RESEARCH

Open Access

Exploration of lung mycobiome in the patients with non-small-cell lung cancer



Abstract

As the Human Microbiome Project (HMP) progresses, the relationship between microbes and human health has been receiving increasing attention. A growing number of reports support the correlation between cancer and microbes. However, most studies have focused on bacteria, rather than fungal communities. In this study, we studied the alteration in lung mycobiome in patients with non-small-cell lung cancer (NSCLC) using metagenomic sequencing and qPCR. The higher fungal diversity and more complex network were observed in the patients with NSCLC. In addition, *Alternaria arborescens* was found as the most relevant fungus to NSCLC, and the enrichment of it in cancerous tissue was also detected. This study proposes that the changes in fungal communities may be closely related to lung cancer, and provides insights into further exploration the relationship between lung cancer and fungi.

Keywords Non-small-cell lung cancer, Mycobiome, Alternaria arborescens

Introduction

Microbes have been implicated in human health [1], and some researches have reported that microbes are also related to lung cancer [2-6]. The microbial composition of the lower airways of lung cancer patients was unique, and this microbial alteration is considered associated

[†]Yiming Zhao and Junqi Yi contributed equally to this work.

*Correspondence:

Zheng Yu

yuzheng@csu.edu.cn

Jingqun Tang

tangjq@csu.edu.cn

⁴ NHC Key Laboratory of Carcinogenesis and the Key Laboratory

of Carcinogenesis and Cancer Invasion of the Chinese Ministry of Education, Cancer Research Institute, School of Basic Medical Science,

Central South University, Changsha, Huna, China

⁵ School of Chinese Medicine, Hong Kong Baptist University, Kowloon Tong, China with lung carcinogenesis [7, 8]. A dysregulated microbiota has a role in propagating and maintaining a chronic inflammatory environment [9, 10]. Some lung microbes affect T helper 17 cells which were key in modulation of lung immune status in health and disease [11]. In addition, microbes can directly affect protumorigenic pathways in epithelial cells [7].

Although the opinion that the lungs are sterile has been abandoned, researches on lung microbes have mainly focused on bacteria [12, 13]. Fungi are often neglected due to their low content. However, the mycobiome can play a beneficial or pathogenic role [14]. Fungal genera that have been detected in the pulmonary mycobiota mainly include *Candida, Malassezia, Neosartorya, Saccharomyces,* and *Aspergillus.* A systematic review showed that the human mycobiome, along with its interactions with the human bacteriome and the host, is implicated in the promotion and progression of carcinogenesis [15]. *Candida albicans* exhibit an oncogenic potential in oral cavity cancer [16–19]. *Malassezia* has been found among colorectal carcinoma patients, whilst an increased number of *Basidiomycota* have been suggested to be related



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

¹ Department of Thoracic Surgery, the Second Xiangya Hospital, Central South University, Changsha, Hunan, China

² Department of Microbiology, School of Basic Medical Science, Central South University, Changsha, Hunan, China

³ Hunan Key Laboratory of Early Diagnosis and Precise Treatment of Lung Cancer, Changsha, Hunan, China

to more advanced stages of this kind of cancer [20-22]. However, the relationship between lung mycobiome and lung cancer remains unclear. In this study, we characterized the alteration of fungal lung communities in patients with NSCLC and the Non-NSCLC groups, and found *A. arborescens* were enriched in NSCLC tissues.

Methods

Study population and sample collection

To study the lung mycobiome in patients with NSCLC and to verify the enrichment of key fungi in cancer tissues, 66 patients were enrolled with following inclusion criteria: (1) 40 < Age < 70; (2) No antibiotics within three months; (3) No immunosuppressive drugs within six months. All individuals included in this study had no evidence of infection, sepsis or active tuberculosis. The research presented here has been performed in accordance with the Declaration of Helsinki, and cases enrolled in this study were collected and approved by the ethical review committees from the second Xiangya hospital, Central South University, China (Ethical approval number: 2019–155). The patients were informed about the sample collection and had signed informed consent forms.

As for metagenomic sequencing and estimation of lung fungal content, the bronchoalveolar lavage (BAL) was collected from 38 individuals, including 24 newly diagnosed NSCLC and 14 Non-NSCLC patients (Table 1). There was no significant difference between the NSCLC and Non-NSCLC in terms of age, smoking status, and gender. To better quantify the cancer progress, the patients were divided into 10 groups according to TNM

 Table 1
 Clinical
 Characteristics
 of
 patients
 for
 metagenomic
 sequence

Feature	NSCLC patients		Non-NSCLC Patients 14(36.8%)	
Patients(n = 38)				
Genders(<i>n</i> = 38) <i>p</i> = 0.929	Male	12	Male	8
	Female	12	Female	6
Age(<i>n</i> =38) <i>p</i> = 0.125	≤50	9	<u>≤</u> 50	7
	> 50	15	> 50	7
Smoking status($n = 38$) $p = 1.000$	Smoker	4	Smoker	2
	Non-smoker	20	Non-smoker	12
TNM staging($n = 22$)	T1aN0M0	2		
	T1bN0M0	7		
	T1cN0M0	1		
	T2aN0M0	8		
	T2bN1M0	1		
	T3N1M0	1		
	T4N0M0	1		
	Tia	1		

All patients are immunocompetent in the cohort

stages (Table 2). The BAL samples were obtained under sterile conditions by instillation and aspiration of 20 ml of 0.9% NaCl from the bronchoscope. The samples were frozen in sterile containers and stored at -80 °C. The same amount of NaCl liquid was also collected as for blank control.

