

Methods and Applications

Targeted Analysis of VOCs in Exhaled Breath of Coal Workers' Pneumoconiosis Patients, An Exploratory Study

Rui Gao^{1,2,3}; Lijia Yu¹; Qingyu Meng¹; Yazhen He^{1,2,3}; Fangda Peng^{1,2}; Xiangfu Cao¹; Qingjun Qian¹; Qikai Liu¹; Gaisheng Liu⁴; Fan Yang⁴; Yiwei Shi⁵; Qingsong Chen³; Chunguang Ding^{1,2,3,5,#}

ABSTRACT

Introduction: Pneumoconiosis represents the most prevalent occupational disease in China, with coal workers' pneumoconiosis (CWP) showing the highest incidence. Analysis of volatile organic compounds (VOCs) in the exhaled breath of CWP patients may provide novel insights into its pathogenesis.

Methods: Study data were collected through questionnaires and medical examinations. Thermal desorption-gas chromatography-mass spectrometry was employed for targeted VOC analysis. Differential VOCs were identified using OPLS-DA, the Mann-Whitney U test, and fold change analysis. The discriminatory efficacy of differential VOCs was evaluated using receiver operating characteristic (ROC) curves. Spearman correlation analysis explored relationships between differential VOCs, lung function indices, and blood cell levels.

Results: The pneumoconiosis group showed elevated concentrations of 10 compounds, including isopentane, n-pentane, and isoprene, while four compounds, including 2,4-dimethylpentane, methylcyclohexane, 2,3,4-trimethylpentane, and 2-methylheptane showed decreased concentrations. Combined univariate and multivariate statistical analyses identified six significant VOCs, including isopentane and pentane. Notably, isopentane and n-pentane demonstrated negative correlations with forced vital capacity and levels, while 2-methylheptane showed positive correlations.

Discussion: Clear metabolic differences in VOCs exist between CWP patients and non-dust-exposed healthy controls. Six compounds — isopentane, n-pentane, 3-methylpentane, n-hexane, cyclohexane, and 2-methylheptane — in exhaled breath demonstrate potential as biomarkers for CWP.

Coal Workers' Pneumoconiosis (CWP) is a chronic occupational lung disease that develops following prolonged exposure to mixed coal dust. In China, approximately 15,000 new CWP cases are reported annually. By the end of 2021, the cumulative number of reported occupational pneumoconiosis cases in China reached 915,000, with CWP accounting for 43% of all pneumoconiosis cases. CWP remains a significant public health concern in China, with prevention and treatment challenges persisting. Early screening and diagnosis of CWP represent critical unmet needs in occupational medicine.

The analysis of volatile organic compounds (VOCs) in exhaled breath has emerged as a promising non-invasive diagnostic approach for clinical disease surveillance. Studies have demonstrated that exhaled breath VOCs reflect diverse metabolic processes within the body, with inflammation, oxidative stress, and immune responses significantly influencing both VOC composition and concentration (1). VOCs can traverse the air-blood barrier to enter the alveoli, suggesting that exhaled breath VOCs may serve as indicators of systemic metabolic changes, particularly within pulmonary tissue (2). This approach has shown promise in biomarker studies for various respiratory conditions, including chronic obstructive pulmonary disease (3), idiopathic pulmonary fibrosis (4), and lung cancer (5). However, research on exhaled breath VOCs in pneumoconiosis remains limited, with previous studies focusing primarily on non-targeted analyses. Notably, no studies have conducted targeted quantitative analyses of exhaled breath from coal workers' pneumoconiosis patients.

Therefore, our study employed thermal desorption-gas chromatography-mass spectrometry (TD-GC-MS) to conduct targeted quantitative analysis of 27 lung disease-associated VOCs in the exhaled breath (alveolar air) of CWP patients. Our objectives were to identify characteristic VOCs for CWP and explore potential biomarkers. We also evaluated the diagnostic potential

of these biomarkers and analyzed their correlations with lung function parameters and blood cell indices.

METHODS

Study Subjects

This study recruited 120 volunteers, of whom 65 met the inclusion and exclusion criteria, with ages ranging from 18 to 80 years. Subjects were excluded if they had: (a) autoimmune diseases, diabetes, cancers, asthma, or other chronic inflammatory conditions; (b) undergone lung lavage or any form of lung surgery; or (c) experienced respiratory infections or taken anti-inflammatory or antibiotic medications within 1 week prior to sampling. Demographic information, including smoking history, alcohol consumption, and previous medical history, was collected for all subjects. Lung function and routine blood test results from physical examinations were obtained, with pulmonary function data calculated as percentages of measured values relative to predicted values. All participants provided written informed consent. This study received approval from the Ethics Committee of the First Hospital of Shanxi Medical University (2020-kk104).

