Research protocol for an exploratory study: VOC biomarkers identification and diagnostic model construction for lung cancer based on exhaled breath analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-028448
Article Type:	Protocol
Date Submitted by the Author:	08-Dec-2018
Complete List of Authors:	Li, Wenwen; Sichuan University, West China School of Public Health Dai, Wei; Sichuan Cancer Hospital & Institute, Department of Thoracic Surgery Liu, Mingxin; Sichuan Cancer Hospital & Institute, Department of Thoracic Surgery Long, Yijing; Sichuan University, The College of Life Sciences Wang, Chunyan; Sichuan University, The College of Life Sciences Xie, Shaohua; Chengdu Medical College, Graduate School Liu, Yuanling; Sichuan University, The College of Life Sciences Zhang, Yinchenxi; Sichuan University, The College of Life Sciences Shi, Qiuling; The University of Texas MD Anderson Cancer Center, Department of Symptom Research Peng, Xiaoqin; Chengdu Medical College, Graduate School Liu, Yifeng; Chengdu Medical College, Graduate School Li, Qiang; Sichuan Cancer Hospital & Institute, Department of Thoracic Surgery Duan, Yixiang; Sichuan University, The College of Life Sciences
Keywords:	lung cancer, early diagnosis, exhaled breath, volatile organic compounds, biomarker

Research protocol for an exploratory study: VOC biomarkers identification and diagnostic model construction for lung cancer based on exhaled breath analysis

Wenwen Li[1, 2#], Wei Dai[3#], Mingxin Liu[3#], Yijing Long[2, 4], Chunyan Wang[2, 4], Shaohua Xie[5], Yuanling Liu[2, 4], Yinchenxi Zhang[2, 4], Qiuling Shi[6], Xiaoqin Peng[5], Yifeng Liu[5], Qiang Li[3*], Yixiang Duan[2, 4*]

[1]West China School of Public Health, Sichuan University, Chengdu, Sichuan, China.

[2]Research Center of Analytical Instrumentation, Sichuan University, Chengdu, Sichuan, China.

[3]Department of Thoracic Surgery, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, Sichuan, China.

[4]Key Laboratory of Bio-resource and Eco-environment, Ministry of Education, The College of Life Sciences, Sichuan University, Chengdu, Sichuan, China.

[5]Graduate School, Chengdu Medical College, Chengdu, Sichuan, China.

[6]Department of Symptom Research, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

***Corresponding author:**

Yixiang Duan, Research Center of Analytical Instrumentation, College of Life Sciences, Sichuan University, 29 Wangjiang Road, Chengdu 610065, Sichuan, China. Tel: +86-028-85418180, Fax: +86-028-85418180, Email: yduan@scu.edu.cn.

Qiang Li, Department of Thoracic Surgery, Sichuan Cancer Hospital & Institute, No. 55, Section 4, South Renmin Road, Chengdu 610041, Sichuan, China, Tel: +86-028-85420366, Fax: +86-028-85420116, Email: liqiang@sichuancancer.org.

#These authors contributed equally to this work.

ABSTRACT

Introduction: Lung cancer is the most common cancer and the leading cause of cancer death in China, as well as in the world. Late diagnosis is the main obstacle to improving survival. Currently, early detection methods for lung cancer have many limitations, e.g. low specificity, risk of radiation exposure and overdiagnosis. Exhaled breath analysis is one of the most promising non-invasive techniques for early detection of lung cancer. The aim of this study is to identify volatile organic compound (VOC) biomarkers in lung cancer and to construct a diagnostic model for lung cancer prediction based on exhaled breath analysis.

Methods and analysis: The study will recruit 389 lung cancer patients in one cancer center and 389 healthy subjects in two lung cancer screening centers. Bio-VOC breath sampler and Tedlar bag will be used to collect breath samples. Gas chromatography-mass spectrometry (GC-MS) coupled with solid phase microextraction (SPME) technique will be used to analyze VOCs in exhaled breath. VOC biomarkers with statistical significance and showing abilities to discriminate lung cancer patients from healthy subjects will be selected for construction of diagnostic model to predict lung cancer.

Ethics and dissemination: The study was approved by the Ethics Committee of Sichuan Cancer Hospital on April 6, 2017 (No.SCCHEC-02-2017-011). The results of this study will be disseminated in presentations at academic conferences, publications in peer-reviewed journals and the news media.

Trials registration number: ChiCTR-DOD-17011134.

KEYWORDS: lung cancer, early diagnosis, exhaled breath, volatile organic compounds, biomarker.

Word count: 2942 words.

ARTICLE SUMMARY

Article focus

To construct a diagnostic model for early prediction of lung cancer based on exhaled biomarkers.

Strengths and limitations of this study

1. This study will select exhaled biomarkers of lung cancer based on the largest sample size from multiple centers ever analyzed by gas chromatography-mass spectrometry.

2. Well-selected healthy controls with high risk but negative of lung cancer on chest CT will be recruited in this study.

3. Gas chromatography-mass spectrometry can only obtain a limited amount of volatile organic compounds (VOCs) in exhaled breath and some VOCs cannot be detected.

INTRODUCTION

In China, lung cancer incidence and related mortality have been increasing annually for the last 30 years[1 2]. Lung cancer is now the most common cancer and the leading cause of cancer death, accounting for 24.9% of all cancer deaths in China in 2010[1 2]. Lung cancer is also the most common cancer and the leading cause of cancer death worldwide[3]. Only 16.8% of all patients with lung cancer survive for 5 years or more after diagnosis[4], mainly because lung cancer is often staged as locally advanced or metastatic disease at the initial diagnosis[4 5]. However, the 5-year survival rate of patients with stage I lung cancer is greater than 60% [5 6], and the estimated 10-year survival rate of patients with stage I lung cancer detected on computed tomography (CT) screening can reach 88%[7]. Therefore, early detection is the key to improving survival rates of patients with lung cancer.

Currently, early detection techniques for lung cancer have many limitations, and a simple, reliable and non-invasive early lung cancer screening technique is urgently needed. Sputum cytology has a very low sensitivity[8]. Chest radiography is radioactive and has a very high rate of false-negative results, especially in the detection of early-stage lung cancers[9]. Screening with low-dose CT can reduce mortality from lung cancer by 20%, and lung cancer screening using low-dose CT for high-risk individuals is now recommended[10 11]. However, there are also many disadvantages of low-dose CT, such as a high false-positive rate, overdiagnosis, the risk of radiation exposure and high cost, which limit its application in population-based screening[12 13].

Exhaled breath analysis is completely non-invasive and has great potential to become a screening and diagnostic method for early detection of cancer[14-17]. The majority of previous studies focusing on lung cancer were conducted on small samples[18 19]. Potential volatile organic compound (VOC) biomarkers for lung cancer have been discussed and summarized[15 19]. However, to date, there are no unified VOC biomarkers for lung cancer, and the sets of VOCs employed vary between studies. Therefore, breath analysis is still in an early stage of clinical application. Several factors account for this situation, such as large variation in sample size, diverse sample collection approaches, different analytical techniques, and different data processing methods[19-21]. Our group has investigated exhaled VOC biomarkers in diabetes and breast cancer[22-25]. We have rich experience in breath sample collection, exhaled VOCs analysis and data processing. In addition, we have a stable and reliable source of lung cancer patients and healthy subjects. We aim to identify exhaled VOC biomarkers of lung cancer and establish a diagnostic model for lung cancer prediction. In subsequent studies, this model will be validated in clinical setting and population-based screening.

METHOD AND ANALYSIS

Main Centers: Sichuan Cancer Hospital, Sichuan University, Nanchong Central Hospital and Chengdu Longquanyi District Center for Disease Control and Prevention.

Dates of the study: From March 1, 2017 to February 28, 2020.

Design

Inclusion criteria: Lung cancer patients and healthy subjects aged 50 to 74 will be included. Lung cancer patients should have a pathological diagnosis of primary lung cancer and should not receive any treatment. Healthy subjects should be negative of lung cancer on chest CT.

Exclusion criteria: Patients and controls with diabetes, other malignancies, active asthma, severe liver dysfunction, end-stage renal disease and acute inflammation will be excluded.

Breath sampling procedures

Each participant recruited into this study will be given an information leaflet explaining the research, and will sign an informed consent form. Participants are required to fast for at least 8 h and rest for at least 10 min in a separate room with good ventilation before breath sampling in the morning. Subjects will be asked to take a normal inhalation followed by a normal exhalation via their mouth. Deep inhalation before sampling and nasal ventilation during sampling will not be allowed. Each subject will exhale three times to complete a breath sample collection. Exhaled breath gas will be collected with Bio-VOC breath sampler (Markes Int. U.K). Subjects will exhale gently into the Bio-VOC syringe until they feel mild resistance. The Bio-VOC sampler collects end-tidal breath gas by expelling the dead space air from the nasopharynx and upper airway using a three-way valve. The end-tidal breath gas will then be transferred to Tedlar bag (500 mL) for storage and transportation. Breath samples will be kept at -40 ℃ until analysis. Samples will be collected from lung cancer patients prior to any clinical treatments.

