Cellular Oncology (2026) 49:69
https://doi.org/10.1007/s13402-026-01193-7

RESEARCH

Bacterial taxa associated with lung cancer cases in Southeast Asians: a pilot case-control study

Adrian Low[1] · Yah Ru Juang[2] · Fransiskus X. Ivan[3] · Lina Ang[2] · Luke Hao Shuan Ooi[3] · Sean Wei Jun Chan[2] · Micheal Mac Aogain[4,5] · Tavleen Kaur Jaggi[3] · Sanjay H. Chotirmall[3,6] · Yann Felix Boucher[2,7,8] · Anthony Chau Ang Yii[9] · Mariko Siyue Koh[10,11] · Darren Wan Teck Lim[11,12,13] · Jonathan Wei Jie Lee[1,14,15] · Wei Jie Seow[1,2]

Received: 28 November 2025 / Accepted: 11 March 2026

Abstract

Purpose Lung cancer is the leading cause of cancer-related mortality worldwide, yet its underlying mechanisms remain unclear. Disruptions in the respiratory microbiome may promote inflammation and carcinogenesis. This study aimed to comprehensively compare genus-level sputum microbiota between lung cancer patients and healthy controls in a multiethnic Southeast Asian population.

Methods Sputum samples were collected from lung cancer patients across three Singapore hospitals. Socio-demographic data were obtained via questionnaire. We analyzed 16S rRNA amplicon sequences from 70 lung cancer patients and 47 healthy controls from a separate local cohort, using identical sequencing protocols to minimize batch effects. Alpha- and beta-diversity metrics, random forest models, and ANCOM-BC2 were used to identify microbial features associated with lung cancer and host characteristics.

Results Lung cancer cases showed significantly reduced genus richness compared to controls. Beta-diversity (Aitchison distance) differed by case-control status, sex, age, and smoking history. ANCOM-BC2 identified *Lactobacillus* as differentially enriched in cases. However, this enrichment did not pass the pseudo-count test among never-smokers. Further sex-stratified analysis revealed that the *Lactobacillus* enrichment was driven primarily by male cases. Conversely, *[Eubacterium] nodatum* group, *Mogibacterium*, and *Campylobacter* emerged as robust inverse signatures for lung cancer, with their depletion consistently supported across stratified and unstratified differential abundance analyses, random forest modeling, and Wilcoxon rank-sum tests. However, these findings may still be subjected to residual confounding by exogenous factors, such as medication use, which could not be adequately adjusted for between case-control groups.

Conclusion This pilot case-control study, conducted in a multiethnic Southeast Asian population, identified distinct respiratory microbiota signatures associated with lung cancer using robust differential abundance and machine-learning methods, providing preliminary evidence for a potential role of the respiratory microbiome to lung carcinogenesis, warranting validation in larger, longitudinal studies.

Keywords Lung cancer · Sputum microbiota · 16S rRNA sequencing · Random forest · Microbial diversity

1 Introduction

Lung cancer is the most commonly diagnosed cancer and the leading cause of cancer-related mortality worldwide, with an estimated 2.5 million new cases and 1.8 million deaths reported in 2022 [1]. While tobacco smoking remains the most well-established risk factor, nearly one-quarter of lung cancer cases occur in never-smokers, and the underlying factors contributing to the development of lung cancer in never-smokers remain unclear [2]. This is particularly relevant in Southeast Asia, where the burden of lung cancer among never-smokers is disproportionately

Adrian Low and Yah Ru Juang contributed equally to this work.

Extended author information available on the last page of the article

Published online: 10 April 2026

high compared to Western populations [3]. These regional differences underscore the need for region-specific studies to better understand the etiology of lung cancer and to identify early detection biomarkers in Southeast Asian cohorts.

Previous studies have described microbial colonization and immunological factors shaping the lung microbiome [4, 5]. Recently, in vivo mouse models have demonstrated that lung microbes can influence immune responses and contribute to lung pathogenesis, such as fibrosis [6, 7]. Although the mechanisms of host-microbiome interaction are still being explored, lung dysbiosis may promote chronic inflammation through IL-17-mediated responses, a known driver of malignancy through DNA damage, increased cellular proliferation, and immune suppression [8, 9]. Hence, identifying taxa associated with lung microbiome perturbation could support the development of novel treatment strategies, particularly those tailored to Asian populations.

Most microbiome studies have focused on pulmonary diseases such as asthma and chronic obstructive pulmonary disease (COPD), and few have examined the microbiomes in lung cancer patients [10, 11]. Leveraging next-generation sequencing technologies, current lung cancer microbiome studies primarily utilize 16S rRNA gene amplicon sequencing to characterize microbial communities [12]. Current evidence suggests that microbial alpha-diversity, encompassing genus richness and abundance, is similarly reduced in lung cancer and other pulmonary diseases when compared to healthy controls [11]. While numerous studies have reported associations between specific microbial taxa and lung cancer, findings remain inconsistent, and there is no consensus on the microbial drivers underlying lung carcinogenesis [7, 10]. For instance, *Rubellimicrobium* and *Fictibacillus* have been reported as enriched in lung tumor tissues compared to healthy controls based on biopsy samples [13]. In bronchial brushings, another study reported significantly higher proportions of *Streptococcus* and *Neisseria* and lower proportions of *Staphylococcus* and *Dialister* in cancer cases relative to controls [14]. In a separate study involving never-smoking Chinese women, higher levels of *Granulicatella*, *Abiotrophia*, and *Streptococcus* were detected in cases versus non-cancer controls [15]. In aggregate, these studies highlight considerable variability in microbial profiles across lung cancer populations. Further research with improved control for intrinsic host factors such as age, body mass index (BMI), and smoking status is needed to clarify these inconsistencies.

Systematic reviews of the existing literature indicate that most lung microbiome studies have been conducted in cohorts from Northern Hemisphere countries, with limited representation from Southeast Asian populations [10, 11]. This study represents the first investigation of the association between the airway microbiome and lung cancer risk in Singapore and the Southeast Asia region, where distinct environmental exposures, lifestyle factors, and host genetics may influence microbiome composition and lung cancer susceptibility, underscoring the importance of conducting region-specific investigations [16]. Specifically, we present a comprehensive analysis of bacterial genus-level diversity in sputum samples, comparing microbial composition among lung cancer patients and between cancer cases and healthy controls, across three major tertiary hospitals in Singapore, a multiethnic city-state whose population composition reflects broader Southeast Asian diversity.

2 Methods

2.1 Biosample processing, DNA extraction and 16S rRNA gene amplicon library construction and sequencing

Sputum samples from lung cancer patients (n = 70) were collected across three hospitals in Singapore: Changi General Hospital (CGH; n = 20), National Cancer Centre (NCC; n = 19) and Singapore General Hospital (SGH; n = 31). The study has been approved by the SingHealth Centralized Institutional Review Board (CIRB Ref: 2018/2524) and informed consent was obtained from all the patients. All samples were thawed on ice for DNA extraction using the High-pure PCR Template Preparation Kit (Roche) following the manufacturer's protocol. Total DNA was quantified using the Qubit 4 fluorometer (Thermo Fisher Scientific, OR, USA) with the dsDNA High Sensitivity Assay (Life Technologies Corporation, Carlsbad, CA, USA). Amplicon libraries targeting the V3–V4 hypervariable regions of the 16S rRNA gene (~ 550 bp) were prepared using Illumina's 16S Metagenomic Sequencing Library Preparation protocol with primers 341F (5'-CCTACGGGNGGCWGCAG) and 785R (5'-GACTACHVGGGTATCTAATCC) for 300 bp paired-end sequencing [17]. Barcoded amplicons were sequenced on the MiSeq sequencer (Illumina) platform. A comparator group of healthy controls (n = 47) was obtained from an independent, local study investigating the association between lung microbiome composition, lung function, and arterial stiffness [18]. Volunteers were deemed healthy based on medical history at the point of recruitment as previously described [18]. The healthy control sputum data were sequenced using the same methods and facility to minimize batch effects [18].