Sample Collection and Detection

Subjects were instructed to avoid consuming foods with strong odors the day before exhaled breath collection and to fast after 22:00 (including abstaining from smoking and drinking). Sample collection was conducted in a clean, ventilated room maintained at 25 °C (± 2 °C) before breakfast. Exhaled breath samples were collected using a Bio-VOC sampler equipped with a one-way valve to capture alveolar air. All thermal desorption (TD) tubes underwent heat pretreatment prior to sampling. During collection, subjects rested seated for 10 minutes before exhaling into a Bio-VOC syringe with a disposable mouthpiece. The collected gas was then transferred to the TD tube, with the process repeated twice to obtain approximately 250 mL of exhaled breath. Two environmental blank samples were collected simultaneously as controls. The thermal desorption tubes were subsequently transported to the laboratory for analysis.

The instrumental methodology has been previously described (6). Detailed methodological parameters are provided in [Supplementary Table S1](#) (available at <https://weekly.chinacdc.cn/>).

Statistical Analysis

Statistical analyses were performed using SPSS 25.0 (IBM, Armonk, New York, USA). The Mann-Whitney U test was employed to detect significant variations in metabolites between groups, with a significance threshold of 0.05. Compounds with $P < 0.05$ were incorporated into the lasso regression model using age as a covariate. For compounds below the limit of detection (LOD), concentrations were substituted with $\text{LOD}/\sqrt{2}$.

Multivariate statistical analyses were conducted using SIMCA 14.1 (Umetrics, Uppsala, Sweden). Initial group differentiation was performed using Principal Component Analysis (PCA) and Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) for visual clustering. Differential VOCs were identified using multiple criteria: variable important in projection (VIP) > 1 , univariate statistical significance ($P < 0.05$), and fold change (FC) > 2 or < 0.5 . Age was included as a covariate, and compounds with univariate statistical significance ($P < 0.05$) were incorporated into the Lasso regression model for adjustment. The discriminatory power of screened VOCs was evaluated through receiver operating characteristic (ROC) analysis, with classification ability assessed by area under curve (AUC). Additionally, relationships between variables were examined using Spearman correlation analysis.

Quality Control

Prior to data collection, surveyors underwent rigorous training and assessment to ensure standardization of collection protocols. Sample analyses strictly adhered to established experimental procedures and operational protocols. During data processing and analysis, regular quality checks were performed to identify and remove invalid, duplicate, or inconsistent data entries.

RESULTS

Characteristics of Subjects

The study included 65 participants, comprising 42 CWP patients and 23 healthy controls. Statistical analysis revealed no significant differences between the groups regarding BMI, smoking status, alcohol consumption, and blood indices ([Table 1](#)). Analysis of VOC content between groups ([Table 2](#)) demonstrated that the CWP group exhibited elevated levels of isopentane, n-pentane, isoprene, 2-methylpentane, 3-methylpentane, 1-hexene, n-hexane, methyl

TABLE 1. Characteristics of subjects.

Variables	CWP (n=42)	Control (n=23)	P
Age (years)	67.6±4.8	47.5±5.7	<0.05
BMI (kg/m ²)	24.8±2.9	26.3±2.5	0.510
Smoking (yes, %)	21 (50.0%)	10 (43.5%)	0.796
Drinking (yes, %)	23 (56.1%)	13 (56.5%)	1.000
FVC (predicted %)	67.083±19.411	85.500±10.117	<0.05
FEV1.0 (predicted %)	79.043±24.699	115.818±11.722	<0.05
WBC (10 ⁹ /L)	6.384±1.306	6.571±1.492	0.602
NEU (10 ⁹ /L)	4.841±6.723	3.657±0.963	0.406
LYM (10 ⁹ /L)	2.059±0.739	2.296±0.755	0.225
NLR	1.948 (1.432, 2.526)	1.559 (1.131, 2.078)	0.072
PLR	107.797 (86.118, 134.303)	116.466 (105.208, 135.795)	0.278

Abbreviation: BMI=body mass index; FVC=forced vital capacity; FEV1.0=forced expiratory volume in one second; WBC=white blood cell; NEU=neutrophil; LYM=lymphocyte; NLR=neutrophil to lymphocyte ratio; PLR=platelet to lymphocyte ratio; CWP=coal workers' pneumoconiosis.

TABLE 2. Quantitative values of VOCs in each group [Median (25th,75th percentile) ppb].

