Chinese Chemical Letters 36 (2025) 110301

Contents lists available at ScienceDirect

Chinese Chemical Letters

journal homepage: www.elsevier.com/locate/cclet

Tedlar bag free: Accurate volatolomics of IA stage non-small cell lung cancer come out in wash

Bohao Liu[a,g,1], Xue Jiang[b,c,1], Ruizhi Ning[d,1], Heng Zhao[a,g], Yanpeng Zhang[a,g], Junnan Zhang[b,c], Tianqing Liu[d], Danyao Qu[b,c], Yinhui Bao[b,c], Zhanchen Guo[b,c], Xiaoyan Zeng[e], Shan Gao[a], Kun Fan[a], Runyi Tao[a], Jian Ji[f,*], Guangjian Zhang[a,g,*], Weiwei Wu[b,c,d,*]

[a] *Department of Thoracic Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710061, China*
[b] *Shaanxi Key Laboratory of High-Orbits-Electron Materials and Protection Technology for Aerospace, School of Advanced Materials and Nanotechnology, Xidian University, Xi'an 710126, China*
[c] *State Key Laboratory of Electromechanical Integrated Manufacturing of High-performance Electronic Equipments, Xidian University, Xi'an 710126, China*
[d] *Key Laboratory of Artificial Olfaction of Shaanxi Higher Education Institutes, Interdisciplinary Research Center of Smart Sensors, Academy of Advanced Interdisciplinary Research, Xidian University, Xi'an 710126, China*
[e] *Department of Laboratory Medicine, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an 710126, China*
[f] *State Key Laboratory of Food Science and Technology, National Engineering Research Center for Functional Foods, Jiangnan University, Wuxi 214122, China*
[g] *Key Laboratory of Enhanced Recovery After Surgery of Integrated Chinese and Western Medicine, Administration of Traditional Chinese Medicine of Shaanxi Province, Xi'an 710061, China*

ARTICLE INFO

Article history:
Received 31 March 2024
Revised 25 June 2024
Accepted 29 July 2024
Available online 31 July 2024

Keywords:
Volatolomics
Lung cancer
Volatile organic compounds
Exhaled breath analysis
GC-MS

ABSTRACT

Breath analysis can be used to diagnose diseases non-invasively. Accurate measurement of volatolomics is critical for breath analysis to be a gold standard. Tedlar bags (TB) are often used to collect breath samples, but they emit contaminants that affect accuracy. This issue was overlooked in previous studies. We found contamination issues with TB (*e.g.*, siloxanes and aromatic impurities) that affect the identification of volatile organic compounds (VOCs) due to impurities. Then, home-designed equipment (HD) made with poly-tetrafluoride (PTFE) and quartz glass for breath collection was developed and employed in clinical trials. 15 healthy individuals and 32 non-small cell lung cancer (NSCLC) patients at IA stage participated in this study. 610 VOCs can be collected through TB, which is less than HD (1109 VOCs), demonstrating that the inner wall of the TB easily adsorbs VOCs, leading to decreased detection concentrations. Otherwise, utilizing orthogonal partial least squares discriminant analysis (OPLS-DA), we identified chemical markers with significant discriminatory power (VIP $>$ 1.5, $P < 0.05$). The HD method identified 12 target VOCs, surpassing the 3 target VOCs discerned by the TB method. A model combined with a machine learning algorithm for distinguishing early-stage lung cancer patients was established based on biomarkers, which were selected based on OPLS-DA. The results showed strong predictive capabilities for the HD-based model. It indicated that 12 biomarkers derived from the HD model were more effective in distinguishing NSCLC patients, with an AUC value of 0.92, compared to the AUC value of 0.5 from 3 markers obtained from the TB model. The sensitivity and specificity in the confusion matrix reached 100% and 80% for the HD test, but TB test reached only 40% and 60%. This work demonstrated that optimizing and standardizing VOCs collection methodology from breath of lung cancer patients is essential to identify actual volatiles, which could promote disease volatolomics worldwide.

Due to the speedy process and highly specific advantages, non-invasive detection will play a landmark role in futural disease diagnosis and therapeutic monitoring [1,2]. Compared with bio-imaging probes and plasma cell-free deoxyribonucleic acid (DNA) next-generation sequencing, VOCs detection, as an easier explementary and cheaper non-invasive technology, is promising for fast disease diagnosis and screening [3,4]. The core principle of VOCs detection is the identification and measurement of the carbon-containing compound, which is detectable in the gas phase at

* Corresponding authors.
E-mail addresses: jijian@jiangnan.edu.cn (J. Ji), michael8039@xjtu.edu.cn (G. Zhang), wwwu@xidian.edu.cn (W. Wu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cclet.2024.110301
1001-8417/

room temperature, in exhaled breath [5]. More specifically, the construction of a disease-related VOCs profile, developing sensor detection technology, and building detection algorithms are the three key components of the present investigation into the detection and application in the respiratory examination for organic volatiles [4,6]. Among these, the cornerstones for the clinical application of this technology are the collection of VOCs and the establishment of a spectrum library of disease volatiles.

