Supplementary Figures

Gut microbiome analysis as a tool towards targeted non-invasive biomarkers for early hepatocellular carcinoma

Running title: Gut microbiome for early hepatocellular carcinoma

Zhigang Ren[1,3,2†], Ang Li[2,4,3†], Jianwen Jiang[1,4,9†], Lin Zhou[1,4†], Zujiang Yu[3,2†], Haifeng Lu[4], Haiyang Xie[1,4], Xiaolong Chen[3,2], Li Shao[4], Ruiqing Zhang[5,8], Shaoyan Xu[1], Hua Zhang[4], Guangying Cui[3,2], Xinhua Chen[1,4], Ranran Sun[3,2], Hao Wen[8], Jan Lerut[6], Quancheng Kan[7]*, Lanjuan Li[4]*, and Shusen Zheng[1,4,10]*

[1] Department of Hepatobiliary and Pancreatic Surgery, the First Affiliated Hospital, School of Medicine, Zhejiang University; Key Laboratory of Combined Multi-organ Transplantation, Ministry of Public Health, Hangzhou 310003, China.

[2] Gene Hospital of Henan Province; Precision Medicine Center, the First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China.

[3] Department of Infectious Diseases, the First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China.

[4] State Key Laboratory for Diagnosis and Treatment of Infectious Disease; Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang University, Hangzhou 310003, China.

[5] Hepatobiliary and Hydatid Department, Digestive and Vascular Surgery Centre, Xinjiang Key Laboratory of Echinococcosis, the First Affiliated Hospital of Xinjiang Medical University, Urumqi, Xinjiang 830054, China.

[6] Starzl Unit Abdominal Transplantation, University Hospitals Saint Luc, Université

catholique Louvain, UCL Brussels, Belgium.

[7] Department of Pharmacy, the First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China.

[8] State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Xinjiang Medical University, Urumqi, Xinjiang 830054, China.

[9] Health Management Center, First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310003, China.

[10] Department of Hepatobiliary and Pancreatic Surgery, Shulan (Hangzhou) Hospital, Hangzhou 310022, China.

[†]These authors contributed equally to this work.

Supplementary Figures

Informed consent form and information collection

(Translated from Chinese)

We are from Key Laboratory of Combined Multi-organ Transplantation, Ministry of Public Health; Department of Hepatobiliary and Pancreatic Surgery, First Affiliated Hospital, School of Medicine, Zhejiang University. We will free of charge help you monitor your gut microbial community, thereby analyzing whether gut microbiota is dysbiosis and the degree of imbalance. These results will provide auxiliary data for clinical diagnosis and treatment. Now, you just provide stool and urine according to our instruction. The whole process keeps free of charge. These results will be used for scientific research. Thank you for your corporation.

Number:

Diagnosis:

Patient Sign：

Date:

Patient information collection

Name	Gender	Birth date	Height (cm)	Weight (kg)	BMI	Tel
Floor ward	Admission number	Dietary habit (vegetarian diet/meat/Mixture)	Antibiotics use within 2 months	Yoghourt and probiotics	Previously critical Illness	Long-term drug use
Drinking	Alcohol type	Drinking quantity	Time of Du ration	Whether or not abstinence	HBV	Etiology

Informed consent form for scientific research

(Translated from Chinese)

Dear participants,

We are from Department of Hepatobiliary and Pancreatic Surgery, First Affiliated Hospital, School of Medicine, Zhejiang University, and from Key Laboratory of Combined Multi-organ Transplantation, Ministry of Public Health. We will free of charge help you monitor your healthy condition and record your clinical information and healthy/disease status or disease progression process. The collected fecal and urea samples from participants in hospital will be used for scientific research. These results and data from the hospital electronic medical records will provide auxiliary data for clinical diagnosis and treatment, and will be used for scientific research. Thank you for your corporation.

Number:	Diagnosis:

The information that we collect from this research project will be kept confidential. Information about you that will be collected during the research will be put away and no-one but the researchers will be able to see it. Any information about you will have a number on it instead of your name. Only the researchers will know what your number is and we will lock that information up with a lock and key. It will not be shared with or given to anyone except our research team.

The knowledge that we get from doing this research will be shared with you through community meetings before it is made widely available to the public. Confidential information will not be shared. There will be small meetings in the community and these will be announced. After these meetings, we will publish the results in order that other interested people may learn from our research.

I have read the foregoing information, or it has been read to me. I have had the opportunity to ask questions about it and any questions that I have asked have been answered to my satisfaction. I consent voluntarily to participate as a participant in this research.

Print Name of Participant_____

Signature of Participant _____

Date _____

Day/month/year

A literate witness must sign (if possible, this person should be selected by the participant and should have no connection to the research team). Participants who are illiterate should include their thumb-print as well.

