Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Smoking-Associated Alterations in Bronchoalveolar Lavage Fluid Microbiota and Metabolomic Profiles in Patients With Lung Cancer

Yanchao Liu
The Second Hospital of Hebei Medical University
Miaomiao Cui
Hebei Provincial People's Hospital
Runlu Wang
Shijiazhuang People's Hospital
Zheng Duan
duanzheng1971@126.com

The Second Hospital of Hebei Medical University

Research Article

Keywords: lung cancer, smoking, bronchoalveolar lavage fluid, lung microbiota, 16S rRNA sequencing, untargeted metabolomics, cross-omics analysis

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10830539/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Smoking is one of the most important modifiable risk factors for lung cancer. However, clinical evidence remains limited regarding whether smoking is associated with remodeling of the lower-airway microbiome and local metabolic networks in patients with lung cancer. Bronchoalveolar lavage fluid (BALF) may reflect the local lower-airway microenvironment, and integrated microbiome-metabolome profiling may help identify smoking-associated ecological and metabolic signatures.

Methods

Thirty-seven patients with pathologically confirmed lung cancer who underwent electronic bronchoscopy at the Second Hospital of Hebei Medical University between November 2022 and November 2023 were enrolled, including 22 smokers and 15 never-smokers. BALF samples were collected for 16S rRNA amplicon sequencing and untargeted metabolomic profiling. Smoking-associated microbial taxa, metabolites, and cross-omics associations were evaluated using LEfSe, KEGG enrichment, ROC curve analysis, and Spearman correlation analysis.

Results

A total of 1,378 operational taxonomic units were detected, including 1,004 shared OTUs. Alpha diversity did not differ significantly between groups, whereas weighted UniFrac PCoA showed significant separation in microbial community structure (PERMANOVA, $P < 0.05$). Bacteroidota, Actinomycetota, Prevotella, Schaalia, and Mobiluncus were relatively enriched in smokers, whereas Pseudomonadota, Finegoldia, Peptoniphilus, and Paracoccus were enriched in never-smokers. Untargeted metabolomics identified 23 differential metabolites, of which 18 were upregulated and 5 were downregulated in smokers. These metabolites were mainly enriched in glutathione metabolism, alanine metabolism, arginine and proline metabolism, bile secretion, ABC transporters, pantothenate and coenzyme A biosynthesis, and fructose and mannose metabolism. ROC analysis suggested potential discriminatory value for (Z)-11-hexadecenal, spermidine, oxosativene, nicotine N-oxide, and 1,2-dimethylimidazole (AUC 0.706–0.776). Mobiluncus was positively correlated with nicotine N-oxide, whereas Finegoldia was negatively correlated with dihydrothymine.

Conclusions

BALF from smokers with lung cancer showed local microenvironmental alterations characterized by microbial compositional remodeling and metabolic pathway perturbation. The observed Mobiluncus-

nicotine N-oxide and Finegoldia-dihydrothymine associations are candidate microbiome-metabolome features that require validation in larger, well-matched cohorts.

Highlights

BALF was used to investigate the lower-airway microenvironment in patients with lung cancer.

Integrated 16S rRNA sequencing and untargeted metabolomics revealed smoking-associated microbiome-metabolome alterations.

Mobiluncus-nicotine N-oxide and Finegoldia-dihydrothymine associations were identified as candidate smoking-related microecological-metabolic links.

Introduction

Lung cancer remains one of the leading causes of cancer incidence and mortality worldwide [1, 2]. Smoking is a major etiological factor for lung cancer and contributes to carcinogenesis through DNA damage, oxidative stress, chronic inflammation, immune escape, and remodeling of the tissue microenvironment [3–6]. However, some smoking-associated biological effects cannot be fully explained by host genomic alterations or conventional inflammatory pathways alone, suggesting that the lower-airway microbiome and its metabolites may also participate in smoking-related lung cancer biology [7–11].

With advances in high-throughput sequencing, the lung is no longer considered a sterile organ [7, 12–14]. Accumulating evidence links the respiratory microbiome to chronic obstructive pulmonary disease, asthma, pulmonary infection, and lung cancer [7, 8, 15–20]. In patients with lung cancer, microbial diversity, dominant taxa, and predicted functional profiles may differ in lower-airway or tumor-tissue samples. Certain microbes may influence the tumor microenvironment by modulating inflammatory mediators, epithelial barriers, immune-cell infiltration, and tumor metabolism [8–11, 21, 22]. Smoking itself can alter oral, nasopharyngeal, lower-airway, and gut microbial communities and their secreted metabolites, resulting in complex interactions among smoking exposure, microbial ecology, and lung cancer [5, 6, 23, 24].

Most previous studies have focused on microbial composition in lung cancer or on the effect of smoking on the respiratory microbiome alone. Studies integrating BALF microbiota and metabolomics to characterize the smoking-associated local lung cancer microenvironment remain limited [8–11, 21, 22]. Metabolites represent important functional outputs of microbe-host interactions and may more directly reflect the biological state of the local ecosystem [25–33]. Therefore, combined microbiome-metabolome analysis can characterize smoking-associated lung cancer microenvironmental features from both compositional and functional perspectives.

In this study, BALF samples from patients with lung cancer were used to compare microbial composition, diversity, and metabolomic profiles between smokers and never-smokers. We further analyzed correlations between differentially abundant genera and differential metabolites. The aim was to identify candidate smoking-associated microbiome-metabolome features and to provide exploratory evidence for local microenvironmental mechanisms linking smoking to lung cancer.

