Project Name: Mycobiome-based Analysis of the Microbial Environment in Non-Small Cell Lung Cancer and Its Clinical Correlation

Project Type: Clinical Research

Project Number: 2025YG013

Department: Department of Anesthesiology

Study Duration: 2026.08–2028.08

(I) Main Research Content

1.Basis of project establishment

Lung cancer is the leading cause of cancer-related mortality worldwide, and its pathogenesis remains complex and incompletely understood. Small cell lung cancer accounts for approximately 15% of lung cancer cases; it is highly malignant, progresses rapidly, and is prone to metastasis, with most patients diagnosed at an advanced stage and thus ineligible for surgical resection. Non-small cell lung cancer (NSCLC) is the most common type of lung cancer, including adenocarcinoma, squamous cell carcinoma, large cell carcinoma, and sarcomatoid carcinoma, accounting for 85% of all lung cancer patients. Since bacteria were first cultured from human tumor tissues in the 1920s, mounting evidence has demonstrated the presence of microbial communities within tumors. Research on the tumor microbiome within the tumor microenvironment (TME) has provided a revolutionary perspective for deeply understanding tumor biology. A large body of evidence indicates that the bacterial microbiome plays key roles in shaping the immune status of the TME, influencing the efficacy of chemotherapeutic agents, and promoting tumor metastasis. However, as another important pillar of microbial communities, the role of the mycobiome in lung cancer has long been neglected.

As eukaryotic microorganisms, fungi possess more complex genetic structures and metabolic networks. Studies have shown that they can potentially influence tumor progression through multiple mechanisms, including secreting bioactive secondary metabolites, directly regulating host immune responses via pattern recognition receptors, and forming cross-kingdom interaction networks with bacterial communities. A series of groundbreaking studies have provided preliminary clues: pan-cancer analyses have confirmed the presence of specific fungal communities in lung cancer tissues that differ from those in adjacent normal tissues, with enrichment of *Aspergillus* and *Candida* significantly associated with malignant tumor progression. Although the prospects for mycobiome research are promising, existing studies remain in their infancy: most reports are limited to small sample sizes with vague clinical correlations. The associations between dynamic changes in fungal communities and key clinical characteristics such as pathological subtype, TNM stage, driver gene mutation status, and immunotherapy response have not been precisely elucidated, limiting their clinical translational potential. Moreover, the relationship between fungal infection and cancer is complex and influenced by environmental factors; observed differences depend on fungal culture and identification techniques, specific cancer types, and host immune status, among other factors. Therefore, more extensive clinical studies with larger sample sizes are needed to further investigate the correlation between the mycobiome and lung cancer.

Accordingly, this project aims to systematically reveal the compositional patterns and structural characteristics of the NSCLC mycobiome, and to deeply explore the correlations between pulmonary fungi and NSCLC pathological subtypes, staging.

The successful implementation of this study will not only deepen our understanding of

the lung cancer microenvironment and enrich the theoretical connotation of tumor microbiome science, but also provide novel biomarkers and intervention strategies for the early diagnosis and prognostic assessment of NSCLC.

2.Scientific Questions to Be Addressed

2.1 Revealing the compositional patterns and structural characteristics of the mycobiome in the NSCLC microbial environment

Current studies have detected fungal genera such as *Aspergillus*, *Candida*, *Malassezia*, *Nexia*, and *Saccharomyces* in sputum, bronchoalveolar lavage fluid, and lung tissue specimens, confirming the colonization of fungi in the lung tumor microenvironment. However, the compositional distribution and structural characteristics of fungal genera remain unclear. Moreover, host-microbe interactions are complex, and fungal detection depends on host immune response, microbial interactions, environmental conditions, sample contamination, and the technical proficiency of researchers, among other factors. Some studies have also identified a small number of lung cancer-specific colonizing strains, but these were based on small sample sizes and different study populations. Furthermore, high-risk fungal strains associated with lung cancer may not be limited to one or two species; their development likely involves multiple pathogenic organisms, multiple factors, and multiple pathways. Therefore, this project plans to collect surgically resected cancer tissue specimens from NSCLC patients, simultaneously collect paired adjacent normal tissues from the same patients, and perform DNA extraction, high-throughput sequencing, and in vitro co-culture to systematically construct an ecological atlas of the NSCLC mycobiome and establish the first well-defined NSCLC mycobiome reference database.