I have witnessed the accurate reading of the consent form to the potential participant, and the individual has had the opportunity to ask questions. I confirm that the individual has given consent freely.

Print name of witness_____ AND

Thumb print of participant

Signature of witness _____

Date _____

Day/month/year

Statement by the researcher/person taking consent

I have accurately read out the information sheet to the potential participant, and to the best of my ability made sure that the participant understands that the following will be done:

1. We will free of charge help you monitor your healthy condition and record your clinical information and healthy/disease status or disease progression process.

2. These data from hospital electronic medical records will be used for scientific research.

3. The collected fecal and urea samples will be used for scientific research.

I confirm that the participant was given an opportunity to ask questions about the study, and all the questions asked by the participant have been answered correctly and to the best of my ability. I confirm that the individual has not been coerced into giving consent, and the consent has been given freely and voluntarily.

A copy of this ICF has been provided to the participant.

Print Name of Researcher/person taking the consent_____

Signature of Researcher /person taking the consent_____

Date _____

Day/month/year

Supplementary Figure 1

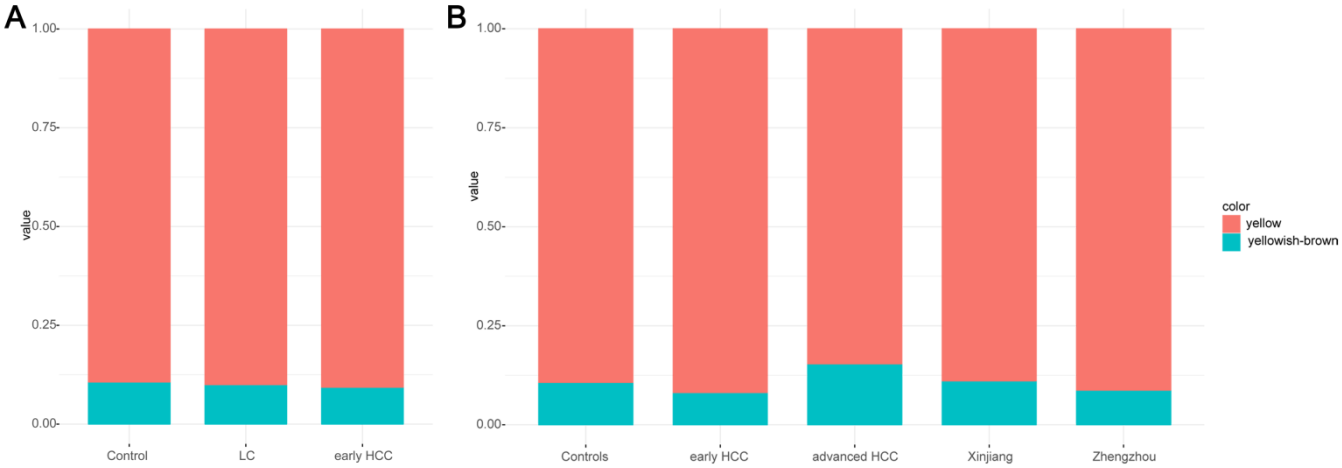

Figure S1. The composition of stool colour from the different cohorts based on stool routine testing. Most of stool samples presented yellow, and showed no significant difference among the different cohorts. **(A)** The composition of stool colour among healthy controls, liver cirrhosis patients and early HCC patients from the discovery phase based on stool routine testing. **(B)** The composition of stool colour among healthy controls, early HCC patients, advanced HCC patients, Xinjiang HCC patients and Zhengzhou HCC patients from the validation phase and the independent diagnosis phase based on stool routine testing. LC, liver cirrhosis.

