gut microbiota signatures of NAFL onset and NAFL-NASH progression, respectively. (HEPATOLOGY 2017;65:451-464)

SEE EDITORIAL ON PAGE 401

The term "gut-liver axis" is used to describe the close relationship that is established between the gut and liver beginning in the very early stages of fetal life. The liver receives most of its blood supply from the intestine through the portal vein and is therefore one of the organs that is most exposed to potentially toxic factors originating in the gut, including all or part of the gut microbiota, and bioactive

Abbreviations: Act-GLP-1, active glucagon-like peptide 1; ALT, alanine aminotransferase; AST, aspartate aminotransferase; AUROC, area under the receiver operating characteristic curve; BMI, body mass index; CTRL, control; FDR, false discovery rate; GGT, gamma-glutamyl transpeptidase; HOMA-IR, homeostasis model assessment of insulin resistance; LPS, lipopolysaccharide; LV, latent variable; MB, metabolomics; MG, metagenomics; NAFL, nonalcoholic fatty liver; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; OTU, operational taxonomic unit; PCA, principal component analysis; PH, phenomics; PLS, partial least squares; PLS-DA, PLS discriminant analysis; Tot-GLP-1, total glucagon-like peptide 1; VOC, volatile organic compound; WC, waist circumference.

Received August 16, 2015; accepted March 19, 2016.

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.28572/suppinfo.

**These authors contributed equally to this work.*

Supported by the Italian Ministry of Health (RC 201402G003251, to L.P.), the Model-Driven Paediatric European Digital Repository, European Commission (under FP7-ICT Programme, to B.D.), and the Fondazione Luca Barbareschi.

components of food that have been processed by the gut microbiota.[1] Thus, quantitative and qualitative variations in the bacteria that compose the gut microbiota may actively contribute to the pathogenesis of several liver diseases, including nonalcoholic fatty liver disease (NAFLD), alcoholic steatohepatitis, and cirrhosis.[2-7]

NAFLD is one of the most common causes of chronic liver disease worldwide.[8] It affects children and adults with genetic and lifestyle (e.g., overnutrition or malnutrition and physical inactivity) risk factors. NAFLD ranges from simple steatosis (also called nonalcoholic fatty liver [NAFL]) to a more severe form defined as nonalcoholic steatohepatitis (NASH) characterized by steatosis, ballooning, inflammation, and eventually fibrosis.[9] In addition to liver-related complications, NAFLD is increasingly recognized as a complication of obesity, and its clinical impact on patient outcome has been demonstrated in large cohort studies showing higher than expected prevalence of cirrhosis and hepatocellular carcinoma in both obese and diabetic patients.[1] In fact, in approximately 20% of cases, NASH develops into cirrhosis and related complications, such as portal hypertension, liver failure, and cancer.[10] Today, it is widely accepted that NAFLD pathogenesis is more complex than the "two-hit hypothesis," which proposes that intrahepatic fat accumulation increases susceptibility to liver damage and allows for deteriorating pathological conditions.[11] Several studies have demonstrated that the gut microbiota affects NAFLD pathogenesis.[12-17]

The widespread epidemic of obesity and diabetes and the diversity in presentation at the individual level might be explained by the relationship between the host and the intestinal microbiota. Indeed, more than 4 million gene products from the microbiome can potentially interact with the immune system to induce a tissue metabolic infection, which is the molecular origin of the low-grade inflammation that characterizes the onset of obesity and diabetes.[18] Considering that obesity is a major risk factor for NAFLD in humans, there is evidence that the gut microbiota can directly influence body weight in several ways: (1) it affects the proportion of calories obtained from the intestinal contents; (2) bile acids have bacteriostatic effects, thereby affecting the absorption and emulsification of fats and lipid-soluble vitamins in the small intestine; (3) it contributes to increased intestinal permeability through loss of epithelial barrier integrity; (4) bacterial translocation into the systemic circulation is increased, allowing more hepatic access for ethanol and bacterial endotoxins such as lipopolysaccharide (LPS); (5) microbial LPS is recognized by Toll-like receptor 4, which triggers nuclear factor κB-mediated proinflammatory cytokine production; (6) dietary choline modulation; and (7) development of insulin resistance in the host.[19]

The gut microbiota meets the definition for a tissue organ developed by Burcelin and colleagues based on its complexity and behavior as it is able to manage the entire framework of metabolic activities associated with the network of microbial communities.[20] Therefore, the well-characterized gut-liver axis might correctly be considered the gut-microbiota-liver network due to the high degree of interconnectedness between

View this article online at wileyonlinelibrary.com.
DOI 10.1002/hep.28572

Potential conflict of interest: Nothing to report.

ARTICLE INFORMATION:

From the [1]Human Microbiome Unit, "Bambino Gesù" Children's Hospital, IRCCS, Rome, Italy; [2]Hepato-Metabolic Disease Unit, "Bambino Gesù" Children's Hospital, IRCCS, Rome, Italy; [3]Liver Research Unit, "Bambino Gesù" Children's Hospital, IRCCS, Rome, Italy; [4]Predictive Models for Biomedicine and Environment Unit, Fondazione Bruno Kessler, Trento, Italy; [5]Institute for Systems Analysis and Computer Science "Antonio Ruberti," National Research Council, 00185 Rome, Italy; [6]SysBio Centre for Systems Biology, 00185 Rome, Italy; [7]Department of Chemistry, Sapienza University of Rome, Italy; [8]Scientific Directorate, "Bambino Gesù" Children's Hospital, IRCCS, Rome, Italy; [9]Parasitology Unit, "Bambino Gesù" Children's Hospital, IRCCS, Rome, Italy.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO:

Prof. Valerio Nobili
Head of Hepato-Metabolic Disease Unit, "Bambino Gesù" Children's Hospital
IRCCS P.le S. Onofrio, 4
00165, Rome, Italy
E-mail: valerio.nobili@opbg.net
Tel: +39-06-68592192

TABLE 1. Anthropometric and Laboratory Parameters of the Subjects Belonging to the Case-Control Study Groups

Parameters	CTRL (n = 54)	Obese (n = 8)	NAFL (n = 27)	NASH (n = 26)
Gender	23 M/31 F	3 M/5 F	21 M/6 F	11 M/15 F
Age (years)	10.24 ± 2.51	11.25 ± 2.65	12.04 ± 2.78*	12.27 ± 2.47*
BMI (kg/m^2)	17.59 ± 1.79	26.15 ± 4.38*	26.46 ± 4.43*	27.42 ± 6.45*
WC (cm)	n.d.	82.38 ± 9.78	85.21 ± 8.32	88.15 ± 6.60
ALT (U/L)	n.d.	41.50 ± 47.70	32.30 ± 22.74	44.46 ± 16.73
AST (U/L)	n.d.	33.75 ± 24.03	28.52 ± 11.71	33.15 ± 18.02
GGT (U/L)	n.d.	20.25 ± 14.14	15.93 ± 7.86	21.15 ± 20.52
Triglycerides (mg/dL)	n.d.	82.25 ± 14.86	100.6 ± 76.01†	111.7 ± 49.13†
Cholesterol (mg/dL)	n.d.	183.40 ± 27.67	151.9 ± 30.74†	168.40 ± 28.33
Glucose (mg/dL)	n.d.	89.56 ± 11.35	90.43 ± 9.99	87.40 ± 27.35
Insulin (μU/mL)	n.d.	47.15 ± 29.49	58.99 ± 33.92	61.95 ± 40.28
HOMA-IR	n.d.	3.22 ± 1.39	4.39 ± 2.95	3.81 ± 1.79
Albumin (g/dL)	n.d.	4.73 ± 0.34	4.70 ± 0.23	4.76 ± 0.24
Bilirubin (mg/dL)	n.d.	0.71 ± 0.21	0.54 ± 0.25	0.71 ± 0.44
Tot-GLP-1 (pmol/L)	n.d.	3.39 ± 0.55	3.65 ± 0.41	3.89 ± 0.68
Act-GLP-1 (pmol/L)	n.d.	1.22 ± 0.44	1.39 ± 0.65	1.20 ± 0.42
LPS (EU/mL)	n.d.	16.42 ± 7.21	14.66 ± 8.52	17.64 ± 8.96
Steatosis	Absent	Absent	Present	Present

Data are mean ± standard deviation.
*$P < 0.05$ versus CTRL.
†$P < 0.05$ versus obese.
Abbreviation: n.d., not determined.

the microbiota and host. Recently developed systems biology approaches are able to integrate gut microbiota targeted metagenomics (MG) and metabolomics (MB) data and clinical phenomics (PH) parameters into a multilayered system of "integrated omics" data.[21] In this study, we took a hypothesis-driven approach, rather than a data-driven approach,[22] to study the structural and functional role of the gut microbiota in the onset and progression of pediatric NAFLD by evaluating the fecal gut microbiome of patients with NAFL, NASH, or obesity in terms of the phylome and metabolome compared to healthy controls (CTRLs).

