Computational and Structural Biotechnology Journal 19 (2021) 5920–5930

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Gut microbiome signatures distinguish type 2 diabetes mellitus from non-alcoholic fatty liver disease

Jiyeon Si [a,1], Giljae Lee [b,1], Hyun Ju You [b,c], Sae Kyung Joo [d], Dong Hyeon Lee [d], Bon Jeong Ku [e], Seoyeon Park [b], Won Kim [d,*], GwangPyo Ko [b,f,g,h,*]

[a] Medical Science Research Institute, School of Medicine, Sungkyunkwan University (SKKU), Suwon 16419, Republic of Korea
[b] Department of Environmental Health Sciences, Graduate School of Public Health, Seoul National University, Seoul 08826, Republic of Korea
[c] Institute of Health and Environment, Seoul National University, Seoul 08826, Republic of Korea
[d] Division of Gastroenterology and Hepatology, Department of Internal Medicine, Seoul National University College of Medicine, Seoul Metropolitan Government Boramae Medical Center, Seoul 07061, Republic of Korea
[e] Department of Internal Medicine, Chungnam National University School of Medicine, Daejeon 35015, Republic of Korea
[f] Center for Human and Environmental Microbiome, Institute of Health and Environment, Seoul National University, Seoul 08826, Republic of Korea
[g] KoBioLabs, Inc., Seoul 08826, Republic of Korea
[h] Bio-MAX/N-Bio, Seoul National University, Seoul 08826, Republic of Korea

ARTICLE INFO

Article history:
Received 1 July 2021
Received in revised form 22 October 2021
Accepted 26 October 2021
Available online 28 October 2021

Keywords:
Type 2 diabetes mellitus
Non-alcoholic fatty liver disease (NAFLD)
Gut microbiome
Biomarker
Enterotype

ABSTRACT

Non-alcoholic fatty liver disease (NAFLD) is closely associated with type 2 diabetes mellitus (T2D), and these two metabolic diseases demonstrate bidirectional influences. The identification of microbiome profiles that are specific to liver injury or impaired glucose metabolism may assist understanding of the role of the gut microbiota in the relationship between NAFLD and T2D. Here, we studied a biopsy-proven Asian NAFLD cohort (n = 329; 187 participants with NAFLD, 101 with NAFLD and T2D, and 41 with neither) and identified *Enterobacter*, *Romboutsia*, and *Clostridium sensu stricto* as the principal taxa associated with the severity of NAFLD and T2D, whereas *Ruminococcus* and *Megamonas* were specific to NAFLD. In particular, the taxa that were associated with both severe liver pathology and T2D were also significantly associated with markers of diabetes, such as fasting blood glucose and Hb1Ac. Enterotype analysis demonstrated that participants with NAFLD had a significantly higher proportion of *Bacteroides* and a lower proportion of *Ruminococcus* than a Korean healthy twin cohort (n = 756). However, T2D could not be clearly distinguished from NAFLD. Analysis of an independent T2D cohort (n = 185) permitted us to validate the T2D-specific bacterial signature identified in the NAFLD cohort. Functional inference analysis revealed that endotoxin biosynthesis pathways were significantly enriched in participants with NAFLD and T2D, compared with those with NAFLD alone. These findings may assist with the development of effective therapeutic approaches for metabolic diseases that are associated with specific bacterial signatures.

1. Introduction

Non-alcoholic fatty liver disease (NAFLD) has a broad spectrum of hepatic manifestations, ranging from non-alcoholic fatty liver to more serious conditions, including non-alcoholic steatohepatitis (NASH), advanced fibrosis, and cirrhosis. Owing to this wide spectrum of disease, lack of specific pharmacotherapy, and requirement for histological examination to make a definitive diagnosis, efforts to identify tangible targets for the treatment of NAFLD have been attempted using genetic, imaging, serological, and Omics approaches [1]. However, it is necessary to further evaluate the potential non-invasive biomarkers identified in this way. In addition, the existence of comorbidities further complicates the charac-

Abbreviations: ALT, alanine aminotransferase; BMI, body mass index; FBS, fasting blood sugar; FDR, false discovery rate; FLI, fatty liver index; HbA1c, glycosylated hemoglobin; LDL, low-density lipoprotein; LPS, lipopolysaccharide; MaAsLin2, microbiome multivariable association with linear models 2; NAFL, non-alcoholic fatty liver; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; NASH-CRN, non-alcoholic steatohepatitis clinical research network; PICRUSt2, phylogenetic investigation of communities by reconstruction of unobserved states 2; T2D, type 2 diabetes mellitus.

* Corresponding authors at: Division of Gastroenterology and Hepatology, Department of Internal Medicine, Seoul Metropolitan Government Boramae Medical Center, 20, Boramae-ro 5-gil, Dongjak-gu, Seoul 07061, Republic of Korea, (W. Kim). Department of Environmental Health Sciences, Graduate School of Public Health, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea (G. Ko).

E-mail addresses: drwon1@snu.ac.kr (W. Kim), gko@snu.ac.kr (G. Ko).

[1] These authors contributed equally.

https://doi.org/10.1016/j.csbj.2021.10.032

terization of the disease. For example, NAFLD is closely associated with type 2 diabetes mellitus (T2D): patients with T2D have a high prevalence of NAFLD (55%), and NAFLD is also a risk factor for T2D [2–4]. Furthermore, the presence of this comorbidity is often associated with a poorer prognosis and a higher incidence of mortality [5]. Emerging evidence indicates that these two metabolic diseases are characterized by alterations to the gut microbiome [6–8]. Previous research regarding these conditions and their associations with the gut microbiome has mostly focused on the identification of non-invasive microbial biomarkers of NAFLD and attempts to identify the roles of gut microbes in the pathophysiology of NAFLD. However, the additional effect of T2D on the gut microbiome that characterizes NAFLD has not been fully investigated.

In the present study, we aimed to identify disease-specific microbial signatures and the additional effect of T2D on the gut microbiome of patients with NAFLD. To this end, we categorized the participants in a biopsy-proven Asian NAFLD cohort on the basis of the presence of NAFLD ± T2D. 16S rRNA amplicon sequencing analysis demonstrated that the gut microbiome includes taxa that 1) change in abundance with the progression of NAFLD alongside T2D, 2) are characteristic of the presence of T2D only, and 3) change with the progression of NAFLD but are not affected by T2D. These findings were validated using an independent T2D cohort. The study of these two independent cohorts has identified specific associations between metabolic diseases and gut bacteria, which suggest unique biomarkers for each condition.

2. Materials & methods

2.1. Study cohort

We analyzed three independent Korean cohorts, with the aim of distinguishing the gut microbiomes of NAFLD and T2D. First, the 'Boramae NAFLD cohort' was recruited by the Seoul Metropolitan Government Seoul National University Boramae Medical Center in South Korea (NCT02206841) [9,10]. Fecal samples collected from 288 participants with biopsy-proven NAFLD and 41 without NAFLD were studied. The clinical and biochemical characteristics of the participants are listed in Tables 1 and 2. This study was performed in accordance with the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Institutional Review Board of Seoul National University Boramae Medical Center (IRB No. 26–2017-48). Written informed consent was obtained from all of the study participants. Second, a validation T2D cohort was recruited from Chungnam National University Hospital (CNUH; n = 185; 36 with T2D and 149 controls). The demographics of this cohort are presented in Table 3. This study was approved by the Institutional Review Board of CNUH (IRB No. 2015–09-942–006). All the participants provided their written informed consent, and all the procedures were performed in accordance with relevant guidelines and regulations. Finally, data from the healthy Korean twin cohort (n = 752) were obtained from the European Nucleotide Archive under study accession number: ERP010289 and used for enterotyping.

2.2. Inclusion/exclusion criteria

2.2.1. The Boramae NAFLD cohort

Participants with radiologic evidence of hepatic steatosis were eligible for inclusion in the study. The subsequent inclusion/exclusion criteria were as follows.

Inclusion criteria: adult (≥18 years old), ultrasonographic findings consistent with fatty infiltration of the liver, and unexplained high serum alanine aminotransferase (ALT) activity within the preceding 6 months.

Exclusion criteria: hepatitis B or C virus infection, autoimmune hepatitis, primary biliary cholangitis or primary sclerosing cholangitis, gastrointestinal cancer or hepatocellular carcinoma, drug-induced steatosis or liver injury, Wilson disease or hemochromatosis, excessive alcohol consumption (men: >30 g/day, women: >20 g/day), antibiotic use within the preceding month, diagnosis of malignancy within the preceding 5 years, human immunodeficiency virus infection, and chronic disorder associated with lipodystrophy or immunosuppression.

Next, those with at least two of the following risk factors underwent liver biopsy: 1) T2D, 2) central obesity (waist circumference ≥ 90 cm for men or ≥ 80 cm for women), 3) high circulating concentration of triglyceride (≥150 mg/dL), 4) low circulating concentration of high-density lipoprotein (HDL)-cholesterol (<40 mg/dL for men or < 50 mg/dL for women), 5) insulin resistance, 6) hypertension, and 7) clinical suspicion of NASH or fibrosis.

Participants without NAFLD also underwent liver biopsy because they were either potential liver donors or required investigation for a liver mass. Of the 41 control participants who did not have NAFLD, 40 were liver donors for transplantation and one required a biopsy because of suspicion of a liver mass. This participant was diagnosed as having focal nodular hyperplasia (FNH) on the basis of histological examination. Participants who had been administered antibiotics within the preceding month were also excluded from the downstream analysis.

2.2.2. T2D validation cohort

Inclusion criteria: adult (≤18 years old) and a diagnosis of T2D made according to the American Diabetes Association criteria [11].

