

Identification of lung cancer breath biomarkers based on perioperative breathomics testing: A prospective observational study

Peiyu Wang,^{a,1} Qi Huang,^{b,1} Shushi Meng,^{c,1} Teng Mu,^b Zheng Liu,^a Mengqi He,^d Qingyun Li,^d Song Zhao,^{b,*} Shaodong Wang,^{a,**} and Mantang Qiu^{a,d,**}

^aDepartment of Thoracic Surgery, Peking University People's Hospital, No. 11 Xizhimen South Street, Beijing 100044, China

^bDepartment of Thoracic Surgery, The First Affiliated Hospital of Zhengzhou University, 1 Jianshe East Road, Zhengzhou, Henan 450003, China

^cDepartment of Thoracic Surgery, Beijing Haidian Hospital, Beijing 100080, China

^dBreax Laboratory, PCAB Research Center of Breath and Metabolism, Beijing 100074, China

Summary

Background Breathomics testing has been considered a promising method for detection and screening for lung cancer. This study aimed to identify breath biomarkers of lung cancer through perioperative dynamic breathomics testing.

Methods The discovery study was prospectively conducted between Sept 1, 2020 and Dec 31, 2020 in Peking University People's Hospital in China. High-pressure photon ionisation time-of-flight mass spectrometry was used for breathomics testing before surgery and 4 weeks after surgery. 28 volatile organic compounds (VOCs) were selected as candidates based on a literature review. VOCs that changed significantly postoperatively in patients with lung cancer were selected as potential breath biomarkers. An external validation was conducted to evaluate the performance of these VOCs for lung cancer diagnosis. Multivariable logistic regression was used to establish diagnostic models based on selected VOCs.

Findings In the discovery study of 84 patients with lung cancer, perioperative breathomics demonstrated 16 VOCs as lung cancer breath biomarkers. They were classified as aldehydes, hydrocarbons, ketones, carboxylic acids, and furan. In the external validation study including 157 patients with lung cancer and 368 healthy individuals, patients with lung cancer showed elevated spectrum peak intensity of the 16 VOCs after adjusting for age, sex, smoking, and comorbidities. The diagnostic model including 16 VOCs achieved an area under the curve (AUC) of 0.952, sensitivity of 89.2%, specificity of 89.1%, and accuracy of 89.1% in lung cancer diagnosis. The diagnostic model including the top eight VOCs achieved an AUC of 0.931, sensitivity of 86.0%, specificity of 87.2%, and accuracy of 86.9%.

Interpretation Perioperative dynamic breathomics is an effective approach for identifying lung cancer breath biomarkers. 16 lung cancer-related breath VOCs (aldehydes, hydrocarbons, ketones, carboxylic acids, and furan) were identified and validated. Further studies are warranted to investigate the underlying mechanisms of identified VOCs.

Funding National Natural Science Foundation of China (82173386) and Peking University People's Hospital Scientific Research Development Funds (RDH2021-07).

Copyright © 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Lung cancer; Breathomics; Volatile organic compounds; Diagnosis

*Corresponding authors at: Department of Thoracic Surgery, the First Affiliated Hospital of Zhengzhou University, 1 Jianshe East Road, Zhengzhou, Henan 450003, China.

**Corresponding authors at: Department of Thoracic Surgery, Peking University People's Hospital, No. 11 Xizhimen South Street, Beijing 100044, China

E-mail addresses: r8339979852@163.com (P. Wang), zhaosongzhu1752@163.com (S. Zhao), 13651217347@163.com (S. Wang), qiumantang@163.com (M. Qiu).

¹ These authors contributed equally to this work.

Research in context

Evidence before this study

Breathomics testing has been considered a promising method for lung cancer screening. We searched PubMed, Embase, and Cochrane Library databases up until Dec 31, 2021 to identify studies focusing on breathomics testing for lung cancer with the search terms “lung cancer”, “breathomics”, “volatile organic compound (VOC)”, and “diagnosis”. Four reviews and 42 cross-sectional studies were finally selected. Among these studies, lung cancer breath biomarkers are highly heterogeneous because of different sample collection, patient condition, test environment, and analysis methods. No breathomics tests are clinically available for lung cancer screening.

Added value of this study

Different from the comparison between patients with lung cancer and healthy individuals or benign nodules in previous studies, this study identified lung cancer breath biomarkers through perioperative dynamic breathomics testing. We focused on 28 VOC candidates that have been previously reported and identified 16 VOCs as lung cancer biomarkers. The external validation confirmed good performance of these biomarkers in lung cancer detection. This study helped to solve the heterogeneity among published studies through establishing groups of 16 VOCs and eight VOCs for lung cancer screening.

Implications of all the available evidence

A large gap exists between breathomics research and clinical practices in lung cancer detection and screening. Breathomics testing for the validated 16 VOCs, mainly aldehydes and hydrocarbon, showed potential for promoting the current lung cancer screening strategy. More scientific studies are warranted to investigate the underlying mechanisms of identified lung cancer VOCs.

Introduction

Lung cancer is the leading cause of cancer-related mortality worldwide.¹ Early diagnosis and treatment are critical for improving lung cancer survival. However, the detection and diagnosis of lung cancer at an early stage is still challenging because of the lack of clinical manifestations and specific biomarkers.² Annual low-dose chest computed tomography (LDCT) screening has been proven to substantially reduce lung cancer specific mortality.^{3,4} However, the application of LDCT screening is challenging in multiple aspects. High costs, radiation exposure, and a high false-positive rate hinder the application of LDCT for large-scale screening.^{2,5} A highly accurate and non-invasive tool for lung cancer screening is urgently warranted.

Breathomics is considered a promising method for lung cancer screening.^{6–8} The altered genome and transcriptome during carcinogenesis and cancer progression will lead to dysregulated metabolic pathways and the accumulation of aberrant metabolites.⁹ Among numerous metabolites, cancer-derived volatile organic compounds (VOCs) can diffuse into alveoli and can be detected in exhaled breath.¹⁰ The VOC concentration in exhaled breath has been demonstrated to be representative of the blood concentration.⁹ Gas chromatography-mass spectrometry (GC–MS) is currently regarded as the gold standard for the identification and quantification of breath biomarkers in human breath.¹⁰ Most studies identify lung cancer-specific VOCs by comparing patients with lung cancer with healthy individuals or benign nodules with cross-sectional designs using GC–MS.^{8–11} However, the results are highly heterogeneous because of the differentiated sample collection strategy, patient condition, and test environment.

Despite the mature technology of GC–MS, the tedious pretreatment steps and time-consuming detection process limit its application.¹² Other direct mass spectrometry, such as secondary electrospray ionization,¹³ selected-ion-flow-tube,¹⁴ and proton-transfer-reaction,¹⁵ has also been used for rapid detection of exhaled breath; however, the vast amount of water vapor in exhaled breath makes the ionization process more intricate and increases the complexity for data analysis. In contrast, the high-pressure photon ionization time-of-flight mass spectrometry (HPPI-TOFMS) is promisingly highlighted for breath testing with advantages of high sensitivity, no sample pretreatment, and good tolerance to humidity.^{16,17} The property of HPPI-TOFMS in compound testing and identification has already been validated in monitoring exhaled propofol concentration during surgery,¹⁸ detecting volatile metabolites in urine,¹⁹ and categorizing mixed flavor compounds.²⁰ We have preliminarily confirmed the application of HPPI-TOFMS in exhaled breath testing for lung cancer and esophageal cancer detection in cross-sectional studies.^{21,22}

In this study, we aimed to identify specific VOCs for lung cancer by dynamically testing perioperative breathomics. An external cross-sectional validation was conducted to evaluate the performance of these VOCs for lung cancer diagnosis.

