MEDICAL SCIENCE MONITOR 25 years

CLINICAL RESEARCH

e-ISSN 1643-3750
© Med Sci Monit, 2019; 25: 2819-2834
DOI: 10.12659/MSM.915332

Received: 2019.01.24
Accepted: 2019.03.25
Published: 2019.04.17

A Preliminary Study of Microbiota Diversity in Saliva and Bronchoalveolar Lavage Fluid from Patients with Primary Bronchogenic Carcinoma

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

AG 1 Ke Wang*
CD 2 Yufen Huang*
B 1 Zhenqiang Zhang
BF 1 Jinling Liao
CD 1 Yudi Ding
CD 2 Xiaodong Fang
BG 1 Lihua Liu
CDEF 1 Jing Luo
AG 1 Jinliang Kong

1 Pulmonary and Critical Care Medicine Ward, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, P.R. China
2 BGI Genomics, BGI-Shenzhen, Shenzhen, Guangdong, P.R. China

* These authors contributed equally to this work

Corresponding Authors: Jing Luo, e-mail: 66875350@qq.com, Jinliang Kong, e-mail: kjl071@126.com

Source of support: This work was supported by the National Natural Science Foundation of China (no. 81460003, no. 81760024, no. 81760419, and no. 81760743), the Innovation Project of Guangxi Graduate Education (no. 201010 LX039), the Guangxi Natural Science Foundation (no. 2016GXNSFAA380297), the Key Research and Development Program of Guangxi (no. AB16380152), the Basic Ability Improvement Project for Young and Middle-aged Teachers in Colleges of Guangxi (no. 2019KY0125) and the Medical Excellence Award Funded by the Creative Research Development Grant from the First Affiliated Hospital of Guangxi Medical University

Background: The present study aimed to evaluate the difference in microbiota diversity in the oral cavity and fluid bronchoalveolar lavage (BALF) of patients with lung cancer.

Material/Methods: Buccal (saliva) and lower respiratory tract BALF samples were collected from 51 patients with primary bronchogenic carcinoma and 15 healthy controls, and bacterial genomic DNA was extracted. High-throughput 16S rDNA amplicon sequencing was performed, and microbial diversity, composition, and functions of microbiota were analyzed by bioinformatics methods.

Results: Patients with lung cancer have lower microbial diversity than healthy controls in both saliva and BALF samples. Significant segregation was observed between the different pathological types of lung cancer groups and the control group regardless of the sampling site. *Treponema* and *Filifactor* were identified as potential bacterial biomarkers in BALF samples, while *Filifactor* was ideal to distinguish healthy controls from lung cancer patients. Moreover, the predictive variation analysis of the KEGG (Kyoto Encyclopedia of Genes and Genomes) metabolic pathway showed that the metabolic differences in microbiota varied by sampling site.

Conclusions: Lung cancer patients carry a different and less diverse microorganism community than healthy controls. Certain bacterial taxa might be associated with lung cancer, but the exact species depends on the sampling site and the pathological type. This study provides basic data on the microbiota diversity in BALF and saliva samples from lung cancer patients. Further investigation with a larger sample size should help validate the enriched species in different pathological types of lung cancers.

MeSH Keywords: **Bronchoalveolar Lavage Fluid • Carcinoma, Bronchogenic • Microbiota • RNA, Ribosomal, 16S • Saliva**

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/915332

4847 2 9 36

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Background

Considered as a terminal illness, primary bronchogenic carcinoma (hereinafter to be referred as "lung cancer") is responsible for the most common cause of cancer-related deaths in worldwide, with a high mortality rate in both men and women. It is reported that the 5-year survival rate of lung cancer is a paltry poor 11% [1,2]. Clinically, most of the lung cancer patients are diagnosed with advanced or distant metastases at their first visit, missing the opportunity for radical surgery in the early stage. Chemoresistance gives rise to unsatisfactory efficacy of chemotherapy, and the high recurrence rate significantly affects the patients' mental state and quality of life [3]. Therefore, an in-depth understanding of the etiology and pathogenesis of lung cancer is necessary to try and achieve early detection, diagnosis, and treatment.

The link between cancer and microbes is well established, and nearly 20% of the global cancer burden is caused by microbial agents. For example, pathogens such as human papilloma virus, Epstein-Barr virus, *Helicobacter pylori*, *Escherichia coli* and *Fusobacterium nucleatum* are all closely associated with cancers [4–7]. Collectively, the ubiquitous bacteria living on and in the human body are described as the microbiota. As the second microbiome habitat behind the alimentary canal in the human body, the respiratory tract harbors numerous microbiota, containing an estimated 500 to 700 different species of bacteria [8]. Until recently, the lower respiratory tract was considered sterile, and the detection of microbes was suggestive of microbial infection [9,10]. However, culture-independent methods have proven that the lungs are not sterile even in healthy controls [11,12]. Bacterial colonization in the respiratory tract is considered normal and comprises a complex microbiome [12]. When present at the mucosal sites, microbes can be attributed to part of the tumor microenvironment of airway malignancies, and their toxic metabolites may damage the local immune barrier. Moreover, intratumoral microbes may directly influence the growth and spread of cancer cells in various ways [13].

To date, several studies have focused on the analysis of buccal sample examination and have reported that microbial population diversity is associated with stomach cancer, pancreatic ductal cancer, and oral or esophagus squamous cell carcinoma, which promotes the occurrence of cancer development accompanied by the changing structure of the oral microbial groups [14–19]. Yan et al. [20] have also demonstrated that there are possible associations of saliva microbiota with lung cancer. Particularly, their research found that levels of *Capnocytophaga* and *Veillonella* were significantly higher in the saliva from lung cancer patients [20]. However, oral microbiota is susceptible to external environmental factors, such as smoking or household coal burning, which has implications in lung tumor etiology [21]. Bronchoalveolar lavage fluid (BALF) is more objective and representative than saliva or sputum in reflecting the microbial environment of the lungs. The potential role of BALF microbiota in lung cancer susceptibility, however, has yet to be defined. To this aim, the present study explored the possible variations of oral and lung microbiota in lung cancer patients and the difference in microbial diversity in samples from the saliva and BALF.

Material and Methods

Study participants and study design

Between December 2014 and February 2016, 51 patients hospitalized in the pulmonary and critical ward in the First Affiliated Hospital of Guangxi Medical University with primary bronchogenic carcinoma (PBC) were enrolled in this study. All of the patients were first examined and clinically diagnosed via case history, chest radiography, and blood tumor marker examinations. Pathology of transbronchial lung biopsy was the most important criterion for patient inclusion in the study because it confirmed the diagnosis and validated the classification of histological pathology. Fifteen healthy controls were recruited as normal controls by advertisement and were reimbursed for their participation. The age composition, lifestyle, and eating habits of the controls were similar to those of the patients. The demographic data of all study participants, such as gender, age, ethnicity, body mass index (BMI), smoking history, pathology type, TNM staging of tumor, and the data from laboratory studies were recorded (Supplementary Table 1). The exclusion criteria for this study were as follows: the participant manifested other basic pulmonary diseases, oral disorders or the presence of removable partial dentures or orthodontic appliances; systemic diseases, such as diabetes mellitus, gastritis, hepatitis and other cancers in addition to PBC; immune-compromising diseases, such as human immunodeficiency virus (HIV) or ongoing immunosuppressive therapy; and other diseases known to affect the oral and airway microbiota. Additionally, none of the participants had received glucocorticoid or antibiotic treatment for at least 30 days before sample collection. The saliva and BALF samples were taken in parallel from each enrolled participant after clinical diagnosis and before treatment. This study was approved by the Medical Ethics Committee at the First Affiliated Hospital of Guangxi Medical University. All samples were collected according to the approved protocol. Based on the patients' full understanding of the purpose of this research, written informed consent was obtained from all participants. Patients were notified if their sample was suitable for our study, and additional verbal consent from the participants was also obtained to undertake additional research.

Procedures and specimen collection

To avoid cross contamination, saliva specimens were collected at the start of the bronchoscopy procedure before sedation or any topical anesthesia. Participants were required to fast overnight and brush their teeth before saliva collection. After gargling with 15 mL of sterile normal saline for 30 seconds, the saliva samples from participants were naturally expectorated without any stimulation into a cryostorage sterile sputum cup. The bronchoscopy with lavage was performed immediately upon completion of the saliva-collected procedure in the endoscopic examination room. After local anesthesia, the flexible fiberoptic bronchoscopy with bronchoalveolar lavage was wedged intranasally into a subsegmental bronchus in the involved focal lobe with the tumor (patients) or in the third-generation bronchus of the lingual lobe (healthy controls) by a single physician. Three aliquots of 50 mL of sterile normal saline were instilled, and the fluid was gently aspirated with a negative pressure of –40 to –50 millimeters of mercury. Suction channel use was avoided until the tip of the bronchoscope extended beyond the carina. All BALF specimens were pooled and collected in a siliconized plastic bottle placed on ice. The saliva and BALF specimens were immediately delivered to the laboratory for microbiological analysis. Before processing for DNA extraction, each BALF specimen was divided into aliquots of 1.5 mL in a sterile Eppendorf tube and stored at –80°C.

DNA Extraction

The collected saliva and BALF specimens were centrifuged at 3000×g for 10 minutes at 4°C. The supernatant was discarded with a sterile pipette, and the remaining pellet was dissolved in 0.5 mL of sterile saline. Bacterial genomic DNA from saliva and BALF samples was extracted using the QIAamp DNA Microbiome Kit (Qiagen, Germany) according to the manufacturer's instructions. DNA concentration was measured by Qubit (Invitrogen, USA).

