microbial biotechnology

Open Access

# Characteristic gut microbiota and metabolic changes in patients with pulmonary tuberculosis

**Shuting Wang† Liya Yang† Haiyang Hu† Longxian Lv, Zhongkang Ji, Yanming Zhao, Hua Zhang, Min Xu, Rongfeng Fang, Lin Zheng, Cheng Ding, Meifan Yang, Kaijin Xu*,† and Lanjuan Li*,†**

*State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Centre for Infectious Diseases, Collaborative Innovation Centre for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310003, China.*

## Summary

**Intestinal flora provides an important contribution to the development of pulmonary tuberculosis (PTB). We performed a cross-sectional study in 52 healthy controls (HCs) and 83 patients with untreated active PTB to assess the differences in their microbiomic and metabolic profiles in faeces via V3-V4 16S rRNA gene sequencing and gas chromatography–mass spectrometry. Patients with PTB had considerable reductions in phylogenetic alpha diversity and the production of short-chain fatty acids, dysbiosis of the intestinal flora and alterations in the faecal metabolomics composition compared with HCs. Significant alterations in faecal metabolites were associated with changes in the relative abundance of specific genera. Our study describes the imbalance of the gut microbiota and altered faecal metabolomics profiles in patients with PTB; the results indicate that the gut microbiota and faecal metabolomic profiles can be used as potential preventive and therapeutic targets for PTB.**

## Introduction

Human tuberculosis (TB) caused by *Mycobacterium tuberculosis* (Mtb) is the leading global cause of death among those attributable to a single pathogen. TB resulted in the deaths of approximately 1.5 million people in 2018. A quarter of the global population is infected with TB, and approximately 3–5% of these individuals will develop active TB during their lifetime. Additionally, China has the second highest burden of multidrug-resistant tuberculosis (MDR-TB), accounting for 14% of the global cases of TB (WHO, 2019). Moreover, the emergence of drug-resistant Mtb and a rise in TB–human immunodeficiency virus (HIV) coinfection poses a substantial threat to public health.

The human gut microbiome (GM) is characterized by diverse microbial communities of multiple phyla of bacteria, archaea, viruses and microbial eukaryotes (Reyes *et al*., 2012; Thomas *et al*., 2017), which play a significant role in human health (Qin *et al*., 2010; Jiang *et al*., 2019). GM is known to influence the energy metabolism of the host (Turnbaugh *et al*., 2009; Robertson *et al*., 2018) and indirectly or directly promote the maturation of immune cells and normal development of immune function (Spiljar *et al*., 2017). For example, modulation of macrophage-inducible C-type lectin (Mincle) by gut bacteria plays important roles in the regulation of the immune response to Mtb that is dependent on dendritic cells (DC) in the lung (Negi *et al*., 2019). Existing evidence led researchers to investigate cross-talk between the intestinal bacteria and the lung, which is defined as the gut–lung axis (Budden *et al*., 2017). Dysbiosis of the intestinal microbiota is relevant to numerous lung diseases, such as chronic obstructive pulmonary disease (COPD) (Rutten *et al*., 2014), asthma (Lee-Sarwar *et al*., 2019), H7N9 virus infection (Qin *et al*., 2015) and *Staphylococcus aureus* pneumonia (Gauguet *et al*., 2015).

Substantial evidence obtained in animals and humans indicates that Mtb infection drives intestinal microbial dysbiosis characterized by changes in the abundance of specific taxa, particularly bacteria that produce short-chain fatty acids (SCFAs), such as *Ruminococcus* and *Bifidobacterium* (Khan et al., 2016; Hu *et al*., 2019, Namasivayam et al., 2019). Potential limitation of Mtb by indole propionic acid produced by the intestinal bacteria has been demonstrated in vitro and in vivo (Negatu et al., 2019). The present study investigated the potential correlations between the host, diet and intestinal bacteria by analysing the faecal metabolic profiles using untargeted gas chromatography–mass spectrometry (GC-MS) combined with high-throughput sequencing of the 16S

Received 31 August, 2020; revised 12 January, 2021; accepted 18 January, 2021.
†These authors contributed equally to this study.
Corresponding authors: Kaijin Xu and Lanjuan Li contributed equally.
*For correspondence. E-mail: ljli@zju.edu.cn, zdyxyxkj@zju.edu.cn; Tel. +86-571-8723-6458; Fax +86-571-8723-6459
*Microbial Biotechnology* (2022) **15**(1), 262–275
doi:10.1111/1751-7915.13761

rRNA gene. This integrated omics analysis was used to assess whether pulmonary tuberculosis (PTB) is related to the changes in the composition and function of microbiota and to determine the mechanism of these associations, aiming to identify the candidate biomarkers for PTB.

## Results

### *Cholesterol levels were decreased in PTB patients*

Eighty-three patients diagnosed with untreated active PTB and 52 age- and sex-matched healthy controls (HCs) were enrolled in this study. Table 1 presents the detailed clinical characteristics of all subjects. With the exception of blood glucose (BG) and albumin (Alb) levels, other characteristics, including the serum levels of total cholesterol (Tch) ($P < 0.001$), high-density lipoprotein cholesterol (HDL) ($P = 0.04997$), low-density lipoprotein cholesterol (LDL) ($P = 0.002$), and very-low-density lipoprotein cholesterol (VLDL) ($P < 0.001$), were significantly lower in patients than in HCs.

### *Mtb infection decreased the alpha diversity and affected the composition of the gut microbiome*

A sequencing run generated 11 448 333 reads, and 94.56% of these reads (10 822 567 reads) passed the quality screening. The reads were clustered into 875 operational taxonomic units (OTUs) based on a similarity of 97%. A Venn diagram showed that 99 OTUs and 74 OTUs were unique in the HC and PTB groups respectively (Fig. 1B).

Microbiota diversity was then evaluated using the following indexes: Chao1, ACE, Shannon, Simpson and observed species. The rarefaction cures for the two groups showed the number of identified OTUs approached a plateau, indicating that the bacterial diversity of the intestinal flora in the HC group was higher than that in the PTB group (Fig. S1B). In addition, the number of observed species in PTB patients was markedly lower than that in HCs ($254.02 \pm 44.65$ vs. $206.17 \pm 52.17$, respectively, $P = 9.78e\text{-}8$) (Fig. S1A), and the alpha diversity, which was assessed by the Chao1, ACE, Shannon and Simpson indexes, was significantly lower in the PTB group than in the HC group ($P = 1.70e\text{-}7$, $1.06e\text{-}7$, $7.90e\text{-}5$, and $0.0021$ respectively) (Fig. 1A). The results of the principal coordinate analysis (PCoA) based on unweighted UniFrac (Adonis, $P < 0.01$) (Fig. 1C) and non-metric multidimensional scaling (NMDS) based on unweighted UniFrac (stress value = 0.191, MRPP, $P < 0.01$) (Fig. S1C) indicated that the dissimilarity maintained a tendency of separation and showed partial overlap between PTB patients and HCs.

### *The microbial abundance was shifted in the PTB group*

Metastat-based and linear discriminant analysis (LDA) effect size (LEfSe) analyses were performed to identify different distributions of the faecal microbiota. The PTB and HC groups had statistically significant differences in five predominant phyla: *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Fusobacteria* and *Verrucomicrobia*. The relative abundances (proportions) of *Firmicutes* and *Bacteroidetes* were 63.81% and 27.82% in the HC group and 48.84% and 44.92% in the PTB group respectively (Fig. 2A).

At the family level, the PTB group had significantly greater abundances of *Bacteroidaceae* (30.36% vs. 17.19%, $q = 0.004$), *Tannerellaceae* (2.38% vs. 0.94%, $q = 0.019$), *Fusobacteriaceae* (2.15% vs. 0.17%, $q = 0.014$) and *Erysipelotrichaceae* (1.44% vs. 0.89%, $q = 0.034$) than the HCs group. The PTB group also had markedly lower abundances of *Bifidobacteriaceae* (0.9%

**Table 1.** Clinical characteristics of the cohort

| Variable | HC ($n = 52$) | PTB ($n = 83$) | *P* value |
|---|---|---|---|
| Gender, Man, *n* (%) | 21 (40.38%) | 47 (56.63%) | 0.078 |
| Age (year) | 32(26-37) | 30(25-48) | 0.486 |
| Laboratory analyses | | | |
| Albumin (g $l^{-1}$) | 48.3(46.75-49.7) | 45.95 (43.83-48) | <0.001 |
| Serum creatinine (mmol $l^{-1}$) | 64 (56-78) | 68 (58-81) | 0.5 |
| Haemoglobin (g $l^{-1}$) | 144 (135-154) | 138 (129.25-150.75) | 0.075 |
| Triglycerides (mmol $l^{-1}$) | 1.13 (0.75-1.39) | 0.98 (0.73-1.43) | 0.785 |
| Total cholesterol (mmol $l^{-1}$) | 4.59 (0.72) | 3.88 (0.75) | <0.001 |
| High-density lipoprotein (mmol $l^{-1}$) | 1.31 (1.11-1.58) | 1.19 (0.98-1.4) | <0.05 |
| Low-density lipoprotein (mmol $l^{-1}$) | 2.60 (0.6) | 2.21 (0.65) | 0.002 |
| Very-low-density lipoprotein (mmol $l^{-1}$) | 0.59 (0.45-0.73) | 0.41 (0.32-0.57) | <0.001 |
| Serum fasting blood glucose (mmol $l^{-1}$) | 4.77 (0.39) | 4.97 (0.59) | 0.042 |
| IFN-$\gamma$ release assay | 52 (−) | 83 (+) | |

Data are represented as mean (SD) or median (interquartile range).