The cancer tissues and matched para-cancerous tissues from additional 28 patients with NSCLC were included for determination of abundance of *A. arborescens*. The cancerous and para-cancerous tissues were separated by surgical scissors and put in the sterile lyophilization tubes. PBS was used as the control. All samples in the study were collected in sterile conditions. The details of clinical data are presented in Table 3.

Table 2 🛛	Cancer	stage
-----------	--------	-------

Staging and Grading Cancer	Quantitative digital	
Non-NSCLC	0	
Tia	1	
T1aN0M0	2	
T1bN0M0	3	
T1cN0M0	4	
T2aN0M0	5	
T2bN0M0	6	
T2cN0M0	7	
T3aN0M0	8	
T3bN0M0, T2bN1M0	9	
T3cN0M0, T3N1M0, T4N0M0	10	

Table 3 Clinical Characteristics of patients for testing A.

 arborescens enrichment

Feature	NSCLC patients	
Genders	Male	15
	Female	13
Age	<u>≤</u> 50	13
	>50	15
Smoking status	Smoker	5
	Non-smoker	23
TNM staging($n = 28$)	Tis	7
	T1aN0M0	1
	T1bN0M0 T1cN0M0 T2aN0M0 T3N0M0 T4N0M0 T4N2M0 T4N1M1a T1cN2M0	4 3 7 1 1 1 1 1
	T1bN0M1a	1

All patients are immunocompetent in the cohort

Metagenomic sequence

The DNA was isolated from samples, and Nanodrop (2000/2000c, American) was used to quantify the DNA. The DNA with 1.8 < A260/A280 < 2.0 and concentration > 20 ng/uL was used for following metagenomics sequencing. The blank controls and isolation kits were used as negative controls to quantify the DNA, but they didn't pass the threshold for sequencing. The metagenomic sequencing was performed using the paired-end sequencing method on the Illumina platform (San Diego, CA, USA). The DNA was sheared by ultrasonication (Covaries, Woburn, MA). The sheared DNA fragments were end-repaired (DNA End Repair Mix) at 20 °C for 30 min. The DNA fragments were purified by QIAquick PCR Purification Kit (Qiagen) and A-tailed using A-Tailing Mix. Libraries were checked using Bioanalyzer 2100 (Agilent) and quantified using the ABI StepOnePlus Real-Time PCR System. Libraries were sequenced on an Illumina platform.

Estimation of lung fungal content

To estimate the lung fungal content, a qPCR approach based on 'FungiQuant' was used [23]. The FungiQuant primers were FungiQuant-F: 5'-GGRAAACTCACC AGGTCCAG-3' and FungiQuant-R: 5'-GSWCTATC CCCAKCACGA-3' yielding products of approximately 350 bp in the fungi 18S rRNA gene. E. penicillium DNA extracted from DNeasy UltraClean Microbial Kit (12,224 Qiamgen, Germany) was used to establish a standard curve in tenfold serial dilutions with 10 ng - 1 pg in each run. BAL samples were used to extract total fungal DNAs (QIAamp UCP Pathogen Mini Kit 50,214, Qiagen, Germany). The optimized conditions included the reaction mixture (20 µl) for qPCR contained ChamQ Universal SYBR Color qPCR Master Mix (Vazyme Biotech, Jiangsu, China), forward and reverse primer (final concentration 400 nM), the templet DNA and molecular-grade water, with all reactions performed in triplicates on the 7900HT Real Time PCR System (Applied Biosystems). We used the following PCR conditions: 3 min at 50 °C for UNG treatment, 10 min at 95 °C for Taq activation, 15 s at 95 °C for denaturation, and 1 min at 65 °C for annealing and extension \times 50 cycles. Fungal content of each BAL sample was calculated according to the standard curve.

Fungal taxonomic profile

Raw sequences were processed to remove low-quality sequences using fastp [24] (Version 0.21.0). Fastuniq [25] (Version 1.1.0) was used to remove duplicates in paired short DNA sequence reads in a FASTQ format. Human sequences were filtered out using the human reference genome (hg37) by bowtie2 [26] (Version 2.3.5).

The remaining high-quality reads were used for taxonomic classification by Kraken2 [27] (Version 2.0.7). We used 'kraken-build' tools to build the fungi-kraken2 database (kraken2-build –download-library fungi), and all fungal nucleotide sequences from NCBI (www.ncbi. nlm.nih.gov) were included. Next, the Bracken was used to obtain the read count for different fungi to estimate relative abundance. The read counts table of several levels (e.g., phylum, class, order, family, genus, species) were rarefied to the minimum fungal read counts to reduce the effects of uneven sampling in the cohort by using R package 'picante' [28] (Version 1.8.2). After the above processing of the raw sequences, we obtained 41,728 reads in each sample, which were annotated to a total of 341 fungal species.

Fungal diversity in the cohort and rarefaction curve analysis

The Shannon–Wiener index and Gini-Simpson index were calculated to determine the fungal alpha diversity. In addition, the Bray–Curtis dissimilarity indices between samples at the species level were calculated to estimate beta diversity by using R package 'vegan' [29]. The permutational multivariate analysis of variance (PERMANOVA) and principal co-ordinates analysis (PCoA) were performed to estimate the between sample (β) diversity. Rarefaction analysis was performed to assess the fungi richness in the NSCLC patients and non-NSCLC. For a given number of samples, we performed random sampling 1000 reads in the cohort with replacement and estimated the total number of species that could be identified from these samples by the richness estimator.

