



Disentangling the Progression of Non-alcoholic Fatty Liver Disease in the Human Gut Microbiota

Tianjiao Wang¹, Xue-Kun Guo¹ and Huji Xu^{1,2,3*}

¹ School of Medicine, Tsinghua University, Beijing, China, ² Department of Rheumatology and Immunology, Shanghai Changzheng Hospital, The Second Military Medical University, Shanghai, China, ³ Peking-Tsinghua Center for Life Sciences, Tsinghua University, Beijing, China

OPEN ACCESS

Edited by:

Ying Wang,
Shanghai Institute of Immunology,
Shanghai Jiao Tong University School
of Medicine, China

Reviewed by:

Zheng Sun,
Brigham and Women's Hospital
and Harvard Medical School,
United States
Yeshi Yin,
Hunan University of Science
and Engineering, China
Grégory Dubourg,
IHU Méditerranée Infection, France

*Correspondence:

Huji Xu
huji_xu@tsinghua.edu.cn

Specialty section:

This article was submitted to
Microbial Immunology,
a section of the journal
Frontiers in Microbiology

Received: 22 June 2021

Accepted: 16 September 2021

Published: 13 October 2021

Citation:

Wang T, Guo X-K and Xu H (2021)
Disentangling the Progression
of Non-alcoholic Fatty Liver Disease
in the Human Gut Microbiota.
Front. Microbiol. 12:728823.
doi: 10.3389/fmicb.2021.728823

Gut microbiome dysbiosis has been known to be associated with all stages of non-alcoholic fatty liver disease (NAFLD), but questions remain about microbial profiles in progression and homogeneity across NAFLD stages. We performed a meta-analysis of three publicly shotgun datasets and built predictive models to determine diagnostic capacity. Here, we found consistently microbiome shifts across NAFLD stages, of which co-occurrence patterns and core sets of new biomarkers significantly correlated with NAFLD progression were identified. Machine learning models that are able to distinguish patients with any NAFLD stage from healthy controls remained predictive when applied to patients with other NAFLD stages, suggesting the homogeneity across stages once again. Focusing on species and metabolic pathways specifically associated with progressive stages, we found that increased toxic metabolites and decreased protection of butyrate and choline contributed to advanced NAFLD. We further built models discriminating one stage from the others with an average of 0.86 of area under the curve. In conclusion, this meta-analysis firmly establishes generalizable microbiome dysbiosis and predictive taxonomic and functional signatures as a basis for future diagnostics across NAFLD stages.

Keywords: liver cirrhosis, liver fibrosis, meta-analysis, microbiome, NAFLD

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD) is defined as the pathological accumulation of lipid droplets in >5% of hepatocytes (Sberna et al., 2018), developing from simple non-alcoholic fatty liver (NAFL), progressing toward non-alcoholic steatohepatitis (NASH), which can also present with liver fibrosis, the main prognostic lesion for disease progression, and ultimately leading to cirrhosis or hepatocellular carcinoma (Fingas et al., 2016). It is rapidly becoming the most prevalent liver disease and also the most increasing cause of hepatocellular cancer and liver transplantation in Western countries. Liver biopsy is the diagnostic gold standard to assess the disease severity, yet it is an invasive, traumatic, inconvenient tool, making it unfeasible for disease screening, diagnosis, or examining progression in routine care. Early identification of the presence of the advanced stage of NAFLD using non-invasive form is a major unmet need in the field.

Evidence is accumulating that the gut microbiome is involved in the etiology of NAFLD, while few studies have focused specifically on microbiota signatures in NAFLD at species level

(Qin et al., 2014; Loomba et al., 2017; Hoyles et al., 2018) and even fewer have examined microbial composition as a function of NAFLD progression (Caussy et al., 2019). Qin et al. (2014) characterized the gut microbiome in liver cirrhosis, but this study involved diverse etiologies of cirrhosis (such as alcoholic liver disease, hepatitis B, and hepatitis C) and did not provide gut microbiome signatures that are specific to NAFLD-related cirrhosis. There are also studies comparing different stages of liver steatosis (Hoyles et al., 2018) or fibrosis severity (Boursier et al., 2016; Loomba et al., 2017; Shen et al., 2017), yet these stages are not exactly the same as the stage of disease progression. Caussy et al. (2019) explored patients with non-NAFLD, NAFLD without advanced fibrosis, or NAFLD cirrhosis, while most of the identified promising diagnostic signatures were unknown genera. Importantly, researches have mainly compared the differences between NAFLD patients and healthy controls or between mild stage and advanced stage of fibrosis, the generalizable microbiome dysbiosis across stages has been ignored. Thus, the alteration of the gut microbiome in NAFLD progression and the generality of microbiome dysbiosis across stages urgently need to be demonstrated.

Here, we collected currently available NAFLD shotgun metagenomic datasets (Boursier et al., 2016; Loomba et al., 2017; Shen et al., 2017), re-classified 107 patients with NAFL, NASH, fibrosis, or cirrhosis and 120 healthy controls according to clinical diagnosis, performed an integrated analysis combining all datasets, and assessed prediction accuracies of the gut microbiome for the detection of the different stages in NAFLD progression. Our study aims to find a panel of gut microbiome-derived biomarkers generally associated with NAFLD across stages or linking NAFLD progression, which help to develop a non-invasive diagnosis of NAFLD.

RESULTS

Gut Microbiome Changes in Non-alcoholic Fatty Liver Disease Progression

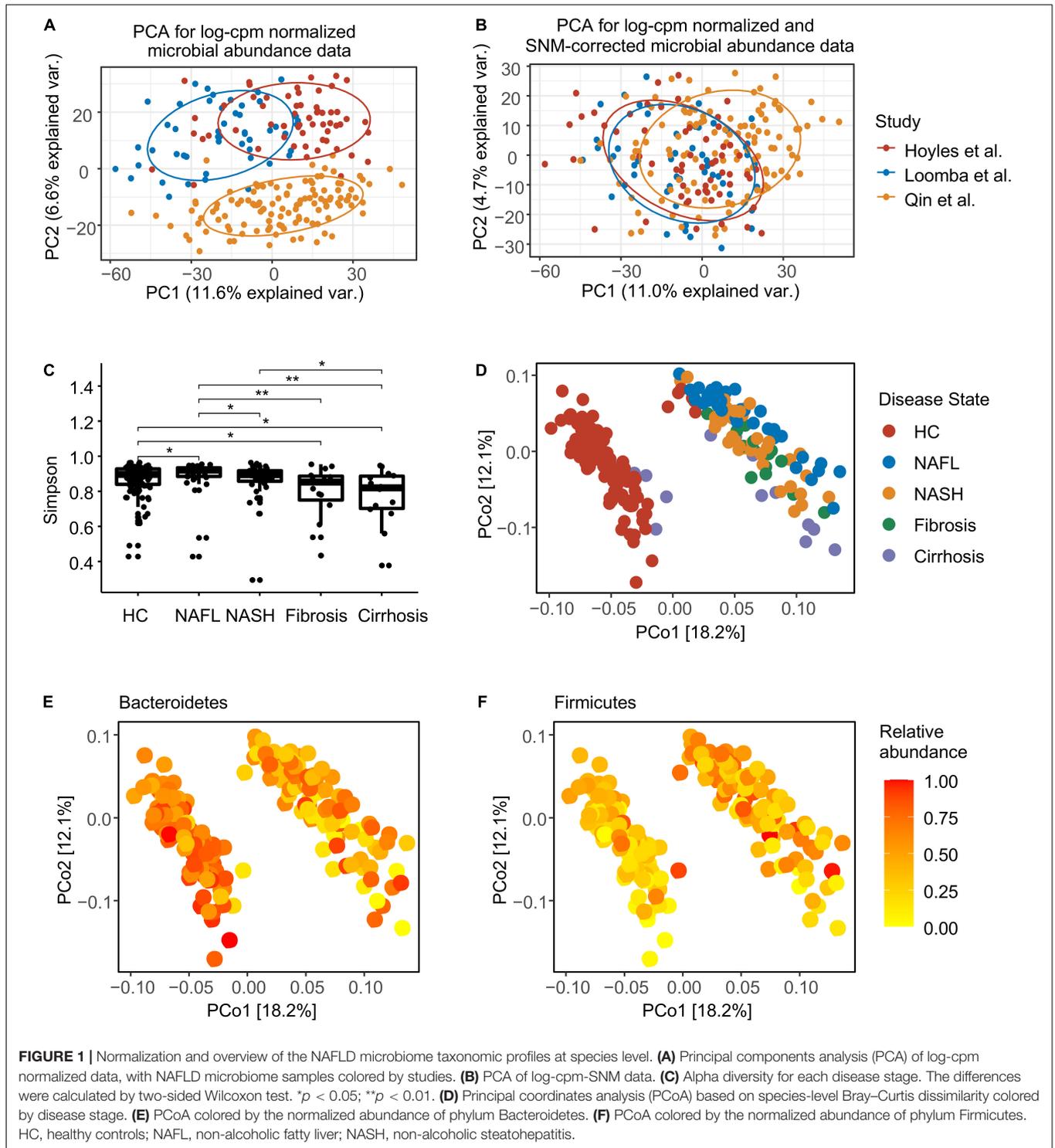
After filtering samples with incomplete diagnostic information, we considered 39 patients with NAFL, 39 patients with NASH, 15 patients with fibrosis, 14 patients with cirrhosis, and 120 healthy controls (Supplementary Table 1). In total, assembled sequences for 493 species (Supplementary Table 2), 1,718,123 gene families (Supplementary Table 3), and 432 metabolic pathways (Supplementary Table 4) were identified in the 227 samples. The combined metagenomic data showed substantial batch effects (Figure 1A), so we converted discrete taxonomical counts into log-counts per million (log-cpm) per sample, and performed supervised normalization (SNM) (Qin et al., 2012) to reduced batch effects (Figure 1B). We then contrasted the effect of disease-associated heterogeneity on microbiome composition with potential confounders [patient age, body mass index (BMI), and sex], and this analysis revealed BMI to have an impact on species composition as predominant as disease (Supplementary Figure 1A), since patients with NAFLD often