2.2 Investigating the correlations between pulmonary fungi and NSCLC pathological subtypes, staging

As eukaryotic microorganisms, fungi possess more complex genetic structures and metabolic networks. Studies have shown that they can potentially influence tumor progression through multiple mechanisms, including secreting bioactive secondary metabolites, directly regulating host immune responses via pattern recognition receptors, and forming cross-kingdom interaction networks with bacterial communities. However, current clinical correlations remain vague. The associations between dynamic changes in fungal communities and key clinical characteristics such

as pathological subtype, TNM stage, driver gene mutation status, and immunotherapy response have not been precisely elucidated, limiting their clinical translational potential. This project plans to integrate clinical pathological information and follow-up data, and utilize bioinformatics and statistical methods to screen for key signature fungi and analyze the correlations between fungal community structure and different NSCLC pathological subtypes, stages.

3.Technical Issues to Be Addressed

3.1 High difficulty of fungal culture

Slow growth: Requires several days to several weeks to form visible colonies.
Specific media required: Sabouraud dextrose agar, potato dextrose agar, etc.
Temperature requirements: Deep fungi need to be cultured at 35–37°C, close to human body temperature, to promote growth.
Humidity requirements: Environmental humidity needs to be maintained above 60%.
pH: Most fungi adapt to weakly acidic environments, with pH generally between 4.0 and 6.0. Some media are adjusted for pH to meet the growth requirements of specific fungi.

3.2 Difficulty of DNA extraction

As eukaryotes, fungi have membrane-bound nuclei with DNA bound to proteins as chromosomes. Their cell walls are difficult to lyse, being mainly composed of chitin, β-glucan, and mannan polysaccharides. Moreover, lung tissue is tough, adding to the difficulty of DNA extraction. Our team has optimized the fungal DNA extraction procedure for lung tissue by multiple grinding steps, adding multiple grinding beads, and modifying lysis conditions (65°C water bath for 1 hour), which has greatly improved the efficiency of fungal DNA extraction from lung tissue.

4.Main Research Content

4.1Multi-dimensional analysis of NSCLC fungal communities and construction of an ecological atlas

Sample collection and processing: Collect surgically resected tissue specimens from patients diagnosed with "non-small cell lung cancer" (confirmed by preoperative needle biopsy, intraoperative frozen section pathology, or postoperative paraffin section) from the Department of Thoracic and Cardiovascular Surgery of the Army 958 Hospital and the Department of Thoracic Surgery of Shenzhen Nanshan People's

Hospital. Using a standardized sampling protocol, two sets of tissues will be simultaneously collected from the same patient under aseptic conditions: lung cancer tissue and paired adjacent normal tissue (>3 cm from the tumor edge), each approximately 0.5 cm in diameter. Sample collection will strictly follow STROBE guidelines to ensure biological representativeness. Detailed clinical and pathological data will also be collected, including TNM stage, driver gene mutation status, treatment history, and follow-up data.

Application of high-throughput multi-omics technologies: ITS2 region amplicon sequencing technology will be employed, using optimized primer pairs (ITS3/ITS4) for nested PCR amplification, followed by paired-end sequencing on the Illumina NovaSeq platform. Sequence denoising will be performed using the DADA2 algorithm to obtain amplicon sequence variants (ASVs). The analysis pipeline includes: α-diversity (Chao1, Shannon, and Simpson indices), β-diversity (PCoA and NMDS analysis based on Bray-Curtis and UniFrac distances), differential species identification (LEfSe analysis, LDA score >3.0), and co-occurrence network analysis (SparCC algorithm). Microbial community reconstruction will further reveal the ecological composition of the NSCLC mycobiome and screen for key fungal biomarkers.