Supplementary Figure 2

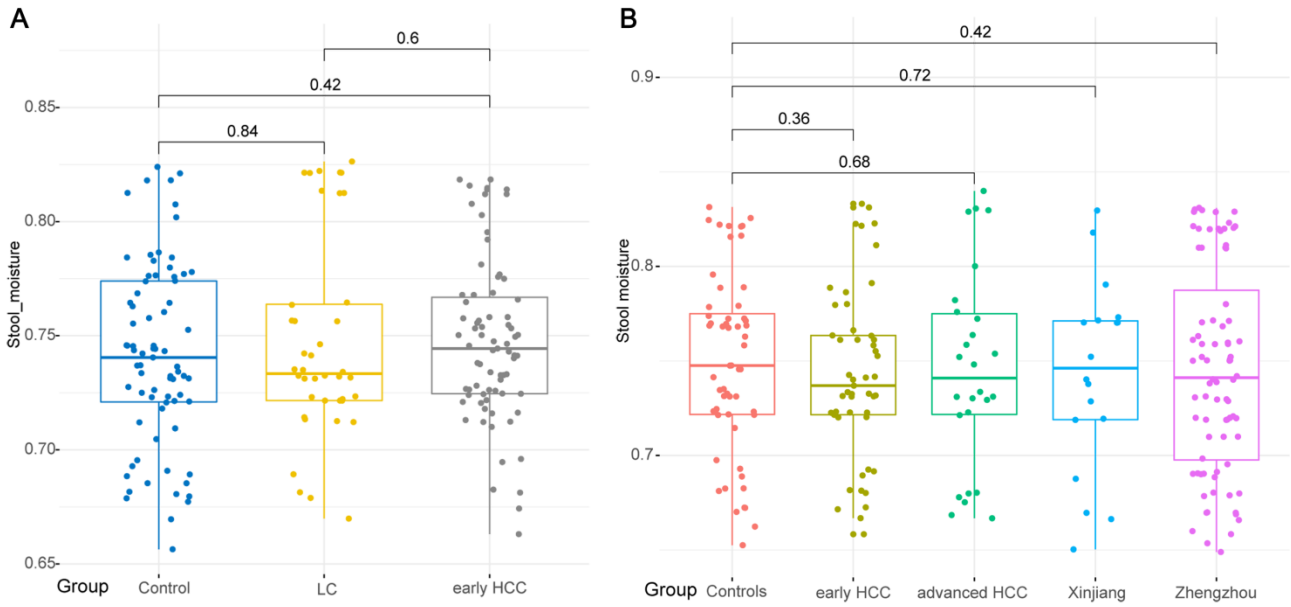

Figure S2. The abundance and distribution of stool moisture among the different cohorts by lyophilization assay on the frozen homogenized faecal material. Stool moisture content was determined in duplicate on the frozen homogenized faecal material (−80 °C) as the percentage of stool mass loss after lyophilization. **(A)** The abundance of stool moisture among healthy controls, liver cirrhosis patients and early HCC patients from the discovery phase. **(B)** The abundance of stool moisture among healthy controls, early HCC patients, advanced HCC patients, Xinjiang HCC patients and Zhengzhou HCC patients from the validation phase and the independent diagnosis phase. LC, liver cirrhosis.