PCR amplification of the bacterial 16S rDNA

To construct the PCR-based 16S rDNA amplicon library for sequencing, PCR enrichment of the V4 hypervariable region of 16S rDNA was performed with the forward primer 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and reverse primer 806R (5'-GGACTACHVGGGTWTCTAAT-3'). PCR cycling conditions were as follows: 95°C for 3 minutes, 30 cycles of 95°C for 45 seconds, 56°C for 45 seconds, 72°C for 45 seconds and final extension for 10 minutes at 72°C for 10 minutes. The PCR products were purified using AmpureXp beads (Agencourt, USA) to remove the unspecific products. The same procedure was also performed with the negative controls: sterile water and the mixture without template. There was no evidence of contamination in the reagents used if no PCR products were amplified in the negative control.

Sequencing and bioinformatics analysis

The qualified amplicon mixture was then sequenced on the MiSeq platform with the PE250 sequencing strategy. Before the 16S rDNA data analysis, raw reads were filtered to remove adaptors and low-quality and ambiguous bases, and then paired-end reads were added to tags by the Fast Length Adjustment of Short reads program (FLASH, v1.2.11) [22]. The tags were clustered into operational taxonomic units (OTUs) with a cutoff value of 97% using UPARSE software (v9.1.13) [23], and the representative sequence from each OTU cluster was obtained. These OTU representative sequences were used to assign taxonomy to the cluster using the Ribosomal Database Project (RDP) Classifier (v.2.2) [24] with a minimum confidence threshold of 0.8, and the training database was the Greengene database (v201305) [25]. Alpha and beta diversity were estimated by MOTHUR (v1.31.2) [26] and QIIME (v1.8.0) [27] at the OTU level, respectively. Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt, v1.1.3) [28] was used to predict KO abundance from OTU data, and differential KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways were identified according to their reporter score from the Z-scores of individual KO groups [29]. Pathway reporter scores over 1.95 (corresponding to 95% confidence to a normal distribution) were considered statistically significant. SourceTracker software [30] was used to estimate the proportions for BALF samples (sink samples) with saliva samples (source samples).

Statistical analysis

Statistical analysis was performed by R software (v3.4.10). For demographic and clinical data, qualitative data were compared via the Kruskal-Wallis test, and quantitative data were determined via Crosstabs with the chi-square test (or Fisher's exact test for sparse counts). PERMANOVA was performed by the R package "vegan" [31] with 9999 permutations. Differential relative abundance of taxonomic groups at the phylum or genus level between lung cancer patients and healthy controls in BALF and saliva samples was calculated by using a 2-tailed Wilcoxon rank-sum test with $P<0.05$. Enrichment in healthy controls or lung cancer patients was determined by the higher mean rank. The random forest model was used to estimate the importance of each differential genus and predict the disease status based on the most importance genus (R package "randomForest") and the receiver-operating characteristic curve (ROC) was drawn using the pROC package [32]. The correlations between clinical index and microbiota were calculated by Spearman's rank correlation (R package "cor.test").

Results

Baseline characteristics of subjects

Demographic data from 51 patients with PBC and 15 healthy controls are displayed in Table 1. There were no significant differences in confounders for lung cancer risk or factors that are known to alter the human microbiome and progression including age, gender, BMI, smoking history, drinking history between the healthy controls and the lung cancer patients. For laboratory studies, differences were mainly generated from classical tumor markers between healthy controls and lung cancer patients, such as carbohydrate antigens, cytokeratin, carcinoembryonic antigen, neuron-specific enolase and squamous cell carcinoma antigen. However, no significant differences were found between infection-related and metabolic indicators, such as white blood cell count, fasting blood glucose, C-reactive protein, eosinophil granulocytes, erythrocyte sedimentation rate, neutrophil granulocytes, and procalcitonin. None of the 66 participants had evidence of respiratory tract infections or had received antibiotic treatment within 1 month prior to the initial saliva collection and the following bronchoscopy. Two types of samples (saliva and bronchoalveolar lavage) were taken from each participant. The bronchoscopy with lavage was performed immediately upon completion of saliva collection.

Alpha and beta diversity between the healthy control and lung cancer groups

In all of the 128 samples (4 BALF samples failed to amplify using PCR and were excluded from further analysis), a total of 28 502 248 raw reads were detected by 16S rDNA pyrosequencing that were approximately 252 bp in length. Each sample contained 222 673 [±69 857 standard deviation (SD)] reads on average. After trimming the sequences and discarding the low-quality reads, adaptors and N bases, 23 969 410 high quality sequences were obtained that take up nearly 84.13% (±18.16% SD) of the raw reads on average. A total of 1709 OTUs were observed across all participants, and the range of OTU numbers varied from 79 to 939 (Supplementary Table 2).

The richness and evenness in each sample were estimated using observed species and the Shannon and Simpson indexes (Supplementary Table 3). Nearly all of the rarefaction curves featured 2-component curves that included a sharp slope in the beginning and a flat slope for the remainder (Figure 1A and Supplementary Figure 1A). This result is a sign of sufficient depth and coverage of sequencing for the number of patients enrolled in the present study. The Shannon and Simpson indexes are 2 indicators commonly used for the quantitative description of alpha diversity in community polymorphisms. The Shannon index is in direct proportion to microbial diversity in samples, whereas the Simpson index is negatively correlated with microbiota diversity. As shown in Figure 1B and 1C, the Shannon index was significantly lower in the lung cancer groups than the healthy control group, while the Simpson index was significantly higher than that in the healthy control group (P=0.002 for Shannon index and P=0.033 for Simpson index in the saliva samples and P=6.55e-07 for Shannon and P=2.46e-07 for Simpson in the BALF samples) for samples from the same sites. There was no difference between saliva and BALF samples in the healthy controls (P=0.683 for Shannon index and P=0.354 Simpson index) for the Shannon or Simpson indexes. Nevertheless, there was a significant difference between the saliva and BALF samples in the lung cancer groups (P=0.029 for Shannon index and P=0.004 Simpson index). When the different pathological lung cancer groups were compared, both the Shannon and Simpson indexes were still significantly different in the lung cancer groups if samples were obtained from the same site as the healthy control group (Supplementary Table 3). When stratifying these results by the pathological type of lung cancer with samples from different sites, the only difference was from the comparison of the lung squamous cell carcinoma (LSCC) group with the healthy control group (P=0.011 for Shannon index, P=0.004 for Simpson index, Supplementary Figure 1B and 1C).

As the 2 sampling sites showed divergence in the alpha diversity between the lung cancer groups and healthy control group, we next performed beta diversity analysis to depict this divergence and assess its accuracy. To determine whether the microbiota in samples distinguished PBC patients from healthy controls and whether the diversity of microbiota was associated with the sampling site, principal coordinate analysis (PCoA) was employed to perform the dimension-reduction treatment of the OTU data set when comparing the overall structure of microbiota collected from different samples and sites. The PCoA analysis results showed that microbiota constitution in both BALF and saliva samples from cancer patients was clearly different than the healthy controls (Figure 1D). PERMANOVA analysis also indicated that the microbial structure in the lung cancer groups was significantly different from that in the healthy control group at the 2 sampling sites (BALF samples: R^2=0.070, P=1e-04; saliva samples: R^2=0.066, P=1e-04). Furthermore, significant segregation of microbial communities was also observed between different pathological types in the lung cancer groups and the healthy control group regardless of the sampling site (Supplementary Figure 1D for BALF samples and Supplementary Figure 1E for saliva samples).

Taxonomy-based characterization of microbiota in saliva and BALF samples

The microbiota profile of samples was identified and quantified at the phylum and genus levels through taxonomic assignment against the reference database using the RDP classifier to reveal their relative abundance in each of the microbiota samples. A total of 37 phyla were responsible for >99% of all sequence reads, and 3 predominant phyla were Firmicutes (30.90%), Bacteroidetes (30.22%), and Proteobacteria (24.25%) in all samples (Figure 2A). A total of 280 genera were identified across all participants. The 7 most abundant

Table 1. The baseline characteristics of enrolled healthy controls (n=15) and lung cancer (n=51) subjects for the present study.

	Characteristics	HC; n=15	LC; n=51			*P* value*
			LAC; n=18	LSCC; n=19	SCLC; n=14	
General data	Age (yeas, mean ±SD)	56.9±6.1	54.8±10.7	62.4±8.4	61.3±5.6	0.057
	Male (n, %)	8 (53.3%)	10 (55.6%)	13 (68.4%)	8 (57.1%)	0.800
	Zhuang (n, %)	9 (60.0%)	10 (55.6%)	4 (21.1%)	9 (64.3%)	**0.041**
	Smoker (n, %)	6 (40.0%)	4 (22.2%)	9 (47.4%)	7 (50.0%)	0.510
	Drinker (n, %)	4 (26.7%)	2 (11.1%)	9 (47.4%)	4 (28.6%)	0.144
	BMI (kg/m^2, mean ±SD)	21.8±2.7	20.2±1.6	21.2±3.2	20.1±2.4	0.229
Tumor TNM staging	T2	N/A	2 (11.1%)	0	1 (7.1%)	0.728
	T3	N/A	6 (33.3%)	6 (31.6%)	4 (25.6%)	
	T4	N/A	10 (55.6%)	13 (68.4%)	9 (64.3%)	
Laboratory studies (mean ±SD)	CRP (mg/L)	25.0±13.7	29.6±22.7	40.3±51.5	49.7±66.6	0.990
	EG (10^9/L)	0.3±0.1	0.6±1.2	0.4±0.3	0.3±0.2	0.789
	ESR (mm/h)	33.0±42.4	47.9±31.4	63.3±39.9	56.5±32.4	0.584
	FBG (U/mL)	4.9±0.8	4.8±1.3	5.4±1.7	5.5±1.8	0.580
	Ferritin (ng/mL)	286.9±95.3	386.6±113.6	545.8±478.9	591.1±407.3	**0.049**
	NG (10^9/L)	5.3±1.9	5.6±2.8	6.4±2.7	6.8±3.7	0.623
	PCT (ng/mL)	0.1±0.0	0.2±0.0	0.3±0.5	0.2±0.2	0.486
	WBC (10^9/L)	8.5±2.5	8.5±4.1	9.5±2.6	9.7±4.2	0.398
	CA 125 (U/mL)	9.7±4.6	298.1±786.1	39.6±28.0	57.6±70.3	**<0.001**
	CA 19-9 (U/mL)	5.8±5.2	202.3±738.2	16.0±17.8	32.6±44.4	**0.004**
	CEA (ng/mL)	1.4±0.8	404.7±1491.5	14.9±53.6	262.7±899.0	**<0.001**
	CK 19 (ng/mL)	2.1±0.9	13.2±28.9	14.2±17.8	7.2±5.9	**<0.001**
	NSE (ng/mL)	13.3±4.4	18.7±7.2	37.8±76.8	81.0±67.9	**<0.001**
	SCCA (ng/L)	0.9±0.8	1.0±0.9	5.1±4.0	0.8±0.8	**<0.001**