Methods

Study design

The overall study design is shown in [Figure 1](#). Briefly, perioperative breath samples were prospectively collected in a discovery study to identify candidate VOCs, then the diagnostic value of these VOCs was tested in an independent validation study consisting of patients with lung cancer and healthy volunteers.

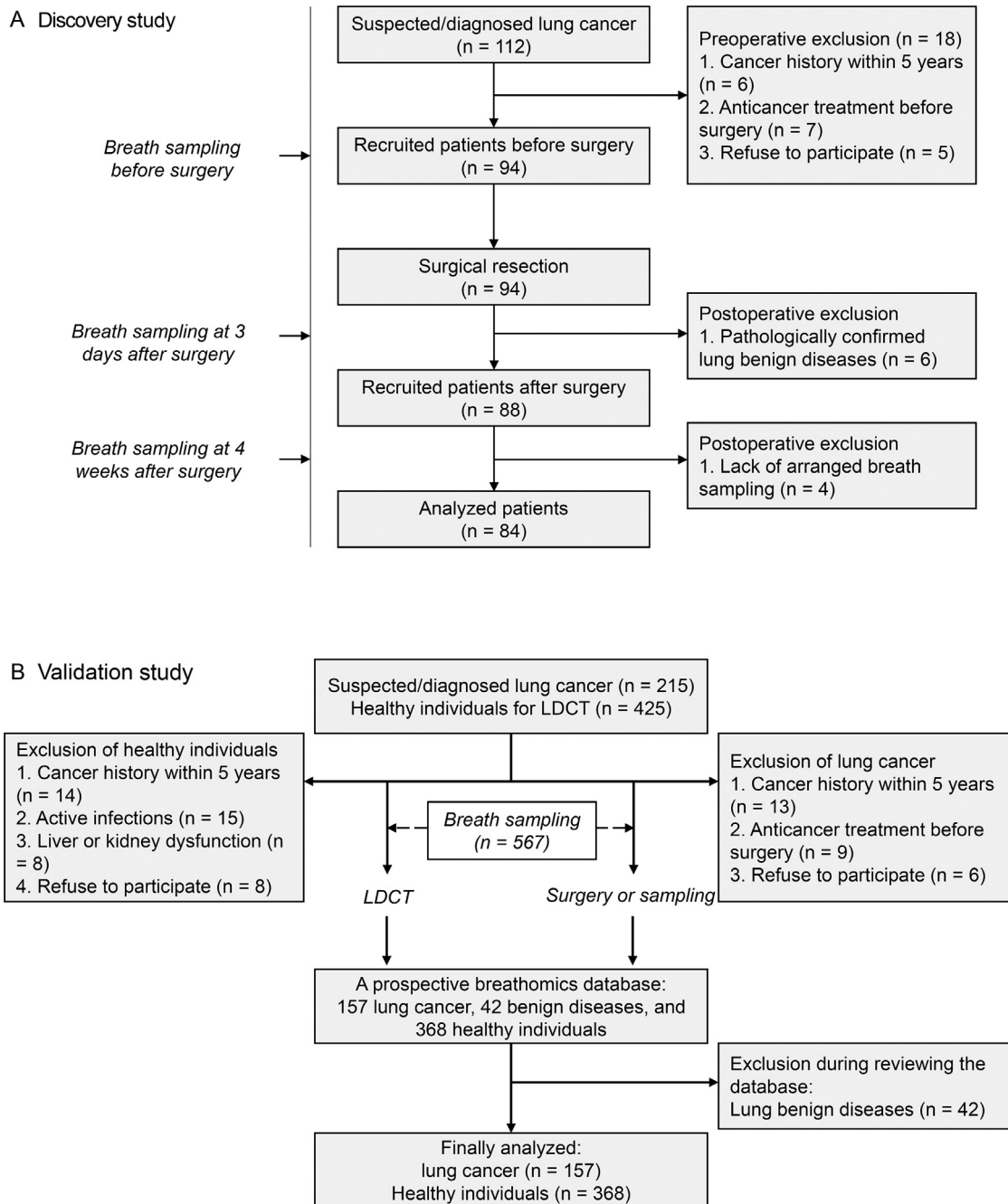


Figure 1. Flowchart of participant recruitment.

The detection study (A) and validation study (B) are shown. LDCT: low-dose chest computed tomography.

The discovery study was conducted at Peking University People's Hospital between Sept 1, 2020 and Dec 31, 2020. Exhaled breath samples were collected at three timepoints, i.e., the morning on the day of surgery, 3 days after surgery, and 4 weeks after surgery. The validation study was performed by reviewing the prospective breathomics database established between July 1, 2020 and Oct 30, 2020 in the First

Affiliated Hospital of Zhengzhou University. This study was approved by the Ethics Committee Board of Peking University People's Hospital (2019PHB095-01). All patients were informed of the study protocol, and written consent was obtained before entering the study. The study was reported following the Standards for Reporting of Diagnostic Accuracy (STARD) reporting guidelines.²³

Participants

Regarding the discovery study, patients with suspected pulmonary lesions, age ≥ 18 years, and with a plan to undergo surgical resection were consecutively recruited. The preoperative exclusion criteria were as follows: (1) cancer history within 5 years; (2) anticancer treatment before surgery; (3) active infections; (4) liver or kidney dysfunction; and (5) lack of written informed consent to participate. Postoperative exclusion criteria were as follows: (1) pathologically confirmed benign lung diseases; (2) lack of planned breath sampling; and (3) incidence of major perioperative complications which may affect breathomics testing. Regarding the validation study, patients with lung cancer were enrolled and excluded according to the same criteria of the discovery study, but breath sample was collected only once before surgery. Healthy individuals consisted of people who underwent LDCT for physical examination but without indicators for lung pathology.

Breath collection

The breath collection protocol has been thoroughly reported.^{21,22} Briefly, all exhaled breath samples were collected by trained investigators using prepared Tedlar (DuPont de Nemours) air bags. On the night before breath collection, the Tedlar bags were baked at 60 °C for 3 h to fully release possible contaminants and continuously purged with high-purity nitrogen four times. Breath samples were collected in a fixed room, and the corresponding environmental air was collected. Participants first gargled with pure water and then performed a single deep nasal inhalation followed by complete exhalation via their mouth into Tedlar bags. A total of 1000 mL of exhaled breath was collected. A CO₂ sensor was used to ensure that alveolar air was collected: exhaled breath collection began once the CO₂ sensor detected a CO₂ concentration exceeding 4%. For patients in the discovery study, exhaled breath samples were collected on the morning of the planned surgery and postoperative day 3 in the hospital. Patients were invited back to the hospital for the third breath sampling four weeks after surgery, or a home-designed sampling equipment was mailed for breath sample collection. For patients with lung cancer in the validation study, exhaled breath samples were collected on the morning of the planned surgery or biopsy in the hospital. Exhaled breath of healthy individuals was collected on the same day of physical examination before LDCT examination at the outpatient department. All participants were required to fast for at least 8 h and not to ingest spicy food, alcohol, or coffee the night before exhaled breath collection.

