# RESEARCH

**Open Access** 

# Distinctive circulating microbial metagenomic signatures in the plasma of patients with lung cancer and their potential value as molecular biomarkers

Huijuan Chen<sup>1,2</sup>, XueNa Yao<sup>1</sup>, Chunyan Yang<sup>3,4,5</sup>, Yiran Zhang<sup>3</sup>, Henan Dong<sup>4</sup>, Jincheng Zhai<sup>4</sup>, Dongjie Fan<sup>6\*</sup> and Qiming Zhou<sup>5\*</sup>

# Abstract

Lung cancer (LC) remains the leading cause of cancer death globally. Recent reports have suggested that circulating microbial nucleic acids have potential as promising biomarkers for cancer liquid biopsies. However, circulating microbial profiles and their potential clinical value in LC patients remained unexplored. In this study, plasma samples from 76 LC patients, 9 liver cancer patients, 11 pancreatic cancer patients, and 53 healthy controls (HCs) were collected and underwent metagenomic analyses by whole genome sequencing. The composition and relative abundance of the microbial profiles were significantly different between the LC patients and HCs. A distinct plasma-based microbial profile was observed in LC patients. By differential analysis using MaAslin, 40 significant species between LC patients and HCs were identified. Five species were selected as optimal circulating microbial biomarkers for LC. The constructed classifier based on these five species showed an AUC of 0.9592, 0.9131, and 0.8077 in the discovery, validation, and additional validation cohorts, respectively. Furthermore, metagenomic profiles of 25 lung tumor tissue and plasma paired samples were analyzed and compared. The microbial diversity was significantly increased in plasma compared with the tumor tissue. Among the 13 shared core microbial species, 10 had no difference between the tumor tissue and paired plasma. In conclusion, circulating microbial nucleic acids in the plasma have potential as biomarkers for LC liquid biopsies. The microbiome in the tumor tissue was one of the possible sources of circulating microbial nucleic acids.

Keywords Lung cancer, Plasma, Whole genome sequencing, Circulating microbial nucleic acids, Biomarkers

\*Correspondence: Dongjie Fan fandongjie@icdc.cn Qiming Zhou genbank@vip.sina.com Beijing CapitalBio Medlab Co. Ltd, Beijing 100176, China <sup>3</sup> Beijing SeqTide Gene Technology Co. Ltd, Beijing 10017, China

- <sup>2</sup> South China Hospital of Shenzhen University, Shenzhen 518111, China
- <sup>4</sup> Beijing Best HealthCare Medical Technology Co. Ltd, Beijing 100176,

China <sup>5</sup> School of Life Science and Technology, Harbin Institute of Technology,

92 West Dazhi Street, Harbin 150001, China

<sup>6</sup> National Key Laboratory of Intelligent Tracking and Forecasting for Infectious Disease, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China



© The Author(s) 2025. Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

# Introduction

Lung cancer (LC) remains the leading cause of malignancy-related deaths worldwide, accounting for approximately 20% of all cancer-related mortality due to its high incidence and late-stage diagnosis [1]. The use of multiple screening tools to diagnose LC at an early stage is an important opportunity to reduce LC mortality. Several randomized control trials, such as the Nederlands Leuvens Screening Onderzoek [2] and the National Lung Screening Trial [3], have shown lower LC-related mortality in patients with low-dose computed tomography (LDCT) screening than no screening or chest radiography. LDCT has been recommended by the National Comprehensive Cancer Network guideline for LC screening [4, 5]. However, the clinical applicability of LDCT is limited by the screening program cost, the radiation dose to the individual, and other factors. LC screening with participant risk stratification or personalized screening intervals could substantially reduce the burden and improve the efficiency of current LC screening. Molecular biomarkers, especially blood-based molecular biomarkers, might have a role in refining selection criteria and improving risk stratification for LC screening [6-9]. In the Asian population, integrating biomarkers could be more important than in other populations because of the high incidence of LC in non-smokers.

Biomarkers for cancer are being developed at a rapid rate with the use of highly sensitive technologies, such as next-generation sequencing (NGS) and omics technologies. Multiple plasma-based molecular biomarkers, such as circulating tumor DNA [10], circulating tumor cells [11], and circulating cell-free RNA [12], have been identified as potential biomarkers for early cancer detection. Serving as a reservoir, cell-free nucleic acids carry human genetic information from all cells, including the non-human genetic information derived from the microbiome, including the cancer microbiome [13, 14]. Multiple studies have identified highly divergent circulating microbial profiles in a variety of non-communicable diseases, including cardiovascular disease [15], autoimmune disease [16], liver disease [17], and cancer [18]. Living bacteria have been found in tumor tissue, and each tumor type had a distinct intra-tumor microbiome composition of bacteria and fungi [19, 20]. Moreover, emerging evidence showed that circulating microbial DNA (cmDNA) was significantly different between tumor patients and healthy controls (HCs) [21]. Theoretically, cmDNA or circulating nucleic acids have potential as promising circulating biomarkers for cancer liquid biopsies. The characteristic composition patterns of cmDNA have been identified, and the potential of circulating microbial profiles for cancer detection has been evaluated in various solid tumors [18, 22-24]. Moreover, several investigations demonstrated different obvious alterations in the bacterial flora of LC patients in different specimens, such as in bronchoalveolar lavage fluid (BALF) of LC, in which levels of *Veillonella* and *Megasphaera* increased significantly [25]. In sputum, *Granulicatella adiacens* and *Streptococcus intermedius* increased [26]. However, few studies have explored the relevance between blood-based circulating microbial metagenomic signatures and LC. Characterizing the circulating microbiome and exploring its source and potential clinical value in LC patients are of great value and interest.

In this study, we analyzed and characterized the circulating microbial metagenomic profile in real-world Chinese LC patients through metagenomic analysis using whole genome sequencing (WGS). Furthermore, we explored the potential of circulating microbial species as biomarkers for LC and selected five species as the optimal marker set using a random forest model. These species showed the potential to effectively differentiate the LC from healthy individuals. In addition, for the first time, we deciphered and compared the microbial profile of surgical tumor tissue and matched plasma in 24 real-world LC patients. We anticipate our results will help provide the scientific evidence support for exploring the potential source of circulating nucleic acids in LC patients.