SD – standard deviation; HC – healthy control; LC – lung cancer; LAC – lung adenocarcinoma; LSCC – lung squamous cell carcinoma; SCLC – small cell lung; BMI – body mass index; FBG – fasting blood glucose; CRP – C-reactive protein; EG – eosinophil granulocyte; ESR – erythrocyte sedimentation rate; NG – neutrophil granulocyte; PCT – procalcitonin; WBC – white blood cells; CA-125 – carbohydrate antigen 125; CA19-9 – carbohydrate antigen 19-9; CEA – carcino-embryonic antigen; CK-19 – cytokeratin-19 antigen; NSE – neuron-specific enolase; SCCA – squamous cell carcinoma antigen. * P value based on Kruskal-Wallis Test (continuous variables) or Chi-Square (categorical variables, or Fisher's exact test for sparse counts) all four groups. Significant values are in blod.

and frequently detected genera that each represented at least 3% in average relative abundance were classified as *Prevotella* (22.46%), *Neisseria* (12.07%), *Veillonella* (10.55%), *Streptococcus* (9.40%), *Haemophilus* (4.44%), *Capnocytophaga* (4.21%) and *Fusobacterium* (3.50%) (Figure 2B). The constitution of microbiota in BALF and saliva showed great homogeneity, as displayed in Figure 3. The Spearman's rank correlation rho of relative abundance for both samples was 0.674 in the healthy control group and 0.656 in the lung cancer group (Figure 3A, 3B).

We further estimated the proportion of BALF microbiota that come from saliva samples by SourceTracker software, and homogeneity analysis demonstrated that a certain number of bacterial species in BALF come from saliva (Figure 3C, 3D). Certainly, species differences existed between the 2 sources of samples in both the healthy control and lung cancer groups. As shown in Figure 2A, relative abundance analysis at the phylum level indicated that Firmicutes was significantly enriched in BALF samples from patients with lung adenocarcinoma (LAC).

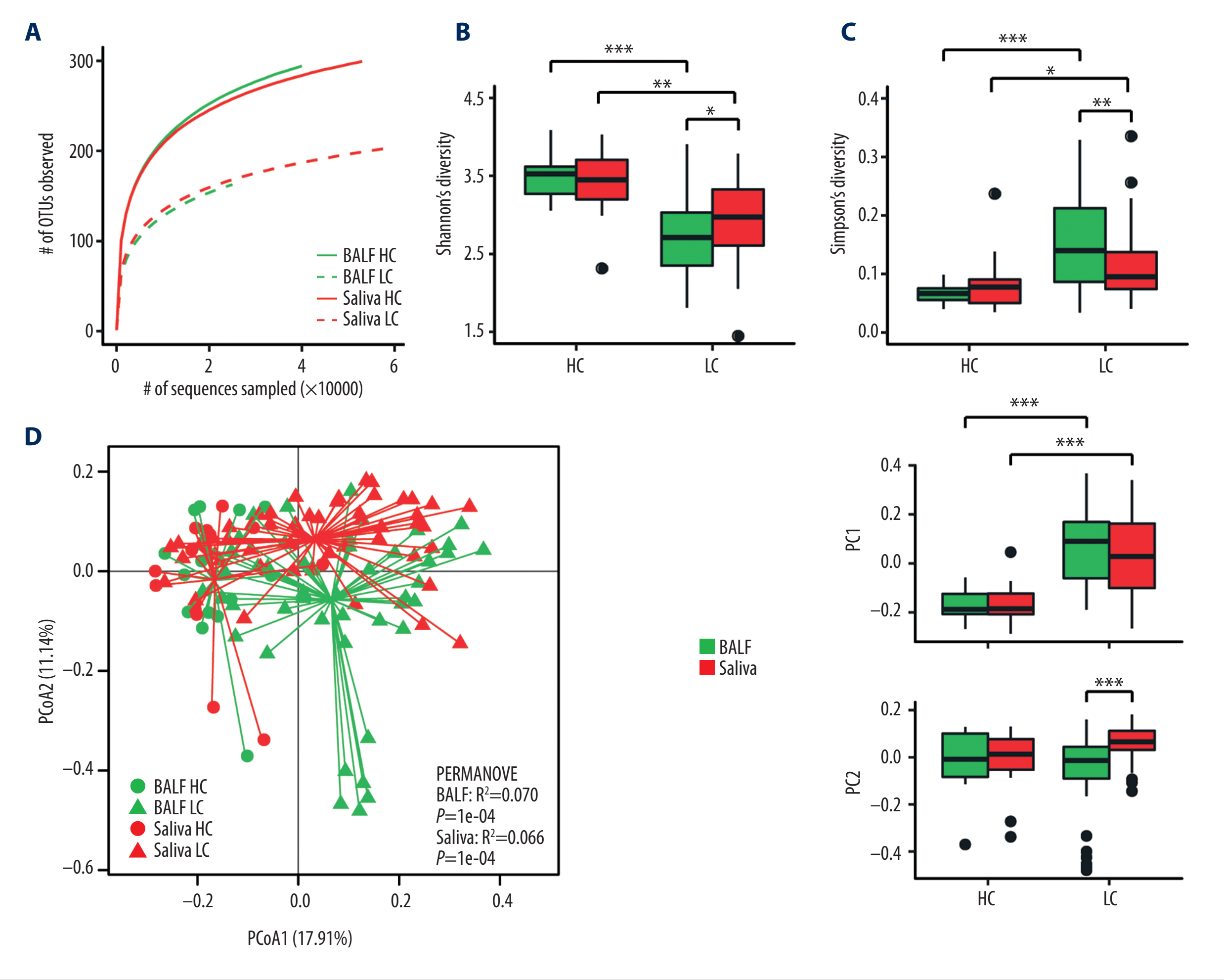


Figure 1. Comparison of the alpha diversity and beta diversity for microbiota from different sampling sites and disease groups. (**A**) The rarefaction curve for samples from different sites and groups was drawn to evaluate the depth and coverage of sequencing and the richness of species. (**B, C**) Shannon and Simpson indexes were used to evaluate the diversity of samples. (**D**) Principal coordinate analysis (PCoA) based on the unweighted UniFrac distance matrix. The green and red colors represent bronchoalveolar lavage (BALF) and saliva samples, respectively. *P<0.05, ** P<0.01, and *** P<0.001 by Wilcoxon rank-sum test.

Actinobacteria was concentrated in saliva samples from LAC and LSCC patients, while Spirochetes was concentrated in saliva samples from small-cell lung cancer (SCLC) patients. At the genus level (Figure 2B), relative abundance analysis indicated that *Pseudomonas* was enriched in the BALF samples from LAC and SCLC patients. *Veillonella* and *Corynebacterium* were abundant in the BALF samples from LSCC patients as well. In saliva samples, *Haemophilus* and *Streptococcus* were found to be enriched in the LSCC patients, *Rothia* and *Actinomyces* were enriched in both LAC and LSCC patients, while *Treponema* was enriched in SCLC patients.

This study tried to gain further insights into the microbial community in BALF and saliva samples from healthy controls and cancer patients to explore whether there were any lung cancer-associated or site-specific taxa. By comparing the bacterial phyla with mean values greater than 0.1% in at least 1 group, taxa with significant differences between healthy controls and lung cancer patients were considered (Table 2). We observed that Spirochetes, Synergistetes and Tenericutes were 3 common differential phyla for cancer patients regardless of whether the sample was from BALF or saliva. Firmicutes and Fusobacteria were identified as the different phyla in BALF samples, while Actinobacteria was identified in saliva samples. Significant differences in species between the healthy control and lung cancer groups in the BALF and saliva samples at the genus level are displayed in Figure 4A and 4D.