VOC	CWP (n=42)	Control (n=23)
Iso-pentane*	1.93 (1.004, 3.910)	0.135 (0.069, 0.302)
1-Pentene	0.152 (0.066, 0.541)	0.17 (0.127, 0.268)
Pentane*	2.186 (1.325, 3.373)	0.072 (0.020, 0.180)
Trans-2-Pentene	0.063 (0.063, 0.545)	0.063 (0.063, 0.141)
Iso-prene*	145.904 (105.775, 188.798)	97.534 (63.288, 105.763)
2,2-Dimethylbutane	13.513 (7.003, 20.313)	7.773 (6.345, 22.390)
2-Methylpentane*	0.152 (0.063, 1.724)	0.063 (0.063, 0.133)
3-Methylpentane*	0.511 (0.135, 2.756)	0.091 (0.012, 0.179)
1-Hexene*	0.135 (0.078, 1.145)	0.078 (0.078, 0.254)
Hexane*	1.104 (0.613, 5.399)	0.570 (0.014, 0.900)
2,4-Dimethylpentane*	0.514 (0.176, 1.023)	1.619 (0.902, 2.187)
Methyl-cyclopentane*	0.250 (0.104, 1.317)	0.105 (0.005, 0.235)
2-Methylhexane	0.051 (0.051, 0.070)	0.061 (0.051, 0.086)
Cyclohexane*	0.178 (0.005, 1.155)	0.005 (0.005, 0.025)
2,3-Dimethylpentane	0.120 (0.012, 0.273)	0.043 (0.012, 0.156)
3-Methylhexane	0.045 (0.045, 0.096)	0.196 (0.045, 0.309)
2,2,4-Trimethylpentane	0.017 (0.017, 0.199)	0.017 (0.017, 0.030)
Heptane	0.190 (0.129, 0.538)	0.283 (0.166, 1.324)
Methylcyclohexane*	0.015 (0.005, 0.093)	0.088 (0.032, 0.135)
Pentanal	0.381 (0.185, 1.195)	0.377 (0.181, 0.748)
2,3,4-Trimethylpentane*	0.013 (0.013, 0.013)	0.022 (0.017, 0.027)
2-Methylheptane*	0.039 (0.039, 0.039)	0.343 (0.320, 0.442)
3-Methylheptane	0.027 (0.027, 0.027)	0.027 (0.027, 0.027)
4-Methyl-2-pentanone	0.003 (0.003, 3.278)	0.003 (0.003, 0.534)
2-Hexanone*	0.013 (0.013, 0.045)	0.013 (0.013, 0.013)
Hexanal	0.761 (0.369, 1.739)	0.726 (0.517, 0.990)
Decane	0.065 (0.050, 0.187)	0.050 (0.050, 0.072)

Abbreviation: VOC=volatile organic compounds; CWP=coal workers' pneumoconiosis.

* P<0.05.

cyclopentane, cyclohexane, and 2-hexanone in exhaled breath compared to controls. Conversely, the CWP group showed decreased levels of 2,4-dimethylpentane, methylcyclohexane, 2,3,4-trimethylpentane, and 2-methylheptane.

Metabolic Profile Analysis

Principal component analysis (PCA), an unsupervised dimensionality reduction technique, was initially employed to visualize the overall data distribution across all samples. The analysis revealed distinct clustering between the pneumoconiosis and control groups based on exhaled VOCs profiles, though with some overlap (Figure 1A). Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA), which excels at group discrimination by removing disease-irrelevant information while emphasizing disease-relevant features, demonstrated

superior group separation ($R^2Y=0.867$, $Q^2=0.826$) as shown in Figure 1B. The model's robustness was validated through 200 permutation tests (Figure 1C), yielding $R^2=(0.0, 0.102)$ and $Q^2=(0.0, -0.464)$, confirming both its validity and reliability.

Screening for Differential VOCs

Differential VOCs were identified using a multi-criteria approach combining VIP values, FC, and statistical significance. The screening criteria included $VIP>1$, $FC>2$ or <0.5 , and $P<0.05$ (Supplementary Table S2, available at <https://weekly.chinacdc.cn/>). This analysis identified six differential VOCs: isopentane, n-pentane, 3-methylpentane, n-hexane, cyclohexane, and 2-methylheptane. Notably, 2-methylheptane maintained statistical significance ($P<0.05$) after age adjustment in the LASSO regression model.