2.2 Sequence processing and microbiota analysis

Paired-end FASTQ files for healthy controls (n = 47; BioProject number PRJNA559069) used the same 16S rRNA

gene primers for sputum microbiota sequencing [18]. Case and control samples were sequenced to depths of 181,719 ± 167,562 (SD) and 146,736 ± 84,368 (SD), respectively (Supplementary Fig. S1). All fastq files for healthy and cancer groups were processed together using QIIME 2 release 2023.7 with default parameters unless stated otherwise [19]. Paired-end reads were denoised, merged and chimera checked using the "qiime dada2 denoise-paired" plugin; forward and reverse reads were truncated at 261 bp and 229 bp, respectively [20]. To remove spurious reads, we filtered amplicon sequences variants (ASVs) shorter than 400 bp, ASVs fewer than 222 reads and are in less than 6% of total samples. Potential PCR contaminants were removed by excluding any ASV with a higher mean read count in the no template controls (n = 6, case; n = 1, control) than in the biological samples (cases or controls). A rarefied table was constructed using scaling with ranked subsampling (SRS'; "qiime srs SRS" plugin) based on the sample with the lowest read frequency (Cmin = 7,730, SGH025) [21]. Rarefaction curves for case-control were calculated using "qiime diversity alpha-rarefaction" command with 100 subsampling iterations to a maximum depth of 35,000 reads and plotted using tidyverse v2.0.0 [22]. The rarefied table was used for alpha- and beta-diversity measures using the commands "qiime diversity alpha" for Shannon, Observed ASVs and Pielou's evenness and "qiime diversity beta" for Aitchison distance. ASVs were classified against a region-specific database (V3-V4 SILVA SSU release 138) using the "qiime feature-classifier classify-sklearn" plugin [23, 24]. The "qiime sample-classifier classify-samples" plugin with random state = 666, estimator = 100, was used to construct the random forest classification model to identify genera that are significantly associated with case-control status [25]. The dataset was automatically split into a training set (80%) for model internal cross-validation and a held-out test set (20%) for independent evaluation of classifier accuracy. Model performance was assessed using the test set by calculating the overall accuracy and the Area Under the Receiver Operating Characteristic (AUROC) curve, which measures the classifier's ability to distinguish between the different cohort groups. The "qiime feature-table filter-samples" command was used to filter samples to the different stratified groups.

2.3 Statistical tests

Beta-diversity and beta-dispersion differences were evaluated using the pairwise permutational analysis of variance (PERMANOVA) and pairwise permutational multivariate analysis of dispersion (PERMDISP) (9,999 permutations), respectively via the respective qiime commands "qiime diversity beta-group-significance" command [26]. To evaluate beta-diversity group significance, the PERMANOVA test was performed using ADONIS "qiime diversity adonis" and the formula=Aitchison distance ~ case-control + age + smoking history + sex + ethnicity [26, 27].

Plots and statistical tests including Wilcoxon rank-sum test with Benjamini Hochberg false discovery rate adjusted p-values (q-values), one-way ANOVA, Chi-square test and Spearman's correlation were performed using RStudio v2024.09.1 + 394 loaded with R v4.4.2 and relevant R libraries tidyverse v2.0.0 for data handling, DT v0.3.3 to handle JavaScript library tables and qiime2R v0.99.6 for heatmaps, principal coordinate analysis (PCoA), boxplots, and stacked bar-plots [22, 28–30]. Inverse probability weighting (IPW) was employed in the comparison of alpha-diversity between cases and controls to adjust for potential confounders. Propensity scores (PS) were estimated via a generalized linear model (GLM) with a logit link function, using cohort status as the outcome and age, sex, and smoking history as covariates. Each sample was assigned a weight equal to the inverse of the probability of receiving its observed cohort assignment: 1/PS for cases and 1/(1-PS) for controls. Inter-group differences were assessed using survey weighted GLMs via the survey package v4.4.8 in R, where a Gaussian family was used for continuous indices (Shannon and Pielou's evenness) and a quasipoisson family was used for Observed ASVs to account for the discrete nature of count data and potential overdispersion [31]. Differential abundance analysis was performed at the genus level using analysis of composition of microbiomes with bias correction v2.8.1 (ANCOM-BC2) method with linear regression and Dunnet's test for multiple-pairwise comparisons, applying Holm-Bonferroni for p-value adjustment [32–34]. Taxa that did not pass the ANCOM-BC2 sensitivity test were regarded as false positives and omitted unless otherwise stated. Non-pseudo-count robust centered-log ratio (RCLR) normalization method based on geometric mean of the observed (non-zero) features was calculated as previously described [35]. To assess the relationship between case-control based on the RCLR normalized data, a linear regression model was fitted using the lm function in R.

3 Results

3.1 Characteristics of the study population

Cases and controls had comparable BMI and ethnic distribution but differed in age, sex, and smoking history (Table 1). Among lung cancer patients, adenocarcinoma and other non-small cell lung cancers were the predominant subtypes, with most patients diagnosed at stage IV (80.9%). Among

Table 1 Characteristics of the study population

Variable	Total ($n=117$)	Cases ($n=70$)	Controls ($n=47$)	p-value
Age, years (mean ± SD)	55.70 ± 15.90	62.00 ± 10.96	46.32 ± 17.54	<0.001[a]
(min – max)	(22.00–84.00)	(36.00–84.00)	(22.00–71.00)	
BMI, kg/m^2 (mean ± SD)	23.25 ± 4.72	23.98 ± 5.40	22.82 ± 3.44	0.423[a]
(min – max)	(16.03–52.52)	(17.26–52.52)	(18.34–30.67)	
Sex (n, %)				
Male	68 (58.1)	52 (74.3)	16 (34.0)	<0.001[b]
Female	49 (41.9)	18 (25.7)	31 (66.0)	
Ethnicity (n, %)				
Chinese	95 (81.9)	51 (73.9)	44 (93.6)	0.096[b]
Malay	16 (13.8)	14 (20.3)	2 (4.3)	
Indian	1 (0.9)	1(1.4)	0 (0)	
Others	4 (3.4)	3 (4.3)	1 (2.1)	
Treatment naïve (n, %)				
No	-	41 (58.6)	-	
Yes	-	29 (41.4)	-	
Smoking history (n, %)				
Ever-smoker	39 (33.3)	36 (51.4)	3 (6.4)	<0.001[b]
Never-smoker	78 (66.7)	34 (48.6)	44 (93.6)	
Lung cancer subtype (n, %)				
Non-small cell lung cancer (NSCLC)[c]	-	56 (80.0)	-	
Adenocarcinoma	-	32 (45.7)	-	
Squamous cell lung cancer	-	4 (5.7)	-	
Other non-small cell cancer	-	20 (28.6)	-	
Small cell lung cancer (SCLC)	-	11 (15.7)	-	
Other/unspecified cell type	-	3 (4.3)	-	
Lung cancer staging (n, %)				
Stage I	-	2 (2.9)	-	
Stage II	-	5 (7.1)	-	
Stage III	-	8 (11.4)	-	
Stage IV	-	55 (78.6)	-	

Abbreviations: SD, standard deviation; BMI, body mass index

[a]One-way ANOVA test

[b]Chi-square test of significance

[c]Includes undifferentiated malignant tumor with uncertain histogenesis; keratinizing squamous cell carcinoma; subtype favor poorly differentiated squamous cell carcinoma; lymphoepithelioma-like carcinoma; mucinous adenoma of lung

lung cancer patients, the distribution of never- and ever-smokers was similar (Table 1).

3.2 Diversity of sputum microbiota differs between case and control groups

The sputum microbial ASV diversity in lung cancer patients ($n=70$) exhibited significantly lower overall genus richness compared to control ($n=47$) group, as evidenced by the Shannon index and observed ASV richness (Fig. 1A-B). However, no significant difference was observed in Pielou's evenness between cases and controls (Fig. 1C). Alpha-diversity did not correlate with increasing BMI or age (Supplementary Fig. S2). Beta-diversity analysis using ADONIS on the Aitchison distance matrix identified case-control status (cancer vs. healthy), sex (male vs. female), age (defined as age at diagnosis for cases or age at sputum collection for controls), and smoking history (never-smoker vs. ever-smoker) as significant contributors to microbiota variation (in order of decreasing effect size) (Fig. 1D). These differences are visualized using principal component analysis (PCoA) plots (Fig. 1E, F, G and H). Notably, beta-dispersion (PERMDISP) differed significantly between never-smokers and ever-smokers, potentially confounding the statistical significance indicated by PERMANOVA (Fig. 1H).

3.3 Random forest microbiota profile models identified bacterial genera that delineate case-control

We performed random forest classifier analysis to identify microbial genera that best distinguish between cases and controls. The top 20 most predictive genera accounted for 53% of the classifier's total importance score (Fig. 2A). The classifier demonstrates an overall accuracy of 87.5% (Cancer = 92.9% and Healthy = 80%) (Supplementary Fig. S3). When comparing relative abundances, 19 of the 20 predictive taxa were significantly different in mean relative abundances, supporting their role as distinguishing features between cases and controls (Fig. 2B). Of these 19 taxa, *Abiotrophia*, and *Clostridia* UCG-014 were the only taxa with significantly higher mean relative abundance in the case than control. Only *Rothia* did not exhibit significant differences in relative abundances between cases and controls (Fig. 2B).

