



OPEN

Meta-omics characteristics of intestinal microbiota associated to HBeAg seroconversion induced by oral antiviral therapy

Yu-Li Zeng^{1,10}, Lei Qin^{2,10}, Wen-Jun Wei^{1,10}, Hong Cai³, Xiao-Fang Yu³, Wei Zhang⁴, Xiao-Lu Wu⁵, Xiao-Bin Liu⁵, Wei-Ming Chen¹, Pan You⁶, Mei-Zhu Hong⁷, Yaming Liu¹, Xuan Dong¹, Ben-Chang Shia⁸✉, Jian-Jun Niu⁶✉ & Jin-Shui Pan⁹✉

Tenofovir and entecavir are currently designated as the preferred oral antiviral drugs for chronic hepatitis B. However, only less than 40% of patients can achieve HBeAg seroconversion. We aim at investigating the role of intestinal microbiome in HBeAg seroconversion induced by oral antiviral therapy and describe multi-omics characteristics of HBeAg seroconversion associated intestinal flora. In this study, we prospectively collected fecal samples at baseline from the patients with HBeAg positive chronic hepatitis B who would have oral antiviral therapy. 16S rDNA sequencing and metabolomics were performed. We identified HBeAg seroconversion-related microbial signature and constructed prediction model for HBeAg seroconversion. Thirty-seven of these subjects achieved HBeAg seroconversion within 156 weeks after the initiation of oral antiviral therapy, while 41 subjects remained HBeAg positive even after over 156 weeks of therapy. A computational statistical and machine learning approach allowed us to identify a microbial signature for HBeAg seroconversion. Using random forest method, we further constructed a classifier based on the microbial signature, with area under curve being 0.749 for the test set. Patients who achieved HBeAg seroconversion tended to have lower abundance of certain fecal metabolites such as essential amino acids, and several dipeptides. By analyzing the fecal microbiota from the patients with and without HBeAg seroconversion, we showed intestinal microbiome play a critical role in HBeAg seroconversion induced by oral antiviral therapy. We also identified intestinal microbial signature that is associated with HBeAg seroconversion after oral antiviral therapy.

Abbreviations

CHB	Chronic hepatitis B
HBV	Hepatitis B virus
TDF	Tenofovir disoproxil fumarate
HBeAg	HBV e-antigen
TLR4	Toll-like receptor 4
OTU	Operational taxonomic unit
AUC	Area under the curve
ALT	Alanine aminotransferase
AST	Aspartate aminotransferase

¹Department of Gastroenterology, Zhongshan Hospital Xiamen University, Xiamen, Fujian, China. ²School of Statistics, University of International Business and Economics, Chaoyang District, Beijing, China. ³Department of Hepatology, Xiamen Hospital of Traditional Chinese Medicine, Xiamen, Fujian, China. ⁴Department of Research Institute, The Fifth Hospital of Shijiazhuang, Shijiazhuang, Hebei, China. ⁵Department of Infectious Diseases, The First Affiliated Hospital of Xiamen University, Xiamen, Fujian, China. ⁶Center of Clinical Laboratory, Zhongshan Hospital Xiamen University, Xiamen, Fujian, China. ⁷Department of Traditional Chinese Medicine, Zhongshan Hospital Xiamen University, Xiamen, Fujian, China. ⁸Graduate Institute of Business Administration, College of Management, Fu Jen Catholic University, New Taipei City 24205, Taiwan. ⁹Liver Research Center, The First Affiliated Hospital of Fujian Medical University, Fuzhou, Fujian, China. ¹⁰These authors contributed equally: Yu-Li Zeng, Lei Qin, and Wen-Jun Wei. ✉email: stat1001@tmu.edu.tw; niujianjun62@163.com; j.s.pan76@gmail.com

PC Principal component
LPS Lipopolysaccharide

For hepatitis B virus (HBV) e-antigen (HBeAg)-positive CHB patients, HBeAg seroconversion is a prerequisite for a definite course of oral antiviral treatment^{1,2}. Unfortunately, the rate of HBeAg seroconversion following 1 year of treatment is less than ideal for currently available antivirals. The seroconversion rates for pegylated interferon and entecavir are 29–36%^{3,4}, and 21%⁵, respectively. After 5 years of therapy, the ratio of HBeAg seroconversion related to entecavir slightly increases to 23%⁶. Similarly, HBeAg seroconversion occurs in 21% of patients following 48 weeks of Tenofovir disoproxil fumarate (TDF) treatment⁷, and the rate increases to 29.9% after 5 years of treatment⁸. Thus, for both entecavir and TDF, HBeAg clearance or seroconversion only occurs in a minority of patients even after multiple years of antiviral therapy. Furthermore, there are few or no additional benefits for extending the antiviral therapy if the patients fail to achieve HBeAg clearance or seroconversion within the first year. Thus, it is critical to screen out patients who may achieve HBeAg seroconversion before the initiation of oral antiviral therapy and provide alternative treatment such as interferon-based therapy for those who show fewer chances for HBeAg seroconversion.

Recently, Chou et al.⁹ found that the intestinal microbiota plays a critical role in the age-related immune clearance of HBV. Consistent with this, the study by Ren et al.¹⁰ indicated that manipulation of gut microbiota might accelerate HBeAg clearance in the patients who remain HBeAg-positive after years of antiviral therapy. The “leakage hypothesis” proposed by Tremaroli et al.¹¹ linked the presence of microbial products in the bloodstream with the onset and progression of liver diseases. Even though intestinal microbiota seems to play a pivotal role in the clinical outcome of CHB, no uniform pattern of gut microbiota related to CHB has been confirmed. In this study, we therefore aimed to determine the link between intestinal microbiota and HBeAg seroconversion in a well-characterized cohort of CHB patients.

Materials and methods

Study approval. The Medical Ethics Committee of the Medical College of Xiamen University approved this survey. All participants provided written informed consents. Patients or the public WERE NOT involved in the design, or conduct, or reporting, or dissemination of our research. All methods were performed in accordance with the relevant guidelines and regulations.

Study design. The current study was designed as prospective specimen collection and retrospective blinded evaluation. From July 2015 to February 2016, we prospectively collected fecal samples at baseline from the patients with HBeAg positive CHB who would have oral antiviral therapy. Data analyst was blinded to clinical outcome of the patients and design of the study. Enrolled patients were randomly divided into training set and test set. We identified microbial signature and constructed classifier for HBeAg seroconversion using training set while validated the results using test set. Three groups were included in this survey: Group H, healthy participants; Group P, who remained HBeAg positive after more than 156 weeks of ETV- or TDF-based therapy; and Group N, who achieved HBeAg seroconversion at 12–156 weeks after ETV- or TDF-based therapy. Patients were ruled out from further analysis in the following cases, for example, who achieved HBeAg clearance while have not anti-HBeAg positive after more than 156 weeks of ETV- or TDF-based therapy. The study design and the construction of prediction model were shown in Supplementary Fig. 1.

Study population. Eligible subjects were 18–50-year-old patients with HBeAg-positive CHB, who were anti-HBeAg negative before oral antiviral therapy and were undergoing ETV- or TDF-based antiviral therapy when they were recruited. Other eligibility criteria included: $18.5 \leq \text{body mass index (BMI)} < 27.0$; achieved HBeAg seroconversion that occurred at 12–156 weeks after ETV- or TDF-based therapy or remained HBeAg positive after more than 156 weeks of ETV- or TDF-based therapy; international normalized ratio ≤ 1.5 . The exclusion criteria for the study were: alcoholism; presence of genetic or metabolic liver diseases, such as Wilson’s disease and hemochromatosis; co-infection with hepatitis C virus, hepatitis D virus, human immunodeficiency virus, Epstein–Barr virus, or cytomegalovirus; concurrent malignant tumors; autoimmune liver diseases; liver cirrhosis; moderate or severe nonalcoholic fatty liver disease or non-alcoholic steatohepatitis confirmed by ultrasound, computed tomography, or FibroScan; pregnant or lactating women; concurrent conditions such as diabetes mellitus, hypertension with blood pressure $\geq 160/100$ mmHg, heart failure, and renal impairment; history of acute gastroenteritis within the last 8 weeks before enrollment. Patients who had a history of other antiviral treatment such as interferon or pegylated interferon, lamivudine, adefovir, or telbivudine were excluded. Patients who reported previous usage of antibiotics, and probiotics within the last 8 weeks were also excluded. Liver cirrhosis or hepatocellular carcinoma (HCC) was diagnosed based on comprehensive integration of clinical symptoms and physical signs, laboratory tests and medical imaging such as ultrasound, computed tomography, magnetic resonance imaging, and liver stiffness measurement, according to widely accepted guidelines. The diagnosis of HCC was further confirmed by histopathological examination. The detailed eligibility and exclusion criteria for the study are provided in Supplementary Table 1.