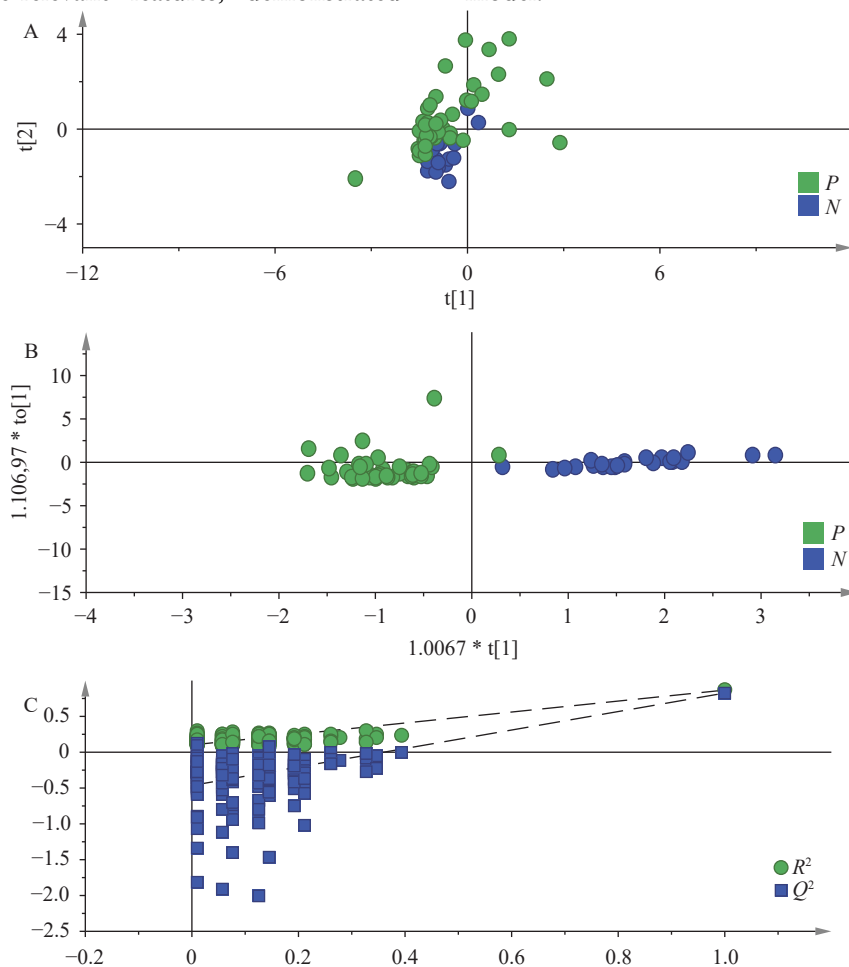


FIGURE 1. Multivariate analysis model. (A) The mode of PCA between coal workers' pneumoconiosis and control group; (B) The mode of OPLS-DA between coal workers' pneumoconiosis and control group; (C) permutation tests of OPLS-DA models for coal workers' pneumoconiosis and control group. Abbreviation: PCA=principal component analysis; OPLS-DA=orthogonal partial least squares-discriminant analysis.

The diagnostic potential of each differential VOC for CWP was assessed using ROC curves. The AUC values were: isopentane 0.940 [95% confidence interval (CI): 0.880, 1.000], n-pentane 0.996 (95% CI: 0.986, 1.000), 3-methylpentane 0.776 (95% CI: 0.663, 0.889), n-hexane 0.753 (95% CI: 0.634, 0.872), cyclohexane 0.777 (95% CI: 0.667, 0.888), and 2-methylheptane 0.948 (95% CI: 0.878, 1.000). The complete ROC curves and detailed analysis results are presented in [Supplementary Figure S1](#) and [Supplementary Table S3](#) (available at <https://weekly.chinacdc.cn/>), respectively.

Correlation Between Differential VOCs and Clinical Parameters

Correlation analysis revealed significant associations between the differential VOCs and pulmonary function parameters. Specifically, isopentane and n-pentane demonstrated negative correlations with both FVC (predicted %) and FEV1.0 (predicted %), while 2-methylheptane showed positive correlations with these same parameters ([Supplementary Table S4](#), available at <https://weekly.chinacdc.cn/>). The relationships between isopentane, n-pentane, and 2-methylheptane versus FVC (predicted %) and FEV1.0 (predicted %) are visually represented in [Figure 2](#).