# **Material and methods**

### Sample collection and study design

This retrospective study included 76 LC patients and 20 other cancer patients (9 liver cancer patients and 11 pancreatic cancer patients) who underwent NGS testing at Chosenmed Lab (Beijing, China) from March 2021 to October 2023. The cancer was diagnosed through the integration of clinical symptoms, imaging, and laboratory testing and confirmed using gold standard pathological examinations by board-certified pathologists, and the Tumor-Node-Metastasis (TNM) staging system was used for staging of LC. For cancer patients, the inclusion criteria were as follows: 1) no evidence of systemic inflammation, such as fever, C-reactive protein; 2) no antibiotics used within 4 weeks before blood sample collection; 3) patients newly diagnosed with one primary cancer and no active preoperative treatment course. The exclusion criteria were: 1) patients under the age of 18; 2) patients with multiple primary cancers; 3) patients that had previously undergone treatment for cancer; 4) patients were pregnant.

Peripheral blood samples (approximately 10 mL) from cancer patients were collected before surgical resection. 25 fresh LC tumor tissue samples were collected concurrently to explore the correlation of the metagenomic profile between LC tumor tissue and plasma. Fifty-three healthy volunteers with normal physical examination results were enrolled as controls. Before drawing blood, the skin surface was sterilized twice using a 0.5% povidone-iodine solution, and peripheral blood was collected into 10 mL cell-free DNA BCT<sup>®</sup> blood collection tubes (Streck, La Vista, NE, USA) according to the user manual. Plasma was prepared within 72 h using a previously reported method [27]. Written informed consent according to the Declaration of Helsinki was obtained from each participant. This study was carried out following The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans and it was approved by the Ethics Committee of Xuanwu hospital (No. [2019]081-R1).

The study was designed as shown in Fig. 1. The samples were divided into a discovery cohort, a validation cohort, and an additional validation cohort. The additional validation cohort consisted of 5 HCs, 5 LC patients, and 20 other cancer patients. The 71 other LC patients and 48 HCs were randomly divided into a discovery cohort and a validation cohort in a 70%:30% ratio. Twenty-five tissue and plasma paired samples were used to explore the correlation of the microbial metagenomic profile between LC tumor tissue and plasma.

A total of 149 peripheral blood samples were collected from 76 lung cancer patients, 53 healthy controls, 9 liver cancer patients, and 11 pancreatic cancer patients. After plasma metagenomic analysis, 5 lung cancer, 5 healthy control, and 20 other cancer samples were selected as an additional validation cohort (the lung cancer and healthy control samples were selected randomly). The remaining 71 lung cancer and 48 healthy controls samples were randomly divided into the discovery cohort and validation cohort by stratified sampling splits of 70% and 30% of the data. Meanwhile, 24 tissue and plasma paired samples randomly selected from validation cohort were used to explore the correlation of the microbial metagenomic profile between lung cancer tumor tissue and plasma.

# Total nucleic acid extraction, library construction, and sequencing

Nucleic acid extraction, library construction, and sequencing were performed in a College of American Pathologists accredited NGS laboratory. The total nucleic acid in plasma and tissue was extracted using a VAMNE magnetic pathogen DNA/RNA kit (Vazyme Biotech, Nanjing, China) and TIANMicrobe magnetic patho-DNA/RNA kit (TIANGEN Biotech, Beijing, China), respectively, according to the manufacturers' instructions. Purified nucleic acid was quantified with a Qubit 4.0 fluorometer using a dsDNA HS assay kit and RNA HS assay kit (Thermo Fisher Scientific, MA, USA). The first strand of cDNA was synthesized using a PureScript 1st strand cDNA synthesis kit (low nucleic acid contamination) (Vazyme Biotech, Nanjing, China).

Sequencing libraries of cell-free total nucleic acids were constructed using an Xgen cfDNA and FFPE DNA library prep kit (IDT, IA, USA) following the kit user manual. In brief, 1–50 ng of DNA was processed by end repair, ligation 1 (ligation adapter to 3' ends), ligation 2 (ligation adapter to 5' ends), and PCR amplification. Libraries of nucleic acid from tissue were prepared using a whole genome sequencing library preparation kit (QIAGEN, DUS, Germany) with the following steps: enzymatic fragmentation, end-repair, dA-tailing, adapter ligation, and library amplification. All the sequencing libraries were quantified using a Qubit 4.0 fluorometer and sequenced using an MGISEQ-2000 sequencing platform (BGI, Shenzhen, China) with 100bp paired-end cycles [28].

Three types of negative controls were included in each batch for sequencing analysis: 1) nucleic acid extraction negative controls, which included reagents from the nucleic acid extraction stage through sequencing; 2) library construction negative controls, which included reagents from the library preparation stage through sequencing; and 3) empty negative control wells, which included UltraPure DNase/RNase-free distilled water from the nucleic acid extraction stage and the library preparation stage through sequencing.

#### Bioinformatic analysis and data decontamination

Raw sequencing data were quality filtered, demultiplexed, and adapter-trimmed using fastq [29]. Clean reads of each sample were mapped to the human genome GRCh38/hg38 using Bowtie 2 with the fastlocal parameter set [30]. Reads were discarded if either mate mapped to hg38, mitochondrial genomes, or bacterial plasmids. The filtered reads were mapped to the microbial reference genome database using the Kraken algorithm with default parameters [31]. A total of 28,330 microbial genomes were in the microbial reference genome database, which was constructed from prior literature [21]. Of the 28,330 microbial genomes, 13,346 were bacterial, 504 were fungal, and 14,480 were viral. The relative abundance at the genus level and species level were estimated for each sample using the Bracken method. All taxa for which the relative abundance was less than 0.01% or the read count was less than 10 were filtered as sequencing or analysis artifacts. Furthermore, we used both the prevalence (negative control-based) and frequency (concentration-based) modes in *decontam* [32] to decontaminate the Kraken count microbial data. A P\*=0.5 hyperparameter value was used for both modes.



