









Distinct signatures of gut microbiome and metabolites associated with significant fibrosis in non-obese NAFLD

Giljae Lee ^{1,10}, Hyun Ju You^{1,2,3,10}, Jasmohan S. Bajaj ⁴, Sae Kyung Joo⁵, Junsun Yu¹, Seoyeon Park¹, Hyena Kang¹, Jeong Hwan Park⁶, Jung Ho Kim⁶, Dong Hyeon Lee ⁵, Seonhwa Lee⁷, Won Kim ⁵  & GwangPyo Ko ^{1,3,8,9} 

Nonalcoholic fatty liver disease (NAFLD) is associated with obesity but also found in non-obese individuals. Gut microbiome profiles of 171 Asians with biopsy-proven NAFLD and 31 non-NAFLD controls are analyzed using 16S rRNA sequencing; an independent Western cohort is used for external validation. Subjects are classified into three subgroups according to histological spectra of NAFLD or fibrosis severity. Significant alterations in microbiome diversity are observed according to fibrosis severity in non-obese, but not obese, subjects. *Ruminococcaceae* and *Veillonellaceae* are the main microbiota associated with fibrosis severity in non-obese subjects. Furthermore, stool bile acids and propionate are elevated, especially in non-obese subjects with significant fibrosis. Fibrosis-related *Ruminococcaceae* and *Veillonellaceae* species undergo metagenome sequencing, and four representative species are administered in three mouse NAFLD models to evaluate their effects on liver damage. This study provides the evidence for the role of the microbiome in the liver fibrosis pathogenesis, especially in non-obese subjects.

¹ Department of Environmental Health Sciences, Graduate School of Public Health, Seoul National University, Seoul 08826, Republic of Korea. ² Institute of Health and Environment, Seoul National University, Seoul 08826, Republic of Korea. ³ Center for Human and Environmental Microbiome, Institute of Health and Environment, Seoul National University, Seoul 08826, Republic of Korea. ⁴ Division of Gastroenterology, Hepatology and Nutrition, Virginia Commonwealth University and McGuire VA Medical Center, Richmond, VA 23249, USA. ⁵ 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. ⁶ Department of Pathology, Seoul National University College of Medicine, Seoul Metropolitan Government Boramae Medical Center, Seoul 07061, Republic of Korea. ⁷ Department of Bio-convergence Engineering, Korea University, Seoul 02841, Republic of Korea. ⁸ KoBioLabs, Inc., Seoul 08826, Republic of Korea. ⁹ Bio-MAX/N-Bio, Seoul National University, Seoul 08826, Republic of Korea. ¹⁰ These authors contributed equally: Giljae Lee, Hyun Ju You. email: drwon1@snu.ac.kr; gko@snu.ac.kr

Nonalcoholic fatty liver disease (NAFLD) is characterized by hepatic manifestations of metabolic disorders, ranging from simple steatosis to nonalcoholic steatohepatitis (NASH), an aggressive histological form, ultimately leading to advanced fibrosis and cirrhosis. The global prevalence of NAFLD is estimated to be 24–30% in most epidemiological studies and is increasing in parallel with obesity and metabolic syndrome¹. Although NAFLD is usually linked to obesity, NAFLD also occurs in non-obese subjects presenting with pathological severities similar to those observed in obese NAFLD patients^{2,3}. In both the West and the East, 3–30% of non-obese populations are consistently reported to have NAFLD, without considering different body mass index (BMI) cut-offs for defining obesity (≥ 25 in Asians vs. ≥ 30 in other ethnicities)⁴. Although visceral fat, dietary composition, and genetic factors can be associated with non-obese NAFLD⁵, additional studies considering other environmental factors are required to elucidate the pathogenesis of non-obese NAFLD.

Increased interest has recently focused on identifying and understanding the specific role of the intestinal microbiota in various metabolic diseases. Gut dysbiosis has been associated with a decrease in beneficial short-chain fatty acid (SCFA)-producing bacteria, changes in the composition of bile acids⁶, activation of immune responses against lipopolysaccharide (LPS)⁷, increased ethanol production by overgrowth of ethanol-producing bacteria^{8,9}, and conversion of phosphatidylcholine to choline and trimethylamine¹⁰. Changes in the gut microbiota, which affect the gut–liver axis, are also associated with the progression of chronic liver disease and advanced fibrosis, such as NAFLD and cirrhosis^{11–15}. However, the microbial taxa associated with disease severity and fibrosis stage are not consistent with those reported in previous NAFLD studies. This inconsistency could be due to the effects of regional differences¹⁶. However, differential underlying BMI status might also account for these inconsistent results.

Here, the aims of this study are to determine whether the histological severity of NAFLD is associated with stool microbial changes in a well-characterized, biopsy-proven Asian NAFLD cohort and to explore microbial markers for the diagnosis of non-obese NAFLD. The 16S rRNA gene amplicon sequencing with 202 subjects stratified by obese status shows specific and distinct changes in the gut microbial community according to fibrosis severity, especially in non-obese subjects. After adjustments for age, sex, BMI, diabetes mellitus, and cirrhosis using multivariate association with linear models (MaAsLin), *Ruminococcaceae* and *Veillonellaceae* are the main microbial taxa associated with significant fibrosis in non-obese subjects. Moreover, stool primary bile acids and propionate levels are significantly elevated in non-obese subjects with worsening fibrosis severity. In bacterial taxametabolites networks according to fibrosis severity, strong inverse interactions between *Veillonellaceae* and *Ruminococcaceae* are noticeable and observed only in non-obese NAFLD subjects, implying the potential of the microbiome-based marker to diagnose fibrosis in non-obese NAFLD. The microbiome–metabolite combination suggested in this study significantly enhances the diagnostic power with an improved AUC of 0.939 (0.584 in all subjects; 0.520 in obese subjects), which is higher than that of fibrosis 4 index in non-obese subjects. We further conduct metagenomic shotgun sequencing of stool samples collected from 38 non-obese subjects, and species of *Megamonas* and *Ruminococcus* are identified as representative fibrosis-related taxa consisting of *Veillonellaceae* and *Ruminococcaceae*.

The metagenome analysis also reveals the relationship between significantly downregulated microbial genes related to primary/secondary bile acid metabolism and increasing amounts of total conjugated bile acids and unconjugated primary bile acids in non-

obese subjects with significant fibrosis. The protective or worsening effect of four strains of bacteria belonging to *Ruminococcaceae* and *Veillonellaceae* on liver damage is investigated using three different mouse NAFLD models. *Ruminococcus faecis* shows an alleviating effect on fibrosis including significantly regressed histological severity and decreased expression of fibrogenic genes (*Colla1*, *Timp1*, and *α -SMA*). Taken together, the intestinal bacteria and related metabolites in the pathogenesis of non-obese NAFLD have the potential to be utilized as a fibrosis marker, as well as a therapeutic target.

Results

Baseline characteristics of human subjects. This study included 171 subjects with biopsy-proven NAFLD (NAFL, $n = 88$; NASH, $n = 83$) and 31 no-NAFLD controls. All of the study subjects were classified into two groups (non-obese, BMI < 25 ; obese, BMI ≥ 25), and each group was divided into three subgroups according to the histological spectra of NAFLD or fibrosis severity. Supplementary Tables 1 and 2 present the detailed characteristics of each group, including clinical, metabolic, biochemical, and histological profiles. Subjects with NASH or significant fibrosis (F2–4) had higher levels of aspartate aminotransferase (AST), alanine aminotransferase (ALT), and insulin resistance in both the obese and non-obese groups. Subjects with significant fibrosis had higher NAFLD activity scores and presented more severe liver histology in terms of the histological classification of NAFLD (Supplementary Table 3 and Supplementary Fig. 1). More detailed baseline characteristics in each fibrosis stage, including well-known NAFLD-associated genetic variants, such as *PNPLA3*, *TM6SF2*, *MBOAT7-TMC4*, and *SREBF-2*, are shown in Supplementary Table 4.

Microbial community changes according to fibrosis severity are different in non-obese and obese NAFLD subjects. We analyzed data from 16S rRNA gene amplicon sequencing and compared the microbial diversity according to the histological spectra of NAFLD or fibrosis severity (Fig. 1). Alpha diversity based on the Shannon metric was plotted, and beta diversity based on the Bray–Curtis distance was plotted for comparison. No significant alterations in diversity among groups stratified by the histological spectra of NAFLD or fibrosis severity were observed (Fig. 1a). We then classified the subjects into two groups according to BMI status. In the non-obese group, a significant decrease in microbial diversity between F1 and F0 ($P = 0.0074$) as well as between F2–4 and F0 ($P = 0.0084$) was observed (Fig. 1b). Moreover, apparent clustering between F0 and F2–4 was observed ($P = 0.038$). In the obese group, no significant differences in diversity were found among groups stratified by the histological spectra of NAFLD or fibrosis severity (Fig. 1c). These results indicate that fibrosis severity rather than necroinflammatory activity is more likely associated with the changes of gut microbiome and underlying BMI status may also be an important factor responsible for alterations of the gut microbiome.