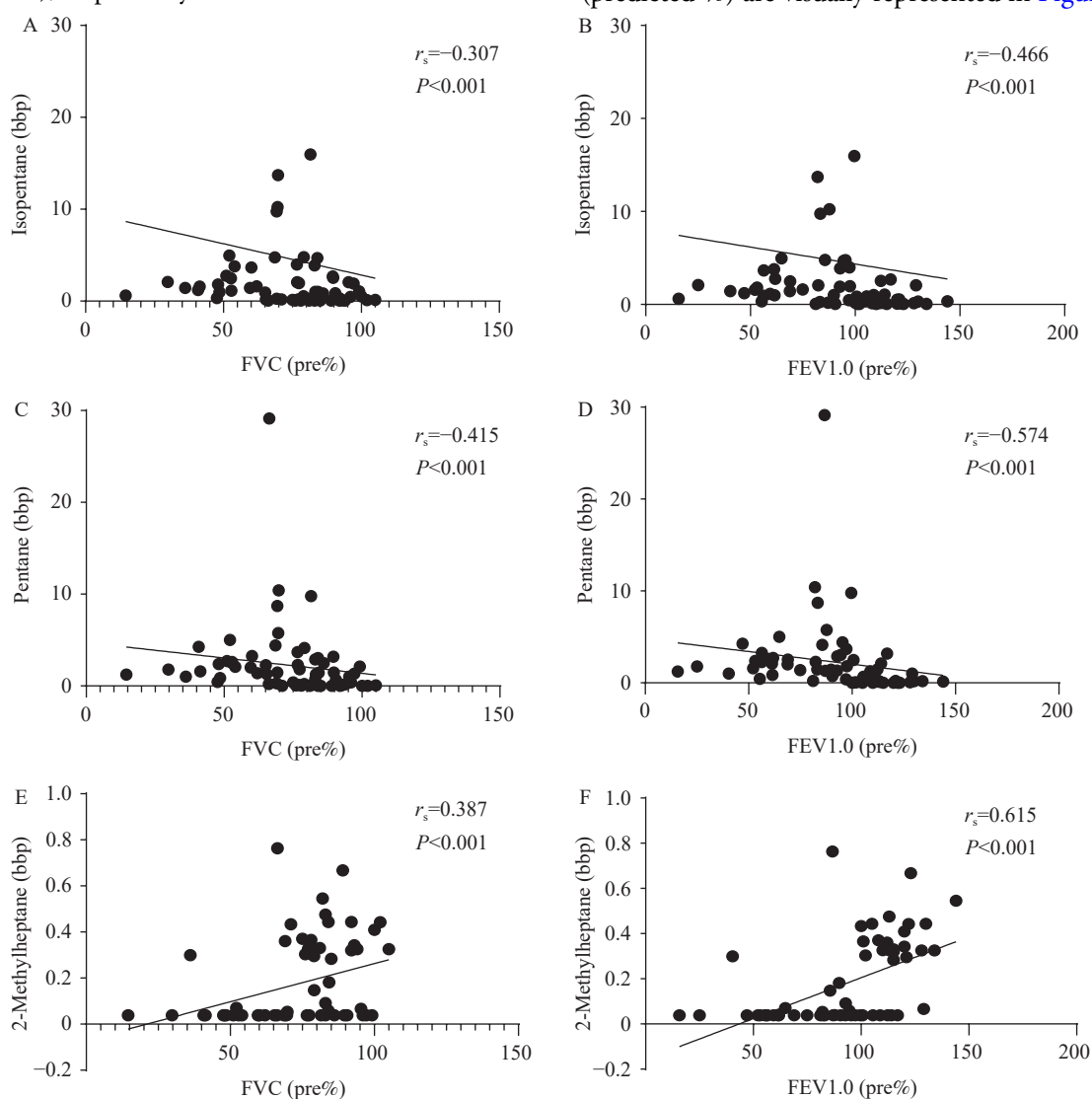


FIGURE 2. Correlation scatterplot. (A) Correlation analysis of isopentane and FVC (pre%); (B) Correlation analysis of isopentane and FEV1.0(pre%); (C) Correlation analysis of pentane and FVC (pre%); (D) Correlation analysis of pentane and FEV1.0 (pre%); (E) Correlation analysis of 2-methylpentane and FVC (pre%); (F) Correlation analysis of 2-methylpentane and FEV1.0 (pre%).

Abbreviation: FVC=forced vital capacity; FEV1.0=forced expiratory volume in one second.

DISCUSSION

In this study, we conducted targeted quantitative analysis using TD-GC-MS technology to analyze exhaled breath samples from 65 subjects. Our findings revealed distinct differences in VOC metabolism between CWP patients and healthy individuals, leading to the identification of six differential VOCs. ROC curve analysis demonstrated that these six VOCs effectively discriminated between the two groups.

In the analysis, elevated isoprene concentrations were observed in the exhaled breath of the coal workers' pneumoconiosis group. Foster (7) and colleagues demonstrated that increased exhaled isoprene may be associated with ROS-mediated oxidative stress, suggesting potentially elevated ROS levels in CWP patients. Similarly, pentane concentrations were significantly higher in the CWP group compared to controls. This finding aligns with Jalali (8) and colleagues' research, which reported elevated exhaled pentane concentrations in workers exposed to crystalline silica dust, indicating increased oxidative damage to ω -6 polyunsaturated fatty acids in both dust-exposed and affected workers. Furthermore, we observed upregulation of five lipid peroxidation markers — isopentane, n-pentane, 3-methylpentane, n-hexane, and cyclohexane — in the CWP group, suggesting elevated oxidative stress levels. Notably, pentane and hexane have been previously identified as differential diagnostic markers for pneumoconiosis (9). The six differential VOCs identified in our study were all alkanes, consistent with these earlier findings.