Assessment of HBV markers and clinical outcomes. HBV markers were determined by several methods. For the HBsAg measurement, detection kit from Boson Biotech (Xiamen, Fujian, China; lower limit of detection of 0.1 IU/mL) was used. For the HBeAg and anti-HBeAg measurement, detection kit from Boson Biotech (Xiamen, Fujian, China) was used. The primary clinical outcome assessed in this study was HBeAg seroconversion, which was defined as HBeAg clearance (< 1.0 S/CO) along with positive anti-HBeAg. HBeAg and anti-HBeAg were doubly verified using a qualitative kit from InTec Products (Xiamen, Fujian, China). The status

of HBeAg and anti-HBeAg was repeatedly determined with an interval of 12 weeks. If there were inconsistent in the two times of tests of HBeAg and anti-HBeAg, the observation time was extended until the results at two time points were consistent. The level of HBV DNA was determined by polymerase chain reaction (PCR) using a kit from Sansure Biotech (Changsha, Hunan, China), with a detection limit of 200 IU/mL. All laboratory tests were performed at the Clinical Laboratory, Zhongshan Hospital Affiliated to Xiamen University.

Assessment and analysis of microbiota. Gut microbiota profiling was conducted by extracting genomic DNA from the stool of participants using the TIANamp Bacteria DNA Kit (TIANGEN, Haidian, Beijing, China). Microbial Sequencing was performed by Xiamen Anjie Medical Data Technology Co., Ltd. The 16S rDNA sequencing was performed on Illumina HiSeq platform. Ambiguous nucleotides, barcode and primer errors, homopolymeric nucleotides, and chimeras were filtered. FLASH was used for DNA assembling and quality control. The OUT is defined based on the similarity threshold of 97%. The OTU clustering was run using QIIME. The rarefaction curve was used to determine whether the sequencing depth was sufficient, which was run on QIIME. After genomic DNA extraction, the gene encoding for the V4 region of the 16S ribosomal DNA was amplified by PCR using universal primers. The universal primer pair contained index sequence and adaptor compatible for sequencing on a HiSeq2500 PE250 platform (Illumina, San Diego, California, USA). Through sequencing we obtained up to 70,453 clean reads of fecal microbiota per sample (median = 46,572). Prevalence of each genus from the gut microbiota signature for HBeAg seroconversion was estimated in both the subjects with and without HBeAg seroconversion.

Metabolic profiling and phenotyping. Metabolic profiling was acquired using UHPLC-QTOF-MS (Agilent, Santa Clara, California, USA). First of all, metabolites were detected under both positive ion (POS) mode and negative ion (NEG) mode through relative standard deviation de-noising method. Half of the minimum value was used to fill up the missing values of raw data. Internal standard normalization method was used in this analysis. In order to perform principal component analysis (PCA), and orthogonal projections to latent structures-discriminate analysis (OPLS-DA), three-dimensional data involving the peak number, sample name, and normalized peak area were imported into SIMCA14.1 software package (V14.1, Sartorius Stedim Data Analytics AB, Umea, Sweden, <https://landing.umetrics.com/downloads-simca>). Metabolomics analysis was performed using the statistical correlation include the false discovery rate (FDR) correction. Figure based on PCA showed the distribution of origin data. In order to obtain a higher level of group separation and get a better understanding of variables responsible for classification, supervised OPLS-DA were applied. Seven-fold cross validation was used to estimate the robustness and the predictive ability of our model, such permutation test was proceeded in order to further validate the model. The R and Q intercept values were calculated after 200 permutations. The robustness of the models was measured by the low values of Q intercept, which had a low risk of over fitting and was reliable. A loading plot was constructed based on OPLS-DA, which was employed to show the contribution of variables to difference between two groups. The loading plot was also used to the important variables situated far from the origin. However, in case of many variables, the loading plot would be complex. The first principal component of variable importance in the projection (VIP) was obtained to refine this analysis. Metabolites with VIP values exceeding 1 were first selected as the changed metabolites. In step 2, the remaining variables were then assessed by Student's t-test ($P > 0.05$), variables were discarded between two comparison groups. In addition, commercial databases including KEGG <http://www.genome.jp/kegg/> and MetaboAnalyst <http://www.metaboanalyst.ca/> were utilized to search for the pathways of metabolites^{12,13}.

Correlation between metabolic profiling and HBeAg seroconversion. For each set of comparisons, Euclidean distance matrix for the quantitative values of the differentiated metabolites was calculated. The differentiated metabolites were clustered by using complete linkage method and was displayed as a heatmap of hierarchical clustering analysis. Using KEGG database, differentiated metabolites were mapped to several specific pathways. The results of the metabolic pathway analysis were shown in a bubble plot.

Correlation between microbiota and metabolic profiling. The correlation of differentiated metabolites between Group P and Group N was calculated by using the “spearman” algorithm. Correlation coefficient (Corr) was stored. Intestinal flora profiling with different abundance at genus level was acquired for subsequent association analysis. Heatmap was employed to indicate the correlation analysis between intestinal flora profiling and metabolic profiling with red indicated a Corr of 1 while blue indicated a Corr equaled to -1. Data of the Corr with statistical difference were marked with “*” in the graph. Figure of correlation network was constructed by using Cytoscape_v3.7.1 software¹⁴.

Statistical analysis. For microbiota profiling, one-way analysis of variance (ANOVA) was used to compare the means of relative abundance of bacterial phylum between the three groups. Least significant difference (LSD) method was applied to conduct multiple comparisons between pairs of groups. Diversity within samples, α -Diversity, was assessed by the richness as measured by number of operational taxonomic units (OTUs, 97% identity) at the same sequencing depth using the vegan R package. Diversity between samples, β -Diversity, was assessed by UniFrac distance, as well as the Bray-Curtis distance using Phyloseq R package¹⁵. Beta diversity between samples was also assessed by PCA and non-metric multidimensional scaling based on Hellinger transformation of relative abundance of all OTUs. One-way ANOVA was used to compare the diversity between the three groups. Microbiota clustering, enterotype stratification, was performed in fecal samples from CHB patients and healthy subjects using the Dirichlet multinomial mixture model¹⁶.

For microbial signature, univariate Wilcoxon test was used to select the original significant genera set between Group N and Group P. Random forest was implemented to select the least number of genera that distinguish Group N and Group P with appropriate accuracy on test set. Robustness of this method was validated by randomly splitting data into training and test sets. The area under the curve (AUC) was used to assess the discriminative ability of this prediction model.

To evaluate the robustness of the relationship between the microbial signatures for HBeAg seroconversion and other clinical parameters, we investigated the relative abundance of significant genera with clinical data in a co-inertia analysis. Relative abundance of microbial signature was transformed by Hellinger transformation, and clinical parameters including age, BMI, HBV DNA level, ALT and AST levels, and time to HBeAg seroconversion were transformed by logarithm function.

PCA was used to explore distribution of significant genera among all the genera and to analyze the relationship between HBeAg seroconversion, clinical data, and microbial signature. The relationship between variables was assessed using Spearman correlation test for continuous variables [e.g., principle component score and clinical data such as age, BMI, and levels of alanine aminotransferase (ALT), aspartate aminotransferase (AST), and HBV DNA]. Network between microbial signatures was plotted among Group N and Group P. For KEGG analysis, significant correlations between pathways of interest and microbial signature were detected to find the genetic cause.