Patients and Methods

HUMAN SUBJECTS

Children and adolescents (n = 61) ranging in age from 7 to 16 years were recruited from the Hepato-Metabolic Disease Unit of the Paediatric Hospital Bambino Gesù (Rome, Italy) during 2013. Patients were stratified based on their diagnosis with NAFL (n = 27), NASH (n = 26), or obesity (n = 8) and matched in a case-control fashion to CTRLs (n = 54) with an overlapping age range enrolled in the Paediatric Hospital Bambino Gesù Human Microbiome Unit.

The Hospital Ethics Research Committee approved the study, which was conducted in accordance with the Declaration of Helsinki (as revised in Seoul, Korea, October 2008). Written informed consent was obtained from parents of the enrolled subjects. The study was approved by the Paediatric Hospital Bambino Gesù Ethics Committee (protocol no. 768.12).

PH DATA

PH data, including body mass index (BMI), waist circumference (WC), alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma-glutamyl transpeptidase (GGT), triglycerides, cholesterol, glucose, insulin, homeostasis model assessment of insulin resistance (HOMA-IR), plasma levels of total and active glucagon-like peptide 1 (Tot-GLP-1 and Act-GLP-1, respectively), and circulating levels of LPS were collected for patients. Gender, age, and BMI were recorded for the CTRLs (Table 1).

TARGETED MG OF GUT MICROBIOTA

A single stool sample was collected and processed from each subject. Genomic DNA was isolated using the QIAamp DNA Stool Mini Kit (Qiagen, Germany). The V1-V3 region of the 16S ribosomal RNA locus was amplified for pyrosequencing using a 454-Junior Genome Sequencer (Roche 454 Life Sciences, Branford, CT). Reads were analyzed by Quantitative Insights into Microbial Ecology (v.1.8.0), grouped

into operational taxonomic units (OTUs) at a sequence similarity level of 97% by PyNAST for taxonomic assignment, and aligned by UCLUST for OTUs matching against Greengenes database (v. 13.8). Significantly different OTUs were selected by false discovery rate (FDR)-adjusted *P* values.

TARGETED MB OF GUT MICROBIOTA

Of the 115 stool samples collected 110 were also processed for MB. Samples were preconditioned according to the manufacturer's instructions, and volatile organic compounds (VOCs) were extracted using carboxen-polydimethylsiloxane-coated fiber (85 μm) and a manual solid-phase microextraction holder (Supelco Inc., Bellefonte, PA). Significantly different molecules were selected by FDR-adjusted *P* values.

STATISTICAL ANALYSIS OF META-OMICS DATA

Multivariate Analysis

The meta-omic data from targeted MG and MB were analyzed using supervised and unsupervised methods including principal component analysis (PCA), the partial least squares (PLS) method, and PLS discriminant analysis (PLS-DA). PCA and PLS-DA were performed using Unscrambler X software, version 10.3 (CAMO Software, Oslo, Norway). The results were validated through full leave-one-out cross-validation procedures and Marten's uncertainty test,[23-25] which couples full cross-validation to the jackknife principle[26] and allows the detection of those original variables that are significantly related to the PLS models. The data matrix was preprocessed before multivariate analysis by mean centering and scaling (i.e., the means of each column were set to 0 and their standard deviations to 1). This common procedure was applied to the data matrix prior to PLS because it allowed us to compare the covariations of the signals independent of their numerical size, while keeping the factorial structure of the data intact.

Univariate Analysis

Data normality was evaluated using the Shapiro-Wilk test. Because the data distribution was not normal, nonparametric tests (Kruskal-Wallis and Mann-Whitney) were applied to compare the data from different groups of patients. The *P* values from the nonparametric statistical analyses were adjusted for multiple testing using FDR.[27]

For further details, see the Supporting Information.

Results

PATIENT CHARACTERISTICS

This study used 61 consecutively enrolled NAFL (n = 27), NASH (n = 26), or obese (n = 8) patients treated in the Hepato-Metabolic Disease Unit of the Paediatric Hospital Bambino Gesù. The PH parameters of the patients compared to the CTRLs (n = 54) are shown in Table 1. The mean age of the NAFL and NASH patients was significantly ($P < 0.05$) higher than that of the CTRLs, and the BMI of the CTRLs was significantly lower than that of the obese, NAFL, and NASH groups ($P < 0.05$). No statistically significant differences were found between groups with regard to the metabolic parameters, with the exception of triglycerides in the NAFL and NASH groups and of cholesterol in NAFL, compared to obese children (Table 1).

A heatmap of the PH data assisted in visualizing the hierarchical clustering of individual properties (Fig. 1). The PH data from all of the patients appeared to fall into three main clusters: (1) WC, BMI, age, Tot-GLP-1, and Act-GLP-1; (2) GGT, ALT, AST, cholesterol, triglycerides, glucose, homeostasis model assessment of insulin resistance, and insulin; and (3) bilirubin, albumin, and LPS (Fig. 1A). Based on their PH parameters, patients were clustered into one of five groups: (1) primarily patients with NAFL or NASH (n = 17) characterized by the highest values for age, Tot-GLP-1, and Act-GLP-1 and the lowest values of WC and BMI, with generally low values for the other PH parameters; (2) primarily NAFL patients (n = 18) characterized by low values for all PH parameters; (3) primarily NAFL patients (n = 8) characterized by high values for WC, BMI, age, Tot-GLP-1, and Act-GLP-1 and low values for the other PH parameters; (4) primarily NAFL patients (n = 10) characterized by high levels of LPS, albumin, and bilirubin and low values for the other PH parameters; and (5) primarily NASH patients (n = 8) characterized by high levels of GGT, ALT, AST, cholesterol, triglycerides, glucose, HOMA-IR, and insulin.

Among the NASH patients, there were two main clusters in the PH data: (1) WC, BMI, age, Tot-GLP-1, Act-GLP-1, GGT, ALT, AST, and cholesterol and (2) triglycerides, bilirubin, albumin, glucose, insulin, and HOMA-IR. The NASH patients were