Materials and Methods

Study population and grouping

This observational study enrolled patients with pathologically confirmed lung cancer who underwent electronic bronchoscopy at the Department of Respiratory and Critical Care Medicine II, the Second Hospital of Hebei Medical University, between November 2022 and November 2023. The study was approved by the institutional ethics committee, and all participants provided written informed consent. Eligible patients had confirmed lung cancer and available BALF samples. Patients were excluded if they had long-term glucocorticoid or immunosuppressive therapy, antibiotic exposure within the preceding 2 months, active pulmonary infectious disease, another systemic malignancy, or prior antitumor treatment, including radiotherapy, chemotherapy, targeted therapy, or immunotherapy. Participants were classified as smokers (S group) or never-smokers (NS group) according to smoking history.

Clinical data and BALF collection

Demographic and clinical data were collected, including age, sex, smoking history, recent antibiotic use, laboratory findings, imaging results, pathological type, and tumor stage. Bronchoscopy was performed under local anesthesia combined with intravenous anesthesia. After inspection of the bronchial tree, sterile saline was instilled into the lesion-related bronchopulmonary segment and aspirated. Approximately 5–10 mL of BALF was collected into sterile containers and stored at -80°C until 16S rRNA amplicon sequencing and untargeted metabolomic analysis.

16S rRNA amplicon sequencing and microbial analysis

BALF samples were transported on dry ice to BGI Genomics (Shenzhen, China). Microbial DNA was extracted using the MagPure Stool DNA KF Kit B according to the standard protocol. After quality control, target-specific primers were used to construct amplicon libraries, which were sequenced on an Illumina paired-end platform. Raw reads were filtered to generate effective sequences. OTUs were clustered at 97% sequence similarity using USEARCH [34], and representative sequences were annotated using the RDP classifier with a confidence threshold of 0.6 [35].

Microbial community structure was assessed using alpha- and beta-diversity analyses. Alpha diversity was evaluated using Chao1, ACE, Shannon, Simpson, and Coverage indices. Beta diversity was assessed by UniFrac-based PCoA [36], and between-group differences were tested using PERMANOVA [37].

Differential taxa were identified using LEfSe with an LDA score threshold > 2 and P < 0.05 [38]. Functional profiles were predicted using PICRUSt2 and mapped to the KEGG pathway database [39, 40].

Untargeted metabolomic profiling

Untargeted metabolomic analysis of BALF samples was performed by BGI Genomics. Frozen samples and pooled quality-control (QC) samples were thawed at 4°C. Each 100-uL aliquot was mixed with 700 uL of pre-cooled extraction solvent (methanol:acetonitrile:water, 4:2:1, v/v/v), vortexed, incubated at -20°C for 2 h, and centrifuged at 25,000 × g for 15 min at 4°C. The supernatant was vacuum-dried, reconstituted in methanol:water (1:1, v/v), vortexed for 10 min, and centrifuged again. QC samples were prepared by pooling 20 uL from each sample.

Metabolites were separated using a Waters UPLC I-Class Plus system and detected with a Q Exactive high-resolution mass spectrometer (Thermo Fisher Scientific, USA). Raw mass-spectrometry data were extracted and normalized using Compound Discoverer 3.3 and annotated against BMDB, mzCloud, and ChemSpider databases. PCA was used to assess sample distribution and data quality. PLS-DA and OPLS-DA were used to evaluate group differences; because the OPLS-DA model showed overfitting, differential metabolites were ultimately screened using PLS-DA VIP values combined with univariate analysis. Differential metabolites were defined as VIP ≥ 1, fold change ≥ 1.2 or ≤ 0.83, and P < 0.05. ROC curve analysis was used to evaluate candidate biomarkers, with AUC > = 0.70 considered potentially informative. Differential metabolic pathways were identified using KEGG enrichment analysis [39, 40].

Cross-omics association and statistical analysis

Spearman rank correlation analysis was performed in R using corr.test to evaluate associations between differentially abundant genera and differential metabolites. Statistical analyses were performed with SPSS 26.0. Categorical variables were compared using the chi-square test or Fisher's exact test. Normally distributed continuous variables were analyzed using the t test or permutation test, whereas non-normally distributed variables were analyzed using the Kruskal-Wallis rank-sum test. A two-sided P < 0.05 was considered statistically significant. Given the imbalance in sex and pathological type between groups, the findings were interpreted as exploratory.

Results

Patient characteristics

A total of 37 patients with lung cancer were included, comprising 22 smokers and 15 never-smokers. The smoking group had a mean age of 65.59 ± 10.15 years (range, 50–86 years), and all participants were male. Histological subtypes included squamous cell carcinoma (n = 8), adenocarcinoma (n = 7), and small-cell lung cancer (n = 7). The non-smoking group had a mean age of 65.47 ± 12.03 years (range,

40–84 years), including 4 men and 11 women; histological subtypes included adenocarcinoma (n = 11), squamous cell carcinoma (n = 2), and small-cell lung cancer (n = 2). Based on the actual counts, age strata did not differ significantly, whereas sex and pathological type were imbalanced between groups, indicating that subsequent findings should be interpreted with caution regarding potential confounding (Table 1).

Table 1
Baseline characteristics of the study population.