Supplementary Figure 3

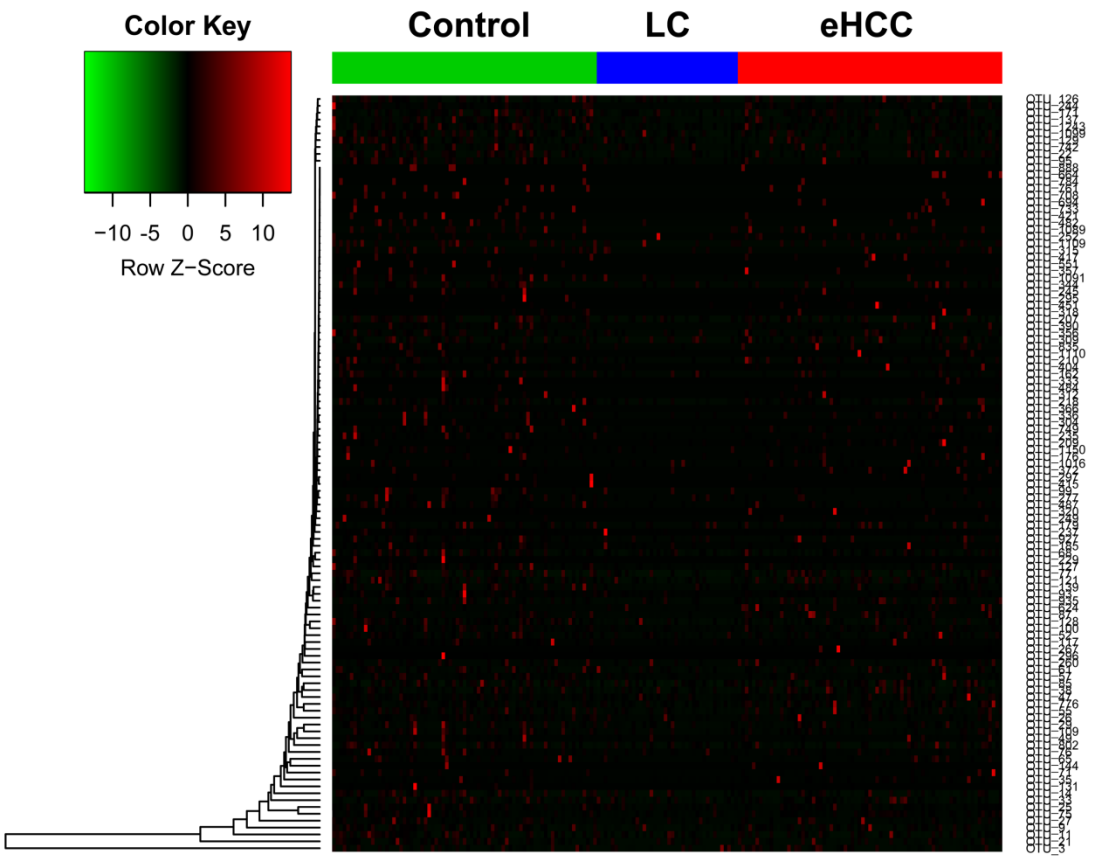

Figure S3. Identification of key OTUs phylotypes among health controls (N=75), liver cirrhosis (N=40) and early HCC with cirrhosis (N=75). To identify key OTUs phylotypes in early HCC, a total of 110 OTUs were identified as key lineages for early HCC by Wilcoxon rank-sum test with Benjamini-Hochberg method, and their abundance and distribution were shown. To show the distribution of OTUs with lower abundance, the colored squares of each row were scaled to denote the relative ratio of each OUT among 190 individuals. LC, liver cirrhosis; eHCC, early HCC with cirrhosis.

Supplementary Figure 4

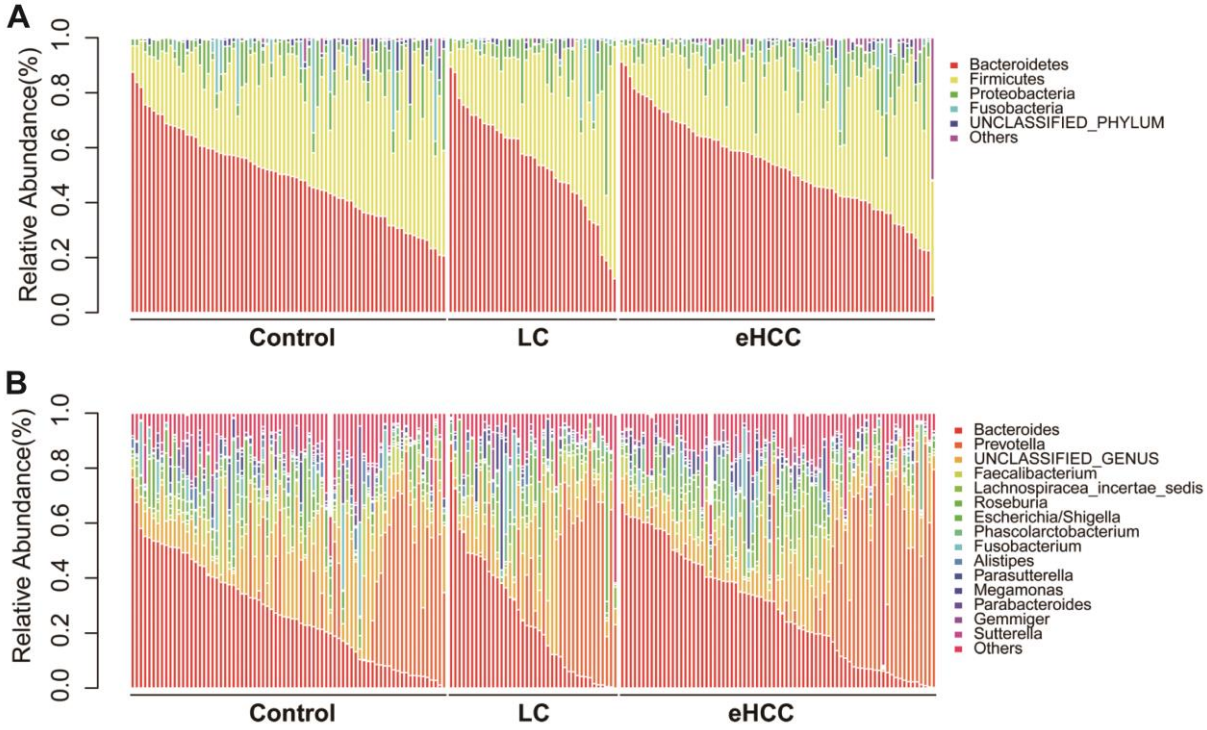

Figure S4. Composition of fecal microbiota on taxonomic level among health controls (N=75), liver cirrhosis (N=40) and early HCC with cirrhosis (N=75). Composition of fecal microbiota among the three groups at the phylum level (A) and the genus level (B). LC, liver cirrhosis; eHCC, early HCC with cirrhosis.