Results

Clinical characteristics of CHB patients. In the enrolled patients, 37 of them achieved HBeAg seroconversion within 156 weeks after the initiation of oral antiviral therapy, while 41 subjects remained HBeAg positive even after over 156 weeks of therapy. Detailed clinical and demographic characteristics of the enrolled patients with CHB with or without HBeAg seroconversion (i.e., Groups N and P, respectively) are summarized in Supplementary Table 2. No differences were observed between Groups N and P with regards to body mass index and antiviral choices. Compared with the patients of Group P, the Group N individuals tended to have lower baseline age and HBV DNA, while higher baseline ALT and AST levels. More women were observed to fall in Group N than in Group P.

Microbiota diversity in fecal samples. Regarding the phylum relative abundance, ANOVA indicated no significant difference between the main bacteria in Group N and Group P. *Bacteroidetes* and *Firmicutes* were the phyla with the most abundance in three groups (Supplementary Fig. 2A). Compared with healthy subjects (Group H), no significant difference was observed in the numbers of OTUs in Groups N and P (Supplementary Fig. 2B). Classical methods of β -diversity such as principal component analysis and Non-Metric Multi-Dimensional Scaling on relative abundance of all OTUs could not find significant differences between Groups H, N, and P (Supplementary Fig. 3).

Enterotype clustering in CHB patients and healthy subjects. Microbiota was found to separate optimally into two distinct microbiota communities as assessed by Laplace parameter and Akaike information criterion (Supplementary Fig. 4). One enterotype was enriched in *Bacteroides* (50% of samples) while the other one was enriched in *Prevotella* (52% of samples) (Fig. 1A). Notably, *Prevotella* enterotyped subjects harbored higher diversity, as compared with *Bacteroides* enterotyped subjects (Fig. 1B). In Group N, the proportion of *Prevotella* enterotyped subjects was higher than that in Group P (Fig. 1C).

Identification of the microbial signature for HBeAg seroconversion. We further explored the association between HBeAg seroconversion and fecal microbiota composition. Wilcoxon test was conducted to identify the significant genera in paired groups (P vs. H, N vs. H, and N vs. P), and this procedure was repeated 100 times on bootstrapped sample sets. About 18 (median) genera could discriminate Group N from Group P, while about 21 (median) genera could distinguish Group N from Group H, and 22 (median) genera distinguished Group P from Group H (Fig. 2A). Wilcoxon test identified microbial signatures between Groups N and P. However, this univariate analysis omitted the correlation between genera and gave a set with a large number of significant genera.

We then tested the random forest method for identification of the microbial signature. Firstly, Wilcoxon test was used to identify 18 genera that showed significant differences between Groups N and P. Then we modelled the HBeAg seroconversion and these significant genera with random forest; the relative importance measured for each genus was given in a descending order. Secondly, we separately assessed the classification accuracy of the first k ($k=1, 2, 3, \dots, 13, 14, 15$) genera. We randomly split all the data into training and test sets in a 3:1 ratio, predicted the HBeAg seroconversion with first k genera on the training set and then validated the model on test set. The process was repeated 100 times to calculate the mean classification accuracy rate on test data. In this step, 13 genera were selected as microbial signature for HBeAg seroconversion to guarantee about 71% accuracy rate on test data (Fig. 2B). The number of significant genera selected by random forest was far less than that obtained by Wilcoxon test, indicating the better performance of variable selection by the random forest method. Boxplot of correct classification rate on training data and test data was also given by 100 times repetition of the calculation at second step. The mean of the accuracy rate on test data was lower than that on training data, and the corresponding standard deviation was higher than that obtained with the training data (Fig. 2C). The distribution of the accuracy rate on test data suggested that this method is robust. Thirdly, receiver operating characteristic was obtained for the evaluation of the classification model based on microbial signature. The area under the curve (AUC) was 74.9%, which indicated a significant discriminative ability (Fig. 2D). Finally, we assessed the classification accuracy with different sample size based on repeat sampling. With increasing

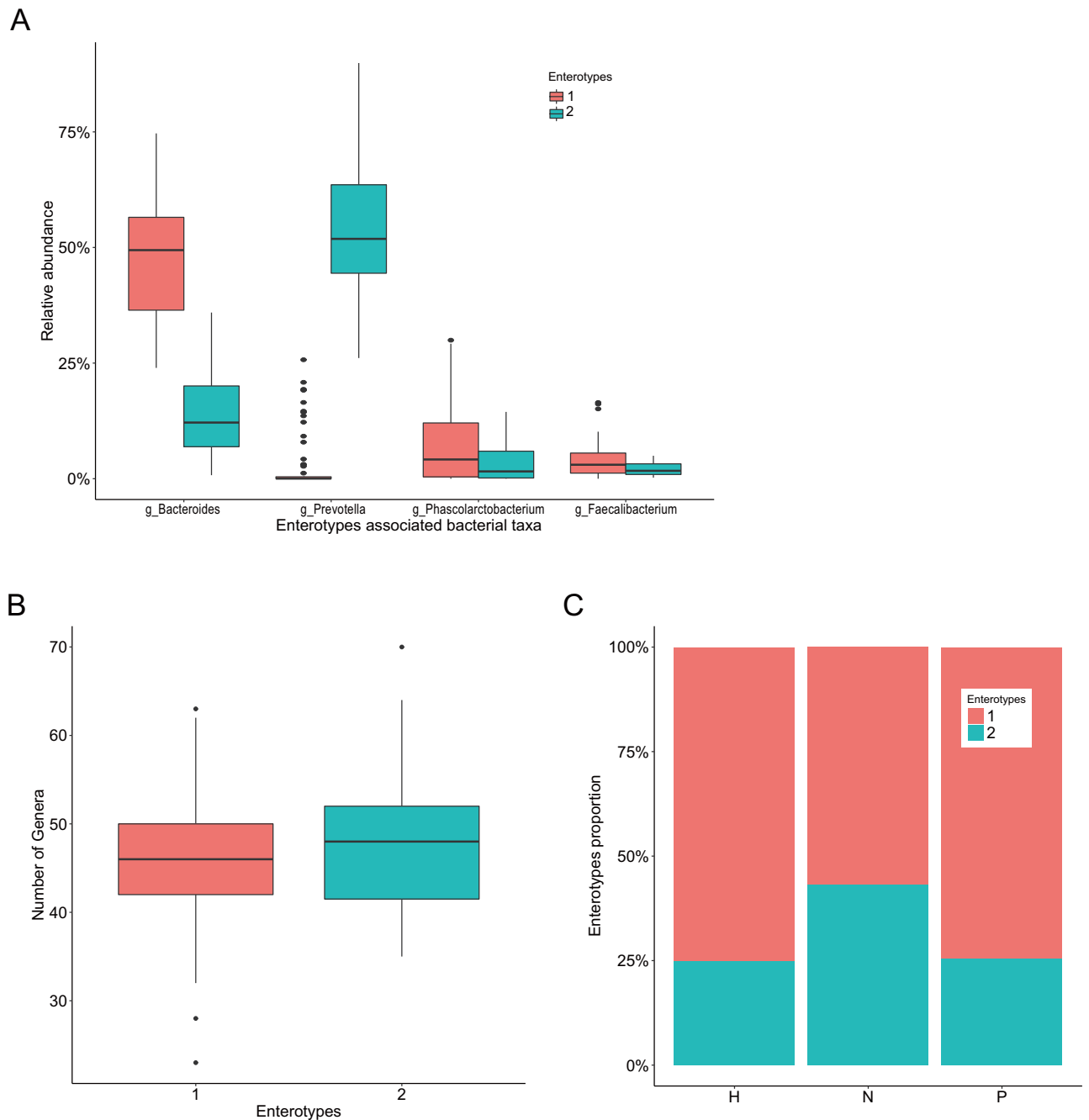


Figure 1. Enterotype clustering in the patients with CHB and healthy controls. (A) Two distinct microbiota communities are clustered with one enterotype is enriched in *Bacteroides* (50% of samples) while the other one is enriched in *Prevotella* (52% of samples). (B) *Prevotella* enterotyped subjects harbor higher diversity, as compared with *Bacteroides* enterotyped subjects. (C) Group N has higher proportion of *Prevotella* enterotyped subjects than that in Group P.

sample size, the mean classification accuracy rate ranged from 67 to 72%, so we concluded that our sample size was adequate (Supplementary Fig. 5).