3.4 *[Eubacterium] nodatum* group, *Mogibacterium*, and *Campylobacter* as inverse taxonomic markers for lung cancer

We performed differential abundance analysis to complement random forest approach to identify distinguishing

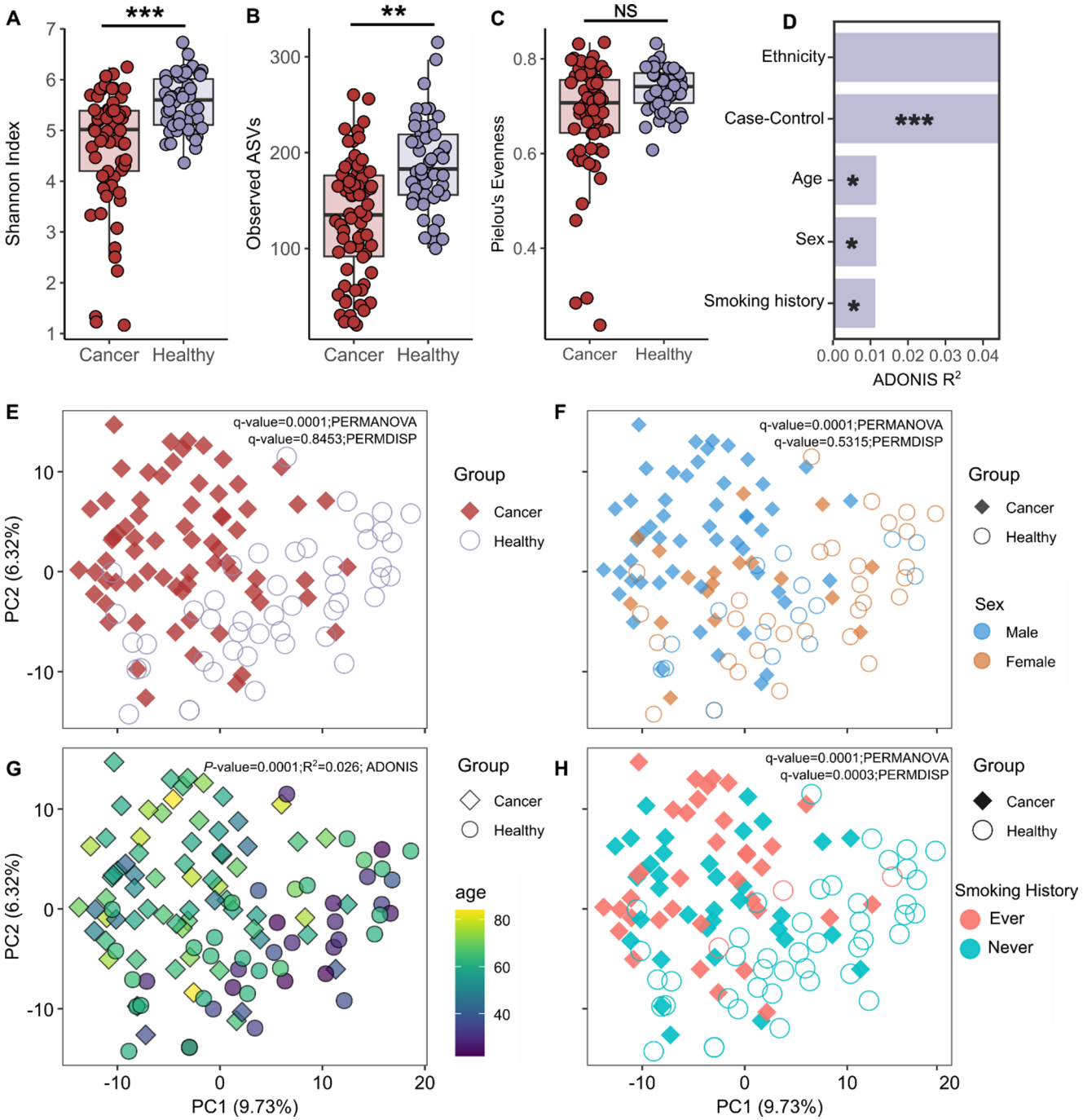

Fig. 1 Lower species richness of lung microbiome in patients with lung cancer compared to healthy controls. **A**) Shannon index, **B**) observed ASVs, and **C**) Pielou's evenness. Differences between groups were assessed using generalized linear models (GLMs). **D**) A barplot showing factors ranked by the proportion of variance (R^2) explained in Aitchison distance (ADONIS). **E**, **F**, **G**, **H**) Principal Coordinate Analysis (PCoA) plots showing the first two principal coordinate axes (PC) that represent the same Aitchison distances among sputum microbiota ($N=117$) samples. The first two axes represent 16.05% of the total variance. **E**) Case-control groups group, Cancer ($n=70$) and Healthy ($n=47$) samples. **F**) Sex group, Male ($n=68$) and Female ($n=49$). **G**) Each color represents the individual age at the time of sample collection. **H**) Smoking history as in ever-smoker ($n=39$) vs. never-smoker ($n=78$). The pairwise PERMANOVA and PERMDISP analyses for the categories are indicated in the PCoA plots except for the age group which was calculated using ADONIS. Statistical significance is denoted as '***' with p-value$\leq$0.001, '**' with p-values$<$0.01. and '*' with p-values$<$0.05, and NS (not significant, p-value$>$0.05)

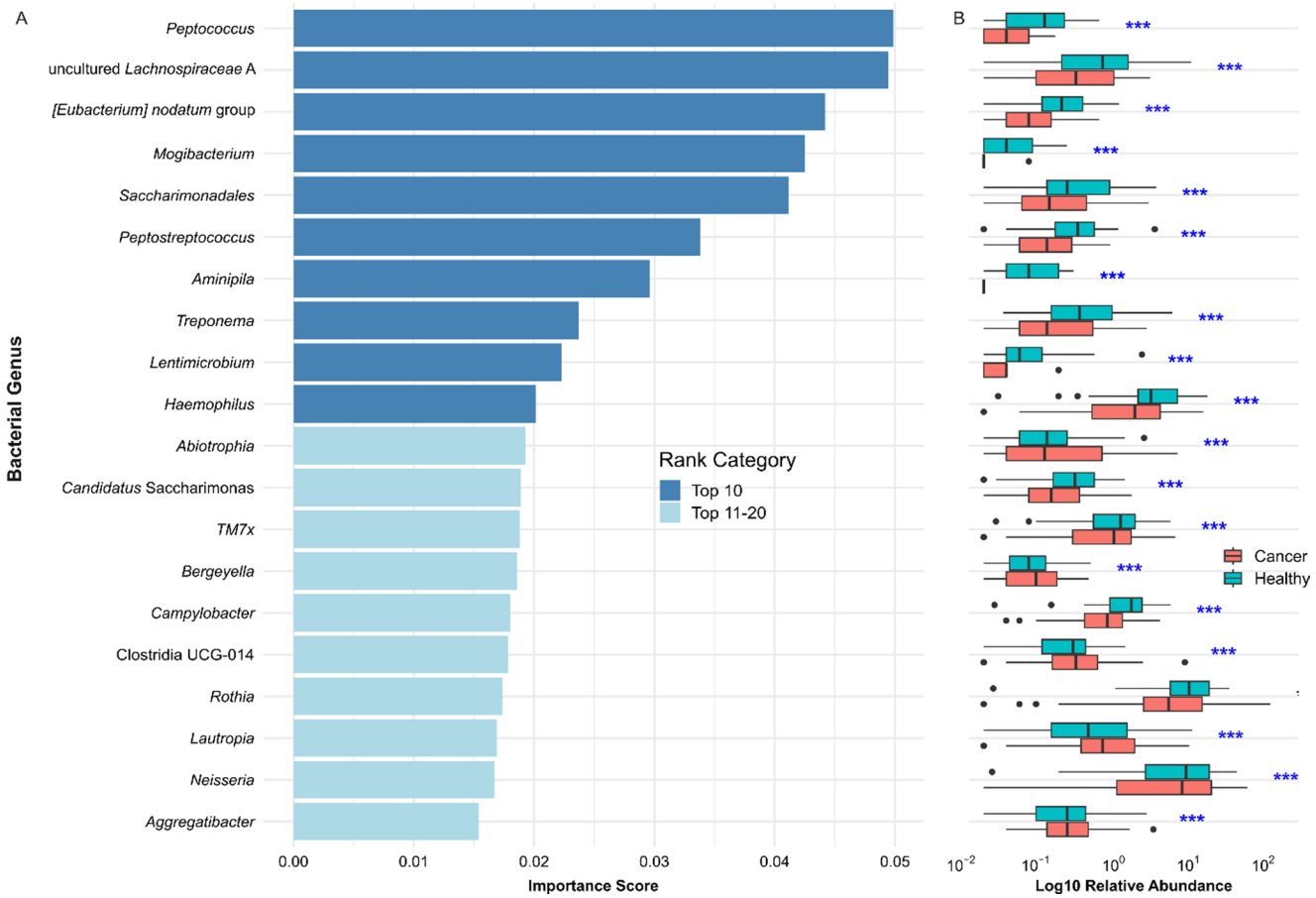

Fig. 2 Top 20 genera predictive of cancer-control status identified by random forest classifier analysis. **A**) Variable importance plot showing the top 20 taxa ranked by importance score. **B**) Boxplot showing relative abundance of the corresponding taxon. Statistical significance is denoted as '*' for q-value < 0.05, '**' for q-value < 0.01, and '***' for q-value < 0.001 based on pairwise Wilcoxon rank-sum test with Benjamini Hochberg false discovery rate adjusted p-values. Genus level taxonomy is based on a SILVA 138 pre-trained rRNA gene classifier