present with obesity. To address this issue, for the present study, a methodology was chosen that explicitly models BMI as a confounder in all applicable tests (Hothorn et al., 2006). Consistent with prior studies (Caussy et al., 2019; Ponziani et al., 2019), samples from individuals with more serious NAFLD stage had significantly lower Simpson alpha diversity (Figure 1C). The stage of health or disease contributed to the first axis of species-based principal coordinates (Figure 1D), and the most variation is driven by a trade-off between phylum Bacteroidetes and Firmicutes (Figures 1E,F and Supplementary Figure 1B).

Univariate Analysis of Species Associated With Non-alcoholic Fatty Liver Disease

At a meta-analysis FDR of 0.05, we identified 99 species, out of 261 species consistently detected across studies, to be associated with general NAFLD microbiome dysbiosis, of which 47 microbial species were identified to be significantly enriched in patients and 52 microbial species were identified to be significantly depleted (Figure 2A). The gut microbial community had consistent alteration patterns across different disease stages (Figure 2B and Supplementary Figure 2), and the majority of significant species were compatible identified as significant in individual disease stages (Figure 2B). Our results are easily reconciled with a model in which increased pathogenic microbes and a lack of protective microbes contribute to NAFLD development (Figure 2C). These species showed relatively low accuracy for cirrhosis, and this may be partially due to the low sample size.

We then focused on a core set of the 38 most significant markers ($FDR < 1 \times 10^{-3}$) for further analysis (Figure 2A). Among these, *Enterococcus faecalis* was previously identified enriched in liver fibrosis patients (Aron-Wisnewsky et al., 2020), and the rest were new microbial biomarkers. Strains enriched in NAFLD such as *Megasphaera unclassified*, *Streptococcus* sp. I G2, *Clostridium spiroforme*, *Coriobacteriaceae bacterium pH1*, *Turicibacter* sp. HGF1, *Veillonella* sp. 3 1 44, *Streptococcus sanguinis*, and *Streptococcus anginosus* negatively correlated with species enriched in healthy controls, including *Weissella cibaria*, *Saccharomyces unclassified*, *Lactobacillus vaginalis*, *Streptococcus* sp. GMD4S, *Lactobacillus crispatus*, *Coprococcus* sp. 8 2 54BFAA, *Coprococcus* sp. HPP0048, and *Klebsiella unclassified* (Figure 3A). All of the classified NAFLD-enriched species are well-known pathogenic bacterium; for example, *C. spiroforme* toxin shows cytotoxicity and causes enteric diseases in human (Uzal et al., 2018), *S. sanguinis* is notorious for a cause of infective endocarditis (Zhu et al., 2018), and *S. anginosus* plays a pathogenic role in cystic fibrosis (Asam and Spellerberg, 2014). The majority of the control-enriched species are beneficial bacterium, for example, *W. cibaria*, as a kimchi lactic acid bacteria, has the ability to prevent cancer (Kwak et al., 2014) and *L. crispatus* is known as a biomarker of the healthy vaginal tract (Lepargneur, 2016; Veščičík et al., 2020). Another cluster of strains enriched in NAFLD including *Rothia dentocariosa*, *Bifidobacterium dentium*, *Actinomyces* sp. ICM47, *E. faecalis*, *Streptococcus* sp. 2 1 36FAA, *Klebsiella*



oxytoca, and *Streptococcus intermedius* negatively correlated with species enriched in healthy controls, including *Alistipes* sp. JC136, *Parabacteroides* sp. ASF519, *Streptococcus infantarius*, and *Dialister succinatiphilus* (Figure 3A). Most of the NAFLD-enriched species are pathogenic; for example, *R. dentocariosa* is a well-known causative agent of dental plaques and periodontal

disease (Yang et al., 2007); *B. dentium* is the only single species of Bifidobacterium recognized as pathogenic (Meile et al., 2008); and *K. oxytoca* and *S. intermedius* were reported to be enriched in patients with angiocardopathy (Jie et al., 2017). To determine whether specific microorganisms were correlated with NAFLD progression, we assessed the abundance