VOCs analysis

Solid phase microextraction (SPME) technique will be utilized in this study to pre-concentrate VOCs in breath samples prior to gas chromatography-mass spectrometry (GC-MS) analysis. Divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) fiber will be used to extract exhaled VOCs. Samples will be analysed using a Thermo Scientific TRACE 1300 gas chromatograph coupled to a TSQ8000 triple quadrupole mass spectrometer. VF-624ms capillary column (60 m × 0.25 mm × 0.25 µm, Agilent Technology) will be used to separate VOCs. The analytical process will be as follows: first, the breath sample kept at -40℃ will be incubated at 37 ℃ for 5 min; then a DVB/CAR/PDMS fiber will be used to pre-concentrate VOCs for 30 min at 37 ℃, and finally the fiber will be desorbed thermally at the front inlet of the gas chromatograph. Split mode and a specific liner for SPME fiber will be used for gas chromatograph. Electron ionization source (70 eV) will be used for mass spectrometer.

Endpoint: Exhaled VOC biomarkers of lung cancer and the accuracy of the diagnostic model for lung cancer prediction.

Quality assurance

Standardization of breath sampling: In order to minimize inter- and intra-observer error caused by researchers during breath sampling, a fixed number of well-trained staff will be appointed to collect breath samples from patients and controls. Operation requirements that may bring about errors in breath sampling and lead to biased results are listed in supplementary. All staff participating in breath sampling will be trained and tested for the process and requirements of sampling. In addition, all researchers involved in this study will go through the certification process, including informed consent signing, breath sampling, samples storage and transportation.

Ambient room air: Breath samples will be collected from patients and controls in a separate room with good ventilation. Ambient room air will be collected simultaneously with each batch of breath samples. The ambient air is used to monitor possible VOC contamination, because it may be inhaled by subjects and become a prominent component of the exhaled breath, leading to anomalous results. Samples with VOCs that appear only in one batch of samples and simultaneously in the corresponding ambient room air at significant concentrations will be regarded as contamination and will be excluded.

Data monitoring: All data collected in this study, including demographic information, clinicopathologic information and raw GC-MS data, will be uploaded to the Clinical Trial Management Public Platform. Data quality will be checked regularly by the principal investigator of this study.

Statistical analysis and plans

Sample size calculation: Exhaled samples of 40 lung cancer patients and 40 healthy subjects were collected and analyzed as preliminary data for sample size calculation. Thirty one VOCs were measured in at least 70% of samples. Among them, ethanol, isoprene, acetone, isopropanol, benzene, ethylbenzene, octanal, nonanal and decane were reported in literature as exhaled biomarkers of lung cancer. The significant level is set as 0.029 based on False Discovery Rate (FDR) procedure for multiple comparisons correction[26]. The FDR procedure is as follows: let $P_{(l)}, .. ., P_{(n)}$ be the ordered p-values for testing hypotheses $H_0 = \{H_{(1)} ... H_{(n)}\}$, and then H_0 is rejected if $P_{(i)} \leqslant ja /n$ for any $i = 1, .. ., n$. In this study, $P_{(18)}$=0.00068 is the largest one that rejected H_0, so p=0.05 × 18/31= 0.029. Finally, 350 subjects for each group are expected to identify the reported 9 biomarkers (with additional 15 markers) that may be significantly different between lung cancer patients and controls, with a power of 90% or greater. From our previous work, 10% breath samples may be invalid due to sampling or storage faults. So the sample size of subjects for this study is estimated to be 778 (700/0.9), with 389 for lung cancer group and 389 for control group. Mann-Whitney U test (SPSS, IBM) was used to evaluate the statistical differences of 31 VOCs in preliminary data. Sample size was calculated by G-power software based on normal distribution of the parent.

Statistical analysis for this study will include the following: Comparisons of VOCs between lung cancer patients and controls will be performed with the Mann-Whitney U test. Multivariate analysis (supervised machine learning) will be utilized to assess the importance of VOCs for distinguishing different disease conditions. Biomarkers showing statistically significant differences and high discriminant value will be selected as lung cancer biomarkers. Subjects will be randomly assigned to training set or prediction set. A logic regression model of exhaled biomarkers will be constructed in the training set and then be tested in prediction set. Receiver operating characteristic (ROC) curve will be constructed and the AUC of the ROC curve will be used to evaluate the diagnostic accuracy of the model. The influence of potential confounders, e.g. demographic factors, will also be investigated using a Mann-Whitney U test or χ^2 test.

Logistic regression will be applied to evaluate the impact of potential factors on the identified biomarkers. A two-sided P value of <0.05 will be considered to indicate statistical significance. All analyses will be performed using SPSS (IBM) and WEKA software (version 3.8.3). Investigators will be blinded to group allocation during the data analysis.

Blind tests: None.

ETHICS

This study was approved by the Ethics Committee of Sichuan Cancer Hospital on April 6, 2017 (No.SCCHEC-02-2017-011).

DISSEMINATION

The results of this study will be disseminated through various channels, including presentations at academic conferences, publications in journals and in the news media.

AKNOWLEDGEMENT

The authors appreciate all the lung cancer patients and healthy subjects who participate in this study. We are also grateful to related staff in Nanchong Central Hospital (Jun Bie) and Chengdu Longquanyi District Center for Disease Control and Prevention (Yang Shi, Honghai Ruan) , who provide a lot of help in recruiting and managing healthy subjects.

Author contributions

Yixiang Duan was involved in the study conception. Qiang Li and Yixiang Duan were involved in acquisition of funding and review of full protocol. Wenwen Li and Wei Dai were involved in the study design and protocol writing. Qiuling Shi and Chunyan Wang were involved in statistical analysis plan. Yijing Long was involved in sorting out *operation requirements for breath sampling. Mingxin Liu, Shaohua Xie, Yuanling Liu, Ming Zhang, Xiaoqin Peng and Yifeng Liu were involved in drafting of the protocol.*

Funding

This study was supported by Sichuan Science and Technology Program (grant No. 2017SZ0013).

Competing interests

None declared.

Ethics approval

Ethics Committee of Sichuan Cancer Hospital (No.SCCHEC-02-2017-011).

Provenance and peer review

Not commissioned; externally peer reviewed.

References

1. Chen W, Zheng R, Baade PD, et al. Cancer statistics in China, 2015. *CA Cancer J Clin* 2016;66:115-32.

2. Chen W, Zheng R, Zeng H, et al. Epidemiology of lung cancer in China. *Thorac Cancer* 2015;6:209-15.

3. Torre LA, Siegel RL, Jemal A. Lung cancer statistics. Lung cancer and personalized medicine. Springer 2016:1-19.

4. Howlader N, Noone A-M, Krapcho M, et al. SEER cancer statistics review (CSR) 1975–2011, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/archive/csr/1975_2011/, based on November 2013 SEER data submission, posted to the SEER web site. 2014

5. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. *Lancet* 2000;355:479-85.

6. Ou SHI, Zell JA, Ziogas A, et al. Prognostic factors for survival of stage I nonsmall cell lung cancer patients: A population ‐ based analysis of 19,702 stage I patients in the California Cancer Registry from 1989 to 2003. *Cancer* 2007;110:1532-41.

7. International Early Lung Cancer Action Program I. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 2006;355:1763-71.

8. Gledhill A, Bates C, Henderson D, et al. Sputum cytology: a limited role. *J Clin Pathol* 1997;50:566-8.

9. Sone S, Takashima S, Li F, et al. Mass screening for lung cancer with mobile spiral computed

tomography scanner. *Lancet* 1998;351:1242-5.

10. National Lung Screening Trial Research T. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395-409.

11. Moyer VA. Screening for lung cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014;160:330-8.

12. Ali MU, Miller J, Peirson L, et al. Screening for lung cancer: a systematic review and meta-analysis. *Prev Med* 2016;89:301-14.

13. Patz EF, Pinsky P, Gatsonis C, et al. Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med* 2014;174:269-74.

14. Davis MD, Fowler SJ, Montpetit AJ. Exhaled breath testing - A tool for the clinician and researcher. *Paediatr Respir Rev* 2018

15. Hakim M, Broza YY, Barash O, et al. Volatile organic compounds of lung cancer and possible biochemical pathways. *Chem Rev* 2012;112:5949-66.

16. Haick H, Broza YY, Mochalski P, et al. Assessment, origin, and implementation of breath volatile cancer markers. *Chem Soc Rev* 2014;43:1423-49.

17. Pereira J, Porto-Figueira P, Cavaco C, et al. Breath analysis as a potential and non-invasive frontier in disease diagnosis: an overview. *Metabolites* 2015;5:3-55.

18. Nardi-Agmon I, Peled N. Exhaled breath analysis for the early detection of lung cancer: recent developments and future prospects. *Lung Cancer: Targets and Therapy* 2017;8:31-8.