 

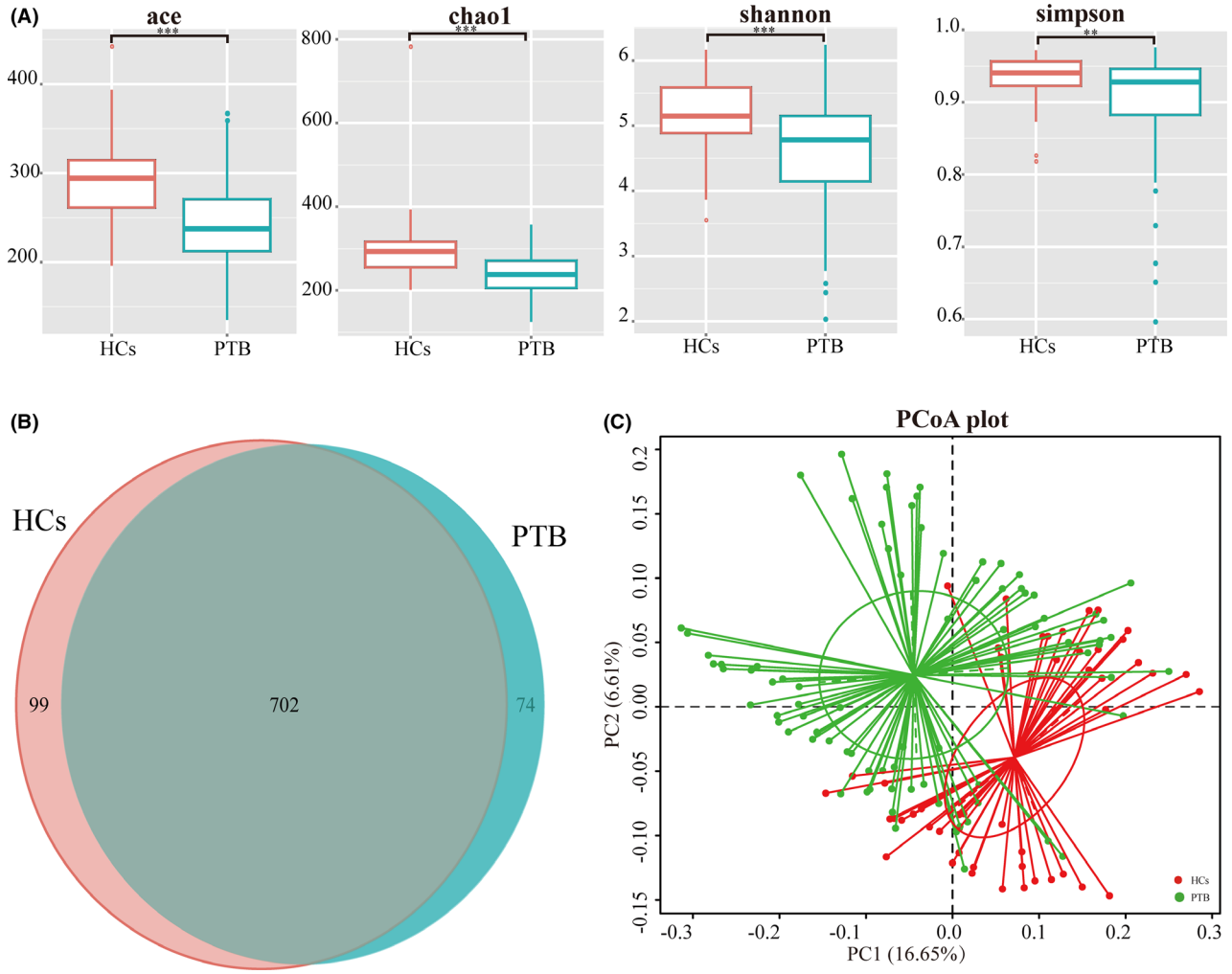


**Fig. 1.** Comparison of the intestinal microbiota richness and diversity in patients with PTB versus healthy controls. A. The alpha diversity was assessed using the ACE, Chao1, Shannon and Simpson indexes, which showed significant differences between the PTB and HCs groups. ***$P < 0.001$. B. Venn diagram showing the shared and unique operational taxonomic units (OTUs) in the flora of the two groups. C. A PCoA based on Unweighted Unifrac distance showed that the distribution of the microbial community in the PTB group was strikingly different from that of the HCs group ($P < 0.01$). PCoA, principal coordinate analysis; PTB: pulmonary tuberculosis; HCs: healthy controls.

vs. 4.99%, $q = 0.012$), *Lachnospiraceae* (23.89% vs. 33.95%, $q = 0.004$), *Ruminococcaceae* (15.51% vs. 21.59%, $q = 0.01$), *Marinifilaceae* (0.27% vs. 0.55%, $q = 0.022$), *Eggerthellaceae* (0.027% vs. 0.19%, $q = 0.004$) and *Barnesiellaceae* (0.091% vs. 0.35%, $q = 0.014$) than the HCs (Fig. 2B).

At the genus level, the abundances of 102 genera, including 10 dominant (≥ mean 1% abundance in either group) and 92 less dominant genera, differed between the two groups. The dominant genera, including *Bacteroides*, *Parabacteroides*, *Fusobacterium* and *Lachnoclostridium*, were notably enriched in the PTB group compared with the HC group, whereas *Blautia*, *Roseburia*, *Bifidobacterium*, unidentified *Ruminococcaceae*, *Fusicatenibacter* and *Romboutsia* were enriched in the HCs group compared with the PTB group (Fig. 2C).

LEfSe analysis was used to identify the key phenotypes contributing to the differences between the groups. According to the LDA score, the optimal-enriched taxa in the stool microbiome of the PTB group were *Bacteroidales, Prevotellaceae* and *Bacteroides vulgatus*, whereas *Firmicutes*, *Clostridiales*, *Lachnospiraceae*, *Ruminococcaceae*, *Actinobacteria*, *Bifidobacteriales* and *Blautia* were more abundant in the HCs (Fig. 2D).

### *Untargeted metabolomics profiles*

Faecal metabolite assessment was performed in all participants, and a total of 744 different features were examined. To further characterize the overall metabolomic changes, we performed PCA, and an OPLS-DA of the faecal profiles showed distinct clusters in the PTB

 