Interestingly, when stratifying these results by different pathological types of lung cancer, a significant enrichment of

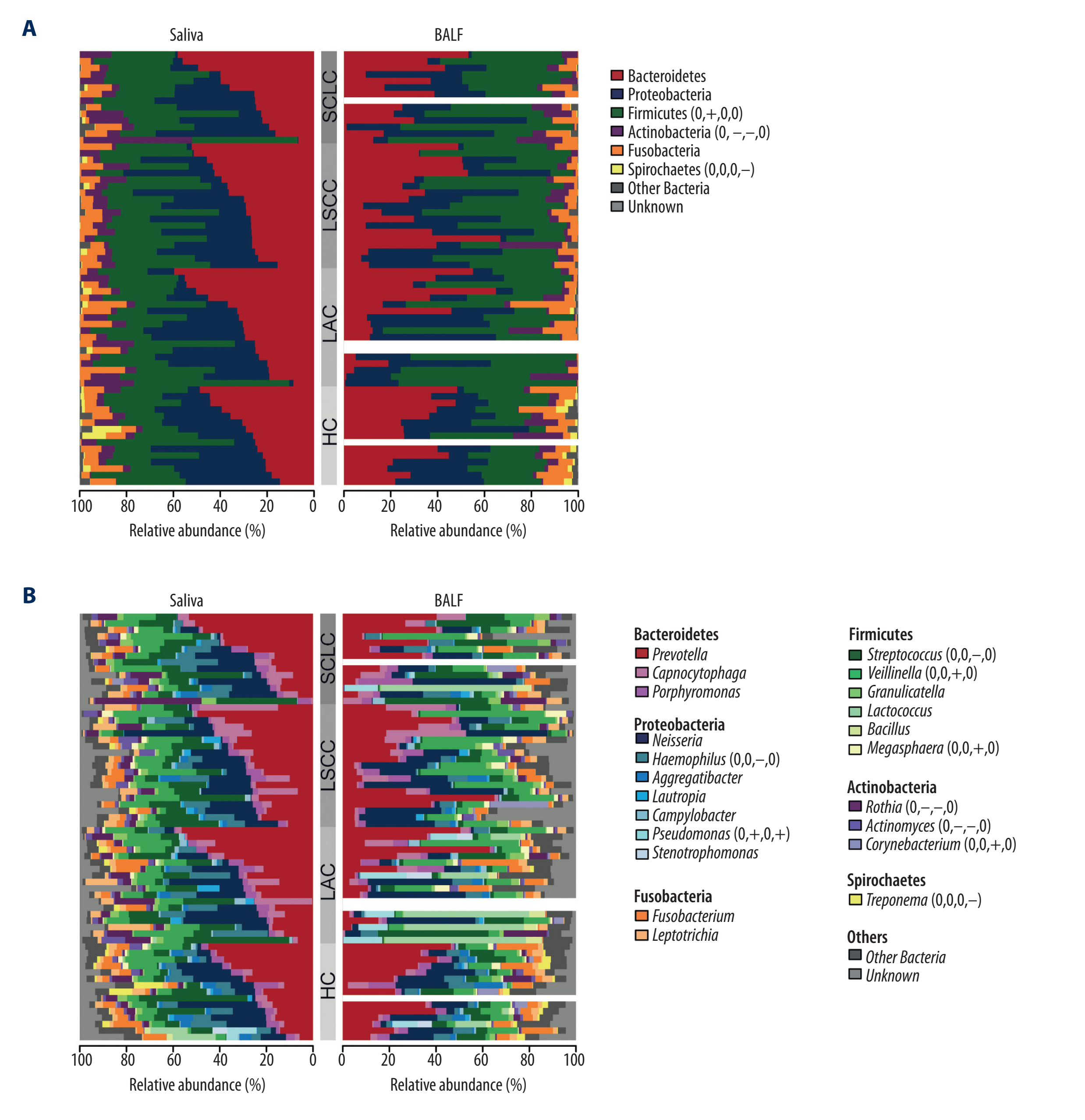


Figure 2. Taxonomical composition of BALF and saliva microbiomes. The predominant bacteria of the phylum (**A**) and genus (**B**) levels are shown. Each bar represents a single sample, and samples were organized by the order of healthy controls, LAC, LSCC, and SCLC patients from the bottom to top. Blank bars representing samples were excluded from further analysis. A paired Wilcoxon rank-sum test was used to detect the different taxa between BALF and saliva samples. Phyla or genera with significantly different relative abundances were marked with '+' or '–' ($P<0.05$), which also corresponded to BALF-enriched or saliva-enriched, respectively. BALF – bronchoalveolar lavage; LAC – lung adenocarcinoma; LSCC – lung squamous cell carcinoma; SCLC – small-cell lung cancer.

Veillonella (16.85% versus 7.17%, P=0.014) and *Capnocytophaga* (6.60% versus 1.86%, P=0.035) was observed when comparing the LSCC patients with the healthy control group. *Lactobacillus* was enriched in the SCLC group (0.43% versus 0.09%, P=0.024). In addition, clear differences in several taxa were also found between the healthy control group and any other pathological types in the lung cancer group. Further details about the relative abundance of taxa at the genus level are displayed in Supplementary Figure 2A and Supplementary Table 4.

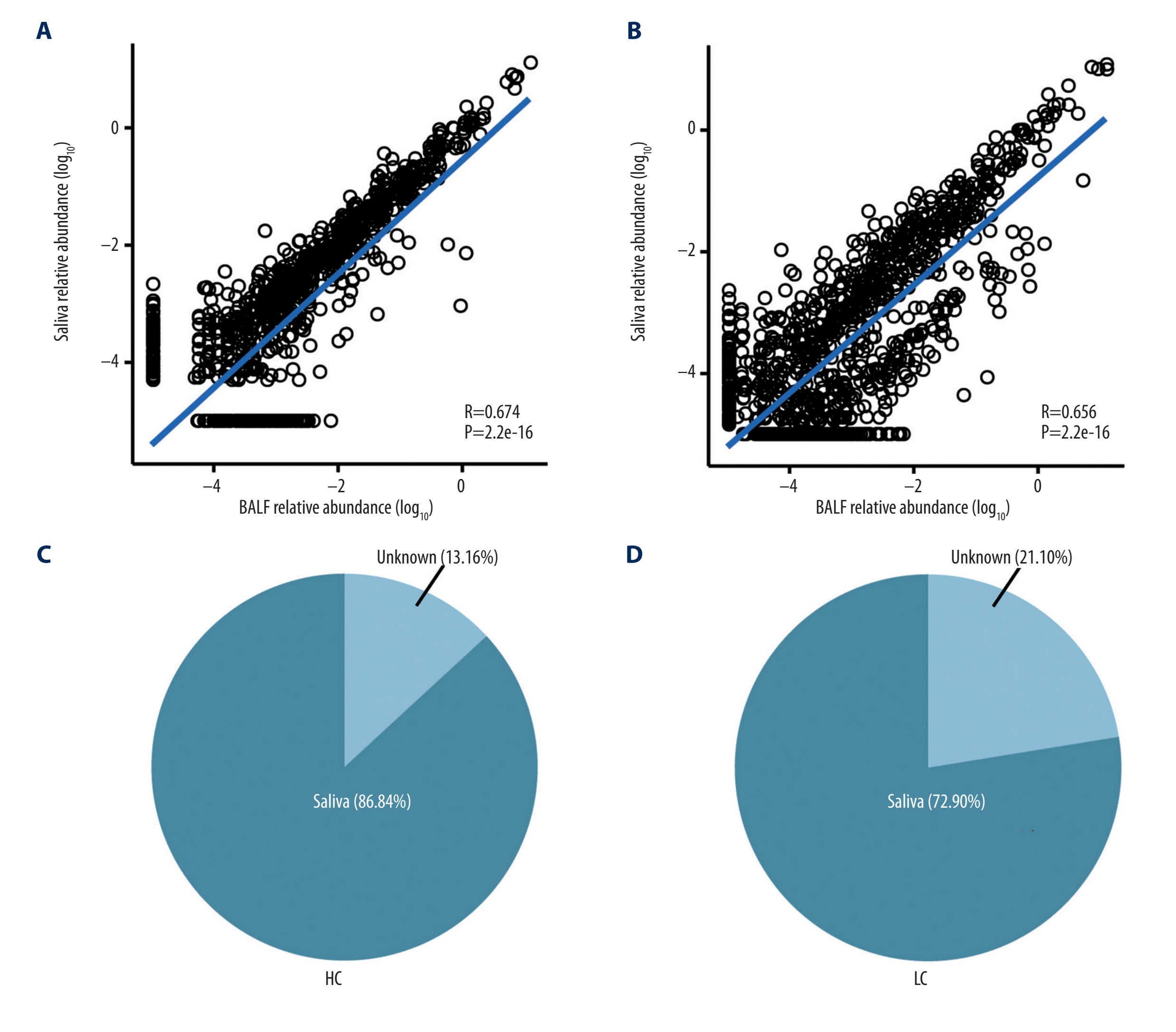


Figure 3. Comparing the microbiota similarities between BALF and saliva samples. (**A, B**) The correlation of BALF and saliva samples based on the relative abundance of OTU level. The horizontal and vertical axes show the log 10 value of the relative abundance. (**C, D**) Predicted proportions for healthy BALF and cancer BALF samples with saliva samples by SourceTracker. BALF – bronchoalveolar lavage; OUT – operational taxonomic units.

In the saliva samples, *Streptococcus* (11.94% versus 9.20%, *P*=0.033), *Capnocytophaga* (5.08% versus 1.82%, *P*=0.018) and *Actinomyces* (2.25% versus 1.3%, *P*=0.039) were found with significantly higher relative abundance in the lung cancer groups than in the healthy control group. Analyses were then stratified by pathological type of lung cancer, and evidence of multiple genus-level differences among different subtypes of lung cancer and healthy controls was observed. The LSCC group had an enrichment of *Capnocytophaga* (6.48% versus 1.82%, *P*=0.030) and *Actinomyces* (2.63% versus 1.31%, *P*=0.033), and the SCLC group showed an enrichment of *Streptococcus* (13.12% versus 9.20%, *P*=0.020) compared with the healthy control group. In addition, *Rothia* also showed a significant difference between the LAC group and the healthy control group (4.77% versus 2.06%, *P*=0.040). Additionally, clear differences in several taxa were also found between the healthy control group and any other pathological types of lung cancer in the saliva samples, and additional details are displayed in Supplementary Figure 2B and Supplementary Table 4.

Selection of potential bacterial biomarkers from BALF and saliva samples

We performed random forest analysis to select the top 10 most important genera from the microbial profiles to discriminate the healthy control and lung cancer groups for BALF and saliva samples (Figure 4B, 4E). The selected genera were synthetically evaluated for their incidence in the microbial community and

Table 2. Differences of phyla between healthy and cancer groups in BALF and saliva samples.