Furthermore, this study compared clinical indicators between populations and analyzed correlations between the six differential VOCs and clinical parameters. Notably, decreased forced vital capacity (FVC) has been associated with mortality in patients with idiopathic pulmonary fibrosis (IPF) (10) and other forms of pulmonary fibrosis, making FVC a critical parameter for assessing disease status in patients with pulmonary fibrosis (11). Correlation analyses revealed that isopentane and n-pentane were negatively correlated with FVC and FEV1.0 levels, while 2-methylheptane showed positive correlations with both FVC and FEV1.0 levels ($P < 0.05$), suggesting these three VOCs are associated with lung function decline. Importantly, the CWP group exhibited increased concentrations of isopentane and n-pentane and decreased concentrations of 2-methylheptane in

exhaled breath, aligning with the correlation analyses and suggesting these three VOCs may serve as potential biomarkers for CWP.

VOCs in the human body originate from both endogenous biochemical processes and environmental exposures, being released through various biological matrices including exhaled breath, urine, and skin (12). Exhalation represents the primary release pathway for VOCs in the human body, and its relationship to disease has garnered significant attention, particularly in lung pathologies (13). In our study, we identified six differential VOCs in exhaled breath. Analysis using the Human Metabolome Database (HMDB) revealed that cell membranes constitute a primary source for these compounds, with pentane, hexane, cyclohexane, and 2-methylheptane specifically identified as endogenous metabolites in exhaled breath. While the mechanisms underlying endogenous VOC production in human exhaled breath remain incompletely understood, elucidating these production pathways could provide crucial insights into the pathogenesis of coal workers' pneumoconiosis (14).

This study has several limitations: First, as an exploratory investigation with a relatively small sample size, future studies with larger cohorts are needed to validate our findings. Second, while previous research has identified methylated alkanes as characteristic VOCs of pneumoconiosis, age significantly influences methylated alkane levels (15). Although we included age as a covariate in our analyses, its precise impact on outcomes remains unclear. Future investigations should consider age-matched case-control populations to better control for age-related effects and ensure result reliability.

Conflict of interest: No conflicts of interest.

Funding: Supported by the National Center for Occupational Safety and Health Self-management Project of China (2019009) and the Central Public-Interest Scientific Institution Basal Research Project of China (2020-PT320-005).

doi: [10.46234/ccdcw2024.278](https://doi.org/10.46234/ccdcw2024.278)

Corresponding author: Chunguang Ding, ding_chg@163.com.

¹ National Center for Occupational Safety and Health, NHC, Beijing, China; ² NHC Key Laboratory for Engineering Control of Dust Hazard, Beijing, China; ³ Guangdong Pharmaceutical University, Guangzhou City, Guangdong Province, China; ⁴ Xishan Coal and Electricity (Group) Co., Ltd. Occupational Disease Prevention and Control Center, Taiyuan City, Shanxi Province, China; ⁵ NHC Key Laboratory of Pneumoconiosis, MOE Key Laboratory of Coal Environmental Pathogenicity and Prevention, Shanxi Key Laboratory of Respiratory Diseases, The First Hospital of Shanxi Medical University, Taiyuan City, Shanxi Province, China.