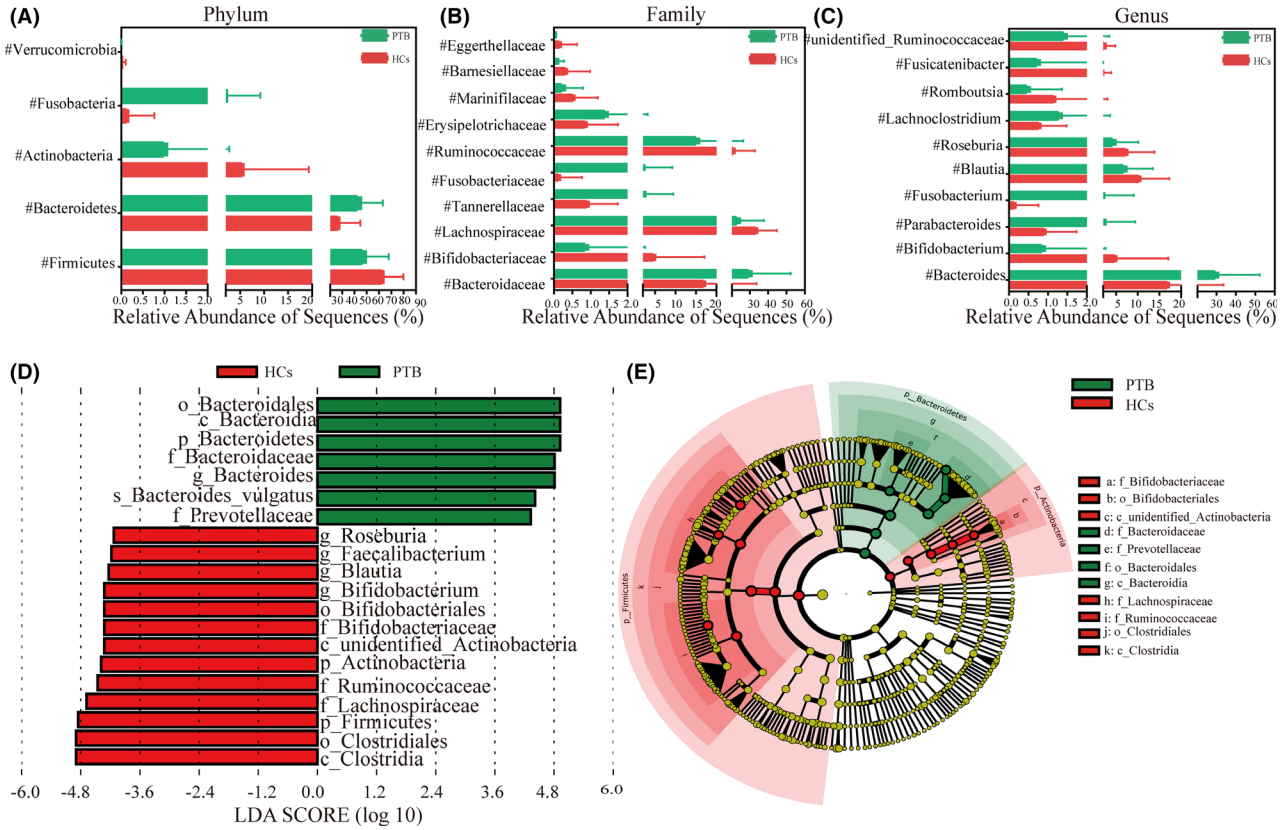


**Fig. 2.** Taxonomic features of the faecal microbiota of patients with PTB and controls. Comparisons of the relative abundances of intestinal microbiota between the PTB and HCs groups were conducted at the phylum (A), family (B) and genus (C) level; #$P < 0.05$. A LEfSe analysis identified the differentially abundant taxa between the PTB and HCs groups. The HC-enriched taxa are indicated with a negative LDA score (red), and the taxa enriched in the patients with PTB are characterized by a positive score (green). Only taxa with an LDA threshold greater than 4.0 are displayed (D). LDA, linear discriminant analysis. The circular cladogram was derived from the LEfSe analysis and showed the relationship between the most differentially abundant taxa in the patients with PTB (green) and the HCs (red). HCs, healthy controls; LEfSe, linear discriminant analysis effect size; PTB, pulmonary tuberculosis.

and HC groups, indicating that Mtb infection alters the metabolomic profile. Moreover, the two groups clearly segregated along PC1, accounting for 35.7% of the variance (Fig. 3A). The OPLS-DA of the faecal metabolomic profile (R2X = 0.456, Q2 = 0.752) demonstrated more pronounced differences between the PTB and HC groups than with those obtained with the PCA models (Fig. 3B). To prevent overfitting, sevenfold cross-validation and 200 response permutation tests were employed to validate the model, indicating that the original models had better predictive power than the newly calculated OPLS-DA models based on randomly assigning class labels (Q2inter = −0.412).

Significant metabolites were identified based on the OPLS-DA model with VIP values > 1 and $P$ values < 0.05 (two-tailed Student's $t$-test). As shown in Fig. 3C, the metabolic phenotype of patients with PTB was characterized by alterations in 26 faecal metabolites (Table S1), including 17 elevated and nine downregulated metabolites, and these metabolites are involved in the following pathways: ABC transporters, phenylalanine metabolism, taurine and hypotaurine metabolism, phosphatidylinositol signalling system, sulfur metabolism, galactose metabolism, inositol phosphate metabolism, primary bile acid biosynthesis, ascorbate and aldarate metabolism, steroid biosynthesis, glycine, serine and threonine metabolism, neuroactive ligand–receptor interaction, aminoacyl-tRNA biosynthesis, cysteine and methionine metabolism, tyrosine metabolism, carbon metabolism, biosynthesis of amino acids and neuroactive ligand–receptor interaction.

### *Mtb infection decreased the production of SCFAs in faeces*

The levels of isobutyric acid, butyric acid, 2-metylbutic acid and valeric acid were markedly lower in the PTB group than in the HCs ($P < 0.0001$, $P < 0.05$, $P < 0.0001$, $P < 0.0001$, respectively, Fig. 4), whereas the level of propionic acid was non-significantly lower in

 

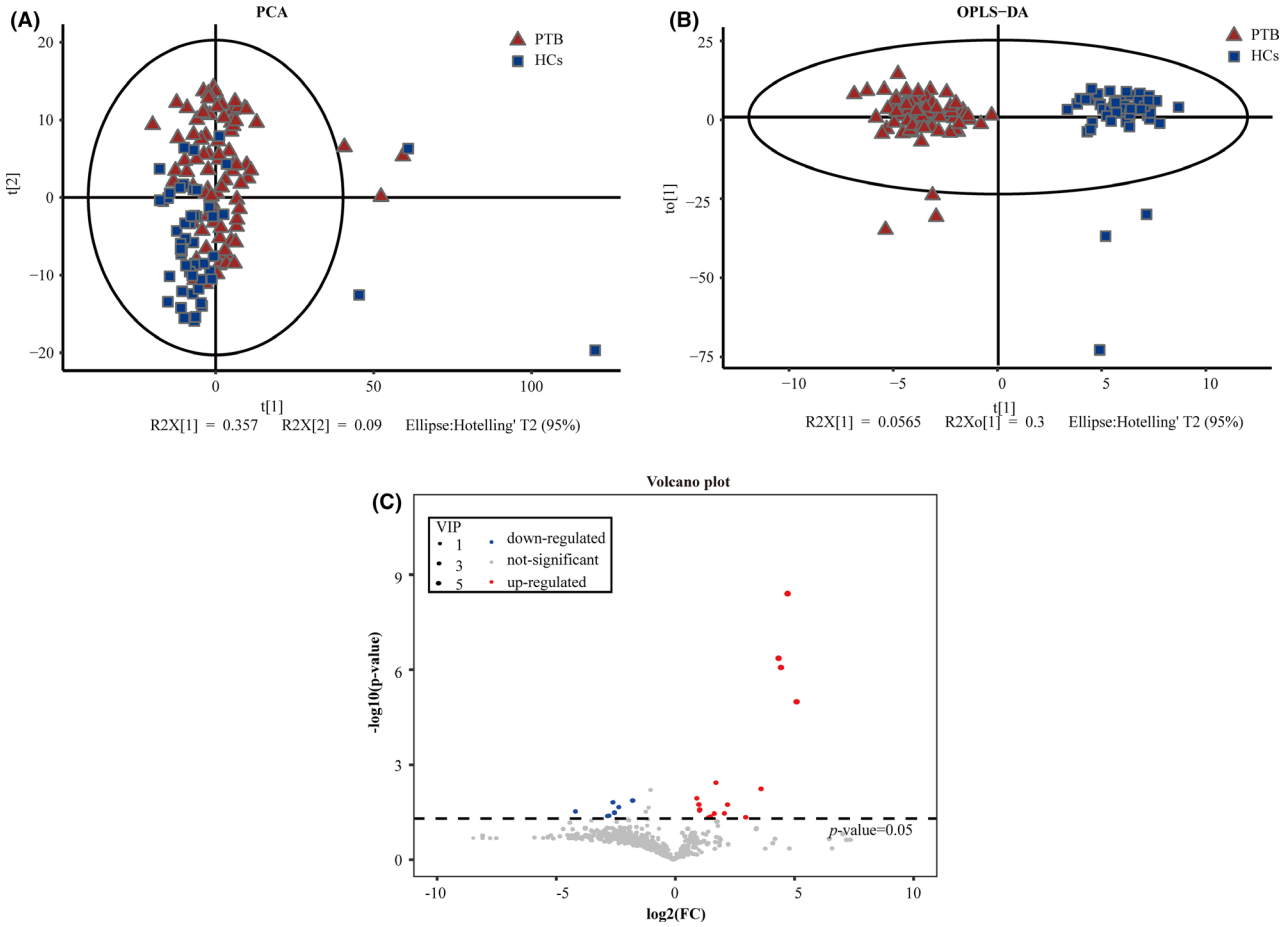


**Fig. 3.** Faecal metabolic changes associated with Mtb infection. A. PCA score plot comparing the patients with PTB (red) and the HCs (blue). B. Score plot of the OPLS-DA showing the patients with PTB (red) and the HCs (blue). C. Volcano plot showing the important metabolites included in the OPLS-DA model. HCs, healthy controls; PTB, pulmonary tuberculosis.

the PTB group than in the HCs ($P > 0.05$). However, no significant elevation in acetic acid concentration was noticed in the PTB group.

### *Discriminant models of PTB based on tree-based features*

To further explore the diagnostic value of the faecal microbiome for PTB, we constructed a random forest-based classification model to accurately distinguish PTB patients from HCs. A 10-fold cross-validation used a training dataset ($n = 94$), which included 58 PTB patients and 36 healthy controls, to select specific genus markers, and 30 genus markers were selected from each sample based on the results. The training of the random forest classifier involved the calculation of the feature importance of these 30 genera and their relative abundances (Table S2), and the features were ranked according to their contribution to the model. The vital genera identified based on the model were described based on a mean decrease in the Gini index (Fig. 5A). The area under the curve (AUC) obtained using the training datasets was 0.872 (95% confidence interval (CI): 0.804–0.94), showing that patients with PTB could be successfully distinguished from the HCs (Fig. S2). Based on the test set, the probability index (POBD) value obtained for patients with PTB was markedly higher than that obtained for the HCs ($P < 0.001$) (Fig. 5B). Consistent with this finding, the AUC of the receiver operating characteristic (ROC) curve for the test dataset ($n = 41$) was 0.851 (95% CI: 0.737–0.965) (Fig. 5C). ROC analysis was used to verify the diagnostic ability of these biomarkers. The AUC values for four genera, *Fusobacterium* (AUC = 0.738), *Fusicatenibacter* (AUC = 0.789), *Tyzzerella* (AUC = 0.732) and *Anaerotruncus* (AUC = 0.707), were higher than 0.7, suggesting that the diagnostic efficacy was adequate (Fig. S3A). The model including all four genera had a relatively better diagnostic ability (AUC = 0.81) (Fig. S3B).