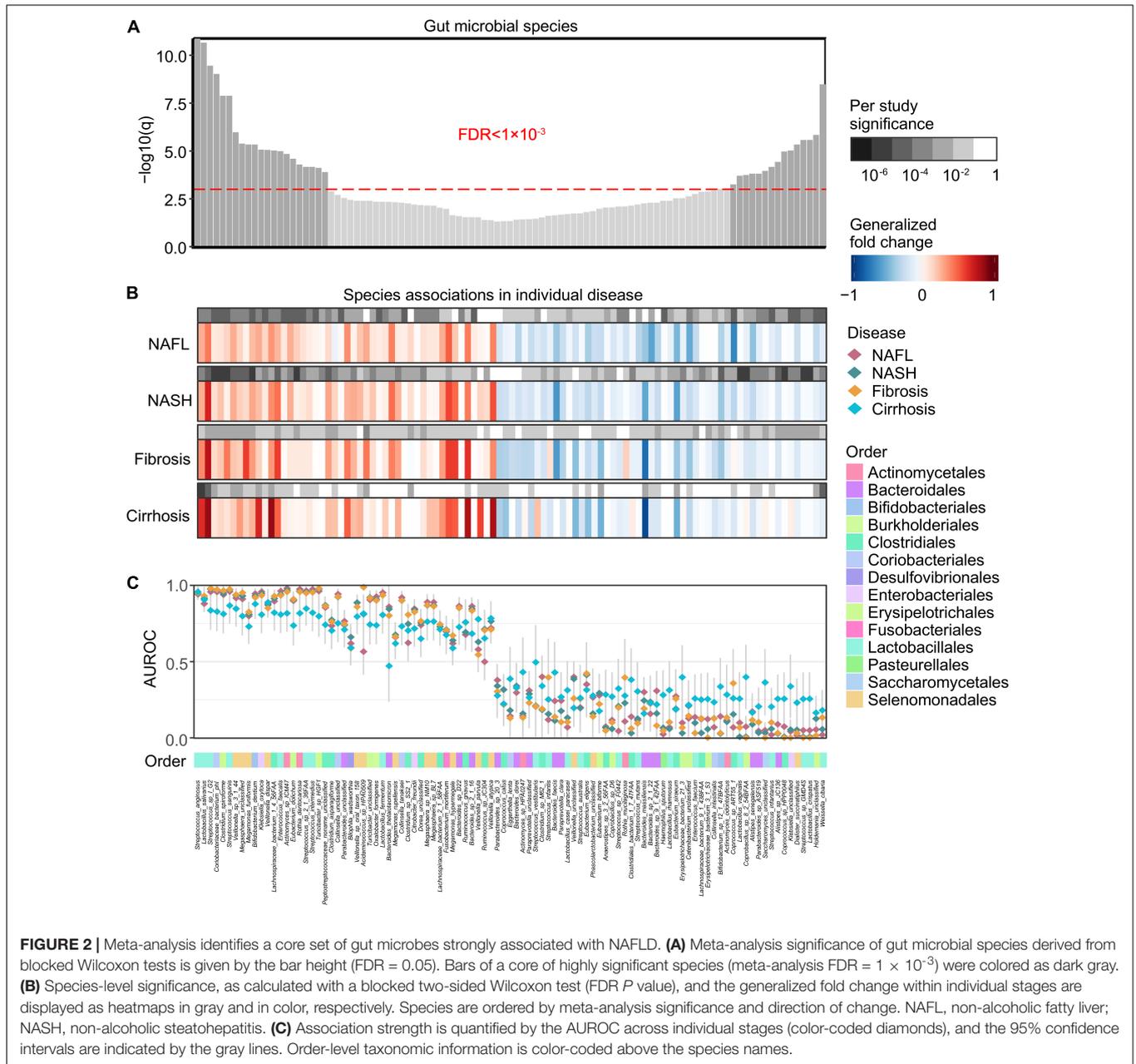


FIGURE 2 | Meta-analysis identifies a core set of gut microbes strongly associated with NAFLD. **(A)** Meta-analysis significance of gut microbial species derived from blocked Wilcoxon tests is given by the bar height (FDR = 0.05). Bars of a core of highly significant species (meta-analysis FDR = 1×10^{-3}) were colored as dark gray. **(B)** Species-level significance, as calculated with a blocked two-sided Wilcoxon test (FDR *P* value), and the generalized fold change within individual stages are displayed as heatmaps in gray and in color, respectively. Species are ordered by meta-analysis significance and direction of change. NAFL, non-alcoholic fatty liver; NASH, non-alcoholic steatohepatitis. **(C)** Association strength is quantified by the AUROC across individual stages (color-coded diamonds), and the 95% confidence intervals are indicated by the gray lines. Order-level taxonomic information is color-coded above the species names.

of prokaryotes across the four stages of NAFLD by partial Spearman’s rank-based correlation (pSRC). At the species level, *C. spiroforme*, *Turicibacter* sp., *E. dolichum*, *L. salivarius*, *M. funiformis*, *S. anginosus*, *S. sanguinis*, *Streptococcus* sp., *V. dispar*, and *Veillonella* sp. were significantly correlated with NAFLD progression (pSRC, FDR < 0.05, **Figure 3B**), whereas *A. senegalensis*, *Coprobacillus* sp., *Coprococcus* sp., *L. crispatus*, *L. vaginalis*, and *W. cibaria* were significantly anti-correlated with NAFLD progression (pSRC, FDR < 0.05, **Figure 3C**).

In addition to the species associated to general microbiome dysbiosis for NAFLD, we also look for species associated with progressive subtype (NASH, fibrosis, and cirrhosis) that can lead to serious consequences, such as hepatocellular

carcinoma and liver-related death. Ten species were enriched in patients with progressive NAFLD, including *Clostridium perfringens*, *Bacteroides coprophilus*, *Flavonifractor unclassified*, *Turicibacter sanguinis*, *Actinomyces graevenitzii*, *Anaerostipes unclassified*, *Coprobacillus* sp. 3 3 56FAA, *Parabacteroides* sp. D25, *Eubacterium cylindroides*, and *Paraprevotella xylaniphila*, while five species were deleted, namely, *Streptococcus tigurinus*, *Slackia piriformis*, *Granulicatella adiacens*, *Roseburia unclassified*, and *Anaerostipes hadrus* (**Figure 3D**). *C. perfringens*, a well-known pathogen with the ability to secrete an arsenal of more than 20 virulent toxins (Kiu and Hall, 2018), is increased in several food poisoning and non-foodborne diseases. In our study, it is significantly and most increased in patients with fibrosis and

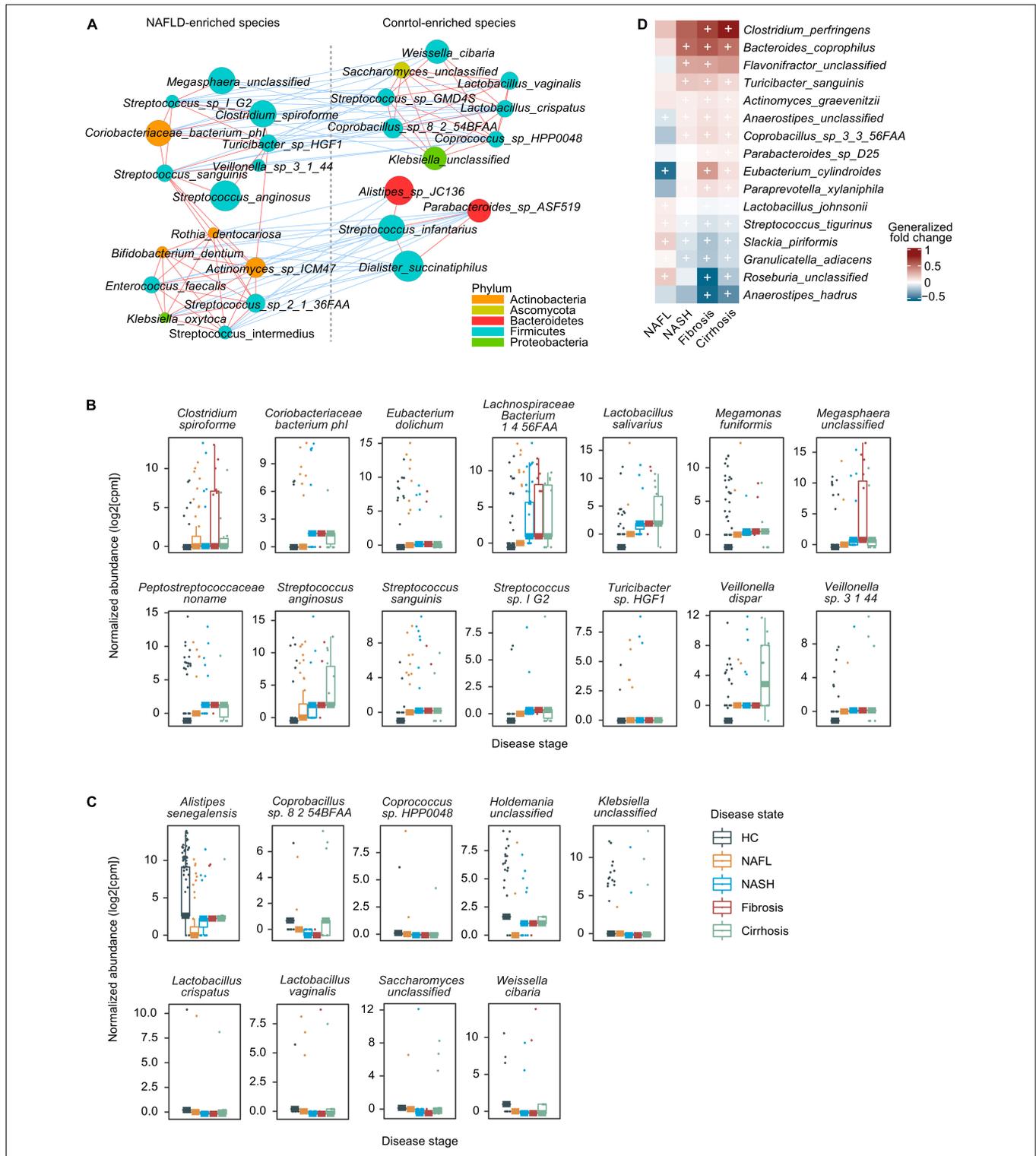


FIGURE 3 | A co-occurrence network and the correlation between gut microbes and NAFLD. **(A)** A co-occurrence network for the core set of highly significant species. The size of the nodes indicates normalized abundance. The color of the nodes indicates their taxonomic assignment. Connecting lines represent Spearman correlation coefficient values above 0.8 (red) or below -0.8 (blue). **(B)** Boxplots showing that species was significantly correlated with NAFLD stages using partial Spearman's rank-based correlation (pSRC), adjusted for BMI (two-sided, FDR < 0.05). **(C)** Boxplots showing that species was significantly anti-correlated with NAFLD stages (pSRC, two-sided, FDR < 0.05). **(D)** Heatmap showing the species that significantly and consistently changed in the last three stages (blocked two-sided Wilcoxon tests, + indicates statistical significance FDR < 0.05). Generalized fold change is colored by direction of the effect, where red indicates higher abundance in patients and blue indicates depletion. HC, healthy controls; NAFL, non-alcoholic fatty liver; NASH, non-alcoholic steatohepatitis.

cirrhosis, implying that the increase in the level of toxins can enter systemic circulation to affect various organs, such as liver. The higher level of *G. adiacens* has been reported to be associated with lung cancer (Shirazi et al., 2019) and pancreatic cancer (Membra et al., 2017), but in our study, it is decreased in patients with advanced NAFLD. Other strains have been little studied so far, so the role of microbe community in the development or prevention of disease is still an ongoing area of research.

