

# The Airway Microbiota of Non-Small-Cell Lung Cancer Patients and Its Relationship to Tumor Stage and EGFR Mutation

**DanHui Huang**

Southern Medical University Nanfang Hospital

**Jing He**

Southern Medical University Nanfang Hospital

**Xiaofang Su**

Southern Medical University Nanfang Hospital

**YaNa Wen**

Southern Medical University Nanfang Hospital

**ShuJia Zhang**

Southern Medical University Nanfang Hospital

**LaiYu Liu**

Southern Medical University Nanfang Hospital

**Haijin Zhao**

Southern Medical University Nanfang Hospital

**CuiPin Ye**

Southern Medical University Nanfang Hospital

**JianHua Wu**

Southern Medical University Nanfang Hospital

**Shaoxi Cai**

Southern Medical University Nanfang Hospital

**Hangming Dong** (✉ [dhm@smu.edu.cn](mailto:dhm@smu.edu.cn))

Southern Medical University <https://orcid.org/0000-0002-9573-0721>

---

## Research

**Keywords:** lung cancer, 16S rRNA sequencing, airway microbiota, metastasis, EGFR gene

**Posted Date:** November 10th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-1033150/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Accumulating studies have suggested the airway microbiota of lung cancer was significantly different from healthy controls. However, little was known about the relationship between airway microbiota and important clinical parameters of lung cancer. In this study, we aimed to explore the association between sputum microbiota and lung cancer stage, lymph node metastasis, intrathoracic metastasis, and Epidermal growth factor receptor (EGFR) gene mutation.

**Methods:** The microbiota of sputum samples from 85 newly diagnosed NSCLC patients were sequenced via 16S rRNA sequencing with V3-V4 region. Sequencing reads were filtered using QIIME2 and clustered against UPARSE.

**Results:** The  $\alpha$  diversity and  $\beta$  diversity was significantly different between patients in stage I to II (early stage, ES) and patients in stage III to IV (advanced stage, AS). Lefse identified that genera *Granulicatella* and *Actinobacillus* were significantly enriched in ES, and genus *Actinomyces* were significantly enriched in AS. PICRUST2 identified NAD salvage pathway was significantly enriched in AS, which was positively associated with *Granulicatella*. Patients with intrathoracic metastasis were associated with increased genus *Peptostreptococcus* and incomplete reductive TCA cycle, which was associated with increased *Peptostreptococcus*. Genera *Parvimonas*, *Pseudomona* and L-valine biosynthesis were positively associated with lymph node metastasis. L-valine biosynthesis was related with increased *Pseudomona*. Finally, genus *Parvimonas* were significantly upregulated in adenocarcinoma patients with EGFR mutation.

**Conclusion:** Taxonomy structure differed between different lung cancer stage. The tumor stage, intrathoracic metastasis, lymph node metastasis, and EGFR mutation were associated with alteration of specific airway genera and metabolic function of sputum microbiota.

## Background

Lung cancer is the second leading malignancy for morbidity and the first for cancer deaths worldwide[1]. Although with the development of target therapy and immunotherapy, the 5-year survival rate of lung cancer remain low, especially in metastatic disease[2]. Historically speaking, the lung has been considered sterile in health. However, with the advent of novel culture-independent techniques, subsequent studies identified that healthy lung was inhabited by distinct commensal microbiota, which was altered under multiple lung diseases [3]. Therefore, it is of great interest to explore the relationship between lung microbiome and lung cancer.

Accumulating studies have suggested that the airway microbiota of lung cancer patients was significantly different from healthy or benign control [4-11], suggesting that airway microbiota may contribute to the development of lung cancer or be affected during the progression of lung cancer. More specifically, the  $\alpha$  diversity[6, 8, 9, 11],  $\beta$  diversity[5, 6, 9-11], and some specific genera [4-11] were changed among non-small cell lung cancer (NSCLC). As of yet, the study of lung microbiota and lung

cancer remained in its infancy and deep knowledge of the interplay between lung cancer with different clinical parameters and lung microbiota needed to be further explored. TNM stage remains the most important prognostic factor in predicting recurrence rates and survival time. The 5-year-survival rate of lung cancer is significantly affected by tumor anatomic stages, from 87%-97% of stage I to 10%-23% of stage IV[12]. Through the analysis of 165 cases of normal tissue adjacent to lung cancer, an early study found that the  $\alpha$  diversity and genus *Thermus* was more abundant in late-stage (stage IIIB and IV) than that in early-stage[13](13)(13)[13], suggesting that lung microbiota participated in the development of different stages of lung cancer. However, the study only included 7 subjects in IIIB stage and 7 subjects in IV stage. Therefore, more studies regarding the association between tumor anatomic stage and lung microbiota should be conducted to find out more potential bacterial markers linked with the stepwise change of lung cancer from early-stage to late-stage.

Lung cancer staging traditionally relies on the TNM staging system. Since the stage of lung cancer was associated with lung microbiota, detailed understanding regarding the association between N, M classifications and lung microbiota should be explored. Previous studies suggested that specific genera might be engaged with the metastasis of lung cancer patients[13, 14]. In vivo mechanistic investigations found that certain species might contribute to the development of extrathoracic or intrathoracic metastasis via enhancement of adhesion of lung cancer cell or regulation of lung immune system[15-17]. Therefore, it is plausible to hypothesize that the lung microbiota may be identified as relevant to N and M classification.

Epidermal growth factor receptor (EGFR) is a paramount therapeutic target for the treatment of lung cancer. Tyrosine kinase inhibitors (TKIs) which target the kinase domain of EGFR are especially effective in NSCLC patients whose tumors harbor activating mutations in the tyrosine kinase domain of the EGFR gene. Bacterium that carried genotoxic markers could promote the accumulation of genetic lesions and initiated cancer development[18]. Current evidences suggested that some pathogens might play a role in driving EGFR gene mutation. A retrospective study found lung adenocarcinoma patients who had tuberculosis lesions had a higher probability of having EGFR gene mutations[19]. Another early study demonstrated an association between human papillomavirus and EGFR gene mutation in lung cancer patients[20]. Conversely, EGFR mutation might also regulate lung microbiome since it played a role in maintaining airway epithelial barrier via activation of Claudin 1, a member of tight junction protein[21]. However, the association between EGFR gene mutation and lung microbiota was unknown. Thus, it is plausible that lung microbiota may have a connection with EGFR gene mutation among NSCLC patients.

In this study, we used next-generation sequencing to identify airway microbiota in spontaneous sputum of NSCLC patients, aiming to characterize airway microbiota in NSCLC patients with different tumor stages (included tumor stage and TNM classification), and EGFR gene mutation.

## Materials And Methods

### 2.1 Patients and samples

The study was approved by the Ethics Committee of Nanfang Hospital, Southern Medical University. First diagnosed NSCLC patients were prospectively admitted in this study at NanFang Hospital, Southern Medical University between April 2017 and September 2019. The inclusion criteria were as follows: pathologically diagnosed of NSCLC; aged 30-80; did not receive any anti-tumor therapy such as surgery, radiotherapy, chemotherapy, targeted therapy or immunotherapy; no evidence of community-acquired pneumonia, acute exacerbation of chronic obstructive pulmonary disease, bronchiectasis with infection, acute bronchitis or asthma; had no fever or purulent or gray sputum; without a history of other malignant diseases or multiple primary lung cancer. We conducted a questionnaire and reviewed the electronic medical records to obtain demographic and clinical data including age, sex, smoking status, antibiotics usage, TNM stage, systemic or pulmonary comorbidities and tumor EGFR mutation. Tumor anatomic stage and TNM classification was based on NCCN clinical practice Guidelines of NSCLC (Version 2020. V1). The EGFR mutation was detected based on the ARMS technology in the pathology department of Nanfang Hospital.

Participants were asked to rinse their mouths before sampling. The first mouthful of phlegm in the morning was collected within 24 hours of hospitalization and transferred into -20°C refrigerators within 2 hours and then transferred into -80°C within 1 week.

## **2.2 DNA extraction, 16S rRNA amplification, 16S rRNA sequencing**

Sputum samples kept on dry ice were transferred to Sagene Biotechnology Company, GuangZhou. DNA was extracted from samples using Hipure Bacterial DNA kit (Mageon, China) using standard techniques. The V3-V4 region of 16S rRNA gene was amplified using specific primers(16S\_341F:5'-CCTAYGGGRBGCASCAG-3';16S\_806R:5-GGACTACNNGGGTATCTAAT). PrimeSTAR HS DNA Polymerase was used for PCR reaction. The concentration and length of the PCR products were detected by 1% agarose gel electrophoresis. Samples with a bright main strip were used for further experiments. Sequencing libraries were conducted using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® sequencing (New England Biolabs, United States). The quality of the library was evaluated under a Qubit® 2.0 Fluorometer (Thermo Scientific) and Agilent Bioanalyzer 2100 system. Sequencing was conducted to generate 250-bp paired-end reads using an Illumina HiSeq 2500 sequencer according to the manufacturer's instructions.

## **2.3 Microbiota analysis**

Raw data was obtained and then further filtered to eliminate reads with adapter pollution and low quality to obtain clean reads by using QIIME2[22]. Clean sequences were clustered by 97% identity into operational taxonomic units (OTUs) using UPARSE[23]. Representative sequence of each OTU was annotated into taxonomy against Greengenes database[24].

We applied OTUs data in online microbiome data analyze platform (MicrobiomeAnalyst) (<https://www.microbiomeanalyst.ca/>) to compare microbiota community structure at both inter-community and  $\alpha$ -diversity level and  $\beta$ -diversity level. For  $\alpha$  diversity, we chose Chao1 value, Simpson

index and Shannon index to evaluate. For  $\beta$  diversity, we estimated using Bray-Curtis distance and visualized by principal coordinate analysis (PCoA). Differential taxonomy was identified by LEfSe (Linear discriminant analysis (LDA) effect size) analysis in an online platform (GALAXY) (<http://huttenhower.sph.harvard.edu/galaxy>). PICRUSt2 was used to predict the functional profiling of microbial communities based on the 16S rRNA sequence[25]. Metabolic function predictions were based on MetaCyc [26] database. Differentially present pathways between groups were analyzed with Welch t test using STAMP[27]. The network analysis on the genus level was carried out with SparCC[28]. P value  $\leq 0.05$  and SparCC correlation scores  $\geq 0.5$  or  $\leq -0.5$  were included for networks inference.