 

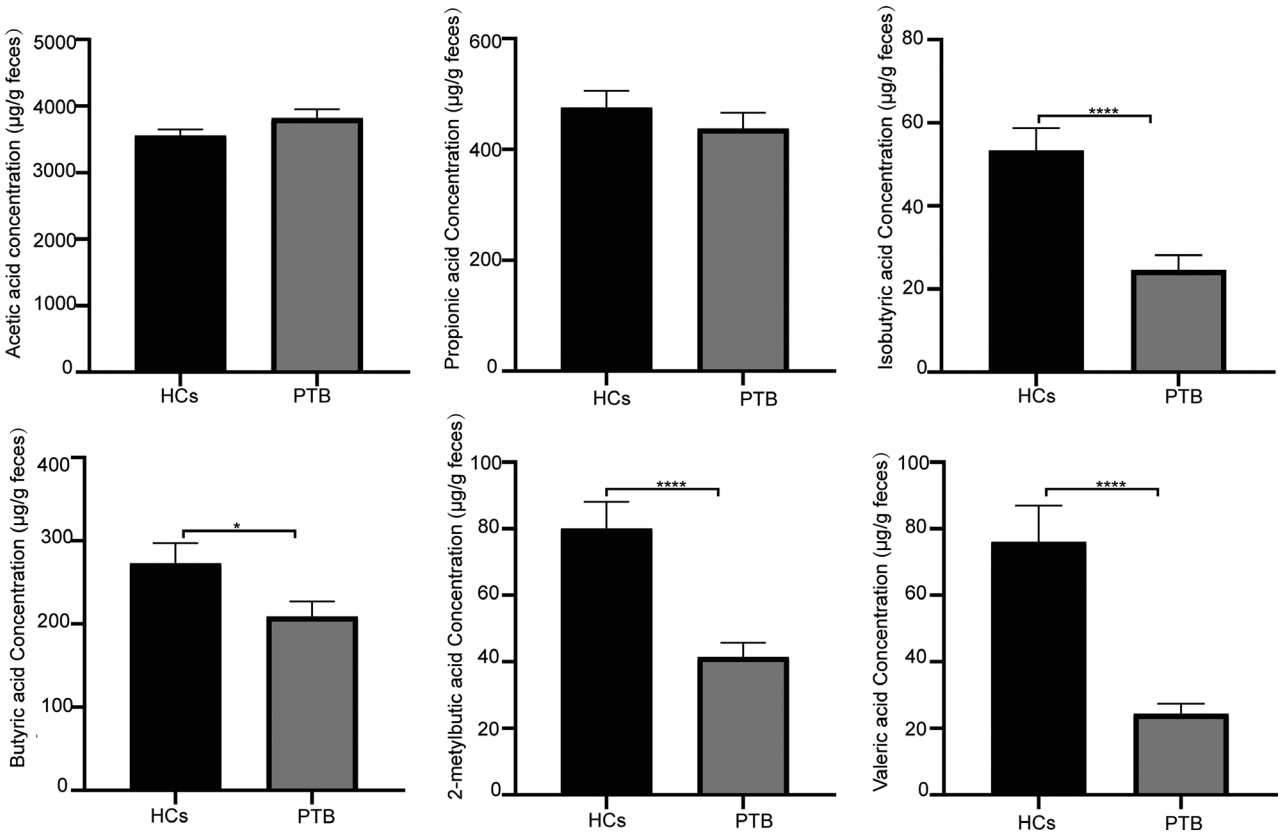


**Fig. 4.** Mycobacterium tuberculosis (Mtb) infection significantly decreased production of short-chain fatty acids (SCFAs) in faeces. The faecal concentrations of SCFAs mainly including acetic acid, propionic acid, isobutyric acid, butyric acid, 2-metylbutic acid and valeric acid were determined by GC-MS. All data are presented as means ± SEM. *$P < 0.05$, and ***< 0.0001 according to non-parametric Mann–Whitney *U* test.

We also constructed a random forest classifier to specify the diagnosis of PTB based on various faecal metabolites. A training set ($n = 94$) was used to determine the significant metabolite markers; the results of 10-fold cross-validation identified 12 differential metabolites (Table S3). As shown in Fig. 5D, a mean decrease in the Gini index-based feature importance was determined for each of the 12 parameters involved in the model. The ROC curve analysis of the training set revealed that the classifier model had an outstanding performance with regard to predicting PTB, with an AUC of 0.993 (95% CI: 0.982–1; Fig. S2). A significantly higher POBD value was obtained in the test set for the PTB patients than for the HCs ($P = 5.62e-08$) (Fig. 4GE). Interestingly, the AUC-ROC obtained in the test set ($n = 41$) was 0.998 (95% CI: 0.991–1; Fig. 4F). A total of 12 metabolites showed potential diagnostic value, with AUC values above 0.9, and the models based on a combination of top five metabolites according to the Gini index, namely 1-tetracosanol, 3-hydroxypicolinic acid, behenic acid, pyrophosphate and tromethamine, demonstrated fair discrimination for PTB (AUC = 0.996) (Fig. 3C and D).

*Correlations between the gut microbiota and other data*

To further explore the correlation between the genera and various indicators, including faecal metabolites and biochemical parameters, we conducted Spearman correlation analysis (Fig. 6A). TG was positively correlated with 4 genera that were enriched in HCs, namely *Blautia*, *Megasphaera, Rhodococcus* and *Bilophila* ($P < 0.05$). The levels of Tch, LDL and VLDL were positively associated with *Rhodococcus*, *Romboutsia*, *Gordonibacter* and *Fusicatenibacter* and negatively associated with *Cetobacterium, Weissella*, *Dorea* and *Campylobacte*r. Moreover, HDL was positively correlated with *Ezakiella* and *Elusimicrobium*, and Glu was negatively correlated with *Dorea* and *Blautia*.

A heat map was used to examine the associations between intestinal *bacteria* and the faecal metabolome (Fig. 6B); the results showed that 5 *metabolites* enriched in HCs, namely indoxyl sulfate (IS), resorcinol, orcinol, inositol and o-phosphoserine, were positively correlated with the relative abundances of *Fusicatenibacter*, unidentified *Lachnospiraceae*, *Barnesiella*, *Raoultibacter*, *Dorea*, *Tyzzerella*, unidentified *Ruminococcaceae*,

 