We also did univariate analysis of genus associated with NAFLD. The raw counts were log-cpm-SNM transformed to reduce batch effect as mentioned above (**Supplementary Figures 1A,B**). At a meta-analysis FDR of 0.05, we identified 6 microbial species to be differentially enriched and 11 microbial species to be differentially depleted in the NAFLD patients out of 81 species consistently detected across studies (**Supplementary Figure 3C**). The gut microbial community also had consistent alteration patterns across different disease stages. *Megasphaera* has been reported to be enriched in patients with liver fibrosis (Chen et al., 2011), and we also got the same results in our study. *Porphyromonas* (Zhu et al., 2013) and *Peptoniphilus* (Del Chierico et al., 2017) were reported to be increased in patients with NASH, which was not repeated in our study, mainly owing to the sequence method and the sample from children. *Phascolarctobacterium* (Aron-Wisniewsky et al., 2020) were reported to be decreased in patients with obesity, a common complication of NAFLD, and we also got the same results in the present study. At the genus level, *Coriobacteriaceae noname*, *Fusobacterium*, *Megasphaera*, and *Sutterellaceae unclassified* were significantly correlated with NAFLD progression (**Supplementary Figure 3D**), and *Leuconostoc*, *Saccharomyces*, and *Weissella* were significantly anti-correlated with NAFLD progression (**Supplementary Figure 3E**).

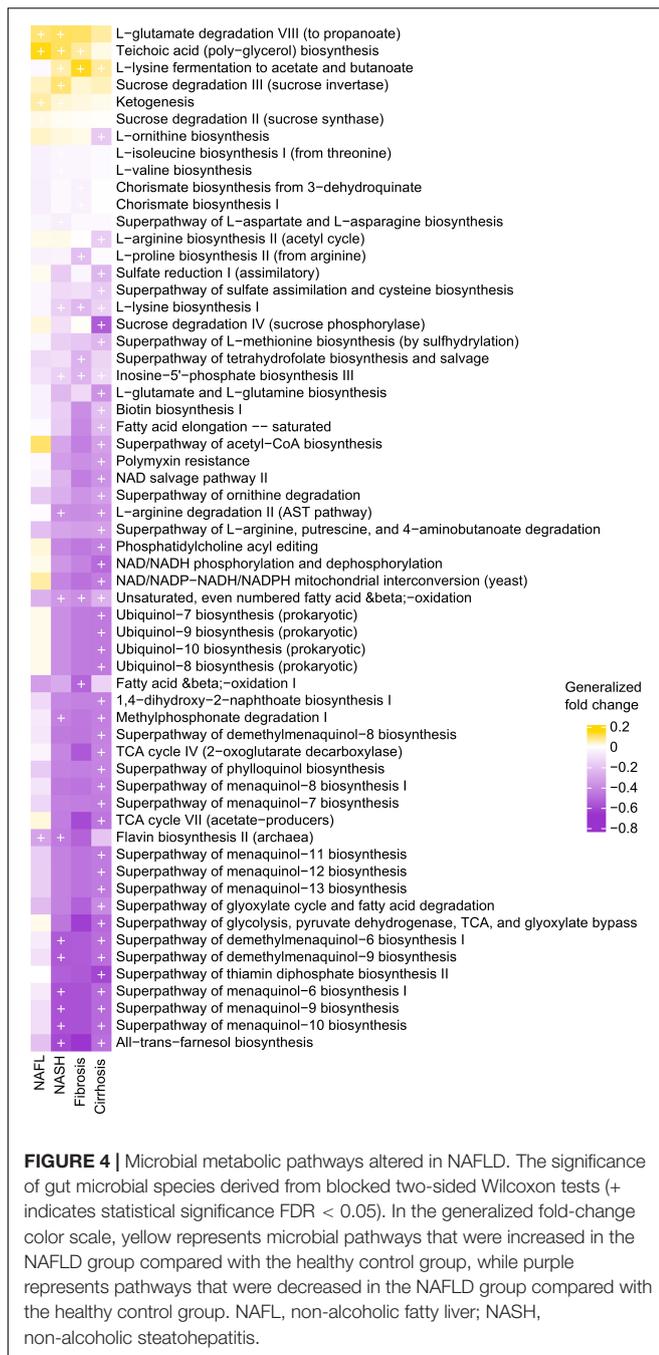
Functional Metagenomic Signatures for Non-alcoholic Fatty Liver Disease

Functional potential of the microbiome was also significantly associated with NAFLD samples when compared to healthy controls. We found 11,941 of the 17,426 single gene families (FDR < 0.05) detected at least once to be enriched in NAFLD patients and 12,969 to be enriched in controls at meta-analysis FDR < 0.05. We further observed 179 out of 189 metagenomically reconstructed microbial functional pathways (FDR < 0.05) to be at least once control-enriched, and only 10 to be enriched in NAFLD patients at all stages of NAFLD. The disordered metabolic pathways showed an abnormal glycolipid metabolism, such as glycolysis, glyoxylate bypass, tricarboxylic acid (TCA) cycle, as well as fatty acid elongation, oxidation, and degradation, and these have been reported in several intestinal and metabolic disorders of multiple etiologies, such as colorectal cancer (Wirbel et al., 2019), obesity (Hoyle et al., 2018), as well as cardiovascular disease (Zhernakova et al., 2018). The NAFLD-enriched pathway is mainly associated with the production of harmful products, such as L-glutamate degradation VIII (to propanoate) and teichoic acid (poly-glycerol) biosynthesis (**Figure 4**). Propanoate generated from L-glutamate degradation is of hypotoxicity

and inhibition of fatty acid synthesis (Hoke et al., 2016). The teichoic acid is a special component of the cell wall of gram-positive (G+) bacterium, helping bacteria stick to the surface of human cells, avoiding phagocytosis by leukocytes and resisting complement, which may be related to pathogenicity (Engholm et al., 2017). Most of the enriched species in patients gut microbiome are gram positive, such as *Streptococcus*, *Lactobacillus*, *Clostridium*, *Bifidobacterium*, and *Actinomyces*, which reinforce our finding. The significant alteration pathways also showed an association with NADH/NAD⁺ balance via glycolysis, TCA, and glyoxylate bypass, which have previously been associated with alcoholic hepatitis (Seronello et al., 2010) and metabolic dysfunction (Alves-Paiva et al., 2018; **Figure 4**). Butyrate is well known not only to be the preferred fuel for the colonic epithelial cells and the major regulator of cell proliferation and differentiation (Guilloteau et al., 2010), but also shown to exert important actions related to cellular homeostasis such as anti-inflammatory, antioxidant, and anti-carcinogenic functions (Hamer et al., 2009). Among pathways involved in butyrate production, the acetyl-CoA biosynthesis, L-glutamate biosynthesis, L-lysine biosynthesis, and L-histidine biosynthesis pathways were decreased in NAFLD patients (**Figure 4**), indicating the reduced protection effects of butyrate. The production of phosphatidylcholine, the main phospholipid in cellular membranes, from lysophosphatidylcholine acylation is a key component of the acyl-editing process, involving recycling of the fatty acids (Agarwal and Garg, 2010). Phosphatidylcholine provides the majority of the exogenous choline, the lack of which is linked to the accumulation of hepatic lipid and induced models of NAFLD in animals (Sherriff et al., 2016). The decreased phosphatidylcholine acyl editing pathway in progressive stages of NAFLD patients implied that the lack of protective effect of phosphatidylcholine and deficiency of choline might contribute to the liver injury (**Figure 4**). In addition, the pathway for synthesizing nutrients, such as menaquinol, flavin, ubiquinol, and phyloquinol, was observed to be reduced in patients (**Figure 4**). In addition, some species that were deleted in patients might affect Type II immune response. For example, *D. succinatiphilus*, known as succinate-utilizing bacteria, might regulate the phenotype and function of immune cells.