Pauling *et al.* [7] first reported the results of the collection and gas chromatography (GC) analysis of 250 volatiles in respiratory samples in 1971. In 2009, Peng *et al.* [8] reported that 9 VOCs were identified as characteristic targets for lung cancer patients through breath diagnosis. In 2014, Fu *et al.* [9] studied the differences in VOCs between lung cancer patients and healthy people with the aim of identifying two or more VOCs as a "fingerprint" for identifying lung cancer. Tsou *et al.* [10] selected an ion flow tube mass spectrometry technique that was used to quantitatively analyze 116 VOCs in breath samples from lung cancer patients. Kamal *et al.* [11] identified a novel VOC signature comprising decane and other long-chain alkane compounds from healthy subjects experimentally challenged with respiratory viruses (RV). Although effective attempts have been made in the above studies, no unanimous and identifiable VOCs were obtained. The broad variance in collecting and analyzing methods is a major impediment to achieving a standardized profile of VOCs. Previous studies have shown that the contents of the collecting bags were found to be polluted, *N*,*N*-dimethylacetamide and phenol were the most abundant external impurity [12,13]. Polar compounds are more likely to be adsorbed onto the inner wall of Tedlar bags (TB) due to their affinity with the surface of TB, leading to a negative deviation. For TB, compound loss is a combined effect of adsorption and diffusion [14,15].

However, few studies report on the effect of non-standardization of breath collection and testing methods on VOCs identification. Variations in how VOCs are collected, such as the use of TB, can obscure certain essential compounds and potentially lead to inaccurate results, especially, if the VOCs are emitted at low concentration when the disease, such as lung cancer, is at very early (IA) stage [16]. Therefore, the application of exhalatory volatile omics in clinical practice could be limited, due to the procedure from patient preparation to gas collection before detection (oral status, insulation mode, collection mode, collection equipment, detection means, *etc.*) might be polluted. Hanna *et al.* [17] analyzed the challenges and problems of exhaled breath tests based on VOCs for cancer diagnosis and suggested that standardization of breath collection methods is needed among the intended population.

This study attempted for achieving to identify the volatolomics in the breath of NSCLC patients at IA stage by investigating a standardized sampling, collection, and analyzing methodologies for the establishment of a VOCs database, volatolomics. Three types of work were completed based on 47 clinical samples. (1) Establishing a standardized collection and detection procedure by controlling each connector of acquisition and detection based on independently designed sample equipment, *i.e.*, gas collection bottle. (2) Demonstrating the contamination in TB, and its great interferon on the VOCs identification. (3) A model to identify patients with IA lung cancer was constructed, and the feasibility of using the sampling technique in this study for early-stage lung cancer patients' expiration diagnosis was evaluated. Therefore, this work demonstrated the great importance of optimizing and standardizing the VOCs collection operation to screen the more actual volatiles in the breath of lung cancer patients, which may shed a light on helping researchers to uniform the methodologies for promoting the establishment of disease volatolomics all over the world.

All materials are spectrally pure grade and used without further purification. The methods include standards of participants in clinical trials, the procedure of breath sampling, preconcentration of VOCs in sorbent tubes and gas chromatography and mass spectrometry (GC-MS) analysis, chromatographic alignment and feature identification, and statistical analysis of target VOCs. All details will be described in supplementary materials.

47 volunteers were recruited in this study. This study was approved by the Ethics Committee of the First Affiliated Hospital of Xi'an Jiaotong University, approval No: XJTU1AF2021LSK-487, and all participants provided written informed consent. Detailed information such as age and gender, is listed in Table S1 (Supporting information) The produce for breath sampling was given in detail in Supporting information. Fig. S1 (Supporting information) displays a flowchart of the process. The method schema of demonstration on the contamination differences caused by exogenous gases between the TB and HD sampling and analysis is shown in Fig. S2 (Supporting information). Fig. S3 (Supporting information) shows the computer-aided design (CAD) of the equipment used for sampling exhaled breath. The breath of NSCLC patients and healthy people are collected by TB and HD, who tested by GC-MS, and discriminated by the OPLS-DA model to find out the chemical markers and specific VOCs of NSCLC IA stage. By comparing the Q^2 of HD test (0.632) and TB test (0.273), the result showed HD test improved the prediction ability of the model. Finally, choose VIP $>$ 1.5, $P <$ 0.05 compounds, HD test process 12 target VOCs, more than TB test which obtained 3 target VOCs.