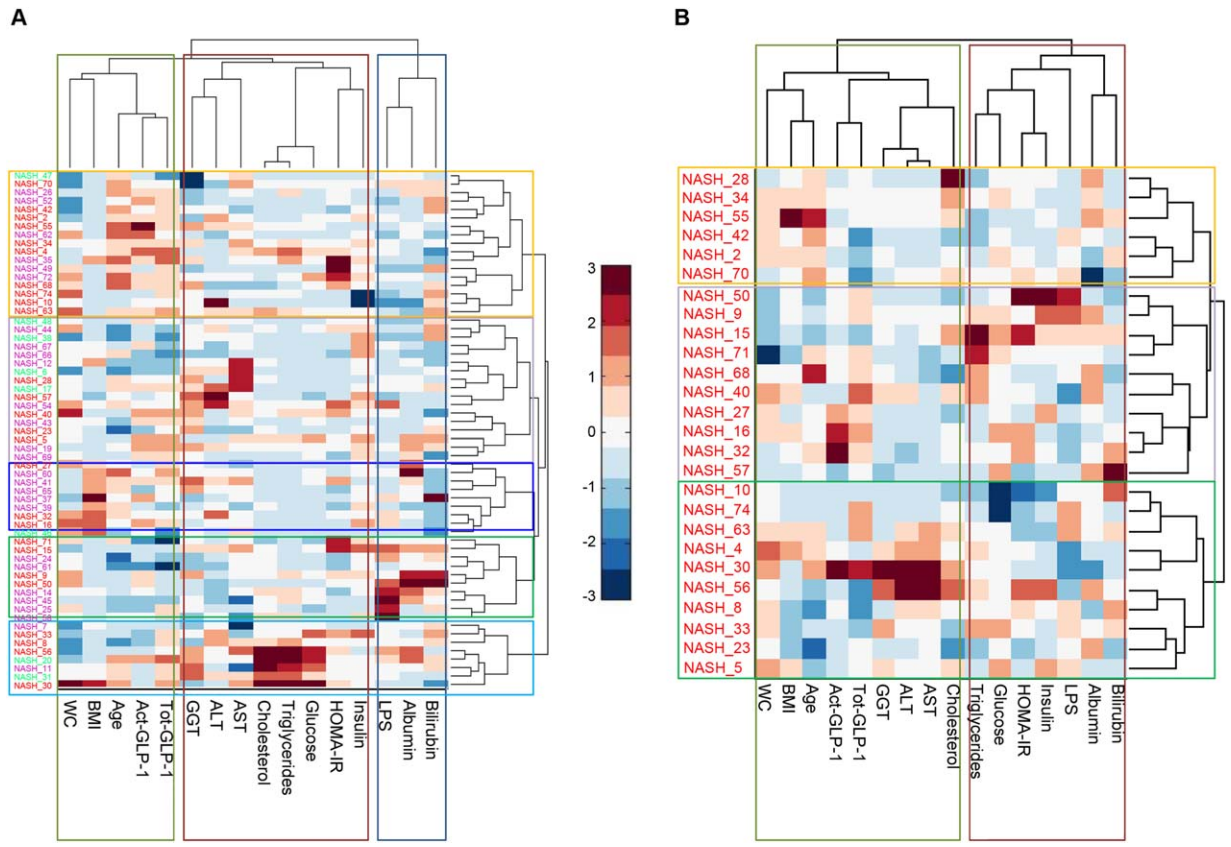

FIG. 1. Correlation heatmap for anamnestic parameters in all NAFLD patients (A) and for anamnestic parameters in NASH patients only (B). Red, NASH; pink, NAFL; green, obese.

grouped into three clusters (Fig. 1B): (1) patients with medium or high values for WC, BMI, age, albumin, and cholesterol and low values for the other PH data (n = 6); (2) patients with medium or high levels of triglycerides, bilirubin, albumin, glucose, insulin, and HOMA-IR (n = 10); and (3) patients with high values for all the PH parameters but particularly group 1 (n = 10).

TARGETED MG OF GUT MICROBIOTA: ECOLOGICAL DIVERSITY OF NAFL, NASH, AND OBESE PATIENTS COMPARED TO CTRLS

A total of 918,876 sequencing reads were obtained from the 115 fecal samples. In terms of ecological diversity, the number of OTUs varied from 1,082 to 37,794 for the entire sample set, the Shannon index was 1.38 ÷ 5.90, the Chao1 index was 58.11 ÷ 899.56, phylogenetic distance was 6.00 ÷ 39.45, and the number of unique OTUs ranged from 43 to 480 with a Good's coverage of 0.95-1 (Supporting Table S1 and Fig. S1). Based on the Shannon index, there was a hierarchy of ecological diversity in which CTRLs > obese > NASH > NAFL (Supporting Fig. S1). An analysis of variance was used to assess significant differences in the α-diversity term for the number of OTUs, Shannon index, and Chao1 index and the entire phylogenetic distance tree for all four groups (NASH, NAFL, obese, and CTRLs). Good's coverage was not included in the analysis. Significantly different ($P < 0.05$) numbers of OTUs were recorded for the CTRL (4,808.56 ± 4,388.67) and NASH (9,206.81 ± 6,446.73) groups, the CTRL and NAFL (13,837.22 ± 10,381.8) groups, and the NAFL and obese (5,779 ± 2,696.25) groups (Supporting Table S1). All of the differences in the Shannon index were statistically significant (Supporting Table S1). The values for the α-Shannon pairs for each patient subset were compared to the CTRLs using Fisher's least significant difference test applied to

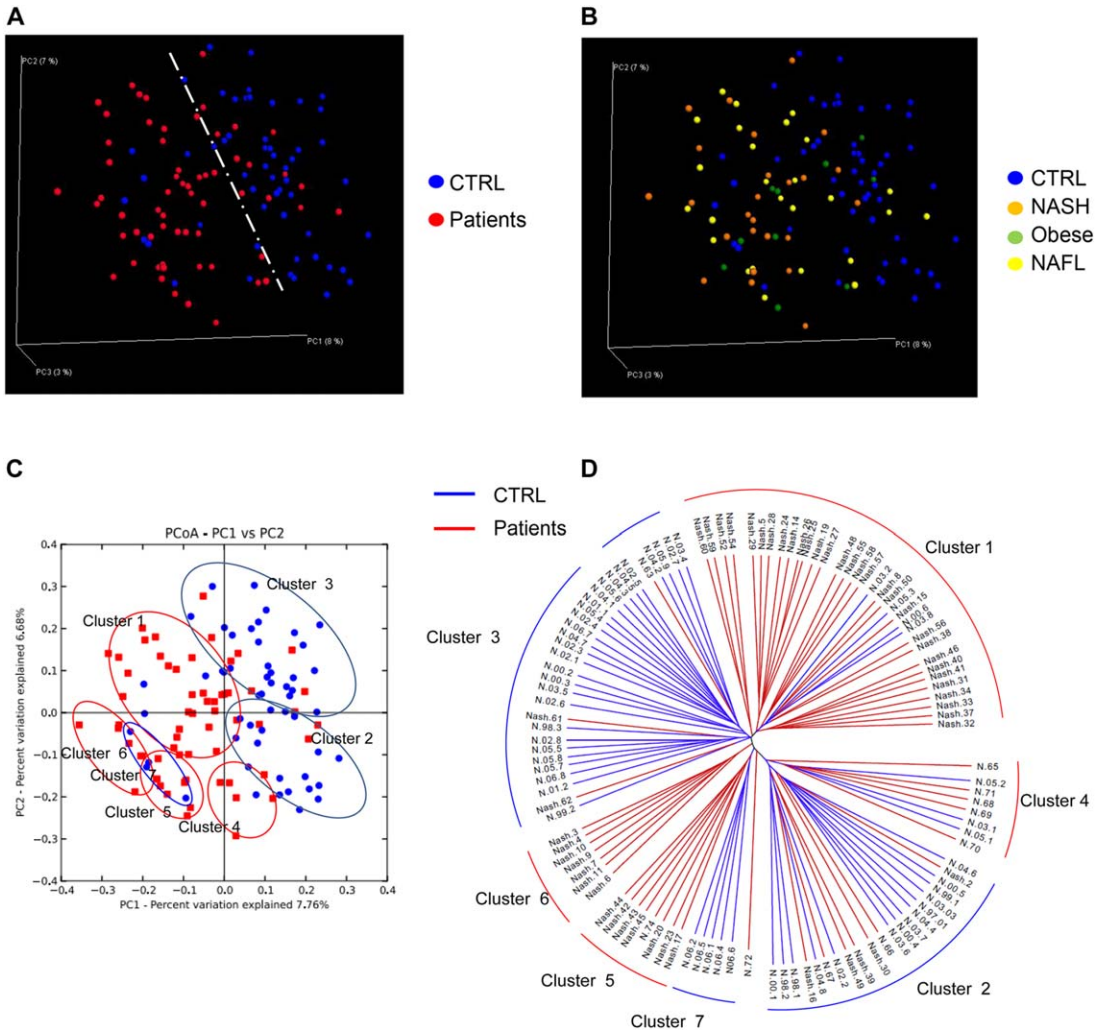

FIG. 2. Metagenomics β-diversity plots and phylogenetic unweighted pair group method with arithmetic mean tree. Principal component plots of the diversity in the microbial communities from fecal samples from all subjects (A) and compared to CTRLs stratified by diagnosis with NAFL, NASH, or obesity (B); jackknifing analysis to measure the robustness of β-diversity by two-dimensional plot (C) and unweighted pair group method with arithmetic mean hierarchical clustering (D). Blue line, CTRL; red line, patient sample set. Abbreviations: PC, principal component; PCoA, principal coordinate analysis.