Validation of Non-alcoholic Fatty Liver Disease Microbiome Signature

To evaluate the utility of the metagenomic gut microbiome signature for the detection of NAFLD, we tested its diagnostic accuracy between patients and healthy controls and cross four stages of NAFLD by training stochastic gradient-boosting machine (GBM) learning models. In patients–controls validation using species-level taxonomic normalized abundances, we observed performances ranging in area under the receiver operating characteristic curve (AUROC) score from 0.9861 to 1.0000 and in area under the precision-recall curve (AUPR) score from 0.8723 to 1.0000 (**Figure 5A** and **Supplementary Figures 4A–D**), and the performances were generally maintained in stage-to-stage transfer (AUROC ranged between 0.8125 and 0.9974) (**Figure 5A**). These results show that within-stage



validation AUROCs and AUPR can be high for prediction of a certain stage of NAFLD with healthy controls, but they can also distinguish other stages from healthy controls, highlighting the homogeneous microbiome dysbiosis across stages. Given the clinical importance of our study, we further assessed whether it could discriminate one stage from the others. Encouragingly, we observed performances ranging in AUROC score from 0.7000 to 0.9352 and in AUPR score from 0.5462 to 0.8332 (Figure 5B and Supplementary Figures 4E–H), which performed well for discriminating between one stage and the others. Random forest

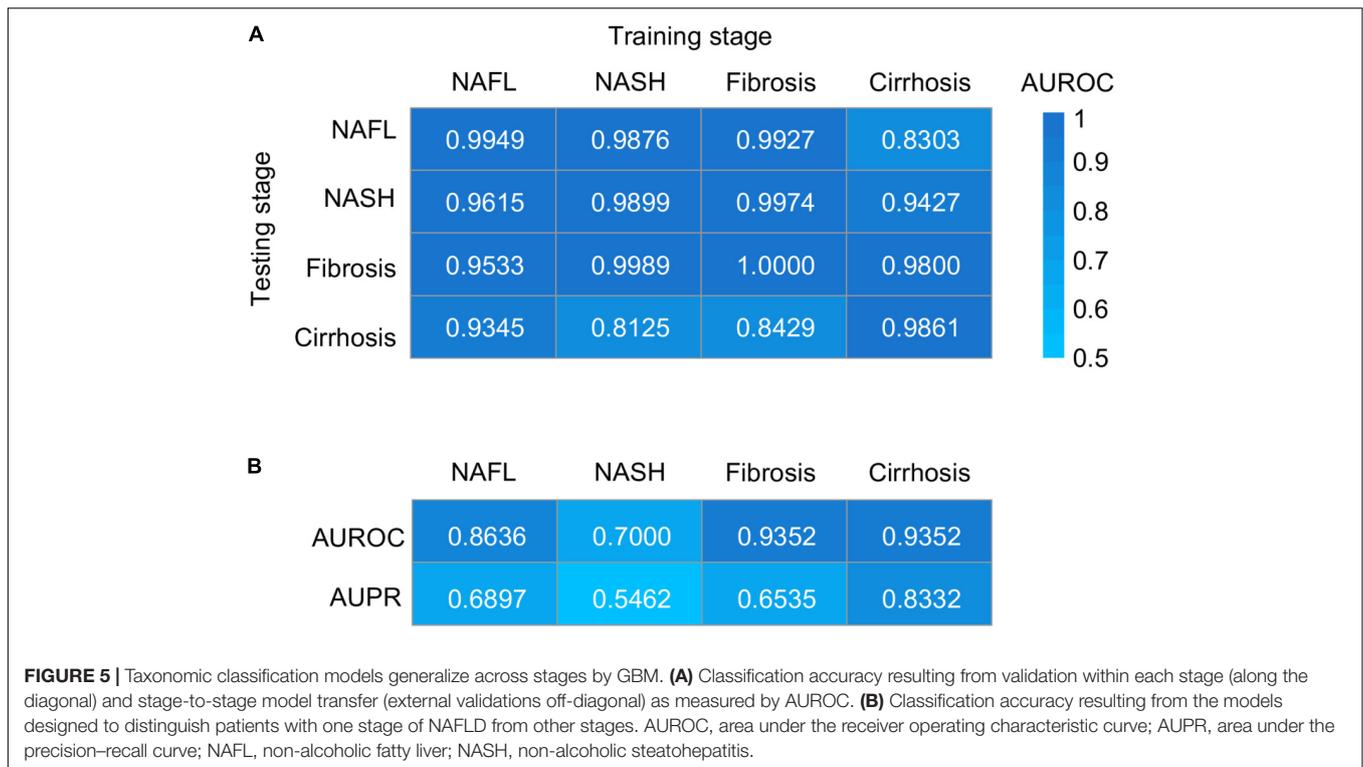
model for verification was also used, and the classification performance was similar to GBM (Supplementary Figure 5). To evaluate the generalizability of our approach across datasets, we randomly sorted raw microbial counts into two batches, repeated all procedures on each independently, tested each independently trained model on the other half of the data, and found highly similar performance (Supplementary Figure 5). The slight reduction in sensitivities and specificities for split data may be partially due to small data sizes.

Among the features used in the model validating patients and controls, *Alistipes* sp. JCI36 was the species with the highest average rank for the importance. As expected, other generally NAFLD-associated species including *Holdemania unclassified*, *Bifidobacterium* sp. 12 1 47BFAA, *Coriobacteriaceae bacterium phI*, *A. odontolyticus*, *Eubacterium siraeum*, *Actinomyces* sp. ICM47, *Bacteroides thetaiotaomicron*, *R. mucilaginosa*, and *S. vestibularis* were also crucial to prediction accuracy (Figures 2, 6A). *A. graevenitzii* and *P. xylaniphila* were identified as species associated with advanced stages of NAFLD and also had good performance for validating patients and controls (Figures 3D, 6A). Among the features used in the model validating four stages of NAFLD, *E. cylindroides*, *Clostridium hathewayi*, *Dorea formicigenerans*, and *Clostridium* sp. HGF2 had the highest rank for the importance, respectively (Figure 6B). Our previous study has shown that *C. hathewayi* was enriched in ankylosing spondylitis (AS) patients (Xu and Yin, 2019; Yin et al., 2020), but in this current study, we found that its abundance was significantly decreased in NASH patients with good discriminating performance.

DISCUSSION

Our study was performed across multiple datasets and populations, through a combined analysis of fecal NAFLD metagenomes from three publicly available datasets. Divergence of metagenomic approaches and study design, such as differences in sample collection and preservation, DNA extraction methodology and sequencing platform, all affect the composition of downstream sequence data. The effect of study-associated heterogeneity on microbiome composition was first quantified. The sequencing platform was the same in all three studies (Illumina HiSeq), while DNA extraction methods were different. Although all three studies stated that the sampling method was rapid freezing to -80°C , there were still technical differences due to human factors. So, the technical variation, such as sampling and DNA extraction for each of the downloaded dataset, was integrally considered as batch effects, which be dealt with at the beginning. Although these effects cannot be completely eliminated, they were greatly reduced (Figure 1B).

Researchers are often more likely to focus on the difference between disease and healthy controls, while the commonality between related diseases is often neglected. Here, we identified a core gut microbiome signature for general NAFLD microbiome dysbiosis instead of disease-stage-specific links (Figure 2). Although most of the classified NAFLD-enriched species have been shown to be pathogenic bacterium for human or animal



models (Yang et al., 2007; Meile et al., 2008; Asam and Spellerberg, 2014; Uzal et al., 2018; Zhu et al., 2018), they were first to be shown as species associated with NAFLD. Among these, we demonstrate an interaction pattern of the most significant set of species (Figure 3A), and most of them were significantly correlated or anti-correlated with disease progression (Figures 3B,C). Boursier et al. (2016) had reported that *Ruminococcus* was significantly enriched in patients with both NASH or fibrosis, and in this current study, we also found a consistent result. Two *Ruminococcus* species (*Ruminococcus gnavus* and *Ruminococcus* sp. JC304) were identified as biomarkers for general NAFLD microbiome dysbiosis, and both were enriched in patients whatever stage (Figure 2). *Coprococcus* were reported to be decreased in NASH patients in three studies (Zhu et al., 2013; Wang et al., 2016; Hoyles et al., 2018). In our present study, three *Coprococcus* species (*Coprococcus catus*, *Coprococcus* sp. ART55 1, and *Coprococcus* sp. HPP0048) were found to be associated with general NAFLD microbiome dysbiosis, and all of them were commonly decreased in all stages of NAFLD patients (Figure 2). Furthermore, we described the species and metabolic pathways that specifically associated with progressive stages of NAFLD, implying that the increased toxic metabolites and decreased protection of butyrate and choline (Figures 3D, 4) together with diseased α -diversity (Figure 1C) contribute to NAFLD progression.