## 2.4 Statistical analysis

The software SPSS (V 23.0) was used for statistical analysis. The continuous variables were compared between two groups by Mann-Whitney U test or independent t test. The categorical variables were compared by chi-square test, continuity-adjusted chi-square test, Fisher's exact test. P value < 0.05 was considered statistically significant.

# Results

## 3.1 Subjects clinical characteristics and sputum microbiota in NSCLC

Spontaneous sputum samples were collected from 116 NSCLC patients preliminarily. After carefully assessment, 85 patients who met eligible criteria were finally taken into further analysis. The procedure of patients' recruitment and exclusion was shown in Figure 1. The average number of trimmed sequences reads number of the 85 subjects was 33271 (7869, 44193). OTU rarefaction curve was constructed to evaluate sequence depth (Supplementary figure S1). The result indicated that sequence depth of sputum samples was sufficient enough to reach a reliable estimate of microbiome structure. The median age of all patients was  $59.21 \pm 8.75$  years. The clinical characteristics of the 85 patients were listed in supplementary table S1. Among the 85 patients, 66 (78%) were adenocarcinoma, 18 (21%) were squamous cell carcinoma, 1 was unidentified type of NSCLC. 13(15%) patients were in tumor stage I, 9 (11%) were in stage II, 11 (13%) were in stage III and 40 (47%) were in stage IV.

Phyla and genera that were  $\geq 1\%$  were considered as dominant. At the phylum level, the dominant phyla in the sputum samples of NSCLC were *Firmicutes* (40%), *Bacteroidetes* (20%), *Actinobacteria* (17%), *Proteobacteria* (13%), *Fusobacteria* (6%), and *TM7* (3%) (Supplementary figure S2 A). At the genus level, the dominant genera in the sputum samples of NSCLC were *Streptococcus* (21%), *Prevotella* (12%), *Rothia* (9%) and *Neisseria* (7%), *Actinomyces* (5%), *Leptotrichia* (4%), *Porphyromona* (4%), *Veillonella* (4%), *Granulicatella* (3%), *Haemophilus* (3%), *Atopobium* (2%), *Peptostreptococcus* (2%), *Capnocytophaga* (2%) and *Fusobacterium* (1%) (Supplementary figure S2 B).

## 3.2 The Association between sputum microbiota and NSCLC clinical stage

Stage III and stage IV lung cancer patients are on a continuum with respect to tumor burden. It is well accepted that a great number of lung cancer patients with anatomical stage III also harbor micro-metastases. A previous study found that the  $\alpha$  diversity of microbiota of non-malignant tissues adjacent to stage IIIB lung tumor tissues was similar with that of stage IV[13], suggesting that the airway microbiota may be similar between stage III and stage IV lung cancer patients. To evaluate the similarity of microbiota between stage III and stage IV patients, we compare the sputum microbiota between these 2 groups via diversity analysis and differential analysis. Among the 11 NSCLC patients in stage III, 4 patients were in stage IIIA, 5 patients were in stage IIIB, 2 patients were in stage IIIC. Baseline information included age (independent samples T test,  $P=0.21$ ), BMI index (Mann-Whitney U test,  $P=0.536$ ), smoking status (continuity-adjusted chi-square test,  $P=0.249$ ), antibiotics treatment before sampling (continuity-adjusted chi-square test,  $P=0.121$ ), pathological type (continuity-adjusted chi-square test,  $P=0.191$ ) was comparable between the groups. Chao1, Simpson index, and Shannon index were selected to estimate the  $\alpha$  diversity of the lung microbiome community.  $\alpha$  diversity between stage III and stage IV patients was similar (Mann-Whitney U test,  $P=0.519$  for Shannon;  $P=0.783$  for chao1;  $P=0.261$  for Simpson index) (Supplementary figure S3 A-C).  $\beta$  diversity based on Bray Curtis distance was used to estimate the  $\beta$  diversity of lung taxonomy community structure in different groups. The result showed that there was no significant difference in taxonomy structure between stage III and stage IV patients (PERMANOVA test,  $P=0.905$ ) (Supplementary figure S3 D). LEfSe analysis was conducted to identify whether differential taxonomy existed between stage III and stage IV patients. Only genus *Paludibacter* was found to be significantly different between the 2 groups (Supplementary figure S3 E). The relative abundance of *Paludibacter* was only 0.01% in stage III and 0.05% in stage IV patients. Taken together, the results above suggested that sputum microbiome of stage III and stage IV patients was similar.

Since the sputum microbiome of stage III and stage IV was similar, we divided the lung cancer patients into 2 groups: stage I and stage II (Early stage, ES) and Stage III and stage IV (Advanced stage, AS) and evaluate the microbiota difference between these 2 groups. Baseline information included demographic and clinical characteristic were comparable between AS and ES groups (supplementary table S2). The relative abundance of phylum level and genus level of ES and AS group were shown in Figure 2 A-B. For the  $\alpha$  diversity, there was a significantly difference in the Chao1 index between ES and AS group. Chao1 index was 221.529 (42.976) in ES group and 198.752398 (42.20770) in AS group (Mann-Whitney U test,  $P=0.038$ ) (Figure 3 A). Simpson index was 0.907 (0.064) in ES patients and 0.919 (0.050) in AS patients ( $P=0.705$ ) (Figure 3 B). Shannon index was 3.398 (0.541) in ES group and 3.419 (0.448) in AS group ( $P=0.815$ ) (Figure 3 C). For the  $\beta$  diversity, Bray Curtis distance based on genus level was performed. The result showed that there was significantly different taxonomy structure between patients in ES group and AS group (genus level, PERMANOVA test,  $P=0.045$ ) (Figure 3D).

Differentia analysis using Lefse identified that phylum *Firmicutes*, genera *Peptoniphilus*, *Granulicatella*, *Hylemonella*, *Actinobacillus*, *SMB53* and *Gemella* were significantly enriched in ES group, and phylum *Actinobacteria*, genus *Actinomyces* were significantly enriched in AS group (Figure 4A). The relative abundance of phyla *Firmicutes* and *Actinobacteria*, genera *Granulicatella*, *Actinomyces* and *Actinobacillus* were  $\geq 0.1\%$  and were shown in Figure 4 B, C.

Functional analysis based on metaCyc database identified 29 differentially abundant pathways (Figure 4 D). The largest 3 pathways which had higher proportion in ES patients were anhydromuropeptides recycling, gondoate biosynthesis (anaerobic) and L-lysine biosynthesis II. For the differential abundant pathways had higher relative abundance in AS group, incomplete reductive tricarboxylic acid (TCA) cycle, NAD salvage pathway I and phosphopantothenate biosynthesis I were the top 3 differentially abundant pathways. Genus *Actinomyces* was positively correlated with the NAD salvage pathway (Spearman rank correlation, P value < 0.0001, r=0.547).

Co-abundance analysis based on SparCC was conducted. The sputum microbiota structure of ES lung cancer patients was more complex and better organized than the taxonomy structure inferred for patients in AS group (Figure 5 A, B). The taxonomy structure of ES group was composed of 33 genera while the structure inferred for AS group was composed of 19 genera. The number of inter-genus correlations in ES group was 78, while only 44 in AS group. The interactions between genus *Streptococcus* and other genera (*Porphyromonas*, *Prevotella*, *Capnocytophaga*, *Veillonella*, *Atopobium*, *Actinomyces*, *Rothia*, *Granulicatella*) were exclusively co-occurrence in the AS group. Co-occurrence between *Actinomyces* and genera *Rothia* and *Atopobium* was ubiquitous among 2 groups, while co-occurrence between *Actinomyces* and genera *Granulicatella*, *Veillonella*, *Prevotella* and *Streptococcus* were exclusive in AS group.

### 3.3 The role of sputum microbiota on NSCLC intrathoracic metastasis and lymph node metastasis

Since tumor stage is associated with the organism metastasis and lymph node metastasis, a further analysis was conducted to explore the linkage between sputum microbiota and these clinical parameters.

Previous mouse studies suggested that the homeostasis of commensal lung microbiota may affect intrathoracic metastasis[17] and extrathoracic metastasis[15] and these 2 phenomena may depend on different mechanisms. Thus, we hypothesized that airway microbiota associated with intrathoracic (ipsilateral or contralateral lung metastasis or pleural metastasis) and extrathoracic metastasis was different. Among the 85 NSCLC patients, 15 were with intrathoracic metastasis and without extrathoracic metastasis (Intra group), only 3 patients were with extrathoracic metastasis and without intrathoracic metastasis, 28 were without neither intrathoracic nor extrathoracic metastasis (Non\_M group). We further explore the characterization of sputum microbiota among Intra and Non\_M patients. Baseline information was comparable between Intra and Non\_M group (Supplementary table S3).  $\alpha$  diversity index between these 2 groups was similar (P=0.6192 for chao1; P=0.1668 for Simpson index; P=0.2193 for Shannon) (Supplementary Figure S4 A-C).  $\beta$  diversity based on Bray Curtis distance was used and the result showed that taxonomy structure between Intra group and Non\_M group was similar (PERMANOVA test, P=0.197) (Supplementary Figure S4 D).

LEfse analysis showed that compared with Non-M patients, genera *Peptostreptococcus*, *Peptococcus*, *Parabacteroides*, and *Escherichia* were significantly enriched in Intra patients, while phylum *Firmicutes* and genus *Streptococcus* were significantly decreased (Supplementary Figure S5 A). The relative

abundance of phylum *Firmicutes* and genera *Peptostreptococcus* and *Streptococcus* were  $\geq 0.1\%$  and were listed in Supplementary Figure S5 B, C.