Taxonomy of the Gut microbial signature for HBeAg seroconversion. Using random forest, 13 genera were screened out as potential microbial signature from the total 169 genera of Group N and Group P (Fig. 3A). Using Kulczynski distance between the representative sequences of all genera, we performed principal component analysis to assess the distribution of the microbial signature for HBeAg seroconversion. As shown in Fig. 3B, the microbial signature was as taxonomically diverse as the genera from the whole microbiota data set. The microbial signature for HBeAg seroconversion was characterized and the dominant genera were identified to be *Bacteroides*, *Sutterella*, and *Oscillospira* (Fig. 3B). The 13 genera for HBeAg seroconversion were ranked according to their importance in the model (Supplementary Table 3 and Supplementary Fig. 6). In this model,

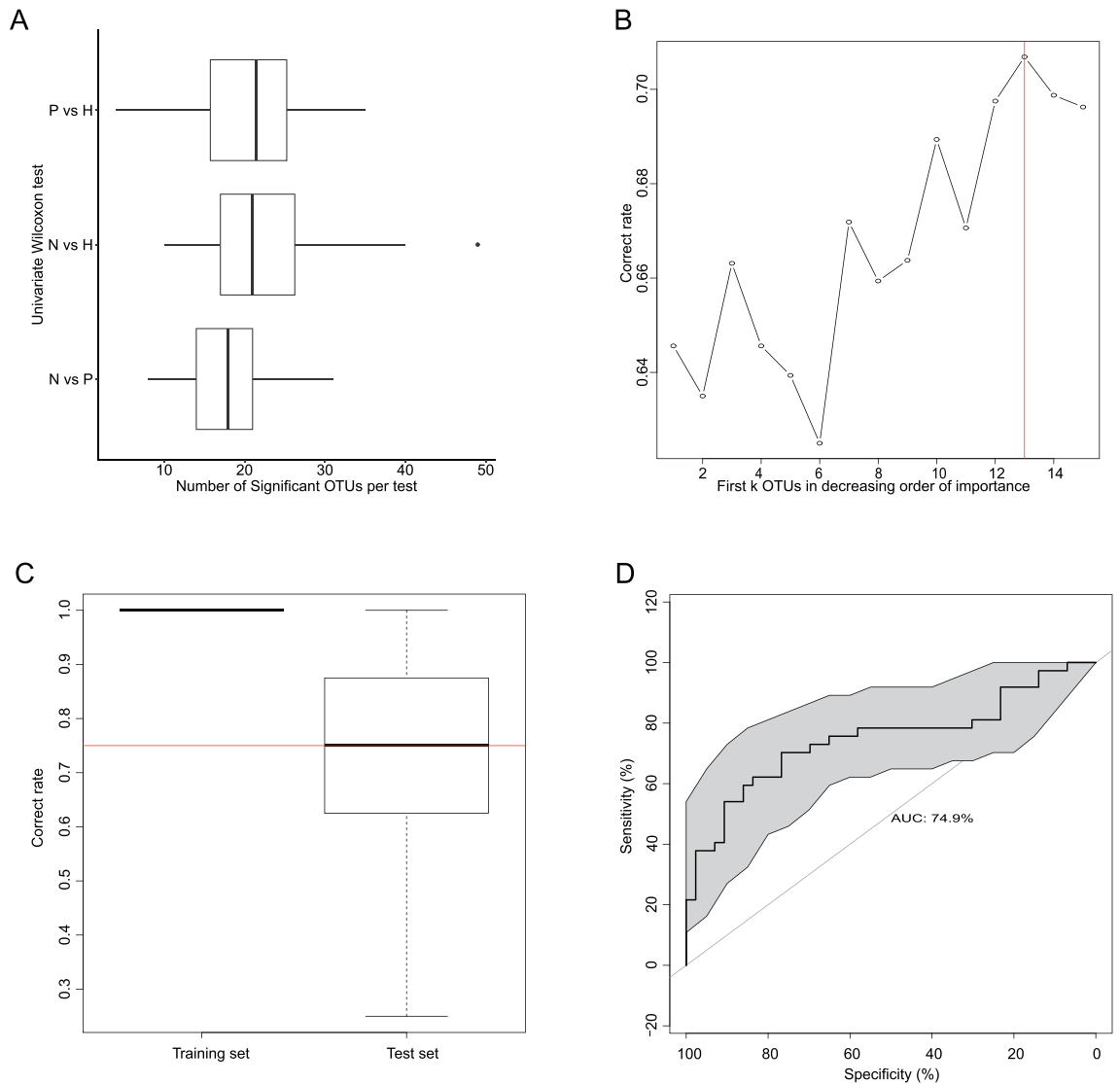


Figure 2. Univariate comparison and machine learning based on HBeAg seroconversion. **(A)** Wilcoxon test is conducted to identify the significant genera in paired groups (P vs. H, N vs. H, N vs. P). This procedure is repeated 100 times on bootstrapped sample set. About 18–22 genera are selected. The number of significant genera for each comparison is reported with boxplot. **(B)** Thirteen genera are selected as microbial signature for HBeAg seroconversion to guarantee about 71% accuracy rate on test data. **(C)** The mean of accuracy rate on test data is about 0.75, indicating this method is robust. **(D)** The area under the ROC curve (AUC) is 74.9%, which indicates the prediction model has a significant discriminative ability.

genera with more importance usually had positive association with HBeAg seroconversion. The importance of the genera was shown in Fig. 3C.

Gut microbial signature for HBeAg seroconversion and association with clinical and microbial parameters.

The first 2 co-inertia principle components (PCs) could explain about 25.3% of co-variation. An undefined *Prevotellaceae*, *un_f_Prevotellaceae*, (PC1 loading -0.27) and *Genus Sutterella* (PC1 loading -0.22) were confirmed to be the most important parameters contributing to variation along the first co-inertia component (PC1), while undefined *Erysipelotrichaceae*, *un_f_Erysipelotrichaceae* (PC2 loading -0.37) and *Genus Bacteroides* (PC2 loading -0.36) were the most important parameters along the second co-inertia component (PC2) (Fig. 4A). The genera that had higher importance in the microbial signature were negatively associated with PC1 (Fig. 4B). Group N showed tendency to have positive PC1 score, while Group P tended to have negative PC1 score (Fig. 4C). Loadings of microbial signatures and clinical parameters along the PC1 and the PC2 were shown in Supplementary Table 4. PC1 was negatively associated with time to HBeAg seroconversion ($\rho = -0.5108$, $P < 0.05$) (Supplementary Fig. 7A). No significant association was observed between HBV DNA and PC2 (Supplementary Fig. 7B). PC1 was positively associated with ALT ($\rho = 0.8060$, $P < 0.05$) and AST ($\rho = 0.8354$, $P < 0.05$) levels (Fig. 4D and Supplementary Fig. 7C, respectively). *Genera un_f_Prevotellaceae* and *un_f_Erysipelotrichaceae* were shown in Supplementary Fig. 8A,B, respectively, while *Genera Sutterella* and