19. Saalberg Y, Wolff M. VOC breath biomarkers in lung cancer. *Clin Chim Acta* 2016;459:5-9.

20. Mathew TL, Pownraj P, Abdulla S, et al. Technologies for Clinical Diagnosis Using Expired Human Breath Analysis. *Diagnostics* 2015;5:27-60.

21. Smolinska A, Hauschild AC, Fijten RRR, et al. Current breathomics-a review on data pre-processing techniques and machine learning in metabolomics breath analysis. *J Breath Res* 2014;8:027105.

22. Yan YY, Wang QH, Li WW, et al. Discovery of potential biomarkers in exhaled breath for diagnosis of type 2 diabetes mellitus based on GC-MS with metabolomics. *Rsc Adv* 2014;4:25430-9.

23. Li J, Peng YL, Liu Y, et al. Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography-mass spectrometry. *Clin Chim Acta* 2014;436:59-67.

24. Li WW, Liu Y, Lu XY, et al. A cross-sectional study of breath acetone based on diabetic metabolic disorders. *J Breath Res* 2015;9:016005.

25. Li WW, Liu Y, Liu Y, et al. Exhaled isopropanol: new potential biomarker in diabetic breathomics and its metabolic correlations with acetone. *RSC Adv* 2017;7:17480-8.

26. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 1986;73:751-4.

Supplementary

Requirements for researchers during breath sampling

Requirements	**Aims**
Pump out air from Tedlar bag fully	Prevent ambient air contamination
Ensure Tedlar bag is tightly sealed	Prevent ambient air contamination
Connect mouthpiece to Bio-VOC sampler tightly	Avoid exhaled gas leakage during expiration
Rotate the three-way valve to an accurate angle	Avoid dead space gas entering the Tedlar bag
Explain to subjects clearly: Do not take a deep breath; Nasal ventilation is not allowed; Blow slowly and entirely.	Guarantee the collection of late expiratory breath; Avoid dead space air contamination.
Adjust the Bio-VOC plunger and the three-way valve immediately	Avoid ambient air entering the Bio-VOC syringe; Avoid sample leakage during transfer from Bio-VOC to Tedlar bag.
Close the intake valve of Tedlar bag quickly and tightly	Avoid exhaled gas leakage; Prevent ambient air contamination.
Place breath sample in a cold container (2-8℃)	Avoid squeezing of sample bags; Avoid volatilization of exhaled gas.
Clean the Bio-VOC syringe before next usage	Prevent cross-contamination

BMJ Open

Research protocol for an exploratory study: VOC biomarkers identification and predictive model construction for lung cancer based on exhaled breath analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-028448.R1
Article Type:	Protocol
Date Submitted by the Author:	21-Feb-2019
Complete List of Authors:	Li, Wenwen; Sichuan University, West China School of Public Health Dai, Wei; Sichuan Cancer Hospital & Institute, Department of Thoracic Surgery Liu, Mingxin; Sichuan Cancer Hospital & Institute, Department of Thoracic Surgery Long, Yijing; Sichuan University, The College of Life Sciences Wang, Chunyan; Sichuan University, The College of Life Sciences Xie, Shaohua; Chengdu Medical College, Graduate School Liu, Yuanling; Sichuan University, The College of Life Sciences Zhang, Yinchenxi; Sichuan University, The College of Life Sciences Shi, Qiuling; The University of Texas MD Anderson Cancer Center, Department of Symptom Research Peng, Xiaoqin; Chengdu Medical College, Graduate School Liu, Yifeng; Chengdu Medical College, Graduate School Li, Qiang; Sichuan Cancer Hospital & Institute, Department of Thoracic Surgery Duan, Yixiang; Sichuan University, The College of Life Sciences
<b>Primary Subject Heading</b>:	Diagnostics
Secondary Subject Heading:	Oncology, Diagnostics
Keywords:	lung cancer, early diagnosis, exhaled breath, volatile organic compounds, biomarker

Research protocol for an exploratory study: VOC biomarkers identification and predictive model construction for lung cancer based on exhaled breath analysis

Wenwen Li[1, 2#], Wei Dai[3#], Mingxin Liu[3#], Yijing Long[2, 4], Chunyan Wang[2, 4], Shaohua Xie[3, 5], Yuanling Liu[2, 4], Yinchenxi Zhang[2, 4], Qiuling Shi[6], Xiaoqin Peng[3, 5], Yifeng Liu[3, 5], Qiang Li[3*], Yixiang Duan[2, 4*]

[1]West China School of Public Health, Sichuan University, Chengdu, Sichuan, China.

[2]Research Center of Analytical Instrumentation, Sichuan University, Chengdu, Sichuan, China.

[3]Department of Thoracic Surgery, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, Sichuan, China.

[4]Key Laboratory of Bio-resource and Eco-environment, Ministry of Education, The College of Life Sciences, Sichuan University, Chengdu, Sichuan, China.

[5]Graduate School, Chengdu Medical College, Chengdu, Sichuan, China.

[6]Department of Symptom Research, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

***Corresponding author:**

Yixiang Duan, Research Center of Analytical Instrumentation, College of Life Sciences, Sichuan University, 29 Wangjiang Road, Chengdu 610065, Sichuan, China. Tel: +86-028-85418180, Fax: +86-028-85418180, Email: yduan@scu.edu.cn.

Qiang Li, Department of Thoracic Surgery, Sichuan Cancer Hospital & Institute, No. 55, Section 4, South Renmin Road, Chengdu 610041, Sichuan, China, Tel: +86-028-85420366, Fax: +86-028-85420116, Email: liqiang@sichuancancer.org.

[#]These authors contributed equally to this work.

ABSTRACT

Introduction: Lung cancer is the most common cancer and the leading cause of cancer death in China, as well as in the world. Late diagnosis is the main obstacle to improving survival. Currently, early detection methods for lung cancer have many limitations, e.g. low specificity, risk of radiation exposure and overdiagnosis. Exhaled breath analysis is one of the most promising non-invasive techniques for early detection of lung cancer. The aim of this study is to identify volatile organic compound (VOC) biomarkers in lung cancer and to construct a predictive model for lung cancer based on exhaled breath analysis.

Methods and analysis: The study will recruit 389 lung cancer patients in one cancer centre and 389 healthy subjects in two lung cancer screening centres. Bio-VOC breath sampler and Tedlar bag will be used to collect breath samples. Gas chromatography-mass spectrometry (GC-MS) coupled with solid phase microextraction (SPME) technique will be used to analyze VOCs in exhaled breath. VOC biomarkers with statistical significance and showing abilities to discriminate lung cancer patients from healthy subjects will be selected for the construction of predictive model for lung cancer.

Ethics and dissemination: The study was approved by the Ethics Committee of Sichuan Cancer Hospital on April 6, 2017 (No.SCCHEC-02-2017-011). The results of this study will be disseminated in presentations at academic conferences, publications in peer-reviewed journals and the news media.

Trials registration number: ChiCTR-DOD-17011134.

KEYWORDS: lung cancer, early diagnosis, exhaled breath, volatile organic compounds, biomarker.

Word count: 2117 words.

ARTICLE SUMMARY

Article focus

To construct a predictive model for early prediction of lung cancer based on exhaled biomarkers.

Strengths and limitations of this study

1. This study will select exhaled biomarkers of lung cancer based on the largest sample size from multiple centres ever analyzed by gas chromatography-mass spectrometry.

2. Well-selected healthy controls with high risk but negative of lung cancer on chest CT will be recruited in this study.

3. Gas chromatography-mass spectrometry can only obtain a limited amount of volatile organic compounds (VOCs) in exhaled breath and some VOCs cannot be detected.

INTRODUCTION

In China, lung cancer incidence and related mortality have been increasing annually for the last 30 years.[1 2] Lung cancer is now the most common cancer and the leading cause of cancer death, accounting for 24.9% of all cancer deaths in China in 2010.[1 2] Lung cancer is also the most common cancer and the leading cause of cancer death worldwide.[3] Only 16.8% of all patients with lung cancer survive for 5 years or more after diagnosis,[4] mainly because lung cancer is often staged as locally advanced or metastatic disease at the initial diagnosis.[4 5] However, the 5-year survival rate of patients with stage I lung cancer is greater than 60%,[5 6] and the estimated 10-year survival rate of patients with stage I lung cancer detected on computed tomography (CT) screening can reach 88%.[7] Therefore, early detection is the key to improving survival rates of patients with lung cancer.