Fig. 1 Study design and flow diagram

# Identification of the circulating microbial markers and construction of the classifier model

MaAslin software was used to identify the differentially abundant microbial species (q value < 0.05) between LC patients and HCs in the discovery cohort. MaAslin software was used to identify the relative abundance of each feature while adjusting for covariables such as age and sex. A random forest classifier was used to predict LC based on the microbial profiles. The classifier model was trained by tenfold cross-validation repeated five times using the carpet R package. The top five microbial species predicting LC were selected as the optimal microbial species for the biomarker set. The receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) values were generated and calculated to evaluate the performance of the random forest classifier model using pROC [33].

# Statistical analyses

The rarefaction analysis between sample size and the number of microbial species was conducted using the R package *amplicon* [34]. Microbial community diversity was determined by the Simpson index or Shannon index, which were calculated by the *vegan* package. Nonmetric multidimensional scaling (NMDS) and principal coordinate analysis (PCoA) were performed using *vegan*. Microbial taxonomic analyses and comparisons between both groups at the genus and species level were performed using the Wilcoxon rank-sum test and visualized using the *pheatmap* package. All the above statistical analyses were performed using R version 4.3.0.

Differences between subjects and sequencing data of the LC and HC samples were compared using *t* tests and either Chi-square or Fisher's exact test using Prism 8 for Windows (GraphPad Software, CA, USA).

# Results

### Characteristics of the studied cohort

A total of 159 plasma samples and 25 LC fresh tumor tissue samples were obtained from the patients and volunteers enrolled in this study. The samples were subjected to WGS analysis. After quality control, the sequencing data of 149 plasma samples (comprising samples from 76 LC patients (24 with paired tumor tissue samples), 53 HC patients, 9 liver cancer patients, and 11 pancreatic cancer patients) and 24 fresh tissue samples were finally subjected to bioinformatic analysis. The patients corresponding to the 149 plasma samples were randomly divided into a discovery cohort, a validation cohort, and an additional validation cohort (Fig. 1).

There was no significant difference between LC patients and HCs in terms of gender and age in all three

groups (Table S1). In the discovery cohort, most LC patients were stage II (20/50, 40%) and stage III (15/50, 30%) patients. Stage information of seven patients was unknown. In the validation cohort, the majority of LC patients were stage I–II patients (15/21, 71.42%), and two patients had unknown stages (Table S1).

# Diversity of circulating microbial metagenomic profile in LC patients and healthy controls

After sequencing data quality control and filtering, the LC plasma samples had  $1.15 \times 10^8$  total clean reads/ sample. An average of 0.57% (0.22%-3.60%) of total reads did not map to the human genome, and 14.62% (9.48%-47.42%) of these non-human mapped reads instead mapped to our microbial database (0.081% of total reads), providing 89,854 microbial reads/sample for downstream analysis. Of these microbial reads, 55.86% were classified as bacterial, 35.85% as fungal, and 1.07% as viral (Table S2). In the healthy controls, an average of  $1.14 \times 10^8$  total clean reads was obtained for each plasma sample. However, 1.57% (0.75%-2.69%) of these clean reads did not map to the human genome, and 4.33% (3.29%-5.04%) of these non-human mapped reads instead mapped to our microbial database (0.068% of total reads), providing 75,368 microbial reads/sample for downstream analysis. Of these microbial reads, 58.87% were classified as bacterial, 32.93% as fungal, and 0.94% as viral (Table S2). There were no significant differences in total sequence reads or microbial reads, including bacterial, fungal, and viral, between the LC and HC groups (Table S2, Figure S1). A rarefaction analysis showed that the species richness nearly approached saturation in both the LC and HC groups (Fig. 2A and B).

After a series of decontamination filters were applied as detailed in the Methods section, a total of 563 species were identified in the plasma samples of LC patients and HCs, and 123 were shared in both LC patients and HCs. The LC and HC groups had 427 and 13 unique species, respectively (Fig. 2C). The observed species number was significantly higher in the LC group than in the HC group (LC: median 60 (49-113) species/sample vs. HC: median 49 (40–98) species/sample, *P*<0.0001) (Fig. 2D). The Shannon diversity showed no difference between the LC and HC groups (P = 0.082 > 0.05) (Fig. 2E). However, the Simpson diversity in the LC group was higher than that in the HC group (mean Simpson index in the LC and HC groups were 0.69 and 0.72, respectively; P = 0.0048 < 0.05, Fig. 2F). In addition, the NMDS and PcoA based on the species relative abundance of each sample revealed that the circulating microbial metagenomic profile was different between the LC and HC groups (Fig. 2G and H).



Fig. 2 Comparison of circulating microbial metagenomic profiles between lung cancer patients and healthy controls. Rarefaction curve in LC (A) and HC (B) samples. C A Venn diagram displaying the overlap between LC patients and HCs showed that 123 of the 563 total species were shared by both groups, while 427 were unique in LC patients, and 13 were unique in HCs. D Number of detected species in each sample. E Shannon diversity index and F Simpson diversity index were computed from all LC and HC samples. G The NMDS and H the PcoA results based on the relative abundance of the detected species showed that the circulating metagenomic profile was significantly different between LC patients and HCs. LC, lung cancer; HC, healthy control

# Circulating microbial metagenomic profile differs between lung cancer patients and healthy controls

We further analyzed the circulating microbial composition and relative abundance in both the LC and HC groups. The composition and abundance of the microbial metagenomic profile in each LC sample at the species level are shown in Fig. 3A. A distinct circulating microbial profile was observed in LC patients. *Klebsiella pneumoniae, Lasiodiplodia theobromae, Salmonella enterica, Cutibacterium acnes,* and *Staphylococcus aureus* were the five most abundant species in LC samples, which accounted for 61.15% of the total microbial relative abundance (Fig. 3A). We also compared the circulating microbial composition between LC and HC samples at the genus and species level. At the genus level, the top 10 most abundant genera accounted for 71.65% and 51.66% of total microbial relative abundance in LC and HC samples, respectively. *Klebsiella, Salmonella, Cutibacterium, Corynebacterium, Staphylococcus,* and *Malassezia* were significantly more abundant in LC samples than in HC samples, whereas *Pseudomonas, Neisseria, Ralstonia,* and *Pasteurella* were less abundant (all P < 0.05) (Figs. 3B and S1). At the species level, six microbial species, *Klebsiella pneumoniae, Salmonella enterica, Cutibacterium acnes, Staphylococcus aureus, Corynebacterium amycolatum,* and *Brevibacterium pigmentatum,* were significantly