Alterations of fibrosis-related microbiota are more prominent in non-obese NAFLD rather than obese NAFLD. Differences in the specific microbial taxa by fibrosis severity in non-obese and obese subjects were compared using univariate and multivariate analyses (Fig. 2). In univariate analysis, the gradual enrichment of *Veillonellaceae*, as well as *Enterobacteriaceae*, was observed according to fibrosis severity in non-obese subjects (Fig. 2a). In obese subjects, *Rikenellaceae* was gradually enriched. In contrast, the abundance of *Ruminococcaceae* significantly decreased as fibrosis became more severe, which was observed only in non-obese subjects. At the genus level, *Faecalibacterium*, *Ruminococcus*

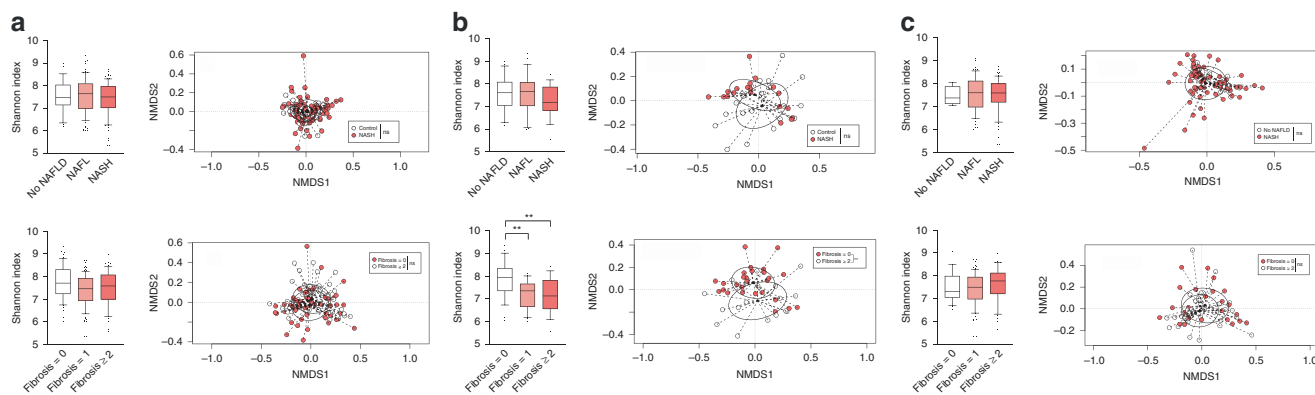


Fig. 1 Comparison of the diversity of gut microbial communities in all, non-obese, and obese subjects. The alpha and beta diversity of **a** all ($n = 202$), **b** non-obese subjects ($n = 64$) (** $P = 0.0074$, FO vs F1; ** $P = 0.0084$, FO vs F2-4), and **c** obese subjects ($n = 138$) divided by the histological spectra of NAFLD or fibrosis severity. Alpha diversity was based on the Shannon index with 12,000 rarefied sequences per sample. The box plots indicate the median, 25th to 75th percentiles (boxes), and 10th to 90th percentiles (whiskers). Statistical analysis was performed using a two-sided nonparametric Mann-Whitney test or a nonparametric Kruskal-Wallis test with Dunn's multiple comparison test. NMDS plots were generated using relative OTU abundance data according to the Bray-Curtis distance, and statistical significance was measured using adonis analysis (panel **b**; * $P = 0.038$, FO vs F2-4).

(*Ruminococcaceae*), *Coprococcus*, and *Lachnospira* (*Lachnospiraceae*) were significantly depleted in the significant fibrosis group (Fig. 2b), while the abundances of *Enterobacteriaceae*_Other (*Enterobacteriaceae*) and *Citrobacter* gradually increased according to fibrosis severity. These alterations were also observed only in non-obese subjects. These results could be also found in the correlation plots (Supplementary Fig. 2). *Enterobacteriaceae* and *Veillonellaceae* had positive correlations ($P = 1.09 \times 10^{-4}$, $P = 2.44 \times 10^{-3}$, respectively), but *Ruminococcaceae* showed an inverse correlation ($P = 4.41 \times 10^{-4}$) with fibrosis severity.

For multivariate analysis, we adjusted for age, sex, and BMI using MaAsLin¹⁷. In the phylum Firmicutes, *Veillonellaceae* showed a steeper increase in relative abundance in non-obese subjects than in obese subjects (non-obese, $P = 0.0002$, $q = 0.0297$), while the abundance of *Ruminococcaceae* was inversely correlated with fibrosis severity in non-obese subjects ($P = 0.0012$, $q = 0.0972$) (Fig. 2c). *Ruminococcus*, a representative genus of *Ruminococcaceae*, also showed a significant inverse association with fibrosis severity only in non-obese subjects ($P = 0.0011$, $q = 0.152$) (Supplementary Fig. 3). In addition, *Veillonellaceae* and *Enterobacteriaceae* showed a significant, positive association with serum free fatty acid (FFA) levels in non-obese subjects ($q = 0.178$, $q = 0.118$, respectively) but not in obese subjects (Supplementary Fig. 4). Adipose tissue insulin resistance (adipo-IR) and glycosylated hemoglobin (HbA1c) were also positively correlated with the abundance of *Veillonellaceae* in non-obese subjects (adipo-IR, $q = 0.142$; HbA1c, $q = 0.157$). In contrast, serum FFA levels were inversely correlated with the abundance of *Ruminococcus* in all subjects ($q = 0.0838$) and in non-obese subjects ($q = 0.144$) but not in obese subjects ($q = 1.00$).

In addition to three variables (age, sex, and BMI), the presence of DM is also known to affect general changes in the microbial community¹⁸. We found that *Enterobacteriaceae* ($P = 0.0002$, $q = 0.039$) and *Faecalibacterium* ($P = 0.0029$, $q = 0.12$) were associated with the presence of DM in all subjects (Supplementary Fig. 5a). The depletion of *Lachnospira* ($P = 4.36 \times 10^{-4}$, $q = 0.316$) (Supplementary Fig. 5b) in the non-obese, as well as the enrichment of *Klebsiella* ($P = 0.0026$, $q = 0.207$) (Supplementary Fig. 5c), which belongs to *Enterobacteriaceae*, in the obese, was observed in subjects with DM. However, the associations of *Veillonellaceae* and *Ruminococcaceae* with fibrosis severity remained significant even after adjustment for DM (Fig. 2d) and the use of metformin (Supplementary Fig. 6). The anti-

diabetic medications prescribed in each subgroup are listed in Supplementary Table 5. The oral hypoglycemic agents include metformin, linagliptin, glimepiride, etc., and some patients received combined treatment. To exclude the confounding effect of cirrhosis, we also performed sensitivity analysis only for study subjects without cirrhosis. Nonetheless, the significant associations of *Veillonellaceae* and *Ruminococcaceae* with fibrosis severity were not affected by cirrhosis (Fig. 2e). In addition, we conducted the same analysis with only NAFLD patients except for healthy control subjects. Similarly, *Veillonellaceae* and *Ruminococcaceae* were significantly associated with fibrosis severity notwithstanding the smaller population (Fig. 2f). To determine whether these notable microbiome changes in non-obese subjects could be attributed to the host gene effect, we adjusted for genetic variants of *PNPL3*, *TM6SF2*, *MBOAT7-TMC4*, and *SREBF-2* in multivariate analysis. The associations between two identified taxa and fibrosis severity in non-obese subjects remained significant even after adjustment for host genetic variants (Supplementary Fig. 7). Taken together, these findings indicate that enrichment of specific taxa according to fibrosis severity might be more prominent in the non-obese group than in the obese group.

Non-obese and obese NAFLD subjects have different stool metabolites levels according to fibrosis severity. We next assessed the stool metabolites that are mainly associated with the gut microbiome using Q-TOF and GC-FID systems. The composition of the bile acid pool varied between non-obese and obese subjects (Fig. 3a and Supplementary Fig. 8). The levels of unconjugated and conjugated bile acids noticeably increased in non-obese subjects with fibrosis, and the total stool bile acid levels were higher (unconjugated bile acids, 2.3 times; conjugated bile acids, 3.6 times) in non-obese subjects with significant fibrosis (F2-4) than in those without fibrosis (F0) (Fig. 3b). In obese subjects, total conjugated bile acid levels decreased with increasing fibrosis severity. Specifically, cholic acid (CA), chenodeoxycholic acid (CDCA), ursodeoxycholic acid (UDCA), glycochenodeoxycholic acid (GCDCA), and glyoursodeoxycholic acid (GUDCA) levels significantly increased in non-obese subjects with worsening fibrosis severity (Fig. 3c). Lithocholic acid (LCA) and deoxycholic acid (DCA) levels were elevated in obese subjects with significant fibrosis, but only LCA reached statistical significance (Fig. 3c). Among three SCFAs, stool propionate levels gradually increased