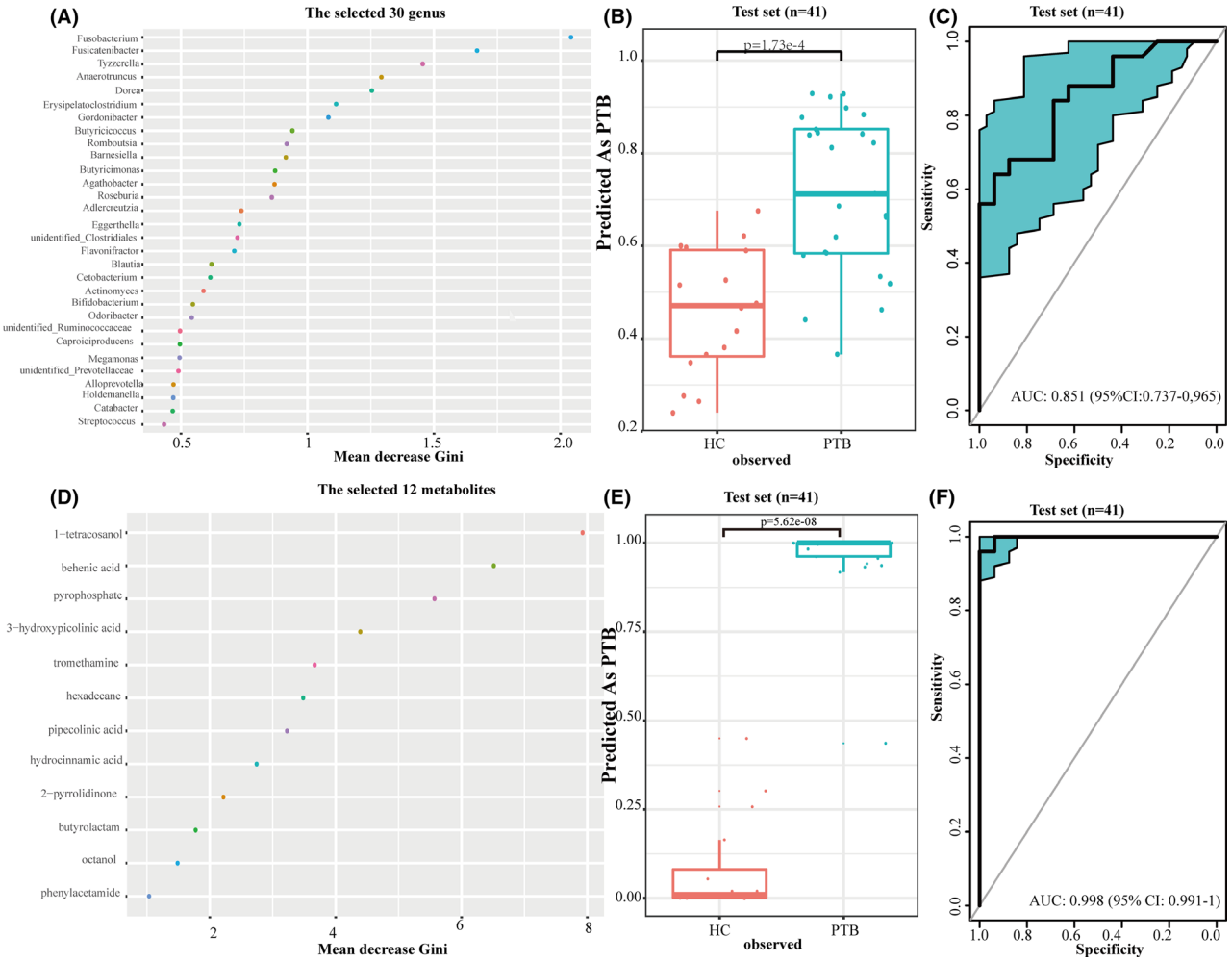


**Fig. 5.** Prediction of PTB using random forest models. A and D. Top 30 genera and 12 metabolites predicting PTB. The x-axis suggests the mean decrease in the Gini index. (B) and (E) The probability index (POBD) value was notably increased in patients with PTB compared with the HCs in the test set ($n$ = 41). The y-axis indicates the probability of samples with a predicted diagnosis of PTB. Each dot represents a single subject. C and F. The area under the curve for the random forest model obtained with the test set ($n$ = 41). PTB: pulmonary tuberculosis; HCs: healthy controls.

*Anaerotruncus*, *Butyricicoccus*, *Bilophila*, *Collinsella*, *Roseburia* and Candidatus *Soleaferrea* ($P < 0.05$). Thirteen upregulated metabolites in the PTB group, namely 2-keto-l-gluconate, ribonic acid, 5-aminolevulinic acid, taurine, 1,4-butanediol, 3-hydroxymethylglutaric acid, 3-hydroxypicolinic acid, 4-hydroxy-3-methoxyphenylglycol, 4-hydroxycyclohexylcarboxylic acid, aconitic acid, erythritol, tromethamine and hydrocinnamic acid, were negatively correlated with the abundances of *Blautia*, *Agathobacter*, *Roseburia*, Candidatus *Soleaferrea*, *Anaerotruncus*, *Tyzzerella* and *Butyricicoccus* ($P < 0.05$).

## Discussion

In general, the diversity and taxonomic numbers of the intestinal microbiota in patients with PTB were significantly lower than those in HCs, which was consistent with the observed loss of metabolic diversity and indicated that the levels of the metabolites, such as SCFAs, and the metabolite classes were considerably reduced in the PTB group. Our study is one of the first to demonstrate PTB-related alterations in the human intestinal bacteria and metabolism using integrated multi-omics data; our findings may provide inspiration for future disease diagnosis and intervention.

Consistent with previous studies, our microbiome analyses revealed a disrupted microbiota in patients with PTB, and this disruption involved a significant decrease in the predominant bacterial species and a reduced bacterial diversity (Hu *et al.*, 2019). Consistent with these findings, Khan *et al.* (2016) previously reported that antibiotic-mediated gut dysbiosis can promote Mtb infection and dissemination, which is closely related to

 

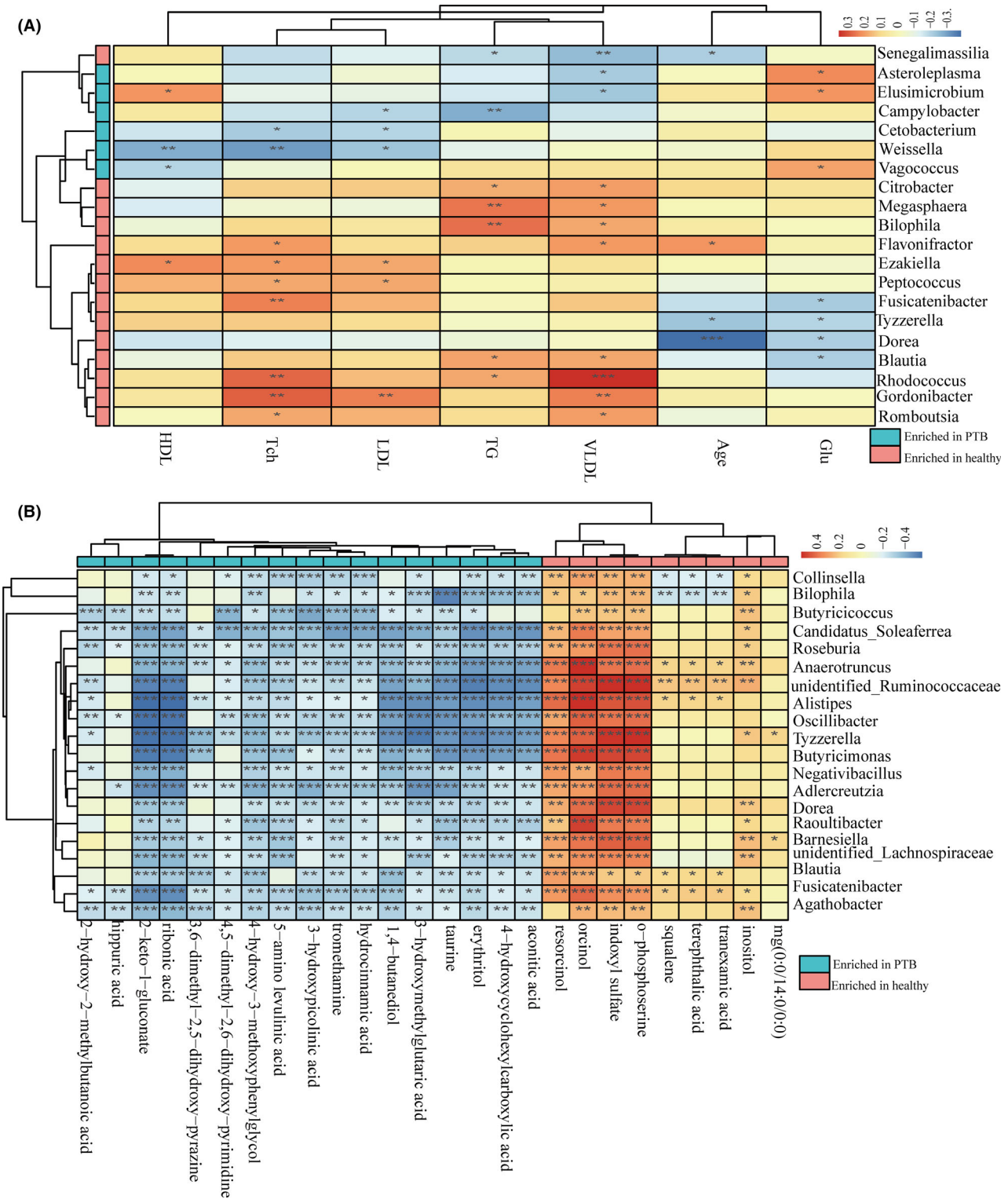


**Fig. 6.** Correlation analysis between gut microbial and biochemical parameters (A) and faecal metabolites (B). The depth of the colour in the heat maps signifies the strength of the correlation: red represents a positive correlation, whereas blue indicates a negative correlation. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Green colour represents features enriched in PTB group, while red colour represents features enriched in HCs. HCs, healthy controls; PTB, pulmonary tuberculosis.

 

decreases in interferon gamma (IFN-γ)- and tumour necrosis factor-alpha (TNF-α)-releasing CD4 T cells and an increase in Tregs observed in a mouse model of infection. The data obtained in our study indicated considerable downregulation of the abundances of beneficial bacterial phylum *Firmicutes* and various genera, such as *Bifidobacterium*, *Blautia*, *Butyricimonas*, *Ruminococcus*, *Roseburia* and *Dorea*, in the PTB samples; these findings suggested that dysregulation occurs in patients with PTB. These genera were reported to influence the production of SCFAs, particularly acetic, propionic and butyric acids (Estaki *et al*., 2016; Rivière *et al*., 2016). In our study, Mtb infection contributed to a decrease in the production of SCFAs, such as propionic acid, butyric acid, isobutyric acid, 2-metylbutic acid and valeric acid, in faeces, which is consistent with the findings of a previous study (Hu *et al*., 2019). SCFAs, particularly butyrate, inhibit inflammation through various effects, such as the induction of Tregs, maintenance of the intestinal epithelial barrier integrity and provision of energy to intestinal cells (Sun *et al*., 2017).