Fig. 3 Distinct circulating microbial metagenomic profiles in the plasma of lung cancer patients and healthy controls. A Descriptive visual representation of microbial taxa at the species level in LC samples. The microbial relative abundance in LC and HC samples at the genus (B) and species (C) levels. D The relative abundance of the top 10 most abundant species in LC and HC samples

more abundant in LC samples than in HC samples, while three species, *Neisseria gonorrhoeae*, *Pasteurella multo-cida*, and *Burkholderia pseudomallei*, were less abundant (all P < 0.05) (Fig. 3C and D).

# Circulating microbial metagenomic panel as a potential novel diagnostic biomarker for patients with lung cancer

In the discovery cohort, MaAslin software was used to further compare the microbial features in LC patients versus those in HCs based on the relative abundance. Twenty species, including *Microvirgula aerodenitrificans*, *Komagataella phaffii*, *Staphylococcus aureus*, *Alternaria incomplexa*, and *Pseudomonas phenolilytica*, were significantly enriched in LC. The top five of the 20 most significantly enriched species in HCs were *Delftia sp WY8*, Burkholderia pseudomallei, Sphingomonas sp., Alternaria arbusti, and Alternaria conjuncta (Fig. 4A). Furthermore, a random forest classifier model between LC and HC samples was constructed to explore the potential circulating microbial biomarkers for LC. Five species, Microvirgula aerodenitrificans (M. aerodenitrificans), Komagataella phaffii (K. phaffii), Alternaria incomplexa (A. incomplexa), Ogataea philodendri (O. philodendri), and Staphylococcus aureus (S. aureus), were selected as the optimal microbial markers for LC by a tenfold crossvalidation of the random forest model repeated five times (Fig. 4B and C).

The ROC curve was used to identify the ability of the five selected species to discriminate between LC and non-LC patient samples. In the discovery cohort, we



**Fig. 4** MaAslin analysis and a random forest model showing the potential of circulating microbes in blood as diagnostic biomarkers. **A** The MaAslin analysis showed that different microbes were enriched in LC and HC samples (all P < 0.05). **B** Five microbial markers were selected as the optimal marker set by tenfold cross-validation in the random forest model. **C** Volcano plot showing the fold change in relative abundance versus the – log (*p*) value of the selected five species in the discovery cohort. Receiver operating curve (ROC) for the discovery cohort (**D**), validation cohort (**E**), and additional validation cohort (**F**)

were able to achieve an AUC of 0.9592 (95% CI 0.8825– 0.9869) between LC patients and HCs (Fig. 4D). An AUC value of 0.9131 with a 95% CI of 0.7694–0.9620 was achieved between LC and HC samples in the validation cohort (Fig. 4E). Furthermore, 5 HC, 5 LC, and 20 other cancer samples (9 liver cancer and 11 pancreatic cancer) formed an additional validation cohort to validate the classifier model for LC. An AUC value of 0.8077 (95% CI 0.6425–0.8342) was observed (Fig. 4F). These data showed that the constructed model based on the circulating microbial biomarkers was able to distinguish LC patients from HCs and those with other types of cancer, which indicated that the circulating microbes in blood had the potential to be non-invasive diagnostic biomarkers for patients with LC.

# Comparison of microbial metagenomic profiles between lung cancer tumor tissue and plasma

Twenty-five paired LC tissue and blood samples were analyzed by WGS to explore the correlation of the microbial metagenomic profile between tumor tissue and plasma. Twenty-four paired samples were included for downstream analysis, and one paired sample was excluded for insufficient raw data from the tissue sample. There was no difference in total sequencing reads between the tissue and plasma samples (Figure S2). However, the non-human reads, microbial reads, and microbial counts, including bacterial, fungal, and viral, were significantly higher in tissue than in plasma (Figure S2).

A total of 514 species were identified in the 24 paired plasma and tissue samples; 66 were shared by both plasma and tissue samples, and 405 and 43 species were unique to plasma and tissue samples, respectively (Fig. 5A). As estimated by the Shannon index (Fig. 5B), the microbial diversity was significantly increased in plasma compared with tissue (P<0.0001). The NMDS analysis showed that there were overlaps in the microbial composition between plasma and tissue (Fig. 5C). The

top 10 most abundant species in plasma and tissue samples were selected, and their relative abundance in each sample is presented in a heatmap (Fig. 5D). The average relative abundance and composition of the top 10 most abundant genera and species in both tissue and plasma groups are shown in Fig. 5E and F, respectively. At the genus level, the top 10 most abundant genera accounted for 91.5% and 52.76% of the total microbial relative abundance in tissue and plasma, respectively. *Staphylococcus, Salmonella, Bacillus, Pasteurella,* and *Ogataea* were significantly more abundant in LC tissue than in paired plasma samples, whereas *Corynebacterium* was less abundant (all P < 0.05) (Figure S3A). Among the top



**Fig. 5** Comparison of microbial metagenomic profiles between lung cancer plasma and tumor tissue samples. **A** Venn diagram. **B** The Shannon diversity index was computed from all 24 tissue and plasma paired samples. **C** NMDS analysis results. **D** The microbial relative abundance and composition in each LC tissue and plasma sample. The average microbial relative abundance and composition in LC tissue and plasma samples. **G** Volcano plot showing a fold change in relative abundance versus the – log (*p*) value of the core microbial species in the 24 paired samples

10 most abundant species in tissue, five species demonstrated a difference in relative abundance between tissue and paired plasma samples. Four species, *Salmonella enterica, Pasteurella multocida, Bacillus thuringiensis,* and *Ogataea philodendri,* were more abundant in tissue than in paired plasma samples, whereas one species, *Corynebacterium amycolatum,* was less abundant (all P < 0.05) (Figure S3B).