Co-occurrence network construction and analysis

To explore the different fungal correlations between NSCLC and Non-NSCLC groups, the co-occurrence networks were constructed based on the relative abundances of different fungi at the species level by using FastSpar [30] (Version 1.0). Only robust (r > 0.8 or r < -0.8) and statistically significant (p < 0.05) correlations were incorporated into network analysis. Network visualization in Fig. 2 and network parameters (i.e., degree, betweenness, diameter, and cluster coefficient) analysis were made with Gephi (version 0.9.2) using the undirected network (where edges have no direction) and the 'Fruchterman-Reingold' layout. Afterwards, we used UMAP algorithm in python 'umap' library to display all data into one plane. The two networks were parsed into modules in the 'igraph' R package (Version 1.2.6). After modularizing, the networks were re-visualized according to module attributes and the two sub-networks were extracted by using Cytoscape (Version 3.8).

Regression model construction

In order to study the relationship between various fungi with the development of NSCLC, we used two regression models in machine learning. The Ordinary Least Square (OLS) model by the function 'lm' in R was used to estimate adjusted R^2 based on cancer stage and relative abundance for each fungus. To investigate the effect of the inclusion of multiple independent variables (fungi) on the regression model, we first reduced the different multiple variables to two variables using the PCA algorithm. The results show that two-dimensional data produced by PCA can explain more than 99.95% of the explainable variance. Next, we introduced a ridge regression model to eliminate the effect of covariance between multiple variables on the model. The PCA and ridge regression was used by python 'sklearn' library.

Key fungi detection between NSCLC and Non-NSCLC groups

The rarefied read counts of fungal taxa at species level were used to achieve Random forests by the R package 'rfPermute' (Version 2.1.81), and different species were identified by R package 'DESeq2' (Version 1.26.0) between NSCLC and Non-NSCLC groups.

Test for A. arborescens enrichment

To test enrichment of A. arborescens, we used the nested PCR on 28 newly diagnosed patients with NSCLC. Nested PCR involves two sequential amplification reactions, each of which uses a different pair of primers. The first amplification primers are the paired primers were the same as the primers (FungiQuant) for estimating fungal content above. The product of the first amplification reaction is used as the template for the second PCR, which is primed by oligonucleotides that are placed internal to the first primer pair. And the second primers are specific for A. arborescens primers which were designed using Primer blast for amplification. Meanwhile, all primers were validated using gradient qPCR to detect the annealing temperature and the specificity of primers. Primers were obtained from Sangon Biotech (Shanghai, China). The primers for A. arborescens were AA-F: 5'- CAAATATGAAGGCGGGCTGGA-3', AA-R: 5'-TGT CCTAGTGGTGGGCGAAC-3'.

Appropriate tissue samples including paired cancer tissue and adjacent cancer tissue were used to extract total fungal DNAs (Qiagen Blood & Tissue Kit, GER). At the same time, an equal volume of PBS DNA was extracted as the negative control and the DNA of Malassezia globosa was extracted as the positive control. The first round of amplification was performed in a 50 µl reaction mix containing 25 µl Premix Taq DNA polymerase (Takara, Dalian, China), 1.2uM for each primer and 5ul of template DNA. To minimize air-borne contamination, all steps were performed in a class 2 laminar flow safety cabinet. The temperature profile for amplification was as follows: initial denaturation at 94 °C for 4 min, denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, and extension at 72 °C for 1 min, for 35 cycles, followed by a final extension at 72 °C for 10 min. Two percent AGAR gel electrophoresis was used (120v 35 min) and the FungiQuant DNA fragments were cut under the UV light and extracted through QIAquick Gel Extraction Kit (Qiagen, GER), whose construction was detected by Nanodrop. In the step of PCR, an equal amount of DEPC treated water was also used to replace the template DNA, which was used to exclude the contamination caused by the PCR experimental system. The reaction mixture (20 μ l) for PCR contained ChamQ Universal SYBR Color PCR Master Mix (Vazyme Biotech, Jiangsu, China), forward and reverse primer (final concentration 400 nM), and the extracted FungiQuant DNA fragments (5 ng). The Vazymecycling program was 40 cycles and consisted of 95 °C for 10 s and 56 °C for 30s and 72 °C for 1 min with an initial cycle of 50 °C for 2 min and 95 °C for 2 min. All steps in nested PCR were described in Additional file 1: Fig. S1. In addition, to test for contamination in the test for A. arborescens enrichment operation, a negative control (possible contamination in the PCR and environment) and a positive control were set up in the pre-amplification step and examined by using agarose gel electrophoresis. In the nested PCR process, a negative control was also set up (containing possible contamination during the sample preservation, preamplification and PCR process), and all results showed that there was no contamination that could affect the results during the test (Additional file 1: Fig. S2).

Assuming that for all templates and primers a cycle equally doubles the number of template DNA, the relative abundance of a certain strain (i) can be calculated as follows:

Relative abundance (i) =
$$\frac{(\frac{1}{2})^{CT_i}}{(\frac{1}{2})^{CT_c}} = (\frac{1}{2})^{CT_i - CT_c} = (\frac{1}{2})^{\Delta CT}$$

The cycle threshold of strain i primer and common primer (total fungus) are represented by CTi and CTc, while Δ CT denotes the difference between them. From the equation, the logarithm of the relative abundance negatively correlates linearly with Δ CT.

Other statistical analysis and data visualization

For all statistical analysis and prediction models, python 3.8.5 and R 3.6.1 were used. The baseline data in Table 1 were obtained by using R package 'compareGroups'. The data visualization process in this article were implemented in python library 'matplotlib' or R package 'ggplot2'.