Characteristic	Smoking (n = 22)	Never-smoking (n = 15)	P value
Male/female, n	22/0	4/11	< 0.001
Age < = 60/>60 years, n	7/15	4/11	0.753
SCLC/SqCC/ADC/other, n	7/8/7/0	2/2/11/0	0.046

Sequencing coverage and microbial composition

The species accumulation curve approached a plateau, indicating sufficient sample size and sequencing depth for community analysis. Across all BALF samples, 1,378 OTUs were identified, including 192 OTUs unique to smokers, 182 unique to never-smokers, and 1,004 shared OTUs. At the phylum level, Bacillota, Bacteroidota, Actinomycetota, Pseudomonadota, and Fusobacteriota were the dominant taxa in both groups. Pseudomonadota showed lower relative abundance in smokers than in never-smokers (11.58% vs. 22.68%, $P < 0.05$). At the genus level, Streptococcus, Veillonella, Prevotella, Neisseria, Schaalia, Rothia, Haemophilus, and Fusobacterium were dominant. Prevotella (9.46% vs. 5.31%) and Schaalia (6.40% vs. 4.02%) were significantly enriched in smokers ($P < 0.05$).

Microbial diversity and differential taxa

Alpha diversity did not differ significantly between groups. Median Chao1, ACE, Shannon, Simpson, and Coverage values were 553.3, 571.8, 3.57, 0.05, and 0.99 in smokers and 560.5, 563.3, 3.35, 0.08, and 0.99 in never-smokers, respectively (all $P > 0.05$; Table 2). Unweighted UniFrac PCoA did not show significant between-group separation, whereas weighted UniFrac PCoA showed significant separation by PERMANOVA ($P < 0.05$), indicating that smoking status was associated with differences in abundance-weighted BALF microbial community structure (Fig. 1A).

Table 2
Alpha-diversity indices of BALF microbiota.

Index	Smoking	Never-smoking	P value
Chao1	553.3	560.5	0.70252
ACE	571.8	563.3	0.96347
Shannon	3.57	3.35	0.28981
Simpson	0.05	0.08	0.08042
Coverage	0.99	0.99	0.21592

LEfSe analysis identified smoking-associated differential taxa. At the phylum level, Bacteroidota and Actinomycetota were enriched in smokers, whereas Pseudomonadota was enriched in never-smokers. At the genus level, Prevotella, Schaalia, Kingella, Asteroleplasma, Pyramidobacter, Sphaerochaeta, Mobiluncus, and Mogibacterium were enriched in smokers; Finegoldia, Vagococcus, Achromobacter, Peptoniphilus, Hymenobacter, Staphylococcus, Ralstonia, and Paracoccus were enriched in never-smokers ($P < 0.05$; Fig. 1B-C). PICRUSt2 prediction showed that carbohydrate metabolism, amino acid metabolism, lipid metabolism, cofactor and vitamin metabolism, energy metabolism, and replication and repair were abundant pathways in both groups, but no significant between-group difference remained after FDR correction.

Metabolomic quality control and global features

Metabolomic quality-control analysis showed that 5,040 of 6,335 detected features had CV < = 30%, yielding a ratio of 0.80 and indicating good analytical reproducibility. PCA showed tight clustering of QC samples, suggesting low systematic error and stable data quality (Table 3). Among 1,849 retained metabolites, the most abundant categories were lipids (16.45%), amino acids/peptides and analogues (15.49%), and benzene derivatives (14.61%). KEGG super-pathway classification showed that amino acid metabolism (26.49%), lipid metabolism (14.59%), biosynthesis of secondary metabolites (14.59%), and carbohydrate metabolism (14.05%) were the leading pathway categories.

Table 3
Quality-control summary for metabolomic detection.

Total features	Features with CV < = 30%	Ratio
6,335	5,040	0.80

Differential metabolites and candidate biomarkers

PCA and PLS-DA score plots indicated metabolic separation between smokers and never-smokers. Using VIP ≥ 1, fold change ≥ 1.2 or ≤ 0.83, and $P < 0.05$, 23 differential metabolites were identified. Compared with never-smokers, smokers had 18 upregulated and 5 downregulated metabolites (Fig. 2A). KEGG enrichment analysis showed significant enrichment of seven pathways: glutathione metabolism,

alanine metabolism, arginine and proline metabolism, bile secretion, ABC transporters, pantothenate and coenzyme A biosynthesis, and fructose and mannose metabolism (all $P < 0.05$; Fig. 2B).

ROC analysis identified five candidate biomarkers with AUC > = 0.70: (Z)-11-hexadecenal (HMDB0243751; AUC = 0.776), spermidine (B100298; AUC = 0.736), oxosativene (HMDB0258180; AUC = 0.718), nicotine N-oxide (HMDB0001411; AUC = 0.712), and 1,2-dimethylimidazole (HMDB0244082; AUC = 0.706; Fig. 3).

Microbiome-metabolome correlations

Spearman correlation analysis showed significant associations between smoking-associated differential genera and differential metabolites (Fig. 4). Prevotella was positively correlated with mannitol (B100102). Schaalia was positively correlated with 1-ethylaziridine, 2-deoxyglucose, 1,2-dimethylimidazole, and perfluorobutanesulfonate. Mobiluncus was positively correlated with nicotine N-oxide (HMDB0001411). Finegoldia, Peptoniphilus, and Paracoccus were negatively correlated with 2-deoxyglucose (B100208), and Finegoldia was also negatively correlated with dihydrothymine (B200100; all $P < 0.05$). These results suggest that smoking-associated metabolic alterations may be coupled with changes in selected BALF genera, particularly Mobiluncus and Finegoldia.