Significantly lower TG, HDL, LDL and VLDL levels are other interesting characteristics of patients with PTB, which may result from a decrease in the protein and fat intake due to a loss of appetite resulting from the infection. Downregulation of these parameters was also correlated with the degree of smear positivity (Deniz *et al*., 2007). Moreover, a cholesterol-rich diet is advantageous for accelerating smear-negative sputum conversion (Pérez-Guzmán *et al*., 2005). In this study, the serum levels of cholesterol (VLDL and LDL) and Tch were positively correlated with the abundances of *Romboutsia* and *Gordonibacter* and negatively associated with the richness of *Cetobacterium, Weissella* and *Campylobacte*r. A previous study suggested that the *Cetobacterium* genus is enriched in individuals with a prudent dietary pattern (Shikany *et al*., 2019). In this study, *Dorea* and *Blautia* were negatively correlated with Glu. As reported previously (Berni Canani *et al*., 2016), *Blautia* is involved in the synthesis of SCFAs, particularly butyric acids. As expected, our study showed that the production of isobutyric acid and butyric acid was significantly reduced in the PTB group. Acetic and butyric acids mediate cholesterol synthesis, fatty acid storage and glucose production (Fang *et al*., 2019). Strengthening nutritional support and prebiotic supplementation may shorten the duration of the treatment and improve medication adherence and even the treatment outcome.

To connect the structure of the microbiome community to its function, we performed an integrated analysis of the faecal metabolomics and intestinal bacteria. The overrepresented metabolites, including ribonic acid, 2-keto-l-gluconate, 4-hydroxycyclohexylcarboxylic acid, 2-hydroxy-2-methylbutanoic acid, taurine, aconitic acid and 3-hydroxymethylglutaric acid, and the underrepresented metabolites, such as o-phosphoserine, mg (0:0/14:0/0:0), terephthalic acid, IS and squalene, participate in the lipid, carbon and amino acid metabolic pathways. Previous studies have reported that squalene, an intermediate in cholesterol synthesis, has pleiotropic activity, such as antioxidant, anti-inflammatory, and drug carrier effects, that enhances immune responses (Kim and Karadeniz, 2012). Squalene has been verified to have antituberculosis activity in vitro (Jiménez *et al*., 2005). A significantly positive correlation between squalene and the abundance of *Blautia* was detected. Aromatic amino acids (tyrosine, phenylalanine and tryptophan) are involved in the production of indoles (IS) and phenols (hydrocinnamic acid) by intestinal bacteria, such as *Clostridium*, which belongs to the order *Clostridiales* (Russell *et al*., 2013; Dodd *et al*., 2017) and was markedly decreased in patients with PTB. In addition, the level of IS, which is a specific bacteria-generated uraemic toxin of tryptophan, was notably altered in the faecal samples from the patients, apparently due to appetite loss. Published evidence indicates that dietary protein intake promotes the production of IS (Poesen *et al*., 2015).

Patients with PTB and healthy individuals demonstrated not identical intestinal bacteria and faecal metabolic profiles. Interestingly, this study is the first to investigate the faecal metabolic profile of PTB patients. Previous studies have found significant differences in the composition of intestinal bacteria between patients with PTB and HCs (Hu *et al*., 2019) and demonstrated the key role of intestinal bacteria in the host immune responses against Mtb (Gupta *et al*., 2018). Early diagnosis and treatment are essential for PTB. In many countries, direct sputum smear microscopy is the most commonly used method for the diagnosis of PTB, which has various disadvantages, including low sensitivity (20–60%) (Steingart *et al*., 2006; Steingart *et al*., 2006), and challenges due to the characteristics or habits of the patients, such as the inability to produce sputum (Alnour, 2018). Additionally, Mtb culture, which is recommended as a reference standard, has a higher sensitivity; however, diagnosis using this test requires 2–6 weeks of analysis in a biosafety level III laboratory facility (Drobniewski *et al*., 2012). Molecular tests such as Cepheid Xpert MTB/RIF (Cepheid, Sunnyvale, CA, USA) and line probe assays directly identify drug-resistant TB through examining specific gene mutations (e.g. the *rpo*B, *kat*G, *inh*A genes), and can be completed in just a few hours and demonstrate good diagnostic performance both in clinical isolates and specimens (Steingart *et al*., 2014). But the application of technology is ultimately limited by the inability to produce sputum. To improve the TB prevention and control, alternative approaches for the diagnosis of sputum-free PTB patients should be explore.

 

Importantly, stool metabolites are considered the products of intestinal bacteria, host cells and diet. Integrative analysis showed the functional status of intestinal bacteria and the integration with the host cells, which further enhances our understanding of the functions of host-resident microbes (Zierer *et al*., 2018). The correlation between gut bacteria and metabolites provides crucial information on the development of PTB, which can be used for potential prevention and the identification of therapeutic targets, particularly for smear- and culture-negative patients.

Several limitations of the present study should be noted. Our cross-sectional, observational study enrolled patients prior to the treatment, and the results of this study need to be confirmed in prospective multicentre studies in different regions through shotgun microbiome metagenomics to improve the resolution of the taxonomic functional composition of the microbiome and immunological studies. The 16S rRNA gene sequencing targets only bacteria and does not target parasites, viruses and fungi that are also involved in microbial-linked faecal metabolites. Shotgun metagenomic analysis can be used to determine the differences in the functional genes present in the gut microbiome of HCs and patients with PTB. Additionally, our study did not involve patients with drug-resistant PTB or patients with non-TB lung diseases as controls; hence, our analytical models require validation in additional studies. The subjects did not adhere to a standardized diet. Diet has a broad influence on the intestinal bacteria in individuals (Gentile and Weir, 2018). Nevertheless, due to the technical shortcomings of GC-MS, the whole metabolomic phenotype of the stool samples from patients with PTB should be explored using a combination of various technologies, such as liquid chromatography–MS and nuclear magnetic resonance (NMR) spectroscopy. Finally, longitudinal studies of patients with cured PTB should address the cause-and-effect relationships between intestinal bacteria and the development of PTB.

## Conclusion

The study provides a detailed description of the disruption of the faecal flora in a cohort of patients with PTB, and the results showed that this disruption is characterized by a decrease in the community diversity and changes in the composition and profile of faecal metabolites. Biomarkers based on the faecal microbiome and metabolome can discriminate patients with PTB from healthy individuals. Thus, it is unclear whether the models can distinguish patients with PTB from patients with other diseases. Additional studies are needed to investigate the link between intestinal bacteria and PTB.

## Experimental procedures

### *Subject recruitment*

The study protocol received ethics approval from the Ethics Committee of The First Affiliated Hospital, College of Medicine, Zhejiang University (no. 1426). Each subject signed a written informed consent form before enrolment, and his/her clinical characteristics were collected.

The Mtb infection status of all participants was assessed by the T-SPOT.TB test (Oxford Immunotec, UK). Healthy controls were recruited from the local community. The diagnosis of confirmed and clinically diagnosed PTB was based on the diagnostic criteria for PTB (WS288-2017) (China). Briefly, all patients with positive T-SPOT.TB results were prospectively recruited if they had at least one of the following symptoms characteristics of PTB: cough ($\geq$2 weeks), fever ($\geq$2 weeks), night sweats, history of haemoptysis, weight loss, loss of appetite, fatigue and radiological features. Active PTB patients met one of the following criteria: (i) positive bacteriological evidence; (ii) positive histological examination; (iii) positive GeneXpert MTB/RIF; and (iv) a good response to anti-TB therapy at the two-month follow-up after the exclusion of other pulmonary infections. Additionally, all healthy subjects had negative T-SPOT.TB tests. Subjects who had other diseases or had recently taken antibiotics were excluded. Finally, the study included 83 patients who were newly diagnosed with PTB prior to chemotherapy and 52 HCs. All faecal and blood samples were freshly collected and stored at $-80^{\circ}$C until analysis.

### *Biochemical analysis*

Blood samples were collected for biochemical analysis using the automated equipment and standard methods. Alb, glucose (Glu), Tch, triglycerides (TG), HDL-c, LDL, VLDL and serum creatinine (Scr) were examined using a Hitachi 7600 automatic biochemical analyser.