Supplementary Figure 5

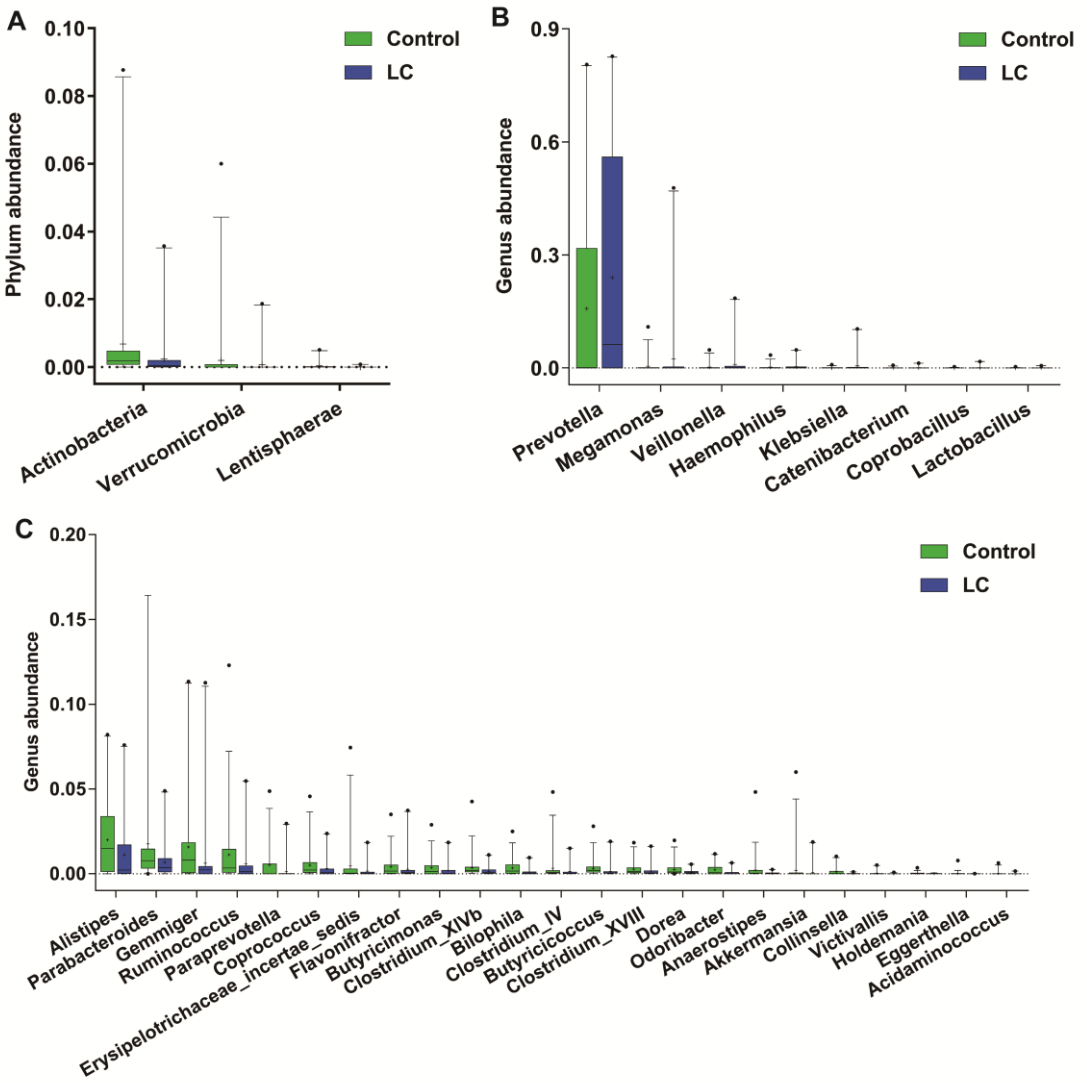

Figure S5. Bacterial difference of gut microbes between liver cirrhosis (N=40) and healthy controls (N=75). (A) The decreased microbial community at the phylum level in liver cirrhosis versus health controls; (B) The increased microbial community at the genus level in liver cirrhosis versus health controls; (C) The decreased microbial community at the genus level in liver cirrhosis versus health controls. LC, liver cirrhosis. The box presented the 95% confidence intervals; the line inside denotes the median, and the symbol "+" denotes the mean value.