Discussion

Using BALF samples and integrated 16S rRNA amplicon sequencing with untargeted metabolomics, this study systematically compared lower-airway microbial and metabolic features between smokers and never-smokers with lung cancer. The main findings were as follows. First, alpha diversity did not differ significantly between groups, whereas weighted beta diversity showed significant separation, suggesting that smoking-associated differences were mainly reflected in the relative abundance of dominant taxa and community structure rather than overall richness. Second, Prevotella, Schaalia, and Mobiluncus were enriched in smokers, whereas Finegoldia, Peptoniphilus, and Paracoccus were enriched in never-smokers. Third, smokers showed metabolic perturbations involving amino acid metabolism, redox-related metabolism, and transport functions. Fourth, Mobiluncus was correlated with nicotine N-oxide, and Finegoldia was correlated with dihydrothymine, suggesting that smoking-related chemical exposure may be linked to local microbiome-metabolome remodeling in lung cancer [11–18, 31–33].

Alterations of the lung microbiome in patients with lung cancer have been reported in multiple studies, although findings vary because of differences in sample type, sequencing region, contamination control, pathological subtype, and treatment status [8–11, 21, 22]. Compared with tumor tissue or sputum, BALF can better reflect the lower-airway luminal microenvironment and is less invasive than tissue sampling [15–20]. In the present study, alpha diversity was comparable between groups, whereas weighted UniFrac PCoA showed significant separation. This suggests that smoking may not simply reduce overall microbial richness but may instead alter the relative abundance and ecological occupancy of selected

dominant taxa. This is consistent with prior observations that smoking can reshape respiratory microbial composition without necessarily changing diversity alone [5, 6, 23, 24].

At the compositional level, Prevotella and Schaalia were enriched in smokers. Prevotella is commonly found in the oral cavity and upper respiratory tract and can reach the lower airways through microaspiration; its increase may reflect smoking-related impairment of airway barriers, mucosal immune alteration, or enhanced oral-lung microbial migration [12–14, 19, 20]. Schaalia belongs to Actinobacteria-related taxa, and some members are associated with mucosal inflammation and anaerobic microenvironments. Smoking may reduce local oxygen tension and alter mucus and redox conditions, thereby favoring colonization by anaerobic or facultative anaerobic taxa [5, 6, 23, 24]. In contrast, Finegoldia, Peptoniphilus, and Paracoccus were relatively enriched in never-smokers, but their biological implications require further validation considering pathological subtype, sex, and tumor microenvironmental context.

Metabolomic analysis showed that differential metabolites in BALF from smokers were enriched in glutathione metabolism, alanine metabolism, arginine and proline metabolism, ABC transporters, and coenzyme A-related pathways. Glutathione metabolism is an important defense system against smoking-induced oxidative stress [5, 6]. Arginine and proline metabolism is associated with immune regulation, collagen remodeling, and tumor-cell proliferation. ABC transporters participate in xenobiotic transport and drug resistance, whereas pantothenate and coenzyme A biosynthesis is closely related to fatty acid oxidation and energy metabolism [11–18, 31–33]. Together, these findings suggest that the local lung cancer microenvironment in smokers may involve enhanced oxidative stress, amino acid metabolic reprogramming, and altered handling of exogenous compounds.

The most novel finding of this study is the microbiome-metabolome association. Nicotine N-oxide is a nicotine-related metabolite, and its positive correlation with Mobiluncus suggests that tobacco-related metabolic exposure may coexist with enrichment of selected microbial taxa [5, 6]. Mobiluncus is an anaerobic or microaerophilic genus more commonly studied in urogenital microbial ecology. Its smoking-associated enrichment in BALF from patients with lung cancer and its correlation with nicotine-related metabolites warrant further investigation. Finegoldia was negatively correlated with dihydrothymine, which may be related to pyrimidine metabolism, cell turnover, or local immune status. Importantly, correlation does not establish causality; these findings should be regarded as candidate hypotheses for future mechanistic studies.

This study has potential translational implications. BALF can be obtained during clinically indicated bronchoscopy. If validated in larger cohorts, smoking-associated microbial and metabolic signatures may help define local microenvironmental subtypes in lung cancer, explain smoking-related tumor heterogeneity, and provide clues for microbiome-based intervention, metabolic pathway targeting, or treatment-response prediction. However, the AUC values of the candidate metabolites were moderate, and the sample size was limited; therefore, these markers should not be considered diagnostic tools at this stage but rather exploratory candidates for validation.

Several limitations should be acknowledged. First, the sample size was small, limiting statistical power. Second, all smokers were male, and pathological subtype distribution was imbalanced between groups; sex and histology may confound microbiome and metabolomic profiles. Third, smoking pack-years, time since cessation, second-hand smoke exposure, and e-cigarette use were not stratified. Fourth, 16S rRNA amplicon sequencing generally resolves taxa at the genus level and cannot provide strain-level or direct functional-gene evidence [34, 35, 41]. Fifth, BALF is a low-biomass sample type; although recent infection and antibiotic exposure were excluded, future studies should incorporate more stringent negative controls and contamination-control procedures [42, 43]. Finally, the cross-sectional design precludes determination of temporal sequence or causality among smoking, microbial alteration, metabolic changes, and the lung cancer microenvironment. Larger studies with matching for sex, age, pathological subtype, and stage, combined with metagenomics, targeted metabolomics, inflammatory profiling, and functional experiments, are needed to validate key microbiome-metabolome axes.