Exclusion criteria: type 1 diabetes mellitus, excessive alcohol consumption (men: >30 g/day, women: >20 g/day), antibiotic or probiotic use within the preceding month, diagnosis of malignancy within the preceding 5 years, acute illness such as infection, atherosclerotic event, bariatric surgery, and gastrointestinal tract infection (e.g., *Clostridium difficile*-associated diarrhea). In addition, the fatty liver index (FLI) [12], a well-established surrogate marker, was calculated using the following formula to exclude individuals with potential NAFLD and permit the study to focus on the T2D-specific characteristics of the cohort.

$$\mathrm{FLI} = \frac{1}{(1+\exp(-\mathrm{x}))} \times 100,$$

$$\mathrm{x} = 0.953 \times \log_e(\text{serum triglyceride, mg/dL}) + 0.139 \times (\text{BMI, kg/m}^2) + 0.718 \times \log_e(\text{serum GGT, IU/L}) + 0.053 \times (\text{waist circumference, cm}) - 15.745$$

As a result, data from 31 participants who had high FLIs (≥60) [13] were excluded from the analysis. Samples from participants who did not have T2D were collected in collaboration with the CNUH health screening center. Written informed consent was obtained from all the participants.

2.3. Liver histology

Liver histology was analyzed by a single pathologist, according to the NASH Clinical Research Network scoring system. NAFLD was diagnosed in the presence of ≥ 5% macrovesicular steatosis, and NASH was diagnosed according to the criteria of Brunt *et al.* [14,15] on the basis of the overall pattern of histological hepatic injury, involving consideration of the extents of steatosis, lobular inflammation, and ballooning [16]. The severity of fibrosis was determined according to the method of Kleiner *et al.* [16]. Significant fibrosis was defined as a score of ≥F2.

Table 1
Clinical characteristics of the NAFLD cohort, classified according to the histological spectra of NAFLD.

	No NAFLD	NAFL	NAFL + T2D	FDR[a]	FDR[b]	NASH	NASH + T2D	FDR[a]	FDR[b]
N (male/female)	41 (11/30)	114 (64/50)	45 (26/19)			73 (34/39)	56 (12/44)		
Age (years)	58.34 ± 10.53	50.26 ± 13.5	57.82 ± 13.23	*	*	49.53 ± 16.09	62.77 ± 10.33	*	*
BMI (kg/m^2)	23.43 ± 2.45	27.96 ± 3.89	26.8 ± 3.1	*		28.62 ± 3.78	27.34 ± 3.43	*	*
Waist circumference (cm)	82.63 ± 7.53	92.4 ± 8.72	89.28 ± 8.46	*		94.84 ± 10.36	93.48 ± 9.32	*	
AST (IU/L)	29.66 ± 18.53	42.09 ± 40.21	33.64 ± 19.78	*		60.67 ± 40.5	63.61 ± 53.9	*	
ALT (IU/L)	27.46 ± 24.74	53.74 ± 43.41	37.69 ± 27.08	*		91.18 ± 74.22	63.89 ± 52.36	*	*
GGT (IU/L)	49.88 ± 57.89	47.51 ± 45.13	46.55 ± 49.94			66.73 ± 51.07	97.71 ± 178.2	*	
HDL (mg/dL)	55.34 ± 12.71	47.56 ± 12.54	45.11 ± 10.72	*		47.38 ± 11.42	44.52 ± 11.26	*	
LDL (mg/dL)	98.72 ± 26.93	108.8 ± 36	100.8 ± 29.59			118.6 ± 27.48	91.15 ± 30.68	*	*
Albumin (g/dL)	4.076 ± 0.29	4.202 ± 0.27	4.164 ± 0.36			4.211 ± 0.32	4.043 ± 0.312		
Platelet count (×10^3/L)	232.3 ± 63.39	252 ± 61.88	240.2 ± 57.8			237.5 ± 59.15	196.3 ± 76.6		
Ferritin (ng/mL)	136.5 ± 260.1	174.1 ± 184.9	133.8 ± 158.9	*		206.6 ± 150.2	217.4 ± 26	*	
HA (ng/mL)	61.96 ± 70.69	39.74 ± 34.31	66.9 ± 90.71			71.81 ± 98.28	117.8 ± 96.55		*
Insulin (μIU/mL)	9.529 ± 3.97	17.09 ± 13.69	13.07 ± 6.2	*		19.17 ± 11.77	16.78 ± 7.75	*	
HbA1c (%)	5.588 ± 0.31	5.696 ± 0.48	7.051 ± 1.06		*	5.928 ± 0.47	7.47 ± 1.48	*	*
C-peptide (ng/mL)	2.035 ± 0.75	3.863 ± 2.29	3.923 ± 4.3	*		4.789 ± 3.74	3.759 ± 1.78	*	
HOMA-IR	2.432 ± 1.12	4.395 ± 3.49	4.392 ± 2.38			5.26 ± 3.63	6.817 ± 4.69		
Adipo-IR	5.497 ± 3.79	10.6 ± 8.88	8.874 ± 4.807			12.23 ± 8.15	13.3 ± 8.81		
FFA (μEq/L)	552.4 ± 235.7	635.2 ± 272.7	682.2 ± 212.5			635 ± 223.8	777.8 ± 276.7	*	*
hsCRP (mg/dL)	0.1685 ± 0.31	0.2002 ± 0.4	0.1529 ± 0.16			0.2695 ± 0.29	0.3184 ± 0.44		
Cholesterol (mg/dL)	176 ± 34.24	188.9 ± 35.97	179.8 ± 46.27			189.9 ± 32.89	161.7 ± 42.82	*	*
TG (mg/dL)	93.2 ± 37.33	157 ± 79.87	166 ± 107.8	*		142.8 ± 50.69	149.1 ± 73.46	*	
FBS (mg/dL)	103.2 ± 20.25	104.4 ± 18.67	134.1 ± 34.57		*	108.9 ± 25.7	159.1 ± 68.52		*
HTN, n (%)	36.59	31.58	48.89			39.73	67.86		*

FDR[a]: no NAFLD *vs.* NASH or NAFL. FDR[b]: no NAFLD *vs.* NASH + T2D or NAFL + T2D. *, Wilcoxon rank sum test (FDR < 0.1).
BMI: body mass index, AST: aspartate aminotransferase, ALT: alanine aminotransferase, GGT: gamma-glutamyl transferase, HDL: high-density lipoprotein, LDL: low-density lipoprotein, HA: hyaluronic acid, HbA1c: hemoglobin A1c (glycated haemoglobin), HOMA-IR: Homeostatic Model Assessment of Insulin Resistance, Adipo-IR: Adipose tissue Insulin Resistance index, FFA: free fatty acid, hsCRP: high-sensitivity C-reactive protein, TG: triglyceride, FBS: fasting blood sugar, HTN: hypertension.

2.4. DNA extraction, 16S rRNA sequencing, and data processing

For the NAFLD and validation T2D cohorts, fecal DNA extraction and microbial profiling were performed as described previously [8]. Briefly, DNA was extracted from 200 mg aliquots of fecal samples using the QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany). Amplicons targeting the V4 region of 16S rRNA were sequenced using the MiSeq platform (2 × 300 reads; Illumina, San Diego, CA, USA) and processed using the DADA2 pipeline (v.1.16.0) [17]. The filtering and trimming parameters were as follows: truncQ = 2, trimLeft = c(10,10), and trunclen = c(230,140). Filtered reads were denoised, merged, and further processed to remove chimeras. Taxonomic classification of the amplicon sequence variants (ASVs) was performed using the Ribosomal

Table 2
Clinical characteristics of the NAFLD cohort, classified according to the severity of hepatic fibrosis.

	Fibrosis = 0			Fibrosis = 1			Fibrosis ≥ 2		
	No T2D	T2D	FDR	No T2D	T2D	FDR	No T2D	T2D	FDR
N (male/female)	60 (27/33)	10 (9/1)		124 (66/58)	46 (19/27)		44 (16/28)	45 (10/35)	
Age (years)	56.3 ± 10.76	61.4 ± 10.91		47.95 ± 14.54	55.77 ± 15.01	*	54.86 ± 15.08	64.88 ± 7.45	*
BMI (kg/m^2)	25.11 ± 3.72	26.75 ± 2.94		29.19 ± 3.5	27.37 ± 3.24		27.85 ± 3.4	26.9 ± 3.45	
Waist circumference (cm)	86.43 ± 10.98	89.36 ± 8.33		93.42 ± 8.73	90.95 ± 8.74		92.25 ± 10.3	92.88 ± 9.76	
AST (IU/L)	30.2 ± 18.69	23.8 ± 4.34		45.94 ± 26.96	48.52 ± 55.59		66.68 ± 68.05	57.92 ± 33.84	
ALT (IU/L)	33.98 ± 35.15	30.1 ± 10.99		68.13 ± 53.73	53.07 ± 51.38		77.75 ± 78.1	56.27 ± 41.18	
GGT (IU/L)	39.03 ± 44.35	53.56 ± 75.91		59.45 ± 53.01	46.54 ± 39.9		59.5 ± 45.55	108.8 ± 197	
HDL (mg/dL)	50.9 ± 13.19	42.8 ± 8.3		48.24 ± 12.89	45.02 ± 11.37		48.05 ± 10.4	44.98 ± 11.24	
LDL (mg/dL)	103.5 ± 33.3	103.1 ± 33.2		112.5 ± 33.23	102.3 ± 28.21		114.2 ± 25.33	87.24 ± 30.48	*
Albumin (g/dL)	4.117 ± 0.28	4.26 ± 0.2		4.241 ± 0.29	4.152 ± 0.36		4.105 ± 0.28	4.004 ± 0.32	
Platelet count (×10^3/L)	241.1 ± 51.27	192.5 ± 47.27		253.9 ± 64.69	249.1 ± 60.54		219.2 ± 59.43	187.1 ± 73.78	
Ferritin (ng/mL)	122.8 ± 82.53	119.5 ± 68.44		178.9 ± 185.6	210.5 ± 295.9		250.6 ± 278.8	163.1 ± 137.2	*
HA (ng/mL)	44.27 ± 52.06	54.16 ± 42.53		47.15 ± 59.43	70.15 ± 88.92		86.45 ± 98.88	129.9 ± 102.6	*
Insulin (μIU/mL)	12.54 ± 8.781	12.04 ± 4.95		17.5 ± 13.69	14.27 ± 7.27		18.55 ± 11.27	16.69 ± 7.54	
HbA1c (%)	5.708 ± 0.46	6.94 ± 0.67	*	5.715 ± 0.44	7.085 ± 1.09	*	5.914 ± 0.52	7.562 ± 1.58	*
C-peptide (ng/mL)	2.663 ± 1.29	3.044 ± 1.3		4.203 ± 3.29	4.224 ± 4.36		4.339 ± 2.46	3.606 ± 1.53	
HOMA-IR	3.178 ± 2.46	4.524 ± 2		4.551 ± 3.59	4.858 ± 3.45		5.22 ± 3.49	6.903 ± 4.59	
Adipo-IR	7.19 ± 5.37	8.578 ± 5.23		10.99 ± 9.07	10.16 ± 6.63		12.1 ± 8.22	13.12 ± 8.62	
FFA (μEq/L)	569.1 ± 248.3	690.8 ± 258.9		642.6 ± 259	706.8 ± 235.8		627 ± 233.6	774 ± 269.3	*
hsCRP (mg/dL)	0.143 ± 0.23	0.081 ± 0.07		0.2234 ± 0.35	0.1983 ± 0.21		0.3005 ± 0.46	0.3284 ± 0.47	
Cholesterol (mg/dL)	181.1 ± 38.34	178 ± 29.88		190.5 ± 34.38	180.7 ± 46.24		184.7 ± 30.89	156.7 ± 44	*
TG (mg/dL)	123.9 ± 65.82	169 ± 65.17		150 ± 74	166.1 ± 106.1		138.7 ± 54.31	144.2 ± 76.53	
FBS (mg/dL)	102.3 ± 19.38	156.7 ± 45.03	*	104.8 ± 18.43	130.1 ± 34.34	*	112.4 ± 29.65	164.2 ± 71.93	*
HTN, n (%)	33.33	50		33.87	54.35	*	40.9	66.67	*