Results

Greater fungal diversity in patients with NSCLC

To explore the mycobiome difference between NSCLC and non-NSCLC groups, we performed metagenomic sequencing in BAL from these two groups (Fig. 1A). The fungal composition in the cohort was dominated by the species *Lasiodiplodia theobromae* and *Malassezia globosa*, representing 64.62% and 11.83% of the fungi, respectively, followed by *Grosmannia clavigera* (4.7%), *Botrytis fragariae* (2.8%) and *Aspergillus flavus* (2.3%). Thus, about 85% of fungi were covered by the top five most abundant species in both Non-NSCLC and NSCLC patients. Among these five species, *Botrytis fragariae* was more abundant in patients with NSCLC than Non-NSCLC (Additional file 1: Fig. S3A). The species rarefaction curve for each sample was performed to approach saturation, indicating that the sequencing depth was adequate and samples in NSCLC exhibited higher richness (Additional file 1: Fig. S3B). Indeed, the higher fungal α diversity in patients with NSCLC was confirmed



Fig. 1 Characterization of lung mycobiome. (**A**) The scheme of the total research design. The *a* diversity between NSCLC and Non-NSCLC described using Shannon–Wiener index (**B**) and Gini-Simpson index (**C**), at species level respectively (Wilcoxon test). (**D**) PCoA based on the Bray–Curtis dissimilarity index shows the β diversity, in which the blue circles and orange triangles represent NSCLC and Non-NSCLC, respectively (PERMANOVA, p = 0.0012). (**E**) The relationship between Gini-Simpson index in different samples and cancer stage, which is calculated by the least-squares linear regressions, with 95% confidence intervals (gray-shaded areas)



Fig. 2 Networks of co-occurrence fungi across two groups. (**A**) The relative abundance of each fungi is used to construct the co-occurrence networks by sparce method. Only the nodes with correlation coefficient (r > 0.8 or r < -0.8 significant at p < 0.05) are considered. The nodes are colored according to phylum. Green edges represent positive correlations and red edges represent negative correlations. Node size is proportional to the betweenness centrality of each species, and edge thickness is proportional to the weight of each correlation. (**B**) UMAP analysis shows each node status from NSCLC to Non-NSCLC. The two plots show different nodes in NSCLC and the Non-NSCLC networks from left to right, which paired to Fig. 2A. Interactive node means the node of which degree is not zero. (**C**) The network parameter is calculated in two networks. Total degree is the sum of edges on each node, representing the number of other nodes (species) in the network which are connected with the given node. Diameter is the largest distance between two nodes in a network. Clustering coefficient shows the extent by which a node is connected to its neighbors. Path length represents the nearest distance between two nodes

by the Shannon–Wiener index and Gini-Simpson index (p=0.013) (Fig. 1B-C). The complex compositions of fungi were visualized on a two-dimensional plane using PCoA analysis and PERMANOVA test, demonstrating the significantly different fungal community composition between the two groups (p=0.0012) (Fig. 1D). The α diversity elevated with the progress of NSCLC (Fig. 1E).

Furthermore, the fungal communities in patients with NSCLC showed a significantly higher β -diversity (p=0.003) (Additional file 1: Fig. S3C). The qPCR was also performed to check the fungal content in the lung, and it showed increased fungal content in patients with NSCLC (Additional file 1: Fig. S3D).





Adjusted ridge regression R² : 0.1099

Fig. 3 Four sub-networks extracted from Non-NSCLC co-occurrence network and NSCLC co-occurrence network. (**A**) Two sub-networks extracted from Non-NSCLC network, which contain two interactive nodes (blue font) only occurred in Non-NSCLC network but not in NSCLC network. (**B**) Two sub-networks extracted from NSCLC network, which pair with the two networks in Fig. 3A. In the Fig. 3, the nodes with red font are the interactive nodes only occurring in NSCLC network and the nodes with black bold font are first neighbor connected with blue font nodes. The grey font nodes represent interactive nodes in both networks. And the numbers under each node labels are the adjusted R² for every species in Ordinary Least Square (OLS) model

More complex co-occurrence fungi network in patients with NSCLC

To study the community differences of fungi between the two groups, we constructed two co-occurrence networks of NSCLC and non-NSCLC groups. The degree distributions of the fungal co-occurrence networks conformed to the power-law distribution, indicating that the fungal community was constructed in a non-random way (Additional file 1: Fig. S4). The pattern of co-occurrence was more obvious in the NSCLC network, and the majority of taxa belonged to *Ascomycota* phylum in both networks (Fig. 2A).

To better observe the distribution of fungi in the cooccurrence network in the overall sample, the UMAP algorithm was used to downscale the relative abundance data. All the fungi are displayed in a two-dimensional plane in Fig. 2B. Some kinds of fungi marked in red points represented the exclusive taxa in the NSCLC cooccurrence network, which concentrated in the lower left corner of the plane after downscaling. Meanwhile, two kinds of fungi marked in blue points presented the exclusive taxa in the Non-NSCLC network. Compared to the Non-NSCLC group, fungal communities in the NSCLC group had a more complex network with higher edges (n=176), a higher total degree (n=352), and a higher clustering coefficient (0.779, Fig. 2C).

Fungal community structure associated with development of NSCLC

We tracked two kinds of fungi (Trichosporon asahii and Schizosaccharomyces octosporus) that were isolated (degree was zero) in the co-occurrence networks after the transition of pattern from the Non-NSCLC to NSCLC, and extracted the sub-networks of them (Fig. 3A-B). In the two sub-networks, Fusarium pseudograminearum and Malassezia globsa were as interactive nodes connecting other fungi. With the disruption of co-occurrence relationship between T. asahii and F. pseudograminearum, more exclusive fungi in the NSCLC co-occurrence network uncovered. Furthermore, with the disruption of co-occurrence relationship between S. octosporus and M. globosa, this situation occurred equally. To explore the effect of fungal community alteration on the development of NSCLC, we constructed a model through ridge regression, which suggested that altered fungal community had a greater potential for the progression of NSCLC (adjusted R^2 from -0.0518 to 0.1099). Notably, the contribution to the development of NSCLC (adjusted R^2) for each fungus did not increase with the degree which is the size of edges for them (Additional file 1: Fig. S5A). It suggested that the fungi isolated in the networks cannot be ignored either.