Conclusion

BALF from smokers with lung cancer showed alterations in microbial composition and metabolomic profiles, characterized by differences in weighted beta diversity, enrichment of smoking-associated taxa, and perturbation of oxidative stress-, amino acid metabolism-, and transport-related pathways. The Mobiluncus-nicotine N-oxide and Finegoldia-dihydrothymine associations may represent candidate features of smoking-related microbiome-metabolome networks in the lung cancer microenvironment. These exploratory cross-omics findings require confirmation in larger, well-matched cohorts and mechanistic studies.

Declarations

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the Second Hospital of Hebei Medical University. All participants provided written informed consent after being informed of the study purpose, procedures, and potential risks. The ethics approval number should be added before submission if required by the institutional approval letter.

Clinical trial number: not applicable.

Consent for publication

Not applicable. No individual-level identifying information or images are included in this manuscript.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request, subject to institutional ethics and data-management requirements. Raw 16S rRNA sequencing data should be deposited in an appropriate public repository, such as the NCBI

Sequence Read Archive, before publication, and the accession number should be added to this statement.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by Hebei Provincial Department of Finance. (Grant numbers ZF2023129)

Authors' contributions

Writing – original draft: Yanchao Liu. Runlu Wang.

Data collection and sorting:Jiawen Liu, Miaomiao Cui

Writing – review & editing: Zhang Duan.

Acknowledgements

Not applicable.

References

1. 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. 10.3322/caac.21660.
2. Xia C, Dong X, Li H, Cao M, Sun D, He S, et al. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. Chin Med J (Engl). 2022;135(5):584–90. 10.1097/CM9.0000000000002108.
3. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene. 2002;21(48):7435–51. 10.1038/sj.onc.1205803.
4. Bade BC, Dela Cruz CS. Lung cancer 2020: epidemiology, etiology, and prevention. Clin Chest Med. 2020;41(1):1–24. 10.1016/j.ccm.2019.10.001.
5. Hecht SS. Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst. 1999;91(14):1194–210. 10.1093/jnci/91.14.1194.
6. Hecht SS. Lung carcinogenesis by tobacco smoke. Int J Cancer. 2012;131(12):2724–32. 10.1002/ijc.27816.
7. Dickson RP, Erb-Downward JR, Huffnagle GB. The role of the bacterial microbiome in lung disease. Expert Rev Respir Med. 2013;7(3):245–57. 10.1586/ers.13.24.

8. Najafi S, Abedini F, Azimzadeh Jamalkandi S, Shariati P, Ahmadi A, Gholami Fesharaki M. The composition of lung microbiome in lung cancer: a systematic review and meta-analysis. BMC Microbiol. 2021;21(1):315. 10.1186/s12866-021-02375-z.
9. Dong Q, Chen ES, Zhao C, Jin C. Host-microbiome interaction in lung cancer. Front Immunol. 2021;12:679829. 10.3389/fimmu.2021.679829.
10. Jin C, Lagoudas GK, Zhao C, Bullman S, Bhutkar A, Hu B, et al. Commensal microbiota promote lung cancer development via γδ T cells. Cell. 2019;176(5):998–e101316. 10.1016/j.cell.2018.12.040.
11. Dong H, Tan Q, Xu Y, Zhu Y, Yao Y, Wang Y, et al. Convergent alteration of lung tissue microbiota and tumor cells in lung cancer. iScience. 2022;25(1):103638. 10.1016/j.isci.2021.103638.
12. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. Am J Respir Crit Care Med. 2011;184(8):957–63. 10.1164/rccm.201104-0655OC.
13. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. Am J Respir Crit Care Med. 2013;187(10):1067–75. 10.1164/rccm.201210-1913OC.
14. Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Gao Z, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. Microbiome. 2013;1(1):19. 10.1186/2049-2618-1-19.
15. Dickson RP, Erb-Downward JR, Huffnagle GB. Homeostasis and its disruption in the lung microbiome. Am J Physiol Lung Cell Mol Physiol. 2015;309(10):L1047–55. 10.1152/ajplung.00279.2015.
16. Huffnagle GB, Dickson RP, Lukacs NW. The respiratory tract microbiome and lung inflammation: a two-way street. Mucosal Immunol. 2017;10(2):299–306. 10.1038/mi.2016.108.
17. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, et al. Analysis of the lung microbiome in the healthy smoker and in COPD. PLoS ONE. 2011;6(2):e16384. 10.1371/journal.pone.0016384.
18. Sze MA, Dimitriu PA, Hayashi S, Elliott WM, McDonough JE, Gosselink JV, et al. The lung tissue microbiome in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2012;185(10):1073–80. 10.1164/rccm.201111-2075OC.
19. Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, et al. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. mBio. 2015;6(2):e00037. 10.1128/mBio.00037-15.
20. Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, Huffnagle GB, Schmidt TM. Application of a neutral community model to assess structuring of the human lung microbiome. mBio. 2015;6(1):e02284–14. 10.1128/mBio.02284-14.
21. Zhao Y, Liu Y, Li S, Peng Z, Liu X, Chen J, et al. Role of lung and gut microbiota on lung cancer pathogenesis. J Cancer Res Clin Oncol. 2021;147(8):2177–86. 10.1007/s00432-021-03644-0.