Currently, early detection techniques for lung cancer have many limitations, and a simple, reliable and non-invasive early lung cancer screening technique is urgently needed. Sputum cytology has a very low sensitivity.[8] Chest radiography is radioactive and has a very high rate of false-negative results, especially in the detection of early-stage lung cancers.[9] Screening with low-dose CT can reduce mortality from lung cancer by 20%, and lung cancer screening using low-dose CT for high-risk individuals is now recommended.[10 11] However, there are also many disadvantages of low-dose CT, such as a high false-positive rate, overdiagnosis, the risk of radiation exposure and high cost, which limit its application in population-based screening.[12 13]

Exhaled breath analysis is completely non-invasive and has great potential to become a screening and diagnostic method for early detection of cancer.[14-17] The majority of previous studies focusing on lung cancer were conducted on small samples.[18 19] Potential volatile organic compounds (VOCs) biomarkers for lung cancer have been discussed and summarized.[15 19] However, to date, there are no unified VOC biomarkers for lung cancer, and the sets of VOCs employed vary between studies. Therefore, breath analysis is still in an early stage of clinical application. Several factors account for this situation, such as large variation in sample size, diverse sample collection approaches, different analytical techniques, and different data processing methods.[19-21] Our group has investigated exhaled VOC biomarkers in diabetes and breast cancer.[22-25] We have rich experience in breath sample collection, exhaled VOCs analysis and data processing. In addition, we have a stable and reliable source of lung cancer patients and healthy subjects. In this study, we aim to identify exhaled VOC biomarkers of lung cancer and establish a predictive model for lung cancer. Our research hypothesis is that the predictive model will reach 80% sensitivity and 80% specificity through cross validation. In subsequent studies, this model will be validated in a clinical setting and population-based screening.

METHOD AND ANALYSIS

Main Centres: Sichuan Cancer Hospital, Sichuan University, Nanchong Central Hospital and Chengdu Longquanyi District Center for Disease Control and Prevention. Lung cancer patients will be recruited from Sichuan Cancer Hospital. Healthy subjects will be recruited from two lung cancer screening centres, including Nanchong Central Hospital and Chengdu Longquanyi District Center for Disease Control and Prevention. Breath sample analysis will be conducted in Sichuan University.

Dates of the study: From March 1, 2017 to February 28, 2020.

Design

Inclusion criteria: Lung cancer patients and healthy subjects were both aged from 50 to 74 years. Lung cancer patients should have a pathological diagnosis of primary lung cancer based on the 2015 World Health Organization Classification of lung tumors,[26] The pathologic stages were based on the eighth edition of the TNM classification for lung cancer.[27] All the recruited patients should not receive any cancer treatment before breath sampling. Healthy subjects should be negative of lung cancer on chest CT.

Exclusion criteria: Patients and controls with diabetes, other malignancies, active asthma, severe liver dysfunction, end-stage renal disease, and acute inflammation will be excluded.

Breath sampling procedures

Each participant recruited into this study will be given an information leaflet explaining the research and will sign an informed consent form. Participants are required to fast for at least 8 h and rest for at least 10 min in a separate room with good ventilation before breath sampling. Exhaled gas will be collected in the morning from 7:00 am to 9:00 am. So, fasting from 23:00 pm the day before can meet the requirement. For lung cancer subjects who will not receive surgery, breath samples collection will be performed after pathologic diagnosis and prior to cancer treatment. For patients undergoing surgery, breath samples will be collected the day before the surgery. If the postoperative pathology is not primary lung cancer, the patient will be excluded.

Subjects will be asked to take a normal inhalation followed by a normal exhalation via their mouth. Deep inhalation before sampling and nasal ventilation during sampling will not be allowed. Each subject will exhale three times to complete a breath sample collection. Exhaled breath gas will be collected with Bio-VOC breath sampler (Markes Int. U.K). Subjects will exhale gently into the Bio-VOC syringe until they feel mild resistance. The Bio-VOC sampler collects end-tidal breath gas by expelling the dead space air from the nasopharynx and upper airway using a three-way valve. The end-tidal breath gas will then be transferred to Tedlar bag (500 mL) through the three-way valve for storage and transportation. Breath samples will be kept at -40 ℃ until analysis (within seven days). The storage stability of VOCs in Tedlar bag at -40 ℃ has been assessed (supplementary figure 1). The results indicated VOCs could remain stable within seven days in Tedlar bags at -40 ℃.

VOCs analysis

Solid phase microextraction (SPME) technique will be utilized in this study to pre-concentrate VOCs in breath samples prior to gas chromatography-mass spectrometry (GC-MS) analysis. Divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) fiber will be used to extract exhaled VOCs. Samples will be analyzed using a Thermo Scientific TRACE 1300 gas chromatograph coupled to a TSQ8000 triple quadrupole mass spectrometer. VF-624ms capillary column (60 m × 0.25 mm × 0.25 μm, Agilent Technology) will be used to separate VOCs. The analytical process will be as follows: first, the breath sample kept at -40℃ will be incubated at 37 ℃ for 5 min; then a DVB/CAR/PDMS fiber will be used to pre-concentrate VOCs for 30 min at 37 ℃, and finally the fiber will be desorbed thermally at the front inlet of the gas chromatograph. Split mode and a specific liner for SPME fiber will be used for the gas chromatograph. Electron ionization source (70 eV) will be used for the mass spectrometer.

Endpoint: Exhaled VOC biomarkers of lung cancer and the accuracy of the predictive model for lung cancer .

Quality assurance

Standardization of breath sampling: In order to minimize inter- and intra-observer error caused by researchers during breath sampling, a fixed number of well-trained staff will be appointed to collect breath samples from patients and controls. Operation requirements that may bring about errors in breath sampling and lead to biased results are listed in supplementary table 1. All staff participating in breath sampling will be trained and tested for the process and requirements of sampling. In addition, all researchers involved in this study will go through the certification process, including informed consent signing, breath sampling, samples storage, and transportation.

Ambient room air: Breath samples will be collected from patients and controls in a separate room with good ventilation. Ambient room air will be collected simultaneously with each batch of breath samples. The ambient air is used to monitor possible VOC contamination, because it may be inhaled by subjects and become a prominent component of the exhaled breath, leading to anomalous results. Samples with VOCs that appear only in one batch of samples and simultaneously in the corresponding ambient room air at significant concentrations will be regarded as contamination and will be excluded.

Data management and monitoring: All data collected in this study, including demographic information, clinicopathologic information and raw GC-MS data, will be uploaded to the Clinical Trial Management Public Platform (http://www.medresman.org). Data quality will be checked regularly by the principal investigator of this study. Data monitoring will be performed regularly by the Ethics Committee of Sichuan Cancer Hospital.

Statistical analysis and plans

Sample size calculation: Exhaled samples of 40 lung cancer patients and 40 healthy subjects were collected and analyzed as preliminary data for sample size calculation. Thirty-one VOCs were measured in at least 70% of samples. Among them, ethanol, isoprene, acetone, isopropanol, benzene, ethylbenzene, octanal, nonanal and decane were reported in literature as exhaled biomarkers of lung cancer. The significant level is set as 0.029 based on False Discovery Rate (FDR) procedure for multiple comparisons correction.[28] The FDR procedure is as follows: let $P_{(1)}, .. ., P_{(n)}$ be the ordered p-values for testing hypotheses $H_0 = \{H_{(1)} ... H_{(n)}\}$, and then H_0 is rejected if $P_{(i)} \leqslant ja /n$ for any $i = 1, .. ., n$. In this study, $P_{(18)}$=0.00068 is the largest one that rejected H_0, so p=0.05 × 18/31= 0.029. Finally, 350 subjects for each group are expected to identify the reported 9 biomarkers (with additional 15 markers) that may be significantly different between lung cancer patients and controls, with a power of 90%

or greater. From our previous work, 10% breath samples may be invalid due to sampling or storage faults. So the sample size of subjects for this study is estimated to be 778 (700/0.9), with 389 for lung cancer group and 389 for the control group. Mann-Whitney U test (SPSS, IBM) was used to evaluate the statistical differences of 31 VOCs in preliminary data. The sample size was calculated by G-power software based on normal distribution of the parent.

Statistical analysis for this study will include the following: Comparisons of VOCs between lung cancer patients and controls will be performed with the Mann-Whitney U test. Principal component analysis (PCA), linear discriminant analysis (LDA), or independent component analysis (ICA), etc. will be used to reduce the dimension of datasets. Afterwards, multivariate analysis (supervised machine learning method such as PLSDA, OPLSDA, sPLSDA) and cross validation (10-fold or leave-one-out) will be utilized to identify exhaled biomarkers. Subjects will be randomly assigned to training set or testing set. A logistic regression model will be constructed based on exhaled biomarkers through the training set and trained by the testing set. Then the model will be verified through leave-one-out cross validation. Receiver operating characteristic (ROC) curve will be constructed and the area under the curve (AUC) of the ROC curve will be used to evaluate the diagnostic accuracy of the model. Correct classification rates (CCRs) of the model will be calculated afterwards. The influence of potential confounders, e.g. age, sex, smoking status, alcohol drinking, will also be investigated through univariate analysis. Logistic regression will be applied to evaluate the impact of potential factors on the identified biomarkers. A two-sided P value of <0.05 will be considered to indicate statistical significance. All analyses will be performed using SPSS (IBM) and WEKA software (version 3.8.3). Data analyst will be blinded to group allocation during the data processing.