HPPI-TOFMS analysis

The design, structure, and characteristics of HPPI-TOFMS have been thoroughly reported.¹⁷ Generally, HPPI-TOFMS consists of a vacuum ultraviolet lamp-based HPPI ion source and an orthogonal acceleration

time-of-flight (TOF) mass analyzer. The TOF mass analyzer had a mass resolution of 4000 (full width half maximum) at a mass-to-charge ratio (m/z) of 92 achieved with a 0.4 m field-free drift tube. A gas-phase exhaled breath sample was directly introduced into the ionization region through a stainless-steel capillary from the air bag. To eliminate condensation of exhaled VOCs and minimize surface adsorption, the stainless-steel capillary was heated to 100 °C and the HPPI ion source was heated to 60 °C. The TOF signals were recorded by a 400-picosecond time-to-digital converter rate at 25 kHz, and all the mass spectra were accumulated for 60 s. All mass spectrometers require mass calibration before they are put to use. Typically, 1,2-dichloroethylene, perchloroethylene, and hexachloro-1,3-butadiene, as well-known m/z values, evenly distributed over the mass range of interest. The calibration formula $y=ax^2+bx+c$ was used to realize the conversion of time of flight to m/z . This calibration process is accomplished by the mass spectrometer's software. Mass spectrum peaks detected by HPPI-TOFMS with m/z less than 500 were recorded. Noise reducing and baseline correction, and VOCs feature detection were further completed for the mass spectrum data pre-processing. Ambient background air data were subtracted from that of exhaled breath samples, and the obtained data were used for further analysis.

Data analysis

Considering the varieties of VOCs in exhaled breath, we reviewed the original studies included in four relevant reviews to select potential breath biomarkers of lung cancer.^{8–11} Lung cancer-related VOCs that had been reported in not less than two original studies were selected as potential breath biomarkers. Finally, a total of 28 VOCs were selected for further validation (Table S1), they are mainly hydrocarbons (aromatic and aliphatic) and oxygenated compounds (aldehydes, alcohols, phenols, carboxylic acids, ethers, and furans). The identification of VOCs in the mass spectrum was based on their m/z values and the ionization model of HPPI-TOFMS.^{16,17} The peak intensity of extracted VOC was calculated and selected as an indicator for concentration. The breathomics testing results before surgery were compared with those after surgery. VOCs showing significant changes in peak intensity were further screened as potential breath biomarkers for lung cancer. Then, we examined the performance of these VOCs in distinguishing patients with lung cancer from healthy individuals in the validation study. We specifically investigated the association between VOCs and cancer types and cancer stage in patients with lung cancer.

Statistical analysis

Categorical data are presented as frequencies (percentages), and continuous data are presented as the means (standard deviations) or medians (interquartile ranges).

Changes of VOCs' perioperative peak intensity were assessed using Wilcoxon matched-pairs signed-rank sum tests. Differences between groups were compared with ANOVA, Pearson's chi-square test, Mann–Whitney U test, or Kruskal–Wallis test. Multivariate logistic regression models with backward conditional methods were used to investigate the association between lung cancer and VOCs peak intensities. Predictive models based on multivariable logistic regression were used to assess the diagnostic performance of VOCs for lung cancer. The Hosmer–Lemeshow test was used to check the goodness of fit of predictive models. Sensitivity, specificity, accuracy, positive predictive value, and negative predictive value were calculated. Receiver operating characteristic curves were generated, and the area under the curve (AUC) was calculated. A two-tailed *P* value < 0.05 was considered significant. All analyses were conducted using IBM SPSS Statistics for Windows (version 24.0, IBM Corp., Armonk, NY, USA) and the R tool (version 4.1.2).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or submitting for publication. All authors had full access to all the data in the study and accept responsibility to submit for publication.

Results

Identification of lung cancer biomarkers by perioperative breathomics

In the discovery study, a total of 112 patients undergoing surgery were enrolled. 18 patients were excluded before surgery, with reasons shown in Figure 1A. Six patients were diagnosed with benign disease based on postoperative pathology and were thus excluded. Another four patients were excluded because of the lack of exhaled breath sampling four weeks after surgery. Major complications or mortalities had not been observed. 84 patients with lung cancer, consisting of 27 male patients and 57 female patients, were finally included for data analysis (Table 1). The average age was 55.0 years (standard deviation: 10.2). Most patients were diagnosed with adenocarcinoma at an early pathological stage. Wedge resection was most frequently adopted, followed by lobectomy and segmentectomy.

The peak intensity of 28 VOCs at three time points are shown in Table 2. Most VOCs showed fluctuant changes among the three timepoints. Perioperative drug metabolism and surgical stress could impact postoperative breathomics testing at three days after surgery. Patients' physiological status mostly returned to normal four weeks after surgery based on our clinical practice. Therefore, we focused on the comparison of breathomics

Characteristics	Data, N (%)
Demographic data	
Age, years, means ± SD	55.0 ± 10.2
Gender (female)	57 (67.9)
Body mass index, kg/m ² , medians (IQR)	24.1 (22.9–26.3)
Smoking history	12 (14.3)
Comorbidities	
Diabetes	9 (10.7)
Cardiovascular disease	22 (26.2)
Cerebrovascular disease	3 (3.6)
Respiratory disease	2 (2.4)
Thyroid disease	10 (11.9)
Surgical procedures	
Lobectomy	35 (41.7)
Segmentectomy	4 (4.8)
Wedge resection	45 (53.6)
Pathological data	
Cancer type	
Adenocarcinoma	80 (95.2)
Squamous cell carcinoma	1 (1.2)
Small cell carcinoma	3 (3.6)
Pathological stage	
IA1/IA2/IA3/II-III	47/18/10/9 (56.0/21.4/11.9/10.7)
Lymph node metastasis	9 (10.7)
Multiple primary cancer	10 (11.9)

Table 1: Patient characteristics (N = 84).
IQR: interquartile range; N: number; SD: standard deviation.

before surgery and four weeks after surgery. An example of the different mass spectra at the two time points is shown in Figure 2A. Based on Wilcoxon matched-pairs signed-rank sum tests, 15 VOCs showed a significant reduction in peak intensity during the period, and they were 2-hydroxyacetaldehyde, isoprene, pentanal, butyric acid, toluene, 2,5-dimethylfuran, cyclohexanone, hexanal, heptanal, acetophenone, propylcyclohexane, octanal, nonanal, decanal, and 2,2-dimethyldecane. However, the peak intensity of acetaldehyde showed a significant increase at four weeks after surgery. An example of the identification of these 16 VOCs in mass spectrum is shown in Figure 2B. Dynamic changes in peak intensities of these VOCs at three time points were shown in Figure 3. These 16 VOCs were selected as candidate lung cancer breath biomarkers for further analysis.

Diagnosis validation

In the validation study, we included 157 patients with lung cancer and 368 healthy individuals (Figure 1B). Not surprisingly, patients with lung cancer were older and had a higher prevalence rate of cardiovascular diseases than healthy individuals (Table 3). Smoking history and alcohol consumption were more frequently observed in healthy individuals. Comparisons of the