We defined the core microbial species if they were observed and shared in 80% of samples [35, 36]. A total of 26 core microbial species were observed in the 24 paired plasma and tissue samples, 13 were shared by both plasma and tissue samples, and 10 and 3 species were unique in plasma and tissue samples, respectively (Table S3). Among the 26 core microbial species, the differential microbial relative abundance based on MaAs-Lin identified 12 species significantly enriched in plasma samples and three species in tissue samples (Fig. 5G). Moreover, among the 13 shared core microbial species, 10 showed no difference between the tissue and paired plasma samples.

### Discussion

With the rapid progression of omics technologies, cancer-related biomarkers for predicting, diagnosing, and prognosticating are being developed and discovered at a rapid rate and are mainly human genomic and proteomic profiles. In recent years, exploration of the cancer microbiome and circulating microbial nucleic acids, especially cmDNA, has produced an emerging avenue for cancerrelated biomarker discovery. Several NGS approaches are capable of profiling circulating microbial nucleic acids, including targeted PCR detection of specific species, 16S rRNA, and shotgun sequencing [37]. Metagenomic analyses using WGS provide a unique opportunity to explore the microbiome using non-human sequencing reads. These are increasingly used to characterize microbial cfDNA in the bloodstream [21, 38], which allow the circulating metagenomes comprising bacteria, fungi, and viruses to be easily deciphered. In this study, we first comprehensively characterized the blood-based circulating microbial profile in LC patients using metagenomic analyses by WGS. Notably, we further elucidated the microbial profiles of 24 surgical LC tumor tissue and plasma paired samples using the same approach.

Reduced microbial diversity has been considered a feature of disease states in various diseases, including cancer in some studies [39–43]. However, there is still no consensus. An increased microbiota diversity was observed in several studies [44–46]. However, most of those studies were focused on the gut microbiome or microbiota in the tissue using 16S rRNA. In a study based on the BALF of LC patients, increased microbial diversity was observed in BALF samples from cancer patients compared with individuals with benign mass-like lesions [47]. In our study, the Shannon diversity showed no difference between LC and HC samples, but the Simpson diversity was higher in LC samples than in HC samples. Our study demonstrated that the circulating microbial composition and relative abundance in patients with LC were different from those of HCs. In terms of the composition of circulating microbial profiles, Klebsiella pneumoniae, Lasiodiplodia theobromae, Salmonella enterica, Cutibacterium acnes, and Staphylococcus aureus were the five species with the five highest relative abundances in LC. All five species are pulmonary opportunistic pathogens or other pathogens [48-50]. Among the top 10 genera and species in terms of relative abundance in LC and HC samples, the abundance of five genera and six species was significantly increased in the LC group, especially that of K. pneumoniae (Fig. 3). The exact microbial taxa identified in LC samples varied in studies depending on the sample type. The abundance of the oral microorganisms Veillonella and Capnocytophaga significantly increased in the saliva samples of LC patients [51], while Granulicatella adiacens and six other opportunistic pathogens increased in sputum samples [36], and Veillonella, Megasphaera increased in BALF samples [47].

Machine learning (ML) is a powerful tool for identifying new potential microbial biomarkers associated with specific cancer types. ML approaches, especially the random forest algorithm, have been widely used in building cancer prediction models based on microbiome data from various cancer types [52-57]. The microbial taxa abundance is assumed to be different in cancer states and HCs and is the most commonly used feature of microbiome data. In our study, five species, M. aerodenitrificans, K. phaffii, A. incomplexa, O. philodendri, and S. aureus were selected as new potential circulating microbial biomarkers for LC using the random forest algorithm based on the relative abundance of the microbial profile. M. aerodenitrificans is a denitrifying Gram-negative organism. M. aerodenitrificans may give rise to clinical disease, particularly in immunocompromised patients [58]. A higher relative abundance was observed in the fecal samples of the umbilical cord blood transplantation engraftment failure inflammatory bowel diseases patients with IL10RA deficiency than engraftment success patients [59]. K. phaffii is a yeast species and a commonly used alternative host for manufacturing therapeutic proteins. K. phaffii-induced intestinal necrosis has been reported, long-term immunosuppressant exposure is one of its predisposing factors [60]. A. incomplexa and O. philodendri are fungi, which may be linked to allergic lower respiratory tract diseases [61]. S. aureus as a common human pathogen, associates with various infections. Chronic inflammation caused by *S. aureus* may link to an elevated risk of cancers by creating a tumor-promoting microenvironment [62]. However, the exact biological processes of *S. aureus* involved in the cancer development and progression are still needed to explore. The classifier model based on these biomarkers showed excellent performance in both the validation and additional validation cohorts.

The source of the circulating microbial nucleic acids in cancer patients is rarely reported and remains unexplored. Hypotheses have been proposed based on the knowledge about the source of cfDNA and ctDNA. Passive release of endogenous tumor microbial nucleic acids following the death of cancer cells, resulting from apoptosis, necrosis, active secretion of vesicles containing microbial nucleic acids, translocation of the intestinal microbiome, translocation of the cancer microbiome along with tumor metastasis were the potential main sources of circulating nucleic acids in cancer patients [63]. Herein, our study demonstrated that the total number of species and the microbial diversity were significantly decreased in LC tissue compared with those of the paired plasma samples. Ten of 13 paired samples that shared core microbial species showed no difference between the tissue and paired plasma. Our findings suggested that the tumor intra-microbiome may be one of the sources of circulating microbial nucleic acids. However, this concept still needs further investigation to explore and confirm the origin of circulating nucleic acids and the mechanisms of their leakage into the bloodstream.