Data availability statement

After the main results of this study are published, anonymous data that support the published articles will be given to the applicant from the corresponding author on reasonable request.

Patient and public involvement statement

Patients, healthy subjects and the public were not involved in the study design, recruitment to and conduct. We do not have a plan to inform the result to the study participants unless they apply for it.

ETHICS AND DISSEMINATION

This study was approved by the Ethics Committee of Sichuan Cancer Hospital on April 6, 2017 (No.SCCHEC-02-2017-011). The results of this study will be disseminated through various channels, including presentations at academic conferences, publications in journals and in the news media.

Acknowledgements The authors appreciate all the lung cancer patients, patient advisers, and healthy subjects who participate in this study. We are also grateful to relate staff in Nanchong Central Hospital (Jun Bie) and Chengdu Longquanyi District Center for Disease Control and Prevention (Yang Shi, Honghai Ruan), who provide a lot of help in recruiting and managing healthy subjects.

Author contributions YXD was involved in the study conception. QL and YXD were involved in acquisition of funding and review of full protocol. WWL and WD were involved in the study design and protocol writing. QLS and CYW were involved in statistical analysis plan. YJL was involved in sorting out operation requirements for breath sampling. MXL, SHX, YLL,YCXZ, XQP and YFL were involved in drafting of the protocol.

Funding Sichuan Science and Technology Program (grant number 2017SZ0013).

Competing interests None declared.

Patient consent obtained.

Ethics approval Ethics Committee of Sichuan Cancer Hospital (No.SCCHEC-02-2017-011).

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Chen W, Zheng R, Baade PD*, et al.* Cancer statistics in China, 2015. *CA Cancer J Clin* 2016;66:115-32.

2. Chen W, Zheng R, Zeng H*, et al.* Epidemiology of lung cancer in China. *Thorac Cancer* 2015;6:209-15.

3. Torre LA, Siegel RL, Jemal A. Lung cancer statistics. Lung cancer and personalized medicine. Springer 2016:1-19.

4. Howlader N, Noone A-M, Krapcho M*, et al.* SEER cancer statistics review (CSR) 1975–2011, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/archive/csr/1975_2011/, based on November 2013 SEER data submission, posted to the SEER web site. 2014

5. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. *Lancet* 2000;355:479-85.

6. Ou SHI, Zell JA, Ziogas A*, et al.* Prognostic factors for survival of stage I nonsmall cell lung cancer patients: A population ‐ based analysis of 19,702 stage I patients in the California Cancer Registry from 1989 to 2003. *Cancer* 2007;110:1532-41.

7. International Early Lung Cancer Action Program I. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 2006;355:1763-71.

8. Gledhill A, Bates C, Henderson D*, et al.* Sputum cytology: a limited role. *J Clin Pathol* 1997;50:566-8.

9. Sone S, Takashima S, Li F*, et al.* Mass screening for lung cancer with mobile spiral computed tomography scanner. *Lancet* 1998;351:1242-5.

10. National Lung Screening Trial Research T. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395-409.

11. Moyer VA. Screening for lung cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014;160:330-8.

12. Ali MU, Miller J, Peirson L*, et al.* Screening for lung cancer: a systematic review and meta-analysis. *Prev Med* 2016;89:301-14.

13. Patz EF, Pinsky P, Gatsonis C*, et al.* Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med* 2014;174:269-74.

14. Davis MD, Fowler SJ, Montpetit AJ. Exhaled breath testing - A tool for the clinician and researcher. *Paediatr Respir Rev* 2018

15. Hakim M, Broza YY, Barash O*, et al.* Volatile organic compounds of lung cancer and possible biochemical pathways. *Chem Rev* 2012;112:5949-66.

16. Haick H, Broza YY, Mochalski P*, et al.* Assessment, origin, and implementation of breath volatile cancer markers. *Chem Soc Rev* 2014;43:1423-49.

17. Pereira J, Porto-Figueira P, Cavaco C*, et al.* Breath analysis as a potential and non-invasive frontier in disease diagnosis: an overview. *Metabolites* 2015;5:3-55.

18. Nardi-Agmon I, Peled N. Exhaled breath analysis for the early detection of lung cancer: recent developments and future prospects. *Lung Cancer: Targets and Therapy* 2017;8:31-8.

19. Saalberg Y, Wolff M. VOC breath biomarkers in lung cancer. *Clin Chim Acta* 2016;459:5-9.

20. Mathew TL, Pownraj P, Abdulla S*, et al.* Technologies for Clinical Diagnosis Using Expired Human Breath Analysis. *Diagnostics* 2015;5:27-60.

21. Smolinska A, Hauschild AC, Fijten RRR*, et al.* Current breathomics-a review on data pre-processing techniques and machine learning in metabolomics breath analysis. *J Breath Res* 2014;8:027105.

22. Yan YY, Wang QH, Li WW*, et al.* Discovery of potential biomarkers in exhaled breath for diagnosis of type 2 diabetes mellitus based on GC-MS with metabolomics. *Rsc Adv* 2014;4:25430-9.

23. Li J, Peng YL, Liu Y*, et al.* Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography-mass spectrometry. *Clin Chim Acta* 2014;436:59-67.

24. Li WW, Liu Y, Lu XY*, et al.* A cross-sectional study of breath acetone based on diabetic metabolic disorders. *J Breath Res* 2015;9:016005.

25. Li WW, Liu Y, Liu Y*, et al.* Exhaled isopropanol: new potential biomarker in diabetic breathomics and its metabolic correlations with acetone. *RSC Adv* 2017;7:17480-8.

26. Travis WD, Brambilla E, Nicholson AG*, et al.* The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* 2015;10:1243-60.

27. Goldstraw P, Chansky K, Crowley J*, et al.* The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol* 2016;11:39-51.

28. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 1986;73:751-4.

Supplementary

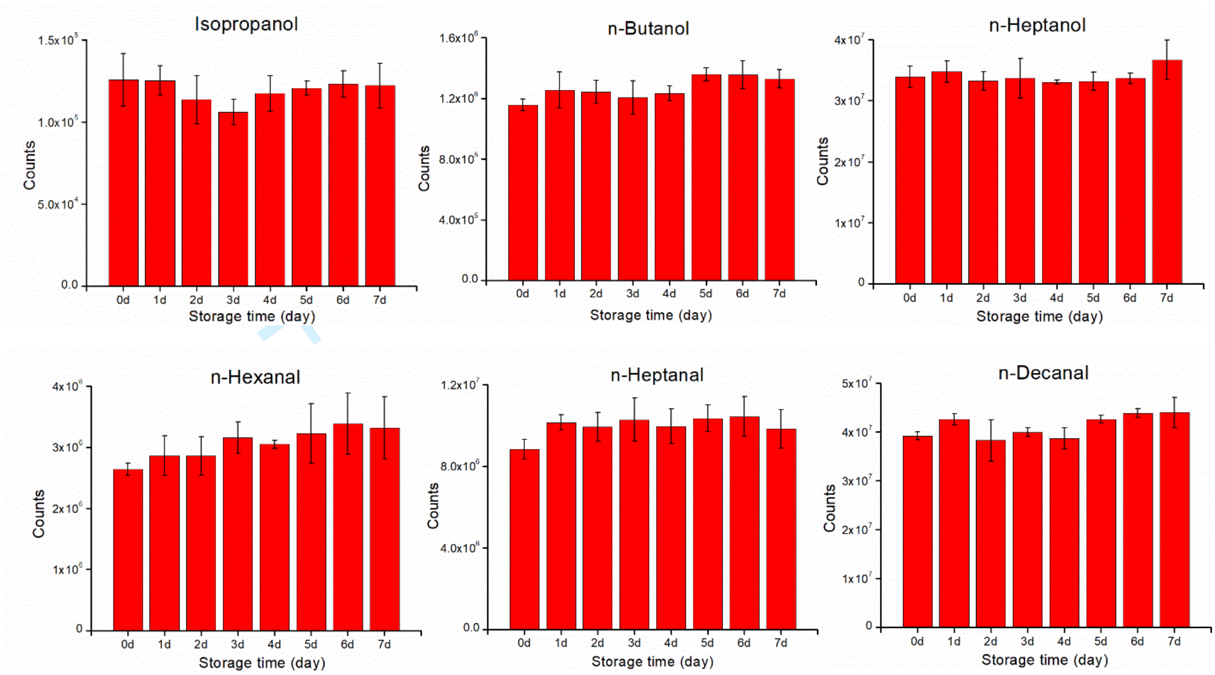

Figure 1. Concentrations of isopropanol, n-butanol, n-heptanol, n-hexanal, n-heptanal, and n-decanal (100 ppbv for each compound) in Tedlar bags at -40 °C are nearly constant within seven days, indicating a good storage stability.