*, Wilcoxon rank sum test (FDR < 0.1).
BMI: body mass index, AST: aspartate aminotransferase, ALT: alanine aminotransferase, GGT: gamma-glutamyl transferase, HDL: high-density lipoprotein, LDL: low-density lipoprotein, HA: hyaluronic acid, HbA1c: hemoglobin A1c (glycated haemoglobin), HOMA-IR: Homeostatic Model Assessment of Insulin Resistance, Adipo-IR: Adipose tissue Insulin Resistance index, FFA: free fatty acid, hsCRP: high-sensitivity C-reactive protein, TG: triglyceride, FBS: fasting blood sugar, HTN: hypertension.

Table 3
Clinical characteristics of the validation T2D cohort.

	No T2D	T2D	FDR
N (male/female)	149 (55/94)	36 (15/21)	
Age (years)	65.52 ± 11.54	72.61 ± 9.02	*
BMI (kg/m^2)	23.26 ± 2.61	23.98 ± 3.4	*
Waist (cm)	80.97 ± 9.69	85.19 ± 10.07	*
AST (IU/L)	24.26 ± 5.6	28.11 ± 12.4	
ALT (IU/L)	19.47 ± 8.01	23.25 ± 12.21	
GGT (IU/L)	23.88 ± 15.69	31.5 ± 33.42	
HDL (mg/dL)	53.3 ± 11.08	45.92 ± 12.3	*
LDL (mg/dL)	125.6 ± 32.41	98.69 ± 27.74	*
Cholesterol (mg/dL)	206 ± 38.56	174.6 ± 37.9	*
TG (mg/dL)	132.8 ± 61.43	152.9 ± 67.43	*
FBS (mg/dL)	91.1 ± 5.54	131.2 ± 24.02	*
Fatty Liver Index	22.18 ± 14.28	32.92 ± 17.84	*

*, Wilcoxon rank sum test (FDR < 0.1).
BMI: body mass index, AST: aspartate aminotransferase, ALT: alanine aminotransferase, GGT: gamma-glutamyl transferase, HDL: high-density lipoprotein, LDL: low-density lipoprotein, TG: triglyceride, FBS: fasting blood sugar.

Database Project (RDP) classifier (RDP trainset 16/release 11.5). For the Korean twin cohort [18], single-ended fastq sequences were processed using the same DADA2 pipeline, except that the following filtering and trimming parameters were used: truncQ = 11, trimLeft = c(20), and truncLen = c(200). Data from four participants were excluded because of low sequencing depth (<10,000 reads). Prior to the downstream analysis (excluding multivariate taxonomic association analysis), bacterial abundances were transformed using the centered log-ratio (CLR) transformation of Aitchison to control the composition of the sequencing data (CoDaSeq R package function codaSeq.clr [19]). Zero counts in the AVS table were replaced by a Geometric Bayesian multiplicative approach using the zCompositions package function cmultRepl [20].

2.5. Identification of overrepresented bacterial metabolic pathways in the NAFLD cohort

The metabolic pathways that were overrepresented in the microbial community in the NAFLD cohort were identified using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States 2 (PICRUSt2, v.2.3.0_b) and the default parameters [21]. Up- and downregulated metabolic genes and pathways were identified by referring to the MetaCyc database (https://metacyc.org/). Associations of NAFLD progression with T2D were analyzed using the Multivariate Association with the Linear Models 2 (MaAsLin2) package in R (v.1.2.0), after adjustment for age and sex [22]. Regression coefficients were used for heatmap analysis.

2.6. Explanatory power of cohort variables

The effect sizes of cohort covariates on variation in the microbial community were evaluated as described previously [23]. Briefly, distance-based RDA (db-RDA) was performed at the ASV and genus levels using Euclidean distance, as implemented in vegan [24]. Covariates (false discovery rate [FDR] < 0.1) identified during this step were entered into forward stepwise model selection to measure their cumulative effect sizes. Prior to this analysis, the collinearity of the variables was checked using phiK and Spearman's rank correlation.

2.7. Statistical analysis

Statistical analysis of the microbiome was performed in R (v.4.0.5) [25]using the vegan [24], phyloseq [26], pairwiseAdonis [2728], CoDaSeq, DirichletMultinomial [29], fifer [30], MaAsLin2 [22], and ppcor [31] packages. Enterotyping (or community typing) was performed using the Dirichlet Multinomial Mixtures (DMM) approach and a genus-abundance RMP matrix in R, as described by Holmes *et al.* [32]. The diversity and composition of the bacterial communities were evaluated using α- (Observed richness, Shannon, and Pielou's indices) and β- (principal components analysis (PCA), based on Euclidean distance) diversity at the ASV and genus levels and vegan [24]. The composition of the bacterial communities were compared between the groups using permutational multivariate analysis of variance (PERMANOVA) pairwise comparisons (pairwise Adonis test) [27]. Multivariate analysis of the associations between metabolic diseases and the abundances of bacterial taxa was performed using the MaAsLin2 package in R [22]. Comparisons between two groups were made using Wilcoxon's rank sum test, and the Kruskal–Wallis test, followed by a *post-hoc* Dunn's test, was used for analysis of more than two groups. Categorical data were analyzed using the chi-square test. Associations of taxa with host parameters were identified by a partial correlation, to adjust for confounders, using the R package ppcor [31]. Statistical testing of more than two features included correction for multiple testing using the Benjamini–Hochberg method, and the results are reported with FDRs. All the statistical tests used were two-sided. The results were visualized using R or GraphPad Prism 9 (GraphPad Software, San Diego, CA, USA).

3. Results

3.1. Description of the cohort

We used a biopsy-proven NAFLD cohort (n = 329) to distinguish the microbial signature of T2D from that of the progression of NAFLD. The participants were categorized into subgroups using their histological characteristics (the NASH-CRN histological scoring system), the severity of fibrosis, and the presence or absence of T2D. First, we compared the anthropometric and biochemical characteristics of control participants (no NAFLD or no fibrosis), those with liver disease alone (NAFL, NASH, or fibrosis), and those with liver disease and T2D (Tables 1 and 2). The participants with both liver disease and T2D had significantly higher serum hyaluronic acid (HA) concentration, HbA1c, free fatty acid (FFA) concentrations, fasting blood sugar (FBS), and blood pressure than the liver disease only group (Wilcoxon rank sum test, FDR < 0.1). By contrast, the participants with liver disease and T2D had significantly lower ALT activity, low-density lipoprotein (LDL)-cholesterol concentration, and total cholesterol concentration than those with NAFLD only (Wilcoxon rank sum test, FDR < 0.1).