### *V3-V4 16S rRNA gene sequencing analysis*

DNA was extracted from faeces using a QIAamp DNA stool mini kit (Hilden, Germany) according to the manufacturer's protocols with minor modifications. The bacterial 16S rRNA gene was amplified by PCR using the forward primer 341F (5'-CCTAYGGGRBGCASCAG-3') and the reverse primer 806R (5'-GGACTACNNGGGTATCTAAT-3') with the barcode and then sequenced using the Ion S5TMXL platform. Initially, the raw reads were generated by removing the barcode and primer sequences using Cutadapt (V1.9.1) (Langille *et al*., 2013). High-throughput sequencing reads were quality-

 

filtered and processed with Quantitative Insights Into Microbial Ecology (QIIME) (version 1.9.1). Briefly, the effective sequences were clustered to the same OTUs with ≥ 97% identity using Uparse in QIIME software (Uparse v7.0.1001). Then, the analysis of the taxonomic assignment of representative sequences of each OTU was conducted using the Mothur method and the SSUrRNA database in SILVA132 (Edgar, 2013). The alpha and beta diversities were analysed based on these normalized output data. QIIME was employed to calculate the alpha diversity (observed OTUs, Chao1, Shannon, Simpson and ACE) and beta diversity. Venn diagrams and rarefaction curves were plotted using R (version 3.6). PCoA was performed using the WGCNA, stat and ggplot2 packages in R, and the differences in the sample groupings obtained by PCoA and NMDS were compared by permutational multivariate analysis of variance (PERMANOVA and MRPP in the vegan package in R). LEfSe analyses were performed using the website http://huttenhower.sph.harvard.edu/galaxy/ with the LDA score threshold of 4. Metastats analyses were conducted at various taxonomic levels (phylum, family and genus), and the differences between the groups were examined with permutation tests. Additionally, *P* values were adjusted using the Benjamin and Hochberg false discovery rate method to obtain the q values. To assess the correlations between these data, Spearman correlation coefficients were calculated using the cor and cor.test functions in R and the correlation matrix. The sequence data of this study have been deposited in the GenBank Sequence Read Archive (SRA) of NCBI under the accession code BioProject PRJNA684468.

### *Sample processing and GC-MS analysis of the metabolomic profile*

Faecal metabolomic profiling was performed as reported previously (Ye *et al.*, 2018). Briefly, the metabolites were extracted by mixing 30-mg stool samples with 800 μl of precooled methanol (Sigma-Aldrich, St. Louis, MO, USA). After centrifugation at 14 000 rpm at 4°C for 15 min, the supernatant was collected through a 0.22-μm Millipore filter membrane and transferred into a new Eppendorf tube containing 20 μl of the internal standard (1 mg $ml^{-1}$ heptadecanoic acid). The samples were dried under a stream of nitrogen (Aosheng, Hangzhou, China). The remaining aliquot of the stool sample was dissolved in 50 μl of 15 mg $ml^{-1}$ methoxyamine hydrochloride (Sigma-Aldrich) in anhydrous pyridine (Sigma-Aldrich) and incubated at 37°C for 24 h. Then, 50 μl of N,O-bistrifluoroacetamide (BSTFA) with 1% trimethylsilyl chloride (TMCS) (Sigma-Aldrich) was added, and the mixture was vortexed for 1 min and incubated at 70°C for 2 h. Untargeted metabolomic analysis was performed using an Agilent 7890A/5975C gas chromatography–mass spectrometer system (Agilent Technologies, Santa Clara, CA, USA).

The data obtained by the GC-MS analysis were preprocessed using ChemStation (version E. 02. 02. 1431, Agilent, CA, USA) and Chroma TOF software (version 4.34, LECO, St. Joseph, MI, USA). Metabolites were annotated according to the untargeted GC-MS database from Lumingbio and NIST. Principal component analysis (PCA) and orthogonal partial least-squares discriminant analysis (OPLS-DA) were performed to visualize the differences between the groups. Significant metabolites were identified according to the OPLS-DA model with variable influence on projection (VIP) values > 1 and *P* values < 0.05 (two-tailed Student's t-test).

SCFAs were extracted from 20 mg of faeces in 500 μl of water containing 10 g $ml^{-1}$ hexanoic acid-d3 as an internal standard. After centrifugation at 150 000 rpm at 4°C for 5 min, the supernatant mixed with the same volumes of ethyl acetate (5% sulfuric acid). Then, the mixture was vortexed, centrifuged and incubated for 30 min at 4°C. Standard mixtures of 6 SCFAs were prepared by the same procedure.

### *Random forest-based classification models*

Microbiome and metabolomic data were used to construct separate random forest-based classification models (Breiman, 2001) using the randomForest R package; the relative abundance of all bacterial genera and relative metabolite contents were used as separate independent variables, and the group (HCs and PTB) was the dependent variable. All datasets were partitioned into a training set (70%) and a test set (30%) using random sampling. Initially, a 10-fold cross-validation was conducted using the training set to select a classifier with the best performance and the lowest cross-validation error. Subsequent training of the random forest classifier involved the calculation of a feature importance measure, which was defined as the ranking based on a mean decrease in the Gini index, for every feature of the training dataset. The ratio of the number of randomly generated decision trees for the predicted samples of PTB to the number of HCs to the actual observed number of the PTB samples based on the test set was defined as the POBD, which was used to evaluate the accuracy of the model. Additionally, the performance of the models was evaluated using the test dataset based on the AUC of ROC curves plotted using the pROC package in R.

### *Statistical analysis*

The data are presented as the means ± standard deviations (SDs), means ± standard error of the mean (SEM)

 

or medians with interquartile ranges, as appropriate. The Shapiro–Wilk test was used to examine the normality of the data. Continuous variables with a normal distribution were assessed using two-tailed independent sample t tests, whereas the data that did not fit a normal distribution were assessed using a non-parametric Mann–Whitney test. The categorical data were analysed using the chi-squared test. Statistical significance was defined as $P < 0.05$, and the analyses were performed with SPSS (version 21, SPSS Inc., Chicago, IL, USA). Images were generated by GraphPad Prism 8.0 and R.

**Acknowledgements**

We sincerely thank all the subjects and our colleagues who contributed to the study.

**Funding Information**

The National Science and Technology Major project of the13th Five-Year Plan (2017ZX10105001006001; 2017ZX10105001006002) and the Medical and Health Science and Technology Project of Zhejiang province (2015KYB137) supported the research.

**Conflicts of Interest**

All authors declare no competing interests related to this study.

**Authors' contributions**

L.J.L., S.T.W., K.J.X., L.Y.Y. and H.Y.H. designed the research. S.T.W., H.Y.H., Y.M.Z., R.F.F., M.F.Y. and M.X. recruited the subjects and collected and retreated the samples. S.T.W., Z.K.J. and H.Z. carried out the experiments. L.X.L., S.T.W., C.D. and L.Z. performed the data analysis. S.T.W., L.Y.Y., K.J.X. and L.J.L. drafted the article. H.Y.H., L.J.L., L.X.L. and K.J.X. revised the article. All authors have approved the final version of the manuscript.

**References**

Alnour, T.M.S. (2018) Smear microscopy as a diagnostic tool of tuberculosis: review of smear negative cases, frequency, risk factors, and prevention criteria. *Indian J Tuberc* **65:** 190–194.

Berni Canani, R., Sangwan, N., Stefka, A.T., Nocerino, R., Paparo, L., Aitoro, R., *et al*. (2016) Lactobacillus rhamnosus GG-supplemented formula expands butyrate-producing bacterial strains in food allergic infants. *ISME J* **10:** 742–750.

Breiman, L. (2001) Random forests. *Mach Learn* **45:** 5–32.

Budden, K.F., Gellatly, S.L., Wood, D.L., Cooper, M.A., Morrison, M., Hugenholtz, P., and Hansbro, P.M. (2017) Emerging pathogenic links between microbiota and the gut-lung axis. *Nat Rev Microbiol* **15:** 55–63.

Deniz, O., Gumus, S., Yaman, H., Ciftci, F., Ors, F., Cakir, E., *et al*. (2007) Serum total cholesterol, HDL-C and LDL-C concentrations significantly correlate with the radiological extent of disease and the degree of smear positivity in patients with pulmonary tuberculosis. *Clin Biochem* **40:** 162–166.

Dodd, D., Spitzer, M.H., Van Treuren, W., Merrill, B.D., Hryckowian, A.J., Higginbottom, S.K., *et al*. (2017) A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature* **551:** 648–652.

Drobniewski, F., Nikolayevskyy, V., Balabanova, Y., Bang, D., and Papaventsis, D. (2012) Diagnosis of tuberculosis and drug resistance: what can new tools bring us? *Int J Tuberculosis Lung Dis* **16:** 860–870.

Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10:** 996–998.

Estaki, M., Pither, J., Baumeister, P., Little, J.P., Gill, S.K., Ghosh, S., *et al*. (2016) Cardiorespiratory fitness as a predictor of intestinal microbial diversity and distinct metagenomic functions. *Microbiome* **4:** 42.

Fang, Q.Y., Hu, J.L., Nie, Q.X., and Nie, S.P. (2019) Effects of polysaccharides on glycometabolism based on gut microbiota alteration. *Trends Food Sci Technol* **92:** 65–70.

Gauguet, S., D'Ortona, S., Ahnger-Pier, K., Duan, B., Surana, N.K., Lu, R., *et al*. (2015) Intestinal microbiota of mice influences resistance to staphylococcus aureus pneumonia. *Infect Immun* **83:** 4003–4014.

Gentile, C.L., and Weir, T.L. (2018) The gut microbiota at the intersection of diet and human health. *Science* **362:** 776–780.