Table1. Requirements for researchers during breath sampling

Requirements	Aims
Pump out air from Tedlar bag fully	Prevent ambient air contamination
Ensure Tedlar bag is tightly sealed	Prevent ambient air contamination
Connect mouthpiece to Bio-VOC sampler tightly	Avoid exhaled gas leakage during expiration
Rotate the three-way valve to an accurate angle	Avoid dead space gas entering the Tedlar bag
Explain to subjects clearly: Do not take a deep breath; Nasal ventilation is not allowed; Blow slowly and entirely.	Guarantee the collection of late expiratory breath; Avoid dead space air contamination.
Adjust the Bio-VOC plunger and the three-way valve immediately	Avoid ambient air entering the Bio-VOC syringe; Avoid sample leakage during transfer from Bio-VOC to Tedlar bag.
Close the intake valve of Tedlar bag quickly and tightly	Avoid exhaled gas leakage; Prevent ambient air contamination.
Place breath sample in a cold container (2-8 °C)	Avoid squeezing of sample bags; Avoid volatilization of exhaled gas.
Clean the Bio-VOC syringe with dry dust-free cloth before the next use	Prevent cross-contamination

BMJ Open

Research protocol for an exploratory study: VOC biomarkers identification and predictive model construction for lung cancer based on exhaled breath analysis

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-028448.R2
Article Type:	Protocol
Date Submitted by the Author:	26-Mar-2019
Complete List of Authors:	Li, Wenwen; Sichuan University, West China School of Public Health Dai, Wei; Sichuan Cancer Hospital & Institute, Department of Thoracic Surgery Liu, Mingxin; Sichuan Cancer Hospital & Institute, Department of Thoracic Surgery Long, Yijing; Sichuan University, The College of Life Sciences Wang, Chunyan; Sichuan University, The College of Life Sciences Xie, Shaohua; Chengdu Medical College, Graduate School Liu, Yuanling; Sichuan University, The College of Life Sciences Zhang, Yinchenxi; Sichuan University, The College of Life Sciences Shi, Qiuling; The University of Texas MD Anderson Cancer Center, Department of Symptom Research Peng, Xiaoqin; Chengdu Medical College, Graduate School Liu, Yifeng; Chengdu Medical College, Graduate School Li, Qiang; Sichuan Cancer Hospital & Institute, Department of Thoracic Surgery Duan, Yixiang; Sichuan University, The College of Life Sciences
<b>Primary Subject Heading</b>:	Diagnostics
Secondary Subject Heading:	Oncology, Diagnostics
Keywords:	lung cancer, early diagnosis, exhaled breath, volatile organic compounds, biomarker

Research protocol for an exploratory study: VOC biomarkers identification and predictive model construction for lung cancer based on exhaled breath analysis

Wenwen Li[1, 2#], Wei Dai[3#], Mingxin Liu[3#], Yijing Long[2, 4], Chunyan Wang[2, 4], Shaohua Xie[3, 5], Yuanling Liu[2, 4], Yinchenxi Zhang[2, 4], Qiuling Shi[6], Xiaoqin Peng[3, 5], Yifeng Liu[3, 5], Qiang Li[3*], Yixiang Duan[2, 4*]

[1]West China School of Public Health, Sichuan University, Chengdu, Sichuan, China.

[2]Research Center of Analytical Instrumentation, Sichuan University, Chengdu, Sichuan, China.

[3]Department of Thoracic Surgery, Sichuan Cancer Hospital & Institute, Sichuan Cancer Center, School of Medicine, University of Electronic Science and Technology of China, Chengdu, Sichuan, China.

[4]Key Laboratory of Bio-resource and Eco-environment, Ministry of Education, The College of Life Sciences, Sichuan University, Chengdu, Sichuan, China.

[5]Graduate School, Chengdu Medical College, Chengdu, Sichuan, China.

[6]Department of Symptom Research, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

***Corresponding author:**

Yixiang Duan, Research Center of Analytical Instrumentation, College of Life Sciences, Sichuan University, 29 Wangjiang Road, Chengdu 610065, Sichuan, China. Tel: +86-028-85418180, Fax: +86-028-85418180, Email: yduan@scu.edu.cn.

Qiang Li, Department of Thoracic Surgery, Sichuan Cancer Hospital & Institute, No. 55, Section 4, South Renmin Road, Chengdu 610041, Sichuan, China, Tel: +86-028-85420366, Fax: +86-028-85420116, Email: liqiang@sichuancancer.org.

[#]These authors contributed equally to this work.

ABSTRACT

Introduction: Lung cancer is the most common cancer and the leading cause of cancer death in China, as well as in the world. Late diagnosis is the main obstacle to improving survival. Currently, early detection methods for lung cancer have many limitations, e.g. low specificity, risk of radiation exposure and overdiagnosis. Exhaled breath analysis is one of the most promising non-invasive techniques for early detection of lung cancer. The aim of this study is to identify volatile organic compound (VOC) biomarkers in lung cancer and to construct a predictive model for lung cancer based on exhaled breath analysis.

Methods and analysis: The study will recruit 389 lung cancer patients in one cancer centre and 389 healthy subjects in two lung cancer screening centres. Bio-VOC breath sampler and Tedlar bag will be used to collect breath samples. Gas chromatography-mass spectrometry (GC-MS) coupled with solid phase microextraction (SPME) technique will be used to analyze VOCs in exhaled breath. VOC biomarkers with statistical significance and showing abilities to discriminate lung cancer patients from healthy subjects will be selected for the construction of predictive model for lung cancer.

Ethics and dissemination: The study was approved by the Ethics Committee of Sichuan Cancer Hospital on April 6, 2017 (No.SCCHEC-02-2017-011). The results of this study will be disseminated in presentations at academic conferences, publications in peer-reviewed journals and the news media.

Trials registration number: ChiCTR-DOD-17011134.

KEYWORDS: lung cancer, early diagnosis, exhaled breath, volatile organic compounds, biomarker.

Word count: 2146 words.

ARTICLE SUMMARY

Article focus

To construct a predictive model for early prediction of lung cancer based on exhaled biomarkers.

Strengths and limitations of this study

1. This study will select exhaled biomarkers of lung cancer based on the largest sample size from multiple centres ever analyzed by gas chromatography-mass spectrometry.

2. Well-selected healthy controls with high risk but negative of lung cancer on chest CT will be recruited in this study.

3. Gas chromatography-mass spectrometry can only obtain a limited amount of volatile organic compounds (VOCs) in exhaled breath and some VOCs cannot be detected.

INTRODUCTION

In China, lung cancer incidence and related mortality have been increasing annually for the last 30 years.[1 2] Lung cancer is now the most common cancer and the leading cause of cancer death, accounting for 24.9% of all cancer deaths in China in 2010.[1 2] Lung cancer is also the most common cancer and the leading cause of cancer death worldwide.[3] Only 16.8% of all patients with lung cancer survive for 5 years or more after diagnosis,[4] mainly because lung cancer is often staged as locally advanced or metastatic disease at the initial diagnosis.[4 5] However, the 5-year survival rate of patients with stage I lung cancer is greater than 60%,[5 6] and the estimated 10-year survival rate of patients with stage I lung cancer detected on computed tomography (CT) screening can reach 88%.[7] Therefore, early detection is the key to improving survival rates of patients with lung cancer.

Currently, early detection techniques for lung cancer have many limitations, and a simple, reliable and non-invasive early lung cancer screening technique is urgently needed. Sputum cytology has a very low sensitivity.[8] Chest radiography is radioactive and has a very high rate of false-negative results, especially in the detection of early-stage lung cancers.[9] Screening with low-dose CT can reduce mortality from lung cancer by 20%, and lung cancer screening using low-dose CT for high-risk individuals is now recommended.[10 11] However, there are also many disadvantages of low-dose CT, such as a high false-positive rate, overdiagnosis, the risk of radiation exposure and high cost, which limit its application in population-based screening.[12 13]

Exhaled breath analysis is completely non-invasive and has great potential to become a screening and diagnostic method for early detection of cancer.[14-17] The majority of previous studies focusing on lung cancer were conducted on small samples.[18 19] Potential volatile organic compounds (VOCs) biomarkers for lung cancer have been discussed and summarized.[15 19] However, to date, there are no unified VOC biomarkers for lung cancer, and the sets of VOCs employed vary between studies. Therefore, breath analysis is still in an early stage of clinical application. Several factors account for this situation, such as large variation in sample size, diverse sample collection approaches, different analytical techniques, and different data processing methods.[19-21] Our group has investigated exhaled VOC biomarkers in diabetes and breast cancer.[22-25] We have rich experience in breath sample collection, exhaled VOCs analysis and data processing. In addition, we have a stable and reliable source of lung cancer patients and healthy subjects. In this study, we aim to identify exhaled VOC biomarkers of lung cancer and establish a predictive model for lung cancer. Our research hypothesis is that the predictive model will reach 80% sensitivity and 80% specificity through cross validation. In subsequent studies, this model will be validated in a clinical setting and population-based screening.