Alternaria arborescens as key fungi related to NSCLC

To detect key fungi between the patients with NSCLC and the non-NSCLC, we used two different methods to reduce the randomness and inaccuracy of the algorithm. 17 and 20 distinct kinds of fungi were screened out by DESeq2 and Random Forest algorithm, respectively. After combining the two results, 5 kinds of fungi (A. arborescens, Eremoyces bilateralis, Aureobasidium namibiae, Tilletiopsis washingtonensis and Paraphaeosphaeria sporulosa) were detected as key fungi (Fig. 4A). In addition, the OLS model was used to predict the effect of correlation between the single fungus and NSCLC development. There were 45 kinds of fungi closely relative to NSCLC development (Fig. 4B), and these key fungi were significantly enriched with the NSCLC development (Fig. 4C, p = 0.0012). Not only was A. arborescens as the key fungus obtained by both algorithms, but also as the fungus with largest adjusted R^2 to the progress of NSCLC in our OLS model. It was detected that its relative abundance was gradually increasing as NSCLC progressed (Fig. 4D). Although other three key fungi showed relationship (Additional file 1: Fig. S5B-D), none of them performed a great gradual upward trend. To test the generality of the correlation between A. arborescens and NSCLC, we performed the nested PCR on another cohorts (n=28), and the results also showed *A. arbores*cens significantly enriched in cancer tissues compared to peri-carcinoma tissues (Fig. 4E, p = 0.00005).

Discussion

In this study, the higher fungal diversity and more complex network were observed in patients with NSCLC compared to Non-NSCLC group. Machine learning model based on the fungi relative abundance and cancer progress was built, and we found the alteration of fungi community relative to the NSCLC. In addition, *A. arborescens* was detected as the most relative fungus with NSCLC, and also showed significantly higher relative abundance in cancer tissues compared to that of para-cancerous tissues. These data suggest an association between a distinct human mycobiome and cancer in the lung.

Firstly, we provided the description of lung mycobiota composition, and the top two abundant kinds

(See figure on next page.)

Fig. 4 Detection of the relative key fungi with NSCLC. (**A**) Two methods are used to find key differential fungi between two groups. The volcano plot demonstrates the differential abundance of fungi between NSCLC and Non-NSCLC. Points are colored according to the number of log2 fold change with passed thresholds (p < 0.05 and |log2 fold change|>1) which are calculated by DEseq2. The bar chart of key fungi between two groups which is determined by Random forest model (p < 0.05). (**B**) The fungi associated with NSCLC progress. We sorted each fungus by the adjusted R^2 calculated from OLS model. Only the fungi with p < 0.05 and FDR < 0.2 are showed. (**C**) A venn graph shows the comprehensive results of the three methods. The hypergeometric test is used to check if the five key fungi detected by two methods (DEseq2 and Random Forest) enrich with the NSCLC development. (**D**) A line chart shows the variation between the log₁₀ mean relative abundance of the *A. arborescens* and cancer stage. (**E**) Another cohort is used to confirm the enrichment of *A. arborescens* in the tissue of NSCLC (Paired Wilcoxon test, p = 0.00005)



Fig. 4 (See legend on previous page.)

of fungi are *L. theobromae* and *M. globosa* in both groups. Although *L. theobromae* is a fungus as yet undescribed as a common resident of the lung microbiome, there are many fungi that do not have apparent

host-specificity and are rather ubiquitous [31, 32]. Moreover, it is reported *L. theobromae* is associated with human health [33, 34]. *M. globosa* is also a skin-associated microbe and it was found relative to cancer

in the recent studies [35, 36]. Secondly, the increased diversity of mycobiota was observed in the lung environment of patients with NSCLC, and the alpha-diversity of mycobiota in NSCLC patients was positively associated with clinical stages. There is a strong link between lung cancer, microbes, and inflammatory status [36–39]. Although the higher bacterial diversity was usually considered as benefits in the gut [40], the diversity of fungi increases significantly after infection [20]. This increased diversity was also detected in patients with Crohn's disease [37, 41].

Co-occurrence network offers an approach to explore the microbial community structure, maintenance and dynamics [7], which can be applied to statistically explore the taxa that are highly connected in the community [42, 43]. As shown in the Fig. 2, the co-occurrence network analysis of fungi in this study suggested the variation in fungal communities and the enhanced complexity of interactions among fungi in patients with NSCLC. Meanwhile, in the UMAP analysis, most of the fungi which were added to the NSCLC network in red color were clustered around these two isolated blue nodes. This is consistent with the results in our extracted sub-networks analysis in Fig. 3. Compared to S. octosporu, there are more exclusive taxa in the NSCLC co-occurrence network occurred, when T. asahii was isolated from the Non-NSCLC network. F. pseudograminearum and M. globosa are not only normally detected in the environment but also detected in BAL samples [44]. In our study, they are also interactive nodes linked to altered nodes between two groups. It reported that the common microbial composition of lungs contains environment-microorganism and the composition of respiratory microbiota is influenced by environmental factors [45, 46].