To identify the contamination emitted by TB which is a commonly employed sampling container for gas collection, a comparative experiment of GC-MS test on the pure N_2 had been conducted by blowing N_2 through Tenax, TB + Tenax, and HD + Tenax, respectively. Fig. S4 (Supporting information) illustrates that nitrogen exhibits trace impurities, predominantly comprising alkane substances and aromatic compounds. The intrinsic impurities emitted by TB and HD are shown in Fig. 1a. Fig. 1b shows the abundance comparison of the typical 16 volatile contaminations emitted from TB and HD of the highest concentrations. Most of them are hydrocarbons, siloxane compounds, and aromatic compounds. The concentrations of hexamethylcyclotrisiloxane, dodecamethylcyclohexasiloxane, octamethyl cyclotetrasilazane, diphenyl, tetradecamethylcycloheptasiloxane, hexadecamethylcyclooctasiloxane, toluene, mesitylene, butanoic acid butyl ester, dodecane, 3-carene, nonane, camphene, dodecane, octanal, benzene(1,1-dimethylpropyl) from TB are much higher (693.4 times, 594.2 times, 294.9 times, 199.0 times, 117.8 times, 105.4 times, 93.3 times, 40.2 times, 37.0 times, 25.1 times, 14.5 times, 10.2 times, 8.7 times, 8.4 times, 5.7 times, 3.9 times) than those from HD.

More species with higher concentrations meanwhile of VOCs contaminations are identified by GC-MS test, in which the proportions of aromatic compounds, siloxanes, and hydrocarbon are 24.4%, 60.1%, and 5.2% (Fig. 1c), respectively. In contrast, the VOCs contaminations from HD are aromatic (content 77.8%) and hydrocarbons (content 3.8%) (Fig. 1d). The details of the peak area are provided in Tables S2 and S3 (Supporting information).

Therefore, TB obviously emits more impurities compared with HD. In some applications of gas collection, such as environment odor characterization, these intrinsic impurities are neglectable because the collected target gas normally has a much higher concentration ($>$100 ppm). Unfortunately, in the applications of breath analysis, especially in the disease diagnosis at an early stage, the VOCs of ultra-low concentrations (tens of ppb-ppm level) are screened by these impurities, leading to an inaccuracy unreliability in volatolomics.

The total ion chromatogram of exhaled breath sampled by HD and TB are shown in Fig. S5 (Supporting information). 610 kinds of VOCs are identified when the breath is collected by TB. After subtracting the intrinsic impurities, *viz.*, VOCs emitted by

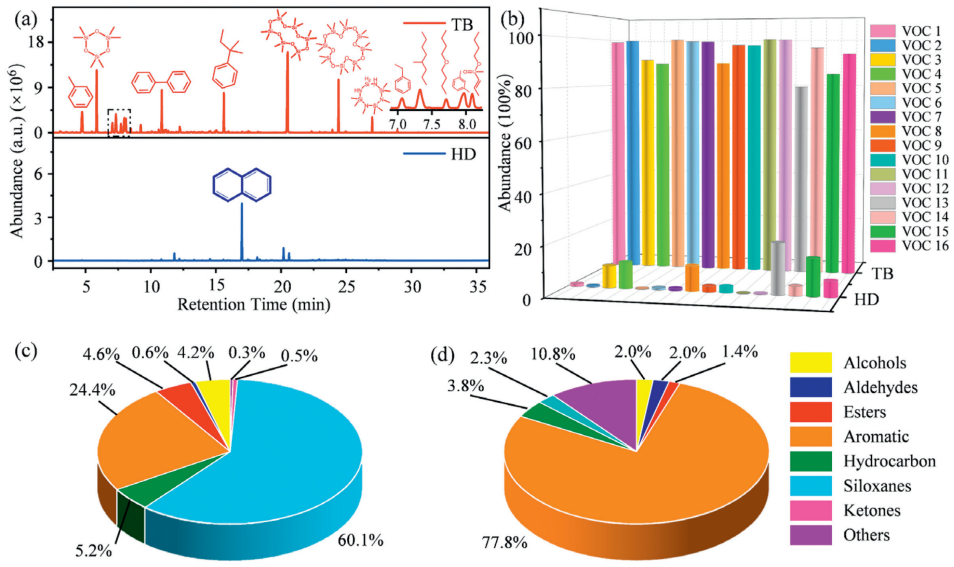

Fig. 1. (a) Total ion chromatogram of blank samples from TB (red line) and HD (blue line). (b) Comparison of impurity gas obtained through TB and HD equipment. VOC1: toluene, VOC2: diphenyl, VOC3: nonane, VOC4: dodecane, VOC5: dodecamethylcyclohexasiloxane, VOC6: tetradecamethylcycloheptasiloxane, VOC7: hexadecamethylcyclooctasiloxane, VOC8: camphene, VOC9: mesitylene, VOC10: butanoic acid butyl ester, VOC11: hexamethylcyclotrisiloxane, VOC12: octamethyl cyclotetrasilazane, VOC13: benzene(1,1-dimethylpropyl), VOC14: decane, VOC15: octanal, VOC16: 3-Carene. Classification pie chart for a blank sample test result: (c) TB equipment and (d) HD equipment.

the TB, including hydrocarbons (caryophyllene (Z-), 1-hexadecene, 3-methylene-7,11,15-trimethyl-1,6,10,14-hexadecatetraene), ketones (4-octylcyclohexanone, 4-octylcyclohexanone, menthenyl ketone), aromatic (1,2-dimethylnaphthalene, *trans*-1,2-diphenylethene, di(2-ethyl hexyl) phthalate), *etc.*, 324 kinds of VOCs with a identification score greater than 70% and mass spectrum matching degree more significant than 70% are selected to be further analyzed. However, 1109 kinds of VOCs are identified when the breath is collected by HD. After subtracting intrinsic impurities, *viz.*, VOCs emitted by HD including hydrocarbons (1-octene, alpha-phellandrene, pentadecane, 1-tetradecene), alcohols (benzyl alcohol, 1-naphthol, nonanal), esters (2-methylbutyl butyrate, ethyl propionate, ethyl ethoxyacetate, isopentyl acetate), 392 kinds of VOCs with an identification score greater than 70% and mass spectrum matching degree greater than 70% are selected for further investigation.