Functional analysis based on metaCyc database identified 31 differentially abundant pathways (Supplementary Figure S5 D). The largest 3 pathways which had higher proportion in Intra patients were incomplete reductive TCA cycle, tetrapyrrole biosynthesis II (from glycine), tetrapyrrole biosynthesis I (from glutamate). For the differential abundant pathways had higher relative abundance in Non\_M group, purine ribonucleosides degradation, lactose and galactose degradation I, L-lysine biosynthesis II were the top 3 differentially abundant pathways. Genus *Peptostreptococcus* was positively correlated with the incomplete reductive TCA cycle (Spearman rank correlation, P value=0.017,  $r=0.5779$ ).

Next, we explored the association between sputum microbiota and lymph node metastasis. Among 28 patients in M0 stage, 12 were in N1-3 stage (LNM\_Y) and 16 were in N0 stage (LNM\_N). Baseline information was comparable between LNM\_Y and LNM\_N group (Supplementary table S4).  $\alpha$  diversity analysis indicated that Chao 1 (P=0.0593), Simpson (P=1.000), and Shannon (P=0.9818) index were similar between the 2 groups (Supplementary Figure S6 A-C).  $\beta$  diversity analysis based on Bray Curtis distance showed that there was no significant difference in the bacterial community between the 2 groups (PERMANOVA test, P=0.091) (Supplementary Figure S6 D).

Compared with LNM\_N, genera *Parvimonas* and *Pseudomonas* were significantly increased in LNM\_Y, while phylum *Proteobacteria* and genera *Neisseria*, *Actinobacillus*, *Eikenella* were significantly declined in LNM\_Y (Supplementary Figure S7 A). All the above-mentioned differential taxonomy except for genus *Eikenella* was  $\geq 0.1\%$ . The relative abundance of each differential genus and phylum were listed in Supplementary Figure S7 B, C

Functional profile prediction based on Metacyc database identified 23 differential metabolic pathways (Supplementary Figure S7 D). L-valine biosynthesis, L-isoleucine biosynthesis I (from threonine), L-isoleucine biosynthesis II were the top 3 differential pathways that were more abundant in LNM\_Y group. Anhydromuropeptides recycling, 8-amino-7-oxononanoate biosynthesis I, biotin biosynthesis I and ppGpp biosynthesis were the top 3 differential pathways that were more enriched in LNM\_N group. Genus *Pseudomonas* was associated with L-valine biosynthesis (Spearman rank correlation, P value= 0.012,  $r=0.468$ ).

### **3.4 The Association between sputum microbiota and NSCLC EGFR gene mutation**

Among the 65 lung adenocarcinoma patients, 44 patients with EGFR mutation testing were available in subgroup analysis. Finally, 21 were with EGFR mutation-positive (EGFR+), 23 were with EGFR mutation-negative (EGFR-). Patients with EGFR mutation were more likely to be never smoker (Fisher exact test, P=0.036) and female (Fisher exact test, P=0.031). Other baseline information included age, BMI, tumor stage, antibiotics usage was comparable between 2 groups (Supplementary table S5).

$\alpha$  diversity between EGFR+ and EGFR- was similar ( $P=0.1054$  for chao1;  $P=0.1532$  for Simpson index;  $P=0.0820$  for Shannon) (Supplementary Figure S8 A-C).  $\beta$  diversity based on Bray Curtis distance was conducted to estimate the bacterial community composition in different groups. The result showed that there was no association between EGFR mutation and airway taxonomy structure (PERMANOVA test,  $P=0.212$ ) (Supplementary Figure S8 D).

LEfse analysis identified that EGFR mutation was associated with significantly enriched level of phyla *Bacteroidetes* and *Tenericutes*, genera *Sharpea*, *Prevotella*, *Porphyromonas*, *Parvimonas*, *Desulfovibrio*, *Mycoplasma*, *Actinobacillus*, *Dialister*, and *Eikenella* (Figure 6 A). Subgroup analysis limited to non-smoker subjects was conducted. The result showed similarly that phylum *Bacteroidetes* and genera *Parvimonas* and *Actinobacillus* were associated with EGFR mutation (Figure 6 B). The relative abundance of both genera *Parvimonas* and *Actinobacillus* and phylum *Bacteroidetes* were  $\geq 0.1\%$  and were shown in Figure 6 C, D

PICRUSt2 based on Metacyc prediction identified that superpathway of L-aspartate and L-asparagine biosynthesis, preQ0 biosynthesis and queuosine biosynthesis were the most 3 significantly abundant pathways in EGFR mutation non-smoking group and L-isoleucine biosynthesis II, L-isoleucine biosynthesis II III and superpathway of branched amino acid biosynthesis were the top 3 pathways that were significantly enriched in EGFR negative non-smoking group (Figure 6 E).

## Discussion

Growing pieces of evidence suggested that the development of cancer is affected by human commensal microbiota through inflammation, immunity, metabolism pathways[29]. Recently, various studies identified the alteration of airway microbiota among NSCLC patients[5, 7, 9, 11, 30-34]. The interplay between microbiota and lung cancer is complex. However, only few studies focused on the association between airway microbiota and tumor clinical parameters, includes tumor anatomic stage, metastasis, and gene mutation. In this study, we reported the characterization of sputum microbiota among NSCLC patients with early stage (stage I and stage II) and advanced stage (stage III and stage IV). More deeply, we explored the association between sputum microbiota and tumor N stage and intrathoracic metastasis. Besides, we investigated the linkage between EGFR mutation of lung adenocarcinoma and sputum microbiota.

Using 16S rRNA sequencing to profile the sputum microbiota in NSCLC patients, we found that the most abundant phylum and genus in NSCLC sputum samples were *Firmicutes* (40%) and *Streptococcus* (21%), which was consistent with the previous 2 studies analyzed sputum microbiota in lung cancer patients [11, 33]. TNM stage is the most predominant factor in predicting NSCLC survival time(35). The stepwise development of NSCLC from early-stage to late-stage was the results of various genetic and epigenetic alterations[36, 37], which may be associated with alteration of airway microbiota. Lung cancer staging system is categorical, however stage III and stage IV lie on a continuum with respect to tumor burden[38]. A great proportion of stage III patients had occult metastasis. The difference of stage III and stage IV lung

cancer patients lie on the tumor burden of distant sites, instead of the tumor burden of local regional sites[38]. Among the 11 III stage NSCLC patients enrolled in this study, 7 (63%) patients were in stage IIIB or IIIC. We found that the  $\alpha$  diversity and  $\beta$  diversity between stage III and stage IV patients were not significantly different, suggesting that the sputum microbiota might not sensitively reflect the tumor burden of distant site. Similarly, Yu et al collected adjacent tumor tissues from lung cancer patients and found that the  $\alpha$  diversity among NSCLC patients in IIIB and IV stage was similar[13]. However, considering that III stage NSCLC is a heterogeneous disease, the difference of sputum or lung tissue microbiota between stage III and stage IV lung cancer should be interpreted in a larger scale study in the future.

Compared with ES stage patients, we found a significant reduction of  $\alpha$  diversity in AS patients. The significant decrease of  $\alpha$  diversity in lung cancer patients compared with healthy or non-malignant control was evident in several studies, among which 2 studies used sputum samples[39, 40], 1 study used protected brush samples[41] and 1 study used surgical lung tissues[7]. Taken above, the results suggested that the reduction of  $\alpha$  diversity might be a potential marked indicated the development and progression of lung cancer.  $\beta$  diversity between ES and AS lung cancer patients were significant different in our study, indicating that the taxonomy community structure differed during the progression of lung cancer. The results of genus network analysis also supported the difference of taxonomy community structure. The SparCC results indicated that the sputum microbiota structure of ES lung cancer patients was more complex and better organized than the taxonomy structure inferred for AS patients.

We reported differential abundant taxonomy among NSCLC patients in AS stage and ES stage. More precisely, phylum *Firmicutes*, genera *Granulicatella*, *Actinobacillus* were significantly enriched in ES group, and phylum *Actinobacteria*, genus *Actinomyces* were significantly enriched in AS group. *Granulicatella* has been previously identified as a member of the normal bacterial flora of the respiratory tract[42] and was implicated in clinical infection such as sinusitis[43]. A study enrolled female lung cancer patients in China and found significantly enriched genus *Granulicatella* in sputum samples of lung cancer patients compared with healthy control[11]. Another pilot study using metagenomic sequencing technology identified *Granulicatella adicens*, a species belongs to genus *Granulicatella*, in sputum of lung cancer patients compared with benign diseases[4]. Taken together, our result and the previous studied mentioned above suggested that genus *Granulicatella* might played a role in the early development of NSCLC. *Actinobacillus* was a common member of human oral commensal microbiota. Previous studies found that *Actinobacillus* might influence the production of inflammatory cytokines[16] and was associated with COPD [44]. COPD is a widely recognized risk factor of lung cancer. Chronic inflammation is a key feature of COPD and could be a potential driver for lung cancer development[45]. Thus, genus *Actinobacillus* might serve as a linkage between COPD and lung cancer. It is plausible that the inhabitation of *Actinobacillus* lead to a chronic inflammation of the lung and enhance the initiation and early development of lung cancer. An early study identified genus *Actinomyces* was a common anaerobe colonizing in the airway of lung cancer patients[46]. It is interesting to note that in our study the co-occurrence of *Actinomyces* and genus *Veillonella* exclusively existed in AS group. Thus, in AS lung cancer patients, the increase of genus *Actinomyces* could increase the abundance of genus *Veillonella*. A

previous study found that the lower airway of lung cancer patients was enriched for genus *Veillonella*, which was further found to be associated with upregulation of ERK and PI3K signaling pathways[47]. It was recognized that PI3K and ERK pathways activation was involved in lung cancer metastasis[48]. Besides, we found genus *Actinomyces* was positively related with NAD salvage pathway, which was significantly enriched in AS patients. Cancer cells have enhanced glycolysis for sustaining rapid proliferation. Increased NAD levels enhance glycolysis and fuel cancer cells and is associated with cancer cell survival and enhanced invasion capacity [49, 50]. In fact, rate-limiting enzyme, such as nicotinamide phosphoribosyltransferase, was frequently amplified in several cancer cells[51]. Thus, in addition to its possible indirect influence on cancer related signaling pathway, genus *Actinomyces* might enhance lung cancer progression partly via enhanced NAD production.