3.2. Variations in the gut microbial community in the NAFLD cohort

In the NAFLD cohort, the variation in the gut microbial community could mostly be explained by age at both the genus and ASV levels (db-RDA, adjusted R^2 [0.51:0.8%], FDR < 0.1; Fig. 1A and Table S1). T2D in combination with fibrosis and NAFLD (NAFL or NASH) was associated with the second highest explanatory power for this community variation (db-RDA, adjusted R^2 [0.35:0.56%], FDR < 0.1). Other significant covariates included circulating parameters that are linked to the liver conditions and T2D; for example, C-peptide, platelet count, FFA, and triiodothyronine (db-RDA, adjusted R^2 [0.14:0.26%], FDR < 0.1). PCA based on the Aitchison transformation confirmed that the variations were related to the metabolic diseases; however, although T2D could be distinguished from NAFLD, it could not be distinguished from fibrosis in this way (pairwise Adonis test, FDR < 0.1; Fig. 1B and C, Fig. S1B, and Table S2). For example, a PERMANOVA test did not differentiate fibrosis from fibrosis combined with T2D (Table S2A). Biodiversity

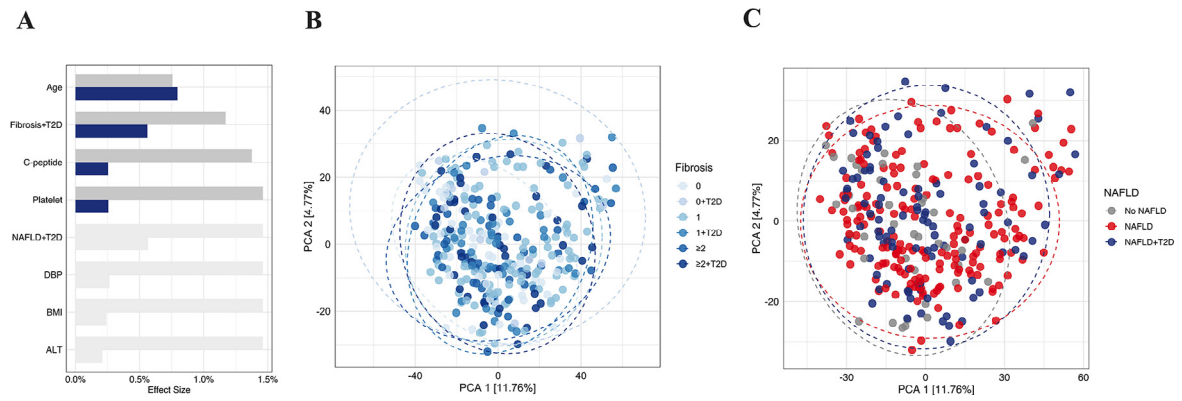

Fig. 1. Variations in the microbial community in the NAFLD-T2D cohort. (A) Explanatory power of the characteristics of the cohort for the variations in the microbial community at the genus level. Blue and dark gray bars indicate the individual and cumulative effect sizes. Light gray bars indicate variables that were significant on an individual basis but were not included in the forward stepwise RDA model. Principal component analysis (PCA) was performed, based on Aitchison transformation, for (B) fibrosis and (C) NAFLD. n = 329; FDR < 0.1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

analysis of the NAFLD cohort using NASH-CRN category and the severity of fibrosis demonstrated a significant decrease in the richness of the community with an increase in the severity of NAFLD in the presence of T2D at the ASV level (Fig. S2A). The same analysis conducted at the genus level also revealed reductions in bacterial diversity and evenness (Fig. S2B). Together, these results suggest an independent effect of T2D on the gut microbiome signature for NAFLD.

3.3. Bacterial signature of T2D in the NAFLD cohort

Multivariate taxonomic association analysis for NAFLD and T2D status identified three distinct bacterial signatures after adjustment for age (FDR < 0.1; Fig. 2 and Tables S3–4): 1) a signature characteristic of T2D during the progression of NAFLD or fibrosis (Fig. 2A and B), 2) a signature characteristic of the presence of T2D in patients with fibrosis (Fig. 2C), and 3) a signature for NAFLD alone (Fig. 2D). The relative abundances of *Enterobacter, Romboutsia,* and *Clostridium sensu stricto* were significantly associated with the progression of NAFLD (NASH and fibrosis) in combination with T2D (pattern 1; Fig. 2A and B). Further investigation of the associations of the gut microbial composition with T2D showed that the abundances of *Enterobacter, Clostridium sensu stricto,* and *Lachnospiraceae* were significantly affected only by T2D in combination with liver disease (pattern 2; Fig. 2C). Following pattern 3, *Megamonas* and *Ruminococcus* were associated with NAFLD alone, with no T2D. These results imply that different gut microbes are independently associated with NAFLD and T2D.

Next, we investigated how the characteristics of the host interact with the abundances of the bacterial taxa that were significantly associated with metabolic disease (Fig. 3). Analysis of the host–microbe interactions identified specific associations of parameters relating to diabetes, such as FBS and HbA1c, with the abundances of *Romboutsia, Clostridium sensu stricto,* and *Escherichia/Shigella* (Spearman's correlation, adjusted for age, FDR < 0.1). In addition, we found a strong positive association between serum LDL-cholesterol concentration and *Faecalibacterium*. This suggests that the differences in the abundances of these bacterial taxa are more closely related to T2D than NAFLD.

3.4. Associations of microbial functional potentials with T2D and NAFLD

To better understand the links between metabolic diseases and microbial function, we characterized the microbial metabolic pathways that were associated with each disease using PICRUSt2 (Fig. 4). Of the 394 metabolic pathways inferred, 36 were significantly upregulated in the NASH or significant fibrosis (≥F2) group (multivariate regression analysis, FDR < 0.2). Of these, the largest number of significant pathways (6 out of 31 for participants with NAFLD and 11 out of 25 for participants with fibrosis) are involved in the degradation of aromatic compounds, which implies that they increase nutrient or energy supply for the growth of specific bacteria with increasing disease severity (Fig. 4A–B) [33]. Moreover, participants with both NAFLD and T2D demonstrated the upregulation of pathways involved in cofactor, carrier, and vitamin biosynthesis (10 out of 31 pathways), which implies greater biosynthesis of small molecules that are involved in enzymatic reactions. We found that ECASYN-PWY (the enterobacterial common antigen biosynthesis pathway) and KDO-NAGLIPASYN-PWY (the superpathway for (Kdo)2-lipid A biosynthesis) were positively associated with the severity of NAFLD and T2D, after adjustment for age (Fig. 4C–D). These results suggest that individuals with both NAFLD and T2D may have greater flux through deleterious microbial metabolic pathways than those with NAFLD alone.

3.5. Effect of antidiabetic medications in the NAFLD cohort

Given that antidiabetic medications may affect the relationship between T2D and the gut microbiome [34], we next assessed the effect of taking antidiabetic medications on host health status (Table S5). Participants who had been treated with antidiabetic medications had significantly lower LDL-cholesterol, ferritin, and cholesterol concentrations, but a higher FBS concentration than those who had not (Tables S6–7). However, a taxonomic association analysis of participants that were taking antidiabetic medications generated the same associations of bacterial taxa with T2D, which implies that such medications did not have effects on the gut microbiome in the present study (Fig. S3, Tables S8–9).

3.6. Enterotypes in the Korean population cohorts

To determine whether an enterotype was present in the selected Korean cohort and whether a specific enterotype is linked to the metabolic disorder, we used Dirichlet Multinomial Mixtures (DMM)-modeling based clustering, which was previously used to identify an inflammatory enterotype in Western population cohorts [35–37]. The DMM community typing of a combination of the NAFLD cohort (n = 335) and the healthy Korean twin cohort (n = 751) identified three optimum clusters on the basis of the minimum Bayesian information criterion (Fig. 5A, S4A). The three community types were labeled on the basis of their bacterial

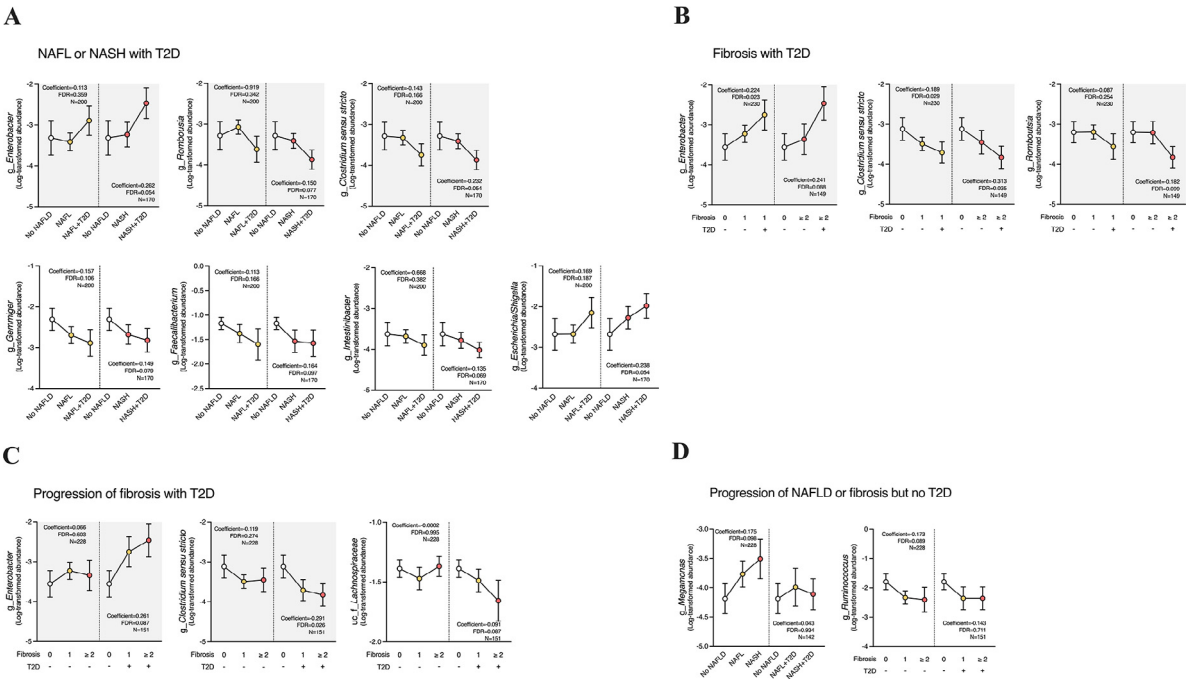

Fig. 2. Comparisons of the abundances of microbial taxa in the NAFLD cohort, categorized according to either the histology of liver biopsies from participants with NAFLD and T2D or the severity of hepatic fibrosis in participants with T2D. The genera that were significantly associated with NAFLD are depicted as mean connected plots. Significant associations were found for (A) NAFL or NASH with T2D, (B) fibrosis with T2D only, (C) the progression of fibrosis with T2D, and (D) the progression of NAFLD or fibrosis but no T2D. Features with the background color indicate the overlapping features between the groups. Associations were identified using the multivariate association analysis, after adjustment for age, and the corrected significance levels were calculated using the Benjamini-Hochberg method (FDR < 0.2). n = 329.