differences in the mean values. The following pairs were statistically significant: NAFL versus NASH (-0.75 ± 0.00), NAFL versus obese (-0.98 ± 0.01), and NAFL versus CTRLs (-0.98 ± 0.00). In contrast, the mean differences in the α-Shannon values between the NASH versus CTRL, NASH versus obese, and obese versus CTRL groups were not significant (Supporting Table S1). The Chao1 and entire phylogenetic distance tree analyses showed statistically significant differences between the CTRLs and both the NASH and NAFL groups. The difference in Good's coverage was statistically different only between the CTRL and NAFL groups (Supporting Table S1). The β-diversity, visualized by PCA, showed clear separation between the patient and CTRL groups (Fig. 2A). However, there were no clearly defined clusters for the patients stratified into the NAFL, NASH, and obese groups (Fig. 2B). This was confirmed using a jackknifing analysis performed on a smaller, randomly chosen number of sequences from each sample plotted in a two-dimensional β-diversity plot (Fig. 2C) and by the unweighted pair group with arithmetic mean hierarchical clustering method, which indicated three main and four secondary clusters (Fig. 2D).

The majority of the OTU sequences were assigned to six main phyla: Actinobacteria, Bacteroidetes,

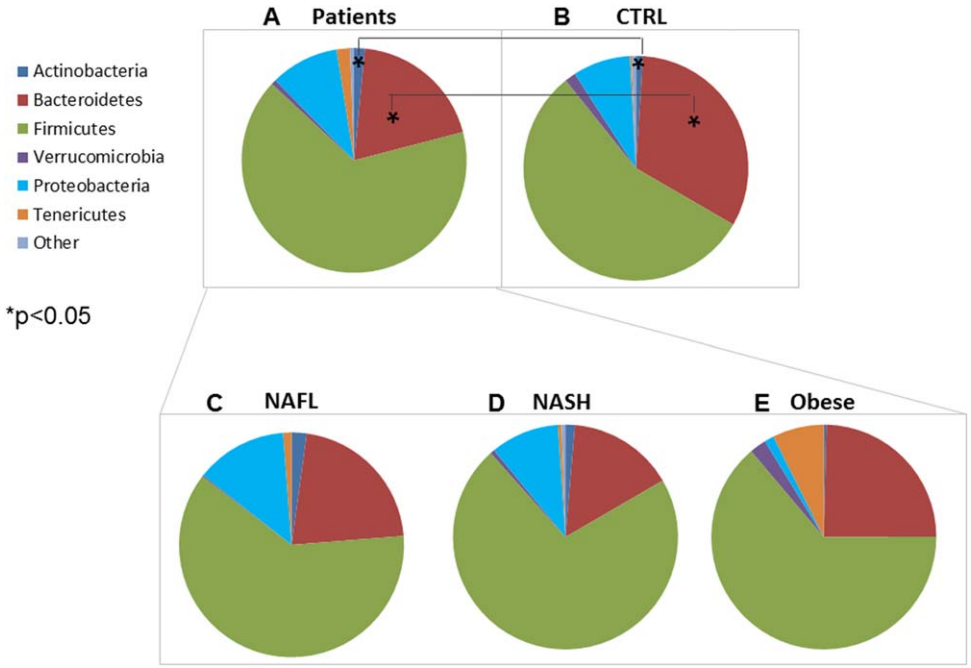

FIG. 3. Pie charts of OTU distribution. The distribution of the OTUs was compared between patients (A) and CTRLs (B) and within the NAFLD patients stratified by diagnosis: NAFL (C), NASH (D), and obese (E).

Firmicutes, Verrucomicrobia, Proteobacteria, and Tenericutes. The most abundant phyla in all of the subjects were Firmicutes, Bacteroidetes, and Proteobacteria. In the NAFL patients, the main phyla were Firmicutes (61%), Bacteroidetes (21%), Proteobacteria (13%), and Actinobacteria (2.1%). In the NASH patients compared to the NAFL patients, Firmicutes were increased (71%) and Bacteroidetes (15%), Proteobacteria (9.8%), and Actinobacteria (1.3%) were decreased (Supporting Table S2). The relative abundance of Firmicutes (63%) and Bacteroidetes (24%) in the obese group was similar to that of the NAFL patients, while Proteobacteria (1.36%) and Actinobacteria (0.4%) decreased compared to the NAFL patients (Supporting Table S2). In the CTRLs compared to the patient groups, Firmicutes (55%) were decreased and Bacteroidetes (32%) were increased (Supporting Table S2). A Kruskal-Wallis test indicated that Actinobacteria (0.0153 versus 0.0099, $P = 0.048$) was significantly increased in all patients compared to the CTRLs and that Bacteroidetes (0.1940 versus 0.3236, $P = 0.029$) was significantly decreased (Fig. 3A,B; Supporting Table S2). There were no significant differences between the CTRLs and the NAFL and NASH patient groups; however, the obese patients were significantly different from the CTRLs and the NAFL group (Fig. 3C-E; Supporting Table S2).

In Table 2 the distribution of the OTUs at the species level is shown for those OTUs that have at least 1% abundance. Among the Actinobacteria, the Coriobacteriaceae appeared to decrease across the groups (NAFL > NASH > obese versus CTRLs), but these findings did not attain statistical significance. Within the Bacteroidetes, the relative abundance of the Bacteroidaceae family and the *Bacteroides* genus was not significantly different between the patient groups and CTRLs, but there was some evidence that they were reduced in the NASH patients. At the species level, only *Bacteroides ovatus* showed some evidence of variation between the groups, but the results did not attain statistical significance, consistent with the trends at higher taxonomic levels. In addition, Rikenellaceae was more abundant in the CTRLs compared to each patient group, but the results were not statistically significant. Among the Firmicutes, the abundance of the *Oscillospira* genus was significantly lower ($P = 0.053$) in all of the patient groups (0.66 in NAFL and NASH, 1.19 in obese) compared to the CTRLs (2.05). Grouping the patients, *Bradyrhizobium*, *Anaerococcus*, *Peptoniphilus*, *Propionibacterium acnes*, *Dorea*, and *Ruminococcus* were statistically ($P < 0.05$) increased, while *Oscillospira* and Rikenellaceae appeared reduced compared to the CTRL group (Supporting Table S2).

TABLE 2. Abundance of Gut Microbiota OTUs at the Species Level (L2) in Patients and Control Subjects