Phylum*	HC mean abundance (SD)		LC mean abundance (SD)		*P* value	Enriched
BALF						
Firmicutes	23.84%	(7.28%)	38.42%	(18.08%)	0.005	LC
Fusobacteria	9.18%	(4.71%)	5.12%	(5.01%)	0.003	HC
Spirochaetes	1.82%	(1.57%)	0.11%	(0.26%)	5.67E-07	HC
Synergistetes	0.16%	(0.33%)	0.03%	(0.08%)	2.56E-04	HC
Tenericutes	0.43%	(0.55%)	0.11%	(0.19%)	0.004	HC
Saliva						
Actinobacteria	4.10%	(2.73%)	7.50%	(7.09%)	0.025	LC
Spirochaetes	2.21%	(3.78%)	0.48%	(0.89%)	0.003	HC
Synergistetes	0.13%	(0.13%)	0.08%	(0.23%)	0.008	HC
Tenericutes	0.35%	(0.59%)	0.06%	(0.11%)	0.001	HC

* Phyla were display with mean values more than 0.1% in at least one groups.

were ordered by using the mean decreasing accuracy and the Gini coefficient. *Treponema* was selected as the bacterial biomarker candidate with the highest mean decreasing accuracy and the Gini coefficient value among 10 selected genera, while *Filifactor* was identified as the potential bacterial biomarker in saliva samples. ROC analysis was performed to evaluate the sensitivity and specificity of these potential bacterial biomarkers to evaluate their preclinic utilities. The overall performance of the 2 potential biomarkers in detecting lung cancer and identifying different pathological types of lung cancer (LAC, LSCC, and SCLC) are displayed in Figure 4C and 4F. *Treponema* generated a ROC value of 85.57% (95% CI: 73.22–97.91%) in BALF samples to distinguish between healthy controls and lung cancer patients, while *Filifactor* obtained a ROC value of 79.74% (95% CI: 68.58–90.90%). Further ROC analysis of different pathological lung cancer subtypes is displayed in Supplementary Figure 3 for BALF and saliva samples, and the ROC values are noted next to the ROC curves. We also performed Spearman's rank correlation analysis to evaluate the potential relationship between the selected genera and the clinical index, which is usually launched as a routine examination for diagnosing lung cancer. The heat map of the Spearman's rank correlation between microbiota genera and clinical index in BALF and saliva samples is displayed as Figure 5A and 5B, respectively. Clearly, the results of the correlation analysis indicated that significant correlations between *Treponema* and clinical lung cancer markers, including SCCA, CA125, CK-19, CA-199, and CEA, were observed in the BALF samples. Similarly, *Filifactor* was associated with CA125, CK-19, and CA-199.

Predictive variation analysis of KEGG metabolic pathway

The metabolic characteristics of the microbial community always vary from the living environment. KEGG pathway enrichment analysis and functional analysis were carried out to predict the metabolic difference in microbiota in different sources of samples. As shown in Figure 6, we explored the metabolic characteristics and differences in the microbiota of lung cancer from BALF and saliva samples. The metabolic differences in the second and third KEGG pathway levels for the microbiota of lung cancer and healthy controls are displayed. We found that the metabolic characteristics of microbiota varied by sampling site. In general, there were 16 and 13 discrepant maps enriched in the healthy control and lung cancer groups, respectively. Microbes in the BALF of cancer patients always exhibited apparent metabolic behaviors for the signaling of amino acid metabolism, metabolism of terpenoids and polyketides and xenobiotic biodegradation and metabolism. Conversely, several metabolic pathways, such as carbohydrate metabolism, energy metabolism, and replication and repair signaling, were activated in healthy individuals. Notably, more functional differences were observed in saliva samples than in healthy controls. Overall, there were 25 and 29 discrepant maps enriched in the saliva samples from the healthy control and lung cancer groups, respectively. Antimicrobial resistance, folding, sorting and degradation, glycan biosynthesis and metabolism, metabolism of cofactors and vitamins and nucleotide metabolism were enriched only in cancer patients. Rather, some metabolic-related signaling pathways, such as amino acid metabolism, carbohydrate metabolism, energy metabolism and lipid metabolism, were activated in the cancer group, while others were depleted.

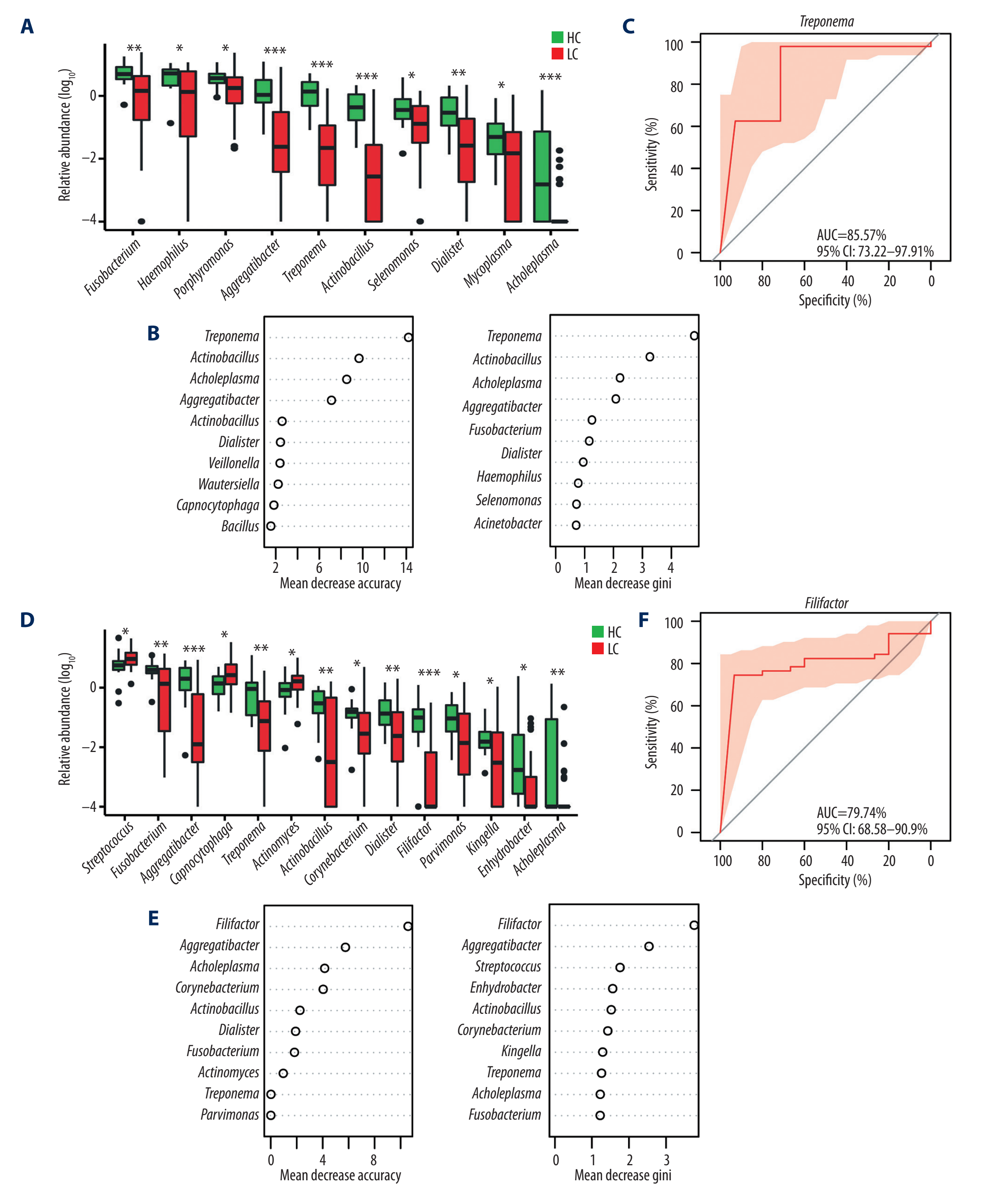


Figure 4. Significant different species between the healthy control and lung cancer groups in bronchoalveolar lavage (BALF) and saliva samples. (**A, D**) Different genera of BALF and saliva microbiota between healthy control and lung cancer groups. The genera were shown only if their abundance was over 1% in at least 1 sample and with a significant difference (Wilcoxon rank-sum test). * *P*<0.05, ** *P*<0.01, and *** *P*<0.001 by Wilcoxon rank-sum test. **(B, E)** The top 10 most important genera to discriminate the healthy control and lung cancer groups using random forest analysis in BALF and saliva samples. The importance of each genus was determined by using the mean decreasing accuracy and the Gini coefficient. (**C, F**) Receiver-operating characteristic curves (ROC) were obtained using the most important genus in BALF and saliva samples.