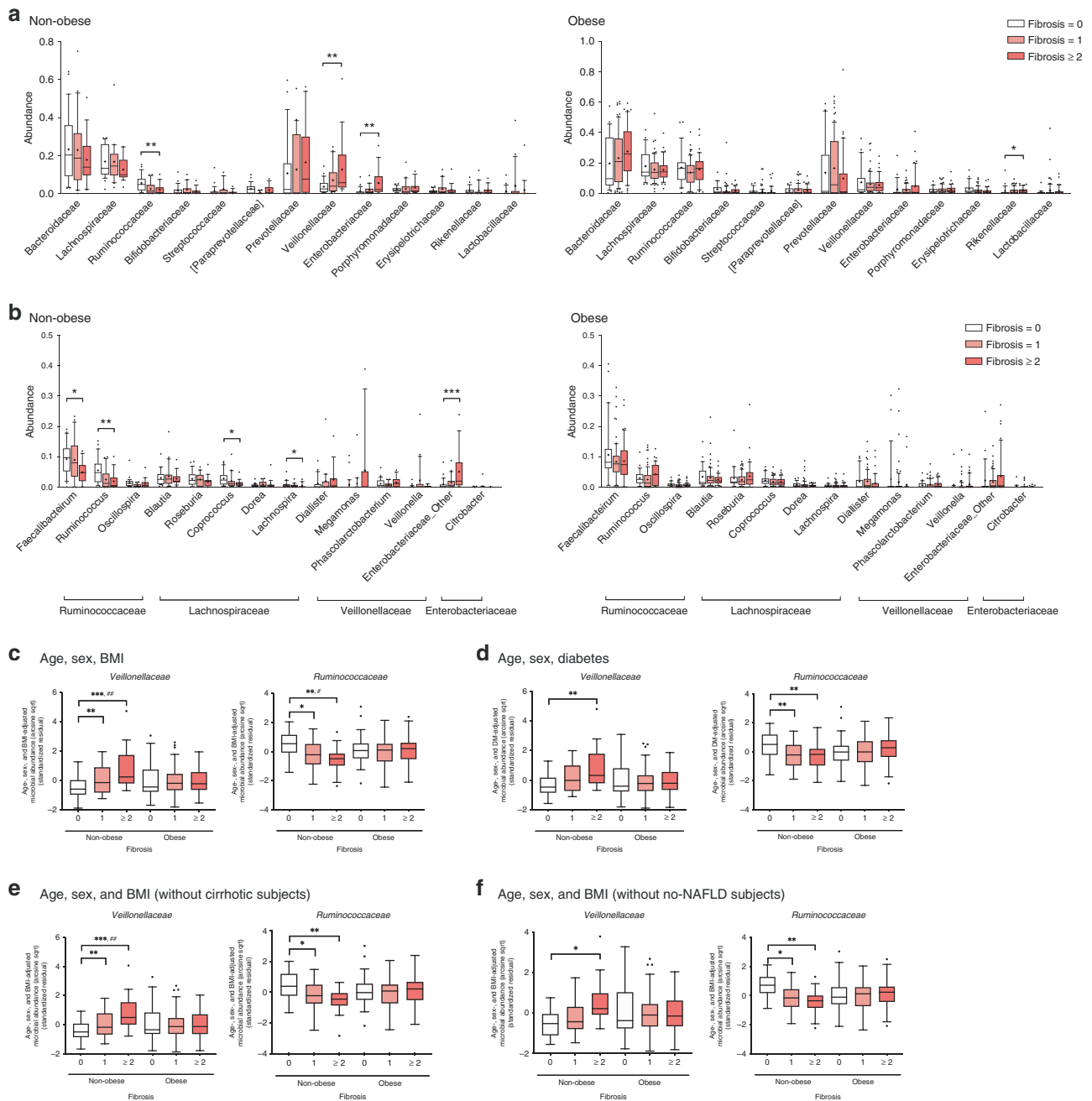


Fig. 2 Phylogenetic comparisons of the gut microbiome in non-obese and obese subjects. **a** The 13 family- and **b** 14 genus-level taxa with top abundances are depicted for clarity. Statistical significance was measured using a nonparametric Kruskal–Wallis test with Dunn’s multiple comparison test. The *P* values are as follows: panel (**a**): ***P* = 0.0013, ***P* = 0.0054, ***P* = 0.0016, and **P* = 0.04; panel (**b**): **P* = 0.0217, ***P* = 0.0013, **P* = 0.0209, **P* = 0.0234, and ****P* = 0.0014. Multivariate associations between specific gut-microbiome components and fibrosis severity stratified by obesity status (**c–f**). Arcsine-root transformed abundances of bacteria were regressed against **c** age, sex, and BMI, **d** age, sex, and diabetes, **e** age, sex, and BMI (without cirrhotic subjects), and **f** age, sex, and BMI (without no-NAFLD subjects). Statistical analyses for multivariate associations were performed using MaAsLin with adjustments for multiple comparisons (*q* value). The *P* and *q* values are as follows: panel (**c**): ***P* = 0.0057, ****P* < 0.001, and ##*q* = 0.0297, **P* = 0.0313, ***P* = 0.0012, and #*q* = 0.0972; panel (**d**): from left, ***P* = 0.0020, ***P* < 0.0100, (F0 vs F1), and ***P* = 0.0013 (F0 vs F2–4); panel (**e**): ***P* = 0.0071, ****P* = 0.0001, and ##*q* = 0.0339, **P* = 0.0191, and ***P* = 0.0015; panel (**f**): **P* = 0.0136, **P* = 0.0292, and ***P* = 0.0023. The box plots indicate the median, 25th to 75th percentiles (boxes), and 10th to 90th percentile (whiskers). Colors inside the box represent fibrosis severity. **a–d** *n* = 64 for non-obese and *n* = 138 for obese; **e** *n* = 60 for non-obese and *n* = 125 for obese; **f** *n* = 43 for non-obese and *n* = 129 for obese. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, #*q* < 0.10, ##*q* < 0.05. Source data are provided as a Source Data file.

as fibrosis became more severe in non-obese subjects (non-obese; *P* = 0.0032, obese; *P* = 0.7979) (Fig. 3d).

Non-obese NAFLD and obese NAFLD show different patterns of bacterial taxa-metabolites networks. To understand the

interaction among gut microbiota in non-obese and obese subjects, the co-occurrence of taxa was measured and depicted with relative abundances in correlation with fibrosis severity (Supplementary Fig. 9). As expected, *Veillonellaceae* and *Enterobacteriaceae* were inversely correlated with *Ruminococcaceae* in non-obese subjects.

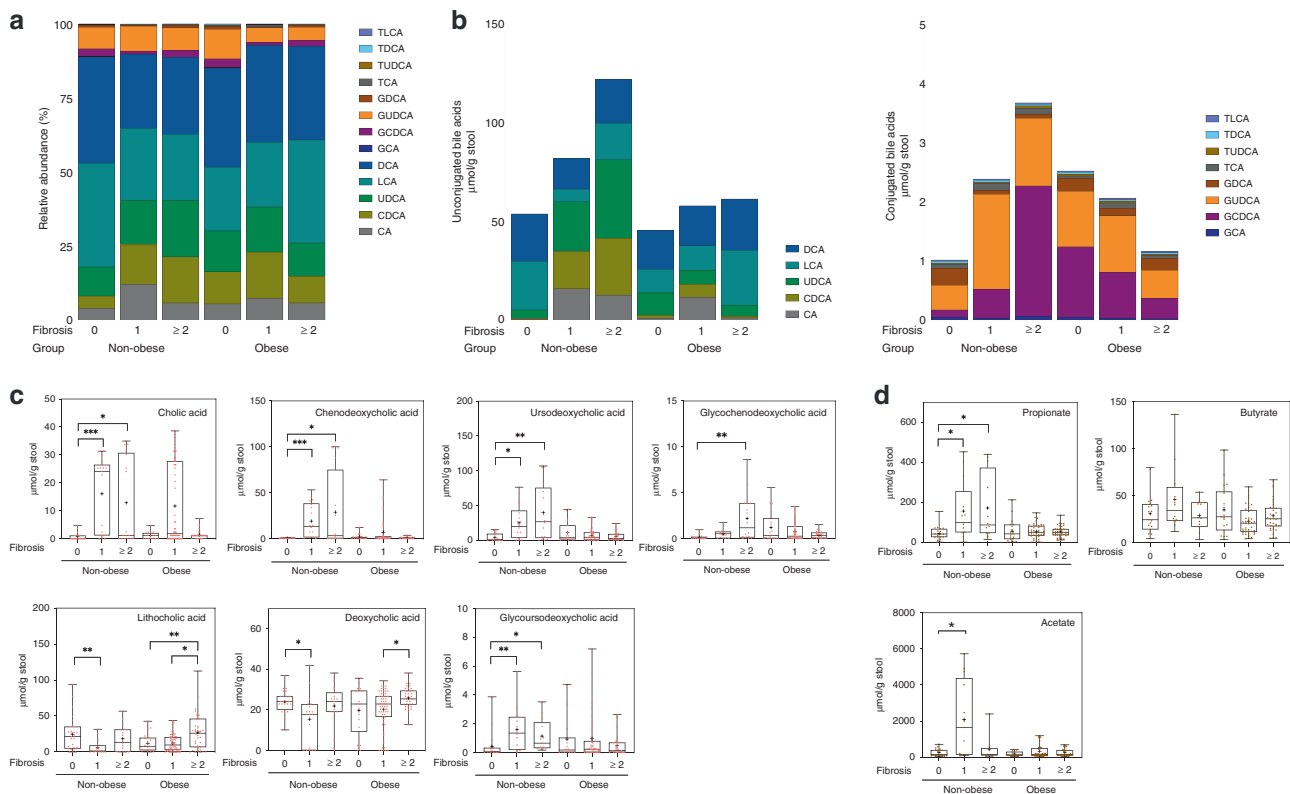


Fig. 3 Comparison of stool metabolites levels stratified by fibrosis severity and obesity status. **a** Composition of bile acid profiles in different clinical settings. Stacked plots are generated using the average abundances of the 13 bile acids. **b** The stacked bars represent (left) unconjugated bile acids levels and (right) conjugated bile acids levels, which are stratified by fibrosis severity and obesity status. **c** Box plots represent the concentrations of the stool bile acids, which are stratified by fibrosis severity and obesity status; cholic acid (CA), *** $P = 0.0007$, * $P = 0.0260$, $n = 57$ for non-obese, and $n = 111$ for obese; chenodeoxycholic acid (CDCA), *** $P < 0.0001$, * $P = 0.0164$, $n = 54$ for non-obese, and $n = 96$ for obese; ursodeoxycholic acid (UDCA), * $P = 0.0107$, ** $P = 0.0012$, $n = 56$ for non-obese, and $n = 95$ for obese; glycochenodeoxycholic acid (GCDCA), ** $P = 0.0023$, $n = 52$ for non-obese, and $n = 110$ for obese; lithocholic acid (LCA), ** $P = 0.0075$, * $P = 0.0265$, ** $P = 0.0045$, $n = 59$ for non-obese, and $n = 122$; deoxycholic acid (DCA), * $P = 0.0119$, * $P = 0.0128$, $n = 60$ for non-obese, and $n = 122$ for obese; glycoursodeoxycholic acid (GUDCA), ** $P = 0.0064$, * $P = 0.0115$, $n = 47$ for non-obese, and $n = 99$ for obese. **d** Box plots represent the most abundant fecal short-chain fatty acids (SCFAs) levels (acetate, propionate, and butyrate), which are stratified by fibrosis severity and obesity status; propionate, * $P = 0.0114$, * $P = 0.0182$, $n = 51$ for non-obese, and $n = 100$ for obese; butyrate, $n = 43$ for non-obese, and $n = 98$ for obese; acetate, * $P = 0.0440$, $n = 42$ for non-obese, and $n = 89$ for obese. The box plots indicate the median, 25th to 75th percentiles (boxes), and minimum to maximum values (whiskers). Outliers were removed by the ROUT method ($Q = 1\%$) and data were analyzed using a nonparametric Kruskal-Wallis test with Dunn’s multiple comparison test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Taurolithocholic acid (TLCA), taurodeoxycholic acid (TDCA), taurochenodeoxycholic acid (TUDCA), taurocholic acid (TCA), glycodeoxycholic acid (GDCA), glycocholic acid (GCA).

However, strong interactions among these bacteria were not observed in the obese and all of the subjects, implying its specific role in the progression of fibrosis in non-obese subjects.

To explore the key drivers responsible for this observation, we analyzed the co-occurrence analysis of taxa and metabolites as illustrated in Fig. 4. The strong interactions among bile acids were observed in both non-obese and obese subjects (Fig. 4a and b). Interestingly, primary bile acids had an inverse relationship with *Ruminococcaceae*, which are known as indicators of healthy intestines, in both non-obese and obese subjects. *Veillonellaceae* showed positive interactions with primary bile acids and UDCA, as well as propionate. Bile acids usually retain the potential to modulate the growth of susceptible bacteria or enrich relatively resistant bacteria independent of obesity status. Nevertheless, the relationship of gut bacterial taxa and stool metabolites with severe fibrosis was more noticeable in non-obese NAFLD subjects than in obese NAFLD subjects.