Broadly applicable, non-invasive methods for diagnosing the stage of NAFLD are currently not available. The identification of microbial biomarkers for NAFLD may enable the design of non-invasive diagnostic tools. We developed machine learning models able to distinguish patients with any stage of NAFLD

from healthy controls with an average performance of 0.99 AUROC when validated on datasets excluded from the training of the model (Figure 5A). The models designed to distinguish patients with one stage of NAFLD from other stages also had an excellent performance with an average AUROC of 0.8585 (Figure 5B). The slight reduction of AUROC for the later model compared to the former indicate more homogeneity in microbiome across four stages of NAFLD than with healthy control, which was also verified by principal coordinate analysis (Figure 1D) and the consistent alteration across NAFLD stages (Figure 2). Therefore, microbiome-based NAFLD prediction models enable a very high diagnostic potential. The integrated data are slightly unbalanced between case and control, so we choose GBM, a decision tree-based integrated learning. Decision trees tend to do well with category disequilibrium data. It uses categorization rules based on class variables to create a classification tree so that samples of different categories can be forcibly separated. The training set of category unbalanced data has little influence on the training results of decision tree algorithms. Furthermore, if these diagnostic features can distinguish non-alcoholic fatty liver and alcoholic fatty liver or other similar diseases such as liver cirrhosis caused by other causes, it will be more significant. However, shotgun sequencing is expensive, and related researches are also very few. Therefore, it is hard to obtain related available data. As the price of sequencing falls and more data become available, this may become possible.

Although this study included the relatively small sizes of the experimental cohorts, analysis of patients with different stage of NAFLD presents a distinct opportunity for studying the

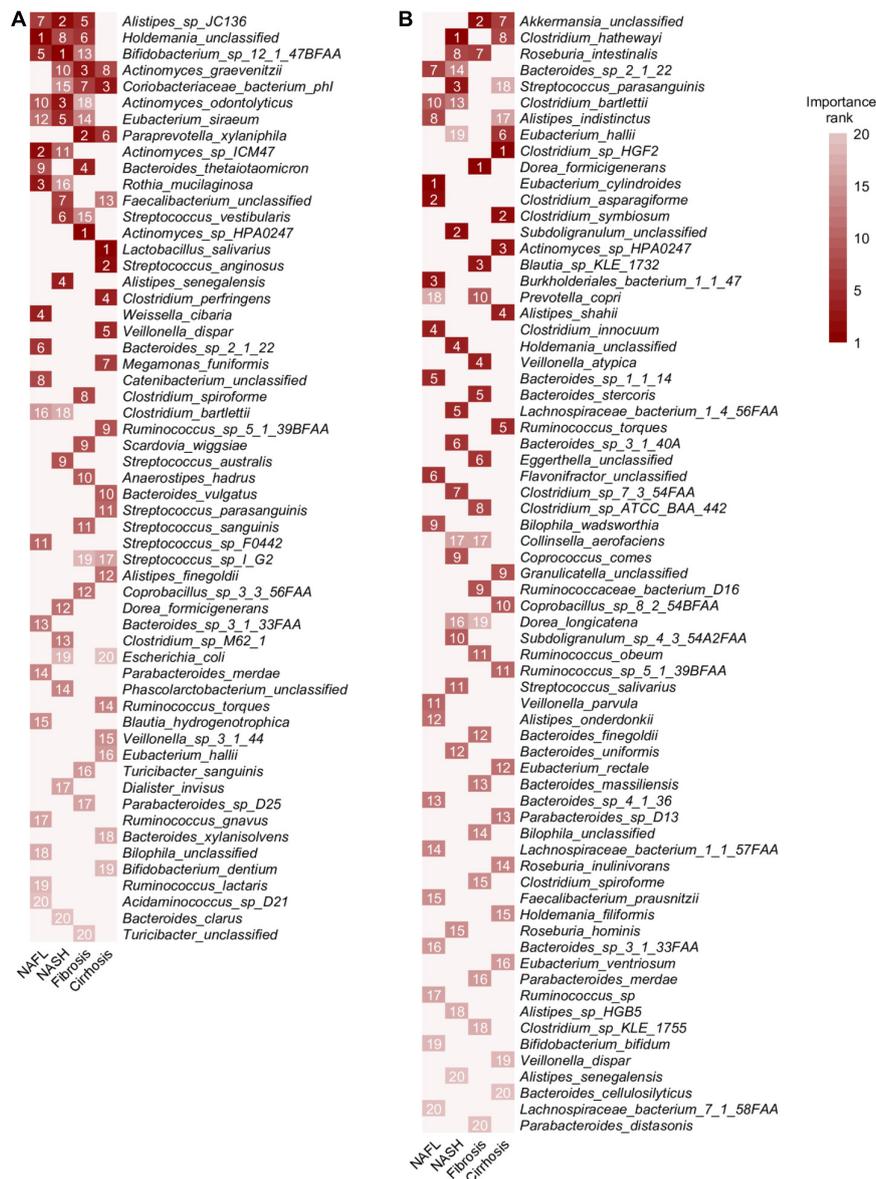


FIGURE 6 | Ranking relevance of each species in the predictive models for each stage. **(A)** The importance of each species for the prediction performance in each dataset estimated using stochastic gradient-boosting machine (GBM) learning models. **(B)** The importance of each species for the prediction performance of the models designed to distinguish one stage of NAFLD from other stages. Only species appearing in the 20 top-ranking features in at least one dataset are reported. HC, healthy controls; NAFL, non-alcoholic fatty liver; NASH, non-alcoholic steatohepatitis.

general NAFLD-associated and stage-specific microbiome. By combining multiple cohorts of potentially low generalizability, it is possible to obtain better representation of the spectrum of NAFLD cases and controls. At present, researches about gut microbes are still very limited, and we still know little about the role of different strains in different situations. Even some known probiotics can be opportunistic pathogens. Therefore, this study combines data from three studies to identify potential candidate bacteria that contribute to disease development. These bacteria have been poorly studied and functional studies are needed to explore their role in disease. With appropriate

methodology, artifactual findings due to batch effects present in any individual dataset can be avoided. In addition, the identification of pathogenic and beneficial microbial species might lead to novel therapies for severe forms of NAFLD. Taking account of limited accuracy of serum markers, the expense of MRE technologies and the invasiveness of liver biopsy, gut microbiome test is more convenient and feasible for disease screening. Our discovery of a gut microbiome-derived signature that accurately identifies the stage of NAFLD lays the foundations and points to the potential for non-invasive microbial diagnostic tests to supplement existing screening.

MATERIALS AND METHODS

Study Inclusion and Data Acquisition

We used PubMed to search for studies that published fecal shotgun metagenomic data of human NAFLD patients and healthy CTRLs. Raw FASTQ files were downloaded for the three included studies from the European Nucleotide Archive (ENA) using the following ENA identifiers: ERP015847 for Hoyles et al. (2018), PRJNA373901 for Loomba et al. (2017), and ERP005860 for Qin et al. (2014).

Sample Processing and Filtration

The stage of NAFLD was diagnosed according to liver biopsy. Biopsies were assessed for the following three parameters: Steatosis was graded 0–3, lobular inflammation was graded 0–3, and ballooning was graded 0–2. Fibrosis stage was classified into five stages from 0 to 4. NAFL patients have fat accumulation in the liver (steatosis) involving at least 5% of hepatocytes on routine stains without lobular inflammation, ballooning, and fibrosis. Presence of NASH was defined as a pattern that was consistent with steatohepatitis including presence of at least 5% steatosis, lobular inflammation, and ballooning with or without peri-sinusoidal fibrosis (fibrosis stage 1). Fibrosis stage consists of periportal fibrosis (fibrosis stage 2) and bridging fibrosis (fibrosis stage 3). Cirrhosis was defined as stage 4 fibrosis.

Participants were included in the study if they met the following criteria: (1) 18 years or older, (2) fat accumulation in the liver (steatosis) involving at least 5% of hepatocytes on routine stains, (3) no evidence of other acute or chronic liver disease, and (4) absence of regular or excessive use of alcohol. Patients were excluded from the study if they met any of the following criteria: (1) clinical or histological evidence of alcoholic liver disease; and (2) clinical or biochemical evidence of liver diseases other than NAFLD, including hepatitis B, hepatitis C, alpha-1 antitrypsin deficiency, hemochromatosis, Wilson's disease, autoimmune hepatitis, polycystic liver diseases, cholestatic liver diseases, and vascular liver diseases. Patients in the three studies who met the above conditions and had a clear liver biopsy diagnosis were included in this study.

The liver imaging and liver biochemistry results of all healthy controls were in the normal range. Physical examination; routine examination of blood, urine, and stools; preoperative serological tests; liver function; renal function; electrolyte; liver ultrasound; electrocardiogram; and chest X-ray results were checked in the healthy controls to exclude any abnormal samples, such as clinical or biochemical evidence of liver diseases, chronic illnesses associated with hepatic steatosis, use of drugs known to cause hepatic steatosis, and presence of systemic infectious illnesses.