Furthermore, the construction of machine learningbased regression models helps us to better understand the impact of different fungi and changing co-occurrence networks on the progression of NSCLC. As one of the key fungi, A. arborescens had the most significant correlation with the progression of NSCLC. Although A. arborescens mainly cause disease in plants, the airborne spores they produce can invade human respiratory tracts and cause respiratory and lung diseases. A. arborescens is one species of Alternaria genus. Alternaria is one kind of fungi that can also lead to infection in humans, and its spores are one of the most effective air allergens [47–50]. In addition, Alternaria is an opportunistic fungus, which can infect immunocompromised patients [51]. A recent study revealed that Alternaria can influence the development of cancer by affecting the immune system [36, 52]. Notably, several papers have reported the enrichment of microbes in cancer tissues and their possible role in the development of cancer by affecting immune and infection status [53-55]. In our study, another batch of samples from tumor and paracancerous tissues was also used to detect A. arborescens enrichment in cancerous tissues of NSCLC, and this enrichment increased credibility and avoided the bias of the results due to fungal infection. The reasons of alteration of lung microbiome can be varied and complex. This change can be mainly caused by microbial migration, elimination and growth rates [56-58]. It has also been reported that microbial metabolites of the gut may cause the immune system alterations thus affecting the respiratory microbiota [11, 59, 60]. Although microbes might directly affect protumorigenic pathways in epithelial cells [7], it is still unclear to determine the causal roles of fungal alterations.

Nevertheless, there are some limitations in our study. For instance, larger clinical cohorts need to be covered in the future. In addition, some microbial alterations in the lung caused by other conditions may compromise the results. Therefore, functional studies on mice are also needed to confirm our proposed effect of altered fungal community on NSCLC. Overall, our present study has taken the steps toward bringing a new perspective to elucidate the potential relationship between NSCLC and mycobiota. Based on detected fungal signatures, novel targeted treatment modalities in personalized medicine may emerge with the modification or restoration of a healthy fungal community in patients with cancer.

Conclusions

In this study, we focused on the alteration of lung mycobiome in patients with NSCLC. Compared to Non-NSCLC group, the mycobiome in the lungs of patients with NSCLC showed greater fungal diversity. By using ridge regression model, we indicated the fungal co-occurrence network structure may be associated with NSCLC progress. In addition, *A. arborescens* was detected as the most relative key fungus to NSCLC, and we also found it enriched in the cancer tissues. Our study provides invaluable insights into further exploration in the relationship between NSCLC and fungi.

Abbreviations

NSCLC	Non-small-cell lung cancer
BAL	Bronchoalveolar lavage
OLS	Ordinary Least Square
HMP	Human Microbiome Project
PCoA	Principal co-ordinates analysis
PERMANOVA	Permutational multivariate analysis of variance
UMAP	Uniform Manifold Approximation and Projection

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12866-023-02790-4

Additional file 1: Fig. S1 This is the scheme from sample processing to nested PCR. Firstly, the DNA of the tissue samples, Fungus positive control and the environment blank control (EBC) are extracted. Next, we performed the first amplification by PCR. And after that, the Agarose gel electrophoresis demonstrates contamination-free amplification. Lane 1: DNA marker, Lane 2: The 18S rRNA fragement DNA of Malassezia globosa, which is the positive control. Lane 3 - 8: Six clinical samples were randomly selected for testing. Lane 9: The DNA of PBS, which is the negative control. In addition, 18S fragments including EBC (located at 300bp) are cut with UV light and extracted DNA again. At last, we performed the nested PCR to estimate the relative abundance of microbes. In this step, we additionally included the Non-template control group (NTC) used to exclude nested PCR reagent contamination. Fig. S2 It shows the amplification and melting curves of the nested PCR. The graphs from top row represent the amplification curve of clinical samples and negative control, respectively. As for graphs from bottom row, they are melting curves. The two peaks in the left column represent the carcinoma tissue and the paracarcinoma group. There is no effective peak on the right column which means Negative control (EBC and NTC group). Fig. S3 (A) barplot shows the fungal composition of both NSCLC and Non-NSCLC groups. The top 15 fungi in terms of relative abundance were shown in the picture, and other were classified as 'Others'. (B) Species rarefaction curves in red and black indicate NSCLC and non-NSCLC groups, respectively. (C) Differences in beta-diversity between the mycobiome in patients with NSCLC and non-NSCLC groups were estimated based on a Bray-Curtisdistance matrix of all 38 samples (Wilcoxon test, p = 0.003). (D) The fungal content of patients with NSCLC and Non-NSCLC groups by using qPCR (t test, p =0.031). Fig. S4 The degree distribution for co-occurrence networks in NSCLC and Non-NSCLC group, respectively. The *p*-values are calculated by using permutation test. Fig. S5 (A) The plot only shows the fungi of which degree is not zero in NSCLC co-occurrence network, and different points mean different fungi's adjusted R². Based on their distribution, we fit these points to a trend line, and calculated p value and R². The p-values are calculated by permutation test. (B, C and D) The line charts show the variation between the \log_{10} mean relative abundance of different fungi and cancer stage.

Acknowledgements

We thank Jing Huang, Kaijuan Wu, Xuan Peng, Jing Li, Siqi Yao, Jingyan Wang, and Xingyu Zhai for their constructive comments on the manuscript.

Authors' contributions

JQT, ZY, JJX and JW designed the experiments, JQY carried out experiments, YMZ prepared the figures. YMZ drafted the manuscript. LYC, WZ, MHW participation in discussion and revised the manuscript. All authors contributed to this manuscript. All authors read and approved the final manuscript.

Funding

This work was funded by the National Natural Science Foundation of China (81972198, 81773147, 32000054, 32170071), Strategic Priority Research Program of Central South University (ZLXD2017004); Key Research and Development Program of Hunan (2019SK2253).

Availability of data and materials

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive in National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences, under accession number CRA006566, which publicly accessible at https://bigd.big.ac.cn/gsa.