Although our study has yielded intriguing results, a number of limitations remain. First, this was a retrospective analysis with a limited sample size, and a prospective study with cross-regional, a larger sample size is needed to validate the findings of this study. Second, the plasma microbial signatures were not identified according the histological subtypes and disease stages of LC. The pathological type, stage and other clinical characteristics of lung cancer may have an impact on the lung microbiome. And, we did not investigate the history of smoking and daily diet. The microbiome profiling differed significantly between smokers and neversmokers [64], and the daily diet may provoke the change of lung microbiome through gut-lung axis [65]. Considering more clinical factors, such as tumor histology, performance status, comorbidities, smoking history, is clearly beneficial for comprehensive understanding the lung cancer's relationship with microbial signature and drawing more evidence-based conclusions. Third, even though we have identified significant microbes' circulating nucleic acids in the plasma as diagnosis biomarkers of LC, the mechanistic role of these microbes in LC development and the origin of these microbes circulating nucleic acids remain unclear, and the potential of circulating microbial nucleic acids as prognostic biomarkers of LC is still unknown. Therefore, further in-depth studies are essential to decipher the precise mechanistic relationship between each microbe and LC development, and, to explore the potential of circulating microbial nucleic acids as prognostic biomarkers of LC.

# Conclusions

In this study, we characterized the circulating microbial profile of plasma samples in patients with LC based on metagenomic analyses using WGS. The composition and abundance of the microbial profile were significantly different between the LC and the HC groups. Five microbial species were identified as an LC prediction biomarker panel using a random forest approach, which showed good performance in distinguishing LC patients from HCs. This indicates that circulating microbial nucleic acids have the potential to be biomarkers for LC liquid biopsies. Furthermore, we illustrated the characteristics of the microbial profile of the tumor tissue and matched plasma samples from LC. Ten of 13 shared core microbial species had no difference between the tumor tissue and matched plasma samples, although reduced microbial species numbers and diversity were observed in the tumor tissue. This result partly illustrates the origin of circulating microbial nucleic acids in cancer patients and provides a new way to explore the source of circulating microbial nucleic acids in cancer patient.

#### Abbreviations

LC	Lung cancer
HC	Healthy control
LDCT	Low-dose computed tomography
NGS	Next-generation sequencing
cmDNA	Circulating microbial DNA
BALF	Bronchoalveolar lavage fluid
WGS	Whole genome sequencing
ROC	Receiver operating characteristic
NMDS	Non-metric multidimensional scaling
PCoA	Principal coordinate analysis

### **Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-06209-8.

Supplementary material 1 Supplementary material 2 Supplementary material 3 Supplementary material 4

#### Acknowledgements

Not applicable

#### Author contributions

CHJ designed the study, prepared the figures, wrote the original and revised the manuscript. XNY, CYY, ZYR, DHN and ZJC acquired resource, collected the samples and conducted the analysis. DJF and QMZ supervised the whole study, acquired the funding, reviewed and revised the manuscript. All authors reviewed and approved the manuscript.

#### Funding

This study was supported by Cancer Genome Atlas of (CGAC) project (YCZYPT[2018]06) of the National Human Genetic Resource Sharing Service Platform (2005DKA213000).

#### Availability of data and materials

Sequence data that support the findings of this study have been deposited in the Genome Sequence Archive (GSA) repository with the primary accession code PRJCA022982 (https://bigd.big.ac.cn/gsa-human/browse/HRA006638).

#### Declarations

#### Ethics approval and consent to participate

This study was approved by the Ethics Committee of Xuanwu hospital (No. [2019]081-R1). Written informed consent was obtained from each participant and the methods were carried out in accordance with approved guidelines.

#### **Consent for publication**

Not applicable.

#### **Competing interests**

The authors declare no competing interests.

Received: 4 November 2024 Accepted: 5 February 2025 Published online: 14 February 2025

#### References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71(3):209–49.
- de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, Lammers JJ, Weenink C, Yousaf-Khan U, Horeweg N, et al. Reduced lung-cancer mortality with volume ct screening in a randomized trial. N Engl J Med. 2020;382(6):503–13.
- Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395–409.
- Adams SJ, Stone E, Baldwin DR, Vliegenthart R, Lee P, Fintelmann FJ. Lung cancer screening. Lancet. 2023;401(10374):390–408.
- Senthil P, Kuhan S, Potter AL, Jeffrey Yang CF. Update on lung cancer screening guideline. Thorac Surg Clin. 2023;33(4):323–31.
- Sozzi G, Boeri M, Rossi M, Verri C, Suatoni P, Bravi F, Roz L, Conte D, Grassi M, Sverzellati N, et al. Clinical utility of a plasma-based miRNA signature classifier within computed tomography lung cancer screening: a correlative MILD trial study. J Clin Oncol. 2014;32(8):768–73.
- Sullivan FM, Mair FS, Anderson W, Armory P, Briggs A, Chew C, Dorward A, Haughney J, Hogarth F, Kendrick D, et al. Earlier diagnosis of lung cancer in a randomised trial of an autoantibody blood test followed by imaging. Eur Respir J. 2021;57(1):2000670.
- Seijo LM, Peled N, Ajona D, Boeri M, Field JK, Sozzi G, Pio R, Zulueta JJ, Spira A, Massion PP, et al. Biomarkers in lung cancer screening: achievements, promises, and challenges. J Thorac Oncol. 2019;14(3):343–57.
- 9. Sears CR, Mazzone PJ. Biomarkers in lung cancer. Clin Chest Med. 2020;41(1):115–27.
- Gao Q, Zeng Q, Wang Z, Li C, Xu Y, Cui P, Zhu X, Lu H, Wang G, Cai S, et al. Circulating cell-free DNA for cancer early detection. Innovation (Camb). 2022;3(4): 100259.