METHOD AND ANALYSIS

Main Centres: Sichuan Cancer Hospital, Sichuan University, Nanchong Central Hospital and Chengdu Longquanyi District Center for Disease Control and Prevention. Lung cancer patients will be recruited from Sichuan Cancer Hospital. Healthy subjects will be recruited from two lung cancer screening centres, including Nanchong Central Hospital and Chengdu Longquanyi District Center for Disease Control and Prevention. Breath sample analysis will be conducted in Sichuan University.

Dates of the study: From March 1, 2017 to February 28, 2020.

Design

Inclusion criteria: Lung cancer patients and healthy subjects are both aged from 50 to 74 years. Lung cancer patients should have a pathological diagnosis of primary lung cancer based on the 2015 World Health Organization Classification of lung tumors.[26] The pathologic stages were based on the eighth edition of the TNM classification for lung cancer.[27] All the recruited patients should not receive any cancer treatment before breath sampling. Healthy subjects recruited from two lung cancer screening centres should be negative of lung cancer on chest CT based on a previous project “Early Diagnosis and Early Treatment of Rural Cancer” in China from 2014.

Exclusion criteria: Patients and controls with diabetes, other malignancies, active asthma, severe liver dysfunction, end-stage renal disease, and acute inflammation will be excluded.

Breath sampling procedures

Each participant recruited into this study will be given an information leaflet explaining the research and will sign an informed consent form. Participants are required to fast for at least 8 h and rest for at least 10 min in a separate room with good ventilation before breath sampling. Exhaled gas will be collected in the morning from 7:00 am to 9:00 am. So, fasting from 23:00 pm the day before can meet the requirement. For lung cancer subjects who will not receive surgery, breath sample collection will be performed after pathologic diagnosis and prior to cancer treatment. For patients undergoing surgery, breath samples will be collected the day before the surgery. If the postoperative pathology is not primary lung cancer, the patient will be excluded.

All the subjects will be asked to take a normal inhalation followed by a normal exhalation via their mouth. Deep inhalation before sampling and nasal ventilation during sampling will not be allowed. Each subject will exhale three times to complete a breath sample collection. Exhaled breath gas will be collected with Bio-VOC breath sampler (Markes Int. U.K). Subjects will exhale gently into the Bio-VOC syringe until they feel mild resistance. The Bio-VOC sampler collects end-tidal breath gas by expelling the dead space air from the nasopharynx and upper airway using a three-way valve. The end-tidal breath gas will then be transferred to Tedlar bag (500 mL) through the three-way valve for storage and transportation. Breath samples will be kept at -40 ℃ until analysis (within seven days). The storage stability of VOCs in Tedlar bag at -40 ℃

has been assessed (supplementary figure 1). And the results indicated that VOCs could remain stable within seven days in Tedlar bags at -40 ℃.

VOCs analysis

Solid phase microextraction (SPME) technique will be utilized in this study to pre-concentrate VOCs in breath samples prior to gas chromatography-mass spectrometry (GC-MS) analysis. Divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) fiber will be used to extract exhaled VOCs. Samples will be analyzed using a Thermo Scientific TRACE 1300 gas chromatograph coupled to a TSQ8000 triple quadrupole mass spectrometer. VF-624ms capillary column (60 m × 0.25 mm × 0.25 μm, Agilent Technology) will be used to separate VOCs. The analytical process will be as follows: first, the breath sample kept at -40℃ will be incubated at 37 ℃ for 5 min; then a DVB/CAR/PDMS fiber will be used to pre-concentrate VOCs for 30 min at 37 ℃, and finally the fiber will be desorbed thermally at the front inlet of the gas chromatograph. Split mode and a specific liner for SPME fiber will be used for the gas chromatograph. Electron ionization source (70 eV) will be used for the mass spectrometer.

Endpoint: Exhaled VOC biomarkers of lung cancer and the accuracy of the predictive model for lung cancer .

Quality assurance

Standardization of breath sampling: In order to minimize inter- and intra-observer error caused by researchers during breath sampling, a fixed number of well-trained staff will be appointed to collect breath samples from patients and controls. Operation requirements that may bring about errors in breath sampling and lead to biased results are listed in supplementary table 1. All staff participating in breath sampling will be trained and tested for the process and requirements of sampling. In addition, all researchers involved in this study will go through the certification process, including informed consent signing, breath sampling, samples storage, and transportation.

Ambient room air: Breath samples will be collected from patients and controls in a separate room with good ventilation. Ambient room air will be collected simultaneously with each batch of breath samples. The ambient air is used to monitor possible VOC contamination, because it may be inhaled by subjects and become a prominent component of the exhaled breath, leading to anomalous results. Samples with VOCs that appear only in one batch of samples and simultaneously in the corresponding ambient room air at significant concentrations will be regarded as contamination and will be excluded.

Data management and monitoring: All data collected in this study, including demographic information, clinicopathologic information and raw GC-MS data, will be uploaded to the Clinical Trial Management Public Platform (http://www.medresman.org). Data quality will be checked regularly by the principal investigator of this study. Data monitoring will be performed regularly by the Ethics Committee of Sichuan Cancer Hospital.

Statistical analysis and plans

Sample size calculation: Exhaled samples of 40 lung cancer patients and 40 healthy subjects were collected and analyzed as preliminary data for sample size calculation. Thirty-one VOCs were measured in at least 70% of samples. Among them, ethanol, isoprene, acetone, isopropanol, benzene, ethylbenzene, octanal, nonanal and decane were reported in literature as exhaled biomarkers of lung cancer. The significant level is set as 0.029 based on False Discovery Rate (FDR) procedure for multiple comparisons correction.[28] The FDR procedure is as follows: let $P_{(l)}, .. ., P_{(n)}$ be the ordered p-values for testing hypotheses $H_0 = \{H_{(1)} ... H_{(n)}\}$, and then H_0 is rejected if $P_{(i)} \leqslant ja\ /n$ for any $i = 1, .. ., n.$ In this study, $P_{(18)}$=0.00068 is the largest one that rejected H_0, so p=0.05 × 18/31= 0.029. Finally, 350 subjects for each group are expected to identify the reported 9 biomarkers (with additional 15 markers) that may be significantly different between lung cancer patients and controls, with a power of 90% or greater. From our previous work, 10% breath samples may be invalid due to sampling or storage faults. Therefore, the sample size of subjects for this study is estimated to be 778 (700/0.9), with 389 for lung cancer group and 389 for the control group. Mann-Whitney U test (SPSS, IBM) was used to evaluate the statistical differences of 31 VOCs in preliminary data. The sample size was calculated by G-power software based on normal distribution of the parent.

Statistical analysis for this study will include the following: Comparisons of VOCs between lung cancer patients and controls will be performed with the Mann-Whitney U test. Principal component analysis (PCA), linear discriminant analysis (LDA), or independent component analysis (ICA), etc. will be used to reduce the dimension of datasets. Afterwards, multivariate analysis (supervised machine learning method such as PLSDA, OPLSDA, sPLSDA) and cross validation (10-fold or leave-one-out) will be utilized to identify exhaled biomarkers. Subjects will be randomly assigned to training set or testing set. A logistic regression model will be constructed based on exhaled biomarkers through the training set and trained by the testing set. Then the model will be verified through leave-one-out cross validation. Receiver operating characteristic (ROC) curve will be constructed and the area under the curve (AUC) of the ROC curve will be used to evaluate the diagnostic accuracy of the model. Correct classification rates (CCRs) of the model will be calculated afterwards. The influence of potential confounders, e.g. age, sex, smoking status, alcohol drinking, will also be investigated through univariate analysis. Logistic regression will be applied to evaluate the impact of potential factors on the identified biomarkers. A two-sided P value of <0.05 will be considered to indicate statistical significance. All analyses will be performed using SPSS (IBM) and WEKA software (version 3.8.3). Data analyst will be blinded to group allocation during the data processing.

Data availability statement

After the main results of this study are published, anonymous data that support the published articles will be given to the applicant from the corresponding author on reasonable request.

Patient and public involvement statement

Patients, healthy subjects and the public were not involved in the study design, recruitment and conduct. We do not have a plan to inform the result to the study participants unless they apply for it.

ETHICS AND DISSEMINATION

This study was approved by the Ethics Committee of Sichuan Cancer Hospital on April 6, 2017 (No.SCCHEC-02-2017-011). The results of this study will be disseminated through various channels, including presentations at academic conferences, publications in journals and in the news media.

Acknowledgements The authors appreciate all the lung cancer patients, patient advisers, and healthy subjects who participate in this study. We are also grateful to related staff in Nanchong Central Hospital (Jun Bie) and Chengdu Longquanyi District Center for Disease Control and Prevention (Yang Shi, Honghai Ruan), who provide a lot of help in recruiting and managing healthy subjects.

Author contributions YXD and QL were involved in the study conception. QL and YXD were involved in acquisition of funding and review of full protocol. WWL and WD were involved in the study design and protocol writing. QLS and CYW were involved in statistical analysis plan. YJL was involved in sorting out operation requirements for breath sampling. MXL, SHX, YLL,YCXZ, XQP and YFL were involved in drafting of the protocol.