VOC	Peak intensity of HPPI-TOFMS			P values	
	Pre-surgery	PO 3 days	PO 4 weeks	Pre vs. PO 3d	Pre vs. PO 4w
Acetaldehyde	137 (91–194)	119 (79–193)	170 (108–273)	0.71	0.024
Ethanol	91 (30–285)	99 (12–345)	65 (12–423)	0.65	0.81
Propionaldehyde	2521 (1822–4103)	2036 (1384–3161)	2334 (1839–3565)	0.007	0.44
Propanol	91 (41–141)	66 (40–145)	80 (40–136)	0.45	0.76
2-Hydroxyacetaldehyde	782 (557–1134)	524 (418–743)	680 (471–879)	<0.001	0.025
Dimethyl sulfide	135 (75–213)	56 (41–86)	156 (89–310)	<0.001	0.058
Isoprene	6062 (3108–7857)	4399 (2156–6978)	3838 (1769–5733)	0.25	0.017
Butanal	377 (246–1080)	487 (186–1201)	256 (115–976)	0.97	0.37
Benzene	97 (73–189)	92 (66–186)	107 (81–179)	0.72	0.82
Pentanal	112 (81–191)	96 (61–164)	76 (49–142)	0.092	0.003
Butyric acid	298 (228–375)	281 (164–383)	188 (132–360)	0.42	0.012
Toluene	3345 (1277–8142)	4929 (1619–18,836)	1091 (360–5563)	0.070	0.034
Phenol	51 (22–77)	42 (23–134)	45 (14–83)	0.44	0.51
2, 5-Dimethylfuran	19 (0–40)	13 (0–28)	0 (0–14)	0.023	<0.001
Cyclohexanone	150 (75–282)	111 (77–211)	38 (21–64)	0.022	<0.001
Hexanal	26 (10–51)	14 (0–35)	0 (0–26)	0.012	0.018
Propyl acetate	56 (32–93)	44 (16–96)	62 (34–114)	0.50	0.21
Styrene	1775 (624–5791)	1512 (503–6369)	756 (155–4502)	0.46	0.20
Benzaldehyde	1990 (888–8328)	1980 (839–7658)	1100 (345–5768)	0.41	0.28
Heptanal	39 (18–65)	15 (0–35)	23 (0–46)	<0.001	0.001
4-hydroxyhexanal	145 (89–290)	84 (42–152)	116 (63–250)	0.001	0.42
Acetophenone	148 (84–242)	146 (68–252)	86 (49–208)	0.48	0.044
Propylcyclohexane	254 (123–484)	196 (115–340)	67 (49–129)	0.13	<0.001
Octanal	58 (19–108)	30 (11–65)	27 (12–52)	0.006	0.001
Benzothiazole	0 (0–22)	0 (0–21)	10 (0–20)	0.23	0.88
Nonanal	31 (15–78)	37 (17–64)	0 (0–15)	0.27	<0.001
Decanal	57 (27–121)	68 (25–114)	10 (0–27)	0.40	<0.001
2, 2-Dimethyldecane	158 (88–357)	183 (117–319)	47 (32–76)	0.96	<0.001

Table 2: Changes of volatile organic compounds in exhaled breath after surgical resection of lung cancers.
Data are the median (interquartile range). P values are from Wilcoxon matched-pairs signed-rank sum tests. Molecular formula, molecular weight, and m/z values are shown in Table S1. HPPI-TOFMS: high-pressure photon ionization time-of-flight mass spectrometry; PO: postoperative; VOC: volatile organic compound.

peak intensity of the 16 mentioned VOCs between the two groups are shown in Figure 4. All these VOCs showed significantly elevated peak intensity in patients with lung cancer compared to healthy individuals. The volcano plot (Figure 5A) shows the fold changes and differences of 16 VOCs between lung cancer and healthy individuals. The correlation analyses demonstrated the differentiated enrichment of VOCs in patients with lung cancer and healthy individuals (Figure 5B and Figure 5C, respectively), suggesting different VOCs correlation modules in two groups. After adjusting for confounding factors, including age, sex, smoking history, alcohol history, and comorbidities, lung cancer status (vs. healthy individuals) was still an independent factor of elevated VOC peak intensity (\geq upper tertiles) (Table 4).

The individual diagnostic performance of 16 VOCs for lung cancer is shown in Figure 5D and summarized in Table S2. Isoprene and hexanal showed the highest AUCs of 0.859 and 0.843, respectively. Using logistic

regression, the diagnostic model including 16 VOCs achieved a diagnostic AUC of 0.952, sensitivity of 89.2%, specificity of 89.1%, and accuracy of 89.1% (Figure 5E). The Hosmer–Lemeshow test indicated good fitting result ($P = 0.64$). Then, we assessed the diagnostic performance of the combination of the top eight VOCs with AUCs over 0.750, namely isoprene, hexanal, pentanal, propylcyclohexane, nonanal, 2,2-dimethyldecane, heptanal, and decanal. As shown in Figure 5F, the diagnostic model including these eight VOCs achieved a diagnostic AUC of 0.931, sensitivity of 86.0%, specificity of 87.2%, and accuracy of 86.9%, with good fitting result (Hosmer–Lemeshow test, $P = 0.45$). When these VOCs were used as classified parameters, the diagnostic performance was still satisfying (Figure 5E,F).

Breath biomarkers and pathological parameters

As shown in Tables 1 and 3, the discovery study mainly consisted of lung adenocarcinoma (95.2%), and the

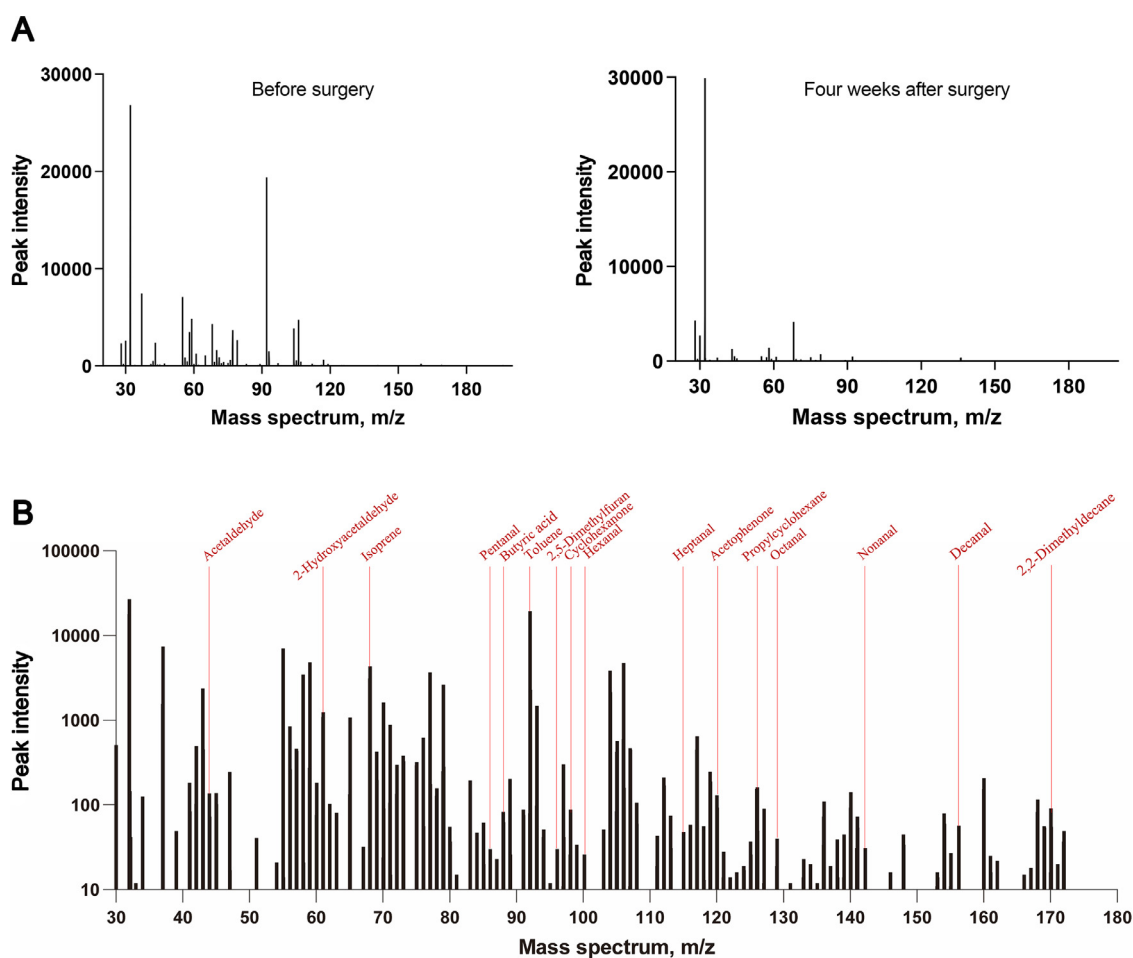


Figure 2. Examples of mass spectrums.