composition, community richness, and host BMI (Fig. S4B–D). In the merged cohort dataset, the *Bacteroides* (B) and *Ruminococcus* (R) enterotypes included *Bacteroides* at similar abundances, but the B enterotype featured a greater abundance of *Escherichia/Shigella*, lower richness, and higher BMI than the R enterotype (Kruskal–Wallis test, FDR < 0.1). Comparison of the enterotypes between the healthy Korean twin and NAFLD cohorts showed that the latter cohort had a significantly higher proportion of the B enterotype and a lower proportion of the R enterotype (chi-square test, FDR < 0.1; Fig. 5A). The greater proportion of the

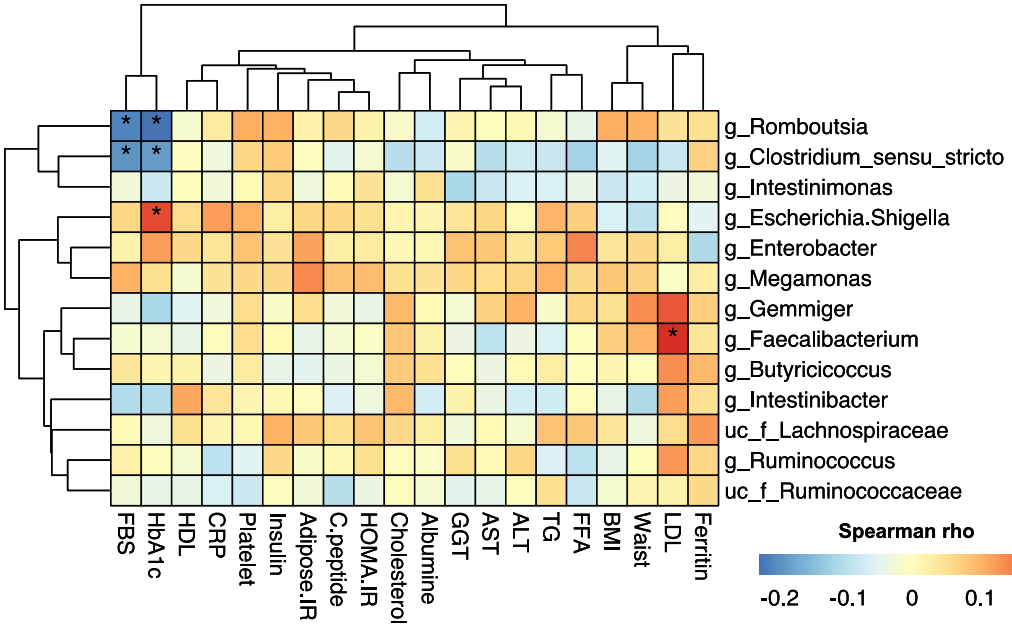

Fig. 3. Relationships between the abundances of disease-associated taxa and host parameters. Correlation coefficients (adjusted for age) for the relationships between the relative abundances of 10 disease-associated taxa and 20 host parameters were calculated using Spearman's correlation method. The color of the plot denotes the magnitude of Spearman's rho (n = 329, * FDR < 0.1). Abbreviations: FBS, fasting blood sugar; HbA1c, glycosylated hemoglobin; HDL, high-density lipoprotein; CRP, high-sensitivity C-reactive protein; Adipose.IR, adipose tissue insulin resistance; C.peptide, connecting peptide; HOMA.IR, homeostasis model assessment of insulin resistance; GGT, gamma-glutamyl transferase; AST, aspartate aminotransferase; ALT, alanine aminotransferase; TG, triglycerides; FFA, free fatty acid; BMI, body mass index; Waist, waist circumference; LDL, low-density lipoprotein.

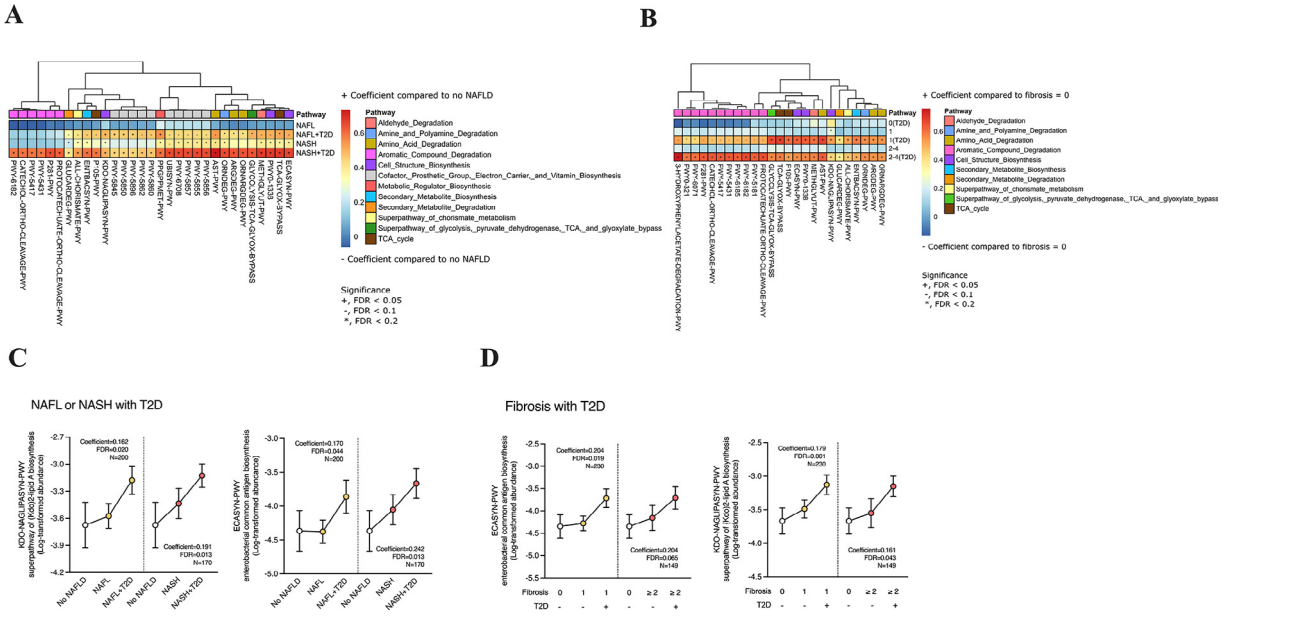

Fig. 4. Inferred microbial metabolic pathways in participants with NAFLD, categorized according to either the histology of a liver biopsy from participants with NAFLD and T2D or the severity of fibrosis in participants with T2D. The activity of metabolic pathways in the microbial community was predicted using PICRUSt2 and annotated using the MetaCyc database. (A) Thirty-one metabolic pathways were significantly upregulated in participants with NASH ± T2D *vs.* those without NAFLD, as shown in a heatmap. (B) Twenty-five metabolic pathways were upregulated in participants with a fibrosis score ≥ 2 ± T2D. The color of the plots denotes the magnitudes of the coefficients. The associations of T2D with two metabolic pathways related to lipopolysaccharide synthesis are also shown according to (C) the liver histology of participants with NAFLD and (D) the severity of hepatic fibrosis. The data were analyzed using the multivariate association analysis method, after adjustment for age, and FDR was calculated using the Benjamini-Hochberg method. n = 329; +, FDR < 0.05; –, FDR < 0.10; *FDR < 0.20.

B enterotype was associated with the presence of NAFLD, but there was no significant contribution of T2D (Fig. 5B–C). Furthermore, there was a trend for the B enterotype to become more abundant with an increase in the severity of fibrosis (chi-square test, p = 0.065; Fig. S4E), but T2D was not significantly associated with the B enterotype (chi-square test, p = 0.24).

3.7. Bacterial taxa associated with T2D in a validation T2D cohort

To validate our findings regarding the bacterial signature of T2D, we analyzed a separate cohort of patients with T2D that were recruited at CNUH (n = 185). Because T2D is frequently accompanied by NAFLD, we excluded patients with a high FLI (≥60) from the present analysis because such individuals are likely to have NAFLD (NAFL or NASH) [38]. We confirmed that there were no significant differences in the serum ALT, aspartate aminotransferase, or γ-glutamyl transferase activities between participants who did or did not have T2D (Table 3). The variations in the gut microbial communities of this validation cohort could be explained by height, the presence of T2D, alcohol consumption, and age (db-RDA, adjusted R^2 0.6 [0.71%], FDR < 0.1; Fig. 6A and Table S10). The fact that variation could be explained by height implied a sex effect in the cohort; therefore, we included this as a potential confounder in the downstream analysis. The composition of the microbial community significantly differed between the controls and participants with T2D (Adonis R^2 = 0.01, p = 0.01; Fig. 6B); however, the biodiversity of the microbiota did not significantly differentiate the groups (Fig. S5A). Enterotype analysis confirmed the absence of an association between a specific enterotype and T2D (Fig. S5B). After adjustment for age, sex, and FLI, we found that *Escherichia/Shigella* (+) and *Clostridium sensu stricto* (–) were significantly associated with T2D (multivariate regression analysis, FDR < 0.1; Fig. 6C and Table S11). By means of inferred functional analysis, we found that two pathways related to LPS biosynthesis were also more abundant in the participants with T2D (Fig. 6D and 6E), as shown in the NAFLD cohort in the present study.