Sequence Preprocessing and Taxonomic and Functional Profiling

Fecal metagenomic shotgun sequences were quality filtered using Trimmomatic (Bolger et al., 2014). Filtered reads were then aligned to the human genome and the PhiX genome for human and contaminant DNA removal using bowtie2 (Langmead and Salzberg, 2012). We used MetaPhlan2 for quantitative profiling

of the taxonomic composition of the microbial communities of all metagenomic samples, and HUMANN2 was chosen to profile pathway abundances using a ChocoPhlan database and gene-family abundances using a UniRef90 database. Microbial profiles were first converted to count per million (cpm) to account for library size. Then, profiles were filtered to focus on a set of species that were detectable in all studies. Specifically, microbial species that did not exceed a cpm of 1 in at least three samples were excluded from further analysis. Functional profiles, such as gene family or metabolic pathway abundance profiles, were preprocessed as the species profiles.

Normalization and Confounder Analysis

Cognizant of how technical variation and heterogeneous ethnicity between studies could confound our results, we made data normalization to remove batch effects before further analysis. In brief, we transformed our discrete taxonomical count data to approximately normally distributed, log-count per million (log-cpm) data, which models and removes the data's heteroskedasticity; and then performed supervised normalization (SNM) on the data to remove significant batch effects while preserving biological effects (Mecham et al., 2010). Confounder analysis was performed by ANOVA to quantify the effect of confounders relative to that of disease state on single microbial species. Variance calculations were performed on ranks of microbiome abundance in non-Gaussian distribution.

Statistical Analyses

Since microbiome data are characterized by non-Gaussian distributions with excessive dispersion, the non-parametric significance testing using blocked Wilcoxon rank-sum testing was implemented in the R “coin” package (Hothorn et al., 2006). Informed by the results of the preceding confounder analysis, BMI was blocked in meta-analysis. The mean difference in a set of predefined quantiles of the logarithmic case and distributions was calculated as a generalized fold change as this has been shown to achieve better resolution for sparse microbiome profiles than other methods (Feng et al., 2012). We used quantiles ranging from 0.1 to 0.9 in increments of 0.1. The non-parametric effect size was measured by Area Under Curve (AUC) obtained from the “pROC” package (Robin et al., 2011) in R. The co-occurrence of significantly different genera was analyzed using Spearman correlation. Partial Spearman's rank-based correlation (pSRC) in the R “coin” package (Hothorn et al., 2006) was used to assess the associations between microbial profile and NAFLD stages with BMI as confounder. To adjust for multiple hypothesis testing, *p*-values were adjusted using the false discovery rate (FDR) method (Benjamini and Hochberg, 1995). Principal coordinate analyses based on species-level Bray–Curtis dissimilarity, together with α -diversity, were calculated using “vegan” in R (Dixon, 2003).

Machine Learning Model

Stochastic GBM learning models were trained, automatically tuned, and tested using the GBM package (Friedman and Analysis, 2002) and Caret package (Kuhn, 2008) in R. Training and testing occurred on separate, randomly selected, stratified

sampling splits of 70 and 30% of the data, respectively, and a fixed random number seed was used to ensure reproducibility of the model results and comparability among models. During model training, the data were first centered and scaled for each sample to have mean zero and unit standard deviation; fourfold cross-validation was used to create multiple subsets of the training data and to perform a basic grid search optimization of GBM parameters, including interaction depth (1, 2, or 3) and number of trees (50, 100, or 150), while maximizing AUROC of the final, fully trained model. Learning rate (shrinkage) was held constant at 0.1 and the number of minimum observations per node was fixed at 5. Final model performances, including ROC curves and PR curves, were generated by applying the final model to the unseen 30% holdout test set. ROC and PR curves as well as AUROC and AUPR values were calculated using the PRROC package (Grau et al., 2015). Variable importance scores of the resultant, non-zero model features were estimated using the GBM and Caret packages (Friedman and Analysis, 2002; Kuhn, 2008). Performance of classifiers within each stage was analyzed by fourfold cross-validation. Performance of classifiers across stages was analyzed by stage-to-stage transfer validation, that is, each stage was taken as training set, and the others were taken as testing sets. All species detected were included as features, and the importance of each feature was analyzed.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories

REFERENCES

- Agarwal, A. K., and Garg, A. (2010). Enzymatic activity of the human 1-acylglycerol-3-phosphate-O-acyltransferase isoform 11: upregulated in breast and cervical cancers. *J. Lipid Res.* 51, 2143–2152. doi: 10.1194/jlr.M004762
- Alves-Paiva, R. M., Kajigaya, S., Feng, X., Chen, J., Desierto, M., Wong, S., et al. (2018). Telomerase enzyme deficiency promotes metabolic dysfunction in murine hepatocytes upon dietary stress. *Liver Int.* 38, 144–154. doi: 10.1111/liv.13529
- Aron-Wisniewsky, J., Vigiotti, C., Witjes, J., Le, P., Holleboom, A. G., Verheij, J., et al. (2020). Gut microbiota and human NAFLD: disentangling microbial signatures from metabolic disorders. *Nat. Rev. Gastroenterol. Hepatol.* 17, 279–297. doi: 10.1038/s41575-020-0269-9
- Asam, D., and Spellerberg, B. (2014). Molecular pathogenicity of *Streptococcus anginosus*. *Mol. Oral Microbiol.* 29, 145–155. doi: 10.1111/omi.12056
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Boursier, J., Mueller, O., Barret, M., Machado, M., Fizanne, L., Araujo-Perez, F., et al. (2016). 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. doi: 10.1002/hep.28356
- Causy, C., Tripathi, A., Humphrey, G., Bassirian, S., Singh, S., Faulkner, C., et al. (2019). A gut microbiome signature for cirrhosis due to nonalcoholic fatty liver disease. *Nat. Commun.* 10:1406. doi: 10.1038/s41467-019-09455-9

and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

TW and HX conceived and supervised the study. TW performed the taxonomic profiling, machine learning, statistical analyses, produced the figures, and wrote the manuscript with contributions from X-KG and HX. All authors discussed and approved the manuscript.

FUNDING

HX was supported by the National Natural Science Foundation of China (Grant 31821003) and the China Ministry of Science and Technology (Grant 2018AAA0100300). Shanghai Municipal Key Clinical Specialty (shslczdzk02602) and Shanghai Science and Technology Development Funds (2020-SH-XY-2).

ACKNOWLEDGMENTS

We are thankful to Zhao's cluster for sequence preprocessing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2021.728823/full#supplementary-material>

- Chen, Y., Yang, F., Lu, H., Wang, B., Chen, Y., Lei, D., et al. (2011). Characterization of fecal microbial communities in patients with liver cirrhosis. *Hepatology* 54, 562–572. doi: 10.1002/hep.24423
- Del Chierico, F., Nobili, V., Vernocchi, P., Russo, A., De Stefanis, C., Gnani, D., et al. (2017). Gut microbiota profiling of pediatric nonalcoholic fatty liver disease and obese patients unveiled by an integrated meta-omics-based approach. *Hepatology* 65, 451–464. doi: 10.1002/hep.28572
- Dixon, P. M. (2003). VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 14, 927–930. doi: 10.1111/j.1654-1103.2003.tb02228.x
- Engholm, D. H., Kilian, M., Goodsell, D. S., Andersen, E. S., and Kjærgaard, R. S. (2017). A visual review of the human pathogen *Streptococcus pneumoniae*. *FEMS Microbiol. Rev.* 41, 854–879. doi: 10.1093/femsre/fux037
- Feng, J., Meyer, C. A., Wang, Q., Liu, J. S., Shirley Liu, X., and Zhang, Y. (2012). GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics* 28, 2782–2788. doi: 10.1093/bioinformatics/bts515
- Fingas, C. D., Best, J., Sowa, J. P., and Canbay, A. (2016). Epidemiology of nonalcoholic steatohepatitis and hepatocellular carcinoma. *Clin. Liver Dis. (Hoboken)* 8, 119–122. doi: 10.1002/cld.585
- Friedman, J. H. J. C. S., and Analysis, D. (2002). Stochastic gradient boosting. *Comput. Stat. Data Anal.* 38, 367–378. doi: 10.1016/S0167-9473(01)00065-2
- Grau, J., Grosse, I., and Keilwagen, J. (2015). PRROC: computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinformatics* 31, 2595–2597. doi: 10.1093/bioinformatics/btv153
- Guilloteau, P., Martin, L., Eeckhaut, V., Ducatelle, R., Zabielski, R., and Van Immerseel, F. (2010). From the gut to the peripheral tissues: the multiple effects of butyrate. *Nutr. Res. Rev.* 23, 366–384. doi: 10.1017/S0954422410000247