Microbiome–metabolite combination reflects significant fibrosis in non-obese subjects with NAFLD. To assess the utility of the

gut microbiome for indicating significant fibrosis, the areas under the receiver-operating characteristic curve (AUCs) for diagnosing significant fibrosis were compared (Fig. 5). *Veillonellaceae* and *Ruminococcaceae* were selected as the most representative and significant fibrosis-related bacterial taxa. The combined bacterial marker to diagnose significant fibrosis yielded an AUC of 0.765 in non-obese subjects (0.559 in all subjects; 0.544 in obese subjects). In addition, we selected four stool metabolites (CA, CDCA, UDCA, and propionate) as fibrosis-related metabolites, and the combination of the four metabolites diagnosed significant fibrosis with an AUC of 0.758 in non-obese subjects (0.574 in all subjects; 0.520 in obese subjects). The addition of stool metabolites to these bacterial taxa significantly enhanced the diagnostic power with an improved AUC of 0.939 (0.584 in all subjects; 0.520 in obese subjects). The diagnostic power of this microbiome–metabolite combination was higher in non-obese subjects than that of FIB-4, which is widely used as a non-invasive fibrosis test in NAFLD. Considering the broad spectrum of pathological changes between healthy controls and cirrhotic subjects, we also assessed the AUCs restricted to study subjects without cirrhosis (Supplementary Fig. 10a) or without healthy controls (Supplementary Fig. 10b).

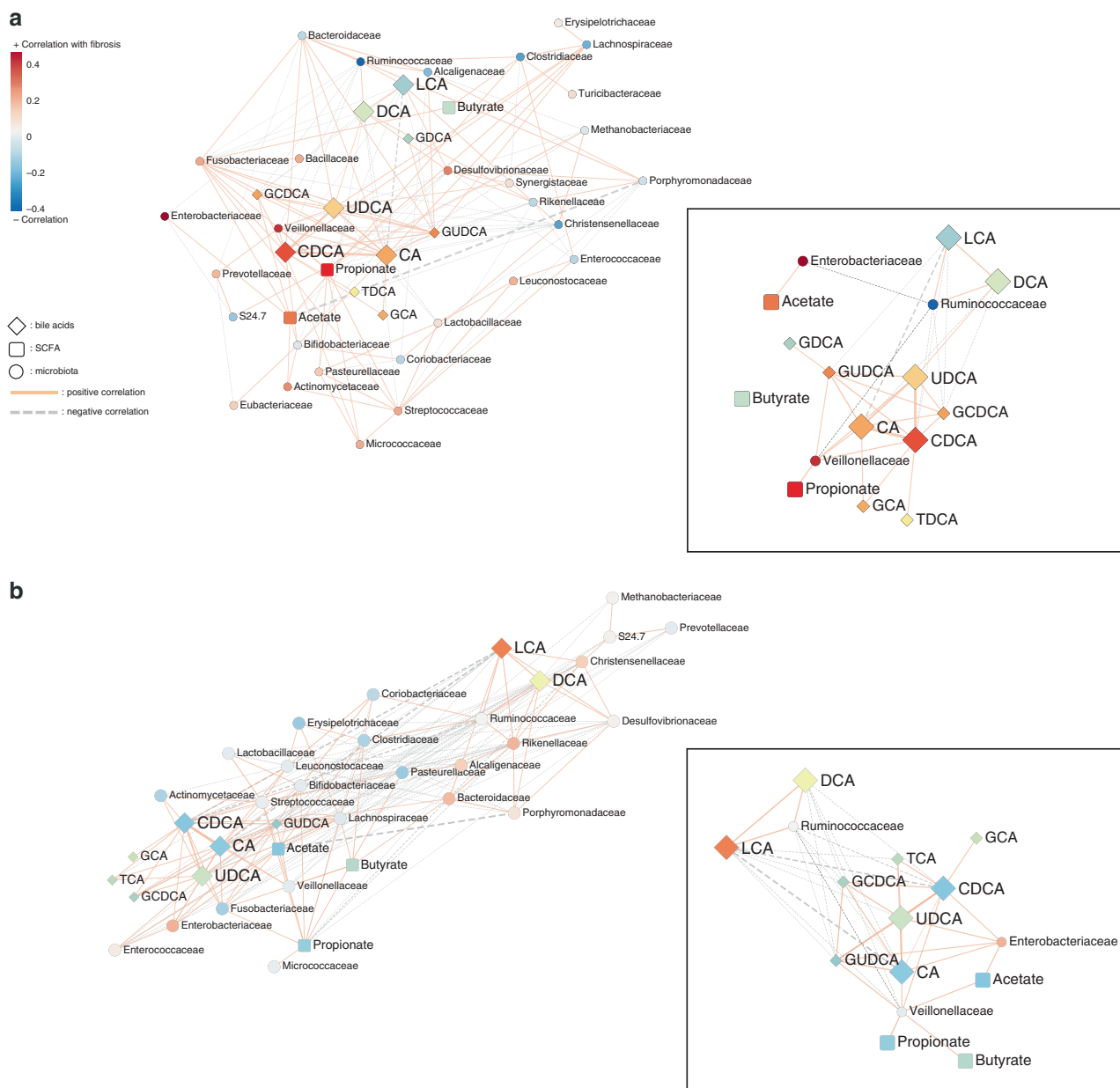


Fig. 4 Network profiles between microbial taxa and stool metabolite components in non-obese and obese subjects. Co-occurrence coefficients among family-level microbiome components and stool metabolites were calculated by SparCC, and networks ($P < 0.05$) are depicted using Cytoscape. **a** Non-obese and **b** obese. The solid line (orange) and dotted line (gray) indicate positive and negative correlations, respectively. The shape of the node denotes the components used in this study (ellipse: microbiome, diamond: stool bile acids, and round rectangle: SCFAs) and the color indicates the degree of correlation with fibrosis severity. The P -value for each coefficient was obtained by bootstrapping the dataset 500 times and applying SparCC to each of those 500 datasets. Source data are provided as a Source Data file.

Nevertheless, the AUCs using six combined markers sustained high diagnostic power in non-obese subjects (0.838 and 0.810 in the populations without cirrhotic subjects and healthy controls, respectively).

The Western NAFLD cohort similarly shows different microbiome patterns between non-obese and obese subjects.

To validate our results using an external independent cohort, we used public datasets. Caussy et al. reported a gut-microbiome-based biomarker for diagnosing NAFLD-related cirrhosis using a well-characterized NAFLD twin cohort¹⁴. A total of 168 subjects were divided into non-obese and obese groups according to BMI (≥ 30 , obese; <30 , non-obese), and each group was divided based on the

presence of advanced fibrosis. Baseline characteristics of the validation cohort are described in Supplementary Table 6. Non-obese subjects with advanced fibrosis were enriched with *Veillonellaceae* ($P = 0.0120$), which was not observed in obese subjects ($P = 0.8818$) (Fig. 6a). In addition, *Ruminococcaceae* showed a tendency to decrease only in non-obese ($P = 0.1350$; obese, $P = 0.9944$) subjects with advanced fibrosis. The combination of two fibrosis-related bacteria diagnosed advanced fibrosis with an AUC of 0.721 in non-obese subjects (0.578 in obese subjects) (Fig. 6b). Because the Western NAFLD twin cohort lacked metabolites data, the diagnostic power of the Western NAFLD twin cohort seemed to be lower than that of our Korean NAFLD cohort using the taxametabolite combination. These results indicate that changes in *Veillonellaceae/Ruminococcaceae* according to fibrosis severity in

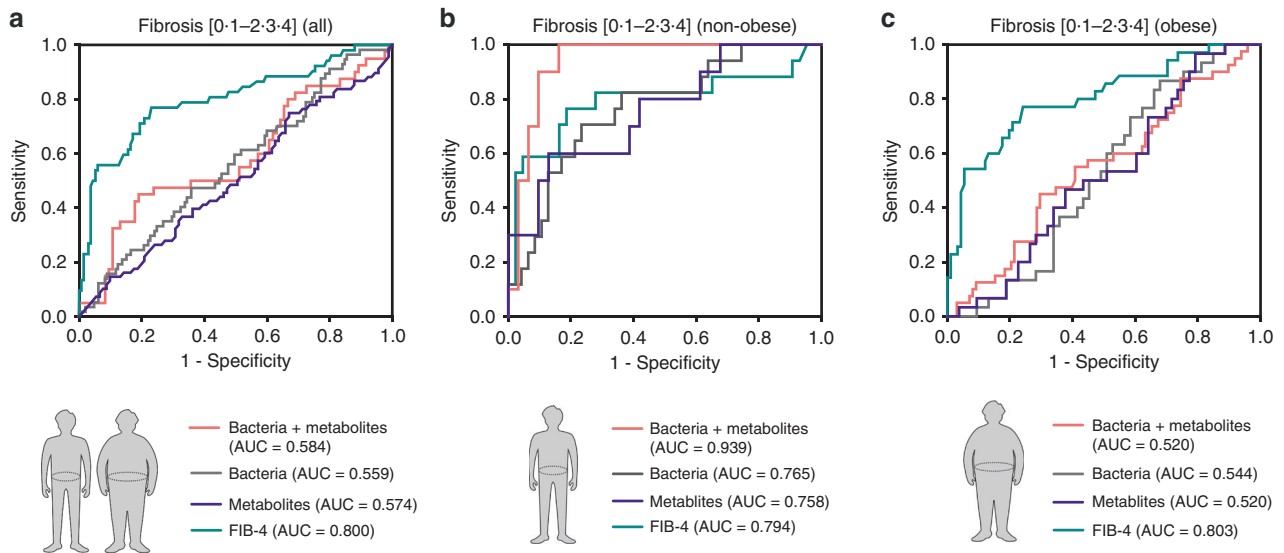


Fig. 5 Receiver-operating characteristic (ROC) curves for the diagnosis of significant fibrosis in all, non-obese, and obese subjects. ROC curves using the combination of two bacteria (*Veillonellaceae* and *Ruminococcaceae*) and four stool metabolites (CA, CDCA, UDCA, and propionate) were plotted for the diagnosis of significant fibrosis in **a** all, **b** non-obese subjects, and **c** obese subjects, and the areas under the ROC curves (AUCs) were calculated. FIB-4, fibrosis 4 index.