Submitted: September 06, 2024

Accepted: December 16, 2024

Issued: December 27, 2024

REFERENCES

- Boots AW, van Berkel JJBN, Dallinga JW, Smolinska A, Wouters EF, van Schooten FJ. The versatile use of exhaled volatile organic compounds in human health and disease. *J Breath Res* 2012;6(2):027108. <https://doi.org/10.1088/1752-7155/6/2/027108>.
- Di Natale C, Macagnano A, Martinelli E, Paolesse R, D'Arcangelo G, Roscioni C, et al. Lung cancer identification by the analysis of breath by means of an array of non-selective gas sensors. *Biosens Bioelectron* 2003;18(10):1209 - 18. [https://doi.org/10.1016/S0956-5663\(03\)00086-1](https://doi.org/10.1016/S0956-5663(03)00086-1).
- Li ZT, Zeng PY, Chen ZM, Guan WJ, Wang T, Lin Y, et al. Exhaled volatile organic compounds for identifying patients with chronic pulmonary aspergillosis. *Front Med* 2021;8:720119. <https://doi.org/10.3389/FMED.2021.720119>.
- Yamada YI, Yamada G, Otsuka M, Nishikiori H, Ikeda K, Umeda Y, et al. Volatile organic compounds in exhaled breath of idiopathic pulmonary fibrosis for discrimination from healthy subjects. *Lung* 2017;195(2):247 - 54. <https://doi.org/10.1007/s00408-017-9979-3>.
- Zou YC, Wang Y, Jiang ZL, Zhou Y, Chen Y, Hu YJ, et al. Breath profile as composite biomarkers for lung cancer diagnosis. *Lung Cancer* 2021;154:206 - 13. <https://doi.org/10.1016/j.lungcan.2021.01.020>.
- Gao R, Peng FD, He YZ, Feng J, Liu WB, Yu LJ, et al. Determination of 27 volatile organic compounds in exhaled breath by thermal desorption-gas chromatography-mass spectrometry. *J Environ Occup Med* 2024;41(1):96 - 102. <https://doi.org/10.11836/JEOM23283>.
- Foster WM, Jiang L, Stekiewicz PT, Risby TH. Breath isoprene: temporal changes in respiratory output after exposure to ozone. *J Appl Physiol* (1985) 1996;80(2):706-10. <http://dx.doi.org/10.1152/jappl.1996.80.2.706>.
- Jalali M, Zare Sakhvidi MJ, Bahrami A, Berijani N, Mahjub H. Oxidative stress biomarkers in exhaled breath of workers exposed to crystalline silica dust by SPME-GC-MS. *J Res Health Sci* 2016;16(3):153 - 61.
- Yang HY, Shie RH, Chang CJ, Chen PC. Development of breath test for pneumoconiosis: a case-control study. *Respir Res* 2017;18(1):178. <https://doi.org/10.1186/s12931-017-0661-3>.
- Collard HR, King Jr TE, Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2003;168(5):538 - 42. <https://doi.org/10.1164/rccm.200211-1311OC>.
- Takei R, Brown KK, Yamano Y, Kataoka K, Yokoyama T, Matsuda T, et al. Prevalence and prognosis of chronic fibrosing interstitial lung diseases with a progressive phenotype. *Respirology* 2022;27(5):333 - 40. <https://doi.org/10.1111/resp.14245>.
- Drabińska N, Flynn C, Ratcliffe N, Belluomo I, Myridakis A, Gould O, et al. A literature survey of all volatiles from healthy human breath and bodily fluids: the human volatilome. *J Breath Res* 2021;15(3):034001. <https://doi.org/10.1088/1752-7163/abf1d0>.
- van der Schee MP, Paff T, Brinkman P, van Aalderen WMC, Haarman EG, Sterk PJ. Breathomics in lung disease. *Chest* 2015;147(1):224 - 31. <https://doi.org/10.1378/chest.14-0781>.
- Magnano MC, Ahmed W, Wang R, Bergant Marušič M, Fowler SJ, White IR. Exhaled volatile organic compounds and respiratory disease: recent progress and future outlook. *TrAC Trends Anal Chem* 2024;176:117739. <https://doi.org/10.1016/j.trac.2024.117739>.
- Phillips M, Cataneo RN, Greenberg J, Gunawardena R, Rahbari-Oskoui F. Increased oxidative stress in younger as well as in older humans. *Clin Chim Acta* 2003;328(1-2):83 - 6. [https://doi.org/10.1016/S0009-8981\(02\)00380-7](https://doi.org/10.1016/S0009-8981(02)00380-7).

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLE S1. Linear equations, correlation coefficients, measurement ranges, detection limits and quantification limits for VOCs.

VOC	LOD/(ppb)	LOQ/(ppb)
Iso-pentane	0.098	0.326
1-Pentene	0.094	0.312
Pentane	0.028	0.092
Trans-2-Pentene	0.089	0.297
Iso-prene	0.015	0.049
2,2-Dimethylbutane	0.133	0.444
2-Methylpentane	0.089	0.297
3-Methylpentane	0.017	0.058
1-Hexene	0.110	0.367
Hexane	0.019	0.065
2,4-Dimethylpentane	0.111	0.370
Methylcyclopentane	0.007	0.025
2-Methylhexane	0.071	0.238
Cyclohexane	0.006	0.022
2,3-Dimethylpentane	0.017	0.057
3-Methylhexane	0.064	0.214
2,2,4-Trimethylpentane	0.024	0.080
Heptane	0.073	0.244
Methylcyclohexane	0.007	0.023
Pentanal	0.067	0.225
2,3,4-Trimethylpentane	0.018	0.061
2-Methylheptane	0.055	0.182
3-Methylheptane	0.038	0.126
4-Methyl-2-pentanone	0.005	0.015
2-Hexanone	0.018	0.060
Hexanal	0.070	0.233
Decane	0.070	0.234

Abbreviation: LOD=limit of detection; LOQ=limit of quantification; VOC=volatile organic compounds.