Lung microbiota was reported to have influence on proliferation or metastasis of intrathoracic cancer via regulation of immune system[17, 52]. In this study, we reported intrathoracic metastasis was associated with enriched sputum genus *peptostreptococcus* and decreased *Streptococcus*. *Peptostreptococcus* was associated with colon cancer progression[53, 54]. However, its relationship with lung cancer remained largely unknown. We noticed that genus *Peptostreptococcus* are obligate anaerobes. It has been suggested that tumor microenvironment condition such as hypoxia may enhance tumor invasion and metastasis[55]. Therefore, it was plausible to speculate that the anoxic lung tumor condition, which can facilitate intrathoracic metastasis, may favor the growth of some obligate anaerobe, such as genus *Peptostreptococcus*. It is of interest to notice that incomplete reductive TCA cycle of sputum microbiota was significantly enriched in Intra group and was positively related with genus *Peptostreptococcus*. Reductive TCA cycle existed in anaerobe, including some deeply rooted bacteria, is one alternative strategy for fixing CO<sub>2</sub>[56]. During this reaction, oxaloacetate is finally produced[57] and may participate in TCA cycle in cancer cell. Current evidences demonstrated that certain cancer cells, including lung cancer with specific genome subtype[58, 59], rely heavily on the TCA cycle for energy production [60]. A recent study reported that enhanced TCA cycle might promote lung metastasis of certain cancer[61].

In the absence of distant metastasis, the existence of lung cancer spread to a regional lymph node affects clinical treatment options and prognosis. In this study, we found that the  $\alpha$  diversity and  $\beta$  diversity were similar between LNM\_Y and LNM\_N, which indicated that the sputum taxonomy structure did not vary during the progression of lymph node metastasis. LEfse analysis revealed genera *Parvimonas*, *Pseudomonas* was positively correlated with lymph node metastasis, while genera *Neisseria* and *Actinobacillus* was associated with depression of lymph node metastasis. Genus *Pseudomonas* showed a correlation with adenocarcinoma[62]. A clinical study identified that genus *Pseudomonas* was positively associated with matrix metalloproteinase in transplant lung patients[63], which was associated with metastasis and invasiveness of cancer cell[64]. Genus *Neisseria* was found to be negatively associated with lymph node metastasis. A previous study discovered that compared with healthy control, the relative abundance of salivary *Neisseria* was significantly decreased among lung cancer patients, which suggested that it might serve as a protective role in lung cancer progression[65]. Metabolic function prediction identified L-valine biosynthesis and L-isoleucine were increased in sputum microbiota

of LNM\_Y patients. Valine and isoleucine belong to branched chain amino acids, which play critical role in the regulation of energy homeostasis, nutrition metabolism, immunity and disease in humans[66]. They can act as signaling molecules regulating metabolism of glucose, lipid, and protein synthesis and serve as potential biomarkers in cancer[66](66)(66)[66]. Since genus *Pseudomonas* was positively associated with L-valine biosynthesis, it was plausible that *Pseudomonas* might apply valine for lung cancer cell and enhance its proliferation and invasiveness.

EGFR mutation was a strong prognostic factor among lung adenocarcinoma patients. The present data here showed that certain sputum bacterium had a close link with EGFR mutation among lung adenocarcinoma. Both in the overall analysis and subgroup analysis limited to non-smoker subjects, the results showed that the relative abundance of phylum *Bacteroidetes* and genera *Parvimonas* and *Actinobacillus* were positively associated with EGFR mutation. The increased EGFR signaling pathway was identified as relevant to airway mucin production, epithelial cell repairment [67], thus may have an influence on the abundance of phylum *Bacteroidetes*, genus *Parvimonas* and genus *Actinobacillus*. On the other hand, other evidences suggested that some specific bacterium such as genus *Parvimonas* may cause the EGFR mutation. Currently, several evidences suggested that *Parvimonas micra*, a member of genus *Parvimonas*, was enriched in patients with colon cancer[68, 69]. Interestingly, in vitro study demonstrated that infection of *Parvimonas micra* could enhance the ability of human inflammatory cells to generate reactive oxygen species and caused DNA damage of human cells[70], which could cause oncogene mutation and carcinogenesis.

Our study provided novel insight into the association between sputum microbiota, its predicted metabolic function and lung cancer stage, intrathoracic metastasis, lymph node metastasis and EGFR mutation. However, there are some limitations in our study. Firstly, the number of patients enrolled in this study is not large enough, so there may be heterogeneity. Secondly, the use of sputum can not surrogate lung cancer tissue. It should be caution to interpret intratumor microbiota using our results. Thirdly, the discovery of specific bacterial genera to distinguish lung cancer with various important clinical parameters hypothesis lacked validation cohorts, which may result in the false positive value and unreliability. Fourthly, the study is a cross-sectional study and only illustrates the phenomenon from microbiology. The mechanism of the microbiota and the causal relationship needed further exploration.

## Conclusions

Collectively, the present data showed association between important clinical parameters of lung cancer and airway microbiota. The taxonomy structure differed between patients in early stage and advanced stage. The tumor stage, intrathoracic metastasis, lymph node metastasis, and EGFR mutation were associated with alteration of specific airway genera and predicted metabolic function of sputum microbiota. Our study shed light that airway microbiota might participate in various pathophysiological processes that were importantly related to lung cancer development. Further studies with large scale and multi-omics are needed to achieve a better understanding of the role of microbiota in the development

and progression of lung cancer could pave a new way for exploring new therapeutic options and biomarkers.

## Abbreviations

NSCLC: non-small-cell lung cancer

EGFR: Epidermal growth factor receptor

ES: early stage

AS: advanced stage

TKIs: Tyrosine kinase inhibitors

OTUs: operational taxonomic units

TCA: reductive tricarboxylic acid

Intra: lung cancer patients with intrathoracic metastasis but without extrathoracic metastasis

Non-M: lung cancer patients without neither intrathoracic nor extrathoracic metastasis

LN<sub>M</sub>\_Y: M0 lung cancer patients with lymph node metastasis

LN<sub>M</sub>\_N: M0 lung cancer patients without lymph node metastasis

EGFR+ lung cancer with EGFR mutation

EGFR-: lung cancer without EGFR mutation

## Declarations

### Data availability statement

Public database with 16S rRNA sequencing data could be obtained online at the Sequence Read Archive (SRA). The BioProject number is PRJNA741774.

### Ethics approve and consent to participate

The study was approved by the Ethics Committee of Nanfang Hospital, Southern Medical University and all analysis were performed after obtaining individual written informed consent.

### Consent for publication

Obtained.

## Declaration of Competing Interest

There are no interests to declare.

## Funding

This work was supported by National Natural Science foundation of China (No. 81970032; 81670026; 8187011256).

## Acknowledgements

Not applicable.

## Contribution

(I) Conceptualization: Hangming Dong, Shaoxi Cai, DanHui Huang. (II) Formal analysis: DanHui Huang, Jing He, XiaoFang Su; (III) Data curation: DanHui Huang, YaNa Wen, ShuJia Zhang. (IV) Project administration: Jing He, XiaoFang Su, YaNa Wen, ShuJia Zhang, LaiYu Liu, Haijin Zhao, CuiPin Ye, JianHua Wu. (V)Funding acquisition: Shaoxi Cai, Hangming Dong; (VI) Writing-original draft: DanHui Huang, Jing He. (VII) Writing-review and editing: Shaoxi Cai, Hangming Dong.

## References

1. Siegel RL, Miller KD, Jemal A: **Cancer statistics, 2020**. *CA Cancer J Clin* 2020, **70**(1):7-30.
2. Herbst RS, Morgensztern D, Boshoff C: **The biology and management of non-small cell lung cancer**. *NATURE* 2018, **553**(7689):446-454.
3. Dickson RP, Erb-Downward JR, Huffnagle GB: **The role of the bacterial microbiome in lung disease**. *Expert Rev Respir Med* 2013, **7**(3):245-257.
4. Cameron S, Lewis KE, Huws SA, Hegarty MJ, Lewis PD, Pachebat JA, Mur L: **A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer**. *PLOS ONE* 2017, **12**(5):e177062.
5. Tsay JJ, Wu BG, Badri MH, Clemente JC, Shen N, Meyn P, Li Y, Yie TA, Lhaxhang T, Olsen E *et al*: **Airway Microbiota Is Associated with Upregulation of the PI3K Pathway in Lung Cancer**. *Am J Respir Crit Care Med* 2018, **198**(9):1188-1198.
6. Lee SH, Sung JY, Yong D, Chun J, Kim SY, Song JH, Chung KS, Kim EY, Jung JY, Kang YA *et al*: **Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions**. *LUNG CANCER* 2016, **102**:89-95.
7. Cheng C, Wang Z, Wang J, Ding C, Sun C, Liu P, Xu X, Liu Y, Chen B, Gu B: **Characterization of the lung microbiome and exploration of potential bacterial biomarkers for lung cancer**. *Transl Lung Cancer Res*