Much fewer kinds and concentrations of VOCs are identified using TB than HD to collect the exhaled breath, which ascribes to more absorption on the inner surface of TB than that of HD. This phenomenon leads to inaccurate concentrations of VOCs with high abundance, even missing information (*i.e.*, missing VOCs with low abundance). It is worth noting that, based on our experience, even though a long-time N_2 flush is done to clean the inner surfaces of TB before reuse, some kinds of the absorbed VOCs canot be totally removed, which interferes with the subsequent experiments. The logic to identify the markers (VOCs) of NSCLC patients is seeking out the main contributors to discriminating the breath of healthy people and patients. Therefore, as shown in Fig. 2a, both in TB and HD tests, the VOCs in the exhaled breath of two groups (healthy people and patients) are collected and analyzed by GC-MS in sequence. After subtracting the patterns of intrinsic impurities, the GC-MS patterns of healthy people and patients are obtained. The heat map (Figs. 2b and d) is generated by hierarchical Pearson, which displays the top 25 VOCs with significant differences (Colors from red to blue represent the relative content of the VOCs in the corresponding exhaled sample from high to low). As shown in Fig. 2b, there was a significant difference in exhaled breath between the healthy and patient groups, which was sampled by HD. On the contrary, exhaled gas collected through TB showed no significant difference between the two groups. To identify target compounds in the exhaled breath of lung cancer patients, OPLS-DA calculations were performed. Figs. 2c and e are the plots of discrimination according to the HD and TB tests respectively. A much better separation is achieved through HD, than TB test. The data of individuals breath through HD had a better fitting factor ($R^2Y = 0.995$) compared to $R^2Y = 0.455$ in TB test. The larger R^2Y indicates a great difference between two Q^2 indicates that the better prediction ability of this model. Generally, it is considered the model exhibited a better ability to groups, which demonstrates that the model established based on the HD test has a better separation ability. In addition, the larger predict when $Q^2 > 0.5$. The prediction effect ($Q^2 = 0.632$) in HD test is higher than that of TB test ($Q^2 = 0.273$), which means the better prediction ability of the model in the HD test, and higher reliability. To further demonstrate the reliability of the model, 900 times permutation tests are carried out. The result for 900 permutation tests is shown in Figs. S6a and b (Supporting information). The R^2 and Q^2 values were 0.93 and -0.364 for HD test, compared to $R^2 = 0.249$ and $Q^2 = -0.179$ for TB test.

Both two models satisfy $Q^2 < 0.05$, although the blue regression line of the Q^2 points is based on two models, which intersect the vertical axis (on the left) below zero. But it is important to note that not all Q^2 values to the left are lower than the original points to the right of the Q^2 points in the TB test, but model based on HD test satisfied the condition. This phenomenon indicated that the model based on HD was highly robust and there was no overfitting [18]. The S-Plots (Figs. S6c and d in Supporting information) also demonstrate that the HD model can distinguish more differential VOCs.

Based on the OPLS-DA analysis, to find the differential VOCs, choose compounds that meet the conditions: VIP > 1.5, The *P*-value (obtained from the Mann-Whitney U test) < 0.05. The differential VOCs were qualitatively determined and are shown in Tables S4 and S5 (Supporting information).

*A*1 represents the average chromatographic peak area of breath VOCs in patients, while *A*2 represents breath VOCs in healthy samples. Fold change (FC) is defined as follows:

$$\text{Foldchange (FC)} = A1/A2. \quad (1)$$

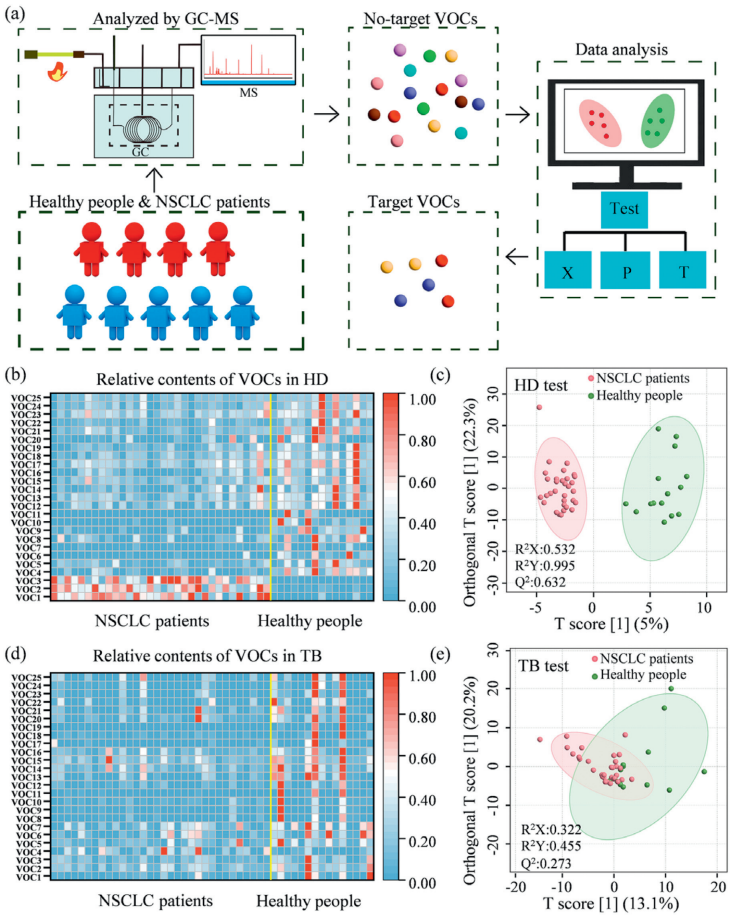

Fig. 2. (a) Schematic diagram of the experimental process. (b) Heatmap for relative contents of VOCs in HD test. VOC1: allyl chloride, VOC2: *p*-cresol, VOC3: (1*S*-*trans*)-2-methyl-5-(1-methylethenyl)-2-cyclohexen-1-ol, VOC4: nonadecanoic acid methyl ester, VOC5: 1,1-diallylhydrazine, VOC6: octadecanoic acid, VOC7: 2,6-bis(1,1-dimethylethyl)-4-methylphenol, VOC8: *tert*-butyldimethylsiloxybutane, VOC9: trichloroethylene, VOC10: beta-myrcene, VOC11: acetone, VOC12: diphenyl, VOC13: 2,6-dimethylnaphthalene, VOC14: acenaphthylene, VOC15: 3-methyl-2,5-piperadinedione, VOC16: alpha-butylcinnamaldehyde, VOC17: tetradecane, VOC18: cyanoalanine, VOC19: 4-ethylphenyl methyl ketone, VOC20: 1,5-dihydroxy-4,6,8-trimethylazulene, VOC21: 3-ethyl-2,6,10-trimethylundecane, VOC22: octyl butyrate, VOC23: (*Z*)-9-heptacosene, VOC24: 2-methyl-nonadecane, VOC25: tricosane. (c) OPLS-DA score plots for HD test. (d) Heatmap for relative contents of VOCs in TB test. VOC1: hexamethylethane, VOC2: 1,2-dichloropropane, VOC3: 2-phenylethyl acetate, VOC4: acrylic acid 2-ethylhexyl ester, VOC5: ethyl acetate, VOC6: *ortho* xylene, VOC7: 3-methylindole, VOC8: dichloromethane, VOC9: 2-methylpentane, VOC10: heptyl formate, VOC11: hexane, VOC12: 1,2,4,5-tetramethylbenzene, VOC13: acetone, VOC14: alpha-phellandrene, VOC15: 1,3-diethylbenzene, VOC16: 3-heptanone, VOC17: 1-octanol, VOC18: isoflavone, VOC19: benzene, VOC20: butyl butyrate, VOC21: *p*-cymene, VOC22: heptadecane, VOC23: limonene, VOC24: isopropyl propionate, VOC25: anethole. (e) OPLS-DA score plots for TB test. R^2X and R^2Y, the interpretation rate of the model; Q^2, the predictive ability of the model.

As shown in Table S4 (Supporting information), we extracted 12 target VOCs from the HD experiment. Among these, compared to healthy individuals, 3 VOCs presented the highest fold change value in the exhaled gas of patients. Those compounds were *p*-cresol (12.92 times), allyl chloride (14.33 times), and (1*S*-*trans*)-2-methyl-5-(1-methylethenyl)-2-cyclohexen-1-ol (30.66 times). For the other 9 VOCs, compared to healthy individuals, a decrease in the exhaled gas of patients, except for acetone (0.03 times) and beta-myrcene (0.08 times), which was reduced obviously. The remaining 7 VOCs, including lipids (nonadecanoic acidmethyl ester), including lipids (nonadecanoic acidmethyl ester), acids (octadecanoic acid), hydrocarbons (2,6-dimethylnaphthalene, 3-ethyl-2,6,10-trimethylundecane, 1,1-diallylhydrazine, trichloroethylene) and aromatic (diphenyl), there were showed decreases of less than 1/10. The result is shown in Fig. 3. Compared with the healthy individuals, *p*-cresol, allyl chloride, and (1*S*-*trans*)-2-methyl-5-(1-methylethenyl)-2-cyclohexen-1-ol were significantly increased in IA patients. Conversely, octadecanoic acid, beta-myrcene and acetone were significantly decreased in IA patients, and those 3 targets showed a relatively concentrated distribution.