Supplementary Figure 6

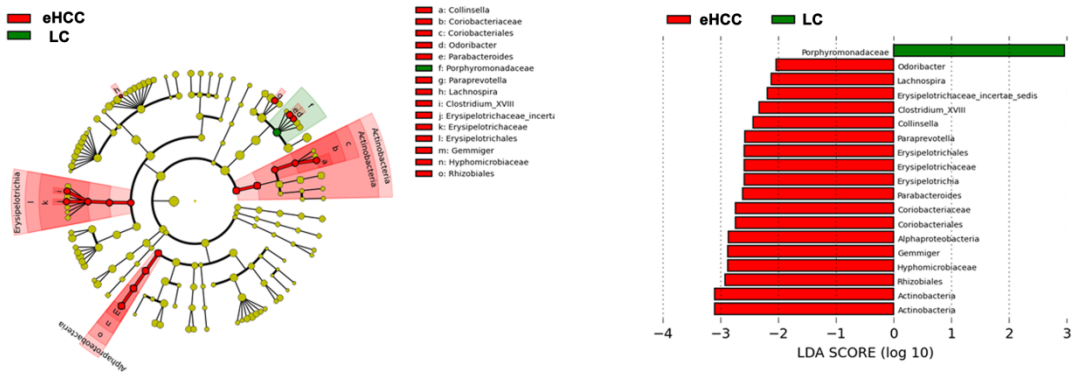

Figure S6. The specific characterization of fecal microbiota to distinguish toxigenic types was analyzed by linear discriminant analysis (LDA) effect size (LEfSe) method (http://huttenhower.sph.harvard.edu/lefse/) **between early HCC patients with cirrhosis (N=75) and liver cirrhosis (N=40).** (A) LEfSe method identified the most differentially abundant taxons between early HCC patients with cirrhosis and liver cirrhosis. The brightness of each dot is proportional to its effect size. (B) The early HCC-enriched taxa are indicated with a negative LDA score (red), and liver cirrhosis-enriched taxa present a positive score (green). Only taxa achieving an LDA significant threshold >2 are shown. eHCC, early HCC with cirrhosis; LC, liver cirrhosis.

Supplementary Figure 7

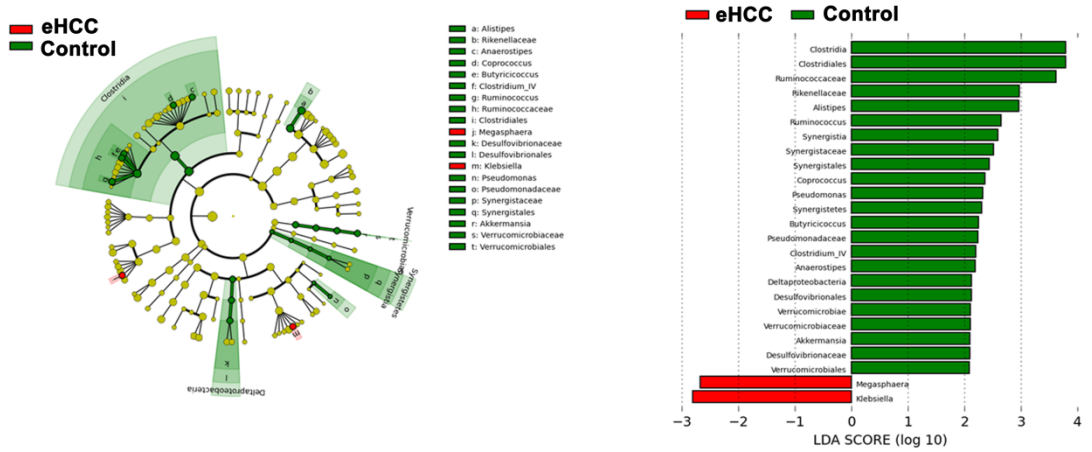

Figure S7. The specific characterization of fecal microbiota to distinguish toxigenic types was analyzed by linear discriminant analysis (LDA) effect size (LEfSe) method between early HCC patients with cirrhosis (N=75) and health controls (N=75). (A) LEfSe method identified the most differentially abundant taxons between early HCC patients with cirrhosis and health controls. The brightness of each dot is proportional to its effect size. (B) The early HCC-enriched taxa are indicated with a negative LDA score (red), and health control-enriched taxa present a positive score (green). Only taxa achieving an LDA significant threshold >2 are shown. eHCC, early HCC with cirrhosis.