Table 2 Differentially abundant taxa between lung cancer and healthy control groups using ANCOM-BC2, adjusting for sex, smoking history, and age

Taxon	Log Fold Change between cancer cases and controls	Dunnet's test (q-value)	Sensitivity test passed?
Lactobacillus	1.727	2.2×10^{-4}	Yes
[Eubacterium] nodatum group	-1.067	0.034	Yes
Mogibacterium	-1.064	0.038	Yes
Campylobacter	-0.970	0.032	Yes
Rothia	-1.252	0.048	Yes

microbial genera between cases and controls. After adjusting for smoking history, sex, and age, five genera showed statistical significant differences between the two groups that passed the pseudo-count sensitivity test (Table 2). Of these, *Lactobacillus* was enriched in lung cancer cases, and four genera were depleted (Table 2). When stratified by smoking status, *Mogibacterium* was significantly more abundant among healthy never-smokers but was not detected in most cancer patients within the same smoking category (Fig. 3). This was supported by differential abundance analysis of stratified analysis within the never-smoker subgroup, where seven genera were identified (Supplementary Table S1). Among them were the same five taxa, including *Mogibacterium,* that showed consistency in the primary adjusted analysis. Notably, while *Lactobacillus* was significantly enriched in the cancer group, it did not pass the pseudo-count sensitivity test within the never-smoker cancer group, likely due to the reduced sample size of lung cancer cases among never-smokers (Supplementary Table S1). A similar analysis was not performed for ever-smokers ($n = 3$) owing to the small sample size among controls. *[Eubacterium] nodatum* group, *Mogibacterium*, and *Campylobacter* were also among the top 20 predictive taxa identified via random-forest that had significantly different relative abundances between case-control, highlighting their potential as inverse taxonomic markers for lung cancer (Fig. 2).

Since pseudo-count normalization methods may bias compositional data [35], we performed secondary analysis

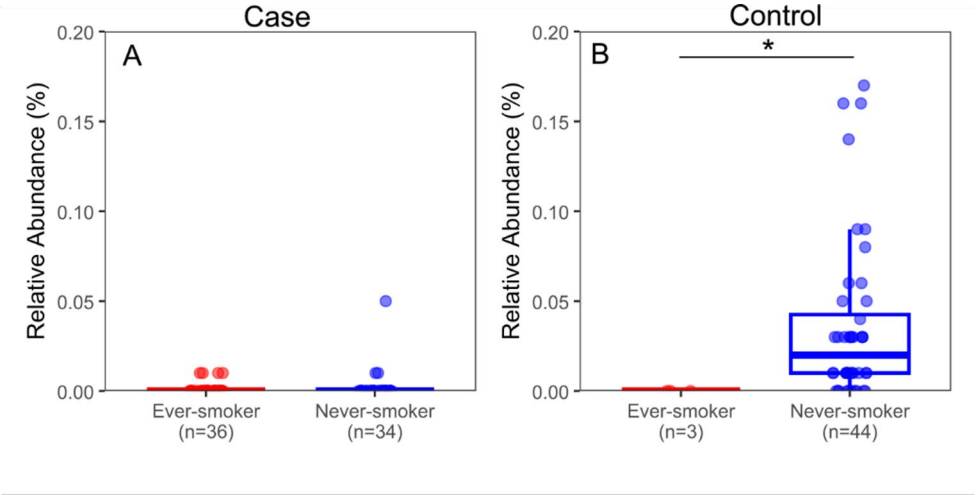

Fig. 3 Relative abundance of *Mogibacterium* in case-control groups, stratified by smoking history. **A**) Lung cancer cases and **B**) Healthy controls. Statistical significance is denoted as '*' for q-value < 0.05 based on Wilcoxon rank-sum test with Benjamini Hochberg false discovery rate adjusted p-values

using RCLR coupled with linear regression to validate the results from ANCOM-BC2. While both methods identified a depletion of *Campylobacter* and *Rothia* in the cancer group, contradictory differential outcomes were observed for *[Eubacterium] nodatum* group and *Mogibacterium*. This discrepancy was driven by the low prevalence of these taxa in the cancer group (Supplementary Table S2 and Supplementary Fig. S4). Based on the abundance of distributions between normalization methods, ANCOM-BC2 provided a more biologically plausible representation of the community shift, especially for taxa sensitive to pseudo-count transformation. Consequently, ANCOM-BC2 was retained as the primary method for all subsequent downstream analyses.

3.5 Cancer subtype accounts for the largest variation in microbiota composition

We further examined the cancer group to assess associations between clinical parameters and sputum microbiota, using permutational multivariate analysis of variance (ADONIS) to determine whether Aitchison distances varied within cases. Of the factors tested, lung cancer subtype (small-cell carcinoma, non-small cell carcinoma and others) accounted for the largest variation in microbial composition, followed by sex and age, based on Adonis R^2 values (Supplementary Table S3). Other variables including hospital, smoking history, cancer stage, ethnicity, treatment status, and BMI did not significantly influence microbiota composition (Supplementary Tables S3 and S4). Furthermore, no statistical differences in alpha-diversity indices were observed between the cancer subtypes (Supplementary Fig. S5) with age and BMI (Supplementary Fig. S2).

We performed differential abundance analysis on the sputum microbiota of cancer patients to identify taxa associated with smoking history. *Actinomycetaceae*_F0332 (LFC=-0.985; q-value = 0.011) was significantly less abundant in ever-smokers. However, after adjusting for sex, and age, no differentially abundant taxa was identified between ever- and never-smokers. Comparing small-cell carcinoma to non-small cell carcinoma subtypes, *Tannerella* (LFC − 1.589; q-value = 0.005) was significantly less abundant in small-cell lung cancer compared to non-small cell lung cancer. No taxa were differentially abundant between the remaining subtypes and non-small cell lung cancer as the reference.

3.6 Sex as a factor that differentiates the sputum microbiota of cancer and control groups

As beta-diversity differed significantly between males and females, regardless of case-control status, we stratified the samples by sex to identify sex-specific microbial signatures. We found that the Aitchison distances between cases and controls remained significantly different for both sexes (Fig. 4).

We compared the top 10 most abundant bacterial phyla and families in male and female participants across cancer and control groups (Supplementary Fig. S6). The community structure at the phylum-level, regardless of sex, was largely represented by the same ten phyla, with *Desulfobacterota* showing low prevalence (~5% of male samples and 11.7% of female samples) and low abundance (median relative abundance ~ 0.2% of positive male samples and ~ 0.02% positive female samples). At the family level, community structures showed modest differences, in the order of a few families such as *Prevotellaceae* and *Neisseriaceae* between sexes (Supplementary Fig. S6). We also compared the top 20 most abundant bacterial genera in sex-stratified cases to identify four genera that differed between male and female cancer cases, namely *Staphylococcus*, and *Lactobacillus* in males and the uncultured TM7x and *Klebsiella* in

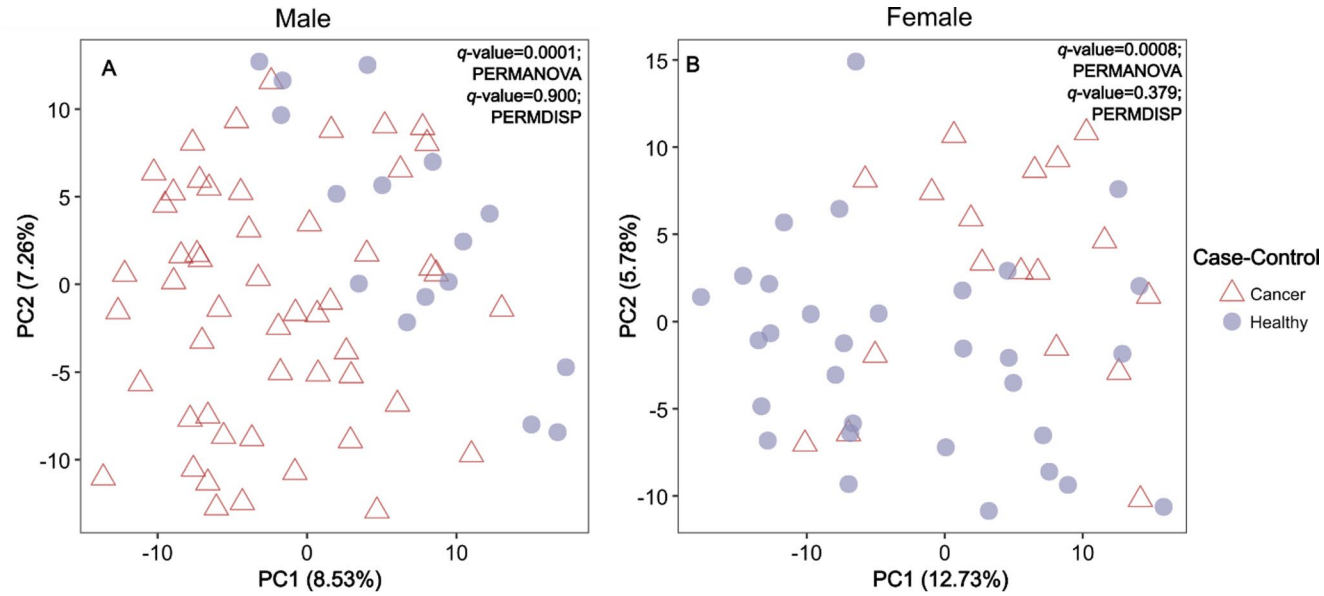

Fig. 4 Beta-diversity (Aitchison distance) of sputum microbiota between case-control, stratified by sex. **A**) Male cancer ($n=52$) and healthy ($n=16$) groups. **B**) Female cancer ($n=18$) and healthy ($n=31$) groups. The global PERMANOVA and PERMDISP tests are shown in the plots. Variance for each axis is shown in parenthesis