To further demonstrate whether the sampling method affected the selection of target VOCs, the top 25 VOCs with significant differences for heat map analysis were selected. Fig. S7 (Supporting information) shown that the concentration of (1*S*-*trans*)-2-methyl-5-(1-methylethenyl)-2-cyclohexen-1-ol and nonadecanoic acid methyl ester decreased when collected exhaled breath using TB compared to HD. This indicates that the concentration of potential target VOCs may change due to adsorption to the inner wall of the TB, which is used as a collection apparatus.

Previous research has found that changes in some compounds are associated with human health. Beta-myrcene exhibits antitumor activity against lung cancer cells by inducing oxidative stress and cell apoptosis mechanisms [19]. Consistent with previous studies, we found reduced concentrations of this compound in exhaled breath of lung cancer patients. *p*-Cresol is a carcinogenic compound that can be metabolized in the human body and produce carcinogenic effects [20]. It is consistent with the findings in this work that concentrations of *p*-cresol are elevated in exhaled breath of lung cancer patients. Previous studies have found that the concentration of acetone in lung cancer exhaled breath is reduced [21]. Our findings are consistent with predecessors, but the specific reasons for this phenomenon need to be further studied. Only 3 VOCs exhibited significant differences between healthy and patient groups in TB test through the same procedure as the HD

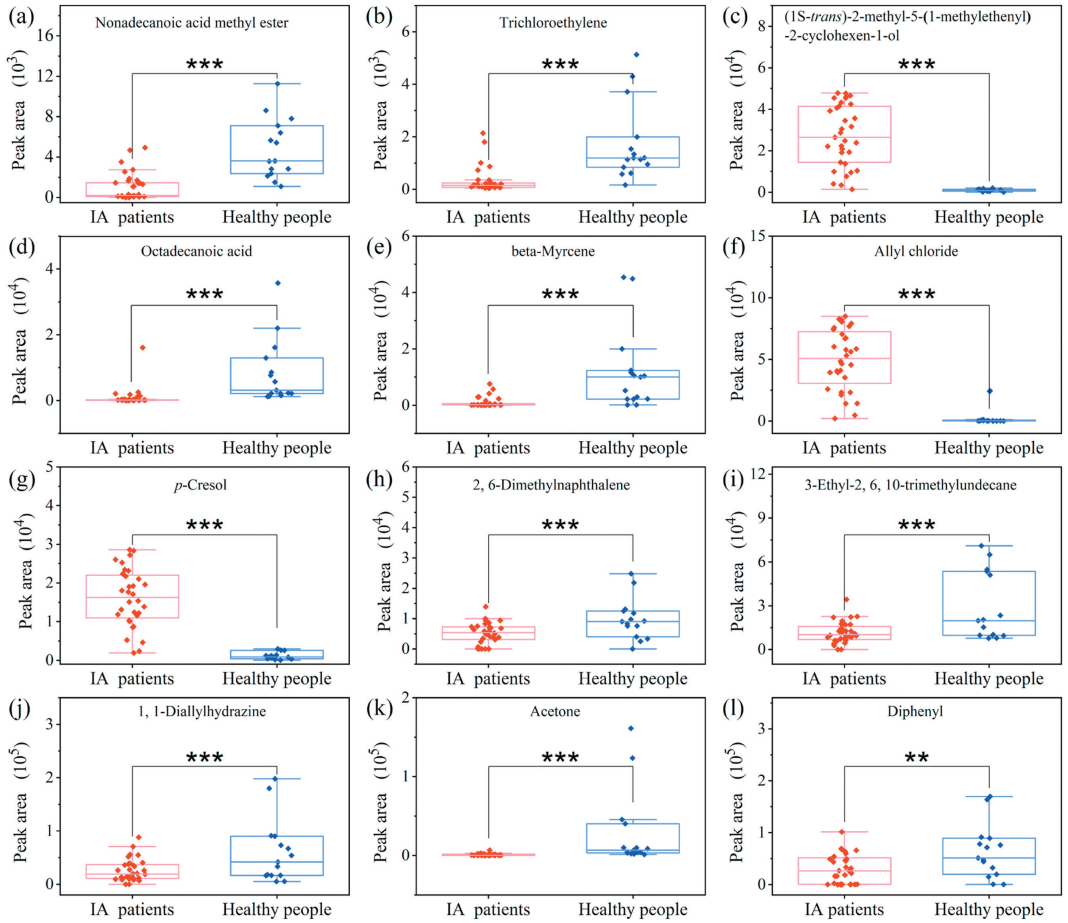

Fig. 3. Box plots about the differential VOCs in exhaled breath from lung cancer and healthy people sampled by HD test (FC > 2.0, VIP > 1.5, $P < 0.05$): (a) nonadecanoic acid methyl ester, (b) trichloroethylene, (c) (1*S*-*trans*)-2-methyl-5-(methylethenyl)-2-cyclohexen-1-ol, (d) octadecanoic acid, (e) beta-myrcence, (f) allyl chloride, (g) *p*-cresol, (h) 2,6-dimethynaphthalene, (i) 3-ethyl-2,6,10-trimethylundecane, (j) 1,1-diallyhydrazine, (k) acetone, (l) diphenyl.

test. The detailed information is provided in Table S5 (Supporting information).