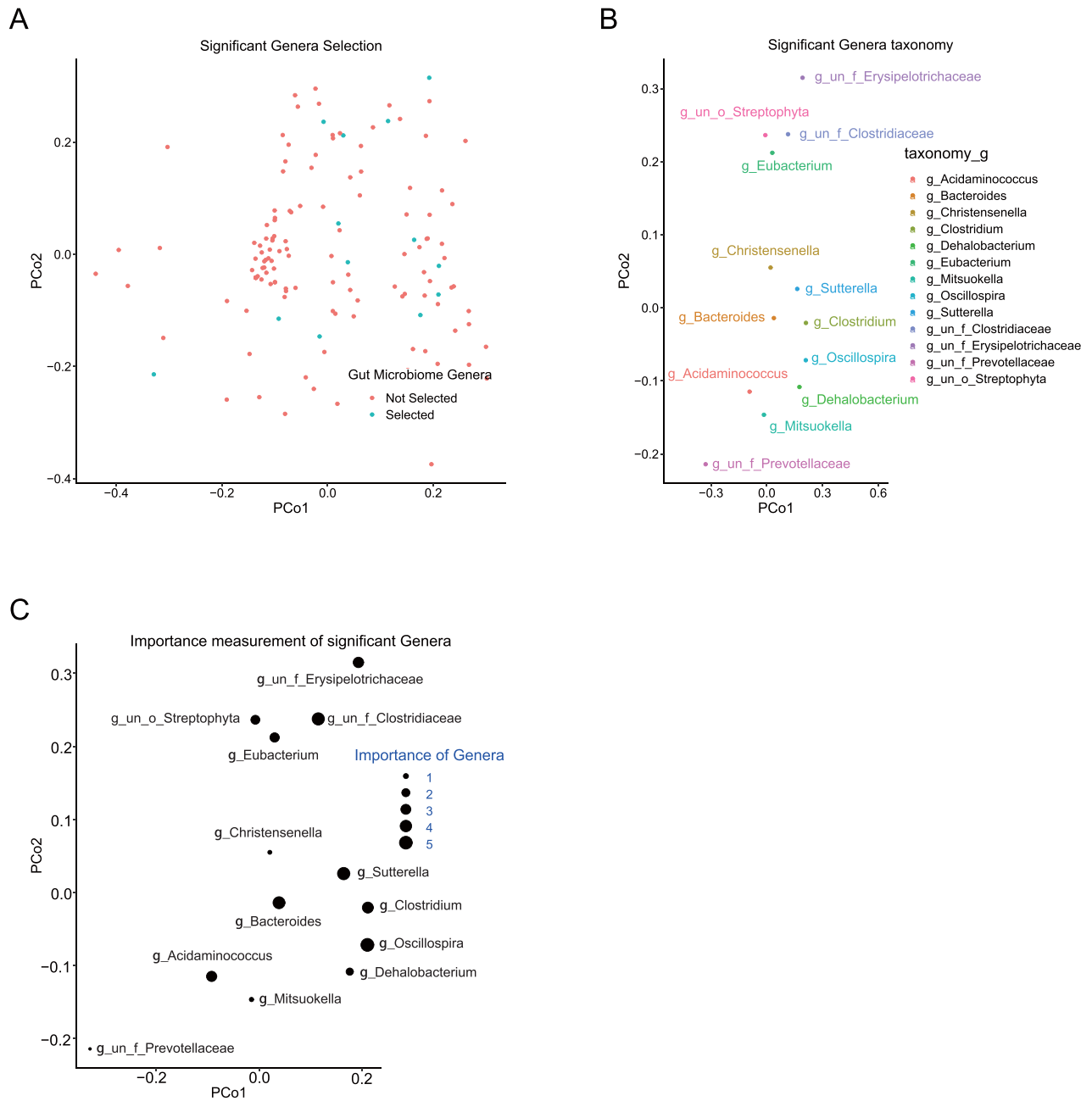


Figure 3. Taxonomic assessment of genus microbiota signature for HBeAg seroconversion. Axes represent the two first components from principal coordinate analysis (PCoA) based on the phylogenetic distance between genus representative sequences. **(A)** Thirteen genera (blue dots) are selected out from 169 genera using random forest while red dots represent unselected genera. **(B)** Taxonomic assignment of the significant genera at genus level. **(C)** The importance of significant genera is sized by the respective dot.

Bacteroides were shown in Supplementary Fig. 8C,D, respectively. Additionally, age is positively associated with time to HBeAg seroconversion ($\rho = 0.2664$, $P < 0.05$). Correlations between microbial signatures and clinical parameters were also shown in Supplementary Table 5.

Network of reciprocal interaction. We separately calculated the correlation network between 13 significant genera with large relative abundance in microbial signatures among Groups N and P. The measure of correlation used was the Spearman Correlation coefficient, and only significant correlations were included in the network (Supplementary Fig. 9). In Group N, significant positive association was observed between *Genus Dehalobacterium* and *Genus Christensenella*. Pronounced positive association was also observed between *Genus Clostridium* and *Genus un_f_Clostridiaceae*. However, negative association existed between *Genus Eubacterium* and *Genus Acidaminococcus*. In Group P, significant positive association was observed between *Genus Dehalobacterium* and *Genus Eubacterium*, *Genus Clostridium* and *Genus un_f_Prevotellaceae*, *Genus Bacteroides* and