Establishment of a multi-medium combined culture system: Isolate culturable strains under anaerobic/microaerophilic conditions. Verify the activity and purity of dominant strains through morphological observation, physiological and biochemical tests, and ITS sequence phylogenetic analysis. Scanning electron microscopy will be used to observe the colonization status of fungi in tissues, and an NSCLC-related fungal resource library will be established.

4.2 Clinical correlation study between the mycobiome and lung cancer

Baseline data collection: Based on the electronic medical record system and clinical history, comprehensive clinical data will be collected, including: (1) Patient gender, age, BMI, smoking history, and comorbidities such as hypertension, diabetes, chronic obstructive pulmonary disease, and chronic pharyngitis; (2) Clinical symptoms including fever, chills, cough, hemoptysis, sputum production, sputum characteristics and volume, and antibiotic use within one month prior to admission; (3) Relevant laboratory and examination reports: blood routine, liver and kidney function, CRP, blood gas analysis, and pulmonary function reports; (4) Imaging data: chest X-ray, chest CT, PET-CT, cranial MRI, etc. The size and location of primary and metastatic lesions, and whether complicated by pleural effusion and its volume and nature will be recorded. TNM stage and clinical stage will be determined; (5) Histopathological data: collection of pathology reports and NSCLC pathological subtypes; (6) Treatment regimens: whether preoperative neoadjuvant chemotherapy or immunotherapy was administered.

Statistical analysis: Statistical analysis will be performed using SPSS and R software. For continuous variables such as age and BMI, if normally distributed, two independent samples t-tests will be used for between-group comparisons, and one-way ANOVA will be used for multi-group comparisons. For non-normally distributed data, non-parametric tests will be used, mainly Wilcoxon rank-sum test or Mann-Whitney U test for two samples, and Kruskal-Wallis test for multiple samples. For categorical variables, percentages will be used, and chi-square test or Fisher's exact test will be applied.

Biomarker mining: A three-stage feature screening strategy will be adopted: first, univariate analysis (Wilcoxon test, Spearman correlation) will be used for initial screening of candidate features; then LASSO regression will be applied for feature compression to eliminate collinearity; finally, machine learning algorithms such as random forest and XGBoost will be used to screen key species closely related to early occurrence, progression, and prognosis of lung cancer from the fungal community data. Innovatively, composite indicators such as fungal abundance and community stability index will be introduced, combined with clinical baseline data to construct a multi-dimensional feature panel. Recursive feature elimination will be used to optimize feature combinations, and feature importance scores will be calculated. For screened key fungal biomarkers, qPCR technology will be used for validation to ensure detection stability and reproducibility.

(II) Research Methods (Methods, principles, mechanisms, algorithms, and models to be adopted to address the project issues; limit: 2,000 words)

1.Research Protocol

This study is a case-control study. The study subjects are patients initially diagnosed with "lung mass/lung nodule/lung cancer" who are planned for surgical resection. All patients will sign informed consent. Two sets of tissues will be simultaneously collected from the same patient under aseptic conditions: lung cancer tissue and paired adjacent normal tissue (>3 cm from the tumor edge), each approximately 0.5 cm in diameter.

Inclusion criteria: (1) Diagnosed as non-small cell lung cancer by preoperative needle biopsy, intraoperative frozen section pathology, or postoperative paraffin section, and planned for surgical resection; (2) Aged 30 to 80 years; (3) No preoperative chemotherapy, radiotherapy, targeted therapy, or immunotherapy or other anti-tumor treatments.

Exclusion criteria:(1) Patients not meeting inclusion criteria; (2) Combined with primary malignancies of other systems or recurrent lung cancer, or combined with

immunodeficiency diseases; (3) Combined with acute exacerbation of chronic obstructive pulmonary disease, bronchial asthma, bronchiectasis with infection, pulmonary tuberculosis, or pneumoconiosis; (4) Recent use of antifungal drugs; (5) Basic clinical information (such as pathological type, clinical stage, treatment regimen) unavailable.