OTUs*			NAFL†		NASH		Obese		CTRL		*P*	FDR
Actinobacteria			2.15	1.34	0.39	0.99					0.01	0.06
	Coriobacteriaceae		2.01	0.89	0.27	0.92					0.03	0.18
		Collinsella aerofaciens	1.48	0.41	0.08	0.53					0.07	0.39
Bacteroidetes			21.45	15.28	24.4	32.36					0.01	0.06
	Bacteroidaceae		14.75	8.39	19.79	18.27					0.01	0.15
		Bacteroides	10.76	6.38	14.67	12.9					0.02	0.22
		Bacteroides caccae	1.01	1.24	0.92	1.16					0.70	0.83
		Bacteroides fragilis	0.23	0.04	2.35	1.15					0.31	0.56
		Bacteroides ovatus	2.38	0.41	1.49	2.44					0.02	0.22
	Porphyromonadaceae		0.54	1.6	0.18	2.03					0.02	0.15
		Parabacteroides	0.23	0.75	0.11	1.59					0.08	0.39
	Prevotellaceae		1.66	2.98	0.27	1.98					0.93	0.95
		Prevotella	0.3	0.14	0.09	1.27					0.62	0.80
		Prevotella copri	1.36	2.83	0.0	0.7					0.40	065
	Rikenellaceae		**3.41**	**1.44**	**3.7**	**6.08**					**0.00**	**0.05**
	(Barnesiellaceae)		0.93	0.48	0.26	2.07					0.06	0.35
Firmicutes			61.36	71.81	63.5	55.91					0.06	0.30
	Lactobacillaceae		0.32	1.14	0.01	0.03					0.14	0.50
	Streptococcaceae		1.72	2.07	0.55	1.50					0.05	0.30
		Streptococcus	1.7	2.06	0.54	1.31					0.05	0.37
	Clostridiaceae		1.47	4.08	4.43	2.42					0.01	0.70
		Clostridium	0.72	1.07	3.23	1.54					0.42	0.67
	Lachnospiraceae		18.22	31.02	12.28	10.75					0.56	0.54
		Blautia	1.27	2.77	1.14	1.3					0.04	0.33
	Ruminococcaceae		27.20	26.63	29.74	28.95					0.81	0.87
		Oscillospira	**0.66**	**0.66**	**1.19**	**2.05**					**0.00**	**0.05**
		Ruminococcus	0.63	1.98	4.75	3.3					0.16	0.56
		Faecalibacterium prausnitzii	6.05	5.83	8.21	5.33					0.72	0.83
	Erysipelotrichaceae		1.74	4.29	0.51	3.44					0.10	0.54
		Eubacterium biforme	0.07	0.61	0.02	1.11					0.12	0.47
	Veillonellaceae		13.64	15.06	13.15	8.88					0.79	0.87
		Dialister	11.73	14.27	12.8	7.81					0.54	0.73
		Phascolarctobacterium	1.41	0.17	0.03	0.66					0.61	0.79
Proteobacteria			13.09	9.88	1.36	8.27					0.56	0.67
	Comamonadaceae		0.01	0.06	0.01	1.06					0.82	0.86
		Comamonas	0.00	0.00	0.00	1.05					0.12	0.47
	Enterobacteriaceae		9.61	9.34	1.12	9.36.042					0.85	0.89
	Pseudomonadaceae		3.26	0.02	0.00	0.01					0.68	0.86
Verrucomicrobia			0.18	0.61	2.4	1.6					0.25	0.62
	Verrucomicrobiaceae	*Akkermansia muciniphila*	0.18	0.61	2.4	1.6					0.23	0.56
Tenericutes	Mollicutes	*RF39*	1.2	0.5	7.3	0.2					0.43	0.68

*OTUs with average abundance ≥1%.
†Numbers listed are percentages. Bold indicates significance.

TARGETED MB OF GUT MICROBIOTA: VOCS DISCRIMINATE BETWEEN PATIENTS AND CTRLS

Bacterial metabolites were isolated from fecal samples as VOCs. The following 292 VOCs were characterized: terpenes (n = 39); indoles (n = 6); azetidines (n = 1); amides (n = 1); amines (n = 3); thiols (n = 1); eterocycles (n = 3); furanones (n = 2); alchohols (n = 47); aldehydes (n = 19); esters (n = 51); acids (n = 13); aromatic hydrocarbons (n = 17); phenols (n = 5); sulfur compounds (n = 2); pyrazines (n = 1); pyridines (n = 5); ketons (n = 38); alkenes (n = 20); alkanes (n = 17); and hydrazines (n = 1). There was substantial variance between samples and low variance between technical replicates. The levels of 26 of the 292 metabolites, primarily alcohols, acids, aldehydes, ketones, amines, and esters that result from microbial actions, were up-regulated in the patients (NASH/NAFLD/obese) compared to CTRLs (Supporting Table S3). Only 2 of the 292 metabolites (aromatic hydrocarbons and hydrazines) were down-regulated in the patients compared to CTRLs. For further details, see Supporting Information.

INTEGRATED PROFILING OF THE TARGETED MG AND MB RESULTS

The data pertaining to the OTUs and VOCs were compiled into a single database and processed using multivariate analyses. Databases compiled using this approach are characterized by a large quantity of null data that in many cases lead to misleading results by affecting the overall variance of the data field.[22] To prevent this, prior to assembling the database, the data sets were iteratively tested to find less populated variables and subjects so that the final database was <25% null data. The final compiled data set included 44 variables (33 OTUs and 11 VOCs) and 89 subjects (45 CTRLs, 5 obese, 16 NAFL, and 23 NASH). PCA was used to explore the data in an unsupervised manner and identify outliers. After excluding 6 patients in the NASH group as outliers, the score plot of the first two components did not show samples clustering depending on the disease status (Supporting Fig. S2), and the variance explained by the first two components was low (19% of total variance). Outliers were still present in the principal component score plot outside the Hotelling T^2 ellipse including 3 of 5 obese children, 1 CTRL, and 1 NASH patient. Therefore, the final analysis was performed excluding the obese children and all of the outliers. To identify an integrated MG/MB profile for CTRLs and patients, PLS-DA was performed using 44 variables and the 79 remaining subjects, using the health status (CTRLs or patients) as an external *Y* variable.

The median values and interquartile range of the data used for the multivariate analysis are reported in Supporting Tables S4 and S5. The PLS-DA score plot showed a significant distinction between the CTRLs and the NAFLD patients along latent variable 1 (LV1) (R^2 = 0.688, Q^2 = 0.328; Fig. 4A). Based on the PLS model, patients had significantly higher levels of Lachnospiraceae, 4-methyl-2-pentanone, 1-butanol, and 2-butanone compared to CTRLs (Supporting Fig. S3A). In addition, patients had a lower abundance of *Oscillospira*, Rikenellaceae, *Parabacteroides*, *Bacteroides fragilis*, and *Sutterella* compared to CTRLs (Supporting Fig. S3A). The PLS-DA model did not appear to discriminate between the profiles of patients with NASH and NAFL (Fig. 4A). The average area under the receiver operating characteristic curve (AUROC = 0.978) confirmed that the model was able to discriminate between patients and CTRLs (Fig. 4B). Thus, the PLS-DA model primarily reflects differences in the MG/MB profile of liver steatosis in NAFL and NASH patients compared to healthy individuals.

To determine whether targeted MG and MB variables could be used as putative biomarkers to discriminate disease progression, we independently analyzed the NAFL and NASH patient groups. However, the PLS-DA model (Fig. 4C) showed that only the NAFL patients could be significantly distinguished from CTRLs (R^2 = 0.768, Q^2 = 0.450) along both significant LVs. Indeed, the average AUROC (0.997) confirmed that the model was able to discriminate the NAFL patients from the CTRL group (Fig. 4D). Based on the PLS model, NAFL patients had higher levels of Lachnospiraceae, 1-pentanol, 2-butanone, and 2-pentanone and low abundance of *Oscillospira*, *Parabacteroides distasonis*, *Parabacteroides*, *Bacteroides fragilis*, *Coprococcus*, and *Anaerostipes* compared to CTRLs (Supporting Fig. S3B). However, as this is based on NAFL patients, this PLS-DA model reflects the targeted MG/MB profile of liver steatosis not complicated by fibrosis.

The PLS-DA models built using data from the CTRLs versus NASH (Supporting Fig. S4) and NAFL versus NASH patients did not show any significant discriminating features between groups (one LV