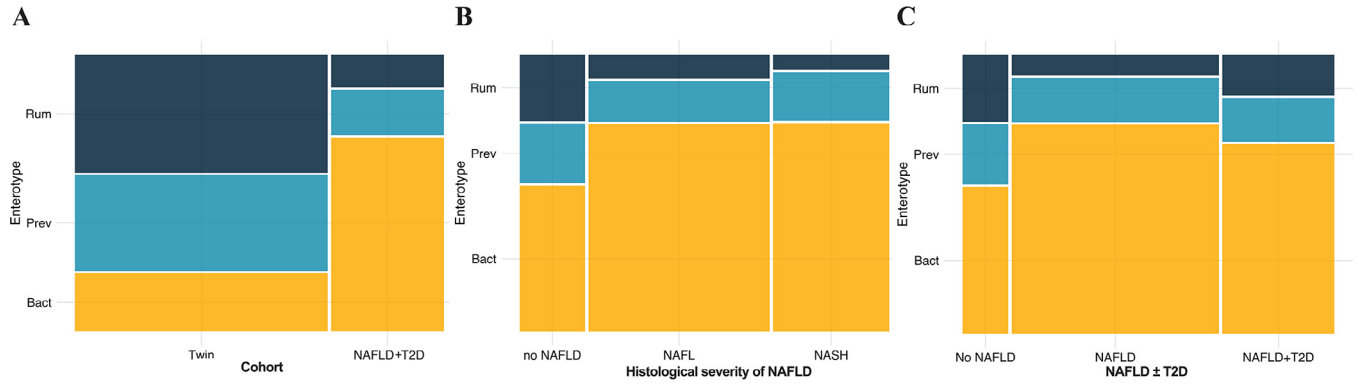

Fig. 5. Enterotypes of the Korean population cohorts. (A) Proportions of each enterotype in the healthy Korean twin (n = 752) and NAFLD-T2D (n = 329) cohorts. (B) Proportions of each enterotype in the NAFLD-T2D cohort, categorized according to the presence of NAFLD alone (n = 228) and (C) NAFLD plus T2D (n = 329). FDR < 0.1.

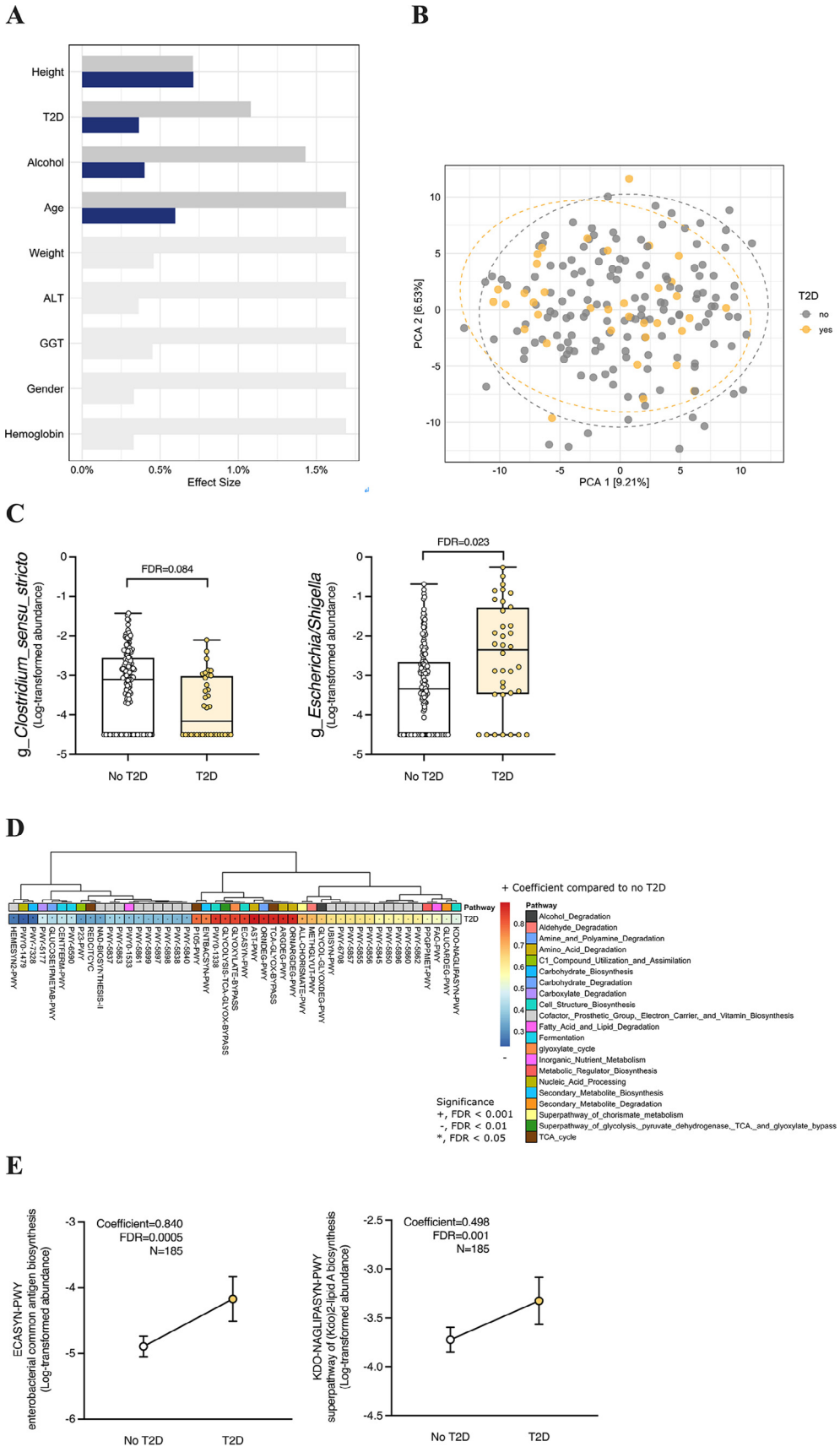

Fig. 6. Validation of the identified NAFLD-T2D-associated bacterial signatures using an independent T2D cohort. (A) The explanatory power of the cohort variables for the variations in the microbial communities at the genus level. Blue and dark gray bars indicate individual and cumulative effect sizes, respectively. Light gray bars indicate variables that were significant on an individual basis but were not included in the forward stepwise RDA model. (B) Principal components analysis (PCA), based on Aitchison transformation, for T2D. (C) Differences in the abundance of bacterial taxa in participants with and without T2D in the validation cohort. (D) Forty-seven inferred metabolic pathways that were significantly upregulated in participants with T2D. (E) Comparisons of the pathways related to LPS biosynthesis. n = 185; FDR < 0.1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

NAFLD is a well-established risk factor for T2D, and *vice versa* [3,4]. We recently identified gut microbiome signatures and metabolites that are associated with the histological severity of NAFLD, but some of these microbial signatures were lost after adjustment for T2D [8]. Therefore, we hypothesized that the two metabolic disorders may be independently linked with the gut microbiome. Indeed, multivariate taxonomic association analysis identified several taxa that followed the distinct patterns linked with 1) disease progression (*Enterobacter, Romboutsia,* and *Clostridium sensu stricto*), 2) the presence/absence of T2D alongside NAFLD (*Enterobacter, Clostridium sensu stricto,* and Lachnospiraceae), and 3) NAFLD only (*Megamonas* and *Ruminococcus*). *Enterobacter*, which belongs to the Enterobacteriaceae family, has previously been reported to be linked to NAFLD [8,39–41]. However, we previously showed that this association disappears when the data are adjusted for the presence of T2D [8]. Using the present, larger NAFLD cohort, and the categorization of the participants according to the presence or absence of T2D, we have confirmed the association of particular bacterial taxa with T2D, rather than NAFLD. In a previous study, *Romboutsia* was found to be overrepresented in lean NAFLD patients and to be closely associated with hepatic triglyceride levels [42]. *Clostridium* was previously shown to be less abundant in individuals with either NAFLD [43] or T2D [44]; however, we found significantly fewer *Clostridium* in patients with fibrosis and T2D, but not if T2D was absent. The link between NAFLD and T2D and the abundance of *Lachnospiraceae* has been previously studied, but inconsistent results have been obtained. Significant associations of this taxon with these diseases have been identified, but both greater and lesser abundance have been reported [43,45–47]. *Megamonas* and *Ruminococcus* have been previously shown to be associated with severe NAFLD (NASH or significant fibrosis) [7,8,48,49], and we found that it was only significantly less abundant in patients with fibrosis but no T2D.

Analysis of host–microbe interactions showed that *Romboutsia* and *Clostridium sensu stricto* were linked with markers of diabetes specifically (FBS and HbA1c). Although we found that these two taxa were associated with NAFLD progression and T2D, this suggests that they are more closely associated with T2D than NAFLD. Interestingly, we also found a strong positive association between serum LDL-cholesterol concentration, which is a risk factor for both cardiovascular disease and T2D [50], and *Faecalibacterium,* which is a well-known gut commensal. There was a significantly lower LDL-cholesterol concentration and a lower abundance of *Faecalibacterium* in participants with both NAFLD and T2D. This is consistent with the results of a study of individuals with severe fibrosis, in whom the hepatic inflammation reduced the production of very-low-density lipoprotein (VLDL), resulting in a reduction in LDL-cholesterol concentration [51–53]. The results of another previous study suggested that a medication (e.g., statin)-induced reduction in LDL-cholesterol concentration or genetic variants might increase the risk of T2D, but the mechanism involved has not been identified [54]. Thus, the unexpected positive correlation between the abundance of *Faecalibacterium* and LDL-cholesterol concentration might be attributable to the lower LDL-cholesterol concentration that characterized participants with severe NAFLD and T2D, in comparison with participants with less severe disease.