8. Liu HX, Tao LL, Zhang J, Zhu YG, Zheng Y, Liu D, Zhou M, Ke H, Shi MM, Qu JM: **Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects.** *INT J CANCER* 2018, **142**(4):769-778.
9. Greathouse KL, White JR, Vargas AJ, Bliskovsky VV, Beck JA, von Muhlinen N, Polley EC, Bowman ED, Khan MA, Robles AI *et al*: **Interaction between the microbiome and TP53 in human lung cancer.** *GENOME BIOL* 2018, **19**(1):123.
10. Apopa PL, Alley L, Penney RB, Arnaoutakis K, Steliga MA, Jeffus S, Bircan E, Gopalan B, Jin J, Patumcharoenpol P *et al*: **PARP1 Is Up-Regulated in Non-small Cell Lung Cancer Tissues in the Presence of the Cyanobacterial Toxin Microcystin.** *FRONT MICROBIOL* 2018, **9**:1757.
11. Hosgood HR, Sapkota AR, Rothman N, Rohan T, Hu W, Xu J, Vermeulen R, He X, White JR, Wu G *et al*: **The potential role of lung microbiota in lung cancer attributed to household coal burning exposures.** *ENVIRON MOL MUTAGEN* 2014, **55**(8):643-651.
12. Rusch VW, Chansky K, Kindler HL, Nowak AK, Pass HI, Rice DC, Shemanski L, Galateau-Salle F, McCaughan BC, Nakano T *et al*: **The IASLC Mesothelioma Staging Project: Proposals for the M Descriptors and for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Mesothelioma.** *J THORAC ONCOL* 2016, **11**(12):2112-2119.
13. Yu G, Gail MH, Consonni D, Carugno M, Humphrys M, Pesatori AC, Caporaso NE, Goedert JJ, Ravel J, Landi MT: **Characterizing human lung tissue microbiota and its relationship to epidemiological and clinical features.** *GENOME BIOL* 2016, **17**(1):163.
14. Huang D, Su X, Yuan M, Zhang S, He J, Deng Q, Qiu W, Dong H, Cai S: **The characterization of lung microbiome in lung cancer patients with different clinicopathology.** *AM J CANCER RES* 2019, **9**(9):2047-2063.
15. Gowing SD, Chow SC, Cools-Lartigue JJ, Chen CB, Najmeh S, Goodwin-Wilson M, Jiang HY, Bourdeau F, Beauchamp A, Angers I *et al*: **Gram-Negative Pneumonia Augments Non-Small Cell Lung Cancer Metastasis through Host Toll-like Receptor 4 Activation.** *J THORAC ONCOL* 2019, **14**(12):2097-2108.
16. Hessel CC, Andersson B, Wold AE: **Gram-positive and Gram-negative bacteria elicit different patterns of pro-inflammatory cytokines in human monocytes.** *CYTOKINE* 2005, **30**(6):311-318.
17. Le Noci V, Guglielmetti S, Arioli S, Camisaschi C, Bianchi F, Sommariva M, Storti C, Triulzi T, Castelli C, Balsari A *et al*: **Modulation of Pulmonary Microbiota by Antibiotic or Probiotic Aerosol Therapy: A Strategy to Promote Immunosurveillance against Lung Metastases.** *CELL REP* 2018, **24**(13):3528-3538.
18. Chen J, Pitmon E, Wang K: **Microbiome, inflammation and colorectal cancer.** *SEMIN IMMUNOL* 2017, **32**:43-53.

19. Luo YH, Wu CH, Wu WS, Huang CY, Su WJ, Tsai CM, Lee YC, Perng RP, Chen YM: **Association between tumor epidermal growth factor receptor mutation and pulmonary tuberculosis in patients with adenocarcinoma of the lungs.** *J THORAC ONCOL* 2012, **7**(2):299-305.
20. Li M, Deng F, Qian LT, Meng SP, Zhang Y, Shan WL, Zhang XL, Wang BL: **Association between human papillomavirus and EGFR mutations in advanced lung adenocarcinoma.** *ONCOL LETT* 2016, **12**(3):1953-1958.
21. Jia Z, Bao K, Wei P, Yu X, Zhang Y, Wang X, Wang X, Yao L, Li L, Wu P *et al*: **EGFR activation-induced decreases in claudin1 promote MUC5AC expression and exacerbate asthma in mice.** *MUCOSAL IMMUNOL* 2021, **14**(1):125-134.
22. Hall M, Beiko RG: **16S rRNA Gene Analysis with QIIME2.** *Methods Mol Biol* 2018, **1849**:113-129.
23. Edgar RC: **UPARSE: highly accurate OTU sequences from microbial amplicon reads.** *NAT METHODS* 2013, **10**(10):996-998.
24. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL: **Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB.** *Appl Environ Microbiol* 2006, **72**(7):5069-5072.
25. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C, Langille M: **PICRUSt2 for prediction of metagenome functions.** *NAT BIOTECHNOL* 2020, **38**(6):685-688.
26. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A *et al*: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases.** *NUCLEIC ACIDS RES* 2014, **42**(Database issue):D459-D471.
27. Parks DH, Tyson GW, Hugenholtz P, Beiko RG: **STAMP: statistical analysis of taxonomic and functional profiles.** *BIOINFORMATICS* 2014, **30**(21):3123-3124.
28. Friedman J, Alm EJ: **Inferring correlation networks from genomic survey data.** *PLOS COMPUT BIOL* 2012, **8**(9):e1002687.
29. Chen J, Domingue JC, Sears CL: **Microbiota dysbiosis in select human cancers: Evidence of association and causality.** *SEMIN IMMUNOL* 2017, **32**:25-34.
30. Druzhinin VG, Matskova LV, Demenkov PS, Baranova ED, Volobaev VP, Minina VI, Apalko SV, Churina MA, Romanyuk SA, Shcherbak SG *et al*: **Taxonomic diversity of sputum microbiome in lung cancer patients and its relationship with chromosomal aberrations in blood lymphocytes.** *Sci Rep* 2020, **10**(1):9681.
31. Reinhold L, Mollering A, Wallis S, Palade E, Schafer K, Dromann D, Rupp J, Graspeuntner S, Dalhoff K: **Dissimilarity of Airway and Lung Tissue Microbiota in Smokers undergoing Surgery for Lung Cancer.**

32. Jin J, Gan Y, Liu H, Wang Z, Yuan J, Deng T, Zhou Y, Zhu Y, Zhu H, Yang S *et al*: **Diminishing microbiome richness and distinction in the lower respiratory tract of lung cancer patients: A multiple comparative study design with independent validation.** *LUNG CANCER* 2019, **136**:129-135.
33. Liu HX, Tao LL, Zhang J, Zhu YG, Zheng Y, Liu D, Zhou M, Ke H, Shi MM, Qu JM: **Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects.** *INT J CANCER* 2018, **142**(4):769-778.
34. Cameron S, Lewis KE, Huws SA, Hegarty MJ, Lewis PD, Pachebat JA, Mur L: **A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer.** *PLOS ONE* 2017, **12**(5):e177062.
35. Woodard GA, Jones KD, Jablons DM: **Lung Cancer Staging and Prognosis.** *Cancer Treat Res* 2016, **170**:47-75.
36. Tessema M, Yingling CM, Liu Y, Tellez CS, Van Neste L, Baylin SS, Belinsky SA: **Genome-wide unmasking of epigenetically silenced genes in lung adenocarcinoma from smokers and never smokers.** *CARCINOGENESIS* 2014, **35**(6):1248-1257.
37. Inamura K: **Clinicopathological Characteristics and Mutations Driving Development of Early Lung Adenocarcinoma: Tumor Initiation and Progression.** *INT J MOL SCI* 2018, **19**(4).
38. Marks LB, Saynak M, Christodouleas JP: **Stage III vs. stage IV lung cancer: "Crossing a Great Divide".** *LUNG CANCER* 2010, **67**(1):1-3.
39. Hosgood HR, Sapkota AR, Rothman N, Rohan T, Hu W, Xu J, Vermeulen R, He X, White JR, Wu G *et al*: **The potential role of lung microbiota in lung cancer attributed to household coal burning exposures.** *ENVIRON MOL MUTAGEN* 2014, **55**(8):643-651.
40. Hosgood HR, Mongodin EF, Wan Y, Hua X, Rothman N, Hu W, Vermeulen R, Seow WJ, Rohan T, Xu J *et al*: **The respiratory tract microbiome and its relationship to lung cancer and environmental exposures found in rural china.** *ENVIRON MOL MUTAGEN* 2019, **60**(7):617-623.
41. Liu HX, Tao LL, Zhang J, Zhu YG, Zheng Y, Liu D, Zhou M, Ke H, Shi MM, Qu JM: **Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects.** *INT J CANCER* 2018, **142**(4):769-778.
42. Harris JK, De Groote MA, Sagel SD, Zemanick ET, Kapsner R, Penvari C, Kaess H, Deterding RR, Accurso FJ, Pace NR: **Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis.** *Proc Natl Acad Sci U S A* 2007, **104**(51):20529-20533.