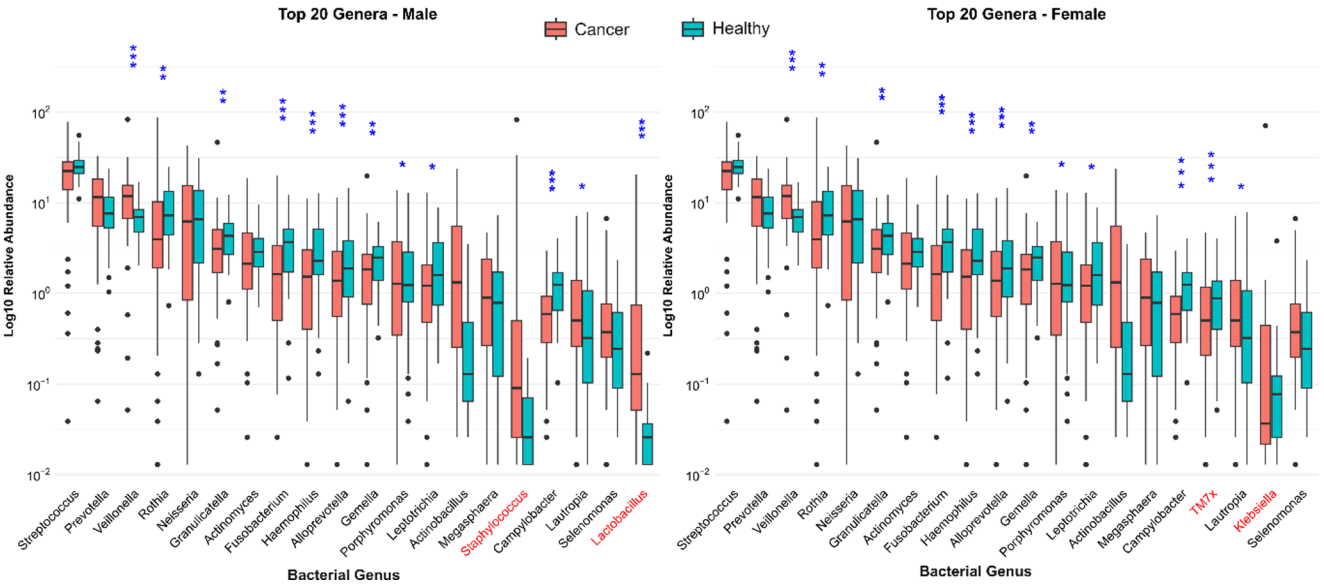

Fig. 5 Top 20 bacterial genera differentially abundant between lung cancer cases and healthy controls, stratified by sex. Each boxplot shows the relative abundance of the corresponding taxon arranged from highest to lowest mean relative abundance. Statistical significance is denoted as '*' for q-value < 0.05), '**' for q-value < 0.01 and '***' for q-value < 0.001 based on pairwise Wilcoxon rank-sum test with Benjamini Hochberg false discovery rate adjusted p-values. Sex-specific taxa are shown in red

females. Most genera were still statistically more abundant in healthy controls compared to cancer cases (Fig. 5).

ANCOM-BC2 analysis stratified by sex was conducted to identify sex-specific differentially abundant taxa between case and control groups. Among females, six genera were significantly less abundant in the cancer group compared to the control group (Table 3). In males, three of the four differentially abundant taxa were also less abundant in the cancer group, with *Lactobacillus* being the only genus enriched among male lung cancer patients (Table 3; Fig. 6).

When examining never-smokers stratified by sex, *Actinobacillus* was the only genus significantly enriched in the male non-smoker cancer group (LFC = 3.57; q-value = 5.5×10^{-4}). However, this taxon did not pass the pseudo-count sensitivity test. No significant taxa were identified when comparing cancer and healthy groups among female never-smokers,

Table 3 ANCOM-BC2 analysis using the control group as reference. Samples were stratified by sex (male/female) before ANCOM-BC2 analysis and results are reported separately

Taxon	Log Fold Change in cancer case-control groups	Dunnet's test (q-value)	Sensitivity test passed?
Female			
Peptococcus	**-1.709**	**0.001**	**Yes**
[Eubacterium] nodatum group	-1.693	0.008	Yes
Aminipila	-1.612	0.001	Yes
Peptostreptococcus	-1.406	0.045	Yes
*Candidatus*_Saccharimonas	-1.366	0.028	Yes
Campylobacter	-1.158	0.027	Yes
Male			
Lactobacillus	2.620	9.6×10^{-8}	Yes
Parvimonas	-1.416	0.0069	Yes
Peptococcus	**-1.348**	**0.0226**	**Yes**
Mogibacterium	-1.246	0.0126	Yes

Taxon common to both sexes are shown in bold

suggesting that smoking status may be a primary driver of microbial shifts in this cohort. Finally, we summarized all taxa found to be differentially abundant between cases and controls, as well as within cases (Fig. 7).

4 Discussion

To our knowledge, this is the first study comparing the respiratory microbiota of lung cancer patients and healthy controls in Singapore and Southeast Asia. Consistent with prior observational studies, we observed significant differences in alpha-diversity specifically genus richness and beta-diversity (overall community composition and structure) between lung cancer cases and controls [14, 36, 37]. Notably, while genus evenness is less commonly analyzed between lung cancer patients and healthy controls, one study showed similar evenness between lung cancer patients and those with pulmonary disease [38], suggesting that the relative distribution of taxa remains similar between lung cancer cases and controls, even as the overall community composition differs. Microbial composition and structure appeared to be influenced by cancer status (case vs. control), demographic factors (age and sex), and tumor subtype, highlighting potential relevance to lung cancer outcomes that warrants further investigation in larger cohorts. While dynamic changes in the gut microbiome across the human lifespan are well-documented, temporal shifts in the lung microbiome during cancer progression remain understudied [39]. Several taxa previously implicated in lung cancer [40, 41], including *Veillonella* and *Megasphaera*, have also been reported in other chronic respiratory diseases such as COPD [42, 43], where they are often associated with airway inflammation, microbial dysbiosis, and disease severity. However, in contrast to some earlier lung cancer studies, we did not observe differential abundance of these genera between cases and controls in either sex, which may likely reflect differences in sample type (sputum versus bronchoalveolar lavage or tissue), analytical approaches, population characteristics, and disease stage. In particular, microbial patterns reported in COPD are often shaped by chronic airway inflammation, frequent corticosteroid use, and recurrent exacerbations, which may not mirror the airway ecology of lung cancer, especially in late-stage disease [44, 45].