SUPPLEMENTARY TABLE S2. Screening for differential VOCs.

VOC	VIP	FC(N/P)	P	Adjusted p-value
Iso-pentane	1.931	0.028	<0.05	1.000
1-Pentene	0.076	0.235	0.454	-
Pentane	2.314	0.035	<0.05	0.530
Trans-2-Pentene	0.529	0.471	0.293	-
Iso-prene	1.237	0.637	<0.05	0.938
2,2-Dimethylbutane	0.129	0.957	0.358	-
2-Methylpentane	0.848	0.019	<0.05	1.000
3-Methylpentane	1.190	0.016	<0.05	1.000
1-Hexene	0.549	0.330	<0.05	0.977
Hexane	1.090	0.017	<0.05	1.000
2,4-Dimethylpentane	1.226	1.511	<0.05	1.000
Methylcyclopentane	0.989	0.035	<0.05	1.000
2-Methylhexane	0.278	0.391	0.292	-
Cyclohexane	1.235	0.015	<0.05	1.000
2,3-Dimethylpentane	0.510	0.237	0.562	-
3-Methylhexane	0.290	0.210	0.158	-
2,2,4-Trimethylpentane	0.696	0.117	0.362	-
Heptane	0.494	0.750	0.079	-
Methylcyclohexane	0.725	0.484	<0.05	1.000
Pentanal	0.109	0.755	0.661	-
2,3,4-Trimethylpentane	0.659	1.047	<0.05	1.000
2-Methylheptane	2.238	5.161	<0.05	0.014*
3-Methylheptane	0.382	0.432	0.355	-
4-Methyl-2-pentanone	0.518	0.055	0.182	-
2-Hexanone	0.652	0.247	<0.05	1.000
Hexanal	0.079	0.700	0.671	-
Decane	0.688	0.388	0.285	-

Abbreviation: VOC=volatile organic compounds; VIP=variable importance in projection; FC=fold change; N=control group; P=coal workers' pneumoconiosis group.

* $P<0.05$.

SUPPLEMENTARY TABLE S3. ROC curve results for six differential VOCs.

VOC	AUC	95% CI	Sensitivity	Specificity	Cut-off value	P
Isopentane	0.940	0.880–1.000	0.881	1.000	0.5733	<0.001
Pentane	0.996	0.986–1.000	0.976	1.000	0.4227	<0.001
3-Methylpentane	0.776	0.663–0.889	0.738	0.783	0.1887	<0.001
Hexane	0.753	0.634–0.872	0.762	0.696	0.6565	<0.001
Cyclohexane	0.777	0.667–0.888	0.643	0.913	0.1101	<0.001
2-Methylheptane	0.948	0.878–1.000	0.952	0.956	0.2330	<0.001
Combined	1.000	1.000–1.000	1.000	1.000	-	<0.001

Abbreviation: VOC=volatile organic compounds; AUC= area under curve; CI=confidence interval.

* $P<0.05$.

SUPPLEMENTARY TABLE S4. Correlation between differential VOCs and clinical parameters.

VOC		WBC	NEU	LYM	NLR	PLR	FVC (predicted %)	FEV1.0 (predicted %)
Isopentane	r_s	-0.017	0.095	-0.08	0.171	-0.033	-0.307	-0.466
	P	0.895	0.454	0.528	0.173	0.793	0.014*	<0.001*
Pentane	r_s	0.039	0.158	-0.003	0.177	-0.111	-0.415	-0.574
	P	0.759	0.208	0.981	0.158	0.377	0.001*	<0.001*
3-Methylpentane	r_s	0.025	0.101	-0.040	0.131	-0.029	-0.143	-0.193
	P	0.844	0.422	0.753	0.297	0.820	0.261	0.127
Hexane	r_s	-0.015	0.08	-0.063	0.129	0.032	-0.064	-0.140
	P	0.904	0.525	0.618	0.306	0.799	0.615	0.270
Cyclohexane	r_s	-0.059	0.008	-0.042	0.061	-0.044	-0.098	-0.160
	P	0.642	0.948	0.739	0.628	0.728	0.439	0.205
2-Methylheptane	r_s	0.053	-0.082	0.126	-0.223	0.115	0.387	0.615
	P	0.677	0.515	0.316	0.074	0.361	0.002*	<0.001*

Abbreviation: WBC=white blood cell; NEU=neutrophil; LYM=lymphocyte; NLR=neutrophil to lymphocyte ratio; PLR=platelet to lymphocyte ratio; FVC=forced vital capacity; FEV1.0=forced expiratory volume in one second; VOC=volatile organic compounds.

* indicates $P<0.05$.