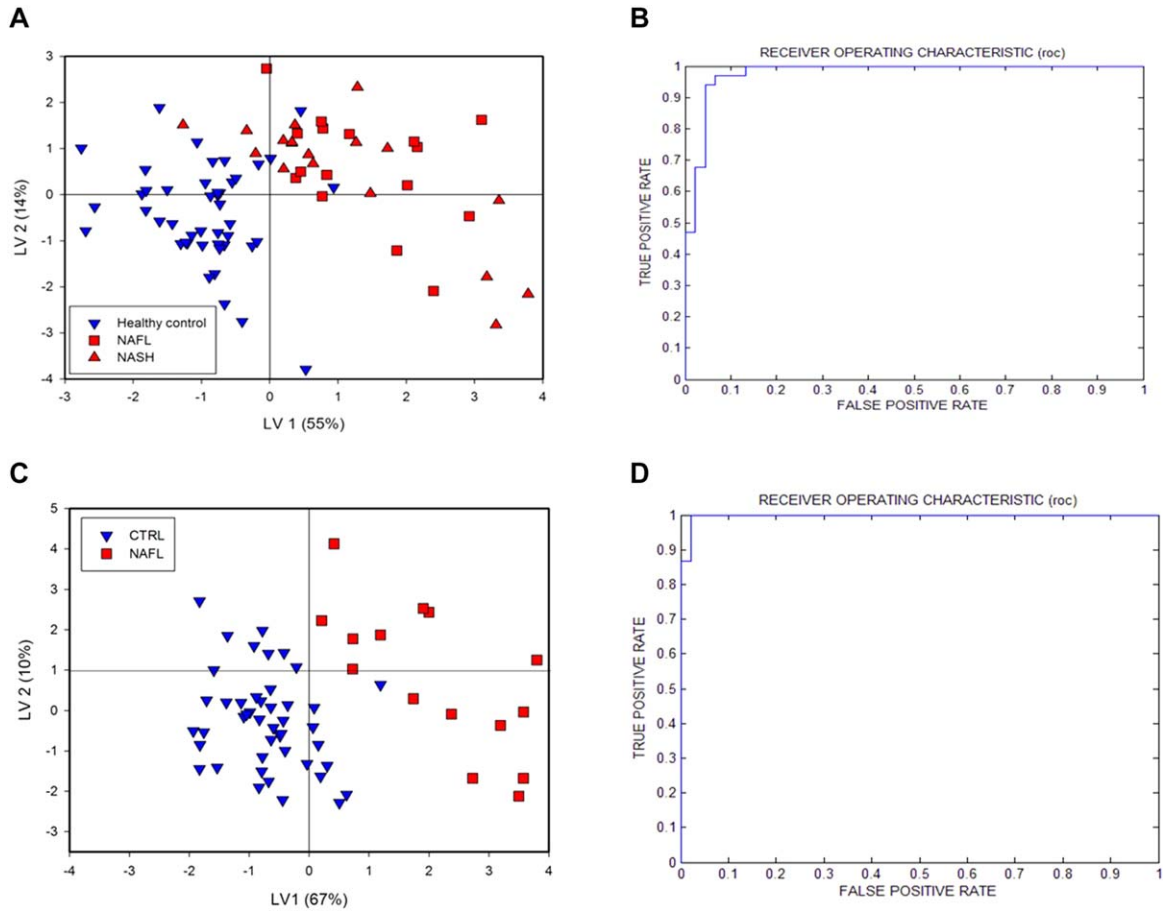

FIG. 4. PLS-DA of CTRLs versus NAFL and NASH patients and related receiver operating characteristic curves. The PLS scores plot shows significant distinctions between the CTRLs and patients along LV1 (R^2 = 0.688, Q^2 = 0.328) (A). The average AUROC (0.9778) confirmed that the model was able to discriminate patients from healthy control (B). PLS-DA of CTRLs versus NAFL. The PLS-DA model based on the CTRL and NAFLD data showed significant distinctions between the two classes (R^2 = 0.768, Q^2 = 0.450, two significant LVs) (C). The average AUROC (0.997) confirmed that the model was able to discriminate NAFL patients from CTRLs (D).

R^2 = 0.480, Q^2 = 0.186; one LV R^2 = 0.444, Q^2 = ND, respectively), meaning that these PLS-DA models were not able to show statistically robust covariations of targeted MG and MB variables with liver fibrosis.

The Spearman rank test was used to assess whether there were any associations between targeted MG and MB variables (Supporting Table S6). The following correlations were observed: (1) a positive correlation between *Streptococcus* and *Blautia* in all subjects and the CTRL, NAFL, and NASH groups; (2) a negative correlation between *Streptococcus* and *Oscillospira* in all subjects and the CTRLs; (3) a positive correlation between 1-butanol and 1-pentanol in all subjects and the CTRL, NAFL, and NASH groups; (4) a positive correlation between 4-methyl-2-pentanone and *Blautia* in the NAFL group; and (5) a negative correlation between 2-butanone and *Ruminococcus* and *Coprococcus* in the NAFL group (Supporting Table S6).

Spearman rank correlations were used to assess the relationships between significant targeted MG/MB variables and some PH variables (Supporting Table S7). In all of the patients with common liver steatosis treatment, the abundance of *Blautia* and *Streptococcus* was positively correlated with LPS levels (R = 0.37 and 0.43, P = 0.0207 and 0.00558, respectively), *Ruminococcus* was correlated with insulin levels (R = 0.37, P = 0.0204), and *Dorea* was correlated with BMI (R = 0.41, P = 0.0101). None of the metabolites correlated with PH variables (Supporting Table S7). Notably, the correlations between *Blautia* and *Streptococcus* and LPS levels were driven by the NASH

TABLE 3. Univariate Statistical Analysis of OTUs and Metabolites Found To Be Significantly Different by the Mann-Whitney U Test

	CTRL		NAFL		NASH	
	Median	IQR	Median	IQR	Median	IQR
Streptococcus[*]	0.080	0.000-0.348	0.158	0.024-0.608	0.382	0.078-0.952
Lachnospiraceae *Ruminococcus*[†]	0.038	0.000-0.264	0.157	0.018-0.402	0.339	0.078-1.337
Lachnospiraceae *Blautia*[‡]	0.376	0.089-1.044	0.129	0.080-0.454	0.728	0.203-3.052
Lachnospiraceae *Coprococcus*[§,‖]	0.251	0.058-0.969	0.071	0.019-0.250	0.356	0.082-0.777
Lachnospiraceae *Dorea*[†]	0.000	0.000-0.060	0.009	0.000-0.086	0.101	0.020-0.237
Oscillospira[†,¶]	1.110	0.506-3.620	0.337	0.087-0.572	0.603	0.105-1.398
1-Butanol[§]	8.884	0.107-35.984	44.535	7.277-91.160	16.357	0.133-34.552
1-Pentanol[‡,¶]	10.881	3.439-34.510	28.590	12.894-86.488	6.588	0.115-18.667
Phenol[§]	16.821	2.678-35.813	57.444	22.069-145.729	24.734	7.223-112.237
2-Butanone[†,¶]	14.980	4.959-29.310	41.789	12.171-140.312	46.521	22.063-120.637
4-Methyl-2-pentanone[†,‡]	264.781	72.536-300.865	171.505	18.233-480.070	532.025	315.450-721.279

Statistical significance expressed as follows: [*]HC versus NASH (uncorrected $P < 0.05$), [†]HC versus NASH (FDR-adjusted P of VOCs and OTUs), [‡]NAFL versus NASH (FDR-adjusted P of VOCs and OTUs), [§]HC versus NAFL (uncorrected $P < 0.05$), [‖]NAFL versus NASH (uncorrected $P < 0.05$), [¶]HC versus NAFL (FDR-adjusted P of VOCs and OTUs).

patients, and the correlation between *Ruminococcus* and insulin levels was mainly driven by the NAFL patients.

To decrease the dimensionality of the data, univariate analyses were performed on the reduced data obtained by multivariate analysis (Table 3). Significantly lower levels of *Oscillospira* and higher levels of 1-pentanol and 2-butanone were reported in patients with NAFL compared to CTRLs (Table 3). NASH patients had lower levels of *Oscillospira* and were independently associated with high abundance of *Dorea* and *Ruminococcus* and with higher levels of 2-butanone and 4-methyl-2-pentanone compared to CTRLs (Table 3).