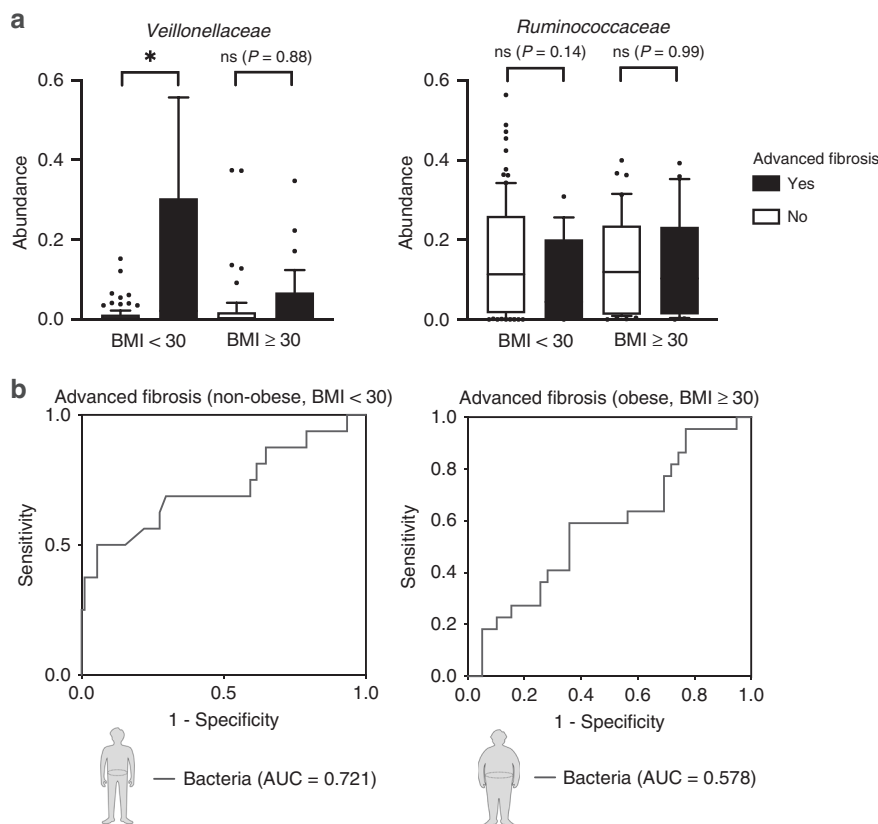


Fig. 6 Validation of gut-microbiome differences between non-obese and obese subjects using a Western NAFLD cohort. **a** Abundance of *Veillonellaceae* and *Ruminococcaceae* stratified by obesity status and advanced fibrosis (* $P = 0.0120$, $n = 107$ for non-obese, and $n = 61$ for obese). **b** ROC curves using the combination of two selected bacteria (*Veillonellaceae* and *Ruminococcaceae*) were plotted for the diagnosis of advanced fibrosis in non-obese and obese subjects, and the areas under the ROC curve (AUCs) were calculated. The box plots indicate the median, 25th to 75th percentiles (boxes), and 10th to 90th percentiles (whiskers). Statistical analysis was performed using a two-sided Mann-Whitney test. * $P < 0.05$. ns, not significant.

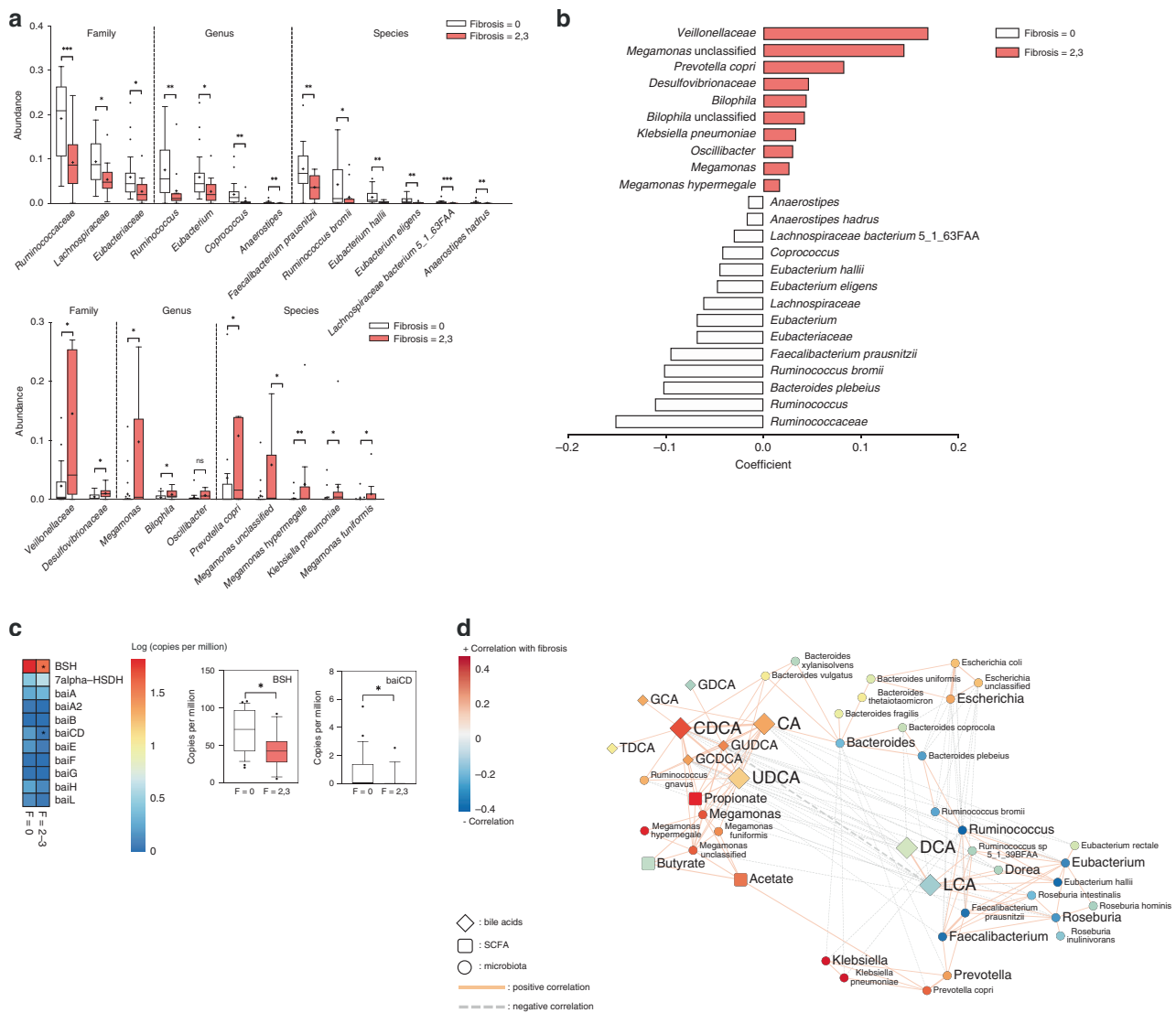


Fig. 7 Identification of fibrosis-related bacteria in non-obese subjects using metagenomic shotgun analysis. Stool metagenome analysis of 38 non-obese subjects was conducted (F0, $n = 25$; F2-3, $n = 13$). **a** Abundances of fibrosis-related taxa are depicted for clarity. Taxa enriched in subjects with fibrosis stage 0 (top) and fibrosis stage of 2 or 3 (bottom). Statistical analysis was performed using a two-sided nonparametric Mann-Whitney test. The P values are as follows: (top) $***P = 0.0009$, $*P = 0.0113$, $*P = 0.0124$, $**P = 0.0024$, $*P = 0.0124$, $**P = 0.0051$, $**P = 0.0049$, $**P = 0.0070$, $*P = 0.0451$, $**P = 0.0041$, $**P = 0.0011$, $***P < 0.0001$, and $**P = 0.0022$; (bottom) $*P = 0.0294$, $*P = 0.0346$, $*P = 0.0136$, $*P = 0.0163$, $*P = 0.0411$, $*P = 0.0164$, $**P = 0.0050$, $*P = 0.0112$, and $*P = 0.0336$. **b** Multivariate associations between specific gut-microbiome components and fibrosis severity. Arcsine-root transformed abundances of bacteria were regressed against age, sex, and BMI. Only significant coefficients ($P < 0.05$) are depicted, and the color inside the box represents the enriched fibrosis group. **c** Heatmap showing the abundances of microbial genes related to bile acid metabolism pathway (left). The box plots indicate the median, 25th to 75th percentiles (boxes), and 10th to 90th percentiles (whiskers) (right) ($*P = 0.0124$ and $*P = 0.0237$). **d** Network profiles between microbial taxa and stool metabolite components. Coefficients among genus, species-level bacteria, and stool metabolite components were calculated by SparCC. Coefficients ($P < 0.05$) are depicted using Cytoscape. The P -value for each coefficient was obtained by bootstrapping the dataset 500 times and applying SparCC to each of those 500 datasets. The solid line (orange) and dotted line (gray) indicate positive and negative correlations, respectively. The shape of the node denotes the components used in this study (ellipse: microbiome, diamond: stool bile acids, and round rectangle: SCFAs) and the color indicates the degree of correlation with fibrosis severity. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$. Source data are provided as a Source data file.

non-obese subjects are replicated in a Western cohort with the potential of the microbiome-based marker to diagnose fibrosis in non-obese subjects with NAFLD.

Species of *Megamonas* and *Ruminococcus* are identified by metagenome sequencing. To identify fibrosis-related species consisting of *Veillonellaceae* and *Ruminococcaceae*, metagenome analysis of stool samples collected from 38 non-obese subjects (F0,

$n = 25$; F2-3, $n = 13$) was further conducted. Consistent with the results of 16S rRNA gene amplicon sequencing, *Veillonellaceae* and *Ruminococcaceae* were the most significant families related to fibrosis in both univariate and multivariate analyses of metagenome data (Fig. 7a, b). At the genus and species levels, *Ruminococcus bromii*, *Megamonas hypermegale*, and *M. funiformis* were the most significant taxa related to fibrosis severity in non-obese subjects. We also assessed the abundances of microbial genes related to primary and secondary bile acid metabolism (Fig. 7c).