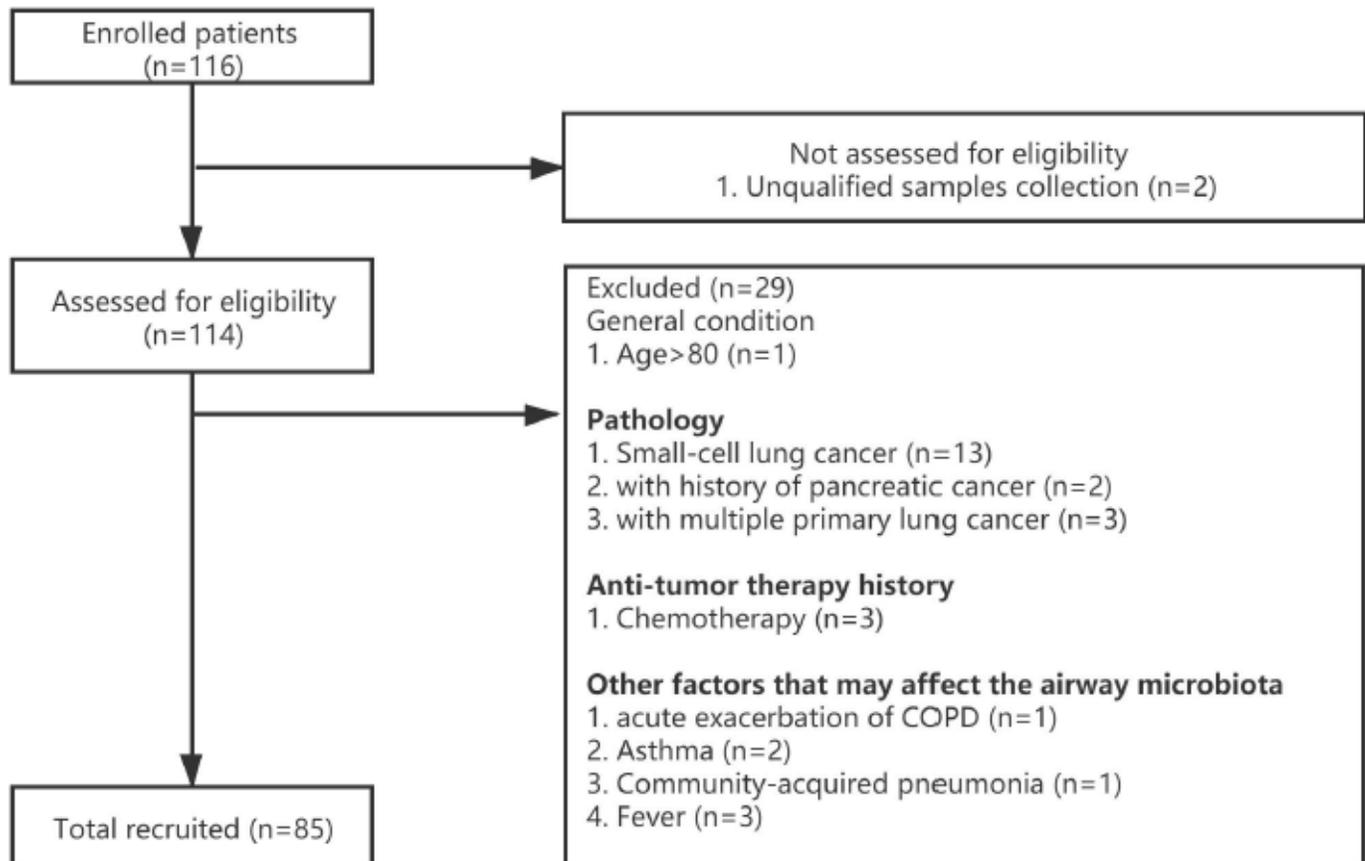
43. De Luca M, Amodio D, Chiurchiu S, Castelluzzo MA, Rinelli G, Bernaschi P, Calo CF, D'Argenio P: **Granulicatella bacteraemia in children: two cases and review of the literature.** *BMC PEDIATR* 2013, **13**:61.
44. Haldar K, George L, Wang Z, Mistry V, Ramsheh MY, Free RC, John C, Reeve NF, Miller BE, Tal-Singer R *et al*: **The sputum microbiome is distinct between COPD and health, independent of smoking history.** *Respir Res* 2020, **21**(1):183.
45. Parris BA, O'Farrell HE, Fong KM, Yang IA: **Chronic obstructive pulmonary disease (COPD) and lung cancer: common pathways for pathogenesis.** *J THORAC DIS* 2019, **11**(Suppl 17):S2155-S2172.
46. Rybojad P, Los R, Sawicki M, Tabarkiewicz J, Malm A: **Anaerobic bacteria colonizing the lower airways in lung cancer patients.** *Folia Histochem Cytobiol* 2011, **49**(2):263-266.
47. Tsay JJ, Wu BG, Badri MH, Clemente JC, Shen N, Meyn P, Li Y, Yie TA, Lhaxhang T, Olsen E *et al*: **Airway Microbiota Is Associated with Upregulation of the PI3K Pathway in Lung Cancer.** *Am J Respir Crit Care Med* 2018, **198**(9):1188-1198.
48. Popper HH: **Progression and metastasis of lung cancer.** *Cancer Metastasis Rev* 2016, **35**(1):75-91.
49. Yaku K, Okabe K, Hikosaka K, Nakagawa T: **NAD Metabolism in Cancer Therapeutics.** *FRONT ONCOL* 2018, **8**:622.
50. Zhu Y, Liu J, Park J, Rai P, Zhai RG: **Subcellular compartmentalization of NAD(+) and its role in cancer: A sereNADe of metabolic melodies.** *Pharmacol Ther* 2019, **200**:27-41.
51. Yaku K, Okabe K, Hikosaka K, Nakagawa T: **NAD Metabolism in Cancer Therapeutics.** *FRONT ONCOL* 2018, **8**:622.
52. Jin C, Lagoudas GK, Zhao C, Bullman S, Bhutkar A, Hu B, Ameh S, Sandel D, Liang XS, Mazzilli S *et al*: **Commensal Microbiota Promote Lung Cancer Development via gammadelta T Cells.** *CELL* 2019, **176**(5):998-1013.
53. Sheng Q, Du H, Cheng X, Cheng X, Tang Y, Pan L, Wang Q, Lin J: **Characteristics of fecal gut microbiota in patients with colorectal cancer at different stages and different sites.** *ONCOL LETT* 2019, **18**(5):4834-4844.
54. Tsoi H, Chu E, Zhang X, Sheng J, Nakatsu G, Ng SC, Chan A, Chan F, Sung J, Yu J: **Peptostreptococcus anaerobius Induces Intracellular Cholesterol Biosynthesis in Colon Cells to Induce Proliferation and Causes Dysplasia in Mice.** *GASTROENTEROLOGY* 2017, **152**(6):1419-1433.
55. Quail DF, Joyce JA: **Microenvironmental regulation of tumor progression and metastasis.** *NAT MED* 2013, **19**(11):1423-1437.

56. Romano AH, Conway T: **Evolution of carbohydrate metabolic pathways.** *RES MICROBIOL* 1996, **147**(6-7):448-455.
57. Evans MC, Buchanan BB, Arnon DI: **A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium.** *Proc Natl Acad Sci U S A* 1966, **55**(4):928-934.
58. Strohecker AM, White E: **Autophagy promotes BrafV600E-driven lung tumorigenesis by preserving mitochondrial metabolism.** *AUTOPHAGY* 2014, **10**(2):384-385.
59. Hoque MO, Kim MS, Ostrow KL, Liu J, Wisman GB, Park HL, Poeta ML, Jeronimo C, Henrique R, Lendvai A *et al*: **Genome-wide promoter analysis uncovers portions of the cancer methylome.** *CANCER RES* 2008, **68**(8):2661-2670.
60. Anderson NM, Mucka P, Kern JG, Feng H: **The emerging role and targetability of the TCA cycle in cancer metabolism.** *PROTEIN CELL* 2018, **9**(2):216-237.
61. Cai Z, Li CF, Han F, Liu C, Zhang A, Hsu CC, Peng D, Zhang X, Jin G, Rezaeian AH *et al*: **Phosphorylation of PDHA by AMPK Drives TCA Cycle to Promote Cancer Metastasis.** *MOL CELL* 2020, **80**(2):263-278.
62. Kovaleva OV, Romashin D, Zborovskaya IB, Davydov MM, Shogenov MS, Gratchev A: **Human Lung Microbiome on the Way to Cancer.** *J IMMUNOL RES* 2019, **2019**:1394191.
63. Mouraux S, Bernasconi E, Pattaroni C, Koutsokera A, Aubert JD, Claustre J, Pison C, Royer PJ, Magnan A, Kessler R *et al*: **Airway microbiota signals anabolic and catabolic remodeling in the transplanted lung.** *J Allergy Clin Immunol* 2018, **141**(2):718-729.
64. Conlon GA, Murray GI: **Recent advances in understanding the roles of matrix metalloproteinases in tumour invasion and metastasis.** *J PATHOL* 2019, **247**(5):629-640.
65. Yan X, Yang M, Liu J, Gao R, Hu J, Li J, Zhang L, Shi Y, Guo H, Cheng J *et al*: **Discovery and validation of potential bacterial biomarkers for lung cancer.** *AM J CANCER RES* 2015, **5**(10):3111-3122.
66. Nie C, He T, Zhang W, Zhang G, Ma X: **Branched Chain Amino Acids: Beyond Nutrition Metabolism.** *INT J MOL SCI* 2018, **19**(4).
67. Burgel PR, Nadel JA: **Epidermal growth factor receptor-mediated innate immune responses and their roles in airway diseases.** *EUR RESPIR J* 2008, **32**(4):1068-1081.
68. Lowenmark T, Lofgren-Burstrom A, Zingmark C, Eklof V, Dahlberg M, Wai SN, Larsson P, Ljuslinder I, Edin S, Palmqvist R: **Parvimonas micra as a putative non-invasive faecal biomarker for colorectal cancer.** *Sci Rep* 2020, **10**(1):15250.

69. Xu J, Yang M, Wang D, Zhang S, Yan S, Zhu Y, Chen W: **Alteration of the abundance of Parvimonas micra in the gut along the adenoma-carcinoma sequence.** *ONCOL LETT* 2020, **20**(4):106.