Discussion

Targeted MG data highlighted values of α-diversity, suggestive of reduced diversity going from CTRLs to obese, NASH, and finally NAFL patients in a pediatric setting. Despite a higher number of sequences for the NAFL samples compared to the other groups, the species richness appeared reduced, suggesting the lowest ecological diversity under NAFL conditions. A pivotal study by Turnbaugh and coworkers[28] suggested that obesity in mouse models can be discussed in terms of a significant decrease in the level of diversity in the gut microbiota. This suggests an analogy between obese mice and gut microbiota of obese patients characterized by reduced diversity of the microbial community and abnormal energy input.[29]

The β-diversity parameter clearly distinguished between patients with NAFLD and CTRLs. In contrast, when patients were stratified based on diagnosis (NAFL, NASH, or obese) there was no net separation between groups. The sum of these analyses indicated that health status was a major effect factor for determining the phylogenetic composition of these samples. Exceptions were present in all four study groups, likely reflecting the effects of other genetic and environmental factors on the microbiome.[12]

With regard to the phylum present, the most abundant OTUs in all subjects were Firmicutes, followed by Bacteroidetes, Proteobacteria, Actinobacteria, Verrucomicrobia, and Tenericutes. Actinobacteria were statistically prevalent in the patients' set versus CTRLs, while the opposite was observed for Bacteroidetes. Differences between our study and previous work likely reflect variations in environmental and dietary factors, as well as age, among the patient cohorts.[12,30]

At lower taxonomic levels, the abundance of Coriobacteriaceae, which has been linked to chronic tissue inflammation and is increased in patients with colon-only Crohn's disease,[31] was higher in patients with NAFLD than CTRLs. Within the NAFLD patients, abundance was greatest in the NAFL patients, followed by NASH and obese patients. These observations suggest that Coriobacteriaceae may have a role in the necroinflammation observed in the NAFL phenotype.

Bacteroidaceae and *Bacteroides* were reduced in NAFL and NASH and increased in obese patients compared to CTRLs. The abundance of *Blautia* was extremely high in NASH but not in NAFL or obese children. Remarkably, *Oscillospira* was significantly less abundant in the NAFL, NASH, and obese patients compared to the CTRLs. This observation was corroborated by multivariate and univariate analyses. These findings are consistent with the microbiome description published by Raman and coworkers.[30] In a study by Lam and coworkers,[32] *Oscillibacter*-like organisms were

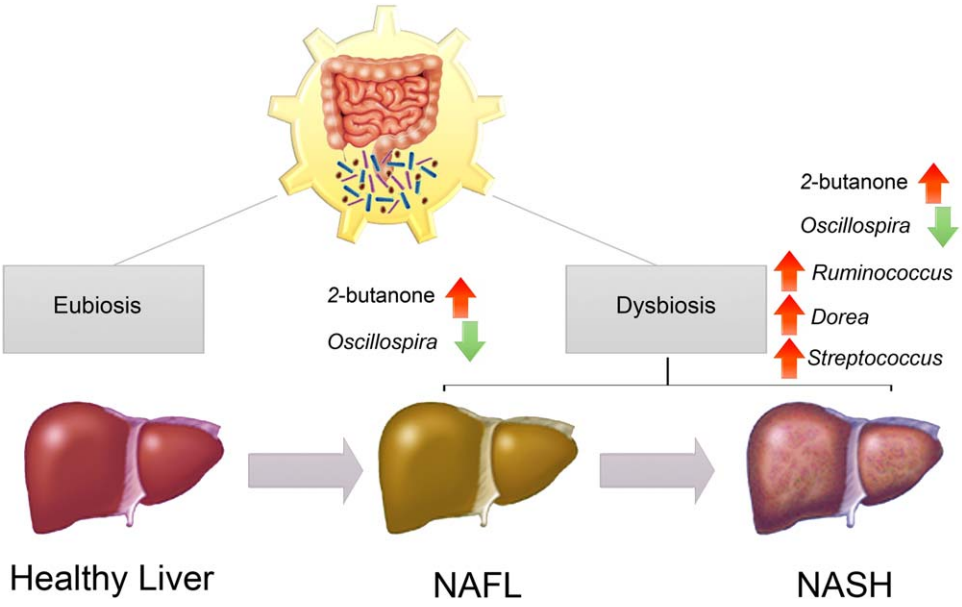

FIG. 5. Graphical picture of microbiota MB and MG signatures at the onset and during progression of pediatric NAFLD.

characterized as a potentially important gut microbe that mediates high saturated fat diet-induced gut dysfunction. This group, including *Oscillibacter* and *Oscillospira*, has been consistently detected in the microbial community of humans.[33,34] Interestingly, in a study by Tims and coworkers,[35] *Oscillospira guillermondii* and related bacteria were more abundant in siblings with a lower BMI, consistent with our data.

Zhu and coworkers highlighted a specific functional interaction between Enterobacteriaceae (e.g., *Escherichia coli*) and ethanol metabolism, through the mixed-acid fermentation pathway.[12,13] While we did not observe significant differences in *E. coli* specifically compared to CTRLs, we did observe a greater abundance of Enterobacteriaceae in NAFL and NASH compared to CTRLs. Ethanol was also up-regulated in 9 out of 27 NASH patients compared to CTRLs. However, we did not measure blood or breath alcohol concentrations, only the volatile fraction by solid-phase microextraction gas chromatography-mass spectrometry. Indeed, Michail and coworkers,[36] who also measured ethanol in the soluble fraction, reported higher values of this metabolite in obese NAFLD patients compared to obese subjects or CTRLs. Moreover, in a pivotal study by Raman and coworkers,[30] ethanol did not emerge as a specific NASH-related metabolite but was a ubiquitous element of the gut microbiota, regardless of disease or health. In addition to *E. coli*, Zhu and coworkers[12] noted that other gut microbial genera, including *Bacteroides*, *Bifidobacterium*, and *Clostridium*, are capable of producing alcohol. Therefore, collectively these genera may cause a significant burden for liver alcohol-scavenging mechanisms.

We were able to identify a specific fecal signature that distinguished patients with NAFL from healthy CTRLs characterized by high levels of 2-butanone and low relative abundance of *Oscillospira*. This signature may be a potential biomarker profile for liver steatosis. In another recent gas chromatography-mass spectrometry study, the levels of 2-butanone in the breath of patients with liver disease were elevated.[37,38] 2-Butanone can be produced in the gut through oxidized nicotinamide adenine dinucleotide-dependent secondary alcohol dehydrogenase activity in a basic (pH) environment by microorganisms including yeast[39] and Enterobacteriaceae,[40] as well as endogenously through liver metabolism.[41] However, production of 2-butanone and 1-butanol has also been linked to *Streptococcus pneumoniae* metabolism,[42] which was elevated in NASH patients but did not attain statistical significance with FDR correction. Intriguingly, the abundance of *Streptococcus* and 1-butanol levels appeared to be negatively correlated with insulin levels in NASH patients, suggesting that this genus may contribute to insulin-resistance preceding type 2 diabetes, a common metabolic syndrome of NAFLD.

Univariate analyses supported further investigations into the relationship between increased 4-methyl-2-

pentanone and NASH in larger cohorts. The increase in *Ruminococcus* and *Dorea* suggests a possible association with the NAFL-NASH transition. However, the positive correlations between *Ruminococcus* and *Dorea* and between insulin levels and BMI underscore the complexity of the metabolic processes in the transition from steatosis to steatohepatitis. The concurrent changes in *Ruminococcus*, *Dorea*, *Streptococcus*, and *Oscillospira* should be further investigated in relation to the evolution of liver steatosis toward fibrosis (Fig. 5). Overall, during the evolution of liver pathology, the variations in the whole microbial ecosystem appear to be more important than single variations of specific genera or species.

To the best of our knowledge, this is the first proof-of-concept study applying an integrated omics-based approach to differentially describe the gut microbiota cometabolism in NAFLD patients compared to healthy controls in a pediatric setting.

In the complex scenario of NAFLD, the interrelated enterotype-metabotype framework appears to contribute to creating a signature that changes during disease progression. In particular, the combination of a low abundance of *Oscillospira* with high levels of 2-butanone may be a specific intestinal MG and MB profile for liver steatosis in children. The high relative abundance of Lachnospiraceae, *Ruminococcus*, and *Dorea* observed in pediatric patients with NASH suggests that changes in the gut microbiota are associated with disease severity.

These findings might provide development of a specific metabolic diagnostic profile of steatosis and suggest a first-line probiotic candidate for treating NAFLD. Future studies in larger, well-characterized cohorts are required to validate the proposed model and to better describe other associations between gut microbiota phylotypes, metabotypes, and disease phenotypes.

Acknowledgment: We thank the native English-speaking experts from BioMed Proofreading LLC for English revision.