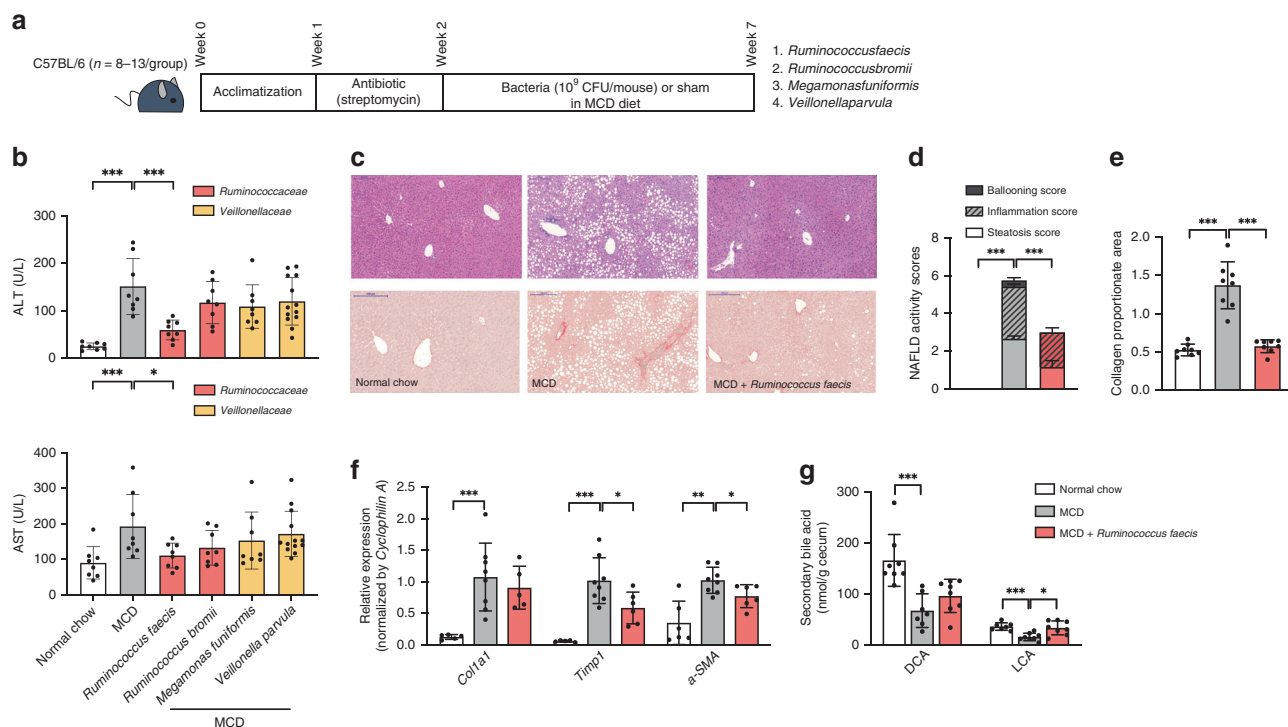


Fig. 8 Effect of fibrosis-associated bacteria on liver damage induced by an MCD diet. Mice were acclimated for 1 week on a standard chow diet. Then, they were treated with streptomycin (1 g/L) in drinking water for colonization of four fibrosis-related bacteria. Following 5 weeks, the mice were given daily 200 μ l of either bacteria (10^9 CFU/mouse in PBS) or sham in an MCD diet. **a** Scheme of the animal experiment. **b** Effects of the MCD diet and bacteria on serum ALT and AST levels (ALT, $***P = 0.0002$ and $***P = 0.0003$; AST, $***P = 0.0047$ and $*P = 0.0281$) $n = 8$ for normal chow, MCD, *R. faecis*, *R. bromii*, and *M. funiformis* group) and $n = 13$ for *V. parvula* group. **c** Representative images of *Ruminococcus faecis*-treated liver tissues stained with H&E and Sirius red. Scale bar indicates 200 μ m. **d** Comparison of histological NAFLD activity scores calculated on H&E stained liver tissues ($***P < 0.0001$ and $***P = 0.0006$; $n = 12$ for all groups). **e** Comparison of collagen proportionate areas measured on Sirius red-stained liver tissues ($***P = 0.0002$ and $***P = 0.0002$; $n = 8$ for all groups). **f** Relative fibrogenic mRNA expression of liver harvested from *Ruminococcus faecis*-treated mice ($**P = 0.0016$, $**P = 0.0016$, $*P = 0.0293$, $**P = 0.0047$, and $*P = 0.0356$; $n = 5-6$ for normal chow and MCD + *Ruminococcus faecis*, $n = 8$ for MCD). **g** Comparison of secondary bile acids levels measured in the cecum of *Ruminococcus faecis*-treated mice ($***P = 0.0003$, $***P = 0.0006$, and $*P = 0.0104$; $n = 7-8$ for normal chow and MCD, $n = 8$ for MCD + *Ruminococcus faecis*). The bar graphs indicate the means with SDs. Statistical analysis was performed using a nonparametric Kruskal-Wallis test with Dunn’s multiple comparison test or a two-sided Mann-Whitney test. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

The abundances of genes encoding bile salt hydrolase (*bsh*) and 7 α -hydroxy-3-oxochol-4-en-24-oyl-CoA dehydrogenase (*baiCD*) were significantly downregulated in non-obese subjects with significant fibrosis, which might be associated with increasing amounts of total conjugated bile acids and unconjugated primary bile acids. Major contributors to *bsh* were species belonging to *Ruminococcaceae*, *Lachnospiraceae*, and *Eubacteriaceae* (Supplementary Fig. 11).

The interaction between fibrosis-related bacterial species and stool metabolites was also revealed by network analysis based on the metagenome data of non-obese NAFLD subjects (Fig. 7d). The distinct co-occurrence pattern and strong correlation between key bacteria and metabolites were reconstructed; *R. bromii*, *Faecalibacterium prausnitzii*, and *Roseburia intestinalis* were inversely correlated with fibrosis severity and primary bile acid level, and *Megamonas* spp. showed a significant, positive correlation with primary bile species and UDCA, along with increasing fibrosis severity. However, *Ruminococcus gnavus*, a member of *Lachnospiraceae* family, was positively associated with primary bile acids and *Megamonas* species.

Administration of *Ruminococcus faecis* alleviates liver damage in NAFLD mouse models. To identify the protective or worsening effect on liver damage, we administered four representative species-level bacteria belonging to *Ruminococcaceae* and

Veillonellaceae-Ruminococcus faecis (*R. faecis*), *R. bromii*, *Megamonas funiformis* (*M. funiformis*), and *Veillonella parvula* (*V. parvula*) to C57BL/6 mice fed methionine- and choline-deficient (MCD) diets for 5 weeks (Fig. 8a). We demonstrated that *R. faecis* feeding decreased serum ALT and AST levels compared to sham feeding (Fig. 8b). No worsening effect was found in mice given *M. funiformis* and *V. parvula*. An alleviating effect of *R. faecis* on fibrosis was shown with H&E and Sirius red staining (Fig. 8c), and the histological severity of NAFLD induced by an MCD diet was significantly regressed in mice fed *R. faecis* (Fig. 8d, e). The mRNA expression of fibrogenic genes (*Col1a1*, *Timp1*, and *α-SMA*) was also downregulated in mice treated with *R. faecis* compared to untreated control mice fed an MCD diet (Fig. 8f). In parallel with the changes in biochemical and histological liver injury markers, the cecal levels of secondary bile acids (DCA and LCA) were also decreased by an MCD diet and increased by *R. faecis* treatment (Fig. 8g). To confirm the alleviating effect of *R. faecis* on liver damage in other mouse NAFLD models, we used a choline-deficient, L-amino acid-defined, high-fat diet (CDAHFD) model which prevents body weight loss in mice and shows no insulin resistance, and a genetic leptin-deficient (*db/db*) model, which develops spontaneously diabetes and fatty liver with insulin resistance. In both models, *R. faecis* administration decreased ALT and AST levels (Supplementary Fig. 12). However, the liver ratio against body weight decreased only in *db/db* mice.

Nevertheless, the fasting insulin levels in serum and insulin resistance measured by ipGTT in *db/db* mice were not affected by *R. faecis* treatment.

In search of the relationship between administered taxa and changes in bile acid or SCFA composition, we further analyzed the compositions of bile acid and SCFA (Fig. 8g, Supplementary Fig. 13, and Supplementary Table 7). We observed that the cecal levels of LCA and DCA were decreased by an MCD diet and were normalized by *R. faecis* treatment. However, an increase in the cecal levels of SCFAs was not observed by *V. parvula* treatment, rather a slight decrease in the level of propionate was observed in the cecum. These results indicated that intestinal bacteria could affect the regression of NAFLD in an insulin-independent manner, supporting our human-associated data found in non-obese NAFLD subjects.

Discussion

In our biopsy-proven Asian NAFLD cohort, we observed significant differences in the changes of gut microbiota according to fibrosis severity between non-obese and obese subjects. Interestingly, alterations in diversity with worsening fibrosis severity were observed only in non-obese subjects. Univariate and multivariate analyses using this cohort allowed us to focus on several microbes as potential targets to diagnose significant fibrosis. Furthermore, total bile acid levels, specifically primary bile acids and UDCA, and propionate levels in stool samples increased with worsening fibrosis severity in non-obese subjects. Utilizing these observations, we suggest that the combination of gut microbiota and stool metabolites might be a more robust diagnostic tool for diagnosing significant fibrosis in non-obese subjects with NAFLD. Moreover, the changes in fibrosis-related microbiota in non-obese subjects were validated using an independent Western NAFLD cohort. To confirm our association and demonstrate the causality, we performed metagenomic shotgun sequencing of 38 non-obese subjects and chose four bacterial species for animal validation. We administered four bacteria to MCD diet-induced NAFLD mice. *R. faecis* had a protective effect on liver damage in MCD diet-, CDAHFD diet-, and *db/db*-induced NAFLD mouse models.

In the current study, we used MaAsLin¹⁷ to identify NAFLD/fibrosis-specific bacteria. An important factor that is usually overlooked in microbiome studies is the removal of confounding effects, which can significantly influence microbiome data¹⁷. In addition, network analysis based on co-occurrence and microbial interactions supported distinct microbiome signatures in non-obese NAFLD patients identified by correlation and multivariate analyses. However, such notable microbiome changes were not related to host genetic traits that are known to be associated with NAFLD. The lack of the associations between genetic variants and bacterial taxa extends the previous study that reported the importance of environmental factors over genetic traits in shaping the human intestinal microbiota¹⁹.

Among three major fibrosis-related taxa, *Enterobacteriaceae* did not remain significant after adjustment for BMI, DM, and cirrhosis. *Enterobacteriaceae* was clustered by the presence of DM, which was consistent with the previous study²⁰. Although *Enterobacteriaceae* was associated with DM and the use of anti-diabetic medications, the overgrowth of Proteobacteria, including *Enterobacteriaceae*, in severe liver disease has been commonly reported^{11,15}.

The *Veillonellaceae* family, including six genera of Gram-negative bacteria (*Megasphaera*, *Veillonella*, *Dialister*, *Allisonella*, *Anaeroglobus*, and *Negativicoccus*)²¹, is known as a group of propionate-producing bacteria that utilize lactate as a substrate²². We first hypothesized that the enrichment of *Veillonellaceae*

produces more propionates, which are absorbed into the liver, ultimately leading to the establishment of NAFLD. However, the administration of *M. funiformis* and *V. parvula* did not worsen liver damage and the cecal propionate levels were not increased in a mouse validation study. Taken together, the strong associations of *Veillonellaceae* with fibrosis severity and stool metabolites in our human studies were not attributed to the causal effects of the tested strains, but could be applied to a diagnostic marker for fibrosis in non-obese subjects with NAFLD.