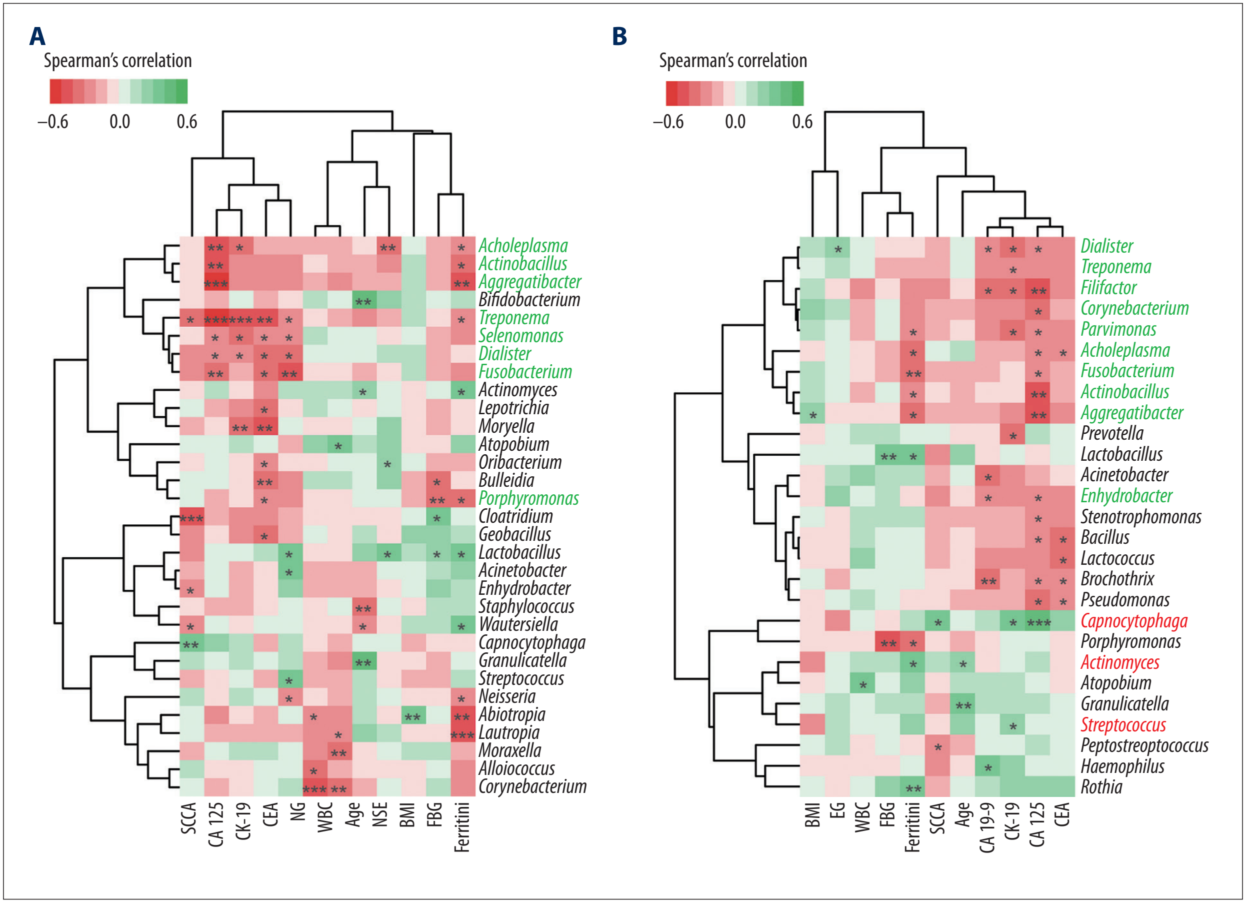


Figure 5. Correlations between microbiota genera and clinical indices. Heat map of Spearman's rank correlation between microbiota genera and clinical indices in bronchoalveolar lavage (BALF) (**A**) and saliva (**B**) samples. Green, positive correlation; red, negative correlation; * *P*<0.05, ** *P*<0.01, and *** *P*<0.001. The genera were colored according to the direction of enrichment. Green – enriched in healthy control; red – enriched in lung cancer; black – no significant difference.

Discussion

To some extent, the microbiota in the lower respiratory tract may be one of the causes of lung cancer etiology, even though lung cancer is generally considered to be a disease caused by the interactions of host genetics and environmental factors, accounting for nearly one-fifth of human malignancies [33]. Several studies have demonstrated that microbes and the microbiota may contribute to tumor occurrence, development and progression [12,13,18,20,21,33]. There are 3 approved categories in which microbiota contribute to carcinogenesis. First, microbiota in the local environment disrupted the equilibrium of host cell apoptosis and proliferation at the genetic and metabolic levels. Second, microbiota may impact the host immune surveillance system, and their secondary metabolites may affect the local inflammatory response. Finally, microbiota could influence the metabolic process of pharmaceuticals and host and environment factors [5,13]. In the current study, pyrosequencing of 16S rDNA was employed to evaluate and compare the structure, diversity and metabolic characteristics of saliva and lower respiratory tract microbiota associated with lung cancer and healthy controls in a Chinese population. To the best of our knowledge, this report is the first to use 16S rDNA approaches to profile and compare the microflora composition and function prediction in saliva and BALF samples and to determine microbiota characteristics, significant discrepancies and biomarkers for the early diagnosis of lung cancer.

As gender, age, obesity, and smoking are considered risk factors for lung cancer development in several reports [12,13], we randomly balanced the enrolled participants with the features aforementioned in both saliva and BALF samples in the cancer patients to exclude the influence of these factors on the 16S rDNA results. Here, we used high-throughput sequencing to investigate saliva and BALF microbiota in lung cancer patients and studied more than 1700 OTUs, including nearly 300 genera. The α-diversity analysis results showed that lung cancer patients had less lung and oral microbiota diversity than healthy controls. Microbes may find a tumor's oxygen tension or carbon sources permissive and take advantage of an underused

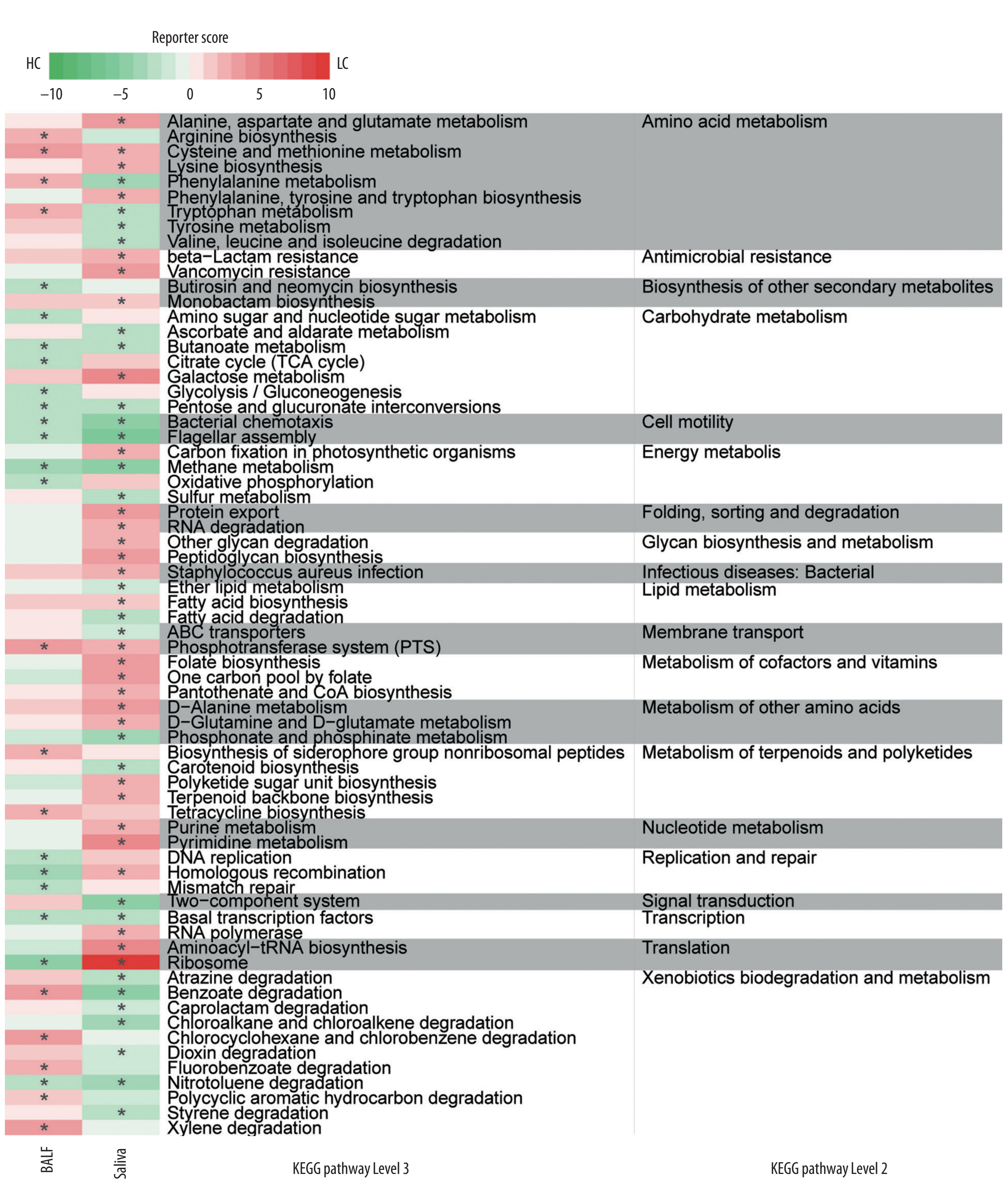


Figure 6. Microbiota function prediction in BALF and saliva samples. Heat map of KEGG pathways differentially enriched between healthy control and lung cancer groups in BALF and saliva. Green, enriched in healthy control; red, enriched in lung cancer. * Denotes a reporter score >1.96 or < −1.96. BALF – bronchoalveolar lavage; KEGG – Kyoto Encyclopedia of Genes and Genomes.

nutritional niche [4]. Decreased abundances of specific microbes may also increase the risk for cancer development in the host at sites that are local or distant from this microbial shift [13]. The detailed mechanism remains unknown. In healthy controls, no obvious differences in the upper and lower respiratory tract [12] were observed. In the current study, our PCoA results also demonstrated a similar flora composition in both the oral and lower respiratory tract, which expanded the current flora known to the digestive tract. PCoA results also consistently showed that distinct microbiota composition can be identified between and lung cancer and healthy control patients regardless of the sampling site, indicating that remarkable microbial taxa may be found and most likely serve as biomarkers or indicators for lung cancer patients.