Declarations

Ethics approval and consent to participate

The research presented here has been performed in accordance with the Declaration of Helsinki, and cases enrolled in this study were collected and approved by the ethical review committees from the second Xiangya hospital, Central South University, China. The patients were informed about the sample collection and had signed informed consent forms.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 5 September 2022 Accepted: 9 February 2023 Published online: 25 March 2023

References

- 1. Lloyd-Price J, Abu-Ali G, Huttenhower C. The healthy human microbiome. Genome Med. 2016;8(1):51.
- Dickson RP. The microbiome and critical illness. Lancet Respir Med. 2016;4(1):59–72.
- Maddi A, Sabharwal A, Violante T, Manuballa S, Genco R, Patnaik S, Yendamuri S. The microbiome and lung cancer. J Thorac Dis. 2019;11(1):280–91.
- 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.
- 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.
- Zheng L, Sun R, Zhu Y, Li Z, She X, Jian X, Yu F, Deng X, Sai B, Wang L, et al. Lung microbiome alterations in NSCLC patients. Sci Rep. 2021;11(1):11736.
- Ramirez-Labrada AG, Isla D, Artal A, Arias M, Rezusta A, Pardo J, Galvez EM. The Influence of Lung Microbiota on Lung Carcinogenesis, Immunity, and Immunotherapy. Trends Cancer. 2020;6(2):86–97.
- Tsay JJ, Wu BG, Badri MH, Clemente JC, Shen N, Meyn P, Li Y, Yie TA, Lhakhang 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):1188–98.
- 9. Kuper H, Adami HO, Trichopoulos D. Infections as a major preventable cause of human cancer. J Intern Med. 2000;248(3):171–83.
- Elinav E, Nowarski R, Thaiss CA, Hu B, Jin C, Flavell RA. Inflammationinduced cancer: crosstalk between tumours, immune cells and microorganisms. Nat Rev Cancer. 2013;13(11):759–71.
- Segal LN, Clemente JC, Tsay JC, Koralov SB, Keller BC, Wu BG, Li Y, Shen N, Ghedin E, Morris A, et al. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. Nat Microbiol. 2016;1:16031.
- Gosens R, Hiemstra PS, Adcock IM, Bracke KR, Dickson RP, Hansbro PM, Krauss-Etschmann S, Smits HH, Stassen FRM, Bartel S. Host-microbe crosstalk in the lung microenvironment: implications for understanding and treating chronic lung disease. Eur Respir J. 2020;56(2):1902320.
- Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, Flores SC, Fontenot AP, Ghedin E, Huang L, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care Med. 2013;187(10):1067–75.
- Cui L, Morris A, Ghedin E. The human mycobiome in health and disease. Genome Med. 2013;5(7):63.

- Vallianou N, Kounatidis D, Christodoulatos GS, Panagopoulos F, Karampela I, Dalamaga M. Mycobiome and Cancer: What Is the Evidence? Cancers (Basel). 2021;13(13):3149.
- Tillonen J, Homann N, Rautio M, Jousimies-Somer H, Salaspuro M. Role of yeasts in the salivary acetaldehyde production from ethanol among risk groups for ethanol-associated oral cavity cancer. Alcohol Clin Exp Res. 1999;23(8):1409–15.
- 17. Han YW, Wang X. Mobile microbiome: oral bacteria in extra-oral infections and inflammation. J Dent Res. 2013;92(6):485–91.
- Marttila E, Bowyer P, Sanglard D, Uittamo J, Kaihovaara P, Salaspuro M, Richardson M, Rautemaa R. Fermentative 2-carbon metabolism produces carcinogenic levels of acetaldehyde in Candida albicans. Mol Oral Microbiol. 2013;28(4):281–91.
- Moritani K, Takeshita T, Shibata Y, Ninomiya T, Kiyohara Y, Yamashita Y. Acetaldehyde production by major oral microbes. Oral Dis. 2015;21(6):748–54.
- Li J, Chen D, Yu B, He J, Zheng P, Mao X, Yu J, Luo J, Tian G, Huang Z, et al. Fungi in gastrointestinal tracts of human and mice: from community to functions. Microb Ecol. 2018;75(4):821–9.
- Richard ML, Liguori G, Lamas B, Brandi G, de Costa G, Hoffmann TW, Pierluigi Di Simone M, Calabrese C, Poggioli G, Langella P, et al. Mucosaassociated microbiota dysbiosis in colitis associated cancer. Gut Microbes. 2018;9(2):131–42.
- 22. Coker OO, Nakatsu G, Dai RZ, Wu WKK, Wong SH, Ng SC, Chan FKL, Sung JJY, Yu J. Enteric fungal microbiota dysbiosis and ecological alterations in colorectal cancer. Gut. 2019;68(4):654–62.
- Liu CM, Kachur S, Dwan MG, Abraham AG, Aziz M, Hsueh PR, Huang YT, Busch JD, Lamit LJ, Gehring CA, et al. FungiQuant: a broad-coverage fungal quantitative real-time PCR assay. BMC Microbiol. 2012;12:255.
- Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
- Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S. FastUniq: a fast de novo duplicates removal tool for paired short reads. PLoS ONE. 2012;7(12):e52249.
- 26. Langdon WB. Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks. Biodata Mining. 2015;8(1):1.
- Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biology. 2014;15(3):R46.
- Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO. Picante: R tools for integrating phylogenies and ecology. Bioinformatics. 2010;26(11):1463–4.
- Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. 2003;14(6):927–30.
- Watts SC, Ritchie SC, Inouye M, Holt KE. FastSpar: rapid and scalable correlation estimation for compositional data. Bioinformatics. 2019;35(6):1064–6.
- Dickson RP, Huffnagle GB. The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease. PLoS Pathog. 2015;11(7):e1004923.
- Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, Bushman FD, Collman RG. Topographical continuity of bacterial populations in the healthy human respiratory tract. Am J Respir Crit Care Med. 2011;184(8):957–63.
- Papacostas LJ, Henderson A, Choong K, Sowden D. An unusual skin lesion caused by Lasiodiplodia theobromae. Med Mycol Case Rep. 2015;8:44–6.
- Woo PC, Lau SK, Ngan AH, Tse H, Tung ET, Yuen KY. Lasiodiplodia theobromae pneumonia in a liver transplant recipient. J Clin Microbiol. 2008;46(1):380–4.
- Crespo Erchiga V, Ojeda Martos A, Vera Casano A, Crespo Erchiga A, Sanchez Fajardo F. Malassezia globosa as the causative agent of pityriasis versicolor. Br J Dermatol. 2000;143(4):799–803.
- Alam A, Levanduski E, Denz P, Villavicencio HS, Bhatta M, Alhorebi L, Zhang Y, Gomez EC, Morreale B, Senchanthisai S, et al. Fungal mycobiome drives IL-33 secretion and type 2 immunity in pancreatic cancer. Cancer Cell. 2022;40(2):153-167 e111.
- Wheeler ML, Limon JJ, Bar AS, Leal CA, Gargus M, Tang J, Brown J, Funari VA, Wang HL, Crother TR, et al. Immunological consequences of intestinal fungal dysbiosis. Cell Host Microbe. 2016;19(6):865–73.
- Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Gao Z, Chen H, Berger KI, Goldring RM, Rom WN, et al. Enrichment of lung microbiome