A: The mass spectrums of a patient before surgery (left) and four weeks after surgery (right). B: Identification of 16 VOCs in mass spectrum before surgery. Patient characteristics: female, 52 years, stage IA3.

validation study consisted of 125 adenocarcinoma, 14 squamous cell carcinoma, 9 small cell lung cancer, and 8 other types of lung cancer. In addition, early-stage lung cancer (IA) was predominant among the participants in both the discovery and validation studies. We then investigated the association between breath VOCs and cancer types and cancer stages. As shown in Figure S1, no significant difference in 16 VOC peak intensity was observed between lung adenocarcinoma and other cancer types. Regarding comparisons of VOCs among different cancer stages (Figure S2), patients of IA1 had lower peak intensity in most VOCs, but the differences were not significant.

Discussion

This study identified breath biomarkers of lung cancer through perioperative breathomics testing using HPPI-TOFMS. A total of 16 VOCs that showed dynamic changes perioperatively were identified and

demonstrated to be capable of distinguishing patients with lung cancer from healthy individuals.

Despite decades of research associated with breath biomarkers, no breath VOC tests are clinically available for lung cancer screening. Previous studies on VOCs for lung cancer diagnosis are heterogeneous in many aspects.^{8–11} Most studies reported an accuracy of over 80% for breath VOCs in diagnosing lung cancer, while several studies reported poor performance of only 60–70%. The studies by Chen et al.²⁴ and Long et al.²⁵ demonstrated good performance of breath biomarkers in discriminating cancer subtypes and cancer stages; however, these findings are not commonly confirmed in other studies. Notably, there has been little consensus on breath VOC groups to diagnose lung cancer among published studies.^{8,9} Although the public mass spectrometry libraries of NIST 05 and NIST 05S are commonly referred to, lung cancer breath VOCs showed high heterogeneity (Table S1). Hydrocarbons, alcohols, and aldehydes are most commonly detected but still

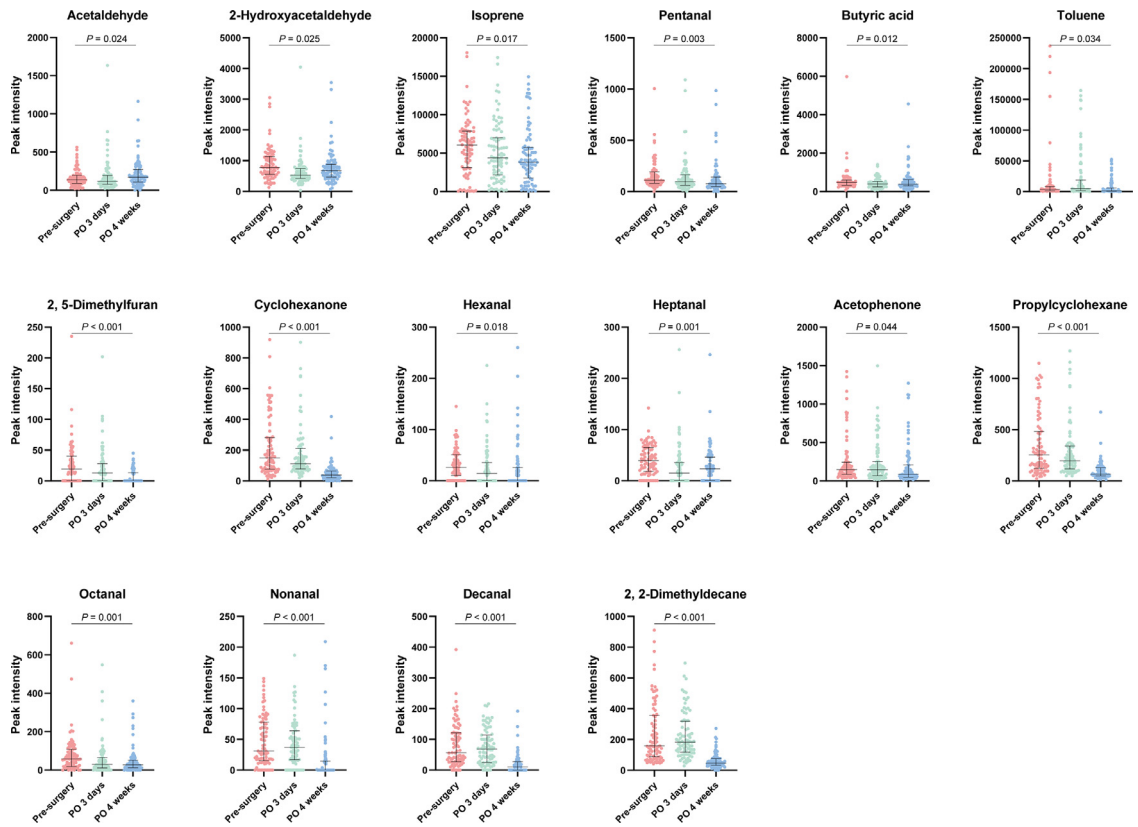


Figure 3. Perioperative dynamic changes of 16 volatile organic compounds in exhaled breath from patients with lung cancer. *P* values representing the difference in peak intensities before surgery and 4 weeks after surgery are from Wilcoxon matched-pairs signed-rank sum tests. PO: postoperative.

Characteristics	Lung cancer (N = 157)	Healthy control (N = 368)	<i>P</i> values
Demographic data			
Age, years	57.0 ± 10.9	44.5 ± 10.6	<0.001
Gender (female)	85 (54.1)	115 (31.3)	<0.001
Smoking history	34 (21.7)	119 (32.3)	0.014
Current smoking	22 (14.0)	101 (27.4)	0.001
Alcohol history	15 (9.6)	61 (16.6)	0.036
Current alcohol drinking	10 (6.4)	60 (16.3)	0.002
Comorbidities			
Diabetes	8 (5.1)	11 (3.0)	0.24
Cardiovascular disease	39 (24.8)	28 (7.6)	<0.001
Respiratory disease	0	5 (1.4)	0.33
Pathological data			
Cancer type	125/14/7/11		
AC/SCC/SCLC/others	(79.6/8.9/4.5/7.0)		
Pathological stage	43/40/18/7/11/15/9		
IA1/IA2/IA3/IB/II/III/IV	(30.1/28.0/12.6/4.9/7.7/10.5/6.3)		

Table 3: Patient characteristics.

Data are the mean ± standard deviation or number (%).

P values are from ANOVA or Pearson's chi-square test.

AC: adenocarcinoma; N: number; SCC: squamous cell carcinoma; SCLC: small cell lung cancer.