Inferred functional analysis revealed that metabolic pathways involved in the degradation of aromatic compounds were significantly more abundant in the NASH + T2D and significant fibrosis + T2D groups. Aromatic compounds (i.e., phenyl-alanine, tyrosine, and tryptophan) are converted to acetyl-CoA, succinyl-CoA, and pyruvate *via* peripheral and central pathways. These converted aromatic compounds can then be utilized as nutrient or energy sources to support microbial growth [33]. Enzymes involved in the degradation of aromatic compounds were mainly encoded by members of Enterobacteriaceae, which suggests that these bacteria may benefit from the gut environment present in patients with NAFLD and T2D. [55]. Furthermore, specific pathways involved in the biosynthesis of LPS, such as ECASYN-PWY (the enterobacterial common antigen biosynthesis pathway) and KDO-NAGLIPASYN-PWY (the superpathway for (Kdo)2-lipid A biosynthesis) were more abundant in the participants with metabolic disease. Given that LPS influx has been proposed to mechanistically link gut dysbiosis with NAFLD in humans and animal models [56,57], the upregulation of these pathways might implicate the microbiome as a risk factor in the pathogenesis of NAFLD and T2D.

The impact of medications on the gut microbiome has been thoroughly studied [58,59]. In particular, metformin treatment has been reported to affect the abundance of specific gut bacteria, such as *Akkermansia muciniphila, Escherichia,* and *Intestinibacter* [60]. However, in the present study, we found that the same set of bacteria were associated with NAFLD and T2D, regardless of the use of antidiabetic medications. Therefore, further studies are warranted to verify whether these types of bacteria are less affected by antidiabetic medications or whether they might not have been effective in the host. In addition to medication, diet also has a substantial impact on the gut microbiome, which in turn can affect the host immune system and metabolic parameters [61]. In particular, the lack of specific pharmacotherapy for NAFLD means that changes in lifestyle, including to the diet, remain the best strategy for the prevention and treatment of the disease [62]. Further research on the effects of diet on NAFLD is necessary to improve the management of the disease.

In recent studies, participants have been classified according to their gut microbial composition, and in this way an inflammatory enterotype has been identified that is associated with multiple diseases (depression, primary sclerosing cholangitis, and inflammatory bowel disease) [63–65]. To determine whether such an inflammatory enterotype was present in the selected Korean cohort, we enterotyped both the study cohort and the validation cohort, and found that the B enterotype was more abundant in the NAFLD cohort than in the healthy Korean twin cohort. However, the resolution of the enterotyping approach was not high enough to distinguish T2D from NAFLD or either disease from individuals who did not have T2D in the validation cohort. Further subtyping of the inflammatory enterotype might help clarify the relationships with each disease and permit the identification of patients who would respond most effectively to particular therapeutics.

Some limitations of the present study include 1) the study design: this cross-sectional study only captures a snapshot of the microbial relationships with the host. 2) The differing definitions of NAFLD that were used for the NAFLD (i.e., biopsy) and T2D (i.e., FLI) cohorts. In a previous study of a large Finnish cohort (n = 6,269), a NAFLD-specific microbial signature was successfully identified by categorizing the participants using this FLI cut-off value [13]. However, the use of this index to define NAFLD may not be as accurate as a diagnosis made using a biopsy. 3) The absence of patients with only a single disease in the NAFLD cohort. Given that T2D always accompanied NAFLD in the present cohort, we were unable to analyze the independent effect of each disease. Despite these limitations, we were able to distinguish the microbial signature of T2D from that of the progression of NAFLD using an additional validation cohort. Further studies with longitudinal cohorts with shotgun metagenomic analysis are warranted to confirm the functional perspective of the gut microbiome with the disease.

5. Conclusions

In conclusion, we have been able to distinguish T2D-specific and NAFLD-specific bacterial signatures. Recent studies of NAFLD have shown that there are distinct alterations to the gut microbiota of individuals with NAFLD, but most of these studies did not consider the presence of T2D as a confounding factor. In the present study, we have identified bacterial taxa that more strongly associated with T2D than NAFLD in a NAFLD-T2D cohort by 1) categorizing the cohort, 2) showing correlations with host parameters that are specific for diabetes and liver disease, and 3) validating the results in an independent T2D cohort. These results should help facilitate further research aimed at improving the diagnosis, prevention, and treatment of these metabolic diseases. In the future, longitudinal studies should be performed to evaluate whether the presence or exacerbation of these bacterial signatures may predict the development or progression of these metabolic diseases.

CRediT authorship contribution statement

Jiyeon Si: Formal analysis, Writing – original draft, Funding acquisition. **Giljae Lee:** Formal analysis, Writing – original draft. **Hyun Ju You:** Funding acquisition, Writing – review & editing. **Sae Kyung Joo:** Data curation, Project administration. **Dong Hyeon Lee:** Data curation, Project administration. **Bon Jeong Ku:** Resources, Funding acquisition. **Seoyeon Park:** Formal analysis. **Won Kim:** Conceptualization, Resources, Writing – review & editing, Funding acquisition. **GwangPyo Ko:** Conceptualization, Funding acquisition, Supervision.

Declaration of Competing Interest

GK is a founder of KoBioLabs, Inc., a company that aims to characterize the role of host–microbiome interactions in chronic diseases. The other authors declare no competing interests.

Acknowledgements

We thank Bioedit (https://www.bioedit.com) for editing drafts of this manuscript. The graphical abstract was created using BioRender.com.

Author Contributions

GL, JS, WK, GBJ, HJY, and GK conceived and designed the study. WK, GBJ, JSK, and DHL coordinated patient recruitment and sample collection. GL and JS conducted bioinformatic analysis. GL, JS, and SP conducted 16S rRNA gene sequencing. The draft manuscript was prepared by GL, JS, and WK.

Funding

This work was supported by a Multidisciplinary Research Grant-in-aid from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI21C0538) and the National Research Foundation of Korea (NRF) (grant numbers: 2021R1C1C2004318, 2015M3C9A4053391, and 2019M3C9A6089879).

Data Availability Statement

Sequences obtained from the NAFLD cohort have been deposited in the European Nucleotide Archive databases with the accession code PRJEB27662, and sequences from the T2D cohort have been deposited with accession code PRJEB45895. The metadata relating to the participants are available from the corresponding author upon request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.10.032.

References

[1] Wong V-S, Adams LA, de Lédinghen V, Wong G-H, Sookoian S. Noninvasive biomarkers in NAFLD and NASH — current progress and future promise. Nature Rev Gastroenterol Hepatol 2018;15(8):461–78.
[2] Younossi ZM, Golabi P, de Avila L, Paik JM, Srishord M, Fukui N, et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: A systematic review and meta-analysis. J Hepatol 2019;71(4):793–801.
[3] Adams LA, Waters OR, Knuiman MW, Elliott RR, Olynyk JK. NAFLD as a risk factor for the development of diabetes and the metabolic syndrome: an eleven-year follow-up study. Am J Gastroenterol 2009;104(4):861–7.
[4] Zelber-Sagi S, Lotan R, Shibolet O, Webb M, Buch A, Nitzan-Kaluski D, et al. Non-alcoholic fatty liver disease independently predicts prediabetes during a 7-year prospective follow-up. Liver Int 2013;33(9):1406–12.
[5] Ferguson D, Finck BN. Emerging therapeutic approaches for the treatment of NAFLD and type 2 diabetes mellitus. Nature Rev Endocrinol 2021;17(8):484–95.
[6] 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(7418):55–60.
[7] Loomba R, Seguritan V, Li W, Long T, Klitgord N, Bhatt A, et al. Gut microbiome-based metagenomic signature for non-invasive detection of advanced fibrosis in human nonalcoholic fatty liver disease. Cell Metab 2017;25(5):1054–1062.e5.
[8] Lee G, You HJ, Bajaj JS, Joo SK, Yu J, Park S, et al. Distinct signatures of gut microbiome and metabolites associated with significant fibrosis in non-obese NAFLD. Nat Commun 2020;11(1). https://doi.org/10.1038/s41467-020-18754-5.
[9] Koo BK, Joo SK, Kim D, Bae JM, Park JH, Kim JH, et al. Additive effects of PNPLA3 and TM6SF2 on the histological severity of non-alcoholic fatty liver disease. J Gastroenterol Hepatol 2018;33(6):1277–85.
[10] Koo BK, Kim D, Joo SK, Kim JH, Chang MS, Kim BG, et al. Sarcopenia is an independent risk factor for non-alcoholic steatohepatitis and significant fibrosis. J Hepatol 2017;66(1):123–31.
[11] Diagnosis and classification of diabetes mellitus. Diabetes Care 2013;36:S67 LP-S74. https://doi.org/10.2337/dc13-S067.
[12] Bedogni G, Bellentani S, Miglioli L, Masutti F, Passalacqua M, Castiglione A, et al. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. BMC Gastroenterol 2006;6(1). https://doi.org/10.1186/1471-230X-6-33.
[13] Ruuskanen MO, Åberg F, Männistö V, Havulinna AS, Méric G, Liu Y, et al. Links between gut microbiome composition and fatty liver disease in a large population sample. Gut Microbes 2021;13(1):1888673. https://doi.org/10.1080/19490976.2021.1888673.
[14] Brunt EM, Janney CG, di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: A proposal for grading and staging the histological lesions. Official Journal of the American College of Gastroenterology | ACG 1999;94.
[15] Brunt EM, Kleiner DE, Wilson LA, Belt P, Neuschwander-Tetri BA, (CRN) for the NCRN. Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. Hepatology 2011;53:810–20. doi: 10.1002/hep.24127.
[16] Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41(6):1313–21.
[17] 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(7):581–3.
[18] Sung J, Cho S-I, Lee K, Ha M, Choi E-Y, Choi J-S, et al. Healthy twin: a twin-family study of Korea — protocols and current status. Twin Res Genetics 2006;9(6):844–8.
[19] Gloor GB, Wu JR, Pawlowsky-Glahn V, Egozcue JJ. It's all relative: analyzing microbiome data as compositions. Ann Epidemiol 2016;26(5):322–9.
[20] Palarea-Albaladejo J, Martín-Fernández JA. zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. Chemometrics and Intelligent Laboratory Systems 2015;143:85–96.
[21] Cui E, Fan X, Li Z, Liu Y, Neal AL, Hu C, et al. Variations in soil and plant-microbiome composition with different quality irrigation waters and biochar supplementation. Appl Soil Ecol 2019;142:99–109.
[22] Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable association discovery in population-scale meta-omics Studies. BioRxiv 2021:2021.01.20.427420. doi: 10.1101/2021.01.20.427420.
[23] Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. Science (New York, NY) 2016;352(6285):560–4. https://doi.org/10.1126/science:aad3503.