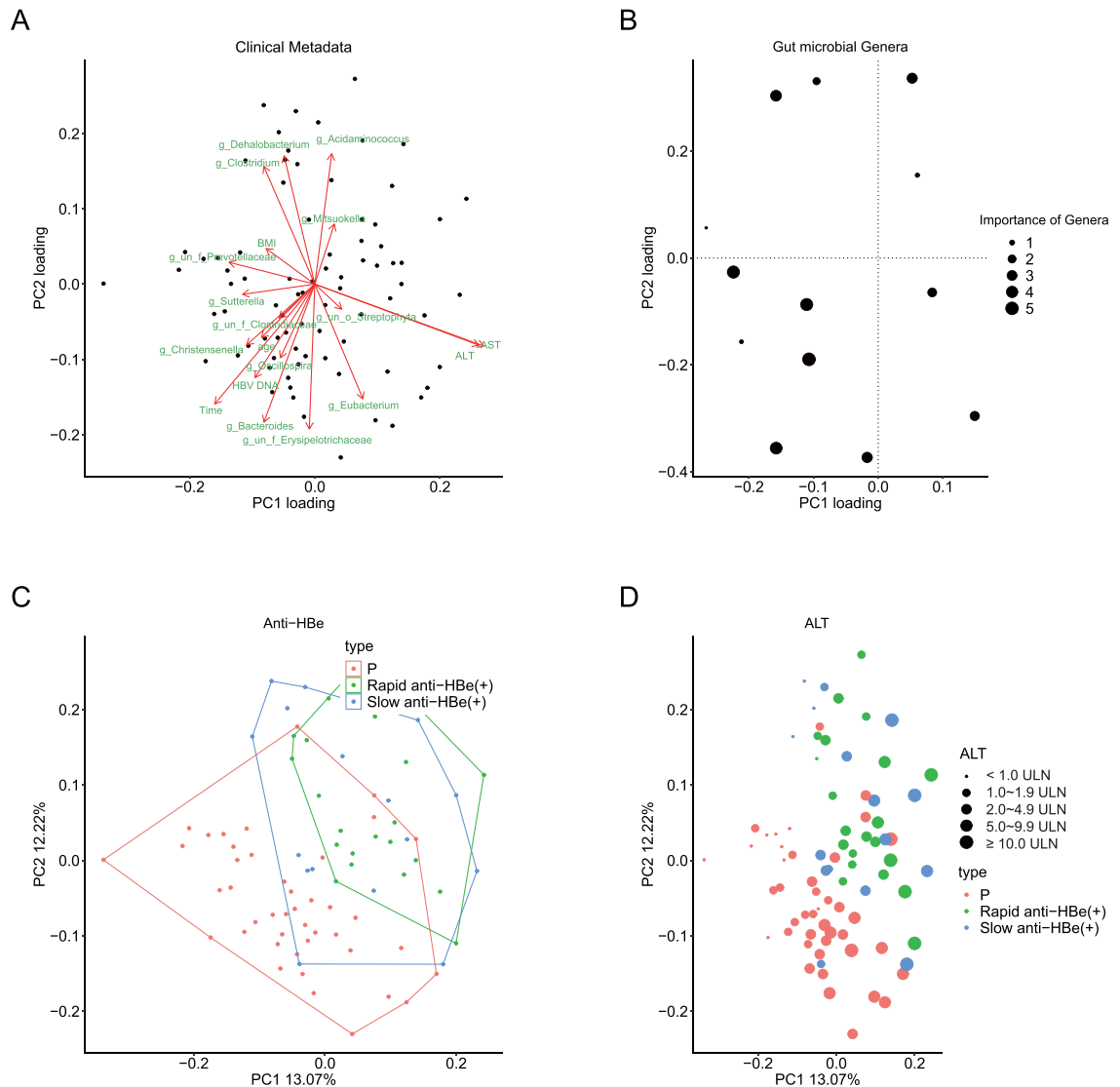


Figure 4. Interactions between clinical, microbial ecology parameter and microbial signatures for HBeAg seroconversion. A co-inertia analysis is performed between the relative abundance of microbial signature genera for HBeAg seroconversion and clinical parameters. (A) Scatter plot of the first two PC loadings of clinical data. Each clinical parameter and microbial signature are labelled in red. (B) Scatter plot of the first two PC loadings for the importance of microbial signatures. Size accounts for the importance. (C,D) Scatter plot of the first two components of co-inertia analysis. Each dot represents a fecal microbiota sample. Orange dots represent the subjects that remained HBeAg positive even after more than 3 years of oral antiviral therapy (Group P), while green dots represent the subjects that achieved HBeAg seroconversion in less than 12 months after the initiation of oral antiviral therapy [Rapid anti-HBe (+)], and blue dots represent the subjects that achieved HBeAg seroconversion later than 12 months after the initiation of oral antiviral therapy [Slow anti-HBe (+)]. (C) Using random forest based on the microbial signature and clinical data of Group P and Group N patients, rapid anti-HBe (+) and Group P are well separated. (D) Baseline ALT has a positive relation with HBeAg seroconversion. Size represents the value of ALT.

Genus Sutterella. On the contrary, negative association existed between *Genus Dehalobacterium* and *Genus Bacteroides*, *Genus Dehalobacterium* and *Genus Sutterella*.

Correlation between metabolic profiling and HBeAg seroconversion. A total of 1831 metabolites were detected under positive ion mode while 1636 metabolites were detected under negative ion mode. Significant difference regarding the intestinal metabolites was observed between and Group P and Group N. Group P tended to have higher abundance of some specific metabolites such as Betaine, essential amino acids, and several dipeptides in fecal samples under both positive ion mode and negative ion mode (Fig. 5A,B). Using KEGG database, differentiated metabolites could be mapped to the following pathways: Aminoacyl-tRNA biosynthesis, and Glycine, serine and threonine metabolism (Fig. 5C,D).

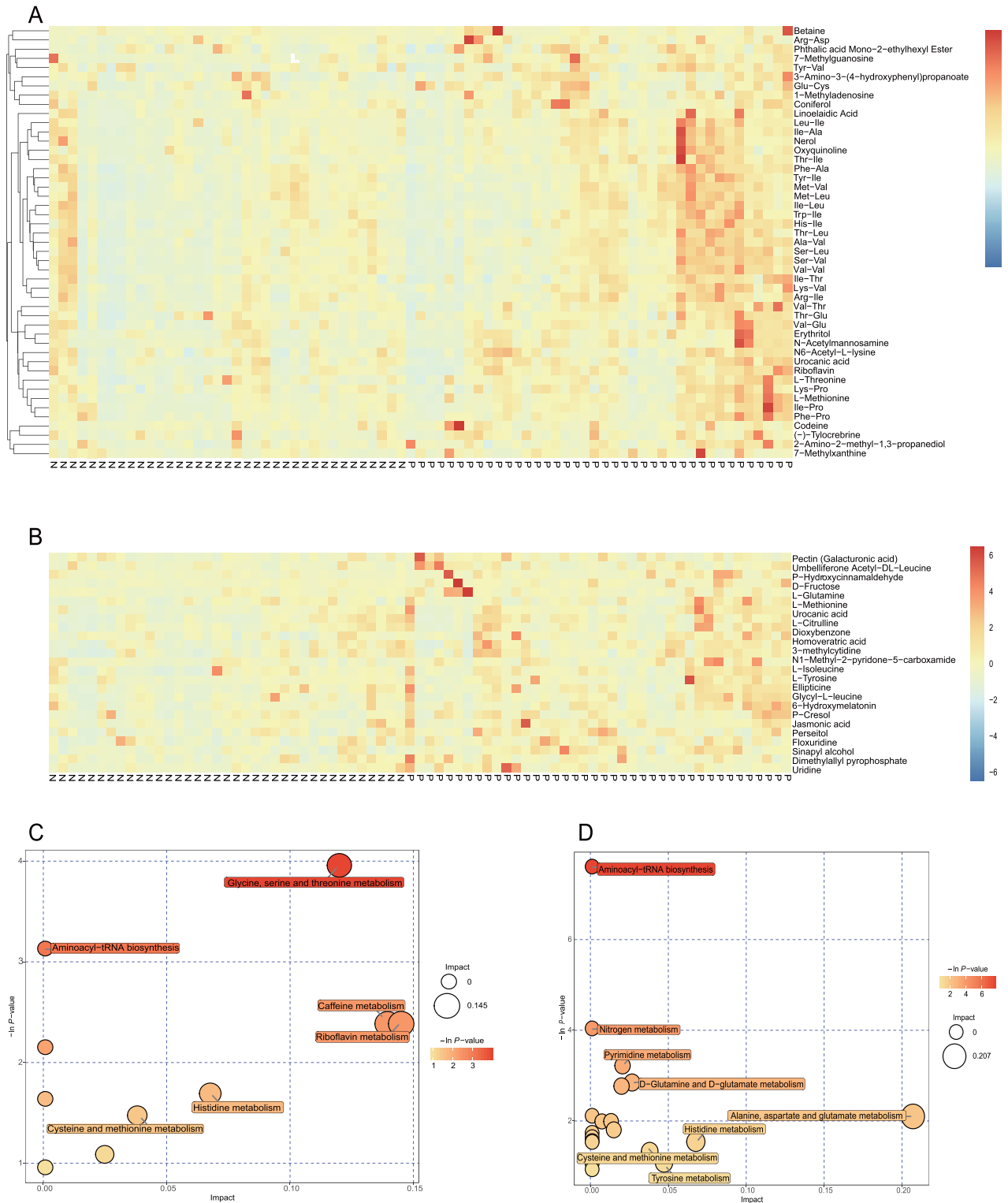


Figure 5. Correlation between differentiated metabolites and HBeAg seroconversion. The differentiated metabolites are defined as $P < 0.05$ (Student's t-test), combined with variable importance in the projection of the first principal component is greater than 1 in the model of orthogonal projections to latent structures-discriminant analysis. The differentiated metabolites are clustered by using complete linkage method and are displayed as a heatmap of hierarchical clustering analysis with the abscissa represents the different experimental cohorts while the ordinate represents the differentiated metabolites. The color blocks at different positions represent the relative expression of the specific metabolites at the corresponding positions. **(A)** Heatmap of hierarchical clustering analysis under positive ion (POS) mode. **(B)** Heatmap of hierarchical clustering analysis under negative ion (NEG) mode. Using KEGG database, differentiated metabolites are mapped to related metabolic pathways, which are shown in a bubble plot. Each bubble in the bubble diagram represents a metabolic pathway. Size of bubble indicates the influence factor of a specific metabolic pathway in topological analysis. The size is larger when the influence factor is higher. The ordinate and the color of bubble indicate the enrichment analysis. The bubble is darker when P value (expressed in minus natural logarithm, i.e., $-\ln P$ -value) is more significant in analysis. **(C)** Enrichment analysis of metabolic pathways under positive ion (POS) mode. **(D)** Enrichment analysis of metabolic pathways under negative ion (NEG) mode.