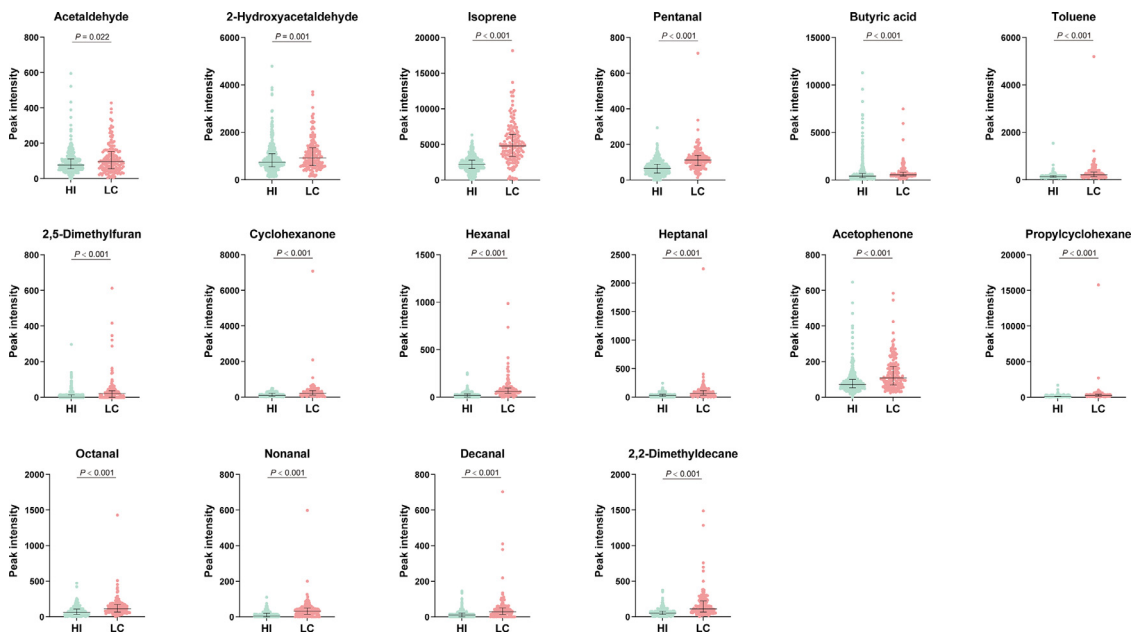


Figure 4. Comparisons of spectrum peak intensity of volatile organic compounds in patients with lung cancer and healthy individuals.

P values are from Mann–Whitney U tests. HI: healthy individual; LC: lung cancer.

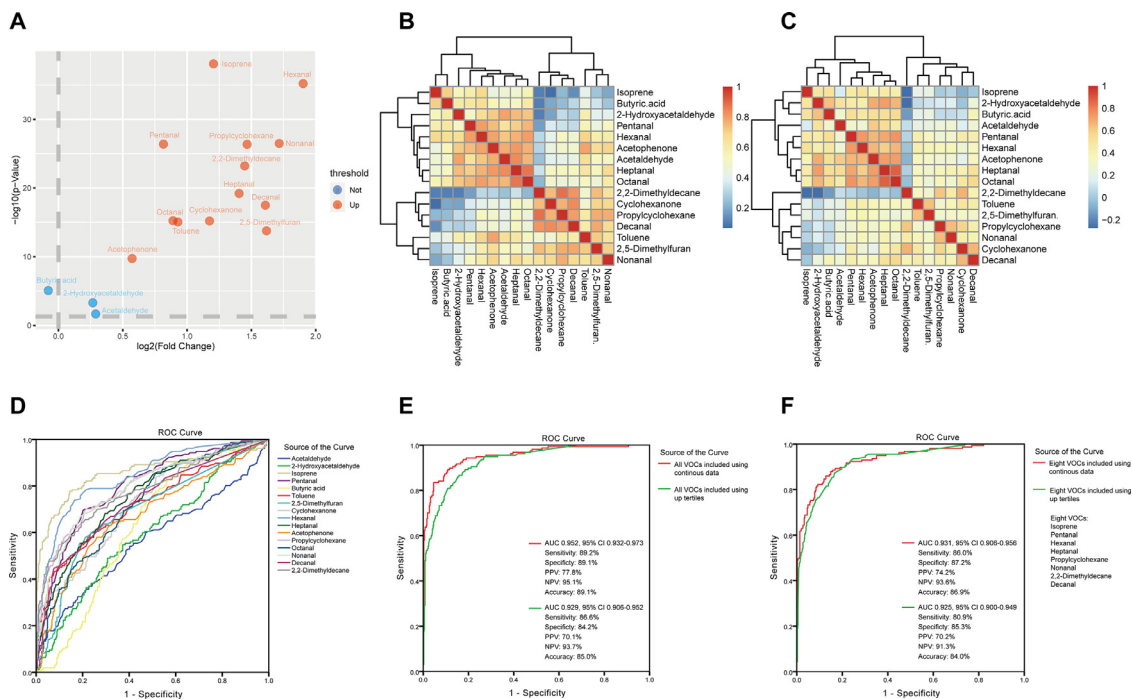


Figure 5. Panels of investigations of breath volatile organic compounds (VOCs) and lung cancers.

A: The volcano plot showing the fold changes and difference in breath VOC peak intensity between patients with lung cancer and healthy individuals. B,C: Correlation analysis of sixteen VOCs in healthy individuals and patients with lung cancer. D: The performance of sixteen VOCs in diagnosing patients with lung cancer from healthy individuals. E: The performance of the combined sixteen VOCs in diagnosing lung cancer. F: The performance of the combined top eight VOCs in diagnosing lung cancer. AUC: area under the curve; CI: confidence interval; NPV: negative predictive value; PPV: positive predictive value; ROC: receiver operating characteristic curve.

Elevated VOCs [†]	Lung cancers vs. Healthy individuals	
	Ad-OR (95% CI)	P values
Acetaldehyde	1.79 (1.20–2.66)	0.002
2-Hydroxyacetaldehyde	1.99 (1.35–2.92)	<0.001
Isoprene	20.5 (12.6–33.3)	<0.001
Pentanal	8.55 (5.60–13.0)	<0.001
Butyric acid	2.06 (1.38–3.07)	<0.001
Toluene	8.24 (5.12–13.3)	<0.001
2,5-Dimethylfuran	8.15 (5.07–13.1)	<0.001
Cyclohexanone	2.86 (1.94–4.22)	<0.001
Hexanal	20.8 (12.4–35.1)	<0.001
Heptanal	5.07 (3.32–7.73)	<0.001
Acetophenone	5.05 (3.24–7.86)	<0.001
Propylcyclohexane	7.31 (4.77–11.2)	<0.001
Octanal	3.64 (2.40–5.51)	<0.001
Nonanal	8.63 (5.64–13.2)	<0.001
Decanal	4.45 (2.98–6.63)	<0.001
2,2-Dimethyldecane	7.17 (4.74–10.8)	<0.001

Table 4: Multivariable analyses investigating the association between lung cancer and elevated peak intensities of volatile organic compounds (VOCs).

Adjusted confounding factors: age, gender, smoking history, alcohol history, diabetes, and cardiovascular diseases.

Multivariate logistic regression models with the backward conditional methods were used.

[†] Elevated VOC was defined as peak intensity over the up tertiles. Ad-OR: adjusted odds ratio; CI: confidence interval.

have differences in specific compounds. The inconsistencies in sample collection, patient condition, test environment, and data analysis could be main obstacles to the generalization of breathomics in clinical practices. The HPPI-TOFMS is promisingly helpful in simplifying breathomics testing procedures and increasing the robustness of breathomics results.^{16–20} Our pilot studies have confirmed the application of HPPI-TOFMS in exhaled breath testing for lung cancer and esophageal cancer detection based on spectrum characteristics.^{21,22} This study further confirmed the property of HPPI-TOFMS in cancer screening through identifying specific VOCs.