Compared to healthy individuals, these VOCs demonstrated a decrease in the patient group. The concentration of acetone, alpha-phellandrene and limonene in the patient group decreased to 0.26 times, 0.45 times, and 0.27 times in the healthy group. The changes in concentration of those VOCs between the two groups were visualized using box plots (Fig. S8 in Supporting information), which depicted the concentration distribution of the sample inadequately.

In addition to acetone mentioned above, alpha-phellandrene and limonene have been found to be associated with lung cancer. alpha-Phellandrene has been reported to have biological capabilities for anti-microbial, anti-inflammatory, and anti-cancer [22]. Meanwhile, limonene has been reported to exhibit antitumor activity [23,24]. This study demonstrates that the concentration of these compounds decreased in the exhaled breath of lung cancer patients. Consequently, there appears to be a connection between the exhaled breath and the health condition. These results indicate that the HD and standardized sampling method can reduce exogenous contamination, achieving more accurate identification of target compounds.

In order to evaluate diagnostic test performance, the ROC curve was analyzed, with AUC representing the area under the ROC curve [25,26]. A larger AUC value indicates better predictive ability of the model. In Fig. S9a (Supporting information), the ROC curve demonstrates the better predicted performance of 12 biomarkers which selected from the HD model, showing an AUC value of 0.92. On the other hand, Fig. S9b (Supporting information) indicated that the AUC value of 3 markers which selected from TB model was only 0.5. It demonstrated that the model established based on TB method has poor predictive ability. The sensitivity and specificity in the confusion matrix reached 100% and 80% for the HD test (Fig. S9c in Supporting information), which are much more precise than TB test. The sensitivity and specificity in the confusion matrix reached 40% and 60% for the TB test (Fig. S9d in Supporting information). Those results confirmed the successful optimization of our method for sampled exhaled breath.

It is worthy to additional note that diagnosing early lung cancer through exhaled breath has yielded varying conclusions across different study methodologies, *i.e.*, the influence of contaminations should be minimized as much as possible to enhance the accuracy. Several factors contribute to this inconsistency. (1) Varied environmental conditions lead to variations in the volatolomics products in the exhaled breath. (2) Diverse food cultures may alter the composition and concentration of VOCs in exhaled breath. (3) Different stages of lung cancer: may result in variations in exhaled breath components. (4) The number of samples: it will affect the sample representativeness and influence the result of the model. (5) Variations in sampling methods: It will mask or reduce the concentration of target compounds when using an improper sampling method which introduces external impurities in the samples. Therefore, besides screening out that interferon, we need to first establish a standard of operation for letting the real volatolomics to come out in the wash by screening volatile contaminations in the collection container. Environmental and dietary differences can be minimized when the experimental group and the control group are

in the same background. To avoid the issue (3), focusing on specific types of lung cancer can enhance the accuracy of the studies. Regarding issue (4), collecting a more significant number of samples can improve the reliability of the results. Crucially, as for issue (5), our experiment concluded that different sampling methods have a significant impact on test results, particularly in reducing the accuracy of detection when introducing more external impurities during the sampling process.

A standard of operation in collection of breath diagnosis is established, through which the volatolomics of NSCLC at IA stage is identified. Two methods of TB (normally used) and HD (our new method) for breath collection were verified and systematically compared towards 47 clinical trials (15 healthy individuals and 32 patients). The TB releases more exogenous gas, which leads less target VOCs (610 VOCs in TB) being detected than those of 1109 VOCs in HD after blank results (absorbed VOCs on the inner surface and released VOCs by TB itself) calibration and deduction. Subsequently, by comparing the OPLS-DA model established by two tests ($R^2Y = 0.995$, $Q^2 = 0.632$ from HD, and $R^2Y = 0.455$, $Q^2 = 0.273$ for TB model), the model established by HD test exhibits higher predictive ability and higher reliability. The result for 900 permutation tests indicated that the model based on HD was highly robust and not overfit. Additionally, select components to satisfy VIP > 1.5, $P < 0.05$, HD test identified 12 target VOCs, while the TB test only identified 3 target VOCs. This result showcases a strong predictive capability of the model utilizing HD. It indicated that 12 biomarkers (obtained from the HD model) were able to distinguish NSCLC patients better, with the AUC value of 0.92, compared to the value of AUC 0.5 by 3 markers (obtained from the TB model). Meanwhile, the sensitivity and specificity in the confusion matrix reached 100% and 80% for the HD test and TB test reached only 40% and 60%. Those results confirmed the successful optimization and demonstrated the following advantages: (1) Reduce exogenous contamination, achieving more accurate identification of target compounds, (2) the inner wall of the TB easily adsorbs VOCs while HD can significantly reduce this situation and increase the concentration of VOCs in the Tenax tube. Therefore, the devices and method in HD are demonstrated to identify the volatolomics of NSCLC at IA stage, which provides valuable insights aiming to establish a general methodology for studying disease volatolomics globally and advance progress in this field.