with supraglottic taxa is associated with increased pulmonary inflammation. Microbiome. 2013;1(1):19.

- 39. Grivennikov SI, Greten FR, Karin M. Immunity, inflammation, and cancer. Cell. 2010;140(6):883–99.
- Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A. 2011;108(Suppl 1):4578–85 Suppl 1.
- Wu X, Xia Y, He F, Zhu C, Ren W. Intestinal mycobiota in health and diseases: from a disrupted equilibrium to clinical opportunities. Microbiome. 2021;9(1):60.
- 42. Fuhrman JA. Microbial community structure and its functional implications. Nature. 2009;459(7244):193–9.
- Berry D, Widder S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Front Microbiol. 2014;5:219.
- Rubio-Portillo E, Orts D, Llorca E, Fernandez C, Anton J, Ferrer C, Galvez B, Esteban V, Revelles E, Perez-Martin C, et al. The Domestic Environment and the Lung Mycobiome. Microorganisms. 2020;8(11):1717.
- Yun Y, Srinivas G, Kuenzel S, Linnenbrink M, Alnahas S, Bruce KD, Steinhoff U, Baines JF, Schaible UE. Environmentally determined differences in the murine lung microbiota and their relation to alveolar architecture. PLoS ONE. 2014;9(12):e113466.
- Man WH, de Steenhuijsen Piters WA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. Nat Rev Microbiol. 2017;15(5):259–70.
- Anaissie EJ, Bodey GP, Rinaldi MG. Emerging fungal pathogens. Eur J Clin Microbiol Infect Dis. 1989;8(4):323–30.
- Zureik M, Neukirch C, Leynaert B, Liard R, Bousquet J, Neukirch F. European community respiratory health s: sensitisation to airborne moulds and severity of asthma: cross sectional study from european community respiratory health survey. BMJ. 2002;325(7361):411–4.
- Lawrence DP, Gannibal PB, Peever TL, Pryor BM. The sections of alternaria: formalizing species-group concepts. Mycologia. 2013;105(3):530–46.
- Pinto VE, Patriarca A. Alternaria Species and their associated mycotoxins. Methods Mol Biol. 2017;1542:13–32.
- Acland KM, Hay RJ, Groves R. Cutaneous infection with Alternaria alternata complicating immunosuppression: successful treatment with itraconazole. Br J Dermatol. 1998;138(2):354–6.
- Hu W, Ran Y, Zhuang K, Lama J, Zhang C. Alternaria arborescens infection in a healthy individual and literature review of cutaneous alternariosis. Mycopathologia. 2015;179(1–2):147–52.
- Peters BA, Hayes RB, Goparaju C, Reid C, Pass HI, Ahn J. The microbiome in lung cancer tissue and recurrence-free survival. Cancer Epidemiol Biomarkers Prev. 2019;28(4):731–40.
- Guo H, Zhao L, Zhu J, Chen P, Wang H, Jiang M, Liu X, Sun H, Zhao W, Zheng Z, et al. Microbes in lung cancer initiation, treatment, and outcome: boon or bane? Semin Cancer Biol. 2022;86(Pt 2):1190–206.
- Nejman D, Livyatan I, Fuks G, Gavert N, Zwang Y, Geller LT, Rotter-Maskowitz A, Weiser R, Mallel G, Gigi E, et al. The human tumor microbiome is composed of tumor type-specific intracellular bacteria. Science. 2020;368(6494):973–80.
- Gately S. Human microbiota and personalized cancer treatments: role of commensal microbes in treatment outcomes for cancer patients. Cancer Treat Res. 2019;178:253–64.
- Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. Lancet. 2014;384(9944):691–702.
- Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, Beck JM, Curtis JL, Huffnagle GB. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. mBio. 2015;6(2):e00037.
- 59. Honda K, Littman DR. The microbiota in adaptive immune homeostasis and disease. Nature. 2016;535(7610):75–84.
- Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, Glickman JN, Siebert R, Baron RM, Kasper DL, et al. Microbial exposure during early life has persistent effects on natural killer T cell function. Science. 2012;336(6080):489–93.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.