- Hamer, H. M., Jonkers, D. M., Bast, A., Vanhoutvin, S. A., Fischer, M. A., Kodde, A., et al. (2009). Butyrate modulates oxidative stress in the colonic mucosa of healthy humans. *Clin. Nutr.* 28, 88–93. doi: 10.1016/j.clnu.2008.11.002
- Hoke, R. A., Ferrell, B. D., Sloman, T. L., Buck, R. C., and Buxton, L. W. (2016). Aquatic hazard, bioaccumulation and screening risk assessment for ammonium 2,3,3,3-tetrafluoro-2-(heptafluoropropoxy)-propanoate. *Chemosphere* 149, 336–342. doi: 10.1016/j.chemosphere.2016.01.009
- Hothorn, T., Hornik, K., De Wiel, M. A. V., and Zeileis, A. J. T. A. S. (2006). A lego system for conditional inference. *Am. Stat.* 60, 257–263. doi: 10.1198/000313006X118430
- Hoyle, L., Fernández-Real, J. M., Federici, M., Serino, M., Abbott, J., Charpentier, J., et al. (2018). Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese women. *Nat. Med.* 24, 1070–1080. doi: 10.1038/s41591-018-0061-3
- Jie, Z., Xia, H., Zhong, S. L., Feng, Q., Li, S., Liang, S., et al. (2017). The gut microbiome in atherosclerotic cardiovascular disease. *Nat. Commun.* 8:845. doi: 10.1038/s41467-017-00900-1
- Kiu, R., and Hall, L. J. (2018). An update on the human and animal enteric pathogen *Clostridium perfringens*. *Emerg. Microbes Infect.* 7:141. doi: 10.1038/s41426-018-0144-8
- Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26. doi: 10.18637/jss.v028.i05
- Kwak, S. H., Cho, Y. M., Noh, G. M., and Om, A. S. (2014). Cancer preventive potential of kimchi lactic acid bacteria (*Weissella cibaria*, *Lactobacillus plantarum*). *J. Cancer Prev.* 19, 253–258. doi: 10.15430/JCP.2014.19.4.253
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Lepargneur, J. P. (2016). *Lactobacillus crispatus* as biomarker of the healthy vaginal tract. *Ann. Biol. Clin. (Paris)* 74, 421–427. doi: 10.1684/abc.2016.1169
- Loomba, R., Seguritan, V., Li, W., Long, T., Klitgord, N., Bhatt, A., et al. (2017). Gut microbiome-based metagenomic signature for non-invasive detection of advanced fibrosis in human nonalcoholic fatty liver disease. *Cell. Metab.* 25, 1054–1062.e1055. doi: 10.1016/j.cmet.2017.04.001
- Mecham, B. H., Nelson, P. S., and Storey, J. D. (2010). Supervised normalization of microarrays. *Bioinformatics* 26, 1308–1315. doi: 10.1093/bioinformatics/btq118
- Meile, L., Le Blay, G., and Thierry, A. (2008). Safety assessment of dairy microorganisms: propionibacterium and bifidobacterium. *Int. J. Food Microbiol.* 126, 316–320. doi: 10.1016/j.ijfoodmicro.2007.08.019
- Memba, R., Duggan, S. N., Ni Chonchubhair, H. M., Griffin, O. M., Bashir, Y., O'Connor, D. B., et al. (2017). The potential role of gut microbiota in pancreatic disease: a systematic review. *Pancreatol.* 17, 867–874. doi: 10.1016/j.pan.2017.09.002
- Ponziani, F. R., Bhoori, S., Castelli, C., Putignani, L., Rivoltini, L., Del Chierico, F., et al. (2019). Hepatocellular carcinoma is associated with gut microbiota profile and inflammation in nonalcoholic fatty liver disease. *Hepatology* 69, 107–120. doi: 10.1002/hep.30036
- Qin, N., Yang, F., Li, A., Prifti, E., Chen, Y., Shao, L., et al. (2014). Alterations of the human gut microbiome in liver cirrhosis. *Nature* 513, 59–64. doi: 10.1038/nature13568
- Qin, S., Kim, J., Arafat, D., and Gibson, G. (2012). Effect of normalization on statistical and biological interpretation of gene expression profiles. *Front. Genet.* 3:160. doi: 10.3389/fgene.2012.00160
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77. doi: 10.1186/1471-2105-12-77
- Sberna, A. L., Bouillet, B., Rouland, A., Brindisi, M. C., Nguyen, A., Mouillot, T., et al. (2018). European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD) and European Association for the Study of Obesity (EASO) clinical practice recommendations for the management of non-alcoholic fatty liver disease: evaluation of their application in people with Type 2 diabetes. *Diabet. Med.* 35, 368–375. doi: 10.1111/dme.13565
- Seronello, S., Ito, C., Wakita, T., and Choi, J. (2010). Ethanol enhances hepatitis C virus replication through lipid metabolism and elevated NADH/NAD+. *J. Biol. Chem.* 285, 845–854. doi: 10.1074/jbc.M109.045740
- Shen, F., Zheng, R. D., Sun, X. Q., Ding, W. J., Wang, X. Y., and Fan, J. G. (2017). Gut microbiota dysbiosis in patients with non-alcoholic fatty liver disease. *Hepatobiliary. Pancreat. Dis. Int.* 16, 375–381. doi: 10.1016/S1499-3872(17)60019-5
- Sherriff, J. L., O'Sullivan, T. A., Properzi, C., Oddo, J. L., and Adams, L. A. (2016). Choline, its potential role in nonalcoholic fatty liver disease, and the case for human and bacterial genes. *Adv. Nutr.* 7, 5–13. doi: 10.3945/an.114.007955
- Shirazi, M. S. R., Al-Alo, K. Z. K., Al-Yasiri, M. H., Lateef, Z. M., and Ghasemian, A. (2019). Microbiome dysbiosis and predominant bacterial species as human cancer biomarkers. *J. Gastrointest. Cancer.* 51, 725–728. doi: 10.1007/s12029-019-00311-z
- Uzal, F. A., Navarro, M. A., Li, J., Freedman, J. C., Shrestha, A., and McClane, B. A. (2018). Comparative pathogenesis of enteric clostridial infections in humans and animals. *Anaerobe* 53, 11–20. doi: 10.1016/j.anaerobe.2018.06.002
- Vešičič, P., Kacerovská Musilová, I., Stránil, J., Štěpán, M., and Kacerovský, M. (2020). *Lactobacillus crispatus* dominant vaginal microbiota in pregnancy. *Ceska Gynekol* 85, 67–70.
- Wang, B., Jiang, X., Cao, M., Ge, J., Bao, Q., Tang, L., et al. (2016). Altered fecal microbiota correlates with liver biochemistry in nonobese patients with non-alcoholic fatty liver disease. *Sci. Rep.* 6:32002. doi: 10.1038/srep32002
- Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., et al. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat. Med.* 25, 679–689. doi: 10.1038/s41591-019-0406-6
- Xu, H., and Yin, J. (2019). HLA risk alleles and gut microbiome in ankylosing spondylitis and rheumatoid arthritis. *Best Pract. Res. Clin. Rheumatol.* 33:101499. doi: 10.1016/j.berh.2020.101499
- Yang, C. Y., Hsueh, P. R., Lu, C. Y., Tsai, H. Y., Lee, P. I., Shao, P. L., et al. (2007). *Rothia dentocariosa* bacteremia in children: report of two cases and review of the literature. *J. Formos. Med. Assoc.* 106, S33–S38. doi: 10.1016/S0929-6646(09)60364-8
- Yin, J., Sternes, P. R., Wang, M., Song, J., Morrison, M., Li, T., et al. (2020). Shotgun metagenomics reveals an enrichment of potentially cross-reactive bacterial epitopes in ankylosing spondylitis patients, as well as the effects of TNFi therapy upon microbiome composition. *Ann. Rheum. Dis.* 79, 132–140. doi: 10.1136/annrheumdis-2019-215763
- Zhernakova, D. V., Le, T. H., Kurilshikov, A., Atanasovska, B., Bonder, M. J., Sanna, S., et al. (2018). Individual variations in cardiovascular-disease-related protein levels are driven by genetics and gut microbiome. *Nat. Genet.* 50, 1524–1532. doi: 10.1038/s41588-018-0224-7
- Zhu, B., Macleod, L. C., Kitten, T., and Xu, P. (2018). *Streptococcus sanguinis* biofilm formation & interaction with oral pathogens. *Future Microbiol.* 13, 915–932. doi: 10.2217/fmb-2018-0043
- Zhu, L., Baker, S. S., Gill, C., Liu, W., Alkhoury, R., Baker, R. D., et al. (2013). Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH) patients: a connection between endogenous alcohol and NASH. *Hepatology* 57, 601–609. doi: 10.1002/hep.26093

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wang, Guo and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.