22. Zheng X, Lu X, Hu Y. Distinct respiratory microbiota associates with lung cancer clinicopathological characteristics. Front Oncol. 2023;13:847182. 10.3389/fonc.2023.847182.
23. Shapiro H, Goldenberg K, Ratiner K, Elinav E. Smoking-induced microbial dysbiosis in health and disease. Clin Sci (Lond). 2022;136(18):1371–87. 10.1042/CS20220175.
24. Wu J, Peters BA, Dominianni C, Zhang Y, Pei Z, Yang L, et al. Cigarette smoking and the oral microbiome in a large study of American adults. ISME J. 2016;10(10):2435–46. 10.1038/ismej.2016.37.
25. Chen Y, Ma Z, Min L, Li H, Wang B, Zhong J, Dai L. Biomarker identification and pathway analysis by serum metabolomics of lung cancer. Biomed Res Int. 2015;2015:183624. 10.1155/2015/183624.
26. Goveia J, Pircher A, Conradi LC, Kalucka J, Lagani V, Dewerchin M, et al. Meta-analysis of clinical metabolic profiling studies in cancer: challenges and opportunities. EMBO Mol Med. 2016;8(10):1134–42. 10.15252/emmm.201606798.
27. Wishart DS. Metabolomics for investigating physiological and pathophysiological processes. Physiol Rev. 2019;99(4):1819–75. 10.1152/physrev.00035.2018.
28. Wishart DS, Guo A, Oler E, Wang F, Anjum A, Peters H, et al. HMDB 5.0: the Human Metabolome Database for 2022. Nucleic Acids Res. 2022;50(D1):D622–31. 10.1093/nar/gkab1062.
29. Wu WS, Wu HY, Wang PH, Chen TY, Chen KR, Chang CW, et al. LCMD: Lung Cancer Metabolome Database. Comput Struct Biotechnol J. 2022;20:65–78. 10.1016/j.csbj.2021.12.002.
30. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol. 2016;17(7):451–9. 10.1038/nrm.2016.25.
31. Fiehn O. Metabolomics–the link between genotypes and phenotypes. Plant Mol Biol. 2002;48(1–2):155–71. 10.1023/A:1013713905833.
32. Nicholson JK, Lindon JC. Systems biology: metabonomics. Nature. 2008;455(7216):1054–6. 10.1038/4551054a.
33. Patti GJ, Yanes O, Siuzdak G, Innovation. Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012;13(4):263–9. 10.1038/nrm3314.
34. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26(19):2460–1. 10.1093/bioinformatics/btq461.
35. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73(16):5261–7. 10.1128/AEM.00062-07.
36. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005;71(12):8228–35. 10.1128/AEM.71.12.8228-8235.2005.
37. Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecol. 2001;26(1):32–46. 10.1111/j.1442-9993.2001.01070.pp.x.
38. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):R60. 10.1186/gb-2011-12-6-r60.

39. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol. 2020;38(6):685–8. 10.1038/s41587-020-0548-6.
40. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28(1):27–30. 10.1093/nar/28.1.27.
41. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol. 2007;45(9):2761–4. 10.1128/JCM.01228-07.
42. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 2014;12:87. 10.1186/s12915-014-0087-z.
43. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in low microbial biomass microbiome studies: issues and recommendations. Trends Microbiol. 2019;27(2):105–17. 10.1016/j.tim.2018.11.003.