Supplementary Figure 8

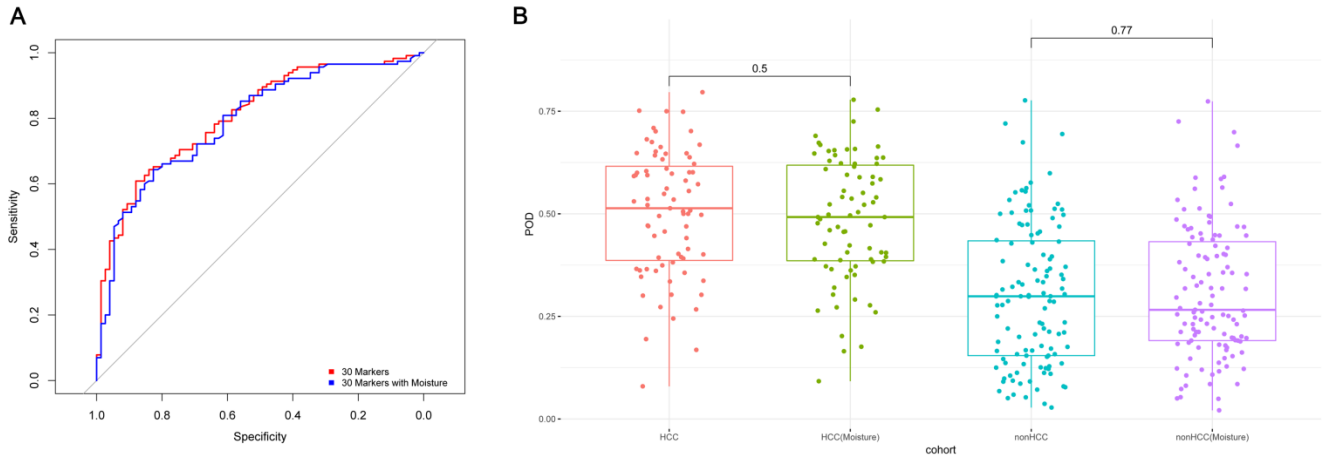

Figure S8. Diagnosis efficacy comparison of the 30 microbial markers with moisture or without moisture between early HCC samples (n=75) and non-HCC samples (40 cirrhosis and 75 healthy controls). (A) The AUC values comparison of the 30 microbial markers with moisture or without moisture between early HCC samples (n=75) and non-HCC samples (40 cirrhosis and 75 healthy controls). **(B)** The POD values comparison based on the 30 microbial markers with moisture or without moisture for early HCC samples (n=75) and non-HCC samples (40 cirrhosis and 75 healthy controls). AUC, the area under the curve. POD, probability of disease.

Supplementary Figure 9

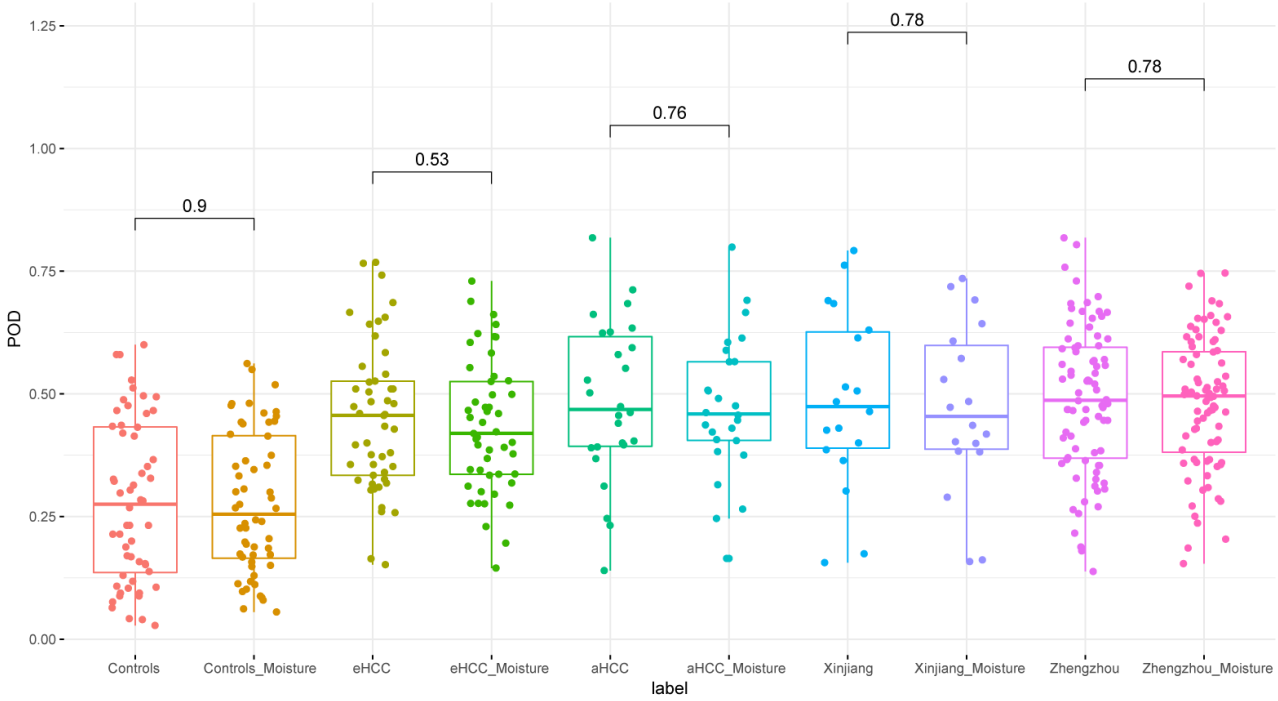

Figure S9. Diagnosis efficacy comparison of the POD values based on the 30 microbial markers with moisture or without moisture among the 56 controls, 30 early HCC samples, 45 advanced HCC samples, 18 Xinjiang HCC samples and 80 Zhengzhou HCC samples, respectively. The POD values based on the 30 OTUs markers with moisture or without moisture showed no obvious difference in the different cohorts from the validation phase and the independent diagnosis phase. POD, probability of disease.