In addition, despite smoking being a major risk factor, we could not conclusively determine whether smoking history has a significant effect on beta-diversity. This is due to the significant dispersion between ever- and never-smokers, suggesting greater inter-individual variability among smokers, which violated the homogeneity assumption and limited the interpretability of PERMANOVA results. Similar challenges have been reported in COPD microbiome studies, where smoking and disease heterogeneity both influence microbial composition and can obscure disease-specific

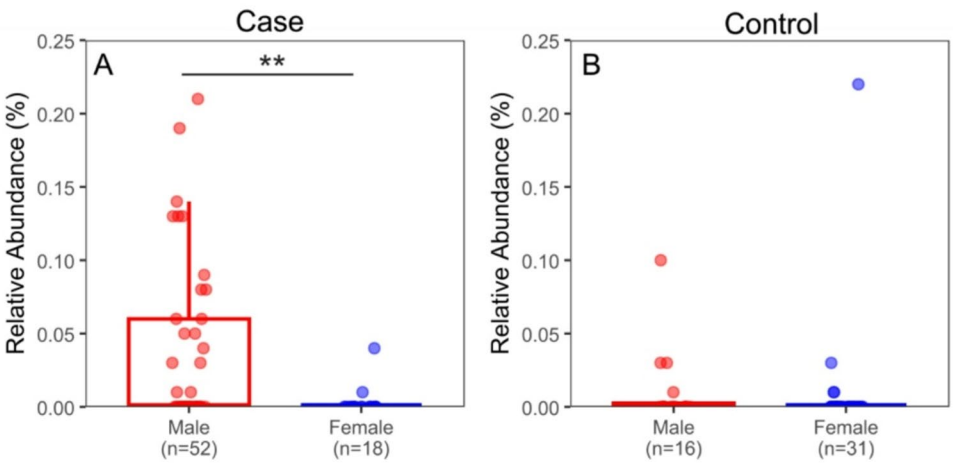

Fig. 6 Boxplots showing the relative abundance of *Lactobacillus* in cases and controls, stratified by sex. **A**) Cancer cases and **B**) Controls. Statistical significance is denoted as '**' for q-value < 0.001 based on pair-wise Wilcoxon rank-sum test with Benjamini Hochberg false discovery rate adjusted p-values. Statistically non-significant results (q-value > 0.05) are not shown

Fig. 7 Statistically significant differentially abundant features by case-control status and within cases analysis using ANCOM-BC2. Bacteria are ordered by increasing *q*-values (Wilcoxon rank-sum test with Benjamini Hochberg false discovery rate adjusted *p*-values)

signals. For instance, COPD sputum microbiomes differ markedly from healthy subjects independent of smoking history, highlighting complex interactions between exposure and disease effects on airway microbiota [46]. Moreover, cigarette smoke-induced dysbiosis in lung microbiomes highlights how exposure-related effects may overlap with disease-associated microbial changes [47]. Nevertheless, an uncultured *Actinomycetaceae* strain F0332, previously identified from oral metagenomes, was differentially less abundant in ever-smokers. Given the known role of Actinomycetaceae (e.g., *Actinomyces israelii*) in pulmonary infections, future culture-based studies are needed to clarify its role in smoking-related lung cancer [48, 49].

In our study, both age and sex appeared to influence airway microbiota composition [50, 51]. Previous research has shown sex-specific variations in lung function and immune responses, though consistent differences in the lung microbiome remain unclear [52, 53]. Sex-based immune dimorphism, known to shape microbiomes at other body sites [54], may partly explain these differences [55]. Specifically, we identified *Lactobacillus,* a genus with known associations to cancer biology, as one of the sex-specific differentially abundant taxa in our study. Males tend to have a higher prevalence of smoking, and smoking is known to alter airway pH, oxidative stress, and mucosal immunity, creating conditions that favor stress-tolerant taxa such as *Lactobacillus* [56–58]. While strongly sex-associated taxa were not observed within never-smokers, this pattern may differ in smokers, where smoking has been associated with differential gene expression in the airways of males and females [55]. Whether changes in gene expression and immune responses between sexes may differentially affect airway microbiota remains to be fully elucidated. Furthermore, the identification of *Lactobacillus* could be

confounded by exogenous factors such as medication and probiotic use. Importantly, adjustment for smoking status may not fully capture cumulative exposure or smoking-related biological effects, and residual confounding is likely. Consequently, these findings likely reflect a combination of smoking-associated microbial selection and sex-specific host factors rather than a direct causal role.

Furthermore, while generally considered beneficial in the gut, *Lactobacillus* has also been linked to progression in different cancers including lung cancer [59–62]. The accumulation of lactate in tumors is now known to influence cancer epigenetics through changes in cell signaling pathways and a post-translational modification known as histone lactylation that may exacerbate cell proliferation [63]. However, a causal link between *Lactobacillus* presence and pathological lactic acid accumulation remains to be definitively proven. *Lactobacillus* also plays a role in chronic respiratory conditions like bronchiectasis [61] and gut-lung axis communication [62, 64], highlighting its capacity to influence airway microbial communities through systemic interactions with the gastrointestinal microbiome. Therefore, its enrichment in lung tumors in our study suggests a potential role in tumor immune modulation and metabolism, warranting further investigation [65]. Understanding these interactions may provide mechanistic insights into how systemic microbiota contributes to lung cancer development and progression and may inform future biomarker or therapeutic strategies targeting both gut and airway microbiomes.

As the oral-lung axis is connected, microbes from the mouth are inevitably present in the sputum microbiota. A prospective study showed that reduced oral microbiome diversity from oral washes may increase lung cancer risk, prompting efforts to identify microbial taxa that could serve as early risk biomarkers [37]. In our study, many bacterial genera identified were common to oral flora, which might not fully represent the lung microbiome. Specifically, *Abiotrophia* was enriched in lung cancer sputum [66], and *Tannerella* was associated with periodontitis and esophageal adenocarcinoma [67, 68]. Interestingly, most differentially abundant genera in our cohort were more prevalent in healthy controls than in lung cancer cases, possibly reflecting the predominance of late-stage disease, limiting insights into microbial changes during early phases of tumorigenesis. Similar stage-dependent microbial shifts have been observed in COPD and bronchiectasis, where advanced disease is often associated with reduced diversity and loss of previously prevalent taxa. For example, while *Campylobacter* is frequently associated with COPD and lung infections, it was depleted in lung cancer cases, suggesting that its role may decline with disease progression due to changes in lung microenvironments [69].

Importantly, while we employed stratified analyses and statistical adjustments to account for key covariates, including age, smoking history, and lifestyle factors, these host-related variables may still partially contribute to the observed microbial patterns between cases and controls. For example, of the three genera that may be inverse taxonomic markers for lung cancer, only *Eubacterium* spp. are associated with beneficial properties in the gut such as modulating gut inflammation through butyrate and propionate production and may also be an important member in the lung-gut axis [70]. *Mogibacterium* and *Campylobacter* tend to be associated with adverse clinical outcomes such as early-stage lung cancer or pulmonary diseases such as COPD and idiopathic pulmonary fibrosis [10, 44, 71–73]. As such, residual confounding cannot be fully excluded, and some of the observed enrichments or depletions may reflect complex interactions between host and environmental factors rather than lung cancer alone. Consequently, while several taxa appear associated with lung cancer status, these findings should be interpreted as associative and influenced by interacting demographic and environmental factors, rather than solely attributable to lung cancer itself. While our random forest classifier achieved high diagnostic accuracy (87.5%) in distinguishing lung cancer cases from healthy controls, we acknowledge that these findings are derived from a multi-center case-control study. The high degree of overlap between the top predictive features of the classifier and those identified via univariate differential abundance testing (ANCOM-BC2) provides greater credence to these taxa. Nevertheless, independent validation in larger, prospective cohorts is required to establish a 'core' microbial signature for lung cancer. Future studies should prioritize the recruitment of balanced cohorts for key covariates, such as age and smoking history, to enable the use of causal inference methods, such as two-sample Mendelian randomization, in validating robust taxonomic signatures of lung cancer [74].

5 Limitations

This study has several limitations that warrant consideration. First, due to the limited sample size, we were unable to match cases and controls by key demographic and clinical variables, including ethnicity, age, sex, and smoking history; however, we adjusted for these potential confounders in our statistical models. Notably, while smoking is a primary driver of lung cancer, the limited number of smokers in our control group constrained our comparative analysis. Furthermore, we acknowledge that metadata regarding antibiotic use was unavailable for the healthy cohort, which remains a potential confounding factor in our microbiome findings. Second, while smoking history

was complete, recall bias from self-reported data may have affected exposure classification. Third, the predominance of stage IV cancer cases limited our ability to explore microbial features associated with cancer stage using differential abundance and random forest analyses. Therefore, to mitigate confounding, linear regression models were adjusted for relevant covariates where appropriate. Fourth, due to the cross-sectional design of our study, causal inferences cannot be made regarding the observed microbiome differences; it remains unclear whether these microbial alterations precede or result from lung cancer development. Longitudinal studies are needed to ascertain the temporal dynamics and potential causal relationships between microbiome composition and lung cancer development. Fifth, potential contamination of sputum samples with oral microbiota limits their utility in characterizing the lower respiratory tract. While invasive methods like bronchoalveolar lavage are routinely used for diagnostic purposes, the degree of oral contamination inherent to these procedures in microbiome studies is still being evaluated [75, 76]. Moreover, these invasive techniques are difficult to implement in large-scale studies of healthy populations, complicating efforts to define a baseline for healthy lung microbiome. Lastly, due to the compositional nature of the data and the high prevalence of sparse taxa, certain data manipulation methods, such as the addition of pseudo-counts or robust centered log-ratio transformation may still introduce false positives, as highlighted here and in other studies [77, 78]. Our analysis and others suggest that leveraging pseudo-count sensitivity methods like ANCOM-BC2, combined with FDR correction, machine learning validation and stratification, provides a more robust approach to limiting false positives and ensuring the reliability of identified biomarkers [79–81].