- 11. Vasseur A, Kiavue N, Bidard FC, Pierga JY, Cabel L. Clinical utility of circulating tumor cells: an update. Mol Oncol. 2021;15(6):1647–66.
- Cheung KWE, Choi SR, Lee LTC, Lee NLE, Tsang HF, Cheng YT, Cho WCS, Wong EYL, Wong SCC. The potential of circulating cell free RNA as a biomarker in cancer. Expert Rev Mol Diagn. 2019;19(7):579–90.
- 13. Sepich-Poore GD, Guccione C, Laplane L, Pradeu T, Curtius K, Knight R. Cancer's second genome: microbial cancer diagnostics and redefining clonal evolution as a multispecies process: humans and their tumors are not aseptic, and the multispecies nature of cancer modulates clinical care and clonal evolution: humans and their tumors are not aseptic, and the multispecies nature of cancer modulates clinical care and clonal evolution. BioEssays. 2022;44(5): e2100252.
- 14. Pietrzak B, Kawacka I, Olejnik-Schmidt A, Schmidt M. Circulating microbial cell-free DNA in health and disease. Int J Mol Sci. 2023;24(3):3051.
- Szeto CC, Kwan BC, Chow KM, Kwok JS, Lai KB, Cheng PM, Pang WF, Ng JK, Chan MH, Lit LC, et al. Circulating bacterial-derived DNA fragment level is a strong predictor of cardiovascular disease in peritoneal dialysis patients. PLoS ONE. 2015;10(5): e0125162.
- Ho HE, Radigan L, Bongers G, El-Shamy A, Cunningham-Rundles C. Circulating bioactive bacterial DNA is associated with immune activation and complications in common variable immunodeficiency. JCI Insight. 2021;6(19): e144777.
- Lelouvier B, Servant F, Païssé S, Brunet AC, Benyahya S, Serino M, Valle C, Ortiz MR, Puig J, Courtney M, et al. Changes in blood microbiota profiles associated with liver fibrosis in obese patients: a pilot analysis. Hepatology. 2016;64(6):2015–27.
- Cho EJ, Leem S, Kim SA, Yang J, Lee YB, Kim SS, Cheong JY, Cho SW, Kim JW, Kim SM, et al. Circulating microbiota-based metagenomic signature for detection of hepatocellular carcinoma. Sci Rep. 2019;9(1):7536.
- 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.
- Narunsky-Haziza L, Sepich-Poore GD, Livyatan I, Asraf O, Martino C, Nejman D, Gavert N, Stajich JE, Amit G, González A, et al. Pan-cancer analyses reveal cancer-type-specific fungal ecologies and bacteriome interactions. Cell. 2022;185(20):3789-3806.e17.
- Rajpoot M, Sharma AK, Sharma A, Gupta GK. Understanding the microbiome: Emerging biomarkers for exploiting the microbiota for personalized medicine against cancer. Semin Cancer Biol. 2018;52(Pt 1):1–8.
- Huang YF, Chen YJ, Fan TC, Chang NC, Chen YJ, Midha MK, Chen TH, Yang HH, Wang YT, Yu AL, et al. Analysis of microbial sequences in plasma cellfree DNA for early-onset breast cancer patients and healthy females. BMC Med Genomics. 2018;11(Suppl 1):16.
- 23. Xiao Q, Lu W, Kong X, Shao YW, Hu Y, Wang A, Bao H, Cao R, Liu K, Wang X, et al. Alterations of circulating bacterial DNA in colorectal cancer and adenoma: a proof-of-concept study. Cancer Lett. 2021;499:201–8.
- 24. Kim JR, Han K, Han Y, Kang N, Shin TS, Park HJ, Kim H, Kwon W, Lee S, Kim YK, et al. Microbiome markers of pancreatic cancer based on bacteriaderived extracellular vesicles acquired from blood samples: a retrospective propensity score matching analysis. Biology. 2021;10(3):219.
- Chung KS, Kim EY, Jung JY, Kang YA, Kim YS, Kim SK, Chang J, Park MS. Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions. Lung Cancer. 2016;102:89–95.
- Cameron SJS, Lewis KE, Huws SA, Hegarty MJ, Lewis PD, Pachebat JA, Mur LAJ. A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer. PLoS ONE. 2017;12(5): e0177062.
- Chen H, Wang A, Wang J, He Z, Mao Y, Liu L. Target-based genomic profiling of ctDNA from Chinese non-small cell lung cancer patients: a result of real-world data. J Cancer Res Clin Oncol. 2020;146(7):1867–76.
- Chen H, Wang B, Zhang Y, Shu Y, Dong H, Zhao Q, Yang C, Li J, Duan X, Zhou Q. A unified DNA- and RNA-based NGS strategy for the analysis of multiple types of variants at the dual nucleic acid level in solid tumors. J Clin Lab Anal. 2023;37(19–20): e24977.
- Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
- 30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.