Correlation between microbiota and metabolic profiling. Different metabolic profiling between group N and group P were shown by OPLS-DA (Supplementary Fig. 10). Correlation analysis indicated that *Genus Christensenella* can explain most of the differentiated metabolites in fecal samples between Group P and Group N under both positive ion mode and negative ion mode (Supplementary Fig. 11A,B). Correlation network analysis also showed that *Genus Christensenella* was closely related to most of the differentiated metabolites (Supplementary Fig. 11C,D).

Discussion

In the present study we characterized the signature of fecal microbiota in a cohort of CHB patients with or without HBeAg seroconversion. Using a machine learning approach, we showed that HBeAg seroconversion is associated with a distinct fecal microbiota signature, which could also be associated with ALT and AST levels. We explored the data using a combination of classical approaches for microbiota analysis. However, these yielded no pronounced differences between the CHB patients with or without HBeAg seroconversion. Group N tended to have higher abundance of *Firmicutes* phylum and a lower abundance of *Bacteroidetes* phylum, although no significant difference was detected when compared with Group P.

In the present study, we further explored our data using Wilcoxon test and random forest that identified a robust microbial signature. At the family level, the microbial signature that was enriched included *Bacteroidaceae*, *Alcaligenaceae*, and *Ruminococcaceae*. *Genus Un_f_Prevotellaceae* and *Genus Sutterella* displayed negative association with PC1. This indicates that patients with higher abundance of *Un_f_Prevotellaceae* and *Genus Sutterella* have less opportunities to get HBeAg seroconversion. On the other hand, *Genus un_f_Erysipelotrichaceae* and *Genus Bacteroides* were the most important factors associated with second co-inertia component and were negatively associated with HBeAg seroconversion.

Several algorithms have been developed to predict the efficacy of interferon-based antiviral therapy for CHB^{17–19}. However, no such algorithms have been reported for testing the efficacy of nucleos(t)ide analogues in the treatment of CHB. Since intestinal microbiota has a critical or even decisive role on the outcome of chronic HBV infection, it seems very likely that Group N would have different microbial signatures from those of Group P, which can serve as the base of a novel prediction model. Using classical ecologic approaches, we found considerable overlap between Groups N and P, with no significant differences. However, when random forest method was applied, Group N could be well distinguished from Group P. Notably, the AUC area of the established prediction model for the test set was 0.749, indicating a robust model that can predict whether HBeAg seroconversion is likely to occur in a specific individual.

In recent years, several studies have reported gut dysbiosis in patients with CHB^{20,21}. Besides CHB, gut dysbiosis also plays a critical role in liver fibrosis and the genesis of hepatocellular carcinoma related to non-alcoholic fatty liver disease^{22,23}. Chen et al.²⁴ have characterized the alteration of *Bacteroidetes* phylum in case of HBV or alcohol-related cirrhosis. However, the relative abundance of *Bacteroidetes* in the patients with HBeAg seroconversion versus without seroconversion remains unconfirmed. Our study shows that there is differential abundance of *Firmicutes* and *Bacteroidetes* phylum in the patients with or without HBeAg seroconversion. *Bacteroides* is the most prevalent Gram-negative gut bacterial taxon, and is the main source of lipopolysaccharide (LPS)²⁵. Pathogen-derived LPS is a potent ligand for host receptor toll-like receptor 4 (TLR4), which plays an important role in sensing bacteria as part of the innate immune response. Research by Chou et al.⁹ suggested that LPS-TLR4 signaling plays a critical role in the immune tolerance to HBV infection in young mice. After binding to its receptor TLR4, LPS triggers the M1 or M2 polarization of Kupffer cells. M1 polarization leads to the clearance of HBV, while M2 polarization of Kupffer cells support HBV persistence by suppressing HBV-specific CD8⁺ T cells²⁶. Elderly individuals usually have higher abundance of *Bacteroidetes*²⁷, which is linked to elevated levels of serum LPS. Thus, it seems reasonable that the elderly tends to have poorer efficacy of antiviral treatment, which is consistent with clinical observations. For both HBeAg-positive and HBeAg-negative patients, younger age (< 30 years) was a significant predictor of ALT normalization after telbivudine treatment²⁸. More than that, age was one of the significant predictive factors for sustained virologic response²⁹. After HBeAg seroconversion, patients aged > 40 relapsed more frequently than those aged ≤ 40³⁰. Therefore, higher abundance of *Bacteroidetes* may, at least partly, explain the lower likelihood of HBeAg seroconversion in older patients.

As is shown in Fig. 5 and Supplementary Fig. 11, patients who achieved HBeAg seroconversion tended to have lower abundance of certain fecal metabolites such as essential amino acids, and several dipeptides, which had positive correlation with the relative abundance of *Genus Christensenella*. *Genus Christensenella* is a kind of newly isolated gram-negative, strictly anaerobic bacterium that inhabits the human intestine. Recent finding indicates that *Genus Christensenella* can only trigger a weak inflammatory response through the NF- κ B signalling pathway, which may partly explain why the patients in Group P with higher abundance of *Genus Christensenella* is unlikely to obtain HBeAg seroconversion³¹. Study has indicated there are certain relationship between plasma amino acid profiles and various stages of HBV infection³². However, no confirmed correlation between fecal amino acid profiles and HBV infection has been reported, which remains to be further elucidated. Amino acids may as metabolic mediators of the cross talk between host and pathogen³³. According to KEGG database, fecal metabolites with different abundance such as essential amino acids and dipeptides could be mapped to several pathways: aminoacyl-tRNA biosynthesis, and glycine, serine and threonine metabolism. Recent discovery has revealed that cytoplasmic aminoacyl-tRNA synthetases can sense and respond to danger signals and induce antiviral immunity³⁴. It seems reasonable that different intestinal flora affects the metabolism of amino acids which affects the activity of aminoacyl-tRNA synthetases, and further regulates the antiviral immunity of host against HBV.

A limitation of the present study is that no mucosal sample was obtained for microbiota analysis. Since there were no significant bowel changes in the CHB patients, the participants were usually not ready to accept

colonoscopy examination. Secondly, for Group P and Group N, only about 40 cases were enrolled in each group. Nonetheless, according to statistical inference, this sample size does not significantly impair the accuracy of the prediction model. As can be seen from Supplementary Fig. 5, if the total number of Group N and Group P is less than 45, increasing the sample size can sharply improve the mean accuracy. However, if the sample size is more than 45, the mean accuracy rate increases slowly. It can be inferred that increasing sample size would have had a further but less significant improving effect on the mean accuracy. Thus, the sample size in our study is adequate.

To conclude, machine learning based random forest allowed us to identify gut microbial signature for HBeAg seroconversion, which also could be reproduced by the test set. Importantly, because of its relatively high accuracy, the prediction model based on the gut microbial signature can be used not only as a clinical predictor of HBeAg seroconversion before the initiation of oral antiviral therapy, but also as a way to explore the relevant genera for HBeAg seroconversion in the future. Our study highlights the difficulty in stratifying patients based on a microbiota profile when only classical ecologic approaches are employed. The use of machine learning has allowed us to better explore the microbial data, allowing us to identify certain links between gut microbiota and the clinical outcome.