Sample size estimation: The study is designed as a paired case-control study (cancer tissue vs. adjacent normal tissue). Based on the latest research published in international authoritative journals, the estimated exposure value P1 for the case group is 0.3, and the estimated odds ratio (OR) is 2. Parameters are set as: significance level $\alpha=0.05$, statistical power $1-\beta=0.9$. According to the following formula, the required sample size per group is M=215. Considering an approximate 15% dropout rate, the final sample size is determined to be approximately 500 cases. Sample size estimation formula:

$$M=\frac{[\frac{u_{\alpha}}{2}+u_{\beta}\sqrt{\frac{OR}{1+OR}(1-\frac{OR}{1+OR})}]^2/(\frac{OR}{1+OR}-0.5)^2}{(p_0q_1+p_1q_0)}$$

The collected data from the two groups (cancer tissue vs. adjacent normal tissue) will be analyzed by bioinformatics methods to explore the differential expression of the NSCLC mycobiome. Subgroup analysis will be performed on cancer tissue samples grouped by different criteria to describe the fungal distribution characteristics within and between different subgroups, and to explore the correlation between mycobiome and NSCLC:

(1) Based on tissue sampling site: The two sets of tissues simultaneously collected from the same patient will be divided into cancer tissue group and adjacent normal tissue group, and intra-group and inter-group mycobiome characteristics will be compared.
(2) Based on different pathological subtypes: According to the 2021 WHO Classification of Thoracic Tumors, cancer tissue samples will be divided into adenocarcinoma group, squamous cell carcinoma group, large cell carcinoma group, and sarcomatoid carcinoma group.

2.Sample Collection, Processing, and High-Throughput Sequencing

A paired design will be adopted, with two sets of tissues simultaneously collected from the same patient: cancer tissue and adjacent normal tissue (>3 cm from the tumor edge). Each patient will provide 2 tissue samples (0.5 cm in diameter), totaling 500 samples.

Sample processing: Intraoperatively, tissue samples will be placed in EP tubes under aseptic conditions and transferred to a -80°C ultra-low temperature freezer within 30

minutes. An electronic sample management system will be established to ensure traceability and quality control.

DNA extraction and quality control: DNA will be extracted using the DNeasy PowerSoil Pro Kit, with optimized lysis conditions (65°C water bath for 1 hour). Negative controls (no template) and positive controls (known fungal standards) will be set up for each batch, and synthetic DNA internal standards (ZymoBIOMICS Spike-in Control) will be added to monitor extraction efficiency.

Quality control indicators: DNA concentration ≥10 ng/μL (measured by Qubit), A260/A280 ratio 1.8–2.0 (measured by NanoDrop), integrity verified by agarose gel electrophoresis and qPCR (ITS region amplification). Unqualified samples will be excluded, with an estimated dropout rate ≤10%.

Fungal isolation, culture, and identification: Multi-medium combinations (Sabouraud dextrose agar, potato dextrose agar, brain heart infusion agar) will be used, cultured at 28°C under anaerobic/microaerophilic conditions for 14 days. Three replicates will be set up per sample. Strain identification will include morphological observation (microscopy), physiological and biochemical tests (API 20C AUX), and molecular identification by ITS sequence sequencing. An NSCLC-related fungal resource library will be established.

High-throughput sequencing: ITS2 amplicon sequencing: Nested PCR amplification of the ITS2 region will be performed using primers ITS3 (5'-GCATCGATG0000000AAGAACGCAGC-3') and ITS4 (5'-TCCTCCGCTTATTGATATGC-3'). PCR conditions: 95°C for 3 min; 35 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 45 s; 72°C for 10 min.

Sequencing platform: Illumina NovaSeq 6000 (2×250 bp paired-end sequencing), target sequencing depth ≥50,000 reads/sample. Library construction will use the KAPA HyperPlus Kit, and quality control will be assessed by FastQC.

3.Data Analysis

Raw data quality control: Trimmomatic will be used to remove low-quality sequences (Q<20) and adapters.

Sequence denoising: The DADA2 algorithm will be used to generate amplicon sequence variants (ASVs).