[24] Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci 2003;14(6):927–30. https://doi.org/10.1111/jvs.2003.14.issue-610.1111/j.1654-1103.2003.tb02228.x.
[25] Team RC. R: a language and environment for statistical computing. R Foundation for Statistical Computing 2019.
[26] McMurdie PJ, Holmes S, Watson M. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 2013;8(4):e61217.
[27] Barnier J. pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.4 2020.
[28] Gloor GB, Reid G. Compositional analysis: a valid approach to analyze microbiome high-throughput sequencing data. Can J Microbiol 2016;62(8):692–703. https://doi.org/10.1139/cjm-2015-0821.
[29] Morgan M. DirichletMultinomial: dirichlet-multinomial mixture model machine learning for microbiome data. R package version 1.26.0. 2019.
[30] Fife D. A Biostatisticians toolbox for various activities, including plotting, data cleanup, and data analysis. R package version 1.1. 2017.
[31] Kim S. ppcor: an R Package for a fast calculation to semi-partial correlation coefficients. Commun Statistical Appl Methods 2015;22(6):665–74. https://doi.org/10.5351/CSAM.2015.22.6.665.
[32] Holmes I, Harris K, Quince C, Gilbert JA. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS ONE 2012;7(2):e30126.
[33] Fuchs G, Boll M, Heider J. Microbial degradation of aromatic compounds — from one strategy to four. Nature Reviews Microbiology 2011 9:11 2011;9:803–16. doi: 10.1038/nrmicro2652.
[34] Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 2015;528(7581):262–6.
[35] Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature 2011;473(7346):174–80. https://doi.org/10.1038/nature09944.
[36] Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. Nature 2014;509(7500):357–60. https://doi.org/10.1038/nature13178.
[37] Wu GD, Chen J, Hoffmann C, Bittinger K, Chen Y-Y, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science 2011;334:105 LP – 108. doi: 10.1126/science.1208344.
[38] Otgonsuren M, Estep MJ, Hossain N, Younossi E, Frost S, Henry L, et al. A single non-invasive model to diagnose non-alcoholic fatty liver disease (NAFLD) and non-alcoholic steatohepatitis (NASH). J Gastroenterol Hepatol 2014;29(12):2006–13.
[39] Shen F, Zheng R-D, Sun X-Q, Ding W-J, Wang X-Y, Fan J-G. Gut microbiota dysbiosis in patients with non-alcoholic fatty liver disease. Hepatobiliary & Pancreatic Diseases International 2017;16(4):375–81.
[40] 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(2):601–9.
[41] Oh TG, Kim SM, Caussy C, Fu T, Guo J, Bassirian S, et al. A universal gut-microbiome-derived signature predicts cirrhosis. Cell Metab 2020;32(5):878–888.e6.
[42] Chen F, Esmaili S, Rogers GB, Bugianesi E, Petta S, Marchesini G, et al. Lean NAFLD: a distinct entity shaped by differential metabolic adaptation. Hepatology 2020;71(4):1213–27. https://doi.org/10.1002/hep.v71.410.1002/hep.30908.
[43] Wang B, Jiang X, Cao M, Ge J, Bao Q, Tang L, et al. Altered fecal microbiota correlates with liver biochemistry in nonobese patients with non-alcoholic fatty liver disease. Sci Rep 2016;6(1). https://doi.org/10.1038/srep32002.
[44] Larsen N, Vogensen FK, van den Berg FWJ, Nielsen DS, Andreasen AS, Pedersen BK, et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PLoS ONE 2010;5(2):e9085. https://doi.org/10.1371/journal.pone.000908510.1371/journal.pone.0009085.s00110.1371/journal.pone.0009085.s00210.1371/journal.pone.0009085.s003.
[45] Guohong-Liu, Qingxi-Zhao, Hongyun-Wei. Characteristics of intestinal bacteria with fatty liver diseases and cirrhosis. Ann Hepatol 2019;18(6):796–803.
[46] Alvarez-Silva C, Kashani A, Hansen TH, Pinna NK, Anjana RM, Dutta A, et al. Trans-ethnic gut microbiota signatures of type 2 diabetes in Denmark and India. Genome Med 2021;13(1). https://doi.org/10.1186/s13073-021-00856-4.
[47] Li Q, Chang Y, Zhang K, Chen H, Tao S, Zhang Z. Implication of the gut microbiome composition of type 2 diabetic patients from northern China. Sci Rep 2020;10:5450.
[48] Oh JH, Lee JH, Cho MS, Kim H, Chun J, Lee JH, et al. Characterization of gut microbiome in Korean patients with metabolic associated fatty liver disease. Nutrients 2021;13(3):1013. https://doi.org/10.3390/nu13031013.
[49] Zhang X, Shen D, Fang Z, Jie Z, Qiu X, Zhang C, et al. Human gut microbiota changes reveal the progression of glucose intolerance. PLoS ONE 2013;8(8):e71108. https://doi.org/10.1371/journal.pone.0071108.
[50] Howard BV, Robbins DC, Sievers ML, Lee ET, Rhoades D, Devereux RB, et al. LDL cholesterol as a strong predictor of coronary heart disease in diabetic individuals with insulin resistance and low LDL. Arterioscler Thromb Vasc Biol 2000;20(3):830–5.
[51] Chen Y, Xu M, Wang T, Sun J, Sun W, Xu B, et al. Advanced fibrosis associates with atherosclerosis in subjects with nonalcoholic fatty liver disease. Atherosclerosis 2015;241(1):145–50.
[52] Treeprasertsuk S, Björnsson E, Enders F, Suwanwalaikorn S, Lindor KD. NAFLD fibrosis score: a prognostic predictor for mortality and liver complications among NAFLD patients. World J Gastroenterol 2013;19:1219–29. https://doi.org/10.3748/wjg.v19.i8.1219.
[53] Park SB, Choi HJ, Doo SH, Jung D, Shim YN, Oh HJ. The association between non-alcoholic fatty liver disease fibrosis score and serum low density lipoprotein-cholesterol levels in adults with non-alcoholic fatty liver disease. Korean J Family Practice 2020;10(2):110–5.
[54] Klimentidis YC, Arora A, Newell M, Zhou J, Ordovas JM, Renquist BJ, et al. Phenotypic and genetic characterization of lower LDL cholesterol and increased type 2 diabetes risk in the UK biobank. Diabetes 2020;69:2194 LP – 2205.
[55] Dubinsky V, Reshef L, Rabinowitz K, Yadgar K, Godny L, Zonensain K, et al. Dysbiosis in Metabolic Genes of the Gut Microbiomes of Patients with an Ileo-anal Pouch Resembles That Observed in Crohn's Disease. MSystems 2021;6. doi: 10.1128/MSYSTEMS.00984-20.
[56] Carpino G, Del Ben M, Pastori D, Carnevale R, Baratta F, Overi D, et al. Increased liver localization of lipopolysaccharides in human and experimental NAFLD. Hepatology 2020;72(2):470–85.
[57] Yuan J, Chen C, Cui J, Lu J, Yan C, Wei X, et al. Fatty liver disease caused by high-alcohol-producing Klebsiella pneumoniae. Cell Metab 2019;30(6):1172. https://doi.org/10.1016/j.cmet.2019.11.006.
[58] Maier L, Pruteanu M, Kuhn M, Zeller G, Telzerow A, Anderson EE, et al. Extensive impact of non-antibiotic drugs on human gut bacteria. Nature 2018 555:7698 2018;555:623–8. doi: 10.1038/nature25979.
[59] Vich Vila A, Collij V, Sanna S, Sinha T, Imhann F, Bourgonje AR, et al. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. Nature Communications 2020 11:1 2020;11:1–11. doi: 10.1038/s41467-019-14177-z.
[60] Wu H, Esteve E, Tremaroli V, Khan MT, Caesar R, Mannerås-Holm L, et al. Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug. Nature Medicine 2017 23:7 2017;23:850–8. doi: 10.1038/nm.4345.
[61] Singh RK, Chang H-W, Yan D, Lee KM, Ucmak D, Wong K, et al. Influence of diet on the gut microbiome and implications for human health. J Transl Med 2017 15:1 2017;15:1–17. doi: 10.1186/S12967-017-1175-Y.
[62] Trovato FM, Castrogiovanni P, Malatino L, Musumeci G. Nonalcoholic fatty liver disease (NAFLD) prevention: role of Mediterranean diet and physical activity. Hepatobiliary Surgery and Nutrition 2019;8(2):167–9.
[63] Vandeputte D, Kathagen G, D'hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al. Quantitative microbiome profiling links gut community variation to microbial load. Nature 2017;551(7681):507–11.
[64] Valles-Colomer M, Falony G, Darzi Y, Tigchelaar EF, Wang J, Tito RY, et al. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat Microbiol 2019;4(4):623–32. https://doi.org/10.1038/s41564-018-0337-x.
[65] Vieira-Silva S, Sabino J, Valles-Colomer M, Falony G, Kathagen G, Caenepeel C, et al. Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. Nat Microbiol 2019;4(11):1826–31. https://doi.org/10.1038/s41564-019-0483-9.