Received: 21 July 2020; Accepted: 25 January 2021

Published online: 05 February 2021

References

- Sarin, S. K. *et al.* Asian-Pacific clinical practice guidelines on the management of hepatitis B: A 2015 update. *Hepatol. Int.* **10**, 1–98. <https://doi.org/10.1007/s12072-015-9675-4> (2016).
- European Association for the Study of the Liver. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J. Hepatol.* **67**, 370–398. <https://doi.org/10.1016/j.jhep.2017.03.021> (2017).
- Brunetto, M. R. *et al.* Hepatitis B virus surface antigen levels: A guide to sustained response to peginterferon alfa-2a in HBeAg-negative chronic hepatitis B. *Hepatology* **49**, 1141–1150. <https://doi.org/10.1002/hep.22760> (2009).
- Marcellin, P. *et al.* Peginterferon alfa-2a alone, lamivudine alone, and the two in combination in patients with HBeAg-negative chronic hepatitis B. *N. Engl. J. Med.* **351**, 1206–1217. <https://doi.org/10.1056/NEJMoa040431> (2004).
- Chang, T. T. *et al.* A comparison of entecavir and lamivudine for HBeAg-positive chronic hepatitis B. *N. Engl. J. Med.* **354**, 1001–1010. <https://doi.org/10.1056/NEJMoa051285> (2006).
- Chang, T. T. *et al.* Entecavir treatment for up to 5 years in patients with hepatitis B e antigen-positive chronic hepatitis B. *Hepatology* **51**, 422–430. <https://doi.org/10.1002/hep.23327> (2010).
- Marcellin, P. *et al.* Tenofovir disoproxil fumarate versus adefovir dipivoxil for chronic hepatitis B. *N. Engl. J. Med.* **359**, 2442–2455. <https://doi.org/10.1056/NEJMoa0802878> (2008).
- Singh, A. K., Sharma, M. K., Hissar, S. S., Gupta, E. & Sarin, S. K. Relevance of hepatitis B surface antigen levels in patients with chronic hepatitis B during 5 year of tenofovir treatment. *J. Viral. Hepat.* **21**, 439–446. <https://doi.org/10.1111/jvh.12149> (2014).
- Chou, H. H. *et al.* Age-related immune clearance of hepatitis B virus infection requires the establishment of gut microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2175–2180. <https://doi.org/10.1073/pnas.1424775112> (2015).
- Ren, Y. D. *et al.* Fecal microbiota transplantation induces hepatitis B virus e-antigen (HBeAg) clearance in patients with positive HBeAg after long-term antiviral therapy. *Hepatology* **65**, 1765–1768. <https://doi.org/10.1002/hep.29008> (2017).
- Tremaroli, V. & Backhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **489**, 242–249. <https://doi.org/10.1038/nature11552> (2012).
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–462. <https://doi.org/10.1093/nar/gkv1070> (2016).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. <https://doi.org/10.1093/nar/28.1.27> (2000).
- Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504. <https://doi.org/10.1101/gr.1239303> (2003).
- McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217. <https://doi.org/10.1371/journal.pone.0061217> (2013).
- Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: Generative models for microbial metagenomics. *PLoS ONE* **7**, e30126. <https://doi.org/10.1371/journal.pone.0030126> (2012).
- Buster, E. H. *et al.* Factors that predict response of patients with hepatitis B e antigen-positive chronic hepatitis B to peginterferon-alfa. *Gastroenterology* **137**, 2002–2009. <https://doi.org/10.1053/j.gastro.2009.08.061> (2009).
- Mao, Q. G. *et al.* Precise prediction model and simplified scoring system for sustained combined response to interferon-alpha. *World J. Gastroenterol.* **16**, 3465–3471. <https://doi.org/10.3748/wjg.v16.i27.3465> (2010).
- Sonneveld, M. J. *et al.* Response-guided peginterferon therapy in hepatitis B e antigen-positive chronic hepatitis B using serum hepatitis B surface antigen levels. *Hepatology* **58**, 872–880. <https://doi.org/10.1002/hep.26436> (2013).
- Chen, Y. *et al.* Gut dysbiosis in acute-on-chronic liver failure and its predictive value for mortality. *J. Gastroenterol. Hepatol.* **30**, 1429–1437. <https://doi.org/10.1111/jgh.12932> (2015).
- Wang, J. *et al.* Gut microbial dysbiosis is associated with altered hepatic functions and serum metabolites in chronic hepatitis B patients. *Front. Microbiol.* **8**, 2222. <https://doi.org/10.3389/fmicb.2017.02222> (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. <https://doi.org/10.1002/hep.28356> (2016).
- Ponziani, F. R. *et al.* Hepatocellular carcinoma is associated with gut microbiota profile and inflammation in nonalcoholic fatty liver disease. *Hepatology* <https://doi.org/10.1002/hep.30036> (2018).
- Chen, Y. *et al.* Characterization of fecal microbial communities in patients with liver cirrhosis. *Hepatology* **54**, 562–572. <https://doi.org/10.1002/hep.24423> (2011).
- Jacobson, A. N., Choudhury, B. P. & Fischbach, M. A. The biosynthesis of lipooligosaccharide from *Bacteroides thetaiotaomicron*. *MBio* <https://doi.org/10.1128/mBio.02289-17> (2018).
- Tsai, K. N., Kuo, C. F. & Ou, J. J. Mechanisms of hepatitis B virus persistence. *Trends Microbiol.* **26**, 33–42. <https://doi.org/10.1016/j.tim.2017.07.006> (2018).
- Mariat, D. *et al.* The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age. *BMC Microbiol.* **9**, 123. <https://doi.org/10.1186/1471-2180-9-123> (2009).
- Zeuzem, S. *et al.* Baseline characteristics and early on-treatment response predict the outcomes of 2 years of telbivudine treatment of chronic hepatitis B. *J. Hepatol.* **51**, 11–20. <https://doi.org/10.1016/j.jhep.2008.12.019> (2009).

29. Lee, H. W. *et al.* Lamivudine maintenance beyond one year after HBeAg seroconversion is a major factor for sustained virologic response in HBeAg-positive chronic hepatitis B. *Hepatology* **51**, 415–421. <https://doi.org/10.1002/hep.23323> (2010).
30. Yoon, S. K. *et al.* Long-term results of lamivudine monotherapy in Korean patients with HBeAg-positive chronic hepatitis B: Response and relapse rates, and factors related to durability of HBeAg seroconversion. *Intervirology* **48**, 341–349. <https://doi.org/10.1159/000086061> (2005).
31. Yang, Y., Gu, H., Sun, Q. & Wang, J. Effects of *Christensenella minuta* lipopolysaccharide on RAW 2647 macrophages activation. *Microb. Pathog.* **125**, 411–417. <https://doi.org/10.1016/j.micpath.2018.10.005> (2018).
32. Yang, J. *et al.* Correlation between plasma amino acid profiles and the various stages of hepatitis B infection. *Eur. J. Clin. Microbiol. Infect. Dis.* **31**, 2045–2052. <https://doi.org/10.1007/s10096-011-1538-x> (2012).
33. Ren, W. *et al.* Amino acids as mediators of metabolic cross talk between host and pathogen. *Front. Immunol.* **9**, 319. <https://doi.org/10.3389/fimmu.2018.00319> (2018).
34. Lee, E. Y. *et al.* Infection-specific phosphorylation of glutamyl-prolyl tRNA synthetase induces antiviral immunity. *Nat. Immunol.* **17**, 1252–1262. <https://doi.org/10.1038/ni.3542> (2016).

Acknowledgements

Writing support was provided by Editage English Editing Services and funded by the National Natural Science Foundation of China (no. 81871645).

Author contributions

J.S.P., and J.J.N.: study concept and design. Y.L.Z., W.J.W., H.C., X.F.Y., W.Z., X.L.W., X.B.L., P.Y., Y.L., W.M.C., and M.Z.H.: acquisition of data. J.S.P., and J.J.N.: supply of resources and administrative support. L.Q., and B.C.S.: analysis and interpretation of data. Y.L.Z., and J.S.P.: drafting of the manuscript. X.D.: assistance in preparation manuscript and submission. J.S.P., J.J.N., and B.C.S.: critical revision of the manuscript for important intellectual content. L.Q.: statistical analysis. J.S.P., L.Q., P.Y., and Y.L.: obtained funding. J.S.P., and J.J.N.: study supervision. All authors approved the final draft submitted.

Funding

This work was supported by the National Basic Research Program of China (973 Program), Grant number 2015CB553800]; the National Natural Science Foundation of China, Grant numbers 81871645, 61603092, 81500451, 81100285; and the Foundation of Department of Science and Technology of Fujian Province, China, Grant number 2015J01540. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-82939-1>.

Correspondence and requests for materials should be addressed to B.-C.S., J.-J.N. or J.-S.P.

Reprints and permissions information is available at 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021