Veillonellaceae has been associated with chronic liver disease, including cirrhosis and primary sclerosing cholangitis^{14,15,23,24}. In an American cohort, Caussy et al. reported that *Megasphaera* was enriched in a cirrhosis group¹⁴. In addition, Qin et al. reported that *Veillonella* was enriched in cirrhotic patients along with *Streptococcus*, in a Chinese cohort, and they hypothesized that altered bile acid production in cirrhotic patients enables *Veillonella* to originate from the mouth and colonize in the gut environment¹⁵. However, Bajaj et al.¹³ reported a decrease in *Veillonellaceae* in Western cirrhotic subjects (BMI 29.2–30.8) compared with BMI-matched healthy controls (BMI 23.9–35.1). The discrepancy in the enrichment of *Veillonellaceae* in cirrhotic subjects between previous studies might be attributed to several confounding factors in cirrhotic subjects such as various etiologies of cirrhosis, use of rifaximin and beta-blockers, the setting of decompensation, and other complications, including infections and metabolic/cardiovascular disorders. Therefore, we further evaluated and confirmed the significance of associations between the identified bacterial taxa and fibrosis severity and the diagnostic power of six markers in a population without cirrhotic subjects. Given the controversial results of association studies, the confounding effects of cirrhosis and other diseases, and the lack of a causal effect on fibrosis progression in the current study, *Veillonellaceae* could be used only as a diagnostic marker in a selected population, such as non-obese NAFLD patients.

In contrast, *Ruminococcaceae* acts as a key component of the microbiota in healthy people by maintaining homeostasis of the gut microenvironment¹⁵. *Ruminococcaceae* was gradually depleted with worsening fibrosis severity in combined and non-obese subjects with NAFLD, as also noted in the previous studies^{9,11,13}. Two different studies compared gut-microbiome profiles according to obesity status in Latin American NASH and Asian NAFLD cohorts, respectively^{25,26}. Similar to our results, the relative abundance of *Ruminococcus* was significantly reduced in non-obese NASH subjects compared to non-obese controls. However, Boursier et al. reported that the abundance of *Ruminococcus* was higher in subjects with significant fibrosis¹². These discordant results could be partly attributed to the fact that Boursier et al. compared gut-microbiome profiles using only obese subjects (average BMI; 31), suggesting that non-obese subjects might have disparate gut-microbiome compositions compared to obese subjects.

Among the *Ruminococcaceae* family, *F. prausnitzii* was already reported to regulate the hepatic fat content and compositions of lipid species and to reduce adipose tissue inflammation in high-fat-fed mice²⁷. To confirm the causal effect of other *Ruminococcus* genus on NAFLD, we administered *R. faecis* and *R. bromii* isolated from human stools²⁸ in an MCD-diet-induced NAFLD model. *R. faecis* had a protective effect on liver damage, implicating the species specificity of *Ruminococcus* strains and that some intestinal bacteria have the potential to be utilized as a fibrosis marker, as well as a therapeutic target.

Herein, the synthesis of bile acids was elevated with worsening fibrosis severity, and it was observed only in non-obese subjects. Previous studies have also reported the elevated synthesis of bile acids in NAFLD patients^{12,29}. Bile acids are crucial for the emulsification of lipids and for maintaining hepatic glucose

homeostasis, and they exert a strong antimicrobial effect owing to their detergent properties, which likely change the environment of gut microbiota by inhibiting the growth of susceptible bacteria. Although bile acid tolerance shows a strain-specific feature, Gram-negative bacteria are generally more resistant to the antimicrobial effects of bile acids than Gram-positive bacteria^{6,30}.

In the current study, *Ruminococcaceae* was depleted in non-obese subjects with significant fibrosis, suggesting the modulation of gut microbiome by altered bile acid composition. We also observed that microbial genes metabolizing bile acids were significantly altered in non-obese subjects with fibrosis. Notably, *bsh* and *baiCD* genes were downregulated and major contributors to the *bsh* gene were identified as species belonging to *Ruminococcaceae*, *Lachnospiraceae*, and *Eubacteriaceae*.

In an MCD-diet-induced NAFLD model, 5-week *R. faecis* and *R. bromii* treatment increased the cecal levels of secondary bile acids, suggesting a potentially causal role of *Ruminococcus* in the synthesis of secondary bile acids in a direct or indirect way.

Sinha et al.³¹ reported that *Ruminococcaceae* were depleted and stool secondary bile acid levels were reduced in patients with ulcerative colitis, and that the supplementation of LCA ameliorated inflammation in the animal colitis model, implicating the potential therapeutic role of *Ruminococcaceae* in gut microenvironment.

We acknowledge several limitations of this study. First, the cross-sectional study allowed us to compare the stool microbiota and metabolites according to disease severity but did not explain the exact reason for changes in gut microbiota. Therefore, further longitudinal studies using a prospective non-obese NAFLD cohort are warranted to accurately predict non-obese NASH with significant fibrosis using the stool taxa-metabolite markers identified in the current study. Second, a substantial number of patients had DM in this cohort. Both NAFLD and DM are closely interrelated since they share insulin resistance as a common mechanism of pathogenesis³². Anti-diabetic medications such as metformin might exert a confounding effect on microbiome analysis, interfering with the discrimination of fibrosis-specific microbes. Although we adjusted for DM and anti-diabetic medications to distinguish the fibrosis-specific microbiome, further research using a non-diabetic NAFLD cohort is required. Third, the diagnosis and classification methods of fibrosis in the validation cohort (i.e., elastography) were different from those in our cohort (i.e., biopsy). Caussy et al. compared gut microbiota among the NAFLD-related cirrhosis, NAFLD without advanced fibrosis, and no-NAFLD control groups¹⁴. Detailed information about granular fibrosis stages and stool metabolites was not available in this public dataset. Nevertheless, our validation results were sufficient to confirm the differences in gut-microbiome changes according to fibrosis severity between non-obese and obese subjects. However, external validation using the combined microbiome-metabolite marker in additional Western studies needs to be conducted, considering the possible overfitting error caused by a number of explanatory variables in our study. Fourth, after stratification by obesity and fibrosis stage, each subgroup ended up with a small number of patients, which might lead to an insufficient statistical power prone to a type II error. Finally, we could not identify any clues about clear differences according to fibrosis severity in obese NAFLD, unlike in non-obese NAFLD. This finding might be attributed to the stronger driving force of obesity regarding changes in microbial community structure and their metabolites, such as bile acid composition, interfering with the identification of fibrosis-specific microbial markers in obese subjects with NAFLD.

Despite these limitations, the results from our well-characterized, biopsy-proven NAFLD cohort emphasize not only the importance of the gut microbiota as a risk factor accounting for the

pathogenesis of non-obese NAFLD but also the diagnostic and potentially therapeutic implications of the microbiome-metabolite combination as a non-invasive biomarker for diagnosing significant fibrosis in patients with non-obese NAFLD.

Methods

Human subjects. We utilized the data from the 'Boramee NAFLD cohort (NCT 02206841)' study^{5,33}. One hundred seventy-one subjects with biopsy-proven NAFLD and 31 no-NAFLD subjects were included in this study. Subjects with histologically confirmed NAFLD and BMI < 25 kg m⁻² were classified as non-obese NAFLD subjects. Clinical, metabolic, biochemical, and epidemiological characteristics are described in Supplementary Tables 1–5. This study was performed in accordance with the ethical guidelines of the 1975 Declaration of Helsinki for the participation of human subjects and was approved by the Institutional Review Board of Boramee Medical Center (IRB No. 26-2017-48). Written informed consent was obtained from all of the study subjects.

Inclusion and exclusion criteria. Briefly, we enrolled eligible subjects from January 2013 to February 2017 prospectively. The inclusion criteria were as follows: (1) adults at least 18 years old; (2) ultrasonographic findings confirming fatty infiltration of the liver; and (3) unexplained elevated ALT levels within the past 6 months. Subjects who met any of the following criteria were excluded: (1) hepatitis B or C virus infection; (2) autoimmune hepatitis, primary biliary cholangitis or primary sclerosing cholangitis; (3) gastrointestinal cancers or hepatocellular carcinoma; (4) drug-induced steatosis or liver injury; (5) Wilson disease or hemochromatosis; (6) excessive alcohol consumption (men: >210 g week⁻¹, women: >140 g week⁻¹); (7) antibiotics use within the prior month; (8) diagnosis of malignancy (<5 yr); (9) human immunodeficiency virus infection; and (10) chronic disorders associated with lipodystrophy or immunosuppression.

Non-obese and obese controls included subjects without any suspicion of NAFLD who received liver biopsy: (1) during the evaluation for living donor liver transplantation or (2) during the characterization of solid liver masses that were suspected to be hepatic adenomas or focal nodular hyperplasia based on imaging studies (1 subject in the current cohort with focal nodular hyperplasia confirmed by liver biopsy)⁵. Details of the study subjects and host genotyping are presented in the Supplementary Methods.

Liver histology. Liver histology was assessed using the NASH CRN histological scoring system by a single liver pathologist. NAFLD was defined as the presence of ≥5% macrovesicular steatosis based on histological examination. NASH was determined based on an overall pattern of histological hepatic injury consisting of steatosis, lobular inflammation, and ballooning according to the criteria of Brunt et al.^{34,35} We also scored steatosis, lobular inflammation, and ballooning according to the NAFLD activity scoring system³⁶. Fibrosis severity was evaluated according to the criteria of Kleiner et al.³⁶.

Microbiome analysis using 16S rRNA sequencing. DNA in stool samples was extracted using a QIAamp DNA Stool Mini Kit (Qiagen, Hilden, Germany). Sequencing targeting the V4 region of the 16S rRNA was performed using the MiSeq platform (Illumina, San Diego, CA, USA), and further processing of raw sequencing data was performed using the QIIME pipeline (v. 1.8.0)³⁷. Details of data processing are presented in the Supplementary Methods.

Metagenomic shotgun sequencing. DNA in stool samples was extracted using a QIAamp DNA Stool Mini Kit (Qiagen). Sequencing libraries were prepared using an Illumina Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA). Sequencing was performed on an Illumina NextSeq 500 platform (Illumina). Taxonomic and functional profiles of the microbiome were analyzed using MetaPhlan2 (v. 2.6.0)³⁸ and HUMAnN2 (v. 0.11.0)³⁹. Details of the analysis are described in the Supplementary Methods.

Stool metabolite measurements using the GC-FID and Q-TOF system. Stool bile acid profile was assessed using a Q-TOF mass spectrometer (Waters Micro-mass Technologies, Manchester, UK), and SCFA was measured by using an Agilent Technologies 7890A GC system (Agilent Technologies, Santa Clara, CA, USA) according to David's method⁴⁰. Details of metabolite analysis are presented in the Supplementary Methods.