Figures

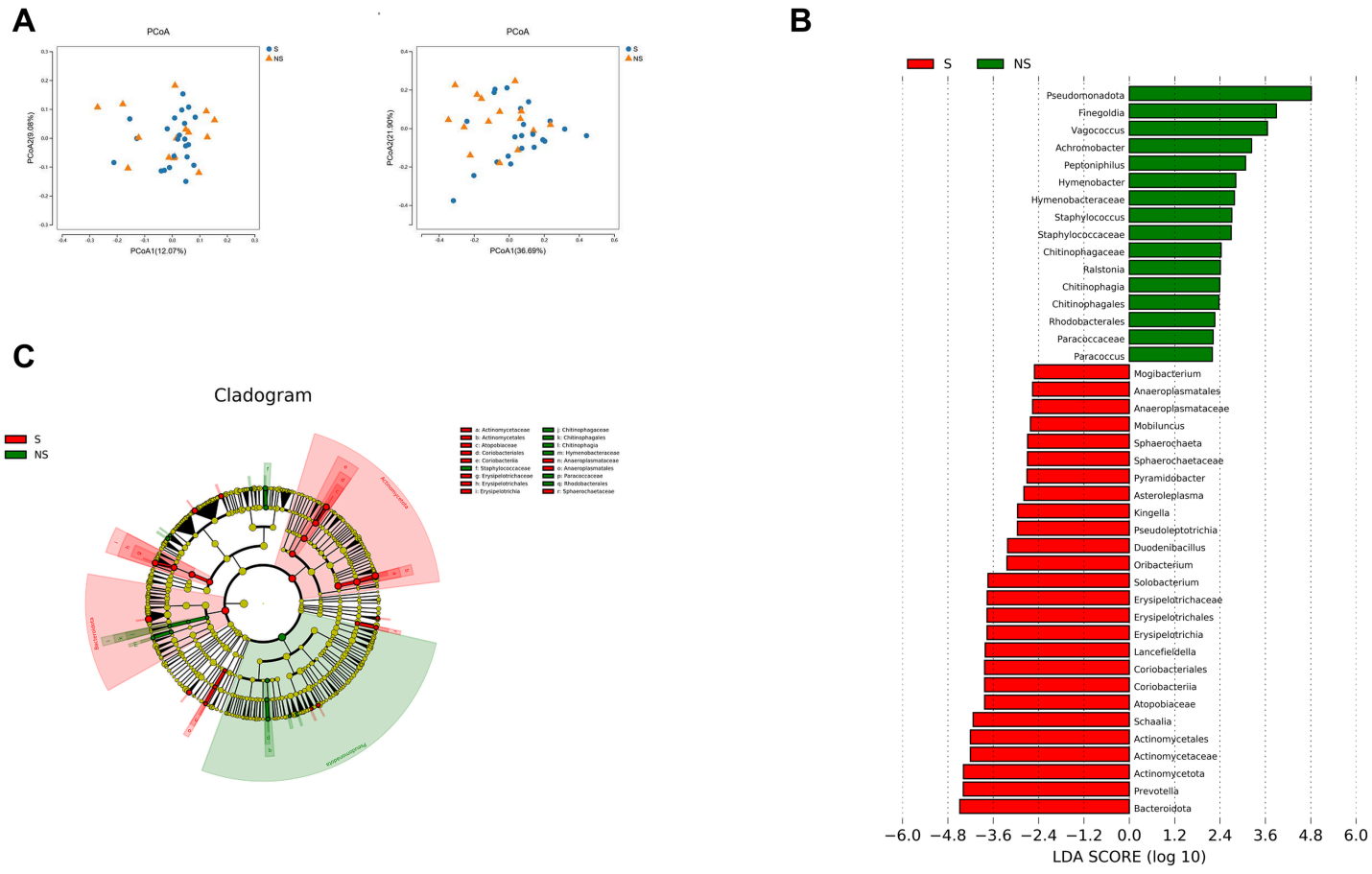


Figure 1

BALF beta diversity and differential microbial taxa. A: UniFrac-based PCoA analysis; B: LEfSe LDA score plot; C: LEfSe cladogram.

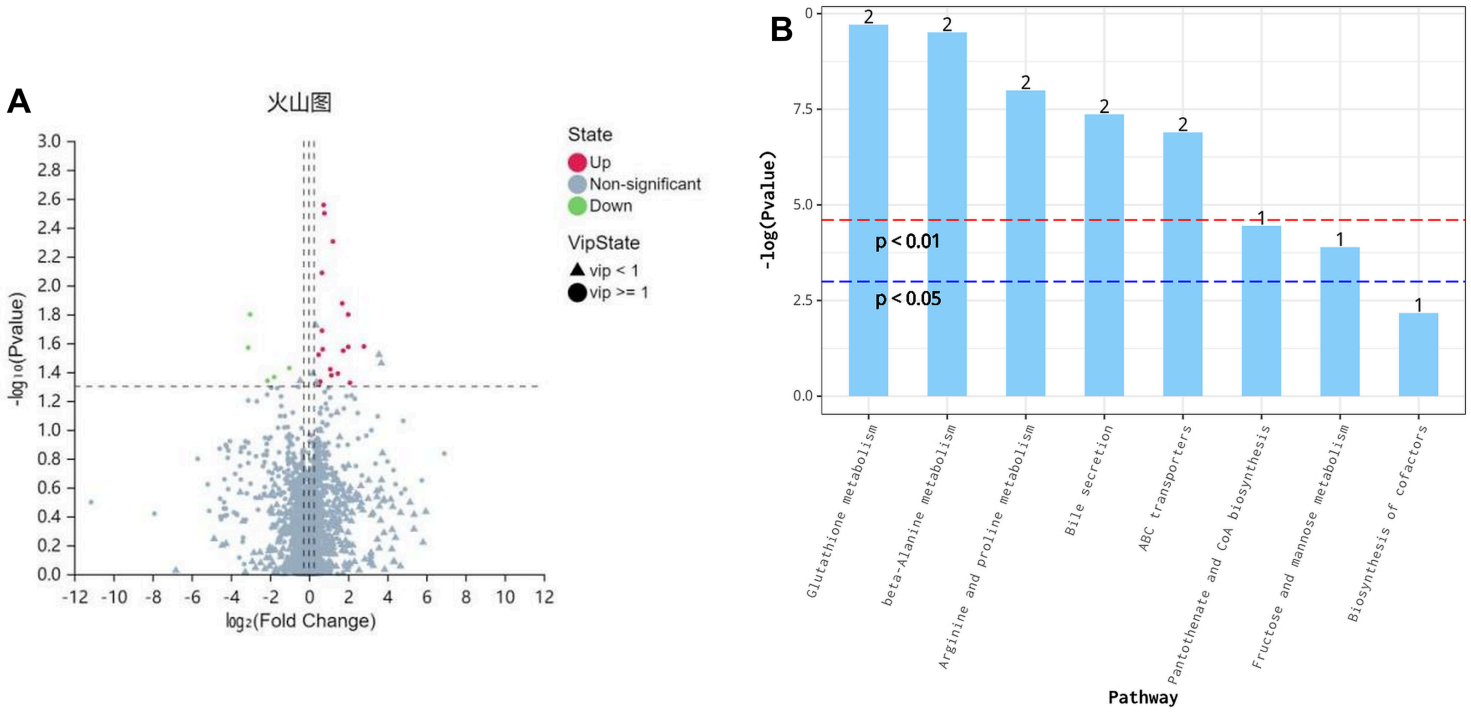


Figure 2

Smoking-associated differential metabolites and pathway enrichment. A: volcano plot of differential metabolites; B: KEGG pathway enrichment analysis.