Taxonomic annotation: Taxonomic assignment will be performed using the UNITE database (v9.0).

Diversity analysis: α-diversity (Chao1, Shannon, and Simpson indices) reflects species quantity (richness) and the evenness of species individual counts; β-diversity (Bray-Curtis distance, PCoA analysis) reflects the structural and compositional differences between pairwise samples.

Differential analysis: LEfSe (LDA score >3.0) will be used to identify differential species between groups.

Data integration: A standardized database (MySQL) will be established, containing clinical data, sequencing data, and experimental data. Data will be de-identified and backed up on encrypted servers.

Statistical analysis:

Between-group comparisons: categorical variables (χ^2 test), continuous variables (ANOVA or Kruskal-Wallis test);
Diversity analysis: PERMANOVA to test between-group differences in β-diversity;
Machine learning: Implemented in Python, with hyperparameters optimized via GridSearchCV.
Software and tools: QIIME2, MetaboAnalyst 5.0, R 4.1.0.

Informed Consent Form · Information Page

Dear Patient:

We would like to invite you to participate in a research study: **"Mycobiome-based Analysis of Microbial Environment in Non-Small Cell Lung Cancer and Its Clinical Correlation Study".** This research protocol has been reviewed and approved by the Ethics Committee for clinical research.

Before you decide whether to participate in this study, please read the following content as carefully as possible. It will help you understand what this study involves, why it is being conducted, the procedures and duration of the study, and the potential benefits, risks, and inconveniences that may result from your participation. If you wish, you may also discuss this with your relatives or friends, or ask your doctor for an explanation to help you make your decision.

I. Background and Objectives of the Study

1.1 Disease Burden and Current Treatment Status

Lung cancer is the leading cause of cancer-related mortality worldwide, and its pathogenesis is complex. Non-small cell lung cancer (NSCLC) is the most common type of lung cancer, accounting for 85% of all lung cancer cases. In recent years, the association between the tumor microenvironment microbiome, particularly the bacteriome, and tumors has received extensive attention. However, the role of the mycobiome, another important component of the microbiome, in lung cancer remains unclear.

1.2 Objectives of This Study

This project aims to investigate the structural characteristics of the fungal microbiome in lung tissue of NSCLC patients and its correlation with patient prognosis. By integrating clinicopathological information and follow-up data, and utilizing bioinformatics and statistical methods, we will screen for key signature fungi and analyze the correlation between fungal community structure and different pathological subtypes, stages, clinical features, and patient prognosis of NSCLC.

1.3 Participating Institutions and Estimated Number of Subjects

This study will be conducted jointly by the 958th Hospital of the Chinese People's Liberation Army (PLA) and the Shenzhen Nanshan People's Hospital. It is estimated that **500 subjects** will be enrolled.

II. Who Should Not Participate in This Study

(1) Patients who do not meet the inclusion criteria;

(2) Patients with concurrent primary malignancies in other systems or recurrent lung cancer, or those with immunodeficiency diseases;

(3) Patients with acute exacerbation of chronic obstructive pulmonary disease, bronchial asthma, bronchiectasis with infection, pulmonary tuberculosis, or pneumoconiosis;

(4) Patients who have recently used antifungal medications;

(5) Patients whose basic clinical information (such as pathological type, clinical stage, and treatment plan) cannot be obtained.

III. What Will Be Required If You Participate in This Study?

1. Before you are enrolled in the study, the doctor will inquire about and record your medical history, and assess your condition. If you meet the inclusion criteria and voluntarily agree to participate in the study, you will sign this informed consent form. If you do not wish to participate in the study, this will not result in any prejudice against you or affect your medical care.
2. If you voluntarily participate in the study, we will collect a small portion of your lung tissue resected during surgery for this study, and collect your basic information.
3. Other Matters Requiring Your Cooperation
   During the course of the study, new information regarding the study methods may become available. If new information arises, your study doctor will inform you promptly and discuss with you whether you are still willing to continue participating in this study. If you decide to continue participating in the study, you may be asked to sign a new informed consent form. During the follow-up phase, the doctor may learn about your condition through telephone calls, outpatient follow-up visits, and other means.