Supplementary Figure 10

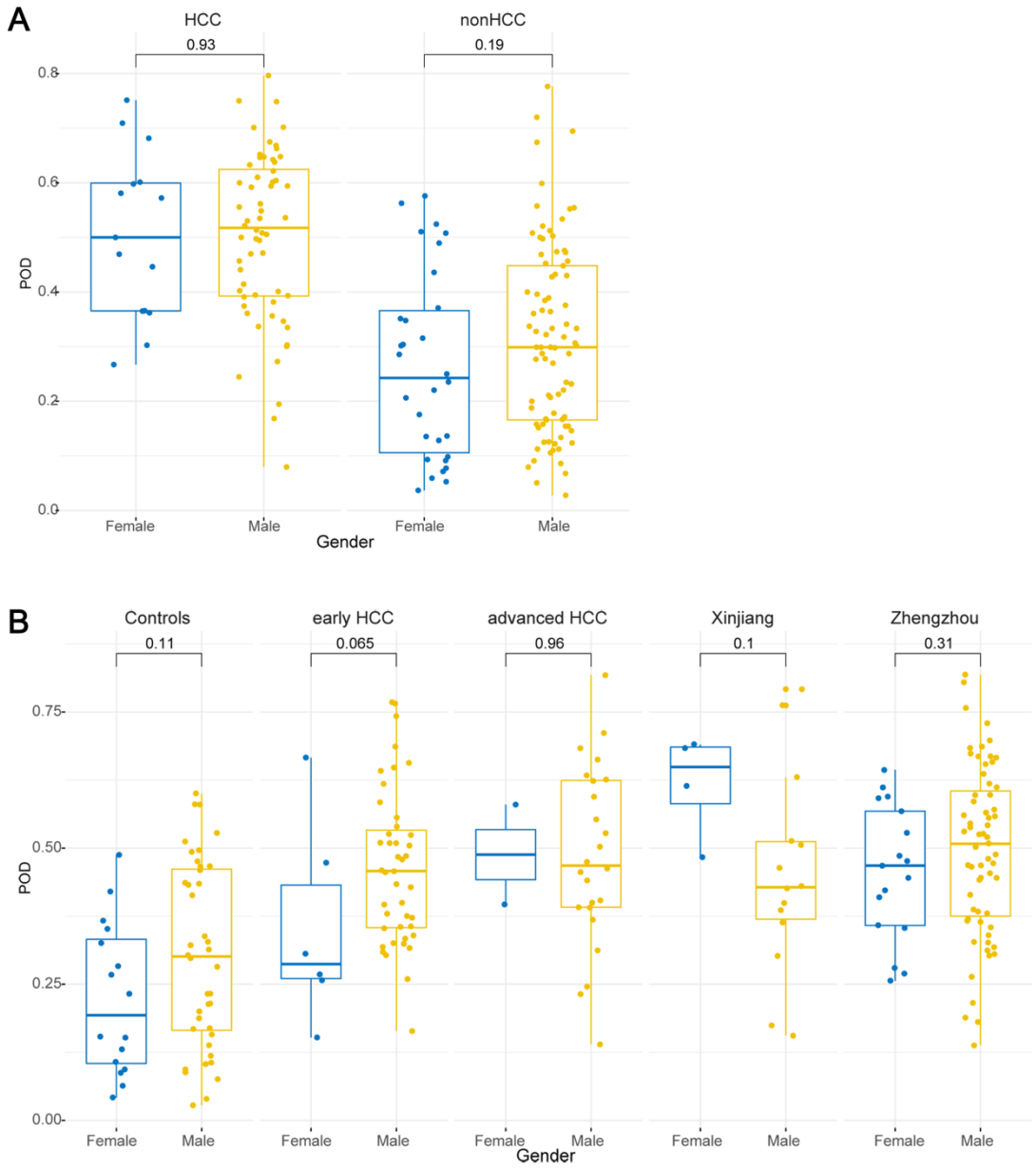

Figure S10. Diagnosis efficacy of the POD based on the 30 microbial markers for HCC between female and male participants in the different cohorts. (A) Diagnosis efficacy of the POD based on the 30 microbial markers for HCC between female and male participants in the early HCC samples (n=75) and non-HCC samples (n=105). **(B)** Diagnosis efficacy of the POD based on the 30 microbial markers for HCC between female and male participants in the 56 controls, 30 early HCC samples, and 45 advanced HCC samples from the validation phase, in the 18 Xinjiang HCC samples and 80 Zhengzhou HCC samples from the independent diagnosis phase. POD, probability of disease.