Bioinformatics analysis and statistical tests. For alpha diversity, the OTU table was rarefied to 12,000 sequences per sample, and the Shannon index was measured in QIIME. Nonparametric multi-dimensional scaling (NMDS) plots were depicted using the Vegan package (v. 2.5-6) in R (v.3.5.0)⁴¹, and the distance was measured using Bray-Curtis methods. Statistical significance between groups was estimated using the adonis function. Multivariate association analysis using microbiome data was performed using MaAsLin (Galaxy v. 1.0)⁴⁷ for the identification of specific taxa associated with the host phenotype without the influence of other metadata. Herein, we adjusted for age, sex, and BMI or clinical factors as fixed variables.

Associations in which the *P*-value adjusted by Benjamini and Hochberg's false discovery rate was <0.25 (default setting) were considered significant. The AUC was measured using SPSS software (v. 25.0) (SPSS Inc., Armonk, NY, USA). For data with non-normal distribution, outliers were removed using the ROUT method⁴² (*Q* = 1%) and statistical comparisons were performed with the Kruskal–Wallis test or the Mann–Whitney test using GraphPad Prism software (v. 8.4.3) (GraphPad Software, San Diego, CA, USA). Details are presented in Supplementary Methods.

Validation of the data using a Western NAFLD cohort. To validate our results, public dataset reporting microbiome-based biomarkers for diagnosing NAFLD-related cirrhosis was obtained from qiita.microbio.me (QIITA Study 11635)¹⁴. A total of 168 subjects with metadata on advanced fibrosis and BMI were used for this study. An obese state was defined by the BMI cut-off value (obese, BMI ≥30; non-obese, BMI <30) currently used in American and European populations^{43,44}. Each group was classified into two subgroups according to the presence or absence of advanced fibrosis.

Interventive animal study. *R. faecis* (JCM no. 15917)²⁸ and *V. parvula* (DSM no. 2008) were distributed from the Korean Collection for Type Cultures (KCTC). *R. bromii* was distributed from the American Type Culture Collection (ATCC, no. 27255). *M. funiformis* was isolated from healthy Korean adult feces. Bacteria (10⁹ CFU/mouse) were gavaged orally in C57BL/6 mice with an MCD or a CDAHFD diet. In *db/db* mice, *R. faecis* was gavaged orally with a normal chow diet. All of the animal procedures were approved by the Institutional Animal Care and Use Committee of the Seoul National University. The mice were housed in a conventional animal facility according to university guidelines at a light/dark cycle of 12 h, temperature range of 21–24 °C, and humidity range of 40–50%. Details are presented in the Supplementary Methods.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The V4 16S rDNA and shotgun metagenome sequence datasets obtained from this study have been deposited in the European Nucleotide Archive databases with the accession codes ERP109777 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB27662>). All other metadata that support the findings of this study are available from the corresponding author upon reasonable request. A validation cohort was obtained from qiita.microbio.me (QIITA Study 11635, <https://www.ebi.ac.uk/ena/browser/view/PRJEB28350>). Source data are provided with this paper.

Received: 10 September 2019; Accepted: 9 September 2020;
Published online: 05 October 2020

References

- Lomba, R. & Sanyal, A. J. The global NAFLD epidemic. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 686–690 (2013).
- Chalasani, N. et al. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Gastroenterological Association, American Association for the Study of Liver Diseases, and American College of Gastroenterology. *Gastroenterology* **142**, 1592–1609 (2012).
- Leung, J. C. et al. Histological severity and clinical outcomes of nonalcoholic fatty liver disease in nonobese patients. *Hepatology* **65**, 54–64 (2017).
- Kim, D. & Kim, W. R. Nonobese fatty liver disease. *Clin. Gastroenterol. Hepatol.* **15**, 474–485 (2017).
- Koo, B. K. et al. Additive effects of PNPLA3 and TM6SF2 on the histological severity of non-alcoholic fatty liver disease. *J. Gastroenterol. Hepatol.* **33**, 1277–1285 (2018).
- Begley, M., Gahan, C. G. & Hill, C. The interaction between bacteria and bile. *FEMS Microbiol. Rev.* **29**, 625–651 (2005).
- Cani, P. D. et al. Metabolic endotoxemia initiates obesity and insulin resistance. *Diabetes* **56**, 1761–1772 (2007).
- Dawes, E. & Foster, S. M. The formation of ethanol in *Escherichia coli*. *Biochim. Biophys. Acta* **22**, 253–265 (1956).
- Zhu, L. et al. Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH. *Hepatology* **57**, 601–609 (2013).
- Wang, Z. et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**, 57–63 (2011).
- Lomba, R. et al. Gut microbiome-based metagenomic signature for non-invasive detection of advanced fibrosis in human nonalcoholic fatty liver disease. *Cell Metab.* **25**, 1054–1062 (2017).
- Boursier, J. et al. The severity of nonalcoholic fatty liver disease is associated with gut dysbiosis and shift in the metabolic function of the gut microbiota. *Hepatology* **63**, 764–775 (2016).
- Bajaj, J. S. et al. Altered profile of human gut microbiome is associated with cirrhosis and its complications. *J. Hepatol.* **60**, 940–947 (2014).
- Causy, C. et al. A gut microbiome signature for cirrhosis due to nonalcoholic fatty liver disease. *Nat. Commun.* **10**, 1406 (2019).
- Qin, N. et al. Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59–64 (2014).
- He, Y. et al. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat. Med.* **24**, 1532–1535 (2018).
- Morgan, X. C. et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* **13**, R79 (2012).
- Qin, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
- Lim, M. Y. et al. The effect of heritability and host genetics on the gut microbiota and metabolic syndrome. *Gut* **66**, 1031–1038 (2016).
- Wu, H. et al. Metformin alters the gut microbiome of individuals with treatment-naïve type 2 diabetes, contributing to the therapeutic effects of the drug. *Nat. Med.* **23**, 850–858 (2017).
- Marchand, H. & Jumas-Bilak, E. In *The Prokaryotes* (Springer, 2014).
- Kara, D., Luppens, S. B. & ten Cate, J. M. Differences between single- and dual-species biofilms of *Streptococcus mutans* and *Veillonella parvula* in growth, acidogenicity and susceptibility to chlorhexidine. *Eur. J. Oral. Sci.* **114**, 58–63 (2006).
- Kummen, M. et al. The gut microbial profile in patients with primary sclerosing cholangitis is distinct from patients with ulcerative colitis without biliary disease and healthy controls. *Gut* **66**, 611–619 (2016).
- Chen, Y. et al. Characterization of fecal microbial communities in patients with liver cirrhosis. *Hepatology* **54**, 562–572 (2011).
- Duarte, S. M. B. et al. Gut microbiome composition in lean patients with NASH is associated with liver damage independent from caloric intake: a prospective pilot study. *Nutr. Metab. Cardiovasc. Dis.* **28**, 369–384 (2017).
- Wang, B. et al. Altered fecal microbiota correlates with liver biochemistry in nonobese patients with non-alcoholic fatty liver disease. *Sci. Rep.* **6**, 32002 (2016).
- Munukka, E. et al. *Faecalibacterium prausnitzii* treatment improves hepatic health and reduces adipose tissue inflammation in high-fat fed mice. *ISME J.* **11**, 1667–1679 (2017).
- Kim, M. S., Roh, S. W. & Bae, J. W. *Ruminococcus faecis* sp. nov., isolated from human faeces. *J. Microbiol.* **49**, 487–491 (2011).
- Jiao, N. et al. Suppressed hepatic bile acid signalling despite elevated production of primary and secondary bile acids in NAFLD. *Gut* **67**, 1881–1891 (2018).
- Tian, Y. et al. The microbiome modulating activity of bile acids. *Gut Microbes* **5**, 1–18 (2020).
- Sinha, S. R. et al. Dysbiosis-induced secondary bile acid deficiency promotes intestinal inflammation. *Cell Host Microbe* **27**, 659–670 (2020).
- Anstee, Q. M., Targher, G. & Day, C. P. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 330–344 (2013).
- Koo, B. K. et al. Sarcopenia is an independent risk factor for non-alcoholic steatohepatitis and significant fibrosis. *J. Hepatol.* **66**, 123–131 (2017).
- Brunt, E. M., Janney, C. G., Di Bisceglie, A. M., Neuschwander-Tetri, B. A. & Bacon, B. R. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. *Am. J. Gastroenterol.* **94**, 2467–2474 (1999).
- Brunt, E. M., Kleiner, D. E., Wilson, L. A., Belt, P. & Neuschwander-Tetri, B. A. Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. *Hepatology* **53**, 810–820 (2011).
- Kleiner, D. E. et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313–1321 (2005).
- Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
- Truong, D. T. et al. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12**, 902–903 (2015).
- Franzosa, E. A. et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* **15**, 962–968 (2018).
- David, L. A. et al. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
- Oksanen, J. et al. The vegan package. *Community Ecology Package* **10**, 719 (2007).
- Motulsky, H. J. & Brown, R. E. Detecting outliers when fitting data with nonlinear regression—a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinforma.* **7**, 123 (2006).
- Organization, W. H. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Consultation. *WHO Technical Report Series* (1995).
- Hedley, A. A. et al. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999–2002. *JAMA* **291**, 2847–2850 (2004).

Acknowledgements

This work was supported by a multidisciplinary research grant-in-aid from the Seoul Metropolitan Government Seoul National University (SMG-SNU) Boramae Medical Center (02-2017-8), 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 (H I17C0912), and the National Research Foundation of Korea (NRF) (Grant No. NRF-2018R1A2A1A05078258).

Author contributions

G.L. and H.J.Y. conducted the human microbiome analysis and wrote the manuscript. G.L., H.J.Y., S.P., and H.K. conducted the animal experiment. G.L. and J.Y. conducted the metabolite analysis. S.K.J., D.H.L., S.L., and W.K. collected specimens from Korean NAFLD subjects and conducted clinical analysis. J.H.P. and J.H.K. analyzed the liver histology. H.J.Y., W.K., J.S.B., and G.P.K. designed the study, analyzed data, and reviewed the paper. All authors read and commented on the manuscript.

Competing interests

G.P.K. is a founder of KoBioLabs, Inc., a company characterizing the role of host–microbiome interaction in chronic diseases. The other authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41467-020-18754-5>.

Correspondence and requests for materials should be addressed to W.K. or G.P.K.

Peer review information *Nature Communications* thanks Vincent Wong and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020