Different from previous studies, this study tried to identify lung cancer breath VOCs through dynamic surveillance of perioperative breathomics. Measures had been taken to minimize the interference of food, medicines, comorbidities, Tedlar bags, and environment on breathomics testing. We compared the breathomics of patients with lung cancer before surgery with that at four weeks after surgery. The detected breath VOCs were further investigated using a large breathomics database consisting of 368 healthy individuals and 157 lung cancers. This study is not a simple validation of previously reported lung cancer breath VOCs but a project to solve the large heterogeneity among published studies by establishing groups of 16 VOCs and eight

VOCs for lung cancer diagnosis. Elevated signals of these VOCs were observed in patients with lung cancer and were demonstrated to be independently associated with lung cancer status. The large sample and limited variables included in the logistic regression-based diagnostic models ensure a low risk of overfitting.^{26,27} Among these VOCs, isoprene was abundant, while hexanal, pentanal, nonanal, and heptanal were micro-scale (Figure 3). All these VOCs showed good performance in diagnosing patients with lung cancer from healthy individuals, which was consistent with previous reports.^{8–11} However, the spectral intensity of these VOCs showed no significant difference among different cancer types or stages. Previous studies have confirmed the capability of these breath VOCs in diagnosing esophageal cancer, gastric cancer, colorectal cancer, and breast cancer.^{10,28–31} Regardless of the distance of the organ where produced, VOCs can be transported by the blood to the lungs and exhaled during breathing. It seems that these reported VOCs may be cancer-specific but not lung cancer-specific. In other words, breathomics analysis of exhaled breath has potential for broad spectrum screening of cancers.

In contrast to the prosperity of breath biomarker discovery in cancer patients, mechanistic investigations on the metabolism and production of VOC biomarkers are dismal. The lipid peroxidation of polyunsaturated fatty acids by reactive oxygen species at the cell membrane is regarded as the main source of alkanes and aldehydes in breath VOCs.^{7,32,33} The peroxidation of polyunsaturated fatty acids could be a cascade amplification program because of the consistent production of reactive oxygen species. Oxidative stress has been considered a potential causative agent in tumorigenesis with enhanced oxidative activity in tumor tissues.³³ Lipid metabolism disorder is widely observed in cancer cells, in which the fatty acid chain tends to be longer and more saturated to resist apoptosis.³⁴ The increased production of isoprene from acetyl-coenzyme A is also reported to be associated with cholesterol biosynthesis.³⁵ Collectively, the aberrant lipidomics profiles in cancer cells could account for the altered breath VOCs. Regarding the specific mediators, cytochrome P450 oxidoreductases (PORs) have the capability to decarboxylate and to desaturate hydroxylated fatty acids.³⁶ It has also been shown to regulate peroxidation of polyunsaturated fatty acids across a wide range of lineages and cell states.³² CYP450 isozymes and fatty acid types may influence the type of VOCs in exhaled breath. Furthermore, reduced ALDH3A2 expression, which is associated with TP53 depletion, has been demonstrated to cause reduced detoxification and endogenous accumulation of aldehyde.³⁷ However, changes in lipidomics profiles and regulator enzymes could occur in multiple cancers. This supports the mentioned speculations that most VOCs could correspond to broad spectrum cancers.

Limitations of this study should be considered when interpreting these findings. The 16 VOCs were selected based on HPPI-TOFMS, and it is highly possible that the diagnostic performance cannot be reproduced with other breath detecting devices. Since, the application of HPPI-TOFMS has not been widely generalized in breathomics testing, and no mass library has been established based on HPPI-TOFMS. This study identified VOCs based on m/z and ionization models; however, the optimal identification of metabolites warrants complex analytical chemistry. The peak intensity was only a substitute indicator for concentration; our future studies will try to achieve accurate concentration of these VOCs. Thus, these findings warrant validation in the application of other mass spectrometry techniques, such as GC-MS. The included lung cancers in the discovery study were mainly adenocarcinoma at an early stage because of the generalization of LDCT, the high proportion of never-smokers, and the good socioeconomic conditions of patients.^{38,39} The difference in breath VOCs among different cancer types and cancer stages demand further investigation with a large sample. More scientific studies are warranted to qualify the utilization of breathomics testing for cancer screening.

Perioperative dynamic breathomics testing was demonstrated to be effective in identifying lung cancer breath biomarkers. Using HPPI-TOFMS, we identified 16 lung cancer-related breath VOCs that were accurate in discriminating patients with lung cancer from healthy individuals. These breath VOCs were mainly isoprene and aldehydes, showing potential for broad-spectrum screening of cancers. This study helped to solve the heterogeneity among published studies through establishing groups of 16 VOCs and eight VOCs for lung cancer diagnosis. More scientific studies are warranted to investigate the mechanisms underlying the production and metabolism of these VOCs.

Funding

National Natural Science Foundation of China (82173386) and Peking University People's Hospital Scientific Research Development Funds (RDH2021-07).

Data sharing statement

All data are shown in the manuscript and supplemental materials.

Contributors

PW, SZ, SW, and MQ contributed to study design, data analysis, and data interpretation. QH, SM, TM and ZL contributed to study design and data collection. MH and QL contributed to data analysis and data interpretation. All authors contributed to manuscript writing and approved the manuscript. MQ, SW, and SZ verified the underlying data. All authors had full access to all the

data in the study and accept responsibility to submit for publication.

Declaration of interests

We declare no competing interests.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.eclinm.2022.101384](https://doi.org/10.1016/j.eclinm.2022.101384).

References

- 1 Global Burden of Disease 2019 Cancer Collaboration. Cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life years for 29 cancer groups from 2010 to 2019: a systematic analysis for the global burden of disease study 2019. *JAMA Oncol*. 2021. <https://doi.org/10.1001/jamaoncol.2021.6987>.
- 2 Dickson JL, Horst C, Nair A, Tisi S, Predecki R, Janes SM. Hesitancy around low-dose CT screening for lung cancer. *Ann Oncol*. 2021. <https://doi.org/10.1016/j.annonc.2021.09.008>. S0923-7534(21)04487-2.
- 3 de Koning HJ, van der Aalst CM, de Jong PA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. *N Engl J Med*. 2020;382(6):503-513. <https://doi.org/10.1056/NEJ-Moa1911793>.
- 4 The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5). <https://doi.org/10.1056/NEJ-Moa1102873>.
- 5 Lin Y, Fu M, Ding R, et al. Patient adherence to lung CT screening reporting & data system-recommended screening intervals in the United States: a systematic review and meta-analysis. *J Thorac Oncol*. 2021. <https://doi.org/10.1016/j.jtho.2021.09.013>.
- 6 Phillips M, Gleeson K, Hughes JMB, et al. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. *Lancet*. 1999;353(9168):1930-1933. [https://doi.org/10.1016/S0140-6736\(98\)07552-7](https://doi.org/10.1016/S0140-6736(98)07552-7).
- 7 Horvath I, Lazar Z, Gyulai N, Kollai M, Losonczy G. Exhaled biomarkers in lung cancer. *Eur Respir J*. 2009;34(1):261-275. <https://doi.org/10.1183/09031936.00142508>.
- 8 Ratiu IA, Ligor T, Bocos-Bintintan V, Mayhew CA, Buszewski B. Volatile organic compounds in exhaled breath as fingerprints of lung cancer, asthma and COPD. *J Clin Med*. 2020;10(1). <https://doi.org/10.3390/jcm10010032>.
- 9 Campanella A, De Summa S, Tommasi S. Exhaled breath condensate biomarkers for lung cancer. *J Breath Res*. 2019;13(4):044002. <https://doi.org/10.1088/1752-7163/ab2f9f>.
- 10 Hanna GB, Boshier PR, Markar SR, Romano A. Accuracy and methodologic challenges of volatile organic compound-based exhaled breath tests for cancer diagnosis: a systematic review and meta-analysis. *JAMA Oncol*. 2019;5(1):e182815. <https://doi.org/10.1001/jamaoncol.2018.2815>.
- 11 Marzorati D, Mainardi L, Sedda G, Gasparri R, Spaggiari L, Cerveri P. A review of exhaled breath: a key role in lung cancer diagnosis. *J Breath Res*. 2019;13(3):034001. <https://doi.org/10.1088/1752-7163/ab0684>.
- 12 Das MK, Bishwal SC, Das A, et al. Investigation of gender-specific exhaled breath volatome in humans by GCxGC-TOF-MS. *Anal Chem*. 2014;86(2):1229-1237. <https://doi.org/10.1021/ac403541a>.
- 13 Gaugg MT, Gomez DG, Barrios-Collado C, et al. Expanding metabolite coverage of real-time breath analysis by coupling a universal secondary electrospray ionization source and high resolution mass spectrometry—a pilot study on tobacco smokers. *J Breath Res*. 2016;10(1):016010. <https://doi.org/10.1088/1752-7155/10/1/016010>.
- 14 Markar SR, Chin ST, Romano A, et al. Breath volatile organic compound profiling of colorectal cancer using selected ion flow-tube mass spectrometry. *Ann Surg*. 2019;269(5):903-910. <https://doi.org/10.1097/SLA.0000000000002539>.