Funding Sichuan Science and Technology Program (grant number 2017SZ0013).

Competing interests None declared.

Patient consent obtained.

Ethics approval Ethics Committee of Sichuan Cancer Hospital (No.SCCHEC-02-2017-011).

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. Chen W, Zheng R, Baade PD, *et al.* Cancer statistics in China, 2015. *CA Cancer J Clin* 2016;66:115-32.

2. Chen W, Zheng R, Zeng H, *et al.* Epidemiology of lung cancer in China. *Thorac Cancer* 2015;6:209-15.

3. Torre LA, Siegel RL, Jemal A. Lung cancer statistics. Lung cancer and personalized medicine. Springer 2016:1-19.

4. Howlader N, Noone A-M, Krapcho M, *et al.* SEER cancer statistics review (CSR) 1975–2011, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/archive/csr/1975_2011/, based on November 2013 SEER data submission, posted to the SEER web site. 2014

5. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. *Lancet* 2000;355:479-85.

6. Ou SHI, Zell JA, Ziogas A, *et al.* Prognostic factors for survival of stage I nonsmall cell lung cancer patients: A population ‐ based analysis of 19,702 stage I patients in the California Cancer Registry from 1989 to 2003. *Cancer* 2007;110:1532-41.

7. International Early Lung Cancer Action Program I. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med* 2006;355:1763-71.

8. Gledhill A, Bates C, Henderson D, *et al.* Sputum cytology: a limited role. *J Clin Pathol* 1997;50:566-8.

9. Sone S, Takashima S, Li F, *et al.* Mass screening for lung cancer with mobile spiral computed tomography scanner. *Lancet* 1998;351:1242-5.

10. National Lung Screening Trial Research T. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395-409.

11. Moyer VA. Screening for lung cancer: US Preventive Services Task Force recommendation statement. *Ann Intern Med* 2014;160:330-8.

12. Ali MU, Miller J, Peirson L, *et al.* Screening for lung cancer: a systematic review and meta-analysis. *Prev Med* 2016;89:301-14.

13. Patz EF, Pinsky P, Gatsonis C, *et al.* Overdiagnosis in low-dose computed tomography screening for lung cancer. *JAMA Intern Med* 2014;174:269-74.

14. Davis MD, Fowler SJ, Montpetit AJ. Exhaled breath testing - A tool for the clinician and researcher. *Paediatr Respir Rev* 2018

15. Hakim M, Broza YY, Barash O, *et al.* Volatile organic compounds of lung cancer and possible biochemical pathways. *Chem Rev* 2012;112:5949-66.

16. Haick H, Broza YY, Mochalski P, *et al.* Assessment, origin, and implementation of breath volatile cancer markers. *Chem Soc Rev* 2014;43:1423-49.

17. Pereira J, Porto-Figueira P, Cavaco C, *et al.* Breath analysis as a potential and non-invasive frontier in disease diagnosis: an overview. *Metabolites* 2015;5:3-55.

18. Nardi-Agmon I, Peled N. Exhaled breath analysis for the early detection of lung cancer: recent developments and future prospects. *Lung Cancer: Targets and Therapy* 2017;8:31-8.

19. Saalberg Y, Wolff M. VOC breath biomarkers in lung cancer. *Clin Chim Acta* 2016;459:5-9.

20. Mathew TL, Pownraj P, Abdulla S, *et al.* Technologies for Clinical Diagnosis Using Expired Human Breath Analysis. *Diagnostics* 2015;5:27-60.

21. Smolinska A, Hauschild AC, Fijten RRR, *et al.* Current breathomics-a review on data pre-processing techniques and machine learning in metabolomics breath analysis. *J Breath Res* 2014;8:027105.

22. Yan YY, Wang QH, Li WW, *et al.* Discovery of potential biomarkers in exhaled breath for diagnosis of type 2 diabetes mellitus based on GC-MS with metabolomics. *Rsc Adv* 2014;4:25430-9.

23. Li J, Peng YL, Liu Y, *et al.* Investigation of potential breath biomarkers for the early diagnosis of breast cancer using gas chromatography-mass spectrometry. *Clin Chim Acta* 2014;436:59-67.

24. Li WW, Liu Y, Lu XY, *et al.* A cross-sectional study of breath acetone based on diabetic metabolic disorders. *J Breath Res* 2015;9:016005.

25. Li WW, Liu Y, Liu Y, *et al.* Exhaled isopropanol: new potential biomarker in diabetic breathomics and its metabolic correlations with acetone. *RSC Adv* 2017;7:17480-8.

26. Travis WD, Brambilla E, Nicholson AG, *et al.* The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* 2015;10:1243-60.

27. Goldstraw P, Chansky K, Crowley J, *et al.* The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol* 2016;11:39-51.

28. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. *Biometrika* 1986;73:751-4.

Supplementary

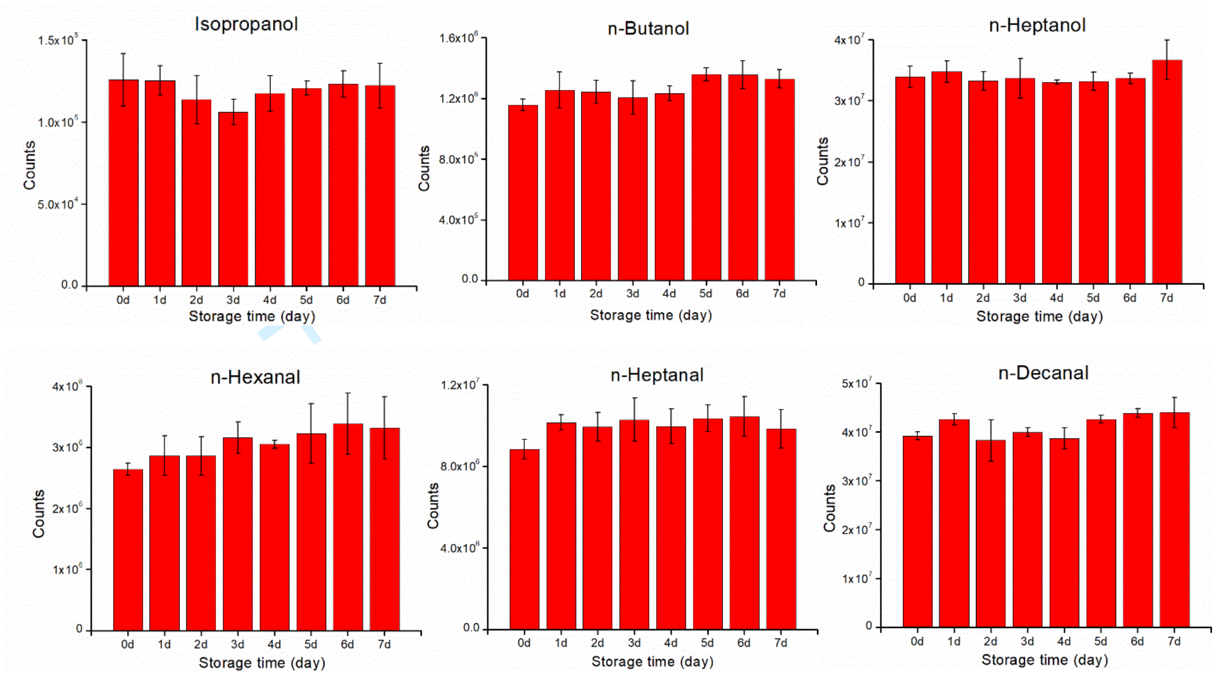

Figure 1. Concentrations of isopropanol, n-butanol, n-heptanol, n-hexanal, n-heptanal, and n-decanal (100 ppbv for each compound) in Tedlar bags at -40 °C are nearly constant within seven days, indicating a good storage stability.

Table1. Requirements for researchers during breath sampling

Requirements	Aims
Pump out air from Tedlar bag fully	Prevent ambient air contamination
Ensure Tedlar bag is tightly sealed	Prevent ambient air contamination
Connect mouthpiece to Bio-VOC sampler tightly	Avoid exhaled gas leakage during expiration
Rotate the three-way valve to an accurate angle	Avoid dead space gas entering the Tedlar bag
Explain to subjects clearly: Do not take a deep breath; Nasal ventilation is not allowed; Blow slowly and entirely.	Guarantee the collection of late expiratory breath; Avoid dead space air contamination.
Adjust the Bio-VOC plunger and the three-way valve immediately	Avoid ambient air entering the Bio-VOC syringe; Avoid sample leakage during transfer from Bio-VOC to Tedlar bag.
Close the intake valve of Tedlar bag quickly and tightly	Avoid exhaled gas leakage; Prevent ambient air contamination.
Place breath sample in a cold container (2-8 °C)	Avoid squeezing of sample bags; Avoid volatilization of exhaled gas.
Clean the Bio-VOC syringe with dry dust-free cloth before the next use	Prevent cross-contamination