Despite these limitations, this work represents the first study in Singapore and Southeast Asia to identify sex-specific lung cancer-associated taxa. Future validation efforts should ideally employ consistent amplicon sequencing methodologies (e.g., identical 16S rRNA primers and thermal cycling conditions) to ensure comparability. Our findings offer preliminary insights into potential airway microbiome–lung cancer associations that may inform future efforts to identify potential early microbial biomarkers in Southeast Asian populations. Validation in larger, prospective cohorts is needed to assess the clinical utility of microbial signatures for early detection, risk stratification, and treatment response, as well as to confirm the robustness and clinical relevance of these findings. Longitudinal studies tracking microbial shifts before and after diagnosis or treatment may also offer insights into microbiota dynamics during disease progression. Lastly, integrating microbial profiling with host genomic, immunologic, and clinical data may further clarify host–microbe interactions and guide clinically actionable microbiome-based strategies for lung cancer prevention and care in Southeast Asian populations.

6 Conclusion

In summary, we identified several differentially abundant bacterial genera in the sputum microbiome of lung cancer cases, as well as between cases and healthy controls in a multiethnic Southeast Asian population. Longitudinal studies with targeted sampling are needed to better understand microbiome dynamics over time and during lung cancer progression. Larger studies with more diverse patient populations will be essential to validate and extend these findings.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13402-026-01193-7.

Author contributions Adrian Low and Yah Ru Juang contributed to formal analysis, methodology, software, visualization, validation, investigation, and writing of the original draft, review, and editing. Sean Wei Jun Chan contributed to formal analysis and writing (review and editing). Fransiskus X. Ivan, Luke Ooi Hao Shuan, Micheal Mac Aogain, Tavleen Kaur Jaggi, and Sanjay H. Chotirmall contributed to data curation, methodology, and writing (review and editing). Wei Jie Seow contributed to conceptualization, formal analysis, methodology, supervision, writing (review and editing). The rest of authors (Lina Ang, Yann Felix Boucher, Anthony Chau Ang Yii, Mariko Koh Siyue, Darren Wan Teck Lim, Jonathan Wei Jie Lee) contributed to writing (review and editing) by providing critical insights to improve the manuscript. All authors were involved in the revision of the manuscript for important intellectual content and approved the final version to be published.

Funding This study is funded by the Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 FRC (FY2017).

Data availability The data that support the findings of this study are analyzed in a secure environment and are not available for external request due to reasons of sensitivity.

Declarations

Ethics approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the SingHealth Centralized Institutional Review Board (CIRB Ref: 2018/2524).

Consent to participate Informed consent was obtained from all individual participants included in the study.

Competing interests The authors declare no competing interests.

References

1. F. Bray et al., Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. **74**(3), 229–263 (2024)
2. J. Ferlay et al., Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int. J. Cancer. **144**(8), 1941–1953 (2019)
3. V. Noronha et al., Uniqueness of lung cancer in Southeast Asia. Lancet Reg. Health Southeast. Asia. **27**, 100430 (2024)
4. R.P. Dickson, J.R. Erb-Downward, G.B. Huffnagle, Homeostasis and its disruption in the lung microbiome. Am. J. Physiol. Lung Cell. Mol. Physiol. **309**(10), L1047–L1055 (2015)
5. G.B. Huffnagle, R.P. Dickson, N.W. Lukacs, The respiratory tract microbiome and lung inflammation: a two-way street. Mucosal Immunol. **10**(2), 299–306 (2017)
6. S.L. Ashley et al. Lung and gut microbiota are altered by hyperoxia and contribute to oxygen-induced lung injury in mice. Sci. Transl Med. **12**(556), (2020)
7. M. Sommariva et al., The lung microbiota: role in maintaining pulmonary immune homeostasis and its implications in cancer development and therapy. Cell. Mol. Life Sci. **77**(14), 2739–2749 (2020)
8. D. Yang et al., The impact of lung microbiota dysbiosis on inflammation. Immunology. **159**(2), 156–166 (2020)
9. S. Majumder, M.J. McGeachy, IL-17 in the Pathogenesis of Disease: Good Intentions Gone Awry. Annu. Rev. Immunol. **39**, 537–556 (2021)
10. X. Yi, J. Gao, Z. Wang, The human lung microbiome-A hidden link between microbes and human health and diseases. Imeta. **1**(3), e33 (2022)
11. S.R. Lucaciu et al. Lung microbiome in lung cancer: a systematic review. Microorganisms. **12**(12), (2024)
12. H. Lin, S.D. Peddada, Analysis of microbial compositions: a review of normalization and differential abundance analysis. NPJ Biofilms Microbiomes. **6**(1), 60 (2020)
13. K. Su, Y. Gao, J. He, A comparison of the microbiome composition in lower respiratory tract at different sites in early lung cancer patients. Transl Lung Cancer Res. **12**(6), 1264–1275 (2023)
14. H.X. Liu et al., Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects. Int. J. Cancer. **142**(4), 769–778 (2018)
15. H.D. Hosgood., 3rd, et al., The potential role of lung microbiota in lung cancer attributed to household coal burning exposures (Environ Mol Mutagen) (2014)
16. S.A. Whiteside, J.E. McGinniss, R.G. Collman. The lung microbiome: progress and promise. J Clin Invest. **131**(15), (2021)
17. A. Klindworth et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. **41**(1), e1 (2013)
18. S.Y. Lee et al., Airway microbiome composition correlates with lung function and arterial stiffness in an age-dependent manner. PLoS One. **14**(11), e0225636 (2019)
19. E. Bolyen et al., Author Correction: Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. **37**(9), 1091 (2019)
20. B.J. Callahan et al., DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods. **13**(7), 581–583 (2016)
21. L. Beule, P. Karlovsky, Improved normalization of species count data in ecology by scaling with ranked subsampling (SRS): application to microbial communities. PeerJ. **8**, e9593 (2020)
22. H. Wickham et al., Welcome to the Tidyverse. J. Open. Source Softw. **4**(43), 1686 (2019)
23. C. Quast et al., The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. **41**(Database issue), D590–D596 (2013)
24. F. Pedregosa et al., Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
25. N.A. Bokulich et al., q2-sample-classifier: machine-learning tools for microbiome classification and regression. J. Open. Res. Softw. **3**(30), 934 (2018)
26. M.J. Anderson, A new method for non-parametric multivariate analysis of variance. Austral Ecol. **26**(1), 32–46 (2001)
27. J. Oksanen et al. vegan: Community Ecology Package. R package version 2.5-6. (2019)
28. J. Allaire, RStudio: integrated development environment for R. MA **770**(394), 165–171 (2012). Boston
29. R. C. Team, R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria. (2021)
30. J. E. Bisanz, qiime2R: Importing QIIME2 artifacts and associated data into R sessions. (2018)
31. T. Lumley, Analysis of complex survey samples. J. Stat. Softw. **9**, 1–19 (2004)
32. H. Lin, S.D. Peddada, Analysis of compositions of microbiomes with bias correction. Nat. Commun. **11**(1), 3514 (2020)
33. C.W. Dunnett, A multiple comparison procedure for comparing several treatments with a control. J. Am. Stat. Assoc. **50**(272), 1096–1121 (1955)
34. S. Holm, A simple sequentially rejective multiple test procedure. Scand. J. Stat. 65–70 (1979)
35. C. Martino et al., A Novel sparse compositional technique reveals microbial perturbations. mSystems, **4**(1), (2019)
36. Y. Qiao, J. Mei, Z. S. Ma, Species diversity and network diversity in the human lung cancer tissue microbiomes. FEMS Microbiol. Lett. **371**, (2024)
37. H.D. Hosgood et al., Variation in oral microbiome is associated with future risk of lung cancer among never-smokers. Thorax. **76**(3), 256–263 (2021)
38. C. Cheng et al., Characterization of the lung microbiome and exploration of potential bacterial biomarkers for lung cancer. Transl Lung Cancer Res. **9**(3), 693–704 (2020)
39. K.Y. Toh et al., Identification of age-associated microbial changes via long-read 16S sequencing. Gut Pathog. **16**(1), 56 (2024)
40. W. Zeng et al., Alterations of lung microbiota in patients with non-small cell lung cancer. Bioengineered. **13**(3), 6665–6677 (2022)
41. S.H. Lee et al., Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions. Lung Cancer. **102**, 89–95 (2016)
42. S. Gupta et al., Comparative analysis of the alveolar microbiome in COPD, ECOPD, Sarcoidosis, and ILD patients to identify respiratory illnesses specific microbial signatures. Sci. Rep. **11**(1), 3963 (2021)
43. L. Millares, E. Monso, The Microbiome in COPD: Emerging Potential for Microbiome-Targeted Interventions. Int. J. Chron. Obstruct Pulmon Dis. **17**, 1835–1845 (2022)