- 31. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15(3):R46.
- Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome. 2018;6(1):226.
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
- 34. Liu Y, Chen L, Ma T, Li X, Zheng M, Zhou X, Chen L, Qian X, Xi J, Lu H, et al. EasyAmplicon: an easy-to-use, open-source, reproducible, and community-based pipeline for amplicon data analysis in microbiome research. iMeta. 2023. https://doi.org/10.1002/imt2.83.
- 35. Dai D, Yang Y, Yang Y, Dang T, Xiao J, Wang W, Teng L, Xu J, Ye J, Jiang H. Alterations of thyroid microbiota across different thyroid microhabitats in patients with thyroid carcinoma. J Transl Med. 2021;19(1):488.
- Neu AT, Allen EE, Roy K. Defining and quantifying the core microbiome: challenges and prospects. Proc Natl Acad Sci U S A. 2021;118(51): e2104429118.
- 37. Kataria R, Shoaie S, Grigoriadis A, Wan JCM. Leveraging circulating microbial DNA for early cancer detection. Trends Cancer. 2023;9(11):879–82.
- Dohlman AB, Arguijo Mendoza D, Ding S, Gao M, Dressman H, Iliev ID, Lipkin SM, Shen X. The cancer microbiome atlas: a pan-cancer comparative analysis to distinguish tissue-resident microbiota from contaminants. Cell Host Microbe. 2021;29(2):281-298.e5.
- Ren Z, Fan Y, Li A, Shen Q, Wu J, Ren L, Lu H, Ding S, Ren H, Liu C, et al. Alterations of the human gut microbiome in chronic kidney disease. Adv Sci (Weinh). 2020;7(20):2001936.
- Ferreira RM, Pereira-Marques J, Pinto-Ribeiro I, Costa JL, Carneiro F, Machado JC, Figueiredo C. Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota. Gut. 2018;67(2):226–36.
- Lepage P, Häsler R, Spehlmann ME, Rehman A, Zvirbliene A, Begun A, Ott S, Kupcinskas L, Doré J, Raedler A, et al. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. Gastroenterology. 2011;141(1):227–36.
- Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, Goedert JJ, Hayes RB, Yang L. Human gut microbiome and risk for colorectal cancer. J Natl Cancer Inst. 2013;105(24):1907–11.
- Zaidi AH, Pratama MY, Omstead AN, Gorbonova A, Mansoor R, Melton-Kreft R, Jobe BA, Wagner PL, Kelly RJ, Goel A. A blood-based circulating microbial metagenomic panel for early diagnosis and prognosis of oesophageal adenocarcinoma. Br J Cancer. 2022;127(11):2016–24.
- Zhou L, Li X, Ahmed A, Wu D, Liu L, Qiu J, Yan Y, Jin M, Xin Y. Gut microbe analysis between hyperthyroid and healthy individuals. Curr Microbiol. 2014;69(5):675–80.
- Feng J, Zhao F, Sun J, Lin B, Zhao L, Liu Y, Jin Y, Li S, Li A, Wei Y. Alterations in the gut microbiota and metabolite profiles of thyroid carcinoma patients. Int J Cancer. 2019;144(11):2728–45.
- Zhao F, Feng J, Li J, Zhao L, Liu Y, Chen H, Jin Y, Zhu B, Wei Y. Alterations of the gut microbiota in Hashimoto's thyroiditis patients. Thyroid. 2018;28(2):175–86.
- 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.
- Gu HJ, Kim YJ, Lee HJ, Dong SH, Kim SW, Huh HJ, Ki CS. Invasive fungal sinusitis by Lasiodiplodia Theobromae in an patient with aplastic anemia: an extremely rare case report and literature review. Mycopathologia. 2016;181(11–12):901–8.
- Samonis G, Maraki S, Kouroussis C, Mavroudis D, Georgoulias V. Salmonella enterica pneumonia in a patient with lung cancer. J Clin Microbiol. 2003;41(12):5820–2.
- Abdullah HM, Waqas Q, Abdalla A, Omar M, Berger P. Cutibacterium acnes pneumonia in an immunocompromised patient: a case report and review of the literature. S D Med. 2021;74(11):523–6.
- 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):3111–22.
- Feng J, Yang K, Liu X, Song M, Zhan P, Zhang M, Chen J, Liu J. Machine learning: a powerful tool for identifying key microbial agents associated with specific cancer types. PeerJ. 2023;11: e16304.

- Xu W, Wang T, Wang N, Zhang H, Zha Y, Ji L, Chu Y, Ning K. Artificial intelligence-enabled microbiome-based diagnosis models for a broad spectrum of cancer types. Brief Bioinform. 2023. https://doi.org/10.1093/ bib/bbad178.
- Zheng Y, Fang Z, Xue Y, Zhang J, Zhu J, Gao R, Yao S, Ye Y, Wang S, Lin C, et al. Specific gut microbiome signature predicts the early-stage lung cancer. Gut Microbes. 2020;11(4):1030–42.
- Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, Fleck JS, Voigt AY, Palleja A, Ponnudurai R, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med. 2019;25(4):679–89.
- Baxter NT, Ruffin MT 4th, Rogers MA, Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med. 2016;8(1):37.
- Yang J, Moon HE, Park HW, McDowell A, Shin TS, Jee YK, Kym S, Paek SH, Kim YK. Brain tumor diagnostic model and dietary effect based on extracellular vesicle microbiome data in serum. Exp Mol Med. 2020;52(9):1602–13.
- Murphy ME, Goodson A, Malnick H, Shah J, Neelamkavil R, Devi R. Recurrent Microvirgula aerodenitrificans bacteremia. J Clin Microbiol. 2012;50(8):2823–5.
- Xue A, Qian X, Gao X, Wang P, Wang L, Zheng C, Huang Z, Hu W, Shi J, Huang Y. Fecal microbial signatures are associated with engraftment failure following umbilical cord blood transplantation in Pediatric Crohns disease patients With IL10RA deficiency. Front Pharmacol. 2020;11: 580817.
- Liu M, Sun J, Wang M, Chen H. Partial ileectomy for intestinal *Komaga-taella phaffii* and *Fusobacterium mortiferum* co-infection. Lancet Infect Dis. 2023;23(2):260.
- Knutsen AP, Bush RK, Demain JG, Denning DW, Dixit A, Fairs A, Greenberger PA, Kariuki B, Kita H, Kurup VP, et al. Fungi and allergic lower respiratory tract diseases. J Allergy Clin Immunol. 2012;129(2):280–91.
- Odunitan TT, Apanisile BT, Akinboade MW, Abdulazeez WO, Oyaronbi AO, Ajayi TM, Oyekola SA, Ibrahim NO, Nafiu T, Afolabi HO, et al. Microbial mysteries: Staphylococcus aureus and the enigma of carcinogenesis. Microb Pathog. 2024;194: 106831.
- You L, Zhou J, Xin Z, Hauck JS, Na F, Tang J, Zhou X, Lei Z, Ying B. Novel directions of precision oncology: circulating microbial DNA emerging in cancer-microbiome areas. Precis Clin Med. 2022. https://doi.org/10.1093/ pcmedi/pbac005.
- 64. Ying KL, Brasky TM, Freudenheim JL, McElroy JP, Nickerson QA, Song MA, Weng DY, Wewers MD, Whiteman NB, Mathe EA, et al. Saliva and lung microbiome associations with electronic cigarette use and smoking. Cancer Prev Res (Phila). 2022;15(7):435–46.
- 65. Marsland BJ, Trompette A, Gollwitzer ES. The gut-lung axis in respiratory disease. Ann Am Thorac Soc. 2015;12(2):S150–6.

### **Publisher's Note**

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