Gupta, N., Kumar, R., and Agrawal, B. (2018) New players in immunity to tuberculosis: the host microbiome, lung epithelium, and innate immune cells. *Front Immunol* **9:** 709.

Hu, Y., Feng, Y., Wu, J., Liu, F., Zhang, Z., Hao, Y., *et al*. (2019) The gut microbiome signatures discriminate healthy from pulmonary tuberculosis patients. *Front Cell Infect Microbiol* **9:** 90.

Hu, Y., Yang, Q., Liu, B., Dong, J., Sun, L., Zhu, Y., *et al*. (2019) Gut microbiota associated with pulmonary tuberculosis and dysbiosis caused by anti-tuberculosis drugs. *J Infect* **78:** 317–322.

Jiang, J.W., Chen, X.H., Ren, Z., and Zheng, S.S. (2019) Gut microbial dysbiosis associates hepatocellular carcinoma via the gut-liver axis. *Hepatobiliary Pancreat Dis Int* **18:** 19–27.

Jiménez, A., Meckes, M., Alvarez, V., Torres, J., and Parra, R. (2005) Secondary metabolites from Chamaedora tepejilote (Palmae) are active against Mycobacterium tuberculosis. *Phytother Res* **19:** 320–322.

Khan, N., Vidyarthi, A., Nadeem, S., Negi, S., Nair, G., and Agrewala, J. N. (2016) Alteration in the gut microbiota provokes susceptibility to tuberculosis. *Front Immunol* **7:** 529.

Kim, S.K., and Karadeniz, F. (2012) Biological importance and applications of squalene and squalane. *Adv Food Nutr Res* **65:** 223–233.

 

Langille, M.G., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., *et al*. (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31:** 814–821.

Lee-Sarwar, K.A., Kelly, R.S., Lasky-Su, J., Zeiger, R.S., O'Connor, G.T., Sandel, M.T., *et al*. (2019) Integrative analysis of the intestinal metabolome of childhood asthma. *J Allergy Clin Immunol* **144:** 442–454.

Namasivayam, S., Kauffman, K. D., McCulloch, J. A., Yuan, W., Thovarai, V., Mittereder, L. R., et al. (2019) Correlation between disease severity and the intestinal microbiome in mycobacterium tuberculosis-infected rhesus macaques, *mBio* **10:** e01018.

National Health and Family Planning Commission of the People's Republic of China. Diagnosis for pulmonary tuberculosis (WS 288-2017). URL http://www.nhc.gov.cn/wjw/s9491/201712/a452586fd21d4018b0ebc00b89c06254.shtml.

Negatu, D. A., Yamada, Y., Xi, Y., Go, M. L., Zimmerman, M., Ganapathy, U., et al. (2019) Gut microbiota metabolite indole propionic acid targets tryptophan biosynthesis in *Mycobacterium tuberculosis*. *MBio* **10:** e02781-18.

Negi, S., Pahari, S., Bashir, H., and Agrewala, J.N. (2019) Gut microbiota regulates mincle mediated activation of lung dendritic cells to protect against mycobacterium tuberculosis. *Front Immunol* **10:** 1142.

Pérez-Guzmán, C., Vargas, M.H., Quiñonez, F., Bazavil-vazo, N., and Aguilar, A. (2005) A cholesterol-rich diet accelerates bacteriologic sterilization in pulmonary tuberculosis. *Chest* **127:** 643–651.

Poesen, R., Mutsaers, H.A., Windey, K., van den Broek, P.H., Verweij, V., Augustijns, P., *et al*. (2015) The influence of dietary protein intake on mammalian tryptophan and phenolic metabolites. *PLoS One* **10:** e0140820.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., *et al*. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464:** 59–65.

Qin, N., Zheng, B., Yao, J., Guo, L., Zuo, J., Wu, L., *et al*. (2015) Influence of H7N9 virus infection and associated treatment on human gut microbiota. *Sci Rep* **5:** 14771.

Reyes, A., Semenkovich, N.P., Whiteson, K., Rohwer, F., and Gordon, J.I. (2012) Going viral: next-generation sequencing applied to phage populations in the human gut. *Nat Rev Microbiol* **10:** 607–617.

Rivière, A., Selak, M., Lantin, D., Leroy, F., and De Vuyst, L. (2016) Bifidobacteria and butyrate-producing colon bacteria: importance and strategies for their stimulation in the human gut. *Front Microbiol* **7:** 979.

Robertson, R.C., Kaliannan, K., Strain, C.R., Ross, R.P., Stanton, C., and Kang, J.X. (2018) Maternal omega-3 fatty acids regulate offspring obesity through persistent modulation of gut microbiota. *Microbiome* **6:** 95.

Russell, W.R., Duncan, S.H., Scobbie, L., Duncan, G., Cantlay, L., Calder, A.G., *et al*. (2013) Major phenylpropanoid-derived metabolites in the human gut can arise from microbial fermentation of protein. *Mol Nutr Food Res* **57:** 523–535.

Rutten, E.P.A., Lenaerts, K., Buurman, W.A., and Wouters, E.F.M. (2014) Disturbed intestinal integrity in patients with COPD: effects of activities of daily living. *Chest* **145:** 245–252.

Shikany, J.M., Demmer, R.T., Johnson, A.J., Fino, N.F., Meyer, K., Ensrud, K.E., *et al*. (2019) Association of dietary patterns with the gut microbiota in older, community-dwelling men. *Am J Clin Nutr* **110:** 1003–1014.

Spiljar, M., Merkler, D., and Trajkovski, M. (2017) The immune system bridges the gut microbiota with systemic energy homeostasis: focus on TLRs, mucosal barrier, and SCFAs. *Front Immunol* **8:** 1353.

Steingart, K.R., Henry, M., Ng, V., Hopewell, P.C., Ramsay, A., Cunningham, J., *et al*. (2006) Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. *Lancet Infect Dis* **6:** 570–581.

Steingart, K.R., Ng, V., Henry, M., Hopewell, P.C., Ramsay, A., Cunningham, J., *et al*. (2006) Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. *Lancet Infect Dis* **6:** 664–674.

Steingart, K.R., Schiller, I., Horne, D.J., Pai, M., Boehme, C.C., and Dendukuri, N. (2014) Xpert(R) MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst Rev*: Cd009593.

Sun, M., Wu, W., Liu, Z., and Cong, Y. (2017) Microbiota metabolite short chain fatty acids, GPCR, and inflammatory bowel diseases. *J Gastroenterol* **52:** 1–8.

Thomas, S., Izard, J., Walsh, E., Batich, K., Chongsathidkiet, P., Clarke, G., *et al*. (2017) The host microbiome regulates and maintains human health: a primer and perspective for non-microbiologists. *Cancer Res* **77:** 1783–1812.

Turnbaugh, P.J., Hamady, M., Yatsunenko, T., Cantarel, B.L., Duncan, A., Ley, R.E., *et al*. (2009) A core gut microbiome in obese and lean twins. *Nature* **457:** 480–484.

WHO. *WHO Global Tuberculosis Report 2019*. Geneva, Switzerland: WHO.

Ye, J., Lv, L., Wu, W., Li, Y., Shi, D., Fang, D., *et al*. (2018) Butyrate protects mice against methionine-choline-deficient diet-induced non-alcoholic steatohepatitis by improving gut barrier function, attenuating inflammation and reducing endotoxin levels. *Front Microbiol* **9:** 1967.

Zierer, J., Jackson, M.A., Kastenmüller, G., Mangino, M., Long, T., Telenti, A., *et al*. (2018) The fecal metabolome as a functional readout of the gut microbiome. *Nat Genet* **50:** 790–795.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Fig. S1**. Dysbiosis of the faecal microbiota in patients with PTB. (A) The number of observed species was lower in the PTB group than in the HC group. (B) Rarefaction curves were constructed based on the amount of sequencing data extracted and the number of corresponding OTUs. (C) A NMDS based on Unweighted Unifrac distance showed that the distribution of the microbial community in the PTB group was different from that of the HCs group (stress value=0.191). NMDS, nonmetric multidimensional scaling.

 

**Fig. S2**. Prediction of PTB using random forest models. (A) and (B) Area under the curve for the random forest model obtained with the training set (n=94). PTB, pulmonary tuberculosis.

**Fig. S3**. Receiver operating characteristic (ROC) analysis of metabolites. (A) ROC curves of *Fusobacterium*, *Fusicatenibacter*, *Tyzzerella*, and *Anaerotruncus*; (B) ROC curves of the combination of *Fusobacterium*, *Fusicatenibacter*, *Tyzzerella*, and *Anaerotruncus*; (C) ROC curves of 1-tetracosanol, 3-hydroxypicolinic acid, behenic acid, pyrophosphate, and tromethamine; (D) ROC curves of the combination of 1-tetracosanol, 3-hydroxypicolinic acid, behenic acid, pyrophosphate, and tromethamine.

**Table S1**. Differential metabolites identified by GC-MS in the PTB group compared with the HCs group.

**Table S2**. Genera used to establish a random forest classifier for the diagnosis of PTB.

**Table S3**. Metabolites involved in the random forest model for PTB diagnosis.

 