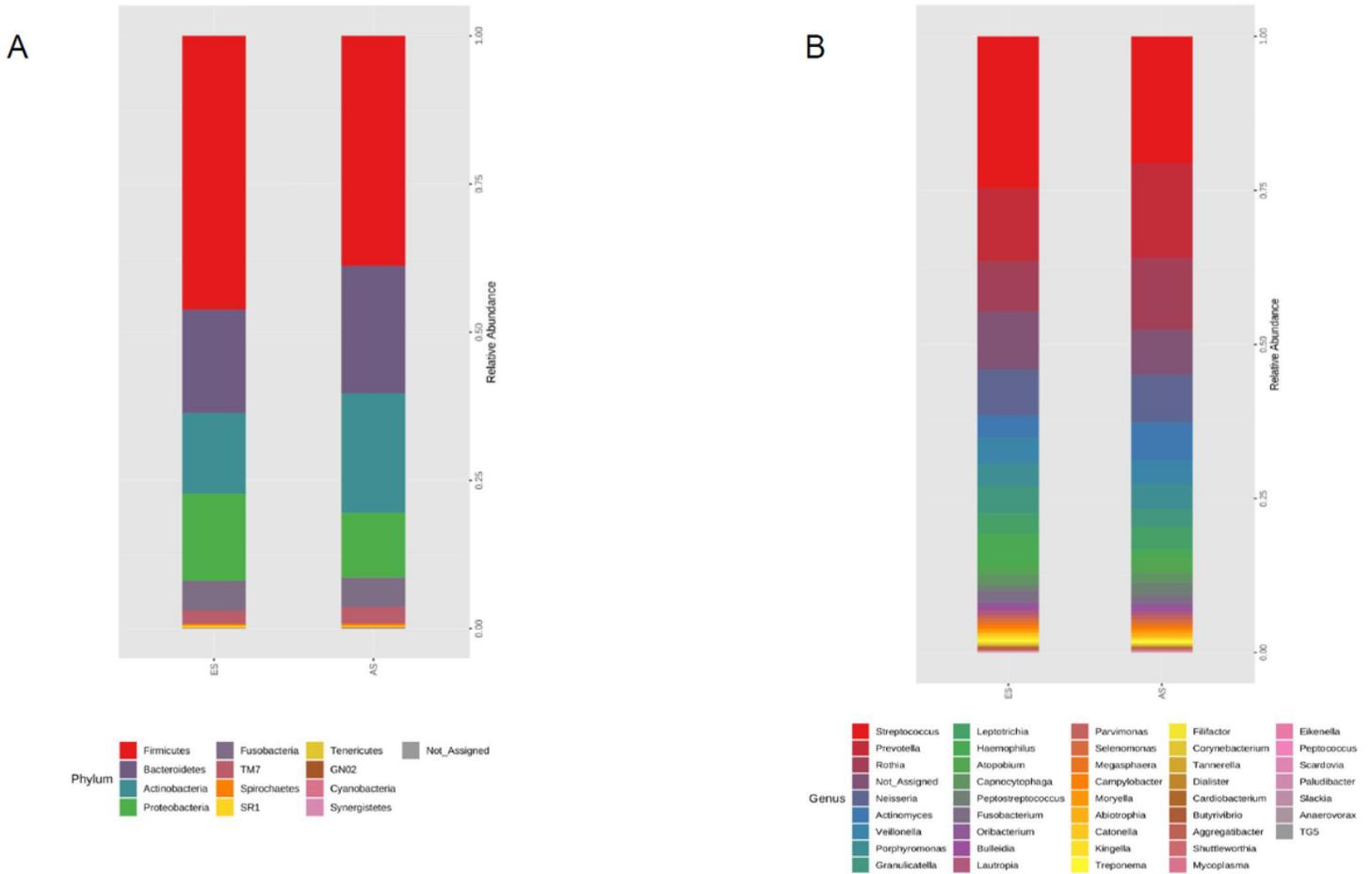
70. Matsui A, Jin JO, Johnston CD, Yamazaki H, Houry-Haddad Y, Rittling SR: **Pathogenic bacterial species associated with endodontic infection evade innate immune control by disabling neutrophils.** *INFECT IMMUN* 2014, **82**(10):4068-4079.

## Figures



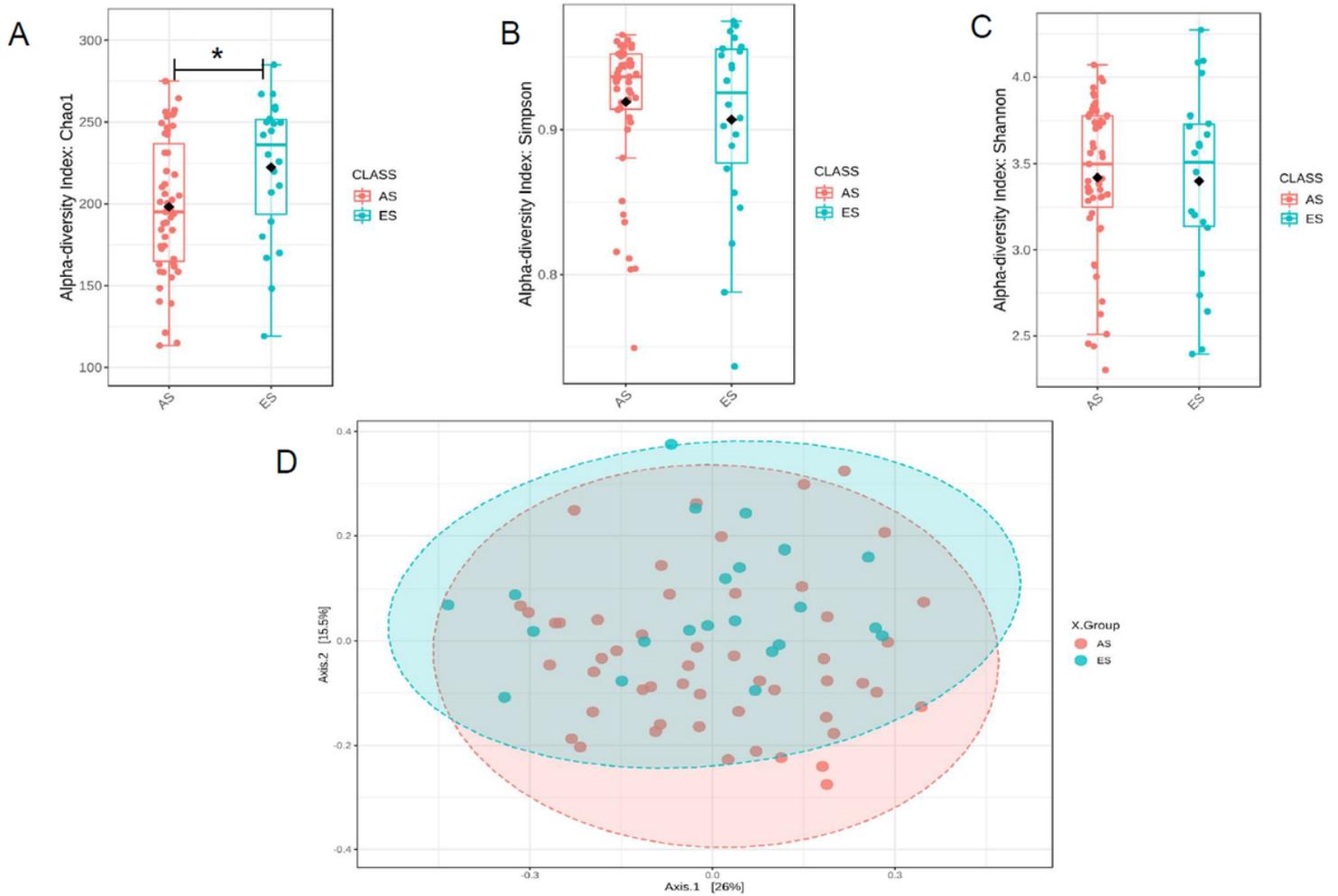
**Figure 1**

Study flow diagram of patients' recruitment and exclusion.



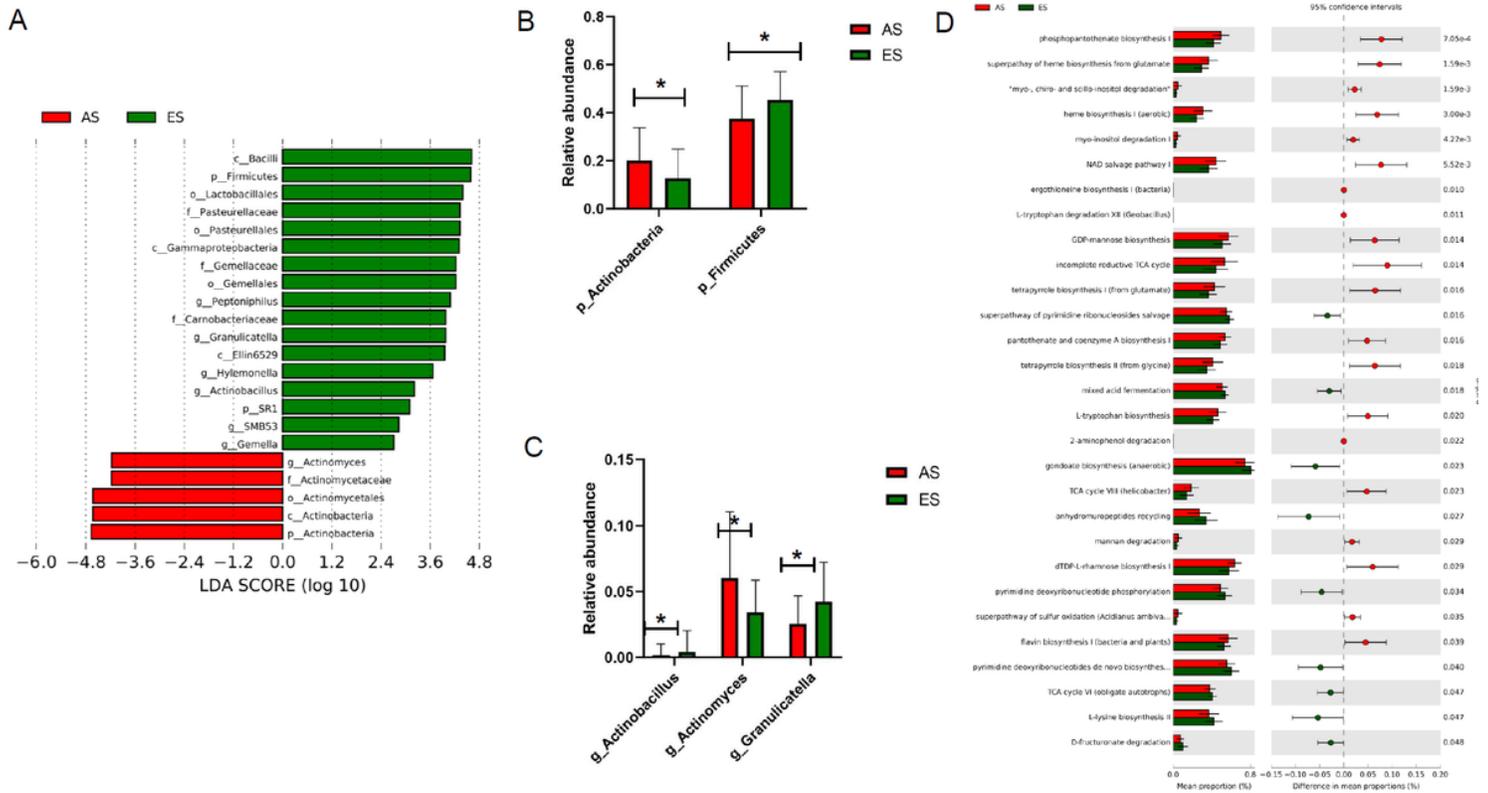
**Figure 2**

Taxonomic composition of sputum microbiota of the patients in ES and AS group. (A) Sputum phyla of the patients in ES and AS group; (B) Sputum Genera of the patients in ES and AS group.