- 15 Trefz P, Schmidt M, Oertel P, et al. Continuous real time breath gas monitoring in the clinical environment by proton-transfer-reaction-time-of-flight-mass spectrometry. *Anal Chem.* 2013;85(21):10321–10329. <https://doi.org/10.1021/ac402298v>.
- 16 Jiang D, Li E, Zhou Q, et al. Online monitoring of intraoperative exhaled propofol by acetone-assisted negative photoionization ion mobility spectrometry coupled with time-resolved purge introduction. *Anal Chem.* 2018;90(8):5280–5289. <https://doi.org/10.1021/acs.analchem.8b00171>.
- 17 Wang Y, Jiang J, Hua L, et al. High-pressure photon ionization source for TOFMS and its application for online breath analysis. *Anal Chem.* 2016;88(18):9047–9055. <https://doi.org/10.1021/acs.analchem.6b01707>.
- 18 Jiang D, Wang X, Chen C, et al. Dopant-assisted photoionization positive ion mobility spectrometry coupled with time-resolved purge introduction for online quantitative monitoring of intraoperative end-tidal propofol. *Anal Chim Acta.* 2018;1032:83–90. <https://doi.org/10.1016/j.aca.2018.06.047>.
- 19 Wang Y, Hua L, Jiang J, et al. High-pressure photon ionization time-of-flight mass spectrometry combined with dynamic purge-injection for rapid analysis of volatile metabolites in urine. *Anal Chim Acta.* 2018;1008:74–81. <https://doi.org/10.1016/j.aca.2018.01.006>.
- 20 Chen X, Hua L, Jiang J. Multi-capillary column high-pressure photoionization time-of-flight mass spectrometry and its application for online rapid analysis of flavor compounds. *Talanta.* 2019;201:33–39. <https://doi.org/10.1016/j.talanta.2019.03.103>.
- 21 Meng S, Li Q, Zhou Z, et al. Assessment of an exhaled breath test using high-pressure photon ionization time-of-flight mass spectrometry to detect lung cancer. *JAMA Netw Open.* 2021;4(3):e213486. <https://doi.org/10.1001/jamanetworkopen.2021.3486>.
- 22 Huang Q, Wang S, Li Q, et al. Assessment of breathomics testing using high-pressure photon ionization time-of-flight mass spectrometry to detect esophageal cancer. *JAMA Netw Open.* 2021;4(10):e2127042. <https://doi.org/10.1001/jamanetworkopen.2021.27042>.
- 23 Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527. <https://doi.org/10.1136/bmj.h5527>.
- 24 Chen X, Muhammad KG, Madeeha C, et al. Calculated indices of volatile organic compounds (VOCs) in exhalation for lung cancer screening and early detection. *Lung Cancer.* 2021;154:197–205. <https://doi.org/10.1016/j.lungcan.2021.02.006>.
- 25 Long Y, Wang C, Wang T, et al. High performance exhaled breath biomarkers for diagnosis of lung cancer and potential biomarkers for classification of lung cancer. *J Breath Res.* 2020;15(1):016017. <https://doi.org/10.1088/1752-7163/abaecb>.
- 26 Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform.* 2002;35(5–6):352–359. [https://doi.org/10.1016/S1532-0464\(03\)00034-0](https://doi.org/10.1016/S1532-0464(03)00034-0).
- 27 Mikolajczyk RT, DiSilvestro A, Zhang J. Evaluation of logistic regression reporting in current obstetrics and gynecology literature. *Obstet Gynecol.* 2008;111(2 Pt 1):413–419. <https://doi.org/10.1097/AOG.0b013e318160f38e>.
- 28 Markar SR, Wiggins T, Antonowicz S, et al. Assessment of a noninvasive exhaled breath test for the diagnosis of oesophagogastric cancer. *JAMA Oncol.* 2018;4(7):970–976. <https://doi.org/10.1001/jamaoncol.2018.0991>.
- 29 Zhang J, Tian Y, Luo Z, Qian C, Li W, Duan Y. Breath volatile organic compound analysis: an emerging method for gastric cancer detection. *J Breath Res.* 2021;15(4). <https://doi.org/10.1088/1752-7163/ac2cde>.
- 30 Chandrapalan S, Bosch S, Cubiella J, et al. Systematic review with meta-analysis: volatile organic compound analysis to improve faecal immunochemical testing in the detection of colorectal cancer. *Aliment Pharmacol Ther.* 2021;54(1):14–23. <https://doi.org/10.1111/apt.16405>.
- 31 Phillips M, Cataneo RN, Cruz-Ramos JA, et al. Prediction of breast cancer risk with volatile biomarkers in breath. *Breast Cancer Res Treat.* 2018;170(2):343–350. <https://doi.org/10.1007/s10549-018-4764-4>.
- 32 Zou Y, Li H, Graham ET, et al. Cytochrome P450 oxidoreductase contributes to phospholipid peroxidation in ferroptosis. *Nat Chem Biol.* 2020;16(3):302–309. <https://doi.org/10.1038/s41589-020-0472-6>.
- 33 Ratcliffe N, Wieczorek T, Drabinska N. A mechanistic study and review of volatile products from peroxidation of unsaturated fatty acids: an aid to understanding the origins of volatile organic compounds from the human body. *J Breath Res.* 2020;14(3):034001. <https://doi.org/10.1088/1752-7163/ab7f9d>.
- 34 Hoy AJ, Nagarajan SR, Butler LM. Tumour fatty acid metabolism in the context of therapy resistance and obesity. *Nat Rev Cancer.* 2021. <https://doi.org/10.1038/s41568-021-00388-4>.
- 35 Stone BG, Besse TJ, Duane WC, Evans CD, DeMaster EG. Effect of regulating cholesterol biosynthesis on breath isoprene excretion in men. *Lipids.* 1993;28(8):705–708. <https://doi.org/10.1007/BF02535990>.
- 36 Munro AW, McLean KJ, Grant JL, Makris TM. Structure and function of the cytochrome P450 peroxxygenase enzymes. *Biochem Soc Trans.* 2018;46(1):183–196. <https://doi.org/10.1042/BST20170218>.
- 37 Antonowicz S, Bodai Z, Wiggins T, et al. Endogenous aldehyde accumulation generates genotoxicity and exhaled biomarkers in esophageal adenocarcinoma. *Nat Commun.* 2021;12(1):1454. <https://doi.org/10.1038/s41467-021-21800-5>.
- 38 Tseng CH, Tsuang BJ, Chiang CJ, et al. The relationship between air pollution and lung cancer in nonsmokers in Taiwan. *J Thorac Oncol.* 2019;14(5):784–792. <https://doi.org/10.1016/j.jtho.2018.12.033>.
- 39 Udelsman BV, Madariaga ML, Chang DC, Kozower BD, Gaissert HA. Concordance of clinical and pathologic nodal staging in resectable lung cancer. *Ann Thorac Surg.* 2021;111(4):1125–1132. <https://doi.org/10.1016/j.athoracsur.2020.06.060>.