IV. Potential Benefits of Participation

Participating in this study may allow for timely assessment of your disease condition and guide subsequent treatment, but this cannot be guaranteed. You will not receive any form of direct financial benefit from this study.

V. Potential Adverse Reactions, Risks, Inconveniences, and Discomforts

If you experience any discomfort, new changes in your condition, or any unexpected events during the study period, regardless of whether they are related to the study, please notify your doctor promptly. He/she will make a judgment and provide appropriate medical treatment.

VI. Related Expenses

The costs associated with specimen collection, specimen processing, data collection, and analysis for this study will be borne by the hospital.

During the study period, the doctor will make every effort to prevent and treat any harm that may result from this study. If an adverse event occurs during the clinical study, the Medical Expert Committee will determine whether it is related to the study. The sponsor will provide corresponding treatment expenses and financial compensation for study-related injuries in accordance with relevant regulations.

VII. Confidentiality of Personal Information

Any information and data about you obtained during the study will be kept strictly confidential. Your lung tissue specimens will be identified by a study code/number rather than your name. Information that could identify you will not be disclosed to anyone outside the research team unless you give permission. Any public reports of the results of this study will not disclose your personal identity. We will make every effort to protect the privacy of your personal medical information within the limits permitted by law.

In accordance with medical research ethics, apart from personal privacy information, study data will be made available for public inquiry and sharing. Such inquiry and sharing will be limited to web-based electronic databases, ensuring that no personal privacy information is leaked.

VIII. How to Obtain More Information?

You may ask any questions about this study at any time and receive corresponding answers. If any important new information arises during the study that may affect your willingness to continue participating, your doctor will notify you promptly.

IX. Voluntary Participation and Right to Withdraw

Whether or not to participate in this study is entirely up to you. You may refuse to participate in this study, or withdraw from it at any time during the study. This will

not affect your relationship with your doctor, nor will it affect your medical care or other interests.

In consideration of your best interests, the doctor or researcher may discontinue your participation in this study at any time during the study.

X. What Should You Do Now?

Whether or not to participate in this study is a decision for you (and your family) to make.

Before you make a decision to participate, please ask your doctor any questions you may have about the study.

Thank you for reading the above materials. If you decide to participate in this study, please inform your doctor, who will arrange all study-related matters for you. Please retain this document.

Informed Consent Form · Signature Page

Clinical Study Project Name: Mycobiome-based Analysis of Microbial Environment in Non-Small Cell Lung Cancer and Its Clinical Correlation Study

Project Undertaking Unit: 958th Hospital of the PLA

Project Task Number: 2025YG013

Statement of Consent

I have read the above introduction to this study and have had the opportunity to discuss this study with the doctor and ask questions. All questions I raised have been answered to my satisfaction.

I understand the potential risks and benefits of participating in this study. I know that participation is voluntary. I confirm that I have had sufficient time to consider this, and I understand that:

I may consult the doctor for more information at any time.

I may withdraw from this study at any time without being subject to discrimination or retaliation, and my medical treatment and rights will not be affected.

I also understand that if I withdraw from the study, especially if I withdraw due to medication reasons, it would be greatly beneficial to the overall study if I inform my doctor of any changes in my condition and complete the corresponding physical and laboratory examinations. If changes in my condition require me to take any other medications, I will seek the doctor's opinion in advance, or inform the doctor truthfully afterwards.

I agree to allow the drug regulatory administration, the Ethics Committee, or the sponsor's representative to review my study data.

I will receive a signed and dated copy of this informed consent form.

Finally, I have decided to agree to participate in this study and will do my best to follow the doctor's advice.

Patient Signature: _____ Date: ____ Year ____ Month ____ Day

Contact Telephone: _____

I confirm that I have explained the details of this trial to the patient, including his/her rights and possible benefits and risks, and have given him/her a signed copy of the informed consent form.

Doctor Signature: _____ Date: ____ Year ____ Month ____ Day

Doctor's Work Telephone: _____