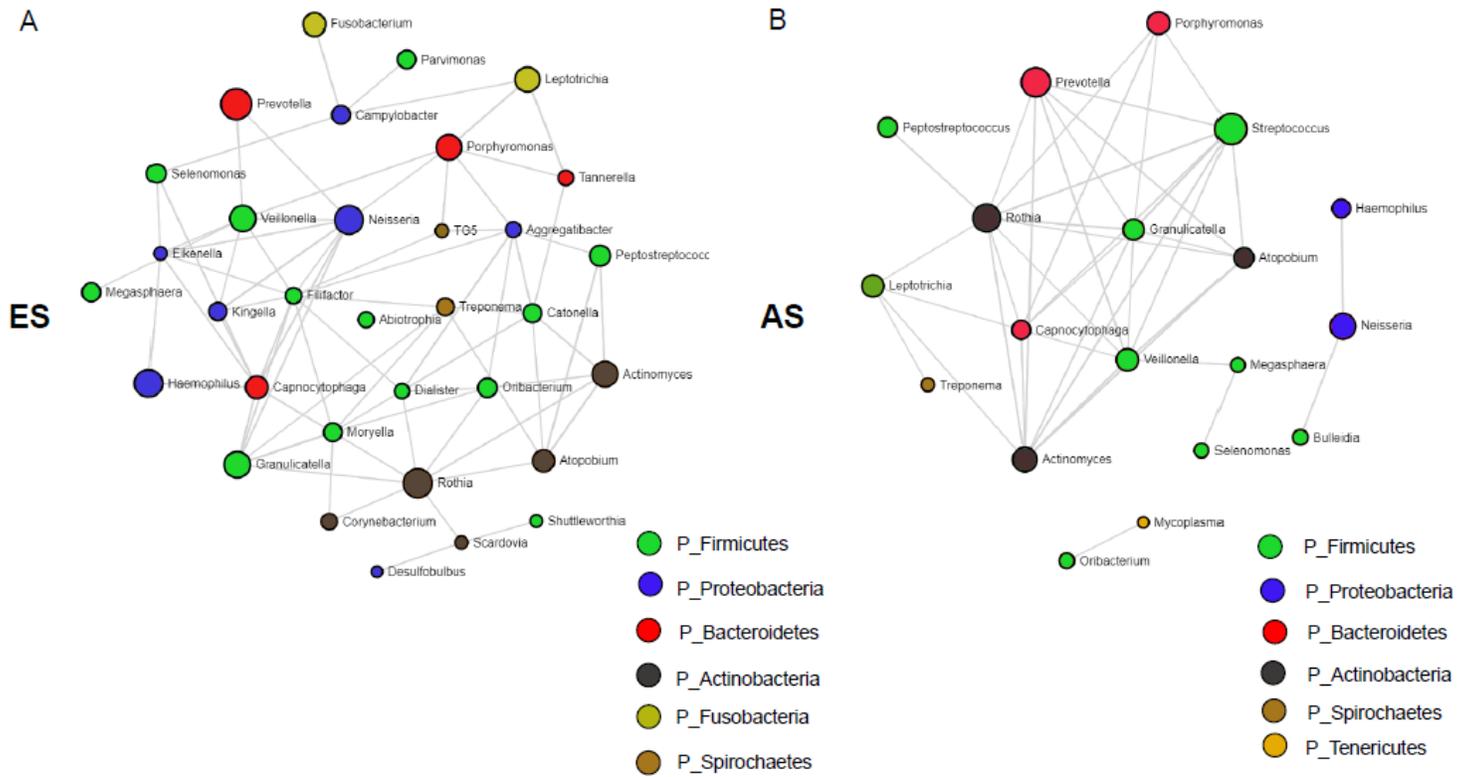
**Figure 3**

Difference of sputum microbiota between NSCLC patients in ES and AS group. (A) Chao 1 index; (B) Simpson index; (C) Shannon index among NSCLC patients in ES and AS group; (D) PCOA plot based on Bray-Curtis distance of sputum genus among NSCLC patients in ES and AS group. \* $P < 0.05$ , P was calculated using Mann-Whitney U test.



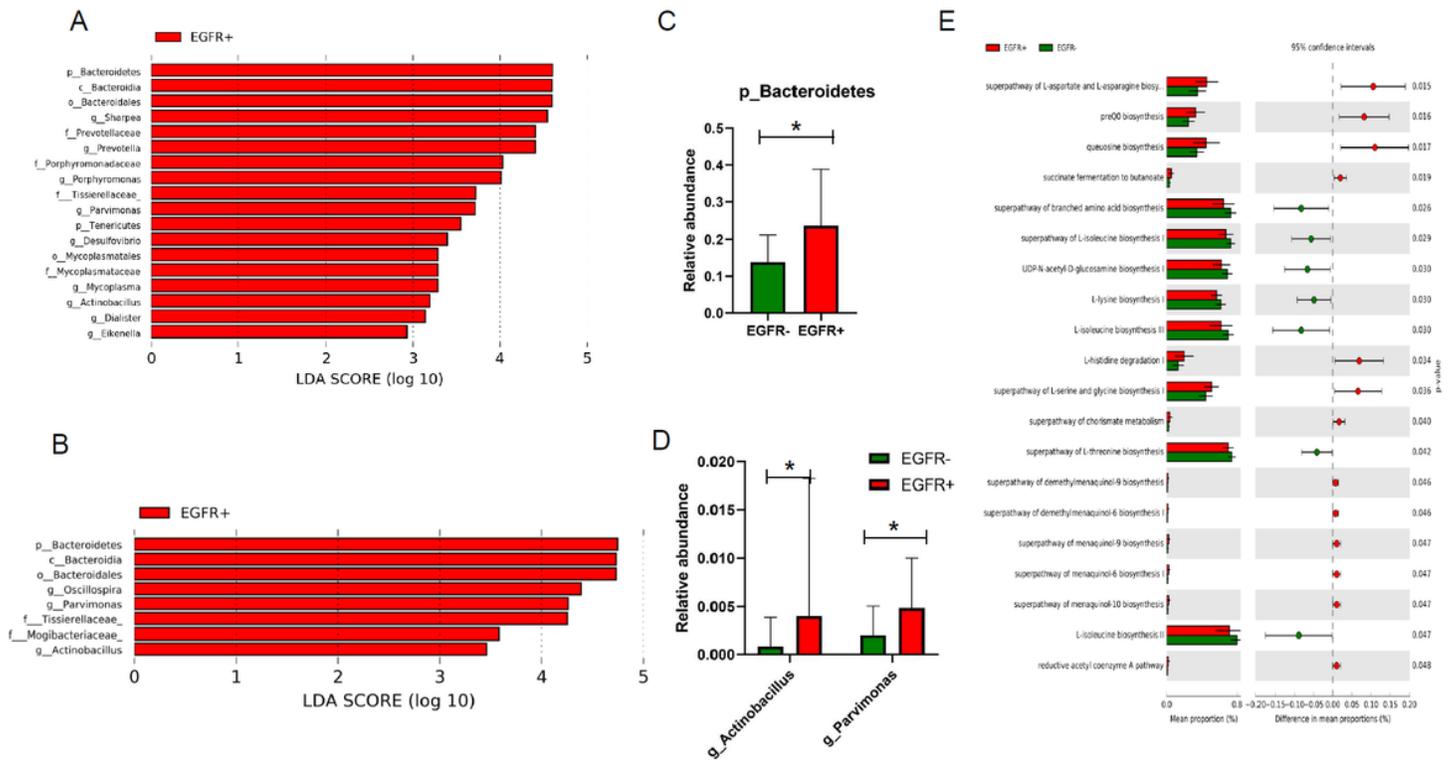
**Figure 4**

Differentially abundant taxonomy and predicted metabolic function of sputum microbiota between NSCLC patients in ES and AS group. (A) Differentially abundant taxonomy between patients in ES and AS group identified by LEFse; (B) Differentially abundant of phyla Actinobacteria, Firmicutes between ES and AS group; (C) Differentially abundant of genera Actinobacillus, Actinomyces and Granulicatella between SCC\_M1 and AD\_M1; (D) Differential predicted metabolic function based on MetaCyc database between patients in ES and AS group. \* $P < 0.05$ , P was calculated using Mann-Whitney U test.



**Figure 5**

Genera Co-occurrence network based on SparCC of patients in (A) ES group; (B) AS group. Only P value  $\leq 0.05$  and SparCC correlation scores  $\geq 0.5$  or  $\leq -0.5$  were included for networks inference. The genus nodes were colored based on phylum level. The size of each node was determined by the relative abundance of each genus.



**Figure 6**

Differentially abundant taxonomy and predicted metabolic function of sputum microbiota between lung adenocarcinoma patients with and without EGFR mutation. (A) Differentially abundant taxonomy between EGFR- lung adenocarcinoma and EGFR+ lung adenocarcinoma identified by LEFse; (B) Differentially abundant taxonomy between EGFR- non-smoker lung adenocarcinoma and EGFR+ non-smoker lung adenocarcinoma identified by LEFse; (C) Differentially abundant of phylum Bacteroidetes between EGFR- non-smoker lung adenocarcinoma and EGFR+ non-smoker lung adenocarcinoma; (D) Differentially abundant of genera Actinobacillus and Parvimonas between EGFR- non-smoker lung adenocarcinoma and EGFR+ non-smoker lung adenocarcinoma; (E) Differential predicted metabolic function based on MetaCyc database between EGFR- non-smoker lung adenocarcinoma and EGFR+ non-smoker lung adenocarcinoma. \* $P \leq 0.05$ , P was calculated using Mann-Whitney test.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFigure.pdf](#)
- [supplementarytable.pdf](#)