44. Z. Wang et al., Inflammatory endotype-associated airway microbiome in chronic obstructive pulmonary disease clinical stability and exacerbations: a multicohort longitudinal analysis. Am. J. Respir Crit. Care Med. **203**(12), 1488–1502 (2021)
45. Y. Yue et al., Sputum microbe community alterations induced by long-term inhaled corticosteroid use are associated with airway function in chronic obstructive pulmonary disease patients based on metagenomic next-generation sequencing (mNGS). Front. Pharmacol. **15**, 1323613 (2024)
46. K. Haldar et al., The sputum microbiome is distinct between COPD and health, independent of smoking history. Respir Res. **21**(1), 183 (2020)
47. V. Laiman et al., Cigarette smoke-induced dysbiosis: comparative analysis of lung and intestinal microbiomes in COPD mice and patients. Respir. Res. **25**(1), 204 (2024)
48. G.B. Hill, O.M. Ayers, A.P. Kohan, Characteristics and sites of infection of *Eubacterium nodatum, Eubacterium timidum, Eubacterium brachy,* and other asaccharolytic eubacteria. J. Clin. Microbiol. **25**(8), 1540–1545 (1987)
49. L.V. Holdeman et al., Descriptions of *Eubacterium timidum* sp. nov., *Eubacterium brachy* sp. nov., and *Eubacterium nodatum* sp. nov. isolated from human periodontitis. Int. J. Syst. Evol. Microbiol. **30**(1), 163–169 (1980)
50. V.T. Chu et al., The antibiotic resistance reservoir of the lung microbiome expands with age in a population of critically ill patients. Nat. Commun. **15**(1), 92 (2024)
51. Z. Zhou, et al., Smoking-induced microbial dysbiosis: a key driver of systemic diseases and emerging therapeutic opportunities. Cell Death Discov. (2025).
52. T. Takahashi, A. Iwasaki, Sex differences in immune responses. Science. **371**(6527), 347–348 (2021)
53. S.L. Klein, K.L. Flanagan, Sex differences in immune responses. Nat. Rev. Immunol. **16**(10), 626–638 (2016)
54. D. Zheng, T. Liwinski, E. Elinav, Interaction between microbiota and immunity in health and disease. Cell. Res. **30**(6), 492–506 (2020)
55. C.X. Yang et al., Widespread Sexual Dimorphism in the Transcriptome of Human Airway Epithelium in Response to Smoking. Sci. Rep. **9**(1), 17600 (2019)
56. C. Huang, G. Shi, Smoking and microbiome in oral, airway, gut and some systemic diseases. J. Transl Med. **17**(1), 225 (2019)
57. K. Kostikas et al., Secondhand smoke exposure induces acutely airway acidification and oxidative stress. Respir Med. **107**(2), 172–179 (2013)
58. A.-R. Koczulla et al., Acute and chronic effects of smoking on inflammation markers in exhaled breath condensate in current smokers. Respiration. **79**(1), 61–67 (2009)
59. Q. Shi, et al., Gut Lactobacillus contribute to the progression of breast cancer by affecting the anti-tumor activities of immune cells in the TME of tumor-bearing mice. Int. Immunopharmacol. **124**, 111039 (2023)
60. L.E. Colbert et al., Tumor-resident *Lactobacillus iners* confer chemoradiation resistance through lactate-induced metabolic rewiring. Cancer Cell. **41**(11), 1945–1962e11 (2023)
61. J.K. Narayana et al., Microbial Dysregulation of the Gut-Lung Axis in Bronchiectasis. Am. J. Respir Crit. Care Med. **207**(7), 908–920 (2023)
62. D. Dora et al., From bench to bedside: an interdisciplinary journey through the gut-lung axis with insights into lung cancer and immunotherapy. Front. Immunol. **15**, 1434804 (2024)
63. J. Chen et al., Lactate and lactylation in cancer. Signal. Transduct. Target. Ther. **10**(1), 38 (2025)
64. S. Rastogi, A. Singh, Gut microbiome and human health: Exploring how the probiotic genus *Lactobacillus* modulate immune responses. Front. Pharmacol. **13**, 1042189 (2022)
65. L. Liu et al., Lung cancer and the Gut-microbiota-lung Axis: emerging evidence and potential clinical implications. Front. Med. (Lausanne). **12**, 1655780 (2025)
66. J. Cheng, L. Zhou, H. Wang, Symbiotic microbial communities in various locations of the lung cancer respiratory tract along with potential host immunological processes affected. Front. Cell. Infect. Microbiol. **14**, 1296295 (2024)
67. B.A. Peters et al., Oral Microbiome Composition Reflects Prospective Risk for Esophageal Cancers. Cancer Res. **77**(23), 6777–6787 (2017)
68. B. Malinowski et al., The role of *Tannerella forsythia* and *Porphyromonas gingivalis* in pathogenesis of esophageal cancer. Infect. Agent Cancer. **14**, 3 (2019)
69. R. Matsumoto et al., Thoracic Empyema Caused by Campylobacter rectus. Intern. Med. **61**(9), 1399–1402 (2022)
70. A. Mukherjee et al., Gut microbes from the phylogenetically diverse genus *Eubacterium* and their various contributions to gut health. Gut Microbes. **12**(1), 1802866 (2020)
71. S. Liu et al., Oral microbial dysbiosis in patients with periodontitis and chronic obstructive pulmonary disease. Front. Cell. Infect. Microbiol. **13**, 1121399 (2023)
72. P.L. Molyneaux et al., Changes in the respiratory microbiome during acute exacerbations of idiopathic pulmonary fibrosis. Respir. Res. **18**, 1–6 (2017)
73. J. Luo et al., Relationship of *Mogibacterium* in the gut microbiota with early-stage lung cancer. J. Appl. Microbiol. **136**(6), (2025)
74. X.Q. Fu et al., Perspectives on the gut microbiota in patients with non-small cell lung cancer-Two-Sample Mendelian Randomization combined with bibliometric analysis (2001–2023). J. Cardiothorac. Surg. **20**(1), 372 (2025)
75. G. Kim et al., Prediction of lung cancer using novel biomarkers based on microbiome profiling of bronchoalveolar lavage fluid. Sci. Rep. **14**(1), 1691 (2024)
76. K. Bahetjan et al., Analysis of the bronchoalveolar lavage fluid microbial flora in COPD patients at different lung function during acute exacerbation. Sci. Rep. **15**(1), 13179 (2025)
77. G.I. Austin, T. Korem, Compositional transformations can reasonably introduce phenotype-associated values into sparse features. mSystems. **10**(5), e0002125 (2025)
78. S.L. Salzberg, Reply to Austin and Korem, Compositional transformations can reasonably introduce phenotype-associated values into sparse features. mSystems. **10**(5), e0024825 (2025)
79. I. Garach Velez et al., Exploring the role of normalization and feature selection in microbiome disease classification pipelines. Gigascience. **14**, (2025)
80. H. Lin, S.D. Peddada, Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. Nat. Methods. **21**(1), 83–91 (2024)
81. J. Pelto et al., Elementary methods provide more replicable results in microbial differential abundance analysis. Brief. Bioinform. **26**(2), (2025)

Authors and Affiliations

Adrian Low[1] · Yah Ru Juang[2] · Fransiskus X. Ivan[3] · Lina Ang[2] · Luke Hao Shuan Ooi[3] · Sean Wei Jun Chan[2] · Micheal Mac Aogain[4,5] · Tavleen Kaur Jaggi[3] · Sanjay H. Chotirmall[3,6] · Yann Felix Boucher[2,7,8] · Anthony Chau Ang Yii[9] · Mariko Siyue Koh[10,11] · Darren Wan Teck Lim[11,12,13] · Jonathan Wei Jie Lee[1,14,15] · Wei Jie Seow[1,2]

✉ Wei Jie Seow
ephswj@nus.edu.sg

1 Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

2 Saw Swee Hock School of Public Health, National University Singapore, Singapore, Singapore

3 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

4 Department of Biochemistry, St. James's Hospital, Dublin, Ireland

5 Clinical Biochemistry Unit, School of Medicine, Trinity College Dublin, Dublin, Ireland

6 Department of Respiratory and Critical Care Medicine, Tan Tock Seng Hospital, Singapore, Singapore

7 Singapore Centre for Environmental Life Sciences Engineering (SCELSE), National University of Singapore, Singapore, Singapore

8 Infectious Diseases Translational Research Program, Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, National University Hospital System, Singapore, Singapore

9 Department of Respiratory and Critical Care Medicine, Changi General Hospital, Singapore, Singapore

10 Department of Respiratory and Critical Care Medicine, Singapore General Hospital, Singapore, Singapore

11 Duke-NUS Medical School, Singapore, Singapore

12 National Cancer Centre Singapore, Singapore, Singapore

13 Institute of Molecular and Cell Biology, A*STAR, Singapore, Singapore

14 Institute for Health Innovation and Technology (iHealthtech), National University of Singapore, Singapore, Singapore

15 Division of Gastroenterology & Hepatology, Department of Medicine, National University Hospital, Singapore, Singapore