Microbiota composition analysis at the phylum and the genus levels through taxonomic assignment were performed, and the results showed that representative flora differed by sampling site. *Veillonella* and *Capnocytophaga* were enriched in the BALF samples, while a significantly high abundance of *Streptococcus*, *Capnocytophaga*, and *Actinomyces* were detected in the saliva samples. We observed that *Capnocytophaga* were enriched in both oral and upper respiratory samples. *Capnocytophaga* species are commonly found in the oropharyngeal tract and the airway of mammals; they are involved in the pathogenesis of periodontal diseases as well as some animal bite wounds [34]. Additional specific studies focus on the expression of *Capnocytophaga* in a large sample of lung cancer patients. *Veillonella* was also found to be enriched in the BALF of lung cancer, and it was previously isolated from the lower airways of lung cancer patients, suggesting that *Veillonella* species may be related to lung cancer [35]. *Streptococcus* is causally linked to chronic lung disease pneumonia and was enriched in the saliva samples [21]. These findings, including ours, seem to suggest that either these bacteria induce a long-term immune response/infection to the organ, or the cancer growth environment favors the growth of these bacteria in the airway or oropharyngeal tracts. When stratifying by pathological type of lung cancer, a significant enrichment in *Veillonella* and *Capnocytophaga* was observed in the BALF samples of LSCC patients. Likewise, *Capnocytophaga* and *Actinomyces* were specifically validated to be enriched in the LSCC group, and an enrichment of *Streptococcus* was observed in the SCLC group, while *Rothia* was also observed to be significantly different in the LAC group. It seems that lung cancer-specific microbial taxa may be pathologically dependent. The microenvironment of the tumor is different from the pathological type of cancer cells, and the great variability of metabolic products and cytokine levels also affected the colonization and growth of bacteria, which contributed to the pathological-specific phenomenon.

Functional metabolic characteristics in bacterial communities vary from the inherent host environment and immunity status [12,13]. Exposure to xenobiotics, such as carcinogens, insecticides, and drugs, and their deposition in pulmonary tissues were considered as initial factors for lung cancer development [36]. In the present study, xenobiotic biodegradation was enriched in the BALF of lung cancer, including the degradation pathway of atrazine, benzoate, chlorocyclohexane, chlorobenzene, fluorobenzoate, nitrotoluene, polycyclic aromatic hydrocarbon and xylene. We can speculate that the enrichment of xenobiotic biodegradation is how microbiota alteration is used to cope with environmentally damaging factors. Although xenobiotic biodegradation is depleted in saliva samples of cancer patients, we cannot entirely deny our previous speculation that oral microbiota is susceptible to multiple influencing factors from the external environment. Further stratification analyses on large sample sizes of different pathological types of cancer patients are promising, and additional research is needed.

Conclusions

We examined the saliva and BALF microbiota in lung cancer patients using high-throughput technology and found that the microbial community changed in lung cancer patients and that the diversity might be site- and pathological-dependent. This study provides basic data on the oral and BALF flora related to lung cancer and provides a hint in the etiology of the disease. The results of our study indicated that lung cancer patients carry a different and less diverse microorganism community than healthy controls. Certain bacterial taxa might be associated with lung cancer, but the exact species depends on the sampling site and the pathological subtype. This study provides basic data on the microbiota diversity in saliva and BALF samples from lung cancer patients.

Acknowledgements

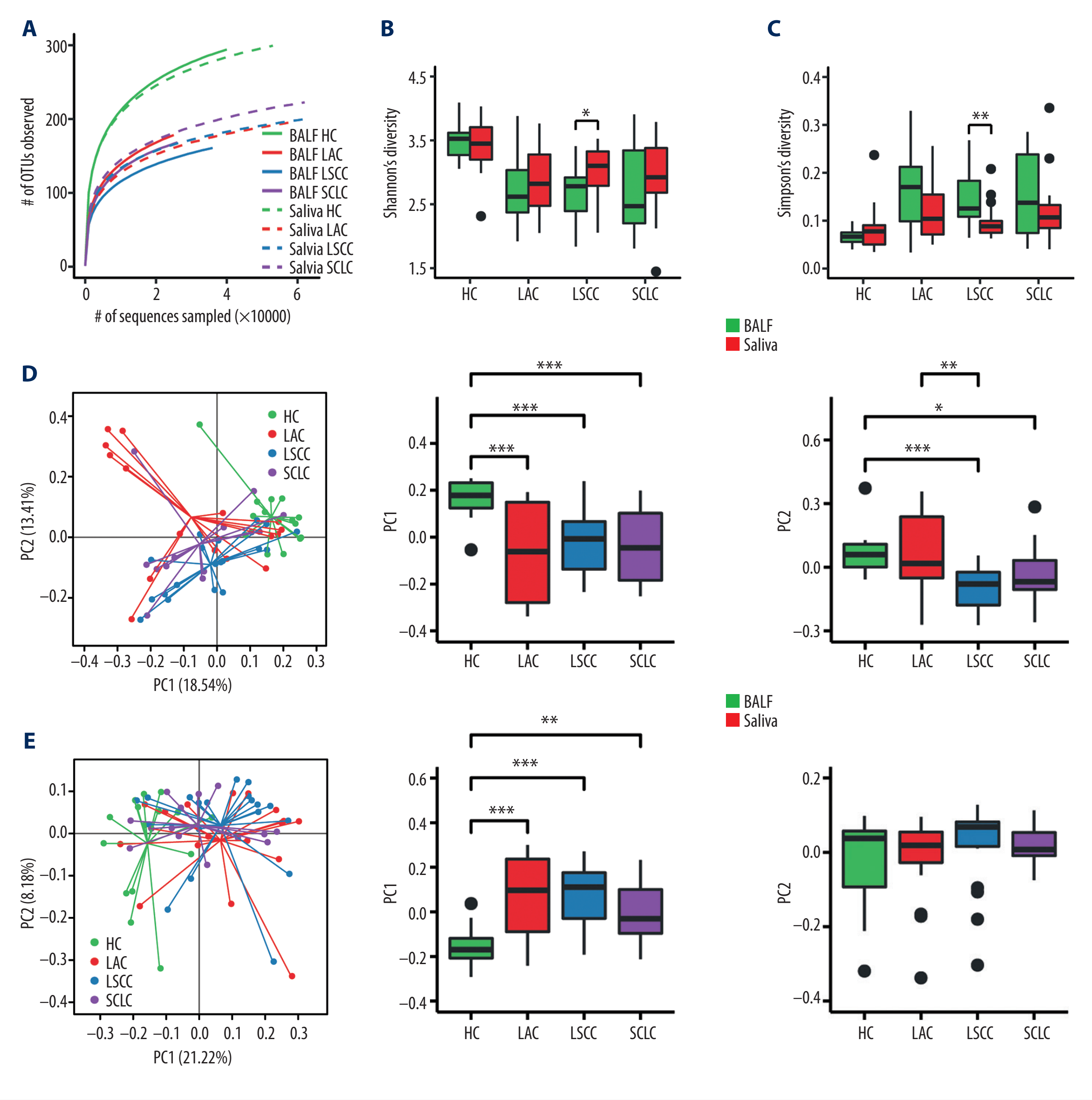
We gratefully acknowledge colleagues at BGI-Shenzhen for DNA quality control, library construction, sequencing and helpful discussions. We also gratefully acknowledge the patients who participated in this study and appreciate their cooperation.

Conflicts of interest

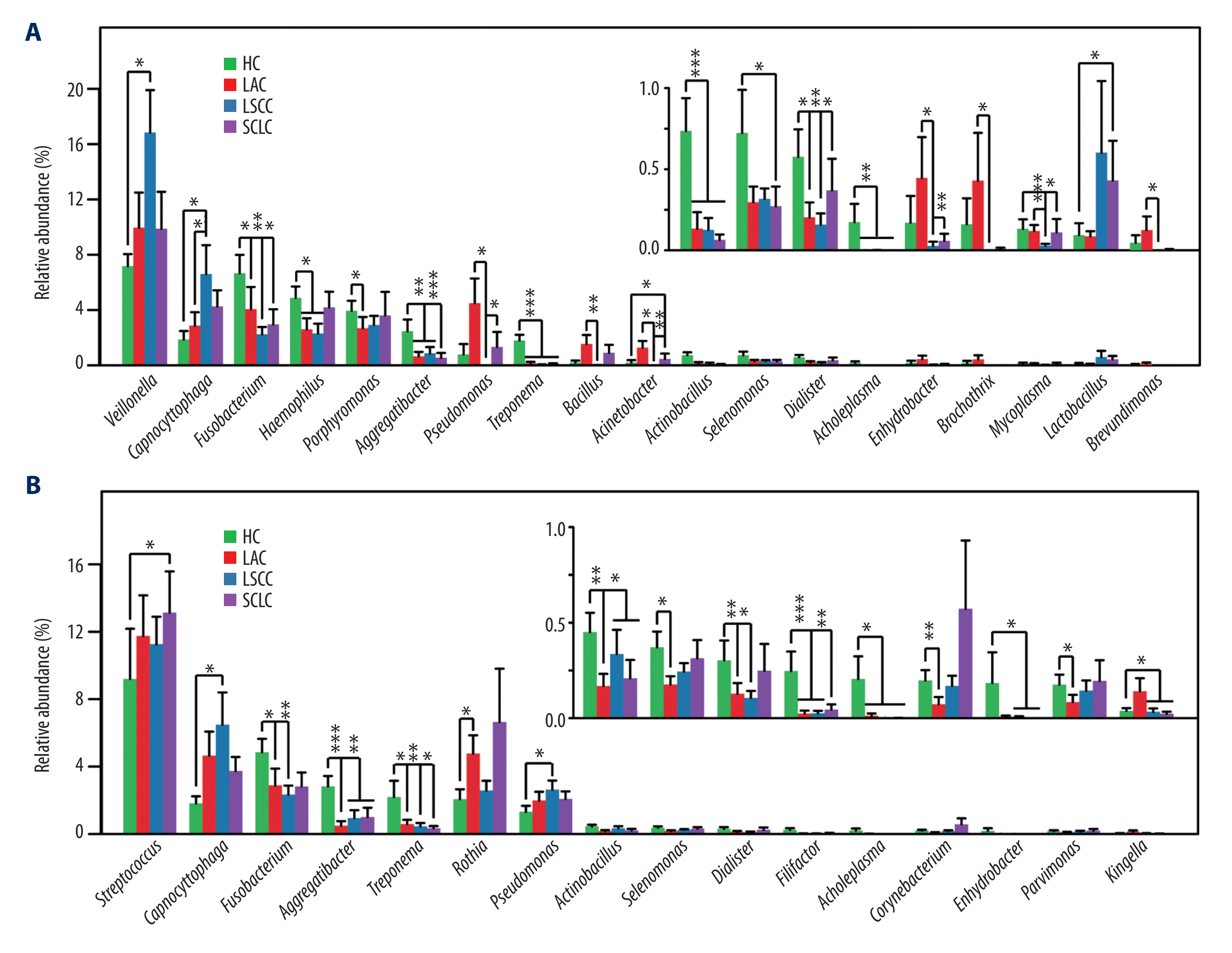
None.