REFERENCES

1) Llorente C, Schnabl B. The gut microbiota and liver disease. Cell Mol Gastroenterol Hepatol 2015;1:275-284.
2) Mehal WZ. The gut–liver axis: a busy two-way street. HEPATOLOGY 2012;55:1647-1649.
3) Compare D, Coccoli P, Rocco A, Nardone OM, De Maria S, Cartenì M, et al. Gut-liver axis: the impact of gut microbiota on non alcoholic fatty liver disease. Nutr Metab Cardiovasc Dis 2012;22:471-476.
4) Malaguarnera M, Di Rosa M, Nicoletti F, Malaguarnera L. Molecular mechanisms involved in NAFLD progression. J Mol Med (Berl) 2009;87:679-695.
5) Abu-Shanab A, Quigley EMM. The role of the gut microbiota in nonalcoholic fatty liver disease. Nat Rev Gastroenterol Hepatol 2010;7:691-701.
6) Vanni E, Bugianesi E. The gut–liver axis in nonalcoholic fatty liver disease: another pathway to insulin resistance? HEPATOLOGY 2009;49:1790-1792.
7) Wigg AJ, Roberts-Thomson IC, Dymock RB, McCarthy PJ, Grose RH, Cummins AG. The role of small intestinal bacterial overgrowth, intestinal permeability, endotoxaemia, and tumour necrosis factor alpha in the pathogenesis of non-alcoholic steatohepatitis. Gut 2001;48:206-211.
8) Milić S, Stimac D. Nonalcoholic fatty liver disease/steatohepatitis: epidemiology, pathogenesis, clinical presentation and treatment. Dig Dis 2012;30:158-162.
9) Chalasani N, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, et al. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Association for the Study of Liver Diseases, American College of Gastroenterology, and the American Gastroenterological Association. HEPATOLOGY 2012;55:2005-2023.
10) Musso G, Gambino R, Cassader M. Non-alcoholic fatty liver disease from pathogenesis to management: an update. Obes Rev 2010;11:430-445.
11) Day CP, James OF. Steatohepatitis: a tale of two "hits"? Gastroenterology 1998;114:842-845.
12) Zhu L, Baker SS, Gill C, Liu W, Alkhouri R, Baker RD, et al. Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH. HEPATOLOGY 2013;57:601-609.
13) **Zhu L**, **Baker RD**, Baker SS. Gut microbiome and nonalcoholic fatty liver diseases. Pediatr Res 2015;77:245-251.
14) Aron-Wisnewsky J, Gaborit B, Dutour A, Clement K. Gut microbiota and non-alcoholic fatty liver disease: new insights. Clin Microbiol Infect 2013;19:338-348.
15) Bajaj JS, Hylemon PB, Younossi Z. The intestinal microbiota and liver disease. Am J Gastroenterol Suppl 2012;1:9-14.
16) Bajaj JS, Hylemon PB, Ridlon JM, Heuman DM, Daita K, White MB, et al. Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. Am J Physiol Gastrointest Liver Physiol 2012;303:G675-G685.
17) Alisi A, Ceccarelli S, Panera N, Nobili V. Causative role of gut microbiota in non-alcoholic fatty liver disease pathogenesis. Front Cell Infect Microbiol 2012;2:132.
18) Burcelin R. Regulation of metabolism: a cross talk between gut microbiota and its human host. Physiology (Bethesda) 2012;27:300-307.
19) Vajro P, Paolella G, Fasano A. Microbiota and gut–liver axis: their influences on obesity and obesity-related liver disease. J Pediatr Gastroenterol Nutr 2013;56:461-468.
20) Burcelin R, Serino M, Chabo C, Garidou L, Pomié C, Courtney M, et al. Metagenome and metabolism: the tissue microbiota hypothesis. Diabetes Obes Metab 2013;15(Suppl. 3):61-70.
21) Del Chierico F, Gnani D, Vernocchi P, Petrucca A, Alisi A, Dallapiccola B, et al. Meta-omic platforms to assist in the understanding of NAFLD gut microbiota alterations: tools and applications. Int J Mol Sci 2014;15:684-711.
22) Biesecker LG. Hypothesis-generating research and predictive medicine. Genome Res 2013;23:1051-1053.

23) Martens H, Martens M. Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). Food Qual Prefer 2000;11:5-16.
24) Martens H, Anderssen E, Flatberg A, Gidskehaug LH, Høy M, Westad F, et al. Regression of a data matrix on descriptors of both its rows and of its columns via latent variables: L-PLSR. Comput Stat Data Anal 2005;48:103-123.
25) Wold S, Høy M, Martens H, Trygg J, Westad F, MacGregor J, et al. The PLS model space revisited. J Chemom 2009;23:67-68.
26) Efron B. The Jackknife, the Bootstrap, and Other Resampling Plans. Philadelphia, PA: Society for Industrial and Applied Mathematics; 1982.
27) Benjamini Y, Hocberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 1995;57:289-300.
28) Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature 2009;457:480-484.
29) Watson SB, Mccauley E, Downing JA. Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status. Limnol Oceanogr 1997;42:487-495.
30) Raman M, Ahmed I, Gillevet PM, Probert CS, Ratcliffe NM, Smith S, et al. Fecal microbiome and volatile organic compound metabolome in obese humans with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2013;11:868-875.
31) Willing BP, Dicksved J, Halfvarson J, Andersson AF, Lucio M, Zheng Z, et al. A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. Gastroenterology 2010;139:1844-1854.
32) Lam YY, Ha CWY, Campbell CR, Mitchell AJ, Dinudom A, Oscarsson J, et al. Increased gut permeability and microbiota change associate with mesenteric fat inflammation and metabolic dysfunction in diet-induced obese mice. PLoS One 2012;7:e34233.
33) Mondot S, Kang S, Furet JP, Aguirre de Carcer D, McSweeney C, Morrison M, et al. Highlighting new phylogenetic specificities of Crohn's disease microbiota. Inflamm Bowel Dis 2011;17:185-192.
34) Claesson MJ, O'Sullivan O, Wang Q, Nikkilä J, Marchesi JR, Smidt H, et al. Comparative analysis of pyrosequencing and a phylogenetic microarray for exploring microbial community structures in the human distal intestine. PloS One 2009;4:e6669.
35) Tims S, Derom C, Jonkers DM, Vlietinck R, Saris WH, Kleerebezem M, et al. Microbiota conservation and BMI signatures in adult monozygotic twins. ISME J 2013;7:707-717.
36) Michail S, Lin M, Frey MR, Fanter R, Paliy O, Hilbush B, et al. Altered gut microbial energy and metabolism in children with non-alcoholic fatty liver disease. FEMS Microbiol Ecol 2015;91:1-9.
37) Boursier J, Diehl AM. Implication of gut microbiota in nonalcoholic fatty liver disease. PLOS Pathog 2015;11:e1004559.
38) Fernández Del Río R, O'Hara ME, Holt A, Pemberton P, Shah T, Whitehouse T, et al. Volatile biomarkers in breath associated with liver cirrhosis—comparisons of pre- and post-liver transplant breath samples. EBioMedicine 2015;2:1243-1250.
39) Van den Velde S, Nevens F, Van Hee P, van Steenberghe D, Quirynen M. GC-MS analysis of breath odor compounds in liver patients. J Chromatogr B Analyt Technol Biomed Life Sci 2008;875:344-348.
40) Patel RN, Hou CT, Laskin AI, Derelanko P, Felix A. Microbial production of methyl ketones. Purification and properties of a secondary alcohol dehydrogenase from yeast. Eur J Biochem 1979;101:401-406.
41) Hou CT, Patel R, Laskin AI, Barnabe N. Microbial oxidation of gaseous hydrocarbons: epoxidation of C2 to C4 n-alkenes by methylotrophic bacteria. Appl Environ Microbiol 1979;38:127-134.
42) Mathews JM, Raymer JH, Etheridge AS, Velez GR, Bucher JR. Do endogenous volatile organic chemicals measured in breath reflect and maintain CYP2E1 levels *in vivo*? Toxicol Appl Pharmacol 1997;146:255-260.

Author names in bold designate shared co-first authorship.

Supporting Information

Additional Supporting Information may be found at onlinelibrary.wiley.com/doi/10.1002/hep.28572/suppinfo.