Ethical statement

The clinical trials in this study are approved by Ethics Committee of the First Affiliation Hospital of Xi'an Jiaotong University (No. XJTUIAF2021LSK-487).

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Bohao Liu: Writing – original draft, Resources, Methodology, Formal analysis, Data curation. **Xue Jiang:** Writing – original draft, Visualization, Resources, Methodology, Investigation, Formal analysis, Data curation. **Ruizhi Ning:** Writing – review & editing, Validation, Data curation, Conceptualization. **Heng Zhao:** Methodology, Data curation. **Yanpeng Zhang:** Project administration, Conceptualization. **Junnan Zhang:** Visualization, Validation. **Tianqing Liu:** Writing – review & editing, Validation. **Danyao Qu:** Validation. **Yinhui Bao:** Validation. **Zhanchen Guo:** Validation, Supervision. **Xiaoyan Zeng:** Formal analysis. **Shan Gao:** Project administration. **Kun Fan:** Project administration. **Runyi Tao:** Resources. **Jian Ji:** Writing – review & editing, Visualization, Supervision, Software, Project administration, Methodology. **Guangjian Zhang:** Writing – review & editing, Supervision, Funding acquisition, Formal analysis, Conceptualization. **Weiwei Wu:** Writing – review & editing, Visualization, Supervision, Project administration, Formal analysis, Conceptualization.

Acknowledgments

This study was financially supported by the Key Research and Development Program of Shaanxi (No. 2022ZDLSF01-04), National Natural Science Foundation of China (No. 22104117), The First Affiliated Hospital of Xi'an Jiaotong University's Medical "Basic-Clinical" Integration Innovation (No. YXJLRH2022033). We sincerely thank the technical support from the engineers (Mr. Yaopeng Zhu, Mr. Changyuan Wang, Mr. Qian Sun) of Shimazu company.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cclet.2024.110301.

References

[1] S. Bratulic, A. Limeta, S. Dabestani, Proc. Natl. Acad. Sci. U. S. A. 119 (2022) e2115328119.
[2] G. Pacchioni, Nat. Rev. Mater. 7 (2022) 679.
[3] M. Phillips, N. Altorki, J.H.M. Austin, J. Clin. Oncol. 23 (2005) 9510.
[4] G. Peng, M. Hakim, Y.Y. Broza, Br. J. Cancer 103 (2010) 542–551.
[5] M. Hakim, Y.Y. Broza, O. Barash, Chem. Rev. 112 (2012) 5949–5966.
[6] N. Kahn, O. Lavie, M. Paz, Nano Lett. 15 (2015) 7023–7028.
[7] L. Pauling, A.B. Robinson, R. Teranishi, Proc. Natl. Acad. Sci. U. S. A. 68 (1971) 2374–2376.
[8] G. Peng, U. Tisch, O. Adams, Nat. Nanotechnol. 4 (2009) 669–673.
[9] X.A. Fu, M.X. Li, R.J. Knipp, Cancer Med. 3 (2014) 174–181.
[10] P.H. Tsou, Z.L. Lin, Y.C. Pan, Cancers 13 (2021) 1431.
[11] F. Kamal, S. Kumar, M.R. Edwards, Am. J. Respir. Crit. Care Med. 204 (2021) 1075–1085.
[12] M.M.L. Steeghs, S.M. Cristescu, F.J.M. Harren, Physiol. Meas. 28 (2007) 73–84.
[13] D. Zanella, A. Henin, S. Mascrez, J. Sep. Sci. 45 (2022) 3542–3555.
[14] P.L. Kasper, A. Oxbøl, M.J. Hansen, J. Environ. Qual. 47 (2018) 246–253.
[15] S. Ghosh, K.-H. Kim, J.R. Sohn, Sci. World J. 11 (2011) 2160–2177.
[16] O. Lawal, W.M. Ahmed, T.M.E. Nijsen, Metabolomics 13 (2017) 110.
[17] G.B. Hanna, P.R. Boshier, S.R. Markar, JAMA Oncol. 5 (2019) 1070.
[18] I. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, Multi- and megavariate data analysis. Principles application, Umetrics Academy, Umeå, 2001.
[19] L. Pinciger, F. Valenti, C. Bergamini, Molecules 28 (2023) 238–264.
[20] P. Vernocchi, T. Gili, F. Conte, Int. J. Mol. Sci. 21 (2020) 8730.
[21] A. Bajtarevic, C. Ager, M. Pienz, BMC Cancer 9 (2009) 348.
[22] S. Thangaleela, B.S. Sivamaruthi, P. Kesika, Sci. Pharm. 90 (2022) 57.
[23] X. Yu, H. Lin, Y. Wang, Onco Targets Ther. 11 (2018) 1833–1847.
[24] F. Gahleitner, C. Guallar-Hoyas, C.S. Beardsmore, Bioanalysis 5 (2013) 2239–2247.
[25] R. Kumar, A. Indrayan, Indian Pediatr. 48 (2011) 277–287.
[26] A.P. Bradley, Pattern Recognit. 30 (1997) 1145–1159.