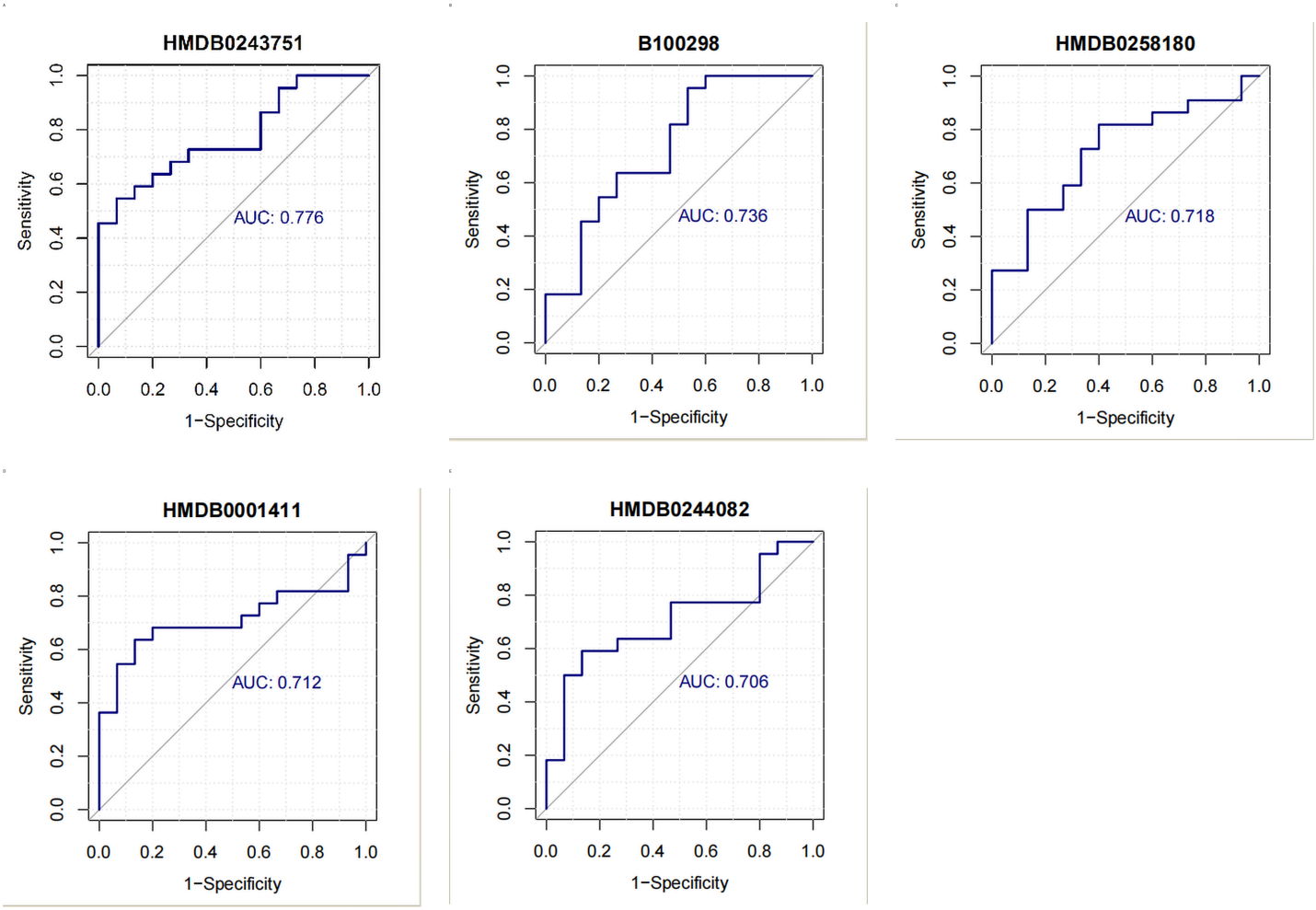


Figure 3

ROC curves of candidate differential metabolites, including (Z)-11-hexadecenal, spermidine, oxosativene, nicotine N-oxide, and 1,2-dimethylimidazole.

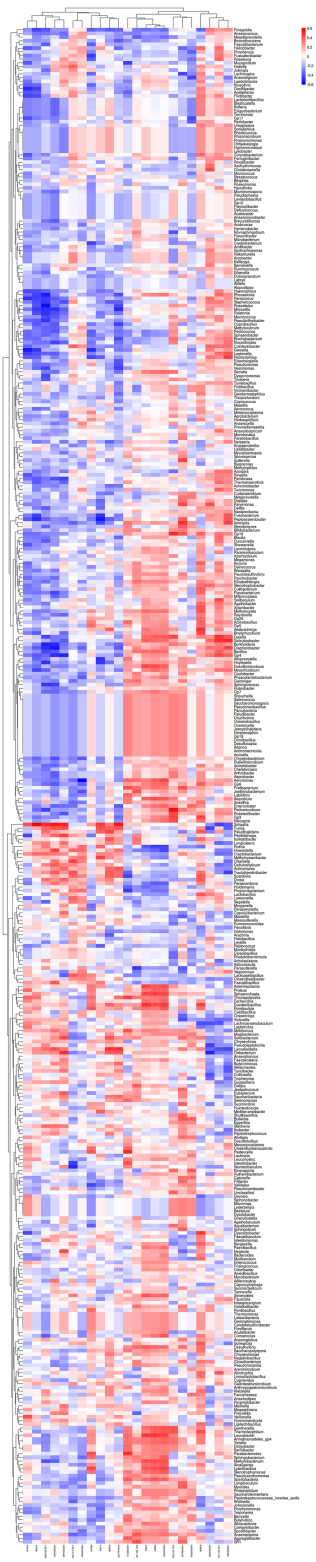


Figure 4

Spearman correlation heatmap of differential genera and differential metabolites.