Supplementary Files

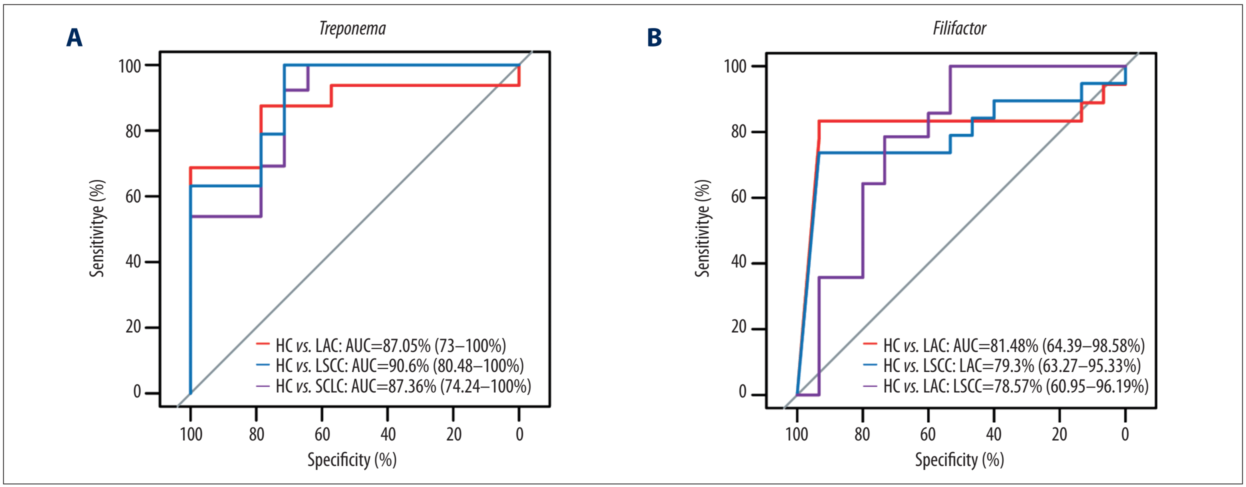
Supplemantary/raw Tables 1–4 available from the corresponding author on request.



Supplementary Figure 1. Comparison of the alpha diversity and beta diversity for microbiota from different sampling sites and different pathological subtype lung cancer groups. (**A**) The rarefaction curve for samples from different sites and different pathological subtype lung cancer groups was drawn to evaluate the depth and coverage of sequencing and the richness of species. (**B, C**) Shannon and Simpson indexes were estimated to evaluate the diversity of different pathological subtype lung cancer samples. (**D, E**) Principal coordinate analysis (PCoA) based on the unweighted UniFrac distance matrix. The green and red colors represent bronchoalveolar lavage (BALF) and saliva samples, respectively. * *P*<0.05, ** *P*<0.01, and *** *P*<0.001 by Wilcoxon rank-sum test.



Supplementary Figure 2. The species with significant differences between healthy control and different pathological subtype lung cancer groups in bronchoalveolar lavage (BALF) (**A**) and saliva (**B**) samples were displayed at the genus level. The species were shown only if its abundance was over 1% in at least 1 sample and with a significant difference between any 2 groups. Bar chart was expressed as the mean ± standard error of the mean (SEM). * *P*<0.05, ** *P*<0.01, and *** *P*<0.001 by Wilcoxon rank-sum test.



Supplementary Figure 3. (**A, B**) Receiver-operating characteristic curves (ROC) for the selected potential bacterial biomarker to distinguish the healthy control and different pathological subtype lung cancer groups.

References:

1. Huo X, Huo B, Wang H et al: Implantation of computed tomography-guided Iodine-125 seeds in combination with chemotherapy for the treatment of stage III non-small cell lung cancer. J Contemp Brachytherapy, 2017; 9: 527–34
2. Nakagiri T, Tokunaga T, Kunoh H et al: Surgical treatment following chemo-targeted therapy with bevacizumab for lung metastasis from colorectal carcinoma: analysis of safety and histological therapeutic effects in patients treated at a single institution. Case Rep Oncol, 2018; 11: 98–108
3. Stenehjem DD, Bellows BK, Yager KM et al: Cost-utility of a prognostic test guiding adjuvant chemotherapy decisions in early-stage non-small cell lung cancer. Oncologist, 2016; 21: 196–204
4. Niederreiter L, Adolph TE, Tilg H: Food, microbiome and colorectal cancer. Dig Liver Dis, 2018; 50: 647–52
5. Shang FM, Liu HL: *Fusobacterium nucleatum* and colorectal cancer: A review. World J Gastrointest Oncol, 2018; 10: 71–81
6. Abreu MT, Peek RM Jr.: Gastrointestinal malignancy and the microbiome. Gastroenterology, 2014; 146: 1534–46.e3
7. Ojesina AI, Lichtenstein L, Freeman SS et al: Landscape of genomic alterations in cervical carcinomas. Nature, 2014; 506: 371–75
8. Hogan DA, Willger SD, Dolben EL et al: Analysis of lung microbiota in bronchoalveolar lavage, protected brush and sputum samples from subjects with mild-to-moderate cystic fibrosis lung disease. PLoS One, 2016; 11: e0149998
9. Costerton JW, Stewart PS, Greenberg EP: Bacterial biofilms: A common cause of persistent infections. Science, 1999; 284: 1318–22
10. Govan JR, Nelson JW: Microbiology of lung infection in cystic fibrosis. Br Med Bull, 1992; 48: 912–30
11. Morris A, Beck JM, Schloss PD et al: Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care Med, 2013; 187: 1067–75
12. Beck JM, Young VB, Huffnagle GB: The microbiome of the lung. Trans Res, 2012; 160: 258–66
13. Garrett WS: Cancer and the microbiota. Science, 2015; 348: 80–86
14. Yu G, Gail MH, Shi J et al: Association between upper digestive tract microbiota and cancer-predisposing states in the esophagus and stomach. Cancer Epidemiol Biomarkers Prev, 2014; 23: 735–41
15. Hu X, Zhang Q, Hua H, Chen F: Changes in the salivary microbiota of oral leukoplakia and oral cancer. Oral Oncol, 2016; 56: e6–8
16. Patel T, Bhattacharya P, Das S: Gut microbiota: An indicator to gastrointestinal tract diseases. J Gastrointest Cancer, 2016; 47: 232–38
17. Hunt RH, Yaghoobi M: The esophageal and gastric microbiome in health and disease. Gastroenterol Clin North Am, 2017; 46: 121–41
18. Olson SH, Satagopan J, Xu Y et al: The oral microbiota in patients with pancreatic cancer, patients with IPMNs, and controls: A pilot study. Cancer Causes Control, 2017; 28: 959–69
19. Wang ZK, Yang YS: Upper gastrointestinal microbiota and digestive diseases. World J Gastroenterol, 2013; 19: 1541–50
20. Yan X, Yang M, Liu J et al: Discovery and validation of potential bacterial biomarkers for lung cancer. Am J Cancer Res, 2015; 5: 3111–22
21. Hosgood HD 3rd, Sapkota AR, Rothman N et al: The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. Environ Mol Mutagen, 2014; 55: 643–51
22. Magoc T, Salzberg SL: FLASH: Fast length adjustment of short reads to improve genome assemblies. Bioinformatics, 2011; 27: 2957–63
23. Edgar RC: UPARSE: Highly accurate OTU sequences from microbial amplicon reads. Nat Methods, 2013; 10: 996–98
24. Wang Q, Garrity GM, Tiedje JM, Cole JR: Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol, 2007; 73: 5261–67
25. DeSantis TZ, Hugenholtz P, Larsen N et al: Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol, 2006; 72: 5069–72
26. Schloss PD, Westcott SL, Ryabin T et al: Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol, 2009; 75: 7537–41
27. Caporaso JG, Kuczynski J, Stombaugh J et al: QIIME allows analysis of high-throughput community sequencing data. Nat Methods, 2010; 7: 335–36
28. Langille MG, Zaneveld J, Caporaso JG et al: Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol, 2013; 31: 814–21
29. Patil KR, Nielsen J: Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proc Natl Acad Sci USA, 2005; 102: 2685–89
30. Knights D, Kuczynski J, Charlson ES et al: Bayesian community-wide culture-independent microbial source tracking. Nat Methods, 2011; 8: 761–63
31. Oksanen J: Multivariate analysis of ecological communities in R: Vegan tutorial. R Package Version, 2011; 1: 11–12
32. Robin X, Turck N, Hainard A et al: pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 2011; 12: 77
33. de Martel C, Ferlay J, Franceschi S et al: Global burden of cancers attributable to infections in 2008: A review and synthetic analysis. Lancet Oncol, 2012; 13: 607–15
34. Jolivet-Gougeon A, Sixou J-L, Tamanai-Shacoori Z, Bonnaure-Mallet M: Antimicrobial treatment of *Capnocytophaga* infections. Int J Antimicrob Agents, 2007; 29: 367–73
35. Rybojad P, Los R, Sawicki M et al: Anaerobic bacteria colonizing the lower airways in lung cancer patients. Folia Histochem Cytobiol, 2011; 49: 263–66
36. Wu J, Peters BA, Dominianni C et al: Cigarette smoking and the oral microbiome in a large study of American adults. ISME J, 2016; 10: 2435–46